Numerical analysis of PDE constrained optimal control problems with pointwise inequality constraints on the state and the control

vorgelegt von

Dipl.-Math. techn. Ira Neitzel

aus Dresden

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

Doktorin der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Rolf H. Möhring
Gutachter:	Prof. Dr. Fredi Tröltzsch
Gutachter:	Prof. Dr. Roland Herzog

Tag der wissenschaftlichen Aussprache: 08.06.2011

Berlin 2011
D 83

Zusammenfassung

Gegenstand dieser Arbeit ist die numerische Analysis von Optimierungsproblemen mit partiellen Differentialgleichungen (PDEs), deren Zustand oder Steuerung punktweisen Ungleichungsbeschränkungen unterliegt. Wir interessieren uns insbesondere für nichtkonvexe Probleme mit semilinearer Zustandsgleichung. Es liegt in der Natur derartiger Probleme, dass eine Lösung oftmals nur numerisch gefunden werden kann. Man interessiert sich deshalb für den Fehler zwischen einer (lokalen) Lösung des kontinuierlichen Problems und einer zugehörigen (lokalen) diskreten Lösung. Eine umfassende Diskussion des kontinuierlichen Problems ist dabei eine Grundvoraussetzung. Insbesondere bei vorhanden punktweisen Zustandsbeschränkungen treten spezifische Schwierigkeiten sowohl analytischer als auch numerischer Art auf, denen man entweder direkt oder mit Hilfe von Regularisierungsansätzen begegnen kann.

Mit dieser Dissertation leisten wir auf verschiedene Weise neue Beiträge zur Diskussion von Optimalsteuerungsproblemen mit punktweisen Zustandsschranken aber auch reinen Kontrollschranken. Nach kurzer Einführung in die Thematik und Bereitstellung gewisser Grundlagen beschäftigen wir uns in Kapitel 3 mit einem elliptischen semiinfiniten Optimalsteuerungsproblem. Bekannte Resultate zu notwendigen und hinreichenden Optimalitätsbedingungen, die vergleichsweise hohe Regularität von Lösungen elliptischer PDEs und die Endlichdimensionalität des Steuerungsraumes lassen eine direkte Diskussion von a priori Diskretisierungsfehlerabschätzungen für dieses Problem zu. Unser Hauptergebnis in diesem Kapitel ist eine a priori Fehlerschranke der Ordnung $\mathcal{O}(h^2|\ln h|)$ für lokale Lösungen einer Finite-Elemente-Diskretisierung des Optimalsteuerungsproblems mit Gitterweite h in einem zweidimensionalen Ortsgebiet. Dazu stellen wir gewisse Annahmen an die Struktur der aktiven Menge.

In Kapitel 4 betrachten wir ein parabolisches Optimalsteuerungsproblem mit punktweise beschränkten Steuerungsfunktionen, semilinearer Zustandsgleichung und punktweisen Zustandsbeschränkungen im gesamten Orts-Zeit-Gebiet. Im Gegensatz zu dem in Kapitel 3 diskutierten elliptischen Problem sind hier u.a. hinreichende Bedingungen zweiter Ordnung nur für eindimensionale Ortsgebiete verfügbar. Auch läßt die Verwendung von Steuerungsfunktionen an Stelle endlich vieler Parameter keine sinnvollen a priori Annahmen an die Struktur der aktiven Menge zu. Wir regularisieren daher das Problem mit der auf Meyer, Rösch und Tröltzsch zurückgehenden Lavrentievregularisierung, und können so unter anderem auf bekannte Resultate zurückgreifen, die eine höhere Regularität der Lagrangeschen Multiplikatoren sichern und eine tiefergehende Analysis ermöglichen. Wir beweisen ein Konvergenzresultat für lokale Lösungen des regularisierten Problems und weisen die lokale Eindeutigkeit regularisierter Lösungen nach.

In Kapitel 5 untersuchen wir die Finite-Element-Diskretisierung eines kontrollbeschränkten parabolischen Optimalsteuerungsproblems mit semilinearer Zustandsgleichung. Wir beweisen Fehlerordnungen für diskrete lokale Lösungen in der L^2-Norm. Dabei erweitern wir Resultate, die für linear-quadratische Probleme bekannt sind, auf den nichtkonvexen Fall. Es müssen insbesondere Beschränktheitsresultate in der L^∞-Norm der semidiskreten und diskreten Zustände gezeigt werden, die unabhängig von den Diskretisierungsparametern gelten. Außerdem erfordert die Diskussion lokal optimaler Lösungen, dass Konvergenzresultate und quadratische Wachstumsbedingungen in denselben Normen betrachtet werden.

Abstract

The purpose of this thesis is the numerical analysis of optimal control problems with partial differential equations (PDEs), whose control or state is subject to pointwise inequality constraints. We are specifically interested in nonconvex problems with semilinear state equation. It is intrinsic to the considered problem class that solutions can often only be found by numerical methods. Consequently, one is interested in estimating the error between a (local) solution of the continuous problem and an associated discrete local solution. A basic requirement for this purpose is a thorough discussion of the continuous problem. In the presence of pointwise state constraints this leads to specific difficulties of analytical and numerical nature. These difficulties have to be approached either directly or by means of regularization.

With this thesis we make several new contributions to the discussion of optimal control problems with pointwise state constraints but also those with pure control constraints. After a short introduction into the field of research and providing some basic theoretical results we discuss in Chapter 3 a semiinfinite elliptic optimal control problem. Known results on necessary and sufficient optimality conditions, the comparably high regularity of solutions to elliptic PDEs, as well as the finite dimensional control space allows to address a priori discretization error estimates for this problem directly, i. e. without further regularization. Our main result in this chapter is an a priori error bound of order $\mathcal{O}(h^2|\ln h|)$ for local solutions of a finite element discretization with mesh size h of this optimal control problem in two space dimensions. For that, we rely on certain assumptions on the structure of the active set.

In Chapter 4 we address a parabolic optimal control problem with L^∞ bounds on the control functions, semilinear state equation, and pointwise state constraints in the whole space-time-domain. In contrast to the elliptic problem discussed in Chapter 3, second order sufficient conditions are only available for one-dimensional spatial domains. Moreover, the use of control functions instead of finitely many control parameters does not allow any a priori assumptions on the structure of the active sets. Therefore, we use a Lavrentiev regularization as suggested originally by Meyer, Rösch and Tröltzsch. Consequently, we can make use of available higher regularity results for the Lagrange multipliers that allow for a deeper analysis. We prove a convergence result for locally optimal solutions of the regularized problem and show local uniqueness of regularized local solutions.

In Chapter 5 we analyze the finite element discretization of a control-constrained parabolic optimal control problem with semilinear state equation. We prove error estimates for discrete local solutions in the L^2-norm. We extend known results for linear-quadratic problems to the nonconvex setting. In particular, we have to prove boundedness results for the semidiscrete and discrete state functions in the L^∞-norm that hold independently of the discretization parameters. Moreover, the discussion of local solutions requires convergence results and quadratic growth conditions to be considered in the same spaces.

Acknowledgments

My thesis advisor Prof. Dr. Fredi Tröltzsch accompanied my research from my first steps in optimal control to the completion of this thesis. I would like to express my gratitude to him for his support and guidance throughout the last years.

My sincere thanks goes to my paper coauthors for the fruitful collaborations, and to all my former and current colleagues for the friendly and supportive working atmosphere. To Prof. Dr. Boris Vexler (Technische Universität München) I would like to express special thanks for sharing his expertise on discontinous Galerkin methods. Moreover, I thank Prof. Dr. Roland Herzog (Technische Universität Chemnitz) for agreeing to be a co-referee for this thesis.

I gratefully acknowledge partial funding of my research by the German Research Foundation (DFG).

Contents

1. Introduction

The theory of optimal control of partial differential equations (PDEs) is a well-established field in applied mathematics that developed rapidly along with the constant increase in computing power. Optimization problems that arise are widely spread, ranging from flow control, [64], over applications in material sciences such as optimal steel cooling or hardening strategies, cf. [43, 76], to applications in medicine: in cancer therapy, for instance, the so-called local hyperthermia is employed to make tumors more susceptible for other forms of treatment, cf. [41], and optimal heating profiles have to be developed. In these examples, the processes are governed by partial differential equations that are as diverse as the applications themselves. Heat transfer equations, Maxwell's equation, or the Navier-Stokes equations, to name a few, may be encountered.

In general, a PDE constrained optimal control problem for a control $u \in U$ and a state $y \in Y$ for some control space U and state space Y to be specified takes the following abstract form:

$$\text{Minimize } J(y,u), \quad \text{subject to } u \in U_{\mathrm{ad}}, \quad y = G(u),$$

where

$$J = J(y,u)\colon Y \times U \to \mathbb{R}$$

is an objective functional that is to be minimized, and

$$G\colon U \to Y$$

denotes a so-called control to state mapping that in context of PDE constrained optimization involves a solution operator of a partial differential equation. Each control $u \in U$ is assigned a unique state $y = y(u) \in Y$. Finally, the set U_{ad} is a set of admissible controls that handles additional constraints such as control bounds or pointwise constraints on the state. In the examples mentioned initially, these constraints are essential to obtain a meaningful problem formulation. For instance, in the steel cooling process the temperature differences in the steel must be bounded in order to avoid cracks, see [43]. In local hyperthermia therapy, the generated temperature in the patient's body must not exceed a certain limit, cf. [41].

A discussion of a general PDE-constrained optimal control problem includes the discussion of the underlying PDE with respect to existence, uniqueness and regularity of solutions, the existence of optimal controls, as well as first order necessary and second order sufficient optimality conditions. In addition, adequate solution algorithms and discretization strategies have to be developed. And at last, it is desirable to estimate the quality of a discrete solution with the help of discretization error estimates. An introduction to these aspects can be found in the textbooks [144] or [74].

The analysis and even more so the development of efficient numerical solution algorithms for PDE constrained optimization problems became a vivid field in mathematics on the verge of functional analysis, optimization theory, and numerical mathematics. One of the founders of this theory is J. L. Lions, whose classical textbook [90] covers a wide-ranging theory for elliptic, parabolic, and hyperbolic optimal control problems. However, up to today there remain interesting challenges, for instance with respect to problems with nonlinear state equation, the presence of pointwise state constraints, and aspects of numerical analysis not covered by [90]. Research projects like the DFG priority program 1253 "Optimization with partial differential equations", are a public sign for this continued interest in the topic.

This thesis is a contribution to the numerical analysis of PDE constrained optimal control problems governed by semilinear elliptic and parabolic PDEs of heat equation type, subject to pointwise control and state constraints. We are especially interested in the discussion of pointwise state constraints, which are known to lead to numerous theoretical and numerical difficulties. Let us give a brief overview about some related questions and results.

One of the challenges related to pointwise state constraints lies within the formulation of first-order optimality conditions of Karush-Kuhn-Tucker-type (KKT) in useful spaces. This theory is often based on the so called Slater condition, which we will state later. Then, the cone of non-negative functions is required to have non-empty interior. This requires continuity of the state functions, because in L^p-spaces with $1 \leq p < \infty$ the cone of nonnegative functions has empty interior, while for $p = \infty$ the dual space has very low regularity. Even if optimality conditions of Karush-Kuhn-Tucker-type can be formulated, the Lagrange multipliers associated with the pointwise state constraints are generally only obtained in the space of regular Borel measures, cf. [17], which in turn leads to low regularity of the adjoint state. We refer to e.g. further investigations by Casas, [18, 19], as well as Bergounioux and Kunisch [11, 12], or Alibert and Raymond as well as Raymond and Zidani [2, 124, 125] for the discussion of pure pointwise state constraints.

Another difficulty is associated with second-order-sufficient conditions (SSC) for nonconvex problems, also a field of active research. There is quite a number of papers devoted to this subject. For PDE-constrained optimal control problems the first contributions are due to Tröltzsch and Goldberg, see [52, 53]. Further results involving state constraints have been obtained by Casas, Tröltzsch, and Unger, [28, 29] or Raymond and Tröltzsch, [123]. Yet, for parabolic problems in particular, the available theory is not as general as desired. For spatio-temporal control functions for instance, and pointwise state constraints given in the whole domain Q, a satisfactory theory of SSC is so far only available for one-dimensional distributed control problems, cf. the recent results by Casas, de los Reyes, and Tröltzsch in [22], or the earlier works by Raymond and Tröltzsch, [123]. Only in special cases, i.e. a setting with finitely many time-dependent controls, SSC have been shown to hold for higher dimensions by de los Reyes, Merino, Rehberg and Tröltzsch, cf. [36].

These theoretical challenges also influence the numerical analysis of optimal control problems, for instance when trying to derive a priori discretization error estimates. There is a number of publications on elliptic control-constrained optimal control problems, [8, 21, 25, 26, 27, 37, 71, 82, 109, 130], deriving error estimates for boundary and distributed control problems with different types of control discretization, for problems with linear and semilinear state equation. Fewer results are known for problems with pointwise state constraints in the whole domain. Here, we mention the convergence results by Casas in [20] for problems with finitely many state constraints, and the error estimates for elliptic state constrained distributed control functions by Casas and Mateos in [23], Deckelnick and Hinze in [38, 39] and Meyer in [107]. We will give a more detailed overview in Chapter 3. Further results involving gradient constraints where obtained by Günther and Hinze in [63] and Ortner and Wollner, [119]. Recently, some progress with respect to optimal error estimates has been made for state-constrained elliptic problems with finite dimensional control space. In [106], Merino, Vexler and Tröltzsch analyzed a problem with finitely many pointwise state constraints and derived error estimates that reflect the error for uncontrolled equations. Then, in [104] and [105] linear quadratic problems of semi-infinite type, i.e. with finite dimensional control space and pointwise state constraints in a domain have been analyzed by Merino, Tröltzsch, and the author, where the same order of convergence is proven under certain conditions. Comparable results for nonconvex problems with semilinear state equation will be presented in this thesis.

For parabolic optimal control problems, the situation is even more difficult. Even for parabolic control-constrained problems the theory is not as developed as in the elliptic setting. There are a number of contributions to the analysis of linear-quadratic problems. A fundamental contribution certainly comes from Malanowski, [92], where error estimates for the control are proven for Galerkin-type approximations of convex control-constrained parabolic problems. In [101, 102], Meidner and Vexler derived error estimates for the control utilizing discontinuous Galerkin schemes that clearly separate the influence of spatial and temporal discretization. We refer also to [87, 129, 150, 103] for further investigations. For nonconvex problems, we mention also the early results of McNight and Bosarge, jr., [98], where a Lagrange multiplier approach has been used. For control-constrained problems with semilinear state equation, we refer to Chrysafinos' results on plain convergence in [33], and to the recent publication of Vexler and the author, cf. [118], for finite element error estimates, extending the linear-quadratic setting from [102]. Only very few attempts have been made to derive a priori error estimates for the finite element discretization of parabolic state constrained optimal control problems. To the author's knowledge, there are only two contributions for linear-quadratic problems. Deckelnick and Hinze considered the variational discretization for finitely many time-dependent controls, cf. [40]; and linear-quadratic problems with pointwise-in-time state constraints have been discussed by Meidner, Rannacher, and Vexler in [100].

For all these reasons, regularization techniques have been a wide field of active research in the recent past and remain to be of interest. We mention for example a Moreau-Yosida regularization approach by Ito and Kunisch, [78], a Lavrentiev-regularization technique by Meyer, Rösch, and Tröltzsch, [110], or for boundary control problems the source-term representation-based Lavrentiev regularization by Tröltzsch and Yousept, [145], or Tröltzsch and the author, [116], as well as the virtual control concept by Krumbiegel and Rösch, [85]. Moreover, barrier methods, have regularizing effect. We will not discuss these methods in this thesis, but we refer for instance to Schiela, [135], Prüfert et. al. [121] or Ulbrich and Ulbrich, [147].

All mentioned types of regularization have been subject to further intensive research. The main interest lies again in the discussion of optimality conditions, but also on convergence issues and estimation of the regularization error. In addition, the question of convergence of solution algorithms is of interest if a problem is to be solved numerically. Then, estimating the discretization error of regularized problems is also desirable.

For Moreau-Yosida related publications, we refer for instance to Bergounioux, Haddou, Hintermüller and Kunisch, [10], Bergounioux and Kunisch, [12], Hintermüller, Ito, and Kunisch, [66], Hintermüller and Kunisch, [67], and Hintermüller, Tröltzsch and Yousept, [69], to name a few. Recent results by Hintermüller and Kunisch, [68], take into account not only control and state constraints but also constraints on the derivative. We also point out [112], where Meyer and Yousept derived convergence for a nonconvex elliptic problem and to [115] for parabolic convergence results for problems with semilinear state equation derived by Tröltzsch and the author.

Meanwhile, Lavrentiev regularization has also attracted much attention. Numerical methods, theoretical properties, and different problem formulations have been investigated. We refer to results by Tröltzsch et. al. in [120, 142] for linear-quadratic problems, [111] for semilinear elliptic problems, the discussion of second order sufficient conditions by Rösch and Tröltzsch in [131, 132], and further investigations by Rösch et. al., [30, 31] as well as Griesse et. al., [61, 60], for the discussion of solution algorithms. Closely related are also problems with constraints of bottleneck type, cf. the analysis of such problems by Bergounioux and Tröltzsch, [13, 14]. A challenge posed by a so called fully constrained problem with lower and upper mixed control-state constraints as well as box constraints concerning the existence of Lagrange multipliers has been solved by Rösch and Tröltzsch in [132]. Linear-quadratic fully constrained parabolic problems have been investigated in [116] by Tröltzsch and the author, including a discussion of boundary control problems motivated by the benchmark problem by Betts and Campbell, cf. [15]. Convergence and error estimates for parabolic problems governed by semilinear equations have to the author's knowledge only been studied in [115].

Finite element error estimates for regularized problems are also of interest. Elliptic problems with regularized pointwise state constraints have been discussed mainly by Hinze et. al., [73], and Rösch et. al. in [32]. The discretization for interior point methods in the presence of pointwise state constraints has been analyzed in [75], and the discretization of Moreau-Yosida regularization is discussed in [65].

In this thesis, we will aim at closing some of the remaining gaps in the theory of PDE-constrained optimal control problems subject to pointwise inequality constraints. The motivating question is the development of a priori-error estimates for the finite element discretization of state-constrained optimal control problems, subject to semilinear elliptic and parabolic heat equations. We will address different aspects. Throughout, we are interested in analyzing local solutions of nonconvex problems. After a short introduction into fundamental regularity theory for the underlying partial differential equations in Chapter 2 we will address three main topics:

- We will discuss an elliptic state constrained problem of semi-infinite type in Chapter 3. The comparably high regularity of solutions to semilinear elliptic state equations, the rather simple control-structure of only finitely many control parameters, and readily available discretization error estimates for uncontrolled equations in two space dimensions allows to focus entirely on the challenges of pointwise state constraints. Under some assumptions to be discussed, involving in particular an assumption on the active set that allows to characterize the structure of the Lagrange multipliers, we prove a convergence result for discrete optimal controls that reflects the convergence order of uncontrolled equations. The main ideas applied in this section have been published by Merino, Tröltzsch, and the author in [104] and [105] for linear-quadratic problems. Additionally, we have to take into account the difficulties posed by nonconvex problem formulations, such as convergence of local solutions, quadratic growth conditions, and facts as simple as not having a superposition principle for nonlinear state equations. We will conduct some numerical tests, but the discussion of efficient solution algorithms shall not be the purpose of this chapter. Therefore, we simply revert to available nonlinear programming solvers for the completely discretized problem formulations.

- In principle, we are also interested in finite element error estimates for parabolic state constrained problems, where we look at problems with distributed control functions that can vary arbitrarily in space and time. This problem class unites numerous difficulties, from more restrictive regularity results for parabolic equations, to the fact that in general the active sets are difficult to characterize, to the challenge of having to develop error estimates in space and time, depending on two essentially different discretization parameters. We will therefore consider a Lavrentiev regularized version of a parabolic control-and-state constrained optimal control problem. In Chapter 4, one key issue will be to provide a convergence result for Lavrentiev parameter tending to zero, as well as a regularization error estimate. Secondly, we dwell on the improved regularity properties of the regularized problem, and discuss properties of the regularized problem including second-order sufficient optimality conditions and a stability analysis of associated linear-quadratic problems, that in particular give rise to a local uniqueness result and may also be used for a convergence discussion of numerical solution algorithms. In contrast to the discussion of comparable elliptic problems, we have to take into account the fact that L^2-controls do not in general generate bounded states, which complicates the analysis in several points. Most of the results presented in Chapter 4 have been published by Tröltzsch and the author in [115]. See also [114] for a short overview.

- Then, in Chapter 5, we develop a priori error estimates for the finite element discretization of control-constrained optimal control problems governed by a semilinear parabolic state equation. While the results of Chapter 5 are interesting in their own right, we also motivate how under certain assumptions the model problem can be interpreted as a Lavrentiev-regularized version of a state-constrained model problem, cf. [73] for a similar elliptic setting, and in that sense also complements the study of state-constrained model problems. We give some results on regularization that have been proven and published for linear-quadratic problems by Tröltzsch and the author in [116]. The results on error estimates presented in this chapter have been published by Vexler and the author in [118] for a slightly different objective function. The extension to the more general problem formulation is straight-forward, but necessary in order to discuss certain Lavrentiev-regularized problems.

A brief overview about the first two items can also be found in [117].

Before we turn to the detailed analysis of the mentioned example problems, let us say a few words about the notation used. In the next chapter, we will agree on some general underlying assumptions and notation, but in the interest of readability, the notation will only be completely consistent chapter by chapter. In particular, the same name may be used for similar but not necessarily identical things. For instance, the control-to-state operator associated with the nonlinear state equations will always be denoted by G. Since the main chapters are self-contained, this will be no cause of confusion. The only overlap will appear between Chapters 4 and 5 when it comes to the properties of G, but it will be immediately apparent that the properties of G proven in Chapter 4 transfer to Chapter 5. Last, let us mention that the first section of each of the Chapters 3-5 contains a short introduction to the model problem and a summary of known results as a benefit for the reader, since most of these properties are used in the following analysis.

2. Fundamentals and model equations

Naturally, existence and regularity issues of the governing PDE play an important role for a thorough analysis of PDE-constrained optimal control problems. For that reason, we begin this thesis with a presentation of some functional analytic basics and known existence and regularity results for the PDEs under consideration. For further information we point out textbooks on functional analysis, e.g. [1], partial differential equations, e.g. [48], or optimal control, [144].

2.1. Notation

For the solution theory of both elliptic and parabolic partial differential equations we require certain regularity assumptions on the involved spatial domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$. Note first that the term *domain* automatically implies that Ω is an open and connected set. Throughout, we will work with convex Lipschitz domains, e.g. domains Ω whose boundary $\Gamma := \partial\Omega$ can locally be expressed as the graph of a Lipschitz continuous function. We refer to e.g. [113] for a more detailed characterization of domains and their boundaries. An important class are polygonally or polyhedrally bounded convex domains in $\mathbb{R}^2$ or $\mathbb{R}^3$.

We use standard notation for the Lebesgue spaces $L^p(\Omega)$, Sobolev spaces $W^{m,p}(\Omega)$, with the short notation $H^m(\Omega) := W^{m,2}(\Omega)$, as well as some spaces of continuous functions, including Hölder continuous functions. For time-dependent problems, we will make use of spaces of abstract functions, or vector valued functions, that are defined on a compact time interval and admit values in Banach spaces. We refer the reader to Appendix A.1 for a short overview, and to standard monographs such as [1] for a detailed discussion. Details can also be found in [144], where functional analytic fundamentals that are applied in optimal control are presented in a concise way.

2.2. Differentiability in function spaces

For each model problem under consideration, we will review first and second-order optimality conditions. This involves some kind of differentiation of what we will call a control-to-state mapping, as well as the objective function. We state precise definitions and conditions collected in [144].

Definition 2.2.1. *Let $\{U, \|\cdot\|_U\}, \{W, \|\cdot\|_W\}$ be normed linear spaces. A mapping $T\colon \mathcal{U} \subset U \to W$ is said to be Fréchet differentiable at $u \in \mathcal{U}$ if there exists an operator $T'(u) \in \mathcal{L}(U, W)$ and a mapping $r(u,\cdot)\colon U \to W$ with the following properties: for all $v \in U$ such that $u + v \in \mathcal{U}$, we have*

$$T(u + v) = T(u) + T'(u)v + r(u, v),$$

where r satisfies the condition

$$\frac{\|r(u, v)\|_W}{\|v\|_U} \to 0 \quad \text{as } \|v\|_U \to 0.$$

The operator $T'(u)$ is then called the Fréchet derivative of T at u. T is said to be continuously Fréchet differentiable if

$$\|u - v\|_U \to 0 \Rightarrow \|T'(u) - T'(v)\|_{\mathcal{L}(U,W)} \to 0$$

is satisfied.

For a continuous linear operator T, differentiability is easily obtained and $T'(u) = T$ for all u is observed. However, our model problems are governed by certain types of nonlinearities, both in the PDE itself and in the objective function. These nonlinearities depend on either or both the control u and the state y. The arising operators are called superposition operators, or Nemytskii operators. We give a short overview about their properties, and refer to [81] or again to [144] for more details.

Definition 2.2.2. *Let $E \subset \mathbb{R}^n$, $n \in \mathbb{N}$, be a bounded set, and let $\phi = \phi(e, \rho)\colon E \times \mathbb{R}^m \to \mathbb{R}$, $m \in \mathbb{N}$ be a function. The mapping Φ given by*

$$\Phi(\rho) = \phi(\cdot, \rho(\cdot)),$$

which assigns to a (possibly vector valued) function $\rho\colon E \to \mathbb{R}^m$ the function $\sigma\colon E \to \mathbb{R}$, $\sigma(e) = \phi(e, \rho(e))$, is called a Nemytskii operator.

We will assume the following Carathéodory-type, boundedness, and Lipschitz continuity conditions of order two for the defining nonlinearity ϕ:

Assumption 2.2.3. *The function $\phi = \phi(e, \rho)\colon E \times \mathbb{R}^m \to \mathbb{R}$ is measurable with respect to $e \in E$ for all fixed $\rho \in \mathbb{R}^m$, and twice continuously differentiable with respect to ρ for almost all $e \in E$. For $\rho = 0$, it is bounded of order two with respect to e by a constant $K > 0$, i.e. ϕ is assumed to satisfy*

$$|\phi(\cdot, 0)| + |\phi'(\cdot, 0)| + |\phi''(\cdot, 0)| \le K, \tag{2.2.1}$$

where ϕ' and ϕ'' denote the first and second order derivative of ϕ with respect to ρ. Also, ϕ and its derivatives with respect to ρ up to order two are uniformly Lipschitz on bounded sets, i.e. for all $S > 0$ there exists $L_S > 0$ such that ϕ satisfies

$$|\phi''(\cdot, \rho_1) - \phi''(\cdot, \rho_2)| \le L_S |\rho_1 - \rho_2| \tag{2.2.2}$$

for all $\rho_i \in \mathbb{R}^m$ with $|\rho_i| \le S$.

In the last assumption and throughout the following,

$$|\cdot|, \quad \text{and} \quad \langle \cdot, \cdot \rangle$$

denote the norm and inner product in $\mathbb{R}^m$, respectively. We will use this for instance for elements $x \in \Omega$ as well as $u \in \mathbb{R}^m$. For brevity of notation, we will also use $|\cdot|$ as a matrix norm for e.g. Φ''. Later on, we will replace E by a spatial domain Ω or a space-time domain Q, with elements $e = x$, or $e = (t, x)$, respectively. Moreover, $\rho \in \mathbb{R}^m$ will be replaced by y or (y, u). Under these assumptions, it is known that the Nemytskii operator Φ is continuous in $L^\infty(E)$, and locally Lipschitz continuous with respect to the $L^r(E)$ norm for every $1 \le r \le \infty$. Moreover, Φ is continuously Fréchet differentiable in $L^\infty(E)$, with derivative

$$(\Phi'(\rho)v)(e) = \phi'(e, \rho(e))v(e)$$

for all $v \in L^\infty(E)$ and almost all $e \in E$, where ϕ' denotes the derivative of ϕ with respect to ρ. For consideration of other L^p-spaces, note that $\Phi(\rho)$ maps $L^p(E)^m$ into $L^q(E)$ for $1 \le q \le p < \infty$ if and only if there are functions $\alpha \in L^q(E)$ and $\beta \in L^\infty(E)$ such that the growth condition

$$|\phi(e, \rho)| \le \alpha(e) + \beta(e)|\rho|^{\frac{p}{q}}$$

is satisfied. Then, Φ is automatically continuous for $q < \infty$. It is differentiable from $L^p(E)$ into $L^q(E)$, if the Nemytskii operator generated by $\phi'(e,\rho)$ maps $L^p(E)$ into $L^r(E)$ with

$$r = \frac{pq}{p-q}.$$

This will turn out useful when discussing different types of objective functionals, and second order sufficient conditions. For more details, we refer to [144, Chapter 4], and the references mentioned therein.

2.3. Basic assumptions

Due to the diverse nature of the model problems considered we will have to specify some assumptions and convenient notation chapter by chapter. There are, however, some fundamental assumptions that can be formulated for all what follows.

Assumption 2.3.1. *Let $\Omega \subset \mathbb{R}^n$, $1 \leq n \leq 3$, be a nonempty convex Lipschitz domain with boundary $\Gamma := \partial\Omega$ if $n > 1$, or a bounded interval for $n = 1$. For a time dependent problem, we denote by $I := (0,T)$ a time interval with fixed final time $T > 0$, and $Q := I \times \Omega$ denotes the space-time-domain with boundary $\Sigma := I \times \Gamma$.*

As a rule of thumb, our results on discretization will be confined to two space dimensions due either to error estimates for uncontrolled equations that are only readily availabe in 2D, or to the regularity of the discrete solutions themselves.

We will basically consider two types of state equations, the semilinear elliptic boundary value problem

$$\begin{aligned} \mathcal{A}y + d(x,y,u) &= 0 \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Gamma, \end{aligned} \tag{2.3.1}$$

where $u \in \mathbb{R}^m$ is a vector of control parameters, and its parabolic analogue

$$\begin{aligned} \partial_t y + \mathcal{A}y + d(t,x,y) &= u \quad \text{in } Q, \\ y(0,\cdot) &= y_0 \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Sigma, \end{aligned} \tag{2.3.2}$$

for space-time-dependent control functions u with regularity yet to be discussed. Note that in (2.3.1), where u is a vector of real numbers, we allow d to depend on u. Both PDEs are governed by a uniformly elliptic and symmetric differential operator $\mathcal{A}$, defined by

$$\mathcal{A}y(x) = -\sum_{i,j=1}^{n} \partial_j(a_{ij}(x)\partial_i y(x)) \tag{2.3.3}$$

with coefficients $a_{ij} \in \mathcal{C}^{0,1+\nu}(\Omega)$, $0 < \nu < 1$, that satisfy

$$m_0|\xi|^2 \leq \sum_{i,j=1}^{n} a_{ij}(x)\xi_i\xi_j \quad \forall \xi \in \mathbb{R}^n, \quad \text{a.e. in } \Omega \tag{2.3.4}$$

for some $m_0 > 0$. Moreover, $\partial_t y$ denotes the partial derivative of y with respect to time. We point out that the choice of homogeneous Dirichlet boundary conditions is mostly due to a unified presentation of results. The theory can be adapted to homogeneous variational boundary conditions as long as uniqueness of solutions is guaranteed in the elliptic setting, but then Poincaré's inequality is not applicable and some proofs become more involved. The results on regularization presented in Chapter 4 transfer also to inhomogeneous Robin-type boundary conditions with sufficiently regular boundary data.

Later on, we will refer to u as the control and to y as the state. For now, let us discuss the existence and regularity of solutions to (2.3.1) and (2.3.2) for fixed u from a respective appropriate control space.

2.4. Elliptic model equations

We first discuss existence and regularity of solutions of the uncontrolled elliptic state equation (2.3.1) as well as linear equations as they will later on appear as adjoint equations or linearized state equations. Therefore, we define the control space

$$U := \mathbb{R}^m,$$

and consider a fixed $u \in \mathbb{R}^m$ throughout this section. The canonical state space will be the space $H_0^1(\Omega)$, and we introduce the short notation

$$V := H_0^1(\Omega).$$

However, H^1-regularity will not be sufficient for our purposes, since we will consider pointwise state constraints in our model problem formulation. We will therefore provide higher regularity results. Throughout, we use the following abbreviations for the inner product and norm on $L^2(\Omega)$:

$$(\cdot,\cdot) := (\cdot,\cdot)_{L^2(\Omega)}, \quad \|\cdot\| := \|\cdot\|_{L^2(\Omega)}.$$

Moreover, for the norms on $L^p(\Omega)$ with $1 \leq p < \infty$ and especially for the $L^\infty(\Omega)$-norm we use the abbreviations

$$\|\cdot\|_p := \|\cdot\|_{L^p(\Omega)}, \quad \|\cdot\|_\infty := \|\cdot\|_{L^\infty(\Omega)}.$$

In addition, for a real number $\rho > 0$ and an element $x \in \Omega$ we define $B_\rho(x)$ to be the open ball in $\mathbb{R}^n$ around x with radius ρ. If instead of $x \in \Omega$ a control vector $u \in \mathbb{R}^m$ is given, then $B_\rho(u)$ denotes the open ball in $\mathbb{R}^m$ around u.

Let first the linear elliptic boundary value problem

$$\begin{aligned} \mathcal{A}y + d_0 y &= f \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Gamma \end{aligned} \tag{2.4.1}$$

with data $f \in L^2(\Omega)$ and $d_0 \in L^\infty(\Omega)$, $d_0 \geq 0$ be given. It is reasonable to look for solutions in weak sense, since the homogeneous Dirichlet boundary conditions do not require i.e. very weak solution concepts. For the convenient discussion of elliptic PDEs, we define the bilinear form

$$\mathbf{a}\colon V \times V \to \mathbb{R}, \quad \mathbf{a}(v,w) = \sum_{i,j=1}^{n} (a_{ij}\partial_i v, \partial_j w).$$

Note that we do not include the term associated with d_0 in the bilinear form $\mathbf{a}$, since this will leave more flexibility to use $\mathbf{a}$ also for e.g. linearized equations when combining it with an appropriate additional term.

Definition 2.4.1. *A function $y \in H_0^1(\Omega)$ is said to be a weak solution of the Dirichlet problem* (2.4.1) *if it fulfills*

$$\mathbf{a}(y,\varphi) + (d_0 y, \varphi) = (f,\varphi) \quad \forall\, \varphi \in H_0^1(\Omega).$$

Proposition 2.4.2. *Let Assumption 2.3.1 be satisfied and let Ω_1 be a subdomain of Ω with $\bar{\Omega}_1 \subset \Omega$. For every right-hand side $f \in L^2(\Omega)$, the linear boundary value problem* (2.4.1) *admits a unique weak solution $y \in H_0^1(\Omega)$. Moreover, the solution exhibits the improved regularity*

$$y \in H_0^1(\Omega) \cap H^2(\Omega) \cap \mathcal{C}(\bar{\Omega}),$$

and there exists a constant $C > 0$ not depending on f, such that the estimate

$$\|y\| + \|\nabla y\| + \|\nabla^2 y\| + \|y\|_{\mathcal{C}(\bar{\Omega})} \leq C\,\|f\|$$

is satisfied. For $f \in C^{0,\nu}$ with a real number $\nu \in (0,1)$, we even have

$$y \in \mathcal{C}^{2,\nu}(\Omega) \cap W^{2,\infty}(\Omega_1).$$

Proof. The existence and uniqueness of $y \in H_0^1(\Omega)$ is a standard result, following from the Lax-Milgram lemma after applying Poincaré's inequality. Since we assume Ω to be a bounded, convex Lipschitz domain, Theorem 3.2.1.2 in [62] guarantees a solution $y \in H^2(\Omega)$. By a Sobolev embedding theorem, cf. e.g. [1], we obtain continuity of the solution by $n \leq 3$. The $\mathcal{C}^{2,\nu}$-regularity follows directly from [51, Theorem 6.13] noting that convex polygonal domains fulfill a so-called exterior sphere condition, i.e. for each $x^* \in \Gamma$, there exists a sphere $B = B_\rho(y)$ satisfying $\bar{B} \cap \bar{\Omega} = x^*$. From $\mathcal{C}^{2,\nu}$-regularity on $\bar{\Omega}_1 \subset \Omega$ the desired $W^{2,\infty}$-regularity on Ω_1 is immediately obtained. □

We also point out the results in [2] and [18], where boundedness and continuity of solutions have been discussed for Neumann boundary value problems in n-dimensional domains for right-hand sides $f \in L^p(\Omega)$, $p > n/2$. Applying results by Griepentrog, cf. [56], continuity is also obtained under weaker regularity assumptions on the coefficients of the differential operator.

Finally, before turning to the nonlinear equation, we present an existence and regularity result for linear elliptic Dirichlet problems with an element $\mu \in \mathcal{C}(\Omega)^*$ in the right-hand side. In the sequel, we identify $\mathcal{C}(\Omega)^*$ with the set of regular Borel measures $\mathcal{M}(\Omega)$, cf. [3]. We will encounter this type of equation when deriving first order optimality conditions for the semi-infinite optimal control problem in Chapter 3. Elliptic problems with mixed boundary conditions and measures in the right-hand side as well as in the boundary term have been discussed by Casas, cf. [18], and Alibert and Raymond, [2]. A problem with $\mathcal{A} = -\Delta$ and homogeneous Dirichlet boundary conditions on smooth domains is discussed in detail in Section 5.3.5 of [16]. We consider the boundary value problem

$$\begin{aligned} \mathcal{A}p + d_0 p &= \mu \quad \text{in } \Omega, \\ p &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{2.4.2}$$

We follow Casas, [17], and call $p \in L^2(\Omega)$ a solution of (2.4.2), if and only if

$$\int_\Omega (-\sum_{i,j=1}^n \partial_j(a_{ij}(x)\partial_i\varphi(x)) + d_0\varphi)(x)p(x)dx = \int_\Omega \varphi(x)d\mu(x), \quad \forall\, \varphi \in H^2(\Omega) \cap H_0^1(\Omega) \tag{2.4.3}$$

is satisfied.

Proposition 2.4.3. *Let Assumptions 2.3.1 be true and let $d_0 \in L^\infty(\Omega)$ be a given nonnegative function. For every $\mu \in \mathcal{C}_0(\Omega)^*$, the boundary value problem* (2.4.2) *admits a unique solution $p \in W_0^{1,s}(\Omega)$ for all $s \in [1, n/(n-1))$, and there exists a constant $C = C(s) > 0$ such that*

$$\|p\|_{W_0^{1,s}(\Omega)} \leq C\|\mu\|_{\mathcal{C}(\Omega)^*}.$$

Proof. This follows from [17]. $\square$

The homogeneous Dirichlet boundary conditions will later be motivated by prescribing state constraints only in the interior of Ω. Now, we consider the nonlinear state equation (2.3.1), where we rely on the following precise setting of Assumption 2.2.3.

Assumption 2.4.4. *In (2.3.1), the function $d = d(x, y, u)\colon \Omega \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ is Hölder continuous with respect to $x \in \Omega$ for all fixed $(y, u) \in \mathbb{R} \times \mathbb{R}^m$, and twice continuously differentiable with respect to (y, u) for all $x \in \Omega$. For $y = u = 0$, it is bounded of order two with respect to x by a constant $K > 0$, i.e. d is assumed to satisfy*

$$|d(\cdot, 0, 0)| + |d'(\cdot, 0, 0)| + |d''(\cdot, 0, 0)| \le K, \tag{2.4.4}$$

where d' and d'' denote the first and second order derivative of d with respect to (y, u). Also, d and its derivatives with respect to (y, u) up to order two are uniformly Lipschitz on bounded sets, i.e. for all $S > 0$ there exists $L_S > 0$ such that d satisfies

$$|d''(\cdot, y_1, u_1) - d''(\cdot, y_2, u_2)| \le L_S(|y_1 - y_2| + |u_1 - u_2|) \tag{2.4.5}$$

for all $y_i, u_i \in \mathbb{R}$ with $|y_i|, |u_i| \le S$. Moreover, d is monotone with respect to y, i.e. suppose that $\partial_y d(\cdot, y, u) > 0$.

Definition 2.4.5. *For a given control vector $u \in \mathbb{R}^m$, a function $y \in H_0^1(\Omega)$ is said to be a weak solution of the Dirichlet problem (2.3.1) if it fulfills*

$$\mathbf{a}(y, \varphi) + (d_0 y, \varphi) + (d(\cdot, y, u), \varphi) = 0 \quad \forall\, \varphi \in H_0^1(\Omega).$$

We obtain an existence and regularity result similar to the linear setting.

Theorem 2.4.6. *Let Assumptions 2.3.1 and 2.4.4 hold and let Ω_1 be a subdomain of Ω with $\bar{\Omega}_1 \subset \Omega$. For each control $u \in \mathbb{R}^m$, there exists a unique weak solution*

$$y \in H_0^1(\Omega) \cap H^2(\Omega) \cap \mathcal{C}(\bar{\Omega}) \cap \mathcal{C}^{2,\nu}(\Omega) \cap W^{2,\infty}(\Omega_1)$$

of (2.3.1). We obtain the existence of a positive constant C, such that

$$\|y\| + \|\nabla y\| + \|\nabla^2 y\| + \|y\|_{\mathcal{C}(\bar{\Omega})} + \|y\|_{W^{2,\infty}(\Omega_1)} \le C(|u| + 1).$$

Proof. The existence of a unique solution $y \in H^1(\Omega) \cap L^\infty(\Omega)$ follows from similar arguments as in [144], where a semilinear problem with Neumann boundary conditions has been discussed. As a preliminary step, assume that the nonlinearity d is globally bounded, and apply the steps of [144, Theorem 4.5], which are based on a method by Stampacchia, cf. [80], to

$$\begin{aligned} \mathcal{A}y + d(\cdot, y, u) - d(\cdot, 0, u) &= -d(\cdot, 0, u) && \text{in } \Omega, \\ y &= 0 && \text{on } \Gamma. \end{aligned}$$

Here, the monotonicity of d with respect to y can be used when testing the weak formulation with y. Due to the Lipschitz properties of d, it is clear that eventually the right-hand side can be estimated by

$$\|d(\cdot, 0, u)\| \le \|d(\cdot, 0, u) - d(\cdot, 0, 0)\| + \|d(\cdot, 0, 0)\| \le c(|u| + 1).$$

By applying a cut-off argument from Casas, [18], to the possibly unbounded nonlinearity d one obtains the existence of a solution $y \in H_0^1(\Omega) \cap L^\infty(\Omega)$ satisfying the estimate

$$\|y\|_{H_0^1(\Omega)} + \|y\|_{C(\bar{\Omega})} \leq C(|u| + 1),$$

taking advantage of the monotonicity and Lipschitz continuity of d. The remaining estimates follow by applying Proposition 2.4.2 to the linear equation

$$\begin{aligned} \mathcal{A}y &= -d(\cdot, y, u) && \text{in } \Omega, \\ y &= 0 && \text{on } \Gamma, \end{aligned}$$

since under Assumption 2.4.4 we have $d(\cdot, y, u) \in \mathcal{C}^{0,\nu}(\Omega)$. We refer also to the proof of Lemma 3 in [106]. □

2.5. Parabolic model equations

We will now build the basis for the further analysis of the parabolic model problems. Throughout the following, it will be helpful to make use of a short notation for inner products and norms on the spaces $L^2(Q)$ and $L^\infty(Q)$:

$$(v, w)_I \;:=\; (v, w)_{L^2(Q)}, \quad \|v\|_I \;:=\; \|v\|_{L^2(Q)}, \quad \|v\|_{\infty,\infty} \;:=\; \|v\|_{L^\infty(Q)}. \tag{2.5.1}$$

The subscript I is motivated by the notation for the time interval I, and will prove to be more legible than e.g. a subscript Q. In some estimates, we will also need the norm

$$\|v\|_{\infty,2} := \|v\|_{L^\infty(I, L^2(\Omega))}.$$

We still use the short notation

$$V := H_0^1(\Omega),$$

and note that by identifying the space $H := L^2(Q)$ with its dual space H^* we have a Gelfand triple

$$V \hookrightarrow H \cong H^* \hookrightarrow V^*,$$

where the embeddings are dense, compact, and continuous. Moreover, we define

$$Y := \mathcal{W}(0, T) = \{y \,|\, y \in L^2(I, V) \text{ and } \partial_t y \in L^2(I, V^*)\}.$$

For future reference, we point out that the embedding $\mathcal{W}(0, T) \hookrightarrow L^2(Q)$ is compact, cf. for instance [89]. Then, let the parabolic initial-boundary value problem

$$\begin{aligned} \partial_t y + \mathcal{A}y + d_0 y &= f && \text{in } Q, \\ y(0, \cdot) &= y_0 && \text{in } \Omega, \\ y &= 0 && \text{on } \Sigma \end{aligned} \tag{2.5.2}$$

be given with data $f \in L^2(Q)$, $y_0 \in V$ and $d_0 \in L^\infty(Q)$.

