Optimal boundary control of quasilinear elliptic partial differential equations: theory and numerical analysis

vorgelegt von

Dipl.-Math. Vili Dhamo

von der Fakultät II - Mathematik und Naturwissenschaften
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender: Prof. Dr. Yuri B. Suris
Berichter: Prof. Dr. Fredi Tröltzsch
Berichter: Prof. Dr. Eduardo Casas

Tag der wissenschaftlichen Aussprache: 21.03.2012

Berlin 2012
D 83

Preface

This thesis is concerned with a class of optimal control problems governed by quasilinear elliptic partial differential equations with inhomogeneous Neumann boundary conditions. The boundary datum will be considered as the control variable which must satisfy given inequality constraints.

Control of equations of this type is interesting, because in many practical applications of optimal control theory to problems in engineering and medical science the underlying PDE's are quasilinear; for instance, in models of heat conduction, where the heat conductivity coefficient depends on the spatial coordinate and on the temperature of the system. The heat conductivity of carbon steel depends on the temperature and also on the alloying additions contained; see Bejan [**7**]. If the different alloys of steel are distributed smoothly in the domain, then the conductivity coefficient should depend in a sufficiently smooth way on both, space variable and temperature. Similarly, this dependence is observed in the growth of silicon carbide bulk single crystals; see Klein et al. [**73**].

The quasilinear equation under consideration is not monotone, because the leading coefficient of the differential operator is dependent on the solution of the equation. Control problems with quasilinear elliptic equations of non-monotone type were recently considered by Casas and Tröltzsch [**36, 37, 38**]. The authors have treated the case of distributed controls and their contributions include not only the derivation of necessary and sufficient optimality conditions but also the analysis of the numerical approximation of those control problems. It is known that in the case of boundary controls the analysis is more difficult, since the regularity of the states is lower than that of distributed controls. The goal of the present work is to extend the results obtained by Casas and Tröltzsch to the case of Neumann boundary controls problems in polygonal domains of dimension two. Most of the material of this thesis can be found in [**20, 18, 19**].

In order to tackle different aspects of the theoretical and numerical analysis of the control problem, it is necessary to perform a comprehensive study of the well-posedness of the state equation and further to analyze the differentiability of the control-to-state mapping. The first chapter deals with these topics as well as contains some useful regularity results concerning the adjoint state equation.

Chapter 2 is dedicated to the finite-element based approximation of the state and adjoint state equation. Our main focus is the error analysis for these approximations. A serious difficulty in this analysis is that the uniqueness of a solution of the discrete quasilinear equation is an open problem. To overcome this difficulty a local uniqueness result can be provided which is sufficient for further investigations.

In Chapter 3, the optimal control problem associated with the quasilinear equation is formulated and the question of existence of solutions is answered positively. Furthermore, first- and second-order optimality conditions are established and some higher regularity results for optimal controls are derived.

Chapter 4 contains the numerical analysis of the control problem. Approximating the state and adjoint state by finite elements of degree one and the control by step functions, the strong convergence of discrete local optimal controls to a strict local optimal control of the continuous problem can be shown. Finally, the error analysis for the optimal controls is carried out and confirmed by numerical experiments.

In the last chapter, some extensions of the results concerning the numerical approximation of the quasilinear equation to three-dimensional polyhedral domains are presented.

Acknowledgments. At first, I would like to express my gratitude to my supervisor Fredi Tröltzsch for always being open for any question, for introducing optimal control theory to me and also for the help in gaining experience in research and teaching.

My special thanks are addressed to Eduardo Casas for suggesting this interesting subject. Without his enduring support and the countless fruitful discussions I would not have been able to realize this thesis. Besides, I want to thank him for the productive and successful collaboration in the last four years.

I wish to thank the whole Optimal Control Group at the Technical University of Berlin for the exchange of ideas and for all the enjoyable time we spent together.

Last but by no means least, I like to thank my parents, my brother and my girlfriend for their never-ending encouragement and ongoing support in all its forms.

General notation

Given a Banach space V, we shall denote by $\|\cdot\|_V$ the standard norm of V and by $\langle\cdot,\cdot\rangle_{V^*,V}$ the duality product between the dual space V^* and V. When no ambiguity arises, we will abbreviate $\langle\cdot,\cdot\rangle_{V^*,V}$ by $\langle\cdot,\cdot\rangle$.

In the sequel, we recall some Banach spaces which are frequently used in this thesis. Let $\Omega \subset \mathbb{R}^n$, $n \geq 1$, be bounded and measurable. We say that a property holds for almost all (for a.a.) $x \in \Omega$ (or a.e. in Ω) if it is valid in Ω except for a measurable set of Lebesgue measure zero. The space of all continuous functions in the closure $\bar{\Omega}$ of Ω is denoted by $C(\bar{\Omega})$. We shall write $C^{0,\alpha}(\bar{\Omega})$ for the space of Hölder functions with Hölder exponent $0 < \alpha \leq 1$ and we shall call $f \in C^{0,1}(\bar{\Omega})$ Lipschitz continuous. Moreover, for $1 \leq q \leq \infty$ we define

$$L^q(\Omega) = \Big\{ f : \Omega \longrightarrow \mathbb{R} \,|\, f \text{ is measurable and } \|f\|_{L^q(\Omega)} \text{ is finite} \Big\}\,,$$

where

$$\|f\|_{L^q(\Omega)} := \begin{cases} \left(\displaystyle\int_\Omega |f(x)|^q \, dx \right)^{1/q} & \text{if } 1 \leq q < \infty\,, \\ \operatorname{ess\,sup}_{x\in\Omega} |f(x)| & \text{if } q = \infty\,. \end{cases}$$

In the last formula dx is the Lebesgue measure in Ω. It is well-known that the dual space to $L^q(\Omega)$, $1 < q < \infty$, can be identified by $L^{q'}(\Omega)$, where q' is the conjugate exponent of q satisfying $(1/q) + (1/q') = 1$. For $1 \leq q \leq \infty$ we set $W^{0,q}(\Omega) := L^q(\Omega)$ and, for $m \in \mathbb{N}$, the Sobolev space $W^{m,q}(\Omega)$ stands for the space of all measurable functions $f : \Omega \longrightarrow \mathbb{R}$, whose weak derivatives $D^\alpha f$ of order α, with $|\alpha| \leq m$, belong to $L^q(\Omega)$, cf. Adams [**1**]. We equip the space $W^{m,q}(\Omega)$ with the norm

$$\|f\|_{W^{m,q}(\Omega)} := \left(\int_\Omega \sum_{|\alpha|\leq m} |D^\alpha f(x)|^q \, dx \right)^{1/q}$$

if $1 \leq q < \infty$ and $\|f\|_{W^{m,\infty}(\Omega)} := \sum_{|\alpha|\leq m} \|D^\alpha f\|_{L^\infty(\Omega)}$ if $q = \infty$. In particular, the space $H^m(\Omega) := W^{m,2}(\Omega)$ is a Hilbert space with the scalar product

$$(f,g)_{H^m(\Omega)} = \sum_{|\alpha|\leq m} (D^\alpha f, D^\alpha g)_{L^2(\Omega)} = \sum_{|\alpha|\leq m} \left(\int_\Omega |D^\alpha f D^\alpha g|^2 \, dx \right)^{1/2}.$$

For a real exponent $s > 0$ with $s \notin \mathbb{N}$ we write $s = [s] + \sigma$, where $[s]$ is an integer and $0 < \sigma < 1$. Then the Sobolev-Slobodetskij space $W^{s,q}(\Omega)$, $1 \leq q < \infty$, consists of all functions $f \in W^{[s],q}(\Omega)$ such that

$$\sum_{|\alpha|=[s]} \int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^q}{|x-y|^{n+\sigma q}} dx\, dy < \infty\,.$$

We equip the space $W^{s,q}(\Omega)$ with the norm

$$\|f\|_{W^{s,q}(\Omega)} := \left(\|f\|^q_{W^{[s],q}(\Omega)} + \sum_{|\alpha|=[s]} \int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^q}{|x-y|^{n+\sigma q}} dx\, dy \right)^{1/q}$$

More details on Sobolev and Sobolev-Slobodetskij spaces can be found, e.g., in Adams [**1**], Nečas [**82**], Wloka [**92**].

Let now Ω have a Lipschitz boundary $\Gamma = \partial\Omega$, cf. Nečas [**82**] or Gajewski et al. [**54**] for a precise definition; σ denotes the usual $(n-1)$-dimensional measure over Γ induced by its parametrization. The spaces $C(\Gamma)$, $C^{0,\alpha}(\Gamma)$ and $L^q(\Gamma)$ are defined analogously. However, the introduction of Sobolev spaces on Γ is more delicate. Let us denote

$$W^{1/q',q}(\Gamma) := \left\{ z|_\Gamma \mid z \in W^{1,q}(\Omega) \right\}$$

endowed with the norm $\|g\|_{W^{1/q',q}(\Gamma)} := \inf\{\|z\|_{W^{1,q}(\Omega)} \mid z|_\Gamma = g\}$, where $z|_\Gamma$ is the trace of z on Γ, cf. Lions [**77**]. To avoid heavy formulas we will mostly write z instead of $z|_\Gamma$. The dual space of $W^{1/q',q}(\Gamma)$ is denoted by $W^{-1/q',q'}(\Gamma)$. Since $1/q' = 1 - 1/q \in (0,1)$, we may equip $W^{1/q',q}(\Gamma)$ with the intrinsic norm

$$\|g\|_{W^{1/q',q}(\Gamma)} = \left(\|g\|^q_{L^q(\Gamma)} + \int_\Gamma \int_\Gamma \frac{|g(x) - g(y)|^q}{|x-y|^{n-2+q}} d\sigma(x)\, d\sigma(y) \right)^{1/q}$$

which is equivalent to the previous one; see Grisvard [**60**, page 20].

For the corresponding trace theorems and embedding theorems for bounded Lipschitz domains the reader is referred to, e.g., Adams [**1**] and Nečas [**82**]; see also Ding [**48**].

We write $S'(v)h$ for the Fréchet derivative of an operator $S : V \longrightarrow Y$, Y being another Banach space, at $v \in V$ in the direction $h \in V$; for the second Fréchet derivative in the directions $h_1, h_2 \in V$, we write $S'(v)[h_1, h_2]$ or simply $S'(v)h^2$ if $h_1 = h_2 = h$.

Given two sets A and B with $A \subset B$, $\chi_A^B : B \longrightarrow \{0,1\}$ denotes the indicator function which is one in A and zero in $B \setminus A$. If there is no risk of notational confusion we will write χ_A.

Throughout the thesis, $B_V(x,\rho)$ is the open ball in V with radius ρ centered at x, and $\overline{B}_V(x,\rho)$ stands for its closure. By C (without index) generic constants are denoted and in some formulas, the partial derivative $\partial/\partial x_j$ is abbreviated by ∂_j.

Contents

CHAPTER 1

Analysis of quasilinear elliptic PDEs

1.1. Introduction

This chapter is concerned with a particular class of quasilinear elliptic equations of the form

$$(1.1.1) \qquad \begin{cases} -\operatorname{div}\,[a(x,y(x))\nabla y(x)] + f(x,y(x)) = 0 & \text{in } \Omega\,, \\ a(x,y(x))\partial_\nu y(x) = u(x) & \text{on } \Gamma\,. \end{cases}$$

Our main goal is to develop a comprehensive analysis of the above problem which will be widely used in the subsequent chapters. To make the setting more flexible for applications we will impose regularity assumptions on the nonlinear terms a and f which are as weak as necessary to study the above equation in the framework of PDE constrained optimal control problems; see Chapter 3 and 4.

Another aim of this chapter is to provide useful results for the numerical analysis of the finite element based approximation of (1.1.1); see Chapter 2. One problem which arises in this study is the approximation of the domain Ω. Typically, Ω is approximated by a new domain Ω_h with piecewise polygonal (in 2D) or polyhedral (in 3D) boundary. In many cases, the simplest and the most convenient choice consists of replacing Ω by a polygonal or polyhedral domain Ω_h. This requires the comparison of the boundary datum u of (1.1.1) and that of the discrete equation which is defined on $\partial\Omega_h$. Let us mention here that the effect of the domain change on the solution of problems associated with semilinear elliptic equations is investigated by Casas and Sokolowski [**33**]. To simplify our analysis we will restrict our consideration to polygonal domains of dimension two. The case when the domain Ω is a polyhedral set of dimension three is discussed in Chapter 5. However, assuming that Ω is a plane polygonal domain not necessary convex, introduces a new difficulty: the regularity of elliptic equations in corner domains needs special care.

In the analysis of the quasilinear equation (1.1.1), we will be faced with several difficulties. In spite that f is considered monotone non-decreasing with respect to y, the above equation is not monotone, because the coefficient a of the divergence term depends on y. The first difficulty caused by the non-monotone character of the equation (1.1.1) is found when deducing regularity properties for its solution. Other difficulties appear when analyzing the linearized equation and the associated adjoint equation which are both non-monotone; see Section 1.6 and 1.7.

In this chapter, we will also focus our attention to the differentiability properties of the solution operator $u \longmapsto y$. These properties are of utmost importance since they allow us to establish necessary and sufficient optimality conditions for control problems governed by PDEs of the type (1.1.1). The discussion of these differentiability properties will require a careful analysis of the corresponding linearized equation. Another issue we address in this chapter is the study of the adjoint problem. There are at least two reasons why it is of high interest to investigate the adjoint equation. First, it has become an essential element in optimal control theory; see Chapter 3 and 4. Secondly, it plays a key role in the proof of error estimates for the finite element approximation of (1.1.1); see Chapter 2.

Up to now, there exist only a few contributions where quasilinear equations have been studied in the context of optimal control problems. We mention Lions [**77**], Casas and Fernández [**21, 22**], Casas and Yong [**41**] and Casas et al. [**23**], for problems with nonlinearity of gradient type. Recently, Casas and Tröltzsch considered in [**36, 37, 38**] the equation (1.1.1) with homogeneous Dirichlet boundary conditions. Here we extend the theory developed in [**36**] to the more delicate case of inhomogeneous Neumann boundary conditions.

This chapter provides the analysis of all partial differential equations which occur when considering optimal control problems governed by (1.1.1). It is organized as follows: In the upcoming section, we state the main assumptions on the data involved in (1.1.1). In Sections 1.3-1.5, we discuss the well-posedness of equation (1.1.1) in different spaces. Section 1.4 contains some preliminary results which are used to obtain higher regularity of the solution of the quasilinear equation. Section 1.6 is devoted to the study of the linearized equation and to the derivation of some differentiability properties of the solution operator $u \longmapsto y$. The analysis of the adjoint problem is the scope of the last section.

1.2. Main assumptions on the nonlinearities

We assume the following hypotheses about the data involved in (1.1.1).

Assumption 1.1. *$\Omega \subset \mathbb{R}^2$ is an open bounded polygonal domain with boundary Γ. We denote the unit outward normal vector to Γ at x by $\nu(x)$.*

Assumption 1.2. *The function $a : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function, i.e. measurable with respect to the first variable and continuous with respect to the second one. Moreover,*

$$\exists \alpha_a > 0 \ \text{ such that } \ a(x,y) \geq \alpha_a \text{ for a.a. } x \in \Omega \text{ and all } y \in \mathbb{R}\,, \tag{1.2.1}$$

$a(\cdot,0) \in L^\infty(\Omega)$ and for any $M > 0$ there exists a constant $C_{a,M} > 0$ such that

$$|a(x,y_2) - a(x,y_1)| \leq C_{a,M}|y_2 - y_1| \text{ for a.a. } x \in \Omega \text{ and all } |y_i| \leq M,\ i = 1,2\,.$$

Assumption 1.3. *The function $f : \Omega\times\mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function and there exists $p > 4/3$ such that $f(\cdot, 0) \in L^p(\Omega)$. Moreover, f is monotone non-decreasing with respect to the second variable for a.a. $x \in \Omega$ and there exist a positive constant $\alpha_f > 0$ and a measurable set $E \subset \Omega$ with Lebesgue measure $|E| > 0$ such that*

$$\frac{f(x, y_2) - f(x, y_1)}{y_2 - y_1} \geq \alpha_f \quad \forall x \in E \ \text{ and } \ \forall y_1, y_2 \in \mathbb{R}\,, \ \text{with } y_1 \neq y_2\,. \tag{1.2.2}$$

Finally, we assume that for any $M > 0$ there exists a function $\phi_M \in L^p(\Omega)$ such that

$$|f(x, y_2) - f(x, y_1)| \leq \phi_M(x)|y_2 - y_1| \ \text{ for a.a. } x \in \Omega \text{ and all } |y_1|, |y_2| \leq M\,. \tag{1.2.3}$$

Example 1.4. *Taking $a(x, y) = \phi_0(x)+y^{2m}$ and $f(x, y) = y+y^3-\zeta(x)$, with $m \in \mathbb{N}$, $\phi_0 \in L^\infty(\Omega)$, $\phi_0(\cdot) \geq \alpha_a > 0$ a.e. in Ω, and $\zeta \in L^p(\Omega)$, $p > 4/3$, we see that the equation*

$$\left\{\begin{aligned} -\operatorname{div}[(\phi_0(x) + y^{2m}(x))\nabla y(x)] + y + y^3(x) =&\ \zeta(x) \quad \text{in } \Omega\,,\\ [\phi_0(x) + y^{2m}(x)]\,\partial_\nu y(x) =&\ u(x) \quad \text{on } \Gamma\,,\end{aligned}\right.$$

satisfies the above assumptions. Other possible choices are $a(x, y) = \phi_0(x) + e^y$ or $f(x, y) = e^y - \zeta(x)$.

Throughout the thesis, p and E, introduced in Assumption 1.3, will be fixed and the solutions of PDEs are understood in the weak sense: We say that $y \in H^1(\Omega)\cap L^\infty(\Omega)$ is a *solution* of (1.1.1) if the following identity holds

$$\int_\Omega \{a(x, y(x))\nabla y(x)\cdot\nabla\phi(x) + f(x, y(x))\phi(x)\}\,dx = \int_\Gamma u(x)\phi(x)\,d\sigma(x) \tag{1.2.4}$$

for any *test function* $\phi \in H^1(\Omega)$.

1.3. Well-posedness of the quasilinear equation

To prove the existence of a weak solution y of (1.1.1) in $H^1(\Omega)$ we cannot apply the Minty-Browder theorem for monotone operators; see, for instance, Gajewski et al. [**54**, Theorem III.2.1]. This is due to the fact that the equation (1.1.1) does not lead to a monotone operator in general. For an example showing the non-monotone character of (1.1.1) in a particular case the reader is referred to Hlaváček et al. [**67**, Remark 2.2].

In the following theorem, we establish the existence and uniqueness of a solution to (1.1.1).

Theorem 1.5. *Under the Assumptions 1.1-1.3, for any $u \in L^s(\Gamma)$ with $s > 1$ equation (1.1.1) has a unique solution $y_u \in H^1(\Omega) \cap L^\infty(\Omega)$. Moreover, there exists $\mu \in (0, 1)$ independent of u such that $y_u \in C^{0,\mu}(\bar\Omega)$ and, for any set U bounded in $L^s(\Gamma)$,*

$$\|y_u\|_{H^1(\Omega)} + \|y_u\|_{C^{0,\mu}(\bar\Omega)} \leq C_U \quad \forall u \in U\,, \tag{1.3.1}$$

with some constant $C_U > 0$.

PROOF. *Existence of a solution.* To show the existence of a solution of (1.1.1) we introduce the truncated functions a_M and f_M as follows. Depending on $M > 0$, the function a_M is given by

$$a_M(x,y) = \begin{cases} a(x,y) & \text{if } |y| \le M\,, \\ a(x,+M) & \text{if } y > +M\,, \\ a(x,-M) & \text{if } y < -M\,. \end{cases}$$

In the same way, we define the truncation f_M of f. Now let $u \in L^s(\Gamma)$ and $\varepsilon > 0$ be arbitrary but fixed. Consider the mapping $F : L^2(\Omega) \longrightarrow L^2(\Omega)$ defined by $F(w) = z_\varepsilon$, where z_ε satisfies the linear equation

$$\begin{cases} -\text{div}\,[a_M(x,w)\nabla z_\varepsilon] + \varepsilon\chi_E z_\varepsilon = -f_M(x,w) & \text{in } \Omega\,, \\ a_M(x,w)\partial_\nu z_\varepsilon = u & \text{on } \Gamma\,. \end{cases} \tag{1.3.2}$$

Thanks to Assumption 1.3, (1.3.2) is monotone, hence by applying the Lax-Milgram theorem, we get the existence of a unique solution $z_\varepsilon \in H^1(\Omega)$ of (1.3.2), therefore F is well-defined. Next we will use the Poincaré inequality (see, for instance, Gajewski et al. [**54**, Lemma II.1.36])

$$\|z\|^2_{H^1(\Omega)} \le C_E \left(\|\nabla z\|^2_{L^2(\Omega)} + \|z\|^2_{L^2(E)}\right) \quad \forall z \in H^1(\Omega)\,, \tag{1.3.3}$$

with $C_E > 0$ being independent of z. Taking z_ε as test function in the weak formulation of (1.3.2), along with the positivity of a_M, there holds

$$\begin{aligned} \min\{\alpha_a, \varepsilon\}\|z_\varepsilon\|^2_{H^1(\Omega)} &\le C_E \int_\Omega \left\{a_M(x,w)|\nabla z_\varepsilon|^2 + \varepsilon\chi_E z_\varepsilon^2\right\} dx \\ &= C_E \left(\int_\Gamma u z_\varepsilon \, d\sigma(s) - \int_\Omega f_M(x,w) z_\varepsilon \, dx\right) \\ &\le C \left(\|u\|_{L^s(\Gamma)} + \|f_M(\cdot,w)\|_{L^p(\Omega)}\right) \|z_\varepsilon\|_{H^1(\Omega)}\,. \end{aligned}$$

In the last inequality, we have used the continuity of the trace operator

$$\gamma : H^1(\Omega) \longrightarrow H^{1/2}(\Gamma)\,,\ \gamma(z)(x) = z|_\Gamma(x)\,,\ \text{for } z \in H^1(\Omega) \cap C(\bar\Omega) \text{ and a.a. } x \in \Gamma\,,$$

as well as the continuous embeddings $H^1(\Omega) \hookrightarrow L^{p'}(\Omega)$ and $H^{1/2}(\Gamma) \hookrightarrow L^{s'}(\Gamma)$, where p' and s' are the conjugate exponents to p and s, respectively. Thus,

$$\|z_\varepsilon\|_{H^1(\Omega)} \le C_{a,\varepsilon} \left(\|u\|_{L^s(\Gamma)} + \|f_M(\cdot,w)\|_{L^p(\Omega)}\right),$$

where $C_{a,\varepsilon}$ depends only on $|\Omega|$, α_a and ε, but neither on a_M nor on f_M. Because of the compact embedding of $H^1(\Omega)$ in $L^2(\Omega)$ we can apply Schauder's theorem to obtain the existence of a fixed point $y_\varepsilon \in H^1(\Omega)$ of F which is a solution of

$$\begin{cases} -\,\text{div}\,[a_M(x,y_\varepsilon)\nabla y_\varepsilon] + \varepsilon\chi_E y_\varepsilon = -f_M(x,y_\varepsilon) & \text{in } \Omega\,, \\ a_M(x,y_\varepsilon)\partial_\nu y_\varepsilon = u & \text{on } \Gamma\,. \end{cases} \tag{1.3.4}$$

Next we show the boundedness of $\{y_\varepsilon\}_{\varepsilon>0}$ in $H^1(\Omega)$. Thanks to (1.2.2) and the definition of f_M, we have $(f_M(\cdot,y_\varepsilon) - f_M(\cdot,0))\,y_\varepsilon = \frac{(f_M(\cdot,y_\varepsilon)-f_M(\cdot,0))My_\varepsilon}{M} \geq \alpha_f M|y_\varepsilon|$ a.e. in Ω if $|y_\varepsilon| > M$, thus we arrive at

$$(f_M(x,y_\varepsilon) - f(x,0))\,y_\varepsilon \geq \begin{cases} \alpha_f y_\varepsilon^2 & \text{if } |y_\varepsilon| \leq M\,, \\ \alpha_f M|y_\varepsilon| & \text{if } |y_\varepsilon| > M\,, \end{cases}$$

for a.a. $x \in \Omega$. This leads to

$$\begin{aligned} \alpha_f M \int_E |y_\varepsilon|\,dx &= \alpha_f M \int_{E\cap\{|y_\varepsilon|>M\}} |y_\varepsilon|\,dx + \alpha_f M \int_{E\cap\{|y_\varepsilon|\leq M\}} |y_\varepsilon|\,dx \\ &\leq \int_{E\cap\{|y_\varepsilon|>M\}} (f_M(x,y_\varepsilon) - f_M(x,0))\,y_\varepsilon\,dx + \alpha_f M^2|E| \\ &\leq \int_\Omega (f_M(x,y_\varepsilon) - f_M(x,0))\,y_\varepsilon\,dx + \alpha_f M^2|E|\,. \end{aligned}$$

Therefore, we have from this inequality and the weak formulation of (1.3.4)

$$\begin{aligned} &\alpha_a\|\nabla y_\varepsilon\|^2_{L^2(\Omega)} + \alpha_f M \int_E |y_\varepsilon|\,dx \\ &\quad\leq \int_\Omega a_M(x,y_\varepsilon)|\nabla y_\varepsilon|^2\,dx + \int_\Omega (f_M(x,y_\varepsilon) - f_M(x,0))\,y_\varepsilon\,dx + \alpha_f M^2|E| \\ &\quad\leq C\left(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\right)\|y_\varepsilon\|_{H^1(\Omega)} - \int_E \varepsilon y_\varepsilon^2\,dx + \alpha_f M^2|E| \\ &\quad\leq C\left(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\right)\|y_\varepsilon\|_{H^1(\Omega)} + \alpha_f M^2|E| \\ &\quad\leq C'\Big(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\Big)\Big(\|\nabla y_\varepsilon\|_{L^2(\Omega)} + \int_E |y_\varepsilon|\,dx\Big) + \alpha_f M^2|E|\,, \end{aligned} \tag{1.3.5}$$

where we have used that

$$\|z\| := \|\nabla z\|_{L^2(\Omega)} + \int_E |z|\,dx$$

is a norm equivalent to the $\|\cdot\|_{H^1(\Omega)}$ norm. Taking

$$M > 2C'\left(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\right)/\alpha_f\,,$$

we deduce from (1.3.5) that

$$\begin{aligned} &\alpha_a\|\nabla y_\varepsilon\|^2_{L^2(\Omega)} + \frac{1}{2}\alpha_f M \int_E |y_\varepsilon|\,dx \\ &\qquad\leq C'\left(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\right)\|\nabla y_\varepsilon\|_{L^2(\Omega)} + \alpha_f M^2|E|\,. \end{aligned} \tag{1.3.6}$$

Now for the term $\Big(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\Big)\|\nabla y_\varepsilon\|_{L^2(\Omega)}$ we use Young's inequality with $q = q' = 2$ and get

$$\begin{aligned}\Big(\|f_M(\cdot,0)\|_{L^p(\Omega)} &+ \|u\|_{L^s(\Gamma)}\Big)\|\nabla y_\varepsilon\|_{L^2(\Omega)} \\ &\le \frac{1}{2\alpha_a}\Big(\|f_M(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)}\Big)^2 + \frac{\alpha_a}{2}\|\nabla y_\varepsilon\|^2_{L^2(\Omega)}\,.\end{aligned}$$

Inserting the last inequality in (1.3.6), it follows

$$\|\nabla y_\varepsilon\|^2_{L^2(\Omega)} + \int_E |y_\varepsilon|\,dx \le C'' \quad \forall \varepsilon > 0\,,$$

where C'' depends on $\|u\|_{L^s(\Gamma)}$ but not on ε. This proves the boundedness of $\{y_\varepsilon\}_{\varepsilon>0}$ in $H^1(\Omega)$. Moreover, applying Stampacchia's truncation method (see, for instance, Stampacchia [**89**] or the exposition for semilinear elliptic equations in the textbook by Tröltzsch [**91**]), we deduce the uniform boundedness of y_ε independently of ε, i.e.

$$\|y_\varepsilon\|_{L^\infty(\Omega)} \le C_\infty\Big(\|u\|_{L^s(\Gamma)} + \|f(\cdot,0)\|_{L^p(\Omega)}\Big)\,, \tag{1.3.7}$$

where the constant C_∞ depends only on α_a and α_f but neither on $a_M(\cdot, y_\varepsilon)$ nor on $f_M(\cdot, y_\varepsilon)$ or on ε. By choosing

$$M \ge C_\infty\Big(\|u\|_{L^s(\Gamma)} + \|f(\cdot,0)\|_{L^p(\Omega)}\Big)\,,$$

(1.3.7) implies that $a_M(x, y_\varepsilon(x)) = a(x, y_\varepsilon(x))$ and $f_M(x, y_\varepsilon(x)) = f(x, y_\varepsilon(x))$ for a.a. $x \in \Omega$, therefore $y_\varepsilon \in H^1(\Omega) \cap L^\infty(\Omega)$ is a solution of

$$\int_\Omega \{a(x, y_\varepsilon)\nabla y_\varepsilon \cdot \nabla\phi + (\varepsilon\chi_E y_\varepsilon + f(x, y_\varepsilon))\,\phi\}\,dx = \int_\Gamma u\phi\,d\sigma(x) \quad \forall \phi \in H^1(\Omega)\,. \tag{1.3.8}$$

By taking subsequences, $\{y_\varepsilon\}_{\varepsilon>0}$ converges weakly* in $L^\infty(\Omega)$ and weakly in $H^1(\Omega)$ to some $y \in H^1(\Omega) \cap L^\infty(\Omega)$ when $\varepsilon \to 0$. Thanks to the compactness of the embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ for every $q \in [1, \infty)$, the convergence $y_\varepsilon \to y$ is strong in every $L^q(\Omega)$. Passing to the limit in (1.3.8), we see that y satisfies (1.2.4), hence it is a solution of (1.1.1). Moreover, (1.3.7) and the weak* convergence $y_\varepsilon \rightharpoonup y$ in $L^\infty(\Omega)$ lead to the inequality

$$\|y\|_{L^\infty(\Omega)} \le C_\infty\Big(\|u\|_{L^s(\Gamma)} + \|f(\cdot,0)\|_{L^p(\Omega)}\Big)\,. \tag{1.3.9}$$

The Hölder regularity is well known; see Murthy and Stampacchia [**81**], Stampacchia [**88**] or Griepentrog and Recke [**59**], Gröger [**61**].

Uniqueness of the solution. Here we follow a comparison principle proposed by Hlaváček et al. [**67**]. Let $y_i \in H^1(\Omega) \cap L^\infty(\Omega)$, $i = 1, 2$, be two solutions of (1.1.1) and $\varepsilon > 0$. Since the regularity result proved above implies that $y_i \in C(\bar\Omega)$, we can define the open sets

$$\Omega_0 = \{x \in \Omega \,|\, y_2(x) - y_1(x) > 0\} \quad \text{and} \quad \Omega_\varepsilon = \{x \in \Omega \,|\, y_2(x) - y_1(x) > \varepsilon\}\,.$$

Now we introduce the function $z_\varepsilon(x) = \min\{\varepsilon, (y_2(x) - y_1(x))^+\}$ which belongs to $H^1(\Omega)$ and has the following properties: $0 \le z_\varepsilon \le \varepsilon$ in Ω, $\nabla z_\varepsilon(x) = 0$ for a.a. $x \notin \Omega_0 \backslash \Omega_\varepsilon$ and $\nabla z_\varepsilon = \nabla(y_2 - y_1)^+ = \nabla(y_2 - y_1)$ a.e. in $\Omega_0 \backslash \Omega_\varepsilon$. Choosing z_ε as test function in the weak formulations of the equations corresponding to y_i, along with (1.2.1) and the monotonicity of f, we get

$$\begin{aligned}
\alpha_a \|\nabla z_\varepsilon\|^2_{L^2(\Omega)} &\le \int_\Omega \Big\{a(x,y_2)|\nabla z_\varepsilon|^2 + (f(x,y_2) - f(x,y_1))\, z_\varepsilon\Big\}\, dx \\
&= \int_\Omega \{a(x,y_2)\nabla(y_2 - y_1)\cdot\nabla z_\varepsilon + (f(x,y_2) - f(x,y_1))\, z_\varepsilon\}\, dx \\
&= \int_\Omega (a(x,y_1) - a(x,y_2))\, \nabla y_1 \cdot \nabla z_\varepsilon \, dx \\
&= \int_{\Omega_0 \backslash \Omega_\varepsilon} (a(x,y_1) - a(x,y_2))\, \nabla y_1 \cdot \nabla z_\varepsilon \, dx \\
&\le C_{a,M} \|y_2 - y_1\|_{L^\infty(\Omega_0\backslash\Omega_\varepsilon)} \|\nabla y_1\|_{L^2(\Omega_0\backslash\Omega_\varepsilon)} \|\nabla z_\varepsilon\|_{L^2(\Omega_0\backslash\Omega_\varepsilon)} \\
&\le C_{a,M}\varepsilon \|\nabla y_1\|_{L^2(\Omega_0\backslash\Omega_\varepsilon)} \|\nabla z_\varepsilon\|_{L^2(\Omega)} \qquad (1.3.10)
\end{aligned}$$

with $M \ge \max\{\|y_1\|_{C(\bar\Omega)}, \|y_2\|_{C(\bar\Omega)}\}$. From Assumption 1.3 it follows

$$0 \le z_\varepsilon \le (y_2 - y_1)^+ = y_2 - y_1 \le \frac{1}{\alpha_f}\,(f(\cdot,y_2) - f(\cdot,y_1)) \quad \text{in } E \cap \bar\Omega_0\,,$$

hence

$$z_\varepsilon^2 \le \frac{1}{\alpha_f}\,(f(\cdot,y_2) - f(\cdot,y_1))\, z_\varepsilon \quad \text{in } E\,,$$

since $z_\varepsilon = 0$ in $E \setminus \bar\Omega_0$. By the Poi ncaré inequality (1.3.3), we then obtain as in (1.3.10)

$$\begin{aligned}
\|z_\varepsilon\|^2_{L^2(\Omega)} &\le C_E \left(\|\nabla z_\varepsilon\|^2_{L^2(\Omega)} + \|z_\varepsilon\|^2_{L^2(E)}\right) \\
&\le \frac{C_E}{\min\{\alpha_a, \alpha_f\}} \int_\Omega \Big\{a(x,y_2)|\nabla z_\varepsilon|^2 + (f(x,y_2) - f(x,y_1))\, z_\varepsilon\Big\}\, dx \\
&\le C\varepsilon \|\nabla y_1\|_{L^2(\Omega_0\backslash\Omega_\varepsilon)} \|\nabla z_\varepsilon\|_{L^2(\Omega)}\,. \qquad (1.3.11)
\end{aligned}$$

Combining (1.3.10) and (1.3.11), we infer

$$\|z_\varepsilon\|^2_{L^2(\Omega)} \le C'\varepsilon^2 \|\nabla y_1\|^2_{L^2(\Omega_0\backslash\Omega_\varepsilon)}\,.$$

Since $\lim_{\varepsilon\to 0} |\Omega_0 \backslash \Omega_\varepsilon| = 0$, we deduce from the last inequality

$$|\Omega_\varepsilon| = \varepsilon^{-2}\int_{\Omega_\varepsilon} \varepsilon^2\, dx = \varepsilon^{-2}\int_{\Omega_\varepsilon} |z_\varepsilon|^2\, dx \le C'\|\nabla y_1\|^2_{L^2(\Omega_0\backslash\Omega_\varepsilon)} \to 0 \quad \text{as } \varepsilon \to 0$$

which implies that $|\Omega_0| = \lim_{\epsilon\to 0} |\Omega_\varepsilon| = 0$ and hence $y_2 \le y_1$. In the same way, we prove that $y_2 \ge y_1$.

Proof of (1.3.1). Following the arguments on page 4 and taking $\phi = y_u$ in the equation (1.2.4) with y_u substituted for y, along with the positivity of a and the monotonicity of f, there holds

$$\begin{aligned}\min\{\alpha_a, \alpha_f\} \|y_u\|^2_{H^1(\Omega)} &\leq \int_\Omega \Big\{ a(x, y_u)|\nabla y_u|^2 + \left(f(x, y_u) - f(x, 0)\right) y_u \Big\}\, dx \\ &\leq C \left(\|u\|_{L^s(\Gamma)} + \|f(\cdot, 0)\|_{L^p(\Omega)} \right) \|y_u\|_{H^1(\Omega)} \,.\end{aligned}$$

Hence,

$$\|y_u\|_{H^1(\Omega)} \leq C_{a,f} \left(\|u\|_{L^s(\Gamma)} + \|f(\cdot, 0)\|_{L^p(\Omega)} \right). \tag{1.3.12}$$

Finally, inequality (1.3.1) follows from (1.3.12), (1.3.9), with y replaced by y_u, and the estimates in [**88**]. □

Remark 1.6. *The Lipschitz property of a w.r.t. the second variable is necessary only for the uniqueness of a solution of (1.1.1), not for its existence. If this property is violated Hlaváček et al. have presented in* [**67**] *an one-dimensional example of non-uniqueness of solutions.*

Remark 1.7. *If f is differentiable w.r.t. the second component then Assumption 1.3 implies that $(\partial f/\partial y) \geq \alpha_f$ in $E \times \mathbb{R}$ and it is dominated by an $L^p(\Omega)$ function. The proof of the existence of a solution in $H^1(\Omega) \cap C(\bar\Omega)$ of (1.1.1) is then easier; see* [**20**, Theorem 2.4]*. Let us briefly sketch the main steps of it. Utilizing the identity*

$$f(x, y(x)) = f(x, 0) + f_0(x, y(x))y(x) \quad \textit{with} \quad f_0(x, y) := \int_0^1 \frac{\partial f}{\partial y}(x, \theta y)\, d\theta\,,$$

it is enough to consider the mapping $F : L^2(\Omega) \longrightarrow L^2(\Omega)$ defined by $F(w) = z$, where z is the unique solution in $H^1(\Omega)$ of

$$\begin{cases} -\mathrm{div}\,[a_M(x, w)\nabla z] + f_{0M}(x, w)z = -f(x, 0) & \textit{in}\, \Omega\,, \\ a_M(x, w)\partial_\nu z = u & \textit{on}\, \Gamma\,. \end{cases}$$

Here, f_{0M} denotes the truncation of f_0 as described in the proof of the previous theorem. Once again, by Schauder's theorem, there exists a fixed point $y_M \in H^1(\Omega)$ of F. The rest of the proof is along the lines of Theorem 1.5.

Remark 1.8. *It is important to remark that in view of the previous theorem, the set $\{y_u \,\big|\, u \in U\}$ is bounded in $C(\bar\Omega)$ for bounded $U \subset L^s(\Gamma)$, $s > 1$.*

If a is continuous in $\bar\Omega \times \mathbb{R}$ then we are going to prove that $y_u \in W^{1,r}(\Omega)$ for some $r > 2$. For equations of the type (1.1.1) with homogeneous Dirichlet boundary conditions such a regularity result is standard; see Morrey [**80**, pp. 156-157] or Giaquinta [**55**, § 18, p. 73]. To our best knowledge, in the case of quasilinear equations with inhomogeneous Neumann boundary conditions, $W^{1,r}(\Omega)$ regularity results have not yet appeared in the literature. To overcome this obstacle we will apply a result by Dauge [**46**] that holds true for equations with constant coefficients. For this reason, we follow the classical approach of freezing the coefficient a around certain points of the domain to perform a reduction from variable coefficient to constant coefficient.

Theorem 1.9. *Suppose that the Assumptions 1.1-1.3 hold and that $a : \bar{\Omega} \times \mathbb{R} \longrightarrow \mathbb{R}$ is continuous. Then there exists $\bar{r} > 3$ such that, for any*

$$2 < r \leq \begin{cases} \min\left\{\bar{r}, \dfrac{2p}{2-p}\right\} & \text{if } p \in \left(\dfrac{4}{3}, 2\right), \\ \bar{r} & \text{if } p \geq 2\,, \end{cases} \tag{1.3.13}$$

and any $u \in L^{r/2}(\Gamma)$, the solution y_u of (1.1.1) belongs to $W^{1,r}(\Omega)$. Moreover, for any bounded set $U \subset L^{r/2}(\Gamma)$ there exists a constant $C_U > 0$ such that

$$\|y_u\|_{W^{1,r}(\Omega)} \leq C_U \quad \forall u \in U\,. \tag{1.3.14}$$

In addition, if Ω is convex then the above conclusions remain valid for some $\bar{r} \geq \frac{6}{3-\sqrt{5}}$.

Proof. Let us first assume that $2 < r \leq (2p)/(2-p)$ if $p < 2$ and $r \in (2, \infty)$ otherwise. The restriction $r \leq \bar{r}$ for some $\bar{r} > 3$ will be imposed later. By virtue of Theorem 1.5, (1.1.1) admits a unique solution y_u in $H^1(\Omega) \cap C(\bar{\Omega})$. We have to prove its $W^{1,r}(\Omega)$ regularity. Note that, thanks to our assumptions and the continuity of y_u, $\tilde{a}(\cdot) := a(\cdot, y_u(\cdot))$ is continuous in $\bar{\Omega}$.

Let $\rho > 0$ then there exists a finite number of boundary points $\{x_j\}_{j=1}^m \subset \Gamma$ such that $\Gamma \subset \bigcup_{j=1}^m B_{\mathbb{R}^2}(x_j, \rho)$. Further, let D be an open set with regular boundary such that $D \subset \bar{D} \subset \Omega$ and $\Omega \subset \bigcup_{j=1}^m B_{\mathbb{R}^2}(x_j, \rho) \cup D$. We also take a partition of unity $\{\psi_j\}_{j=0}^m \subset C^\infty(\mathbb{R}^2)$ with $\sum_{j=0}^m \psi_j(x) = 1$ and $0 \leq \psi_j(x) \leq 1$ $\forall x \in \bar{\Omega}$ and $j = 0, ..., m$, supp $\psi_0 \subset D$ and supp $\psi_l \subset B_{\mathbb{R}^2}(x_l, \rho)$ for $l = 1, ..., m$. Then $y_u = \sum_{j=0}^m y_j$, where $y_j := \psi_j y_u$. We prove that $y_j \in W^{1,r}(\Omega)$ for every $j = 0, ..., m$.

For $j = 0$ we have $y_0 = \psi_0 y_u$, hence

$$\begin{aligned} -\operatorname{div}[\tilde{a}(x)\nabla y_0] + y_0 &= -\operatorname{div}[\tilde{a}(x)\psi_0 \nabla y_u] - \operatorname{div}[\tilde{a}(x) y_u \nabla\psi_0] + \psi_0 y_u \\ &= -\psi_0 \operatorname{div}[\tilde{a}(x)\nabla y_u] - \tilde{a}(x)\nabla y_u \cdot \nabla\psi_0 - \operatorname{div}[\tilde{a}(x) y_u \nabla\psi_0] + \psi_0 y_u \\ &= -\psi_0 f(x, y_u) - \tilde{a}(x)\nabla y_u \cdot \nabla\psi_0 - \operatorname{div}[\tilde{a}(x) y_u \nabla\psi_0] + \psi_0 y_u \\ &= G \quad \text{in } \Omega \end{aligned} \tag{1.3.15}$$

and $y_0 = 0$ on Γ. It suffices to show that $G \in W^{-1,r}(\Omega) := W_0^{1,r'}(\Omega)^*$, where $W_0^{1,r'}(\Omega)$ denotes the closure in $W^{1,r'}(\Omega)$ of the space $C_c^\infty(\Omega)$ consisting of all infinitely differentiable functions with compact support in Ω. Then the $W^{1,r}(\Omega)$ regularity of y_0 follows from Morrey [**80**, pp. 156-157]. To prove that $\psi_0 f \in W^{-1,r}(\Omega)$ we make use of the Sobolev embedding $W_0^{1,r'}(\Omega) \hookrightarrow L^{p'}(\Omega)$ which holds true when $1/p' \geq 1/r' - 1/2$ or equivalently $1/r \geq 1/p - 1/2$. The last inequality is valid due to our assumption on r. Moreover, $\tilde{a}(\cdot)\nabla y_u\cdot\nabla\psi_0 \in L^2(\Omega)$, $\operatorname{div}[\tilde{a}(\cdot)y_u\nabla\psi_0] \in H^1(\Omega)^*$ and the inclusions $L^2(\Omega) \subset H^1(\Omega)^* \subset W^{-1,r}(\Omega)$, hence we conclude that $G \in W^{-1,r}(\Omega)$.

We fix now $j = 1, ..., m$ and $x_j \in \Gamma$. Employing (1.2.4) with y replaced by y_u, we have for arbitrary $z \in H^1(\Omega)$

$$\begin{aligned}\int_\Omega \tilde{a}(x)\left(z\nabla y_u\cdot\nabla\psi_j + \psi_j\nabla y_u\cdot\nabla z\right)dx &= \int_\Omega \tilde{a}(x)\nabla y_u\cdot\nabla(\psi_j z)\,dx\\ &= -\int_\Omega f(x,y_u)\psi_j z\,dx + \int_\Gamma u\psi_j z\,d\sigma(x)\,,\end{aligned}$$

therefore

$$\begin{aligned}\int_\Omega \{\tilde{a}(x)\nabla y_j\cdot\nabla z + y_j z\}dx &= \int_\Omega \{\tilde{a}(x)\left(y_u\nabla\psi_j\cdot\nabla z + \psi_j\nabla y_u\cdot\nabla z\right) + y_u\psi_j z\}\,dx\\ &= \int_\Omega \{\tilde{a}(x)\left(y_u\nabla\psi_j\cdot\nabla z - z\nabla y_u\cdot\nabla\psi_j\right) + \left(y_u - f(x,y_u)\right)\psi_j z\}\,dx\\ &\quad+\int_\Gamma u\psi_j z\,d\sigma(x) =: F(z)\,. \qquad (1.3.16)\end{aligned}$$

F is a linear continuous functional on $W^{1,r'}(\Omega)$. To verify this consider, for instance, the terms $z\nabla y_u$ and $uz|_\Gamma$ with $z \in W^{1,r'}(\Omega)$:

$$\|z\nabla y_u\|_{L^1(\Omega)} \le C\|z\|_{L^{\frac{2r}{r-2}}(\Omega)}\|\nabla y_u\|_{L^{\frac{2r}{r+2}}(\Omega)} \le C\|z\|_{W^{1,r'}(\Omega)}\|\nabla y_u\|_{L^2(\Omega)}$$

is a consequence of the embedding $W^{1,r'}(\Omega) \hookrightarrow L^{\frac{2r}{r-2}}(\Omega)$ and the fact that the conjugate number of $\frac{2r}{r-2}$ is $\frac{2r}{r+2} < 2$. Moreover, $z|_\Gamma \in W^{1-1/r',r'}(\Gamma) \hookrightarrow L^{\frac{r}{r-2}}(\Gamma)$, cf. Grisvard [**60**, Theorem 1.5.1.3], hence Hölder's inequality yields

$$\begin{aligned}\|uz|_\Gamma\|_{L^1(\Gamma)} &\le \|u\|_{L^{r/2}(\Gamma)}\|z|_\Gamma\|_{L^{\frac{r}{r-2}}(\Gamma)}\\ &\le C\|u\|_{L^{r/2}(\Gamma)}\|z|_\Gamma\|_{W^{1-1/r',r'}(\Gamma)}\\ &\le C\|u\|_{L^{r/2}(\Gamma)}\|z\|_{W^{1,r'}(\Omega)}\,.\end{aligned}$$

From (1.3.16) we get for $z \in W^{1,r'}(\Omega)$

$$\int_\Omega \{\tilde{a}(x_j)\nabla y_j\cdot\nabla z + y_j z\}\,dx = \int_{\Omega\cap B_{\mathbb{R}^2}(x_j,\rho)} \left(\tilde{a}(x_j) - \tilde{a}(x)\right)\nabla y_j\cdot\nabla z\,dx + F(z)\,.$$

Consider now the mapping $\mathcal{F} : W^{1,r}(\Omega) \longrightarrow W^{1,r}(\Omega)$, $\mathcal{F}(w) = y_w$, where y_w is the solution of the problem

$$\int_\Omega \{\tilde{a}(x_j)\nabla y_w\cdot\nabla z + y_w z\}\,dx = \int_{\Omega\cap B_{\mathbb{R}^2}(x_j,\rho)} \left(\tilde{a}(x_j) - \tilde{a}(x)\right)\nabla w\cdot\nabla z\,dx + F(z)\,, \qquad (1.3.17)$$

for every $z \in W^{1,r'}(\Omega)$. According to Dauge [**46**, Corollary 3.10], there exists $\bar{r} > 3$ such that, for any r satisfying (1.3.13), the equation (1.3.17) admits a unique solution

$y_w \in W^{1,r}(\Omega)$. Consequently, $\mathcal{F}$ is well-defined for these values of r. If Ω is convex then $\bar{r} \geq \frac{6}{3-\sqrt{5}}$, cf. [**46**, Corollary 3.12].

Next we prove that $\mathcal{F}$ is a contraction so that Banach's fixed-point theorem is applicable and yields $y_j \in W^{1,r}(\Omega)$. For $w_i \in W^{1,r}(\Omega)$, $i = 1, 2$, we find

$$\|\mathcal{F}(w_2) - \mathcal{F}(w_1)\|_{W^{1,r}(\Omega)} \leq C \|\tilde{a}(x_j) - \tilde{a}(\cdot)\|_{L^\infty(\Omega \cap B_{\mathbb{R}^2}(x_j,\rho))} \|w_2 - w_1\|_{W^{1,r}(\Omega)} \,,$$

where C depends only on α_a and not on x_j. Let us show this estimate. Consider first the mapping $T : W^{1,r}(\Omega) \longrightarrow W^{1,r'}(\Omega)^*$, $\langle Tw, z\rangle = \int_\Omega \{\nabla w \cdot \nabla z + wz\}\, dx$. Then T is bijective, cf. [**46**, p. 233], and continuous, hence it is an isomorphism. The inverse of T is also an isomorphism, thus, given $\xi \in W^{1,r'}(\Omega)^*$, the solution w of the equation $Tw = \xi$ satisfies $\|w\|_{W^{1,r}(\Omega)} \leq C\|\xi\|_{W^{1,r'}(\Omega)^*}$. This fact can be used to deduce from the identity

$$\begin{aligned}
&\int_\Omega \{a(x_j)\nabla(y_{w_2} - y_{w_1}) \cdot \nabla z + (y_{w_2} - y_{w_1})z\}\, dx \\
&\qquad = \int\limits_{\Omega \cap B_\rho(x_j)} (\tilde{a}(x_j) - \tilde{a}(x))\, \nabla(w_2 - w_1) \cdot \nabla z \, dx =: L(z) \quad \forall z \in W^{1,r'}(\Omega)\,,
\end{aligned}$$

the following inequality

$$\begin{aligned}
\|\mathcal{F}(w_2) - \mathcal{F}(w_1)\|_{W^{1,r}(\Omega)} &\leq C\|L\|_{W^{1,r'}(\Omega)^*} \\
&\leq C \sup_{\|z\|_{W^{1,r'}(\Omega)} \leq 1} \; \sup_{x \in \Omega \cap B_{\mathbb{R}^2}(x_j,\rho)} |\tilde{a}(x_j) - \tilde{a}(x)|\, \|w_2 - w_1\|_{W^{1,r}(\Omega)} \|z\|_{W^{1,r'}(\Omega)} \\
&\leq \tilde{C}\|\tilde{a}(x_j) - \tilde{a}(\cdot)\|_{L^\infty(\Omega \cap B_{\mathbb{R}^2}(x_j,\rho))} \|w_2 - w_1\|_{W^{1,r}(\Omega)} \,.
\end{aligned}$$

Now we can choose ρ sufficiently small such that $\tilde{C}\|\tilde{a}(x_j) - \tilde{a}(\cdot)\|_{L^\infty(\Omega \cap B_{\mathbb{R}^2}(x_j,\rho))} < 1$, hence $\mathcal{F}$ is a contraction.

We finish the proof by verifying (1.3.14) which follows from the same estimate for y_j, $j = 0, ..., m$. For y_0 we get the desired estimate from known results for linear elliptic equations, cf. Morrey [**80**, pp. 156-157]. To estimate $\|y_j\|_{W^{1,r}(\Omega)}$, with $j \geq 1$, we are going to use an estimate for the linear case and the fact that $\mathcal{F}$ is a contraction. Let us take $w^0 = 0$ and $w^{k+1} = \mathcal{F}(w^k)$, $k = 0, 1, 2, ...$, then $\|w^{k+1}\|_{W^{1,r}(\Omega)} \leq \frac{\Lambda_\mathcal{F}}{1-\Lambda_\mathcal{F}}\|w^1\|_{W^{1,r}(\Omega)}$, where $\Lambda_\mathcal{F}$ is the contractivity constant of $\mathcal{F}$. Since $\{w^k\}_{k=1}^\infty$ converges to the unique fixed point y_j, $\|y_j\|_{W^{1,r}(\Omega)} \leq \frac{\Lambda_\mathcal{F}}{1-\Lambda_\mathcal{F}}\|w^1\|_{W^{1,r}(\Omega)}$. Moreover, w^1 is the solution of the linear equation

$$\int\limits_\Omega \Big\{\tilde{a}(x_j)\nabla w^1 \cdot \nabla z + w^1 z\Big\}\, dx = \underbrace{\int\limits_{\Omega \cap B_{\mathbb{R}^2}(x_j,\rho)} (\tilde{a}(x_j) - \tilde{a}(x))\, \nabla w^0 \cdot \nabla z \, dx}_{=0} + F(z)$$

$\forall z \in W^{1,r'}(\Omega)$, hence as above $\|w^1\|_{W^{1,r}(\Omega)} \leq C\|F\|_{W^{1,r'}(\Omega)^*}$ and the proof is complete. □

Remark 1.10. *The paper by Dauge* [**46**], *cited in the proof of the previous theorem, deals only with problems in 3D but the result is also valid for dimension two. This can be seen as follows. Given the solution $y \in H^1(\Omega)$ of the problem*

$$\begin{cases} -\Delta y + cy = f & \text{in } \Omega\,, \\ \partial_\nu y = u & \text{on } \Gamma\,, \end{cases}$$

with a positive constant c, we introduce the prism $\tilde{\Omega} = \Omega \times (0,1)$ and consider the problem

$$\begin{cases} -\Delta \tilde{y} + c\tilde{y} = \tilde{f} & \text{in } \tilde{\Omega}\,, \\ \partial_\nu \tilde{y} = \tilde{u} & \text{on } \Gamma \times (0,1)\,, \\ \partial_\nu \tilde{y} = 0 & \text{on } \Omega \times \{0,1\}\,, \end{cases}$$

where $\tilde{f} : \tilde{\Omega} \times \mathbb{R} \longrightarrow \mathbb{R}$ and $\tilde{u} : \Gamma \times (0,1) \longrightarrow \mathbb{R}$ are defined by $\tilde{f}(\tilde{x}) = f(x_1, x_2)$ and $\tilde{u}(\tilde{x}) = u(x_1, x_2)$, with $\tilde{x} = (x_1, x_2, x_3) \in \tilde{\Omega}$. Then there holds $\tilde{y}_{\tilde{u}} \in W^{1,r}(\tilde{\Omega})$ and $\tilde{y}_{\tilde{u}}(\tilde{x}) = y_u(x_1, x_2)$, therefore $y_u \in W^{1,r}(\Omega)$.

Remark 1.11. *Theorems 1.5 and 1.9 are still valid if we require in Assumption 1.3 that ϕ_M and $f(\cdot, 0)$ belong to $L^p(\Omega)$ with $p \geq 2q/(2+q) > 1$ and $q > 2$. The reason for assuming $p > 4/3$ will become clear in Theorem 1.18.*

1.4. Regularity of solutions of elliptic PDEs in domains with corners

This section contains some auxiliary results concerning the regularity of solutions of the following Neumann problem

$$\begin{cases} -\Delta y = f & \text{in } \Omega\,, \\ \partial_\nu y = g & \text{on } \Gamma\,, \end{cases} \tag{1.4.1}$$

where $\Omega \subset \mathbb{R}^2$ satisfies Assumption 1.1. Further, we assume that $f \in L^p(\Omega)$ with $p > 4/3$, $g \in L^2(\Gamma)$ satisfying the compatibility condition

$$\int_\Omega f(x)\,dx + \int_\Gamma g(x)\,d\sigma(x) = 0\,. \tag{1.4.2}$$

It is a well-known consequence of the Lax-Milgram theorem that (1.4.1) has a solution in $H^1(\Omega)$ that is unique up to an additive constant; see Grisvard [**60**, Theorem 4.4.3.1] when $g = 0$.

Theorem 1.12. *Suppose that Assumption 1.1 holds. There exists $s_\Omega > 3/2$ depending on Ω such that for every $3/2 < s < s_\Omega$ the problem*

$$\begin{cases} -\Delta z = \zeta & \text{in } \Omega\,, \\ \partial_\nu z = g & \text{on } \Gamma\,, \end{cases}$$

with $\zeta \in H^{s-2}(\Omega)$, $g \in L^2(\Gamma)$ and

$$\langle \zeta, 1 \rangle_{H^{s-2}(\Omega), H^{2-s}(\Omega)} + \int_\Gamma g(x)\,d\sigma(x) = 0\,, \tag{1.4.3}$$

has a unique solution $z \in H^{3/2}(\Omega)$ up to an additive constant. Moreover, there exists a constant $C_s > 0$ independent of ζ and g such that

$$\|z\|_{H^{3/2}(\Omega)} \le C_s \left(\|\zeta\|_{H^{s-2}(\Omega)} + \|g\|_{L^2(\Gamma)} + \left| \int_\Omega z(x)\, dx \right| \right) . \tag{1.4.4}$$

The last term in (1.4.4) is a consequence of the uniqueness of z up to an additive constant. We should remark that $H^{s-2}(\Omega) := H_0^{2-s}(\Omega)^*$ if $s < 2$; the latter space is defined analogous to $W_0^{1,r'}(\Omega)$ on page 9. However, because $0 < 2 - s < 1/2$, there holds $H_0^{2-s}(\Omega) = H^{2-s}(\Omega)$, hence $H^{s-2}(\Omega) = H^{2-s}(\Omega)^*$; see, for instance, Grisvard [**60**, Theorem 1.4.5.2-(c)]. Let us also mention that $\langle \zeta, 1 \rangle_{H^{s-2}(\Omega), H^{2-s}(\Omega)} = \int_\Omega \zeta\, dx$ if $\zeta \in L^p(\Omega)$ with $p \ge 4/3$; see Corollary 1.13 below.

PROOF. To simplify the notation we will write $\langle \cdot, \cdot \rangle$ instead of $\langle \cdot, \cdot \rangle_{H^{s-2}(\Omega), H^{2-s}(\Omega)}$. Let us consider the problems

$$\begin{cases} -\Delta z_1 = \frac{1}{|\Omega|} \langle \zeta, 1 \rangle & \text{in } \Omega\,, \\ z_1 = 0 & \text{on } \Gamma\,, \end{cases} \tag{1.4.5}$$

$$\begin{cases} -\Delta z_2 = \zeta - \frac{1}{|\Omega|} \langle \zeta, 1 \rangle & \text{in } \Omega\,, \\ \partial_\nu z_2 = 0 & \text{on } \Gamma\,, \end{cases} \tag{1.4.6}$$

$$\begin{cases} -\Delta z_3 = 0 & \text{in } \Omega\,, \\ \partial_\nu z_3 = g - \partial_\nu z_1 & \text{on } \Gamma\,. \end{cases} \tag{1.4.7}$$

According to Dauge [**44**, Theorem 23.3] or [**45**, Theorem 3], there exists $s_\Omega > 3/2$, depending on the angles of Ω and the minimum positive eigenvalue of the Laplace operator in Ω, such that the problem (1.4.5) has a unique solution z_1 in $H^s(\Omega)$ if $\zeta \in H^{s-2}(\Omega)$ and $3/2 < s < s_\Omega$. Moreover, the following estimate holds

$$\|z_1\|_{H^s(\Omega)} \le C_{s,1} \|\zeta\|_{H^{s-2}(\Omega)}\,. \tag{1.4.8}$$

The regularity of z_1 implies that $\partial_\nu z_1 \in L^2(\Gamma)$ (see also page 14) and by integrating (1.4.5),

$$-\int_\Gamma \partial_\nu z_1\, d\sigma(x) = -\int_\Omega \Delta z_1\, dx = \langle \zeta, 1 \rangle\,. \tag{1.4.9}$$

The existence of a unique solution $z_2 \in H^s(\Omega)$ of (1.4.6), up to an additive constant, follows from Dauge [**44**, Corollary 23.5] for $3/2 < s < s_\Omega$ and $\zeta \in H^{s-2}(\Omega)$. Now we have the estimate

$$\begin{aligned} \|z_2\|_{H^s(\Omega)} &\le C_{s,2} \left(\left\| \zeta - \frac{1}{|\Omega|} \langle \zeta, 1 \rangle \right\|_{H^{s-2}(\Omega)} + \left| \int_\Omega z_2(x)\, dx \right| \right) \\ &\le C'_{s,2} \left(\|\zeta\|_{H^{s-2}(\Omega)} + \left| \int_\Omega z_2(x)\, dx \right| \right) . \end{aligned}$$

Finally, taking into account (1.4.3) and (1.4.9), we get

$$\int_\Gamma (g - \partial_\nu z_1)\, d\sigma(x) = \int_\Gamma g\, d\sigma(x) + \langle \zeta, 1\rangle = 0\,.$$

Thus, from Kenig [**72**, p. 121] we deduce the existence of a solution z_3 in $H^{3/2}(\Omega)$ of (1.4.7) that is unique up to the addition of a constant. Similarly, we obtain with the help of the continuity of the trace operator and (1.4.8)

$$\begin{aligned}\|z_3\|_{H^{3/2}(\Omega)} &\le C\left(\|g - \partial_\nu z_1\|_{L^2(\Gamma)} + \left|\int_\Omega z_3(x)\, dx\right|\right)\\ &\le C\left(\|g\|_{L^2(\Gamma)} + \|z_1\|_{H^s(\Omega)} + \left|\int_\Omega z_3(x)\, dx\right|\right)\\ &\le C_{s,3}\left(\|g\|_{L^2(\Gamma)} + \|\zeta\|_{H^{s-2}(\Omega)} + \left|\int_\Omega z_3(x)\, dx\right|\right).\end{aligned}$$

Consequently, $z_1 + z_2 + z_3 \in H^{3/2}(\Omega)$ and $z = z_1 + z_2 + z_3 + \mathit{constant}$. Inequality (1.4.4) follows immediately from the above estimates for z_i, $i = 1, 2, 3$. □

Corollary 1.13. *Suppose that Assumption 1.1 holds, $f \in L^p(\Omega)$ with $p > 4/3$, $g \in L^2(\Gamma)$ and the condition (1.4.2) is satisfied. If $y \in H^1(\Omega)$ is a solution of (1.4.1) then $y \in H^{3/2}(\Omega)$ and there exists a constant $C > 0$, independent of f and g, such that*

$$\|y\|_{H^{3/2}(\Omega)} \le C\left(\|f\|_{L^p(\Omega)} + \|g\|_{L^2(\Gamma)} + \left|\int_\Omega y(x)\, dx\right|\right). \tag{1.4.10}$$

Proof. It suffices to prove that $L^p(\Omega) \subset H^{s-2}(\Omega)$ for some $3/2 < s < s_\Omega$. Then we can apply the previous theorem to obtain the desired regularity of y and (1.4.10). Thanks to the inclusion $H^{1/2}(\Omega) \subset L^4(\Omega)$, we can take $\varepsilon > 0$ small enough and s close to $3/2$ such that $p > (4-\varepsilon)/(3-\varepsilon)$ and $H^{2-s}(\Omega) \subset L^{4-\varepsilon}(\Omega)$. Then it follows $L^p(\Omega) \subset L^{\frac{(4-\varepsilon)}{(3-\varepsilon)}}(\Omega) = L^{4-\varepsilon}(\Omega)^* \subset H^{2-s}(\Omega)^* = H^{s-2}(\Omega)$. □

The next proposition deals with the regularity of the trace and normal derivative of a function belonging to $H^{3/2}(\Omega)$. It is known that for Lipschitz domains the trace operator is linear and continuous from $H^s(\Omega)$ to $H^{s-1/2}(\Gamma)$ if $1/2 < s < 3/2$, as well as from $H^s(\Omega)$ to $H^1(\Gamma)$ if $s > 3/2$, cf. Ding [**48**]. However, Jerison and Kenig [**71**, § 3] have constructed a function in $H^{3/2}(\Omega)$, whose trace is not in $H^1(\Gamma)$.

Furthermore, if Ω is polygonal and $y \in H^s(\Omega)$ with $3/2 < s < 2$ then $\partial_\nu y \in L^2(\Gamma)$. Indeed, $\partial_i y \in H^{s-1}(\Omega)$ for $i = 1, 2$, hence there exists the trace $(\partial_i y)|_\Gamma \in H^{s-3/2}(\Gamma)$. Now $\partial_\nu y = \nabla y \cdot \nu = (\partial_1 y)|_\Gamma \nu_1 + (\partial_2 y)|_\Gamma \nu_2$. The difficulty comes from the discontinuity of the normal vector ν. In polygons, ν is constant on every edge e of Γ, therefore $\partial_\nu y|_e \in H^{s-3/2}(e)$ and $0 < s - 3/2 < 1/2$. Finally, we can apply Theorem 1.5.2.3-(a) in Grisvard's book [**60**] to deduce, without additional conditions, that $\partial_\nu y \in L^2(\Gamma)$.