Definition 2.5.1. *A function $y \in \mathcal{W}(0, T)$ is called a weak solution of (2.5.2) if it fulfills*

$$\int_0^T \langle \partial_t y, \varphi \rangle_{V^*,V}\, dt + \iint_Q \sum_{i,j=1}^n a_{ij} \partial_i y \partial_j \varphi \; dxdt, + \iint_Q d_0 y \varphi \, dxdt = \iint_Q f\varphi \; dxdt \quad \forall\, \varphi \in L^2(I, V)$$
$$y(0, \cdot) = y_0 \quad \textit{in } \Omega,$$

where $\langle \cdot, \cdot, \rangle_{V^*,V}$ denotes the duality pairing between V^* and V.

We will provide an existence result for solutions of the linear equation in the state space Y. We will also provide higher regularity of the solution depending on the given data. On the one hand, we are interested in continuity, to eventually derive a continuity result for the nonlinear state equation which we will need to discuss state constraints appropriately. On the other hand, we will need a higher regularity result for the derivative with respect to time, as well as an estimate for the second order spatial derivative in order to derive discretization error estimates.

Proposition 2.5.2. *Let Assumption 2.3.1 hold and let $d_0 \in L^\infty(Q)$ be given. For every function $f \in L^2(Q)$ and initial state $y_0 \in L^2(\Omega)$, the initial boundary value problem* (2.5.2) *admits a unique solution $y \in \mathcal{W}(0,T)$. In addition, the estimate*

$$\|y\|_I + \|y\|_{\infty,2} + \|y\|_{\mathcal{W}(0,T)} \le C\left(\|f\|_I + \|y_0\|\right)$$

is fulfilled for a constant $C > 0$. If the initial state admits the regularity $y_0 \in V$, y exhibits the improved regularity

$$y \in L^2(I, H^2(\Omega) \cap V) \cap L^\infty(I,V) \cap H^1(I, L^2(\Omega)),$$

and the estimate

$$\|\nabla y\|_{\infty,2} + \|\nabla^2 y\|_I + \|\partial_t y\|_I \le C\left(\|f\|_I + \|y_0\|_V\right)$$

is satisfied. For $f \in L^p(\Omega)$, $p > n/2+1$ and $y_0 \in \mathcal{C}(\bar\Omega)$, y belongs to $\mathcal{C}(\bar Q)$, and the estimate

$$\|y\|_{\mathcal{C}(\bar Q)} \le C(p)(\|f\|_{L^p(Q)} + \|y_0\|_{\mathcal{C}(\bar\Omega)})$$

is satisfied.

Proof. For the existence of a unique solution $y \in \mathcal{W}(0,T)$ we refer to [86], or [144] for homogeneous variational boundary conditions. The main idea is to first consider a weak formulation of the initial boundary value problem with test functions $\varphi \in H^1(I, L^2(\Omega)) \cap L^2(I,V)$, that satisfy $\varphi(\cdot,T) = 0$. The existence of a unique weak solution $y \in L^2(I,V)$ is obtained by a Galerkin approximation and is proven to be a function from $\mathcal{W}(0,T)$. Cf. also [90] or [151]. The continuity result is proven in [144] for Robin-type boundary conditions based on results on maximal parabolic regularity for problems with zero initial state from [57, 58], which also hold for homogeneous Dirichlet boundary conditions. Finally, the regularity

$$y \in L^2(I, H^2(\Omega) \cap V) \cap L^\infty(I,V) \cap H^1(I, L^2(\Omega))$$

is proven in [48]. □

Boundedness in the norm $\|\cdot\|_{\infty,\infty}$ is obviously also obtained. For that result, we require the initial state y_0 to be bounded, only. Continuity of solutions was also discussed by Casas in [19], or by Raymond and Zidani, cf. [125].

The semilinear equation (2.3.2) is discussed next, under assumptions on the nonlinearity d that fit into the setting of Assumption 2.2.3.

Assumption 2.5.3. *In* (2.3.2), *the function $d = d(t,x,y)\colon Q \times \mathbb{R} \to \mathbb{R}$ is measurable with respect to $(t,x) \in Q$ for all fixed $y \in \mathbb{R}$, and twice continuously differentiable with respect to y for almost all $(t,x) \in Q$. For $y = 0$, it is bounded of order two with respect to (t,x) by a constant $K > 0$, i.e. d is assumed to satisfy*

$$|d(\cdot,0)| + |d'(\cdot,0)| + |d''(\cdot,0)| \le K, \tag{2.5.3}$$

where d' and d'' denote the first and second order derivative of d with respect to y. Also, d and its derivatives with respect to y up to order two are uniformly Lipschitz on bounded sets, i.e. for all $S > 0$ there exists $L_S > 0$ such that d satisfies

$$|d''(\cdot, y_1) - d''(\cdot, y_2)| \leq L_S|y_1 - y_2| \tag{2.5.4}$$

for all $y_i \in \mathbb{R}$ with $|y_i| \leq S$. Last, suppose that $\partial_y d(\cdot, y) > 0$.

A weak formulation is given in the following definition.

Definition 2.5.4. *A function $y \in \mathcal{W}(0,T)$ is a weak solution of* (2.3.2) *if it fulfills*

$$\int_0^T \langle \partial_t y, \varphi \rangle_{V^*,V}\, dt + \iint_Q \sum_{i,j=1}^n a_{ij}\partial_i y \partial_j \varphi\, dxdt, + \iint_Q d(\cdot, y)\varphi\, dxdt = \iint_Q f\varphi\, dxdt \quad \forall \varphi \in L^2(I, V)$$
$$y(0, \cdot) = y_0 \quad in\ \Omega.$$

If data in L^2-spaces is given, the following theorem provides solvability of the nonlinear initial-boundary value problem in useful spaces, yet relying on more restrictive assumptions on the nonlinearity.

Theorem 2.5.5. *Let Assumption 4.2.1 hold with global boundedness and Lipschitz properties of d, i.e. the properties formulated in Assumption 2.2.3 hold for all $y \in \mathbb{R}$. For every function $f \in L^2(Q)$, and initial state $y_0 \in L^2(\Omega)$, the initial boundary value problem* (5.2.1b) *admits a unique solution $y \in \mathcal{W}(0,T)$, and the estimate*

$$\|y\|_{\mathcal{W}(0,T)} \leq C(\|f\|_I + \|y_0\| + 1)$$

holds for a constant $C > 0$. In addition, if $y_0 \in V$, the solution exhibits the improved regularity

$$y \in L^2(I, H^2(\Omega) \cap V) \cap L^\infty(I, V) \cap H^1(I, L^2(\Omega)) \hookrightarrow \mathcal{C}(\bar{I}, V)$$

and the estimate

$$\|y\|_{L^\infty(I,V)} + \|y\|_{L^2(I,H^2(\Omega))} + \|\partial_t y\|_I \leq C(\|f\|_I + \|y_0\|_V + 1)$$

is fulfilled for a constant $c > 0$.

Proof. Existence of such a solution is obtained by a Galerkin method, taking into account appropriate convergence results to cope with the nonlinearity. For the complete proof, we refer e.g. to [144] for an analogous problem with Robin-type boundary condition. An essential ingredient is the monotonicity of d. Adding the term $-d(\cdot, 0)$ to both sides of (2.3.2) and noting that

$$(d(\cdot, y) - d(\cdot, 0), y)_I \geq 0$$

eventually yields the estimate

$$\|y\|_{\infty,2} + \|y\|_{\mathcal{W}(0,T)} \leq \|f\|_I + \|d(\cdot, 0)\|_I,$$

which explains the term $+1$ in the a priori estimate. The higher regularity of y then follows from applying Proposition 2.5.2 to the linear equation

$$\begin{aligned} \partial_t y + \mathcal{A}y &= f - d(\cdot, y) && \text{in } Q \\ y(0, \cdot) &= y_0 && \text{in } \Omega \\ y &= 0 && \text{on } \Sigma, \end{aligned}$$

estimating

$$\|d(\cdot, y)\|_{L^p(Q)} \leq \|d(\cdot, y) - d(\cdot, 0)\|_{L^p(Q)} + \|d(\cdot, 0)\|_{L^p(Q)} \leq L\|y\|_{L^p(Q)} + \|d(\cdot, 0)\|_{L^p(Q)}$$

by the Lipschitz continuity of d, if d is globally Lipschitz continuous. □

For higher regularity of the given data we obtain the following existence and regularity result under weaker assumptions on the nonlinearity:

Theorem 2.5.6. *Let Assumptions 2.3.1 and 2.4.4 hold. For every function $f \in L^p(Q)$, $p > n/2 + 1$, and initial state $y_0 \in \mathcal{C}_0(\bar{\Omega})$, the initial boundary value problem* (2.5.2) *admits a unique solution $y \in \mathcal{W}(0,T) \cap \mathcal{C}(\bar{Q})$, and the estimates*

$$\begin{aligned} \|y\|_{\mathcal{C}(\bar{Q})} &\leq C\,(\|f\|_{L^p(Q)} + \|y_0\|_{\mathcal{C}(\bar{\Omega})} + 1), \\ \|y\|_I + \|y\|_{\infty,2} + \|y\|_{\mathcal{W}(0,T)} &\leq C\,(\|f\|_I + \|y_0\| + 1) \end{aligned}$$

hold for a constant $C > 0$. In addition, if $y_0 \in V$, the solution exhibits the improved regularity

$$y \in L^2(I, H^2(\Omega) \cap V) \cap L^\infty(I, V) \cap H^1(I, L^2(\Omega)) \hookrightarrow \mathcal{C}(\bar{I}, V)$$

and the estimate

$$\|\nabla y\|_I + \|\nabla y\|_{\infty,2} + \|\nabla^2 y\|_I + \|\partial_t y\|_I \leq C\,(\|f\|_I + \|y_0\|_V + 1)$$

is fulfilled for a constant $C > 0$.

Proof. Existence and continuity of a solution y is shown in [144] for mixed boundary conditions. In a first step, the existence of a solution in $\mathcal{W}(0,T)$ under stronger assumptions on the nonlinearity d, i.e. global boundedness and Lipschitz continuity, is shown for data in L^2-spaces, cf. Theorem 2.5.5. A cut-off argument applied to a possibly unbounded nonlinearity d shows boundedness of the solutions if the data is regular enough. We refer to [144] and the references therein for details. The higher regularity of y, including continuity, follows again by applying Proposition 2.5.2 to the linear equation

$$\begin{aligned} \partial_t y + \mathcal{A}y &= f - d(\cdot, y) && \text{in } Q \\ y(0,\cdot) &= y_0 && \text{in } \Omega \\ y &= 0 && \text{on } \Sigma, \end{aligned}$$

estimating

$$\|d(\cdot,y)\|_{L^p(Q)} \leq \|d(\cdot,y) - d(\cdot,0)\|_{L^p(Q)} + \|d(\cdot,0)\|_{L^p(Q)} \leq L\|y\|_{L^p(Q)} + \|d(\cdot,0)\|_{L^p(Q)}.$$

Here, local Lipschitz continuity of d is sufficient since y is already known to be bounded. □

Again, if $y_0 \in L^\infty(Q)$, we obtain boundedness of y in the $L^\infty(Q)$-norm. Note that for right-hand sides $f \in L^2(Q)$ Theorem 2.5.6 is only applicable for one-dimensional spatial domains.

We devote a short paragraph to the final-boundary value problem

$$\begin{aligned} -\partial_t p + \mathcal{A}p + d_0 p &= f && \text{in } Q, \\ p(T,\cdot) &= p_T && \text{in } \Omega, \\ p &= 0 && \text{on } \Sigma. \end{aligned} \tag{2.5.5}$$

We will see later that this type of equation appears in the optimality system of the control problem. The solution theory for this type of problem does not differ from initial-boundary value problems. This becomes clear when considering the time-transformation $t = T - \tau$, $\tau \in [0,T]$, which transfers (2.5.5) into an initial boundary value problem with transformed right-hand side.

Proposition 2.5.7. *For every $f, d_0 \in L^\infty(Q)$ and $p_T \in V \cap L^\infty(\Omega)$ there exists a unique solution $p \in Y \cap L^\infty(Q)$ of the equation* (2.5.5)*. Moreover, the solution exhibits the improved regularity*

$$p \in L^2(I, H^2(\Omega) \cap V) \cap H^1(I, L^2(\Omega)) \cap L^\infty(Q) \hookrightarrow C(\bar{I}, V)$$

and the estimates

$$\|p\|_{\infty,\infty} \leq C\left(\|g\|_{\infty,\infty} + \|p_T\|_\infty\right),$$
$$\|\partial_t p\|_I + \|p\|_I + \|\nabla p\|_I + \|\nabla^2 p\|_I + \|p\|_{\infty,2} \leq C\left(\|f\|_I + \|p_T\|\right)$$

are satisfied for a constant $C > 0$.

Proof. By the time transformation $t = T - \tau$, $\tau \in [0, T]$, this is an immediate consequence of Proposition 2.5.2. □

In state-constrained problems a measure will appear on the right-hand side:

$$\begin{aligned} -\partial_t p + \mathcal{A}p + d_0 p &= \mu_Q && \text{in } Q, \\ p(T, \cdot) &= \mu_T && \text{in } \Omega, \\ p &= 0 && \text{on } \Sigma. \end{aligned} \tag{2.5.6}$$

We proceed as in [9].

Definition 2.5.8. *A function $p \in L^1(I, W_0^{1,1}(\Omega))$ is called a weak solution of* (2.5.6)*, if*

$$\iint_Q \partial_t \varphi p \, dxdt + \iint_Q \sum_{i,j=1}^n a_{ij} \partial_i \varphi \partial_j p \, dxdt, + \iint_Q d_0 \varphi p \, dxdt = \iint_Q \varphi \, d\mu_Q(t,x) + \int_\Omega \varphi(T, \cdot) \, d\mu_T(T, x) \tag{2.5.7}$$

is satisfied for all $\varphi \in \mathcal{C}^1(\bar{Q}) \cap \mathcal{C}_0(Q \cup \Omega_T)$, with $\Omega_T := \{T\} \times \Omega$.

The homogeneous Dirichlet boundary conditions will again be motivated by the fact that the state constraints are not active on Σ by assumption. The following existence result can be found in [19] or [9]:

Proposition 2.5.9. *Let Assumption 2.3.1 be true and let $d_0 \in L^\infty(Q)$ be a given function. For every $\mu \in \mathcal{C}_0(\bar{Q})^*$,* (2.5.6) *admits a unique solution $p \in L^\tau(I, W_0^{1,s}(\Omega))$ for all $\tau, s \in [1, 2)$ with $(2/\tau) + (n/s) > n + 1$.*

In conclusion of this chapter, let us point out that the general Assumptions 2.3.1 and 2.4.4 or 2.5.3, respectively, are valid throughout the remainder of this thesis without explicit notice.

3. An elliptic control problem of semi-infinite type

We start the analysis of PDE-constrained optimal control problems subject to pointwise state constraints with the discussion of problems with finitely many control parameters. This is not an unusual situation in practice, since finitely many control parameters are easier to implement than control functions that can vary arbitrarily in space and time. In [36], some typical practical applications have been mentioned, like the finitely many spray nozzles in steel cooling, [43], or finitely many microwave antennas in local hyperthermia, cf. [41]. Of course, in some applications these finitely many controls will depend on time. In this chapter, however, our main goal will be the derivation of a priori error estimates for the finite element discretization of stationary elliptic semi-infinite problems. Before we begin the detailed analysis, let us point out that we extend results published by Merino, Tröltzsch, and the author for linear quadratic problems in [104] and [105].

3.1. The optimal control problem and its analysis

In this first section, we present the problem under consideration, agree on the notation used throughout this chapter, and discuss the problem on the continuous, i.e. undiscretized level. We will address questions concerned with the existence of solutions and collect known results on first and second order optimality conditions.

3.1.1. Problem formulation, assumptions, and notation

The problem formulation for a control vector $u \in \mathbb{R}^m$ and an associated state y reads:

$$\text{Minimize } J(y,u) := \int_\Omega \Psi(x,y,u)\; dx \tag{3.1.1a}$$

subject to the semilinear elliptic PDE constraint

$$\begin{aligned} \mathcal{A}y + d(\cdot,y,u) &= 0 \quad \text{in } \Omega,\\ y &= 0 \quad \text{on } \Gamma, \end{aligned} \tag{3.1.1b}$$

the pointwise state constraints

$$y(x) \le b \quad \forall\, x \in K, \tag{3.1.1c}$$

in a compact interior subset K of Ω with nonempty interior to be characterized below, and the control bounds

$$u_a \le u \le u_b, \tag{3.1.1d}$$

which are to be understood componentwise in all what follows. More precisely, we will analyse a setting where the conditions stated below hold.

Assumption 3.1.1.

(A.1) *Let $\Omega \subset \mathbb{R}^2$ be a convex and polygonally bounded domain.*

(A.2) *The differential operator $\mathcal{A}$ satisfies* (2.3.3) *and* (2.3.4) *on page 19 with $n = 2$. Moreover, let the nonlinearity d fulfill the conditions stated in Assumption 2.4.4 on page 22.*

(A.3) *The control constraints are defined by two vectors $u_a, u_b \in \mathbb{R}^m$ with $u_a < u_b$, where the inequalities are to be understood componentwise. The state constraint is prescribed in a compact subset K of Ω with nonempty interior that satisfies the following condition for subdomains Ω_0, Ω_1 of Ω:*

$$K \subset \Omega_0, \quad \bar{\Omega}_0 \subset \Omega_1, \quad \bar{\Omega}_1 \subset \Omega.$$

(A.4) *The function $\psi = \psi(x, y, u) \colon \Omega \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ is measurable and Hölder continuous with respect to $x \in \Omega$ for all fixed $(y, u) \in \mathbb{R} \times \mathbb{R}^m$, and twice continuously differentiable with respect to (y, u) for all $x \in \Omega$. Its first and second order derivative will be denoted by Ψ' and Ψ'', respectively. Moreover, for $(y, u) = (0, 0)$, the function ψ and its derivatives up to order two are bounded and Lipschitz continuous on bounded sets with respect to (y, u), i.e. they fulfill conditions* (2.4.4) *and* (2.4.5) *accordingly.*

Note that the choice of Ω_0, Ω_1 in $(A.3)$ matches the notation and conditions from [106], such that the finite element error estimate for semilinear state equations shown there can be applied. Eventually, we will assume that the state constraints are only active in finitely many points. By the above construction, these active points have a positive distance to the boundary of Ω.

Since Ω is assumed to be twodimensional there is no need to reserve n for the spatial dimension of the problem. Moreover, we denote by $\|\cdot\|_{\infty,\Omega_0}$ the norm $\|\cdot\|_{L^\infty(\Omega_0)}$ for the subset $\Omega_0 \subset \Omega$, where the state constraints are prescribed. Recall that we agreed to set

$$V := H_0^1(\Omega).$$

In addition, we set

$$U := \mathbb{R}^m.$$

To handle the control constraints in a comfortable manner we introduce the set of admissible controls

$$U_{\text{ad}} := \{u \in \mathbb{R}^m \ : \ u_a \le u \le u_b\}.$$

Problem (3.1.1) belongs to a class of semi-infinite optimization problems. This classification comes from the fact that the control space is finite-dimensional, whereas the state constraints are prescribed in infinitely many points. The theory of semi-infinite optimization is quite well-established. We point out for example [16, 126, 149] and the references therein for an overview, as well as [55, 77, 137] where numerical aspects of semi-infinite programming are discussed. First order necessary and second order sufficient optimality conditions are known from the discussion of semi-infinite programming problems in [16]. Recently, they have been exploited for a model problem closer to Problem (3.1.1) in [36], relying on a constraint qualification, cf. [153]. Still challenging are aspects of finite-element error analysis that are not usually found in semi-infinite programming, since in our problem formulation the objective function as well as the constraints are implicitly defined via the solution of a PDE. In this context, we mention the results by Alt, [4], who derives error estimates for the approximation of infinite optimization problems and applies them to a finite difference discretization of nonlinear optimal control problems with ordinary differential equation.

Therefore, it will be our main aim to provide a priori finite element error estimates for the control problem and make new contributions to the numerical analysis of state-constrained optimal control problems. Let us first give a brief overview about related available results, most of which hold for control functions.

There are rather many publications that deal with purely control-constrained problems. After the first investigations by Falk in [49] and Geveci in [50], optimal error estimates for problems with linear and semilinear state equation, distributed or boundary control, and different types of control approximation have been analyzed in the recent past. To name a few results, consider for instance the estimate for problems with semilinear state equation and distributed control with piecewise constant control discretization of order $\mathcal{O}(h)$ both in the L^2- as well as the L^∞-norm by Arada, Casas, and Tröltzsch from [8]. For boundary control problems with constant control approximation we refer to results by Casas, Mateos, and Tröltzsch in [25]. For piecewise linear controls, Rösch was the first to improve the order of convergence in the L^2-norm to $\mathcal{O}(h^{\frac{3}{2}})$ under some additional structural assumptions in [128], cf. also [130]. Later, the same order of convergence has even been shown for semilinear boundary control problems in e.g. [24] by Casas and Mateos. Further results include a superconvergence property and a postprocessing technique by Meyer and Rösch, cf. [109], as well as so called variational discretization without explicit control discretization by Hinze, cf. [71].

For error estimates for state-constrained problems we mention that convergence for problems with control functions and only finitely many pointwise state constraints has been obtained by Casas in [20]. A linear-quadratic problem with pure pointwise state constraints in a domain has been analyzed by Deckelnick and Hinze in [38] for variational discretization of the controls, and an error estimate of the order $\mathcal{O}(h^{2-\frac{n}{2}-\varepsilon})$ has been obtained in spatial dimensions $n = 2, 3$. The same order of convergence is proven by Meyer for a problem with pointwise state and control constraints and piecewise constant control discretization in [107]. Deckelnick and Hinze later obtained $\mathcal{O}(h|\ln h|)$ in two space dimensions and $\mathcal{O}(h^{\frac{1}{2}})$ in three space dimensions, cf. [39].

Recently, in [106], Merino, Tröltzsch, and Vexler considered a nonconvex problem with finitely many control parameters and only finitely many pointwise state constraints. The analysis there was essentially based on a finite element error estimate in the L^∞-norm of order $h^2|\ln h|$ for uncontrolled equations, which in turn uses a corresponding result for linear equations by Rannacher and Vexler from [122]. By a transformation of the control problem from [106] into a finite-dimensional nonlinear programming problem, the order $h^2|\ln h|$ could also be obtained for the error in the controls. This was proven with the help of nontrivial stability estimates.

In this chapter, we aim at providing these higher order estimates for our semi-infinite problem (3.1.1) under certain assumptions. In this setting, the approach from [106] is not applicable. Even though a typical situation for semi-infinite optimization problems is for the optimal state to touch the bound in only finitely many points, the analysis is essentially complicated by the fact that these points are unknown. Then, one effect of prescribing the state constraints in a domain rather than in fixed points is a possible change of the location of the discrete active points in each refinement level of the discretization. In turn, this allows for the controls to vary more freely. It is therefore not a priori clear whether or not the optimal error estimate of [106] is valid for semi-infinite problems. Indeed, the low regularity of the Lagrange multiplier associated with the state constraints may even implicate that the full order of convergence cannot be achieved. For a linear-quadratic setting, such semi-infinite model problems have been analyzed by Merino, Tröltzsch, and the author in [104] and [105]. There, an order of $h^2|\ln h|$ is proven under certain structural assumptions on the active set, whereas an estimate of lower order was proven sharp for other situations. Now, we aim at extending these results to a more general nonconvex problem formulation, relying on the principal ideas published in [104] and [105].

We proceed by a brief discussion of the optimal control problem (3.1.1) with respect to existence of solutions, as well as first and second order optimality conditions. We provide these known results on elliptic state-constrained optimal control problems in the remainder of this subsection. Eventually, in Section 3.2, we exploit in more detail some assumptions and properties of importance to our convergence result. The main focus of this chapter is on the finite element discretization of the model problem, starting in Section 3.3 with the main result being presented in Section 3.4. This discussion is followed by some generalizations, 3.4, and some numerical experiments in Section 3.6.

3.1.2. The control-to-state operator and the reduced objective functional

The boundary value problem (3.1.1b) determines the properties of the optimal control problem (3.1.1) in a fundamental way, since it determines the dependence between a control $u \in \mathbb{R}^m$ and the state y. By the existence, uniqueness, and regularity results from Chapter 2 for solutions of the state equation (3.1.1b) for fixed $u \in \mathbb{R}^m$, the following definition is meaningful.

Definition 3.1.2. *The mapping*

$$G\colon U \to H_0^1(\Omega) \cap \mathcal{C}(\bar{\Omega}),$$

which assigns to each $u \in \mathbb{R}^m$ a unique state $y = G(u)$ that solves (3.1.1b) *in weak sense is called the control-to-state operator associated with Problem* (3.1.1). *In addition, we introduce the control-reduced objective function*

$$f\colon U \to \mathbb{R}, \quad f(u) := J(G(u), u).$$

The analysis of Problem (3.1.1) is essentially based on the properties of G and f. We therefore provide some useful auxiliary results.

Proposition 3.1.3. *[106, Lemma 2]. Under our general Assumption 3.1.1 the control-to-state operator G and the reduced objective function f are of class $\mathcal{C}^2$. For $u \in \mathbb{R}^m$ and arbitrary elements $v, v_1, v_2 \in \mathbb{R}^m$, the function $\tilde{y} := G'(u)v$ is given by the unique weak solution of*

$$\begin{aligned} \mathcal{A}\tilde{y} + \partial_y d(\cdot, y, u)\tilde{y} &= -\partial_u d(\cdot, y, u)v \quad && \text{in } \Omega, \\ \tilde{y} &= 0 && \text{on } \Gamma. \end{aligned} \tag{3.1.2}$$

The function $z_{v_1,v_2} := G''(u)[v_1, v_2]$ is the unique solution of

$$\begin{aligned} \mathcal{A}z + \partial_y d(\cdot, y, u)z &= -(\tilde{y}_1, v_1^T)d''(\cdot, y, u)(\tilde{y}_2, v_2^T)^T \quad && \text{in } \Omega, \\ z &= 0 && \text{on } \Gamma, \end{aligned} \tag{3.1.3}$$

with $\tilde{y}_i = G'(u)v_i$, $i = 1, 2$. The first and second order derivatives of f are given by

$$f'(u)v = \int\limits_{\Omega} \left(\partial_u \Psi(x, y, u)v + \partial_y \Psi(x, y, u)\tilde{y}\,\right) dx \tag{3.1.4}$$

as well as

$$f''(u)[v_1, v_2] = \int\limits_{\Omega} \left((\tilde{y}_1, v_1^T)\Psi''(x, y, u)(\tilde{y}_2, v_2^T)^T\right)\, dx \tag{3.1.5}$$

For G, f, and their derivatives, we obtain a Lipschitz result that will be used frequently in the following.

Lemma 3.1.4. *Let $u_1, u_2 \in U_{ad}$ and $v \in U$ be arbitrary control vectors. There exists a constant $C > 0$ such that*

$$\begin{aligned} \|G(u_1) - G(u_2)\|_{H^1(\Omega)\cap\mathcal{C}(\bar\Omega)} &\leq C|u_1 - u_2| && (3.1.6)\\ \|G'(u_1)v - G'(u_2)v\|_{H^1(\Omega)\cap\mathcal{C}(\bar\Omega)} &\leq C|u_1 - u_2||v| && (3.1.7)\\ \|G''(u_1)[v,v] - G''(u_2)[v,v]\|_{H^1(\Omega)\cap\mathcal{C}(\bar\Omega)} &\leq C|u_1 - u_2||v|^2 && (3.1.8)\end{aligned}$$

as well as

$$\begin{aligned} |f(u_1) - f(u_2)| &\leq C|u_1 - u_2| && (3.1.9)\\ |f'(u_1)v - f'(u_2)v| &\leq C|u_1 - u_2||v| && (3.1.10)\\ |f''(u_1)[v,v] - f''(u_2)[v,v]| &\leq C|u_1 - u_2||v|^2 && (3.1.11)\end{aligned}$$

is satisfied.

Proof. The proof, though technical, follows by standard methods. For associated results for control functions, we refer for instance to [144]. We postpone the proof to Appendix A.2. □

3.1.3. The control-reduced problem formulation and existence of solutions

After these preliminary considerations, we now prove the existence of at least one global solution to the model problem. It is convenient to reformulate the problem with the help of the control-to-state operator G and the reduced objective function f. We obtain the control-reduced formulation

$$\text{Minimize } f(u) \text{ subject to } u \in U_{\text{ad}}, \quad G(u) \leq b \text{ in } K. \tag{$\mathbb{P}$}$$

In addition to the set of admissible controls, U_{ad}, we define the set of feasible controls

$$U_{\text{feas}} = \{u \in U_{\text{ad}} \ : \ G(u) \leq b \text{ in } K\}$$

which simplifies further discussion. Then, the existence of an optimal control is easily obtained due to the finite dimensional control space.

Theorem 3.1.5. *Let the set of feasible controls U_{feas} be nonempty. Then, under Assumptions 3.1.1, there exists an optimal control $\bar u \in U_{feas}$ with associated optimal state $\bar y = G(\bar u)$ for Problem ($\mathbb{P}$).*

Proof. The proof of existence benefits from the fact that the control space is finite dimensional. The set U_{feas} is compact, since it is bounded and closed, and the Weierstrass theorem yields the assertion. □

As was also pointed out in [106], this result is not generally true if control functions instead of control vectors appear nonlinearly in the state equation. Since the control problem considered is not necessarily convex, uniqueness of the solution is not guaranteed. Therefore, we will consider local solutions in the sense of the following definition.

Definition 3.1.6. *A feasible control $\bar u \in U_{feas}$ is called a local solution of ($\mathbb{P}$), if there exists a positive real number ε such that $f(\bar u) \leq f(u)$ holds for all feasible controls $u \in U_{feas}$ of ($\mathbb{P}$) with $|u - \bar u| \leq \varepsilon$.*

3.1.4. First and second order optimality conditions

We already know that the state functions are regular enough to expect a Lagrange multiplier at least in the space $\mathcal{C}(\bar{\Omega})^*$. However, to guarantee its existence, a constraint qualification is required. We rely on a linearized Slater condition.

Assumption 3.1.7. *We say that $\bar{u}$ satisfies the linearized Slater condition for* $(\mathbb{P})$, *if there exist a real number $\gamma > 0$ and a control vector $u_\gamma \in U_{ad}$ such that*

$$G(\bar{u})(x) + G'(\bar{u})(u_\gamma - \bar{u})(x) \le b - \gamma \quad \forall\, x \in K \tag{3.1.12}$$

is satisfied, which, because of continuity, is equivalent to

$$G(\bar{u})(x) + G'(\bar{u})(u_\gamma - \bar{u})(x) < b \quad \forall\, x \in K.$$

Remark 3.1.8. *When deriving convergence results for a discretized version of Problem* $(\mathbb{P})$, *we will eventually consider auxiliary problem formulations that require the controls to be in the vicinity of $\bar{u}$, i.e. $|u - \bar{u}| \le \varepsilon$ for a small positive parameter ε. Then, we will need that the Slater point u_γ fulfills the same closeness condition. Indeed, it is reasonable to assume this, since by choosing*

$$u_\gamma^\varepsilon = \bar{u} + t(u_\gamma - \bar{u}), \quad t = \min\left\{1, \frac{\varepsilon}{|u_\gamma - \bar{u}|}\right\},$$

this closesness condition is obviously fulfilled, and the Slater point property

$$G(\bar{u}) + G'(\bar{u})(u_\gamma^\varepsilon - \bar{u}) = (1-t)G(\bar{u}) + t(G(\bar{u}) + G'(\bar{u})(u_\gamma - \bar{u})) \le (1-t)b + t(b-\gamma) \le b - \gamma_\varepsilon$$

is fulfilled in K with a distance parameter $\gamma_\varepsilon = t\gamma$. For later purposes, we point out that γ_ε depends only linearly on ε.

Then, the following Karush-Kuhn-Tucker conditions are obtained from [153] or by adapting the theory of [18, 2].

Theorem 3.1.9. *Let Assumptions 3.1.1 and 3.1.7 be satisfied. If $\bar{u}$ is a locally optimal control of* $(\mathbb{P})$ *with associated optimal state $\bar{y}$, then there exists a regular Borel measure $\bar{\mu} \in \mathcal{M}(K)$ and an adjoint state $\bar{p} \in W^{1,s}(\Omega)$, $1 \le s < 2$ with*

$$\begin{aligned} \mathcal{A}\bar{p} + \partial_y d(\cdot, \bar{y}, \bar{u})\bar{p} &= \partial_y \psi(\cdot, \bar{y}, \bar{u}) + \bar{\mu} && \text{in } \Omega \\ \bar{p} &= 0 && \text{on } \Gamma, \end{aligned} \tag{3.1.13}$$

in the sense of (2.4.3),

$$\langle g, u - \bar{u} \rangle \ge 0 \quad \forall u \in U_{ad}, \tag{3.1.14}$$

where $g \in \mathbb{R}^m$ is defined by $g = (g_i)$ with

$$g_i := \int_\Omega \left(\partial_{u_i}\psi(x, \bar{y}, \bar{u}) - \bar{p}\,\partial_{u_i} d(x, \bar{y}, \bar{u})\right) dx, \quad i = 1, \ldots, m,$$

as well as the complementary slackness condition

$$\int_K (\bar{y} - b)\, d\bar{\mu}(x) = 0, \quad \bar{\mu} \ge 0, \quad \bar{y}(x) \le b \quad \forall x \in K. \tag{3.1.15}$$

Remark 3.1.10. *In [12], Bergounioux and Kunisch have shown for linear-quadratic problems that the irregular part of the measure $\bar{\mu}$ is situated at the boundary of the active set. In our situation, this implies that no part of $\bar{\mu}$ acts on the boundary Γ, since no constraints are imposed there.*

Let us point out that instead of using the variational inequality (3.1.14) to handle the control constraints, it is possible to use Lagrange multipliers $\bar{\eta}_a, \bar{\eta}_b \in \mathbb{R}^m$ and formulate a gradient equation. This will be helpful when discussing the discrete optimal control problem.

Lemma 3.1.11. *Let the tuple $(\bar{u}, \bar{y}, \bar{p}, \bar{\mu}) \in U_{ad} \times V \times W^{1,s}(\Omega) \times \mathcal{M}(K)$ fulfill the setting of Theorem 3.1.9. There exist nonnegative Lagrange multipliers $\bar{\eta}_a, \bar{\eta}_b \in \mathbb{R}^m$ such that the variational inequality* (3.1.14) *can equivalently be written with the help of additional Lagrange multipliers. There exist (componentwise) nonnegative Lagrange multipliers $\bar{\eta}_a, \bar{\eta}_b \in \mathbb{R}^m$ such that*

$$g_i + \bar{\eta}_{b,i} - \bar{\eta}_{a,i} = 0 \tag{3.1.16}$$

and

$$\langle u_a - \bar{u}, \bar{\eta}_a \rangle = \langle \bar{u} - u_b, \bar{\eta}_b \rangle = 0. \tag{3.1.17}$$

Proof. This is a well-known result in optimal control. The existence of multipliers $\bar{\eta}_a, \bar{\eta}_b \in \mathbb{R}^m$ follows by construction, i.e. setting

$$\bar{\eta}_{a,i} := \max(0, g_i), \qquad \bar{\eta}_{b,i} = -\min(0, g_i),$$

which are obviously nonnegative. Then, it is easy to prove that the complementary slackness conditions (3.1.17) are fulfilled. We refer to [144, Theorem 2.29] for a related discussion of a linear-quadratic control problem with control function. □

For our further analysis, we define the reduced Lagrangian

$$\mathcal{L} : \mathbb{R}^m \times \mathcal{M}(K) \to \mathbb{R}$$

for Problem $(\mathbb{P})$ by

$$\mathcal{L}(u, \mu) := f(u) + \int_K (G(u) - b)\, d\mu(x). \tag{3.1.18}$$

By Proposition 3.1.3, it is clear that the Lagrangian is twice continuously differentiable with respect to u. For brevity, we write $\mathcal{L}'$ instead of $\partial\mathcal{L}/\partial u$ as well as $\mathcal{L}''$ instead of $\partial^2\mathcal{L}/\partial u^2$. It is known that the first order necessary optimality conditions of Theorem 3.1.9 can now be expressed as

$$\mathcal{L}'(\bar{u}, \bar{\mu})(u - \bar{u}) \geq 0 \quad \forall\, u \in U_{\mathrm{ad}}, \tag{3.1.19}$$

or, with the Lagrange multipliers $\bar{\eta}^a, \bar{\eta}^b \in \mathbb{R}^m$ from Lemma 3.1.11,

$$\mathcal{L}'(\bar{u}, \bar{\mu}) + \bar{\eta}^b - \bar{\eta}^a = 0, \tag{3.1.20}$$

combined with the complementary slackness condition (3.1.15) for the nonnegative Lagrange multiplier $\bar{\mu}$, and in case of (3.1.20), the complementary slackness condition (3.1.17) for $\bar{\eta}^a, \bar{\eta}^b \geq 0$. We refer to [91]. In fact, the optimality conditions from Theorem 3.1.9 are often derived by means of the formal

Lagrange method. For $u \in \mathbb{R}^m$, $y = G(u)$ and arbitrary elements $v, v_1, v_2 \in \mathbb{R}$ as well as $\tilde{y}_1 := G'(u)v_1$ and $\tilde{y}_2 := G'(u)v_2$, the first and second order derivative of the Lagrangian takes the form

$$\begin{aligned}\mathcal{L}'(u,\mu)v &= \int_\Omega (\partial_u\psi(x,y,u) - p(x)\partial_u d(x,y,u))v\,dx\\ \mathcal{L}''(u,\mu)[v_1,v_2] &= \int_\Omega ((\tilde{y}_1, v_1^T)(\psi''(x,y,u) - p(x)d''(x,y,u))(\tilde{y}_2, v_2^T)^T)\,dx,\end{aligned}$$

where p solves the adjoint equation (3.1.13) with u, y, μ substituted into the right-hand side. This is obtained by straight-forward computations, defining the Lagrangian as

$$\mathcal{L}(u,y,p,\mu) := J(y,u) + \int_K (y-b)\,d\mu(x).$$

For control functions appearing linearly in the right-hand side, we refer to [144], where these calculations have been carried out in detail.

For non-convex problems, stationary points, i.e. solutions of the optimality system stated in Theorem 3.1.9, are not automatically solutions of the optimal control problem. Sufficient optimality conditions can be obtained by a classical second-order analysis involving the Lagrangian. Let us state a Lipschitz property of the second derivative of $\mathcal{L}''$ with respect to the control u. This follows by straight-forward computations following the steps in [144].

Corollary 3.1.12. *Let $\mu \in \mathcal{M}(K)$ and $u_1, u_2 \in \mathbb{R}^m$ be given. The second derivative $\mathcal{L}''$ of the Lagrangian is Lipschitz continuous with respect to u, i.e. there exists a constant $c > 0$ such that*

$$|\mathcal{L}''(u_1,\mu)[v,v] - \mathcal{L}''(u_2,\mu)[v,v]| \leq c|u_1-u_2||v|^2.$$

for all $v \in \mathbb{R}^m$.

Proof. The proof is rather straigh-forward, and follows by the definition of $\mathcal{L}$ and Lemma 3.1.4. We only need to point out that

$$\int_K (G''(u_1)[v,v] - G''(u_2)[v,v])\,d\mu(x) \leq \|G''(u_1)[v,v] - G''(u_2)[v,v]\|_{C(K)}\|\mu\|_{C(K)^*}$$

by the definition of the dual norms. □

In order to obtain a sufficient optimality condition we make the following assumption:

Assumption 3.1.13. *There exists a constant $\alpha > 0$ such that*

$$\mathcal{L}''(\bar{u},\bar{\mu})[v,v] \geq \alpha|v|^2. \tag{3.1.21}$$

is satisfied for all $v \in \mathbb{R}^m$.

Note that Assumption 3.1.13 is equivalent to

$$\int_\Omega ((\tilde{y}, v)(\psi''(x,\bar{y},\bar{u}) - \bar{p}(x)d''(x,\bar{y},\bar{u}))(\tilde{y}, v^T)^T)\,dx \geq \alpha|v|^2$$

for all $v \in \mathbb{R}^m$ and all $\tilde{y}$ satisfying the linearized state equation

$$\begin{aligned} \mathcal{A}\tilde{y} + \partial_y d(\cdot, y, u)\tilde{y} &= -\partial_u d(\cdot, y, u)v \quad \text{in } \Omega, \\ \tilde{y} &= 0 \quad \text{on } \Gamma. \end{aligned}$$

The last assumption guarantees, together with the first order necessary optimality conditions, local optimality of a control $\bar{u} \in U_{\text{feas}}$. This follows by standard arguments and will be proven in the following. Assumption 3.1.13 is a strong second order sufficient condition. It is possible to formulate a weaker condition along the lines of e.g. [16], where the motivation is to formulate sufficient conditions that are as close to the associated necessary conditions as possible. However, strong assumptions are often used in e.g. the analysis of numerical solution algorithms. Since we are more interested in the resulting quadratic growth condition we will rely on the quite strong condition.

Theorem 3.1.14. *Let a control vector $\bar{u} \in U_{feas}$ satisfy the first order optimality conditions of Theorem 3.1.9 as well as the second order sufficient condition from Assumption 3.1.13. Then, $\bar{u}$ is a locally optimal control for Problem ($\mathbb{P}$), and there exist constants $\varepsilon > 0$ and $\beta > 0$ such that the quadratic growth condition*

$$f(\bar{u}) \leq f(u) - \beta|u - \bar{u}|^2 \tag{3.1.22}$$

is satisfied for all $u \in U_{feas}$ with $|u - \bar{u}| \leq \varepsilon$.

Proof. For a feasible control u, we proceed by Taylor expansion of $\mathcal{L}$ in $\bar{u}$ and obtain

$$\mathcal{L}(u, \bar{\mu}) = \mathcal{L}(\bar{u}, \bar{\mu}) + \mathcal{L}'(\bar{u}, \bar{\mu})(u - \bar{u}) + \frac{1}{2}\mathcal{L}''(u_\xi, \bar{\mu})[u - \bar{u}, u - \bar{u}]$$

with a control $u_\xi = \bar{u} + \xi(u - \bar{u})$ with $\xi \in (0, 1)$. By the optimality of $\bar{u}$, we have

$$\begin{aligned} \mathcal{L}(u, \bar{\mu}) &\geq \mathcal{L}(\bar{u}, \bar{\mu}) + \frac{1}{2}\mathcal{L}''(u_\xi, \bar{\mu})[u - \bar{u}, u - \bar{u}] \\ &= \mathcal{L}(\bar{u}, \bar{\mu}) + \frac{1}{2}\mathcal{L}''(\bar{u}, \bar{\mu})[u - \bar{u}, u - \bar{u}] + \frac{1}{2}\left(\mathcal{L}''(u_\xi, \bar{\mu}) - \mathcal{L}''(\bar{u}, \bar{\mu})\right)[u - \bar{u}, u - \bar{u}]. \end{aligned} \tag{3.1.23}$$

In addition, the complementary slackness conditions (3.1.15) and the positivity of $\bar{\mu}$ combined with the feasibility of u imply

$$f(\bar{u}) = \mathcal{L}(\bar{u}, \bar{\mu}), \quad f(u) \geq \mathcal{L}(u, \bar{\mu}). \tag{3.1.24}$$

Together with (3.1.21) and Corollary 3.1.12 the statements (3.1.23) and (3.1.24) imply

$$f(u) \geq f(\bar{u}) + \frac{\alpha}{2}|u - \bar{u}|^2 - \frac{c}{2}|\bar{u} - u_\xi||u - \bar{u}|^2.$$

Clearly, by $|\bar{u} - u_\xi| = \xi|u - \bar{u}|$ the assertion is obtained for $|u - \bar{u}| \leq \varepsilon$ with ε sufficiently small. □

3.2. Structural assumptions and properties

Now, we formulate an assumption on the active set of the control problem that is quite natural for semi-infinite optimization and will later on play an important role in our error estimates. Based on this assumption, we derive some helpful properties of the optimal state as well as the associated Lagrange multipliers.

3.2.1. Structure of the active set

Before we continue, let us motivate the following steps with the help of a linear-quadratic example, cf. [105]. Consider the following linear state equation

$$\begin{aligned} \mathcal{A}y &= \sum_{i=1}^{m} u_i e_i \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Gamma, \end{aligned}$$

where the e_i, $i = 1, \ldots, m$ are fixed, smooth basis functions $e_i \colon \Omega \to \mathbb{R}$. Then, suppose that $y = b$ holds on an open set $\tilde{\Omega} \subset K$ of positive measure. This implies

$$\mathcal{A}y = 0 = \sum_{i=1}^{m} u_i e_i \quad \text{in } \tilde{\Omega}.$$

If we assume the functions e_i to be linearly independent on every open subset of Ω, this cannot be true. Therefore, it seems reasonable also for nonlinear problems to assume that the optimal state is active only on sets of measure zero. Still, the so-called active sets can be quite diverse, but a situation quite often observed in semi-infinite programming is that constraints become active in only finitely many points.

We therefore make the following assumption:

Assumption 3.2.1. *Let $\bar{u}$ be a locally optimal control of* ($\mathbb{P}$). *For the optimal state $\bar{y} = G(\bar{u})$, there are exactly n points $\bar{x}_1, \ldots, \bar{x}_n \in \operatorname{int} K$ where the state constraints become active, i.e. where $\bar{y}(\bar{x}_j) = b$. Moreover, there exists $\omega > 0$ such that*

$$-\langle \theta, \nabla^2 \bar{y}(\bar{x}_j)\theta \rangle \geq \omega |\theta|^2 \qquad \forall \theta \in \mathbb{R}^2, \ \forall j = 1, \ldots, n. \tag{3.2.1}$$

Note that $\bar{y} \in \mathcal{C}^{\nu,2}(\Omega)$. We will call the points $\bar{x}_1, \ldots, \bar{x}_n$ active points of Problem ($\mathbb{P}$), or simply active points of $\bar{y}$. Let us elaborate on some consequences that will help to characterize discrete active points. Before, though, we clearly emphasize again that n is not to be confused with the spatial dimension of Ω, since we consider a two-dimensional setting by Assumption 3.1.1.

3.2.2. Properties of the optimal state

Eventually, in our final error estimates for Problem (3.1.1), we will make use of a characterization of discrete and continuous active points. We now provide all properties of the optimal state needed in later estimates. Assumption 3.2.1 implies that the considered optimal state $\bar{y}$ admits strict local maxima in the active points. Moreover, a condition of quadratic decrease of $\bar{y}$ can be deduced in the neighborhood of such an active point $\bar{x}_j$. We can also prove that the optimal state $\bar{y}$ has a positive distance to the bound b outside some neighborhood of all active points.

Lemma 3.2.2 ([105]). *Let Assumptions 3.1.1 and 3.2.1 be valid. There exists a real number $R > 0$ such that for each $j \in \{1, \ldots, n\}$,*

$$\bar{y}(x) \leq b - \frac{\omega}{4}|x - \bar{x}_j|^2 \quad \forall x \in K \ \textit{with} \ |x - \bar{x}_j| \leq R \tag{3.2.2}$$

is satisfied. Moreover, there exists a constant $\delta > 0$ *such that*

$$\bar{y}(x) \leq b - \delta \quad \forall x \in K \setminus \bigcup_{j=1}^{n} B_R(\bar{x}_j). \tag{3.2.3}$$

Proof. Recall that $\bar{y} \in \mathcal{C}^{2,\nu}(\Omega)$, so that for each active point $\bar{x}_j \in K$ of $\bar{y}$, $j = 1, \dots, n$, and arbitrary $x \in K$ a second-order Taylor expansion of $\bar{y}$ in $\bar{x}_j$ is possible. We obtain

$$\bar{y}(x) = \bar{y}(\bar{x}_j) + \langle \nabla \bar{y}(\bar{x}_j), x - \bar{x}_j \rangle + \frac{1}{2} \langle x - \bar{x}_j, \nabla^2 \bar{y}(x_\xi)(x - \bar{x}_j) \rangle \tag{3.2.4}$$

with an $x_\xi = \bar{x}_j + \xi(x - \bar{x}_j) \in \mathbb{R}^2$, $\xi \in (0,1)$. Now notice that $\bar{y}(\bar{x}_j) = b$ and $\nabla \bar{y}(\bar{x}_j) = 0$, since $\bar{x}_j$ is a local maximum of $\bar{y}$. Consequently, (3.2.4) implies

$$\bar{y}(x) = b + \frac{1}{2} \langle x - \bar{x}_j, \nabla^2 \bar{y}(\bar{x}_j)(x - \bar{x}_j) \rangle - \frac{1}{2} \langle x - \bar{x}_j, (\nabla^2 \bar{y}(x_\xi) - \nabla^2(\bar{y}(\bar{x}_j)))(x - \bar{x}_j) \rangle. \tag{3.2.5}$$

Finally, using Assumption 3.2.1 and the Hölder continuity of $\nabla^2 \bar{y}$ in (3.2.5) leads to

$$\bar{y}(x) \leq b - \frac{\omega}{2} |x - \bar{x}_j|^2 + \frac{c}{2} |x - \bar{x}_j|^\nu |x - \bar{x}_j|^2. \tag{3.2.6}$$

Then, there exists a real number $R_j > 0$ such that (3.2.2) is satisfied. Outside of all the balls $B_{R_j}(\bar{x}_j)$, $\bar{y}$ is inactive by assumption. Since the problem admits only finitely many active points, we can define R as the maximum over all R_j, and by continuity of $\bar{y}$ as well as Assumption 3.2.1 we conclude that there exists a $\delta > 0$ such that (3.2.3) is satisfied. □

3.2.3. Structure of Lagrange multipliers

Due to the structural assumption on the optimal state $\bar{y}$, we can deduce a helpful property for the Lagrange multiplier $\bar{\mu}$ associated with the state constraints in Problem ($\mathbb{P}$), cf. also [17, Corollary 1].

Proposition 3.2.3. *Let* $\bar{\mu} \in \mathcal{M}(K)$ *be a Lagrange multiplier fulfilling the optimality system in Theorem 3.1.9 for a locally optimal control* $\bar{u}$, *and let Assumption 3.2.1 be satisfied. Then* $\bar{\mu}$ *admits the form*

$$\bar{\mu} = \sum_{j=1}^{n} \bar{\mu}_j \delta_{\bar{x}_j}, \quad \bar{\mu}_j \in \mathbb{R},$$

where $\delta_{\bar{x}_j}$ *denotes the Dirac measure located at the active point* $\bar{x}_j$ *of the optimal state* $\bar{y} = G(\bar{u})$.

Proof. From the complementary slackness conditions (3.1.15), the feasibility of $\bar{y}$, and the nonnegativity of $\bar{\mu}$ we observe

$$\begin{aligned} 0 = \int_K (\bar{y}(x) - b)\, d\bar{\mu}(x) &= \sum_{j=1}^{n} \int_{B_\rho(\bar{x}_j)} (\bar{y}(x) - b)\, d\bar{\mu}(x) + \int_{K \setminus \bigcup_{j=1}^{n} B_\rho(\bar{x}_j)} (\bar{y}(x) - b)\, d\bar{\mu}(x) \\ &\leq \int_{K \setminus \bigcup_{j=1}^{n} B_\rho(\bar{x}_j)} (\bar{y}(x) - b)\, d\bar{\mu}(x) \leq 0 \end{aligned}$$

for all $\rho > 0$ small enough such that the $B_\rho(\bar{x}_j)$ are pairwise disjoint. Since $\bar{y}(x) - b$ is strictly negative on $K \setminus \bigcup_{j=1}^{n} B_\rho(\bar{x}_j)$, this implies that $\bar{\mu}(x) = 0$ must hold on $K \setminus \bigcup_{j=1}^{n} B_\rho(\bar{x}_j)$. Then, we obtain for all functions $g \in \mathcal{C}(K)$:

$$\begin{aligned} \int_K g(x)\,d\bar{\mu}(x) &= \sum_{j=1}^{n} \int_{B_\rho(\bar{x}_j)} g(x)\,d\bar{\mu}(x) \\ &= \sum_{j=1}^{n} \int_{B_\rho(\bar{x}_j)} g(\bar{x}_j)\,d\bar{\mu}(x) + \sum_{j=1}^{n} \int_{B_\rho(\bar{x}_j)} (g(x) - g(\bar{x}_j))\,d\bar{\mu}(x) \\ &= \sum_{j=1}^{n} g(\bar{x}_j) \int_{B_\rho(\bar{x}_j)} 1\,d\bar{\mu}(x) + \sum_{j=1}^{n} \int_{B_\rho(\bar{x}_j)} (g(x) - g(\bar{x}_j))\,d\bar{\mu}(x). \end{aligned} \tag{3.2.7}$$

We point out that

$$\lim_{\rho \to 0} \int_{B_\rho(\bar{x}_j)} 1\,d\bar{\mu}(x) =: \bar{\mu}_j$$

exists, since the associated sequence is monotonically decreasing by the nonnegativity of $\bar{\mu}$, and bounded from below by 0. Moreover, for $\rho \to 0$, we obtain that $g(x) - g(\bar{x}_j)$ tends uniformly to zero in each $B_\rho(\bar{x}_j)$ by the continuity of g. Taking the limit $\rho \to 0$ in (3.2.7) hence yields:

$$\int_K g(x)\,d\bar{\mu}(x) = \sum_{j=1}^{n} g(\bar{x}_j)\bar{\mu}_j = \sum_{j=1}^{n} \bar{\mu}_j \delta_{\bar{x}_j}(g),$$

which proves the assertion. □

In the sequel, we identify the multiplier $\bar{\mu} \in \mathcal{M}(K)$ with the associated n-vector of coefficients, which we denote by $\bar{\mu}$ as well. Note that we have neither assumed nor formulated a uniqueness result for the Lagrange multiplier. In fact, the discussion of whether or not Lagrange multipliers are uniquely determined can be quite involved, cf. for instance the discussion in [16].

3.2.4. Strongly active constraints

For later reference, we introduce the following characterization of active constraints.

Definition 3.2.4. *We call a control constraint strongly active in $\bar{u}_i$ if either $\bar{u}_i = u_{a,i}$ and $\bar{\eta}_{a,i} \geq \eta_0 > 0$ or $\bar{u}_i = u_{b,i}$ and $\bar{\eta}_{b,i} \geq \eta_0 > 0$ for all Lagrange multipliers $\bar{\eta}_a, \bar{\eta}_b \in \mathbb{R}^m$ fulfilling the gradient equation from Lemma 3.1.11. Moreover, we define the index set of the strongly active control constraints associated with $\bar{u}$ by*

$$\mathcal{A}_{\bar{u}} = \{i \in \{1, \dots m\} \mid \bar{\eta}_{a,i} > 0 \text{ or } \bar{\eta}_{b,i} > 0 \text{ for all Lagrange multipliers } \bar{\eta}_a \text{ and } \bar{\eta}_b\},$$

and denote by $m_A = \sharp\mathcal{A}_u \leq m$ the number of strongly active control constraints. By the index set

$$\mathcal{I}_{\bar{u}} = \{1, \dots m\} \setminus \mathcal{A}_{\bar{u}},$$

we cover the remaining (inactive or weakly active) control constraints.

We will eventually be able to remove the strongly active components of the control from the problem formulation by showing that associated components of the discrete optimal control are also active for small discretization parameters $h > 0$. Then, we can focus in detail on the state constraints, which are our main concern.

Definition 3.2.5. *Let $\bar{x}_j \in K$, $j \in \{1, \dots, n\}$ be an active point of $\bar{y}$, i.e. $\bar{y}(\bar{x}_j) = b$. We say that the state constraint is strongly active in $\bar{x}_j$, if the associated component $\bar{\mu}_j$ of all corresponding Lagrange multipliers is strictly positive. Then, we call $\bar{x}_j$ a strongly active point of $(\mathbb{P})$, or just a strongly active point of $\bar{y}$. Let*

$$\mathcal{A}_{\bar{y}} = \{j \in \{1, \dots, n\} \mid \bar{\mu}_j > \mu_0 \text{ for all Lagrange multipliers } \bar{\mu}\}$$

denote the index set for the strongly active state constraints. We define $n_A := \sharp \mathcal{A}_{\bar{y}} \leq n$ as the number of strongly active points of $\bar{y}$. Finally, define

$$\mathcal{I}_{\bar{y}} = \{1, \dots n\} \setminus \mathcal{A}_{\bar{y}}.$$

as the set of inactive or weakly active state constraints.

Remark 3.2.6. *Definitions 3.2.4 and 3.2.5 help to present the convergence result in Section 3.4.3 in a concise way. We will eventually also comment on situations where all Lagrange multiplier triples admit in sum at least as many positive components as there are controls, but these positive components may be associated with different control components or different active points, cf. the discussion in Section 3.5.*

3.3. The finite element discretization of the model problem

The discretization of the model problem requires on the one hand the discretization of the underlying state equation, which we treat by standard conforming finite elements. On the other hand, the infinitely many state constraints need to be discretized. We will obtain this by simply prescribing the constraints in the nodes contained in the set K, but need to discuss whether or not this influences the feasibility of controls in the neighborhood of a local solution $\bar{u}$. In contrast to problems with control functions, no control discretization is necessary, that could influence the rate of convergence. This is contrary to problems with control functions whose discretization will result in different admissible sets on the one hand and an additional error component on the other hand. There, special discretization techniques, cf. [71], have to be employed to avoid this.

3.3.1. Preliminaries

In this section, we provide some fundamental and to the most part known facts and results for discrete problems involving elliptic state equations. We consider a family of triangulations $\{\mathcal{T}_h\}_{h>0}$ of $\bar{\Omega}$, consisting of nonoverlapping triangles $T \in \mathcal{T}_h$ such that $\bigcup_{T \in \mathcal{T}_h} T = \bar{\Omega}$. Here, we benefit from the fact that the domain Ω is polygonally bounded. By

$$\mathcal{N}_h := \{x_h \mid x_h \text{ is a vertex of } \mathcal{T}_h\}$$

we denote the set of nodes defining the triangulation, and for later use we introduce the set of nodes contained in the subset K by

$$\mathcal{N}_h^K := K \cap \mathcal{N}_h.$$

For each triangle $T \in \mathcal{T}_h$, we introduce the outer diameter $\rho_o(T)$ of T, and the diameter $\rho_i(T)$ of the largest circle contained in T. The mesh size h is defined by $h = \max_{T \in \mathcal{T}_h} \rho_o(T)$. The grid fulfills the following regularity assumption, cf. for instance [34]:

Assumption 3.3.1. *There exist positive constants ρ_o and ρ_i such that*

$$\frac{\rho_o(T)}{\rho_i(T)} \leq \rho_i \text{ and } \frac{h}{\rho_o(T)} \leq \rho_o, \qquad \forall T \in \mathcal{T}_h,$$

are fulfilled for all $h > 0$.

Definition 3.3.2. *Associated with the given triangulation $\mathcal{T}_h$, we introduce the discrete state space as the set of cellwise linear and continuous functions*

$$V_h = \{v_h \in \mathcal{C}(\bar{\Omega}) \mid v_{h|_T} \in \mathcal{P}_1(T) \;\; \forall T \in \mathcal{T}_h, \, v_h = 0 \text{ on } \Gamma\},$$

where $\mathcal{P}_1(T)$ denotes the set of affine real-valued functions defined on T.

3.3.2. Discretization of the state equation

In this section we discretize the state equation and collect known properties of the control-to-state operator and the objective function on the discrete level. Discretizing the state equation (3.1.1b) yields

$$\mathbf{a}(y_h, \varphi) + (d(\cdot, y_h, u), \varphi) = 0 \quad \forall \varphi \in V_h, \tag{3.3.1}$$

and we obtain the following result on existence and discretization error estimates.

Theorem 3.3.3. *Let Assumptions 3.1.1 and 3.3.1 be satisfied. Then, for each $u \in \mathbb{R}^m$ the equation (3.3.1) has a unique solution $y_h \in V_h$. There exists a constant $C > 0$ independent of h such that*

$$||y_h||_{H^1(\Omega)} \leq C(|u| + 1) \tag{3.3.2}$$

is satisfied. Moreover, with $y := G(u)$, the error estimates

$$||y_h - y|| \leq Ch^2 ||y||_{H^2(\Omega)} \tag{3.3.3}$$

$$||y_h - y||_\infty \leq Ch ||y||_{H^2(\Omega)} \tag{3.3.4}$$

$$||y_h - y||_{\infty, \Omega_0} \leq Ch^2 |\ln h| (||y||_{W^{2,\infty}(\Omega_1)} + ||y||_{H^2(\Omega)}) \tag{3.3.5}$$

are satisfied.

Proof. The existence of $y_h \in V_h$ and the error estimates (3.3.3) and (3.3.4) follow from [23], where they were proven for control functions. The improved L^∞ estimate (3.3.5) in Ω_0 was proven in [106] based on a result for linear equations shown in [122]. The stability estimate (3.3.2) is a standard result obtained from testing (3.3.1) with y_h: The H_0^1-ellipticity of $\mathbf{a}$ and the monotonicity of d yield

$$c||y_h||_V^2 \leq \mathbf{a}(y_h, y_h) + (d(\cdot, y_h, u) - d(\cdot, 0, u), y_h) = -(d(\cdot, 0, u), y_h) \leq ||d(\cdot, 0, u)|| \, ||y_h||,$$

and therefore

$$||y_h||_V \leq c(||d(\cdot, 0, u) - d(\cdot, 0, 0) + d(\cdot, 0, 0)||),$$

and the Lipschitz continuity of d combined with the boundedness of U_{ad} yields the assertion. □

Corollary 3.3.4. *For $u \in U_{ad}$, the associated discrete states $y_h := G_h(u)$ are uniformly bounded in $L^\infty(\Omega)$ independently of u.*

From now on, we will sometimes speak of the discrete problem or the discrete setting, as opposed to the continuous problem, or continuous settting. Note that the definition of the control space and the set of admissible controls remains unchanged. We define a discrete control-to-state-operator and a discrete control-to-state mapping, and supply analogous stability and differentiability results as in the last section.

Definition 3.3.5. *We call the mapping $G_h \colon \mathbb{R}^m \to V_h$, which assigns a unique discrete state $y_h = G_h(u)$ to each $u \in \mathbb{R}^m$ the discrete control-to-state operator. In addition, we introduce the discrete control-reduced objective function*

$$f_h \colon U \to \mathbb{R}, \quad f_h(u) := J(G_h(u), u).$$

Remark 3.3.6. *In the following, we will quite frequently make use of the boundedness of U_{ad} and the resulting uniform boundedness of the discrete states from Corollary 3.3.4. In particular, Corollary 3.3.4 allows to make use of the Lipschitz properties of the nonlinearity d and its derivatives with Lipschitz constants that do not depend on either the discretization parameter h nor the control $u \in U_{ad}$.*

Combining the Lipschitz properties of Ψ and the discretization error estimate from Theorem 3.3.3 yields an error estimate for the objective function:

Corollary 3.3.7. *There exists a constant $C > 0$ such that*

$$|f(u) - f_h(u)| \leq Ch^2$$

is satisfied for all h sufficiently small and all $u \in U_{ad}$.

Proof. Let $y = G(u)$ and $y_h = G_h(u)$ denote the continuous and discrete states associated with $u \in U_{\text{ad}}$. Due to Theorems 2.4.6 and 3.3.3, we know that $\|y\|_{C(\Omega)}$ and $\|y_h\|_{C(\Omega)}$ are uniformly bounded for all $u \in U_{\text{ad}}$. Then, the assertion follows immediately from the local Lipschitz continuity of Ψ:

$$|f(u) - f_h(u)| \leq \int_\Omega |\Psi(x, y, u) - \Psi(x, y_h, u)| \; dx \leq C\|y - y_h\| \leq ch^2$$

by Theorem 3.3.3. $\square$

Proposition 3.3.8. *The discrete control-to-state operator $G_h \colon \mathbb{R}^m \to V_h$ and the discrete reduced objective function $f_h \colon U \to \mathbb{R}$ are of class $\mathcal{C}^2$. For $u \in \mathbb{R}^m$ and arbitrary elements $v, v_1, v_2 \in \mathbb{R}^m$, the function $\tilde{y}_h := G_h'(u)v$ is given by the unique solution of*

$$\mathbf{a}(\tilde{y}_h, \varphi) + (\partial_y d(\cdot, y_h, u)\tilde{y}_h, \varphi) = -(\partial_u d(\cdot, y_h, u)v, \varphi) \quad \forall \varphi \in V_h, \tag{3.3.6}$$

and the function $z_h^{v_1,v_2} := G_h''(u)[v_1, v_2]$ is the unique solution of

$$\mathbf{a}(z_h, \varphi) + (\partial_y d(\cdot, y_h, u)z_h, \varphi) = -((\tilde{y}_{v_1,h}, v_1^T)d''(\cdot, y_h, u)(\tilde{y}_{v_2,h}, v_2^T)^T, \varphi) \quad \forall \varphi \in V_h \tag{3.3.7}$$

with $\tilde{y}_{v_i,h} = G_h'(u)v_i$, $i = 1, 2$. The first and second order derivatives of f_h are given by

$$f_h'(u)v = \int_\Omega (\partial_u \Psi(x, y_h, u)v + \partial_y \Psi(x, y_h, u)\tilde{y}_h) \; dx \tag{3.3.8}$$

as well as

$$f_h''(u)[v_1, v_2] = \int_\Omega \left((\tilde{y}_{h,1}, v_1^T) \Psi''(x, y_h, u) (\tilde{y}_{h,2}, v_2^T)^T \right) \, dx. \tag{3.3.9}$$

Proof. This is analogous to the continuous setting and has also been used for semilinear elliptic control-constrained problems in [21]. □

Lemma 3.3.9. *Let $u_1, u_2 \in U_{ad}$ be arbitrary control vectors. For an arbitrary direction $v \in \mathbb{R}^m$ there exists a constant $C > 0$ independent of h and u_1, u_2, such that*

$$\|G_h(u_1) - G_h(u_2)\|_{H^1(\Omega)} \leq C|u_1 - u_2|$$

as well as

$$|f_h(u_1) - f_h(u_2)| \leq C|u_1 - u_2|$$

is satisfied.

Proof. This is shown analogously to the continuous setting. □

We will need an error estimate for the linearized state equation.

Proposition 3.3.10. *Let $u \in U_{ad}$, and $v \in \mathbb{R}^m$ be given and denote by $\tilde{y} := G'(u)v$ and $\tilde{y}_h := G_h'(u)v$ the solutions of the linearized state equations* (3.1.2) *and* (3.3.6). *Moreover, Ω_0 denotes the interior subdomain of Ω containing the set K where the state constraints are prescribed. Then the estimate*

$$\|\tilde{y} - \tilde{y}_h\|_{L^\infty(\Omega_0)} \leq ch^2 |\ln h| |v|$$

is fulfilled with a constant $c > 0$ not depending on h.

Proof. Noting that $\tilde{y}$ fulfills the equation

$$\begin{aligned} \mathcal{A}\tilde{y} + \partial_y d(\cdot, y, u)\tilde{y} &= -\partial_u d(\cdot, y, u)v, && \text{in } \Omega, \\ \tilde{y} &= 0 && \text{on } \Gamma, \end{aligned}$$

we define the auxiliary function $\tilde{z}$ as the unique weak solution of

$$\begin{aligned} \mathcal{A}\tilde{z} + \partial_y d(\cdot, y_h, u)\tilde{z} &= -\partial_u d(\cdot, y_h, u)v, && \text{in } \Omega, \\ \tilde{z} &= 0 && \text{on } \Gamma, \end{aligned}$$

with $y = G(u)$ and $y_h = G_h(u)$. Notice that $\partial_y d(\cdot, y, u)$ as well as $\partial_y d(\cdot, y_h, u)$ are bounded independently of h and u. We split the error into

$$\|\tilde{y} - \tilde{y}_h\|_{L^\infty(\Omega_0)} \leq \|\tilde{y} - \tilde{z}\|_{L^\infty(\Omega_0)} + \|\tilde{z} - \tilde{y}_h\|_{L^\infty(\Omega_0)}. \tag{3.3.10}$$

Then, it is clear that the first term in (3.3.10) accounts for the linearization of d at different states y and y_h, and that the second term in (3.3.10) is a pure discretization error for linear equations. The estimate

$$\|\tilde{z} - \tilde{y}_h\|_{L^\infty(\Omega_0)} \leq ch^2 |\ln h| |v| \tag{3.3.11}$$

follows by applying Theorem 3.3.3 or the results from [122] combined with the regularity and stability estimate for elliptic equations from Proposition 2.4.2. For the first term in (3.3.10), we observe that the difference $\tilde{y} - \tilde{z}$ fulfills the equation

$$\begin{aligned} \mathcal{A}(\tilde{y} - \tilde{z}) + \partial_y d(\cdot, y, u)(\tilde{y} - \tilde{z}) &= (\partial_y(\cdot, y_h, u) - \partial_y(\cdot, y, u))\tilde{z} + (\partial_u d(\cdot, y_h, u) - \partial_u d(\cdot, y, u))v \quad \text{in } \Omega, \\ \tilde{y} - \tilde{z} &= 0 \quad \text{on } \Gamma, \end{aligned}$$

which results in

$$\|\tilde{y} - \tilde{z}\|_{L^\infty(\Omega_0)} \leq \|\tilde{y} - \tilde{z}\|_{L^\infty(\Omega)} \leq c(\|y_h - y\|_{L^2(\Omega)})(\|\tilde{z}\|_{L^\infty(\Omega)} + |v|) \leq ch^2|v| \tag{3.3.12}$$

by Theorem 2.4.6, the Lipschitz continuity of $\partial_y d$ and $\partial_u d$, and the boundedness of U_{ad} as well as y and y_h. Combining (3.3.11) and (3.3.12) yields the assertion. □

Similar to the continuous problem, we reformulate the problem in a convenient way.

$$\text{Minimize } f_h(u) := J(G_h(u), u) \text{ subject to } u \in U_{\text{ad}}, \quad G_h(u)(x) \leq b \quad \forall x \in K, \tag{3.3.13}$$

where the state constraints are still prescribed in infinitely many points. It remains to discretize the pointwise state constraints. Let us motivate that we can simply prescribe the constraints for the discrete problem in the grid points contained in K and at the same time take into account the difficulties caused by the existence of local solutions in the next section.

3.3.3. Discretization of the state constraints

For a given locally optimal control $\bar{u} \in U_{\text{ad}}$ of Problem ($\mathbb{P}$) satisfying the first order necessary condition of Theorem 3.1.9 and the second order sufficient condition of Assumption (3.1.13), as well as the structural Assumption 3.2.1, we consider the auxiliary problems

$$\text{Minimize } f_h(u) \text{ subject to } u \in U^\varepsilon_{\text{ad}}, \quad G_h(u)(x_j) \leq b \quad \forall x_j \in \mathcal{N}_h^K, \tag{3.3.14}$$

as well as

$$\text{Minimize } f_h(u) \text{ subject to } u \in U^\varepsilon_{\text{ad}}, \quad G_h(u)(x) \leq b \quad \forall x \in K, \tag{3.3.15}$$

where

$$U^\varepsilon_{\text{ad}} := \{u \in U_{\text{ad}} \colon |u - \bar{u}| \leq \varepsilon\}.$$

Note that the idea of considering auxiliary problems comes originally from [26], where convergence of local solutions of purely control-constrained problems has been shown. We will argue that for ε small enough, Problem (3.3.14) and (3.3.15) are locally equivalent, i.e. it does not matter for local optimality whether the constraints are prescribed in all K or in the grid points contained in K, only. We define the auxiliary sets

$$U^{h,\varepsilon}_{\text{feas}} := \left\{u \in U^\varepsilon_{\text{ad}} \;:\; y_h(x_j) \leq b \quad \forall x_j \in \mathcal{N}_h^K\right\}$$

and

$$\widehat{U}^{h,\varepsilon}_{\text{feas}} := \{u \in U^\varepsilon_{\text{ad}} \;:\; y_h(x) \leq b \quad \forall x \in K\}$$

where $y_h = G_h(u)$.

Lemma 3.3.11. *For $\varepsilon > 0$ and h sufficiently small, $\widehat{U}^{h,\varepsilon}_{feas} = U^{h,\varepsilon}_{feas}$ is satisfied.*

Proof. Clearly, $\widehat{U}^{h,\varepsilon}_{\text{feas}} \subset U^{h,\varepsilon}_{\text{feas}}$ holds. We prove the inclusion $U^{h,\varepsilon}_{\text{feas}} \subset \widehat{U}^{h,\varepsilon}_{\text{feas}}$. Let therefore $v \in U^{h,\varepsilon}_{\text{feas}}$ be given and denote by

$$z := G(v), \quad z_h := G_h(v)$$

the state and discrete state associated with v, respectively. In any triangle T contained completely in K we know that $z_h(x) \le b$ for all $x \in T$ if and only if $z_h(x_j) \le b$ for all corners x_j of T, since z_h is cellwise linear and the bound b is a constant. The triangles $T \subset \Omega \setminus K$ do not have to be considered. All remaining triangles T intersect the boundary ∂K, and for h sufficiently small we can assume $T \subset \Omega_0$, since K is a compact subdomain of Ω_0. In these triangles, we then observe

$$\begin{aligned} z_h = G_h(v) &= G(\bar u) + G(v) - G(\bar u) + G_h(v) - G(v) \\ &\le G(\bar u) + \|G(v) - G(\bar u)\|_{L^\infty(T)} + \|G_h(v) - G(v)\|_{L^\infty(T)} \\ &\le G(\bar u) + \|G(v) - G(\bar u)\|_{L^\infty(\Omega_0)} + \|G_h(v) - G(v)\|_{L^\infty(\Omega_0)} \\ &\le G(\bar u) + c|v - \bar u| + ch^2 |\ln h|. \end{aligned} \tag{3.3.16}$$

Here we made use of the Lipschitz stability result (3.1.6), as well as the improved L^∞-error estimate for the state equation from Theorem 3.3.3. Now, recall that by Assumption 3.2.1 any active point $\bar x_j$ of $\bar y$ has a positive distance to the boundary of K. Then, for h small enough, a triangle T that intersects the boundary K will not intersect the neighborhoods $B_R(\bar x_j)$ of the active points, cf. Lemma 3.4.14, and we can assume that

$$T \cap K \subset K \setminus \bigcup_{j=1}^{n} B_R(\bar x_j).$$

Then, Lemma 3.4.14 ensures $G(\bar u)(x) = \bar y(x) \le b - \delta$ for some $\delta > 0$ and all $x \in T \cap K$ for the considered triangles. Consequently, (3.3.16) yields

$$z_h(x) \le b - \delta + c\varepsilon + ch^2 |\ln h| \le b - \delta/2$$

for h and ε sufficiently small. This implies $v \in \hat{U}^{h,\varepsilon}_{\text{feas}}$. Note that in $T \setminus K$ we find $z_h(x) \le b - \delta/4$ by the continuity of z_h, and hence even if constraints are imposed outside K due to possible nonconvexity, they will not become active. ☐

With the last result in mind, we formulate a completely discretized problem

$$\text{Minimize } f_h(u) := J(G_h(u), u) \text{ subject to } u \in U_{\text{ad}}, \quad G_h(u)(x_j) \le b \quad \forall x_j \in \mathcal{N}_h^K, \tag{$\mathbb{P}_h$}$$

for which we introduce the set of discrete feasible controls

$$U^h_{\text{feas}} := \left\{ u \in U_{\text{ad}} : y_h(x_j) \le b \quad \forall x_j \in \mathcal{N}_h^K \right\}.$$

As before, we introduce the notation of a local solution to $(\mathbb{P}_h)$.

Definition 3.3.12. *A feasible control $\bar u_h \in U^h_{feas}$ is called a local solution of $(\mathbb{P}_h)$, if there exists a positive real number ε such that $f_h(\bar u_h) \le f_h(u)$ holds for all feasible controls $u \in U^h_{feas}$ of $(\mathbb{P}_h)$ with $|u - \bar u_h| \le \varepsilon$.*

3.3.4. Intermediate convergence analysis

Our first convergence result can now be obtained by standard methods, since the discretization of the constraints does not cause any difficulties with respect to the feasibility of controls, as long as a sufficiently small neighborhood of $\bar{u}$ is considered. In the linear-quadratic setting from [105, 104], we have used the Slater point u_γ to construct auxiliary feasible controls that were used as test functions in the variational inequalities for the continuous and the discrete optimal control, cf. also [49] or [107]. Now, we construct feasible auxiliary functions and use standard arguments involving the quadratic growth condition in the neighborhood of $\bar{u}$, cf. also [143].

We will analyse Problem (3.3.15), i.e. we consider

$$\text{Minimize } f_h(u) \text{ subject to } u \in U_{\text{ad}}^\varepsilon, \quad G_h(u)(x) \le b \quad \forall x \in K, \tag{$\mathbb{P}_h^\varepsilon$}$$

where the short notation $(\mathbb{P}_h^\varepsilon)$ is justified by Lemma 3.3.11. Let us first provide some auxiliary results.

Lemma 3.3.13. *Let Assumption 3.1.1 hold, let $\bar{u}$ be a locally optimal control of $(\mathbb{P})$, and let $u_\gamma \in U_{ad}^\varepsilon$ be the Slater point from Assumption 3.1.7. There exists a sequence $\{u_{t(h)}\}_{t(h)}$ of controls that are feasible for $(\mathbb{P}_h^\varepsilon)$ for h and ε sufficiently small, and that converge to $\bar{u}$ with order $h^2|\ln h|$ as h tends to zero.*

Proof. The proof follows in a standard way. Consider

$$u_t := \bar{u} + t(u_\gamma - \bar{u})$$

with $t = t(h)$ tending to zero as h tends to zero. Obviously, $\{u_t\}_{t(h)}$ converges to $\bar{u}$ as h tends to zero, and the order of convergence is defined by $t(h)$. It is also clear that $|\bar{u} - u_t| \le \varepsilon$ if h is sufficiently small. To prove feasibility of u_t for $(\mathbb{P}_h^\varepsilon)$, note that

$$\begin{aligned} G_h(u_t) &= G(u_t) + G_h(u_t) - G(u_t) \\ &= (1-t)G(\bar{u}) + tG(\bar{u}) + tG'(u_\xi)(u_\gamma - \bar{u}) + G_h(u_t) - G(u_t) \\ &= (1-t)G(\bar{u}) + t(G(\bar{u}) + G'(\bar{u})(u_\gamma - \bar{u})) + t(G'(u_\xi) - G'(\bar{u}))(u_\gamma - \bar{u}) + G_h(u_t) - G(u_t) \end{aligned}$$

which follows by Taylor expansion of $G(u_t)$ at $\bar{u}$ with a $u_\xi = \bar{u} + \xi t(u_\gamma - \bar{u})$, $\xi \in (0,1)$. By the feasibility of $\bar{y} = G(\bar{u})$ for $(\mathbb{P})$, the Lipschitz continuity of G', (3.1.7), and the error estimate of Theorem 3.3.3, as well as the boundedness of $U_{\text{ad}}^\varepsilon$, we then obtain

$$\begin{aligned} G_h(u_t) &\le (1-t)b + t(b-\gamma) + c_1 t|u_\xi - \bar{u}||u_\gamma - \bar{u}| + c_2 h^2|\ln h| \\ &= b - t\gamma + c_1 t^2 \xi |u_\gamma - \bar{u}|^2 + c_2 h^2 |\ln h| \\ &\le b - t\gamma + c_1 t\varepsilon^2 + c_2 h^2|\ln h| \end{aligned}$$

for $0 < t, \xi < 1$ with the Slater point properties of Lemma 3.3.16 and $|u_\gamma - \bar{u}| \le \varepsilon$ by Remark 3.1.8. We obtain

$$G_h(u_t) \le b + t(c_1\varepsilon^2 - \gamma) + c_2 h^2 |\ln h|. \tag{3.3.17}$$

Choosing $t = t(h) = \frac{c_2 h^2 |\ln h|}{\gamma - c_1 \varepsilon^2} > 0$ for ε small enough yields

$$G_h(u_t) \le b \text{ in } K,$$

and obviously $t(h) = \mathcal{O}(h^2|\ln h|)$. □

Remark 3.3.14. *We verify that the proof of the last lemma remains valid using the Slater point u^ε_γ constructed in Remark 3.1.8 in case $|u_\gamma - \bar u| > \varepsilon$. We choose the sequence $u^\varepsilon_t := \bar u + t(h)(u^\varepsilon_\gamma - \bar u)$. Since $|u^\varepsilon_\gamma - \bar u| \le \varepsilon$, the only difference in the proof becomes obvious in equation* (3.3.17)*, which then reads*

$$G_h(u_t) \le b + t(c_1\varepsilon^2 - \gamma_\varepsilon) + c_2 h^2 |\ln h| = b + t(c_1\varepsilon^2 - c_3\varepsilon) + c_2 h^2|\ln h|$$

with $c_3 = \frac{1}{|u_\gamma - \bar u|}$. Obviously, the choice of $t(h) = \frac{c_2 h^2 |\ln h|}{c_3\varepsilon - c_1\varepsilon^2}$ yields a desired positive

$$t(h) = \mathcal{O}(h^2|\ln h|),$$

if ε is sufficiently small.

As a side effect of the last lemma, we can deduce a solvability result for Problem $(\mathbb{P}^\varepsilon_h)$.

Theorem 3.3.15. *Under Assumption 3.1.1, there exists at least one globally optimal control $\bar u^\varepsilon_h \in U_{ad}$ with associated discrete optimal state $\bar y^\varepsilon_h = G_h(\bar u^\varepsilon_h)$ for Problem $(\mathbb{P}^\varepsilon_h)$ for all ε sufficiently small.*

Proof. This follows by standard arguments, since Lemma 3.3.13 guarantees that the set $U^{h,\varepsilon}_{\text{feas}}$ is not empty. □

With the existence of an optimal control to Problem $(\mathbb{P}^\varepsilon_h)$ verified, the next step towards an error estimate is the construction of an auxiliary control sequence $\{u_{\tau(h)}\}_{\tau(h)} \subset U^{h,\varepsilon}_{\text{feas}}$ that converges to $\bar u^\varepsilon_h$. For that purpose, we prove that the Slater point u_γ from Assumption 3.1.7 is also a Slater point for the discrete problem.