PROPOSITION 1.14. *Under the Assumption 1.1, if $y \in H^{3/2}(\Omega)$ and $\Delta y \in L^p(\Omega)$ with $p > 4/3$ then $y|_\Gamma \in H^1(\Gamma)$, $\partial_\nu y \in L^2(\Gamma)$ and the following estimate holds*

$$\|y|_\Gamma\|_{H^1(\Gamma)} + \|\partial_\nu y\|_{L^2(\Gamma)} \leq C \left(\|y\|_{H^{3/2}(\Omega)} + \|\Delta y\|_{L^p(\Omega)}\right). \tag{1.4.11}$$

PROOF. Let $y_1 \in H^1(\Omega)$ satisfy $-\Delta y_1 = -\Delta y$ in Ω and $y_1 = 0$ on Γ. As in the proof of Theorem 1.12, $y_1 \in H^s(\Omega)$ with some $s > 3/2$, hence $\partial_\nu y_1 \in L^2(\Gamma)$ and

$$\|\partial_\nu y_1\|_{L^2(\Gamma)} \leq C\|y_1\|_{H^s(\Omega)} \leq C\|\Delta y\|_{L^p(\Omega)}. \tag{1.4.12}$$

Setting $y_2 = y - y_1$, it follows that y_2 is harmonic and $y_2 \in H^{3/2}(\Omega)$. Therefore, according to Jerison and Kenig [**71**, Theorem 5.6 and Corollary 5.7], $y_2|_\Gamma \in H^1(\Gamma)$, $\partial_\nu y_2 \in L^2(\Gamma)$ and

$$\|y_2|_\Gamma\|_{H^1(\Gamma)} + \|\partial_\nu y_2\|_{L^2(\Gamma)} \leq C\|y_2\|_{H^{3/2}(\Omega)}. \tag{1.4.13}$$

Furthermore, $y_1|_\Gamma = 0$ and $y|_\Gamma = y_2|_\Gamma \in H^1(\Gamma)$. Since $\partial_\nu y = \partial_\nu y_1 + \partial_\nu y_2 \in L^2(\Gamma)$, (1.4.11) follows from (1.4.12)-(1.4.13) and the continuity of the trace operator:

$$\begin{aligned}\|y|_\Gamma\|_{H^1(\Gamma)} + \|\partial_\nu y\|_{L^2(\Gamma)} &\leq \|y_2|_\Gamma\|_{H^1(\Gamma)} + \|\partial_\nu y_1\|_{L^2(\Gamma)} + \|\partial_\nu y_2\|_{L^2(\Gamma)} \\ &\leq C \left(\|y_2\|_{H^{3/2}(\Omega)} + \|\Delta y\|_{L^p(\Omega)}\right) \\ &\leq C \left(\|y\|_{H^{3/2}(\Omega)} + \|\Delta y\|_{L^p(\Omega)}\right).\end{aligned}$$

□

COROLLARY 1.15. *Let Assumption 1.1 be fulfilled, $p > 4/3$ and $c \in L^p(\Omega)$ satisfy $c(\cdot) \geq 0$ a.e. in Ω and $c(\cdot) \geq \alpha > 0$ at least on a subset of Ω with positive measure. Then the problem*

$$\begin{cases} -\Delta y + c(x)y = f & \textit{in}\, \Omega, \\ \partial_\nu y = g & \textit{on}\, \Gamma, \end{cases} \tag{1.4.14}$$

with $f \in L^p(\Omega)$ and $g \in L^2(\Gamma)$, has a unique solution $y \in H^{3/2}(\Omega)$ and

$$\|y\|_{H^{3/2}(\Omega)} \leq C_c \left(\|f\|_{L^p(\Omega)} + \|g\|_{L^2(\Gamma)}\right), \tag{1.4.15}$$

where $C_c > 0$ depends on $\|c\|_{L^p(\Omega)}$ but it is independent of f and g.

PROOF. Under the assumptions of the corollary, the existence and uniqueness of a solution $y \in H^1(\Omega) \cap C(\bar{\Omega})$ of (1.4.14) follows, for instance, from Alibert and Raymond [**3**, Theorem 2]. Moreover,

$$\|y\|_{H^1(\Omega)} + \|y\|_{C(\bar{\Omega})} \leq C_1 \left(\|f\|_{L^p(\Omega)} + \|g\|_{L^2(\Gamma)}\right)$$

with $C_1 > 0$ independent of f, g and c. The $H^{3/2}(\Omega)$ regularity of y follows from Corollary 1.13 when replacing f by $f - cy \in L^p(\Omega)$. To show (1.4.15) we apply

(1.4.10) and get

$$\begin{aligned}\|y\|_{H^{3/2}(\Omega)} &\le C\left(\|f-cy\|_{L^p(\Omega)}+\|g\|_{L^2(\Gamma)}\right)\\ &\le C\left(\|f\|_{L^p(\Omega)}+\|c\|_{L^p(\Omega)}\|y\|_{C(\bar\Omega)}+\|g\|_{L^2(\Gamma)}\right)\\ &\le C_2\left(\|f\|_{L^p(\Omega)}+C_1\|c\|_{L^p(\Omega)}\left(\|f\|_{L^p(\Omega)}+\|g\|_{L^2(\Gamma)}\right)+\|g\|_{L^2(\Gamma)}\right)\\ &= C_2\left(1+C_1\|c\|_{L^p(\Omega)}\right)\left(\|f\|_{L^p(\Omega)}+\|g\|_{L^2(\Gamma)}\right).\end{aligned}$$

Thus, (1.4.15) is obtained by setting $C_c = C_2\left(1+C_1\|c\|_{L^p(\Omega)}\right)$. □

Remark 1.16. *The $H^{3/2}(\Omega)$ regularity of the solution y of (1.4.1) if $f \in L^2(\Omega)$ is studied by Casas et al.* [**29**, Lemma 2.2]. *Similarly to our technique, the Neumann problem is decomposed in two different problems and a combination of the results by Jerison and Kenig* [**70, 71**] *is used. For convex domains in $\mathbb{R}^n$, $n \ge 1$, an analogous result to Proposition 1.14 is proved in Casas et al.* [**28**, Lemma A.1].

1.5. Higher regularity of solutions of the quasilinear equation

In this section, we will see that stronger assumptions on the coefficient a of the differential operator yield higher regularity of the solution of (1.1.1). This higher regularity is crucial to confirm a high rate of convergence for the finite element approximation of (1.1.1); see Chapter 2. Since we will need in the sequel the differentiability of a a.e. in Ω, we impose the following assumption.

Assumption 1.17. *For every $M > 0$ there exists a constant $C_M > 0$ such that, for all $x_i \in \bar\Omega$, $|y_i| \le M$, $i = 1,2$, the following local Lipschitz property is satisfied:*

$$|a(x_2,y_2)-a(x_1,y_1)| \le C_M\left(|x_2-x_1|+|y_2-y_1|\right).$$

Since a is Lipschitz in x and y and $C^{0,1}(\bar\Omega) = W^{1,\infty}(\Omega)$, a is differentiable a.e. in Ω with uniformly bounded weak partial derivatives and, given $y \in H^1(\Omega) \cap C(\bar\Omega)$, the chain rule is valid:

$$\partial_j\left[a(x,y(x))\right] = \left[\nabla_x a\right]_j(x,y(x)) + \frac{\partial a}{\partial y}(x,y(x))\partial_j y(x) \text{ for } j = 1,2, \text{ and a.a. } x \in \Omega.$$

Theorem 1.18. *Suppose that the Assumptions 1.1-1.3 and 1.17 hold. Then for any $u \in L^2(\Gamma)$ the solution y_u of (1.1.1) belongs to $H^{3/2}(\Omega)$. Moreover, for any bounded set $U \subset L^2(\Gamma)$ there exists a constant $C_U > 0$ such that*

$$\|y_u\|_{H^{3/2}(\Omega)} \le C_U \quad \forall u \in U. \tag{1.5.1}$$

Proof. From Theorem 1.9 we get that $y_u \in W^{1,r}(\Omega)$ for some $r > 3$ (notice that $\frac{2p}{2-p} > 3$ if $4/3 < p < 2$). Let us show that $y_u \in H^{3/2}(\Omega)$. Thanks to the assumptions on a and the continuity of y_u, it follows that $a(\cdot, y_u(\cdot)) \in C(\bar\Omega)$ and

$(\partial a/\partial y)(\cdot, y_u(\cdot)) \in L^\infty(\Omega)$. By expanding the divergence term of the equation (1.1.1) and dividing by $a = a(\cdot, y_u(\cdot)) \geq \alpha_a > 0$, we verify in Remark 1.20 below that

$$-\Delta y_u = \underbrace{\frac{1}{a}}_{L^\infty(\Omega)} \left(-\underbrace{f(\cdot, y_u)}_{L^p(\Omega)} + \sum_{j=1}^{2} \underbrace{\partial_j a(\cdot, y_u)}_{L^\infty(\Omega)} \underbrace{\partial_j y_u}_{L^r(\Omega)} + \underbrace{\frac{\partial a}{\partial y}(\cdot, y_u)}_{L^\infty(\Omega)} \underbrace{|\nabla y_u|^2}_{L^{r/2}(\Omega)} \right) \quad \text{in } \Omega\,, \tag{1.5.2}$$

$$\partial_\nu y_u = \frac{u}{a} \quad \text{on } \Gamma\,. \tag{1.5.3}$$

The right-hand sides of (1.5.2) and (1.5.3) are in $L^{\min\{p, r/2\}}(\Omega)$ and $L^2(\Gamma)$, respectively. Hence, we can apply Corollary 1.13 to obtain the $H^{3/2}(\Omega)$ regularity of y_u. Inequality (1.5.1) simply follows from (1.4.10) along with the Assumptions 1.2-1.3, Remark 1.8 and (1.3.14). □

CONCLUSION 1.19. In dimension two, there holds $H^{3/2}(\Omega) \subset W^{1,4}(\Omega)$. Therefore, if Assumption 1.17 is valid and $u \in L^2(\Gamma)$ then we can improve the regularity of y_u to $W^{1,4}(\Omega)$, no matter if $\bar{r} < 4$, where $\bar{r}$ is given in Theorem 1.9.

REMARK 1.20. *In the proof of the previous theorem we made use of the fact that the solution $y_u \in W^{1,r}(\Omega)$ $(r > 3)$ of (1.1.1) is also a solution of (1.5.2)-(1.5.3). Let us verify this. For $s = \min\{p, r/2\}$ we define the functional $F : L^{s'}(\Omega) \longrightarrow \mathbb{R}$ by*

$$F(z) = \int_\Omega \frac{1}{a} \left(-f(x, y_u) + \sum_{j=1}^{2} \partial_j a(x, y_u) \partial_j y_u + \frac{\partial a}{\partial y}(x, y_u) |\nabla y_u|^2 \right) z\, dx\,.$$

We prove that y_u is a solution of the equation

$$\begin{cases} -\Delta y = F & \text{in } \Omega\,, \\ \partial_\nu y = \dfrac{u}{a} & \text{on } \Gamma\,. \end{cases} \tag{1.5.4}$$

For that purpose we pass to the weak formulation of (1.5.4) and consider the problem of finding $y \in H^1(\Omega)$ such that, for every $z \in H^1(\Omega) \subset L^{s'}(\Omega)$,

$$\begin{aligned} \int_\Omega \nabla y \cdot \nabla z\, dx &= \int_\Gamma \frac{u}{a} z\, d\sigma(x) + F(z) \\ &= \int_\Gamma \frac{u}{a} z\, d\sigma(x) + \int_\Omega \frac{1}{a} \left(-f(x, y_u) + \sum_{j=1}^{2} \partial_j a(x, y_u) \partial_j y_u + \frac{\partial a}{\partial y}(x, y_u) |\nabla y_u|^2 \right) z\, dx\,. \end{aligned} \tag{1.5.5}$$

For arbitrary $z \in H^1(\Omega)$ we have $z\partial_j y_u \in L^2(\Omega)$, $j = 1, 2$, since $H^1(\Omega) \subset L^6(\Omega)$ and $\partial_j y_u \in L^3(\Omega)$. Then owing to the boundedness of $a(\cdot, y_u(\cdot))$, $[\nabla_x a](\cdot, y_u(\cdot))$ and $(\partial a/\partial y)(\cdot, y_u(\cdot))$, we deduce

$$\partial_j[a(\cdot, y_u)z] = [\nabla_x a]_j(\cdot, y_u)z + \frac{\partial a}{\partial y}(\cdot, y_u) z \partial_j y_u + a(\cdot, y_u)\partial_j z \in L^2(\Omega) \quad \textit{for } j = 1, 2.$$

Hence, $a(\cdot, y_u)z \in H^1(\Omega)$ and substituting this for the test function z in (1.5.5), we see that y_u satisfies (1.5.5). This proves that y_u is a solution of (1.5.4) and any other solution y of (1.5.4) is of the form $y = y_u + constant$.

The last regularity result of this section is only valid for convex polygonal domains. To our best knowledge, for non-convex corner domains, such a result does not exist in general. Related results for domains with smooth boundaries are addressed in Remark 1.22 below.

THEOREM 1.21. *Let the Assumptions 1.1-1.3 and 1.17 be satisfied. We also assume that Ω is convex and $p \geq 2$. Then there exists $r_0 > 2$ depending on the measure of the angles in Γ such that, for any*

$$2 \leq r \leq \min\{\bar{r}, r_0, p\}$$

($\bar{r} \geq \frac{6}{3-\sqrt{5}}$; see Theorem 1.9) and any $u \in W^{1-1/r,r}(\Gamma)$, the solution y_u of (1.1.1) belongs to $W^{2,r}(\Omega)$. Moreover, for any bounded set $U \subset W^{1-1/r,r}(\Gamma)$ there exists a constant $C_U > 0$ such that

$$\|y_u\|_{W^{2,r}(\Omega)} \leq C_U \quad \forall u \in U\,.$$

PROOF. First, we prove the result for $2 \leq r \leq \min\{\bar{r}/2, p\}$. Since $r \geq 2$, we have $u \in W^{1-1/r,r}(\Gamma) \subset H^{1/2}(\Gamma) \subset L^q(\Gamma)$ for any $1 \leq q < \infty$. Taking $q = \bar{r}/2$, we can apply Theorem 1.9 to deduce that $y_u \in W^{1,\bar{r}}(\Omega)$. We have to show the $W^{2,r}(\Omega)$ regularity of y_u. Repeating the steps of the proof of Theorem 1.18, we get that the right-hand side of (1.5.2) is in $L^r(\Omega)$. Hence, it is enough to prove that $u/a = (u/a(\cdot, y_u(\cdot))) \in W^{1-1/r,r}(\Gamma)$. Then a well-known result by Grisvard [**60**, Corollary 4.4.3.8] on maximal regularity yields the existence of $r_0 > 2$ such that $y_u \in W^{2,r}(\Omega)$ if $2 \leq r \leq \min\{r_0, \bar{r}/2, p\}$. This r_0 depends on the measure of the angles in Γ. To verify the $W^{1-1/r,r}(\Gamma)$ regularity of u/a we will use the following two facts:

$$\text{If } b \in C^{0,\mu}(\Gamma),\ u \in W^{1-1/r,r}(\Gamma) \text{ with } \mu > 1 - 1/r, \text{ then } bu \in W^{1-1/r,r}(\Gamma)\,; \tag{1.5.6}$$

cf. Grisvard [**60**, Theorem 1.4.1.1], and

$$W^{k,q}(\Omega) \subset C^{0,\mu}(\bar{\Omega}) \quad \text{with} \quad \mu \begin{cases} = k - 2/q & \text{if } k - 2/q < 1\,, \\ < 1 & \text{if } k - 2/q = 1\,, \\ = 1 & \text{if } k - 2/q > 1\,, \end{cases} \tag{1.5.7}$$

provided that $kq > 2$, cf. Necǎs [**82**, §2, Theorem 3.8]. Thanks to Assumption 1.17 and (1.2.1), $1/a$ is Lipschitz w.r.t. both variables. Furthermore, from (1.5.7) we get $y_u \in W^{1,\bar{r}}(\Omega) \subset C^{0,1-\frac{2}{\bar{r}}}(\bar{\Omega})$, therefore $(1/a) \in C^{0,1-\frac{2}{\bar{r}}}(\Gamma)$. Now there are two possibilities:

(1) If $r < \bar{r}/2$ or equivalently if $1 - 1/r < 1 - 2/\bar{r}$ then it follows from (1.5.6) that $(u/a) \in W^{1-1/r,r}(\Gamma)$. In particular, $y_u \in H^2(\Omega)$ if $u \in H^{1/2}(\Gamma)$.

(2) If $r = \bar{r}/2$ we cannot apply (1.5.6) directly. Nevertheless, because of the inclusion $u \in W^{1-1/r,r}(\Gamma) \subset H^{1/2}(\Gamma)$ we can use the above argumentation to obtain, in an intermediate step, that $y_u \in H^2(\Omega)$. This fact, together with $H^2(\Omega) \subset C^{0,\mu}(\bar{\Omega})$ for any $\mu \in (1-\frac{2}{\bar{r}}, 1)$, yields $(1/a) \in C^{0,\mu}(\Gamma)$. By applying (1.5.6), we deduce $(u/a) \in W^{1-1/r,r}(\Gamma)$ and $y_u \in W^{2,r}(\Omega)$.

Finally, we assume that $\bar{r}/2 < r \leq \min\{r_0, \bar{r}, p\}$ which implies $p > 2$; notice that $\bar{r} \geq \frac{6}{3-\sqrt{5}} > 4$, hence $\bar{r}/2 > 2$. From the first part of the proof we know that $y_u \in W^{2,\min\{r_0,\bar{r}/2,p\}}(\Omega) \subset C^{0,1}(\bar{\Omega})$. Then $(1/a) \in C^{0,1}(\Gamma)$ and from (1.5.6) we get $(u/a) \in W^{1-1/r,r}(\Gamma)$, once again. Moreover, the right-hand side of (1.5.2) is in $L^r(\Omega)$. By applying again Grisvard's result, we conclude $y_u \in W^{2,r}(\Omega)$. □

Remark 1.22. *The proof of Theorem 1.5 on the existence and uniqueness of a solution of (1.1.1) is valid for arbitrary Lipschitz domains in $\mathbb{R}^n$, $n \in \mathbb{N}$, provided that $s > n-1$ and $p > n/2$. On the other hand, the statement of Theorem 1.5 holds true, assuming that Γ is of class C^1 and that the leading coefficients of the elliptic differential operator are continuous, cf. Agmon et al.* [**2**, Theorems 15.3′ and 15.1″]. *If the boundary Γ is of class $C^{1,1}$ then the $W^{2,r}(\Omega)$ regularity result of Theorem 1.21 is a consequence of Theorem 2.4.2.7 in Grisvard* [**60**].

1.6. Differentiability of the solution operator

This section is devoted to the analysis of a class of linear elliptic differential equations of non-monotone type. This class includes, in particular, the linearization of the equation (1.1.1) around a solution y_u. The necessity of studying the linearization of the quasilinear equation arises when differentiability properties of the solution operator $u \longmapsto y_u$ have to be investigated.

Let us first consider the following linear elliptic equation in divergence form

$$\begin{cases} -\operatorname{div}\left[\tilde{a}(x)\nabla z(x) + \mathbf{b}(x)z(x)\right] + c(x)z(x) = \zeta(x) & \text{in } \Omega\,, \\ \left[\tilde{a}(x)\nabla z(x) + \mathbf{b}(x)z(x)\right]\cdot\nu = v(x) & \text{on } \Gamma\,, \end{cases} \tag{1.6.1}$$

where $\zeta \in H^1(\Omega)^*$, $v \in H^{-1/2}(\Gamma)$ are given and $\tilde{a}, c : \Omega \longrightarrow \mathbb{R}$, $\mathbf{b} : \Omega \longrightarrow \mathbb{R}^2$ are specified in the next theorem. We say that $z \in H^1(\Omega)$ is a solution of (1.6.1) if $\forall w \in H^1(\Omega)$ there holds

$$\int_\Omega \{(\tilde{a}(x)\nabla z + z\mathbf{b}(x))\cdot\nabla w + c(x)zw\}\,dx = \langle\zeta, w\rangle_{H^1(\Omega)^*,H^1(\Omega)} + \langle v, w\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)}\,.$$

In view of the occurrence of the non-monotone part $\operatorname{div}[\mathbf{b}z]$, the well-posedness of (1.6.1) is not obvious. The next theorem establishes the existence and uniqueness of a solution $z \in H^1(\Omega)$ of the previous equation.

Theorem 1.23. *Suppose that Assumption 1.1 holds. We also assume that $\beta > 2$ and $\gamma > 1$. Further, let $\tilde{a} \in L^\infty(\Omega)$ with $\tilde{a}(x) \geq \alpha > 0$ for a.a. $x \in \Omega$, $\mathbf{b} \in L^\beta(\Omega)^2$ and $c \in L^\gamma(\Omega)$, satisfying $c(x) \geq 0$ for a.a. $x \in \Omega$ and $c(x) \geq \alpha$ for all $x \in E$, where E is* *a measurable subset of* Ω *such that* $|E| > 0$. *Then the operator* $S : H^1(\Omega) \longrightarrow H^1(\Omega)^*$ *defined by*

$$\langle Sz, w\rangle_{H^1(\Omega)^*,H^1(\Omega)} = \int_\Omega \{\tilde{a}(x)\nabla z\cdot\nabla w + z\mathbf{b}(x)\cdot\nabla w + c(x)zw\}\, dx \tag{1.6.2}$$

is an isomorphism.

Proof. We follow here the ideas by Casas and Tröltzsch [**36**, Theorem 2.7], who study the linearized equation of (1.1.1) with homogeneous Dirichlet boundary conditions. Although (1.6.1) is a more general equation than the one considered by the previous authors and contains Neumann boundary data, the modification of the proof of [**36**, Theorem 2.7] is basically straightforward. Let us note the main differences.

Since the operator S is linear, the continuity of S is equivalent to its boundedness. By Hölder's inequality and the embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ for every $1 \le q < \infty$, we have for every $z \in H^1(\Omega)$

$$\begin{aligned}\|Sz\|_{H^1(\Omega)^*} &= \sup_{\|w\|_{H^1(\Omega)}\le 1} \left|\langle Sz, w\rangle_{H^1(\Omega)^*,H^1(\Omega)}\right| \\ &\le \sup_{\|w\|_{H^1(\Omega)}\le 1} \Big\{\|\tilde{a}\|_{L^\infty(\Omega)}\|\nabla z\|_{L^2(\Omega)}\|\nabla w\|_{L^2(\Omega)} \\ &\qquad + \|\mathbf{b}\|_{L^\beta(\Omega)}\|z\|_{L^{\frac{2\beta}{\beta-2}}(\Omega)}\|\nabla w\|_{L^2(\Omega)} + \|c\|_{L^\gamma(\Omega)}\|z\|_{L^{\frac{2\gamma}{\gamma-1}}(\Omega)}\|w\|_{L^{\frac{2\gamma}{\gamma-1}}(\Omega)}\Big\} \\ &\le C\left(\|\tilde{a}\|_{L^\infty(\Omega)} + \|\mathbf{b}(x)\|_{L^\beta(\Omega)} + \|c\|_{L^\gamma(\Omega)}\right)\|z\|_{H^1(\Omega)}\,,\end{aligned}$$

thus S is bounded. It remains to prove that S is bijective.

Injectivity of S. In order to prove that S is injective or equivalently that the equation $Sz = 0$ admits only the solution $z = 0$, we can follow the same steps of the proof of [**36**, Theorem 2.7] with obvious modifications.

Surjectivity of S. To verify the surjectivity of S we modify conveniently the arguments of [**36**, Theorem 2.7]. For every $t \in [0, 1]$ we define the linear operator $S_t : H^1(\Omega) \longrightarrow H^1(\Omega)^*$ by

$$\langle S_t z, w\rangle_{H^1(\Omega)^*,H^1(\Omega)} = \int_\Omega \{\tilde{a}(x)\nabla z\cdot\nabla w + tz\mathbf{b}(x)\cdot\nabla w + c(x)zw\}\, dx\,.$$

Obviously, $S_1 = S$ and the operator S_0 is monotone, hence S_0 is an isomorphism as follows from the Lax-Milgram theorem. Let D be the set of points $t \in [0, 1]$ for which S_t defines an isomorphism. Then $D \ne \emptyset$, since $0 \in D$. Let t_{max} be the supremum of D. We are going to prove that $S_{t_{\max}}$ is an isomorphism between $H^1(\Omega)$ and $H^1(\Omega)^*$, i.e. $t_{max} \in D$. We note that $S_{t_{\max}}$ is continuous and injective; the fact of including t_{max} in the equation does not matter for the proof of the continuity and injectivity. It remains to show that $S_{t_{\max}}$ is surjective. Given $\xi \in H^1(\Omega)^*$, we have to find an element $z \in H^1(\Omega)$ such that $S_{t_{\max}} z = \xi$. Let $\{t_k\}_{k=1}^{\infty} \subset D$ be a sequence such that $t_k \to t_{max}$ when $k \to \infty$ and denote by z_k the function in $H^1(\Omega)$ such that $S_{t_k} z_k = \xi$. Then, by the Poincaré inequality (1.3.3) and the assumptions on $\tilde{a}$ and c, we have

$$\begin{aligned}
\alpha \|z_k\|^2_{H^1(\Omega)} &\leq \int_\Omega \left\{ \tilde{a}(x) |\nabla z_k|^2 + c(x) z_k^2 \right\} dx \\
&= \left(\langle \xi, z_k \rangle_{H^1(\Omega)^*, H^1(\Omega)} - t_k \int_\Omega z_k \mathbf{b}(x) \cdot \nabla z_k \, dx \right) \\
&\leq C \left(\|\xi\|_{H^1(\Omega)^*} + t_k \|z_k \mathbf{b}\|_{L^2(\Omega)} \right) \|z_k\|_{H^1(\Omega)} \, , \\
&\leq C \left(\|\xi\|_{H^1(\Omega)^*} + t_k \|z_k\|_{L^{\frac{2\beta}{\beta-2}}(\Omega)} \|\mathbf{b}\|_{L^\beta(\Omega)} \right) \|z_k\|_{H^1(\Omega)}
\end{aligned}$$

which implies

$$\|z_k\|_{H^1(\Omega)} \leq C \left(\|\xi\|_{H^1(\Omega)^*} + \|z_k\|_{L^{\frac{2\beta}{\beta-2}}(\Omega)} \right) .$$

Arguing as in [**36**, Theorem 2.7], we get that $\{z_k\}_{k=1}^{\infty}$ is bounded in $H^1(\Omega)$ and the weak limit z of a subsequence of $\{z_k\}_{k=1}^{\infty}$ satisfies $S_{t_{max}} z = \xi$. Therefore, we conclude that $t_{\max} \in S$. Finally, we prove that $t_{\max} = 1$. If it is false then we consider the operators $S_{t_{\max}+\varepsilon}, S_{t_{\max}} \in \mathcal{L}(H^1(\Omega), H^1(\Omega)^*)$, $\forall \varepsilon > 0$ with $t_{\max} + \varepsilon \leq 1$ and obtain

$$\begin{aligned}
&\|S_{t_{\max}+\varepsilon} - S_{t_{\max}}\|_{\mathcal{L}(H^1(\Omega), H^1(\Omega)^*)} \\
&\qquad \leq \sup_{\|z\|_{H^1(\Omega)} \leq 1} \sup_{\|w\|_{H^1(\Omega)} \leq 1} \left| \int_\Omega \varepsilon z \mathbf{b}(x) \cdot \nabla w \, dx \right| \\
&\qquad \leq \varepsilon \sup_{\|z\|_{H^1(\Omega)} \leq 1} \sup_{\|w\|_{H^1(\Omega)} \leq 1} \|\mathbf{b}\|_{L^\beta(\Omega)} \|z\|_{L^{\frac{2\beta}{\beta-2}}(\Omega)} \|\nabla w\|_{L^2(\Omega)} \leq \tilde{C} \varepsilon \, .
\end{aligned}$$

Thus, taking $0 < \varepsilon < 1/(\tilde{C} \|S^{-1}_{t_{\max}}\|_{\mathcal{L}(H^1(\Omega)^*, H^1(\Omega))})$, we have by standard arguments of linear algebra that $S_{t_{max}+\varepsilon}$ is also an isomorphism, which contradicts the fact that t_{max} is the supremum of D. □

Before establishing the analysis of the linearized equation of (1.1.1), we need some differentiability of the coefficients a and f.

ASSUMPTION 1.24. *The functions a and f are of class C^2 with respect to the second variable and for any number $M > 0$ there exist constants $D_M, D_{M,a} > 0$, such that*

$$\begin{aligned}
(1.6.3) \qquad (1) \quad & \sum_{j=1}^{2} \left| \frac{\partial^j a}{\partial y^j}(x, y) \right| + \left| \frac{\partial^j f}{\partial y^j}(x, y) \right| \leq D_M \, , \\
(2) \quad & \left| \frac{\partial^k a}{\partial y^k}(x_1, y_1) - \frac{\partial^k a}{\partial y^k}(x_2, y_2) \right| \leq D_{M,a} \left(|x_1 - x_2| + |y_1 - y_2| \right) ,
\end{aligned}$$

for a.a. $x \in \Omega$, all $x_i \in \bar{\Omega}$ and $|y|, |y_i| \leq M$, $i = 1, 2$, $k = 1, 2$.

Taking into account the differentiability of f w.r.t. the second variable and Assumption 1.3, we find that $(\partial f/\partial y)(x,y) \geq 0$ for a.a. $x \in \Omega$ and all $y \in \mathbb{R}$. Moreover,

$$\frac{\partial f}{\partial y}(x,y) \geq \alpha_f > 0 \quad \forall (x,y) \in E \times \mathbb{R}\,, \tag{1.6.4}$$

where $E \subset \Omega$ is given in Assumption 1.3.

Let us now come back to our quasilinear equation (1.1.1). The linearized equation of (1.1.1) around a solution y has the following form

$$\begin{cases} -\operatorname{div}\left[a(x,y)\nabla z + \dfrac{\partial a}{\partial y}(x,y)z\nabla y\right] + \dfrac{\partial f}{\partial y}(x,y)z = \zeta & \text{in } \Omega\,, \\ \left[a(x,y)\nabla z + \dfrac{\partial a}{\partial y}(x,y)z\nabla y\right]\cdot\nu = v & \text{on } \Gamma\,. \end{cases} \tag{1.6.5}$$

By setting

$$\tilde{a}(x) = a(x,y(x))\,,\ \mathbf{b}(x) = \frac{\partial a}{\partial y}(x,y(x))\nabla y(x) \ \text{ and } \ c(x) = \frac{\partial f}{\partial y}(x,y(x))\,, \tag{1.6.6}$$

we see that (1.6.5) is contained in the class of equations of the type (1.6.1). In analogy to Theorem 1.23, we may state the following existence and uniqueness theorem for a solution of (1.6.5).

Theorem 1.25. *Suppose that the Assumptions 1.1-1.3 and 1.24-*(1) *hold. Given $y \in W^{1,r}(\Omega)$ with $r > 2$, for any $v \in H^{-1/2}(\Gamma)$ and $\zeta \in H^1(\Omega)^*$, the linearized equation (1.6.5) has a unique solution $z \in H^1(\Omega)$ in the sense that*

$$\begin{aligned} \int_\Omega \left\{ a(x,y)\nabla z\cdot\nabla w + \frac{\partial a}{\partial y}(x,y)z\nabla y\cdot\nabla w + \frac{\partial f}{\partial y}(x,y)zw \right\} dx \\ = \langle \zeta, w\rangle_{H^1(\Omega)^*,H^1(\Omega)} + \langle v, w\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)} \quad \forall w \in H^1(\Omega)\,. \end{aligned} \tag{1.6.7}$$

Moreover, there holds the inequality

$$\|z\|_{H^1(\Omega)} \leq C\left(\|\zeta\|_{H^1(\Omega)^*} + \|v\|_{H^{-1/2}(\Gamma)}\right),$$

with some positive constant C which may depend on a, f and y but not on ζ and v.

Proof. In order to deduce the existence and uniqueness of a solution $z \in H^1(\Omega)$ of (1.6.5), we will apply Theorem 1.23. Therefore, it is enough to check if the assumptions of Theorem 1.23 are satisfied. For a fixed $y \in W^{1,r}(\Omega)$, let us set the coefficients $\tilde{a}$, $\mathbf{b}$ and c, as in (1.6.6). Since $W^{1,r}(\Omega) \subset C(\bar{\Omega})$ is valid for $r > 2$, y is uniformly bounded and we have, together with Assumption 1.2 and (1.6.3), that $a(\cdot,y(\cdot))$, $(\partial a/\partial y)(\cdot,y(\cdot)) \in L^\infty(\Omega)$ and $(\partial f/\partial y)(\cdot,y(\cdot)) \in L^\infty(\Omega)$, hence $\tilde{a} \in L^\infty(\Omega)$, $\mathbf{b} \in L^r(\Omega)^2$ and $c \in L^\infty(\Omega)$. Moreover, by defining $\alpha = \min\{\alpha_a, \alpha_f\}$, it follows from (1.2.1) and (1.6.4) that $\tilde{a}(\cdot) \geq \alpha$ and $c(\cdot) \geq 0$ a.e. in Ω as well as $c(\cdot) \geq \alpha$ in E. Now Theorem 1.23 implies that the left-hand side of (1.6.7) defines an isomorphism between $H^1(\Omega)$ and $H^1(\Omega)^*$. Given $\xi \in H^1(\Omega)^*$, this is equivalent to the existence and uniqueness of a solution $z \in H^1(\Omega)$ to the equation

$$\int_\Omega \left\{ a(x,y)\nabla z\cdot\nabla w + \frac{\partial a}{\partial y}(x,y) z\nabla y\cdot\nabla w + \frac{\partial f}{\partial y}(x,y)zw \right\} dx = \langle \xi, w\rangle_{H^1(\Omega)^*,H^1(\Omega)}$$

$\forall w \in H^1(\Omega)$. The proof of the well-posedness of (1.6.7) in $H^1(\Omega)$ is then complete when taking $\xi = \zeta + Bv$, where the operator $B : H^{-1/2}(\Gamma) \longrightarrow H^1(\Omega)^*$ is defined by

$$\langle Bv, w\rangle_{H^1(\Omega)^*,H^1(\Omega)} = \langle v, w|_\Gamma\rangle_{H^{-1/2}(\Gamma),H^{1/2}(\Gamma)}\ . \tag{1.6.8}$$

The inequality in the assertion of the theorem is an immediate consequence of the linearity and continuity of the solution operator $(\zeta, v) \longmapsto z$ associated with (1.6.5). □

Remark 1.26. *The proof of Theorem 1.25 can be modified in an obvious way to verify that, for any given function $y \in W^{1,r}(\Omega)$, with $r > 2$, and $y_i \in L^\infty(\Omega)$, $i = 1, 2, 3$, the equation*

$$\begin{cases} -\operatorname{div}\left[a(x,y_1)\nabla z(x) + \dfrac{\partial a}{\partial y}(x,y_2)z\nabla y\right] + \dfrac{\partial f}{\partial y}(x,y_3)z = \zeta & in\,\Omega\,, \\ \left[a(x,y_1)\nabla z(x) + \dfrac{\partial a}{\partial y}(x,y_2)z\nabla y\right]\cdot\nu = v & on\,\Gamma\,, \end{cases}$$

has a unique solution $z \in H^1(\Omega)$.

In the next step, we will prove that the solution z of (1.6.5) is continuous if v and ζ are regular functions. For this purpose, we state in the following lemma a useful result based on the well-known Stampacchia truncation method, cf. Stampacchia [**88**] or Murthy and Stampacchia [**81**]. Further, this continuity will allow us to deduce higher regularity of z. This will be shown in Theorem 1.29 below.

Lemma 1.27. *(*[**3**, Theorem 2]*, see also* [**81**, Theorem 2.9]*) Suppose that Assumption 1.1 holds and that $\tilde a \in L^\infty(\Omega)$ with $\tilde a(x) \geq \alpha > 0$ for a.a. $x \in \Omega$, $c \in L^\gamma(\Omega)$ with $\gamma > 1$, satisfying $c(x) \geq 0$ for a.a. $x \in \Omega$ and $c(x) \geq \alpha$ for all $x \in E$, where E is a measurable subset of Ω with $|E| > 0$. We also assume that $\zeta \in L^\gamma(\Omega)$, $v \in L^\theta(\Gamma)$ and $f_i \in L^\beta(\Omega)$, $i = 1, 2$, with $\theta > 1$ and $\beta > 2$. Then the unique solution $z \in H^1(\Omega)$ of the equation*

$$\int_\Omega \{\tilde a(x)\nabla z\cdot\nabla w + c(x)zw\}\,dx = \int_\Omega \left\{\zeta w + \sum_{i=1}^{2} f_i\partial_i w\right\} dx + \int_\Gamma vw\,d\sigma \quad \forall w \in H^1(\Omega)$$

is continuous in $\bar\Omega$.

Proposition 1.28. *Under the Assumptions 1.1-1.3 and 1.24-(1) and supposing that $\zeta \in L^\gamma(\Omega)$, $v \in L^\theta(\Gamma)$ and $y \in W^{1,r}(\Omega)$ with $\gamma, \theta > 1$, and $r > 2$, the unique solution z of the equation (1.6.5) is continuous in $\bar\Omega$.*

Proof. Obviously, the solution $z \in H^1(\Omega)$ of (1.6.5) satisfies the equation

$$\int_\Omega \left\{ a(x,y)\nabla z\cdot\nabla w + \frac{\partial f}{\partial y}(x,y)zw \right\} dx$$
$$= \int_\Omega \left\{ \zeta w - \sum_{i=1}^{2} \int_\Omega \frac{\partial a}{\partial y}(x,y) z \partial_i y \partial_i w \right\} dx + \int_\Gamma vw \, d\sigma(x) \quad \forall w \in H^1(\Omega)\,.$$

Since $H^1(\Omega) \subset L^q(\Omega)$ for all $1 \le q < \infty$, the functions $(\partial a/\partial y)(\cdot, y(\cdot))z\partial_i y$, $i = 1, 2$, belong to $L^\beta(\Omega)$ with some $\beta > 2$. Indeed, $(\partial a/\partial y)(\cdot, y(\cdot)) \in L^\infty(\Omega)$, $\partial_i y \in L^r(\Omega)$ and there exists $q < \infty$ such that $\frac{1}{2} > \frac{1}{r} + \frac{1}{q}$. Hence, setting $\frac{1}{\beta} := \frac{1}{r} + \frac{1}{q}$, we have $\beta > 2$ and $z\partial_i y \in L^\beta(\Omega)$. The continuity of z follows then immediately from Lemma 1.27. □

The next theorem states the $W^{1,r}(\Omega)$ regularity of the solution to the linearized equation (1.6.5). As we have seen in Theorem 1.9, in presence of non-zero Neumann boundary data, the proof of the $W^{1,r}(\Omega)$ regularity result requires a quite different approach compared to the case of homogeneous Dirichlet boundary conditions; see Casas and Tröltzsch [**36**, Theorem 2.4].

Theorem 1.29. *Suppose that the Assumptions 1.1-1.3 and 1.24-(1) hold. Assume further that $a : \bar\Omega \times \mathbb{R} \to \mathbb{R}$ is continuous. For any $\zeta \in L^{\frac{2r}{r+2}}(\Omega)$, $v \in L^{r/2}(\Gamma)$ and $y \in W^{1,r}(\Omega)$ with $r > 2$ satisfying the condition (1.3.13), the solution z of (1.6.5) belongs to $W^{1,r}(\Omega)$.*

Proof. Since $v \in L^{r/2}(\Gamma) \subset H^{-1/2}(\Gamma)$ and $\zeta \in L^{\frac{2r}{r+2}}(\Omega) \subset H^1(\Omega)^*$, the existence and uniqueness of $z \in H^1(\Omega)$ follow from Theorem 1.25. Moreover, by applying Proposition 1.28, we deduce that $z \in C(\bar\Omega)$. To obtain the $W^{1,r}(\Omega)$ regularity of z we follow the same steps as in Theorem 1.9 with obvious modifications. We only remark that equation (1.3.16) on page 10 has to be replaced by

$$\int_\Omega \{\tilde a(x)\nabla z_j\cdot\nabla w + z_j w\}\, dx = F(w)\,,$$

for arbitrary $w \in H^1(\Omega)$, $z_j := z\psi_j$, $\tilde a(\cdot) := a(\cdot, y(\cdot))$ and

$$F(w) := \int_\Omega \left\{ \tilde a(x)\,(z\nabla w - w\nabla z)\cdot\nabla\psi_j - \frac{\partial a}{\partial y}(x,y) z \nabla y\cdot\nabla(\psi_j w) \right.$$
$$\left. + \zeta\psi_j w + \left(1 - \frac{\partial f}{\partial y}(x,y)\right) z\psi_j w \right\} dx + \int_\Gamma v\psi_j w \, d\sigma(x)\,.$$

By the same arguments as in the proof of Theorem 1.9 and thanks to the continuity of z and the embedding $W^{1,r}(\Omega) \hookrightarrow C(\bar\Omega)$, it can be easily checked that $F \in W^{1,r'}(\Omega)^*$. □

The next theorem states the differentiability of the relation between the boundary datum u and the associated solution y_u of (1.1.1).

THEOREM 1.30. *Suppose that the Assumptions 1.1-1.3 and 1.24-(1) are fulfilled and* $a : \bar{\Omega} \times \mathbb{R} \longrightarrow \mathbb{R}$ *is continuous. The mapping* $G : L^{r/2}(\Gamma) \longrightarrow W^{1,r}(\Omega)$, $G(u) = y_u$, *is twice continuously Fréchet-differentiable, i.e. of class* C^2, *for all* r *satisfying the condition (1.3.13). Moreover, for any* $v, v_1, v_2 \in L^{r/2}(\Gamma)$, *the functions* $z_v = G'(u)v$ *and* $z_{v_1,v_2} = G''(u)[v_1, v_2]$ *are the unique solutions in* $W^{1,r}(\Omega)$ *of the equations*

$$\begin{cases} -\operatorname{div}\left[a(x,y_u)\nabla z + \dfrac{\partial a}{\partial y}(x,y_u)z\nabla y_u\right] + \dfrac{\partial f}{\partial y}(x,y_u)z = 0 & \text{in } \Omega\,, \\ \left[a(x,y_u)\nabla z + \dfrac{\partial a}{\partial y}(x,y_u)z\nabla y_u\right]\cdot\nu = v & \text{on } \Gamma\,, \end{cases} \tag{1.6.9}$$

and

$$\begin{cases} -\operatorname{div}\left[a(x,y_u)\nabla z + \dfrac{\partial a}{\partial y}(x,y_u)z\nabla y_u\right] + \dfrac{\partial f}{\partial y}(x,y_u)z = -\dfrac{\partial^2 f}{\partial y^2}(x,y_u)z_1z_2 \\ \qquad + \operatorname{div}\left[\dfrac{\partial a}{\partial y}(x,y_u)\left(z_1\nabla z_2 + z_2\nabla z_1\right) + \dfrac{\partial^2 a}{\partial y^2}(x,y_u)z_1z_2\nabla y_u\right] \quad \text{in } \Omega\,, \\ \left[a(x,y_u)\nabla z + \dfrac{\partial a}{\partial y}(x,y_u)z\nabla y_u\right]\cdot\nu = \\ \qquad = -\left[\dfrac{\partial a}{\partial y}(x,y_u)\left(z_1\nabla z_2 + z_2\nabla z_1\right) + \dfrac{\partial^2 a}{\partial y^2}(x,y_u)z_1z_2\nabla y_u\right]\cdot\nu \quad \text{on } \Gamma\,, \end{cases} \tag{1.6.10}$$

respectively, where $z_i = G'(u)v_i$, $i = 1, 2$.

PROOF. Let us introduce the mapping $F : W^{1,r}(\Omega) \times L^{r/2}(\Gamma) \longrightarrow W^{1,r'}(\Omega)^*$,

$$\langle F(y,u), w\rangle = \int_\Omega \{a(x,y)\nabla y\cdot\nabla w + f(x,y)w\}\,dx - \int_\Gamma uw\,d\sigma(x)\,.$$

As a consequence of the Assumptions 1.2, 1.3 and 1.24-(1), and the embedding $W^{1-1/r',r'}(\Gamma) \hookrightarrow L^{\frac{r}{r-2}}(\Gamma) = L^{(r/2)'}(\Gamma)$, F is well defined, of class C^2 and $F(y_u, u) = 0$ for every $u \in L^{r/2}(\Gamma)$. Our goal is to prove that $\partial_y F(y_u, u) : W^{1,r}(\Omega) \longrightarrow W^{1,r'}(\Omega)^*$ defined by

$$\langle \partial_y F(y_u,u)z, w\rangle = \int_\Omega \left\{a(x,y_u)\nabla z\cdot\nabla w + \frac{\partial a}{\partial y}(x,y_u)z\nabla y_u\cdot\nabla w + \frac{\partial f}{\partial y}(x,y_u)zw\right\}dx$$

is an isomorphism. Then we can apply the implicit function theorem (see Cartan [**12**]) to deduce the differentiability properties of G stated in the theorem. The representations (1.6.9) and (1.6.10) for G' and G'' are then obtained by simple computations.

According to Theorem 1.25, for any $v \in H^{-1/2}(\Gamma)$ there exists a unique element $z_v \in H^1(\Omega)$ such that

$$\partial_y F(y_u, u) z_v = Bv\,,$$

where the operator $B : H^{-1/2}(\Gamma) \longrightarrow H^1(\Omega)^*$ is defined in (1.6.8). It suffices to show that $z_v \in W^{1,r}(\Omega)$ if $v \in L^{r/2}(\Gamma)$. More precisely, this means that the unique solution $z_v \in H^1(\Omega)$ of (1.6.5), with y replaced by y_u, associated with $v \in L^{r/2}(\Gamma)$ and $\zeta = 0$, belongs to $W^{1,r}(\Omega)$. But this regularity follows directly from Theorem 1.29.

The existence, uniqueness and $W^{1,r}(\Omega)$ regularity of the solution of (1.6.10) can be deduced analogously. □

If we assume that a is locally Lipschitz continuous w.r.t. both variables then we know from Theorem 1.18 that the solution y_u of (1.1.1), associated with the boundary datum $u \in L^2(\Gamma)$, belongs to $H^{3/2}(\Omega)$. The next theorem shows that this regularity also holds for the solution of (1.6.5) provided that $v \in L^2(\Gamma)$ and $\zeta \in L^p(\Omega)$, $p > 4/3$.

Theorem 1.31. *Suppose that the Assumptions 1.1-1.3, 1.17 and 1.24 hold. We also assume that $v, u \in L^2(\Gamma)$, $\zeta \in L^p(\Omega)$ ($p > 4/3$) and $y = y_u$ is the solution of (1.1.1). Then the solution z of (1.6.5) belongs to $H^{3/2}(\Omega)$ and there holds the estimate*

$$\|z\|_{H^{3/2}(\Omega)} \le C(\|v\|_{L^2(\Gamma)} + \|\zeta\|_{L^p(\Omega)})\,, \tag{1.6.11}$$

with some $C > 0$ which depends on $\|y_u\|_{H^{3/2}(\Omega)}$ but not on v or ζ. Moreover, the mapping $G : L^2(\Gamma) \longrightarrow H^{3/2}(\Omega)$, $G(u) = y_u$, is of class C^2.

Proof. According to Theorem 1.18, $y_u \in H^{3/2}(\Omega) \subset W^{1,4}(\Omega)$, hence from Theorem 1.29 it follows $z \in W^{1,\min\{4,\bar{r}\}}(\Omega)$; notice that $\frac{2r}{r+2} = p > \frac{4}{3}$ implies that $r > 4$. Let us show that $z \in H^{3/2}(\Omega)$. Since the function a is locally Lipschitz w.r.t. both variables and strictly positive, we may expand the divergence in the first equality of (1.6.5) and divide by a to deduce

$$\begin{aligned}
-\Delta z + \frac{1}{a}\frac{\partial f}{\partial y} z &= \frac{1}{a}\left(\zeta + \sum_{j=1}^{2} (\partial_j a \partial_j z) + \frac{\partial a}{\partial y}\nabla z \cdot \nabla y_u + \operatorname{div}\left[\frac{\partial a}{\partial y} z \nabla y_u\right]\right)\\
&= \frac{1}{a}\left(\zeta + \sum_{j=1}^{2}\left(\partial_j a \partial_j z + \frac{\partial^2 a}{\partial_j \partial y} z \partial_j y_u\right) + 2\frac{\partial a}{\partial y}\nabla z \cdot \nabla y_u\right.\\
&\qquad \left. + \frac{\partial^2 a}{\partial y^2} z |\nabla y_u|^2 + \frac{\partial a}{\partial y} z \Delta y_u\right).
\end{aligned} \tag{1.6.12}$$

Analogously, from (1.1.1) we have

$$\Delta y_u = \frac{1}{a}\left(-f + \sum_{j=1}^{2} \partial_j a \partial_j y_u + \frac{\partial a}{\partial y}\left|\nabla y_u\right|^2\right) \in L^{\min\{p,2\}}(\Omega)\,.$$

By combining this with the $W^{1,\min\{4,\bar{r}\}}(\Omega)$ regularity of z, we find that the right-hand side of (1.6.12) is in $L^q(\Omega)$ with some $q > 4/3$. On the other hand, taking into account that $\partial_\nu y_u = (u/a) \in L^2(\Gamma)$ and $z|_\Gamma \in L^\infty(\Gamma)$, we get from (1.6.5) that $a(\cdot, y_u)\partial_\nu z = v - (\partial a/\partial y)(\cdot, y_u) z \partial_\nu y_u \in L^2(\Gamma)$. Finally, the $H^{3/2}(\Omega)$ regularity of z is a consequence of Corollary 1.15.

Now let us prove (1.6.11). By virtue of Corollary 1.15 and (1.6.12), we easily obtain

$$\|z\|_{H^{3/2}(\Omega)} \leq C \left(\|v\|_{L^2(\Gamma)} + \|\zeta\|_{L^p(\Omega)} + \|z\|_{W^{1,r}(\Omega)} \right),$$

where $r := \min\{4, \bar{r}\}$. It remains to estimate $\|z\|_{W^{1,r}(\Omega)}$. For this purpose, we make use of the linear operator $\partial_y F(y_u, u) : W^{1,r}(\Omega) \longrightarrow W^{1,r'}(\Omega)^*$ defined in the proof of Theorem 1.30 and $B : L^2(\Gamma) \longrightarrow W^{1,r'}(\Omega)^*$ given by

$$\langle Bv, w \rangle = \int_\Gamma vw \, d\sigma(x) \, .$$

Setting $\xi = \zeta + Bv$, we have the identity

$$\partial_y F(y_u, u) z = \xi \, .$$

Since $\partial_y F(y_u, u)$ is an isomorphism, it follows that its inverse is an isomorphism, too. This fact yields the estimate

$$\|z\|_{W^{1,r}(\Omega)} \leq C \|\xi\|_{W^{1,r'}(\Omega)^*} \leq C \left(\|\zeta\|_{L^p(\Omega)} + \|v\|_{L^2(\Gamma)} \right).$$

The differentiability of the mapping G as stated in the Theorem is then a consequence of Theorem 1.30 and the preceding regularity result. □

From the mean value theorem and the differentiability of the mapping G from $L^{r/2}(\Gamma)$ to $W^{1,r}(\Omega)$ (Theorem 1.30) or from $L^2(\Gamma)$ to $H^{3/2}(\Omega) \subset W^{1,4}(\Omega)$ (Theorem 1.31) we obtain

Corollary 1.32. *For any bounded set $U \subset L^{r/2}(\Gamma)$ with r satisfying the condition (1.3.13) or $r = 4$ there exists a constant $C_U > 0$ such that*

$$\|y_{u_2} - y_{u_1}\|_{W^{1,r}(\Omega)} \leq C_U \|u_2 - u_1\|_{L^{r/2}(\Gamma)} \quad \textit{for any } \; u_i \in U, i = 1, 2 \, .$$

If Ω is convex then from Theorem 1.21 we know that the solution y_u corresponding to $u \in W^{1-1/r,r}(\Gamma)$, with $2 \leq r \leq \min\{\bar{r}, r_0, p\}$ for some $r_0 > 2$ and $\bar{r} \geq \frac{6}{3-\sqrt{5}}$, belongs to $W^{2,r}(\Omega)$. In this context, a natural question arises: Can we prove a result analogous to Theorem 1.30 with $G : W^{1-1/r,r}(\Gamma) \longrightarrow W^{2,r}(\Omega)$? The next theorem gives a positive answer to it.

Theorem 1.33. *Let the Assumptions 1.1-1.3, 1.17 and 1.24 be satisfied and assume that Ω is convex and $p \geq 2$. There exists $r_0 > 2$ depending on the measure of the angles in Γ such that the mapping $G : W^{1-1/r,r}(\Gamma) \longrightarrow W^{2,r}(\Omega)$, $G(u) = y_u$, is of class C^2 for $2 \leq r \leq \min\{r_0, \bar{r}, p\}$ ($\bar{r} \geq \frac{6}{3-\sqrt{5}}$; see Theorem 1.9). Moreover, for any $v, v_1, v_2 \in W^{1-1/r,r}(\Gamma)$, the functions $z_v := G'(u)v$ and $z_{v_1,v_2} := G''(u)[v_1, v_2]$ are the unique solutions in $W^{2,r}(\Omega)$ of the equations (1.6.9) and (1.6.10), respectively.*

Proof. Consider the Banach space

$$V(\Omega) = \{y \in W^{2,r}(\Omega) \,|\, \partial_\nu y \in W^{1-1/r,r}(\Gamma)\}$$

endowed with the graph norm $\|y\|_{V(\Omega)} = \|y\|_{W^{2,r}(\Omega)} + \|\partial_\nu y\|_{W^{1-1/r,r}(\Gamma)}$. Then all the elements $y_u = G(u)$ belong to this space provided that $u \in W^{1-1/r,r}(\Gamma)$; see Theorem 1.21. Let us introduce the mapping

$$F : V(\Omega) \times W^{1-1/r,r}(\Gamma) \longrightarrow L^r(\Omega) \times W^{1-1/r,r}(\Gamma)\,,$$
$$F(y,u) = (-\mathrm{div}[a(\cdot,y)\nabla y] + f(\cdot,y), a(\cdot,y)\partial_\nu y - u)\,.$$

Next we verify that F is well defined. By expanding the divergence term, we find

$$\mathrm{div}\,[a(\cdot,y)\nabla y] = [\nabla_x a]\,(\cdot,y)\cdot\nabla y + \frac{\partial a}{\partial y}(\cdot,y)\,|\nabla y|^2 + a(\cdot,y)\Delta y\,.$$

Since $y \in W^{2,r}(\Omega)$ with $r \geq 2$, the right-hand side of the previous equality is in $L^r(\Omega)$, therefore it remains to show that $a(\cdot,y)\partial_\nu y \in W^{1-1/r,r}(\Gamma)$ to deduce that F is well defined. To this end, we use the Lipschitz property of a with respect to x and y as well as the embedding $H^2(\Omega) \hookrightarrow C^{0,\mu}(\bar{\Omega})$ for every $\mu \in (0,1)$; see Eq. (1.5.7). In particular, $a(\cdot,y) \in C^{0,\mu}(\Gamma)$ for all $\mu \in (1-\frac{1}{r},1)$, which along with (1.5.6), yields $a(\cdot,y)\partial_\nu y \in W^{1-1/r,r}(\Gamma)$.

On the other hand, it is obvious that F is a C^2 mapping. Next we are going to apply the implicit function theorem. Thus, we need to prove that the linear operator $\partial_y F(y,u) : V(\Omega) \longrightarrow L^r(\Omega) \times W^{1-1/r,r}(\Gamma)$,

$$\partial_y F(y,u)z =$$
$$= \left(-\mathrm{div}\left[a(\cdot,y)\nabla z + \frac{\partial a}{\partial y}(x,y)z\nabla y\right] + \frac{\partial f}{\partial y}(\cdot,y)z, \left[a(\cdot,y)\nabla z + \frac{\partial a}{\partial y}(\cdot,y)z\nabla y\right]\cdot\nu\right),$$

is an isomorphism. To this aim, we have to show for any $\zeta \in L^r(\Omega)$ and any $v \in W^{1-1/r,r}(\Gamma)$ the existence of a unique solution $z \in V(\Omega)$ of (1.6.5). The existence and uniqueness of a solution z in $H^1(\Omega) \cap C(\bar{\Omega})$ follows from Theorem 1.25 and Proposition 1.28. Further, due to the inclusion $W^{1-1/r,r}(\Gamma) \subset L^q(\Gamma)$ for every $q < \infty$, the $W^{1,\bar{r}}(\Omega)$ regularity of z is deduced from Theorem 1.29. Finally, we prove that $z \in W^{2,r}(\Omega)$ and $\partial_\nu z \in W^{1-1/r,r}(\Gamma)$ by transforming (1.6.5) to

$$-\Delta z = \frac{1}{a}\left(\zeta + \mathrm{div}\left[\frac{\partial a}{\partial y}(\cdot,y)z\nabla y\right] + [\nabla_x a](\cdot,y)\cdot\nabla z + \frac{\partial a}{\partial y}(\cdot,y)\nabla z\cdot\nabla y - \frac{\partial f}{\partial y}(\cdot,y)z\right)$$
$$= \frac{1}{a}\left(\zeta + z\left[\nabla_x\frac{\partial a}{\partial y}\right](\cdot,y)\cdot\nabla y + \frac{\partial^2 a}{\partial y^2}(\cdot,y)z\,|\nabla y|^2 + 2\frac{\partial a}{\partial y}(\cdot,y)\nabla z\cdot\nabla y\right.$$
$$\left. + \frac{\partial a}{\partial y}(\cdot,y)z\Delta y + [\nabla_x a]\,(\cdot,y)\cdot\nabla z - \frac{\partial f}{\partial y}(\cdot,y)z\right) \quad \text{in } \Omega\,, \tag{1.6.13}$$

$$\partial_\nu z = \frac{1}{a}\left(v - \frac{\partial a}{\partial y}(\cdot,y)z\partial_\nu y\right) \quad \text{on } \Gamma\,. \tag{1.6.14}$$

In the sequel, we will follow the steps of the proof of Theorem 1.21 and consider first the case when $r \le \bar{r}/2$.

The right-hand side of (1.6.13) belongs to $L^{\min\{p,r\}}(\Omega)$. Let us show that the right-hand side of (1.6.14) is in $W^{1-1/r,r}(\Gamma)$. Once this is proved, the well-known regularity result by Grisvard [**60**, Corollary 4.4.3.8] yields the existence of $r_0 > 2$ introduced in Theorem 1.21 such that $z \in W^{2,r}(\Omega)$ for $2 \le r \le \min\{r_0, \bar{r}/2, p\}$. In particular, the proof would be complete if $p = 2$.

Thanks to the inclusion (1.5.7), we have $z \in W^{1,\bar{r}}(\Omega) \subset C^{0,1-2/\bar{r}}(\bar{\Omega})$ as well as $y \in H^2(\Omega) \subset C^{0,\mu}(\bar{\Omega})$ for every $\mu \in [1-2/\bar{r}, 1)$. Together with Assumption 1.24-(2) and (1.2.1) we find that the functions $\frac{1}{a(\cdot,y(\cdot))}$ and $\frac{1}{a(\cdot,y(\cdot))}\frac{\partial a}{\partial y}(\cdot,y(\cdot))z$ are in $C^{0,1-2/\bar{r}}(\bar{\Omega})$. Now if $r < \bar{r}/2$ we can apply (1.5.6) to deduce

$$\frac{1}{a}\left(v - \frac{\partial a}{\partial y}(\cdot,y) z \partial_\nu y\right) \in W^{1-1/r,r}(\Gamma)\,. \tag{1.6.15}$$

In particular, $z \in H^2(\Omega)$ if $u, v \in H^{1/2}(\Gamma)$. If $r = \bar{r}/2$ we will improve the regularity of z to $W^{2,r}(\Omega)$. Since $z \in H^2(\Omega)$ and taking into account (1.5.7), we obtain $\frac{1}{a(\cdot,y(\cdot))}$, $\frac{1}{a(\cdot,y(\cdot))}\frac{\partial a}{\partial y}(\cdot,y(\cdot))z \in C^{0,\mu}(\bar{\Omega})$ with some $\mu > 1 - 2/\bar{r}$, hence once again (1.5.6) yields (1.6.15).

Finally, let us discuss the case when $\bar{r}/2 < r \le \min\{r_0, \bar{r}, p\}$ which implies that $p > 2$. From the above arguments we know that $y_u \in W^{2,\min\{r_0,\bar{r}/2,p\}}(\Omega) \subset C^{0,1}(\bar{\Omega})$. Therefore, the right-hand sides of (1.6.13) and (1.6.14) are in $L^r(\Omega)$ and $W^{1-1/r,r}(\Gamma)$, respectively. Consequently, as above, we conclude that $z \in W^{2,r}(\Omega)$. □

REMARK 1.34. *In bounded polygonal domains, the space $V(\Omega)$ defined in the proof of the previous theorem is in general a proper subspace of $W^{2,r}(\Omega)$. In particular, the normal derivative of a function in $H^2(\Omega)$ is not necessarily an element of $H^{1/2}(\Gamma)$ but it belongs only to $H^{1/2-\varepsilon}(\Gamma)$ for all $\varepsilon > 0$, cf. Grisvard* [**60**, § 1.5.3]. *A simple counterexample is $y(x) = |x|^2$ on the unit square $\Omega = (0,1)^2$, cf. Casas et al.* [**28**, p. 800]. *This is the reason why we have introduced the Banach space $V(\Omega)$. If Ω is of class $C^{1,1}$ we do not have this difficulty, cf.* [**60**, Theorem 1.5.2.1], *hence the functional F given on page 28 can be defined in $W^{2,r}(\Omega) \times W^{1-1/r,r}(\Gamma)$.*

1.7. The adjoint problem

In the previous section, we analyzed the differentiability of the mapping $u \longmapsto y_u$. As pointed out at the beginning of this chapter, there are several reasons which motivate the study of the adjoint operator associated to the one defined in (1.6.2).

We will proceed as in the previous section, by stating first a general result on the well-posedness of the adjoint problem. Further, we will turn to our specific problem (1.1.1) and consider the corresponding adjoint equation.

Theorem 1.35. *Supposing that the assumptions of Theorem 1.23 hold true, the operator $T : H^1(\Omega) \longrightarrow H^1(\Omega)^*$ defined by*

$$\langle T\varphi, z\rangle_{H^1(\Omega)^*,H^1(\Omega)} = \int_\Omega \{\tilde{a}(x)\nabla\varphi\cdot\nabla z + z\mathbf{b}(x)\cdot\nabla\varphi + c(x)\varphi z\}\, dx \tag{1.7.1}$$

is an isomorphism. Moreover, the following regularity properties hold:

(1) *Assume that $\tilde{a} \in C(\bar{\Omega})$. For any $2 < r \le \bar{r}$ with $\bar{r} > 3$ given in Theorem 1.9, $\mathbf{b} \in L^r(\Omega)^2$ and $c, \zeta \in L^{\frac{2r}{r+2}}(\Omega)$, there exists a unique element $\varphi \in W^{1,r}(\Omega)$ satisfying*

$$\langle T\varphi, z\rangle_{H^1(\Omega)^*,H^1(\Omega)} = \int_\Omega \zeta z\, dx \quad \forall z \in H^1(\Omega)\,. \tag{1.7.2}$$

(2) *If $\tilde{a} \in W^{1,r}(\Omega)$, $\mathbf{b} \in L^r(\Omega)^2$ and $c, \zeta \in L^{r/2}(\Omega)$ with some $r > 8/3$, then the solution φ of (1.7.2) belongs to $H^{3/2}(\Omega)$ and there exists a constant $C' > 0$, dependent on $\tilde{a}$, $\mathbf{b}$ and c, but not on ζ, such that*

$$\|\varphi\|_{H^{3/2}(\Omega)} \le C'\|\zeta\|_{L^{r/2}(\Omega)}\,. \tag{1.7.3}$$

(3) *If Ω is convex, $\tilde{a} \in W^{1,4}(\Omega)$, $\mathbf{b} \in L^4(\Omega)^2$ and $c, \zeta \in L^2(\Omega)$, then the solution φ of (1.7.2) is in $H^2(\Omega)$ and there exists $C'' > 0$ independent of ζ such that*

$$\|\varphi\|_{H^2(\Omega)} \le C''\|\zeta\|_{L^2(\Omega)}\,. \tag{1.7.4}$$

Proof. Since T is the adjoint of the operator S defined in (1.6.2) and S is an isomorphism, this property holds also true for T.

Proof of (1). The existence and uniqueness of a solution $\varphi \in H^1(\Omega)$ of (1.7.2) is an immediate consequence of the fact that T is an isomorphism. Moreover, we know that the inverse of an isomorphism is also an isomorphism, hence

$$\|\varphi\|_{H^1(\Omega)} \le C_1\|\zeta\|_{L^{r/2}(\Omega)} \tag{1.7.5}$$

with some $C_1 > 0$ independent of ζ. Rewriting (1.7.2) in the form

$$\int_\Omega \{\tilde{a}(x)\nabla\varphi\cdot\nabla z + c(x)\varphi z\}\, dx = \int_\Omega (\zeta - \mathbf{b}(x)\cdot\nabla\varphi)\, z\, dx \quad \forall z \in H^1(\Omega)\,,$$

we see that Lemma 1.27 is applicable here, yielding $\varphi \in C(\bar{\Omega})$. In order to prove the $W^{1,r}(\Omega)$ regularity of φ, we follow the steps of the proof of Theorem 1.9. Let us point out the main differences. This time, we consider the equation

$$\int_\Omega \{\tilde{a}(x)\nabla\varphi_j\cdot\nabla z + \varphi_j z\}\, dx = F(z)$$

for arbitrary $z \in H^1(\Omega)$, $\varphi_j := \varphi\psi_j$ and

$$F(z) := \int_\Omega \{\tilde{a}(x)\,(\varphi\nabla z - z\nabla\varphi)\cdot\nabla\psi_j + (\zeta + \varphi - \mathbf{b}(x)\cdot\nabla\varphi - c(x)\varphi)\,\psi_j z\}\, dx\,. \tag{1.7.6}$$

By the same arguments as in the proof of Theorem 1.9, along with the continuity of φ, one can check that $F \in W^{1,r'}(\Omega)^*$. To verify this, we consider exemplarily only the term $(\mathbf{b}\cdot\nabla\varphi)z$:

$$\|(\mathbf{b}\cdot\nabla\varphi)z\|_{L^1(\Omega)} \le C\|\mathbf{b}\cdot\nabla\varphi\|_{L^{\frac{2r}{r+2}}(\Omega)}\|z\|_{L^{\frac{2r}{r-2}}(\Omega)} \le C\|\mathbf{b}\|_{L^r(\Omega)}\|\nabla\varphi\|_{L^2(\Omega)}\|z\|_{W^{1,r'}(\Omega)}\,.$$

Proof of (2). Repeating the arguments of the first part of the theorem, there holds $F \in W^{1,s'}(\Omega)^*$ with $s = \min\{r, \bar{r}\} > 8/3$, consequently $\varphi \in W^{1,s}(\Omega)$. Now using the assumptions on $\tilde{a}$ and $\frac{s}{2} > \frac{1}{2}\frac{8}{3} = \frac{4}{3}$ as well as $\nabla\tilde{a} - \mathbf{b} \in L^r(\Omega)^2 \subset L^s(\Omega)^2$ and $\nabla\varphi \in L^s(\Omega)^2$, we can pass to the Neumann formulation

$$\text{(1.7.7)} \qquad \begin{cases} -\Delta\varphi + \dfrac{1}{\tilde{a}}c\varphi = & \dfrac{1}{\tilde{a}}\left(\zeta + (\nabla\tilde{a} - \mathbf{b})\cdot\nabla\varphi\right) \in L^{s/2}(\Omega) \quad \text{in } \Omega\,, \\ \partial_\nu\varphi = & 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad \text{on } \Gamma\,, \end{cases}$$

and obtain the desired $H^{3/2}(\Omega)$ regularity of φ from Corollary 1.15. Let us now prove (1.7.3). Exploiting Corollary 1.15, we get

$$\text{(1.7.8)} \qquad \|\varphi\|_{H^{3/2}(\Omega)} \le C_2\left(\|\zeta\|_{L^{r/2}(\Omega)} + \left(\|\tilde{a}\|_{W^{1,r}(\Omega)} + \|\mathbf{b}\|_{L^r(\Omega)}\right)\|\varphi\|_{W^{1,s}(\Omega)}\right)$$

with $C_2 > 0$ dependent on c. Let us now derive an estimate for $\|\varphi\|_{W^{1,s}(\Omega)}$. Passing to the weak formulation of (1.7.7) and rearranging the terms, there holds

$$\int_\Omega \{\nabla\varphi\cdot\nabla z + \varphi z\}\,dx = \int_\Omega \left[\frac{1}{\tilde{a}}\left(\zeta + \nabla\tilde{a}\cdot\nabla\varphi - \mathbf{b}\cdot\nabla\varphi - c\varphi\right) + \varphi\right] z\,dx =: \tilde{F}(z)$$

$\forall z \in H^1(\Omega)$. As pointed out on page 11, the left-hand side of the previous identity defines an isomorphism from $W^{1.s}(\Omega)$ to $W^{1.s'}(\Omega)^*$. Consequently, its inverse is also an isomorphism and

$$\begin{aligned} \|\varphi\|_{W^{1,s}(\Omega)} &\le C\|\tilde{F}\|_{W^{1.s'}(\Omega)^*} \\ &\le C_3\left(\|\zeta\|_{L^{r/2}(\Omega)} + \left(\|\tilde{a}\|_{W^{1,r}(\Omega)} + \|\mathbf{b}\|_{L^r(\Omega)} + \|c\|_{L^{r/2}(\Omega)} + 1\right)\|\varphi\|_{H^1(\Omega)}\right). \end{aligned}$$

Combining the last inequality with (1.7.5) and (1.7.8), we conclude (1.7.3).