Lemma 3.3.16. *Let Assumption 3.1.7 be satisfied. For all sufficiently small $\varepsilon, h > 0$ the Slater point u_γ from Assumption 3.1.7 satisfies*

$$G_h(\bar u^\varepsilon_h) + G'_h(\bar u^\varepsilon_h)(u_\gamma - \bar u^\varepsilon_h) \le b - \frac{\gamma}{2} \quad \text{in } K.$$

Proof. Due to Remark 3.1.8 we assume that $|u_\gamma - \bar u| \le \varepsilon$ is satisfied. The rest of the proof follows again by straight-forward calculations. In K, we obtain

$$\begin{aligned} &G_h(\bar u^\varepsilon_h) + G'_h(\bar u^\varepsilon_h)(u_\gamma - \bar u^\varepsilon_h) \\ &\quad = G(\bar u) + G'(\bar u)(u_\gamma - \bar u) + G_h(\bar u^\varepsilon_h) - G(\bar u^\varepsilon_h) + (G'_h(\bar u^\varepsilon_h) - G'(\bar u^\varepsilon_h))(u_\gamma - \bar u^\varepsilon_h) \\ &\qquad + G(\bar u^\varepsilon_h) - G(\bar u) - G'(\bar u)(\bar u^\varepsilon_h - \bar u) + (G'(\bar u^\varepsilon_h) - G'(\bar u))(u_\gamma - \bar u^\varepsilon_h). \end{aligned} \tag{3.3.18}$$

Now, we can estimate

$$G(\bar u) + G'(\bar u)(u_\gamma - \bar u) \le b - \gamma \tag{3.3.19}$$

by Assumption 3.1.7,

$$G_h(\bar u^\varepsilon_h) - G(\bar u^\varepsilon_h) + (G'_h(\bar u^\varepsilon_h) - G'(\bar u^\varepsilon_h))(u_\gamma - \bar u^\varepsilon_h) \le c_1 h^2 |\ln h| \tag{3.3.20}$$

by Theorem 3.3.3 and Proposition 3.3.10, as well as

$$G(\bar u^\varepsilon_h) - G(\bar u) - G'(\bar u)(\bar u^\varepsilon_h - \bar u) \le \frac{1}{2} G''(u_\xi)[\bar u^\varepsilon_h - \bar u, \bar u^\varepsilon_h - \bar u] \le c_2\varepsilon^2 \tag{3.3.21}$$

by Taylor expansion with $u_\xi = \bar{u} + \xi(\bar{u}_h^\varepsilon - \bar{u})$ for some $\xi \in (0,1)$, and finally

$$(G'(\bar{u}_h^\varepsilon) - G'(\bar{u}))(u_\gamma - \bar{u}_h^\varepsilon) \le c_3\varepsilon^2 \tag{3.3.22}$$

by Lemma 3.1.4. Insertion of (3.3.19)-(3.3.22) into (3.3.18) yields

$$G_h(\bar{u}_h^\varepsilon) + G_h'(\bar{u}_h^\varepsilon)(u_\gamma - \bar{u}_h^\varepsilon) \le b - \gamma + c_1 h^2|\ln h| + c_4\varepsilon^2 \le b - \frac{\gamma}{2} \tag{3.3.23}$$

for ε, h sufficiently small. Note again that using the Slater point u_γ^ε with parameter $\gamma_\varepsilon = c\varepsilon\gamma$ from Remark 3.1.8 does not cause any difficulties because γ_ε depends only linearly on ε. □

Corollary 3.3.17. *Let $\varepsilon > 0$ be given sufficiently small, let $\bar{u}$ be a locally optimal control of $(\mathbb{P})$, and let $\bar{u}_h^\varepsilon$ be any globally optimal control of $(\mathbb{P}_h^\varepsilon)$. Moreover, let $u_\gamma \in U_{ad}^\varepsilon$ be the Slater point from Assumption 3.1.7. There exists a sequence $\{u_{\tau(h)}\}_{\tau(h)}$ of controls that are feasible for $(\mathbb{P})$ and that converges to $\bar{u}_h^\varepsilon$ with order $h^2|\ln h|$ for all sufficiently small ε.*

Proof. The existence of $\{u_{\tau(h)}\}_{\tau(h)}$ follows as in Lemma 3.3.13, with a slight modification to avoid Lipschitz stability of G_h' in the $L^\infty(Q)$-norm. Consider

$$u_\tau := \bar{u}_h + \tau(u_\gamma - \bar{u}_h)$$

with $\tau = \tau(h)$ tending to zero as h tends to zero. Obviously, $\{u_\tau\}_{\tau(h)}$ converges to $\bar{u}_h$ as h tends to zero, and the order of convergence is defined by $\tau(h)$. To prove feasibility of u_τ for $(\mathbb{P})$, note that

$$\begin{aligned}
G(u_\tau) =& G_h(u_\tau) + G(u_\tau) - G_h(u_\tau)\\
=&(1-\tau)G_h(\bar{u}_h) + \tau G_h(\bar{u}_h) + \tau G_h'(u_\xi)(u_\gamma - \bar{u}_h) + G(u_\tau) - G_h(u_\tau)\\
=&(1-\tau)G_h(\bar{u}_h) + \tau(G_h(\bar{u}_h) + G_h'(\bar{u}_h)(u_\gamma - \bar{u}_h))\\
&+ \tau(G_h'(u_\xi) - G'(u_\xi) + G'(u_\xi) - G'(\bar{u}_h) + G'(\bar{u}_h) - G_h'(\bar{u}_h))(u_\gamma - \bar{u}_h) + G(u_\tau) - G_h(u_\tau)
\end{aligned}$$

which follows by Taylor expansion of $G_h(u_\tau)$ at $\bar{u}_h$ with a $u_\xi = \bar{u}_h + \xi(u_\gamma - \bar{u}_h)$, $\xi \in (0,1)$. Applying Proposition 3.3.10 to the terms $G'(u) - G_h'(u)$, $u = u_\xi, \bar{u}_h$, the rest of the proof is completely analogous to Lemma 3.3.13. □

We can now state our first convergence result and error estimate.

Lemma 3.3.18. *Let $\bar{u}$ be a locally optimal solution of Problem $(\mathbb{P})$ and let $\bar{u}_h^\varepsilon$ be any globally optimal control for $(\mathbb{P}_h^\varepsilon)$ with $\varepsilon > 0$ small enough such that the quadratic growth condition (3.1.22) as well as Lemmas 3.3.13, 3.3.16, and Corollary 3.3.17 hold. Then, there exists a constant $C > 0$ independent of h such that*

$$|\bar{u} - \bar{u}_h^\varepsilon| \le Ch\sqrt{|\ln h|}$$

for all sufficiently small $h > 0$.

Proof. The proof is in principle identical with the one in [143] for finitely many state constraints. Let $u_{t(h)} \in U_{\text{feas}}^{h,\varepsilon}$ be the control constructed in Lemma 3.3.13. Then $|\bar{u} - u_{t(h)}| \le ch^2|\ln h|$ holds. Since it is feasible for $(\mathbb{P}_h^\varepsilon)$ we know $f_h(\bar{u}_h^\varepsilon) \le f_h(u_t)$ due to the optimality of $\bar{u}_h^\varepsilon$. This can be estimated as follows:

$$f_h(\bar{u}_h^\varepsilon) \le f_h(u_t) \le |f_h(u_t) - f_h(\bar{u})| + |f_h(\bar{u}) - f(\bar{u})| + f(\bar{u}). \tag{3.3.24}$$

By the Lipschitz property of f_h from Lemma 3.3.9 as well as the properties of u_t from Lemma 3.3.13 we obtain

$$|f_h(u_t) - f_h(\bar{u})| \leq c|u_t - \bar{u}| \leq ch^2|\ln h|. \tag{3.3.25}$$

Moreover, Corollary 3.3.7 guarantees

$$|f_h(\bar{u}) - f(\bar{u})| \leq ch^2. \tag{3.3.26}$$

We insert (3.3.25) and (3.3.26) in (3.3.24) and can estimate the discrete optimal value against the continuous optimal value,

$$f_h(\bar{u}_h^\varepsilon) \leq f(\bar{u}) + c_1 h^2|\ln h|. \tag{3.3.27}$$

Now, we choose $v := u_\tau^h$ from Corollary 3.3.17 and insert it in the quadratic growth condition (3.1.22), which then reads

$$f(\bar{u}) \leq f(u_\tau^h) - \beta|u_\tau^h - \bar{u}|^2. \tag{3.3.28}$$

Moreover, using the Lipschitz property (3.1.9) for f as well as the properties of u_τ^h leads to

$$f(u_\tau^h) - f(\bar{u}_h^\varepsilon) \leq c|u_\tau^h - \bar{u}_h^\varepsilon| \leq c_2 h^2|\ln h|. \tag{3.3.29}$$

Combining (3.3.28) and (3.3.29), we deduce

$$f(\bar{u}) \leq f(u_\tau^h) - \beta|u_\tau^h - \bar{u}|^2 \leq f(\bar{u}_h^\varepsilon) + c_2 h^2|\ln h| - \beta|\bar{u}_h^\varepsilon - \bar{u}|^2, \tag{3.3.30}$$

which yields

$$\beta|\bar{u} - \bar{u}_h^\varepsilon|^2 \leq f(\bar{u}_h^\varepsilon) - f(\bar{u}) + c_2 h^2|\ln h|. \tag{3.3.31}$$

Finally, combining (3.3.31) and (3.3.27) leads to the estimate

$$\beta|\bar{u} - \bar{u}_h^\varepsilon|^2 \leq c_1 h^2|\ln h| + c_2 h^2|\ln h|, \tag{3.3.32}$$

which implies the assertion. □

3.4. Convergence analysis for the discrete problem

In this section, we prove the main result of this chapter, i.e. we provide an a priori estimate for the error between a local solution $\bar{u}$ of Problem $(\mathbb{P})$ and an associated discrete local solution $\bar{u}_h$ of Problem $(\mathbb{P}_h)$. Adapting the steps from the linear-quadratic setting in [104, 105], we will divide this into three steps. At first, we collect the results available for the auxiliary problems in the last section and apply them to $(\mathbb{P}_h)$. In a second step, we make use of this intermediate convergence result to derive some properties of the discrete problem that will be used in our final error estimate. In particular, we characterize the discrete optimal state as well as the location of discrete active points. Third and last, we discuss if and how the intermediate error estimate can be improved.

3.4.1. A convergence result of order $h\sqrt{|\ln h|}$

With the convergence error estimate from Lemma 3.3.18 it is clear that $\bar{u}_h^\varepsilon$ converges to $\bar{u}$ as h tends to zero. Consequently, $\bar{u}_h^\varepsilon$ is not situated at the boundary of $U_{\mathrm{ad}}^\varepsilon$, which implies that it is a local solution of $(\mathbb{P}_h)$. We can formulate our first convergence result for the discrete problem formulation.

Theorem 3.4.1. *Let $\bar{u}$ be a locally optimal solution of Problem $(\mathbb{P})$ fulfilling the first order necessary optimality conditions of Theorem 3.1.9 and let Assumptions 3.1.1-3.2.1 hold. There exists a sequence $\{\bar{u}_h\}$ of locally optimal controls for $(\mathbb{P}_h)$ that converge to $\bar{u}$ as h tends to zero. There is a constant $C > 0$ independent of h, such that*

$$|\bar{u} - \bar{u}_h| \leq Ch\sqrt{|\ln h|}$$

holds for all sufficiently small $h > 0$.

Proof. This is a direct consequence of Lemma 3.3.18. □

At this point, we know that the locally optimal controls of Problem $(\mathbb{P})$ can be approximated by local solutions of the discrete problem $(\mathbb{P}_h)$. Note that this result could even be obtained under less restrictive conditions than Assumption 3.2.1, where we required the active set to consist of only finitely many active points in the interior of K. We have used this assumption to ensure the equivalence of the auxiliary problems (3.3.14) and (3.3.15) and to be able to use the improved L^∞-error estimate for the states from 3.3.3. Yet, this could also be obtained for more general active sets as long as they have a positive distance to the boundary ∂K so that the improved error estimate (3.3.5) can be applied.

The question remains, whether or not the error estimate of Theorem 3.4.1 can be improved to reflect the order of the error due to discretization of the state equation of Theorem 3.3.3. Before we aim at improving the error estimate, let us state the discrete first order optimality system, which is now easily obtained due to Lemma 3.3.16. For brevity, we agree on the following:

Assumption 3.4.2. *Throughout the following, let $\{\bar{u}_h\}$ denote the sequence of discrete locally optimal controls from Theorem 3.4.1 approximating $\bar{u}$ with associated states $\{\bar{y}_h\}$.*

Proposition 3.4.3. *Let Assumptions 3.1.1-3.4.2 be satisfied and let $h > 0$ be sufficiently small. Moreover, we denote by $N_p = \sharp\mathcal{N}_h^K$ the number of grid points contained in K. Then there exists a regular Borel measure $\bar{\mu}_h \in \mathcal{M}(K)$ that can be represented by*

$$\bar{\mu}_h = \sum_{k=1}^{N_p} \bar{\mu}_{k,h}\delta_{x_k}, \quad \bar{\mu}_{k,h} \in \mathbb{R}, \quad k = 1, \ldots, N_p$$

and an adjoint state $\bar{p}_h \in W_0^{1,s}(\Omega)$, $1 \leq s < 2$ such that the optimal control $\bar{u}_h$ of $(\mathbb{P}_h)$ with associated discrete state $\bar{y}_h$ fulfills

$$a(\varphi, \bar{p}_h) + (\partial_y d(\cdot, \bar{y}_h, \bar{u}_h)p_h, \varphi) = (\partial_y \Psi(\cdot, \bar{y}_h, \bar{p}_h), \varphi) + \int_K \varphi d(\bar{\mu}_h), \quad \forall \varphi \in V_h, \tag{3.4.1}$$

$$\langle g_h, u - \bar{u}_h \rangle \geq 0 \quad \forall u \in U_{ad}, \tag{3.4.2}$$

where $g_h \in \mathbb{R}^m$ is defined by $g_h = (g_{h,i})$ with

$$g_{h,i} := \int_\Omega (\partial_{u_i}\psi(x, \bar{y}_h, \bar{u}_h) - \bar{p}_h \partial_{u_i} d(x, \bar{y}_h, \bar{u}_h))\, dx, \quad i = 1, \ldots, m,$$

as well as the complementary slackness condition

$$\int_K (\bar{y}_h - b)\, d\bar{\mu}_h = 0, \quad \bar{\mu}_h \geq 0. \tag{3.4.3}$$

Proof. This follows as in [18] or [2], where semilinear problems with Robin or Neumann-type boundary conditions have been considered. A linear-quadratic problem with Dirichlet boundary conditions and control functions has been taken into account in [17]. □

Remark 3.4.4. *In all what follows, we identify $\bar{\mu}_h$ with the associated N_p-vector of real numbers $\bar{\mu}_{k,h}$, $k = 1, \dots, N_p$.*

Again, a formulation with Lagrange multipliers associated with the control constraints is easily obtained. If the tuple $(\bar{u}_h, \bar{y}_h, \bar{p}_h, \bar{\mu}_h) \in U_{\mathrm{ad}} \times V_h \times W^{1,s}(\Omega) \times \mathcal{M}(K)$ fulfills the setting of Proposition 3.4.3, there exist nonnegative Lagrange multipliers $\bar{\eta}_h^a, \bar{\eta}_h^b \in \mathbb{R}^m$ such that the variational inequality (3.4.2) can equivalently be written

$$g_{h,i} + \bar{\eta}_{h,i}^b - \bar{\eta}_{h,i}^a = 0 \tag{3.4.4}$$

and

$$\langle u_a - \bar{u}_h, \bar{\eta}_h^a \rangle = \langle \bar{u}_h - u_b, \bar{\eta}_h^b \rangle = 0, \tag{3.4.5}$$

cf. again [144]. Alternatively, the first order necessary optimality conditions of Proposition 3.4.3 can be expressed with the help of the discrete Lagrangian

$$\mathcal{L}_h \colon U \times \mathcal{M}(K) \to \mathbb{R}, \quad \mathcal{L}_h(u, \mu_h) := f_h(u) + \iint_K (G_h(u) - b)\, d\mu_h(x).$$

As an equivalent to Proposition 3.4.3 we then obtain

$$\mathcal{L}_h'(\bar{u}_h, \bar{\mu}_h)(u - \bar{u}_h) \geq 0 \quad \forall\, u \in U_{\mathrm{ad}}, \tag{3.4.6}$$

or

$$\mathcal{L}_h'(\bar{u}_h, \bar{\mu}_h) + \bar{\eta}_h^b - \bar{\eta}_h^a = 0, \tag{3.4.7}$$

combined with the complementary slackness condition (3.4.3), and (3.4.5), respectively, for the nonnegative Lagrange multipliers $\bar{\mu}_h, \bar{\eta}_h^a, \bar{\eta}_h^b$, as in the continuous setting.

3.4.2. Some properties of the discrete problem

In this section, we will provide all technical results that are necessary to improve the intermediate error estimate. The steps from the linear-quadratic setting in e.g. [105] need only minor modifications due to the nonlinear control-to-state whereever a derivative of the control-to-state operator is used. The overall goal of this section is to provide an estimate for the distance of discrete and continuous active points. This will be a key argument in the later proof of convergence, and it is also the main difference to the completely finite dimensional problem in [106], where the constraints were considered in fixed points. Eventually, we will make use of the fact that the active points on the discrete and the continuous level have a similar distance as the discrete and continuous controls. This will be useful in Taylor-type arguments. In Section 3.2.2, we have already discussed some structural properties of the continuous optimal state $\bar{y}$. With the convergence result of Theorem 3.4.1 let us now discuss the structure of the discrete state.

Corollary 3.4.5. *Let Assumption 3.4.2 hold. The discrete state $\bar{y}_h$ converges uniformly to the associated state $\bar{y}$ as h tends to zero.*

Proof. This is obtained from

$$\|G_h(\bar{u}_h) - G(\bar{u})\|_\infty \le \|G(\bar{u}_h) - G(\bar{u})\|_\infty + \|G_h(\bar{u}_h) - G(\bar{u}_h)\|_\infty \le c|\bar{u}_h - \bar{u}| + ch \le ch\sqrt{|\ln h|},$$

where we applied the Lipschitz result (3.1.6), as well as Theorems 3.3.3 and 3.4.1. □

To obtain a convergence result in the whole domain Ω we have used the lower order L^∞-error estimate from Theorem 3.3.3. A similar result holds for the linearized states.

Corollary 3.4.6. *Let Assumption 3.4.2 be satisfied and let $v \in \mathbb{R}^m$ be given. The linearized discrete state $\tilde{y}_h := G_h'(\bar{u}_h)v$ converges uniformly to the linearized state $\tilde{y} := G'(\bar{u})v$ in $L^\infty(\Omega)$.*

Proof. This follows easily from

$$\|G_h'(\bar{u}_h)v - G'(\bar{u})v\|_\infty \le \|G_h'(\bar{u}_h)v - G'(\bar{u}_h)v\|_\infty + \|G'(\bar{u}_h)v - G'(\bar{u})v\|_\infty.$$

Making use of the error estimate for linearized equations from Proposition 3.3.10 allows to estimate

$$\|G_h'(\bar{u}_h)v - G'(\bar{u}_h)v\|_\infty \le ch,$$

and the Lipschitz result of Lemma 3.1.4 combined with the error estimate for the control from Theorem 3.4.1 leads to

$$\|G'(\bar{u}_h)v - G'(\bar{u})v\|_\infty \le c\,|u_h - \bar{u}| \le ch\sqrt{|\ln h|},$$

Combining all estimates yields the assertion. □

We proceed in several steps, beginning with the proof of a suboptimal estimate for the distance between the discrete and continuous active points. Before, we define an auxiliary state that will come in helpful in the further discussion.

Definition 3.4.7. *Let $\bar{u}_h$ be as stated in Assumption 3.4.2. Then we define the auxiliary state*

$$\tilde{y}_h := G(\bar{u}_h).$$

Lemma 3.4.8. *Let $\bar{u}_h$ be as stated in Assumption 3.4.2. In particular, let Assumption 3.2.1 hold, that guarantees the active sets to consist of finitely many points, only. There exists $h_0 > 0$ such that, for all $h \le h_0$, we have*

$$\bar{y}_h(x) \le b - \delta/2 \quad \forall x \in K \setminus \bigcup_{j=1}^{n} B_R(\bar{x}_j), \tag{3.4.8}$$

where δ, R are chosen as in the analogous Lemma 3.4.14 on the continuous level. Moreover, if $\bar{x}_j^h \in B_R(\bar{x}_j)$ is an active point of an optimal state $\bar{y}_h$ of $(\mathbb{P}_h)$, there exists a constant $c > 0$ such that

$$|\bar{x}_j^h - \bar{x}_j| \le ch^{\frac{1}{2}}|\ln h|^{\frac{1}{4}}. \tag{3.4.9}$$

Proof. Inequality (3.4.8) is a direct consequence of the uniform convergence stated in Corollary 3.4.5 and Assumption 3.2.1 on the structure of the active set. This implies in particular that the discrete state can only be active in a neighborhood of the continuous active points $\bar{x}_j$, $j = 1, \ldots, n$, with finite radius R. Assume therefore $\bar{x}_j^h \in B_R(\bar{x}_j)$ to be an active point of $(\mathbb{P}_h)$, i.e. $\bar{y}_h(\bar{x}_j^h) = b$. We observe

$$\bar{y}_h(\bar{x}_j^h) = \bar{y}(\bar{x}_j^h) + (\tilde{y}_h - \bar{y})(\bar{x}_j^h) + (\bar{y}_h - \tilde{y}_h)(\bar{x}_j^h), \tag{3.4.10}$$

where $\tilde{y}_h$ is defined in Definition 3.4.7. Then, by the Lipschitz stability result (3.1.6) for the continuous control-to-state operator G and the error estimate of Theorem 3.4.1, we obtain

$$\|\tilde{y}_h - \bar{y}\|_\infty = \|G(\bar{u}_h) - G(\bar{u})\|_\infty \leq c|\bar{u}_h - \bar{u}| \leq ch\sqrt{|\ln h|}. \tag{3.4.11}$$

In addition, Theorem 3.3.3 guarantees

$$\|\bar{y}_h - \tilde{y}_h\|_{\infty,K} = \|G_h(\bar{u}_h) - G(\bar{u}_h)\|_{\infty,K} \leq ch^2|\ln h| \tag{3.4.12}$$

in K. Inserting both estimates into (3.4.10) delivers

$$\bar{y}_h(\bar{x}_j^h) \leq \bar{y}(\bar{x}_j^h) + ch\sqrt{|\ln h|} + ch^2|\ln h|. \tag{3.4.13}$$

Finally, applying (3.2.2) to estimate $\bar{y}(\bar{x}_j^h)$ in (3.4.13) leads to

$$\bar{y}_h(\bar{x}_j^h) \leq b - \frac{\omega}{4}|\bar{x}_j^h - \bar{x}_j|^2 + ch\sqrt{|\ln h|}, \tag{3.4.14}$$

and with $\bar{y}_h(\bar{x}_j^h) = b$, this can be transformed to

$$|\bar{x}_j - \bar{x}_j^h| \leq ch^{\frac{1}{2}}|\ln h|^{\frac{1}{4}},$$

and the assertion is obtained. $\square$

In the following, our main goal is to improve this suboptimal estimate. This involves distance estimates on two levels. First, we compare active points of $\bar{y}$ (or maxima of $\bar{y}$) with maxima of $\tilde{y}_h$. Noting that $\bar{y} = G(\bar{u})$ and $\tilde{y}_h = G(\bar{u}_h)$ makes clear that we only take into account the influence of the controls. In a second step, we will estimate the distance between active points of $\tilde{y}_h$ and $\bar{y}_h$, which estimates the influence of the discretization of the state equation.

Lemma 3.4.9. *For each active point $\bar{x}_j$, $j = 1, \ldots, n$, of $\bar{y} = G(\bar{u})$, there exists a unique local maximum $\tilde{x}_j^h \in B_R(\bar{x}_j)$ of $\tilde{y}_h = G(\bar{u}_h)$. Moreover, the estimate*

$$|\bar{x}_j - \tilde{x}_j^h| \leq Ch\sqrt{|\ln h|}$$

is satisfied for a constant $C > 0$ independent of h.

Proof. For a control $u \in U$ we denote by $y = G(u)$ the associated state on the continuous level, and define the function

$$F(x, u) := \nabla y(x).$$

Notice that $F(x, u) = 0$ describes a necessary condition for a local maximum of y at x. Hence, we observe that $F(\bar{x}_j, \bar{u}) = 0$ holds for all $j = 1, \ldots, n$, since due to Assumption 3.2.1, $\bar{y}$ admits a local maximum in $\bar{x}_j$ for all $j = 1, \ldots, n$. Notice further that by the same assumption the matrix

$$\frac{\partial F}{\partial x}(\bar{x}_j, \bar{u}) = \nabla^2 \bar{y}(\bar{x}_j)$$

is not singular. This allows to apply the implicit function theorem, and we obtain the existence of constants $r, \tau, c > 0$ such that for all $u \in U$ with $|u - \bar{u}| \leq \tau$, there exists a unique $\tilde{x}_j(u) \in B_r(\bar{x}_j)$ fulfilling

$$F(\tilde{x}_j(u), u) = 0, \quad \text{as well as} \quad |\tilde{x}_j(u) - \bar{x}_j| \leq c|u - \bar{u}|.$$

Applying this to $u := \bar{u}_h$ yields the existence of $\tilde{x}_j^h := \tilde{x}_j(\bar{u}_h)$ with

$$|\tilde{x}_j^h - \bar{x}_j| \leq ch\sqrt{|\ln h|}$$

by the convergence result of Theorem 3.4.1. For h small enough, we then have $\tilde{x}_j^h \in B_R(\bar{x}_j)$. It remains to prove that $\tilde{x}_j^h$ is indeed a local maximum of $\tilde{y}_h$. We estimate the Hessian matrix of $\tilde{y}_h$ at $\tilde{x}_j^h$ by

$$-\nabla^2 \tilde{y}_h(\tilde{x}_j^h) = -\nabla^2 \bar{y}(\tilde{x}_j^h) - \nabla^2(\tilde{y}_h - \bar{y})(\tilde{x}_j^h) = -\nabla^2 \bar{y}(\tilde{x}_j^h) - \nabla^2 z_{u_\xi}(\tilde{x}_j^h) \tag{3.4.15}$$

with $z_{u_\xi} = G'(u_\xi)(\bar{u}_h - \bar{u})$, $u_\xi = \bar{u} + \xi(\bar{u}_h - \bar{u})$, $\xi \in (0,1)$. Multiplying (3.4.15) from the left and right-hand side by $\theta \in \mathbb{R}^2$ yields

$$-\langle \theta, \nabla^2 \tilde{y}_h(\tilde{x}_j^h)\theta \rangle = -\langle \theta, \nabla^2 \bar{y}(\tilde{x}_j^h)\theta \rangle - \langle \theta, \nabla^2 z_{u_\xi}(\tilde{x}_j^h)\theta \rangle. \tag{3.4.16}$$

Now, the first item in (3.4.16) can be estimated by

$$\omega|\theta|^2 \leq -\langle \theta, \nabla^2 \bar{y}(\tilde{x}_j^h)\theta \rangle \tag{3.4.17}$$

by inequality (3.2.1). Moreover, from Lemma 3.4.9 we know that $\tilde{x}_j^h$ tends to $\bar{x}_j$ as h tends to zero, from which we conclude that $\nabla^2 z_{u_\xi}(\tilde{x}_j^h)$ is bounded. Hence, with the convergence result of Theorem 3.4.1, we obtain for the second term in (3.4.16) that it can be estimated by

$$-ch\sqrt{|\ln h|}|\theta|^2 \leq \langle \theta, \nabla^2 z_{u_\xi}(\tilde{x}_j^h)\theta \rangle. \tag{3.4.18}$$

Inserting both estimates (3.4.17) and (3.4.18) in (3.4.16) yields

$$-\langle \theta, \nabla^2 \tilde{y}_h(\tilde{x}_j^h)\theta \rangle \geq (\omega - ch\sqrt{|\ln h|})|\theta|^2 \geq \frac{\omega}{2}|\theta|^2 \tag{3.4.19}$$

for h sufficiently small. This implies coercivity of the Hessian matrix $-\nabla^2 \tilde{y}_h(\tilde{x}_j^h)$ so that $\tilde{y}_h$ admits a strict local maximum in $\tilde{x}_j^h$. Thanks to the coercivity derived above, there is a small ball around $\tilde{x}_j^h$ such that this local maximum is unique in this ball. Without limitation of generality we can assume that in Lemma 3.4.8 R was taken small enough for this to hold in $B_R(\bar{x}_j)$. □

We proceed by proving some properties of the auxiliary function $\tilde{y}_h$, that are of the same structure as the ones of $\bar{y}$ and $\bar{y}_h$.

Lemma 3.4.10. *There exist positive real numbers r and C such that for all sufficiently small h the auxiliary function $\tilde{y}_h$ defined in Lemma 3.4.9 satisfies the following properties:*

$$\tilde{y}_h(x) \leq b + Ch^2|\ln h| \quad \text{for all } x \in K \tag{3.4.20}$$

$$\tilde{y}_h(x) \leq b - \frac{\delta}{2} \quad \text{for all } x \in K \setminus \bigcup_{j=1}^n B_R(\bar{x}_j) \tag{3.4.21}$$

$$\tilde{y}_h(x) \leq \tilde{y}_h(\tilde{x}_j^h) - \frac{\omega}{8}|x - \tilde{x}_j^h|^2 \quad \text{for all } x \in B_r(\bar{x}_j),\ j = 1, \ldots, n. \tag{3.4.22}$$

Proof. The proof is straight-forward. To prove (3.4.20), we observe that

$$\tilde{y}_h(x) = \bar{y}_h(x) + (\tilde{y}_h(x) - \bar{y}_h(x)) \le b + \|G(\bar{u}_h) - G_h(\bar{u}_h)\|_{\infty,K} \le b + ch^2|\ln h|$$

by the feasibility of $\bar{y}_h$ and the accuracy result from Theorem 3.3.3. From Theorem 3.4.1 we already know that $\bar{u}_h$ converges to $\bar{u}$ as h tends to zero. Then, by the Lipschitz continuity of G in $L^\infty(\Omega)$ we know that $\tilde{y}_h$ converges uniformly towards $\bar{y}$ as h tends to zero. We immediately obtain (3.4.21) as an analogue to (3.2.3) and (3.4.8). The last inequality is obtained by a Taylor expansion of $\tilde{y}_h$ in $\tilde{x}_j^h$, given by

$$\begin{aligned}\tilde{y}_h(x) &= \tilde{y}_h(\tilde{x}_j^h) + \nabla\tilde{y}_h(\tilde{x}_j^h)(x-\tilde{x}_j^h) + \frac{1}{2}\langle x-\tilde{x}_j^h, \nabla^2\tilde{y}_h(x_j^\xi)(x-\tilde{x}_j^h)\rangle \\ &\le \tilde{y}_h(\tilde{x}_j^h) + \frac{1}{2}\langle x-\tilde{x}_j^h, \nabla^2\tilde{y}_h(\tilde{x}_j^h)(x-\tilde{x}_j^h)\rangle + \frac{1}{2}\langle x-\tilde{x}_j^h, (\nabla^2\tilde{y}_h(x_\xi) - \nabla^2\tilde{y}_h(\tilde{x}_j^h))(x-\tilde{x}_j^h)\rangle\end{aligned}$$

with some $x_j^\xi = \tilde{x}_j^h + \xi(x - \tilde{x}_j^h)$, $\xi \in (0,1)$, since $\nabla\tilde{y}_h(\tilde{x}_j^h) = 0$. The Hölder continuity of $\nabla^2\tilde{y}_h$ due to Theorem 2.4.6 and the coercivity of $-\nabla^2\bar{y}(\bar{x}_j^h)$ from inequality (3.4.19) imply

$$\tilde{y}_h(x) \le \tilde{y}_h(\tilde{x}_j^h) - \frac{\omega}{4}|x-\tilde{x}_j^h|^2 + \frac{c}{2}|x_\xi - \tilde{x}_j^h|^\nu |x-\tilde{x}_j^h|^2$$

with some $0 < \nu < 1$. Hence, with $|x_\xi - \tilde{x}_j^h| = \xi|x-\tilde{x}_j^h|$, we obtain the existence of a real number $r > 0$ not depending on h such that

$$\tilde{y}_h(x) \le \tilde{y}_h(\tilde{x}_j^h) - \frac{\omega}{8}|x-\tilde{x}_j^h|^2$$

is satisfied if $|x - \tilde{x}_j^h| \le r$ holds. □

With these properties of $\tilde{y}_h$ at hand, we can provide an estimate for the distance $|\tilde{x}_j^h - \bar{x}_j^h|$.

Lemma 3.4.11. *Assume that there exists a discrete active point $\bar{x}_j^h \in B_R(\bar{x}_j)$. Let $\tilde{x}_j^h$ be an associated maximum of $\tilde{y}_h$. There exists a constant $C > 0$ such that*

$$|\bar{x}_j^h - \tilde{x}_j^h| \le Ch\sqrt{|\ln h|}.$$

Proof. By Lemmas 3.4.8 and 3.4.9 we find that

$$|\bar{x}_j^h - \tilde{x}_j^h| \le |\bar{x}_j^h - \bar{x}_j| + |\bar{x}_j - \tilde{x}_j^h| \le ch^{\frac{1}{2}}|\ln h|^{\frac{1}{4}}$$

is satisfied. This suboptimal estimate guarantees that $\bar{x}_j^h \in B_r(\tilde{x}_j^h)$ for h small enough, i.e.

$$\tilde{y}_h(\bar{x}_j^h) \le \tilde{y}_h(\tilde{x}_j^h) - \frac{\omega}{8}|\bar{x}_j^h - \tilde{x}_j^h|^2. \tag{3.4.23}$$

We also observe that

$$b = \bar{y}_h(\bar{x}_j^h) = \tilde{y}_h(\bar{x}_j^h) + (\bar{y}_h - \tilde{y}_h)(\bar{x}_j^h) \le \tilde{y}_h(\bar{x}_j^h) + ch^2|\ln h|$$

by Theorem 3.3.3. From that, we conclude that for a discrete active point $\bar{x}_j^h$

$$\tilde{y}_h(\bar{x}_j^h) \ge b - ch^2|\ln h| \tag{3.4.24}$$

must hold. By the uniform estimate

$$\tilde{y}_h(x) \le b - \delta/2 \quad \forall x \in K \setminus \bigcup_{j=1}^n B_R(\bar{x}_j),$$

stated in Lemma 3.4.10, this can only be true inside the balls $B_R(\bar{x}_j)$. Hence, combining (3.4.23) and (3.4.24) gives a necessary condition for activity of the discrete state $\bar{y}_h$ at $\bar{x}_j^h$, which reads

$$b - ch^2|\ln h| \le \tilde{y}_h(\bar{x}_j^h) \le \tilde{y}_h(\tilde{x}_j^h) - \frac{\omega}{8}|\bar{x}_j^h - \tilde{x}_j^h|^2.$$

This can be transferred into

$$|\bar{x}_j^h - \tilde{x}_j^h|^2 \le ch^2|\ln h| + \frac{8}{\omega}(\tilde{y}_h(\tilde{x}_j^h) - b) \le ch^2|\ln h| + \frac{8}{\omega}\left(\bar{y}_h(\tilde{x}_j^h) - b + (\tilde{y}_h - \bar{y}_h)(\tilde{x}_j^h)\right) \le ch^2|\ln h|$$

where the last inequality follows from $\bar{y}_h(\tilde{x}_j^h) \le b$ and Theorem 3.3.3. □

We now directly obtain an improved estimate for the distance $|\bar{x}_j - \bar{x}_j^h|$:

Corollary 3.4.12. *For any discrete active point $\bar{x}_j^h \in B_R(\bar{x}_j)$ we obtain the distance estimate*

$$|\bar{x}_j - \bar{x}_j^h| \le Ch\sqrt{\ln h}$$

for some $C > 0$.

Proof. By Lemmas 3.4.9 and 3.4.10 this follows from $|\bar{x}_j - \bar{x}_j^h| \le |\bar{x}_j - \tilde{x}_j^h| + |\tilde{x}_j^h - \bar{x}_j^h|$. □

Note that neither the existence of a discrete active point $\bar{x}_j^h$ in the neighborhood of $\bar{x}_j$, nor its uniqueness is guaranteed at this time. The strong activity property will provide existence of active points of $\bar{y}_h$ in the neighborhoods of all strongly active points $\bar{x}_j$, $j \in \mathcal{A}_{\bar{y}}$, for all sufficiently small h. To prepare this statement, we proceed with an intermediate result.

Lemma 3.4.13. *Let $\{h_n\}_{n\in\mathbb{N}}$ be a sequence of positive real numbers converging to zero as n tends to infinity, and let $\{\bar{u}_h\}$ be the sequence of locally optimal solutions of $(\mathbb{P}_h)$ from Assumption 3.4.2. Any sequence $\{\bar{\mu}_{h_n}\}_{n\in\mathbb{N}}$ of Lagrange multipliers for $(\mathbb{P}_h)$ is bounded in $\mathcal{M}(K)$.*

Proof. This is a standard conclusion from the Slater condition. The proof is given for convenience. We have already pointed out in Lemma 3.3.16 that the Slater point u_γ is also a Slater point for the optimal control of Problem $(\mathbb{P}_h^\varepsilon)$ and by the intermediate convergence result also for $(\mathbb{P}_h)$. For simplicity, we omit the subscript n in h_n and the associated optimal controls, states, and Lagrange multipliers. Inserting $u_\gamma \in U_{ad}$ into the variational inequality (3.4.6) for $\bar{u}_h$ and making use of the complementary slackness condition (3.4.3) and the Slater point property of Lemma 3.3.16, we find

$$\begin{aligned}
0 &\le \mathcal{L}_h'(\bar{u}_h, \bar{\mu}_h)(u_\gamma - \bar{u}_h) \\
&= f_h'(\bar{u}_h)(u_\gamma - \bar{u}_h) + \int_K G_h'(\bar{u}_h)(u_\gamma - \bar{u}_h)d\bar{\mu}_h(x) \\
&= f_h'(\bar{u}_h)(u_\gamma - \bar{u}_h) + \int_K (G_h(\bar{u}_h) + G_h'(\bar{u}_h)(u_\gamma - \bar{u}_h) - b)d\bar{\mu}_h(x) + \int_K (b - G_h(\bar{u}_h))d\bar{\mu}_h(x) \\
&= f_h'(\bar{u}_h)(u_\gamma - \bar{u}_h) + \int_K (G_h(\bar{u}_h) + G_h'(\bar{u}_h)(u_\gamma - \bar{u}_h) - b)d\bar{\mu}_h(x) \\
&\le f_h'(\bar{u}_h)(u_\gamma - \bar{u}_h) - \frac{\gamma}{2}\int_K 1 \cdot d\bar{\mu}_h(x)
\end{aligned}$$

for all h sufficiently small, which yields

$$\frac{\gamma}{2}\int_K 1 \cdot d\bar{\mu}_h(x) \leq f_h'(\bar{u}_h)(u_\gamma - \bar{u}_h).$$

Moreover, the sequence $\{\bar{u}_h\}$ is bounded as h tends to zero. This follows directly from the boundedness of U_{ad}. Therefore, we obtain

$$\|\bar{\mu}_h\|_{\mathcal{M}(K)} = \int_K d\bar{\mu}_h(x) \leq \frac{c}{\gamma}.$$

$\square$

Lemma 3.4.14. *Let Assumptions 3.1.1-3.4.2 hold. For any $j \in \mathcal{A}_{\bar{y}}$, there exists a constant $C > 0$ and at least one point $\bar{x}_j^h \in B_R(\bar{x}_j)$ of Problem $(\mathbb{P}_h)$ where $\bar{y}_h$ is active, i.e. $\bar{y}_h(\bar{x}_j^h) = b$, with*

$$|\bar{x}_j - \bar{x}_j^h| \leq Ch\sqrt{|\ln h|}. \tag{3.4.25}$$

Proof. Let $\{h_n\} > 0$ be a sequence of mesh sizes converging to zero, and denote by $\{\bar{u}_n\}$, $\{\bar{y}_n\}$, and $\{\bar{\mu}_n\}$ the sequences of discrete optimal control, state, and an associated Lagrange multiplier associated with h_n in a neighborhood of $\bar{u}$, cf. Assumption 3.4.2. Now let us assume the contrary: then, there exists an index $j \in \mathcal{A}_{\bar{y}}$ and for all n a positive $h_n < \frac{1}{n}$ such that $\bar{y}_n(x) := \bar{y}^{h_n}(x) < b$ holds for all $x \in K \cap B_R(\bar{x}_j)$. Consequently, we can assume

$$\bar{y}_n(x) < b \quad \forall\, x \in B_R(\bar{x}_j),$$

and by the complementary slackness condition (3.4.3), we have

$$\bar{\mu}_{n_{|B_R(\bar{x}_j)}} = 0 \tag{3.4.26}$$

for all n. By Lemma 3.4.13, the sequence $\{\bar{\mu}_n\}$ is bounded in $\mathcal{M}(K)$. Therefore, we can select a subsequence converging weakly* to some $\hat{\mu} \in \mathcal{M}(K)$. Let, w.l.o.g., $\{\bar{\mu}_n\}$ be this sub-sequence. Moreover, we already know that $\bar{u}_n \to \bar{u}$ in $\mathbb{R}^m$ and

$$G_h'(\bar{u}_n)(u - \bar{u}_n) \to G'(\bar{u})(u - \bar{u})$$

in $\mathcal{C}(K)$ by Corollary 3.4.6. We now verify that $\hat{\mu}$ is a Lagrange multiplier associated with $\bar{y}$: we have

$$f_{h_n}'(\bar{u}_n)(u - \bar{u}_n) + \int_K G_h'(\bar{u}_n)(u - \bar{u}_n) d\bar{\mu}_n(x) \geq 0 \tag{3.4.27}$$

for all $u \in U_{ad}$ and all $n \in \mathbb{N}$. Notice that all sequences except $\{\bar{\mu}_h\}$ converge strongly. Passing to the limit in (3.4.27) yields

$$f'(\bar{u})(u - \bar{u}) + \int_K G'(\bar{u})(u - \bar{u}) d\hat{\mu}(x) \geq 0,$$

i.e. $\hat{\mu}$ satisfies the variational inequality (3.1.19). Moreover, we obviously have $\hat{\mu} \geq 0$. Finally, passing to the limit in

$$\int_K (\bar{y}_n - b) d\bar{\mu}_n(x) = 0,$$

we see that the complementary slackness condition is fulfilled by $\hat{\mu}$. Therefore, $\hat{\mu}$ fulfills all conditions to be satisfied by a Lagrange multiplier. Selecting a $y \in \mathcal{C}(\bar{\Omega})$ with $y(x) = 1$ in $B_{\frac{R}{2}}(\bar{x}_j)$ and $y(x) \equiv 0$ in $K \setminus B_R(\bar{x}_j)$, we find

$$\int_{B_{\frac{R}{2}}(\bar{x}_j)} 1 d\bar{\mu}_n(x) = 0$$

for all n by (3.4.26). Passing to the limit, we find that the restriction of $\hat{\mu}$ to $B_{\frac{R}{2}}(\bar{x}_j)$ vanishes, contradicting our Assumption of strict positivity of all Lagrange multipliers associated with $\bar{y}(\bar{x}_j)$. The estimate (3.4.25) has already been stated in Corollary 3.4.12 as a conclusion of Lemmas 3.4.9 and 3.4.11. Therefore, the assertion of the Lemma is obtained. □

Now, we proceed by removing the strongly active control components from the problem formulation, to focus completely on the difficulties associated with pointwise state constraints. In principle, we repeat the arguments of Lemmas 3.4.13 and 3.4.14 and apply them to the control constraints. In that way, we obtain that for each strongly active control constraint the associated component of the discrete control is also active, and hence known for all h sufficiently small.

Lemma 3.4.15. *Let $\bar{u}$ be a locally optimal control of Problem $(\mathbb{P})$ satisfying Assumption 3.4.2 and let $\mathcal{A}_{\bar{u}}$ be the associated set of strongly active control constraints. There exists $h_0 > 0$ such that for all $h \le h_0$ and all $i \in \mathcal{A}_{\bar{u}}$ the respective control constraint is active in the associated component of the discrete optimal control $\bar{u}_h$, i.e. for $i \in \mathcal{A}_{\bar{u}}$ we observe*

$$\begin{aligned} \bar{u}_i &= u_{a,i} \quad \Rightarrow \quad \bar{u}_{h,i} = u_{a,i}, \\ \bar{u}_i &= u_{b,i} \quad \Rightarrow \quad \bar{u}_{h,i} = u_{b,i}. \end{aligned}$$

The proof is similar to the one of Lemma 3.4.14. In a preparatory step, we show a boundedness result for the Lagrange multipliers associated with the control constraints.