Proof of (3). Thanks to the second part of the theorem, $\varphi \in H^{3/2}(\Omega) \subset W^{1,4}(\Omega)$. The $H^2(\Omega)$ regularity of φ is deduced by applying the well-known result by Grisvard for elliptic problems in convex domains; cf. [**60**, Corollary 4.4.3.8]. It is enough to remark that $\Delta\varphi \in L^2(\Omega)$ as follows after passing the term $c\varphi$ in (1.7.7) to the right-hand side.

Prior to demonstrating (1.7.4), let us recall a classical result for convex domains which can be found e.g. in [**60**]. Let $w \in H^2(\Omega)$ be the unique solution of the equation

$$\begin{cases} -\Delta w + w = & f \quad \text{in } \Omega\,, \\ \partial_\nu w = & 0 \quad \text{on } \Gamma\,, \end{cases}$$

where $f \in L^2(\Omega)$. Then there exists $C_\Omega > 0$, depending only on Ω, such that

$$\|w\|_{H^2(\Omega)} \le C_\Omega\|f\|_{L^2(\Omega)}\,.$$

Adding $(1 - \dfrac{1}{\tilde{a}}c)\varphi$ to both sides of the first equation in (1.7.7), we can apply the last result to (1.7.7) and get, along with the positivity of $\tilde{a}$, Hölder's inequality, the embedding $H^{3/2}(\Omega) \hookrightarrow W^{1,4}(\Omega)$ and (1.7.3),

$$\begin{aligned}
\|\varphi\|_{H^2(\Omega)} &\leq C\left(\|\varphi\|_{L^2(\Omega)} + \frac{1}{\alpha}\|\zeta + (\nabla\tilde{a} - \mathbf{b})\cdot\nabla\varphi - c\varphi\|_{L^2(\Omega)}\right) \\
&\leq C\left(\|\zeta\|_{L^2(\Omega)} + \left(1 + \|c\|_{L^2(\Omega)} + \|\tilde{a}\|_{W^{1,4}(\Omega)} + \|\mathbf{b}\|_{L^4(\Omega)}\right)\|\varphi\|_{H^{3/2}(\Omega)}\right) \\
&\leq C''\|\zeta\|_{L^2(\Omega)}\,.
\end{aligned}$$

□

Let us formulate now the adjoint problem associated with the quasilinear equation (1.1.1). We have seen in the previous section that the linearization of (1.1.1) around a solution leads to the equation (1.6.5). The following equation is the *adjoint equation* of (1.6.5):

$$\text{(1.7.9)} \qquad \left\{\begin{aligned} -\text{div}[a(x,y)\nabla\varphi] + \frac{\partial a}{\partial y}(x,y)\nabla y\cdot\nabla\varphi + \frac{\partial f}{\partial y}(x,y)\varphi &= \zeta \quad \text{in } \Omega\,, \\ a(x,y)\partial_\nu\varphi &= v \quad \text{on } \Gamma\,. \end{aligned}\right.$$

In the next theorem, we establish the existence, uniqueness and regularity of the solution of (1.7.9).

Theorem 1.36. *Suppose that the Assumptions 1.1-1.3 and 1.24-*(1) *hold. Given $y \in W^{1,r}(\Omega)$ with $r > 2$, for any $v \in H^{-1/2}(\Gamma)$ and $\zeta \in H^1(\Omega)^*$, the equation (1.7.9) has a unique solution $\varphi \in H^1(\Omega)$.*

(1) *Assume that $a : \bar{\Omega} \times \mathbb{R} \to \mathbb{R}$ is continuous. If r satisfies the condition (1.3.13), $\zeta \in L^{\frac{2r}{r+2}}(\Omega)$ and $v \in L^{r/2}(\Gamma)$, then there holds $\varphi \in W^{1,r}(\Omega)$.*

(2) *Supposing that Assumption 1.17 holds, $r > 8/3$, $\zeta \in L^p(\Omega)$ ($p > 4/3$) and $v \in L^2(\Gamma)$, then $\varphi \in H^{3/2}(\Omega)$.*

(3) *Suppose that Assumption 1.17 is satisfied and Ω is convex. Further, let $2 \leq r \leq r_0$, with r_0 given as in Theorem 1.21, $\zeta \in L^r(\Omega)$ and either $v = 0$ and $y \in W^{1,2r}(\Omega)$ or $v \in W^{1-1/r,r}(\Gamma)$ and $y \in W^{1,q}(\Omega)$ with $q > 2r$. Then the function φ belongs to $W^{2,r}(\Omega)$.*

Proof. From Theorem 1.35 we know that $T \in \mathcal{L}(H^1(\Omega), H^1(\Omega)^*)$ is an isomorphism. Setting the coefficients $\tilde{a}$, $\mathbf{b}$ and c as in (1.6.6), this is equivalent to the well-posedness of the adjoint equation (1.7.9) in $H^1(\Omega)$, since φ satisfies the equation $T\varphi = \zeta + Bv \in H^1(\Omega)^*$, where B is defined in (1.6.8).

Proof of (1). Thanks to the inclusion $W^{1,r}(\Omega) \subset C(\bar{\Omega})$, the function y is bounded, $a = a(\cdot, y(\cdot)) \in C(\bar{\Omega})$ and $(\partial a/\partial y)(\cdot, y(\cdot))\nabla y \in L^r(\Omega)^2$. The result follows then by similar arguments as in Theorem 1.35-(1). Let us only remark that, taking into account (1.6.6), the functional F defined in (1.7.6) has now the form

$$F(z) = \int_\Omega \left\{ a(x,y)\,(\varphi\nabla z - z\nabla\varphi)\cdot\nabla\psi_j + \left(\zeta - \frac{\partial a}{\partial y}(x,y)\nabla y\cdot\nabla\varphi\right)\psi_j z \right. \\ \left. + \left(1 - \frac{\partial f}{\partial y}(x,y)\right)\varphi\psi_j z\right\} dx + \int_\Gamma v\psi_j z\, d\sigma(x)$$

and that the functional $z \longmapsto \int_\Gamma v\psi_j z\, d\sigma(x)$ is an element of $W^{1,r'}(\Omega)^*$; compare also page 10.

Proof of (2). The $H^{3/2}(\Omega)$ regularity of φ is an immediate consequence of Theorem 1.35-(2); the presence of the boundary datum $v \in L^2(\Gamma)$ does not matter for the proof.

Proof of (3). To prove that $\varphi \in W^{2,r}(\Omega)$ we reformulate (1.7.9) as follows

$$(1.7.10)\quad \begin{cases} -\Delta\varphi = \dfrac{1}{a}\left(\zeta - \dfrac{\partial f}{\partial y}(\cdot,y)\varphi + [\nabla_x a](\cdot,y)\cdot\nabla\varphi - 2\dfrac{\partial a}{\partial y}(\cdot,y)\nabla y\cdot\nabla\varphi\right) & \text{in } \Omega\,, \\ \partial_\nu\varphi = \dfrac{v}{a} & \text{on } \Gamma\,. \end{cases}$$

If we show that $\Delta\varphi \in L^r(\Omega)$ and $\partial_\nu\varphi \in W^{1-1/r,r}(\Gamma)$ then the result follows from Grisvard [**60**, Corollary 4.4.3.8]. Let us consider first the case when $r = 2$. From the second part of the theorem we know that $\varphi \in H^{3/2}(\Omega) \subset W^{1,4}(\Omega)$ which, together with $y \in W^{1,4}(\Omega)$, leads to $\Delta\varphi \in L^2(\Omega)$. This finishes the proof if $v = 0$. Consider now the case when $v \in H^{1/2}(\Gamma)$, $v \neq 0$ and $y \in W^{1,q}(\Omega)$ with $q > 4$. From (1.5.7) we have $W^{1,q}(\Omega) \subset C^{0,1-2/q}(\bar\Omega)$, hence $1/a(\cdot,y(\cdot)) \in C^{0,1-2/q}(\bar\Omega)$. Then due to the fact that $1 - \frac{2}{q} > 1 - \frac{1}{r} = \frac{1}{2}$, we can apply (1.5.6) to conclude $(v/a) \in H^{1/2}(\Gamma)$.

If $r > 2$ then $\varphi \in H^2(\Omega) \subset W^{1,2r}(\Omega)$, as proved above. Together with $y \in W^{1,2r}(\Omega)$, this yields $\Delta\varphi \in L^r(\Omega)$ which is sufficient to deduce the $W^{2,r}(\Omega)$ regularity of φ when $v = 0$. Further, let us assume $v \neq 0$, $v \in W^{1-1/r,r}(\Gamma)$, and $y \in W^{1,q}(\Omega)$ with $q > 2r$. Once again, $1/a(\cdot,y(\cdot)) \in C^{0,1-2/q}(\bar\Omega)$ and (1.5.6) leads to $(v/a) \in W^{1-1/r,r}(\Gamma)$ which completes the proof. □

The following result for the adjoint problem is the analog to Corollary 1.32.

Corollary 1.37. *Suppose that the Assumptions 1.1-1.3, 1.17 and 1.24-*(1) *hold. Given a bounded set $Y \subset W^{1,r}(\Omega)$ with $r > 8/3$ and $v_i \in L^2(\Gamma)$, $\zeta_i \in L^p(\Omega)$ ($p > 4/3$), $i = 1,2$, there exists $C_Y > 0$ such that, for any $y_i \in Y$, there holds the estimate*

$$(1.7.11)\quad \|\varphi_2 - \varphi_1\|_{H^{3/2}(\Omega)} \le C_Y\left(\|y_2 - y_1\|_{W^{1,r}(\Omega)} + \|\zeta_2 - \zeta_1\|_{L^p(\Omega)} + \|v_2 - v_1\|_{L^2(\Gamma)}\right),$$

where $\varphi_i \in H^{3/2}(\Omega)$ *is the solution to*

$$\begin{cases} -\operatorname{div}[a(x,y_i)\nabla\varphi_i(x)] + \dfrac{\partial a}{\partial y}(x,y_i)\nabla y_i\cdot\nabla\varphi_i + \dfrac{\partial f}{\partial y}(x,y_i)\varphi_i = \zeta_i & \textit{in}\,\Omega\,, \\ a(x,y_i)\partial_\nu\varphi_i = v_i & \textit{on}\,\Gamma\,. \end{cases}$$

Proof. Since $W^{1,r}(\Omega) \subset C(\bar\Omega)$, the set Y is bounded in $C(\bar\Omega)$. Subtracting the equations satisfied by φ_i, we get

$$\begin{cases} -\operatorname{div}\left[a(x,y_2)\nabla\left(\varphi_2-\varphi_1\right)\right] + \dfrac{\partial a}{\partial y}(x,y_2)\nabla y_2\cdot\nabla\left(\varphi_2-\varphi_1\right) & \\ \qquad\qquad + \dfrac{\partial f}{\partial y}(x,y_2)(\varphi_2-\varphi_1) = g_\Omega & \text{in}\,\Omega\,, \\ \qquad\qquad a(x,y_2)\nabla(\varphi_2-\varphi_1)\cdot\nu = g_\Gamma & \text{on}\,\Gamma\,, \end{cases}$$

where $g_\Omega \in L^s(\Omega)$, with some $\frac{4r}{r+4} \geq s > \frac{4}{3}$, and $g_\Gamma \in L^2(\Gamma)$ are given by

$$\begin{aligned} g_\Omega = \zeta_2 - \zeta_1 - \operatorname{div}\left[\left(a(\cdot,y_1)-a(\cdot,y_2)\right)\nabla\varphi_1\right] \\ + \left(\frac{\partial a}{\partial y}(\cdot,y_1)\nabla y_1 - \frac{\partial a}{\partial y}(\cdot,y_2)\nabla y_2\right)\cdot\nabla\varphi_1 + \left(\frac{\partial f}{\partial y}(\cdot,y_1) - \frac{\partial f}{\partial y}(\cdot,y_2)\right)\varphi_1 \end{aligned}$$

and

$$g_\Gamma = v_2 - v_1 + \left(a(\cdot,y_1)-a(\cdot,y_2)\right)\partial_\nu\varphi_1\,,$$

respectively. Notice that $\nabla y_i\cdot\nabla\varphi_j \in L^s(\Omega)$, $i,j = 1,2$, because $\nabla y_i \in L^r(\Omega)^2$, $\nabla\varphi_j \in L^4(\Omega)^2$ and $\frac{1}{4}+\frac{1}{r} \leq \frac{1}{s} < \frac{3}{4}$. Now it is easy to modify the proof of (1.7.3) such that (1.7.11) holds true. □

CHAPTER 2

Finite element approximation

2.1. Introduction

This chapter is devoted to the finite element based approximation of the quasilinear equation (1.1.1) and the adjoint equation (1.7.9). We impose minimal regularity assumptions on the data, see Assumptions 1.2-1.3, and use finite elements of degree one to approximate the equations under consideration.

One major difficulty in the analysis is that the uniqueness of solutions of the discrete quasilinear equation is an open problem, even though the continuous equation has a unique solution. This is due to the non-monotone character of the equation (1.1.1). In contrast to the continuous case, the comparison principle that we exploited in the proof of Theorem 1.5 cannot be applied to the discrete equation. Moreover, the nonlinear term a introduces a strong difficulty in the analysis of error estimates for the discretization of the equations (1.1.1) and (1.7.9).

Our main aims of this chapter are threefold. First, we derive estimates in different norms for the error between the solution of the continuous problem (1.1.1) and solutions of the corresponding approximate problems. We assume again that Ω is a polygonal set of $\mathbb{R}^2$; see Barrett and Elliot [**6**] for a finite element approximation of a Neumann type problem in a curved domain. Some results for polyhedral domains of dimension three are presented in Chapter 5. We will distinguish two cases: whether Ω is convex or not. These two different situations yield different orders of convergence in the $L^2(\Omega)$ norm of solutions of the discrete problem to the solution of (1.1.1); see Section 2.3.

Second, we focus on the uniqueness of the solution of the discrete version of (1.1.1). Applying Brouwer's fixed point theorem, we can prove the existence of a discrete solution, but, as far as we know, the uniqueness is an open question until now. However, for a wide class of equations, we are able to prove uniqueness of the discrete solutions, provided that the discretization is fine enough. For equations which do not fall into this class, we give a useful classification of the behavior of solutions of the discrete version of (1.1.1); see Section 2.4. The reader is also referred to Hlaváček [**66**] and Hlaváček et al. [**67**] for some uniqueness results if h is large enough or y and $f(\cdot, 0)$ are sufficiently small.

Third, we study the differentiability of the mapping which associates with each boundary datum a solution of the discrete quasilinear equation which is, in some sense, locally unique. An important consequence of this result is that the discrete adjoint equation has a unique solution in spite of its non-monotone character; see Theorem 2.27. Furthermore, we prove some error estimates for the numerical approximation of the adjoint equation.

All these results play a crucial role in the proof of error estimates for optimal control problems associated with (1.1.1); see Chapter 4.

Let us relate our results to the previous ones in the literature. The corresponding Dirichlet problem associated with equation (1.1.1) was first studied by Douglas and Dupont [**50**] in dimension $n \leq 3$, with f independent on y. They assumed a to be a C^2 function in $\bar{\Omega} \times \mathbb{R}$ such that $(\partial^j a / \partial y^j)$ was bounded in $\bar{\Omega} \times \mathbb{R}$ for $j = 0, 1, 2$. The proof of the error estimates is based on the Aubin-Nitsche trick. To this end, they consider the equation

$$L^* \varphi = \zeta \quad \text{in } \Omega \,, \quad \varphi = 0 \ \text{ on } \Gamma \,, \tag{2.1.1}$$

where $\zeta \in L^2(\Omega)$ and L^* is the adjoint operator of

$$(Lw)(x) = -\text{div} \left[a(x, y(x)) \nabla w(x) + \frac{\partial a}{\partial y}(x, y(x)) w(x) \nabla y(x) \right].$$

Then they use the regularity $\varphi \in H^2(\Omega)$ which holds true because of their assumptions about the $C^{2+\alpha}(\bar{\Omega})$ regularity of y and the smooth property of Γ.

Later, Liu et al. [**78**] considered an extension of equation (1.1.1) to functions $a(x, y) = (a_{ij}(x, y))_{i,j=1}^n$. They still assumed a to be C^2 and bounded along with their derivatives, as mentioned above. They also extended the estimates to solutions $y \in H_0^1(\Omega) \cap H^{k+1}(\Omega)$, $k \geq 1$ (an homogeneous Dirichlet problem was studied), hence they did not require classical solutions as in [**50**]. Moreover, they assumed Ω to be a polygonal or polyhedral domain. In the proof of the error estimates, they followed the Aubin-Nitsche approach, too. The difficulty was that under their assumptions they could not deduce the $H^2(\Omega)$ regularity of the solution φ of (2.1.1), even the existence and uniqueness was not proved. They just assumed the existence, uniqueness and regularity of φ.

Brenner and Scott [**11**, pp. 188–191] studied the Dirichlet problem for small data, proving existence and uniqueness of a solution of the discrete quasilinear equation and deriving error estimates. Unfortunately, their method relies deeply on a fix-point method that cannot be extended to general data.

Casas and Tröltzsch [**37**] focused on the piecewise linear finite element approximation of the equation (1.1.1) with homogeneous Dirichlet boundary condition in the context of optimal control problems. They obtained $L^r(\Omega)$ and $W^{1,r}(\Omega)$ $(r \geq 2)$ error estimates without the boundedness assumption on a and its derivatives. They also proved existence, uniqueness and $H^2(\Omega)$ regularity for the solution of (2.1.1).

In Section 2.3, we will see that even the C^2 regularity of a w.r.t. y is not necessary to derive error estimates; we only require a local Lipschitz property of a.

The outline of this chapter is as follows: In the next section, we recall standard results concerning the finite element method. In Section 2.3, the discrete version of (1.1.1) is introduced and error estimates in different norms are proved in the case of a non-convex and convex domain, respectively. The uniqueness of discrete solutions is the topic of Section 2.4. In Section 2.5, some numerical experiments are presented which confirm the theoretical results of Section 2.3. Section 2.6 is devoted to the study of some useful properties of the discrete solution operator. Finally, this chapter is ended with the error analysis of the finite element approximation of the adjoint equation (1.7.9).

2.2. Assumptions and preliminary results

In this section, we introduce some useful notations and recall some preliminary results concerning the finite element method.

We start with the introduction of a family of triangulations $\{\mathcal{T}_h\}_{h>0}$ of $\bar{\Omega}$, where Ω is supposed to satisfy Assumption 1.1. The mesh $\mathcal{T}_h$ consists of open and pairwise disjoint triangles $T \in \mathcal{T}_h$ such that $\bar{\Omega} = \bigcup_{T\in\mathcal{T}_h} \bar{T}$. With each element $T \in \mathcal{T}_h$ we associate two parameters $\rho(T)$ and $\delta(T)$, where $\rho(T)$ denotes the diameter of the triangle T and $\delta(T)$ is the diameter of the largest ball contained in T. Define the size of the mesh by $h := \max_{T\in\mathcal{T}_h} \rho(T)$. This triangulation is supposed to be regular in the following sense (Brenner and Scott [**11**], see also Casas et al. [**29**]): There exist two positive constants ρ and δ such that

$$\frac{\rho(T)}{\delta(T)} \leq \delta \quad \text{and} \quad \frac{h}{\rho(T)} \leq \rho$$

hold for any $T \in \mathcal{T}_h$ and any $h > 0$. Associated with this triangulation we set

$$Y_h = \Big\{ y_h \in C(\bar{\Omega}) \,|\, y_h|_T \in \mathcal{P}_1(T) \ \text{ for all } \ T \in \mathcal{T}_h \Big\}\,,$$

where $\mathcal{P}_1(T)$ stands for the space of polynomials of degree at most one defined in T. Clearly, the finite element space Y_h consists of continuous functions which are linear in every triangle $T \in \mathcal{T}_h$.

Throughout this chapter, we denote by $\Pi_h : H^1(\Omega) \to Y_h$ the interpolation operator introduced by Scott and Zhang in [**87**] having the following two important properties:

$$\|z - \Pi_h z\|_{H^k(\Omega)} \leq Ch^{m-k}\|z\|_{H^m(\Omega)} \quad \text{for } \ z \in H^m(\Omega) \ \text{ and } \ 2 \geq m \geq k \geq 0 \tag{2.2.1}$$

and

$$\|\Pi_h z\|_{W^{1,q}(\Omega)} \leq C_q \|z\|_{W^{1,q}(\Omega)} \quad \text{for } \ z \in W^{1,q}(\Omega) \ \text{ and } \ q \in (1,\infty)\,, \tag{2.2.2}$$

where $C > 0$ and $C_q > 0$ are independent of h and z. Moreover, for $2 > m \geq k \geq 0$ there holds the stronger result

$$\lim_{h\to 0} \frac{1}{h^{k-m}} \|z - \Pi_h z\|_{H^k(\Omega)} = 0\,, \tag{2.2.3}$$

cf. Brenner and Scott [**11**, §4.8 and Theorem 12.4.2] or Bramble and Scott [**10**]. For the particular case when $m = 3/2$ and $k = 1$ we are going to deduce a formula which is easier to apply than the previous one.

Lemma 2.1. *Let $K \subset H^{3/2}(\Omega)$ be a compact set. Then there exists a sequence $\{\varepsilon_h\}_{h>0}$ of positive real numbers, dependent on K, with $\varepsilon_h \to 0$ as $h \to 0$ such that*

$$\|z - \Pi_h z\|_{H^1(\Omega)} \le \varepsilon_h h^{1/2} \|z\|_{H^{3/2}(\Omega)} \quad \forall z \in K\,. \tag{2.2.4}$$

Proof. Let us consider the linear mapping $A_h : H^{3/2}(\Omega) \longrightarrow H^1(\Omega)$ defined by

$$A_h(z) = \frac{1}{h^{1/2}}\,(z - \Pi_h z)\,.$$

In view of (2.2.1), the mapping $z \longmapsto \|A_h z\|_{H^1(\Omega)}$ is continuous in $H^{3/2}(\Omega)$ and, together with the compactness of K, we deduce the existence of an element $\bar{z}_h$ in K such that

$$\varepsilon_h := \sup_{z\in K} \|A_h z\|_{H^1(\Omega)} = \|A_h \bar{z}_h\|_{H^1(\Omega)}\,.$$

Moreover, (2.2.1) yields that $0 \le \varepsilon_h \le C$ for any $h > 0$. Let us denote by ε_0 the upper limit of $\{\varepsilon_h\}_{h>0}$. Again, due to the compactness of K, there exists a subsequence of $\{\bar{z}_h\}_{h>0}$, denoted in the same way, such that $\{\bar{z}_h\}_{h>0}$ converges strongly in $H^{3/2}(\Omega)$ to some $\bar{z} \in H^{3/2}(\Omega)$. Now from (2.2.3) we obtain

$$\begin{aligned} 0 = \lim_{h\to 0} \|A_h \bar{z}\|_{H^1(\Omega)} &\ge \limsup_{h\to 0} \Big\{ \|A_h \bar{z}_h\|_{H^1(\Omega)} - \|A_h(\bar{z}_h - \bar{z})\|_{H^1(\Omega)} \Big\} \\ &\ge \limsup_{h\to 0} \{\varepsilon_h\} - C \limsup_{h\to 0} \Big\{ \|\bar{z}_h - \bar{z}\|_{H^{3/2}(\Omega)} \Big\} = \varepsilon_0\,, \end{aligned}$$

hence $\varepsilon_0 = 0$ and consequently the whole sequence $\{\varepsilon_h\}_{h>0}$ converges to zero. □

We will rename $C\varepsilon_h$ by ε_h, therefore all the constants are included in ε_h. Obviously, this convention does not change the fact that $\varepsilon_h \to 0$ when $h \to 0$.

Let us also consider the standard nodal interpolation operator $\mathcal{I}_h : C(\bar{\Omega}) \longrightarrow Y_h$ defined by

$$(\mathcal{I}_h z)(x_i) = z(x_i) \quad \text{at every node } x_i \text{ of the triangulation } \mathcal{T}_h\,.$$

Lemma 2.2. *(*[**42**, Theorem 3.1.6]*) Let $k \ge 0$, $m \ge 0$ and $t, q \in [1, \infty]$, be given such that the embeddings $W^{m,q}(T) \hookrightarrow C(\bar{T})$ and $W^{m,q}(T) \hookrightarrow W^{k,t}(T)$ hold for every $T \in \mathcal{T}_h$. Then there exists a constant $C > 0$ independent of h such that*

$$\|z - \mathcal{I}_h z\|_{W^{k,t}(T)} \le C h^{2(1/t - 1/q) + m - k} \|z\|_{W^{m,q}(T)} \quad \forall z \in W^{k,t}(T)\,. \tag{2.2.5}$$

Finally, we recall an inverse estimate for functions belonging to Y_h:

$$\|z_h\|_{W^{m,q}(\Omega)} \le \frac{C}{h^{2\max\{0, 1/t - 1/q\} + m - k}} \|z_h\|_{W^{k,t}(\Omega)} \quad \forall z_h \in Y_h \tag{2.2.6}$$

if $k \le m$ and $t, q \in [1, \infty]$, where $C > 0$ is independent of h, cf. Ciarlet and Lions [**43**, Theorem 17.2] or Bramble and Scott [**10**, Theorem 4.4.24].

In the subsequent sections, we will suppose that the Assumptions 1.1-1.3 and 1.17 hold and, unless otherwise said, $u \in L^2(\Gamma)$ is arbitrary but fixed. In view of Theorem 1.18 on page 16, the solution y_u of (1.1.1) belongs to $H^{3/2}(\Omega) \subset W^{1,4}(\Omega)$. Since u is fixed, we will only write y instead of y_u.

We close this preparatory section by recalling an inequality that will be often used in the sequel.

LEMMA 2.3. *(*[**19**, Lemma A.1]*) There exists a constant $C > 0$ such that*

$$\|z\|_{L^4(\Omega)} \le C\|z\|_{L^2(\Omega)}^{1/2}\|z\|_{H^1(\Omega)}^{1/2} \quad \forall z \in H^1(\Omega)\,.$$

2.3. Numerical analysis of the quasilinear equation

In this section, we study the approximation of the quasilinear equation (1.1.1) by finite elements of degree one. The discrete version of (1.1.1) is given by

$$\text{(2.3.1)} \quad \begin{cases} \text{Find } y_h \in Y_h \text{ such that, for all } \phi_h \in Y_h, \\ \displaystyle\int_\Omega \{a(x,y_h(x))\nabla y_h(x)\cdot\nabla\phi_h(x)+f(x,y_h(x))\phi_h(x)\}\,dx=\int_\Gamma u(x)\phi_h(x)\,d\sigma(x)\,. \end{cases}$$

The next theorem shows that this problem has at least one solution and that all solutions of (2.3.1) are bounded in $H^1(\Omega)$.

THEOREM 2.4. *The equation (2.3.1) has at least one solution. Moreover, there exists a positive constant $C > 0$ such that, for any solution $y_h \in Y_h$ of (2.3.1), the following inequality holds*

$$\text{(2.3.2)} \qquad \|y_h\|_{H^1(\Omega)} \le C\left(\|f(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^2(\Gamma)}\right) \quad (p > 4/3)\,.$$

PROOF. To show the existence of a solution y_h of (2.3.1) we are going to apply Brouwer's fixed point theorem. Let us take $h > 0$ arbitrary but fixed, $M > 0$, $\varepsilon > 0$ and consider the mapping $L_\varepsilon : Y_h \longrightarrow Y_h$, $L_\varepsilon w_h =: z_{h,\varepsilon}$, where $z_{h,\varepsilon}$ satisfies the equation

$$\int_\Omega \{a_M(x,w_h)\nabla z_{h,\varepsilon}\cdot\nabla\phi_h + \varepsilon\chi_E z_{h,\varepsilon}\phi_h\}\,dx$$
$$= \int_\Gamma u\phi_h\,d\sigma(x) - \int_\Omega f_M(x,w_h)\phi_h\,dx \quad \forall\phi_h \in Y_h\,.$$

Here, a_M and f_M are the truncations of a and f, respectively, defined on page 4 and $E \subset \Omega$ is introduced in Assumption 1.3. Taking into account that $Y_h \subset H^1(\Omega)\cap C(\bar\Omega)$, the mapping L_ε is uniquely determined as follows from the Lax-Milgram theorem. Next we show that the mapping L_ε has a fixed point. Arguing as on page 4, we have

$$\|z_{h,\varepsilon}\|_{H^1(\Omega)} \le C_{a,\varepsilon}\left(\|u\|_{L^2(\Gamma)} + \|f_M(\cdot,w_h)\|_{L^p(\Omega)}\right),$$

where $C_{a,\varepsilon}$ depends only on $|\Omega|$, α_a and ε, but neither on a_M nor on f_M. Let us now prove that L_ε is continuous. Once this is verified, we can apply Brouwer's fixed point theorem to obtain a function $y_{h,\varepsilon} \in Y_h$ such that $y_{h,\varepsilon} = L_\varepsilon y_{h,\varepsilon}$, i.e. satisfying

$$\begin{aligned}(2.3.3)\quad &\int_\Omega \{a_M(x, y_{h,\varepsilon})\nabla y_{h,\varepsilon}\cdot\nabla\phi_h + \varepsilon\chi_E y_{h,\varepsilon}\phi_h\}\,dx \\ &\qquad = \int_\Gamma u\phi_h\,d\sigma(x) - \int_\Omega f_M(x, y_{h,\varepsilon})\phi_h\,dx \quad \forall\phi_h \in Y_h\,.\end{aligned}$$

Let us take $w_h \in Y_h$ and a sequence $\{w_{h,k}\}_{k=1}^\infty \subset Y_h$ converging to w_h. It is enough to prove that $z_{h,\varepsilon}^k := L_\varepsilon w_{h,k} \to L_\varepsilon w_h =: z_{h,\varepsilon}$ as $k\to\infty$. Since $\{z_{h,\varepsilon}^k\}_{k=1}^\infty$ is bounded in Y_h and we are in finite dimensions, there exists a subsequence, denoted in the same way, converging strongly in any norm to some $\tilde z_{h,\varepsilon}$. Now we can pass to the limit in the equation satisfied by $z_{h,\varepsilon}^k$ and observe that $\tilde z_{h,\varepsilon} = L_\varepsilon w_h = z_{h,\varepsilon}$. Since any converging subsequence of $\{z_{h,\varepsilon}^k\}_{k=1}^\infty$ converges to $z_{h,\varepsilon}$, it is a standard consequence that the whole sequence converges to $z_{h,\varepsilon}$, too; see e.g. Gajewski et al. [**54**, Lemma 5.4 on page 10]. The convergence $L_\varepsilon w_{h,k} \to L_\varepsilon w_h$ implies the continuity of L_ε and this concludes the existence of $y_{h,\varepsilon} \in Y_h$ satisfying (2.3.3).

Furthermore, the boundedness of $\{y_{h,\varepsilon}\}_{\varepsilon>0}$ in $H^1(\Omega)$ independent of ε can be obtained as in the proof of Theorem 1.5; see pages 5-6. Again, since all norms are equivalent in Y_h, the sequence $\{\|y_{h,\varepsilon}\|_{C(\bar\Omega)}\}_{\varepsilon>0}$ is bounded by some constant $C_\infty > 0$. Hence, we can choose $M \geq C_\infty$ to deduce that $y_{h,\varepsilon}$ satisfies (2.3.3) with a and f instead of a_M and f_M, respectively.

On the other hand, there exists a subsequence $\{y_{h,\varepsilon_k}\}_{k=1}^\infty$ ($\varepsilon_k \to 0$ as $k\to\infty$) of $\{y_{h,\varepsilon}\}_{\varepsilon>0}$ which converges strongly in Y_h to some y_h when $k\to\infty$. By passing to the limit in (2.3.3), with ε replaced by ε_k, it follows that y_h is a solution of (2.3.1).

It remains to prove inequality (2.3.2). To this end, we consider the following identity

$$\int_\Omega\Big\{a(x,y_h)|\nabla y_h|^2 + (f(x,y_h) - f(x,0))\,y_h\Big\}\,dx = \int_\Gamma u y_h\,d\sigma(x) - \int_\Omega f(x,0)y_h\,dx\,.$$

From this equality, along with the positivity of a and the monotonicity of f, we infer

$$\begin{aligned}C_0\|y_h\|_{H^1(\Omega)}^2 &\leq \alpha_a\|\nabla y_h\|_{L^2(\Omega)}^2 + \alpha_f\|y_h\|_{L^2(E)}^2 \\ &\leq C_1\Big(\|f(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^2(\Gamma)}\Big)\,\|y_h\|_{H^1(\Omega)}\end{aligned}$$

which implies (2.3.2) with $C = C_1/C_0$. □

Remark 2.5. *The existence of a solution of the corresponding problem with mixed Dirichlet and Neumann boundary conditions was proved by Hlaváček et al.* [**67**] *under stronger assumptions on a and f, namely supposing that a and f are uniformly bounded.*

The issue of the uniqueness of solutions of (2.3.1) is much more difficult and it is the scope of Section 2.4.

2.3.1. Error estimates in non-convex domains. In this subsection, we consider the case when Ω is not convex and derive estimates of the error $y - y_h$ in the $L^2(\Omega)$ and $H^1(\Omega)$ norms.

THEOREM 2.6. *There exists $h_0 > 0$ such that, for any $h < h_0$, the equation (2.3.1) has at least one solution y_h that obeys*

$$\|y - y_h\|_{L^2(\Omega)} + h^{1/2}\|y - y_h\|_{H^1(\Omega)} \le \varepsilon_h h \,, \tag{2.3.4}$$

where $\varepsilon_h \to 0$ when $h \to 0$. If $\{y_h\}_{h>0}$ is a family of solutions of (2.3.1) that is bounded in $L^\infty(\Omega)$, then (2.3.4) holds as well.

PROOF. Let us take

$$M = 1 + \|y\|_{L^\infty(\Omega)} \,,$$

$$c_0(x) = \frac{\alpha_f}{2}\chi_E(x) \quad \text{and} \quad \tilde f(x,t) = f(x,t) - \frac{\alpha_f}{2}\chi_E(x)t \quad \text{for } x \in \Omega \text{ and } t \in \mathbb{R} \,.$$

The monotonicity of $\tilde f$ w.r.t. the second component can be seen in the following way. If $x \in E$ then for any real $t_1 \neq t_2$ we have from inequality (1.2.2) on page 3 that

$$\frac{\tilde f(x,t_2) - \tilde f(x,t_1)}{t_2 - t_1} = \frac{f(x,t_2) - f(x,t_1)}{t_2 - t_1} - \frac{\alpha_f}{2}\frac{t_2 - t_1}{t_2 - t_1} \ge \frac{\alpha_f}{2} \,.$$

If $x \in \Omega \setminus E$ the monotonicity of $\tilde f$ follows immediately from the monotonicity of f. Let $y_h \in Y_h$ be any function satisfying $\forall \phi_h \in Y_h$ the equation

$$\int_\Omega \Big\{a_M(x,y_h)\nabla y_h\cdot\nabla\phi_h + \Big(c_0(x)y_h + \tilde f_M(x,y_h)\Big)\,\phi_h\Big\}\,dx = \int_\Gamma u\phi_h\,d\sigma \,. \tag{2.3.5}$$

Here, a_M and $\tilde f_M$ are the truncations of a and $\tilde f$, respectively, described on page 4. As in the proof of Theorem 2.4, the existence of y_h follows from Brouwer's fixed point theorem. Observing that $\|\tilde f_M(\cdot,y_h)\|_{L^p(\Omega)} \le \Big(\|\phi_M\|_{L^p(\Omega)} + \alpha_f/2\Big)\,M + \|f(\cdot,0)\|_{L^p(\Omega)}$, with $p > 4/3$ and ϕ_M introduced in Assumption 1.3, the boundedness of $\{\|y_h\|_{H^1(\Omega)}\}_{h>0}$ follows by the same arguments as (2.3.2).

In the sequel, we prove that y_h satisfies (2.3.4), provided that h is small enough. Certainly, if $\|y_h\|_{L^\infty(\Omega)} \le M$ then y_h is a solution of (2.3.1). We will split the proof of (2.3.4) in several steps.

Step 1: Preparations. Consider the function $B_M : L^2(\Omega) \times H^1(\Omega) \times H^1(\Omega) \longrightarrow \mathbb{R}$ defined by

$$B_M(w,z,\phi) = \int_\Omega \{a_M(x,w(x))\nabla z\cdot\nabla\phi + (c_0(x) + c_M(x))\,z\phi\}\,dx \,,$$

where

$$c_M(x) := \begin{cases} \dfrac{\tilde f_M(x,y(x)) - \tilde f_M(x,y_h(x))}{y(x) - y_h(x)} & \text{if } y(x) \neq y_h(x), \\ \alpha_f & \text{otherwise}\,. \end{cases} \tag{2.3.6}$$

From the definition of M and the monotonicity of f w.r.t. the second variable it holds $c_M \geq 0$ a.e. in Ω and

$$c_0(x) + c_M(x) \geq \frac{\alpha_f}{2} \quad \forall x \in E\,. \tag{2.3.7}$$

To see (2.3.7) we consider the most difficult case when $y_h(x) > M$ and $x \in E$. Then (1.2.2) implies

$$\begin{aligned}
c_M(x) &= \frac{\tilde{f}_M(x, y(x)) - \tilde{f}_M(x, y_h(x))}{y(x) - y_h(x)} \\
&= \frac{f(x, y(x)) - f(x, M)}{y(x) - y_h(x)} - \frac{\alpha_f}{2} \frac{y(x) - M}{y(x) - y_h(x)} \\
&= \frac{y(x) - M}{y(x) - y_h(x)} \left(\frac{f(x, y(x)) - f(x, M)}{y(x) - M} - \frac{\alpha_f}{2} \right) \geq \frac{y(x) - M}{y(x) - y_h(x)} \frac{\alpha_f}{2} \geq 0\,.
\end{aligned}$$

If $y_h(x) < -M$ we can proceed analogously. If $x \in E$ and $|y_h(x)| \leq M$ then

$$c_M(x) = \frac{\tilde{f}_M(x, y(x)) - \tilde{f}_M(x, y_h(x))}{y(x) - y_h(x)} = \frac{f(x, y(x)) - f(x, y_h(x))}{y(x) - y_h(x)} - \frac{\alpha_f}{2} \chi_E(x) \geq \frac{\alpha_f}{2}\,.$$

On the other hand, analogous arguments and inequality (1.2.3) on page 3 lead to

$$c_M(x) \leq \phi_M(x) + \frac{\alpha_f}{2} =: \tilde{\phi}_M(x) \;\text{ and }\; \tilde{\phi}_M \in L^p(\Omega)\,. \tag{2.3.8}$$

Let us now provide some useful properties of the function B_M. From the positivity of a, (2.3.7) and the Poincaré inequality (1.3.3) on page 4, we get

$$B_M(w, z, z) \geq \alpha_a \int_\Omega |\nabla z|^2 \, dx + \frac{\alpha_f}{2} \int_E z^2 \, dx \geq C_0 \|z\|^2_{H^1(\Omega)}$$

and, together with (2.3.8),

$$\begin{aligned}
|B_M(w, z, \phi)| &\leq \sup_{x \in \Omega,\, |t| \leq M} |a(x, t)| \|\nabla z\|_{L^2(\Omega)} \|\nabla \phi\|_{L^2(\Omega)} + \frac{\alpha_f}{2} \|z\|_{L^2(\Omega)} \|\phi\|_{L^2(\Omega)} \\
&\quad + \|\tilde{\phi}_M\|_{L^p(\Omega)} \|z\|_{L^{\frac{2p}{p-1}}(\Omega)} \|\phi\|_{L^{\frac{2p}{p-1}}(\Omega)} \\
&\leq C_M \|z\|_{H^1(\Omega)} \|\phi\|_{H^1(\Omega)}
\end{aligned}$$

with $C_M > 0$ only dependent on M but not on h. Moreover, invoking the definition of M, (1.1.1) and (2.3.5), one finds for all $\phi \in H^1(\Omega)$ and $\phi_h \in Y_h$

$$B_M(y, y, \phi) = \int_\Gamma u \phi \, d\sigma(x) + \int_\Omega \Big(c_M(x) y - \tilde{f}_M(x, y) \Big) \phi \, dx\,, \tag{2.3.9}$$

$$B_M(y_h, y_h, \phi_h) = \int_\Gamma u \phi_h \, d\sigma(x) + \int_\Omega \Big(c_M(x) y_h - \tilde{f}_M(x, y_h) \Big) \phi_h \, dx\,. \tag{2.3.10}$$

Furthermore, with the aid of (2.3.6), (2.3.9) and (2.3.10), we obtain for every $\phi_h \in Y_h$

$$\begin{aligned} B_M(y,y,\,\phi_h) &- B_M(y_h,y_h,\phi_h) \\ &= \int_\Omega \Big\{ c_M(x)(y-y_h) - \Big(\tilde{f}_M(x,y) - \tilde{f}_M(x,y_h)\Big)\Big\}\,\phi_h\,dx = 0\,. \end{aligned} \tag{2.3.11}$$

Step 2: Proof of (2.3.4) with y_h satisfying (2.3.5).

Step 2.1: $H^1(\Omega)$ error estimate. Using (2.3.11), the Lipschitz property of a, (2.2.4) (we take here $K = \{y\}$) and (2.2.2) we have the estimate

$$\begin{aligned} C_0\|y_h{-}\Pi_h y\|^2_{H^1(\Omega)} &\le B_M(y_h, y_h-\Pi_h y, y_h-\Pi_h y) \\ &= B_M(y_h,y_h,y_h-\Pi_h y) - B_M(y_h,\Pi_h y, y_h-\Pi_h y) \\ &= B_M(y,y,y_h-\Pi_h y) - B_M(y_h,\Pi_h y, y_h-\Pi_h y) \\ &= B_M(y,y-\Pi_h y, y_h-\Pi_h y) \\ &\quad + (B_M(y,\Pi_h y,y_h-\Pi_h y) - B_M(y_h,\Pi_h y,y_h-\Pi_h y)) \\ &= B_M(y,y-\Pi_h y, y_h-\Pi_h y) \\ &\quad + \int_\Omega (a_M(x,y)-a_M(x,y_h))\,\nabla\Pi_h y\cdot\nabla(y_h-\Pi_h y)\,dx \\ &\le C_M\|y-\Pi_h y\|_{H^1(\Omega)}\|y_h-\Pi_h y\|_{H^1(\Omega)} \\ &\quad + C_{a,M}\int_\Omega |y-y_h||\nabla\Pi_h y\cdot\nabla(y_h-\Pi_h y)|\,dx \\ &\le \Big(\varepsilon_h h^{1/2}\|y\|_{H^{3/2}(\Omega)} + C\|y-y_h\|_{L^4(\Omega)}\|y\|_{W^{1,4}(\Omega)}\Big)\,\|y_h-\Pi_h y\|_{H^1(\Omega)}\,. \end{aligned} \tag{2.3.12}$$

Thus,

$$\|y_h-\Pi_h y\|_{H^1(\Omega)} \le \varepsilon_h h^{1/2} + C\|y-y_h\|_{L^4(\Omega)}\,.$$

From this inequality and the estimate (2.2.4) we infer

$$\begin{aligned} \|y-y_h\|_{H^1(\Omega)} &\le \|y-\Pi_h y\|_{H^1(\Omega)} + \|y_h-\Pi_h y\|_{H^1(\Omega)} \\ &\le \varepsilon_h h^{1/2} + C\|y-y_h\|_{L^4(\Omega)}\,. \end{aligned} \tag{2.3.13}$$

Further, according to Lemma 2.3, it follows

$$\|y-y_h\|_{L^4(\Omega)} \le C\|y-y_h\|^{1/2}_{L^2(\Omega)}\|y-y_h\|^{1/2}_{H^1(\Omega)}$$

and, for $q = q' = 2$ and any $\varepsilon > 0$, Young's inequality leads to

$$\|y-y_h\|^{1/2}_{L^2(\Omega)}\|y-y_h\|^{1/2}_{H^1(\Omega)} \le \frac{1}{2\varepsilon^2}\|y-y_h\|_{L^2(\Omega)} + \frac{\varepsilon^2}{2}\|y-y_h\|_{H^1(\Omega)}\,.$$

Taking ε small enough in the previous inequality, we conclude from (2.3.13) that

$$\|y-y_h\|_{H^1(\Omega)} \le \varepsilon_h h^{1/2} + C\|y-y_h\|_{L^2(\Omega)}\,. \tag{2.3.14}$$

Step 2.2: $L^2(\Omega)$ error estimate. In order to estimate $\|y-y_h\|_{L^2(\Omega)}$, we use a duality argument based on the Aubin-Nitsche trick. To this aim, we introduce the function

$$b_M(x) = \begin{cases} \dfrac{a_M(x,y(x)) - a_M(x,y_h(x))}{y(x)-y_h(x)} & \text{if } y(x) \neq y_h(x)\,, \\ \alpha_a & \text{otherwise}\,. \end{cases}$$

Exploiting Assumption 1.2, we easily see that $|b_M(x)| \leq C_M\ \forall x \in \Omega$, with $C_M > 0$ being independent of h. Next we apply Theorem 1.35-(2) on page 30, by setting $\tilde{a}(x) = a_M(x,y(x))$, $\mathbf{b}(x) = b_M(x)\nabla y(x)$ and $c(x) = c_0(x) + c_M(x)$, to obtain a function $\varphi \in H^{3/2}(\Omega)$ such that, for every $w \in H^1(\Omega)$

$$\int_\Omega \{a_M(x,y)\nabla\varphi\cdot\nabla w + b_M w\nabla y\cdot\nabla\varphi + (c_0+c_M)\,\varphi w\}\,dx = \int_\Omega (y-y_h)w\,dx\,. \tag{2.3.15}$$

By taking $w = y - y_h$ in (2.3.15) and using the definitions of c_M and b_M as well as the positivity of a and (2.3.5), we derive

$$\begin{aligned}
\|y-y_h\|^2_{L^2(\Omega)} = &\int_\Omega \{a_M(x,y)\nabla\varphi\cdot(\nabla y - \nabla y_h) + b_M(x)\,(y-y_h)\,\nabla y\cdot\nabla\varphi \\
&\quad + (c_0(x)+c_M(x))\,\varphi(y-y_h)\}\,dx \\
= &\int_\Omega \Big\{a_M(x,y)\nabla\varphi\cdot\nabla y + \Big(c_0(x)y + \tilde{f}_M(x,y)\Big)\,\varphi\Big\}\,dx \\
&- \int_\Omega \Big\{a_M(x,y)\nabla\varphi\cdot\nabla y_h + \Big(c_0(x)y_h + \tilde{f}_M(x,y_h)\Big)\,\varphi\Big\}\,dx \\
&+ \int_\Omega (a_M(x,y) - a_M(x,y_h))\,\nabla y\cdot\nabla\varphi\,dx \\
= &\int_\Omega \Big\{a_M(x,y)(\nabla\varphi - \nabla\Pi_h\varphi)\cdot\nabla y + \Big(c_0(x)y + \tilde{f}_M(x,y)\Big)(\varphi - \Pi_h\varphi)\Big\}\,dx \\
&+ \int_\Omega \Big\{a_M(x,y_h)\nabla\Pi_h\varphi\cdot\nabla y_h + \Big(c_0(x)y_h + \tilde{f}_M(x,y_h)\Big)\,\Pi_h\varphi\Big\}\,dx \\
&- \int_\Omega \Big\{a_M(x,y)\nabla\varphi\cdot\nabla y_h + \Big(c_0(x)y_h + \tilde{f}_M(x,y_h)\Big)\,\varphi\Big\}\,dx \\
&+ \int_\Omega (a_M(x,y) - a_M(x,y_h))\,\nabla y\cdot\nabla\varphi\,dx \\
= &\int_\Omega a_M(x,y)\,(\nabla\varphi - \nabla\Pi_h\varphi)\cdot(\nabla y - \nabla y_h)\,dx \\
&+ \int_\Omega \Big(c_0(x)\,(y-y_h) + \tilde{f}_M(x,y) - \tilde{f}_M(x,y_h)\Big)\,(\varphi - \Pi_h\varphi)\,dx \\
&+ \int_\Omega (a_M(x,y) - a_M(x,y_h))\,(\nabla y\cdot\nabla\varphi - \nabla y_h\cdot\nabla\Pi_h\varphi)\,dx
\end{aligned}$$

$$\begin{aligned}&= \int_\Omega a_M(x,y)\,(\nabla\varphi - \nabla\Pi_h\varphi)\cdot(\nabla y - \nabla y_h)\,dx\\
&\quad + \int_\Omega \Big(c_0(x)\,(y-y_h) + \tilde f_M(x,y) - \tilde f_M(x,y_h)\Big)\,(\varphi - \Pi_h\varphi)\,dx\\
&\quad + \int_\Omega (a_M(x,y) - a_M(x,y_h))\,(\nabla y - \nabla y_h)\cdot\nabla\Pi_h\varphi\,dx\\
&\quad + \int_\Omega (a_M(x,y) - a_M(x,y_h))\,(\nabla\varphi - \nabla\Pi_h\varphi)\cdot\nabla y\,dx\,. \qquad (2.3.16)\end{aligned}$$

With the aid of the assumptions on a and f and the estimates (2.2.1) and (2.2.2), it follows

$$\|y-y_h\|^2_{L^2(\Omega)} \le C\left(h^{1/2} + \|y-y_h\|_{H^1(\Omega)}\right)\|y-y_h\|_{H^1(\Omega)}\|\varphi\|_{H^{3/2}(\Omega)}\,.$$

In view of inequality (1.7.3) on page 30, we have $\|\varphi\|_{H^{3/2}(\Omega)} \le C\|y-y_h\|_{L^2(\Omega)}$, hence

$$\|y-y_h\|_{L^2(\Omega)} \le C\left(h^{1/2} + \|y-y_h\|_{H^1(\Omega)}\right)\|y-y_h\|_{H^1(\Omega)}\,. \qquad (2.3.17)$$

Utilizing (2.3.14), we get from the previous inequality

$$\|y-y_h\|_{L^2(\Omega)} \le \varepsilon_h h + \tilde C\|y-y_h\|^2_{L^2(\Omega)}\,. \qquad (2.3.18)$$

In the next step, we prove the strong convergence $y_h \to y$ in $L^2(\Omega)$ as $h \to 0$, implying

$$\tilde C\|y-y_h\|^2_{L^2(\Omega)} \le \frac{1}{2}\|y-y_h\|_{L^2(\Omega)}\,,$$

for sufficiently small h. Thus, we deduce from (2.3.14) and (2.3.18) that y_h satisfies (2.3.4) for every $h>0$ small enough.

Step 3: Convergence $y_h \to y$ in $L^2(\Omega)$. The boundedness of $\{y_h\}_{h>0}$ in $H^1(\Omega)$ implies the existence of an element $w \in H^1(\Omega)$ and a subsequence of $\{y_h\}_{h>0}$, denoted again by $\{y_h\}_{h>0}$, such that $y_h \rightharpoonup w$ weakly in $H^1(\Omega)$. Now we show that $w = y$. To this end, let $z \in H^2(\Omega)$ be arbitrarily chosen. Then $\Pi_h z \in Y_h$ and there holds $\|z - \Pi_h z\|_{H^1(\Omega)} \to 0$ as $h \to 0$. Knowing that y_h satisfies

$$\int_\Omega \Big\{a_M(x,y_h)\nabla y_h\cdot\nabla\Pi_h z + \Big(c_0(x)y_h + \tilde f_M(x,y_h)\Big)\,\Pi_h z\Big\}\,dx = \int_\Gamma u\Pi_h z\,d\sigma(x)\,,$$

and due to the boundedness of $a_M(\cdot,y_h)$ and the domination of $\tilde f_M(\cdot,y_h)$ by a function in $L^p(\Omega)$, the strong convergence $\Pi_h z \to z$ in $H^1(\Omega)$, (2.2.2) and the weak convergence $y_h \rightharpoonup w$ in $H^1(\Omega)$, we can pass to the limit in the previous equation and find that w satisfies

$$\int_\Omega \Big\{a_M(x,w)\nabla w\cdot\nabla z + \Big(c_0(x)w + \tilde f_M(x,w)\Big)\,z\Big\}\,dx = \int_\Gamma uz\,d\sigma(x)\,. \qquad (2.3.19)$$

From this equality and the density of $H^2(\Omega)$ in $H^1(\Omega)$ we deduce that (2.3.19) holds for any $z \in H^1(\Omega)$. It is immediate to check that (2.3.19) also holds with y instead of w. Hence, from the uniqueness of the solution of (2.3.19), see Lemma 2.9 below, we conclude that $w = y$. Now, as argued in the proof of Theorem 2.4, the whole original sequence $\{y_h\}_{h>0}$ converges weakly to y in $H^1(\Omega)$. Since $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$, the convergence $y_h \to y$ is strong in $L^2(\Omega)$.

Step 4: Convergence $y_h \to y$ in $L^\infty(\Omega)$. In the last step of the proof, we demonstrate the convergence $y_h \to y$ in $L^\infty(\Omega)$ which implies that y_h is a solution of (2.3.1) for h sufficiently small. To this aim, we use the estimates (2.2.5) and (2.2.6) to obtain

$$\|z - \mathcal{I}_h z\|_{L^2(\Omega)} + h\|z - \mathcal{I}_h z\|_{L^\infty(\Omega)} \leq Ch^{3/2}\|z\|_{H^{3/2}(\Omega)} \quad \forall z \in H^{3/2}(\Omega) \tag{2.3.20}$$

and

$$\|z_h\|_{L^\infty(\Omega)} \leq \frac{C}{h}\|z_h\|_{L^2(\Omega)} \quad \forall z_h \in Y_h\,. \tag{2.3.21}$$

From (2.3.4), (2.3.20) and (2.3.21), we deduce

$$\begin{aligned}\|y - y_h\|_{L^\infty(\Omega)} &\leq \|y - \mathcal{I}_h y\|_{L^\infty(\Omega)} + \|\mathcal{I}_h y - y_h\|_{L^\infty(\Omega)}\\ &\leq Ch^{1/2}\|y\|_{H^{3/2}(\Omega)} + \frac{C}{h}\|\mathcal{I}_h y - y_h\|_{L^2(\Omega)}\\ &\leq Ch^{1/2}\|y\|_{H^{3/2}(\Omega)} + \frac{C}{h}\Big(\|\mathcal{I}_h y - y\|_{L^2(\Omega)} + \|y - y_h\|_{L^2(\Omega)}\Big) \to 0\,.\end{aligned}$$

The above uniform convergence ensures the existence of a value $h_0 > 0$ such that $\|y_h\|_{L^\infty(\Omega)} \leq \|y\|_{L^\infty(\Omega)} + 1 = M$ for all $h < h_0$. This implies $a_M(x, y_h) = a(x, y_h)$ and $f_M(x, y_h) = f(x, y_h)$, thus y_h is a solution of (2.3.1) for every $h < h_0$ and (2.3.4) holds.

Finally, if $\{y_h\}_{h>0}$ is a family of solutions of (2.3.1) bounded in $L^\infty(\Omega)$, then we can argue as above taking

$$M = \sup_{h>0}\|y_h\|_{L^\infty(\Omega)} + \|y\|_{L^\infty(\Omega)} + 1\,. \tag{2.3.22}$$

□

REMARK 2.7. *In step 2.2 of the previous proof, we can also use (2.2.4) to estimate the error $\|\varphi - \Pi_h\varphi\|_{H^1(\Omega)}$, although φ is dependent on h because of its dependence on y_h. This is due to the fact that $\{y - y_h\}_{h>0}$ is bounded in $H^1(\Omega)$, hence relatively compact in $L^2(\Omega)$. This implies that the set containing the solutions $\varphi = \varphi(h)$ of (2.3.15) corresponding to $y - y_h$, $h > 0$, is relatively compact in $H^{3/2}(\Omega)$. According to Lemma 2.1, ε_h can be chosen independently of $y - y_h$ in the interpolation estimate of φ.*

REMARK 2.8. *The introduction of the equation (2.3.5) is not standard. The reason for this is that the truncation f_M of the function f does not satisfy in general inequality (1.2.2) on page 3. To preserve this property we have replaced $f_M(x, t)$ by $c_0(x)t + \tilde{f}_M(x, t)$, $(x, t) \in \Omega \times \mathbb{R}$. This difficulty does not occur in the case of homogeneous Dirichlet boundary conditions, since (1.2.2) is not a necessary assumption, cf. Casas and Tröltzsch* [**37**].

Next we prove a result that we have already used in the proof of the previous theorem.

Lemma 2.9. *The solution of the equation (2.3.19) is unique.*

Proof. First, the solution $y = y_u$ of (1.1.1) is also a solution of (2.3.19), with M given on page 41. To deduce its uniqueness we follow the lines of Theorem 1.5 on page 3. Let $w \in H^1(\Omega)$ be another solution of (2.3.19). The continuity of w in $\bar\Omega$ can be shown as in the proof of Theorem 1.5.

Let $\varepsilon > 0$. By setting $y_1 = y$ and $y_2 = w$, we define the sets Ω_0, Ω_ε, and the function $z_\varepsilon \in H^1(\Omega)$ as on pages 6-7. Thanks to the monotonicity of f, we find $\left(\tilde f_M(x,w) - \tilde f_M(x,y)\right) z_\varepsilon \geq 0$. Indeed, if $z_\varepsilon(x) \neq 0$ it follows by computations similar to the ones on page 42 that

$$\left(\tilde f_M(x,w) - \tilde f_M(x,y)\right) z_\varepsilon = \frac{\tilde f_M(x,w) - \tilde f_M(x,y)}{z_\varepsilon} z_\varepsilon^2 \geq \frac{\tilde f_M(x,w) - \tilde f_M(x,y)}{w-y} z_\varepsilon^2 \geq 0\,.$$

Now a straightforward modification of the estimates (1.3.10) and (1.3.11) on page 7 yields $\|z_\varepsilon\|^2_{L^2(\Omega)} \leq C'\varepsilon^2 \|\nabla y\|^2_{L^2(\Omega_0\setminus\Omega_\varepsilon)}$. The rest of the proof is identical with that of Theorem 1.5. □

As a consequence of the estimate (2.3.4), we obtain the following result.

Corollary 2.10. *For every sequence $\{y_h\}_{h<h_0}$ of solutions to (2.3.1) satisfying (2.3.4) it holds*

$$\|y - y_h\|_{W^{1,q}(\Omega)} \leq Ch^{2/q-1/2} \quad \forall q \in (2,4]\,. \tag{2.3.23}$$

Proof. Invoking Theorem 2.6 and the inequalities (2.2.5) and (2.2.6), it follows that

$$\begin{aligned}
\|y - y_h\|_{W^{1,q}(\Omega)} &\leq \|y - \mathcal{I}_h y\|_{W^{1,q}(\Omega)} + \|\mathcal{I}_h y - y_h\|_{W^{1,q}(\Omega)} \\
&\leq C_1 h^{2/q-1/2}\|y\|_{H^{3/2}(\Omega)} + \frac{C_2}{h^{\frac{q-2}{q}}}\|\mathcal{I}_h y - y_h\|_{H^1(\Omega)} \\
&\leq C_1 h^{2/q-1/2}\|y\|_{H^{3/2}(\Omega)} + \frac{C_2}{h^{\frac{q-2}{q}}}\left(\|\mathcal{I}_h y - y\|_{H^1(\Omega)} + \|y - y_h\|_{H^1(\Omega)}\right) \\
&\leq C_1 h^{2/q-1/2}\|y\|_{H^{3/2}(\Omega)} + C_2 h^{\frac12 - \frac{q-2}{q}}\left(C_1\|y\|_{H^{3/2}(\Omega)} + \varepsilon_h\right) \leq Ch^{2/q-1/2}
\end{aligned}$$

and (2.3.23) is concluded. □

2.3.2. Error estimates in convex domains. Our goal in this subsection is to derive error estimates in the $L^2(\Omega)$ and $H^1(\Omega)$ norms in the case when Ω is convex. These estimates will improve the ones obtained in the non-convex case, due to the higher regularity of the solution φ of the adjoint problem (2.3.15) in convex domains.

THEOREM 2.11. *Assuming that Ω is convex, the conclusions of Theorem 2.6 remain valid with*

$$\|y - y_h\|_{L^2(\Omega)} + h\|y - y_h\|_{H^1(\Omega)} \le \varepsilon_h h^{3/2} \tag{2.3.24}$$

instead of (2.3.4).

PROOF. By Theorem 2.6, there exists $h_0 > 0$ such that, for any $h < h_0$, the equation (2.3.1) has at least one solution y_h that obeys (2.3.4). The proof will be complete if we show the error estimate in the $L^2(\Omega)$ norm as stated in the theorem.

According to Corollary 2.10 and to the embedding $W^{1,q}(\Omega) \hookrightarrow C(\bar{\Omega})$ for any $q > 2$, y_h converges uniformly to y as h tends to zero. Choosing M as in (2.3.22) (the supremum is now taken over $\{y_h\}_{h<h_0}$), we can skip the truncation of the coefficients a and f and consider equation (2.3.1) directly.

In the convex case, Theorem 1.35-(3) on page 30 (remember that $y \in W^{1,4}(\Omega)$) ensures the $H^2(\Omega)$ regularity of the solution φ of the equation

$$\int_\Omega \{a(x,y)\nabla\varphi\cdot\nabla w + (b\nabla y\cdot\nabla\varphi + c\varphi)\, w\}\, dx = \int_\Omega (y - y_h) w\, dx \quad \forall w \in H^1(\Omega)\,, \tag{2.3.25}$$

where

$$b(x) := \frac{a(x,y(x)) - a(x,y_h(x))}{y(x) - y_h(x)} \quad \text{and} \quad c(x) := \frac{f(x,y(x)) - f(x,y_h(x))}{y(x) - y_h(x)}$$

if $y(x) \ne y_h(x)$ and $b(x) := \alpha_a$ and $c(x) := \alpha_f$ otherwise. Moreover, inequality (1.7.4) on page 30 implies the existence of $C > 0$ such that $\|\varphi\|_{H^2(\Omega)} \le C\|y - y_h\|_{L^2(\Omega)}$. Furthermore, from (2.2.1) it follows that $\varphi - \Pi_h\varphi$ is of order h^2 and h in the $L^2(\Omega)$ and $H^1(\Omega)$ norms, respectively.

In the sequel, we show that the above results lead to the desired error estimate in the $L^2(\Omega)$ norm. From inequality (2.3.4) we get

$$\|y - y_h\|_{L^2(\Omega)} \le \varepsilon_h h \tag{2.3.26}$$

and

$$\|y - y_h\|_{L^6(\Omega)} \le \varepsilon_h h^{1/2}\,, \tag{2.3.27}$$

where we have used that $H^1(\Omega) \hookrightarrow L^6(\Omega)$. Now (2.3.16) can be replaced by

$$\begin{aligned}
\|y - y_h\|^2_{L^2(\Omega)} &= \int_\Omega a(x,y)\nabla(\varphi - \Pi_h\varphi)\cdot\nabla(y - y_h)\, dx \\
&\quad + \int_\Omega (f(x,y) - f(x,y_h))\,(\varphi - \Pi_h\varphi)\, dx \\
&\quad + \int_\Omega (a(x,y) - a(x,y_h))\,\nabla(y - y_h)\cdot\nabla\Pi_h\varphi\, dx \\
&\quad + \int_\Omega (a(x,y) - a(x,y_h))\,\nabla(\varphi - \Pi_h\varphi)\cdot\nabla y\, dx\,.
\end{aligned} \tag{2.3.28}$$

With the aid of (2.3.28), (2.3.4) and (2.2.1), we are going to show that

$$\|y-y_h\|_{L^2(\Omega)} \le \varepsilon_h h^{3/2} + \varepsilon_h h^{1/2}\|y-y_h\|_{L^3(\Omega)}\,. \tag{2.3.29}$$

To prove this inequality let us estimate each integral on the right-hand side of (2.3.28). The boundedness of a and the Cauchy-Schwarz inequality lead to

$$\begin{aligned}\int_\Omega a(x,y)\nabla(\varphi-\Pi_h\varphi)\cdot\nabla(y-y_h)\,dx &\le C\|\nabla(\varphi-\Pi_h\varphi)\|_{L^2(\Omega)}\|\nabla(y-y_h)\|_{L^2(\Omega)}\\ &\le C\|\varphi-\Pi_h\varphi\|_{H^1(\Omega)}\|y-y_h\|_{H^1(\Omega)}\\ &\le \varepsilon_h h^{3/2}\|\varphi\|_{H^2(\Omega)} \le \varepsilon_h h^{3/2}\|y-y_h\|_{L^2(\Omega)}\,.\end{aligned}$$

Using (1.2.3) on page 3 and Hölder's inequality, we can estimate the second term on the right-hand side of (2.3.28) as follows

$$\begin{aligned}\int_\Omega (f(x,y)-f(x,y_h))\,(\varphi-\Pi_h\varphi)\,dx &\le \int_\Omega |\phi_M(x)||y-y_h||\varphi-\Pi_h\varphi|\,dx\\ &\le \|\phi_M\|_{L^{4/3}(\Omega)}\||y-y_h||\varphi-\Pi_h\varphi|\|_{L^4(\Omega)}\\ &\le \|\phi_M\|_{L^{4/3}(\Omega)}\|y-y_h\|_{L^8(\Omega)}\|\varphi-\Pi_h\varphi\|_{L^8(\Omega)}\,.\end{aligned}$$

The embedding $H^1(\Omega)\hookrightarrow L^8(\Omega)$, along with (2.3.4), yields

$$\int_\Omega (f(x,y)-f(x,y_h))\,(\varphi-\Pi_h\varphi)\,dx \le \varepsilon_h h^{3/2}\|\varphi\|_{H^2(\Omega)} \le \varepsilon_h h^{3/2}\|y-y_h\|_{L^2(\Omega)}\,.$$

Thanks to the stability of the interpolation operator Π_h in $W^{1,6}(\Omega)$ (Eq. (2.2.2)) and the embedding $H^2(\Omega)\hookrightarrow W^{1,6}(\Omega)$, we obtain for the third term on the right-hand side of (2.3.28)

$$\begin{aligned}&\int_\Omega (a(x,y)-a(x,y_h))\nabla(y-y_h)\cdot\nabla\Pi_h\varphi\,dx\\ &\qquad\le C_{a,M}\|\,|y-y_h|\,\nabla\Pi_h\varphi\|_{L^2(\Omega)}\|\nabla(y-y_h)\|_{L^2(\Omega)}\\ &\qquad\le \varepsilon_h h^{1/2}\|y-y_h\|_{L^3(\Omega)}\|\nabla\Pi_h\varphi\|_{L^6(\Omega)}\\ &\qquad\le \varepsilon_h h^{1/2}\|y-y_h\|_{L^3(\Omega)}\|y-y_h\|_{L^2(\Omega)}\,.\end{aligned}$$

Concerning the last integral on the right-hand side of (2.3.28), we have, along with (2.3.27),

$$\begin{aligned}&\int_\Omega (a(x,y)-a(x,y_h))\nabla(\varphi-\Pi_h\varphi)\cdot\nabla y\,dx\\ &\qquad\le C_{a,M}\|y-y_h\|_{L^6(\Omega)}\|\nabla\left(\varphi-\Pi_h\varphi\right)\|_{L^2(\Omega)}\|\nabla y\|_{L^3(\Omega)}\\ &\qquad\le \varepsilon_h h^{3/2}\|y-y_h\|_{L^2(\Omega)}\,.\end{aligned}$$

Summing up all inequalities above, we conclude (2.3.29).