Lemma 3.4.16. *Let $\{h_n\}_{n\in\mathbb{N}}$ be a sequence of positive real numbers converging to zero as n tends to infinity and let $\{\bar{u}_n\}$ denote the associated sequence of discrete optimal controls in the vicinity of $\bar{u}$. Any sequence $\{\bar{\eta}^a_{h_n}\}_{n\in\mathbb{N}}$, $\{\bar{\eta}^b_{h_n}\}_{n\in\mathbb{N}}$ of Lagrange multipliers associated with the control constraints for $(\mathbb{P}_h)$ is bounded in $\mathbb{R}^m$.*

Proof. Componentwise evaluation of the first-order optimality conditions (3.4.4) for Problem $(\mathbb{P}_h)$ yields

$$\bar{\eta}^a_{h,i} - \bar{\eta}^b_{h,i} = g_{h,i}, \quad \forall i = 1, \dots, m.$$

Because at least one of the components $\bar{\eta}^a_{h,i}, \bar{\eta}^b_{h,i}$ is zero, this implies

$$|\bar{\eta}^a_{h,i}| + |\bar{\eta}^b_{h,i}| \le |g_{h,i}| \le \|\partial_{u_i}\Psi(x, \bar{y}_h, \bar{u}_h)\| + \|\partial_{u_i} d(\cdot, \bar{y}_h, \bar{u}_h)\|_\infty \|\bar{p}_h\|_{L^1(\Omega)} \le c + c\|\bar{\mu}_h\|_{\mathcal{M}(K)} \le c$$

where we used the fact that $\|\bar{\mu}_h\|_{\mathcal{M}(K)}$ is uniformly bounded due to Lemma 3.4.13. □

With this auxiliary boundedness result we can now prove Lemma 3.4.15.

Proof of Lemma 3.4.15. Let $\{h_n\} > 0$ be a sequence of mesh sizes converging to zero, and denote by $\{\bar{u}_n\}$, $\{\bar{y}_n\}$, and $\{\bar{\mu}_n\}$, $\{\bar{\eta}^a_n\}, \{\bar{\eta}^b_n\}$ the control, state, and associated Lagrange multiplier sequence for the control and state constraints associated with h_n, respectively. Now let us assume the contrary:

then, there exists an index $i \in \mathcal{A}_{\bar{u}}$ and for all n a positive $h_n < \frac{1}{n}$ such that without loss of generality $\bar{u}_{n,i} > u_{a,i}$ where $\bar{u}_i = u_{a,i}$. By the complementary slackness condition (3.4.5), we have

$$\bar{\eta}^a_{n,i} = 0 \tag{3.4.28}$$

for all n. Moreover, by Lemma 3.4.16, the sequence $\{\bar{\eta}^a_n\}$ is bounded in $\mathbb{R}^m$. Therefore, we can select a sub-sequence converging to some $\hat{\eta}^a \in \mathbb{R}^m$. Let, w.l.o.g., $\{\bar{\eta}^a_n\}$ be this sub-sequence. We already know that $\bar{u}_n \to \bar{u}$ in $\mathbb{R}^m$ and $G'_h(\bar{u}_n)(u - \bar{u}_n) \to G'(\bar{u})(u - \bar{u})$ in $\mathcal{C}(K)$, as well as $\bar{\mu}_n \rightharpoonup^* \hat{\mu}$ in $\mathcal{M}(K)$, where $\hat{\mu}$ is a Lagrange multiplier associated with $\bar{y}$. We use this to verify that $\hat{\eta}^a$ is a Lagrange multiplier associated with $\bar{u}$: we have

$$f'_{h_n}(\bar{u}_n)(u - \bar{u}_n) + \int_K G'_h(\bar{u}_n)(u - \bar{u}_n)d\bar{\mu}_n(x) + \langle \eta^b_n - \eta^a_n, u - \bar{u}_n\rangle = 0$$

for all $u \in U_{ad}$ and all $n \in \mathbb{N}$. Passing to the limit yields

$$f'(\bar{u})(u - \bar{u}) + \int_K G'(\bar{u})(u - \bar{u})d\hat{\mu}(x) + \langle \hat{\eta}^b - \hat{\eta}^a, u - \bar{u}\rangle = 0,$$

i.e. the gradient equation (3.1.19) is satisfied. Moreover, we obviously have $\hat{\eta}^a, \hat{\eta}^b \geq 0$. Finally, passing to the limit in

$$\langle u_a - \bar{u}^a_n, \eta^a_n\rangle = 0$$

yields that the complementary slackness condition is fulfilled by $\hat{\eta}^a$. Therefore, $\hat{\eta}^a$ fulfills all conditions to be satisfied by a Lagrange multiplier. Selecting a $u \in U_{\mathrm{ad}}$ with $u_i = \frac{u_{b,i} - u_{a,i}}{2} \neq 0$, we find by

$$\eta^a_{n,i}(u_{a,i} - u_i) = 0$$

that $\eta^a_{n,i} = 0$ for all n by (3.4.28). Passing to the limit, we see that $\hat{\eta}^a_i$ vanishes, contradicting our assumption of strict positivity of all Lagrange multipliers associated with $\bar{u}_i$. □

3.4.3. An improved error estimate of order $h^2|\ln h|$

We have now collected all ingredients but one to improve the error estimate of Theorem 3.4.1. First of all, we know that local solutions to $(\mathbb{P})$ can be approximated by discrete local solutions $\bar{u}_h$ of $(\mathbb{P}_h)$, and the discretization error can be estimated with order $\mathcal{O}(h|\ln h|)$. In addition, we know that the distance between related active points of the continuous and the discrete state fulfills an error estimate of the same order. Under the assumption of strong activity, we moreover know that discrete active points exist in the neighborhoods of associated continuous active points. To improve the intermediate error estimate from Theorem 3.4.1 we need an additional assumption that in some sense keeps the controls in line.

Assumption 3.4.17. *If m_a and n_a denote the number of strongly active control constraints and strongly active state constraints from Definitions 3.2.4 and 3.2.5, respectively, and m is the number of controls, the following equality is fulfilled:*

$$m = m_a + n_a.$$

Assumption 3.4.17 implies that there are exactly as many strongly active state and control constraints as there are controls. Let us elaborate on this assumption. In [104], it has been motivated by a simple non-PDE-related example by F. Tröltzsch that the estimate from Theorem 3.4.1 is indeed sharp in certain situations with $m > m_a + n_a$. Numerical tests have also shown this in [105]. On the other hand, a known result for semi-infinite programming problems implies that there can exist at most m strongly active constraints, if the control is m-dimensional.

Lemma 3.4.18. *Let Assumptions 3.1.1, 3.1.7, and 3.2.1 be satisfied. Then, the following inequality holds for m, m_a, n_a:*

$$n_a + m_a \leq m.$$

Proof. By Lemma 3.4.15, we know that we can disregard the strongly active control components and consider a problem formulation with only $m - m_a$ control parameters. Then, from [16, Lemma 4.98], we deduce that there exists a Lagrange multiplier $\hat{\mu} \in \mathcal{M}(K)$, which is nonnegative in at most $m - m_a$ many components. This implies the assertion noting that strong activity of n_a points requires positivity of each Lagrange multiplier in the associated n_a components. □

Naturally, we are interested in a setting where $n_a > 0$, and implicitely assume this in the following. Then, we define the $n_a \times n_a$-matrix Y with entries

$$Y_{l,k} = \tilde{y}_{i_k}(\bar{x}_{j_l}), \quad i_k \in \mathcal{A}_{\bar{u}},\, j_l \in \mathcal{A}_{\bar{y}},$$

and $\tilde{y}_{i_k}$ solves

$$\begin{aligned} \mathcal{A}\tilde{y}_{i_k} + \partial_y d(\cdot,\bar{y},\bar{u})\tilde{y}_{i_k} &= -\partial_{u_{i_k}} d(\cdot,\bar{y},\bar{u}) && \text{in } \Omega \\ \tilde{y}_{i_k} &= 0 && \text{on } \Gamma. \end{aligned}$$

Theorem 3.4.19. *Let $\bar{u}$ be a locally optimal solution of Problem $(\mathbb{P})$ with Assumptions 3.1.1-3.4.2 and the first-order necessary optimality conditions from Theorem 3.1.9 satisfied. Moreover, let the matrix Y be regular. There exists a sequence $\bar{u}_h$ of locally optimal controls for $(\mathbb{P}_h)$ that converge to $\bar{u}$ as h tends to zero. With a constant $C > 0$ independent of h, there holds*

$$|\bar{u} - \bar{u}_h| \leq Ch^2|\ln h|$$

for all sufficiently small $h > 0$.

Proof. The existence of a sequence $\{\bar{u}_h\}_h$ of optimal solutions to $(\mathbb{P}_h)$ converging to $\bar{u}$ with order $h\sqrt{|\ln h|}$ is already known from Theorem 3.4.1. Now, consider the n_a strongly active points $\bar{x}_{j_k}$, $j_k \in \mathcal{A}_{\bar{y}}$, $k = 1, \dots n_a$, and for each such point choose one associated discrete active point $\bar{x}^h_{j_k} \in B_R(\bar{x}_{j_k})$, which exists according to Lemma 3.4.14. We observe

$$\bar{y}(\bar{x}_{j_k}) = b = \bar{y}_h(\bar{x}^h_{j_k}) = \tilde{y}_h(\bar{x}^h_{j_k}) + (\bar{y}_h - \tilde{y}_h)\,(\bar{x}^h_{j_k}),$$

which can be rewritten as

$$(\bar{y}_h - \tilde{y}_h)(\bar{x}^h_{j_k}) = \bar{y}(\bar{x}_{j_k}) - \tilde{y}_h(\bar{x}_{j_k}) + \tilde{y}_h(\bar{x}_{j_k}) - \tilde{y}(\bar{x}^h_{j_k}). \tag{3.4.29}$$

We proceed by Taylor expansion of $\tilde{y}_h(\bar{x}^h_{j_k})$ at $\bar{x}_{j_k}$, which, with $x_\xi = \bar{x}_{j_k} + \xi(\bar{x}^h_{j_k} - \bar{x}_{j_k})$ for a $\xi \in (0,1)$, yields

$$\tilde{y}_h(\bar{x}^h_{j_k}) \;=\; \tilde{y}_h(\bar{x}_{j_k}) + \langle \nabla\tilde{y}_h(\bar{x}_{j_k}), \bar{x}^h_{j_k} - \bar{x}_{j_k}\rangle + \frac{1}{2}\langle \bar{x}^h_{j_k} - \bar{x}_{j_k}, \nabla^2\tilde{y}_h(x_\xi)(\bar{x}^h_{j_k} - \bar{x}_{j_k})\rangle. \tag{3.4.30}$$

Inserting (3.4.30) in (3.4.29) and making use of $\nabla\bar{y}(\bar{x}_{j_k}) = 0$ implied by Assumption 3.2.1 leads to

$$|(\bar{y}_h - \tilde{y}_h)(\bar{x}_{j_k})| = |(\bar{y} - \tilde{y}_h)(\bar{x}_{j_k}) - \langle \nabla(\tilde{y}_h(\bar{x}_{j_k}), \bar{x}^h_{j_k} - \bar{x}_{j_k}\rangle - \frac{1}{2}\langle \bar{x}^h_{j_k} - \bar{x}_{j_k}, \nabla^2\tilde{y}_h(x_\xi)(\bar{x}^h_{j_k} - \bar{x}_{j_k})\rangle| \tag{3.4.31}$$

$$\begin{aligned} &= |(\bar{y} - \tilde{y}_h)(\bar{x}_{j_k}) + \langle \nabla(\bar{y} - \tilde{y}_h)(\bar{x}_{j_k}), \bar{x}^h_{j_k} - \bar{x}_{j_k}\rangle && (3.4.32) \\ &\quad - \frac{1}{2}\langle \bar{x}^h_{j_k} - \bar{x}_{j_k}, \nabla^2\tilde{y}_h(x_\xi)(\bar{x}^h_{j_k} - \bar{x}_{j_k})\rangle| \\ &\leq \|\bar{y}_h - \tilde{y}_h\|_\infty + |\langle \nabla(\bar{y} - \tilde{y}_h)(\bar{x}_{j_k}), \bar{x}^h_{j_k} - \bar{x}_{j_k}\rangle| + c|\bar{x}^h_{j_k} - \bar{x}_{j_k}|^2 && (3.4.33) \\ &\leq \|\bar{y}_h - \tilde{y}_h\|_\infty + \|\nabla(\bar{y} - \tilde{y}_h)\|_{\infty,K}|\bar{x}^h_{j_k} - \bar{x}_{j_k}| + c|\bar{x}^h_{j_k} - \bar{x}_{j_k}|^2. && (3.4.34) \end{aligned}$$

To estimate (3.4.34) further, we derive

$$\|\nabla(\bar{y}-\tilde{y}_h)\|_{\infty,K} = \|\nabla(G(\bar{u})-G(\bar{u}_h))\|_{\infty,K} = \|\nabla G'(u_\xi)(\bar{u}_h-\bar{u})\|_{\infty,K} \le c_3|\bar{u}-\bar{u}_h|, \tag{3.4.35}$$

again with $u_\xi = \bar{u}+\xi(\bar{u}_h-\bar{u})$ for some $\xi \in (0,1)$, where we applied Proposition 2.4.2 to make use of the interior $W^{2,\infty}$-regularity of solutions to linearized state equations. Then, using (3.4.35), the error estimate for the control from Theorem 3.4.1, as well as the estimate for the distance of the active points from Lemma 3.4.14 allows to estimate

$$\|\nabla(\bar{y}-\tilde{y}_h)\|_{\infty,K}|\bar{x}^h_{j_k}-\bar{x}_{j_k}| \le c\,h\sqrt{|\ln h|}\cdot h\sqrt{|\ln h|} \le c\,h^2|\ln h|.$$

Inserting this in (3.4.34) implies

$$|\bar{y}(\bar{x}_{j_k})-\tilde{y}_h(\bar{x}_{j_k})| \le ch^2|\ln h| \quad \forall j_k \in \mathcal{A}_{\bar{y}}, \quad k=1,\dots,n_a. \tag{3.4.36}$$

Noting that

$$\begin{aligned}
|\bar{y}(\bar{x}_{j_k})-\tilde{y}_h(\bar{x}_{j_k})| &= |G'(\bar{u})(\bar{u}-\bar{u}_h)(\bar{x}_{j_k}) + \frac{1}{2}G''(u_\xi)[\bar{u}-\bar{u}_h,\bar{u}-\bar{u}_h](\bar{x}_{j_k})| \\
&\ge |G'(\bar{u})(\bar{u}-\bar{u}_h)(\bar{x}_{j_k})| - \frac{1}{2}\|G''(u_\xi)[\bar{u}-\bar{u}_h,\bar{u}-\bar{u}_h]\|_\infty \\
&\ge |G'(\bar{u})(\bar{u}-\bar{u}_h)(\bar{x}_{j_k})| - c|\bar{u}-\bar{u}_h|^2
\end{aligned}$$

we obtain

$$|G'(\bar{u})(\bar{u}-\bar{u}_h)(\bar{x}_{j_k})| \le c|\bar{u}-\bar{u}_h|^2 + |\bar{y}(\bar{x}_{j_k})-\tilde{y}_h(\bar{x}_{j_k})| \le ch^2|\ln h| \quad \forall j_k \in \mathcal{A}_{\bar{y}}, \quad k=1,\dots,n_a \tag{3.4.37}$$

with the help of estimate (3.4.36) and Theorem 3.4.1. Now recall that $|\bar{u}_i-\bar{u}_{h,i}|=0$ for $i \in \mathcal{A}_{\bar{u}}$ by Lemma 3.4.15. Then, obviously only the remaining control components can contribute to the error. We define

$$\bar{u}_{\mathcal{I},i} = \begin{cases} \bar{u}_i & \text{if } i \in \mathcal{I}_{\bar{u}} \\ 0 & \text{else} \end{cases}, \quad \bar{u}^h_{\mathcal{I},i} = \begin{cases} \bar{u}^h_i & \text{if } i \in \mathcal{I}_{\bar{u}} \\ 0 & \text{else} \end{cases}, \quad i=1,\dots,m$$

as well as

$$\bar{u}_{\mathcal{A}} := \bar{u}-\bar{u}_{\mathcal{I}}, \quad \bar{u}^h_{\mathcal{A}} := \bar{u}_h-\bar{u}^h_{\mathcal{I}},$$

thus collecting all components that belong to strongly active constraints of $\bar{u}$ into $\bar{u}_{\mathcal{A}}$ and $\bar{u}^h_{\mathcal{A}}$. Then, we rewrite (3.4.37) as

$$\left|G'(\bar{u})(\bar{u}_{\mathcal{I}}-\bar{u}^h_{\mathcal{I}})(\bar{x}_{j_k}) + G'(\bar{u})(\bar{u}_{\mathcal{A}}-\bar{u}^h_{\mathcal{A}})(\bar{x}_{j_k})\right| \le ch^2|\ln h|, \tag{3.4.38}$$

where $G'(\bar{u})(\bar{u}_{\mathcal{A}}-\bar{u}^h_{\mathcal{A}})(\bar{x}_{j_k})$ vanishes. Therefore, (3.4.38) simplifies to

$$\left|\sum_{i\in\mathcal{I}_{\bar{u}}}(\bar{u}_i-\bar{u}_{h,i})\tilde{y}_i(\bar{x}_{j_k})\right| \le ch^2|\ln h|, \tag{3.4.39}$$

which we rewrite as

$$|Y(\bar{u}_{\mathcal{I}_{\bar{u}}}-\bar{u}^h_{\mathcal{I}_{\bar{u}}})| \le ch^2|\ln h|.$$

Then, by the regularity of Y, we obtain

$$\left|\bar{u}_{\mathcal{I}_{\bar{u}}}-\bar{u}^h_{\mathcal{I}_{\bar{u}}}\right| \le ch^2|\ln h|.$$

Noting again that $\bar{u}_i=\bar{u}_{h,i}$ for all $i \in \mathcal{A}_{\bar{u}}$, the assertion is obtained. □

3.5. Generalization

The theory presented in this chapter can be generalized to a larger class of optimal control problems. Let us motivate and demonstrate this. First of all, we have seen in the proof of Theorem 3.4.19 that the control constraints do not influence the main idea, since the strongly active control components are removed from the problem formulation, and weakly active control constraints are treated like inactive components. Let us therefore also comment on a setting that includes one or more unrestricted control parameters $u_i \in \mathbb{R}$. Moreover, we have pointed out that the Lagrange multipliers may not necessarily be unique and in Remark 3.2.6 a generalization of the presented theory has been announced. We will discuss this now.

3.5.1. The absence of control bounds

Let us first concentrate on a setting where some or all control components are unconstrained, i.e. we redefine the bounds

$$u_a < u_b, \quad u_{a,i} \in \mathbb{R} \cup -\infty, \quad u_{b,i} \in \mathbb{R} \cup \infty.$$

Still, for a given fixed control $u \in \mathbb{R}^m$, Theorem 2.4.6 guarantees the existence of an associated state

$$y = G(u) \in H_0^1(\Omega) \cap \mathcal{C}^{2,\nu}(\Omega) \cap W^{2,\infty}(\Omega_1).$$

and the definition of the reduced objective function f remains meaningful. One of the key ingredients in many proofs was the boundedness of U_{ad}, which obviously is no longer guaranteed. We have to compensate this by an additional assumption.

Assumption 3.5.1. *The reduced objective function f fulfills the condition*

$$\lim_{|u| \to \infty} f(u) = \infty.$$

Then, we consider problem ($\mathbb{P}$) with an appropriately redefined set of admissible and feasible controls U_{ad} and U_{feas}, respectively.

A simple objective function that falls in this category is a tracking type function

$$f(u) = \frac{1}{2}\|G(u) - y_d\|^2 + \frac{\kappa}{2}|u|^2$$

with quadratic Tikhonov term with a positive real number κ.

Under Assumption 3.5.1, Theorem 3.1.5 remains valid and provides the existence of at least one optimal control of Problem ($\mathbb{P}$). This is true because it is sufficient to consider controls u in the compact set $\overline{B(0,\tau)} \cap U_{\mathrm{feas}}$, where the $\overline{B(0,\tau)}$ denotes the closed ball in $\mathbb{R}^m$ around $0 \in R^m$ with radius τ. Moreover, for a given locally optimal control $\bar{u} \in \mathbb{R}^m$ all previously shown results remain valid when considering e.g. test functions in $B(\bar{u},\varepsilon) \cap U_{\mathrm{ad}}$ or $B(\bar{u},\varepsilon) \cap U_{\mathrm{feas}}$ whereever appropriate. This guarantees in particular that all appearing constants c only depend on ε, which is a finite number.

The previously shown results, in particular the ones of Theorems 3.4.1 and 3.4.19 remain valid in their present form if we agree that the components $\bar{\eta}_{a,i}, \bar{\eta}_{b,i}$ of the Lagrange multipliers associated with the control constraints are automatically set to zero. Obviously, the same convention must hold for the discrete multipliers.

3.5.2. A further discussion of active constraints

In conclusion of this chapter, we want to weaken the assumptions on the strongly active constraints used in Theorem 3.4.19. Consider first the following motivating example with only one control parameter in a one-dimensional spatial domain $\Omega := (0, 2\pi)$, to which Theorem 3.4.19 cannot be applied.

$$\text{Minimize } J(y,u) := \frac{1}{2}\int_0^{2\pi} \left(y(x) - \sin^2 x\right)^2 dx + \frac{1}{2}\left(u - \left(2 - \frac{4}{3\pi}\right)\right)^2 \tag{3.5.1a}$$

subject to the linear elliptic PDE constraint

$$\begin{aligned} -\Delta y &= u(\sin^2 x - \cos^2 x) && \text{in } \Omega, \\ y &= 0 && \text{on } \Gamma, \end{aligned} \tag{3.5.1b}$$

as well as the the pointwise bound

$$y(x) \leq 1 \quad \forall\, x \in \Omega, \tag{3.5.1c}$$

and the control bound

$$u \geq 2. \tag{3.5.1d}$$

The adjoint and gradient equation to be fulfilled are given by

$$\begin{aligned} -\Delta \bar{p} &= \bar{y} - \sin^2 x + \bar{\mu} && \text{in } \Omega, \\ \bar{p} &= 0 && \text{on } \Gamma, \end{aligned} \tag{3.5.2}$$

as well as

$$\left(\bar{u} - \left(2 - \frac{4}{3\pi}\right)\right) + \int_0^{2\pi} \bar{p}(\sin^2 x - \cos^2 x)\, dx - \bar{\eta}_a = 0, \tag{3.5.3}$$

for a nonnegative Lagrange multiplier $\bar{\eta}_a \in \mathbb{R}$. The complementary slackness condition for the nonnegative Lagrange multiplier $\bar{\mu}$ reads

$$\int_0^{2\pi} (\bar{y} - 1)\, d\bar{\mu}(x) = 0. \tag{3.5.4}$$

This convex optimal control problem has been constructed such that $\bar{u} = 2$ is the unique optimal control with associated state $\bar{y} = \sin^2 x$. It is easily verified that then the state equation (3.5.1b) is fulfilled and that the optimal state satisfies the constraints (3.5.1c) with two active points $\bar{x}_1 = \frac{\pi}{2}$ and $\bar{x}_2 = \frac{3}{2}\pi$. Note that these points are situated in the interior of $\Omega = (0, 2\pi)$ and therefore the constraints could also be prescribed in an interior compact set K. According to [16, Proposition 4.92], there exists a Lagrange multiplier vector $\bar{\mu} \in \mathbb{R}^2$ with at most one positive component. From the symmetry of the optimal state with two active points one would expect that there exist at least two such multipliers $\bar{\mu}_1$ and $\bar{\mu}_2$, associated with two adjoint states $\bar{p}_1$ and $\bar{p}_2$. In fact, let $\bar{p}_1$ and $\bar{p}_2$ be given by

$$\bar{p}_1(x) = \begin{cases} \frac{2}{\pi}x & \text{if } x \in (0, \frac{\pi}{2}] \\ -\frac{2}{3\pi}x + \frac{4}{3} & \text{if } x \in (\frac{\pi}{2}, 2\pi] \end{cases}, \qquad \bar{p}_2(x) = \begin{cases} \frac{2}{3\pi}x & \text{if } x \in (0, \frac{3}{2}\pi] \\ -\frac{2}{\pi}x + 4 & \text{if } x \in (\frac{3}{2}\pi, 2\pi] \end{cases}.$$

Both functions are depicted in Figure 3.1. Straight-forward calculations show that

$$-\Delta \bar{p}_1 = \frac{8}{3\pi}\delta_{x=\frac{\pi}{2}}, \quad -\Delta \bar{p}_2 = \frac{8}{3\pi}\delta_{x=\frac{3}{2}\pi}.$$

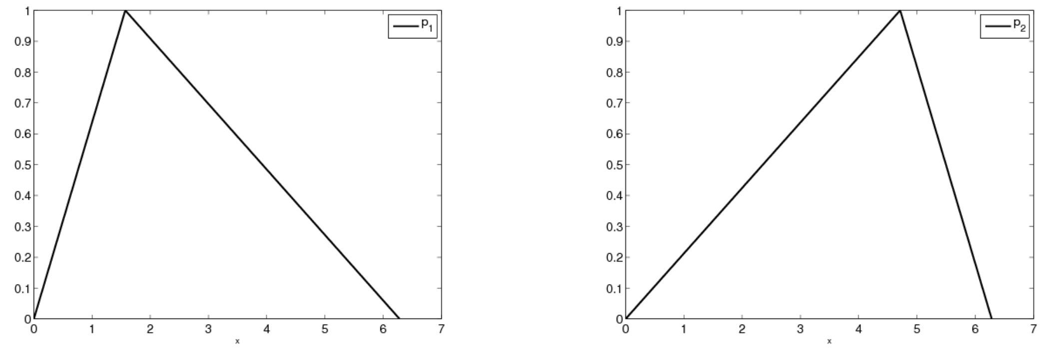


Figure 3.1.: Two different adjoint states for Problem (3.5.1)

Keeping in mind that we agreed to identify Lagrange multipliers that are associated with finitely many active points with vectors from $\mathbb{R}^n$ we define

$$\bar{\mu}_1 = \left(\frac{8}{3\pi}, 0\right)^T \geq 0, \quad \bar{\mu}_2 = \left(0, \frac{8}{3\pi}\right)^T \geq 0.$$

Then, with $\bar{y} = \sin^2 x$ both the tuples $(\bar{p}_1, \bar{\mu}_1)$ as well as $(\bar{p}_2, \bar{\mu}_2)$ fulfill the adjoint equation (3.5.2). It is equally straight-forward to verify that the gradient equation (3.5.3) is fulfilled for both adjoint states $\bar{p}_1$ and $\bar{p}_2$ with $\bar{\eta}_a^{(1,2)} = 0$.

A third triple of dual variables is given by $\bar{p}_3 = 0$, $\bar{\mu}_3 = (0,0)^T$, and $\bar{\eta}_a^{(3)} = \frac{4}{3\pi}$.

In this example problem, the number of control parameters is $m = 1$ and each pair of Lagrange multipliers $(\mu, \eta) \in \mathbb{R}^2 \times \mathbb{R}$ admits in sum exactly one positive component, but according to the Definitions 3.2.4 and 3.2.5, which only accounts for positive components associated with the same control parameter or active point, we have $m_a = n_a = 0$, and consequently

$$m_a + n_a < m.$$

Nevertheless, one would expect to be able to prove the higher order of convergence, $h^2|\ln h|$, with the help of subsequences that account for the three different situations. Let us demonstrate this in a more general setting.

In the following, let $\bar{u}$ be a local solution of $(\mathbb{P})$ and let $\{\bar{u}_h\}$ denote the sequence of discrete local solutions approximating $\bar{u}$ from Assumption 3.4.2. In particular, we already know that $\bar{u}_h$ converges to $\bar{u}$ as h tends to zero. For brevity, we collect the Lagrange multipliers $\bar{\eta}_a$ and $\bar{\eta}_b$ into a single vector

$$\bar{\eta} = \bar{\eta}_a - \bar{\eta}_b \in \mathbb{R}^m.$$

This causes no ambiguity, since at most one of the upper or lower bound can be active in each control component, which implies that either or both $\bar{\eta}_{a,i}$ and $\bar{\eta}_{b,i}$ are zero for all $i = 1, \ldots, m$.

Definition 3.5.2. *Let $\bar{u}$ be a local solution of $(\mathbb{P})$. We introduce the set of Lagrange multipliers for $\bar{u}$ by*

$$\Lambda(\bar{u}) := \{(\eta, \mu) \in \mathbb{R}^m \times \mathcal{M}(K) \,|\, \eta, \mu \text{ fullfill the conditions in Lemma 3.1.11}\}.$$

In addition, for every $\lambda \in \Lambda(\bar{u})$, we define

$$\mathcal{A}_{\bar{u}}^{\lambda} := \{i \mid \eta_i > 0\} \quad \text{and} \quad \mathcal{A}_{\bar{y}}^{\lambda} := \{j \mid \mu_j > 0\}$$

and introduce

$$m_\lambda := \sharp \mathcal{A}_{\bar{u}}^{\lambda}, \qquad n_\lambda := \sharp \mathcal{A}_{\bar{y}}^{\lambda}.$$

Moreover, we define a real number l_λ by

$$l_\lambda := n_\lambda + m_\lambda. \tag{3.5.5a}$$

In other words, l_λ denotes the number of positive components of each Lagrange multiplier tuple, and the formerly defined sets of strongly active control or state constraints, respectively, have been redefined for each $\lambda \in \Lambda(\bar{u})$. That allows to focus on the number of positive components, only. For that purpose, let

$$\bar{l} := \min_{\lambda \in \Lambda} \{l_\lambda\}$$

be given, i.e. each multiplier λ admits in sum at least $\bar{l}$ positive components. Then, in preparation of the next theorem, let the $n_\lambda \times n_\lambda$-matrix Y^λ be given with entries

$$Y_{l,k}^{\lambda} = \tilde{y}_{i_k}(\bar{x}_{j_l}),$$

where $i_k \in \mathcal{A}_{\bar{u}}^{\lambda}$, $j_l \in \mathcal{A}_{\bar{y}}^{\lambda}$, and $\tilde{y}_{i_k}$ solves

$$\begin{aligned} \mathcal{A}\tilde{y}_{i_k} + \partial_y d(\cdot, \bar{y}, \bar{u}) &= -\partial_{u_{i_k}} d(\cdot, \bar{y}, \bar{u}) && \text{in } \Omega \\ \tilde{y}_{i_k} &= 0 && \text{on } \Gamma. \end{aligned}$$

A generalization of Theorem 3.4.19 now reads:

Theorem 3.5.3. *Let $\bar{u}$ be a locally optimal solution of Problem $(\mathbb{P})$ fulfilling the first order necessary optimality conditions of Theorem 3.1.9 as well as Assumptions 3.1.1-3.4.2. Moreover suppose*

$$m = \bar{l},$$

and let all matrices Y^λ be uniformly regular. Last, let $\bar{u}_h$ denote the sequence of locally optimal controls for $(\mathbb{P}_h)$ that converges to $\bar{u}$ as h tends to zero from Assumption 3.4.2. With a constant $C > 0$ independent of h, there holds

$$|\bar{u} - \bar{u}_h| \le Ch^2 |\ln h|$$

for all sufficiently small $h > 0$.

Proof. It is already known from Theorem 3.4.1 that the sequence $\{\bar{u}_h\}$ converges to $\bar{u}$ as h tends to zero with order $h\sqrt{|\ln h|}$. From Lemmas 3.4.13 and 3.4.16, we know that all sequences $\{\bar{\mu}_h\}$ and $\{\bar{\eta}_h\}$ are bounded in $\mathcal{M}(K)$ or $\mathbb{R}^m$, respectively. Then, there exist subsequences which we again denote by $\{\bar{\eta}_h\}$ and $\{\bar{\mu}_h\}$, converging to $\hat{\eta}, \hat{\mu} \in \mathbb{R}^m \times \mathcal{M}(K)$. Since everything in the sequel is true for any subsequence the assertion is obtained.

As in the proofs of Lemmas 3.4.14 and 3.4.15 we deduce that

$$\hat{\lambda} := (\hat{\eta}, \hat{\mu}) \in \Lambda(\bar{u}),$$

i.e. $\hat{\lambda}$ is a tuple of Lagrange multipliers for $(\mathbb{P})$. We define

$$\hat{m}_\lambda := \sharp\{\hat{\eta}_i > 0\}, \quad i = 1, \ldots, m, \qquad \hat{n}_\lambda := \sharp\{\hat{\mu}_j > 0\}, \quad j = 1, \ldots, n,$$

as well as

$$\hat{l} := \hat{m}_\lambda + \hat{n}_\lambda.$$

It is clear that $\hat{l} \geq \bar{l}$, since $\hat{\lambda} \in \Lambda(\bar{u})$ and $\bar{l}$ denotes the minimum number of positive components of all multipliers.

Moreover, we define the sequences $\{m_{a,h}\}, \{n_{a,h}\}$, and $\{l_h\}$ that count the active control components and points on the discrete level, where the index a indicates activity:

$$\begin{aligned} m_{a,h} &:= \sharp\{\bar{u}_{h,i} = u_{a,i} \text{ or } \bar{u}_{h,i} = u_{b,i}\}, \\ n_{a,h} &:= \sharp\{\bar{x}_j \,|\, \text{there exists at least one point } \bar{x}_j^h \in B_R(\bar{x}_j) \text{ s.t. } \bar{y}_h(\bar{x}_j^h) = b\}. \end{aligned}$$

Note that this means that there are $m_{a,h}$ active control components of the discrete problem, and at least $n_{a,h}$ active points of the continuous problem in whose neighborhoods there exist discrete active points. It does not necessarily imply that the discrete multipliers admit just as many positive components, nor does it imply that $\bar{\mu}_j > 0$ for all $j = 1, \ldots, n_{a,h}$, since we did not claim strict complementarity. We particularly point out that the problem admits only finitely many controls as well as only finitely many active points. This leaves only finitely many possibilities for the number of positive components in each multiplier (η_h, μ_h), i.e. only finitely many choices for $(m_{a,h}, n_{a,h})$.

Now, after possibly reverting to another subsequence, assume that for all $n > 0$ there exists $h < \frac{1}{n}$ such that

$$m_{a,h} + n_{a,h} < \bar{l}.$$

Consider first the $m - m_{a,h}$ remaining inactive control components on the discrete level. As in the proof of Lemma 3.4.15, we find that the associated components of $\hat{\eta}$ vanish, which implies in particular that there are at most $m_{a,h}$ positive components of $\hat{\eta}$, i.e.

$$\hat{m}_\lambda \leq m_{a,h}. \tag{3.5.6}$$

Now we turn to the analysis of the $n - n_{a,h}$ active points of Problem $(\mathbb{P})$, in whose neighborhood there is no discrete active point. Adapting the proof of Lemma 3.4.14, we eventually obtain that $\hat{\mu} = 0$ in all these neighborhoods, i.e.

$$\hat{n}_\lambda \leq n_{a,h}. \tag{3.5.7}$$

Adding (3.5.6) and (3.5.7), and noting that $\hat{l} \leq \hat{m}_\lambda + \hat{n}_\lambda$ we obtain

$$\hat{l} \leq m_{a,h} + n_{a,h} < \bar{l},$$

which is a contradiction to the definition of $\hat{l}$ and the fact that $(\hat{\eta}, \hat{\mu}) \in \Lambda(\bar{u})$.

We have thus proven that $m_{a,h} + n_{a,h} \geq \bar{l}$, i.e. we have at least as many active constraints on the discrete level as $\bar{l}$ indicates. The proof can now be continued as in Theorem 3.4.19, reverting to converging subsequences $\{\bar{u}_{h_k}\}$. First of all, for each subsequence the $m_{a,h}$ active control components can be removed from the problem formulation, and at least $m - m_{a,h} = \bar{l} - m_{a,h} \leq n_{a,h}$ pairs of discrete and continuous active points can be chosen. All of these pairs fulfill the distance estimate of order $h\sqrt{|\ln h|}$ from Corollary 3.4.12. For each subsequence we hence obtain

$$|\bar{u} - \bar{u}_{h_k}| \leq c_k h^2 |\ln h|.$$

Since these arguments are valid for any subsequence, we obtain the assertion by taking the maximum over all the constants c_k. This is possible since we only have finitely many active points $\bar{x}_j$ as well as controls $\bar{u}_i$, leaving only finitely many constellations for positive multiplier components. □

3.6. Numerical results

In this section, we provide numerical experiments that illustrate our proven orders of convergence. On the one hand, we will demonstrate that the lower order estimate of Theorem 3.4.1 is sharp in certain situations, on the other hand, we will also show examples for the higher order convergence of Theorem 3.4.19. Since our main focus in this thesis is on the presence of state constraints, we will focus on examples without control constraints. We will transform the model problems to be considered into finite dimensional nonlinear programming problems which we solve with the help of the MATLAB optimization routines `quadprog` or `fmincon`, respectively. The routine `quadprog` solves quadratic programming problems with linear equality and inequality constraints with the help of an active set strategy, whereas `fmincon` solves more general nonlinear programming problems with nonlinear equality and inequality constraints with an active set method, cf. [97]. The underlying PDEs are solved with the help of the routines `assempde` or `pdenonlin`, respectively, of MATLAB's PDE toolbox. In our following computations we will use examples where the exact solution is not readily available. Instead, we will determine a numerical solution $\hat{u}_h$ on a fine grid with mesh size $\hat{h}$ as a replacement. We will determine the experimental order of convergence by

$$EOC = \frac{\ln(|\hat{u}_h - \bar{u}_{h_1}|) - \ln(|\hat{u}_h - u_{h,2}|)}{\ln(h_1) - \ln(h_{h_2})},$$

with h_1, h_2 being two consecutive mesh sizes with $h_1, h_2 > \hat{h}$.

3.6.1. A problem with approximation of order $h\sqrt{\ln h}$

Very simple examples can show that the error estimate of Theorem 3.4.1 cannot always be improved. In [104] and [105], problems with and without PDE constraints prove this. Here, we consider a simple linear-quadratic model problem of the form

$$\text{Minimize } \frac{1}{2}\int_0^1\int_0^1 (y-y_d)^2\,dx_1dx_2 + \frac{1}{2}|u-u_d|^2 \tag{E_1}$$

subject to

$$\begin{aligned} -\Delta y + y &= \sum_{i=1}^{3} u_i e_i(x) && \text{in } \Omega := (0,1)^2 \\ y &= 0 && \text{on } \Gamma \\ y &\le b && \text{in } \Omega \end{aligned}$$

with the specific choice

$$\begin{aligned} y_d(x) &= (x_1 + x_1^2 - 2x_1^3)(x_2 + x_2^2 - 2x_2^3), && (3.6.1)\\ e_1(x) &= 25x_1x_2 - 2x_1 - 2x_2, && (3.6.2)\\ e_2(x) &= -2x_1^2 - 2x_2^2 + 13x_1^2x_2 + 13x_1x_2^2 + x_1^2x_2^2, && (3.6.3)\\ e_3(x) &= 4x_1^3 + 4x_2^3 - 26x_1x_2^3 - 26x_1^3x_2 - 2x_1^2x_2^3 - 2x_1^3x_2^2 + 4x_1^3x_2^3, && (3.6.4) \end{aligned}$$

as well as

$$u_d = (-1,\,1,\,1)^T, \quad b = \frac{1}{10}. \tag{3.6.5}$$

An exact solution of this problem is not known. The problem is constructed such that without the presence of the pointwise state constraints a solution would be given by $\bar{y} = y_d$ and $\bar{u} = u_d$, where the optimal state admits also values that are greater than $1/10$. Note that due to the linear PDE we can make use of the superposition principle and determine three states $y_1, \ldots, y_3$ that correspond to the PDE with right-hand sides $e_1, \ldots, e_3$. Then, we obtain

$$y(u) = \sum_{i=1}^{3} u_i y_i(x),$$

and the problem can be transformed into a quadratic problem

$$\text{Minimize } \frac{1}{2} u^T \left((y_i, y_j)\right)_{i,j} u + \frac{1}{2} u^T u - \left((y_d, y_i)\right)_i^T u - u_d^T u$$

subject to

$$\sum_{i=1}^{3} u_i y_i(x) \leq b.$$

For the numerical approximation, we precompute the approximate states y_i^h, $i = 1, \ldots, 3$ and determine a solution

$$\hat{u}_h \approx (0.7488,\, 0.8416,\, 1.1981)^T$$

on a grid with mesh size $\hat{h} = 2^{-10}$ with the help of the optimization routine `quadprog`. We then solve the optimal control problem on a series of meshes with mesh size $h = 2^{-k}$, $k = 0, \ldots, 8$ and compare the solutions $\bar{u}_h^{(k)}$ with $\hat{u}_h$. The set of meshes is obtained by iteratively refining an initial mesh obtained by MATLAB's routine `initmesh` with parameter `hmax`$= 1$, using the mesh-refinement routine `refinemesh`. In Figure 3.2, we display the discrete optimal state and Lagrange multiplier computed on mesh with meshsize $h = 2^{-6}$. The optimal state and Lagrange multiplier clearly indicate that the problem admits only one active point. Therefore, the number of active constraints is smaller than the number of controls and our theory provides the order $\mathcal{O}(h\sqrt{\ln h})$, only. Figure 3.3 shows the error $|\bar{u}_h^{(k)} - \hat{u}_h|$ plotted against $h_k := 2^{-k}$ in logarithmic scale. For comparison, the straight line indicates linear order of convergence, proving that the worst case estimate of order $\mathcal{O}(h\sqrt{\ln |h|})$ can actually be observed. Note that we do not expect to see the influence of the logarithmic term in our numerical computations. The experimental order of convergence EOC is also shown in Figure 3.3.

3.6.2. Problems with approximation of order $h^2|\ln h|$

We now turn to example problems where we expect to observe the higher order of convergence proven in Theorem 3.4.19. We begin with a linear-quadratic problem closely related to Example (E_1).