Further, let us take h small enough such that $h < 1$ and $\varepsilon_h < 1$. We prove now the following inequality

$$\|y - y_h\|_{L^2(\Omega)} \le \frac{\varepsilon_h}{1-\varepsilon_h} h^{3/2} \tag{2.3.30}$$

which yields the desired error estimate in the $L^2(\Omega)$ norm. To verify (2.3.30) we deduce that

$$\|y - y_h\|_{L^2(\Omega)} \le \varepsilon_h h^{3/2} \left[\sum_{j=0}^{k-1} \varepsilon_h^j\right] h^{-\frac{1}{2^k}} \quad \text{for every } k \in \mathbb{N}\,. \tag{2.3.31}$$

Taking the limit $k \to \infty$ in the last inequality, we arrive at (2.3.30). To obtain (2.3.31) we proceed by induction on k. For $k = 1$ (2.3.31) is the same as (2.3.26). Let us assume that (2.3.31) holds for some integer k and let us prove it for $k+1$. First, employing Hölder's inequality, we find

$$\|y - y_h\|_{L^3(\Omega)} \le \|y - y_h\|_{L^2(\Omega)}^{1/2} \|y - y_h\|_{L^6(\Omega)}^{1/2}\,. \tag{2.3.32}$$

Next using (2.3.31) and (2.3.27), we get from (2.3.32)

$$\begin{aligned} \|y - y_h\|_{L^3(\Omega)} &\le \varepsilon_h^{1/2} h^{3/4} \left[\sum_{j=0}^{k-1} \varepsilon_h^j\right]^{1/2} h^{-\frac{1}{2^{k+1}}} \varepsilon_h^{1/2} h^{1/4} \\ &= \varepsilon_h h \left[\sum_{j=0}^{k-1} \varepsilon_h^j\right]^{1/2} h^{-\frac{1}{2^{k+1}}} \le \varepsilon_h h \left[\sum_{j=0}^{k-1} \varepsilon_h^j\right] h^{-\frac{1}{2^{k+1}}}\,. \end{aligned}$$

Inserting this estimate in (2.3.29), it follows

$$\|y - y_h\|_{L^2(\Omega)} \le \varepsilon_h h^{3/2} \left(1 + \varepsilon_h \left[\sum_{j=0}^{k-1} \varepsilon_h^j\right] h^{-\frac{1}{2^{k+1}}}\right) \le \varepsilon_h h^{3/2} \left[\sum_{j=0}^{k} \varepsilon_h^j\right] h^{-\frac{1}{2^{k+1}}}$$

which yields (2.3.31) and concludes the proof. □

2.3.3. Further results.

In many applications, the regularity of the solution of the equation (1.1.1) is better than $H^{3/2}(\Omega)$. This is the case when, e.g., Ω is convex and u is sufficiently regular; see Theorem 1.21 on page 18. In such a situation, we obtain better error estimates as the following theorem shows.

THEOREM 2.12. *Let $y \in H^2(\Omega)$ be the solution of (1.1.1). Then the conclusions of Theorem 2.6 remain valid with*

$$\|y - y_h\|_{L^2(\Omega)} + h^{1/2} \|y - y_h\|_{H^1(\Omega)} \le C h^{3/2} \tag{2.3.33}$$

instead of (2.3.4), where $C > 0$ is independent of h. Moreover, if Ω is convex then

$$\|y - y_h\|_{L^2(\Omega)} + h \|y - y_h\|_{H^1(\Omega)} \le C h^2\,. \tag{2.3.34}$$

Proof. According to Theorem 2.6, for any $h < h_0$ there exists at least one solution y_h of (2.3.1) satisfying (2.3.4). Since $\{y_h\}_{h<h_0}$ is bounded in $W^{1,4}(\Omega) \subset C(\bar{\Omega})$ (Corollary 2.10), we can choose M as in (2.3.22) (the supremum is now taken over $\{y_h\}_{h<h_0}$) to skip the truncation of the coefficients a and f. Therefore, we can consider equation (2.3.1) directly. Thanks to the higher regularity of y, we have by (2.2.1) the following estimate for the interpolation error

$$\|y - \Pi_h y\|_{L^2(\Omega)} + h\|y - \Pi_h y\|_{H^1(\Omega)} \le Ch^2 \|y\|_{H^2(\Omega)}\,. \tag{2.3.35}$$

Let us first prove (2.3.33) if Ω is non-convex. By the same arguments as in the proof of Theorem 2.6, the previous estimate allows us to deduce

$$\|y - y_h\|_{H^1(\Omega)} \le C\left(h + \|y - y_h\|_{L^2(\Omega)}\right) \quad \forall h < h_0\,. \tag{2.3.36}$$

We apply (2.3.36) to improve the order of convergence up to that given in (2.3.33). Combining (2.3.36) with (2.3.17), we find for sufficiently small h that

$$\|y - y_h\|_{L^2(\Omega)} + h^{1/2}\|y - y_h\|_{H^1(\Omega)} \le C\left(h^{3/2} + \|y - y_h\|^2_{L^2(\Omega)}\right).$$

Finally, the convergence $\|y - y_h\|_{L^2(\Omega)} \to 0$ as $h \to 0$ leads to (2.3.33).

If Ω is convex then (2.3.24) holds true. Moreover, the fact that $y \in H^2(\Omega) \subset W^{1,4}(\Omega)$ and Theorem 1.35-(3) on page 30 imply that the solution φ of (2.3.25) is in $H^2(\Omega)$ and $\|\varphi\|_{H^2(\Omega)} \le C\|y - y_h\|_{L^2(\Omega)}$ is valid with some constant $C > 0$. With the aid of the assumptions on a and f, it follows from (2.3.28) that

$$\|y - y_h\|_{L^2(\Omega)} \le C\left(h\|y - y_h\|_{H^1(\Omega)} + \|y - y_h\|^2_{H^1(\Omega)}\right).$$

This inequality, combined with (2.3.36), gives

$$\|y - y_h\|_{L^2(\Omega)} \le C\left(h^2 + \|y - y_h\|^2_{L^2(\Omega)}\right).$$

The last estimate, along with the convergence $\|y - y_h\|_{L^2(\Omega)} \to 0$ as $h \to 0$, completes the proof of (2.3.34). □

Remark 2.13. *For an extension of the statement of Theorem 2.12 to higher order finite elements the reader is referred to* [**18**, Theorem 3.7].

The next result is a simple consequence of the previous theorem in the case when $u \in H^{1/2}(\Gamma)$, $p \ge 2$ and Ω convex. Then from Theorem 1.21 it follows that $y \in H^2(\Omega)$.

Corollary 2.14. *Let $u \in H^{1/2}(\Gamma)$ and assume that Ω is convex and $p \ge 2$. Then the conclusions of Theorem 2.6 remain valid with*

$$\|y - y_h\|_{L^2(\Omega)} + h\|y - y_h\|_{H^1(\Omega)} \le Ch^2$$

instead of (2.3.4).

Up to now we have established error estimates only for a fixed boundary datum u. Notice that the sequence $\{\varepsilon_h\}_{h>0}$ given in the statement of Theorem 2.6 is dependent on u. With the aid of this sequence, we were able to deduce the convergence in $L^\infty(\Omega)$ of solutions of (2.3.5) to y. Therefore, h_0 given in Theorem 2.6 depends also on u.

In the context of optimal control problems, we frequently are faced with the situation in which we have to apply error estimates for the solution of (1.1.1) with u being an arbitrary element of a given bounded set K of $L^2(\Gamma)$. In such a situation, an essential assumption in Lemma 2.1 does not hold in general, namely the compactness of the set $G(K)$, where $G : u \longrightarrow y_u$ is the solution operator introduced in Theorem 1.31 on page 26. However, we should remark that Lemma 2.1 was only important to obtain the convergence in $L^\infty(\Omega)$ of solutions of (2.3.5) to y_u for every $u \in K$. Having this converge in mind, we can proceed as follows: we truncate a and f with

$$M = 1 + \sup_{u\in K} \|y_u\|_{C(\bar\Omega)} < \infty$$

(Eq. (1.3.1) on page 3) and we get error estimates in the $H^1(\Omega)$ and $W^{1,q}(\Omega)$ norms, $q \in (2, 4)$, using (2.2.1) instead of (2.2.4), hence replacing ε_h in both estimates by a constant C dependent on K. Then a posteriori the convergence in $W^{1,q}(\Omega)$ implies the uniform convergence of discrete solutions to the solution of the continuous equation. Consequently, the truncation does not play any role for $h < h_0$ which depends now only on K.

The following result extends the estimates obtained so far to the above situation.

Corollary 2.15. *Let $K \subset L^2(\Gamma)$ be bounded. Then there exist $h_0 > 0$ and $C_{K_i} > 0$, $i = 0, 1$, such that, for any $u \in K$ and $h < h_0$, equation (2.3.1) has at least one solution $y_h(u)$ that obeys*

$$\|y_u - y_h(u)\|_{L^2(\Omega)} + h^{1/2}\|y_u - y_h(u)\|_{H^1(\Omega)} \le C_{K_0} h\,, \tag{2.3.37}$$

$$\|y_u - y_h(u)\|_{W^{1,q}(\Omega)} \le C_{K_1} h^{2/q-1/2} \quad \forall q \in (2, 4]\,, \tag{2.3.38}$$

where y_u is the solution of (1.1.1). If Ω is convex then there holds

$$\|y_u - y_h(u)\|_{L^2(\Omega)} + h\|y_u - y_h(u)\|_{H^1(\Omega)} \le C_{K_2} h^{3/2} \tag{2.3.39}$$

instead of (2.3.37). Moreover, if Ω is convex, $p \ge 2$ and $u \in H^{1/2}(\Gamma)$, then

$$\|y_u - y_h(u)\|_{L^2(\Omega)} + h\|y_u - y_h(u)\|_{H^1(\Omega)} \le C_{K_3} h^2\,. \tag{2.3.40}$$

2.4. Uniqueness of solutions of the discrete equation

In this section, we are going to study the uniqueness of solutions of (2.3.1). As far as we know, the issue of uniqueness is an open problem until now. In the sequel, we establish uniqueness results in two different situations. If a is bounded in $\Omega \times \mathbb{R}$ and f is dominated by a $L^p(\Omega)$ function then we are able to prove uniqueness of (2.3.1) provided that h is small enough; see Theorem 2.16. If the above assumptions on a and f are not fulfilled, then we have uniqueness in a more restrictive class of functions; see Theorem 2.18 for a precise formulation of the latter result.

As mentioned at the end of Section 2.2, $u \in L^2(\Omega)$ is fixed and $y := y_u$.

THEOREM 2.16. *Suppose that there exist a constant* $C_\infty > 0$ *and a function* ϕ_∞ *in* $L^p(\Omega)$ *(*$p > 4/3$*) such that*

$$|a(x,t)| \le C_\infty \quad \textit{and} \quad |f(x,t)| \le \phi_\infty(x) \quad \textit{for a.a. } x \in \Omega\,,\ \forall t \in \mathbb{R}\,.$$

Then there exists $h_0 > 0$ *such that (2.3.1) has a unique solution for every* $h < h_0$*.*

PROOF. According to Corollary 2.10, there exists $h_0 > 0$ such that, for any $h < h_0$, (2.3.1) has at least one solution $\hat{y}_h$ with $\hat{y}_h \to y$ in $L^\infty(\Omega)$ as h tends to zero. To show the uniqueness of $\hat{y}_h$ we will argue by contradiction. To this aim, we assume the existence of a sequence $\{h_k\}_{k=1}^\infty$, with $0 < h_k < h_0$ and $h_k \to 0$ as $k \to \infty$, such that (2.3.1) has another solution y_{h_k} with $y_{h_k} \ne \hat{y}_{h_k}$. The function y_{h_k} satisfies for every $\phi_{h_k} \in Y_{h_k}$ the equation

$$\int_\Omega \{a(x,y_{h_k})\nabla y_{h_k}\cdot\nabla\phi_{h_k} + f(x,y_{h_k})\phi_{h_k}\}\,dx = \int_\Gamma u\phi_{h_k}\,d\sigma(x)\,. \tag{2.4.1}$$

By virtue of Theorem 2.4, $\{y_{h_k}\}_{k=1}^\infty$ is bounded in $H^1(\Omega)$, hence we can extract a subsequence, denoted in the same way, such that $y_{h_k} \rightharpoonup \tilde{y}$ weakly in $H^1(\Omega)$ as $k \to \infty$. Let us show that $y = \tilde{y}$. Because of the compactness of the embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ $\forall q \in [1,\infty)$, the convergence y_h to $\tilde{y}$ is strong in $L^q(\Omega)$. Moreover, $y_{h_k}(x) \to \tilde{y}(x)$ for a.a. $x \in \Omega$ and, by the continuity of a and f w.r.t. the second component, $f(x,y_{h_k}(x))$ and $a(x,y_{h_k}(x))$ converge to $f(x,\tilde{y}(x))$ and $a(x,\tilde{y}(x))$, respectively. Furthermore, thanks to the assumptions of the theorem on a and f, we have

$$f(\cdot,y_{h_k}) \to f(\cdot,\tilde{y}) \quad \text{in } L^p(\Omega)\,,$$
$$a(\cdot,y_{h_k})\nabla\phi_{h_k} \to a(\cdot,\tilde{y})\nabla\phi \quad \text{in } L^2(\Omega) \quad \forall\phi,\phi_{h_k} \in H^1(\Omega) \text{ with } \phi_{h_k} \to \phi \text{ in } H^1(\Omega)\,.$$

By taking $\phi \in H^2(\Omega)$ arbitrarily and putting $\phi_{h_k} = \Pi_{h_k}\phi$ in (2.4.1), using the convergence $\Pi_{h_k}\phi \to \phi$ in $H^1(\Omega)$ as $h_k \to 0$ and passing to the limit in (2.4.1), we end up with the equation

$$\int_\Omega \{a(x,\tilde{y})\nabla\tilde{y}\cdot\nabla\phi + f(x,\tilde{y})\phi\}\,dx = \int_\Gamma u\phi\,d\sigma(x)\,.$$

Finally, as a result of the density of $H^2(\Omega)$ in $H^1(\Omega)$ and the uniqueness of the solution y of (1.1.1), we conclude that $y = \tilde{y}$.

Let us now set $z_{h_k} = y_{h_k} - \hat{y}_{h_k}$ and $w_{h_k} = z_{h_k}/\|z_{h_k}\|_{L^6(\Omega)}$. Our next goal is to show that $\{w_{h_k}\}_{k=1}^\infty$ is bounded in $H^1(\Omega)$. Subtracting both equations satisfied by y_{h_k} and $\hat{y}_{h_k}$ and dividing by $\|z_{h_k}\|_{L^6(\Omega)}$, we arrive at

$$\int_\Omega \{a(x,y_{h_k})\nabla w_{h_k}\cdot\nabla\phi_{h_k} + b_{h_k}w_{h_k}\nabla\hat{y}_{h_k}\cdot\nabla\phi_{h_k} + c_{h_k}w_{h_k}\phi_{h_k}\}\,dx = 0 \tag{2.4.2}$$

for all $\phi_{h_k} \in Y_{h_k}$, where

$$b_{h_k}(x) := \frac{a(x, y_{h_k}(x)) - a(x, \hat{y}_{h_k}(x))}{y_{h_k}(x) - \hat{y}_{h_k}(x)} \quad \text{and} \quad c_{h_k}(x) := \frac{f(x, y_{h_k}(x)) - f(x, \hat{y}_{h_k}(x))}{y_{h_k}(x) - \hat{y}_{h_k}(x)}$$

if $y_{h_k}(x) \neq \hat{y}_{h_k}(x)$ and $b_{h_k}(x) := \alpha_a$ and $c_{h_k}(x) := \alpha_f$ otherwise. The boundedness of $\{b_{h_k}\}_{k=1}^{\infty}$ in $L^\infty(\Omega)$ and $\{c_{h_k}\}_{k=1}^{\infty}$ in $L^p(\Omega)$ can be seen in the following way. Define

$$M = \sup_{h_k} \|\hat{y}_{h_k}\|_{L^\infty(\Omega)} \quad \text{and} \quad \Omega_{h_k} = \{x \in \Omega \,|\, |y_{h_k}(x)| \leq M + 1\}.$$

Then by the assumptions of the theorem on a and f, we find

$$|b_{h_k}(x)| \leq C_{a,M+1}, \quad |c_{h_k}(x)| \leq \phi_{M+1}(x) \quad \text{if } x \in \Omega_{h_k},$$

and taking into account that $|y_{h_k}(x) - \hat{y}_{h_k}(x)| \geq 1 \; \forall x \in \Omega \setminus \Omega_{h_k}$,

$$\left.\begin{aligned} |b_{h_k}(x)| &\leq |a(x, y_{h_k}(x)) - a(x, \hat{y}_{h_k}(x))| \leq 2C_\infty, \\ |c_{h_k}(x)| &\leq |f(x, y_{h_k}(x)) - f(x, \hat{y}_{h_k}(x))| \leq 2\phi_\infty(x), \end{aligned}\right\} \text{if } x \in \Omega \setminus \Omega_{h_k}.$$

By taking subsequences if necessary, $b_{h_k} \rightharpoonup b$ weakly* in $L^\infty(\Omega)$ and $c_{h_k} \rightharpoonup c$ weakly in $L^p(\Omega)$. Moreover, c is non-negative and $c(x) \geq \alpha_f \; \forall x \in E$.

From (2.4.2), along with the monotonicity of f and Hölder's inequality, we get

$$\begin{aligned} C_0 \|w_{h_k}\|^2_{H^1(\Omega)} &\leq \int_\Omega \left\{ a(x, y_{h_k}(x)) |\nabla w_{h_k}|^2 + c_{h_k}(x) w_{h_k}^2 \right\} dx \\ &= -\int_\Omega b_{h_k}(x) w_{h_k} \nabla \hat{y}_{h_k} \cdot \nabla w_{h_k} \, dx \\ &\leq C \|w_{h_k}\|_{L^6(\Omega)} \|\hat{y}_{h_k}\|_{W^{1,3}(\Omega)} \|w_{h_k}\|_{H^1(\Omega)}. \end{aligned}$$

Thus, $\|w_{h_k}\|_{H^1(\Omega)} \leq C \|\hat{y}_{h_k}\|_{W^{1,3}(\Omega)}$ which shows the boundedness of $\|w_{h_k}\|_{H^1(\Omega)}$, because $\{\hat{y}_{h_k}\}_{k=1}^{\infty}$ is bounded in $W^{1,3}(\Omega)$; see Corollary 2.10. Hence, taking a subsequence, denoted again by $\{w_{h_k}\}_{k=1}^{\infty}$, there exists $w \in H^1(\Omega)$ such that $w_{h_k} \rightharpoonup w$ weakly in $H^1(\Omega)$ as $k \to \infty$. Once again, due to the compactness of the embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$, the convergence $w_{h_k} \to w$ is strong in $L^6(\Omega)$. Since $\|w_{h_k}\|_{L^6(\Omega)} = 1$, it follows that $\|w\|_{L^6(\Omega)} = 1$, too. Taking in (2.4.2) $\phi_{h_k} = \Pi_{h_k}\phi$, with $\phi \in H^2(\Omega)$ arbitrarily chosen, and passing to the limit, we obtain, along with the strong convergences $\hat{y}_h \to y$ in $W^{1,3}(\Omega)$ (Corollary 2.10), $\Pi_{h_k}\phi \to \phi$ in $H^1(\Omega)$ and the convergence $y_{h_k} \to y$ a.e. in Ω, that

$$\int_\Omega \{a(x, y(x)) \nabla w \cdot \nabla \phi + b(x) w \nabla y \cdot \nabla \phi + c(x) w \phi\} \, dx = 0. \tag{2.4.3}$$

Finally, since $H^2(\Omega)$ is dense in $H^1(\Omega)$, equation (2.4.3) holds for every $\phi \in H^1(\Omega)$. The left-hand side of (2.4.3) equals $\langle Sw, \phi \rangle_{H^1(\Omega)^*, H^1(\Omega)}$, where S is the isomorphism defined on page 20 with $\tilde{a} = a(\cdot, y(\cdot))$ and $\mathbf{b} = b\nabla y$. Consequently, $w = 0$ which contradicts the fact that $\|w\|_{L^6(\Omega)} = 1$. □

PROPOSITION 2.17. *For any $M > \|y\|_{L^\infty(\Omega)}$, there exists $h_M > 0$ such that (2.3.5) has a unique solution for every $h < h_M$.*

Proof. We know already from the last step of the proof of Theorem 2.6 that every solution of (2.3.5) converges to y in $L^\infty(\Omega)$. In the sequel, we will follow the lines of the proof of Theorem 2.16 and assume again the existence of a sequence $\{h_k\}_{k=1}^\infty$ of positive real numbers, with $h_k \to 0$ as $k \to \infty$, such that (2.3.5) has two solutions y_{h_k} and $\hat{y}_{h_k}$ with $y_{h_k} \neq \hat{y}_{h_k}$. Because of the uniform convergence mentioned above, there exists $\tilde{k} \in \mathbb{N}$ such that $\|y_{h_k}\|_{L^\infty(\Omega)} \leq M$ and $\|\hat{y}_{h_k}\|_{L^\infty(\Omega)} \leq M$ $\forall k \geq \tilde{k}$. Thus, for any $k \geq \tilde{k}$ the functions y_{h_k} and $\hat{y}_{h_k}$ satisfy (2.3.1). Setting $z_{h_k} = y_{h_k} - \hat{y}_{h_k}$ and $w_{h_k} = z_{h_k}/\|z_{h_k}\|_{L^6(\Omega)}$, we consider again equation (2.4.2). The boundedness of $\{b_{h_k}\}_{k\geq\tilde{k}}$ in $L^\infty(\Omega)$ and $\{c_{h_k}\}_{k\geq\tilde{k}}$ in $L^p(\Omega)$ are easy to verify. By taking subsequences, it follows that $b_{h_k} \rightharpoonup b$ weakly* in $L^\infty(\Omega)$ and $c_{h_k} \rightharpoonup c$ weakly in $L^p(\Omega)$. Finally, the boundedness of $\{w_{h_k}\}_{k\geq\tilde{k}}$ in $H^1(\Omega)$ and the contradiction is obtained in the same way as in the proof of Theorem 2.16. □

Theorem 2.18. *For every $M > \|y\|_{L^\infty(\Omega)}$ there exists $h_M > 0$ such that the equation (2.3.1) has at most one solution y_h with the property $\|y_h\|_{L^\infty(\Omega)} \leq M$ $\forall h < h_M$. Moreover, if for any $h < h_M$ there is another solution $\tilde{y}_h$ of (2.3.1) such that $\|\tilde{y}_h\|_{L^\infty(\Omega)} > M$ then $\lim_{h\to 0} \|\tilde{y}_h\|_{L^\infty(\Omega)} = +\infty$.*

Proof. Let us assume that y_h and $\tilde{y}_h$ are solutions of (2.3.1) with $\|y_h\|_{L^\infty(\Omega)} \leq M$ and $\|\tilde{y}_h\|_{L^\infty(\Omega)} \leq M$. Then y_h and $\tilde{y}_h$ are also solutions of the equation (2.3.5). According to the preceding proposition, there exists $h_M > 0$ such that $y_h = \tilde{y}_h$ $\forall h < h_M$.

In order to show that $\lim_{h\to 0} \|\tilde{y}_h\|_{L^\infty(\Omega)} = +\infty$ if $\tilde{y}_h$ is a solution of (2.3.1) such that $\|\tilde{y}_h\|_{L^\infty(\Omega)} > M$, we assume the contrary. Then there exists a subsequence $\{\tilde{y}_{h_k}\}_{k=1}^\infty$ bounded in $L^\infty(\Omega)$ with $h_k < h_M$ and $h_k \to 0$ as $k \to +\infty$. For

$$\tilde{M} := \sup_k \|\tilde{y}_{h_k}\|_{L^\infty(\Omega)}$$

the function $\tilde{y}_{h_k}$ satisfies (2.3.5) with $\tilde{M}$ and $\tilde{y}_{h_k}$ instead of M and y_{h_k}, respectively. By virtue of the last step of the proof of Theorem 2.6, $\tilde{y}_{h_k}$ converges to y strongly in $L^\infty(\Omega)$. Thus, $\|\tilde{y}_{h_k}\|_{L^\infty(\Omega)} < M$ for h_k sufficiently small which contradicts the assumption that $\|\tilde{y}_{h_k}\|_{L^\infty(\Omega)} > M$ for all $h_k < h_M$. □

2.5. Numerical experiments

This section is devoted to the verification of the error estimates obtained in Section 2.3 by three numerical test examples. In the first two examples, we consider the Neumann problem

$$\left\{ \begin{aligned} -\Delta y + y &= 0 \quad \text{in } \Omega\,, \\ \partial_\nu y &= u \quad \text{on } \Gamma\,, \end{aligned} \right. \tag{2.5.1}$$

where Ω is convex or not. In both examples, the boundary data u belong to $L^2(\Gamma)$ but $u \notin L^s(\Gamma)$ for any $s > 2$. In the construction of such a function u, we have incorporated the function $g : (0, 1/2) \longrightarrow \mathbb{R}$, $g(t) = 1/(t^{1/2}\log(t))$. Making use of the substitution $z = \log(t)$, we have

$$\int_0^{1/2} \left| \frac{1}{t^{1/2}\log(t)} \right|^s dt = \int_{-\infty}^{-\log(2)} \frac{e^{-\frac{s-2}{2}z}}{|z|^s} dz \begin{cases} < \infty & \text{if } s = 2\,, \\ = \infty & \text{if } s > 2\,. \end{cases}$$

Hence, the function g, and consequently u, too, has exactly the regularity mentioned above. Therefore, the solution y of (2.5.1) is in $H^{3/2}(\Omega)$; see Corollary 1.15 on page 15. Since we do not know y exactly, we take as reference solution the numerical solution y_{fine} of (2.5.1) computed on a very fine mesh whose mesh size is denoted in the sequel by h_{fine}. Hence, instead of studying the behavior of the error $\|y - y_h\|_X$ as h becomes small, we investigate the error $\|y_{fine} - y_h\|_X$ for $X = L^2(\Omega)$ and $X = H^1(\Omega)$. The experimental order of convergence is computed by

$$EOC_X(y_{fine}) := \frac{\log(\|y_{fine} - y_{h_1}\|_X) - \log(\|y_{fine} - y_{h_2}\|_X)}{\log(h_1) - \log(h_2)}$$

for two consecutive mesh sizes h_1 and h_2. To motivate the above formula let us assume that $\|y_{fine} - y_h\|_X$ is of order $\mathcal{O}(h^\sigma)$ which is denoted by $\|y_{fine} - y_h\|_X \sim Ch^\sigma$. Then $\|y_{fine} - y_{h_i}\|_X \sim Ch_i^\sigma$, $i = 1, 2$, hence $\log(\|y_{fine} - y_{h_i}\|_X) \sim \log(C) + \sigma\log(h_i)$ and finally $\log(\|y_{fine} - y_{h_1}\|_X) - \log(\|y_{fine} - y_{h_2}\|_X)/(\log(h_1) - \log(h_2)) \sim \sigma$.

The third example is concerned with a particular quasilinear equation of the form (1.1.1). We follow the same procedure as above and report on both situations, convex and non-convex domains, where we consider the same boundary datum u as in the first and second example, respectively.

Let us briefly describe how we have performed the computations.

To solve problem (2.5.1) numerically we make use of the finite element solver of the MATLAB PDE Toolbox. For the evaluation of the singular integrals involving the function g we have used the adaptive Gauss-Kronrod quadrature `quadgk` of MATLAB.

Concerning the quasilinear equation (1.1.1), a Newton method is implemented. To this aim, we transform the problem of finding y satisfying (1.1.1) into the problem of finding a solution of the equation $F(y) = 0$, with F appropriately chosen. To solve this problem we set up the following iterative method: Given an initial iterate y_0, determine the iterate y_{n+1} by solving the equation $F(y_n) + F'(y_n)(y_{n+1} - y_n) = 0$, $n = 0, 1, 2....$ More precisely, the algorithm is given by the following steps.

Algorithm 2.19. *(Algorithm for (1.1.1))*

(I) Initialization: Choose an initial datum $y_0 \in H^{3/2}(\Omega)$ *and set* $n = 0$.

(II) Compute the solution y *of the linearized equation*

$$\int_\Omega \left\{ a(x, y_n)\nabla y \cdot \nabla\phi + \frac{\partial a}{\partial y}(x, y_n) y \nabla y_n \cdot \nabla\phi + \frac{\partial f}{\partial y}(x, y_n) y\phi \right\} dx = F_n(\phi) \tag{2.5.2}$$

$\forall \phi \in H^1(\Omega)$, where

$$F_n(\phi) := \int_\Omega \left\{ \left(\frac{\partial f}{\partial y}(x, y_n) y_n - f(x, y_n) \right) \phi + \frac{\partial a}{\partial y}(x, y_n) y_n \nabla y_n \cdot \nabla \phi \right\} dx + \int_\Gamma u \phi \, d\sigma(x) \, .$$

(III) Stop if $\|y - y_n\|_{L^\infty(\Omega)} < TOL$ or set $y_{n+1} = y$, $n = n + 1$ and go to (II).

For our purposes, we have chosen $TOL = 10^{-8}$ and $y_0 \equiv 1$. Again, we have taken the finite element solver of the MATLAB PDE Toolbox to deal with the linear equation (2.5.2). This solver had to be extended, since non-standard expressions are present such as the term $\int_\Omega \frac{\partial a}{\partial y}(x, y_n) y \nabla y_n \cdot \nabla \phi \, dx$. A study of the previous algorithm as well as convergence issues are not discussed here, since these would go beyond the scope of this thesis.

EXAMPLE 2.20. *We fix $\Omega = (0,1)^2$ and denote by Γ_1 to Γ_4 the four sides of the square, starting at the bottom side and moving counterclockwise. We construct the boundary datum as follows. Let $c_i : (0,1) \longrightarrow \Gamma_i$ be the parametrization of Γ_i, $i = 1, ..., 4$, with $c_1(t) = \{t\} \times \{0\}$ and the remaining c_i are defined analogously, taking into account the aforementioned running direction on Γ. Then we define*

$$u(c_i(t)) = \frac{1}{t^{1/2} \log(t)} \, , \; t \in (0, 0.5] \, , \; \text{and} \; u(c_i(1-t)) = \frac{1}{t^{1/2} \log(t)} \, , \; t \in (0, 0.5) \, ,$$

i.e. on every boundary side Γ_i, u has a peak shape concentrated at both endpoints of Γ_i.

Figure 2.1 shows u and y_h for $h = 2^{-7}$. Table 2.1 below illustrates the distance between y_h and the reference solution y_{fine} corresponding to the mesh size $h_{ref} = 2^{-10}$ as well as the convergence speed represented by $EOC_X(y_{fine})$ for $X = L^2(\Omega)$ and $X = H^1(\Omega)$. We point out that h_{ref} is in the range of mesh sizes associated with the finest meshes we have been able to manage with a PC with MATLAB. Apparently,

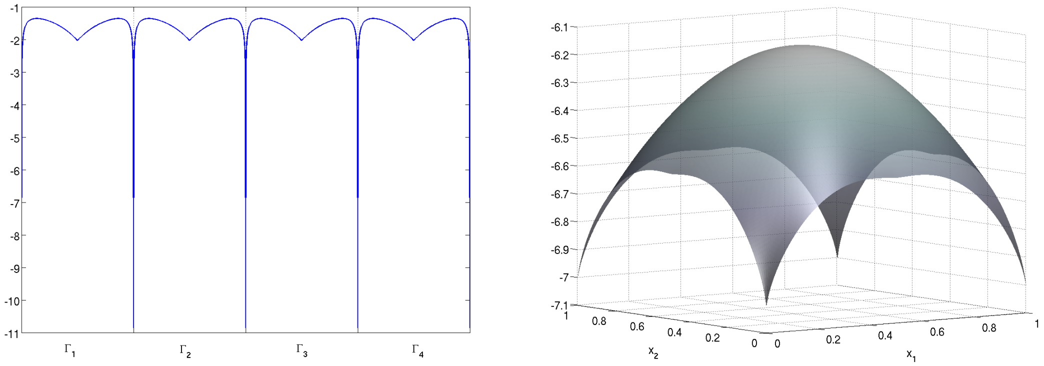

FIGURE 2.1. Data from Example 2.20; u (left frame) and y_h with $h = 2^{-7}$ (right frame).

h	$\|y_{fine} - y_h\|_{L^2(\Omega)}$	$EOC_{L^2(\Omega)}(y_{fine})$	$\|y_{fine} - y_h\|_{H^1(\Omega)}$	$EOC_{H^1(\Omega)}(y_{fine})$
2^{-3}	$2.0961e-03$	-	$1.4545e-01$	-
2^{-4}	$6.2538e-04$	1.7449	$8.7239e-02$	0.7375
2^{-5}	$1.9148e-04$	1.7076	$5.4156e-02$	0.6879
2^{-6}	$6.0283e-05$	1.6673	$3.4438e-02$	0.6531
2^{-7}	$1.9365e-05$	1.6383	$2.2154e-02$	0.6364
2^{-8}	$6.2774e-06$	1.6252	$1.4199e-02$	0.6418
2^{-9}	$2.0136e-06$	1.6404	$8.8210e-03$	0.6868

Table 2.1. Convergence behavior of $\|y_{fine}-y_h\|_X$ and experimental order of convergence $EOC_X(y_{fine})$ for $X = L^2(\Omega)$ and $X = H^1(\Omega)$.

the order of convergence in the $L^2(\Omega)$ and $H^1(\Omega)$ norms are better than $\mathcal{O}(h^{3/2})$ and $\mathcal{O}(h^{1/2})$, respectively, as expected from (2.3.24). However, this is not surprising, since our theoretical results exhibit a worst-case scenario which is hardly to realize numerically. Nevertheless, it is obvious that the convergence rates in the $L^2(\Omega)$ and $H^1(\Omega)$ norms are worse than $\mathcal{O}(h^2)$ and $\mathcal{O}(h)$ as predicted in Corollary 2.14 in the case of regular boundary data. This indicates that $y \notin H^2(\Omega)$.

Example 2.21. *Let Ω be the L-shaped domain $\Omega = (0,1) \times (0,0.5) \cup (0.5,1) \times (0,1)$. Analogous to Example 2.20, we denote by Γ_1 to Γ_6 the six boundary sides of Ω, starting at the bottom side and moving counterclockwise. Now define the mappings $c_i : (0,1) \longrightarrow \Gamma_i$, $i = 1,2$, and $c_j : (0,0.5) \longrightarrow \Gamma_j$, $j = 3,...,6$, by $c_1(t) = \{t\} \times \{0\}$, $c_2(t) = \{1\} \times \{t\}$, $c_3(t) = \{1-t\} \times \{1\}$, $c_4(t) = \{0.5\} \times \{1-t\}$, $c_5(t) = \{0.5-t\} \times \{0.5\}$ and $c_6(t) = \{0\} \times \{0.5-t\}$. We set*

$$u(c_1(t)) = -\frac{1}{t^{1/2}\log(t)},\ t \in (0,0.5]\,, \quad \textit{and} \ \ u(c_1(1-t)) = \frac{1}{t^{1/2}\log(t)},\ t \in (0,0.5)\,,$$

$$u(c_2(t)) = \frac{1}{t^{1/2}\log(t)},\ t \in (0,0.5]\,, \quad \textit{and} \ \ u(c_2(1-t)) = -\frac{1}{t^{1/2}\log(t)},\ t \in (0,0.5)\,,$$

and

$$u(c_j(t)) = -\frac{1}{(0.5-t)^{1/2}\log(0.5-t)} - \frac{1}{t^{1/2}\log(t)}\,,\ t \in (0,0.5)\,,\ j = 3,...,6\,.$$

Certainly, on every boundary side Γ_i, u has a peak shape concentrated at both endpoints of Γ_i.

The function u and the numerical solution y_h of (2.5.1) for $h = 2^{-7}$ are shown in Figure 2.2. In Table 2.2 we report on the convergence history for this test example. Again, it can be seen that the numerical computations provide a better convergence

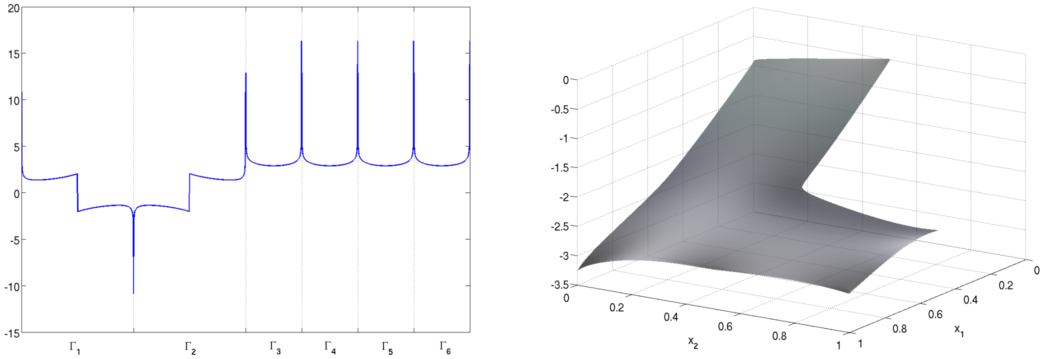

FIGURE 2.2. Data from Example 2.21; u (left frame) and y_h with $h = 2^{-7}$ (right frame).

h	$\|y_{fine} - y_h\|_{L^2(\Omega)}$	$EOC_{L^2(\Omega)}(y_{fine})$	$\|y_{fine} - y_h\|_{H^1(\Omega)}$	$EOC_{H^1(\Omega)}(y_{fine})$
2^{-3}	$7.8942e-03$	-	$1.6861e-01$	-
2^{-4}	$3.2540e-03$	1.2786	$1.0375e-01$	0.7006
2^{-5}	$1.4021e-03$	1.2146	$6.7363e-02$	0.6231
2^{-6}	$6.1544e-04$	1.1880	$4.5273e-02$	0.5733
2^{-7}	$2.6922e-04$	1.1928	$3.0771e-02$	0.5571
2^{-8}	$1.1449e-04$	1.2336	$2.0711e-02$	0.5712
2^{-9}	$4.5043e-05$	1.3458	$1.3401e-02$	0.6281

TABLE 2.2. Convergence behavior of $\|y_{fine}-y_h\|_X$ and experimental order of convergence $EOC_X(y_{fine})$ for $X = L^2(\Omega)$ and $X = H^1(\Omega)$.

behavior than our prediction for non-convex domains; theoretically, we expect to see the orders $\mathcal{O}(h)$ and $\mathcal{O}(h^{1/2})$ in the $L^2(\Omega)$ and $H^1(\Omega)$ norms, respectively (Eq. (2.3.4)). However, compared to the first example, it is obvious that the non-convexity of L-shaped domain has indeed an effect on the order of convergence of the solutions in the $L^2(\Omega)$ norm, thus confirming our theoretical investigations in Section 2.3.

EXAMPLE 2.22. *We consider the quasilinear equation*

$$\left\{ \begin{aligned} -\operatorname{div}\left[(1+(x_1+x_2)^2+y^2(x))\nabla y(x)\right]+y+y^3(x) &= 0 && \textit{in } \Omega\,, \\ (1+(x_1+x_2)^2+y^2(x))\partial_\nu y(x) &= u(x) && \textit{on } \Gamma\,, \end{aligned} \right. \tag{2.5.3}$$

where Ω and u are first given as in Example 2.20 and next as in Example 2.21.

Figure 2.3 shows the numerical solution y_h for $h = 2^{-7}$ in both cases. Tables 2.3 and 2.4 give a detailed insight into the behavior of $\|y_{fine} - y_h\|_{L^2(\Omega)}$ and $\|y_{fine} - y_h\|_{H^1(\Omega)}$

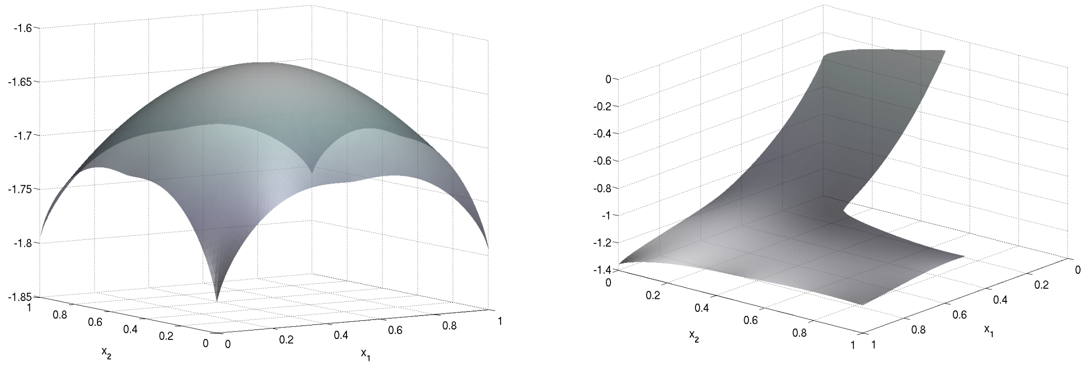

FIGURE 2.3. Data from Example 2.22; y_h with $h = 2^{-7}$ when Ω is convex (left frame) and non-convex (right frame).

for $h = 2^{-i}$, $i = 2, ..., 8$. Now y_{fine} is the solution of the discrete version of (2.5.3)

h	$\lVert y_{fine} - y_h \rVert_{L^2(\Omega)}$	$EOC_{L^2(\Omega)}(y_{fine})$	$\lVert y_{fine} - y_h \rVert_{H^1(\Omega)}$	$EOC_{H^1(\Omega)}(y_{fine})$
2^{-2}	$1.4061e-03$	-	$4.9945e-02$	-
2^{-3}	$4.1403e-04$	1.7639	$2.8409e-02$	0.8140
2^{-4}	$1.2146e-04$	1.7693	$1.6810e-02$	0.7570
2^{-5}	$3.6688e-05$	1.7271	$1.0309e-02$	0.7054
2^{-6}	$1.1435e-05$	1.6819	$6.4740e-03$	0.6712
2^{-7}	$3.6351e-06$	1.6534	$4.0879e-03$	0.6633
2^{-8}	$1.1521e-06$	1.6578	$2.5168e-03$	0.6998

TABLE 2.3. Convergence behavior of $\lVert y_{fine} - y_h \rVert_X$ and experimental order of convergence $EOC_X(y_{fine})$ for convex Ω, $X = L^2(\Omega)$ and $X = H^1(\Omega)$.

corresponding to the mesh with mesh size $h_{ref} = 2^{-9}$. In order to accelerate the convergence of the Newton method for computing y_{h_i} with $h_i = 2^{-i}$, $i = 3, ..., 8$, we have taken as initial iterate y_0 the solution $y_{h_{i-1}}$, since the latter is close to the solution of the continuous equation (2.5.3). A close look at the Tables 2.3 and 2.4 shows that the convergence behavior in both cases, when Ω is convex or not, is the same as that observed in Example 2.20 and Example 2.21, respectively.

2.6. The discrete solution operator

In this section, we establish some important properties of the discrete solution operator $u \longmapsto y_h(u)$. Since this mapping is possibly multivalued, our analysis has only local character. However, we are able to prove that for a fixed datum $\bar{u} \in L^2(\Gamma)$

h	$\|y_{fine} - y_h\|_{L^2(\Omega)}$	$EOC_{L^2(\Omega)}(y_{fine})$	$\|y_{fine} - y_h\|_{H^1(\Omega)}$	$EOC_{H^1(\Omega)}(y_{fine})$
2^{-2}	$7.7661e-03$	-	$1.7221e-01$	-
2^{-3}	$3.0020e-03$	1.3712	$9.4984e-02$	0.8584
2^{-4}	$1.2527e-03$	1.2609	$5.3957e-02$	0.8159
2^{-5}	$5.4571e-04$	1.1989	$3.2124e-02$	0.7482
2^{-6}	$2.3880e-04$	1.1923	$1.9965e-02$	0.6862
2^{-7}	$1.0175e-04$	1.2307	$1.2673e-02$	0.6557
2^{-8}	$4.0109e-05$	1.3431	$7.9089e-03$	0.6802

TABLE 2.4. Convergence behavior of $\|y_{fine}-y_h\|_X$ and experimental order of convergence $EOC_X(y_{fine})$ for non-convex Ω, $X = L^2(\Omega)$ and $X = H^1(\Omega)$.

there exists a $L^2(\Gamma)$ ball centered at $\bar{u}$ such that, for any element u of it, there exists a unique discrete solution $y_h(u)$ in a certain $W^{1,q}(\Omega)$ ball, $q \in (2, 4)$, centered at $y_{\bar{u}}$.

Throughout this section, we suppose that the Assumptions 1.1-1.3, 1.17 and 1.24-(1) hold. Further, we fix $\bar{u} \in L^2(\Gamma)$ and $\bar{y} := y_{\bar{u}} \in H^{3/2}(\Omega)$; see Theorem 1.18 on page 16.

THEOREM 2.23. *Let $\bar{q} \in (2, 4)$ be fixed. There exist $h_0 > 0$ and constants $\rho > 0$ and $\kappa_\rho > 0$, dependent on h_0 and $\bar{q}$, such that, for any $h < h_0$ and any $u \in \overline{B}_{L^2(\Gamma)}(\bar{u}, \rho)$, the discrete quasilinear equation (2.3.1) has a unique solution $y_h(u) \in Y_h$ in the closed ball $\overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y}, \kappa_\rho)$.*

PROOF. Let us first assume $\rho = 1$. A smaller radius will be introduced later. Applying Corollary 2.15, we deduce the existence of $h_0 > 0$ such that, for any $\|\bar{u} - u\|_{L^2(\Gamma)} \leq \rho$ and $h < h_0$, there exists a solution $y_h(u) \in Y_h$ of (2.3.1) verifying the estimate

$$\|y_u - y_h(u)\|_{W^{1,\bar{q}}(\Omega)} \leq C_1 h^{2/\bar{q}-1/2} \leq C_1 h_0^{2/\bar{q}-1/2}.$$

Then, due to the Lipschitz continuity of G (Corollary 1.32 on page 27), there holds

$$\begin{aligned}\|\bar{y} - y_h(u)\|_{W^{1,\bar{q}}(\Omega)} &\leq \|\bar{y} - y_u\|_{W^{1,\bar{q}}(\Omega)} + \|y_u - y_h(u)\|_{W^{1,\bar{q}}(\Omega)} \\ &\leq C_2\|\bar{u} - u\|_{L^2(\Gamma)} + C_1 h_0^{2/\bar{q}-1/2}.\end{aligned}$$

Thus, taking $\hat{C} = \max\{C_1, C_2\}$ and $\kappa_\rho = \hat{C}\left(\rho + h_0^{2/\bar{q}-1/2}\right)$, there exists at least one element $y_h(u) \in \overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y}, \kappa_\rho)$ for any $u \in \overline{B}_{L^2(\Gamma)}(\bar{u}, \rho)$ and $h < h_0$.

In the rest of the proof, we show the existence of a number ρ such that $y_h(u)$ is the unique solution of (2.3.1) in the ball $\overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y}, \kappa_\rho)$ for any $u \in \overline{B}_{L^2(\Gamma)}(\bar{u}, \rho)$. To this end, we will argue by contradiction. Let us assume that there exists a sequence $\{h_k\}_{k=1}^\infty$ of positive real numbers with $h_k \to 0$ as $k \to \infty$, a sequence $\{u_{h_k}\}_{k=1}^\infty$ in

$L^2(\Gamma)$ with $\|u_{h_k}-\bar{u}\|_{L^2(\Gamma)} < \frac{1}{k}$ and functions $y^1_{h_k}(u_{h_k}), y^2_{h_k}(u_{h_k}) \in \overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y},\eta_k)$ with $\eta_k = \hat{C}\left(\frac{1}{k} + h_k^{2/\bar{q}-1/2}\right)$ and $y^1_{h_k}(u_{h_k}) \neq y^2_{h_k}(u_{h_k})$, such that for any $\phi_{h_k} \in Y_{h_k}$, $i=1,2$,

$$\int_\Omega \left\{a(x,y^i_{h_k}(u_{h_k}))\nabla y^i_{h_k}(u_{h_k})\cdot\nabla\phi_{h_k} + f(x,y^i_{h_k}(u_{h_k}))\phi_{h_k}\right\}dx = \int_\Gamma u_{h_k}\phi_{h_k}\,d\sigma(x)\,.$$

To simplify the notation we set $y^i_k = y^i_{h_k}(u_{h_k})$. Notice that the existence of y^i_k with the property $y^i_k \to \bar{y}$ in $W^{1,\bar{q}}(\Omega)$ is a consequence of Corollary 2.15. Further, define

$$y_k = \frac{y^2_k - y^1_k}{\|y^2_k - y^1_k\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)}}\,.$$

We show first that $\{y_k\}_{k=1}^\infty$ is bounded in $H^1(\Omega)$. To this end, we subtract both equations satisfied by y^2_k and y^1_k, divide by $\|y^2_k - y^1_k\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)}$ and apply the mean value theorem, to get the existence of measurable functions θ_{h_k} and η_{h_k} with values in $[0,1]$ such that

$$\text{(2.6.1)}\quad \int_\Omega\left\{a(x,y^1_k)\nabla y_k\cdot\nabla\phi_{h_k} + \frac{\partial a}{\partial y}(x,v_{h_k})y_k\nabla y^2_k\cdot\nabla\phi_{h_k} + \frac{\partial f}{\partial y}(x,w_{h_k})y_k\phi_{h_k}\right\}dx = 0\,,$$

where $v_{h_k} := y^1_k + \theta_{h_k}(y^2_k - y^1_k)$ and $w_{h_k} := y^1_k + \eta_{h_k}(y^2_k - y^1_k)$. Notice that the measurability of θ_{h_k} and η_{h_k} can be shown by applying [**53**, Theorem 1.2 on page 236 and Proposition 1.1 on page 234] to the positive functions

$$\begin{aligned}&g_1 : \bar{\Omega}\times[0,1]\to\mathbb{R}\,,\\ &g_1(x,t) = \left|a(x,y^2_k(x)) - a(x,y^1_k(x)) - \frac{\partial a}{\partial y}(x,y^1_k(x) + t(y^2_k(x) - y^1_k(x)))\right|\,,\end{aligned}$$

and

$$\begin{aligned}&g_2 : \Omega\times[0,1]\to\mathbb{R}\,,\\ &g_2(x,t) = \left|f(x,y^2_k(x)) - f(x,y^1_k(x)) - \frac{\partial f}{\partial y}(x,y^1_k(x) + t(y^2_k(x) - y^1_k(x)))\right|\,,\end{aligned}$$

respectively. Exploiting the uniform boundedness of y^i_k, v_{h_k} and w_{h_k}, and the assumptions on a and f, we get from (2.6.1)

$$\begin{aligned}C_0\|y_k\|^2_{H^1(\Omega)} &\le \alpha_a\|\nabla y_k\|^2_{L^2(\Omega)} + \alpha_f\|y_k\|^2_{L^2(\Omega)}\\ &\le C\int_\Omega |y_k||\nabla y^2_k||\nabla y_k|\,dx\\ &\le C\|y_k\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)}\|\nabla y^2_k(u_{h_k})\|_{L^{\bar{q}}(\Omega)}\|y_k\|_{H^1(\Omega)}\,.\end{aligned}$$

Hence, the boundedness of $\|y_k\|_{H^1(\Omega)}$ is an immediate consequence of the boundedness of y^2_k in $W^{1,\bar{q}}(\Omega)$ and the identity $\|y_k\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)} = 1$. Taking a subsequence, denoted in the same way, there exists $\hat{y}\in H^1(\Omega)$ such that $y_k \rightharpoonup \hat{y}$ weakly in $H^1(\Omega)$ as $k\to\infty$. Due to the compactness of the embedding $H^1(\Omega)\hookrightarrow L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)$, the convergence y_k

to $\hat{y}$ is strong in $L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)$ and, because of $\|y_k\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)} = 1$, we have $\|\hat{y}\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)} = 1$. Using similar arguments as on page 54, we pass to the limit in (2.6.1) and arrive at

$$\int_\Omega \left\{ a(x,\bar{y})\nabla\hat{y}\cdot\nabla\phi + \frac{\partial a}{\partial y}(x,\bar{y})\hat{y}\nabla\bar{y}\cdot\nabla\phi + \frac{\partial f}{\partial y}(x,\bar{y})\hat{y}\phi \right\} dx = 0 \tag{2.6.2}$$

$\forall \phi \in H^1(\Omega)$. According to Theorem 1.25 on page 22, (2.6.2) implies that $\hat{y} = 0$ which contradicts the fact that $\|\hat{y}\|_{L^{\frac{2\bar{q}}{\bar{q}-2}}(\Omega)} = 1$. □

Remark 2.24. *Exploiting the arguments on page 61, it is easy to see that, given $h_0 > 0$ and $\bar{q} \in (2,4)$, there exist $\rho > 0$ and $\kappa_\rho > 0$ such that, for any $h < h_0$ and any $u \in \overline{B}_{L^2(\Gamma)}(\bar{u},\rho)$, the equation (2.3.1) has a unique solution $y_h(u) \in Y_h$ in the closed ball $\overline{B}_{W^{1,q}(\Omega)}(\bar{y},\kappa_\rho)$ $\forall q \in [\bar{q},4)$. In particular $\rho > 0$ and $\kappa_\rho > 0$ are the same for all $q \in [\bar{q},4)$.*

As a result of Theorem 2.23 we may define the discrete counterpart of the solution operator G, namely

$$G_h : B_{L^2(\Gamma)}(\bar{u},\rho) \longrightarrow \overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y},\kappa_\rho) \cap Y_h\,, \quad G_h(u) = y_h(u)\,,$$

where $\bar{q} \in (2,4)$ and $y_h(u)$ is the unique solution of (2.3.1) in $\overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y},\kappa_\rho)$. Our next goal is to study the differentiability of G_h.

Theorem 2.25. *There exists $h_1 \le h_0$ such that, for any $h < h_1$, the mapping $G_h : B_{L^2(\Gamma)}(\bar{u},\rho) \longrightarrow \overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y},\kappa_\rho) \cap Y_h$, $u \longmapsto y_h(u)$, is of class C^2. If we denote $z_h(v) = G_h'(u)v$, with $v \in L^2(\Gamma)$, then $z_h(v)$ is the unique solution of the problem*

$$\begin{cases} \text{Find } z_h(v) \in Y_h \text{ such that, for all } \phi_h \in Y_h, \\ \displaystyle\int_\Omega \left\{ a(x,y_h(u))\nabla z_h(v)\cdot\nabla\phi_h + \frac{\partial a}{\partial y}(x,y_h(u))z_h(v)\nabla y_h(u)\cdot\nabla\phi_h \right. \\ \displaystyle\qquad\qquad \left. + \frac{\partial f}{\partial y}(x,y_h(u))z_h(v)\phi_h \right\} dx = \int_\Gamma v\phi_h\, d\sigma(x)\,. \end{cases} \tag{2.6.3}$$

Proof. To achieve the regularity of G_h stated in the theorem we will use the implicit function theorem by considering $F_h : L^2(\Gamma) \times Y_h \longrightarrow Y_h^*$,

$$\langle F_h(u,y_h),\phi_h\rangle = \int_\Omega \{a(x,y_h)\nabla y_h\cdot\nabla\phi_h + f(x,y_h)\phi_h\}\, dx - \int_\Gamma u\phi_h\, d\sigma(x)\,.$$

It is obvious that F_h is of class C^2 in $B_{L^2(\Gamma)}(\bar{u},\rho) \times Y_h$ and $F_h(u,y_h(u)) = 0$ for every $u \in B_{L^2(\Gamma)}(\bar{u},\rho)$ and $h < h_0$. Therefore, it remains to prove that the mapping

$(\partial F_h/\partial y_h)(u, y_h(u)) : Y_h \longrightarrow Y_h^*$ defined by

$$\left\langle \frac{\partial F_h}{\partial y_h}(u, y_h(u))z_h, \phi_h \right\rangle =$$
$$= \int_\Omega \left\{ a(x, y_h(u))\nabla z_h \cdot \nabla \phi_h + \frac{\partial a}{\partial y}(x, y_h(u))z_h \nabla y_h(u) \cdot \nabla \phi_h + \frac{\partial f}{\partial y}(x, y_h(u))z_h \phi_h \right\} dx$$

is an isomorphism. The representation (2.6.3) of G_h' is obtained by a simple computation.

The mapping $(\partial F_h/\partial y_h)(u, y_h(u))$ is linear and Y_h is finite dimensional, thus it suffices to prove its injectivity or equivalently that the equation $(\partial F_h/\partial y_h)(u, y_h(u))z_h = 0$ admits only the solution $z_h = 0$. For this purpose, we will follow the approach proposed by Schatz [**86**]. In order to shorten the notation, we denote $y_h = y_h(u)$. Introduce the function $B_h : H^1(\Omega) \times H^1(\Omega) \longrightarrow \mathbb{R}$ defined by

$$B_h(w, \phi) = \int_\Omega \left\{ a(x, y_h)\nabla w \cdot \nabla \phi + \frac{\partial a}{\partial y}(x, y_h) w \nabla y_h \cdot \nabla \phi + \frac{\partial f}{\partial y}(x, y_h) w \phi \right\} dx\,. \tag{2.6.4}$$

Since $\{y_h\}_{h<h_0}$ is bounded in $W^{1,4}(\Omega) \subset C(\bar{\Omega})$ (Eq. (2.3.38)), we get from the Assumptions 1.2 and 1.24-(1) that

$$|B_h(w, \phi)| \leq C\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)} \quad \forall w, \phi \in H^1(\Omega)\,. \tag{2.6.5}$$

We divide the remaining part of the proof of $z_h = 0$ in three steps.

Step 1: Proof of a Gårding's inequality. We prove that there exist constants C_1 and C_2 with $C_1 > 0$ such that

$$|B_h(w, w)| \geq C_1\|w\|^2_{H^1(\Omega)} - C_2\|w\|^2_{L^2(\Omega)} \quad \forall w \in H^1(\Omega)\,. \tag{2.6.6}$$

If $C_2 \leq 0$, thanks to (2.6.5)-(2.6.6), an application of the Lax-Milgram theorem yields that the unique solution z_h of $(\partial F_h/\partial y_h)(u, y_h(u))z_h = 0$ is zero. Let us now study the more general case when $C_2 > 0$. Using the assumptions on a and f, we have

$$\begin{aligned} |B_h(w, w)| &= \left| \int_\Omega \left\{ a(x, y_h)|\nabla w|^2 + \frac{\partial a}{\partial y}(x, y_h) w \nabla w \cdot \nabla y_h + \frac{\partial f}{\partial y}(x, y_h) w^2 \right\} dx \right| \\ &\geq \min\{\alpha_a, \alpha_f\}\|w\|^2_{H^1(\Omega)} - \tilde{C} \int_\Omega |w \nabla w \cdot \nabla y_h|\, dx \\ &\geq \min\{\alpha_a, \alpha_f\}\|w\|^2_{H^1(\Omega)} - \tilde{C}\|w\|_{L^4(\Omega)}\|w\|_{H^1(\Omega)}\|y_h\|_{W^{1,4}(\Omega)}\,. \end{aligned} \tag{2.6.7}$$

Further, Lemma 2.3 and Young's inequality with $q = q' = 2$ imply

$$\|w\|_{L^4(\Omega)} \leq C_3\|w\|^{1/2}_{L^2(\Omega)}\|w\|^{1/2}_{H^1(\Omega)} \leq C_3 \left(\frac{1}{2\varepsilon^2}\|w\|_{L^2(\Omega)} + \frac{\varepsilon^2}{2}\|w\|_{H^1(\Omega)} \right)$$

for any $\varepsilon > 0$. Combining the last inequality with (2.6.7), it follows

$$(2.6.8)\quad |B_h(w,w)| \geq \min\{\alpha_a, \alpha_f\}\|w\|^2_{H^1(\Omega)} - C_4\left(\frac{1}{2\varepsilon^2}\|w\|_{L^2(\Omega)} + \frac{\varepsilon^2}{2}\|w\|_{H^1(\Omega)}\right)\|w\|_{H^1(\Omega)}\,.$$

Again, by Young's inequality, there holds

$$\|w\|_{L^2(\Omega)}\|w\|_{H^1(\Omega)} \leq \frac{1}{4\sigma}\|w\|^2_{L^2(\Omega)} + \sigma\|w\|^2_{H^1(\Omega)} \quad \forall \sigma > 0$$

which, along with (2.6.8), leads to (2.6.6) by setting $C_1 = \min\{\alpha_a, \alpha_f\} - C_4(\frac{\varepsilon^2}{2} + \frac{\sigma}{2\varepsilon^2})$, $C_2 = \frac{C_4}{8\sigma\varepsilon^2}$, taking $\sigma = \varepsilon^3$ and ε small enough such that $C_1 > 0$.

Step 2: $z_h = 0$. By the definition of B_h and inequality (2.6.6), we have

$$C_1\|z_h\|^2_{H^1(\Omega)} - C_2\|z_h\|^2_{L^2(\Omega)} \leq |B_h(z_h, z_h)| = \left\langle \frac{\partial F_h}{\partial y_h}(u, y_h(u))z_h, z_h \right\rangle = 0\,,$$

hence

$$(2.6.9)\qquad \|z_h\|_{H^1(\Omega)} \leq \left(\frac{C_2}{C_1}\right)^{1/2}\|z_h\|_{L^2(\Omega)}\,.$$

The last step of the proof consists in verifying the following result

$$(2.6.10)\qquad \exists \hat{h}_1 > 0 \text{ such that } \|z_h\|_{L^2(\Omega)} \leq C_5 h^{1/2}\|z_h\|_{H^1(\Omega)} \quad \forall h < \hat{h}_1\,,$$

where $C_5 > 0$ is independent of z_h. Once (2.6.10) is shown, taking

$$h_1 = \min\left\{\hat{h}_1, h_0, \frac{C_1}{C_2 C_5^2}\right\},$$

we deduce from (2.6.9) and (2.6.10) that $z_h = 0$ for all $h < h_1$.