A problem with linear PDE

We consider a modification of Example (E_1) with only one control parameter. Then, we obtain the following formulation:

$$\text{Minimize } \frac{1}{2} \int_0^1 \int_0^1 (y - y_d)^2 \, dx_1 dx_2 + \frac{1}{2} |u - u_d|^2 \qquad (E_2)$$

subject to

$$\begin{aligned} -\Delta y + y &= ue(x) && \text{in } \Omega := (0,1)^2 \\ y &= 0 && \text{on } \Gamma \\ y &\leq b && \text{in } \Omega \end{aligned}$$

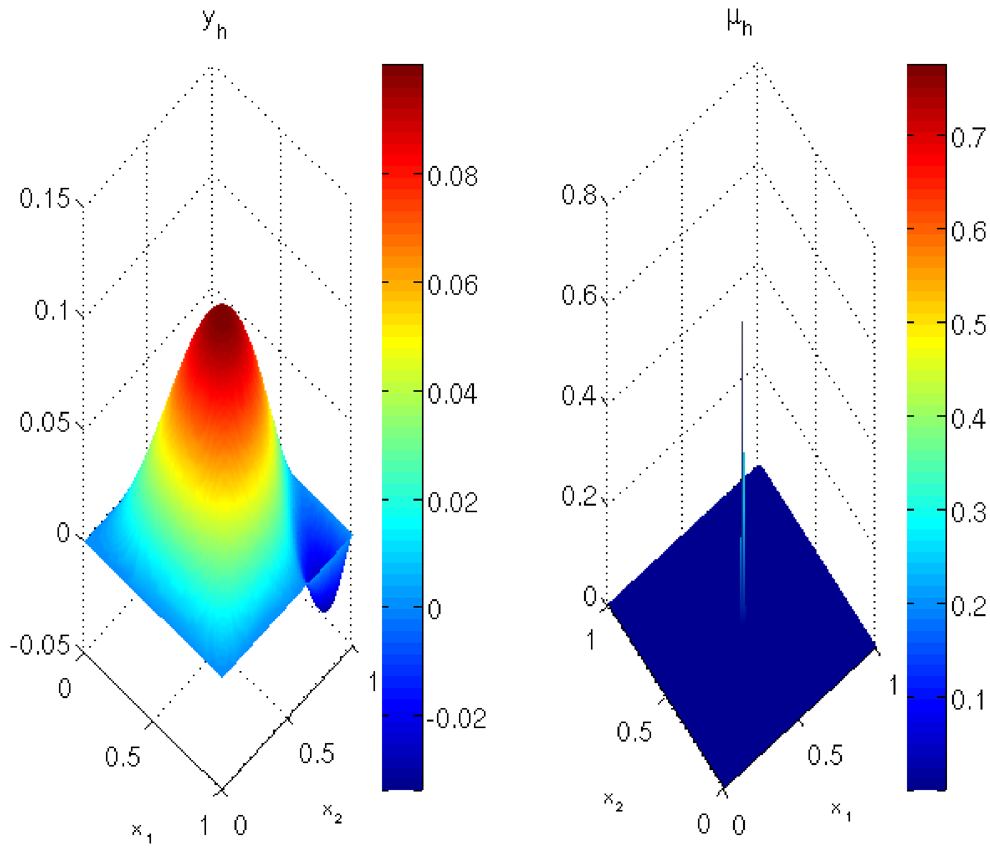


Figure 3.2.: Example (E_1): Discrete optimal state and Lagrange multiplier

with

$$y_d(x) = (x_1 + x_1^2 - 2x_1^3)(x_2 + x_2^2 - 2x_2^3), \quad e(x) = e_1(x) + e_2(x) + e_3(x), \quad u_d = 1, \quad b = \frac{1}{10},$$

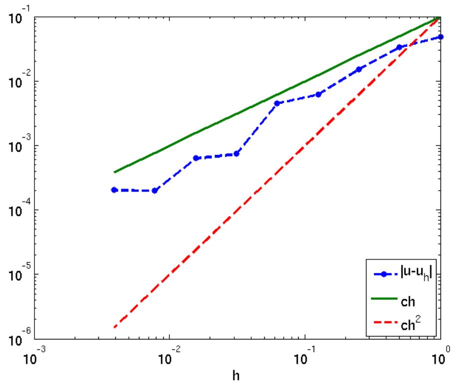
cf. (3.6.1)–(3.6.4) as well as (3.6.5). Again, the exact solution of this problem is not known, and we therefore compute a reference solution

$$\hat{u}_h \approx 0.35849$$

on a fine grid with mesh size $\hat{h} = 2^{-10}$ and proceed as before. We calculate discrete optimal solutions on a set of iteratively refined meshes with mesh size $h_k = 2^{-k}$, $0 = 1, \dots, 9$, where we again precompute a state y_e^h that solves the PDE with right-hand side $e(x)$. We then obtain $\bar{y}_h = \bar{u}_h \cdot y_e^h$. Figure 3.4 shows the optimal state and Lagrange multiplier computed on a mesh with $h = 2^{-6}$. Figure 3.5 shows the error in the optimal control, indicating quadratic convergence, both in numbers and graphically.

Comparison of Examples (E_1) and (E_2)

Before we continue with a further example involving a nonlinear PDE we briefly compare Example (E_1) and (E_2). Both have been constructed in a similar manner. Example (E_1) admits three controls u_1,



h	$\vert\hat{u} - \bar{u}_h\vert$	EOC
2^0	4.8466e-02	-
2^{-1}	3.3791e-02	0.52
2^{-2}	1.5241e-02	1.15
2^{-3}	6.1636e-03	1.31
2^{-4}	4.5072e-03	0.45
2^{-5}	7.4279e-04	2.60
2^{-6}	6.3083e-04	0.24
2^{-7}	2.0057e-04	1.65
2^{-8}	2.0373e-04	-0.02
-	-	$\varnothing \approx 1$

Figure 3.3.: Example (E_1): Experimental order of convergence

u_2, u_3, and the optimal state $\bar{y}$ is a linear combination of the three states $y_1, \dots, y_3$, each associated with a right-hand side e_i, i.e.

$$\bar{y} = \bar{u}_1 y_1 + \bar{u}_2 y_2 + \bar{u}_3 y_3.$$

On the other hand, in Example (E_2), we find that the optimal state $\bar{y}$ is simply a multiple of a state y_e associated with a right-hand side $e = e_1 + e_2 + e_3$, i.e.

$$\bar{y} = \bar{u} y_e = \bar{u}(y_1 + y_2 + y_3).$$

In Example (E_2), the optimal control $\bar{u}_h$ is therefore simply determined by $\bar{u}_h = b/\max(\bar{y}_e^h)$, where y_e^h is the finite element approximation of y_e. In Example (E_1) there seems to be more freedom for the controls (yet there uniqueness is guaranteed by the Tikhonov term in the objective function). This may serve as a descriptive explanation for the lower convergence rate. For convenience, we show the states y_i, $i = 1, \dots, 3$, as well as y_e in Figure 3.6.

We also point out the discussion in [104], where the nonuniform decrease in the error that can be observed for both Examples (E_1) and (E_2) has been explained by the specific choice of the mesh. More precisely, the distance between continuous and true active points seems to reflect itself in different constants. Consequently, not only the mesh size but also the location of the grid points seem to influence the convergence behavior. In [104], a simple one-dimensional example has been analyzed in this context.

We should also mention that for classical semi-infinite programming the specific choice of grid points is known to influence also the order of convergence, see. [136].

A nonconvex problem with semilinear PDE

Finally, we present an example with semilinear state equation and two control parameters, where we can observe the proven higher order of convergence.

$$\text{Minimize } \frac{1}{2}\int_{(0,1)^2} (y - y_d)^2 \, dx + \frac{1}{2}|u - u_d|^2 \qquad (E_3)$$

subject to

$$\begin{aligned} -\Delta y + y^3 &= u_1 e_1(x) + u_2 e_2(x) && \text{in } \Omega := (0,1)^2 \\ y &= 0 && \text{on } \Gamma \\ y &\leq b && \text{in } \Omega \end{aligned}$$

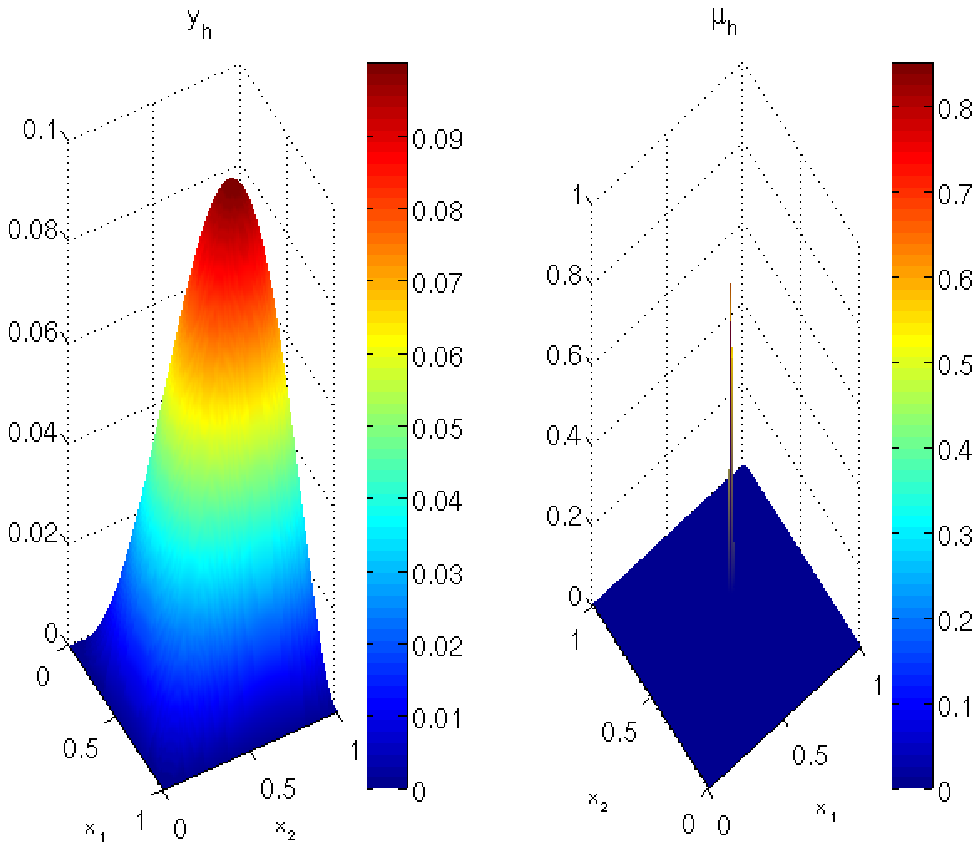


Figure 3.4.: Example (E_2): Discrete optimal state and Lagrange multiplier

with

$$y_d(x) = \sin(2\pi x_1)\sin(2\pi x_2), \quad e_1(x) = \sin(2\pi x_1)\sin(2\pi x_2), \quad e_2(x) = \sin^3(2\pi x_1)\sin^3(2\pi x_2)$$

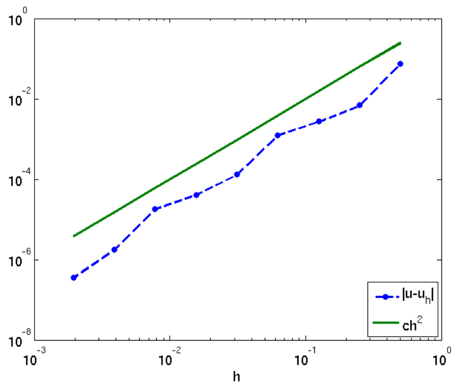
as well as

$$u_d = (8\pi^2,\, 1)^T \quad b = \frac{1}{2}.$$

Once again, the exact solution of this problem is not known, but constructed in a way that a problem without the pointwise state constraints would admit a solution $\bar{u} = u_d$ with the associated state $\bar{y} = y_d$ also admitting values larger than $1/2$. We proceed similar as before and compute a reference solution

$$\hat{u}_h \approx (50.6078,\, -17.2691)^T$$

on a fine grid with mesh size $h = 2^{-8}$. For this problem, we have to make use of a nonlinear PDE solver rather than precomputing approximate states y_i^h, $i = 1, 2$. We calculate discrete optimal solutions on a set of iteratively refined meshes with mesh size $h_k = 2^{-k}$, $k = 1, \ldots, 7$, with the help of `fmincon`. Figure 3.7 shows the optimal state and Lagrange multiplier computed on a mesh with $h = 2^{-6}$. Clearly, two active points are visible. From Theorem 3.4.19, we then expect the order of convergence $\mathcal{O}(h^2 \ln h)$. Figure 3.8 shows the error in the optimal control both graphically and in numbers, indicating quadratic convergence and thus underlining our theoretical results.



h	$\lvert\bar{u}-\bar{u}_h\rvert$	EOC
2^{-1}	7.3361e-02	-
2^{-2}	6.8567e-03	3.42
2^{-3}	2.7442e-03	1.32
2^{-4}	1.2353e-03	1.15
2^{-5}	1.3105e-04	3.24
2^{-6}	4.0846e-05	1.68
2^{-7}	1.8271e-05	1.16
2^{-8}	1.7941e-06	3.35
2^{-9}	3.5882e-07	2.32
-	-	$\varnothing \approx 2.2$

Figure 3.5.: Example (E_2): Experimental order of convergence

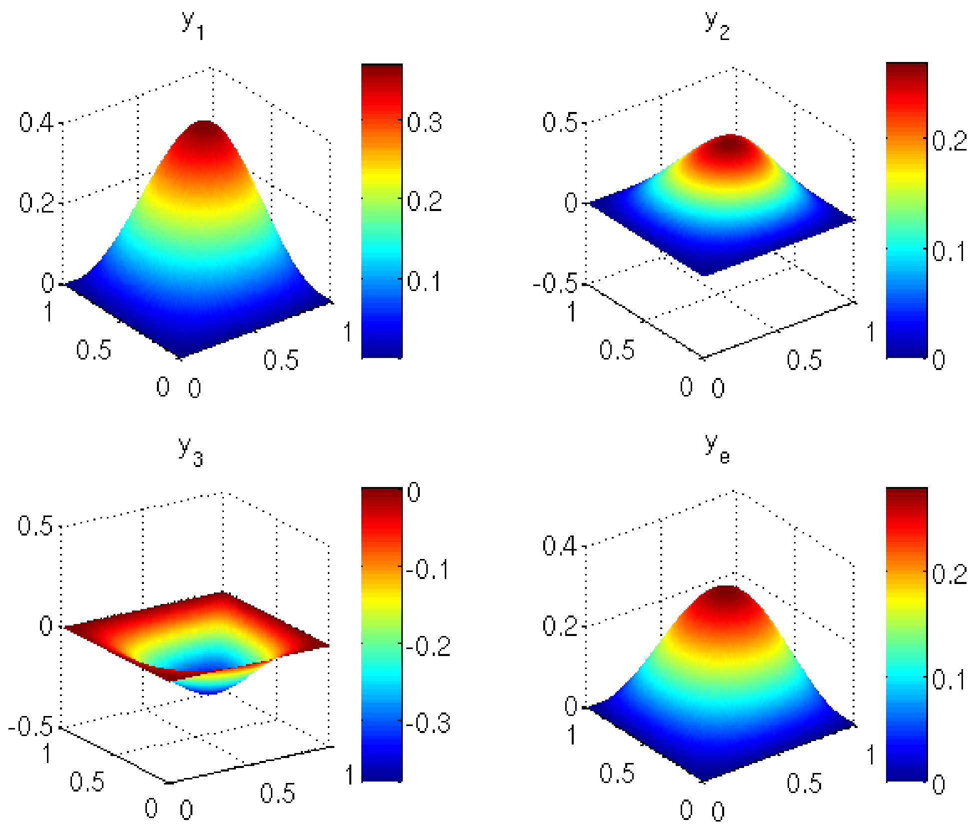


Figure 3.6.: Examples (E_1) and (E_2): States y_1, y_2, y_3 and $y_e := y_1 + y_2 + y_3$

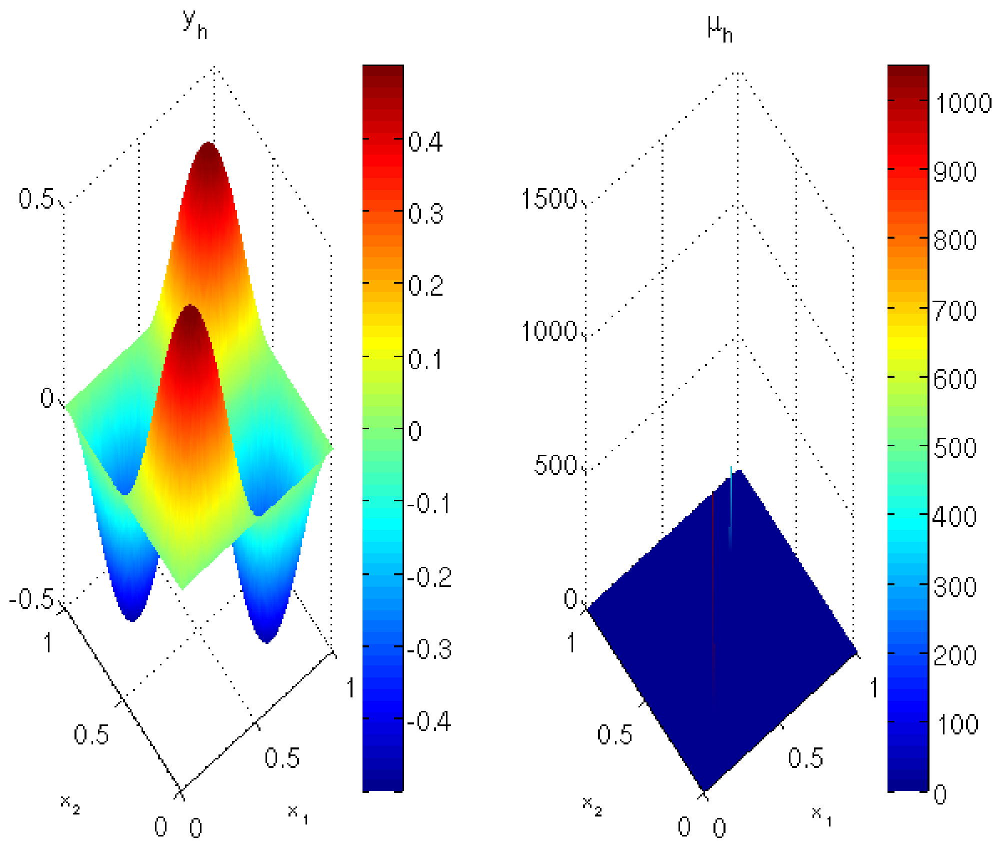
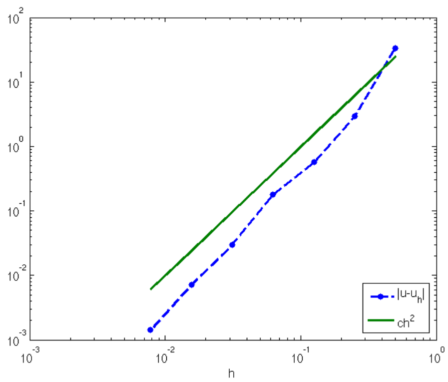


Figure 3.7.: Example (E_3): Discrete optimal state and Lagrange multiplier



h	$\lvert\bar{u} - \bar{u}_h\rvert$	EOC
2^{-1}	3.3726e+01	-
2^{-2}	2.9126e+00	3.53
2^{-3}	5.6799e-01	2.36
2^{-4}	1.8039e-01	1.65
2^{-5}	3.0145e-02	2.58
2^{-6}	7.1513e-03	2.08
2^{-7}	1.4294e-03	2.32
-	-	$\varnothing \approx 2.4$

Figure 3.8.: Example (E_3): Experimental order of convergence

4. A parabolic control problem with pointwise state and control constraints

4.1. Introduction

We now consider a model problem governed by a semilinear parabolic PDE and spatio-temporal control functions, subject to pointwise bounds on the state and the control. In the introduction and in the last chapter we have already discussed how pointwise state constraints present a challenge in the analysis of PDE-constrained optimal control problems. Compared to the semi-infinite problem (3.1.1) we now observe additional challenges. First, we now deal with control functions rather than finitely many control parameters. From the analytic point of view, the parabolic initial-boundary value problem demands higher regularity of the given data for comparable regularity results for the states. Due to the presence of pointwise bounds on the control, we have at least sufficient regularity of the states for a meaningful discussion of first order necessary optimality conditions, yet only limited results on second order sufficient conditions are available. From the viewpoint of finite element error estimates additional challenges come into play which we will comment on in more detail in the next chapter.

For these reasons we do not aim directly at proving error estimates for a finite element discretization, but discuss a regularization of the problem first. Let us give a brief overview about three regularization methods that have been studied frequently in the recent past. A first approach is the *Moreau-Yosida type regularization method* by Ito and Kunisch [78], where the pointwise state constraints

$$y_a \le y \le y_b$$

are penalized by a quadratic penalty functional and a regularized control constrained problem is obtained. The resulting objective function in its simplest form contains terms of the type

$$\frac{1}{2\gamma}\iint\limits_Q \max(0,\gamma(y_\gamma - y_b))^2\,dxdt + \frac{1}{2\gamma}\iint\limits_Q \max(0,\gamma(y_a - y_\gamma))^2\,dxdt$$

added to the unregularized objective function $J(y,u)$, where γ denotes a regularization parameter that is chosen large. After differentiation, this leads to a nondifferentiable term

$$\gamma \max(0, y_\gamma - y_b) - \gamma \max(0, y_a - y_\gamma)$$

in the right-hand side of the adjoint equation. A semi-smooth Newton approach has proven to work well in numerical computations, and allows a function space analysis of this problem, cf. [78]. However, since the objective function is not twice differentiable, a classical second-order analysis is not possible. A workaround has recently been presented by Krumbiegel, Rösch, and the author in [84]; we will comment on this shortly.

Later, Meyer, Rösch, and Tröltzsch, cf. [110], suggested a *Lavrentiev type method* that in some sense preserves the nature of state-constraints, replacing them by mixed control-state constraints

$$y_a \le \lambda u + y \le y_b,$$

where the regularization parameter λ is chosen small. We refer also to the nonlinear setting in [111]. Note that this method requires the control and state to be defined on the same sets. In e.g. boundary control problems, it cannot directly be applied. A possible extension was proven by Tröltzsch and the author in [116] for parabolic problems, and by Tröltzsch and Yousept, in [145, 146] for elliptic boundary control problems. We will focus in detail on Lavrentiev regularization for distributed control problems in this thesis.

A third regularization technique developed in the recent past is the *virtual control concept* by Krumbiegel and Rösch, cf. [85] originally developed for boundary control problems that has also proven useful for distributed control problems with additional control constraints, cf. [30]. This concept also transfers pure state constraints into mixed control-state constraints

$$y_a \leq \varepsilon v + y \leq y_b,$$

but in contrast to Lavrentiev regularization this is done for an additional distributed control, that may or may not enter the PDE as a source term. To the author's knowledge, there are no publications related to parabolic problems available.

All techniques have been discussed in more or less detail in the literature, with the main focus on elliptic problems with linear state equation and quadratic objective functional. Special emphasis was laid on the convergence analysis for the regularization parameter tending to the limit, cf. also the overview in Chapter 1. Not so many contributions were devoted to this issue in the nonlinear case. We mention [72], where some related questions for the Lavrentiev type regularization in the semilinear elliptic case are discussed, [31], where in addition a regularization error estimate has been proven, and [112], where the convergence of the Moreau-Yosida type regularization has been studied for a semilinear elliptic problem that arises from the control of the growth of SiC bulk single crystals. Convergence for the Moreau-Yosida regularization for control problems governed by semilinear parabolic equations has been studied by Tröltzsch and the author in [115]. For the virtual control concept, a contribution to the analysis of problems governed by semilinear elliptic equations has recently been published by Krumbiegel, Rösch, and the author, which not only focuses on convergence but also on second-order sufficient conditions, cf. [83]. For mixed pointwise control-state constraints, SSC have first been studied by Rösch and Tröltzsch et. al. in [133] for elliptic problems, and even earlier for mixed constraints of bottleneck type in [131].

To put the three regularization approaches into perspective, we point out the known fact that Lavrentiev regularization with a virtual control that does not enter the PDE itself, is equivalent to the penalization method by Ito and Kunisch. This fact has recently been used by Krumbiegel, Rösch, and the author to derive sufficient optimality conditions for the formally less regular Moreau-Yosida regularization in an elliptic setting, cf. [84]. When comparing the virtual control concept with classical Lavrentiev regularization, it stands out that the virtual control approach keeps the control constraints separate from the mixed control-state constraints, i.e. one obtains

$$u_a \leq u \leq u_b, \quad y_a \leq \varepsilon v + y \leq y_b, \tag{4.1.1}$$

where u and y do not appear in the same inequality. In contrast, for Lavrentiev regularization with

$$u_a \leq u \leq u_b, \quad y_a \leq \lambda u + y \leq y_b, \tag{4.1.2}$$

the control appears by itself in the control constraints and together with the state in the mixed control-state constraints. We will comment on some consequences of this difference between Lavrentiev type regularization and the virtual control concept later on.

In this chapter, we focus on Lavrentiev regularization of problems with semilinear parabolic equations and bilateral pointwise control and state constraints. The afore mentioned fact that the control and mixed control-state constraints are not strictly separated leads to interesting theoretical questions. We proceed as follows:

First, we give a brief overview about known results of the unregularized problem to motivate the reason for regularization. Along the way, we show properties of the solution operator that are needed in this and the following chapter. Secondly, we discuss in more detail the properties of the Lavrentiev-regularized problem formulation to demonstrate what can be gained from a regularized problem. That includes the regularity of Lagrange multipliers as well as the formulation of second-order sufficient optimality conditions. Third, we conduct a stability analysis for associated linear-quadratic problems. This is motivated by a number of reasons. First of all, the results are interesting by themselves, but they also give rise to a deeper analysis of nonlinear problems. We give more details later on. We end the chapter with a convergence result for vanishing Lavrentiev parameter and also state a regularization error estimate.

After all, the purpose of this chapter is twofold. While all results are worthwile to be discussed by themselves we also gain a better understanding of what makes the discussion of a parabolic-state-constrained problem especially difficult, what can be expected from regularization, and what complicates the discussion of a regularized problem. This will then be used in Chapter 5 to regularize a purely state-constrained problem with the ultimate goal to provide a priori error estimates for the finite element discretization of the resulting control-constrained problem. In view of this, we can interpret this chapter as an intermediate step towards an error estimate for the discretization.

Let us emphasize that the results presented in this chapter have been published by Tröltzsch and the author in [115] and [114].

4.2. The optimal control problem and its analysis

4.2.1. Problem formulation, assumptions, and notation

The problem under consideration is given by

$$\text{Minimize } J(y,u) := \iint\limits_Q \Psi(t,x,y,u)\,dxdt \tag{4.2.1a}$$

subject to the PDE constraint

$$\begin{aligned} \partial_t y + \mathcal{A}y + d(\cdot,y) &= u && \text{in } Q,\\ y(0,\cdot) &= y_0 && \text{in } \Omega,\\ y &= 0 && \text{on } \Sigma, \end{aligned} \tag{4.2.1b}$$

the pointwise state constraints

$$y_a(t,x) \le y(t,x) \le y_b(t,x) \quad \forall\,(t,x) \in \bar{Q}, \tag{4.2.1c}$$

and the control bounds

$$u_a \le u \le u_b \quad \text{a.e. in } Q. \tag{4.2.1d}$$

We rely on the following general assumption:

Assumption 4.2.1.

(A.1) *Let Assumption 2.3.1 on page 19 hold, and let the differential operator $\mathcal{A}$ satisfy* (2.3.3) *and* (2.3.4) *on page 19. Moreover, let the nonlinearity d fulfill the conditions stated in Assumption 2.5.3.*

(A.2) *The initial state y_0 is a function from $\mathcal{C}(\bar{\Omega}) \cap H_0^1(\Omega)$. The constraints on the control and state are given by $u_a, u_b \in L^\infty(Q)$, $u_b - u_a \geq c > 0$ a.e. in Q, and $y_a, y_b \in \mathcal{C}(\bar{Q})$, $y_a \leq y_b$ in $\bar{Q}$, respectively, such that $y_a(x,0) < y_0(x) < y_b(x,0)$ holds for all $x \in \bar{\Omega}$. We additionally assume $y_a(t,x) < 0 < y_b(t,x)$ for all $(t,x) \in \Sigma$, as well as $y_0(x,0) = 0$ on Γ.*

(A.3) *The function $\Psi = \Psi(t,x,y,u)\colon Q \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ fulfills Assumption 2.2.3, i.e. it is measurable with respect to $(t,x) \in Q$ for all fixed $(y,u) \in \mathbb{R}^2$, and twice continuously differentiable with respect to (y,u) for almost all $(t,x) \in Q$. Its first and second order derivative will be denoted by Ψ' and Ψ'', respectively. Moreover, for $y = u = 0$, the function Ψ, and its derivatives up to order two are bounded and Lipschitz continuous on bounded sets with respect to y and u, i.e. they fulfill conditions* (2.2.1) *and* (2.2.2) *accordingly.*

(A.4) *Moreover, the function Ψ is assumed to fulfill the Legendre-Clebsch condition*

$$\partial_{uu}\Psi(t,x,y,u) \geq \nu_0 > 0 \tag{4.2.2}$$

for almost all $(t,x) \in Q$, all $y \in \mathbb{R}$, and all $u \in [\inf u_a, \sup u_b]$.

(A.5) *Last, Ψ is convex in u, i.e.*

$$\Psi(t,x,y,\xi u + (1-\xi)v) \leq \xi\Psi(t,x,y,u) + (1-\xi)\Psi(t,x,y,v)$$

for almost all $(t,x) \in Q$, all $u, v \in \mathbb{R}$, and every $\xi \in (0,1)$.

The general representation of the objective function has immediate consequences when discussing e.g. the differentiability or Lipschitz continuity properties, which affects for instance the discussion of second order sufficient conditions for the optimal control problem and its regularized version to be discussed in the next chapter. In particular, we will generally not obtain L^2-differentiability of the objective function with respect to u.

Last, we again agree that $V := H_0^1(\Omega)$ and $Y := \mathcal{W}(0,T)$ is given, and for short notation we introduce the control space

$$U := L^\infty(Q)$$

and the set of admissible controls

$$U_{\mathrm{ad}} := \{u \in L^\infty(Q)\colon u_a \leq u \leq u_b\}.$$

We will also use the general notation introduced in Chapter 2 and in particular point out the short notation (2.5.1) on page 23 for some frequently appearing norms. In the remainder of this section, we proceed as in the elliptic case. We introduce a control-to-state operator and a reduced objective function, and with the help of their properties discuss the existence of optimal solutions, as well as first and second order optimality conditions.

4.2.2. The control-to-state operator and the reduced objective functional

The state equation (4.2.1b) fits into the general setting of the initial-boundary value problem (2.3.2) discussed in Chapter 2. Hence, Theorem 2.5.6 ensures that the following definition is meaningful:

Definition 4.2.2. *The mapping*

$$G\colon U \to \mathcal{W}(0,T)\cap \mathcal{C}(\bar{Q}), \quad u \mapsto y = G(u),$$

where y solves (4.2.1b) *in weak sense, is called the control-to-state-operator associated with Problem* (4.2.1). *In addition, we introduce the reduced objective functional*

$$f\colon U \to \mathbb{R}, \quad f(u) := J(G(u),u).$$

We proceed by collecting differentiability properties of G and f.

Proposition 4.2.3. *The control-to-state-operator G is of class $\mathcal{C}^2$. Let $v, v_1, v_2 \in L^\infty(Q)$ and $y = G(u)$ be given. Then, the first and second order derivatives of G are given by $\tilde{y} := G'(u)v$ and $\tilde{z} := G''(u)v_1v_2$ being the solutions of*

$$\begin{aligned} \partial_t \tilde{y} + \mathcal{A}\tilde{y} + \partial_y d(\cdot,y)\tilde{y} &= v \quad \text{in } Q \\ \tilde{y}(0,\cdot) &= 0 \quad \text{in } \Omega \\ \tilde{y} &= 0 \quad \text{on } \Sigma, \end{aligned} \tag{4.2.3}$$

$$\begin{aligned} \partial_t \tilde{z} + \mathcal{A}\tilde{z} + \partial_y d(\cdot,y)\tilde{z} &= -\partial_{yy} d(\cdot,y)\tilde{y}_1\tilde{y}_2 \quad \text{in } Q \\ \tilde{z}(0,\cdot) &= 0 \quad \text{in } \Omega \\ \tilde{z} &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{4.2.4}$$

where $\tilde{y}_1 = G'(u)v_1$ and $\tilde{y}_2 = G'(u)v_2$. The reduced objective function f is of class $\mathcal{C}^2$ with respect to the L^∞-topology. For $u, v, v_1, v_2 \in L^\infty$ and $\tilde{y}_i := G'(u)v_i$, $i = 1,2$, as well as $z = G''(u)[v_1,v_2]$ its first and second order derivative are given by

$$\begin{aligned} f'(u)v &= \iint_Q \left(\partial_y \Psi(t,x,y,u)\tilde{y} + \partial_u \Psi(t,x,y,u)v\right) dxdt \\ f''(u)[v_1,v_2] &= \iint_Q \begin{bmatrix} \tilde{y}_1 \\ v_1 \end{bmatrix}^T \begin{bmatrix} \partial_{yy}\Psi(t,x,y,u) & \partial_{yu}\Psi(t,x,y,u) \\ \partial_{uy}\Psi(t,x,y,u) & \partial_{uu}\Psi(t,x,y,u) \end{bmatrix} \begin{bmatrix} \tilde{y}_2 \\ v_2 \end{bmatrix} dxdt \\ &+ \iint_Q \partial_y \Psi(t,x,y,u) z \, dxdt. \end{aligned}$$

Proof. For the differentiability of G, we refer to Theorems 5.9, 5.15, and 5.16 in [144]. Moreover, by [144, Lemma 4.12], we know that the objective function $J\colon L^\infty(Q)\times L^\infty(Q) \to \mathbb{R}$ is twice continuously differentiable. Then, by the properties of G and the chain rule, the assertion is obtained. □

The estimates from Proposition 2.5.2 hold for $\tilde{y}$ and $\tilde{z}$, since $\partial_y d(\cdot,y)$ is bounded by Assumption 4.2.1. The following Lipschitz stability results will be helpful in the sequel.

Lemma 4.2.4. *Let $u_1, u_2 \in U_{ad}$ and $v \in L^\infty(Q)$ be given. Then there exists a constant $C > 0$ such that*

$$\|G(u_1) - G(u_2)\|_{L^2(Q)\cap L^\infty(I,V)} \le C\,\|u_1 - u_2\|_I \tag{4.2.5}$$

$$\|G(u_1) - G(u_2)\|_{\infty,\infty} \le C\,\|u_1 - u_2\|_{L^p(Q)} \tag{4.2.6}$$

$$\|G'(u_1)v - G'(u_2)v\|_I \le C\,\|u_1 - u_2\|_I \|v\|_I \tag{4.2.7}$$

$$\|G'(u_1)v - G'(u_2)v\|_{\infty,\infty} \le C\,\|u_1 - u_2\|_{\infty,\infty} \|v\|_{L^p(Q)} \tag{4.2.8}$$

$$\|G''(u_1)[v,v] - G''(u_2)[v,v]\|_I \le C\,\|u_1 - u_2\|_I \|v\|_I^2 \tag{4.2.9}$$

is fulfilled for $p > n/2 + 1$. Moreover, there exists a constant C, such that for all $u_1, u_2 \in U_{ad}$ and all $v \in L^\infty(Q)$

$$\begin{aligned} |f(u_1) - f(u_2)| &\leq C\,\|u_1 - u_2\|_I & (4.2.10)\\ |f'(u_1)v - f'(u_2)v| &\leq C\,\|u_1 - u_2\|_I \|v\|_I & (4.2.11)\\ |f''(u_1)[v,v] - f''(u_2)[v,v]| &\leq C\,\|u_1 - u_2\|_{\infty,\infty} \|v\|_I^2 & (4.2.12) \end{aligned}$$

is satisfied.

Proof. As in Lemma 3.1.4, we postpone the proof to the Appendix, since it is rather technical. □

4.2.3. The control reduced problem formulation and existence of solutions

With the control-to-state operator at hand the optimal control problem (4.2.1a)–(4.2.1d) is equivalent to

$$\text{Minimize } f(u) \text{ subject to } u \in U_{\text{ad}}, \quad y_a \leq G(u) \leq y_b, \tag{$\mathbb{P}$}$$

The existence of optimal controls is now easily discussed. In addition to the set of admissible controls we define the set of feasible controls

$$U_{\text{feas}} := \{u \in U_{\text{ad}} \,|\, y_a \leq G(u) \leq y_b \text{ in } \bar{Q}\}.$$

Then the following theorem is a standard conclusion under our general Assumption 4.2.1.

Theorem 4.2.5. *If the set of feasible controls, U_{feas}, is not empty, the optimal control problem ($\mathbb{P}$) admits at least one (globally) optimal control $\bar{u} \in U_{feas}$ with associated optimal state $\bar{y} = G(\bar{u})$.*

Proof. The proof follows with the technique of proof used in [144]. Let us only point out the main ingredients and differences to e.g. the model problem in Chapter 3. Since the control is not finite dimensional, the weak compactness of the feasible set does not automatically imply the existence of an optimal control. When discussing a minimizing sequence of controls, one important step in the proof of existence is to show that the sequence of associated states converges accordingly. For that reason, we only choose nonlinearities $d = d(t, x, y)$ that do not depend on u. Let us mention, though, that control functions appearing linearly in d could be handled. An additional issue is the convergence of the objective function. Here, the convexity of Ψ with respect to u stated in Assumption 4.2.1 is needed. □

Since we may encounter the existence of multiple locally optimal controls we introduce the notation of a local solution, which we will also denote by $\bar{u}$.

Definition 4.2.6. *A feasible control $\bar{u} \in U_{feas}$ is called a local solution of (4.2.1) in the sense of $L^p(Q)$ if there exists a positive real number ε such that $f(\bar{u}) \leq f(u)$ holds for all $u \in U_{feas}$ of ($\mathbb{P}$) with $\|u - \bar{u}\|_{L^p(Q)} \leq \varepsilon$.*

It is reasonable to consider local solutions in L^p-spaces, since all controls from U_{ad} are functions in $L^p(Q)$ for any $1 \leq p \leq \infty$. However, the formulation of sufficient optimality conditions for the general setting considered in this chapter is only available for $p = \infty$. The reason for this lies in the low regularity properties of parabolic initial boundary value problems, cf. Chapter 2.

4.2.4. First and second order optimality conditions

We again assume a constraint qualification of Slater type.

Assumption 4.2.7. *We say that $\bar{u}$ satisfies the linearized Slater condition for* $(\mathbb{P})$ *if there exist a point $u_\gamma \in U_{ad}$ and a fixed positive real number γ such that*

$$y_a(t,x) + \gamma \leq G(\bar{u})(t,x) + G'(\bar{u})(u_\gamma - \bar{u})(t,x) \leq y_b(t,x) - \gamma \quad \forall (t,x) \in \bar{Q}. \tag{4.2.13}$$

In addition, let

$$u_a(t,x) + \gamma \leq \bar{u}(t,x) \leq u_b(t,x) - \gamma$$

be satisfied.

The Slater condition with respect to the control constraints will be useful in the future discussion of the regularized problem. Now, we expect Lagrange multipliers with respect to the state constraints to exist in the space $\mathcal{M}(\bar{Q})$ of regular Borel measures on $\bar{Q}$. First order optimality conditions in form of the Pontryagin maximum principle have been derived by Casas [19] and Raymond and Zidani [125]. The following theorem follows immediately from the Pontryagin principle in [19]:

Theorem 4.2.8. *Let $\bar{u} \in U_{feas}$ with associated optimal state $\bar{y} = G(\bar{u})$ in $\mathcal{W}(0,T) \cap C(\bar{Q})$ be an optimal solution of the control problem* $(\mathbb{P})$ *in the sense of Definition 4.2.6. Assume that $u_\gamma \in U_{ad}$ exists such that the linearized Slater condition* (4.2.13) *from Assumption 4.2.7 is satisfied. Then there exist regular Borel measures $\bar{\mu}_a, \bar{\mu}_b \in \mathcal{M}(\bar{Q})$ and an adjoint state $\bar{p} \in L^r(0,T;W^{1,s}(\Omega))$, for all $r,s \in [1,2)$ with $\frac{2}{r} + \frac{n}{s} > n+1$ with*

$$\begin{aligned} -\partial_t \bar{p} + \mathcal{A}\bar{p} + \partial_y d(\cdot,\bar{y})\bar{p} &= \partial_y \Psi(\cdot,\bar{y},\bar{u}) + \bar{\mu}_{b_Q} - \bar{\mu}_{a_Q} && \text{in } Q \\ \bar{p}(\cdot,T) &= \bar{\mu}_{b_T} - \bar{\mu}_{a_T} && \text{in } \Omega \\ \bar{p} &= 0 && \text{on } \Sigma, \end{aligned}$$

in the sense of (2.5.7),

$$\iint\limits_Q (\partial_u \Psi(t,x,\bar{y},\bar{u}) + \bar{p})(u - \bar{u})\, dxdt \geq 0 \quad \forall u \in U_{ad}, \tag{4.2.14}$$

$$\iint\limits_{\bar{Q}} (y_a - \bar{y}) d\bar{\mu}_a(t,x) = 0, \quad \bar{\mu}_a \geq 0, \qquad \iint\limits_{\bar{Q}} (\bar{y} - y_b) d\bar{\mu}_b(t,x) = 0, \quad \bar{\mu}_b \geq 0, \tag{4.2.15}$$

where $\bar{\mu}_{i_Q} = \bar{\mu}_{i_{|Q}}$, $\bar{\mu}_{i_T} = \bar{\mu}_{i_{|\bar{\Omega}\times T}}$, $i \in \{a,b\}$, denote the restrictions of $\bar{\mu}_a$, $\bar{\mu}_b$ to the indicated sets.

Note that $\bar{\mu}_{a_{|0}} = \bar{\mu}_{b_{|0}} = 0$ as well as $\bar{\mu}_{a_{|\Sigma}} = \bar{\mu}_{b_{|\Sigma}} = 0$ follows from Assumption 4.2.1. We again introduce a Lagrange functional $\mathcal{L}\colon L^\infty(Q) \times \mathcal{M}(\bar{Q}) \times \mathcal{M}(\bar{Q}) \to \mathbb{R}$,

$$\mathcal{L}(u,\mu_a,\mu_b) := f(u) + \iint\limits_{\bar{Q}} (y_a - G(u))\, d\mu_a(t,x) + \iint\limits_{\bar{Q}} (G(u) - y_b)\, d\mu_b(t,x). \tag{4.2.16}$$

By Proposition 4.2.3, $\mathcal{L}$ is twice continuously differentiable with respect to u. Again, we write $\mathcal{L}'$ instead of $\partial\mathcal{L}/\partial u$ as well as $\mathcal{L}''$ instead of $\partial^2\mathcal{L}/\partial u^2$. For arbitrary $u, v, v_1, v_2 \in L^\infty(Q)$ and $\mu_a, \mu_b \in \mathcal{M}(\bar{Q})$, its first and second order derivatives can be determined to fulfill

$$\begin{aligned}
\mathcal{L}'(u,\mu_a,\mu_b)v &= f'(u)v - \iint_{\bar{Q}} G'(u)v\,d\mu_a(t,x) + \iint_{\bar{Q}} G'(u)v\,d\mu_b(t,x)\\
&= \iint_{\bar{Q}} (\partial_u \Psi(t,x,y,u)v + pv)\,dxdt\\
\mathcal{L}''(u,\mu_a,\mu_b)[v_1,v_2] &= f''(u)[v_1,v_2] - \iint_{\bar{Q}} G''(u)[v_1,v_2]\,d\mu_a(t,x) + \iint_{\bar{Q}} G''(u)[v_1,v_2]\,d\mu_b(t,x)\\
&= \iint_{Q} \begin{bmatrix} \tilde{y}_1 \\ v_1 \end{bmatrix}^T \begin{bmatrix} \partial_{yy}\Psi(t,x,y,u) & \partial_{yu}\Psi(t,x,y,u) \\ \partial_{uy}\Psi(t,x,y,u) & \partial_{uu}\Psi(t,x,y,u) \end{bmatrix} \begin{bmatrix} \tilde{y}_2 \\ v_2 \end{bmatrix} dxdt\\
&\quad - \iint_{Q} p\partial_{yy} d(t,x,y)[\tilde{y}_1,\tilde{y}_2]\,dxdt,
\end{aligned}$$

where p solves (4.2.14). We refer to [144] for a detailed discussion.