Step 3: Proof of (2.6.10). Let us denote $y = y_u$ and by φ the unique solution in $H^{3/2}(\Omega)$ of the equation

$$(2.6.11)\qquad \begin{cases} -\text{div}\,[a(x,y)\nabla\varphi] + \dfrac{\partial a}{\partial y}(x,y)\nabla y\cdot\nabla\varphi + \dfrac{\partial f}{\partial y}(x,y)\varphi = z_h & \text{in } \Omega\,, \\ a(x,y)\partial_\nu\varphi = 0 & \text{on } \Gamma\,. \end{cases}$$

Let us also define the function $B : H^1(\Omega) \times H^1(\Omega) \longrightarrow \mathbb{R}$ by

$$(2.6.12)\qquad B(w,\phi) = \int_\Omega \left\{a(x,y)\nabla w\cdot\nabla\phi + \frac{\partial a}{\partial y}(x,y)w\nabla y\cdot\nabla\phi + \frac{\partial f}{\partial y}(x,y)\phi w\right\}dx\,.$$

Passing to the weak formulation of equation (2.6.11), along with (2.3.37), (2.2.1), (2.2.2) and the fact that $B_h(z_h, \Pi_h\varphi) = 0$, we find

$$\begin{aligned}\|z_h\|^2_{L^2(\Omega)} &= B(z_h, \varphi)\\ &= B(z_h, \varphi - \Pi_h\varphi) + B(z_h, \Pi_h\varphi)\\ &= B(z_h, \varphi - \Pi_h\varphi) + (B(z_h, \Pi_h\varphi) - B_h(z_h, \Pi_h\varphi))\\ &\le C\left(\|\varphi - \Pi_h\varphi\|_{H^1(\Omega)} + \|y - y_h\|_{H^1(\Omega)}\|\Pi_h\varphi\|_{W^{1,4}(\Omega)}\right)\|z_h\|_{H^1(\Omega)}\\ &\le Ch^{1/2}\|\varphi\|_{H^{3/2}(\Omega)}\|z_h\|_{H^1(\Omega)}\,. \qquad (2.6.13)\end{aligned}$$

Then (2.6.10) follows from (2.6.13) and inequality $\|\varphi\|_{H^{3/2}(\Omega)} \le C\|z_h\|_{L^2(\Omega)}$ (Eq. (1.7.3) on page 30). □

Proposition 2.26. *Let $U \subset B_{L^2(\Gamma)}(\bar{u}, \rho)$ be bounded. Then there exist $C_U > 0$ and $h_2 \le h_1$ such that for any $h < h_2$, $u \in U$ and $v \in L^2(\Gamma)$, there holds*

$$\|G'_h(u)v\|_{W^{1,4}(\Omega)} \le C_U\|v\|_{L^2(\Gamma)}\,. \qquad (2.6.14)$$

Proof. Let us introduce the following auxiliary problem of finding $z \in H^1(\Omega)$ such that

$$\begin{aligned}&\int_\Omega\left\{a(x, y_h)\nabla z\cdot\nabla\phi + \frac{\partial f}{\partial y}(x, y_h)z\phi\right\}dx\\ &\qquad = \int_\Gamma v\phi\,d\sigma(x) - \int_\Omega \frac{\partial a}{\partial y}(x, y_h)z_h(v)\nabla y_h\cdot\nabla\phi\,dx =: F(\phi) \quad \forall\phi \in H^1(\Omega)\,.\end{aligned}$$

From Theorem 2.25 we know that $\{z_h(v)\}_{h<h_1}$ is bounded in $W^{1,\bar{q}}(\Omega) \subset C(\bar{\Omega})$ for $\bar{q} \in (2, 4)$. This fact, along with the boundedness of $\{y_h\}_{h<h_0}$ in $W^{1,4}(\Omega)$, implies that $F \in W^{1,4/3}(\Omega)^*$. Hence, $z \in W^{1,4}(\Omega)$, as follows from Theorem 1.9 on page 9. Now we apply a well-known estimate by Brenner and Scott [**11**, p. 171] which is also valid in the case of Neumann boundary conditions and leads to the estimate (2.6.14): there exist a constant $C > 0$ and $0 < h_2 \le h_1$ such that for any $h < h_2$

$$\|z_h(v)\|_{W^{1,4}(\Omega)} \le C\|z\|_{W^{1,4}(\Omega)}\,.$$

We should remark that the linear elliptic operators in the equations satisfied by z and $z_h(v)$ depend on y_h and consequently on h. However, we may use the result by Brenner and Scott [**11**, p. 171] due to the following reason. The previous inequality holds true for $y_{\tilde{h}}$ instead of y_h, with $\tilde{h} < h_1$ arbitrary but fixed and C may depend on the norm of $y_{\tilde{h}}$ in $W^{1,4}(\Omega)$. Thanks to the boundedness of $\{y_h\}_{h<h_1}$ in $W^{1,4}(\Omega)$, this inequality remains valid, in particular, when taking $h = \tilde{h}$. □

2.7. Numerical analysis of the adjoint equation

Throughout this section, we suppose that the Assumptions 1.1-1.3, 1.17 and 1.24 hold.

Let us denote by $\bar{u} \in L^2(\Gamma)$ a fixed boundary datum and by $\bar{y} \in H^{3/2}(\Omega)$ the associated solution of (1.1.1). According to Theorem 2.23, there exist $h_0 > 0$ and, for $\bar{q} \in (2,4)$ fixed, $\rho > 0$ and $\kappa_\rho > 0$, such that, for any $h < h_0$ and any $u \in \bar{B}_{L^2(\Gamma)}(\bar{u},\rho)$, the equation (2.3.1) has a unique solution $y_h = y_h(u) \in Y_h \cap \overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y},\kappa_\rho)$. Notice that $\{y_h\}_{h<h_0}$ is bounded in $W^{1,4}(\Omega)$ because of the boundedness of $\overline{B}_{L^2(\Gamma)}(\bar{u},\rho)$ in $L^2(\Gamma)$. Let us also denote by y the function $y_u = G(u)$ associated with a fixed element $u \in \overline{B}_{L^2(\Gamma)}(\bar{u},\rho)$. From Theorem 1.18 on page 16 we know that $y \in H^{3/2}(\Omega)$.

Our next step is to carry out the numerical analysis of the adjoint equation

$$\text{(2.7.1)} \qquad \begin{cases} -\text{div}\,[a(x,y)\nabla\varphi] + \dfrac{\partial a}{\partial y}(x,y)\nabla y \cdot \nabla\varphi + \dfrac{\partial f}{\partial y}(x,y)\varphi = \zeta & \text{in}\,\Omega\,, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad a(x,y)\partial_\nu\varphi = v & \text{on}\,\Gamma\,, \end{cases}$$

for $v \in L^2(\Gamma)$ and $\zeta \in L^p(\Omega)$ with $p > 4/3$. Using the triangulation $\mathcal{T}_h$ introduced in Section 2.3, we approximate φ by solutions of the discrete version of (2.7.1):

$$\text{(2.7.2)} \qquad \begin{cases} \text{Find } \varphi_h \in Y_h \text{ such that for all } \phi_h \in Y_h \\ \displaystyle\int_\Omega \left\{ a(x,y_h)\nabla\varphi_h \cdot \nabla\phi_h + \frac{\partial a}{\partial y}(x,y_h)\phi_h \nabla y_h \cdot \nabla\varphi_h + \frac{\partial f}{\partial y}(x,y_h)\varphi_h\phi_h \right\} dx \\ \displaystyle\qquad\qquad\qquad\qquad\qquad\qquad = \int_\Omega \zeta\phi_h\,dx + \int_\Gamma v\phi_h\,d\sigma(x)\,. \end{cases}$$

Theorem 2.27. *For every $h < h_1$, with h_1 given in Theorem 2.25, $v \in L^2(\Gamma)$ and $\zeta \in L^p(\Omega)$ ($p > 4/3$), the variational problem (2.7.2) has a unique solution $\varphi_h \in Y_h$.*

Proof. Since the mapping $(\partial F_h/\partial y_h)(u,y_h)$ defined in the proof of Theorem 2.25 is an isomorphism, the same holds true for its adjoint. This leads immediately to the existence and uniqueness of a solution $\varphi_h \in Y_h$ of (2.7.2) for any $h < h_1$. □

Before establishing error estimates for the approximation of (2.7.1), we prove an auxiliary result concerning a partial discretization of equation (2.7.1). In a first step, we consider a continuous problem of the form (2.7.1) with y_h substituted for y. In a second step, we pass into the fully discretized problem (2.7.2).

In the sequel, we make use of the following inequality

$$\text{(2.7.3)} \quad \|z\|_{L^2(\Gamma)} \le C_\Omega \left(\varepsilon^{1/2}\|\nabla z\|^2_{L^2(\Omega)} + \varepsilon^{-1/2}\|z\|^2_{L^2(\Omega)}\right)^{1/2} \quad \forall z \in H^1(\Omega)\,,\ \varepsilon \in (0,1)\,,$$

cf. Grisvard [**60**, Theorem 1.5.1.10]).

Lemma 2.28. *For any $h < h_0$, $v \in L^q(\Gamma)$ and $\zeta \in L^p(\Omega)$, with $q > 2$ and $p > 4/3$, the equation*

$$\text{(2.7.4)} \qquad \begin{cases} -\text{div}\,[a(x,y_h)\nabla\hat{\varphi}] + \dfrac{\partial a}{\partial y}(x,y_h)\nabla y_h \cdot \nabla\hat{\varphi} + \dfrac{\partial f}{\partial y}(x,y_h)\hat{\varphi} = \zeta & \textit{in}\,\Omega\,, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad a(x,y_h)\partial_\nu\hat{\varphi} = v & \textit{on}\,\Gamma\,, \end{cases}$$

has a unique solution $\hat\varphi \in H^{3/2}(\Omega)$ and it obeys the estimate

$$\|\hat\varphi - \varphi\|_{H^1(\Omega)} \le Ch^{3/4}\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right), \tag{2.7.5}$$

where $C > 0$ is independent of h. Moreover, if Ω is convex then

$$\|\hat\varphi - \varphi\|_{H^1(\Omega)} \le Ch\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right). \tag{2.7.6}$$

Finally, if Ω is convex, $p \ge 2$ and $u \in H^{1/2}(\Gamma)$, then (2.7.5) is replaced by

$$\|\hat\varphi - \varphi\|_{H^1(\Omega)} \le Ch^{3/2}\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right). \tag{2.7.7}$$

We point out that $q > 2$ in the statement of the lemma is only necessary for deriving the error estimates; for the regularity result it is enough to take $q = 2$.

Proof. Taking into account the boundedness of $\{y_h\}_{h<h_0}$ in $W^{1,4}(\Omega)$, the existence, uniqueness and regularity of $\hat\varphi$ is a consequence of Theorem 1.36 on page 32. Let us derive two useful estimates concerning $\hat\varphi$. In view of $\hat\varphi \in H^{3/2}(\Omega) \hookrightarrow W^{1,4}(\Omega)$ and replacing in (1.7.10) on page 33 y and φ by y_h and $\hat\varphi$, respectively, we obtain

$$\|\Delta\hat\varphi\|_{L^{\min\{p,2\}}(\Omega)} \le C\left(\|\zeta\|_{L^p(\Omega)} + \|\hat\varphi\|_{H^{3/2}(\Omega)}\right).$$

Moreover, we get by a straightforward modification of the proof of (1.7.3) on page 30 that

$$\|\hat\varphi\|_{H^{3/2}(\Omega)} \le C\left(\|\zeta\|_{L^p(\Omega)} + \|v\|_{L^q(\Gamma)}\right). \tag{2.7.8}$$

Let us now prove (2.7.5). Subtracting (2.7.1) and (2.7.4), we have

$$\left\{\begin{aligned} &-\mathrm{div}\,[a(x,y)\nabla(\varphi-\hat\varphi)] + \frac{\partial a}{\partial y}(x,y)\nabla y\cdot\nabla(\varphi-\hat\varphi) \\ &\qquad\qquad + \frac{\partial f}{\partial y}(x,y)(\varphi-\hat\varphi) = g_\Omega \quad \text{in } \Omega, \\ &\qquad\qquad a(x,y)\partial_\nu(\varphi-\hat\varphi) = g_\Gamma \quad \text{on } \Gamma, \end{aligned}\right. \tag{2.7.9}$$

where the functions g_Ω and g_Γ are given by

$$\begin{aligned} g_\Omega(x) = \mathrm{div}\,[(a(x,y) - a(x,y_h))\,\nabla\hat\varphi] &+ \left(\frac{\partial a}{\partial y}(x,y_h)\nabla y_h - \frac{\partial a}{\partial y}(x,y)\nabla y\right)\cdot\nabla\hat\varphi \\ &+ \left(\frac{\partial f}{\partial y}(x,y_h) - \frac{\partial f}{\partial y}(x,y)\right)\hat\varphi \end{aligned} \tag{2.7.10}$$

and

$$g_\Gamma(x) = [a(x,y_h) - a(x,y)]\partial_\nu\hat\varphi,$$

respectively. It is easy to check that $g_\Omega \in L^{\min\{p,2\}}(\Omega)$ and $g_\Gamma \in L^q(\Gamma)$ as a consequence of the $W^{1,4}(\Omega)$ regularity of $\hat\varphi$, y and y_h, and the facts $\Delta\hat\varphi \in L^{\min\{p,2\}}(\Omega)$ and

$\partial_\nu \hat\varphi \in L^q(\Gamma)$; see the second equation of (2.7.4). From Theorem 1.36 we know that $\varphi - \hat\varphi$ is the unique solution of (2.7.9) and it satisfies

$$\|\varphi - \hat\varphi\|_{H^1(\Omega)} \le C\left(\|g_\Omega\|_{H^1(\Omega)^*} + \|g_\Gamma\|_{H^{-1/2}(\Gamma)}\right) \tag{2.7.11}$$

with $C > 0$ being independent of φ or $\hat\varphi$. Next we derive an estimate for $\|g_\Omega\|_{H^1(\Omega)^*}$. An integration by parts yields for arbitrary $\phi \in H^1(\Omega)$

$$\begin{aligned}\int_\Omega \operatorname{div}\left[(a(x,y) - a(x,y_h))\,\nabla\hat\varphi\right]\phi\,dx &= \int_\Gamma (a(x,y) - a(x,y_h))\,\partial_\nu\hat\varphi\phi\,d\sigma(x)\\ &\quad - \int_\Omega (a(x,y) - a(x,y_h))\,\nabla\hat\varphi\cdot\nabla\phi\,dx\,.\end{aligned} \tag{2.7.12}$$

Let us estimate both terms. From the assumptions on a, using (2.3.37), (2.7.3) with $\varepsilon = h$, as well as the embedding $H^{1/2}(\Gamma) \hookrightarrow L^s(\Gamma)$ for all $s \in [1,\infty)$, we obtain

$$\begin{aligned}\left|\int_\Gamma (a(x,y) - a(x,y_h))\,\partial_\nu\hat\varphi\phi\,d\sigma(x)\right| &\le C\int_\Gamma |y - y_h||\partial_\nu\hat\varphi\phi|\,d\sigma(x)\\ &\le C\|y - y_h\|_{L^2(\Gamma)}\|v\phi\|_{L^2(\Gamma)}\\ &\le \varepsilon_h h^{3/4}\|v\|_{L^q(\Gamma)}\|\phi\|_{H^1(\Omega)}\,.\end{aligned} \tag{2.7.13}$$

Invoking Lemma 2.3 and (2.3.37) again, we have

$$\|y - y_h\|_{L^4(\Omega)} \le C\|y - y_h\|_{L^2(\Omega)}^{1/2}\|y - y_h\|_{H^1(\Omega)}^{1/2} \le Ch^{3/4}\,, \tag{2.7.14}$$

hence

$$\begin{aligned}\left|\int_\Omega (a(x,y) - a(x,y_h))\,\nabla\hat\varphi\cdot\nabla\phi\,dx\right| &\le C\int_\Omega |y - y_h||\nabla\hat\varphi\cdot\nabla\phi|\,dx\\ &\le C\|y - y_h\|_{L^4(\Omega)}\|\hat\varphi\|_{W^{1,4}(\Omega)}\|\phi\|_{H^1(\Omega)}\\ &\le Ch^{3/4}\|\hat\varphi\|_{H^{3/2}(\Omega)}\|\phi\|_{H^1(\Omega)}\,.\end{aligned} \tag{2.7.15}$$

Thus, from (2.7.12), (2.7.13) and (2.7.15), we deduce the following estimate for the first term of g_Ω

$$\begin{aligned}&\left|\int_\Omega \operatorname{div}\left[(a(x,y) - a(x,y_h))\,\nabla\hat\varphi\right]\phi\,dx\right|\\ &\qquad \le Ch^{3/4}\left(\|\hat\varphi\|_{H^{3/2}(\Omega)} + \|v\|_{L^q(\Gamma)}\right)\|\phi\|_{H^1(\Omega)}\,.\end{aligned} \tag{2.7.16}$$

Concerning the second term in the definition of g_Ω, we write

$$\begin{aligned}&\int_\Omega \left(\frac{\partial a}{\partial y}(x,y_h)\nabla y_h - \frac{\partial a}{\partial y}(x,y)\nabla y\right)\cdot\nabla\hat\varphi\phi\,dx =\\ &= \int_\Omega \left(\frac{\partial a}{\partial y}(x,y_h) - \frac{\partial a}{\partial y}(x,y)\right)\nabla y_h\cdot\nabla\hat\varphi\phi\,dx + \int_\Omega \frac{\partial a}{\partial y}(x,y)\nabla\left(y_h - y\right)\cdot\nabla\hat\varphi\phi\,dx\,.\end{aligned} \tag{2.7.17}$$

Now we get, along with (2.7.14),

$$\begin{aligned}\left|\int_\Omega\left(\frac{\partial a}{\partial y}(x,y_h)-\frac{\partial a}{\partial y}(x,y)\right)\nabla y_h\cdot\nabla\hat\varphi\phi\,dx\right| &\le C\|(y_h-y)\phi\|_{L^2(\Omega)}\|\nabla y_h\cdot\nabla\hat\varphi\|_{L^2(\Omega)}\\ &\le C\|y_h-y\|_{L^4(\Omega)}\|\hat\varphi\|_{H^{3/2}(\Omega)}\|\phi\|_{H^1(\Omega)}\\ &\le Ch^{3/4}\|\hat\varphi\|_{H^{3/2}(\Omega)}\|\phi\|_{H^1(\Omega)}\,. \qquad (2.7.18)\end{aligned}$$

To estimate the second term of (2.7.17) we take into account that

$$\frac{\partial a}{\partial y}(x,y)\nabla\hat\varphi\in L^2(\Omega)^2 \text{ and } \operatorname{div}\left[\frac{\partial a}{\partial y}(x,y)\nabla\hat\varphi\right]\in L^{\min\{p,2\}}(\Omega)\,.$$

Hence, we can perform again an integration by parts

$$\begin{aligned}(2.7.19)\quad \int_\Omega\frac{\partial a}{\partial y}(x,y)\nabla(y_h-y)\cdot\nabla\hat\varphi\phi\,dx &= \int_\Gamma\frac{\partial a}{\partial y}(x,y)\partial_\nu\hat\varphi(y_h-y)\phi\,d\sigma(x)\\ -\int_\Omega\frac{\partial a}{\partial y}(x,y)(y_h-y)\nabla\hat\varphi\cdot\nabla\phi\,dx &- \int_\Omega\operatorname{div}\left[\frac{\partial a}{\partial y}(x,y)\nabla\hat\varphi\right](y_h-y)\phi\,dx\,.\end{aligned}$$

For the first two terms we proceed as in (2.7.13) and (2.7.15). For the third one we find

$$\begin{aligned}&\left|\int_\Omega\operatorname{div}\left[\frac{\partial a}{\partial y}(x,y)\nabla\hat\varphi\right](y_h-y)\phi\,dx\right|\\ &\qquad\le C\left\|\operatorname{div}\left[\frac{\partial a}{\partial y}(x,y)\nabla\hat\varphi\right]\right\|_{L^{\min\{2,p\}}(\Omega)}\|(y_h-y)\phi\|_{L^{\max\{2,p'\}}(\Omega)}\\ &\qquad\le C\left(\|\Delta\hat\varphi\|_{L^{\min\{2,p\}}(\Omega)}+\|\hat\varphi\|_{H^{3/2}(\Omega)}\right)\|y_h-y\|_{L^4(\Omega)}\|\phi\|_{H^1(\Omega)}\\ (2.7.20)\quad&\qquad\le Ch^{3/4}\left(\|\zeta\|_{L^p(\Omega)}+\|\hat\varphi\|_{H^{3/2}(\Omega)}\right)\|\phi\|_{H^1(\Omega)}\,,\end{aligned}$$

where we used (2.7.14), the facts that $p'<4$ (because $p>4/3$) and $H^1(\Omega)\hookrightarrow L^s(\Omega)$ for any $s\in[1,\infty)$. Therefore, the inequalities (2.7.19) and (2.7.20) lead to

$$\begin{aligned}(2.7.21)\quad&\left|\int_\Omega\frac{\partial a}{\partial y}(x,y)\nabla\left(y_h-y\right)\cdot\nabla\hat\varphi\phi\,dx\right|\\ &\qquad\le Ch^{3/4}\left(\|\zeta\|_{L^p(\Omega)}+\|v\|_{L^q(\Gamma)}+\|\hat\varphi\|_{H^{3/2}(\Omega)}\right)\|\phi\|_{H^1(\Omega)}\,.\end{aligned}$$

Finally, (2.7.17), (2.7.18) and (2.7.21), imply

$$\begin{aligned}(2.7.22)\quad&\left|\int_\Omega\left(\frac{\partial a}{\partial y}(x,y_h)\nabla y_h-\frac{\partial a}{\partial y}(x,y)\nabla y\right)\cdot\nabla\hat\varphi\phi\,dx\right|\\ &\qquad\le Ch^{3/4}\left(\|\zeta\|_{L^p(\Omega)}+\|v\|_{L^q(\Gamma)}+\|\hat\varphi\|_{H^{3/2}(\Omega)}\right)\|\phi\|_{H^1(\Omega)}\,.\end{aligned}$$

The last term on the right-hand side of (2.7.10) is easy to estimate:

$$\begin{aligned}\left|\int_\Omega \left(\frac{\partial f}{\partial y}(x,y_h) - \frac{\partial f}{\partial y}(x,y)\right)\hat\varphi\phi\,dx\right| &\le C\|y_h - y\|_{L^2(\Omega)}\|\hat\varphi\|_{L^4(\Omega)}\|\phi\|_{L^4(\Omega)}\\ &\le Ch\|\hat\varphi\|_{H^{3/2}(\Omega)}\|\phi\|_{H^1(\Omega)}\,.\end{aligned}$$

Let us now estimate $\|g_\Gamma\|_{H^{-1/2}(\Gamma)}$. Taking $\psi \in H^{1/2}(\Gamma)$, we get analogous to (2.7.13)

$$\begin{aligned}\left|\int_\Gamma (a(x,y_h) - a(x,y))\,\partial_\nu\hat\varphi\psi\,d\sigma(x)\right| &\le \|y_h - y\|_{L^2(\Gamma)}\|\partial_\nu\hat\varphi\|_{L^q(\Gamma)}\|\psi\|_{L^{\frac{2q}{q-2}}(\Gamma)}\\ &\le Ch^{3/4}\|v\|_{L^q(\Gamma)}\|\psi\|_{H^{1/2}(\Gamma)}\,.\end{aligned}\tag{2.7.23}$$

From (2.7.11), along with the inequalities (2.7.16) and (2.7.22)-(2.7.23), we arrive at

$$\begin{aligned}\|\varphi - \hat\varphi\|_{H^1(\Omega)} &\le Ch^{3/4}\left(\|\zeta\|_{L^p(\Omega)} + \|v\|_{L^q(\Gamma)} + \|\hat\varphi\|_{H^{3/2}(\Omega)}\right)\\ &\le Ch^{3/4}\left(\|\zeta\|_{L^p(\Omega)} + \|v\|_{L^q(\Gamma)}\right)\end{aligned}$$

and conclude (2.7.5). It remains to prove (2.7.6) and (2.7.7). To deduce (2.7.6) we repeat the above steps. Some estimates can be improved because we can use (2.3.39) for the error $y_h - y$. As a consequence of (2.3.39) and Lemma 2.3, we have

$$\|y - y_h\|_{L^4(\Omega)} \le C\|y - y_h\|^{1/2}_{L^2(\Omega)}\|y - y_h\|^{1/2}_{H^1(\Omega)} \le Ch\,.$$

Further, an application of (2.7.3) with $\varepsilon = h^2$ yields

$$\|y - y_h\|_{L^2(\Gamma)} \le C_\Omega\left(h\|y - y_h\|^2_{H^1(\Omega)} + h^{-1}\|y - y_h\|^2_{L^2(\Omega)}\right)^{1/2} \le Ch\,.$$

Taking into account these inequalities, we get the order $\mathcal{O}(h)$ for the expressions in (2.7.16) and (2.7.22)-(2.7.23), and finally for $\|\varphi - \hat\varphi\|_{H^1(\Omega)}$.

Let us finish the proof by considering the case of a convex domain, $p \ge 2$ and a boundary datum $u \in H^{1/2}(\Gamma)$. With the aid of (2.3.40), we observe that (2.7.3) applied with $\varepsilon = h^2$ leads this time to the inequality

$$\|y - y_h\|_{L^2(\Gamma)} \le C_\Omega\left(h\|y - y_h\|^2_{H^1(\Omega)} + h^{-1}\|y - y_h\|^2_{L^2(\Omega)}\right)^{1/2} \le Ch^{3/2}\,.$$

Further, we have

$$\|y - y_h\|_{L^4(\Omega)} \le C\|y - y_h\|^{1/2}_{L^2(\Omega)}\|y - y_h\|^{1/2}_{H^1(\Omega)} \le Ch^{3/2}\,.$$

With these inequalities we can improve the previous estimates to conclude (2.7.7). □

THEOREM 2.29. *For any $h < h_1$, $v \in L^q(\Gamma)$ and $\zeta \in L^p(\Omega)$, with h_1 given in Theorem 2.25, $q > 2$ and $p > 4/3$, we have*

$$\|\varphi - \varphi_h\|_{L^2(\Omega)} + h^{1/4}\|\varphi - \varphi_h\|_{H^1(\Omega)} \le Ch^{3/4}\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right)\tag{2.7.24}$$

with $C > 0$ independent of h. Moreover, if Ω is convex then

$$\|\varphi - \varphi_h\|_{L^2(\Omega)} + h^{1/2}\|\varphi - \varphi_h\|_{H^1(\Omega)} \le Ch\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right). \tag{2.7.25}$$

Finally, if Ω is convex, $p \ge 2$ and $v, u \in H^{1/2}(\Gamma)$, then (2.7.24) is replaced by

$$\|\varphi - \varphi_h\|_{L^2(\Omega)} + h^{1/2}\|\varphi - \varphi_h\|_{H^1(\Omega)} \le Ch^{3/2}\left(\|v\|_{H^{1/2}(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right). \tag{2.7.26}$$

Proof. First, we show that

$$\|\hat\varphi - \varphi_h\|_{L^2(\Omega)} \le Ch^{1/2}\|\hat\varphi - \varphi_h\|_{H^1(\Omega)} \tag{2.7.27}$$

and

$$\|\hat\varphi - \varphi_h\|_{H^1(\Omega)} \le Ch^{1/2}\|\hat\varphi\|_{H^{3/2}(\Omega)}. \tag{2.7.28}$$

The first estimate can be deduced in an analogous way as we proved (2.6.10). Let us only comment the differences. Let $z \in H^{3/2}(\Omega)$ be the unique solution of the equation

$$\left\{\begin{aligned} -\mathrm{div}\left[a(x,y)\nabla z + \frac{\partial a}{\partial y}(x,y) z \nabla y\right] + \frac{\partial f}{\partial y}(x,y) z &= \hat\varphi - \varphi_h \quad \text{in } \Omega, \\ \left[a(x,y)\nabla z + \frac{\partial a}{\partial y}(x,y) z \nabla y\right]\cdot\nu &= 0 \qquad\quad \text{on } \Gamma. \end{aligned}\right. \tag{2.7.29}$$

Notice that the $H^{3/2}(\Omega)$ regularity of z follows directly from Theorem 1.31 on page 26. Making use of the mappings B and B_h defined in (2.6.12) and (2.6.4), respectively, (2.3.37) and (2.2.1), and taking into account that $B_h(\Pi_h z, \hat\varphi - \varphi_h) = 0$, we obtain

$$\begin{aligned} \|\hat\varphi - \varphi_h\|^2_{L^2(\Omega)} &= B(z, \hat\varphi - \varphi_h) \\ &= B(z - \Pi_h z, \hat\varphi - \varphi_h) + B(\Pi_h z, \hat\varphi - \varphi_h) \\ &= B(z - \Pi_h z, \hat\varphi - \varphi_h) + (B(\Pi_h z, \hat\varphi - \varphi_h) - B_h(\Pi_h z, \hat\varphi - \varphi_h)) \\ &\le C\left(\|z - \Pi_h z\|_{H^1(\Omega)} + \|y - y_h\|_{H^1(\Omega)}\|\Pi_h z\|_{W^{1,4}(\Omega)}\right)\|\hat\varphi - \varphi_h\|_{H^1(\Omega)} \\ &\le Ch^{1/2}\|z\|_{H^{3/2}(\Omega)}\|\hat\varphi - \varphi_h\|_{H^1(\Omega)}. \qquad (2.7.30) \end{aligned}$$

Hence, (2.7.27) follows from (2.7.30) and the inequality $\|z\|_{H^{3/2}(\Omega)} \le C\|\hat\varphi - \varphi_h\|_{L^2(\Omega)}$; compare inequality (1.6.11) on page 26. To show (2.7.28) let us introduce the function

$$S_h : H^1(\Omega) \times H^1(\Omega) \longrightarrow \mathbb{R},\ S_h(w,\phi) = \int_\Omega \left\{a(x,y_h)\nabla w\cdot\nabla\phi + \frac{\partial f}{\partial y}(x,y_h) w\phi\right\} dx.$$

Utilizing the assumptions on a and f, it is immediate that

$$S_h(w,w) \ge C_0\|w\|^2_{H^1(\Omega)} \text{ and } |S_h(w,\phi)| \le C_1\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)} \quad \forall w,\phi \in H^1(\Omega).$$

Consequently,

$$\begin{aligned} C_0\|\hat\varphi - \varphi_h\|^2_{H^1(\Omega)} &\le S_h(\hat\varphi - \varphi_h, \hat\varphi - \varphi_h) \\ &= S_h(\hat\varphi - \varphi_h, \hat\varphi - \Pi_h\hat\varphi) + S_h(\hat\varphi - \varphi_h, \Pi_h\hat\varphi - \varphi_h) =: I_1 + I_2. \end{aligned}$$

Exploiting (2.2.1), the estimate for I_1 is obvious:

$$|I_1| = |S_h(\hat{\varphi} - \varphi_h, \hat{\varphi} - \Pi_h\hat{\varphi})| \le Ch^{1/2}\|\hat{\varphi}\|_{H^{3/2}(\Omega)}\|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)} .$$

Invoking the equations satisfied by $\hat{\varphi}$ and φ_h, we infer

$$\begin{aligned} |I_2| &= |S_h(\hat{\varphi} - \varphi_h, \Pi_h\hat{\varphi} - \varphi_h)| \\ &= \left|\int_\Omega \frac{\partial a}{\partial y}(x, y_h)\,(\Pi_h\hat{\varphi} - \varphi_h)\,\nabla y_h\cdot\nabla\,(\hat{\varphi} - \varphi_h)\,dx\right| \\ &\le C\|\,(\Pi_h\hat{\varphi} - \varphi_h)\,|\nabla y_h|\|_{L^2(\Omega)}\|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)} \\ &\le C\left(\|\Pi_h\hat{\varphi} - \hat{\varphi}\|_{L^4(\Omega)} + \|\hat{\varphi} - \varphi_h\|_{L^4(\Omega)}\right)\|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)} . \end{aligned}$$

From the last three inequalities it follows

$$\|\hat{\varphi} - \varphi_h\|^2_{H^1(\Omega)} \le C\left(h^{1/2}\|\hat{\varphi}\|_{H^{3/2}(\Omega)} + \|\hat{\varphi} - \varphi_h\|_{L^4(\Omega)}\right)\|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)}$$

which, as already shown on page 43, leads to

$$\|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)} \le C\left(h^{1/2}\|\hat{\varphi}\|_{H^{3/2}(\Omega)} + \|\hat{\varphi} - \varphi_h\|_{L^2(\Omega)}\right) . \tag{2.7.31}$$

Combining (2.7.31) and (2.7.27), we finally conclude (2.7.28).

The estimate for $\|\varphi - \varphi_h\|_{H^1(\Omega)}$ as given in (2.7.24) and (2.7.25) follows from (2.7.28), (2.7.8) and Lemma 2.28:

$$\begin{aligned} \|\varphi - \varphi_h\|_{H^1(\Omega)} &\le \|\varphi - \hat{\varphi}\|_{H^1(\Omega)} + \|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)} \\ &\le \|\varphi - \hat{\varphi}\|_{H^1(\Omega)} + Ch^{1/2}\|\hat{\varphi}\|_{H^{3/2}(\Omega)} \\ &\le Ch^{1/2}\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right) . \end{aligned}$$

To complete the proof of (2.7.24) and (2.7.25) we have to estimate $\|\varphi - \varphi_h\|_{L^2(\Omega)}$. To this end, we use again the previous lemma, the embedding $H^1(\Omega) \hookrightarrow L^2(\Omega)$, (2.7.27) and the estimate in the $H^1(\Omega)$ norm already proved, and get

$$\begin{aligned} \|\varphi - \varphi_h\|_{L^2(\Omega)} &\le \|\varphi - \hat{\varphi}\|_{L^2(\Omega)} + \|\hat{\varphi} - \varphi_h\|_{L^2(\Omega)} \\ &\le C\left(\|\varphi - \hat{\varphi}\|_{H^1(\Omega)} + h^{1/2}\|\hat{\varphi} - \varphi_h\|_{H^1(\Omega)}\right) \\ &\le C\left(\|\varphi - \hat{\varphi}\|_{H^1(\Omega)} + h\|\hat{\varphi}\|_{H^{3/2}(\Omega)}\right) \\ &\le Ch^\sigma\left(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\right) \end{aligned}$$

with $\sigma = 3/4$ if Ω is non-convex and $\sigma = 1$ if Ω is convex.

It remains to show (2.7.26). Under the assumptions of the theorem, we know that $y = y_u \in H^2(\Omega)$. Moreover, $\hat{\varphi} \in H^2(\Omega)$ which can be deduced from Theorem 1.36-(3) on page 32, because $\{y_h\}_{h<h_1}$ is bounded in $W^{1,6}(\Omega)$. The latter boundedness result

follows from (2.2.2), the inverse inequality (2.2.6), (2.3.40) and (2.3.35),

$$\begin{aligned}\|y_h - \Pi_h y\|_{W^{1,6}(\Omega)} &\le \frac{C}{h^{2/3}}\|y_h - \Pi_h y\|_{H^1(\Omega)}\\ &\le \frac{C}{h^{2/3}}\Big(\|y_h - y\|_{H^1(\Omega)} + \|y - \Pi_h y\|_{H^1(\Omega)}\Big)\\ &\le \frac{C}{h^{2/3}}h\Big(1 + \|y\|_{H^2(\Omega)}\Big) < \infty\,.\end{aligned}$$

Furthermore, the solution z of (2.7.29) is in $H^2(\Omega)$, too. This can be shown along the lines of the proof of Theorem 1.33; see pages 28-29. The higher regularity of y and z allows us to use (2.3.40) and (2.3.35). Therefore, we can improve (2.7.27) to

$$\|\hat\varphi - \varphi_h\|_{L^2(\Omega)} \le Ch\|\hat\varphi - \varphi_h\|_{H^1(\Omega)}\,.$$

On the other hand, thanks to the $H^2(\Omega)$ regularity of $\hat\varphi$, we can replace (2.7.28) by

$$\|\hat\varphi - \varphi_h\|_{H^1(\Omega)} \le Ch\|\hat\varphi\|_{H^2(\Omega)}\,.$$

This is easily obtained, by repeating the proof of inequality (2.7.28) and taking into account (2.3.35). Combining the last two inequalities and

$$\|\hat\varphi\|_{H^2(\Omega)} \le C\Big(\|v\|_{H^{1/2}(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\Big)\,,$$

we get

$$\|\hat\varphi - \varphi_h\|_{L^2(\Omega)} + h\|\hat\varphi - \varphi_h\|_{H^1(\Omega)} \le Ch^2\Big(\|v\|_{H^{1/2}(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\Big)\,.$$

By following the steps above and using (2.7.7), the proof of (2.7.26) is complete. □

REMARK 2.30. *Lemma 2.28 contains estimates for the error $\hat\varphi - \varphi$ in the $H^1(\Omega)$ norm only. Unfortunately, we are not able to prove a higher order of convergence in the $L^2(\Omega)$ norm. For this reason we have used on page 73 error estimates in the $H^1(\Omega)$ norm to obtain estimates for $\|\varphi - \varphi_h\|_{L^2(\Omega)}$. Indeed, this procedure delivers very rough estimates in general.*

The following result will be used in the context of error estimates for control problems associated with the quasilinear equation (1.1.1); see Chapter 4.

COROLLARY 2.31. *For any $h < h_1$, $v \in L^q(\Gamma)$ and $\zeta \in L^p(\Omega)$, with h_1 given in Theorem 2.25, $q > 2$ and $p > 4/3$, we have*

$$\|\varphi_h\|_{W^{1,4}(\Omega)} \le C\Big(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\Big)\,, \tag{2.7.32}$$

where C is only dependent on u but not on h, v or ζ.

PROOF. By a simple modification of the proof of estimate (1.7.3) on page 30, we see that the solution φ of (2.7.1) satisfies $\|\varphi\|_{H^{3/2}(\Omega)} \le C'\Big(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\Big)$. Combining this result with the inequality $\|\varphi - \varphi_h\|_{W^{1,4}(\Omega)} \le C''\Big(\|v\|_{L^q(\Gamma)} + \|\zeta\|_{L^p(\Omega)}\Big)$, which can be deduced along the lines of the proof of (2.3.23), we conclude (2.7.32). □

CHAPTER 3

The optimal control problem

3.1. Introduction

In this chapter, we investigate a wide class of optimal boundary control problems governed by quasilinear elliptic equations of the type (1.1.1). Our main goal is to establish first- and second-order conditions for local optimality in the presence of pointwise constraints on the control.

For convex problems first-order necessary optimality conditions are even sufficient for global optimality. In contrast to this, for nonlinear problems higher order conditions such as second-order sufficient conditions should be employed to verify local optimality. The latter ones are proved to be indispensable for several reasons. First, they play an important role in the stability and numerical analysis of the optimal control problems, in particular in the error analysis for local solutions of the finite element approximation of the control problems. Secondly, the convergence analysis of higher order numerical optimization algorithms such as SQP-type methods rely heavily on second-order sufficient conditions, see Alt and Malanowski [**4**], Dontchev et al. [**49**] or Ito and Kunisch [**68**]. Likewise, second-order necessary conditions should also be studied since they serve to measure the gap between them and the sufficient ones. In turn, this gap shows how restrictive the sufficient conditions under consideration are. This explains why we are concerned in formulating sufficient second-order conditions which are the closest to the associated necessary ones.

There are two common techniques to verify that certain second-order conditions are sufficient for local optimality. The first way is to apply some abstract methods for optimization problems in function spaces, see Casas and Tröltzsch [**35**], while the other method uses Pontryagin's principle, cf. Casas and Mateos [**25**]. In [**30**], Casas, Mateos and Tröltzsch, have shown that both methods are equivalent.

Meanwhile, there exists a very extensive literature devoted to second-order optimality conditions for control problems governed by partial differential equations. We mention only the textbook by Tröltzsch [**91**] for an overview, Goldberg and Tröltzsch [**56, 57, 58**] for boundary control of parabolic equations, Casas, Tröltzsch and Unger [**39, 40**], as well as Casas and Tröltzsch [**34**], for elliptic boundary control problems with nonlinear boundary conditions.

The list of contributions concerning the Pontryagin's principle is very large. For elliptic problems this principle was investigated by Bonnans and Casas [**9**] and Casas

[**13**], while the parabolic case was studied by Casas [**15**], Casas et al. [**32**] and Raymond and Zidani [**83**]. In the context of quasilinear equations with nonlinearity of gradient type, Pontryagin's principle was considered by Casas [**14**] and Casas and Yong [**41**].

There is some recent progress in the case of optimal control problems governed by quasilinear equations. The first step towards a corresponding analysis was recently made by Casas and Tröltzsch in [**36**], where first- and second-order optimality conditions as well as a Pontryagin's principle for the distributed optimal control of quasilinear elliptic equations are discussed. Other publications concerning quasilinear equations, in which the leading coefficient of the differential operator depends on the gradient of the solution, have already been commented in Section 1.1.

The plan of this chapter is as follows. In the first section, the control problem is formulated and the existence of an optimal solution is shown. Next first-order necessary optimality conditions are derived. These conditions lead to a useful characterization for optimal controls, which in turn allows us to deduce a corresponding regularity result; see Theorem 3.10. For the derivation of the second-order optimality conditions a Pontryagin's principle is proved in Section 3.4. Finally, in Section 3.5 we establish necessary and sufficient second-order optimality conditions.

3.2. Problem formulation and main assumptions

We associate with the quasilinear equation (1.1.1) the following optimal control problem

$$(\mathcal{P})\quad \begin{cases} \min J(u) = \displaystyle\int_\Omega L(x, y_u(x))\, dx + \int_\Gamma l(x, y_u(x), u(x))\, d\sigma(x)\,, \\ \text{subject to}\quad u \in U_{ad} := \{v \in L^\infty(\Gamma) \,|\, u_a(x) \le v(x) \le u_b(x) \text{ a.e. } x \in \Gamma\}\,, \\ \qquad\qquad\quad (y_u, u) \text{ satisfying the equation (1.1.1)}\,, \end{cases}$$

where $L : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ and $l : \Gamma \times (\mathbb{R} \times \mathbb{R}) \longrightarrow \mathbb{R}$ are Carathéodory functions and $u_a, u_b \in L^\infty(\Gamma)$, with $u_a \le u_b$ a.e. on Γ. In the context of optimal control, the PDE (1.1.1) is denoted as *state equation*, u is called *control* function and y_u is the associated *state*, J the *cost function* or *objective function* and U_{ad} the *set of admissible controls.*

Example 3.1. *In control theory, a frequent example for the choice of the functional is that of tracking type (see, for instance, Ito and Kunisch* [**68**]*)*

$$J(u) = \frac{1}{2}\int_\Omega (y_u(x) - y_d(x))^2\, dx + \frac{\lambda}{2}\int_\Gamma (u(x) - u_d(x))^2\, d\sigma(x)$$

with given functions $y_d \in L^2(\Omega)$, $u_d \in L^2(\Gamma)$ and $\lambda \ge 0$. The functions y_d and u_d denote the desired state and control of the problem, respectively. If $\lambda > 0$ then *it is well-known that the so-called Tikhonov term* $\frac{\lambda}{2}\int_\Gamma (u(x)-u_d(x))^2\,d\sigma(x)$ *has a regularizing effect on the optimal control.*

Here we do not consider the case when l does not depend on u. In such a situation optimal controls are often of bang-bang type.

Definition 3.2. A control $\bar u \in U_{ad}$ is said to be *(globally) optimal* with *optimal state* $\bar y = y_{\bar u}$ if

$$J(\bar u) \le J(u) \quad \forall u \in U_{ad}\,.$$

The control $\bar u$ is called *locally optimal* in the sense of $L^\infty(\Gamma)$ if there exists $\varepsilon > 0$ such that the previous inequality holds for all $u \in U_{ad}$ with $\|\bar u - u\|_{L^\infty(\Gamma)} < \varepsilon$.

The next theorem concerns the existence of a global solution for problem $(\mathcal{P})$. Although the proof of this theorem is standard (compare also Casas et al. [**23**] or Casas and Mateos [**26**]), we will sketch it here for convenience of the reader.

Theorem 3.3. *Suppose that the Assumptions 1.1-1.3 hold true,* $a : \bar\Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ *is continuous and, for every* $(x,y) \in \Gamma \times \mathbb{R}$, $l(x,y,\cdot) : \mathbb{R} \longrightarrow \mathbb{R}$ *is convex. Assume also that, for any* $M > 0$*, there exist functions* $\psi_{L,M} \in L^1(\Omega)$ *and* $\psi_{l,M} \in L^1(\Gamma)$ *such that*

$$|L(x,y)| \le \psi_{L,M}(x) \quad \textit{and} \quad |l(s,y,u)| \le \psi_{l,M}(s)$$

for a.a. $x \in \Omega$, $s \in \Gamma$ *and* $|y|,|u| \le M$*. Then* $(\mathcal{P})$ *admits at least one optimal solution* $\bar u$.

Proof. Let $\{u_k\}_{k=1}^\infty \subset U_{ad}$ be a minimizing sequence for $(\mathcal{P})$, i.e. $J(u_k) \to \inf(\mathcal{P})$ when $k \to \infty$. Since $\{u_k\}_{k=1}^\infty$ is bounded in $L^\infty(\Gamma)$, it is also bounded in $L^{r/2}(\Gamma)$ for any $r > 2$ satisfying the condition (1.3.13) on page 9. Therefore, we can take a subsequence, denoted in the same way, which converges weakly in $L^{r/2}(\Gamma)$ to some $\bar u$. Moreover, $\bar u$ is a admissible control for $(\mathcal{P})$, because the set U_{ad} is closed and convex in $L^{r/2}(\Gamma)$, hence it is weakly closed. Further, Theorem 1.9 on page 9 implies that $\{y_{u_k}\}_{k=1}^\infty$ is bounded in $W^{1,r}(\Omega)$. Again, we can extract a subsequence, denoted again in the same way, such that $y_{u_k} \rightharpoonup \tilde y$ weakly in $W^{1,r}(\Omega)$ and strongly in $C(\bar\Omega)$, due to the compactness of the embedding $W^{1,r}(\Omega) \hookrightarrow C(\bar\Omega)$. It suffices to show that $y_{\bar u} = \tilde y$ and that $\bar u$ is optimal for $(\mathcal{P})$. The solution y_{u_k} of the state equation corresponding to u_k satisfies

$$\int_\Omega \{a(x,y_{u_k})\nabla y_{u_k}\cdot\nabla\phi + f(x,y_{u_k})\phi\}\,dx = \int_\Gamma u_k\phi\,d\sigma(x) \quad \forall\phi \in H^1(\Omega)\,. \tag{3.2.1}$$

Thanks to the Assumptions 1.2-1.3 and the previous convergences established, we can pass to the limit in (3.2.1) and obtain

$$\int_\Omega a(x,\tilde y)\nabla\tilde y\cdot\nabla\phi\,dx + \int_\Omega f(x,\tilde y)\phi\,dx = \int_\Gamma \bar u\phi\,d\sigma(x) \quad \forall\phi \in H^1(\Omega)\,.$$

The uniqueness of the solution of the previous equation implies $y_{\bar u} = \tilde y$. Finally, we can prove that $J(\bar u) = \inf(\mathcal{P})$ by applying Mazur's theorem (see, for instance, Ekeland and Temam [**53**]). Since this proof is very similar to that of Casas and Mateos [**26**, Theorem 8], we omit it here. □

Remark 3.4. *An important property of the admissible set U_{ad} is that it is closed and convex. In the event that U_{ad} is not bounded, for instance in the unconstrained case $U_{ad} = L^2(\Gamma)$, some coercivity assumption on J is necessary to ensure the existence of a solution for $(\mathcal{P})$.*

Remark 3.5. *The convexity of l with respect to the control variable plays an essential role in the proof of Theorem 3.3. In the absence of the convexity, some compactness arguments are necessary to deduce the existence of a minimum. Otherwise, this existence is not guaranteed in general.*

In the sequel, we will consider locally optimal solutions, since they are very interesting from the numerical point of view; optimization algorithms mostly provide local minima. Moreover, from the theoretical point of view there are no criteria to distinguish local and global minima except their definitions.

Throughout this chapter, we suppose that $a : \bar{\Omega} \times \mathbb{R} \to \mathbb{R}$ is continuous and the Assumptions 1.1-1.3 and 1.24-(1) are satisfied.

3.3. First-order optimality conditions

The goal of this section is to derive first-order necessary optimality conditions. These optimality conditions satisfied by $\bar{u} \in U_{ad}$ can be obtained from the standard variational inequality

$$J'(\bar{u})(u - \bar{u}) \geq 0 \quad \forall u \in U_{ad}\,. \tag{3.3.1}$$

To proceed in this way the differentiability of J is needed. Since we also aim to discuss second-order optimality conditions, we shall impose the following assumption.

Assumption 3.6. *Let*

$$r > 2 \ \ \textit{satisfy (1.3.13)}\,, \quad \beta \geq \frac{2r}{r+2} \ \ \textit{and} \ \ \gamma \geq \frac{r}{2}\,.$$

The functions $L : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ and $l : \Gamma \times (\mathbb{R} \times \mathbb{R}) \longrightarrow \mathbb{R}$ are of class C^2 with respect to the second variable and to the last two variables, respectively. For any $M > 0$ there exist constants $C_{L,M}, C_{l,M} > 0$, and functions $\psi_{\Omega,M} \in L^\beta(\Omega)$, $\psi_{1,\Gamma,M} \in L^\gamma(\Gamma)$

and $\psi_{2,\Gamma,M} \in L^2(\Gamma)$, such that

$$\left|\frac{\partial L}{\partial y}(x,y)\right| \le \psi_{\Omega,M}(x)\,, \quad \left|\frac{\partial^2 L}{\partial y^2}(x,y)\right| \le C_{L,M}\,,$$

$$\left|\frac{\partial^2 L}{\partial y^2}(x,y_2) - \frac{\partial^2 L}{\partial y^2}(x,y_1)\right| \le C_{L,M}\,|y_2 - y_1|\,,$$

$$\left|\frac{\partial l}{\partial y}(x',y,u)\right| \le \psi_{1,\Gamma,M}(x')\,, \ \left|\frac{\partial l}{\partial u}(x',y,u)\right| \le \psi_{2,\Gamma,M}(x')\,, \ \left\|D^2_{(y,u)} l(x',y,u)\right\|_{\mathbb{R}^{2\times 2}} \le C_{l,M}\,,$$

$$\left\|D^2_{(y,u)} l(x',y_2,u_2) - D^2_{(y,u)} l(x',y_1,u_1)\right\|_{\mathbb{R}^{2\times 2}} \le C_{l,M}\left(|y_2-y_1| + |u_2-u_1|\right),$$

for a.a. $x \in \Omega$, $x' \in \Gamma$ and $|y|, |y_i|, |u|, |u_i| \le M$, $i = 1, 2$, where $D^2_{(y,u)} l$ denotes the second derivative of l w.r.t. (y, u), i.e. the associated Hessian matrix, and $\|\cdot\|_{\mathbb{R}^{2\times 2}}$ is any matricial norm.

THEOREM 3.7. *The functional $J : L^\infty(\Gamma) \longrightarrow \mathbb{R}$ is of class C^2 and for every $u, v, v_1, v_2 \in L^\infty(\Gamma)$, we have*

$$J'(u)v = \int_\Gamma \left(\frac{\partial l}{\partial u}(x, y_u, u) + \varphi_u\right) v \, d\sigma(x) \tag{3.3.2}$$

and

$$\begin{aligned} J''(u)v_1v_2 = &\int_\Gamma \left\{\frac{\partial^2 l}{\partial y^2}(x,y_u,u) z_{v_1} z_{v_2} + \frac{\partial^2 l}{\partial y \partial u}(x,y_u,u)(z_{v_1} v_2 + z_{v_2} v_1)\right. \\ &\left. + \frac{\partial^2 l}{\partial u^2}(x,y_u,u) v_1 v_2\right\} d\sigma(x) + \int_\Omega \left(\frac{\partial^2 L}{\partial y^2}(x,y_u) - \varphi_u \frac{\partial^2 f}{\partial y^2}(x,y_u)\right) z_{v_1} z_{v_2}\, dx \\ &- \int_\Omega \nabla\varphi_u \cdot \left(\frac{\partial^2 a}{\partial y^2}(x,y_u) z_{v_1} z_{v_2} \nabla y_u + \frac{\partial a}{\partial y}(x,y_u)\left(z_{v_1}\nabla z_{v_2} + z_{v_2} \nabla z_{v_1}\right)\right) dx\,, \end{aligned} \tag{3.3.3}$$

where $\varphi_u \in W^{1,r}(\Omega)$ is the unique solution of the adjoint equation

$$\begin{cases} -\operatorname{div}[a(x,y_u)\nabla\varphi_u] + \dfrac{\partial a}{\partial y}(x,y_u)\nabla y_u \cdot \nabla\varphi_u + \dfrac{\partial f}{\partial y}(x,y_u)\varphi_u = \dfrac{\partial L}{\partial y}(x,y_u) & \text{in } \Omega\,, \\ a(x,y_u)\partial_\nu \varphi_u = \dfrac{\partial l}{\partial y}(x,y_u,u) & \text{on } \Gamma\,, \end{cases} \tag{3.3.4}$$

and $z_{v_i} = G'(u)v_i$ is the solution of (1.6.9) with $v = v_i$, $i = 1, 2$. If, in addition, Assumption 1.17 holds, $\beta > 4/3$ and $\gamma \ge 2$ (see Assumption 3.6), then $\varphi_u \in H^{3/2}(\Omega)$.

PROOF. The first- and second-order derivatives of J can be obtained from Theorem 1.30 on page 25 and the chain rule. By taking into account that $y_u \in W^{1,r}(\Omega)$ (Theorem 1.9 on page 9), the existence, uniqueness and $W^{1,r}(\Omega)$ regularity of φ_u follow from Theorem 1.36-(1) on page 32. The last statement of the theorem is a consequence of Theorem 1.36-(2) and the fact that $y \in H^{3/2}(\Omega)$; see Theorem 1.18 on page 16. □

The function φ_u is called the *adjoint state* associated with u. As one can see from the previous theorem, it allows us to get a simple expression of the derivatives of J.

Remark 3.8. *From the expressions (3.3.2)-(3.3.3) for $J'(u)$ and $J''(u)$ it is easy to check that the functionals $J'(u)$ and $J''(u)$ can be extended from $L^\infty(\Gamma)$ to $L^2(\Gamma)$. Indeed, it is enough to remark that, thanks to the $W^{1,r}(\Omega)$ regularity of z_{v_i}, the integrals in (3.3.3) are well defined for every $v_i \in L^2(\Gamma)$, $i = 1, 2$. Moreover, these integrals are continuous w.r.t. the topology of $L^2(\Gamma)$ because of the continuous dependence of z_{v_i} on v_i; notice that the inverse operator of $\partial_y F(y_u, u)$ given on page 25 is an isomorphism.*

The first-order necessary optimality conditions stated in the next theorem follow from the variational inequality (3.3.1), along with the expression of the derivative of J given in (3.3.2) and (3.3.4).

Theorem 3.9. *Assume that $\bar u$ is a local minimum of $(\mathcal{P})$ and $\bar y$ the associated state. Then there exists $\bar\varphi \in W^{1,r}(\Omega)$ such that*

$$(3.3.5)\qquad \begin{cases} -\mathrm{div}[a(x,\bar y)\nabla\bar\varphi] + \dfrac{\partial a}{\partial y}(x,\bar y)\nabla\bar y\cdot\nabla\bar\varphi + \dfrac{\partial f}{\partial y}(x,\bar y)\bar\varphi = \dfrac{\partial L}{\partial y}(x,\bar y) \quad in\ \Omega\,, \\ \hfill a(x,\bar y)\partial_\nu\bar\varphi = \dfrac{\partial l}{\partial y}(x,\bar y,\bar u)\ on\ \Gamma\,, \end{cases}$$

$$(3.3.6)\qquad \int_\Gamma \left(\frac{\partial l}{\partial u}(x,\bar y(x),\bar u(x)) + \bar\varphi(x)\right)(u(x)-\bar u(x))\,d\sigma(x) \ge 0 \quad \forall u \in U_{ad}\,.$$

If we define the Riesz representation of J' by

$$(3.3.7)\qquad \bar d(x) = \frac{\partial l}{\partial u}(x,\bar y(x),\bar u(x)) + \bar\varphi(x) \quad \text{for a.a. } x \in \Gamma$$

then we get from the variational inequality (3.3.6) that

$$(3.3.8)\qquad \bar d(x) = \begin{cases} = 0 & \text{if } u_a(x) < \bar u(x) < u_b(x)\,, \\ \le 0 & \text{if } \bar u(x) = u_b(x)\,, \\ \ge 0 & \text{if } \bar u(x) = u_a(x)\,. \end{cases}$$

We finish this section by proving a regularity result for optimal solutions of $(\mathcal{P})$ which is deduced from the first-order necessary conditions. For a slightly more general result we refer to [**20**].

Theorem 3.10. *Suppose that*

$$(3.3.9)\qquad \exists\Lambda_l > 0 \ \ such\ that\ \ \frac{\partial^2 l}{\partial u^2}(x,y,u) \ge \Lambda_l \ \ for\ a.a.\ \ x \in \Gamma \ \ and\ \ \forall (y,u) \in \mathbb{R}^2\,.$$

Let us also assume that, for every $M > 0$, there exists $C_{l,M} > 0$ such that

$$\left|\frac{\partial l}{\partial u}(x_2, y, u) - \frac{\partial l}{\partial u}(x_1, y, u)\right| + \left|\frac{\partial l}{\partial y}(x_2, y, u) - \frac{\partial l}{\partial y}(x_1, y, u)\right| \leq C_{l,M} \, |x_2 - x_1| \tag{3.3.10}$$

$\forall x_1, x_2 \in \Gamma$ and $|y|, |u| \leq M$. *Then for every $x \in \Gamma$ the equation*

$$\frac{\partial l}{\partial u}(x, \bar{y}(x), t) + \bar{\varphi}(x) = 0 \tag{3.3.11}$$

has a unique solution $\bar{t} =: \bar{s}(x)$. The function $\bar{s} : \Gamma \longrightarrow \mathbb{R}$ is related to $\bar{u}$ by the formula

$$\bar{u}(x) = \mathit{Proj}_{[u_a(x), u_b(x)]} \left\{\bar{s}(x)\right\} = \max\left\{\min\left\{u_b(x), \bar{s}(x)\right\}, u_a(x)\right\}. \tag{3.3.12}$$

Moreover, under the Assumption 1.17, the following regularity results are valid:

1. *If $\beta > 4/3$ and $\gamma \geq 2$ (see Assumption 3.6) then $\bar{s} \in H^1(\Gamma)$. Moreover, if $u_a, u_b \in C^{0,1/2}(\Gamma)$ (respectively $H^1(\Gamma)$), then $\bar{u} \in C^{0,1/2}(\Gamma)$ (respectively $H^1(\Gamma)$).*
2. *If Ω is convex, $2 < \min\{p, \beta, \gamma\}$ (see Assumption 3.6) and $u_a, u_b \in C^{0,1}(\Gamma)$, then $\bar{s}, \bar{u} \in C^{0,1}(\Gamma)$, and $\bar{y}, \bar{\varphi} \in W^{2,\varrho}(\Omega)$, with some $\varrho > 2$.*

PROOF. Let us recall that $\bar{y}, \bar{\varphi} \in W^{1,r}(\Omega) \subset C(\bar{\Omega})$. We fix $x \in \Gamma$ and consider the real function $g : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

$$g(t) = \bar{\varphi}(x) + \frac{\partial l}{\partial u}(x, \bar{y}(x), t).$$

Then g is a C^1 function with $g'(t) \geq \Lambda_l > 0$. Therefore, g is strictly increasing and

$$\lim_{t \to -\infty} g(t) = -\infty \quad \text{and} \quad \lim_{t \to +\infty} g(t) = +\infty.$$

Hence, there exists a unique element $\bar{t} \in \mathbb{R}$ satisfying $g(\bar{t}) = 0$, i.e. $\bar{s}$ is well defined. According to the definition of $\bar{d}$ in (3.3.7) and using (3.3.8) as well as the strict monotonicity of $(\partial l/\partial u)$ with respect to the third variable, we obtain

$$\begin{cases} \text{if } \bar{d}(x) = 0 & \text{then } \bar{u}(x) = \bar{s}(x), \\ \text{if } \bar{d}(x) < 0 & \text{then } u_b(x) = \bar{u}(x) < \bar{s}(x), \\ \text{if } \bar{d}(x) > 0 & \text{then } u_a(x) = \bar{u}(x) > \bar{s}(x), \end{cases}$$

which implies (3.3.12).

Proof of (1). First we observe that $\bar{y}, \bar{\varphi} \in H^{3/2}(\Omega) \subset C(\bar{\Omega})$. The continuity of $\bar{s}$ at every point $x \in \Gamma$ follows from the continuity of the functions $\bar{y}$, $\bar{\varphi}$ and $(\partial l/\partial u)$, by using the estimate

$$
\begin{aligned}
(3.3.13)\quad |\bar{s}(x)-\bar{s}(x')| &\le \frac{1}{\Lambda_l}\left|\frac{\partial l}{\partial u}(x',\bar{y}(x'),\bar{s}(x))-\frac{\partial l}{\partial u}(x',\bar{y}(x'),\bar{s}(x'))\right| \\
&\le \frac{1}{\Lambda_l}\left(|\bar{\varphi}(x')-\bar{\varphi}(x)|+\left|\frac{\partial l}{\partial u}(x',\bar{y}(x'),\bar{s}(x))-\frac{\partial l}{\partial u}(x,\bar{y}(x),\bar{s}(x))\right|\right) \quad \text{for } x'\in\Gamma\,.
\end{aligned}
$$

Next we show that $\bar{s}\in H^1(\Gamma)\subset C^{0,1/2}(\Gamma)$. Remark that Γ is a 1-dimensional manifold, hence a function v is in $H^1(\Gamma)$ if and only if v is continuous at every vertex x_j of Ω, $j=0,...,N(\Gamma)$ and $x_0=x_{N(\Gamma)}$, and on every edge e of Ω there holds $v|_e\in H^1(e)$. Therefore, it is enough to show that $\bar{s}|_e\in H^1(e)$ on every edge e of Ω. To this aim, we will prove that $\bar{s}$ is absolutely continuous on e. Then it is known that $\bar{s}$ is differentiable a.e. on e and that $\bar{s}'$ coincides with the weak derivative of $\bar{s}$. Let us take $a,b\in e$, arbitrarily. From (3.3.9)-(3.3.11) we get

$$
\begin{aligned}
|\bar{s}(b)-\bar{s}(a)| &\le \frac{1}{\Lambda_l}\left|\frac{\partial l}{\partial u}(a,\bar{y}(a),\bar{s}(b))-\frac{\partial l}{\partial u}(a,\bar{y}(a),\bar{s}(a))\right| \\
&\le \frac{C}{\Lambda_l}\left(|b-a|+|\bar{\varphi}(b)-\bar{\varphi}(a)|+|\bar{y}(b)-\bar{y}(a)|\right).
\end{aligned}
$$

Hence, the absolute continuity of $\bar{s}$ follows from the absolute continuity of the restriction of $\bar{y}$ and $\bar{\varphi}$ to Γ. Here we have used the fact that $\bar{y},\bar{\varphi}\in H^{3/2}(\Omega)$, thus by Proposition 1.14 on page 15, $\bar{y}|_\Gamma,\bar{\varphi}|_\Gamma\in H^1(\Gamma)$, and every function belonging to $H^1(I)$ on a given interval $I\subset\mathbb{R}$ is absolutely continuous; see, for instance, Rudin [**85**]. Therefore, $\bar{s}$ is differentiable a.e. on Γ and by differentiating (3.3.11) with $\bar{s}(x)$ substituted for t, we find for a.a. $x\in\Gamma$ that

$$
|\bar{s}'(x)|\le\frac{1}{\Lambda_l}\left|\frac{\partial l^2}{\partial x\partial u}(x,\bar{y}(x),\bar{s}(x))+\frac{\partial l^2}{\partial y\partial u}(x,\bar{y}(x),\bar{s}(x))\bar{y}'(x)+\bar{\varphi}'(x)\right|.
$$

Since the tangential derivatives $\bar{y}'$ and $\bar{\varphi}'$ belong to $L^2(\Gamma)$, we deduce from the previous inequality that $\bar{s}'\in L^2(\Gamma)$, consequently $\bar{s}\in H^1(\Gamma)$. The rest of the statement (1) follows immediately from the identity (3.3.12) and the regularity of $\bar{s}$, provided that u_a and u_b belong to $C^{0,1/2}(\Gamma)$ (respectively $H^1(\Gamma)$).

Proof of (2). Let us take $\varrho=\min\{r_0,r,p,\beta,\gamma,4\}>2$, where r_0 is introduced in Theorem 1.21 on page 18 and r is given in Assumption 3.6. Now let us prove that $\bar{s},\bar{u}\in W^{1-1/\varrho,\varrho}(\Gamma)$. The traces of $\bar{y}$ and $\bar{\varphi}$ belong to $W^{1-1/\varrho,\varrho}(\Gamma)$. Taking the intrinsic norm in $W^{1-1/\varrho,\varrho}(\Gamma)$, the $W^{1-1/\varrho,\varrho}(\Gamma)$ regularity of $\bar{u}$ and $\bar{s}$ follows from (3.3.12)-(3.3.13). Indeed, (3.3.13) implies that

$$
\frac{|\bar{s}(x)-\bar{s}(x')|^\varrho}{|x-x'|^\varrho}\le C\left(1+\frac{|\bar{\varphi}(x)-\bar{\varphi}(x')|}{|x-x'|}+\frac{|\bar{y}(x)-\bar{y}(x')|}{|x-x'|}\right)^\varrho =: z(x,x')
$$

and $\displaystyle\int_\Gamma\int_\Gamma z(x,x')\,d\sigma(x)d\sigma(x') < \infty$, hence $\bar{s} \in W^{1-1/\varrho,\varrho}(\Gamma)$. On the other hand, the projection $\mathrm{Proj}_{[u_a(\cdot),u_b(\cdot)]}$ is Lipschitz with Lipschitz constant 1, therefore

$$\begin{aligned} |\bar{u}(x)-\bar{u}(x')| &= |\mathrm{Proj}_{[u_a(x),u_b(x)]}\{\bar{s}(x)\} - \mathrm{Proj}_{[u_a(x'),u_b(x')]}\{\bar{s}(x')\}| \\ &\le |\mathrm{Proj}_{[u_a(x),u_b(x)]}\{\bar{s}(x)\} - \mathrm{Proj}_{[u_a(x),u_b(x)]}\{\bar{s}(x')\}| \\ &\quad + |\mathrm{Proj}_{[u_a(x),u_b(x)]}\{\bar{s}(x')\} - \mathrm{Proj}_{[u_a(x'),u_b(x')]}\{\bar{s}(x')\}| \\ &\le |\bar{s}(x)-\bar{s}(x')| + |u_a(x)-u_a(x')| + |u_b(x)-u_b(x')|\,. \end{aligned} \tag{3.3.14}$$

Consequently,

$$\begin{aligned} \frac{|\bar{u}(x)-\bar{u}(x')|^\varrho}{|x-x'|^\varrho} &\le \left(\frac{|\bar{s}(x)-\bar{s}(x')|}{|x-x'|} + \frac{|u_a(x)-u_a(x')|}{|x-x'|} + \frac{|u_b(x)-u_b(x')|}{|x-x'|}\right)^\varrho \\ &\le \left(\frac{|\bar{s}(x)-\bar{s}(x')|}{|x-x'|} + L_a + L_b\right)^\varrho, \end{aligned}$$

where L_a and L_b are the Lipschitz constants of u_a and u_b, respectively. This leads to the $W^{1-1/\varrho,\varrho}(\Gamma)$ regularity of $\bar{u}$.

Since Ω is assumed convex, Theorem 1.21 yields $\bar{y} \in W^{2,\varrho}(\Omega) \subset C^{0,1}(\bar{\Omega})$ (Eq. (1.5.7) on page 18). The function $\bar{\varphi}$ has also the same regularity. Next we show that $(\partial l/\partial y)(\cdot,\bar{y},\bar{u}) \in W^{1-1/\varrho,\varrho}(\Gamma)$ which implies, along with $(\partial L/\partial y)(\cdot,\bar{y}) \in L^\varrho(\Omega)$ and Theorem 1.36-(3) on page 32, the $W^{2,\varrho}(\Omega)$ regularity of $\bar{\varphi}$. From (3.3.10) and Assumption 3.6 we get $(\partial l/\partial y)(\cdot,\bar{y},\bar{u}) \in L^\varrho(\Gamma)$ and

$$\begin{aligned} &\left|\frac{\partial l}{\partial y}(x,\bar{y}(x),\bar{u}(x)) - \frac{\partial l}{\partial y}(x',\bar{y}(x'),\bar{u}(x'))\right| \\ &\qquad\qquad \le C\left(|x-x'| + |\bar{y}(x)-\bar{y}(x')| + |\bar{u}(x)-\bar{u}(x')|\right), \end{aligned}$$

for a.a. $x, x' \in \Gamma$. As above, it follows

$$\begin{aligned} &\frac{\left|\dfrac{\partial l}{\partial y}(x,\bar{y}(x),\bar{u}(x)) - \dfrac{\partial l}{\partial y}(x',\bar{y}(x'),\bar{u}(x'))\right|^\varrho}{|x-x'|^\varrho} \\ &\qquad\qquad \le C_\varrho\left(1 + \frac{|\bar{y}(x)-\bar{y}(x')|^\varrho}{|x-x'|^\varrho} + \frac{|\bar{u}(x)-\bar{u}(x')|^\varrho}{|x-x'|^\varrho}\right), \end{aligned}$$

therefore $(\partial l/\partial y)(\cdot,\bar{y},\bar{u}) \in W^{1-1/\varrho,\varrho}(\Gamma)$. Since $2 < \varrho$, we can use again the inclusion $W^{2,\varrho}(\Omega) \subset C^{0,1}(\bar{\Omega})$ and the inequalities (3.3.13)-(3.3.14) to deduce the Lipschitz regularity of $\bar{s}$ and $\bar{u}$. □

Remark that inequality (3.3.9) implies the strict convexity of l with respect to the third variable.

REMARK 3.11. *The previous theorem is also valid for non-convex and non-polygonal domains Ω, assuming the $C^{1,1}$-regularity of Γ; see Grisvard* [**60**].

3.4. Pontryagin's principle

In this section, we derive the Pontryagin's principle satisfied by a local solution of $(\mathcal{P})$. This principle is needed for the second-order analysis which we will carry out in Section 3.5. In contrast to the first-order necessary optimality conditions of integral form (Eq. (3.3.6)), Pontryagin's principle does not require neither the convexity nor the differentiability of the cost function J w.r.t. the controls.

Since equation (1.1.1) is not monotone, we have to adapt the known results for monotone equations to our situation. To overcome the difficulties arising from the lack of monotonicity we shall rely on the following assumption.

Assumption 3.12. *The functions $L : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ and $l : \Gamma \times (\mathbb{R} \times \mathbb{R}) \longrightarrow \mathbb{R}$ are of class C^1 with respect to the second variable, $2 < r$ satisfies (1.3.13) and, for any $M > 0$, there exist functions $\psi_{\Omega,M} \in L^{\frac{2r}{r+2}}(\Omega)$ and $\psi_{\Gamma,M} \in L^{r/2}(\Gamma)$ such that*

$$\left|\frac{\partial L}{\partial y}(x,y)\right| \le \psi_{\Omega,M}(x) \quad \textit{and} \quad \left|\frac{\partial l}{\partial y}(s,y,u)\right| \le \psi_{\Gamma,M}(s)$$

hold for a.a. $x \in \Omega$ and all $s \in \Gamma$, $|y|, |u| \le M$.

Let us introduce the Hamiltonian H associated with the control problem $(\mathcal{P})$:

$$H : \Gamma \times (\mathbb{R} \times \mathbb{R} \times \mathbb{R}) \longrightarrow \mathbb{R}\,, \quad H(x,y,u,\varphi) = l(x,y,u) + \varphi u\,.$$

Pontryagin's principle says that a local solution of $(\mathcal{P})$ minimizes the Hamiltonian at almost every point $x \in \Gamma$, hence it is a stronger condition than the variational inequality (3.3.6).

Theorem 3.13. *Let $\bar{u}$ be a local solution of $(\mathcal{P})$ with associated state $\bar{y}$ and suppose that Assumption 3.12 holds. Then there exists $\bar{\varphi} \in W^{1,r}(\Omega)$ satisfying the adjoint equation (3.3.5) and the minimum condition*

$$H(x,\bar{y}(x),\bar{u}(x),\bar{\varphi}(x)) = \min_{s \in [u_{a_{\bar{\varepsilon}}}(x), u_{b_{\bar{\varepsilon}}}(x)]} H(x,\bar{y}(x),s,\bar{\varphi}(x)) \quad \textit{for a.a. } x \in \Gamma\,, \tag{3.4.1}$$

where

$$u_{a_{\bar{\varepsilon}}}(x) := \max\{u_a(x), \bar{u}(x) - \bar{\varepsilon}\}\,, \quad u_{b_{\bar{\varepsilon}}}(x) := \min\{u_b(x), \bar{u}(x) + \bar{\varepsilon}\}\,,$$

and $\bar{\varepsilon} > 0$ is the radius of the ball in $L^\infty(\Gamma)$ where J achieves the (local) minimum value at $\bar{u}$ among all admissible controls.

Remark 3.14. *If l is convex with respect to the control variable then (3.4.1) follows immediately from the variational inequality (3.3.6).*

To prove the preceding theorem first the sensitivity of the state with respect to certain pointwise perturbations of the control is studied. The following auxiliary results are crucial to accomplish these perturbations.

LEMMA 3.15. ([**14**, Proposition 2]) *For $\rho \in (0,1)$ there exists a sequence of σ-measurable sets $\{E_k\}_{k=1}^{\infty}$, with $E_k \subset \Gamma$ and $\sigma(E_k) = \rho\sigma(\Gamma)$, such that $(1/\rho)\chi_{E_k} \rightharpoonup 1$ weakly* in $L^\infty(\Gamma)$ when $k \to \infty$.*

LEMMA 3.16. *Under the assumptions of Theorem 3.13, for any $u \in L^\infty(\Gamma)$ there exists a number $\hat{\rho} \in (0,1)$ and σ-measurable sets E_ρ, with $\sigma(E_\rho) = \rho\sigma(\Gamma)$ for all $\rho \in (0,\hat{\rho})$, having the following properties: If we define*

$$u_\rho(x) = \begin{cases} \bar{u}(x) & \text{if } x \in \Gamma \setminus E_\rho\,, \\ u(x) & \text{if } x \in E_\rho\,, \end{cases}$$

then

$$y_\rho = \bar{y} + \rho z + r_\rho \quad \text{with} \quad \lim_{\rho\searrow 0} \frac{1}{\rho}\|r_\rho\|_{W^{1,r}(\Omega)} = 0\,, \tag{3.4.2}$$

$$J(u_\rho) = J(\bar{u}) + \rho J^0 + r_\rho^0 \quad \text{with} \quad \lim_{\rho\searrow 0} \frac{1}{\rho}|r_\rho^0| = 0\,, \tag{3.4.3}$$

hold true, where y_ρ is the state associated with u_ρ, z is the unique element of $W^{1,r}(\Omega)$ satisfying the linearized equation

$$\left\{ \begin{aligned} -\operatorname{div}\left[a(x,\bar{y})\nabla z + \frac{\partial a}{\partial y}(x,\bar{y})z\nabla\bar{y}\right] + \frac{\partial f}{\partial y}(x,\bar{y})z &= \ 0 && \text{in } \Omega\,, \\ \left[a(x,\bar{y})\nabla z + \frac{\partial a}{\partial y}(x,\bar{y})z\nabla\bar{y}\right]\cdot\nu &= \ u - \bar{u} && \text{on } \Gamma\,, \end{aligned} \right. \tag{3.4.4}$$

and

$$J^0 := \int_\Omega \frac{\partial L}{\partial y}(x,\bar{y})z\,dx + \int_\Gamma \left\{\frac{\partial l}{\partial y}(x,\bar{y},\bar{u})z + l(x,\bar{y},u) - l(x,\bar{y},\bar{u})\right\} d\sigma(x)\,.$$

PROOF. Since the proof is similar to that of Proposition 4.3 in Casas and Tröltzsch [**36**], we only comment upon the main differences. We define $g \in L^1(\Gamma)$ by

$$g(x) = l(x,\bar{y}(x),u(x)) - l(x,\bar{y}(x),\bar{u}(x))\,.$$

Given $\rho \in (0,1)$, we take a sequence $\{E_k\}_{k=1}^{\infty}$ as in Lemma 3.15. Since $L^\infty(\Gamma)$ is compactly embedded in $W^{-1/r,r}(\Gamma)$, we have $(1/\rho)\chi_{E_k} \to 1$ strongly in $W^{-1/r,r}(\Gamma)$ when $k \to \infty$. Hence, there exists $k_\rho \in \mathbb{N}$ such that

$$\left|\int_\Gamma\left(1 - \frac{1}{\rho}\chi_{E_k}(x)\right)g(x)\,d\sigma(x)\right| + \left\|(1 - \frac{1}{\rho}\chi_{E_k})(u - \bar{u})\right\|_{W^{-1/r,r}(\Gamma)} < \rho \quad \forall k \geq k_\rho\,. \tag{3.4.5}$$

Inequality (3.4.5) is the analog of [**36**, Eq. (4.7)]. The same argumentation as in the proof of [**36**, Proposition 4.3] yields (3.4.2). In order to prove (3.4.3), we introduce $z_\rho := (y_\rho - \bar{y})/\rho$,

$$L_\rho(x) := \int_0^1 \frac{\partial L}{\partial y}(x,\bar{y}(x) + \theta(y_\rho(x) - \bar{y}(x)))\,d\theta$$

and

$$l_\rho(x) := \int_0^1 \frac{\partial l}{\partial y}(x, \bar{y}(x) + \theta(y_\rho(x) - \bar{y}(x)), u_\rho(x))\, d\theta\,.$$

Now recalling the definition of g and using (3.4.2) and (3.4.5), we have

$$\begin{aligned}
&\frac{J(u_\rho) - J(\bar{u})}{\rho} = \\
&= \int_\Omega \frac{L(x, y_\rho(x)) - L(x, \bar{y}(x))}{\rho} dx + \int_\Gamma \frac{l(x, y_\rho(x), u_\rho(x)) - l(x, \bar{y}(x), \bar{u}(x))}{\rho} d\sigma(x) \\
&= \int_\Omega \frac{L(x, y_\rho(x)) - L(x, \bar{y}(x))}{\rho} dx + \int_\Gamma \frac{l(x, y_\rho(x), u_\rho(x)) - l(x, \bar{y}(x), u_\rho(x))}{\rho} d\sigma(x) \\
&\quad + \int_\Gamma \frac{l(x, \bar{y}(x), u_\rho(x)) - l(x, \bar{y}(x), \bar{u}(x))}{\rho} d\sigma(x) \\
&= \int_\Omega L_\rho(x) z_\rho(x)\, dx + \int_\Gamma \left\{ l_\rho(x) z_\rho(x) + \frac{1}{\rho}\chi_{E_\rho}(x) g(x) \right\} d\sigma(x) \\
&\to \int_\Omega \frac{\partial L}{\partial y}(x, \bar{y}(x)) z(x)\, dx + \int_\Gamma \left\{ \frac{\partial l}{\partial y}(x, \bar{y}(x), \bar{u}(x)) z(x) + g(x) \right\} d\sigma(x) = J^0
\end{aligned}$$

as $\rho \to 0$ which implies (3.4.3). □

Proof of Theorem 3.13. Since $\bar{u}$ is a local solution of problem $(\mathcal{P})$, there exists $\bar{\varepsilon} > 0$ such that J achieves the minimum at $\bar{u}$ among all admissible controls of $\overline{B}_{L^\infty(\Gamma)}(\bar{u}, \bar{\varepsilon})$. Let u be one of these controls, i.e. $u \in \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \bar{\varepsilon}) \cap U_{ad}$. By virtue of Lemma 3.16, we consider sets $E_\rho \subset \Gamma$, $\rho > 0$, such that (3.4.2) and (3.4.3) hold. Then $u_\rho \in \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \bar{\varepsilon})$ and (3.4.3) leads to

$$0 \leq \lim_{\rho \to 0} \frac{J(u_\rho) - J(\bar{u})}{\rho} = J^0\,.$$

By using the definition of J^0, the variational formulation of (3.4.4) and the adjoint state equation (3.3.5), we get from the previous inequality

$$\begin{aligned}
0 &\leq \int_\Omega \frac{\partial L}{\partial y}(x, \bar{y}) z\, dx + \int_\Gamma \left\{ \frac{\partial l}{\partial y}(x, \bar{y}, \bar{u}) z + l(x, \bar{y}, u) - l(x, \bar{y}, \bar{u}) \right\} d\sigma(x) \\
&= \int_\Omega \left\{ a(x, \bar{y}) \nabla\bar{\varphi} \cdot \nabla z + \frac{\partial a}{\partial y}(x, \bar{y}) z \nabla\bar{\varphi} \cdot \nabla\bar{y} + \frac{\partial f}{\partial y}(x, \bar{y}) \bar{\varphi} z \right\} dx \\
&\quad + \int_\Gamma \{ l(x, \bar{y}, u) - l(x, \bar{y}, \bar{u}) \} d\sigma(x) \\
&= \int_\Gamma \{ \bar{\varphi}(x) \left(u(x) - \bar{u}(x) \right) + l(x, \bar{y}, u) - l(x, \bar{y}, \bar{u}) \} d\sigma(x) \\
&= \int_\Gamma \{ H(x, \bar{y}(x), u(x), \bar{\varphi}(x)) - H(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x)) \} d\sigma(x)\,.
\end{aligned} \tag{3.4.6}$$

Since $u \in \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \bar{\varepsilon}) \cap U_{ad}$ is arbitrary and taking into account the definitions of $u_{a_{\bar{\varepsilon}}}$ and $u_{b_{\bar{\varepsilon}}}$ given in the statement of Theorem 3.13, we deduce from (3.4.6)

$$\int_\Gamma H(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x))\, d\sigma(x) = \min_{u_{a_{\bar{\varepsilon}}} \le u \le u_{b_{\bar{\varepsilon}}}} \int_\Gamma H(x, \bar{y}(x), u(x), \bar{\varphi}(x))\, d\sigma(x)\,. \tag{3.4.7}$$

It remains to prove that (3.4.7) implies (3.4.1). To this end, we follow Casas and Tröltzsch [**36**, § 4] and Casas [**14**]. Knowing that Γ is a Lipschitz manifold, we consider a finite collection of σ-measurable sets $\{\Gamma_k\}_{k=1}^d$ and functions $\{\psi_k\}_{k=1}^d$ with the following properties:

(1) $\bigcup_{k=1}^d \Gamma_k = \Gamma$, $\mathring{\Gamma}_i \cap \mathring{\Gamma}_j = \emptyset$ if $i \neq j$, and $\sigma(\Gamma) = \sum_{k=1}^d \sigma(\mathring{\Gamma}_k)$.
(2) The functions $\psi_k : (0,1) \longrightarrow \mathbb{R}$ are Lipschitz and for some coordinate system $(x'_k, x_{k,2}) = (x_{k,1}, x_{k,2})$ in $\mathbb{R}^2$ we have that $\mathring{\Gamma}_k = \{(x'_k, \psi_k(x'_k)) \,|\, x'_k \in (0,1)\}$ and, for every set $E = \{(x'_k, \psi_k(x'_k)) \,|\, x'_k \in F\}$ with $F \subset (0,1)$ Lebesgue measurable, there holds $\sigma(E) = \int_F \sqrt{1 + |\psi'_k(x'_k)|^2}\, dx'_k$.

Now let the sequence $\{q_j\}_{j=1}^\infty$ exhaust the rational numbers contained in $[0,1]$. For every j we set $u_j(x) = q_j u_{a_{\bar{\varepsilon}}}(x) + (1 - q_j) u_{b_{\bar{\varepsilon}}}(x)$, $x \in \Gamma$. Then every function u_j belongs to $L^\infty(\Gamma)$ and $u_{a_{\bar{\varepsilon}}}(x) \le u_j(x) \le u_{b_{\bar{\varepsilon}}}(x)$ for every $x \in \Gamma$. Next we consider the functions $F_0, F_j : \Gamma \longrightarrow \mathbb{R}$ given by

$$F_0(x) = H(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x)) \text{ and } F_j(x) = H(x, \bar{y}(x), u_j(x), \bar{\varphi}(x))\,,\ j \in \mathbb{N}\,.$$

According to [**14**, Lemma 3], associated with these integrable functions, there exist σ-measurable sets S_0 and $\{S_j\}_{j=1}^\infty$ of Lebesgue regular points which satisfy $S_i \subset \bigcup_{k=1}^d \mathring{\Gamma}_k$, $\sigma(S_i) = \sigma(\Gamma)$ for $i = 0, 1, ...,$ and

$$\lim_{\varepsilon \searrow 0} \frac{1}{\sigma(\Gamma_\varepsilon(x_0))} \int_{\Gamma_\varepsilon(x_0)} F_i(x)\, d\sigma(x) = F_i(x_0) \quad \forall x_0 \in S_i\,, \tag{3.4.8}$$

where, given $x_0 = (x'_{0k}, \psi_k(x'_{0k})) \in \mathring{\Gamma}_k$, $1 \le k \le d$, and $\varepsilon > 0$ small enough, we set $\Gamma_\varepsilon(x_0) = \{(x'_k, \psi_k(x'_k)) \,|\, x'_k \in (x'_{0k} - \varepsilon, x'_{0k} + \varepsilon) \subset (0,1)\}$. Setting $S = \bigcap_{i=0}^\infty S_i$, we have $\sigma(S) = \sigma(\Gamma)$ and (3.4.8) for every $x_0 \in S$. Taking $x_0 \in S$ and $\varepsilon > 0$, we define

$$u_{j,\varepsilon}(x) = \begin{cases} \bar{u}(x) & \text{if } x \notin \Gamma_\varepsilon(x_0)\,, \\ u_j(x) & \text{otherwise}\,, \end{cases} \qquad j \in \mathbb{N}\,.$$

Then from (3.4.7) and the definition of $u_{j,\varepsilon}$ we deduce

$$\int_\Gamma H(x, \bar{y}(x), \bar{u}(x), \bar{\varphi}(x))\, d\sigma(x) \le \int_\Gamma H(x, \bar{y}(x), u_{j,\varepsilon}(x), \bar{\varphi}(x))\, d\sigma(x)\,,$$

hence

$$\frac{1}{\sigma(\Gamma_\varepsilon(x_0))}\int_{\Gamma_\varepsilon(x_0)} H(x,\bar y(x),\bar u(x),\bar\varphi(x))\,d\sigma(x) \\ \leq \frac{1}{\sigma(\Gamma_\varepsilon(x_0))}\int_{\Gamma_\varepsilon(x_0)} H(x,\bar y(x),u_{j,\varepsilon}(x),\bar\varphi(x))\,d\sigma(x)\,.$$

Passing to the limit in the last inequality when $\varepsilon \to 0$, we get, along with (3.4.8),

$$H(x_0,\bar y(x_0),\bar u(x_0),\bar\varphi(x_0)) \leq H(x_0,\bar y(x_0),u_j(x_0),\bar\varphi(x_0))\,.$$

Since the function $s \longmapsto H(x_0,\bar y(x_0),s,\bar\varphi(x_0))$ is continuous (l is a Carathéodory function; see page 76) and $\{u_j(x_0)\}_{j=1}^\infty$ is dense in $[u_{a_{\bar\varepsilon}}(x_0),u_{b_{\bar\varepsilon}}(x_0)]$, we infer

$$H(x_0,\bar y(x_0),\bar u(x_0),\bar\varphi(x_0)) \leq H(x_0,\bar y(x_0),s,\bar\varphi(x_0)) \quad \forall s \in [u_{a_{\bar\varepsilon}}(x_0),u_{b_{\bar\varepsilon}}(x_0)]\,.$$

Finally, (3.4.1) follows from the previous inequality and the fact that x_0 is an arbitrary point of S. □

3.5. Second-order optimality conditions

In this section, we prove necessary and sufficient second-order optimality conditions for the problem $(\mathcal{P})$. As pointed out at the beginning of this chapter, second-order optimality conditions are very important to analyze the convergence properties of numerical optimization algorithms which are used to solve the control problem. On the other hand, they are also a key tool to carry out error estimates for local solutions of the finite element approximated optimal control problem; see Chapter 4.

Throughout this section, we suppose that Assumption 3.6 holds which implies Assumption 3.12, therefore we can apply Pontryagin's principle; see Theorem 3.13.

If $\bar u$ is an admissible control for problem $(\mathcal{P})$, with associated state $\bar y$ and adjoint state $\bar\varphi$ satisfying (3.3.5) and (3.3.6), then the so-called *cone of critical directions* $C_{\bar u}$ is given by

$$C_{\bar u} = \left\{ v \in L^\infty(\Gamma)\,|\,v(x) \begin{cases} \geq 0 & \text{if } \bar u(x) = u_a(x) \\ \leq 0 & \text{if } \bar u(x) = u_b(x) \\ = 0 & \text{if } \bar d(x) \neq 0 \end{cases} \quad \text{for } x \in \Gamma \right\},$$

where $\bar d$ is defined in (3.3.7). In the previous section, we introduced the Hamiltonian H associated with our problem $(\mathcal{P})$ which obviously satisfies

$$\frac{\partial H}{\partial u}(x,\bar y(x),\bar u(x),\bar\varphi(x)) = \bar d(x)\,.$$

In the sequel, we will use the shorter notations

$$\bar H_u(x) = \frac{\partial H}{\partial u}(x,\bar y(x),\bar u(x),\bar\varphi(x)) \quad \text{and} \quad \bar H_{uu}(x) = \frac{\partial^2 H}{\partial u^2}(x,\bar y(x),\bar u(x),\bar\varphi(x))\,.$$

The following theorem deals with the necessary second-order optimality conditions.

THEOREM 3.17. *Let $\bar{u}$ be a local optimal solution of $(\mathcal{P})$. Then the following inequalities hold:*

$$\begin{cases} J''(\bar{u})v^2 \geq 0 & \forall v \in C_{\bar{u}}\,, \\ \bar{H}_{uu}(x) \geq 0 & \textit{for a.a. } x \in \Gamma \textit{ such that } \bar{H}_u(x) = 0\,. \end{cases} \tag{3.5.1}$$

PROOF. To prove the first inequality of (3.5.1) we will argue as in the proof of Theorem 5.1 in Casas and Tröltzsch [**36**]. Given $v \in C_{\bar{u}}$ arbitrary and $\varepsilon \in (0, \bar{\varepsilon})$, with $\bar{\varepsilon}$ chosen as in Theorem 3.13, we define

$$v_\varepsilon(x) = \begin{cases} 0 & \text{if} \quad u_a(x) < \bar{u}(x) < u_a(x) + \varepsilon \;\text{ or }\; u_b(x) - \varepsilon < \bar{u}(x) < u_b(x)\,, \\ \max\left\{-\dfrac{1}{\varepsilon}, \min\left\{+\dfrac{1}{\varepsilon}, v(x)\right\}\right\} & \qquad \text{otherwise}\,. \end{cases}$$

Obviously, $v_\varepsilon \in C_{\bar{u}}$ and $v_\varepsilon \to v$ in $L^2(\Gamma)$ when $\varepsilon \to 0$. Moreover,

$$u_a(x) \leq \bar{u}(x) + t v_\varepsilon(x) \leq u_b(x) \quad \text{for a.a.} \;\; x \in \Gamma \;\text{ and }\; 0 < t < \varepsilon^2\,.$$

Hence, as a consequence of the local optimality of $\bar{u}$, we have for $g_\varepsilon : [0, \varepsilon^2] \longrightarrow \mathbb{R}$, $g_\varepsilon(t) := J(\bar{u} + t v_\varepsilon)$, that

$$g_\varepsilon(0) = \min_{t \in [0, \varepsilon^2]} g_\varepsilon(t)\,.$$

Thanks to our assumptions, it is clear that g_ε is of class C^2. From the fact that $v_\varepsilon \in C_{\bar{u}}$ we deduce that

$$g'_\varepsilon(0) = J'(\bar{u})v_\varepsilon = \int_\Gamma \bar{d}(x) v_\varepsilon(x)\, d\sigma(x) = 0\,.$$

Since the first derivative of g_ε is zero, the following second-order necessary optimality condition must hold:

$$\begin{aligned} 0 \leq g''_\varepsilon(0) &= J''(\bar{u})v_\varepsilon^2 \\ &= \int_\Gamma \left\{ \frac{\partial^2 l}{\partial y^2}(x, \bar{y}, \bar{u}) z_{v_\varepsilon}^2 + 2\frac{\partial^2 l}{\partial y \partial u}(x, \bar{y}, \bar{u}) z_{v_\varepsilon} v_\varepsilon + \frac{\partial^2 l}{\partial u^2}(x, \bar{y}, \bar{u}) v_\varepsilon^2 \right\} d\sigma(x) \\ &\quad + \int_\Omega \left\{ \left(\frac{\partial^2 L}{\partial y^2}(x, \bar{y}) - \bar{\varphi} \frac{\partial^2 f}{\partial y^2}(x, \bar{y}) \right) z_{v_\varepsilon}^2 \right. \\ &\quad \left. - \nabla \bar{\varphi} \cdot \left(\frac{\partial^2 a}{\partial y^2}(x, \bar{y}) z_{v_\varepsilon}^2 \nabla \bar{y} + 2 \frac{\partial a}{\partial y}(x, \bar{y}) z_{v_\varepsilon} \nabla z_{v_\varepsilon} \right) \right\} dx\,, \end{aligned} \tag{3.5.2}$$

where the last equality in (3.5.2) follows from (3.3.3) with $z_{v_\varepsilon} := G'(\bar{u})v_\varepsilon$. In order to prove the first inequality of (3.5.1), we will pass to the limit in (3.5.2). First, the convergence $v_\varepsilon \to v$ in $L^2(\Gamma)$ implies that $z_{v_\varepsilon} \to z_v$ in $H^1(\Omega)$ with $z_v := G'(\bar{u})v$; notice that the inverse operator of $\partial_y F(y_u, u)$ introduced on page 25 is an isomorphism.

Next we estimate each term on the right-hand side of (3.5.2). Taking into account the embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$ $\forall q \in [1, \infty)$ and Assumption 1.24-(1), we get

$$\begin{aligned}\int_\Omega \left| \frac{\partial^2 a}{\partial y^2}(x,\bar{y}) z_{v_\varepsilon}^2 \nabla\bar{y} \cdot \nabla\bar{\varphi} \right| dx &\le D_M \|z_{v_\varepsilon}\|^2_{L^{\frac{2r}{r-2}}(\Omega)} \|\nabla \bar{y}\|_{L^r(\Omega)} \|\nabla\bar{\varphi}\|_{L^r(\Omega)} \\ &\le C \|z_{v_\varepsilon}\|^2_{H^1(\Omega)} \|\bar{y}\|_{W^{1,r}(\Omega)} \|\bar{\varphi}\|_{W^{1,r}(\Omega)} ,\end{aligned}$$

with r given in Assumption 3.6, and

$$\begin{aligned}\int_\Omega \left| \frac{\partial a}{\partial y}(x,\bar{y}) z_{v_\varepsilon} \nabla z_{v_\varepsilon} \cdot \nabla\bar{\varphi} \right| dx &\le D_M \|z_{v_\varepsilon} \nabla\bar{\varphi}\|_{L^2(\Omega)} \|\nabla z_{v_\varepsilon}\|_{L^2(\Omega)} \\ &\le D_M \|z_{v_\varepsilon}\|_{L^{\frac{2r}{r-2}}(\Omega)} \|\nabla\bar{\varphi}\|_{L^r(\Omega)} \|\nabla z_{v_\varepsilon}\|_{L^2(\Omega)} \\ &\le C \|z_{v_\varepsilon}\|^2_{H^1(\Omega)} \|\bar{\varphi}\|_{W^{1,r}(\Omega)} .\end{aligned}$$

The remaining terms in (3.5.2) can be estimated analogously, with the help of the Assumptions 1.24-(1) and 3.6. We only remark that for the integral terms over Γ we make additionally use of the embedding $H^{1/2}(\Gamma) \hookrightarrow L^q(\Gamma)$ for every $q \in [1, \infty)$. Finally, we pass to the limit in (3.5.2) and deduce

$$0 \le \lim_{\varepsilon \to 0} J''(\bar{u}) v_\varepsilon^2 = J''(\bar{u}) v^2 .$$

This yields the first inequality of (3.5.1). The second inequality follows directly from (3.4.1). Indeed, it is an easy and well known conclusion of (3.4.1) that

$$\bar{H}_u(x) \begin{cases} \ge 0 & \text{if } \bar{u}(x) = u_a(x) , \\ \le 0 & \text{if } \bar{u}(x) = u_b(x) , \\ = 0 & \text{if } u_a(x) < \bar{u}(x) < u_b(x) , \end{cases} \qquad \text{for a.a. } x \in \Gamma \tag{3.5.3}$$

and $\bar{H}_{uu}(x) \ge 0$ for a.a. $x \in \Gamma$ such that $\bar{H}_u(x) = 0$. □

In optimization theory, second-order optimality conditions are conveniently expressed in terms of a Lagrange function associated with $(\mathcal{P})$

$$\mathcal{L} : W^{1,r}(\Omega) \times L^\infty(\Gamma) \times W^{1,r}(\Omega) \longrightarrow \mathbb{R}$$

defined by

$$\begin{aligned}\mathcal{L}(y,u,\varphi) &= \mathcal{J}(y,u) - \int_\Omega \{a(x,y)\nabla y \cdot \nabla\varphi + \varphi f(x,y)\}\, dx + \int_\Gamma \varphi u \, d\sigma(x) \\ &= \int_\Gamma H(x, y(x), u(x), \varphi(x))\, d\sigma(x) + \int_\Omega \{L(x,y) - a(x,y)\nabla y \cdot \nabla\varphi - \varphi f(x,y)\}\, dx ,\end{aligned}$$

where r is given in Assumption 3.6 and

$$\mathcal{J}(y,u) := \int_\Omega L(x,y)\, dx + \int_\Gamma l(x,y,u)\, d\sigma(x) .$$

In the next step, we shall formulate the second-order necessary optimality conditions involving the Lagrange function and Hamiltonian. Defining $\bar{H}_y$, $\bar{H}_{yy}$ and $\bar{H}_{yu}$, similarly to $\bar{H}_u$ and $\bar{H}_{uu}$, we can write the first and second order derivatives of $\mathcal{L}$ w.r.t. (y,u) as follows

$$\begin{aligned} D_{(y,u)}\mathcal{L}(\bar{y},\bar{u},\bar{\varphi})(z,v) &= \int_\Gamma \Big\{\bar{H}_y(x)z(x) + \bar{H}_u(x)v(x)\Big\}\,d\sigma(x) \\ &+ \int_\Omega \left\{\frac{\partial L}{\partial y}(x,\bar{y})z - \bar{\varphi}\frac{\partial f}{\partial y}(x,\bar{y})z - \nabla\bar{\varphi}\cdot\left(a(x,\bar{y})\nabla z + \frac{\partial a}{\partial y}(x,\bar{y})z\nabla\bar{y}\right)\right\}dx\,. \end{aligned}$$

Invoking the adjoint equation (3.3.5), we obtain

$$D_{(y,u)}\mathcal{L}(\bar{y},\bar{u},\bar{\varphi})(z,v) = \int_\Gamma \bar{H}_u(x)v(x)\,d\sigma(x)\,.$$

Moreover, there holds

$$\begin{aligned} D^2_{(y,u)}\mathcal{L}(\bar{y},\bar{u},\bar{\varphi})(z,v)^2 &= \int_\Gamma \Big\{\bar{H}_{yy}(x)z^2(x) + 2\bar{H}_{yu}(x)zv + \bar{H}_{uu}(x)v^2(x)\Big\}\,d\sigma(x) \\ + \int_\Omega \bigg\{\frac{\partial^2 L}{\partial y^2}(x,\bar{y})z^2 &- \bar{\varphi}\frac{\partial f^2}{\partial y^2}(x,\bar{y})z^2 - \nabla\bar{\varphi}\cdot\left(\frac{\partial^2 a}{\partial y^2}(x,\bar{y})z^2\nabla\bar{y} + 2\frac{\partial a}{\partial y}(x,\bar{y})z\nabla z\right)\bigg\}\,dx\,. \end{aligned}$$

If we take $z = G'(\bar{u})v$ we deduce from (3.3.3)

$$J''(\bar{u})v^2 = D^2_{(y,u)}\mathcal{L}(\bar{y},\bar{u},\bar{\varphi})(z,v)^2\,, \tag{3.5.4}$$

hence we can rewrite the condition (3.5.1) as follows

$$\begin{cases} D^2_{(y,u)}\mathcal{L}(\bar{y},\bar{u},\bar{\varphi})(z,v)^2 \geq 0 & \forall (z,v) \in H^1(\Omega)\times C_{\bar{u}} \text{ satisfying } z = G'(\bar{u})v \\ \bar{H}_{uu}(x) \geq 0 & \text{for a.a. } x \in \Gamma \text{ such that } \bar{H}_u(x) = 0\,. \end{cases}$$

The next theorem is the main result of this section and it provides the second-order sufficient optimality conditions for $(\mathcal{P})$. We will employ here the technique devised by Casas et al. [**17**] to obtain optimality conditions having a form similar to the ones in the theory of nonlinear optimization in finite-dimensional spaces. For an analogous distributed control problem this technique is used by Casas and Tröltzsch [**36**].

Theorem 3.18. *Let $\bar{u}$ be an admissible control for problem $(\mathcal{P})$ and $\bar{\varphi} \in W^{1,r}(\Omega)$ satisfying (3.3.5) and (3.3.6). We also assume that there exist $\mu > 0$ and $\tau > 0$ such that*

$$\begin{cases} J''(\bar{u})v^2 > 0 & \forall v \in C_{\bar{u}} \setminus \{0\}\,, \\ \bar{H}_{uu}(x) \geq \mu & \textit{if } |\bar{H}_u(x)| \leq \tau \textit{ for a.a. } x \in \Gamma\,. \end{cases} \tag{3.5.5}$$

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

$$J(\bar{u}) + \frac{\delta}{2}\|u - \bar{u}\|^2_{L^2(\Gamma)} \leq J(u)$$

for every admissible control u for $(\mathcal{P})$ with $\|u-\bar{u}\|_{L^\infty(\Gamma)} \le \varepsilon$.

PROOF. The proof follows basically the ideas by Casas and Tröltzsch [**36**, Theorem 5.2].

Step 1: Preparations. We will argue by contradiction. Let us assume that there exists a sequence $\{u_k\}_{k=1}^\infty \subset L^\infty(\Gamma)$ of admissible controls for $(\mathcal{P})$ such that

$$\|u_k-\bar{u}\|_{L^\infty(\Gamma)} < \frac{1}{k} \quad \text{and} \quad J(\bar{u}) + \frac{1}{k}\|u_k-\bar{u}\|^2_{L^2(\Gamma)} > J(u_k) \quad \forall k \in \mathbb{N}\,. \tag{3.5.6}$$

Let us define

$$y_k = G(u_k) = y_{u_k}\,,\ \bar{y} = G(\bar{u}) = y_{\bar{u}}\,,\ \rho_k = \|u_k-\bar{u}\|_{L^2(\Gamma)} \ \text{ and } \ v_k = \frac{1}{\rho_k}(u_k-\bar{u})\,,$$

then

$$\lim_{k\to\infty} \|y_k-\bar{y}\|_{W^{1,r}(\Omega)} = 0\,,\ \lim_{k\to\infty}\rho_k = 0 \ \text{ and } \ \|v_k\|_{L^2(\Gamma)} = 1 \quad \forall k \in \mathbb{N}\,, \tag{3.5.7}$$

where r is given in Assumption 3.6. By taking a subsequence if necessary, we can assume that $v_k \rightharpoonup v$ weakly in $L^2(\Gamma)$. In the second step of this proof, we are going to show that $v \in C_{\bar{u}}$. To this aim, we prove first the following convergence result

$$\lim_{k\to\infty} \frac{1}{\rho_k}(y_k-\bar{y}) = z \quad \text{in } H^1(\Omega)\,,$$

where $z = G'(\bar{u})v$. By setting $z_k = (y_k-\bar{y})/\rho_k$, subtracting the state equations satisfied by y_k and $\bar{y}$, dividing by ρ_k and applying the mean value theorem, we find

$$\left\{\begin{aligned} &-\operatorname{div}\left[a(x,y_k)\nabla z_k + \frac{\partial a}{\partial y}(x,\bar{y}+\theta_k(y_k-\bar{y}))z_k\nabla\bar{y}\right] \\ &\qquad\qquad\qquad\qquad + \frac{\partial f}{\partial y}(x,\bar{y}+\eta_k(y_k-\bar{y}))z_k = 0 \quad \text{in } \Omega\,, \\ &\qquad \left[a(x,y_k)\nabla z_k + \frac{\partial a}{\partial y}(x,\bar{y}+\theta_k(y_k-\bar{y}))z_k\nabla\bar{y}\right]\cdot\nu = v_k \quad \text{on } \Gamma\,. \end{aligned}\right. \tag{3.5.8}$$

Notice that θ_k and η_k are functions depending on the space variable and their measurability can be shown by following the argumentation on page 62. We multiply (3.5.8) by z_k and make an integration by parts to get with the assumptions on a and f and the Poincaré inequality (1.3.3) on page 4 that

$$\begin{aligned} \min\{\alpha_a,\alpha_f\}\|z_k\|^2_{H^1(\Omega)} &\le \int_\Omega \left\{a(x,y_k)|\nabla z_k|^2 + \frac{\partial f}{\partial y}(x,\bar{y}+\eta_k(y_k-\bar{y}))z_k^2\right\} dx \\ &= \int_\Gamma v_k z_k\, d\sigma(x) - \int_\Omega \frac{\partial a}{\partial y}(x,\bar{y}+\theta_k(y_k-\bar{y}))z_k\nabla z_k\cdot\nabla\bar{y}\, dx \\ &\le \|v_k\|_{L^2(\Gamma)}\|z_k\|_{L^2(\Gamma)} + C\|z_k\|_{L^{\frac{2r}{r-2}}(\Omega)}\|\nabla\bar{y}\|_{L^r(\Omega)}\|\nabla z_k\|_{L^2(\Omega)}\,. \end{aligned}$$

From the previous estimate and (3.5.7) we deduce

$$\|z_k\|_{H^1(\Omega)} \leq C\left(1 + \|z_k\|_{L^{\frac{2r}{r-2}}(\Omega)}\right).$$

As indicated on page 21, the previous inequality implies that $\{z_k\}_{k=1}^{\infty}$ is bounded in $H^1(\Omega)$. This leads to the existence of a subsequence, denoted in the same way, such that $z_k \rightharpoonup z$ weakly in $H^1(\Omega)$. Thanks to the compact embedding $H^1(\Omega) \hookrightarrow L^{\frac{2r}{r-2}}(\Omega)$, the convergence $z_k \to z$ is strong in $L^{\frac{2r}{r-2}}(\Omega) \subset L^2(\Omega)$,

$$\frac{\partial a}{\partial y}(x, \bar{y} + \theta_k(y_k - \bar{y}))z_k \nabla \bar{y} \to \frac{\partial a}{\partial y}(x, \bar{y})z\nabla\bar{y} \quad \text{in} \ \ L^2(\Omega)^2$$

and

$$\frac{\partial f}{\partial y}(x, \bar{y} + \eta_k(y_k - \bar{y}))z_k \to \frac{\partial f}{\partial y}(x, \bar{y})z \quad \text{in} \ \ L^2(\Omega)\,.$$

Therefore, we can pass to the limit in (3.5.8) and deduce

$$\text{(3.5.9)} \qquad \left\{\begin{aligned} -\text{div}\left[a(x,\bar{y})\nabla z + \frac{\partial a}{\partial y}(x,\bar{y})z\nabla\bar{y}\right] + \frac{\partial f}{\partial y}(x,\bar{y})z = 0 \quad \text{in}\ \Omega\,, \\ \left[a(x,\bar{y})\nabla z + \frac{\partial a}{\partial y}(x,\bar{y})z\nabla\bar{y}\right]\cdot\nu = v \quad \text{on}\ \Gamma\,, \end{aligned}\right.$$

hence $z = G'(\bar{u})v$. It remains to show the strong convergence $z_k \to z$ in $H^1(\Omega)$. To this aim, it is enough to prove that

$$\int_\Omega a(x,\bar{y})\,|\nabla z_k|^2\,dx \to \int_\Omega a(x,\bar{y})\,|\nabla z|^2\,dx \quad \text{ as } \ k \to \infty\,,$$

which, along with the strong convergence $z_k \to z$ in $L^2(\Omega)$, implies the strong convergence $z_k \to z$ in $H^1(\Omega)$. Taking into account (3.5.8) and (3.5.9), we have

$$\begin{aligned} \int_\Omega a(x,\bar{y})\left||\nabla z_k|^2 - |\nabla z|^2\right| dx &\leq \int_\Omega |a(x,\bar{y}) - a(x,y_k)|\,|\nabla z_k|^2\,dx \\ &+ \int_\Gamma |v_k z_k - vz|\,d\sigma(x) + \int_\Omega \left|\frac{\partial f}{\partial y}(x,\bar{y}+\eta_k(y_k-\bar{y}))z_k^2 - \frac{\partial f}{\partial y}(x,\bar{y})z^2\right| dx \\ &+ \int_\Omega |\nabla\bar{y}|\left|\frac{\partial a}{\partial y}(x,\bar{y}+\theta_k(y_k-\bar{y}))z_k\nabla z_k - \frac{\partial a}{\partial y}(x,\bar{y})z\nabla z\right| dx \to 0\,, \end{aligned}$$

as an easy consequence of the strong convergences $y_k \to \bar{y}$ in $W^{1,r}(\Omega) \subset C(\bar{\Omega})$, $z_k \to z$ in $L^{\frac{2r}{r-2}}(\Omega)$ and the weak convergences $z_k \rightharpoonup z$ in $H^1(\Omega)$ and $v_k \rightharpoonup v$ in $L^2(\Gamma)$.

Step 2: $v \in C_{\bar{u}}$. From the fact that $u_a(x) \leq u_k(x) \leq u_b(x)$ for a.a. $x \in \Gamma$, we have that $v_k \geq 0$ if $\bar{u}(x) = u_a(x)$ and $v_k(x) \leq 0$ if $\bar{u}(x) = u_b(x)$ a.e on Γ. Since the set of functions satisfying these sign conditions is convex and closed in $L^2(\Gamma)$, it is weakly closed. Therefore, the weak limit v of $\{v_k\}_{k=1}^{\infty}$ satisfies these sign conditions, too. It remains to prove that $v(x) = 0$ for a.a. $x \in \Gamma$ such that $\bar{d}(x) \neq 0$. By using the mean value theorem, we obtain from (3.5.6)

$$\begin{aligned}\frac{\rho_k}{k} = \frac{1}{k\rho_k}\|u_k - \bar{u}\|^2_{L^2(\Gamma)} > \frac{J(u_k) - J(\bar{u})}{\rho_k} &= \int_\Omega \frac{\partial L}{\partial y}(x, \bar{y} + \nu_k(y_k - \bar{y}))z_k\,dx \\ &+ \int_\Gamma \frac{\partial l}{\partial y}(x, \bar{y} + \theta_k(y_k - \bar{y}), \bar{u} + \theta_k(u_k - \bar{u}))z_k\,d\sigma(x) \\ &+ \int_\Gamma \frac{\partial l}{\partial u}(x, \bar{y} + \theta_k(y_k - \bar{y}), \bar{u} + \theta_k(u_k - \bar{u}))v_k\,d\sigma(x)\,,\end{aligned}$$

where ν_k and θ_k are measurable functions with values in $[0, 1]$. Taking the limits in both sides of the previous inequality, using the convergences $z_k \to z$ in $H^1(\Omega)$, $y_k \to \bar{y}$ in $C(\bar{\Omega})$, $u_k \to \bar{u}$ in $L^\infty(\Gamma)$, Assumption 3.6, the adjoint equation (3.3.5), (3.5.9), and integrating by parts, we infer

$$\begin{aligned}0 &\geq \int_\Omega \frac{\partial L}{\partial y}(x, \bar{y})z\,dx + \int_\Gamma \left\{\frac{\partial l}{\partial y}(x, \bar{y}, \bar{u})z + \frac{\partial l}{\partial u}(x, \bar{y}, \bar{u})v\right\}d\sigma(x) \\ &= \int_\Gamma \left(\frac{\partial l}{\partial u}(x, \bar{y}, \bar{u}) + \bar{\varphi}\right)v\,d\sigma(x) = \int_\Gamma \bar{d}(x)v(x)\,d\sigma(x) = \int_\Gamma |\bar{d}(x)|\,|v(x)|\,d\sigma(x)\,.\end{aligned}$$

The last equality is a consequence of the signs of v and $\bar{d}$ (Eq. (3.3.8)). The previous inequality implies that $|\bar{d}(x)v(x)| = 0$ holds a.e. on Γ, hence $v(x) = 0$ whenever $\bar{d}(x) \neq 0$, as we wanted to prove.

Step 3: $v = 0$. In this step, our goal is to prove that v does not satisfy the first condition of (3.5.5). This leads immediately to the identity $v = 0$ and then to the final contradiction. By using the definition of the Lagrange function $\mathcal{L}$, (3.5.6) and the fact that y_k and $\bar{y}$ are the states corresponding to u_k and $\bar{u}$, respectively, we get

$$\begin{aligned}\mathcal{L}(y_k, u_k, \bar{\varphi}) = \mathcal{J}(y_k, u_k) &< \mathcal{J}(\bar{y}, \bar{u}) + \frac{1}{k}\|u_k - \bar{u}\|^2_{L^2(\Gamma)} \\ &= \mathcal{L}(\bar{y}, \bar{u}, \bar{\varphi}) + \frac{1}{k}\|u_k - \bar{u}\|^2_{L^2(\Gamma)}\,.\end{aligned} \tag{3.5.10}$$

Performing a Taylor expansion up to the second order, we obtain

$$\begin{aligned}&\mathcal{L}(y_k, u_k, \bar{\varphi}) = \mathcal{L}(\bar{y} + \rho_k z_k, \bar{u} + \rho_k v_k, \bar{\varphi}) \\ &= \mathcal{L}(\bar{y}, \bar{u}, \bar{\varphi}) + \rho_k D_{(y,u)}\mathcal{L}(\bar{y}, \bar{u}, \bar{\varphi})(z_k, v_k) + \frac{\rho_k^2}{2}D^2_{(y,u)}\mathcal{L}(\bar{y} + \nu_k\rho_k z_k, \bar{u} + \theta_k\rho_k v_k, \bar{\varphi})(z_k, v_k)^2\end{aligned}$$

with functions ν_k and θ_k having the same properties as in the second step of the proof. Employing (3.5.10) and (3.5.6), the last equality leads to

$$\rho_k D_{(y,u)}\mathcal{L}(\bar{y}, \bar{u}, \bar{\varphi})(z_k, v_k) + \frac{\rho_k^2}{2}D^2_{(y,u)}\mathcal{L}(\xi_k, w_k, \bar{\varphi})(z_k, v_k)^2 < \frac{1}{k}\|u_k - \bar{u}\|^2_{L^2(\Gamma)} = \frac{\rho_k^2}{k}\,,$$

where $\xi_k := \bar{y} + \nu_k \rho_k z_k$ and $w_k := \bar{u} + \theta_k \rho_k v_k$. Obviously, $\xi_k \to \bar{y}$ in $W^{1,r}(\Omega)$ and $w_k \to \bar{u}$ in $L^\infty(\Gamma)$ as $k \to \infty$. Taking into account the expressions obtained for the derivatives of $\mathcal{L}$, it follows, after dividing the previous inequality by ρ_k^2,

$$\begin{aligned} \frac{1}{\rho_k}\int_\Gamma \bar{H}_u(x) v_k \, d\sigma(x) &+ \frac{1}{2}\left(\int_\Gamma \left\{ H_{yy}^k(x) z_k^2 + 2H_{yu}^k(x)(z_k, v_k) + \ H_{uu}^k(x) v_k^2 \right\} d\sigma(x) \right. \\ &+ \int_\Omega \left\{ \frac{\partial^2 L}{\partial y^2}(x, \xi_k) z_k^2 - \bar{\varphi}\frac{\partial f^2}{\partial y^2}(x, \xi_k) z_k^2 \right. \\ &\left.\left. - \nabla\bar{\varphi}\cdot\left(\frac{\partial^2 a}{\partial y^2}(x,\xi_k) z_k^2 \nabla \xi_k + 2\frac{\partial a}{\partial y}(x,\xi_k) z_k \nabla z_k\right)\right\} dx\right) < \frac{1}{k}\,, \end{aligned} \tag{3.5.11}$$

where

$$H_{yy}^k(x) := H_{yy}(x, \xi_k(x), w_k(x), \bar{\varphi}(x))$$

with analogous definitions for $H_{yu}^k(x)$ and $H_{uu}^k(x)$. In view of the properties of $D^2_{(y,u)}l$ given in Assumption 3.6, it is easy to check that

$$\begin{aligned} &(H_{yy}^k(x), H_{yu}^k(x), H_{uu}^k(x)) \to (\bar{H}_{yy}(x), \bar{H}_{yu}(x), \bar{H}_{uu}(x)) \quad \text{as } k \to \infty \\ &\text{and} \quad |H_{yy}^k(x)| + |H_{yu}^k(x)| + |H_{uu}^k(x)| \le C \quad \text{for a.a. } x \in \Gamma \end{aligned}$$

and some constant $C > 0$. The following convergence properties can also be verified easily

$$\begin{cases} \dfrac{\partial^j a}{\partial y^j}(\cdot, \xi_k) z_k \nabla\bar{\varphi} \to \dfrac{\partial^j a}{\partial y^j}(\cdot, \bar{y}) z \nabla\bar{\varphi}\,,\ j = 1, 2, & \\ \nabla z_k \to \nabla z \quad \text{and} \quad z_k \nabla \xi_k \to z \nabla \bar{y} & \text{in } L^2(\Omega)^2 \quad \text{and} \\ \bar{\varphi}\dfrac{\partial^2 f}{\partial y^2}(\cdot, \xi_k) z_k \to \bar{\varphi}\dfrac{\partial^2 f}{\partial y^2}(\cdot, \bar{y}) z & \text{in } L^2(\Omega)\,. \end{cases}$$

Using the above properties, we can pass to the limit in (3.5.11) as follows

$$\begin{aligned} \limsup_{k\to\infty} &\left\{ \frac{1}{\rho_k}\int_\Gamma \bar{H}_u(x) v_k(x)\, d\sigma(x) + \frac{1}{2}\int_\Gamma H_{uu}^k(x) v_k^2(x)\, d\sigma(x) \right\} \\ &+ \frac{1}{2}\left(\int_\Gamma \left\{ \bar{H}_{yy}(x) z^2 + 2\bar{H}_{yu}(x)(z, v) \right\} d\sigma(x) + \int_\Omega \left\{ \frac{\partial^2 L}{\partial y^2}(x, \bar{y}) z^2 \right.\right. \\ &- \left.\left. \bar{\varphi}\frac{\partial f^2}{\partial y^2}(x, \bar{y}) z^2 - \ \nabla\bar{\varphi}\cdot\left(\frac{\partial^2 a}{\partial y^2}(x, \bar{y}) z^2 \nabla\bar{y} + 2\frac{\partial a}{\partial y}(x, \bar{y}) z \nabla z\right)\right\} dx \right) \le 0\,. \end{aligned} \tag{3.5.12}$$

Now we prove that $\frac{1}{2}\int_\Gamma \bar{H}_{uu}(x) v^2(x)\, d\sigma(x)$ is a lower bound of the above upper limit. Then from (3.5.12) and (3.5.4) it follows that $J''(\bar{u})v^2 = D^2_{(y,u)}\mathcal{L}(\bar{y}, \bar{u}, \bar{\varphi})(z, v)^2 \le 0$. Finally, according to (3.5.5), this is possible only if $v = 0$. To show the lower estimate mentioned above we make use of a convexity argument for which the second condition of (3.5.5) plays an essential role. The difficulty in deducing this estimate is due to the fact that we only have a weak convergence $v_k \rightharpoonup v$.

Using the convergence $H^k_{uu} \to \bar{H}_{uu}$ in $L^\infty(\Gamma)$, the identity $\bar{H}_u(x)v_k(x) = |\bar{H}_u(x)||v_k(x)|$ (Eq. (3.5.3)) and $\|v_k\|_{L^2(\Gamma)} = 1$, we get

$$\begin{aligned}
&\limsup_{k\to\infty}\left\{\frac{1}{\rho_k}\int_\Gamma \bar{H}_u(x)v_k(x)\,d\sigma(x) + \frac{1}{2}\int_\Gamma H^k_{uu}(x)v_k^2(x)\,d\sigma(x)\right\}\\
&\qquad = \limsup_{k\to\infty}\left\{\frac{1}{\rho_k}\int_\Gamma |\bar{H}_u(x)||v_k(x)|\,d\sigma(x) + \frac{1}{2}\int_\Gamma \bar{H}_{uu}(x)v_k^2(x)\,d\sigma(x)\right\}\\
&\qquad \geq \limsup_{k\to\infty}\left\{\int_{\{|\bar{H}_u(x)|>\tau\}}\left\{\frac{1}{\rho_k}|\bar{H}_u(x)||v_k(x)| + \frac{1}{2}\bar{H}_{uu}(x)v_k^2(x)\right\}d\sigma(x)\right.\\
&\qquad\qquad\qquad\qquad\left. + \frac{1}{2}\int_{\{|\bar{H}_u(x)|\leq\tau\}}\bar{H}_{uu}(x)v_k^2(x)\,d\sigma(x)\right\},
\end{aligned}\tag{3.5.13}$$

where τ is given in (3.5.5). Now thanks to $\rho_k\|v_k\|_{L^\infty(\Gamma)} = \|u_k - \bar{u}\|_{L^\infty(\Gamma)} < 1/k$,

$$\exists k_0 > 0 \text{ such that } \frac{\|\bar{H}_{uu}\|_{L^\infty(\Gamma)}\rho_k\|v_k\|_{L^\infty(\Gamma)}}{\tau} < \frac{\|\bar{H}_{uu}\|_{L^\infty(\Gamma)}}{k\tau} < 1 \quad \forall k \geq k_0\,,$$

therefore $\dfrac{\tau}{\rho_k}|v_k(x)| \geq \|\bar{H}_{uu}\|_{L^\infty(\Gamma)}v_k^2(x)$ for a.a. $x \in \Gamma$ and $\forall k \geq k_0$. Then we have

$$\begin{aligned}
&\limsup_{k\to\infty}\left\{\int_{\{|\bar{H}_u|>\tau\}}\left(\frac{1}{\rho_k}|\bar{H}_u||v_k| + \frac{1}{2}\bar{H}_{uu}v_k^2\right)d\sigma(x) + \frac{1}{2}\int_{\{|\bar{H}_u|\leq\tau\}}\bar{H}_{uu}v_k^2\,d\sigma(x)\right\}\\
&\geq \limsup_{k\to\infty}\left\{\int_{\{|\bar{H}_u|>\tau\}}\left(\|\bar{H}_{uu}\|_{L^\infty(\Gamma)} + \frac{1}{2}\bar{H}_{uu}\right)v_k^2\,d\sigma(x) + \frac{1}{2}\int_{\{|\bar{H}_u|\leq\tau\}}\bar{H}_{uu}v_k^2\,d\sigma(x)\right\}\\
&\qquad \geq \int_{\{|\bar{H}_u|>\tau\}}\left(\|\bar{H}_{uu}\|_{L^\infty(\Gamma)} + \frac{1}{2}\bar{H}_{uu}\right)v^2\,d\sigma(x) + \frac{1}{2}\int_{\{|\bar{H}_u|\leq\tau\}}\bar{H}_{uu}v^2\,d\sigma(x)\\
&\qquad \geq \frac{1}{2}\int_\Gamma \bar{H}_{uu}v^2\,d\sigma(x)\,.
\end{aligned}\tag{3.5.14}$$

Combining (3.5.13) and (3.5.14), we obtain the desired lower estimate.

Step 4: Final contradiction. Since $v = 0$, there holds $z = 0$. This fact, along with $\|v_k\|_{L^2(\Gamma)} = 1$, (3.5.12)-(3.5.14) and the second condition in (3.5.5), leads to

$$\begin{aligned}
0 &\geq \limsup_{k\to\infty}\left\{\int_{\{|\bar{H}_u|>\tau\}}\left(\|\bar{H}_{uu}\|_{L^\infty(\Gamma)} + \frac{1}{2}\bar{H}_{uu}\right)v_k^2\,d\sigma(x) + \frac{1}{2}\int_{\{|\bar{H}_u|\leq\tau\}}\bar{H}_{uu}v_k^2\,d\sigma(x)\right\}\\
&\geq \limsup_{k\to\infty}\left\{\frac{\|\bar{H}_{uu}\|_{L^\infty(\Gamma)}}{2}\int_{\{|\bar{H}_u|>\tau\}}v_k^2\,d\sigma(x) + \frac{\mu}{2}\int_{\{|\bar{H}_u|\leq\tau\}}v_k^2\,d\sigma(x)\right\}\\
&\geq \frac{\min\{\|\bar{H}_{uu}\|_{L^\infty(\Gamma)},\mu\}}{2}\limsup_{k\to\infty}\left\{\int_\Gamma v_k^2\,d\sigma(x)\right\}\\
&= \frac{\min\{\|\bar{H}_{uu}\|_{L^\infty(\Gamma)},\mu\}}{2} > 0\,,
\end{aligned}$$

yielding the contradiction we were looking for. □

By virtue of the identity (3.5.4), the second-order sufficient optimality condition (3.5.5) can be formulated in terms of the Lagrange function as follows

$$\begin{cases} D^2_{(y,u)}\mathcal{L}(\bar{y},\bar{u},\bar{\varphi})(z,v)^2 > 0 & \forall (z,v)\in (H^1(\Omega)\times C_{\bar{u}})\setminus\{(0,0)\} \text{ verifying } z = G'(\bar{u})v\,, \\ \bar{H}_{uu}(x) \ge \mu & \text{if } |\bar{H}_u(x)| \le \tau \text{ for a.a. } x\in\Gamma\,. \end{cases}$$

Remark 3.19. *A comparison of the first inequality of (3.5.5) with the analogous of (3.5.1) shows that the gap between the necessary and sufficient conditions is minimal. On the other hand, the second inequality of (3.5.5) is stronger than the corresponding one of (3.5.1). In general, we cannot take $\tau = 0$ in (3.5.5). The reader is referred to Dunn* [**52**]*, see also Casas et al.* [**30**, page 24]*, for a simple example showing the impossibility of taking $\tau = 0$.*

Remark 3.20. *The statement of Theorem 3.18 involves both the $L^2(\Gamma)$ and $L^\infty(\Gamma)$ norms. This phenomenon is called two-norm discrepancy: the differentiability of the cost function J requires the $L^\infty(\Gamma)$ norm (see Theorem 3.7), while the $L^2(\Gamma)$ norm is the natural one to deduce sufficient optimality condition for strict local minima.*

Remark 3.21. *Let us note that $\bar{H}_{uu}(x) = (\partial^2 l/\partial u^2)(x,\bar{y}(x),\bar{u}(x))$. Therefore, if the second derivative of l w.r.t. the third variable has a positive lower bound for a.a. $x\in\Gamma$ then the second condition of (3.5.5) is satisfied automatically. A standard example is given by the function*

$$l(x,y,u) = l_0(x,y) + \frac{\lambda}{2}u^2 \quad \textit{with } \lambda > 0\,.$$

In this case, controls satisfying the assumptions of Theorem 3.18 are locally optimal even in the sense of $L^2(\Gamma)$, i.e. the problem of the two-norm discrepancy is resolved. To prove this we follow the lines of the proof of Theorem 3.18 with the following differences. We assume that there exists a sequence $\{u_k\}_{k=1}^\infty \subset U_{ad}$ with

$$\|u_k - \bar{u}\|_{L^2(\Gamma)} < \frac{1}{k} \quad \textit{and} \quad J(\bar{u}) + \frac{1}{k}\|u_k-\bar{u}\|^2_{L^2(\Gamma)} > J(u_k) \quad \forall k\in\mathbb{N}\,,$$

instead of (3.5.6). Then, by using the identity $\bar{H}_{uu}(x)\equiv\lambda$ and $\|v_k\|_{L^2(\Gamma)} = 1$, we can considerably shorten the previous proof in the following way

$$\begin{aligned} 0 &\ge \limsup_{k\to\infty}\left\{\frac{1}{\rho_k}\int_\Gamma \bar{H}_u(x)v_k(x)\,d\sigma(x) + \frac{1}{2}\int_\Gamma H^k_{uu}(x)v_k^2(x)\,d\sigma(x)\right\} \\ &= \limsup_{k\to\infty}\left\{\frac{1}{\rho_k}\int_\Gamma |\bar{H}_u(x)||v_k(x)|\,d\sigma(x) + \frac{1}{2}\int_\Gamma \bar{H}_{uu}(x)v_k^2(x)\,d\sigma(x)\right\} \\ &\ge \limsup_{k\to\infty}\left\{\frac{\lambda}{2}\int_\Gamma v_k^2(x)\,d\sigma(x)\right\} = \frac{\lambda}{2}\,. \end{aligned}$$

This yields the desired contradiction.

We finish this section by giving another sufficient optimality condition equivalent to (3.5.5). The proof of this equivalence is carried out in Casas and Mateos [**25**, Theorem 4.4].

Theorem 3.22. *Let $\bar{u}$ be an admissible control for problem $(\mathcal{P})$ and $\bar{\varphi} \in W^{1,r}(\Omega)$ satisfying (3.3.5) and (3.3.6). Then (3.5.5) holds if and only if there exist $\delta > 0$ and $\rho > 0$ such that*

$$J''(\bar{u})h^2 \geq \delta \|h\|^2_{L^2(\Gamma)} \quad \forall h \in C^\rho_{\bar{u}}\,, \tag{3.5.15}$$

where

$$C^\rho_{\bar{u}} = \left\{ h \in L^2(\Gamma) \,|\, h(x) \begin{cases} \geq 0 & \textit{if } \ \bar{u}(x) = u_a(x) \\ \leq 0 & \textit{if } \ \bar{u}(x) = u_b(x) \\ = 0 & \textit{if } \ |\bar{d}(x)| > \rho \end{cases} \quad \textit{for a.a. } x \in \Gamma \right\}.$$

Remark 3.23. *Since $C^\rho_{\bar{u}}$ is strictly contained in $C_{\bar{u}}$, the reader might expect that the condition (3.5.15) is stronger than (3.5.5). The fact is that they are equivalent.*

Remark 3.24. *There are at least two advantages for studying sufficient conditions of the form (3.5.5). First, we can compare it with the necessary one given by (3.5.1). Second, in contrast to the equivalent condition (3.5.15), the proof of its sufficiency is close to the method known from the theory of nonlinear optimization in finite-dimensional spaces. Nevertheless, condition (3.5.15) is the one used for numerical purposes; in order to deduce error estimates for the control problem $(\mathcal{P})$, we will make explicit use of (3.5.15); see Section 4.3.*

CHAPTER 4

Numerical approximation of the control problem

4.1. Introduction

In this chapter, we focus on the numerical analysis of the following optimal boundary control problem

$$(\mathcal{P}) \begin{cases} \min J(u) = \displaystyle\int_\Omega L(x, y_u(x))\, dx + \int_\Gamma l(x, y_u(x), u(x))\, d\sigma(x)\,, \\ \text{subject to} \quad u \in U_{ad} = \{u \in L^\infty(\Gamma) \,|\, \alpha \leq u(x) \leq \beta \text{ a.e. } x \in \Gamma\}\,, \\ \qquad\qquad\quad (y_u, u) \text{ satisfying the state equation (1.1.1).} \end{cases}$$

To simplify the discussion below the box constraints α and β in the problem under consideration are now real numbers and $\alpha < \beta$.

Based on a standard finite element approximation, we will introduce a finite dimensional control problem $(\mathcal{P}_h)$ with $h > 0$. The state functions are discretized by linear finite element ansatz functions, i.e. by functions belonging to the set Y_h defined on page 37. Concerning the discretization of control functions, we consider piecewise constant controls.

At this point, we mention two other approaches for the approximation of $(\mathcal{P})$. The continuous piecewise linear approximation of Neumann controls of semilinear equations was studied carefully by Casas and Mateos [**27**]. Other contributions to this kind of approximation were made by Casas [**16**], concerning distributed problems, and by Casas and Raymond [**31**] for Dirichlet boundary control problems. Another approach is the variational discretization concept suggested by Hinze in [**63**] that was applied to linear quadratic distributed problems and extended in [**27**] to Neumann controls of semilinear equations. The idea is to discretize the state but not the control, thus an infinite dimensional optimization problem has to be solved. Theoretical results for both techniques in the context of distributed quasilinear control problems were recently obtained by Casas and Tröltzsch [**38**]. For a particular Neumann control problem in a convex domain some numerical results are presented in Section 4.4. In our tests, we observe the same rate of convergence of the approximations as related numerical tests for semilinear problems; cf. [**27**] and [**65**].

Our main aims in this chapter are twofold. First, we are going to study the convergence of a sequence $\{\bar{u}_h\}_{h>0}$ of local optimal controls for the discretized problem $(\mathcal{P}_h)$ to a local solution $\bar{u}$ of the continuous problem $(\mathcal{P})$; see Section 4.2. Second, we prove estimates for the error $\bar{u} - \bar{u}_h$ in the $L^2(\Gamma)$ and $L^\infty(\Gamma)$ norms; see Section 4.3. The chapter ends with some numerical experiments shown in Section 4.4.

The preceding chapters have revealed that the analysis of control problems governed by quasilinear equations of the type (1.1.1) is much more complicated than expected at first sight. In fact, they are much more difficult than the corresponding control problems governed by semilinear and monotone equations. Indeed, although we were able to prove that, under quite general assumptions, (1.1.1) has a unique solution (see Chapter 1), the non-monotone character of the equation introduced many difficulties which range from the linearization of (1.1.1) to the analysis of the discrete equation. To deal with the fact that solutions of the discrete equation are probably not unique we proved only local uniqueness as well as the differentiability of the mapping associating to each control its corresponding locally unique discrete state; see Section 2.6. These results, along with first- and second-order optimality conditions and some extra regularity of solutions of $(\mathcal{P})$ to be approximated, are essential ingredients in the proof of error estimates for optimal controls.

The case of a distributed control problem associated with a quasilinear equation similar to (1.1.1) was studied by Casas and Tröltzsch in [**36, 37, 38**]. However, as it is well known, the analysis for boundary controls is often more complicated than for the distributed ones because of the lower regularity of the states corresponding to boundary controls. Therefore, it has required a different approach to the one developed in [**37**] to carry out the numerical analysis of the control problem $(\mathcal{P})$. Our study is limited to polygonal domains in the plane, since the case of curved domains introduces some additional difficulties that are beyond the scope of this thesis. The reader is referred to Casas et al. [**24**], Casas and Sokolowski [**33**] and Deckelnick et al. [**47**] for the analysis of boundary control problems in curved domains. However, we do not assume the convexity of the domain as required in most of the previous papers dealing with error estimates for boundary control problems; see Casas and Mateos [**27**], Casas et al. [**29**], Casas and Raymond [**31**] and Krumbiegel et al. [**74**], all of them devoted to linear or semilinear equations. Because of the higher regularity of the state and adjoint state in convex domains, we obtain better error estimates for optimal controls under the convexity assumption; see Theorems 4.10 and 4.12. Though for the sake of brevity, we have considered only the approximation of the controls by using piecewise constant functions, other possibilities as described above are open; see Casas [**16**], Hinze [**63**] and Meyer and Rösch [**79**].