Now, Theorem 4.2.8 is equivalent to the existence of nonnegative Lagrange multipliers $\bar{\mu}_a, \bar{\mu}_b \in \mathcal{M}(\bar{Q})$ such that

$$\mathcal{L}'(\bar{u},\bar{\mu}_a,\bar{\mu}_b)(u-\bar{u}) \geq 0 \quad \forall\, u \in U_{\text{ad}} \tag{4.2.17}$$

and

$$\iint_{\bar{Q}} (y_a - G(\bar{u}))\,d\bar{\mu}_a(t,x) = \iint_{\bar{Q}} (G(\bar{u}) - y_b)\,d\bar{\mu}_b(t,x) = 0. \tag{4.2.18}$$

Due to the nonconvexity of the problem it is desirable to consider the formulation of second order sufficient optimality conditions. There are very few publications concerned with parabolic problems with pure pointwise state constraints. For unregularized parabolic semilinear problems, SSC have only been proven in full generality for one-dimensional distributed control, cf. [123] and [22]. For higher dimensions, SSC have been established in the special setting with finitely many time-dependent controls in the recent work [36].

Assumption 4.2.9. *Let $\bar{u} \in U_{ad}$ be a control satisfying the first order necessary optimality conditions of Theorem 4.2.8 and let $\bar{\mu}_a, \bar{\mu}_b$ be Lagrange multipliers with respect to the state constraints* (4.2.1c). *We assume that there exists a constant $\alpha > 0$, such that*

$$\frac{\partial^2 \mathcal{L}}{\partial u^2}(\bar{u},\bar{\mu}_a,\bar{\mu}_b)[v,v] \geq \alpha \|v\|_I^2$$

is valid for all $v \in L^\infty(Q)$.

Just as in the elliptic setting it is a quite strong requirement to demand this definiteness property for all directions $v \in L^\infty(Q)$, and it is possible to weaken this assumption along the lines of [36]. However, there are dimensional restrictions that apply even in this setting.

The following result is a consequence of Lemma 4.2.4. It will be helpful in the sequel and at the same time demonstrate the dimensional limits in the parabolic setting.

Corollary 4.2.10. *Let the space dimension n be equal to one. For all fixed $\mu_a, \mu_b \in \mathcal{M}(\bar{Q})$, arbitrary $u_1, u_2 \in U_{ad}$ and $v \in L^\infty(Q)$, there exists a constant $C > 0$ such that*

$$|\mathcal{L}''(u_1,\mu_a,\mu_b)[v,v] - \mathcal{L}''(u_2,\mu_a,\mu_b)[v,v]| \leq C\|u_1 - u_2\|_{\infty,\infty}\|v\|_I^2.$$

Proof. Due to Lemma 4.2.4, we only need to consider the terms associated with the Lagrange multipliers. Observe that

$$\begin{aligned} \left| \iint\limits_{\bar{Q}} \left(G''(u_1) - G''(u_2)\right)[v,v]\, d\mu_a(t,x) \right| &\leq c\|\left(G''(u_1) - G''(u_2)\right)[v,v]\|_{\mathcal{C}(\bar{Q})}\|\mu_a\|_{\mathcal{M}(Q)} \\ &\leq c\|\left(G''(u_1) - G''(u_2)\right)[v,v]\|_{\mathcal{C}(\bar{Q})} \end{aligned}$$

since μ_a is fixed. Now, since $n = 1$, we obtain

$$\|(G''(u_1) - G''(u_2))[v,v]\|_{\mathcal{C}(\bar{Q})} \leq c\|u_1 - u_2\|_I\|v\|_I^2,$$

as in Lemma 4.2.4, and the assertion is obtained after an analogous discussion of the term containing μ_b. □

Let us point out that for higher spatial dimensions, $p = 2$ is not sufficient for estimation of the $\mathcal{C}(\bar{Q})$ norms. In view of second-order sufficient conditions, choosing stronger norms for v is not satisfactory, either.

Theorem 4.2.11. *Let $n = 1$ be satisfied and let the control $\bar{u} \in U_{feas}$ and the Lagrange multipliers $\bar{\mu}_a, \bar{\mu}_b$ fulfill the conditions stated in Assumption 4.2.9. Then there exist constants $\beta > 0$ and $\varepsilon > 0$ such that*

$$f(\bar{u}) \leq f(u) - \beta\|u - \bar{u}\|_I^2$$

is satisfied for all $u \in U_{feas}$ with $\|u - \bar{u}\|_{\infty,\infty} \leq \varepsilon$.

Proof. The steps of the proof are the same as in the elliptic setting in Theorem 5.3.12. Note that by the feasibility of u, the positivity of $\bar{\mu}_a$ and $\bar{\mu}_b$ and the complementary slackness conditions (4.2.18) from Theorem 4.2.8 we have

$$\iint\limits_{\bar{Q}} (G(u) - y_b)\, d\bar{\mu}_b(t,x) - \iint\limits_{\bar{Q}} (y_a - G(u))\, d\bar{\mu}_a(t,x) \leq 0,$$

and

$$\iint\limits_{\bar{Q}} (G(\bar{u}) - y_b)\, d\bar{\mu}_b(t,x) - \iint\limits_{\bar{Q}} (y_a - G(\bar{u}))\, d\bar{\mu}_a(t,x) = 0,$$

and therefore

$$f(\bar{u}) = \mathcal{L}(\bar{u}, \bar{\mu}_a, \bar{\mu}_b), \quad f(u) \geq \mathcal{L}(u, \bar{\mu}_a, \bar{\mu}_b).$$

Consequently, we obtain by Taylor expansion of $\mathcal{L}$ in $\bar{u}$

$$\begin{aligned} f(\bar{u}) - f(u) \leq\; & \mathcal{L}(\bar{u}, \bar{\mu}_a, \bar{\mu}_b) - \mathcal{L}(u, \bar{\mu}_a, \bar{\mu}_b) \\ &= -\mathcal{L}'(\bar{u}, \bar{\mu}_a, \bar{\mu}_b)(u - \bar{u}) - \frac{1}{2}\mathcal{L}''(u_\xi, \bar{\mu}_a, \bar{\mu}_b)[u - \bar{u}, u - \bar{u}] \end{aligned}$$

with $u_\xi = \bar{u} + \xi(u - \bar{u})$ for some $\xi \in (0,1)$. Making use of (4.2.17) and Assumption 5.3.9 we obtain

$$\begin{aligned} f(\bar{u}) - f(u) &\leq -\frac{1}{2}\mathcal{L}''(\bar{u}, \bar{\mu}_a, \bar{\mu}_b)[u - \bar{u}, u - \bar{u}] - \frac{1}{2}(\mathcal{L}''(\bar{u}, \bar{\mu}_a, \bar{\mu}_b) - \mathcal{L}''(u_\xi, \bar{\mu}_a, \bar{\mu}_b))[u - \bar{u}, u - \bar{u}] \\ &\leq -\frac{\alpha}{2}\|u - \bar{u}\|_I^2 + \frac{c}{2}\|\bar{u} - u_\xi\|_{\infty,\infty}\|u - \bar{u}\|_I^2, \end{aligned}$$

where the last inequality follows from Corollary 4.2.10. The fact that $u_\xi - \bar{u} = \xi(u - \bar{u})$ yields the assertion for $\|u - \bar{u}\|_{\infty,\infty} \leq \varepsilon$. We refer also to the formulation of weaker second order sufficient conditions in [22]. □

4.3. Lavrentiev regularization

In the last section, as well as in the semi-infinite model problem in Chapter 3, we have seen that the boundedness, and even better continuity of the state is essential to guarantee the existence of a Lagrange multiplier. The situation is different for mixed control-state constraints, such as e.g.

$$y_a \leq \lambda u + y \leq y_b.$$

Even though this constraint has to be considered in the sense of L^p with $p < \infty$ if u is from such a space, the existence of regular Lagrange multipliers can be shown by e.g. duality arguments, see for instance [110, 142] in certain situations. If $\lambda > 0$ takes the role of a regularization parameter tending to zero, we call these mixed control-state constraints *Lavrentiev regularized constraints*. This is motivated for an elliptic problem in [110] by considering a control-to-state operator with range in $L^2(Q)$, i.e. $S := i_0 G$, where $i_0 \colon Y \to L^2(Q)$ denotes the usual embedding operator. The operator equation

$$S(u) = y_b,$$

which is observed wherever the state constraints are active, is ill-posed, since S is compact. Lavrentiev regularization, [88], leads instead to the well-posed equation

$$\lambda u + S(u) = y.$$

We will consider a Lavrentiev regularized version of Problem $(\mathbb{P})$, which we will denote in short by $(\mathbb{P}_\lambda)$. Note that Lavrentiev regularization does not imply any changes in the underlying PDE, and we can make use of the control-to-state-mapping G and all its previously shown properties. Also, there are no changes in the objective function, and we obtain the regularized problem formulation

$$\text{Minimize } f(u) \quad \text{subject to } u \in U_{\text{ad}}, \quad y_a \leq \lambda u + G(u) \leq y_b. \tag{$\mathbb{P}_\lambda$}$$

We begin by analyzing Problem $(\mathbb{P}_\lambda)$ for fixed $\lambda > 0$ and we will not make use of any properties of the unregularized Problem (4.2.1). This is to illustrate the properties gained by regularization. In Section 4.6, we will address questions of convergence as well as regularization error estimates. Then, we will also show how and under which conditions certain assumptions on the family of regularized problems can be obtained from associated properties of $(\mathbb{P})$.

4.3.1. Existence of solutions

The question of existence of optimal controls does not differ much from the unregularized setting. In principle, only the definition of the set of feasible controls needs to be adapted.

Definition 4.3.1. *For fixed $\lambda > 0$, we denote by*

$$U_{feas}^{\lambda} = \{u \in U_{ad} \mid y_a \leq \lambda u + G(u) \leq y_b \text{ a.e. in } Q\}$$

the set of feasible controls for $(\mathbb{P}_\lambda)$.

The existence of an optimal control now follows by similar arguments as in the proof of Theorem 4.2.5.

Theorem 4.3.2. *If the set of feasible controls, U_{feas}^{λ}, is not empty, the optimal control problem $(\mathbb{P}_\lambda)$ admits at least one (globally) optimal control $\bar{u}_\lambda \in U_{ad}$ with associated optimal state $\bar{y}_\lambda = G(\bar{u}_\lambda)$.*

Proof. We again refer to the technique of proof used in [144]. □

Once again, we may not expect uniqueness of the solution and and introduce a definition of local solutions.

Definition 4.3.3. *Let $\lambda > 0$ be given. A function $\bar{u}_\lambda \in U^\lambda_{feas}$ is called a local solution of $(\mathbb{P}_\lambda)$ in the sense of $L^p(Q)$, if $f(\bar{u}_\lambda) \leq f(u)$ is satisfied for all $u \in U^\lambda_{feas}$ with $\|u - \bar{u}_\lambda\|_{L^p(Q)} \leq \varepsilon$.*

4.3.2. First and second order optimality conditions

The existence of regular Lagrange multipliers associated with mixed control-state constraints is guaranteed by [132] under certain conditions. The proof of existence therein follows two principal steps. First of all, existence of irregular multipliers in $L^\infty(Q)^*$ is shown under the assumption of a linearized Slater condition. Then, under a structural assumption on the active sets, these multipliers are shown to belong to $L^\infty(Q)$. For that purpose, we introduce the following definition.

Definition 4.3.4. *Let σ be a positive real number. For a control $\tilde{u} \in U^\lambda_{feas}$ we introduce the σ-active sets for Problem $(\mathbb{P}_\lambda)$:*

$$\begin{aligned}
M^{\sigma,\lambda}_{u,a}(\tilde{u}) &:= \{(t,x) \in Q : \ \tilde{u}(t,x) \leq u_a(t,x) + \sigma\} \\
M^{\sigma,\lambda}_{u,b}(\tilde{u}) &:= \{(t,x) \in Q : \ \tilde{u}(t,x) \geq u_b(t,x) - \sigma\} \\
M^{\sigma,\lambda}_{y,a}(\tilde{u}) &:= \{(t,x) \in Q : \ \lambda\tilde{u}(t,x) + G(\tilde{u})(t,x) \leq y_a(t,x) + \sigma\} \\
M^{\sigma,\lambda}_{y,b}(\tilde{u}) &:= \{(t,x) \in Q : \ \lambda\tilde{u}(t,x) + G(\tilde{u})(t,x) \geq y_b(t,x) - \sigma\}.
\end{aligned}$$

Assumption 4.3.5. *We assume that there exists $\sigma > 0$ such that the $\sigma-$active sets associated with $\bar{u}_\lambda$ according to Definition (4.3.4) fulfill the condition*

$$(M^{\sigma,\lambda}_{u,a}(\bar{u}_\lambda) \cup M^{\sigma,\lambda}_{y,a}(\bar{u}_\lambda)) \cap (M^{\sigma,\lambda}_{u,b}(\bar{u}_\lambda) \cup M^{\sigma,\lambda}_{y,b}(\bar{u}_\lambda)) = \emptyset$$

for all λ sufficiently small.

Note that for σ sufficiently small, we automatically have

$$M^{\sigma,\lambda}_{u,a}(\bar{u}_\lambda) \cap M^{\sigma,\lambda}_{u,b}(\bar{u}_\lambda) = \emptyset = M^{\sigma,\lambda}_{y,a}(\bar{u}_\lambda) \cap M^{\sigma,\lambda}_{y,b}(\bar{u}_\lambda)$$

due to the conditions on the constraints stated in Assumption 4.2.1, as well as the continuity of the state $\bar{y}_\lambda$. In Assumption 4.3.5, we additionally require that the upper control constraint may not be active, or almost active, wherever the mixed control-state constraints are close to the lower bound, and vice versa. This assumption is somewhat strong in two ways. On the one hand, this assumption must hold for the whole family of regularized solutions, i.e. we have to explicitely assume this for all $\lambda \leq \lambda_0$ for some $\lambda_0 > 0$, even with the convergence results from Section 4.6. This is due to the fact that the optimal controls $\bar{u}_\lambda$ will not necessarily converge in $L^\infty(Q)$, which in turn makes it impossible to characterize the almost active sets $M^{\sigma,\lambda}_{u,a}(\bar{u}_\lambda)$ and $M^{\sigma,\lambda}_{u,b}(\bar{u}_\lambda)$ associated with the control constraints for λ tending to zero. On the other hand, there might of course be cases were Assumption 4.3.5 seems unrealistic. However, there is also a wide range of problems that satisfy Assumption 4.3.5, the simplest being problems with solution that remains bounded even without explicit control constraints and any (artificial) bounds are chosen so that they remain inactive.

Additionally, we assume a Slater type constraint qualification similar to the unregularized setting.

Assumption 4.3.6. *Let $\lambda > 0$ be fixed. We say that $\bar{u}_\lambda$ satisfies the linearized Slater condition for $(\mathbb{P}_\lambda)$ if there exist a point $u_\gamma^\lambda \in U_{ad}$ and a fixed positive real number γ_λ such that*

$$u_a + \gamma_\lambda \le u_\gamma^\lambda \le u_b - \gamma_\lambda, \quad y_a + \gamma_\lambda \le \lambda u_\gamma^\lambda + G(\bar{u}_\lambda) + G'(\bar{u}_\lambda)(u_\gamma^\lambda - \bar{u}_\lambda) \le y_b - \gamma_\lambda$$

are satisfied.

We point out that for controls $\bar{u}_\lambda$ in the vicinity of a control $\bar{u}$ that satisfies Assumption 4.2.7, this can easily be proven to hold. We will do this in Section 4.6. Then we obtain the following theorem concerning first order optimality conditions by applying the results from [134], where for the first time a completely constrained problem, i.e. a problem with bilateral control and state constraints has been considered. The main statement is that the Lagrange multipliers associated with the regularized state constraints are regular functions in $L^\infty(Q)$. This is obtained after an intermediate existence result for a Lagrange multiplier in $L^\infty(Q)^*$ relying on the Slater condition from Assumption 4.3.6, that is proven to be more regular under the separation condition from Assumption 4.3.5.

Theorem 4.3.7 ([134]). *Let $\lambda > 0$ be fixed and let $\bar{u}_\lambda$ be a fixed local solution to $(\mathbb{P}_\lambda)$ in the sense of Definition 4.3.3. If Assumptions 4.3.6 and 4.3.5 are satisfied, then there exist regular nonnegative Lagrange multipliers $\bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda \in L^\infty(Q)$ and an adjoint state $\bar{p}_\lambda \in \mathcal{W}(0,T) \cap \mathcal{C}(\bar{Q})$, such that*

$$\begin{array}{rcll} -\partial_t \bar{p}_\lambda + \mathcal{A}\bar{p}_\lambda + \partial_y d(\cdot, \bar{y}_\lambda)\bar{p}_\lambda &=& \left(\partial_y \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda\right) & \text{in } Q \\ \bar{p}_\lambda(T, \cdot) &=& 0 & \text{in } \Omega \\ \bar{p}_\lambda &=& 0 & \text{on } \Sigma, \end{array} \tag{4.3.1}$$

$$\iint\limits_Q (\partial_u \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{p}_\lambda + \lambda(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda))(u - \bar{u}_\lambda)\, dxdt \ge 0 \ \forall u \in U_{ad} \tag{4.3.2}$$

$$\iint\limits_Q \bar{\mu}_a^\lambda(\lambda \bar{u}_\lambda + \bar{y}_\lambda - y_a)\, dxdt = 0 = \iint\limits_Q \bar{\mu}_b^\lambda(\lambda \bar{u}_\lambda + \bar{y}_\lambda - y_b)\, dxdt \tag{4.3.3}$$

is satisfied.

Note that $\bar{\mu}_{a_{|0}} = \bar{\mu}_{b_{|0}} = 0$ as well as $\bar{\mu}_{a_{|\Sigma}} = \bar{\mu}_{b_{|\Sigma}} = 0$ follows from Assumption 4.2.1. We introduce the regularized Lagrange functional $\mathcal{L}_\lambda \colon L^\infty(Q) \times L^\infty(Q) \times L^\infty(Q) \to \mathbb{R}$,

$$\mathcal{L}_\lambda(u, \mu_a^\lambda, \mu_b^\lambda) := f(u) + \iint\limits_Q (y_a - \lambda u - G(u))\mu_a^\lambda\, dx\, dt + \iint\limits_Q (\lambda u + G(u) - y_b)\mu_b^\lambda\, dxdt, \tag{4.3.4}$$

which is of class $\mathcal{C}^2$ with respect to u by Proposition 4.2.3. The derivatives for arbitrary $u \in U$ and

$v, v_1, v_2 \in L^\infty(Q)$ and fixed $\mu_a^\lambda, \mu_b^\lambda \in L^\infty(Q)$ are given by

$$\begin{aligned}
\mathcal{L}'_\lambda(u, \mu_a^\lambda, \mu_b^\lambda)v = &\ f'(u)v + \iint\limits_Q G'(u)v(\mu_b^\lambda - \mu_a^\lambda)\,dxdt \\
= &\ \iint\limits_{\bar{Q}} (\partial_u \Psi(t, x, y_\lambda, u)v + p_\lambda v)\ dxdt \\
\mathcal{L}''_\lambda(u, \mu_a^\lambda, \mu_b^\lambda)[v_1, v_2] = &\ f''(u)[v_1, v_2] + \iint\limits_Q G''(u)[v_1, v_2](\mu_b^\lambda - \mu_a^\lambda)\,dxdt \\
= &\ \iint\limits_Q \begin{bmatrix} \tilde{y}_1^\lambda \\ v_1 \end{bmatrix}^T \begin{bmatrix} \partial_{yy}\Psi(t, x, y_\lambda, u) & \partial_{yu}\Psi(t, x, y_\lambda, u) \\ \partial_{uy}\Psi(t, x, y_\lambda, u) & \partial_{uu}\Psi(t, x, y_\lambda, u) \end{bmatrix} \begin{bmatrix} \tilde{y}_2^\lambda \\ v_2 \end{bmatrix} dxdt \\
&- \iint\limits_Q p_\lambda \partial_{yy} d(\cdot, y_\lambda)[\tilde{y}_1^\lambda, \tilde{y}_2^\lambda]\,dxdt,
\end{aligned}$$

with $y_\lambda := G(u_\lambda), \tilde{y}_i^\lambda := G'(u_\lambda)v_i$, and p_λ fulfills (4.3.1) with y_λ, u as well as $\mu_a^\lambda, \mu_b^\lambda$ substituted in the right-hand side. The following is a consequence of Lemma 4.2.4 and the regularized counterpart to Corollary 4.2.10.

Corollary 4.3.8. *For all fixed $\mu_a^\lambda, \mu_b^\lambda \in L^\infty(Q)$, arbitrary $u_1, u_2 \in U_{ad}$ and $v \in L^\infty(Q)$, there exists a constant $C > 0$ such that*

$$|\mathcal{L}''_\lambda(u_1, \mu_a^\lambda, \mu_b^\lambda)[v, v] - \mathcal{L}''_\lambda(u_2, \mu_a^\lambda, \mu_b^\lambda)[v, v]| \le c\|u_1 - u_2\|_{\infty,\infty}\|v\|_I^2.$$

Proof. Again, we only need to consider the terms associated with the Lagrange multipliers, due to Lemma 4.2.4. Here, we can make use of the higher regularity of $\mu_a^\lambda, \mu_b^\lambda$. Observe that

$$\begin{aligned}
&\left| \iint\limits_{\bar{Q}} \left(G''(u_1) - G''(u_2)\right)[v, v](\mu_b^\lambda - \mu_a^\lambda)\,dxdt \right| \\
&\qquad \le c\|\left(G''(u_1) - G''(u_2)\right)[v, v]\|_I \|\mu_b^\lambda - \mu_a^\lambda\|_I \le c\|u_1 - u_2\|_I \|v\|_I^2
\end{aligned}$$

by Lemma 4.2.4, since $\mu_a^\lambda, \mu_b^\lambda$ are fixed. □

Theorem 4.3.7 is equivalent to the existence of nonnegative Lagrange multipliers $\bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda \in L^\infty(Q)$ such that

$$\mathcal{L}'_\lambda(\bar{u}_\lambda, \bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda)(u - \bar{u}_\lambda) \ge 0 \quad \forall\, u \in U_{\text{ad}} \tag{4.3.5}$$

and

$$\iint\limits_Q (y_a - \lambda\bar{u}_\lambda - G(\bar{u}_\lambda))\bar{\mu}_a^\lambda\,dxdt = \iint\limits_Q (\lambda\bar{u}_\lambda + G(\bar{u}_\lambda) - y_b)\bar{\mu}_b^\lambda\,dxdt = 0. \tag{4.3.6}$$

With the help of the Lagrangian, we now discuss second order sufficient conditions.

Assumption 4.3.9. *Let $\bar{u}_\lambda \in U_{feas}$ be a control satisfying the first order necessary optimality conditions of Theorem 4.3.7. We assume that there exists a constant $\alpha > 0$, such that*

$$\frac{\partial^2 \mathcal{L}_\lambda}{\partial u^2}(\bar{u}_\lambda, \bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda)[v, v] \ge \alpha\|v\|_I^2$$

is valid for all $v \in L^\infty(Q)$ and all associated Lagrange multipliers $\bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda \in L^\infty(Q)$.

Theorem 4.3.10. *Let the control $\bar{u}_\lambda \in U^\lambda_{feas}$ and the Lagrange multipliers $\bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b$ fulfill the conditions stated in Assumption 4.3.9. Then there exist constants $\beta > 0$ and $\varepsilon > 0$ such that*

$$f(u) \geq f(\bar{u}_\lambda) + \beta\|u - \bar{u}_\lambda\|^2_I$$

is satisfied for all $u \in U^\lambda_{feas}$ with $\|u - \bar{u}_\lambda\|_{\infty,\infty} \leq \varepsilon$.

Proof. The proof follows the same principal steps as the one for Theorem 4.2.11, but we can make use of the higher regularity of the Lagrange multipliers. For a feasible control $u \in U^\lambda_{\text{feas}}$, we proceed by Taylor expansion of $\mathcal{L}_\lambda$ in $\bar{u}_\lambda$ and obtain

$$\mathcal{L}_\lambda(u, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b) = \mathcal{L}_\lambda(\bar{u}_\lambda, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b) + \mathcal{L}'_\lambda(\bar{u}_\lambda, \bar{\mu}^\lambda_a)(u - \bar{u}_\lambda) + \frac{1}{2}\mathcal{L}''_\lambda(u_\xi, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b)[u - \bar{u}_\lambda, u - \bar{u}_\lambda]$$

with a control $u_\xi = \bar{u}_\lambda + \xi(u - \bar{u}_\lambda)$ with $\xi \in (0,1)$. By the optimality of $\bar{u}_\lambda$, we obtain

$$\begin{aligned}
\mathcal{L}_\lambda(u, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b) \;&\geq\; \mathcal{L}_\lambda(\bar{u}_\lambda, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b) + \frac{1}{2}\mathcal{L}''_\lambda(u_\xi, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b)[u - \bar{u}_\lambda, u - \bar{u}_\lambda] \\
&=\; \mathcal{L}_\lambda(\bar{u}_\lambda) + \frac{1}{2}\mathcal{L}''_\lambda(\bar{u}_\lambda, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b)[u - \bar{u}_\lambda, u - \bar{u}_\lambda] \\
&\quad + \frac{1}{2}\left(\mathcal{L}''_\lambda(u_\xi, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b) - \mathcal{L}''_\lambda(\bar{u}_\lambda, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b)\right)[u - \bar{u}_\lambda, u - \bar{u}_\lambda].
\end{aligned}$$

In addition, the complementary slackness conditions (4.3.3) and the nonnegativity of $\bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b$ combined with the feasibility of u, imply

$$f(\bar{u}_\lambda) = \mathcal{L}_\lambda(\bar{u}_\lambda, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b), \quad f(u) \geq \mathcal{L}_\lambda(u, \bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b). \tag{4.3.7}$$

Together with Assumption 4.3.9 and Corollary 4.3.8, (4.3.7) implies

$$f(u) \geq f(\bar{u}_\lambda) + \frac{\alpha}{2}\|u - \bar{u}_\lambda\|^2 - \frac{c}{2}\|\bar{u}_\lambda - u_\xi\|_{\infty,\infty}\|u - \bar{u}_\lambda\|^2_I.$$

Clearly, by $\|\bar{u}_\lambda - u_\xi\|_{\infty,\infty} = \xi\|u - \bar{u}_\lambda\|_{\infty,\infty}$, the assertion is obtained for $\|u - \bar{u}_\lambda\|_{\infty,\infty} \leq \varepsilon$ with ε sufficiently small. □

Let us mention again that the Lipschitz constant c used for estimation of the second derivative of the Lagrange functional depends on $\bar{\mu}^\lambda_a$ and $\bar{\mu}^\lambda_b$, more precisely: on their $L^2(Q)$ norm, and is expected to become large when λ tends to zero. To elaborate more on the structure of the Lagrange multipliers, we point out that so far we have neither proven nor assumed their uniqueness. So far, we have taken into account possible nonuniqueness simply by making assumptions for arbitrary multipliers. Now, we discuss conditions that guarantee uniqueness of the multipliers as well as the adjoint state.

Assumption 4.3.11. *Let $\bar{u}_\lambda$ be a locally optimal control of Problem $(\mathbb{P}_\lambda)$ in the sense of Definition 4.3.3. We assume that there exists $\sigma > 0$ such that the σ-active sets according to Definition 4.3.4 are pairwise disjoint.*

Comparing this to the separation conditions formulated in Assumption 4.3.5 we now additionally exclude situations where e.g. both the lower control bound and the lower mixed control-state constraint are (almost) active simultaneously.

Lemma 4.3.12. *Let $\bar{u}_\lambda$ be a locally optimal control of Problem $(\mathbb{P}_\lambda)$ fulfilling the first order optimality conditions from Theorem 4.3.7. Suppose that Assumption 4.3.11 holds. Then the Lagrange multipliers $\bar{\mu}^\lambda_a, \bar{\mu}^\lambda_b$ and the adjoint state $\bar{p}_\lambda$ are uniquely determined.*

Proof. We prove the uniqueness result following [7] for linear-quadratic elliptic problems. From the complementary slackness conditions (4.3.3) and the nonnegativity of $\bar{\mu}_a^\lambda$ and $\bar{\mu}_b^\lambda$, we know that

$$\bar{\mu}_a^\lambda = 0 \text{ on } Q \setminus M_{y,a}^{\sigma,\lambda}(\bar{u}_\lambda)$$

as well as

$$\bar{\mu}_b^\lambda = 0 \text{ on } Q \setminus M_{y,b}^{\sigma,\lambda}(\bar{u}_\lambda).$$

Due to our separation condition from Assumption 4.3.11, the control constraints cannot be active on $M_{y,a}^{\sigma,\lambda}(\bar{u}_\lambda) \cup M_{y,b}^{\sigma,\lambda}(\bar{u}_\lambda)$, so that the variational inequality (4.3.2) implies

$$\partial_u \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{p}_\lambda + \lambda(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda) = 0 \quad \text{on } M_{y,a}^{\sigma,\lambda}(\bar{u}_\lambda) \cup M_{y,b}^{\sigma,\lambda}(\bar{u}_\lambda).$$

This pointwise interpretation of the variational inequality leads to

$$\bar{\mu}_a^\lambda = \begin{cases} \frac{1}{\lambda}(\partial_u \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{p}_\lambda) & \text{on } M_{y,a}^{\sigma,\lambda}(\bar{u}_\lambda) \\ 0 & \text{else} \end{cases} \tag{4.3.8a}$$

$$\bar{\mu}_b^\lambda = \begin{cases} -\frac{1}{\lambda}(\partial_u \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{p}_\lambda) & \text{on} M_{y,b}^{\sigma,\lambda}(\bar{u}_\lambda) \\ 0 & \text{else} \end{cases} \tag{4.3.8b}$$

since $M_{y,a}^{\sigma,\lambda} \cap M_{y,b}^{\sigma,\lambda} = \emptyset$. Inserting these expressions into the adjoint equation (4.3.1), we obtain

$$\begin{array}{rcll} \partial_t \bar{p}_\lambda + \mathcal{A}\bar{p}_\lambda + (\partial_y d(\cdot \bar{y}_\lambda) + \bar{c}_a + \bar{c}_b)\bar{p}_\lambda &=& \partial_y \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{m}_b - \bar{m}_a & \text{in } Q \\ \bar{p}_\lambda(T, \cdot) &=& 0 & \text{in } \Omega \\ \bar{p}_\lambda &=& 0 & \text{on } \Sigma, \end{array} \tag{4.3.9}$$

where $\bar{c}_a(t,x)$, $\bar{c}_b(t,x)$ are given as

$$\bar{c}_a = \begin{cases} \frac{1}{\lambda} & \text{on } M_{y,a}^{\sigma,\lambda}(\bar{u}_\lambda) \\ 0 & \text{else} \end{cases}, \quad \bar{c}_b = \begin{cases} -\frac{1}{\lambda} & \text{on } M_{y,b}^{\sigma,\lambda}(\bar{u}_\lambda) \\ 0 & \text{else} \end{cases},$$

and $\bar{m}_a, \bar{m}_b$ are defined by

$$\begin{aligned} \bar{m}_a &= \begin{cases} \frac{1}{\lambda}\partial_u \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) & \text{on } M_{y,a}^{\sigma,\lambda}(\bar{u}_\lambda) \\ 0 & \text{else} \end{cases} \\ \bar{m}_b &= \begin{cases} -\frac{1}{\lambda}\partial_u \Psi(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) & \text{on } M_{y,b}^{\sigma,\lambda}(\bar{u}_\lambda) \\ 0 & \text{else} \end{cases}. \end{aligned}$$

Theorem 2.5.6 yields the existence of a unique solution $\bar{p}_\lambda$ to (4.3.9). Then, (4.3.8) yields unique Lagrange multipliers $\bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda \in L^\infty(Q)$. □

4.4. Stability analysis for an associated linear quadratic problem

4.4.1. Model problem and motivation

The results so far obtained in this chapter guarantee that the optimality system associated with the regularized optimal control problem $(\mathbb{P}_\lambda)$ is well behaved in the sense that first order necessary as well as second order sufficient conditions can be discussed and formulated. Yet, imagine a situation where a local optimum is an accumulation point of a sequence of local solutions. Then, in each neighborhood of the optimal solution there exist infinitely many other stationary points. Such a situation has to be excluded in order for numerical algorithms to work well. Also, it is interesting to discuss whether or not a solution algorithm will converge, how perturbations in the data affect the optimal solution, and how sensitive the solution is with respect to numerical errors when e.g. solving the first order optimality system.

Such questions can be answered with the help of the so called strong regularity property of the optimality system, or rather of the generalized equation that equivalently expresses the first order optimality system. Proving this property involves showing a Lipschitz stability result for a second-order Taylor approximation of the problem.

Discussing strong regularity is meanwhile a standard approach in optimal control. The convergence of SQP methods for control-constrained optimal control problems with semilinear state equation was discussed in [148] for elliptic problems and [54, 140] for parabolic problems. We also mention the results in [79], where the Newton method for generalized equations in finite-dimensional spaces is considered, as well as to generalizations in [5] and [42]. Moreover, the Lipschitz stability results from [93] for problems governed by ordinary differential equations, or [95] for elliptic problems, or [94], [141] for parabolic problems should be mentioned. The first one to discuss Lipschitz stability for purely state-constrained problems was Griesse in [59]. Regarding the convergence of numerical solution algorithms for nonlinear state-constrained problems we refer also to recent results in [61, 60] for a regularized elliptic problem.

In this section we will therefore analyze a linear quadratic model problem of the form

$$\text{Minimize } J_\delta(u,y) := \iint\limits_Q \left[\frac{1}{2}(\psi_1 y^2 + 2\psi_2 yu + \psi_3 u^2) + \psi_4 y + (\psi_5 - \delta_1)u\right] dxdt \tag{$\mathbb{P}_\delta$}$$

subject to

$$\begin{aligned} \partial_t y + \mathcal{A}y + d_0 y &= u \quad \text{in } Q \\ y(0,\cdot) &= 0 \quad \text{in } \Omega \\ y &= 0 \quad \text{on } \Sigma, \end{aligned} \tag{4.4.1a}$$

as well as

$$y_a + \delta_2 \le \lambda u + y \le y_b - \delta_3, \tag{4.4.1b}$$

and

$$u_a \le u \le u_b \tag{4.4.1c}$$

for functions $\psi_1, \dots, \psi_5 \in L^\infty(Q)$ and perturbations $\delta_1, \dots, \delta_3 \in L^\infty(Q)$, where we assume

$$\iint\limits_Q \left(\psi_1 y^2 + 2\psi_2 yu + \psi_3 u^2\right) dxdt \ge \nu \|u\|_I^2 \tag{4.4.2}$$

for some $\nu > 0$ as well as $\psi_3 \ge \beta_0 > 0$. Note that the control constraints remain unperturbed since they are handled by an admissible set. We will derive Lipschitz stability for (4.4.1), that is, the optimal control and state, as well as associated adjoint state and Lagrange multipliers, with respect to the perturbations δ_i, $i = 1, \dots, 3$, by thorough consideration of an auxiliary problem.

Before doing this, let Section 4.4.2 serve as a further motivation.

4.4.2. Generalized equations and strong regularity

Assumption 4.4.1. *Throughout this section, let $\lambda > 0$ still be fixed, and consider a locally optimal solution $\bar{u}_\lambda$ of Problem $(\mathbb{P}_\lambda)$ satisfying the first order necessary conditions of Theorem 4.3.7 with associated Lagrange multipliers $\bar{\mu}_a^\lambda$ and $\bar{\mu}_b^\lambda$, as well as the second order sufficient condition from Assumption 4.3.9. Moreover, suppose that the separation condition stated in Assumption 4.3.11 holds, and that the Slater point property from Assumption 4.2.7 is satisfied.*

For a control $u \in U$ and Lagrange multipliers $\mu_a, \mu_b \in L^\infty(Q)$, we introduce the cones

$$\begin{aligned}
\mathcal{N}_{U_{\mathrm{ad}}}(u) &:= \{g \in L^\infty(Q) : (g, v-u) \le 0 \;\forall v \in U_{\mathrm{ad}}\} \\
\mathcal{K}(\mu_a) &:= \begin{cases} \{g \in L^2(Q) : \; (g, \tilde{\mu}_a - \mu_a) \le 0 \;\forall \tilde{\mu}_a \in L^2(Q),\, \tilde{\mu}_a \ge 0\} & \text{if } \mu_a \ge 0 \\ \emptyset & \text{else.} \end{cases} \\
\mathcal{K}(\mu_b) &:= \begin{cases} \{g \in L^2(Q) : \; (g, \tilde{\mu}_b - \mu_b) \le 0 \;\forall \tilde{\mu}_b \in L^2(Q),\, \tilde{\mu}_b \ge 0\} & \text{if } \mu_b \ge 0 \\ \emptyset & \text{else,} \end{cases}
\end{aligned}$$

and write the first order optimality conditions for $(\mathbb{P}_\lambda)$ as the nonlinear generalized equation

$$0 \in \begin{pmatrix} f'(\bar{u}_\lambda) + \lambda(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda) + G'(\bar{u}_\lambda)^*(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda) \\ \lambda \bar{u}_\lambda + G(\bar{u}_\lambda) - y_a \\ y_b - \lambda \bar{u}_\lambda - G(\bar{u}_\lambda) \end{pmatrix} + \begin{pmatrix} \mathcal{N}_{U_{\mathrm{ad}}}(\bar{u}_\lambda) \\ \mathcal{K}(\bar{\mu}_a^\lambda) \\ \mathcal{K}(\bar{\mu}_b^\lambda) \end{pmatrix}. \tag{4.4.3}$$

Following Robinson, [127], we specify the term strong regularity.

Definition 4.4.2. *Let $\delta = (\delta_1, \delta_2, \delta_3) \in (L^\infty(Q))^3$ be a perturbation. We say that (4.4.3) is strongly regular at $(\bar{u}_\lambda, \bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda)$ with associated Lipschitz constant $L_\lambda > 0$, if there exist real numbers $\rho_\delta > 0$ and $\rho > 0$ such that for every $\delta \in (L^\infty(Q))^3$ with $\|\delta\|_{(L^\infty(Q))^3} \le \rho_\delta$, the perturbed system*

$$\delta \in \begin{pmatrix} f'(\bar{u}_\lambda) + f''(\bar{u}_\lambda)(\bar{u}_\delta - \bar{u}_\lambda) + G''(\bar{u}_\lambda)(\bar{u}_\delta - \bar{u}_\lambda)^*(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda) \\ +\lambda(\bar{\mu}_b^\delta - \bar{\mu}_a^\delta) + G'(\bar{u}_\lambda)^*(\bar{\mu}_b^\delta - \bar{\mu}_a^\delta) + \mathcal{N}_{U_{ad}}(\bar{u}_\delta) \\ \\ \lambda \bar{u}_\delta + G'(\bar{u}_\lambda)(\bar{u}_\delta - \bar{u}_\lambda) + G(\bar{u}_\lambda) - y_a + \mathcal{K}(\bar{\mu}_a^\delta) \\ \\ y_b - \lambda \bar{u}_\delta - G'(\bar{u}_\lambda)(\bar{u}_\delta - \bar{u}_\lambda) - G(\bar{u}_\lambda) + \mathcal{K}(\bar{\mu}_b^\delta) \end{pmatrix} \tag{4.4.4}$$

admits exactly one solution $(\bar{u}_\delta, \bar{\mu}_a^\delta, \bar{\mu}_b^\delta)$ with

$$\|\bar{u}_\delta - \bar{u}_\lambda\|_{\infty,\infty} + \|\bar{\mu}_a^\delta - \bar{\mu}_a^\lambda\|_{\infty,\infty} + \|\bar{\mu}_b^\delta - \bar{\mu}_b^\lambda\|_{\infty,\infty} \le \rho,$$

that depends Lipschitz continuously on the perturbation $\delta \in (L^\infty(Q))^3$. More precisely, for perturbations $\delta, \delta' \in (L^\infty(Q))^3$ sufficiently small, we obtain

$$\|\bar{u}_\delta - \bar{u}_{\delta'}\|_{\infty,\infty} + \|\bar{\mu}_a^\delta - \bar{\mu}_a^{\delta'}\|_{\infty,\infty} + \|\bar{\mu}_b^\delta - \bar{\mu}_b^{\delta'}\|_{\infty,\infty} \le L_\lambda \|\delta - \delta'\|_{(L^\infty(Q))^3}. \tag{4.4.5}$$

The generalized equation (4.4.4) is developed by linearization of (4.4.3) at $(\bar{u}_\lambda, \bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda)$ in the direction $(u_\delta - \bar{u}_\lambda, \mu_a^\delta - \bar{\mu}_a^\lambda, \mu_b^\delta - \bar{\mu}_b^\lambda)$. We recognize (4.4.4) as the first order necessary optimality conditions for the linear-quadratic problem

$$\text{Minimize } f_\delta(u) := \frac{1}{2}\mathcal{L}_\lambda''(\bar{u}_\lambda, \bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda)[u - \bar{u}_\lambda, u - \bar{u}_\lambda] + f'(\bar{u}_\lambda)(u - \bar{u}_\lambda) - (\delta_1, u - \bar{u}_\lambda) \tag{4.4.6a}$$

subject to

$$y_a + \delta_2 - \lambda \bar{u}_\lambda - G(\bar{u}_\lambda) \leq \lambda(u - \bar{u}_\lambda) + G'(\bar{u}_\lambda)(u - \bar{u}_\lambda) \leq y_b - \delta_3 - \lambda \bar{u}_\lambda - G(\bar{u}_\lambda), \tag{4.4.6b}$$

as well as

$$u_a \leq u \leq u_b \tag{4.4.6c}$$

with optimal control $\bar{u}_\delta$ and associated Lagrange multipliers $\bar{\mu}_a^\delta, \bar{\mu}_b^\delta$, if these exist in regular spaces. To obtain strong regularity of the generalized equation (4.4.3) we will therefore have to

1. prove existence of a unique optimal control $\bar{u}_\delta$ of Problem (4.4.6)

2. show the existence of an associated Slater point and obtain the existence of (irregular) Lagrange multipliers

3. prove a separation condition in the spirit of Assumption 3.4.2 to obtain higher regularity and uniqueness of the Lagrange multipliers

4. obtain the desired Lipschitz property (4.4.5)

for sufficiently small $\delta \in L^\infty(Q)^3$. Steps 1) and 2) are easily shown:

Lemma 4.4.3. *Let Assumption 4.4.1 hold. Then the linear-quadratic optimal control problem* (4.4.6) *admits a unique optimal control $\bar{u}_\delta$. Moroever, for δ_2, δ_3 small enough, the Slater point u_γ^λ from Assumption 4.2.7 is also a Slater point for Problem* (4.4.6)

Proof. First observe that the feasible set for Problem (4.4.6) is not empty, since $\bar{u}_\lambda$ satisfies all constraints. The existence of a unique solution to this problem can then be deduced from the second order sufficient condition for $\bar{u}_\lambda$ stated in Assumption (4.3.9). More precisely, Assumption (4.3.9) guarantees that the objective function f_δ is strictly convex for every $\delta \in (L^\infty(Q))^3$ and tends to infinity as $\|u\|_I \to \infty$, which yields the existence of a unique solution by standard arguments. Moreover, the upper state constraint can be rewritten as

$$\lambda u + G(\bar{u}_\lambda) + G'(\bar{u}_\lambda)(u - \bar{u}_\lambda) - y_b + \delta_3 \leq 0.$$

To check the Slater point property, note that the derivatives have to be taken with respect to u, and we obtain

$$\begin{aligned} & \lambda \bar{u}_\delta + G(\bar{u}_\lambda) - y_b + \delta_3 + G'(\bar{u}_\lambda)(\bar{u}_\delta - \bar{u}_\lambda) + \lambda(u_\gamma^\lambda - \bar{u}_\delta) + G'(\bar{u}_\lambda)(u_\gamma^\lambda - \bar{u}_\delta) \\ = \; & \lambda u_\gamma^\lambda + G(\bar{u}_\lambda) + G'(\bar{u}_\lambda)(u_\gamma^\lambda - \bar{u}_\lambda) - y_b + \delta_3 \\ \leq \; & \gamma_\lambda / 2 \end{aligned}$$

for δ_3 sufficiently small. The lower state constraint and the control constraints are treated analogously. Consequently, there exist Lagrange multipliers $\bar{\mu}_a^\delta, \bar{\mu}_b^\delta \in L^\infty(Q)^*$, cf. the discussion in [134]. $\square$

Steps 3) and 4) will be the heart of the proof of strong regularity. For $\delta_1 = \delta_2 = \delta_3 = 0$, the separation condition and therefore uniqueness and higher regularity of the multipliers is easily obtained noting that $\bar{u}_\lambda$ with associated Lagrange multipliers solves the linearized unperturbed problem, since the first inclusion in (4.4.4) simplifies to

$$0 \in f'(\bar{u}_\lambda) + \lambda(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda) + G'(\bar{u}_\lambda)^*(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda) + \mathcal{N}_{U_{\mathrm{ad}}}(\bar{u}_\lambda).$$

Moreover, inserting $\bar{u}_\lambda$ in the linearized state constraints with $\delta_2 = \delta_3 = 0$ yields precisely the original mixed control-state constraints

$$y_a \leq \lambda \bar{u}_\lambda + G(\bar{u}_\lambda) \leq y_b,$$

Consequently, the separation condition from Assumption 4.3.11 is valid and higher regularity as well as uniqueness of the Lagrange multipliers is obtained for $\delta = 0$.