The control $\bar{u}$ will stand for a *strict* local minimum of the control problem $(\mathcal{P})$, i.e. there exists $\varepsilon > 0$ such that

$$J(u) > J(\bar{u}) \quad \forall u \in U_{ad} \cap \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon) \quad \text{with } u \neq \bar{u}\,. \tag{4.1.1}$$

Let us denote by $\bar{y}$ the state associated with $\bar{u}$.

Throughout this chapter, we suppose that the Assumptions 1.1-1.3, 1.17, 1.24 and 3.6 hold, with $\beta \geq p > 4/3$ and $\gamma > 2$ (see Assumption 3.6). Moreover, we assume that (3.3.9)-(3.3.10) on pages 80-81 are satisfied.

Let us remark that the assumption $\beta \geq p > 4/3$ is only made to avoid heavy notations and can be relaxed to $\min\{\beta, p\} > 4/3$. In the sequel, we will often consider the case when $p > 2$. Hence, $\beta \geq p > 2$ but, in fact, it is sufficient to require $\min\{\beta, p\} > 2$.

4.2. Approximation of the controls by piecewise constant functions

Let $\{\mathcal{T}_h\}_{h>0}$ be a family of triangulations of $\bar{\Omega}$, as defined on page 37, with maximum mesh size $h > 0$. For fixed h we denote by $\{T_j\}_{j=1}^{N(h)}$ the family of triangles of $\mathcal{T}_h$ with one side on Γ. If the vertices of $\bar{T}_j \cap \Gamma$ are x_j and x_{j+1} then $[x_j, x_{j+1}] := \bar{T}_j \cap \Gamma$, $1 \leq j \leq N(h)$, with $x_{N(h)+1} := x_1$.

Associated with the previous notation, we denote by $\mathcal{U}_h \subset L^\infty(\Gamma)$ the set of piecewise constant control functions, i.e.

$$\mathcal{U}_h = \left\{ u_h \in L^\infty(\Gamma) \,|\, u_h = \sum_{j=1}^{N(h)} u_j \chi_{(x_j, x_{j+1})}\,,\ u_j \in \mathbb{R} \right\}.$$

Let us take $\bar{q} \in (2, 4)$ arbitrary but fixed and h_0, ρ and κ_ρ, as in Theorem 2.23 on page 61. Further, $h_1 \leq h_0$ is as in Theorem 2.25 on page 63. Let $\varepsilon > 0$ be chosen small enough such that (4.1.1) and $0 < \varepsilon < \frac{\rho}{\sigma(\Gamma)^{1/2}}$ hold. Then we have

$$u \in \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon) \implies \|u - \bar{u}\|_{L^2(\Gamma)} \leq \varepsilon \sigma(\Gamma)^{1/2} < \rho \implies u \in B_{L^2(\Gamma)}(\bar{u}, \rho)\,.$$

Hence, according to Theorem 2.23, if $u \in \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon)$ the discrete state equation (2.3.1) on page 39 has, for every $h < h_0$, a unique solution $y_h(u) \in \overline{B}_{W^{1,\bar{q}}(\Omega)}(\bar{y}, \kappa_\rho) \cap Y_h$. Let us set

$$U_{ad,h} = \mathcal{U}_h \cap U_{ad} = \{u_h \in \mathcal{U}_h \,|\, \alpha \leq u_h \leq \beta \ \text{ a.e. on } \ \Gamma\}$$

and consider for $h < h_1$ the auxiliary discrete control problem

$$(\mathcal{P}_h^\varepsilon)\ \begin{cases} \min J_h(u_h) = \displaystyle\int_\Omega L(x, y_h(u_h))\, dx + \int_\Gamma l(x, y_h(u_h), u_h)\, d\sigma(x)\,, \\ \text{s.t. } \ u_h \in U_{ad,h} \cap \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon)\,, \\ \qquad (y_h(u_h), u_h) \ \text{ satisfying the discrete state equation (2.3.1) with } u = u_h\,. \end{cases}$$

We should underline that $\overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon)$ is an additional constraint on the controls. However, we will see later that this constraint is not active if h is small enough. Consequently, we will remove it and finally we will introduce below the discrete problem in a standard way.

By defining

$$\alpha_{h,j} = \max\{\alpha, \max_{x \in [x_j, x_{j+1}]} \bar{u}(x) - \varepsilon\} \quad \text{and} \quad \beta_{h,j} = \min\{\beta, \min_{x \in [x_j, x_{j+1}]} \bar{u}(x) + \varepsilon\}\,,$$

we have the following equivalence for $u_h \in U_{ad,h}$:

$$u_h \in \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon) \iff \alpha_{h,j} \le u_h|_{(x_j,x_{j+1})} \le \beta_{h,j} \quad \forall 1 \le j \le N(h)\,.$$

Theorem 4.1. *The discrete problem ($\mathcal{P}_h^\varepsilon$) admits at least one optimal solution.*

Proof. Since we are in finite dimension, the set of admissible controls is compact and non-empty in $\mathcal{U}_h$. This fact, along with the continuity of the functional J_h, yields the existence of a minimum of ($\mathcal{P}_h^\varepsilon$). □

The next two theorems are the counterparts of Theorems 3.7 and 3.9 and their proofs are analogous.

Theorem 4.2. *For every $h < h_1$ the functional $J_h : B_{L^\infty(\Gamma)}(\bar{u}, \varepsilon) \longrightarrow \mathbb{R}$ is of class C^2 and its derivative is given by*

$$J_h'(u)v = \int_\Gamma \left(\frac{\partial l}{\partial u}(x, y_h(u), u) + \varphi_h(u) \right) v \, d\sigma(x)\,, \tag{4.2.1}$$

where $\varphi_h(u) \in Y_h$ is the unique solution of the problem

$$\int_\Omega \Bigg\{ a(x, y_h(u)) \nabla\varphi_h(u) \cdot \nabla\phi_h + \frac{\partial a}{\partial y}(x, y_h(u)) \phi_h \nabla y_h(u) \cdot \nabla\varphi_h(u) \tag{4.2.2}$$
$$+ \frac{\partial f}{\partial y}(x, y_h(u)) \varphi_h(u) \phi_h \Bigg\} dx = \int_\Omega \frac{\partial L}{\partial y}(x, y_h(u)) \phi_h \, dx + \int_\Gamma \frac{\partial l}{\partial y}(x, y_h(u), u) \phi_h \, d\sigma(x)$$

for every $\phi_h \in Y_h$.

The existence and uniqueness of $\varphi_h(u) \in Y_h$ satisfying (4.2.2) is an immediate consequence of Theorem 2.27 on page 67.

Theorem 4.3. *For every $h < h_1$ let $\bar{u}_h$ be a local minimum of the problem ($\mathcal{P}_h^\varepsilon$) and $\bar{y}_h := y_h(\bar{u}_h)$. Then there exists a unique solution $\bar{\varphi}_h \in Y_h$ of the problem*

$$\int_\Omega \left\{ a(x, \bar{y}_h) \nabla\bar{\varphi}_h \cdot \nabla\phi_h + \left(\frac{\partial a}{\partial y}(x, \bar{y}_h) \nabla\bar{y}_h \cdot \nabla\bar{\varphi}_h + \frac{\partial f}{\partial y}(x, \bar{y}_h) \bar{\varphi}_h \right) \phi_h \right\} dx$$
$$= \int_\Omega \frac{\partial L}{\partial y}(x, \bar{y}_h) \phi_h \, dx + \int_\Gamma \frac{\partial l}{\partial y}(x, \bar{y}_h, \bar{u}_h) \phi_h \, d\sigma(x) \quad \forall \phi_h \in Y_h$$

and

$$\sum_{j=1}^{N(h)} \int_{x_j}^{x_{j+1}} \left\{ \frac{\partial l}{\partial u}(x, \bar{y}_h, \bar{u}_j) + \bar{\varphi}_h \right\} d\sigma(x) \, (u_j - \bar{u}_j) \ge 0 \quad \forall u_j \in [\alpha_{h,j}, \beta_{h,j}]\,, \tag{4.2.3}$$

where $\bar{u}_j = \bar{u}_h|_{(x_j,x_{j+1})}$. Moreover, there exist unique real numbers $\{\bar{s}_j\}_{j=1}^{N(h)}$ such that

$$\int_{x_j}^{x_{j+1}} \left\{ \frac{\partial l}{\partial u}(x, \bar{y}_h(x), \bar{s}_j) + \bar{\varphi}_h(x) \right\} d\sigma(x) = 0 \quad \forall 1 \le j \le N(h)\,. \tag{4.2.4}$$

Finally, the function $\bar{u}_h$ satisfies the following projection formula

$$\bar{u}_j = Proj_{\left[\alpha_{h,j},\beta_{h,j}\right]}\left\{\bar{s}_j\right\} = \max\left\{\min\left\{\beta_{h,j},\bar{s}_j\right\},\alpha_{h,j}\right\} \quad \forall 1 \le j \le N(h)\,. \tag{4.2.5}$$

The existence and uniqueness of the numbers $\bar{s}_j$ follow easily from inequality (3.3.9) on page 80.

Our next goal is to prove the convergence of solutions of $(\mathcal{P}_h^\varepsilon)$ to $\bar{u}$ in $L^\infty(\Gamma)$. For this purpose, we need some preparatory lemmas.

Lemma 4.4. *Let $0 < \tau < 1/2$ be fixed. Suppose that $h < h_2$, where h_2 is given as in Proposition 2.26, and $v, u \in U_{ad} \cap \overline{B}_{L^2(\Gamma)}(\bar{u},\rho)$. Then there holds the estimate*

$$\|y_u - y_h(v)\|_X + \|\varphi_u - \varphi_h(v)\|_X \le C\left(h^\sigma + \|u-v\|_{L^2(\Gamma)}\right), \tag{4.2.6}$$

with $C > 0$ being dependent on τ, and σ takes the following values

	Ω *non-convex*	Ω *convex*
$X = L^2(\Omega)$	$3/4$	1
$X = H^1(\Omega)$	$1/2$	$1/2$
$X = L^2(\Gamma)$	$5/8$	$3/4$
$X = L^\infty(\Gamma)$	$1/2-\tau$	$1/2-\tau$

Finally, if Ω is convex, $p > 2$ and $u \in W^{1-1/\varrho,\varrho}(\Gamma)$ with $\varrho > 2$, then

$$\|y_u - y_h(v)\|_{H^1(\Omega)} + \|\varphi_u - \varphi_h(v)\|_{H^1(\Omega)} \le C\left(h + \|u-v\|_{L^2(\Gamma)}\right), \tag{4.2.7}$$

and

$$\|y_u - y_h(v)\|_{L^\infty(\Gamma)} + \|\varphi_u - \varphi_h(v)\|_{L^\infty(\Gamma)} \le C\left(h^{1-\tau} + h^{-\tau}\|u-v\|_{L^2(\Gamma)}\right).$$

Proof. We split the proof in two parts.

Part 1: Estimates for the states. Let us consider the first two cases when Ω is convex or non-convex and deduce estimates for $y_u - y_h(v)$ in the norms of $L^2(\Omega)$ and $H^1(\Omega)$. Using the estimate (2.3.37) on page 52 we get

$$\begin{aligned}\|y_u - y_h(v)\|_{L^2(\Omega)} &\le \|y_u - y_h(u)\|_{L^2(\Omega)} + \|y_h(u) - y_h(v)\|_{L^2(\Omega)}\\ &\le C\left(h + \|u-v\|_{L^2(\Gamma)}\right).\end{aligned}$$

The estimation of the last term follows from the mean value theorem, Theorem 2.25 on page 63 as well as inequality (2.6.14) on page 66:

$$\begin{aligned}&\|y_h(u) - y_h(v)\|_{W^{1,\bar{q}}(\Omega)}\\ &\qquad \le \sup_{t\in[0,1]} \|G_h'(v+t(u-v))\|_{\mathcal{L}(L^2(\Gamma),W^{1,\bar{q}}(\Omega))}\|u-v\|_{L^2(\Gamma)}\,.\end{aligned} \tag{4.2.8}$$

Analogously, we prove the estimate in the $H^1(\Omega)$ norm. Let us focus now on the estimates in the norms defined on Γ. Taking in inequality (2.7.3) on page 67 $\varepsilon = h$ if Ω is non-convex and $\varepsilon = h^2$ if Ω is convex and arguing as above, it is a immediate consequence of Corollary 2.15 on page 52 that

$$\|y_u - y_h(v)\|_{L^2(\Gamma)} \leq C\left(h^\sigma + \|u - v\|_{L^2(\Gamma)}\right),$$

with $\sigma = 3/4$ if Ω is non-convex and $\sigma = 1$ if Ω is convex. This implies the estimates for the states in the $L^2(\Gamma)$ norm as given in the lemma. To derive the estimate for $\|y_u - y_h(v)\|_{L^\infty(\Gamma)}$ we make use of the embedding $W^{1,q}(\Omega) \hookrightarrow C(\bar{\Omega})$ $\forall q > 2$. For $q = \min\{\bar{q}, 2/(1-\tau)\}$ it follows $q \in (2,4)$, $(2/q) - (1/2) \geq (1/2) - \tau$ and there exists $C > 0$, dependent on q and consequently on τ, such that

$$\|y_u - y_h(v)\|_{L^\infty(\Gamma)} \leq C\|y_u - y_h(v)\|_{W^{1,q}(\Omega)}.$$

Using the triangle inequality, we obtain

$$\|y_u - y_h(v)\|_{W^{1,q}(\Omega)} \leq \|y_u - y_h(u)\|_{W^{1,q}(\Omega)} + \|y_h(u) - y_h(v)\|_{W^{1,q}(\Omega)}.$$

Again, for the second term we can apply (4.2.8). For the first one, we deduce from the estimate (2.3.38) on page 52

$$\|y_u - y_h(u)\|_{W^{1,q}(\Omega)} \leq Ch^{2/q-1/2} \leq Ch^{1/2-\tau}.$$

It remains to consider the case when Ω is convex, $p > 2$ and $u \in W^{1-1/\varrho,\varrho}(\Gamma)$ ($\varrho > 2$). Now $y_u \in W^{2,q}(\Omega)$ for some $2 < q \leq \min\{p, \varrho, \bar{q}, 2/(1-\tau)\}$; see Theorem 1.21 on page 18. Obviously, the estimate (4.2.7) for the states follows from (2.3.40) and the above argumentation. Finally, the $W^{2,q}(\Omega)$ regularity of y_u and an obvious modification of the proof of inequality (2.3.23) on page 47 yield in the same way as in the non-convex case that

$$\|y_u - y_h(v)\|_{L^\infty(\Gamma)} \leq C\left(h^{1-\tau} + \|u - v\|_{L^2(\Gamma)}\right).$$

Part 2: Estimates for the adjoint states. Once again we start by deducing the estimates for the adjoint states in the norms of $L^2(\Omega)$ and $H^1(\Omega)$ when Ω is convex or non-convex. From the inequalities (2.7.24)-(2.7.25) on page 71, Corollary 1.37 on page 33 and taking into account the estimates for the states proved above, we obtain

$$\begin{aligned}\|\varphi_u - \varphi_h(v)\|_{L^2(\Omega)} &\leq \|\varphi_u - \varphi_v\|_{L^2(\Omega)} + \|\varphi_v - \varphi_h(v)\|_{L^2(\Omega)}\\ &\leq C\left(\|u - v\|_{L^2(\Gamma)} + h^\sigma\right),\end{aligned}$$

where σ attains the values given in the statement of the lemma. The estimate in the $H^1(\Omega)$ norm is obtained in the same way. For the $L^2(\Gamma)$ error we can proceed as for the states to deduce the desired estimates. Finally, let us consider the error in the $L^\infty(\Gamma)$ norm. Taking once again $q = \min\{\bar{q}, 2/(1-\tau)\} > 2$, along with the triangle inequality,

$$\begin{aligned}\|\varphi_u - \varphi_h(v)\|_{L^\infty(\Gamma)} &\leq C\|\varphi_u - \varphi_h(v)\|_{W^{1,q}(\Omega)}\\ &\leq C\left(\|\varphi_u - \varphi_v\|_{W^{1,q}(\Omega)} + \|\varphi_v - \varphi_h(v)\|_{W^{1,q}(\Omega)}\right).\end{aligned}$$

Following the proof of (2.3.23), we observe that the order of convergence of the second summand in the previous inequality is $\mathcal{O}(h^{1/2-\tau})$; notice that $\varphi_v \in H^{3/2}(\Omega)$. This fact, along with Corollary 1.37, yields the estimates we were looking for.

Let us study now the more delicate case when Ω is convex, $u \in W^{1-1/\varrho,\varrho}(\Gamma)$ ($\varrho > 2$) and $p > 2$. The $W^{2,\hat{q}}(\Omega)$ regularity of φ_u, with some $\hat{q} > 2$, can be shown as in the proof of Theorem 3.10-(2); see pages 82-83. However, in contrast to the states, we do not have an inequality analogous to (4.2.8) for the adjoint states. In order to benefit from the higher regularity of φ_u, we will argue in a different way. Let us consider first the error in the $H^1(\Omega)$ norm. The triangle inequality

$$\|\varphi_u - \varphi_h(v)\|_{H^1(\Omega)} \leq \|\varphi_u - \varphi_h(u)\|_{H^1(\Omega)} + \|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)}$$

delivers the order $\mathcal{O}(h)$ of convergence for the first summand; compare also the estimate (2.7.26) on page 72. For the second one we are going to prove that

$$\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)} \leq C\left(h + \|u - v\|_{L^2(\Gamma)}\right). \tag{4.2.9}$$

Once this is shown, we conclude (4.2.7). For the estimation of $\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)}$ we subtract the equations satisfied by both functions, insert $\varphi_h(u) - \varphi_h(v)$ as test function and arrive at

$$\begin{aligned}
&\|\varphi_h(u) - \varphi_h(v)\|^2_{H^1(\Omega)} \\
&\leq C\int_\Omega\left(a(x, y_h(u))\,|\nabla\varphi_h(u) - \nabla\varphi_h(v)|^2 + \frac{\partial f}{\partial y}(x, y_h(u))\,(\varphi_h(u) - \varphi_h(v))^2\right)dx \\
&= C\Bigg(\int_\Omega (a(x, y_h(v)) - a(x, y_h(u)))\,\nabla\varphi_h(v)\cdot(\nabla\varphi_h(u) - \nabla\varphi_h(v))\,dx \\
&\quad + \int_\Omega\left(\frac{\partial a}{\partial y}(x, y_h(v))\nabla y_h(v)\cdot\nabla\varphi_h(v) - \frac{\partial a}{\partial y}(x, y_h(u))\nabla y_h(u)\cdot\nabla\varphi_h(u)\right) \\
&\times(\varphi_h(u) - \varphi_h(v))\,dx + \int_\Omega\left(\frac{\partial f}{\partial y}(x, y_h(v)) - \frac{\partial f}{\partial y}(x, y_h(u))\right)\varphi_h(v)\,(\varphi_h(u) - \varphi_h(v))\,dx \\
&\quad + \int_\Omega\left(\frac{\partial L}{\partial y}(x, y_h(u)) - \frac{\partial L}{\partial y}(x, y_h(v))\right)(\varphi_h(u) - \varphi_h(v))\,dx \\
&\quad + \int_\Gamma\left(\frac{\partial l}{\partial y}(x, y_h(u), u) - \frac{\partial l}{\partial y}(x, y_h(v), v)\right)(\varphi_h(u) - \varphi_h(v))\,d\sigma(x)\Bigg) \\
&\leq C\Big(\|y_h(v) - y_h(u)\|_{W^{1,\bar{q}}(\Omega)} + \|u - v\|_{L^2(\Gamma)} \\
&\quad +\|\varphi_h(u) - \varphi_h(v)\|_{L^4(\Omega)}\Big)\,\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)}\,,
\end{aligned}$$

hence

$$\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)} \leq C\left(\|u - v\|_{L^2(\Gamma)} + \|\varphi_h(u) - \varphi_h(v)\|_{L^4(\Omega)}\right). \tag{4.2.10}$$

At this point, let us demonstrate only the estimation of the most difficult term

$$\int_\Omega \left| \frac{\partial a}{\partial y}(x, y_h(v))\nabla y_h(v) \cdot \nabla \varphi_h(v) - \frac{\partial a}{\partial y}(x, y_h(u))\nabla y_h(u) \cdot \nabla \varphi_h(u) \right| |\varphi_h(u) - \varphi_h(v)|\, dx$$
$$\leq C \left(\|y_h(v) - y_h(u)\|_{W^{1,\bar{q}}(\Omega)} + \|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)} \right) \|\varphi_h(u) - \varphi_h(v)\|_{L^4(\Omega)}\,,$$

having used the fact that $\{y_h(u)\}_{h<h_2}$, $\{y_h(v)\}_{h<h_2}$ and the sequences of the corresponding adjoint states are bounded in $W^{1,4}(\Omega)$; see Corollary 2.31 on page 74. From Lemma 2.3 on page 39 and inequality (4.2.10) we infer

$$\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)} \leq C \Big(\|u - v\|_{L^2(\Gamma)} + \|\varphi_h(u) - \varphi_h(v)\|^{1/2}_{L^2(\Omega)} \|\varphi_h(u) - \varphi_h(v)\|^{1/2}_{H^1(\Omega)} \Big)$$

and by Young's inequality,

$$\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)} \leq C\|u - v\|_{L^2(\Gamma)} + \frac{C^2}{2}\|\varphi_h(u) - \varphi_h(v)\|_{L^2(\Omega)}$$
$$+ \frac{1}{2}\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)}\,.$$

Next let us get an estimate for $\|\varphi_h(u) - \varphi_h(v)\|_{L^2(\Omega)}$. Since Ω is convex, it follows from inequality (2.7.25) and Corollary 1.37 that

$$\|\varphi_h(u) - \varphi_h(v)\|_{L^2(\Omega)} \leq \|\varphi_u - \varphi_v\|_{L^2(\Omega)} + \|\varphi_h(u) - \varphi_u\|_{L^2(\Omega)} + \|\varphi_v - \varphi_h(v)\|_{L^2(\Omega)}$$
$$\leq C \left(h + \|u - v\|_{L^2(\Gamma)} \right).$$

Collecting all these results, we conclude (4.2.9). It remains to estimate the error $\varphi_u - \varphi_h(v)$ in the $L^\infty(\Gamma)$ norm. Using the same triangle inequality as for the $H^1(\Omega)$ norm and taking $2 < q = \min\{\hat{q}, \bar{q}, 2/(1-\tau)\}$, we obtain

$$\|\varphi_u - \varphi_h(v)\|_{L^\infty(\Gamma)} \leq C\|\varphi_u - \varphi_h(v)\|_{W^{1,q}(\Omega)}$$
$$\leq C \left(\|\varphi_u - \varphi_h(u)\|_{W^{1,q}(\Omega)} + \|\varphi_h(u) - \varphi_h(v)\|_{W^{1,q}(\Omega)} \right).$$

For the first term we deduce as for the state functions

$$\|\varphi_u - \varphi_h(u)\|_{W^{1,q}(\Omega)} \leq Ch^{2/q} \leq Ch^{1-\tau}.$$

For the second term we use the inverse inequality (2.2.6) on page 38

$$\|\varphi_h(u) - \varphi_h(v)\|_{W^{1,q}(\Omega)} \leq Ch^{2/q-1}\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)}$$
$$\leq Ch^{-\tau}\|\varphi_h(u) - \varphi_h(v)\|_{H^1(\Omega)}\,.$$

Inserting (4.2.9) in the last inequality, we get

$$\|\varphi_h(u) - \varphi_h(v)\|_{W^{1,q}(\Omega)} \leq Ch^{-\tau} \left(h + \|u - v\|_{L^2(\Gamma)} \right) = C \left(h^{1-\tau} + h^{-\tau}\|u - v\|_{L^2(\Gamma)} \right),$$

hence

$$\|\varphi_u - \varphi_h(v)\|_{L^\infty(\Gamma)} \leq C \left(h^{1-\tau} + h^{-\tau}\|u - v\|_{L^2(\Gamma)} \right).$$

□

LEMMA 4.5. *For every $h < h_1$ let $u_h \in U_{ad,h} \cap \overline{B}_{L^2(\Gamma)}(\bar u, \rho)$ and $u \in U_{ad}$ be given.*

(1) *Under the weak convergence $u_h \rightharpoonup u$ in $L^1(\Gamma)$ as $h \to 0$, there holds*

$$\lim_{h\to 0} \|y_h(u_h) - y_u\|_{C(\bar\Omega)\cap H^1(\Omega)} = 0 \quad \textit{and} \quad \liminf_{h\to 0} J_h(u_h) \ge J(u)\,. \tag{4.2.11}$$

(2) *If $u_h \to u$ strongly in $L^1(\Gamma)$ as $h \to 0$ then $\lim_{h\to 0} J_h(u_h) = J(u)$.*

PROOF. Thanks to the weak convergence $u_h \rightharpoonup u$ in $L^1(\Gamma)$, $\{u_h\}_{h<h_1}$ converges weakly to u in $L^2(\Gamma)$ and $u \in \overline{B}_{L^2(\Gamma)}(\bar u, \rho)$. In view of the definition of G_h, we have for $y_h(u_h) = G_h(u_h)$ that $\{y_h(u_h)\}_{h<h_1} \subset \overline{B}_{W^{1,\bar q}(\Omega)}(\bar y, \kappa_\rho) \cap Y_h$, therefore there exists a subsequence $\{y_{h_k}(u_{h_k})\}_{k=1}^{\infty}$ converging weakly to some element $\tilde y \in \overline{B}_{W^{1,\bar q}(\Omega)}(\bar y, \kappa_\rho)$. Since the embedding $W^{1,\bar q}(\Omega) \hookrightarrow C(\bar\Omega)$ is compact, this convergence is strong in $C(\bar\Omega)$. To show that $\tilde y = y_u$ we consider the equation satisfied by $y_{h_k}(u_{h_k})$:

$$\int_\Omega \{a(x, y_{h_k}(u_{h_k}))\nabla y_{h_k}(u_{h_k})\cdot\nabla\phi_{h_k} + f(x, y_{h_k}(u_{h_k}))\phi_{h_k}\}\,dx = \int_\Gamma u_{h_k}\phi_{h_k}\,d\sigma(x) \tag{4.2.12}$$

$\forall \phi_{h_k} \in Y_{h_k}$. Let us take in (4.2.12) $\phi_{h_k} = \Pi_{h_k}\phi$, with $\phi \in H^2(\Omega)$ arbitrarily chosen and Π_h introduced on page 37. Then by passing to the limit, we obtain, along with the strong convergences $y_{h_k}(u_{h_k}) \to \tilde y$ in $C(\bar\Omega)$ and $\Pi_{h_k}\phi \to \phi$ in $H^1(\Omega)$, the weak convergences $y_{h_k}(u_{h_k}) \rightharpoonup \tilde y$ in $W^{1,\bar q}(\Omega)$ and $u_{h_k} \rightharpoonup u$ in $L^2(\Gamma)$, and the assumptions on a and f, that

$$\int_\Omega \{a(x,\tilde y)\nabla\tilde y\cdot\nabla\phi + f(x,\tilde y)\phi\}\,dx = \int_\Gamma u\phi\,d\sigma(x)\,.$$

Finally, the density of $H^2(\Omega)$ in $H^1(\Omega)$ and the uniqueness of the solution of (1.1.1) imply $\tilde y = y_u$. Since all the subsequences have the same limit, we conclude that $\lim_{h\to0}\|y_h(u_h) - y_u\|_{C(\bar\Omega)} = 0$. Let us prove now the convergence $y_h(u_h) \to y_u$ in the $H^1(\Omega)$ norm. To this end, we write

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

Now since $y_h(u_h)$ and y_{u_h} are the discrete and continuous states associated with the same control and $\{u_h\}_{h<h_1}$ is bounded in $L^2(\Gamma)$, we can use the error estimate (2.3.37) on page 52 to deduce that the first summand tends to zero as $h \to 0$. Next we show that the second summand converges to zero, too. At first, similar to the proof of Theorem 3.3 on page 77 it can be shown that $y_{u_h} \rightharpoonup y_u$ weakly in $H^{3/2}(\Omega)$; notice that $\{y_{u_h}\}_{h<h_1}$ is bounded in $H^{3/2}(\Omega)$ (Theorem 1.18 on page 16). Then the proof of the convergence of the states in $H^1(\Omega)$ is completed due to the compactness of the embedding $H^{3/2}(\Omega) \hookrightarrow H^1(\Omega)$. For the proof of the inequality in (4.2.11) the reader is referred to Casas and Mateos [**26**, Lemma 11].

Finally, let us prove the last assertion of the lemma. From the dominated convergence theorem, along with the fact that $\{u_h\}_{h<h_1}$ is bounded in $U_{ad,h} \cap \overline{B}_{L^2(\Gamma)}(\bar u, \rho)$ and $u_h \to u$ strongly in $L^1(\Gamma)$, we deduce the strong convergence $u_h \to u$ in $L^2(\Gamma)$. The

convergence of the cost functions is then an immediate consequence of the continuity of the control-to-state mapping $G : L^2(\Gamma) \longrightarrow H^{3/2}(\Omega)$, $u \longmapsto y_u$, and (4.2.11). $\square$

The next theorem, which is the main result of this section, establishes the convergence in $L^\infty(\Gamma)$ of solutions of $(\mathcal{P}_h^\varepsilon)$ to $\bar{u}$.

Theorem 4.6. *Let $\{\bar{u}_h\}_{h<h_1}$ be a family of solutions of $(\mathcal{P}_h^\varepsilon)$ and $\bar{y}_h := y_h(\bar{u}_h)$. Then there holds the convergence*

$$\lim_{h\to 0}\Big(\|\bar{u}_h - \bar{u}\|_{L^\infty(\Gamma)} + \|\bar{y}_h - \bar{y}\|_{L^\infty(\Omega)} + \|\bar{y}_h - \bar{y}\|_{H^1(\Omega)}\Big) = 0\,.$$

Proof. Since $\{\bar{u}_h\}_{h<h_1}$ is bounded in $L^\infty(\Gamma)$, there exists a weakly* convergent subsequence in $L^\infty(\Gamma)$, still indexed by h, to some $\tilde{u} \in U_{ad} \cap \overline{B}_{L^\infty(\Gamma)}(\bar{u}, \varepsilon)$. Although the proof of the identity $\bar{u} = \tilde{u}$ is standard, see for instance Casas et al. [**29**, Theorem 4.4], we will briefly sketch it for convenience of the reader. From Theorem 3.10-(1) on page 81 we know that $\bar{u} \in C^{0,1/2}(\Gamma)$. Now consider the projection operator $P_h : L^1(\Gamma) \to \mathcal{U}_h$ defined by

$$(P_h u)|_{(x_j,x_{j+1})} = \frac{1}{|x_{j+1} - x_j|}\int_{x_j}^{x_{j+1}} u(x)\, d\sigma(x)$$

which satisfies the inequality

$$\|\bar{u} - P_h\bar{u}\|_{L^\infty(\Gamma)} \le Ch^{1/2}\|\bar{u}\|_{C^{0,1/2}(\Gamma)}\,,$$

with some positive constant C independent of $\bar{u}$ and h. Hence, for all h small enough $P_h\bar{u}$ is admissible for $(\mathcal{P}_h^\varepsilon)$ and from Lemma 4.5-(1) (in particular, $\bar{u}_h \rightharpoonup \tilde{u}$ weakly in $L^1(\Gamma)$) we get

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

Finally, this inequality, along with (4.1.1), yields $\tilde{u} = \bar{u}$ and $\lim_{h\to 0} J_h(\bar{u}_h) = J(\bar{u})$. In view of Lemma 4.5, this implies the desired convergence properties for the states.

The rest of the proof is split into two steps.

Step 1: $L^2(\Gamma)$ convergence of the controls. We will follow here Arada et al. [**5**]. By the definitions of J_h and J, the convergence $\lim_{h\to 0} J_h(\bar{u}_h) = J(\bar{u})$ implies that

$$\lim_{h\to 0}\int_\Gamma \{l(x, \bar{y}_h, \bar{u}_h) - l(x, \bar{y}, \bar{u})\}\, d\sigma(x) = 0\,.$$

On the other hand, since $\bar{u}_h \rightharpoonup \bar{u}$ weakly* in $L^\infty(\Gamma)$, there holds

$$\int_\Gamma \frac{\partial l}{\partial u}(x, \bar{y}, \bar{u})(\bar{u}_h - \bar{u})\, d\sigma(x) \to 0 \quad \text{as } h \to 0\,.$$

Then invoking inequality (3.3.9) on page 80, Assumption 3.6 and (4.2.11), we have for some measurable function $\theta_h(x) \in [0,1]$ and $v_h(x) := \bar{u}(x) + \theta_h(x)(\bar{u}_h(x) - \bar{u}(x))$ that

$$\begin{aligned}
\frac{\Lambda_l}{2}\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)} &\le \frac{1}{2}\int_\Gamma \frac{\partial^2 l}{\partial u^2}(x,\bar{y},v_h)(\bar{u}_h - \bar{u})^2\, d\sigma(x) \\
&= \int_\Gamma \{l(x,\bar{y},\bar{u}_h) - l(x,\bar{y},\bar{u})\}\, d\sigma(x) - \int_\Gamma \frac{\partial l}{\partial u}(x,\bar{y},\bar{u})(\bar{u}_h - \bar{u})\, d\sigma(x) \\
&= \int_\Gamma \{l(x,\bar{y},\bar{u}_h) - l(x,\bar{y}_h,\bar{u}_h)\}\, d\sigma(x) \\
&\quad + \int_\Gamma \{l(x,\bar{y}_h,\bar{u}_h) - l(x,\bar{y},\bar{u})\}\, d\sigma(x) - \int_\Gamma \frac{\partial l}{\partial u}(x,\bar{y},\bar{u})(\bar{u}_h - \bar{u})\, d\sigma(x) \\
&\le C\|\bar{y}_h - \bar{y}\|_{L^2(\Gamma)} + \int_\Gamma \{l(x,\bar{y}_h,\bar{u}_h) - l(x,\bar{y},\bar{u})\}\, d\sigma(x) \\
&\quad - \int_\Gamma \frac{\partial l}{\partial u}(x,\bar{y},\bar{u})(\bar{u}_h - \bar{u})\, d\sigma(x) \to 0 \quad \text{when} \;\; h \to 0\,.
\end{aligned}$$

Step 2: $L^\infty(\Gamma)$ convergence of the controls. Given $x \in \Gamma$, let $1 \le j \le N(h)$ be such that $x \in [x_j, x_{j+1}]$. From the projection formulas (3.3.12) on page 81 and (4.2.5) for the optimal controls and the contractivity of the projection we obtain

$$\begin{aligned}
|\bar{u}(x) - \bar{u}_j| &= \left|\mathrm{Proj}_{[\alpha,\beta]}\{\bar{s}(x)\} - \mathrm{Proj}_{[\alpha_{h,j},\beta_{h,j}]}\{\bar{s}_j\}\right| \\
&\le \left|\mathrm{Proj}_{[\alpha,\beta]}\{\bar{s}(x)\} - \mathrm{Proj}_{[\alpha_{h,j},\beta_{h,j}]}\{\bar{s}(x)\}\right| \\
&\quad + \left|\mathrm{Proj}_{[\alpha_{h,j},\beta_{h,j}]}\{\bar{s}(x)\} - \mathrm{Proj}_{[\alpha_{h,j},\beta_{h,j}]}\{\bar{s}_j\}\right| \\
&\le \left|\mathrm{Proj}_{[\alpha,\beta]}\{\bar{s}(x)\} - \mathrm{Proj}_{[\alpha_{h,j},\beta_{h,j}]}\{\bar{s}(x)\}\right| + |\bar{s}(x) - \bar{s}_j|\,. \qquad (4.2.13)
\end{aligned}$$

Next we demonstrate that the first summand in (4.2.13) is zero for sufficiently small h. Assume that $h^{1/2}\Lambda_{\bar{u}} < \varepsilon/2$, where $\Lambda_{\bar{u}}$ is the Hölder constant of $\bar{u}$. Because $\alpha \le \alpha_{h,j} \le \beta_{h,j} \le \beta$, we have that

$$\left|\mathrm{Proj}_{[\alpha,\beta]}\{\bar{s}(x)\} - \mathrm{Proj}_{[\alpha_{h,j},\beta_{h,j}]}\{\bar{s}(x)\}\right| = 0 \qquad (4.2.14)$$

if $\alpha_{h,j} \le \bar{s}(x) \le \beta_{h,j}$. Let us now assume that $\alpha_{h,j} > \bar{s}(x)$. Under this assumption, we will show that $\alpha = \alpha_{h,j}$ and consequently (4.2.14) holds true. We will argue by contradiction. If $\alpha < \alpha_{h,j}$ then $\bar{u}(x) < \alpha_{h,j} = \max_{\tilde{x}\in[x_j,x_{j+1}]} \bar{u}(\tilde{x}) - \varepsilon$, hence there exists $\hat{x} \in [x_j, x_{j+1}]$ such that $\bar{u}(x) < \bar{u}(\hat{x}) - \varepsilon$. But this leads to the following contradiction

$$\varepsilon < \bar{u}(\hat{x}) - \bar{u}(x) \le \Lambda_{\bar{u}}|\hat{x} - x|^{1/2} \le \Lambda_{\bar{u}} h^{1/2} < \frac{\varepsilon}{2}\,.$$

Analogously, we argue if $\beta_{h,j} < \bar{s}(x)$. Now from (4.2.13) it follows

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Gamma)} \le \|\bar{s} - \bar{s}_h\|_{L^\infty(\Gamma)}\,, \text{ where } \bar{s}_h := \sum_{j=1}^{N(h)} \bar{s}_j \chi_{(x_j,x_{j+1})}\,. \qquad (4.2.15)$$

For the proof of the uniform convergence $\bar{s}_h \to \bar{s}$ as $h \to 0$ we refer to Casas et al. [**29**, pages 204-205]. We should remark that the proof of the latter result relies on the $C^{0,1/2}(\Gamma)$ regularity of $\bar{s}$ and the convergences $\bar{y}_h \to \bar{y}$ and $\varphi_h(\bar{u}_h) \to \bar{\varphi}$ in $L^\infty(\Omega)$ as a consequence of the convergence $\bar{u}_h \to \bar{u}$ in $L^2(\Gamma)$ and Lemma 4.4. □

Because of the uniform convergence $\bar{u}_h \to \bar{u}$ as $h \to 0$, $\bar{u}_h$ is a local solution of the problem

$$(\mathcal{P}_h) \begin{cases} \min J_h(u_h) = \displaystyle\int_\Omega L(x, y_h(u_h))\,dx + \int_\Gamma l(x, y_h(u_h), u_h)\,d\sigma(x)\,, \\ \text{s.t. } \ u_h \in U_{ad,h}\,, \\ \qquad (y_h(u_h), u_h) \ \text{ satisfying the discrete state equation (2.3.1) with } u = u_h\,, \end{cases}$$

for every sufficiently small h. Therefore, (4.2.3) can be rewritten as

$$\sum_{j=1}^{N(h)} \int_{x_j}^{x_{j+1}} \left\{ \frac{\partial l}{\partial u}(x, \bar{y}_h, \bar{u}_j) + \bar{\varphi}_h \right\} d\sigma(x)\,(u_j - \bar{u}_j) \geq 0 \quad \forall u_j \in [\alpha, \beta]$$

and the expression (4.2.5) can be formulated as

$$\bar{u}_j = \operatorname{Proj}_{[\alpha,\beta]}\{\bar{s}_j\} \quad \forall 1 \leq j \leq N(h)\,. \tag{4.2.16}$$

4.3. Error estimates for optimal controls

In this section, $\bar{u}$ is supposed to be a local minimum of $(\mathcal{P})$ satisfying the second-order sufficient condition for optimality (3.5.15) on page 98 and $\{\bar{u}_h\}_{h<h_1}$ ($h_1 > 0$ is given in Theorem 2.25 on page 63) is a sequence of local solutions of $(\mathcal{P}_h)$ converging uniformly to $\bar{u}$; see Theorem 4.6. As usual, $\bar{y}$, $\bar{y}_h := y_h(\bar{u}_h)$, and $\bar{\varphi}$, $\bar{\varphi}_h := \varphi_h(\bar{u}_h)$, stand for the states and adjoint states corresponding to $\bar{u}$ and $\bar{u}_h$. Our goal is to estimate $\|\bar{u}_h - \bar{u}\|_{L^2(\Gamma)}$ and $\|\bar{u}_h - \bar{u}\|_{L^\infty(\Gamma)}$. To this aim, we will state and prove three auxiliary lemmas.

Lemma 4.7. *Let $\delta > 0$ be given as in Theorem 3.22. Then there exists $h_3 \leq h_1$ such that*

$$\frac{\delta}{2}\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)} \leq (J'(\bar{u}_h) - J'(\bar{u}))(\bar{u}_h - \bar{u}) \quad \forall h < h_3\,. \tag{4.3.1}$$

Proof. The proof is exactly the same as in Casas et al. [**29**, Lemma 4.6], therefore we will only sketch the main steps. Setting

$$\bar{d}_h(x) = \frac{\partial l}{\partial u}(x, \bar{y}_h(x), \bar{u}_h(x)) + \bar{\varphi}_h(x) \ \text{ for } x \in \Gamma\,,$$

we deduce from Lemma 4.4 that $\bar{d}_h \to \bar{d}$ uniformly on Γ. With the help of this convergence and taking ρ as in Theorem 3.22 on page 98, it can be shown that there exists $0 < \tilde{h} \leq h_1$ such that $(\bar{u}_h - \bar{u}) \in C^\rho_{\bar{u}}$ $\forall h < \tilde{h}$. Hence, according to Theorem

3.22, it follows $J''(\bar{u})(\bar{u}_h - \bar{u})^2 \geq \delta\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)}$. Now by applying the mean value theorem, we get for some measurable function θ_h with values in $[0,1]$ that

$$\begin{aligned}(J'(\bar{u}_h) - J'(\bar{u}))(\bar{u}_h - \bar{u}) &= J''(\bar{u} + \theta_h(\bar{u}_h - \bar{u}))(\bar{u}_h - \bar{u})^2 \\ &= \left(J''(\bar{u} + \theta_h(\bar{u}_h - \bar{u})) - J''(\bar{u})\right)(\bar{u}_h - \bar{u})^2 + J''(\bar{u})(\bar{u}_h - \bar{u})^2 \\ &\geq \delta\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)} - \left|\left(J''(\bar{u} + \theta_h(\bar{u}_h - \bar{u})) - J''(\bar{u})\right)(\bar{u}_h - \bar{u})^2\right|.\end{aligned}$$

Finally, it is enough to choose $0 < h_3 \leq \tilde{h}$ such that

$$\left|\left(J''(\bar{u} + \theta_h(\bar{u}_h - \bar{u})) - J''(\bar{u})\right)(\bar{u}_h - \bar{u})^2\right| \leq \frac{\delta}{2}\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)} \quad \forall h < h_3$$

to prove (4.3.1). The last inequality is obtained from the representation of J'' (Eq. (3.3.3) on page 79) after long but easy computations. Let us demonstrate exemplarily the estimation of one of the terms involved in (3.3.3). Setting $\hat{u}_h = \bar{u} + \theta_h(\bar{u}_h - \bar{u})$, $v_h = \bar{u}_h - \bar{u}$ and exploiting inequality (1.6.11) on page 26, we deduce

$$\begin{aligned}\int_\Omega \left|\nabla\varphi_{\hat{u}_h}\cdot\nabla y_{\hat{u}_h}\frac{\partial^2 a}{\partial y^2}(x, y_{\hat{u}_h})\left(G'(\hat{u}_h)v_h\right)^2 - \nabla\varphi_{\bar{u}}\cdot\nabla y_{\bar{u}}\frac{\partial^2 a}{\partial y^2}(x, y_{\bar{u}})\left(G'(\bar{u})v_h\right)^2\right| dx \\ \leq C\Big(\left(\|\varphi_{\hat{u}_h} - \varphi_{\bar{u}}\|_{W^{1,4}(\Omega)} + \|y_{\hat{u}_h} - y_{\bar{u}}\|_{W^{1,4}(\Omega)}\right)\|v_h\|^2_{L^2(\Gamma)} \\ + \|\left(G'(\hat{u}_h) - G'(\bar{u})\right)v_h\|_{L^4(\Omega)}\|v_h\|_{L^2(\Gamma)}\Big).\end{aligned}$$

Moreover, by subtracting the equations satisfied by $G'(\hat{u}_h)v_h$ and $G'(\bar{u})v_h$ and using again (1.6.11), we get

$$\|\left(G'(\hat{u}_h) - G'(\bar{u})\right)v_h\|_{H^{3/2}(\Omega)} \leq C\|y_{\hat{u}_h} - y_{\bar{u}}\|_{W^{1,4}(\Omega)}\|v_h\|_{L^2(\Gamma)}.$$

Hence, given $\gamma > 0$, these inequalities, together with the uniform convergence $\bar{u}_h \to \bar{u}$ (and consequently $\hat{u}_h \to \bar{u}$), Corollary 1.32 on page 27 and the estimate (1.7.11) on page 33, yield for h small enough that

$$\begin{aligned}\int_\Omega \Big|\nabla\varphi_{\hat{u}_h}\cdot\nabla y_{\hat{u}_h}\frac{\partial^2 a}{\partial y^2}(x, y_{\hat{u}_h})\left(G'(\hat{u}_h)v_h\right)^2 \\ -\nabla\varphi_{\bar{u}}\cdot\nabla y_{\bar{u}}\frac{\partial^2 a}{\partial y^2}(x, y_{\bar{u}})\left(G'(\bar{u})v_h\right)^2\Big| dx \leq \gamma\|v_h\|^2_{L^2(\Gamma)}.\end{aligned}$$

□

In the next lemma, we estimate the convergence of J'_h to J'.

Lemma 4.8. *Let $h_2 > 0$ be given as in Proposition 2.26. If Ω is non-convex there exists $C_0 > 0$ such that, for all $v \in L^2(\Gamma)$ and $h < h_2$,*

$$|(J'_h(\bar{u}_h) - J'(\bar{u}_h))v| \leq C_0 h^{5/8}\|v\|_{L^2(\Gamma)}. \tag{4.3.2}$$

If Ω is convex and $p > 2$ then for every $\gamma > 0$ there exists $C_\gamma > 0$ such that, for all $v \in L^2(\Gamma)$ and all $h < h_2$,

$$|(J_h'(\bar u_h) - J'(\bar u_h))\, v| \le \Big(C_\gamma h + \gamma \|\bar u_h - \bar u\|_{L^2(\Gamma)}\Big) \|v\|_{L^2(\Gamma)}\,. \tag{4.3.3}$$

Proof. Invoking the expressions of the derivatives J_h' and J' given by the formula (3.3.2) on page 79 and (4.2.1), respectively, it follows

$$\begin{aligned} |(J_h'(\bar u_h) - J'(\bar u_h))v| &\le \int_\Gamma \left| \frac{\partial l}{\partial u}(x, \bar y_h, \bar u_h) - \frac{\partial l}{\partial u}(x, y_{\bar u_h}, \bar u_h) + \bar\varphi_h - \varphi_{\bar u_h} \right| |v|\, d\sigma(x) \\ &\le C_1 \Big(\|\bar y_h - y_{\bar u_h}\|_{L^2(\Gamma)} + \|\bar\varphi_h - \varphi_{\bar u_h}\|_{L^2(\Gamma)}\Big) \|v\|_{L^2(\Gamma)}\,. \end{aligned} \tag{4.3.4}$$

The result in the non-convex case is then an easy application of Lemma 4.4:

$$\|\bar y_h - y_{\bar u_h}\|_{L^2(\Gamma)} + \|\bar\varphi_h - \varphi_{\bar u_h}\|_{L^2(\Gamma)} \le C h^{5/8}\,.$$

To show (4.3.3) we use the following well known property, compare also Grisvard [**60**, Theorem 1.4.3.3]: For every $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that

$$\|z\|_{L^2(\Gamma)} \le \varepsilon \|z\|_{H^1(\Omega)} + C_\varepsilon \|z\|_{L^2(\Omega)} \quad \forall z \in H^1(\Omega)\,,$$

thus we get with the aid of (4.2.6)

$$\begin{aligned} \|\bar y_h - y_{\bar u_h}\|_{L^2(\Gamma)} &= \|y_h(\bar u_h) - y_{\bar u_h}\|_{L^2(\Gamma)} \\ &\le \varepsilon \|y_h(\bar u_h) - y_{\bar u_h}\|_{H^1(\Omega)} + C_\varepsilon \|y_h(\bar u_h) - y_{\bar u_h}\|_{L^2(\Omega)} \\ &\le \varepsilon \|y_h(\bar u_h) - y_{\bar u_h}\|_{H^1(\Omega)} + C_\varepsilon C_2 h\,. \end{aligned}$$

By virtue of Corollary 1.32 on page 27, along with the embedding $W^{1,4}(\Omega) \hookrightarrow H^1(\Omega)$ and the boundedness of the sequence $\{\bar u_h\}_{h<h_2}$ in $L^2(\Gamma)$, we have

$$\|\bar y - y_{\bar u_h}\|_{H^1(\Omega)} \le C_3 \|\bar u - \bar u_h\|_{L^2(\Gamma)}\,.$$

Taking into account that $\bar u \in C^{0,1}(\Gamma) \subset W^{1-1/\varrho,\varrho}(\Gamma)$ with $\varrho > 2$ (Theorem 3.10-(2) on page 81), (4.2.7) leads to

$$\|\bar y - y_h(\bar u_h)\|_{H^1(\Omega)} \le C_4 \Big(h + \|\bar u - \bar u_h\|_{L^2(\Gamma)}\Big)\,,$$

Combining the last two inequalities, we deduce

$$\|y_h(\bar u_h) - y_{\bar u_h}\|_{H^1(\Omega)} \le C_5 \Big(h + \|\bar u - \bar u_h\|_{L^2(\Gamma)}\Big)\,.$$

The same arguments can be applied to the adjoint state. Hence, by setting $\gamma = \varepsilon C_5 C_1$ and $C_\gamma = \gamma + C_\varepsilon C_2 C_1$, (4.3.4) implies (4.3.3). □

One key point in the proof of the error estimates for the controls is to find a function $u_h \in U_{ad,h}$ that approximates $\bar u$ and satisfies $J'(\bar u)u_h = J'(\bar u)\bar u$. Making use of the definition of $\bar{d}$ given on page 80, we define $u_h = \sum_{j=1}^{N(h)} u_j \chi_{(x_j,x_{j+1})} \in \mathcal{U}_h$ by

$$u_j = \begin{cases} \dfrac{1}{I_j} \displaystyle\int_{x_j}^{x_{j+1}} \bar{d}(x)\bar{u}(x)\, d\sigma(x) & \text{if } I_j \neq 0\,, \\ \dfrac{1}{|x_{j+1}-x_j|} \displaystyle\int_{x_j}^{x_{j+1}} \bar{u}(x)\, d\sigma(x) & \text{if } I_j = 0\,, \end{cases} \qquad I_j = \int_{x_j}^{x_{j+1}} \bar{d}(x)\, d\sigma(x)\,. \tag{4.3.5}$$

The next lemma shows that u_h fulfills the required conditions.

Lemma 4.9. *There exists $h_4 > 0$ such that, for every $h < h_4$, the following properties hold:*

(1) $u_h \in U_{ad,h}$.
(2) $J'(\bar{u})u_h = J'(\bar{u})\bar{u}$.
(3) *There exists $C > 0$ independent of h such that*

$$\|\bar{u} - u_h\|_{L^\infty(\Gamma)} \le Ch^\sigma\,,$$

where $\sigma = 1/2$ if Ω is non-convex and $\sigma = 1$ if Ω is convex and $p > 2$.

Proof. The proof is almost identical to the one of Casas et al. [**29**, Lemma 4.8]. The only difference is that $\bar{u} \in C^{0,1/2}(\Gamma)$ if Ω is non-convex. □

Finally, we derive the main error estimate.

Theorem 4.10. *There exists a constant $\tilde{C} > 0$ independent of h such that*

$$\|\bar{u} - \bar{u}_h\|_{L^2(\Gamma)} \le \tilde{C}h^\sigma \tag{4.3.6}$$

with $\sigma = 1/2$ if Ω is non-convex and $\sigma = 1$ if Ω is convex and $p > 2$.

Proof. Since $\bar{u}_h$ is a local minimum of $(\mathcal{P}_h)$, we know that $J_h'(\bar{u}_h)(u_h - \bar{u}_h) \ge 0$, where u_h is defined by (4.3.5). The last inequality is equivalent to

$$J'(\bar{u}_h)(\bar{u} - \bar{u}_h) + (J_h'(\bar{u}_h) - J'(\bar{u}_h))\,(\bar{u} - \bar{u}_h) + J_h'(\bar{u}_h)(u_h - \bar{u}) \ge 0\,.$$

On the other hand, $J'(\bar{u})(\bar{u}_h - \bar{u}) \ge 0$. Adding these inequalities, we get

$$\begin{aligned} (J'(\bar{u}_h) - J'(\bar{u}))\,(\bar{u}_h - \bar{u}) &\le (J_h'(\bar{u}_h) - J'(\bar{u}_h))\,(\bar{u} - \bar{u}_h) + J_h'(\bar{u}_h)(u_h - \bar{u}) \\ &= (J_h'(\bar{u}_h) - J'(\bar{u}_h))\,(\bar{u} - \bar{u}_h) + (J_h'(\bar{u}_h) - J'(\bar{u}_h))\,(u_h - \bar{u}) \\ &\quad + (J'(\bar{u}_h) - J'(\bar{u}))\,(u_h - \bar{u}) + J'(\bar{u})(u_h - \bar{u})\,. \end{aligned} \tag{4.3.7}$$

From (4.3.1), (4.3.2), Corollaries 1.32 and 1.37 on page 27 and 33, respectively, as well as Lemmas 4.4 and 4.9, we deduce in the non-convex case

$$\frac{\delta}{2}\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)} \le C\left(h^{5/8}\left(\|\bar{u} - \bar{u}_h\|_{L^2(\Gamma)} + h^{1/2}\right) + h^{1/2}\|\bar{u}_h - \bar{u}\|_{L^2(\Gamma)}\right),$$

which leads to (4.3.6). If Ω is convex and $p > 2$ we use (4.3.3) instead of (4.3.2). Similarly, by taking $\gamma = \delta/4$, it follows from (4.3.7) that

$$\frac{\delta}{4}\|\bar{u}_h - \bar{u}\|^2_{L^2(\Gamma)} \leq C_\gamma h\left(\|\bar{u}_h - \bar{u}\|_{L^2(\Gamma)} + Ch\right) + C\left(\frac{\delta}{4} + 1\right) h\|\bar{u}_h - \bar{u}\|_{L^2(\Gamma)}\,.$$

Finally, the latter result and Young's inequality imply (4.3.6) with $\sigma = 1$. □

REMARK 4.11. *The error estimates deduced in Theorem 4.10 seem to be optimal. This opinion is based on the fact that the interpolation error of the approximation of functions in $C^{0,\sigma}(\Gamma)$ by piecewise constant functions is of order $\mathcal{O}(h^\sigma)$. Taking $\sigma = 1/2$ if Ω is non-convex and $\sigma = 1$ if Ω is convex and $p > 2$, this is the same order that we have obtained in the previous theorem.*

At the end, we prove the error estimate in $L^\infty(\Gamma)$.

THEOREM 4.12. *Let $0 < \tau < 1/2$ be fixed. There exists a constant $C_\tau > 0$, being dependent on τ but not on h, such that*

$$\|\bar{u} - \bar{u}_h\|_{L^\infty(\Gamma)} \leq C_\tau h^{\sigma-\tau}$$

with $\sigma = 1/2$ if Ω is non-convex and $\sigma = 1$ if Ω is convex and $p > 2$.

PROOF. From (4.2.4) and the continuity of the integrand with respect to x (see Eq. (3.3.10) on page 81) we deduce, for every $1 \leq j \leq N(h)$, the existence of a point $\xi_j \in (x_j, x_{j+1})$ such that

$$\frac{\partial l}{\partial u}(\xi_j, \bar{y}_h(\xi_j), \bar{s}_j) + \bar{\varphi}_h(\xi_j) = 0\,. \tag{4.3.8}$$

Moreover, denoting by $\Lambda_{\bar{s}}$ the Hölder constant of $\bar{s}$ (Theorem 3.10 on pages 80-81), we get for $x \in (x_j, x_{j+1})$

$$\begin{aligned} |\bar{s}(x) - \bar{s}_j| &\leq |\bar{s}(x) - \bar{s}(\xi_j)| + |\bar{s}(\xi_j) - \bar{s}_j| \\ &\leq \Lambda_{\bar{s}}|x - \xi_j|^\sigma + |\bar{s}(\xi_j) - \bar{s}_j| \\ &\leq \Lambda_{\bar{s}} h^\sigma + |\bar{s}(\xi_j) - \bar{s}_j|\,. \end{aligned}$$

Employing this inequality in (4.2.15), along with the properties (3.3.9) and (3.3.11) given on pages 80-81 and (4.3.8), yields

$$\begin{aligned} \|\bar{u} - \bar{u}_h\|_{L^\infty(\Gamma)} \leq \max_{1\leq j\leq N(h)} \|\bar{s} - \bar{s}_j\|_{L^\infty(x_j,x_{j+1})} &\leq \Lambda_{\bar{s}} h^\sigma + \max_{1\leq j\leq N(h)} |\bar{s}(\xi_j) - \bar{s}_j| \\ &\leq \Lambda_{\bar{s}} h^\sigma + \frac{1}{\Lambda_l}\max_{1\leq j\leq N(h)}\left|\frac{\partial l}{\partial u}(\xi_j, \bar{y}_h(\xi_j), \bar{s}(\xi_j)) - \frac{\partial l}{\partial u}(\xi_j, \bar{y}_h(\xi_j), \bar{s}_j)\right| \\ &\leq \Lambda_{\bar{s}} h^\sigma + \frac{1}{\Lambda_l}\max_{1\leq j\leq N(h)}\left\{\left|\frac{\partial l}{\partial u}(\xi_j, \bar{y}_h(\xi_j), \bar{s}(\xi_j)) - \frac{\partial l}{\partial u}(\xi_j, \bar{y}(\xi_j), \bar{s}(\xi_j))\right|\right. \\ &\qquad\left. + |\bar{\varphi}(\xi_j) - \bar{\varphi}_h(\xi_j)|\right\}. \end{aligned}$$

Finally, taking into account Assumption 3.6, Lemma 4.4 and (4.3.6), we conclude

$$\begin{aligned}\|\bar u - \bar u_h\|_{L^\infty(\Gamma)} &\le \Lambda_{\bar s} h^\sigma + C\left(\|\bar y - \bar y_h\|_{L^\infty(\Gamma)} + \|\bar\varphi - \bar\varphi_h\|_{L^\infty(\Gamma)}\right)\\ &\le \Lambda_{\bar s} h^\sigma + C\left(h^{\sigma-\tau} + h^{-\tau}\|\bar u - \bar u_h\|_{L^2(\Gamma)}\right) \le C_\tau h^{\sigma-\tau}\,.\end{aligned}$$

□

Remark 4.13. *The error estimates for the discrete controls given in Theorems 4.10 and 4.12 can also be derived by using an abstract result proved recently by Casas and Tröltzsch* [**38**].

4.4. Numerical experiments

In this section, we present some numerical experiments for the discrete approach of $(\mathcal{P})$ and compare the order of convergence proved in Theorems 4.10 and 4.12 with the experimental ones. We have tested the convergence theory by two examples. In both cases, we consider the following optimal control problem

$$(E)\ \begin{cases}\min\, J(u) = \dfrac{1}{2}\|y_u - y_d\|^2_{L^2(\Omega)} + \dfrac{1}{2}\|u - u_d\|^2_{L^2(\Gamma)} + \displaystyle\int_\Gamma \eta(x)u(x)\,d\sigma(x)\\ \text{subject to }\ u \in U_{ad} := \{u \in L^\infty(\Gamma)\,|\,\alpha \le u(x) \le \beta \text{ a.e. } x \in \Gamma\}\,,\end{cases}$$

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

In the first example, Ω is supposed to be convex and we have specified the functions contained in (E) such that we know a local solution of it. The second example is constructed in such a way that we do not know any local minimum of (E).

For the generation of the mesh for Ω we use the MATLAB PDE Toolbox. For the optimization a standard SQP method is implemented; see for instance Heinkenschloss and Tröltzsch [**62**], Kunisch and Sachs [**76**], Hinze and Kunisch [**64**] and Tröltzsch [**90**]. Given (y_k,u_k,φ_k), $k \in \mathbb{N}$, in step $k+1$ we have to solve the following linear-quadratic problem to find $(y_{k+1},u_{k+1},\varphi_{k+1})$:

$$(QE)_{k+1}\begin{cases}\min\, \tilde J(u_{k+1}) = \mathcal{J}'(y_k,u_k)(y_{k+1}-y_k,u_{k+1}-u_k)\\ \qquad\qquad +\dfrac{1}{2}D^2_{(y,u)}\mathcal{L}(y_k,u_k,\varphi_k)(y_{k+1}-y_k,u_{k+1}-u_k)^2\\ \text{subject to }\ u_{k+1} \in U_{ad}\,,\\ \qquad\qquad (y_{k+1},u_{k+1}) \text{ satisfying the linearized equation}\end{cases}$$

$$\int_\Omega \left\{ \left(a(x,y_k)\nabla y_{k+1} + \frac{\partial a}{\partial y}(x,y_k) y_{k+1} \nabla y_k \right) \cdot \nabla \phi + \frac{\partial f}{\partial y}(x,y_k) y_{k+1} \phi \right\} dx$$

$$= \int_\Omega \left\{ \left(\frac{\partial f}{\partial y}(x,y_k) y_k - f(x,y_k) \right) \phi + \frac{\partial a}{\partial y}(x,y_k) y_k \nabla y_k \cdot \nabla \phi \right\} dx$$

$$+ \int_\Gamma (u_{k+1} + g)\,\phi\, d\sigma(x) \quad \forall \phi \in H^1(\Omega)\,,$$

where $\mathcal{J}(y,u) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{1}{2}\|u - u_d\|^2_{L^2(\Gamma)} + \int_\Gamma \eta(x) u(x)\, d\sigma(x)$ and

$$\mathcal{L}(y,u,\varphi) := \mathcal{J}(y,u) - \int_\Omega \{ a(x,y)\nabla y \cdot \nabla \varphi + \varphi f(x,y) \}\, dx + \int_\Gamma \varphi\,(u+g)\, d\sigma(x)\,.$$

The new iterate φ_{k+1} is the solution of the associated adjoint equation

$$\int_\Omega \left\{ a(x,y_k)\nabla \varphi_{k+1} \cdot \nabla \phi + \frac{\partial a}{\partial y}(x,y_k) \phi \nabla y_k \cdot \nabla \varphi_{k+1} + \frac{\partial f}{\partial y}(x,y_k) \varphi_{k+1} \phi \right\} dx$$

$$= \int_\Omega \left\{ (y_{k+1} - y_\Omega)\,\phi - \frac{\partial a}{\partial y}(x,y_k)(y_{k+1} - y_k) \nabla \varphi_k \cdot \nabla \phi - \varphi_k \frac{\partial^2 f}{\partial y^2}(x,y_k)(y_{k+1} - y_k)\phi \right.$$

$$\left. - \nabla \varphi_k \cdot \left[\frac{\partial^2 a}{\partial y^2}(x,y_k)(y_{k+1} - y_k) \nabla y_k + \frac{\partial a}{\partial y}(x,y_k) \nabla (y_{k+1} - y_k) \right] \phi \right\} dx$$

$\forall \phi \in H^1(\Omega)$. The choice of the initial data (y_1, u_1, φ_1) is indicated in the examples below. To solve each of the linear-quadratic problems $(QE)_k$ numerically we have applied a primal-dual active set strategy, according to Kunisch and Rösch [**75**]; see also Bergounioux and Kunisch [**8**]. The linear equations for y_{k+1} and φ_{k+1} are solved in the same way as equation (2.5.2) on page 56. In order to deal with the convection term $(\partial a/\partial y)(\cdot, y_k)\nabla y_k \cdot \nabla \varphi_{k+1}$ in the adjoint state equation, we have additionally written our own routine. As an alternative to the above SQP method a *semismooth Newton method* can be used; see Ito and Kunisch [**69**] for more details.

We have solved the problem (E) using different mesh sizes. Similarly to Section 2.5, the experimental order of convergence for z is given by

$$EOC_X(z) = \frac{\log(\|z - \bar{z}_{h_1}\|_X) - \log(\|z - \bar{z}_{h_2}\|_X)}{\log(h_1) - \log(h_2)}$$

In the previous formula z, stands for the state, the adjoint state or the control function. We will replace z by $\bar{z}$ if a locally optimal solution is known (Example 4.14) and by z_{fine} if not (Example 4.15); z_{fine} is the numerically computed optimal (state, adjoint state or control) function on the finest mesh with mesh size $h_{fine} = 2^{-8}$.

EXAMPLE 4.14. *We fix the following data:* $\Omega = (0,\pi)^2$, $\alpha = -20$, $\beta = -2$, $a(x,y(x)) = 1 + (x_1 + x_2)^2 + y^2(x)$, $f(x,y(x)) = 2y(x)(\sin^2(x_1) + \sin^2(x_2)) - \xi(x)$, *where* $x = (x_1, x_2)$ *and*

$$\xi(x) = 2\sin(x_1)\sin(x_2)(1 + (x_1 + x_2)^2) + 6\sin^3(x_1)\sin^3(x_2) \\ - (x_1 + x_2)(\sin(x_1)\cos(x_2) + \cos(x_1)\sin(x_2)) \,.$$

Further, we set $y_d(x) = \sin(x_1)\sin(x_2) - 2(\sin^2(x_1) + \sin^2(x_2))$ *in* Ω *and*

$$\begin{aligned} e(x) &= -(1 + (x_1 + x_2)^2)\sin(x_1 + x_2)\,, \\ g(x) &= \min\{0, e(x) - \alpha\} + \max\{0, e(x) - \beta\}\,, \\ \eta(x) &= -1 - Proj_{[\alpha,\beta]}\{e(x)\} \end{aligned}$$

and $u_d = 0$ *on* Γ*. A strict local minimum of* (E) *is given by* $\bar u(x) = Proj_{[\alpha,\beta]}\{e(x)\}$*. The associated state and adjoint state are* $\bar y(x) = \sin(x_1)\sin(x_2)$ *and* $\bar\varphi \equiv 1$ *in* Ω*, respectively. Indeed, it can be easily checked that* $\bar d(x) = \bar\varphi(x) + \eta(x) + \bar u(x) = 0$ *on* Γ*, hence the first-order optimality condition (3.3.6) on page 80 is satisfied. Moreover, in view of equality (3.3.3) on page 79, we have for* $u \in L^\infty(\Gamma)$

$$J''(u)v^2 = \int_\Gamma v^2\, d\sigma(x) + \int_\Omega z_v^2\, dx \quad \forall v \in L^2(\Gamma)\,,$$

where $z_v \in H^1(\Omega)$ *is the solution of equation (1.6.9) on page 25. Consequently,* $J''(\bar u)v^2 > 0$ *for every* $v \in C^0_{\bar u} \setminus \{0\}$*, hence the second-order sufficient condition (3.5.5) on page 91 is satisfied.*

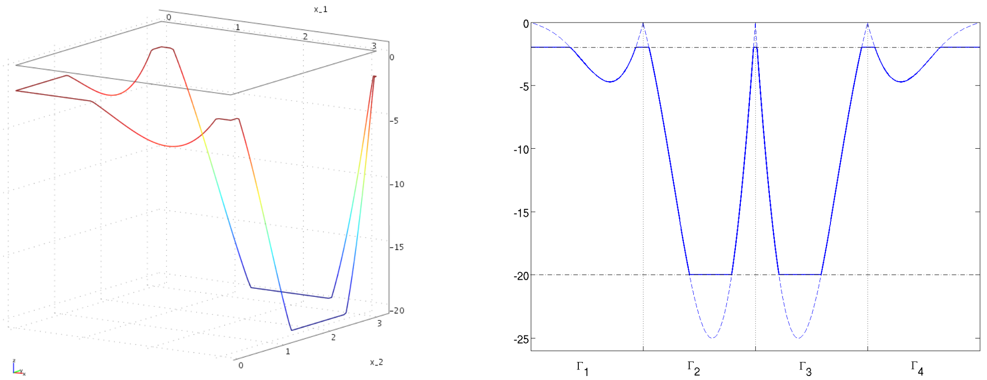

FIGURE 4.1. Optimal control $\bar u$ computed with COMSOL Multiphysics (left frame); Solid line: $\bar u$, dashed line: e (right frame).

The finest grid on which we can run the SQP method has a mesh size of order 2^{-8}; the range of the degrees of freedom for the state and adjoint state function is 10^5, the corresponding one for the controls is 10^3. We have taken $y_1 \equiv 1$, $\varphi_1 \equiv 2$ and $u_1 \equiv \beta$ as initial data. We have used finer grids for a more accurate numerical evaluation of the errors. The results are collected in the table below. For the controls we observe linear convergence for the error in $L^2(\Gamma)$ and $L^\infty(\Gamma)$. However, we cannot expect to obtain numerically the convergence order $\mathcal{O}(h^{1-\tau})$, for some $0 < \tau < 1/2$, in the

h	$EOC_{L^2(\Omega)}(\bar{y})$	$EOC_{H^1(\Omega)}(\bar{y})$	$EOC_{L^2(\Omega)}(\bar{\varphi})$	$EOC_{L^2(\Gamma)}(\bar{u})$	$EOC_{L^\infty(\Gamma)}(\bar{u})$
$2^{-1} \to 2^{-2}$	1.3263	0.9301	1.2267	0.6900	0.0641
$2^{-2} \to 2^{-3}$	1.7321	1.0224	1.8366	1.0634	0.8040
$2^{-3} \to 2^{-4}$	1.9174	1.0162	1.9467	0.9999	0.9534
$2^{-4} \to 2^{-5}$	1.9724	1.0052	1.9542	1.0436	1.0227
$2^{-5} \to 2^{-6}$	1.9818	1.0014	1.8819	0.9948	0.9369
$2^{-6} \to 2^{-7}$	1.9475	1.0004	1.5973	0.9978	0.9907

TABLE 4.1. Experimental order of convergence for the optimal values.

$L^\infty(\Gamma)$ norm. Moreover, the experimental order of convergence is $\mathcal{O}(h^2)$ for both, $\|\bar{y} - \bar{y}_h\|_{L^2(\Omega)}$ and $\|\bar{\varphi} - \bar{\varphi}_h\|_{L^2(\Omega)}$, and $\mathcal{O}(h)$ for $\|\bar{y} - \bar{y}_h\|_{H^1(\Omega)}$.

Next we report on two other discretization techniques known from the literature. In the first approach, the finite dimensional set $\mathcal{U}_h$ of discrete controls consists of piecewise linear continuous functions, i.e.

$$\mathcal{U}_h = U_h^{lin} := \{u \in L^\infty(\Gamma) \,|\, u \in C(\Gamma) \text{ and } u \in \mathcal{P}_1((x_j, x_{j+1})) \text{ for } 1 \le j \le N(h)\}\,.$$

In the context of semilinear elliptic control problems, this kind of discretization was studied already by Casas and Mateos [**27**] for Neumann controls, and Casas [**16**] for distributed controls. We mention also the paper by Rösch [**84**]. In view of the analysis, this approach is more delicate, since there is no pointwise projection formula for the optimal control and its approximation such as (3.3.12) on page 81 and (4.2.16).

In both cases, Neumann and distributed controls, superlinear convergence was proved for the optimal controls in the L^2-norm. Moreover, if the cost function J is quadratic w.r.t. the control variable then convergence of order $\mathcal{O}(h^{3/2})$ and $\mathcal{O}(h)$ for the controls in L^2- and L^∞-norms were obtained, cf. [**27**]. For the numerical realization of this approach we have used the same optimization code as for the piecewise constant approximation, with obvious modifications.

The second approach is the so-called *variational discretization*, where only the state and the adjoint state are discretized, i.e. $\mathcal{U}_h = L^\infty(\Gamma)$, cf. Hinze [**63**]. Nevertheless, the optimal control is obtained by a projection of the adjoint state to the set of admissible controls, therefore it is piecewise linear. Replacing the optimal control by this projection, the optimality system reduces to the system containing the state and adjoint state equation. For linear-quadratic distributed or Neumann control problems numerous numerical tests have shown a quadratic convergence for the optimal controls in the L^2-norm, cf. Hinze [**63**], Hinze and Matthes [**65**]. To solve numerically the optimality system we have used the commercial finite element solver of the equation-based modeling and simulation environment COMSOL Multiphysics[1]. We had two reasons for choosing this solver for our purposes. First, it allows a very simple implementation of the coupled system. Secondly, it is much faster than our own optimization routines, since it follows the so-called *all-at-once approach*, where the optimality system is solved at once.

Finally, we refer to Casas and Tröltzsch [**38**], where both strategies are treated for the distributed control of (1.1.1) with homogeneous Dirichlet boundary conditions. The authors have derived the order $\mathcal{O}(h^\sigma)$ of the L^2-error for the optimal controls, with $\sigma = 3/2$ when J is quadratic w.r.t u and the controls are approximated by piecewise linear functions and $\sigma = 2$ for the case of variational discretization.

By applying these discretization methods to Example 4.14, we want to check if we can observe convergence rates for the optimal controls similar to those for semilinear problems in convex domains. Table 4.2 illustrates the results obtained from the tests of the approaches described above. Obviously, these results reflect exactly the order of convergence observed in the case of semilinear Neumann control problems in convex domains, cf. [**27**] and [**65**].

$\mathcal{U}_h = U_h^{lin}$		
h	$EOC_{L^2(\Gamma)}(\bar{u})$	$EOC_{L^\infty(\Gamma)}(\bar{u})$
$2^{-2} \to 2^{-3}$	1.1743	0.6466
$2^{-3} \to 2^{-4}$	1.4382	1.1106
$2^{-4} \to 2^{-5}$	1.6496	1.5556
$2^{-5} \to 2^{-6}$	1.5295	1.0026
$2^{-6} \to 2^{-7}$	1.4678	1.0233
$2^{-7} \to 2^{-8}$	1.5570	1.0662

$\mathcal{U}_h = L^\infty(\Gamma)$	
h	$EOC_{L^2(\Gamma)}(\bar{u})$
$2^{-1} \to 2^{-2}$	1.7850
$2^{-2} \to 2^{-5}$	1.8981
$2^{-3} \to 2^{-4}$	2.1661
$2^{-4} \to 2^{-5}$	2.2123
$2^{-5} \to 2^{-6}$	1.9922
$2^{-6} \to 2^{-7}$	2.0948

TABLE 4.2. Experimental convergence rates for the optimal controls.

EXAMPLE 4.15. *We fix* $\Omega = (0,1) \times (0,0.5) \cup (0.5,1) \times (0,1)$ *(compare also Example 2.21 on page 58),* $\alpha = 0$, $\beta = 1.8$, $\eta \equiv 0$, $g \equiv 0$ *on* Γ, $a(x, y(x)) = 1+(x_1+x_2)^2+y^2(x)$ *and* $f(x, y(x)) = y + y^3(x)$. *Furthermore, we set*

$$u_d(x) = u_d(x_1, x_2) = Proj_{[\alpha,\beta]} \left\{ |x_1 - x_2|^{1/6} + x_1 \right\}$$

and y_d *is equal to the numerical approximation of* y_{u_d} *on the same mesh on which we have made the computations, i.e.* $y_d := y_h(u_d)$.

We have taken $y_1 \equiv 1$, $\varphi_1 \equiv 1$ and $u_1 \equiv (\alpha + \beta)/2$ as initial data to run the SQP method. Figure 4.2 shows u_d and $y_h(u_d)$ with $h = 2^{-7}$. Obviously, $y_d \to y_{u_d}$ as

[1] COMSOL Multiphysics, www.comsol.com.

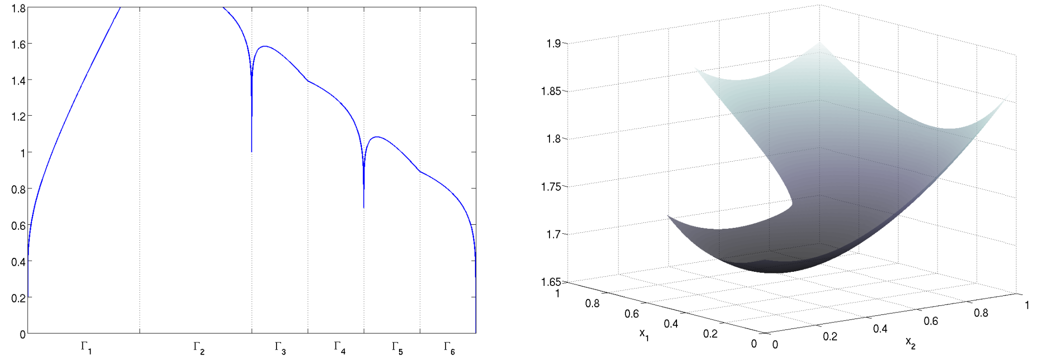

FIGURE 4.2. Desired control u_d (left frame) and associated numerical solution $y_h(u_d)$ for $h = 2^{-7}$ (right frame).

h tends to zero and the rate of this convergence is $\mathcal{O}(h)$; see Theorem 2.6 on page 41. On the other hand, since we do not know any solution $\bar{u}$ of the control problem (E), we have taken as reference solution the numerical solution $\bar{u}_{h_{fine}}$ computed on a very fine grid with mesh size $h_{fine} = 2^{-8}$. Hence, instead of studying the behavior of the error $\|\bar{u} - \bar{u}_h\|_X$ when $h \to 0$, we have considered the error $\|\bar{u}_{h_{fine}} - \bar{u}_h\|_X$, where $X = L^2(\Gamma)$ or $X = L^\infty(\Gamma)$. We should remark that the order of convergence of the optimal controls in both norms is not affected by the fact that y_d is not fixed but depends on h. This is due to the fact that the order of convergence of y_d to y_{u_d} in the $L^2(\Omega)$ norm is better than the expected ones for the optimal controls. More precisely, according to Theorems 4.10 and 4.12, we expect to see numerically the orders $\mathcal{O}(h^{1/2})$ and $\mathcal{O}(h^{1/2-\tau})$ for the optimal controls in the $L^2(\Gamma)$ and $L^\infty(\Gamma)$ norms, respectively, while $\|y_d - y_{u_d}\|_{L^2(\Omega)} = \mathcal{O}(h)$. Table 4.3 shows the convergence speed of $\bar{u}_{h_{fine}} - \bar{u}_h$ in the $L^2(\Gamma)$ and $L^\infty(\Gamma)$ norms. We do not report on the convergence rates for the optimal states and adjoint states because of their dependence on the order of the convergence $y_d \to y_{u_d}$ as h tends to zero.

h	$EOC_{L^2(\Gamma)}(\bar{u}_{h_{fine}})$	$EOC_{L^\infty(\Gamma)}(\bar{u}_{h_{fine}})$
$2^{-1} \to 2^{-2}$	0.7199	0.4427
$2^{-2} \to 2^{-3}$	0.7157	0.4366
$2^{-3} \to 2^{-4}$	0.7133	0.4640
$2^{-4} \to 2^{-5}$	0.7342	0.5365
$2^{-5} \to 2^{-6}$	0.7977	0.6928

TABLE 4.3. Experimental convergence rates for the optimal controls.

A close look at Table 4.3 reveals that the convergence behavior for the optimal controls is comparable to the one predicted in Theorem 4.12, while the convergence rate in the $L^2(\Gamma)$ norm is slightly better than the one obtained in Theorem 4.10.

CHAPTER 5

Extensions to polyhedral domains of dimension three

In this chapter, we briefly comment on some extensions of the results of Chapter 1 and 2 to dimension three. We restrict our discussion to the theoretical and numerical analysis of the quasilinear equation (1.1.1). We neither carry out the numerical analysis for the adjoint equation, nor consider optimal control problems governed by (1.1.1). Both issues are open for quasilinear equations with inhomogeneous boundary conditions in 3D. The corresponding Dirichlet problem in dimension three has already been studied by Casas and Tröltzsch in [**36, 37, 38**].

For elliptic problems in three dimensional domains with corners there does not exist a $W^{2,r}(\Omega)$ regularity result similar to the one of Theorem 1.21 on page 18. To our best knowledge, such a regularity result is missing even in the case when Ω is convex. Therefore, the maximal regularity of y_u, when $u \neq 0$, that we can work with is $H^{3/2}(\Omega)$; see Theorem 5.4.

Throughout this chapter, we will assume the following hypothesis.

Assumption 5.1. *Let $\Omega \subset \mathbb{R}^3$ be an open bounded polyhedral domain with a Lipschitz boundary Γ, $\nu(x)$ is the unit outward normal vector to Γ at x and suppose that the Assumptions 1.2-1.3 hold true with $p > 3/2$.*

This chapter follows the plan of the Chapters 1 and 2. First, we discuss the existence and uniqueness of a solution of (1.1.1). Then we give some regularity results of this solution. Later, in order to carry out the error analysis for the finite element approximation of (1.1.1), we will prove a preliminary result similar to Theorem 1.35 on page 30. All these results are contained in Section 5.1. Finally, in Section 5.2, we derive error estimates for the numerical approximation of (1.1.1) by linear finite elements.

5.1. Well-posedness of the quasilinear equation

In this section, we establish the well-posedness of the quasilinear PDE (1.1.1). Moreover, we study the regularity of its solution as we did in Chapter 1 in dimension two. The next theorem is the counterpart of Theorem 1.5 on page 3.

Theorem 5.2. *For any $u \in L^s(\Gamma)$ with $s > 2$ the equation (1.1.1) has a unique solution $y_u \in H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$ with some $\mu \in (0,1)$ independent of u.*

Proof. The proof of this theorem is almost identical to that of Theorem 1.5. The only difference to the two-dimensional case is that the embeddings $H^1(\Omega) \hookrightarrow L^q(\Omega)$ and $H^{1/2}(\Gamma) \hookrightarrow L^r(\Gamma)$ hold only if $1 \le q \le 6$ and $1 \le r \le 3$. These embeddings, along with $s > 2$ and $p > 3/2$, allow us to deduce the continuity of the solution y_u, see Stampacchia [**88**] or Murthy and Stampacchia [**81**, Theorem 2.9], as well as to prove inequality (1.3.12) on page 8 even in dimension three. □

Theorem 5.3. *Assuming that $a : \bar\Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is continuous, there exists $\bar r > 3$ such that, for any*

$$3 < r \le \begin{cases} \min\left\{\bar r, \dfrac{3p}{3-p}\right\} & \textit{if}\, p \in \left(\dfrac{3}{2}, 3\right), \\ \bar r & \textit{if}\, p \ge 3\,, \end{cases}$$

and any $u \in L^{2r/3}(\Gamma)$, the solution y_u of (1.1.1) belongs to $W^{1,r}(\Omega)$. Finally, if Ω is convex then $\bar r \ge \frac{6}{3-\sqrt{5}}$.

Proof. The proof is basically along the lines of the proof of Theorem 1.9; see pages 9-11. We will only comment on the differences in dimension three. First, let us consider the case when $3 < r \le \min\{(3p)/(3-p), 6\}$ if $p < 3$ and $r \in (3, 6]$ otherwise. Taking $u \in L^{2r/3}(\Gamma)$ arbitrarily, we have to show that G defined in (1.3.15) on page 9 belongs to $W^{-1,r}(\Omega)$. For this purpose, we use the embedding $W^{1,r'}(\Omega) \hookrightarrow L^{p'}(\Omega)$ which holds true under our assumption on r. On the other hand, we have to check whether the functional F defined in (1.3.16) on page 10 belongs to $W^{1,r'}(\Omega)^*$. For $z \in W^{1,r'}(\Omega)$ we have, for instance, that

$$\|z\nabla y_u\|_{L^1(\Omega)} \le C\|z\|_{L^{\frac{3r'}{3-r'}}(\Omega)} \|\nabla y_u\|_{L^{\frac{3r}{r+3}}(\Omega)} \le C\|z\|_{W^{1,r'}(\Omega)} \|\nabla y_u\|_{L^2(\Omega)}\,, \tag{5.1.1}$$

thanks to the embedding $W^{1,r'}(\Omega) \hookrightarrow L^{\frac{3r'}{3-r'}}(\Omega)$ and the identity $\left(\frac{3r'}{3-r'}\right)' = \frac{3r}{r+3} \le 2$ (notice that $3 < r \le 6$). Moreover, $z|_\Gamma \in W^{1-1/r',r'}(\Gamma) \hookrightarrow L^{\frac{2r}{2r-3}}(\Gamma)$, hence Hölder's inequality yields

$$\begin{aligned} \|z|_\Gamma u\|_{L^1(\Gamma)} \le \|u\|_{L^{\frac{2r}{3}}(\Gamma)} \|z|_\Gamma\|_{L^{\frac{2r}{2r-3}}(\Gamma)} &\le C\|u\|_{L^{\frac{2r}{3}}(\Gamma)} \|z|_\Gamma\|_{W^{1-1/r',r'}(\Gamma)} \\ &\le C\|u\|_{L^{\frac{2r}{3}}(\Gamma)} \|z\|_{W^{1,r'}(\Omega)}\,. \end{aligned}$$

As in the proof of Theorem 1.9, we apply a result by Dauge [**46**, Corollary 3.10] to obtain the existence of $\bar r > 3$ depending of the angles of Ω such that, for any r with $3 < r \le \min\{\bar r, (3p)/(3-p), 6\}$ if $p < 3$ and $3 < r \le \min\{\bar r, 6\}$ otherwise, the solution y_u of (1.1.1) is an element of $W^{1,r}(\Omega)$.

Finally, let us discuss the case when $6 < r \le \min\{\bar r, (3p)/(3-p)\}$ if $p < 3$ or $6 < r \le \bar r$ otherwise. The above arguments yield $y_u \in W^{1,6}(\Omega)$, hence (5.1.1) can be replaced now by

$$\|z\nabla y_u\|_{L^1(\Omega)} \le C\|z\|_{L^{\frac{3r'}{3-r'}}(\Omega)} \|\nabla y_u\|_{L^{\frac{3r}{r+3}}(\Omega)} \le C\|z\|_{W^{1,r'}(\Omega)} \|\nabla y_u\|_{L^6(\Omega)}\,.$$

Here we have used the fact that $\left(\frac{3r'}{3-r'}\right)' = \frac{3r}{r+3} \leq 6$ for all $1 \leq r < \infty$. The rest of the proof is the same as that of Theorem 1.9. If Ω is convex then $\bar{r} \geq \frac{6}{3-\sqrt{5}}$; see [**46**, Corollary 3.12]. □

Similar to the two-dimensional case, we are going to prove the $H^{3/2}(\Omega)$ regularity of y_u. In the proof of Theorem 1.18 on page 16, we have made explicit use of some results stated in Section 1.4. Analogously, in dimension three, Casas and the author [**18**] have shown that a solution $y \in H^1(\Omega)$ of the equation (1.4.1) on page 12 belongs to $H^{3/2}(\Omega)$ provided that $f \in L^p(\Omega)$ with $p > 3/2$ and $g \in L^2(\Gamma)$. With this result at hand we are able to state and prove the next theorem.

THEOREM 5.4. *Suppose that Assumption 1.17 holds true and let $u \in L^s(\Gamma)$ with $s > 2$. Then $y_u \in H^{3/2}(\Omega)$.*

PROOF. From Theorem 5.3 we know that $y_u \in W^{1,r}(\Omega)$ with some $r > 3$. Further, here holds the inclusion $W^{1,q}(\Omega) \subset C(\bar{\Omega})$ $\forall q > 3$. Hence, we can follow the proof of Theorem 1.18 and arrive at (1.5.2)-(1.5.3) with right-hand sides in $L^{\min\{p,r/2\}}(\Omega)$ and $L^s(\Gamma)$, respectively. This yields the assertion of the theorem. □

The next theorem summarizes the results analogous to Theorem 1.23 on page 19 and Theorem 1.35 on page 30 for the adjoint problem. We have seen already in Chapter 2 that the introduction of the adjoint equation (2.3.25) on page 48 turned out to be of great importance for the derivation of error estimates for the finite element approximation of (1.1.1). Depending on the regularity of its solution, we were able to get higher order of convergence in the $L^2(\Omega)$ norm than in the $H^1(\Omega)$ norm; see step 2.2 of the proof of Theorem 2.6 on page 44-45.

THEOREM 5.5. *Let $\tilde{a} \in L^\infty(\Omega)$ with $\tilde{a}(x) \geq \alpha > 0$ for all $x \in \Omega$, $\mathbf{b} \in L^3(\Omega)^3$ and $c \in L^p(\Omega)$, satisfying $c(x) \geq 0$ for a.a. $x \in \Omega$ and $c(x) \geq \alpha$ for all $x \in E$, where E is a measurable subset of Ω with $|E| > 0$. Then the operators $S : H^1(\Omega) \longrightarrow H^1(\Omega)^*$ and $T : H^1(\Omega) \longrightarrow H^1(\Omega)^*$ defined in (1.6.2) and (1.7.1), respectively, are isomorphisms.*

(1) *If $\tilde{a} \in W^{1,r}(\Omega)$, $\mathbf{b} \in L^r(\Omega)^3$ and $\zeta \in L^p(\Omega)$, with r given in Theorem 5.3, then there exists a unique element $\varphi \in H^{3/2}(\Omega)$ satisfying*

$$\langle T\varphi, v\rangle_{H^1(\Omega)^*,H^1(\Omega)} = \int_\Omega \zeta v \, dx \quad \forall z \in H^1(\Omega)\,. \tag{5.1.2}$$

Moreover, there exists a constant $C' > 0$, dependent on $\tilde{a}$, $\mathbf{b}$ and c, but not on ζ, such that

$$\|\varphi\|_{H^{3/2}(\Omega)} \leq C' \|\zeta\|_{L^{r/2}(\Omega)}\,. \tag{5.1.3}$$

(2) *If Ω is convex, $\tilde{a} \in W^{1,4}(\Omega)$, $\mathbf{b} \in L^4(\Omega)^3$ and $c, \zeta \in L^2(\Omega)$, then $\varphi \in H^2(\Omega)$ and there exists $C'' > 0$ independent of ζ such that*

$$\|\varphi\|_{H^2(\Omega)} \leq C'' \|\zeta\|_{L^2(\Omega)}\,. \tag{5.1.4}$$

Proof. Concerning the operators S and T, the assertion is deduced along the lines of the proofs of Theorems 1.23 and 1.35, with the obvious modifications. Notice that, in dimension three, the space $H^1(\Omega)$ is compactly embedded in $L^6(\Omega)$ which is sufficient to follow the proofs of Theorems 1.23 and 1.35.

Proof of (1). The existence and uniqueness of a solution $\varphi \in H^1(\Omega)$ of (5.1.2) is equivalent to the fact that T is an isomorphism. Let us prove first that $\varphi \in W^{1,r}(\Omega)$; the $H^{3/2}(\Omega)$ regularity will be an easy consequence of this fact; see below. From (5.1.2) if follows for every $z \in H^1(\Omega)$ that

$$\int_\Omega \{\tilde{a}(x)\nabla\varphi\cdot\nabla z + c(x)\varphi z\}\,dx = \int_\Omega (\zeta - \mathbf{b}(x)\cdot\nabla\varphi)\,z\,dx =: F(z)\,. \tag{5.1.5}$$

To deduce that $\varphi \in W^{1,r}(\Omega)$ we will apply the regularity result by Dauge [**46**, Corollary 3.10] to (5.1.5); see also the proof of Theorem 1.35-(1). To this aim, we have to show that $F \in W^{1,r'}(\Omega)^*$. The difficulty in deducing the latter result comes from the term $z\mathbf{b}\cdot\nabla\varphi$ in (5.1.5). Therefore, we will comment only on this term. For $2 \le l < r$ we define the functional $G : L^{\frac{rl}{rl-r-l}}(\Omega) \longrightarrow \mathbb{R}$ by $G(z) = \int_\Omega z\mathbf{b}(x)\cdot\nabla\varphi\,dx$. From Hölder's inequality we infer

$$|G(z)| \le \|\mathbf{b}\|_{L^r(\Omega)}\|\nabla\varphi\|_{L^l(\Omega)}\|z\|_{L^{\frac{rl}{rl-r-l}}(\Omega)}\,. \tag{5.1.6}$$

Moreover, from the Sobolev embedding theorem we have

$$W^{1,t'}(\Omega) \hookrightarrow L^{\frac{3t}{2t-3}}(\Omega) = L^{\frac{rl}{rl-r-l}}(\Omega) \ \text{ if } \ t = \frac{3rl}{3r + l(3-r)}\,. \tag{5.1.7}$$

Hence, if $\varphi \in W^{1,l}(\Omega)$ then G is a linear continuous functional on $W^{1,t'}(\Omega)$ and consequently $F \in W^{1,t'}(\Omega)^*$, provided that t satisfies the condition in (5.1.7). Now our goal is to construct a strongly monotone increasing finite sequence $\{t_i\}_{i=0}^k$ of real numbers such that if $\varphi \in W^{1,t_i}(\Omega)$, with $t_i < r$, then $\varphi \in W^{1,t}(\Omega)$, with $t = \min\{t_{i+1}, r\}$. Define recursively

$$t_0 = 2 \ \text{ and } \ t_{i+1} = \frac{3rt_i}{3r + t_i(3-r)} \quad i = 0, 1, \dots$$

It is immediate to check that $\{t_i\}_{i\ge 0}$ is a strictly increasing sequence of positive numbers, therefore $t_i \ge 2$ for every $i \ge 0$. Consequently, because $r > 3$, we have

$$t_{i+1} = \frac{3rt_i}{3r + t_i(3-r)} \ge \frac{3r}{3r + 2(3-r)}t_i = \frac{3r}{r+6}t_i\,,$$

thus we obtain

$$t_{i+1} - t_i \ge \left(\frac{3r}{r+6} - 1\right)t_i = \frac{2r-6}{r+6}t_i \ge 2\frac{2r-6}{r+6} > 0\,.$$

Since the difference between t_{i+1} and t_i is bounded from below by $2\frac{2r-6}{r+6} > 0$, there exists $k \in \mathbb{N}$ such that $t_k \ge r$ and we conclude that $\varphi \in W^{1,r}(\Omega)$. It remains to prove the $H^{3/2}(\Omega)$ regularity of φ and (5.1.3). By passing to the Neumann formulation of

(5.1.2), expanding the divergence term and dividing by $\tilde{a} = \tilde{a}(\cdot) > 0$ we end up with equation (1.7.7) on page 31, where s is replaced by r. By the arguments on page 123, the $H^{3/2}(\Omega)$ regularity of φ is then completed, while (5.1.3) is deduced as (1.7.3) on page 30.

Proof of (2). The $H^2(\Omega)$ regularity of φ is an immediate consequence of [**60**, Theorem 3.2.1.3] applied to (1.7.7). It is enough to check that the right-hand side of the first equation of (1.7.7) is in $L^2(\Omega)$. This can be deduced easily from the assumptions of our theorem and the fact that $\varphi \in W^{1,4}(\Omega)$. Let us prove this $W^{1,4}(\Omega)$ regularity. Since $\mathbf{b} \in L^4(\Omega)^3$ and $\varphi \in H^{3/2}(\Omega) \subset W^{1,3}(\Omega)$, the estimate (5.1.6) is replaced by

$$|G(z)| \leq \|\mathbf{b}\|_{L^4(\Omega)} \|\nabla\varphi\|_{L^3(\Omega)} \|z\|_{L^{\frac{12}{5}}(\Omega)}\,. \tag{5.1.8}$$

From the embedding

$$W^{1,4/3}(\Omega) \hookrightarrow L^{\frac{12}{5}}(\Omega)$$

and (5.1.8) it follows that G is a linear continuous functional on $W^{1,4/3}(\Omega)$, therefore F, defined in (5.1.5), belongs to $W^{1,4/3}(\Omega)^*$. Finally, since Ω is convex, we can apply [**46**, Corollary 3.12] to deduce $\varphi \in W^{1,4}(\Omega)$. The estimate (5.1.4) is obtained following the lines of the proof of (1.7.4) on page 30. □

5.2. Numerical analysis of the quasilinear equation

This subsection is concerned with the finite element based approximation of (1.1.1) and its error analysis. In what follows $u \in L^s(\Gamma)$ with $s > 2$ is fixed and we denote by $y = y_u$ the solution of (1.1.1) corresponding to u. From Theorem 5.3 we know that $y \in W^{1,r}(\Omega)$ with some $r > 3$. We also suppose that Assumption 1.17 holds true, hence $y \in H^{3/2}(\Omega)$; see Theorem 5.4.

By using the triangulation $\mathcal{T}_h$ introduced in Chapter 2, we approximate (1.1.1) by the discrete problem (2.3.1) on page 39. In order to derive error estimates as we did for dimension two in Section 2.3, we use the embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$ and the inverse inequality

$$\|z_h\|_{L^\infty(\Omega)} \leq \frac{C}{h^{1/2}} \|z_h\|_{L^6(\Omega)} \quad \forall z_h \in Y_h\,, \tag{5.2.1}$$

with $C > 0$ being independent of h. The last estimate is a consequence of the following modification of the inverse estimate (2.2.6) on page 38 for dimension three:

$$\|z_h\|_{W^{m,q}(\Omega)} \leq \frac{C}{h^{3\max\{0,1/t-1/q\}+m-k}} \|z_h\|_{W^{k,t}(\Omega)} \quad \forall z_h \in Y_h\,, \tag{5.2.2}$$

if $k \leq m$ and $t, q \in [1, \infty]$, cf. Ciarlet and Lions [**43**, Theorem 17.2]. On the other hand, the estimate (2.2.4) on page 38 remains valid in 3D and we will rename again $C\varepsilon_h$ by ε_h.

THEOREM 5.6. *There exists $h_0 > 0$ such that, for any $h < h_0$, the discrete quasilinear equation (2.3.1) has at least one solution y_h that obeys*

$$\|y - y_h\|_{L^2(\Omega)} + h^{1/2}\|y - y_h\|_{H^1(\Omega)} \leq \varepsilon_h h\,, \tag{5.2.3}$$

where $\varepsilon_h \to 0$ when $h \to 0$. If $\{y_h\}_{h>0}$ is a family of solutions of (2.3.1) that is bounded in $L^\infty(\Omega)$, then (5.2.3) holds as well.

PROOF. First of all, the existence of a solution of (2.3.1) follows by the same techniques as in 2D; see the proof of Theorem 2.4 on pages 39-40. Moreover, we have

$$\|y_h\|_{H^1(\Omega)} \leq C(\|f(\cdot,0)\|_{L^p(\Omega)} + \|u\|_{L^s(\Gamma)})\,,$$

with some $C > 0$ independent of $h > 0$. Further, we will follow the lines of Theorem 2.6 on page 41 and comment only on the differences in dimension three. Let $y_h \in Y_h$ be any solution of the equation (2.3.5). For some $r > 3$ the following inequality is the analog to (2.3.12):

$$\begin{aligned}
C_0\|y_h - \Pi_h y\|^2_{H^1(\Omega)} &\leq C_M\|y - \Pi_h y\|_{H^1(\Omega)}\|y_h - \Pi_h y\|_{H^1(\Omega)} \\
&\qquad + C_{a,M}\int_\Omega |y - y_h||\nabla \Pi_h y \cdot \nabla(y_h - \Pi_h y)|\,dx \\
&\leq \varepsilon_h h^{1/2}\|y\|_{H^{3/2}(\Omega)}\|y_h - \Pi_h y\|_{H^1(\Omega)} \\
&\qquad + C\|y - y_h\|_{L^{\frac{2r}{r-2}}(\Omega)}\|y\|_{W^{1,r}(\Omega)}\|y_h - \Pi_h y\|_{H^1(\Omega)}\,.
\end{aligned}$$

Hence, arguing as in (2.3.13), we deduce

$$\|y - y_h\|_{H^1(\Omega)} \leq \varepsilon_h h^{1/2} + C\|y - y_h\|_{L^{\frac{2r}{r-2}}(\Omega)}\,. \tag{5.2.4}$$

Next we show that

$$\|y - y_h\|_{H^1(\Omega)} \leq \varepsilon_h h^{1/2} + C\|y - y_h\|_{L^2(\Omega)}\,. \tag{5.2.5}$$

Take $t = \frac{r-3}{r-2}$ then $\frac{2r}{r-2} = 2t + 6(1-t)$. Therefore, by using Hölder's inequality with $q = \frac{1}{t}$ and $q' = \frac{1}{1-t}$, we get

$$\begin{aligned}
\left(\int_\Omega |y - y_h|^{\frac{2r}{r-2}}\,dx\right)^{\frac{r-2}{2r}} &= \left(\int_\Omega |y - y_h|^{2t}|y - y_h|^{6(1-t)}\,dx\right)^{\frac{r-2}{2r}} \\
&\leq \left(\int_\Omega |y - y_h|^2\,dx\right)^{\frac{t(r-2)}{2r}} \left(\int_\Omega |y - y_h|^6\,dx\right)^{\frac{(r-2)(1-t)}{2r}} \\
&= \|y - y_h\|^{1-\frac{3}{r}}_{L^2(\Omega)}\|y - y_h\|^{\frac{3}{r}}_{L^6(\Omega)}\,.
\end{aligned}$$

Now for any $\varepsilon > 0$ the embedding $H^1(\Omega) \hookrightarrow L^6(\Omega)$ and Young's inequality, with $q = (1 - \frac{3}{r})^{-1}$ and $q' = \frac{r}{3}$, lead to

$$\|y - y_h\|^{1-\frac{3}{r}}_{L^2(\Omega)}\|y - y_h\|^{\frac{3}{r}}_{H^1(\Omega)} \leq \frac{1}{q\varepsilon^{q/q'}}\|y - y_h\|_{L^2(\Omega)} + \frac{\varepsilon}{q'}\|y - y_h\|_{H^1(\Omega)}\,.$$

Taking ε small enough in the previous inequality, (5.2.5) follows from (5.2.4). With the aid of Theorem 5.5-(1), the estimation of $\|y-y_h\|_{L^2(\Omega)}$ is identical to that done in the step 2.2 of the proof of Theorem 2.6, therefore the estimate (2.3.18) holds true. Finally, since the strong convergence $y_h \to y$ in $L^2(\Omega)$ is valid and can be shown as in 2D (see the third step of the proof of Theorem 2.6), inequality (2.3.18) and (5.2.5) imply (5.2.3). At this point, we remark that Lemma 2.9 on page 47 remains valid in dimension three.

To conclude the proof of the theorem as in the two dimensional case we have to show the convergence $y_h \to y$ in $L^\infty(\Omega)$ as $h \to 0$. To this aim, let us denote by $P_h : L^2(\Omega) \to Y_h$ the projection operator in the $L^2(\Omega)$ sense. Then it is known that there exists a constant $C_0 > 0$ independent of y such that $\|P_h y\|_{L^q(\Omega)} \leq C_0^\theta \|y\|_{L^q(\Omega)}$ and

$$\|y - P_h y\|_{L^q(\Omega)} \leq (1 + C_0)^\theta \inf_{w_h \in Y_h} \|y - w_h\|_{L^q(\Omega)} \,,$$

for $1 \leq q \leq \infty$ and $\theta = |1 - \frac{2}{q}|$. These results are due to Douglas, Dupont and Wahlbin [**51**]. Then we have

$$\|y - P_h y\|_{L^\infty(\Omega)} \leq (1 + C_0) \inf_{w_h \in Y_h} \|y - w_h\|_{L^\infty(\Omega)} \leq (1 + C_0) \|y - \mathcal{I}_h y\|_{L^\infty(\Omega)} \,, \tag{5.2.6}$$

where $\mathcal{I}_h$ is the nodal interpolation operator defined on page 38. In dimension three, the exponent of h in the formula (2.2.5) on page 38 is $3\,(1/t - 1/q) + m - k$, cf. Ciarlet [**42**, Theorem 3.1.6]. Consequently,

$$\|y - \mathcal{I}_h y\|_{L^\infty(\Omega)} \leq C_1 h^{1-3/r} \|y\|_{W^{1,r}(\Omega)} \,,$$

and inserting this inequality in (5.2.6), it follows that

$$\|y - P_h y\|_{L^\infty(\Omega)} \leq C_1 (1 + C_0) h^{1-3/r} \|y\|_{W^{1,r}(\Omega)} \,.$$

On the other hand, using (5.2.1), we get

$$\begin{aligned}
\|P_h y - y_h\|_{L^\infty(\Omega)} &\leq \frac{C_2}{h^{1/2}} \|P_h y - y_h\|_{L^6(\Omega)} \\
&\leq \frac{C_2}{h^{1/2}} \Big(\|P_h y - y\|_{L^6(\Omega)} + \|y - y_h\|_{L^6(\Omega)} \Big) \\
&\leq \frac{C_2}{h^{1/2}} \Big((1 + C_0)^{2/3} \inf_{w_h \in Y_h} \|y - w_h\|_{L^6(\Omega)} + \|y - y_h\|_{L^6(\Omega)} \Big) \\
&\leq \frac{C_2}{h^{1/2}} \Big((1 + C_0)^{2/3} + 1 \Big) \|y - y_h\|_{L^6(\Omega)} \,.
\end{aligned}$$

Collecting the above results, we obtain, along with (5.2.1) and (5.2.3),

$$\begin{aligned}\|y-y_h\|_{L^\infty(\Omega)} &\le \|y-P_hy\|_{L^\infty(\Omega)}+\|P_hy-y_h\|_{L^\infty(\Omega)}\\ &\le C_1(1+C_0)h^{1-3/r}\|y\|_{W^{1,r}(\Omega)}+\frac{C_2}{h^{1/2}}\left((1+C_0)^{2/3}+1\right)\|y-y_h\|_{L^6(\Omega)}\\ &\le C_1(1+C_0)h^{1-3/r}\|y\|_{W^{1,r}(\Omega)}+\frac{C_2}{h^{1/2}}\left((1+C_0)^{2/3}+1\right)\|y-y_h\|_{H^1(\Omega)}\\ &\le C_1(1+C_0)h^{1-3/r}\|y\|_{W^{1,r}(\Omega)}+\varepsilon_h\to 0 \ \text{ when } \ h\to 0\,.\end{aligned}$$

The rest of the proof follows by the same procedure as in the 2D case. □

Corollary 5.7. *For every sequence* $\{y_h\}_{h<h_0}$ *of solutions of (2.3.1) satisfying (5.2.3) there holds*

$$\|y_h-y\|_{L^\infty(\Omega)}+\|y_h-y\|_{W^{1,3}(\Omega)}\to 0 \ \text{ when } \ h\to 0\,. \tag{5.2.7}$$

Proof. The convergence in $L^\infty(\Omega)$ is shown in the proof of the previous theorem. Let us prove the convergence $y_h\to y$ in $W^{1,3}(\Omega)$. Using (5.2.2), (5.2.3) and inequality (2.2.4) on page 38, we find

$$\begin{aligned}\|y_h-\Pi_hy\|_{W^{1.3}(\Omega)} &\le \frac{C}{h^{1/2}}\|y_h-\Pi_hy\|_{H^1(\Omega)}\\ &\le \frac{C}{h^{1/2}}\left(\|y_h-y\|_{H^1(\Omega)}+\|y-\Pi_hy\|_{H^1(\Omega)}\right)\le\varepsilon_h\,.\end{aligned}$$

Finally, (5.2.7) is obtained from this inequality and the convergence

$$\|y-\Pi_hy\|_{W^{1,3}(\Omega)}\le C\|y-\Pi_hy\|_{H^{3/2}(\Omega)}\to 0 \ \text{ when } \ h\to 0$$

(Eq. (2.2.3) on page 37 with $m=k=3/2$) as follows:

$$\|y_h-y\|_{W^{1,3}(\Omega)}\le\|y_h-\Pi_hy\|_{W^{1.3}(\Omega)}+\|\Pi_hy-y\|_{W^{1,3}(\Omega)}\to 0 \ \text{ when } \ h\to 0\,.$$

□

Theorem 5.8. *Assuming that* Ω *is convex,* $u\in L^{8/3}(\Gamma)$ *and* $p\ge 12/7$*, the conclusions of Theorem 5.6 remain valid with*

$$\|y-y_h\|_{L^2(\Omega)}+h\|y-y_h\|_{H^1(\Omega)}\le\varepsilon_h h^{3/2}$$

instead of (5.2.3).

Proof. According to Theorem 5.6, there exists $h_0>0$ such that, for any $h<h_0$, equation (2.3.1) has at least one solution y_h satisfying (5.2.3). To complete the proof we have only to improve the $L^2(\Omega)$ error estimate from $\mathcal{O}(h)$ to $\mathcal{O}(h^{3/2})$. For this purpose, we will follow the lines of Theorem 2.11 on page 48.

By virtue of Corollary 5.7, y_h converges to y in $L^\infty(\Omega)$, hence by choosing M as in (2.3.22) on page 46 (the supremum is now taken over $\{y_h\}_{h<h_0}$), we do not need to truncate the coefficients a and f.

Since Ω is convex, Theorem 5.5-(2) implies that the solution φ of the equation (2.3.25) on page 48 is in $H^2(\Omega)$. Notice that from Theorem 5.3 we have $y \in W^{1,4}(\Omega)$.

With these results at hand, we can proceed as in the proof of Theorem 2.11 and estimate the right-hand side of equality (2.3.28) similarly. The only difference in 3D consists in the estimation of the second term of (2.3.28):

$$\begin{aligned}\int_\Omega (f(x,y)-f(x,y_h))(\varphi-\Pi_h\varphi)\,dx &\le \int_\Omega |\phi_M(x)||y-y_h||\varphi-\Pi_h\varphi|\,dx\\ &\le \|\phi_M\|_{L^{3/2}(\Omega)}\|(y-y_h)(\varphi-\Pi_h\varphi)\|_{L^3(\Omega)}\\ &\le \|\phi_M\|_{L^{3/2}(\Omega)}\|y-y_h\|_{L^6(\Omega)}\|\varphi-\Pi_h\varphi\|_{L^6(\Omega)}\\ &\le C\|y-y_h\|_{H^1(\Omega)}\|\varphi-\Pi_h\varphi\|_{H^1(\Omega)}\\ &\le \varepsilon_h h^{3/2}\|y-y_h\|_{L^2(\Omega)}\,.\end{aligned}$$

Finally, we obtain inequality (2.3.29) which leads to (2.3.30) by the same arguments as in the proof of Theorem 2.11. This completes the proof of the $L^2(\Omega)$ error estimate. □

REMARK 5.9. *Unlike in dimension two, we cannot assure more regularity for y by increasing the regularity of u. This is due to the fact that the boundary Γ is not regular. However, Theorem 2.12 on page 50 is valid even in dimension three, cf.* [**19**, Theorem 3.7]

REMARK 5.10. *Concerning the uniqueness of solutions of the discrete equation (2.3.1), the results of Section 2.4 are still valid in 3D. Indeed, thanks to Theorem 5.6 and Corollary 5.7, it is easy to modify the proofs of Theorem 2.16 (with $p > 3/2$) and Theorem 2.18; see Section 2.4.*

Conclusions and outlook

In this thesis, we considered Neumann boundary control problems subject to a class of quasilinear elliptic equations and to box-constraints on the control. The coefficient of the main part of the operator depends on the state function, as a consequence the state equation is not monotone. Due to several difficulties we encountered, we restricted the discussion to polygonal domains of dimension two. Contributions have been made to the theoretical and numerical analysis of these problems and in particular of the underlying quasilinear equation. Although this equation is of particular type, the control of it is - with respect to the analysis - of model character for optimal boundary control problems with more general quasilinear equations or systems.

The main focus of Chapter 1 was to establish a rigorous analysis of the quasilinear equation which was necessary for further theoretical investigations concerning associated control problems. The study included the well-posedness of this equation in different spaces as well as the differentiability of the control-to-state mapping. We payed special attention to the linearization of the quasilinear equation which led again to a non-monotone equation. In the last part of this chapter, the adjoint problem was considered and existence, uniqueness and regularity results for its solutions were proved. Most of the results discussed here have been published in [**20, 18, 19**].

Chapter 2 was dedicated to the numerical approximation of the quasilinear equation by finite elements of degree one. In particular, error estimates in different norms were proved in the case of a non-convex and convex domain, respectively. These two different situations yield a different order of convergence in the $L^2(\Omega)$ norm. The theoretical results were confirmed by some numerical tests. A major difficulty is that the uniqueness of a solution to the discrete approximate equation is not guaranteed because of the non-monotone character of the state equation. This forced us to prove a local uniqueness result only. Further, the differentiability of the discrete control-to-state mapping was discussed and error estimates for the adjoint state equation were established. All results of this chapter except the numerical experiments have been published in [**18, 19**].

In Chapter 3, a Neumann boundary control problem associated with the quasilinear equation was introduced and conditions for the existence of optimal solutions were given. Furthermore, first-order necessary optimality conditions were obtained leading to a higher regularity of local solutions. Finally, a Pontryagin principle as well as second-order necessary and sufficient optimality conditions were derived. All these results are contained in [**20**].

Chapter 4 was devoted to the numerical approximation of the optimal control problem. For discretizing the state equation linear finite elements were used, while controls were approximated by piecewise constant ansatz functions. First of all, we proved that strict local minima of the continuous problem can be approximated in the sense of $L^\infty(\Gamma)$ by local minima of discrete control problems and we got estimates for the rate of this convergence. To obtain these estimates we made use of the higher regularity of optimal controls and assumed a second-order sufficient optimality condition to hold. Finally, the theoretical results were illustrated by some numerical experiments. All theoretical results and some numerical tests can be found in [**19**].

The last chapter is concerned with some possible extensions of the theory developed in the preceding chapters to dimension three. In particular, the well-posedness of the quasilinear equation was studied as well as the analysis of its approximation by linear finite elements was carried out. Following the same goals as in Chapter 2, we derived some error estimates and addressed the difficult issue of the uniqueness of discrete solutions of the equations. Most of these results have been published in [**18**].

Zusammenfassung

Diese Arbeit beschäftigt sich mit Optimalsteuerungsproblemen von quasilinearen elliptischen partiellen Differentialgleichungen mit inhomogenen Neumann-Randbedingungen. Das Randdatum wird als Kontrollvariable aufgefasst und muss vorgegebenen Ungleichungsrestriktionen genügen.

Die Steuerung von quasilinearen Gleichungen ist interessant, denn in vielen praktischen Anwendungen der Theorie der optimalen Steuerung in Ingenieur- und Medizinwissenschaften die zugrunde liegenden partiellen Differentialgleichungen sind quasilinear. Zum Beispiel in Modellen der Wärmeleitung, in denen der Wärmeleitfähigkeitskoeffizient von räumlichen Koordinaten aber auch von der Temperatur des Systems abhängt. Die Wärmeleitfähigkeit vom Kohlenstoffstahl ist sowohl von der Temperatur als auch von den Legierungen abhängig, vgl. Bejan. Sind nun die verschiedenen Stahllegierungen gleichmäßig im Gebiet verteilt, so muss der Wärmeleitfähigkeitskoeffizient auf einer hinreichend gleichmäßigen Weise von der Raumvariablen und der Temperatur abhängen. Eine ähnliche Abhängigkeit kann man auch bei der Züchtung von Siliziumkarbid-Einkristallen beobachten, vgl. Klein et al.

Die hier untersuchte quasilineare Gleichung ist nicht monoton, weil der Hauptkoeffizient des Differentialoperators selbst von der Lösung der Differentialgleichung abhängt. Die optimale Steuerung von nicht monotonen quasilinearen elliptischen Gleichungen ist erst vor Kurzem von Casas und Tröltzsch untersucht worden. Die Autoren haben den Fall verteilter Steuerungen behandelt und ihr Beitrag umfasst nicht nur die Herleitung von notwendigen und hinreichenden Optimalitätsbedingungen, sondern auch die Analysis der numerischen Approximation solcher Probleme. Es ist aber bekannt, dass im Falle der Randsteuerung die Analysis komplizierter ist, da die Zustandsfunktionen eine niedrigere Regularität aufweisen als jene verteilter Steuerungen. Das Ziel der vorliegenden Arbeit ist die Erweiterung der Ergebnisse von Casas und Tröltzsch auf den Fall von optimalen Neumann-Randsteuerungsproblemen in zweidimensionalen, polygonal berandeten Gebieten.

Um verschiedene Aspekte der theoretischen und numerischen Analysis von Optimalsteuerungsproblemen zu diskutieren, ist eine umfangreiche Untersuchung der Wohldefiniertheit der Zustandsgleichung und der Differenzierbarkeit des Steuerungs-Zustands-Operators notwendig. Diese Themen sind Gegenstand des ersten Kapitels, welches auch einige nützliche Ergebnisse zur Regularität der adjungierten Zustandsgleichung enthält.

Kapitel 2 widmet sich der Finite-Elemente-Approximation der Zustands- und adjungierten Zustandsgleichung. Der Schwerpunkt wird dabei auf die Fehleranalysis dieser Approximation gelegt. Dass die Eindeutigkeit der Lösung der diskreten quasilinearen Gleichung ein offenes Problem ist, stellt eine ernste Schwierigkeit dar. Um diese Schwierigkeit zu überwinden, wird eine lokale Eindeutigkeitsaussage bewiesen, welche für die weitere Diskussion ausreichend ist.

Im dritten Kapitel wird nun das zur quasilinearen Gleichung gehörige Optimalsteuerungsproblem formuliert und die Frage der Lösbarkeit des Problems positiv beantwortet. Darüber hinaus werden Optimalitätsbedingungen erster und zweiter Ordnung hergeleitet sowie eine Regularitätsaussage für optimale Steuerungen bewiesen.

Kapitel 4 beschäftigt sich mit der numerischen Analysis des Optimalsteuerungsproblems. Für die Approximation des Zustands und adjungierten Zustands durch finite Elemente erster Ordnung und der Steuerung durch stückweise konstante Funktionen wird die Konvergenz von diskreten lokalen Lösungen gegen eine strikte lokal optimale Steuerung des kontinuierlichen Problems gezeigt. Abschließend wird die Fehleranalysis für optimale Steuerungen aufgeführt und durch numerische Experimente bestätigt.

Im letzten Kapitel werden einige Erweiterungen der Ergebnisse der numerischen Approximation der quasilinearen Gleichung auf dreidimensionale, polyedrisch berandete Gebiete präsentiert.

Bibliography

[1] R. A. Adams. *Sobolev Spaces*, volume 65 of *Pure and applied mathematics: a series of monographs and textbooks*. Academic Press, New York – San Francisco – London, 1975.

[2] S. Agmon, A. Douglis, and L. Nirenberg. Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions. *Comm. Pure Appl. Math.*, 12:623–727, 1959.

[3] J.-J. Alibert and J.-P. Raymond. Boundary control of semilinear elliptic equations with discontinuous leading coefficients and unbounded controls. *Numer. Funct. Anal. and Optimization*, 3 and 4:235–250, 1997.

[4] W. Alt and K. Malanowski. The Lagrange-Newton Method for Nonlinear Optimal Control Problems. *Computational Optimization and Applications*, 2:77–100, 1993.

[5] N. Arada, E. Casas, and F. Tröltzsch. Error estimates for the numerical approximation of a semilinear elliptic control problem. *Computational Optimization and Applications*, 23:201–229, 2002.

[6] J. Barrett and C. Elliot. A practical finite element approximation of a semi-definite Neumann problem on a curved domain. *Numer. Math.*, 51:23–36, 1987.

[7] A. Bejan. *Convection heat transfer.* J. Wiley & Sons, New York, 1995.

[8] M. Bergounioux and K. Kunisch. Primal-dual active set strategy for state-constrained optimal control problems. *Computational Optimization and Applications*, 22:193–224, 2002.

[9] F. Bonnans and E. Casas. Une principe de Pontryagine pour le contrôle des systèmes semilinéaires elliptiques. *J. Diff. Equations*, 90:288–303, 1991.

[10] J. Bramble and J. Scott. Simultaneous approximation in scales of Banach spaces. *Math. Comp.*, 32:947–954, 1978.

[11] S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods.* Springer-Verlag, New York, Berlin, Heidelberg, 1984.

[12] H. Cartan. *Calcul différentiel. Formes différentielles.* Hermann, Paris, 1967.

[13] E. Casas. Control of an elliptic problem with pointwise state contraints. *SIAM J. Control and Optimization*, 4:1309–1318, 1986.

[14] E. Casas. Boundary control problems for quasi-linear elliptic equations: A Pontryagin's principle. *Appl. Math. Optimization*, 33, No.3:256–291, 1996.

[15] E. Casas. Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control and Optimization*, 35:1297–1327, 1997.

[16] E. Casas. Using piecewise linear functions in the numerical approximation of semilinear elliptic control problems. *Advances in Computational Mathematics*, 26:137–153, 2007.

[17] E. Casas, J. C. de Los Reyes, and F. Tröltzsch. Sufficient second-order optimality conditions for semilinear control problems with pointwise state constraints. *Siam J. Optim.*, 19, No. 2:616–643, 2008.

[18] E. Casas and V. Dhamo. Error estimates for the numerical approximation of a quaslinear Neumann problem under minimal regularity of the data. *Numerische Mathematik*, 117:115–145, 2010.

[19] 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.*, online first, DOI 10.1007/s10589-011-9440-0, 2011.

[20] E. Casas and V. Dhamo. Optimality Conditions for a class of optimal boundary control problems with quasilinear elliptic equations. *Control & Cybernetics*, 40:457–490, 2011.

[21] E. Casas and L. Fernández. Distributed control of systems governed by a general class of quasilinear elliptic equations. *J. of Diff. Equations*, 104:20–47, 1993.

[22] E. Casas and L. Fernández. Dealing with integral state constraints in boundary control of quasilinear elliptic equations. *SIAM Control Optim.*, 33(2):568–589, 1995.

[23] E. Casas, L. Fernández, and J. Yong. Optimal control of quasilinear parabolic equations. *Proc. Roy. Soc. Edinburgh Sect. A*, 125:545–565, 1995.

[24] E. Casas, A. Günther, and M. Mateos. A paradox in the approximation of Dirichlet control problems in curved domains. *Submitted*, 2010.

[25] E. Casas and M. Mateos. Second order sufficient optimality conditions for semilinear elliptic control problems with finitely many state constraints. *SIAM J. Control and Optimization*, 40:1431–1454, 2002.

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

[27] E. Casas and M. Mateos. Error estimates for the numerical approximation of Neumann control problems. *Comput. Optim. Appl.*, 39:265–295, 2008.

[28] E. Casas, M. Mateos, and J. P. Raymond. Penalization of Dirichlet optimal control problems. *ESAIM, Control Optim. Calc. Var.*, 15, No. 4:782–809, 2009.

[29] E. Casas, M. Mateos, and F. Tröltzsch. Error estimates for the numerical approximation of boundary semilinear elliptic control problems. *Computational Optimization and Applications*, 31:193–220, 2005.

[30] E. Casas, M. Mateos, and F. Tröltzsch. Necessary and sufficient optimality conditions for optimization problems in function spaces and applications to control theory. In *ESAIM: Proc.*, Volume 13:18-30, 2003.

[31] E. Casas and J.-P. Raymond. Error estimates for the numerical approximation of Dirichlet boundary control for semilinear elliptic equations. *SIAM J. Control Optim.*, 45(5):1586–1611, 2006.

[32] E. Casas, J.-P. Raymond, and H. Zidani. Pontryagins principle for local solutions of control problems with mixed control-state constraints. *SIAM J. Control and Optimization*, 39:1182–1203, 2000.

[33] E. Casas and J. Sokolowski. Approximation of boundary control problems on curved domains. *SIAM J. Control Optim.*, 48 (6):3746–3780, 2010.

[34] E. Casas and F. Tröltzsch. Second order necessary optimality conditions for some state-constrained control problems of semilinear elliptic equations. *Applied Math. and Optimization*, 39:211–227, 1999.

[35] E. Casas and F. Tröltzsch. Second order necessary and sufficient optimality conditions for optimization problems and applications to control theory. *SIAM J. Optimization*, 13:406–431, 2002.

[36] 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 and Optimization*, 48, No. 2:688–718, 2009.

[37] E. Casas and F. Tröltzsch. Numerical analysis of some optimal control problems governed by a class of quasilinear elliptic equations. *ESAIM: COCV, online first.*, 2010.

[38] E. Casas and F. Tröltzsch. A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. *Comput. Optim. Appl.*, pages online first, DOI 10.1007/s10589–011–9453–8, 2011.

[39] E. Casas, F. Tröltzsch, and A. Unger. Second order sufficient optimality conditions for a nonlinear elliptic control problem. *J. for Analysis and its Applications (15)*, 15:687–707, 1996.

[40] E. Casas, F. Tröltzsch, and A. Unger. Second order sufficient optimality conditions for some state-constrained control problems of semilinear elliptic equations. *SIAM J. Control and Optimization*, 38(5):1369–1391, 2000.

[41] E. Casas and J. Yong. Maximum principle for state-constrained optimal control problems governed by quasilinear elliptic equations. *Diff. and Integral Equations*, 8(1):1–18, 1995.

[42] P. Ciarlet. *The Finite Element Method for Elliptic Problems.* North-Holland, Amsterdam, 1978.

[43] P. Ciarlet and L. L. Lions. *Handbook of Numerical Analysis, Vol. II, Part I – Finite Element Methods.* North-Holland, Amsterdam, 1991.

[44] M. Dauge. *Elliptic boundary Value Problems on Corner Domains - Smoothness and Asymtotics of Solutions.* Lecture Notes in Mathematics, Vol. 1341, Springer-Verlag, 1988.

[45] M. Dauge. Problèmes mixtes pour le laplacien dans des domaines polyédraux courbes. *C. R. Acad. Sci. Paris*, t. 309, Série I:553–558, 1989.

[46] M. Dauge. Neumann and mixed problems on curvilinear polyhedra. *Integral Equations Oper. Theory*, 15, No.2:227–261, 1992.

[47] K. Deckelnick, A. Günther, and M. Hinze. Finite element approximation of Dirichlet boundary control for elliptic PDEs on two- and three-dimensional curved domains. *SIAM J. Control Optim.*, 48 (4):2798–2819, 2009.

[48] Z. Ding. A proof of the trace theorem of Sobolev spaces on Lipschitz domains. *Proceedings of the American Mathematical Society*, Vol. 124, No. 2:591–600, 1996.

[49] A. L. Dontchev, W. W. Hager, A. B. Poore, and B. Yang. Optimality, stability, and convergence in nonlinear control. *Applied Math. and Optimization*, 31:297–326, 1995.

[50] J. J. Douglas and T. Dupont. A Galerkin Method for a Nonlinear Dirichlet problem. *Math. Comp.*, 29, No.131:689–696, 1975.

[51] J. J. Douglas, T. Dupont, and L. Wahlbin. The stability in L^q of the L^2-projection into finite element function spaces. *Numerische Mathematik*, 23:193–197, 1975.

[52] J. C. Dunn. On second order sufficient optimality conditions for structured nonlinear programs in infinite-dimensional function spaces. In A. Fiacco, editor, *Mathematical Programming with Data Perturbations*, pages 83–107. Marcel Dekker, 1998.

[53] I. Ekeland and R. Temam. *Convex analysis and variational problems.* North Holland, Amsterdam, 1976.

[54] H. Gajewski, K. Gröger, and K. Zacharias. *Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen.* Akademie–Verlag, Berlin, 1974.

[55] M. Giaquinta. *Introduction to regularity theory for nonlinear elliptic systems.* Birkhäuser, Basel, 1993.

[56] H. Goldberg and F. Tröltzsch. Second order optimality conditions for a class of control problems governed by nonlinear integral equations with application to parabolic boundary control. *SIAM J. Control and Optimization*, 20:687–698, 1989.

[57] H. Goldberg and F. Tröltzsch. Second order optimality conditions for nonlinear parabolic boundary control problems. In *Lecture Notes Contr.Inf.Sci.*, pages 93–103. Proc. Int.Conference on optimal control of partial differential equations, 1991.

[58] H. Goldberg and F. Tröltzsch. Second order sufficient optimality conditions for a class of non–linear parabolic boundary control problems. *SIAM J. Control and Optimization*, 31(4):1007–1025, 1993.

[59] J. A. Griepentrog and L. Recke. Linear elliptic boundary value problems with non-smooth data: Normal solvability on Sobolev-Campanato spaces. *Math. Nachr.*, 225:39–74, 2001.

[60] P. Grisvard. *Elliptic Problems in Nonsmooth Domains.* Pitman, Boston, 1985.

[61] K. Gröger. A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Math. Ann.*, 283, No.4:679–687, 1989.

[62] M. Heinkenschloss and F. Tröltzsch. Analysis of the Lagrange-SQP-Newton method for the control of a phase field equation. *Control and Cybernetics*, 28(2):178–211, 1999.

[63] M. Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *J. Computational Optimization and Applications*, 30:45–63, 2005.

[64] M. Hinze and K. Kunisch. Second order methods for optimal control of time-dependent fluid flow. *SIAM J. Control and Optimization*, 40:925–946, 2001.

[65] M. Hinze and U. Matthes. A note on variational discretization of elliptic Neumann boundary control. *Control and Cybernetics*, 38:577–591, 2009.

[66] I. Hlaváček. Reliable solution of a quasilinear nonpotential elliptic problem of a nonmonotone type with respect to the uncertainty in coeffcients. *J. Math. Anal. Appl.*, 212:452–466, 1997.

[67] I. Hlaváček, M. Křížek, and J. Malý. On Galerkin approximations of a quasilinear nonpotential elliptic problem of a nonmonotone type. *Journal of Mathematical Analysis and Applications*, 184:168–189, 1994.

[68] K. Ito and K. Kunisch. Augmented Lagrangian-SQP methods for nonlinear optimal control problems of tracking type. *SIAM J. Control and Optimization*, 34:874–891, 1996.

[69] K. Ito and K. Kunisch. Semi-smooth Newton methods for state-constrained optimal control problems. *Systems and Control Letters*, 50:221–228, 2003.

[70] D. Jerison and C. Kenig. The Neumann problem on Lipschitz domains. *Bull. Amer. Math. Soc. (N.S.)*, 4:203–207, 1981.

[71] D. Jerison and C. Kenig. The inhomogeneous Dirichlet problem in Lipschitz domains. *J. Funct. Anal.*, 130:161–219, 1995.

[72] C. Kenig. *Harmonic Analysis Techniques for Second Order Elliptic Boundary Value Problems.* vol. 83 of CBMS, American Mathematical Society, Providence, Rhode Island, 1994.

[73] O. Klein, P. Philip, J. Sprekels, and K. Wilmański. Radiation- and convection-driven transient heat transfer during sublimation growth of silicon carbide single crystals. *J. of Crystal Growth*, 222:832–851, 2001.

[74] K. Krumbiegel, C. Meyer, and A. Rösch. A priori error analysis for linear quadratic elliptic Neumann boundary control problems with control and state constraints. *SIAM J. Control Optim.*, 48 (8):5108–5142, 2010.

[75] K. Kunisch and A. Rösch. Primal-Dual Active Set Strategy for a General Class of Constrained Optimal Control Problems. *SIAM J. on Optimization*, 13:321–334, 2002.

[76] K. Kunisch and E. W. Sachs. Reduced SQP-methods for parameter identification problems. *SIAM Journal Numerical Analysis*, 29:1793–1820, 1992.

[77] J. L. Lions. *Quelques mèthodes des rèsolution des problèmes aux limites non linèaires.* Dunod, Gauthier–Villars, Paris, 1969.

[78] L. Liu, M. Křížek, and P. Neittaanmäki. Higher order finite element approximation of a quasilinear elliptic boundary value problem of a non-monotone type. *Appl. Math., Praha*, 41, No.6:467–478, 1996.

[79] C. Meyer and A. Rösch. Superconvergence properties of optimal control problems. *SIAM J. Control and Optimization*, 43:970–985, 2004.

[80] C. Morrey. *Multiple integrals in the calculus of variations.* Springer-Verlag, New York, 1966.

[81] M. K. V. Murthy and G. Stampacchia. A variational inequality with mixed boundary conditions. *Israel J. Math.*, 13:188–224, 1972.

[82] J. Nečas. *Les méthodes directes an théorie des equations elliptiques.* Editeurs Academia, Prague, 1967.

[83] J.-P. Raymond and H. Zidani. Hamiltonian Pontryagin's Principles for control Problems governed by Semilinear Parabolic Equations. *Applied Mathematics and Optimization*, 39:143–177, 1999.

[84] A. Rösch. Error estimates for linear-quadratic control problems with control constraints. *Optimization Methods and Software*, 21, No. 1:121–134, 2006.

[85] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1970.

[86] A. H. Schatz. An Observation Concerning Ritz–Galerkin Methods with Indefinite Bilinear Forms. *Mathematics of Computation*, Vol. 28, Nr. 128:959–962, 1974.

[87] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54, no. 190:483–493, 1990.

[88] G. Stampacchia. Problemi al contorno ellittici, con dati discontinui, dotati di soluzioni hölderiane. (Italian). *Ann. Mat. Pura Appl., IV. Ser.*, 51:1–38, 1960.

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

[90] F. Tröltzsch. On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. *SIAM J. Control and Optimization*, 38:294–312, 1999.

[91] F. Tröltzsch. *Optimal Control of Partial Differential Equations. Theory, Methods and Applications.* To appear in the Graduate Studies Series, AMS, 2010.

[92] J. Wloka. *Partielle Differentialgleichungen.* Teubner-Verlag, Leipzig, 1982.