For small $\delta_i > 0$, $i = 1, \ldots, 3$, a separation condition and therefore higher regularity and uniqueness of the Lagrange multiplier can be proven with the help of an auxiliary problem, that is also used to prove the desired Lipschitz properties.

We will do this for the more general problem $(\mathbb{P}_\delta)$ in the following Sections 4.4.3-4.4.5. Before, let us show that Problem (4.4.6) fits into this more general setting. Defining the functions $\psi_i \in L^\infty(Q), i = 1, \ldots, 5$,

$$\begin{aligned}
\psi_1 &:= \Psi_{yy}(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) - \bar{p}_\lambda d_{yy}(\cdot, \bar{y}_\lambda)\\
\psi_2 &:= \Psi_{yu}(\cdot, \bar{y}_\lambda, \bar{u}_\lambda)\\
\psi_3 &:= \Psi_{uu}(\cdot, \bar{y}_\lambda, \bar{u}_\lambda)\\
\psi_4 &:= \Psi_{y}(\cdot, \bar{y}_\lambda, \bar{u}_\lambda)\\
\psi_5 &:= \Psi_{u}(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) + \bar{p}_\lambda + \lambda(\bar{\mu}_b^\lambda - \bar{\mu}_a^\lambda)\\
d_0 &:= \partial_y d(\cdot, \bar{y}_\lambda),
\end{aligned}$$

where $\bar{p}_\lambda$ solves the adjoint equation from Theorem 4.3.7, and the transformed constraints

$$\tilde{y}_a := y_a - \bar{y}_\lambda - \lambda \bar{u}_\lambda, \quad \tilde{y}_b := y_b - \bar{y}_\lambda - \lambda \bar{u}_\lambda, \quad \tilde{u}_a := u_a - \bar{u}_\lambda, \quad \tilde{u}_b := u_b - \bar{u}_\lambda,$$

and

$$\tilde{U}_{\mathrm{ad}} = \{u \in L^2(Q) \mid \tilde{u}_a \leq u \leq \tilde{u}_b\},$$

it is clear that (4.4.6) fits into the setting of Problem $(\mathbb{P}_\delta)$. Due to the second order sufficient condition stated in Assumption (4.3.9), we find that (4.4.2) is fulfilled $\psi_3 = \Psi_{uu}(\cdot, \bar{y}_\lambda, \bar{u}_\lambda) \geq \beta_0 > 0$ on Q due to our general assumptions.

4.4.3. Analysis of the linear-quadratic model problem $(\mathbb{P}_\delta)$

After this short motivation, let us return to the analysis of Problem $(\mathbb{P}_\delta)$. We define a linear control-to-state mapping

$$\mathcal{G}\colon L^2(Q) \to \mathcal{W}(0,T), \quad u \mapsto y,$$

cf. Proposition 2.5.2 on page 24.

For $(\mathbb{P}_\delta)$, let us just point out that we obtain the existence of a unique solution, which we will for simplicity call $\bar{u}_\delta$ with associated state $\bar{y}_\delta$ as in Lemma 4.4.3. Then, for every $\delta = (\delta_1, \delta_2, \delta_3) \in (L^\infty(Q))^3$ the objective function is strictly convex with respect to u and the objective value tends to ∞ as $\|u\|_I$ tends to infinity. We define the linear-quadratic analogue to Definition 4.3.4:

Definition 4.4.4. *Let τ be a positive real number. For a control u that is feasible for Problem $(\mathbb{P}_\delta)$,* *we introduce the τ-active sets:*

$$\begin{array}{lcl} M_{u,a}^{\tau,\delta}(u) & := & \{(t,x) \in Q : \ u(t,x) \leq u_a(t,x) + \tau\} \\ M_{u,b}^{\tau,\delta}(u) & := & \{(t,x) \in Q : \ u(t,x) \geq u_b(t,x) - \tau\} \\ M_{y,a}^{\tau,\delta}(u) & := & \{(t,x) \in Q : \ \lambda u(t,x) + \mathcal{G}u(t,x) \leq y_a(t,x) + \delta_2 + \tau\} \\ M_{y,b}^{\tau,\delta}(u) & := & \{(t,x) \in Q : \ \lambda u(t,x) + \mathcal{G}u(t,x) \geq y_b(t,x) - \delta_3 - \tau\}. \end{array}$$

For $\delta \equiv 0$, in particular, the optimal solution of the unperturbed Problem $(\mathbb{P}_0)$, is referred to as $\bar{u}_0$ with associated optimal state $\bar{y}_0$, and we define and assume the following:

$$\begin{array}{lcl} M_1 & := & \{(t,x) \colon \bar{u}_0(t,x) \leq u_a(t,x) + \tau\} \\ M_2 & := & \{(t,x) \colon \bar{u}_0(t,x) \geq u_b(t,x) - \tau\} \\ M_3 & := & \{(t,x) \colon \lambda\bar{u}_0(t,x) + \bar{y}_0(t,x) \geq y_a(t,x) + \tau\} \\ M_4 & := & \{(t,x) \colon \lambda\bar{u}_0(t,x) + \bar{y}_0(t,x) \leq y_b(t,x) - \tau\} \end{array}$$

Moreover, set

$$M_5 := Q \setminus \{M_1 \cup M_2 \cup M_3 \cup M_4\}.$$

Assumption 4.4.5. *Suppose that there exists a real number $\tau > 0$ such that the sets $M_1, \ldots, M_5$ are pairwise disjoint, and that there exists a Slater point $u_\gamma^0 \in L^\infty(Q)$ with $u_a \leq u_\gamma^0 \leq u_b$ and a constant $\gamma_0 > 0$ such that*

$$y_a + \gamma_0 \leq \lambda u_\gamma^0 + \mathcal{G}u_\gamma^0 \leq y_b - \gamma_0.$$

In terms of the motivating example of the last section, Assumption 4.4.5 pertains to the properties of the linearized unperturbed problem. Under the last assumption, a linear-quadratic analogue to Theorem 4.3.7 yields the existence of regular Lagrange multipliers and first-order optimality conditions for the optimal control $\bar{u}_0$ for $(\mathbb{P}_0)$ that can be formulated as Karush-Kuhn-Tucker system.

Relying on Assumption 4.4.5, the Slater point property is also easily verified for small perturbations $\delta \in L^\infty(Q)^3 \neq 0$, cf. Lemma 4.4.3. We will also prove that then there exists $\tau > 0$ such that the sets from Definition 4.4.4 are pairwise disjoint. Then, the linear-quadratic analogue to Theorem 4.3.7 guarantees the existence of unique nonnegative Lagrange multipliers $\bar{\mu}_a^\delta \in L^2(Q)$, $\bar{\mu}_b^\delta \in L^2(Q)$ as well as an adjoint state $\bar{p}_\delta \in \mathcal{W}(0,T)$ solving

$$\begin{array}{rcl} -\partial_t \bar{p}_\delta + \mathcal{A}\bar{p}_\delta + d_0\bar{p}_\delta & = & \psi_1\bar{y}_\delta + \psi_2\bar{u}_\delta + \psi_4 + \bar{\mu}_b^\delta - \bar{\mu}_a^\delta \\ \bar{p}_\delta(T,\cdot) & = & 0 \\ \bar{p}_\delta & = & 0 \end{array} \tag{4.4.7}$$

such that

$$\iint\limits_Q \left(-\delta_1 + \psi_3\bar{u}_\delta + \psi_5 + \psi_2\bar{y}_\delta + \bar{p}_\delta + \lambda(\bar{\mu}_b^\delta - \bar{\mu}_a^\delta\right)(u - \bar{u}_\delta)\, dxdt \geq 0 \quad \forall u \in U_{\text{ad}}, \tag{4.4.8}$$

and the complementarity conditions

$$(\bar{\mu}_a^\delta, \lambda\bar{u}_\delta + \bar{y}_\delta - y_a - \delta_2)_I = 0, \quad (\bar{\mu}_b^\delta, \lambda\bar{u}_\delta + \bar{y}_\delta - y_b + \delta_3)_I = 0. \tag{4.4.9}$$

These conditions build necessary and sufficient optimality conditions for Problem $(\mathbb{P}_\delta)$.

A Lipschitz stability result for the control, state, adjoint state, as well as the multipliers will be deduced from associated results for an auxiliary problem, that will be investigated next. We follow in principle the same steps as in [7], where an elliptic optimal control problem is considered. We point out that the problem considered therein admits only lower bounds on the control and lower mixed-control state constraints, such that the Lagrange multipliers associated with this problem are automatically regular, cf. also Theorem 4.3.7. The separation condition on the active sets is only needed for the uniqueness of the multipliers such that Lipschitz stability results can be proven. In contrast, we need the separation condition also for the desired regularity of the multipliers so that we can apply pointwise arguments.

4.4.4. Analysis of an auxiliary problem

As in [93], an auxiliary problem is introduced with the constraints restricted to the almost active sets $M_1, \ldots, M_4$ of the optimal control $\bar{u}_0$ of the unperturbed problem. Throughout, we have to take into account the lower regularity properties of the parabolic solution operator, which requires special attention when e.g. proving that the optimal controls of the auxiliary problem are bounded. We define

$$U_{\text{ad}}^{\text{aux}} = \{u \in L^2(Q) \mid u_a \leq u \text{ a.e. in } M_1,\ u \leq u_b \text{ a.e. in } M_2\},$$

and consider an auxiliary problem

$$\min_{u \in U_{\text{ad}}^{aux}} J_\delta(u, y) := \iint_Q \left[\frac{1}{2}(\psi_1 y^2 + 2\psi_2 y u + \psi_3 u^2) + \psi_4 y + (\psi_5 - \delta_1) u\right] dxdt \tag{$\mathbb{P}_\delta^{\text{aux}}$}$$

subject to

$$\begin{aligned} \partial_t y + \mathcal{A} y + d_0 y &= u && \text{in } Q \\ y(0, \cdot) &= 0 && \text{in } \Omega \\ y &= 0 && \text{on } \Sigma, \end{aligned}$$

$$y_a + \delta_2 \leq \lambda u + y \quad \text{a.e. in } M_3, \quad \lambda u + y \leq y_b - \delta_3 \quad \text{a.e. in } M_4.$$

The existence of a unique solution $\tilde{u}_\delta$ in $U_{\text{ad}}^{\text{aux}}$ follows just like for $(\mathbb{P}_\delta)$. What needs special attention before even trying to prove Lipschitz stability is the fact that $U_{\text{ad}}^{\text{aux}}$ is not a subset of $L^\infty(Q)$, and a priori $\tilde{u}_\delta$ is an element of $L^2(Q)$, only. A typical technique to obtain higher regularity is a so-called bootstrapping method, involving the representation of the optimal control via the adjoint state. To pursue this here, we first need to verify that regular Lagrange multipliers exist, so that the adjoint state is regular enough. The results from [134] are not directly applicable since there the controls are assumed to be in $L^\infty(Q)$. We will see, however, that the separation condition for the sets $M_1, \ldots, M_4$ allows for the required regularity.

Theorem 4.4.6. *Let $\tilde{u}_\delta$ with associated state $\tilde{y}_\delta$ denote the solution of $(\mathbb{P}_\delta^{\text{aux}})$. There exist unique nonnegative Lagrange multipliers $\tilde{\mu}_a^\delta \in L^2(Q)$, $\tilde{\mu}_b^\delta \in L^2(Q)$ with $\tilde{\mu}_a^\delta = 0$ on $Q \setminus M_3$ and $\tilde{\mu}_b^\delta = 0$ on $Q \setminus M_4$, as well as an adjoint state $\tilde{p}_\delta \in \mathcal{W}(0, T)$ solving*

$$\begin{aligned} -\partial_t \tilde{p}_\delta + \mathcal{A}\tilde{p}_\delta + d_0 \tilde{p}_\delta &= \psi_1 \tilde{y}_\delta + \psi_2 \tilde{u}_\delta + \psi_4 + \tilde{\mu}_b^\delta - \tilde{\mu}_a^\delta && \textit{in } Q \\ \tilde{p}_\delta(T, \cdot) &= 0 && \textit{in } \Omega \\ \tilde{p}_\delta &= 0 && \textit{on } \Sigma \end{aligned} \tag{4.4.10}$$

such that

$$\iint_Q \left(-\delta_1 + \psi_3 \tilde{u}_\delta + \psi_5 + \psi_2 \tilde{y}_\delta + \tilde{p}_\delta + \lambda(\tilde{\mu}_b^\delta - \tilde{\mu}_a^\delta)\right)(u - \tilde{u}_\delta)\, dxdt \geq 0 \quad \forall u \in U_{ad}^{aux}, \tag{4.4.11}$$

and the complementarity conditions

$$(\tilde{\mu}_a^\delta, \lambda\tilde{u}_\delta + \tilde{y}_\delta - y_a - \delta_2)_I = 0, \quad (\tilde{\mu}_b^\delta, \lambda\tilde{u}_\delta + \tilde{y}_\delta - y_b + \delta_3)_I = 0 \tag{4.4.12}$$

are satisfied. The Lagrange multipliers associated with the solution of $(\mathbb{P}_\delta^{\mathrm{aux}})$ *admit the form*

$$\tilde{\mu}_a^\delta = \max\{0, \frac{\psi_3}{\lambda^2}(y_a + \delta_2 - \tilde{y}_\delta) + \frac{1}{\psi_3}(-\delta_1 + \psi_5 + \psi_2\tilde{y}_\delta + \tilde{p}_\delta)\} \quad \text{on } M_3 \tag{4.4.13}$$

$$\tilde{\mu}_b^\delta = \max\{0, \frac{\psi_3}{\lambda^2}(\tilde{y}_\delta + \delta_3 - y_b) - \frac{1}{\psi_3}(-\delta_1 + \psi_5 + \psi_2\tilde{y}_\delta + \tilde{p}_\delta)\} \quad \text{on } M_4. \tag{4.4.14}$$

Proof. Let us first express the constraints of $(\mathbb{P}_\delta^{\mathrm{aux}})$

$$\begin{array}{rclrcl} -u & \leq & -u_a \ \text{ on } M_1, & -\lambda u - \mathcal{G}u & \leq & -y_a - \delta_2 \ \text{ on } M_3, \\ u & \leq & u_b \ \text{ on } M_2, & \lambda u + \mathcal{G}u & \leq & y_b - \delta_3 \ \text{ on } M_4, \\ & & & u(t,x) \ \in \ \mathbb{R} \ \text{ on } M & & \end{array} \tag{4.4.15}$$

in another form. Define the linear operator

$$\mathbb{G} : L^2(Q) \to L^2(Q), \quad \mathbb{G}u = (\chi_2 - \chi_1 + \chi_5)u + (\chi_4 - \chi_3)(\lambda u + \mathcal{G}u),$$

where the χ_i, $i = 1, \ldots, 5$ are the characteristic functions of the sets M_i. Then (4.4.15) is equivalent to

$$(\mathbb{G}u)(t,x) \begin{cases} \leq b(t,x) & \text{on } Q \setminus M_5 \\ \text{arbitrary} & \text{on } M_5, \end{cases}$$

where

$$b(t,x) = -\chi_1 u_a + \chi_2 u_b - \chi_3(-y_a - \delta_2) + \chi_4(y_b - \delta_3).$$

We now show that this system satisfies the well-known regularity condition by Zowe and Kurcyusz, [153]. To introduce it, we need the convex cone

$$K(\bar{v}) = \{\alpha(v - \bar{v}) \mid \alpha \geq 0, \quad v \geq 0 \text{ on } Q \setminus M_5, \quad v \in L^2(Q)\}.$$

Notice that v is arbitrary on M_5, since no constraints are given there. There is no further constraint imposed on u, hence the Zowe-Kurcyusz-regularity condition is

$$\mathbb{G}(L^2(Q)) + K(-\mathbb{G}\tilde{u}_\delta) = L^2(Q),$$

i.e. each $z \in L^2(Q)$ can be represented in the form

$$z = \mathbb{G}u + \alpha(v + \mathbb{G}\tilde{u}_\delta),$$

with $v \geq 0$ on $Q \setminus M_5$, $u \in L^2(Q)$, $\alpha \geq 0$. It turns out that $\alpha = 0$ can be taken and also $v = 0$, i.e. $\mathbb{G}u = z$. A comparison with (4.4.15) shows that we can take

$$u = \begin{cases} -z & \text{on } M_1 \\ z & \text{on } M_2 \cup M_5. \end{cases} \tag{4.4.16}$$

It remains to find u on $M_3 \cup M_4$. Define $\hat{u}$ as the function that satisfies (4.4.16) on $M_1 \cup M_2 \cup M_5$ and is zero on $M_3 \cup M_4$. Then, we can decompose u into

$$u = u_3 + u_4 + \hat{u},$$

where $u_3 = 0$ on $Q \setminus M_3$ and $u_4 = 0$ on $Q \setminus M_4$. We obtain the equation $\mathbb{G}(u_3 + u_4 + \hat{u}) = z$ on $M_3 \cup M_4$, hence

$$\begin{aligned} \lambda u_3 + \mathcal{G}(u_3 + u_4 + \hat{u}) &= -z \quad \text{on } M_3 \\ \lambda u_4 + \mathcal{G}(u_3 + u_4 + \hat{u}) &= z \quad \text{on } M_4. \end{aligned} \tag{4.4.17}$$

Given $\hat{u}$, $y = \mathcal{G}(u_3 + u_4 + \hat{u})$ is the solution to

$$\begin{aligned} \partial_t y + \mathcal{A}y + d_0 y &= u_3 + u_4 + \hat{u}, \quad \text{in } Q \\ y(0,\cdot) &= 0 \quad \text{in } \Omega \\ y &= 0 \quad \text{on } \Sigma. \end{aligned} \tag{4.4.18}$$

Therefore, (4.4.17) can be written as $u_3 = \frac{1}{\lambda}(-z - y)$, $u_4 = \frac{1}{\lambda}(z - y)$. Inserting this in (4.4.18), y has to solve the equation

$$\begin{aligned} \partial_t y + \mathcal{A}y + d_0 y + \tfrac{1}{\lambda}\chi_{M_3 \cup M_4} y &= \tfrac{1}{\lambda}(\chi_4 - \chi_3)z + \hat{u} \quad \text{in } Q \\ y(0,\cdot) &= 0 \quad \text{in } \Omega \\ y &= 0 \quad \text{on } \Sigma. \end{aligned} \tag{4.4.19}$$

This equation has a unique solution. On the other hand, given the solution of (4.4.19), $u_3 = \frac{1}{\lambda}(-z - y)$ and $u_4 = \frac{1}{\lambda}(z - y)$ satisfy, together with $\hat{u}$, the system (4.4.17). Therefore, the Kurcyusz-Zowe condition is satisfied. From the associated Lagrange multiplier rule, we obtain at least one Lagrange multiplier function $\tilde{\mu}_\delta \in L^2(Q)$ with $\tilde{\mu}_\delta \geq 0$ on $Q \setminus M_5$. Now, the Lagrange multipliers to the associated single constraints are obtained by restriction of $\tilde{\mu}_\delta$ to the appropriate sets. Their uniqueness follows from the fact that the sets $M_1, \ldots, M_4$ are pairwise disjoint. On $Q \setminus M_3 \cup M_4$ we choose to eliminate the Lagrange multiplier associated with the control constraints and handle the control constraints by means of an admissible set. This leads to the variational inequality (4.4.11).

The projection formulas for $\tilde{\mu}_a^\delta$ and $\tilde{\mu}_b^\delta$ can be shown analogously to [145]. We state the main ideas. The proof is based on the fact that the multipliers are represented by

$$\tilde{\mu}_a^\delta = \max\{0, \tilde{\mu}_a^\delta + c(y_a + \delta_2 - \lambda\tilde{u}_\delta - \tilde{y}_\delta)\} \quad \text{on } M_3 \tag{4.4.20}$$

$$\tilde{\mu}_b^\delta = \max\{0, \tilde{\mu}_b^\delta + c(\lambda\tilde{u}_\delta + \tilde{y}_\delta + \delta_3 - y_b)\} \quad \text{on } M_4, \tag{4.4.21}$$

for an arbitrary $c = c(t,x) > 0$, which is known from e.g. [66] and easy to prove. The main idea is to represent $\tilde{u}_\delta$ in terms of the other quantities, especially containing $\tilde{\mu}_a^\delta$ and $\tilde{\mu}_b^\delta$. With an adequate choice of c, the multipliers inside the max-function cancel out. The variational inequality is given as a gradient equation on M_3 and M_4. Hence,

$$-\delta_1 + \psi_3\tilde{u}_\delta + \psi_5 + \psi_2\tilde{y}_\delta + \lambda(\tilde{\mu}_b^\delta - \tilde{\mu}_a^\delta) + \tilde{p}_\delta = 0 \quad \text{on } M_3 \cup M_4. \tag{4.4.22}$$

Note that $\psi_3 \geq \beta_0 > 0$. Then, we obtain

$$\lambda\tilde{u}_\delta = -\frac{\lambda}{\psi_3}(-\delta_1 + \psi_5 + \psi_2\tilde{y}_\delta + \lambda(\tilde{\mu}_b^\delta - \tilde{\mu}_a^\delta) + \tilde{p}_\delta) \text{ on } M_3 \cup M_4.$$

The last equation can be inserted into the maximum representation of the multipliers, since they are nonzero only on their respective active sets. Choosing $c = \frac{\psi_3}{\lambda^2}$ then yields the assertion. □

Remark 4.4.7. *For future reference, we replace the variational inequality* (4.4.11) *by a pointwise projection formula on the admissible set U_{ad}^{aux}, cf. [144]. Keeping in mind that $\tilde{\mu}_a^\delta = 0$ on $Q \setminus M_3$ and*

$\tilde{\mu}_b^\delta = 0$ *on* $Q \setminus M_4$ *we obtain*

$$\tilde{u}_\delta = P_{[u_a,\infty]}\left(-\frac{1}{\Psi_3}(\Psi_2\tilde{y}_\delta + \tilde{p}_\delta + \Psi_5 - \delta_1)\right) \qquad \text{on } M_1 \tag{4.4.23a}$$

$$u_\delta = P_{[-\infty,u_b]}\left(-\frac{1}{\Psi_3}(\Psi_2\tilde{y}_\delta + \tilde{p}_\delta + \Psi_5 - \delta_1)\right) \qquad \text{on } M_2 \tag{4.4.23b}$$

$$u_\delta = -\frac{1}{\Psi_3}(\Psi_2\tilde{y}_\delta + \tilde{p}_\delta - \lambda\tilde{\mu}_a^\delta + \Psi_5 - \delta_1) \qquad \text{on } M_3 \tag{4.4.23c}$$

$$u_\delta = -\frac{1}{\Psi_3}(\Psi_2\tilde{y}_\delta + \tilde{p}_\delta + \lambda\tilde{\mu}_b^\delta + \Psi_5 - \delta_1) \qquad \text{on } M_4 \tag{4.4.23d}$$

$$u_\delta = -\frac{1}{\Psi_3}(\Psi_2\tilde{y}_\delta + \tilde{p}_\delta + \Psi_5 - \delta_1) \qquad \text{on } M_5. \tag{4.4.23e}$$

Remark 4.4.8. *We would like to point out here that projection formulas similar to* (4.4.20) *and* (4.4.21) *can also be shown for the Lagrange multipliers* $\bar{\mu}_a^\lambda, \bar{\mu}_b^\lambda$ *of the nonconvex regularized optimal control problem* $(\mathbb{P}_\lambda)$ *on their* σ*-active sets, cf. the proof of Lemma 4.3.12. However, the sets* M_3 *and* M_4 *remain fixed for all possible perturbations, and therefore give rise to a Lipschitz stability result of the multipliers* $\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta$ *with respect to perturbations. On the contrary, the* σ*-active sets from Lemma 4.3.12, will change with different controls, and a Lipschitz stability result cannot be proven directly.*

Theorem 4.4.9. *The optimal control* $\tilde{u}_\delta$ *of* $(\mathbb{P}_\delta^{\text{aux}})$*, the state* $\tilde{y}_\delta = \mathcal{G}\tilde{u}_\delta$*, and the associated adjoint state* $\tilde{p}_\delta$ *and Lagrange multipliers* $\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta$ *are functions in* $L^\infty(Q)$.

Proof. This follows from a classical bootstrapping argument and a regularity result for the parabolic initial boundary value problem. Initially, we know that

$$\tilde{u}_\delta, \tilde{y}_\delta, \tilde{p}_\delta, \tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta \in L^2(Q),$$

and that $\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta$ are zero, hence bounded, on $Q \setminus \{M_3 \cup M_4\}$. It remains to show boundedness of $\tilde{u}_\delta$ on Q as well as boundedness of $\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta$ on $M_3 \cup M_4$, then the desired regularity of $\tilde{y}_\delta$ and $\tilde{p}_\delta$ is automatically fulfilled.

In [141], it has been shown for a parabolic initial-boundary value problem with Robin-type boundary conditions, that

$$\tilde{u}_\delta \in L^p(Q)$$

yields

$$\tilde{y}_\delta \in L^{p+s}(Q),$$

for some $s > s_0 > 0$. This can analogously be obtained for homogenous Dirichlet boundary conditions, since Proposition 2.5.2 yields the necessary regularity to apply interpolation arguments from e.g. [139]. Initially, $\tilde{u}_\delta \in L^2(Q)$ yields

$$\tilde{y}_\delta \in L^{2+s}(Q).$$

Since $\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta \in L^2(Q)$, and all other expressions appearing in the right-hand side of the adjoint equation (4.4.10) are L^∞-functions by our assumptions, the same result ensures

$$\tilde{p}_\delta \in L^{2+s}(Q).$$

On M_3, M_4, respectively, we have the projection formulas (4.4.13) and (4.4.14) for the Lagrange multipliers, where all appearing functions are at least $L^{2+s}(Q)$ functions. Since the max-function preserves this regularity, we obtain L^{2+s}-regularity of $\tilde{\mu}_a^\delta$ on M_3 and $\tilde{\mu}_b^\delta$ on M_4. Recalling that the multipliers are zero outside M_3 or M_4, respectively, we have

$$\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta \in L^{2+s}(Q).$$

Consequently, from representation (4.4.23), we obtain that

$$\tilde{u}_\delta \in L^{2+s}(Q).$$

Repeating these arguments, we obtain after finitely many steps that

$$\tilde{u}_\delta, \tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta \in L^r(Q), \quad r > n/2 + 1,$$

which yields continuity of the state $\tilde{y}_\delta$, hence also continuity of the adjoint state $\tilde{p}_\delta$ by Theorem 2.5.6. This implies in return boundedness of the Lagrange multipliers $\tilde{\mu}_a^\delta$ and $\tilde{\mu}_b^\delta$ by the projection formulas and finally boundedness of the optimal control $\tilde{u}_\delta$ due to the gradient equation. □

4.4.5. Stability analysis with respect to perturbations

Now, with the optimality conditions of Theorem 4.4.6 as well as the regularity result of Theorem 4.4.9 at hand we can aim at proving Lipschitz stability estimates following [7]. As an intermediate step, let us start with the stability analysis of $(\mathbb{P}_\delta^{\text{aux}})$ in $L^2(Q)$. We choose two perturbation vectors $\delta = (\delta_1, \delta_2, \delta_3) \in (L^\infty(Q))^3$ and $\delta' = (\delta_1', \delta_2', \delta_3') \in (L^\infty(Q))^3$ with associated optimal solutions $\tilde{u}_\delta$ and $\tilde{u}_{\delta'}$, and introduce the following short notation:

$$\delta u = \tilde{u}_\delta - \tilde{u}_{\delta'}, \quad \delta y = \tilde{y}_\delta - \tilde{y}_{\delta'}, \quad \delta p = \tilde{p}_\delta - \tilde{p}_{\delta'}, \quad \delta\mu_a = \tilde{\mu}_a^\delta - \tilde{\mu}_a^{\delta'}, \quad \delta\mu_b = \tilde{\mu}_b^\delta - \tilde{\mu}_b^{\delta'}.$$

Theorem 4.4.10. *Let δ and δ' be two perturbation vectors. Then*

$$\|\tilde{u}_\delta - \tilde{u}_{\delta'}\|_I + \|\tilde{y}_\delta - \tilde{y}_{\delta'}\|_I + \|\tilde{p}_\delta - \tilde{p}_{\delta'}\|_I + \|\tilde{\mu}_a^\delta - \tilde{\mu}_a^{\delta'}\|_I + \|\tilde{\mu}_b^\delta - \tilde{\mu}_b^{\delta'}\|_I \le L_2 \|\delta - \delta'\|_{L^2(Q)^3}$$

holds for the optimal controls $\tilde{u}_\delta$ and $\tilde{u}_{\delta'}$ of $(\mathbb{P}_\delta^{\text{aux}})$ and the associated states, adjoint states, and Lagrange multipliers $\tilde{y}_\delta, \tilde{y}_{\delta'}$, $\tilde{p}_\delta, \tilde{p}_{\delta'}$, $\tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta$ as well as $\tilde{\mu}_a^{\delta'}, \tilde{\mu}_b^{\delta'}$.

Proof. We proceed in four steps.

- Inserting $\tilde{u}_{\delta'}$ into the variational inequality (4.4.11) for $\tilde{u}_\delta$, as well as $\tilde{u}_\delta$ in the variational inequality for $\tilde{u}_{\delta'}$, obtained from (4.4.11) by substituting δ' for δ, yields

$$\iint_Q \left(-\delta_1 + \psi_3 \tilde{u}_\delta + \psi_5 + \psi_2 \tilde{y}_\delta + \tilde{p}_\delta + \lambda(\tilde{\mu}_b^\delta - \tilde{\mu}_a^\delta\right)(\tilde{u}_{\delta'} - \tilde{u}_\delta)\, dxdt \ge 0 \tag{4.4.24}$$

as well as

$$-\iint_Q \left(-\delta_1' + \psi_3 \tilde{u}_{\delta'} + \psi_5 + \psi_2 \tilde{y}_{\delta'} + \tilde{p}_{\delta'} + \lambda(\tilde{\mu}_b^{\delta'} - \tilde{\mu}_a^{\delta'}\right)(\tilde{u}_{\delta'} - \tilde{u}_\delta)\, dxdt \ge 0. \tag{4.4.25}$$

Adding both inequalities delivers

$$(\psi_3 \delta u, \delta u)_I + (\psi_2 \delta y, \delta u)_I + (\delta p, \delta u)_I \le (\delta_1 - \delta_1', \delta u)_I - \lambda(\delta\bar{\mu}_b - \delta\bar{\mu}_a, \delta u)_I. \tag{4.4.26}$$

From the definition of the adjoint operator $\mathcal{G}^*$, we obtain

$$(\delta p, \delta u)_I = (\psi_1 \delta y + \psi_2 \delta u + \delta\mu_b - \delta\mu_a, \delta y)_I,$$

which inserted in (4.4.26) leads to

$$(\psi_3 \delta u, \delta u)_I + (\psi_1 \delta y, \delta y)_I + 2(\psi_2 \delta y, \delta u)_I \leq (\delta_1 - \delta_1', \delta u)_I - (\lambda \delta u + \delta y, \delta\mu_b - \delta\mu_a)_I.$$

With (4.4.2), the last inequality yields

$$\nu \|\delta u\|_I^2 \leq \|\delta_1 - \delta_1'\|_I \|\delta u\|_I - (\lambda \delta u + \delta y, \delta\mu_b - \delta\mu_a)_I. \tag{4.4.27}$$

- To eliminate the Lagrange multipliers, we follow an idea by Griesse, cf. [59]. From the complementary slackness conditions (4.4.12), the nonnegativity of μ_a^δ and $\mu_a^{\delta'}$, and the feasibility of $\tilde{u}_\delta$ and $\tilde{u}_{\delta'}$ we obtain

$$(y_a + \delta_2 - \lambda\tilde{u}_\delta - \tilde{y}_\delta, \tilde{\mu}_a^{\delta'} - \tilde{\mu}_a^\delta)_I \leq 0, \text{ as well as } (y_a + \delta_2' - \lambda\tilde{u}_{\delta'} - \tilde{y}_{\delta'}, \tilde{\mu}_a^\delta - \tilde{\mu}_a^{\delta'})_I \leq 0.$$

Adding these inequality yields

$$(\lambda \delta u + \delta y, \delta\mu_a)_I \leq (\delta_2 - \delta_2', \delta\mu_a)_I \leq \|\delta_2 - \delta_2'\|_I \|\delta\mu_a\|_I. \tag{4.4.28}$$

Inserting (4.4.28) and an analogous estimate for $\lambda\|\delta\mu_b\|_I$ into (4.4.27) yields

$$\nu\|\delta u\|_I^2 \leq \|\delta_1 - \delta_1'\|_I \|\delta u\|_I + \|\delta_2 - \delta_2'\|_I \|\delta\mu_a\|_I + \|\delta_3 - \delta_3'\|_I \|\delta\mu_b\|_I. \tag{4.4.29}$$

In addition, note that by (4.4.22) and the fact that $M_3 \cap M_4 = \emptyset$ we have

$$\lambda\|\delta\mu_a\|_I = \lambda\|\delta\mu_a\|_{L^2(M_3)} \leq \|\delta_1 - \delta_1'\|_I + \|\psi_3\|_{\infty,\infty}\|\delta u\|_I + \|\psi_2\|_{\infty,\infty}\|\delta y\|_I + \|\delta p\|_I. \tag{4.4.30}$$

as well as

$$\lambda\|\delta\mu_b\|_I = \lambda\|\delta\mu_b\|_{L^2(M_4)} \leq \|\delta_1 - \delta_1'\|_I + \|\psi_3\|_{\infty,\infty}\|\delta u\|_I + \|\psi_2\|_{\infty,\infty}\|\delta y\|_I + \|\delta p\|_I. \tag{4.4.31}$$

- To estimate $\|\delta p\|_I$ in (4.4.30) and (4.4.31), we note that representation (4.4.23) yields

$$\delta\mu_a = \chi_3(-(\delta_1 - \delta_1') + \psi_2\delta y + \psi_3 \delta u + \delta p),$$

$$\delta\mu_b = -\chi_4(-(\delta_1 - \delta_1') + \psi_2\delta y + \psi_3 \delta u + \delta p),$$

where χ_i denotes the characteristic function of M_i. With (4.4.10), δp hence satisfies

$$\begin{aligned} -\partial_t \delta p + \mathcal{A}\delta p + (d_0 + \chi_3 + \chi_4)\delta p &= \psi_1 \delta y + \psi_2 \delta u + (\chi_3 + \chi_4)((\delta_1 - \delta_1') - \psi_2 \delta y - \psi_3 \delta u) \\ \delta p(T, \cdot) &= 0 \\ \delta p &= 0. \end{aligned} \tag{4.4.32}$$

Applying the stability result from Theorem 2.5.6 on page 26 to (4.4.32), we obtain

$$\|\delta p\|_I \leq c(\|\delta y\|_I + \|\delta u\|_I + \|\delta_1 - \delta_1'\|_I) \tag{4.4.33}$$

for some $c > 0$, where

$$\|\delta y\|_I \leq c\|\delta u\|_I \tag{4.4.34}$$

can be estimated by the same theorem.

- Inserting (4.4.34) and (4.4.33) in (4.4.30) and (4.4.31) yields

$$\|\delta\mu_a\|_I \leq c(\|\delta_1 - \delta_1'\|_I + \|\delta u\|_I) \tag{4.4.35}$$
$$\|\delta\mu_b\|_I \leq c(\|\delta_1 - \delta_1'\|_I + \|\delta u\|_I). \tag{4.4.36}$$

These estimates can be inserted in (4.4.29), and Young's inequality yields

$$\|\delta u\|_I \leq c\|\delta - \delta'\|_{(L^2(Q))^3}.$$

In turn, (4.4.33)-(4.4.36) yield the same estimate for $\delta y, \delta p$, and $\delta\mu_a, \delta\mu_b$.

□

With the L^2-stability at hand, we are able to derive an associated L^∞-result.

Theorem 4.4.11. *There exists a constant L_∞ such that for any given $\delta, \delta' \in (L^\infty(Q))^3$ the corresponding solutions of the auxiliary problem satisfy*

$$\|\tilde{u}_\delta - \tilde{u}_{\delta'}\|_{\infty,\infty} + \|\tilde{y}_\delta - \tilde{y}_{\delta'}\|_{\infty,\infty} + \|\tilde{p}_\delta - \tilde{p}_{\delta'}\|_{\infty,\infty} + \|\tilde{\mu}_a^\delta - \tilde{\mu}_a^{\delta'}\|_{\infty,\infty} + \|\tilde{\mu}_b^\delta - \tilde{\mu}_b^{\delta'}\|_{\infty,\infty} \leq L_\infty\|\delta - \delta'\|_{L^\infty(Q)^3}.$$

Proof. The proof requires a bootstrapping argument. We first prove a stability estimate for $\|\bar{\mu}_a^\delta - \bar{\mu}_a^{\delta'}\|_{2+s}$, with $s > s_0 > 0$ as used in the proof of Theorem 4.4.9. From the projection formula, we obtain on M_3

$$\begin{aligned}
\bar{\mu}_a^\delta - \bar{\mu}_a^{\delta'} = \; & \max\{0, \frac{\psi_3}{\lambda^2}(y_a + \delta_2 - \tilde{y}_\delta) + \frac{1}{\psi_3}(-\delta_1 + \psi_5 + \psi_2\tilde{y}_\delta + \tilde{p}_\delta)\} && (4.4.37)\\
& - \max\{0, \frac{\psi_3}{\lambda^2}(y_a + \delta_2' - \tilde{y}_{\delta'}) + \frac{1}{\psi_3}(-\delta_1' + \psi_5 + \psi_2\tilde{y}_{\delta'} + \tilde{p}_{\delta'})\} && (4.4.38)\\
\leq \; & \max\{0, \frac{\psi_3}{\lambda^2}(\delta_2 - \delta_2' - \delta y) + \frac{1}{\psi_3}(-(\delta_1 - \delta_1') + \psi_2\delta y + \delta p)\} && (4.4.39)
\end{aligned}$$

due to the properties of the max-function. By considering the corresponding inequality for $\bar{\mu}_a^{\delta'} - \bar{\mu}_a^\delta$ we obtain

$$\|\delta\mu_a\|_{L^{2+s}(Q)} \leq c\left(\|\delta_1 - \delta_1'\|_{\infty,\infty} + \|\delta_2 - \delta_2'\|_{\infty,\infty} + \|\delta y\|_{L^{2+s}(Q)} + \|\delta p\|_{L^{2+s}(Q)}\right) \tag{4.4.40}$$

for a constant $c > 0$. Noting that the adjoint states fulfill (4.4.32), we obtain

$$\|\delta p\|_{L^{2+s}(Q)} \leq c(\|\delta_1 - \delta_1'\|_{\infty,\infty} + \|\delta u\|_I + \|\delta y\|_I). \tag{4.4.41}$$

Similarly, $\|\delta y\|_{L^{2+s}(Q)}$ can be estimated by

$$\|\delta y\|_{L^{2+s}(Q)} \leq c\|\delta u\|_I. \tag{4.4.42}$$

Now, we apply Theorem 4.4.10 to estimate $\|\delta u\|_I$ and $\|\delta y\|_I$ in (4.4.41) and (4.4.42). After insertion in (4.4.40), this yields

$$\|\tilde{\mu}_a^\delta - \tilde{\mu}_a^{\delta'}\|_{L^{2+s}(Q)} \leq c(\|\delta - \delta'\|_{L^\infty(Q)^3} + \|\delta - \delta'\|_{L^2(Q)^3}) \leq c\|\delta - \delta'\|_{L^\infty(Q)^3}. \tag{4.4.43}$$

An analogous estimate holds for $\|\delta\mu_b\|_{L^{2+s}(Q)}$. Now, from the gradient equation (4.4.23) we deduce

$$\delta u = -\frac{1}{\psi_3}(-(\delta_1 - \delta_1') + \psi_2\delta y + \delta p + \lambda(\delta\mu_b - \delta\mu_a)) \quad \text{on } Q \setminus M_1 \cup M_2,$$

where $\delta\mu_a, \delta\mu_b \equiv 0$ outside M_3, M_4, respectively, and therefore

$$||\delta u||_{L^{2+s}(Q\setminus(M_1\cup M_2))} \le c(||\delta_1 - \delta_1'||_{\infty,\infty} + ||\psi_2||_{\infty,\infty}||\delta y||_{L^{2+s}(Q)} + ||\delta p||_{L^{2+s}(Q)} + \lambda(||\delta\mu_b||_{L^{2+s}(Q)} + ||\delta\mu_a||_{L^{2+s}(Q)}).$$

Inserting the estimates (4.4.41), (4.4.42), and (4.4.43) and its analogue for the upper bound and reapplying the previous steps leads to

$$||\delta u||_{L^{2+s}(Q\setminus(M_1\cup M_2))} \le c(||\delta - \delta'||_{L^\infty(Q)^3} + ||\delta u||_I) \le c||\delta - \delta'||_{L^\infty(Q)^3}.$$

It remains to estimate $||\delta u||_{L^{2+s}(Q)}$ on $M_1 \cup M_2$. On M_1, we observe from (4.4.23) that $\tilde{u}_\delta$ satisfies the projection formula

$$\tilde{u}_\delta = P_{[u_a,\infty]}(-\frac{1}{\psi_3}(-\delta_1 + \psi_5 + \psi_2 y_\delta + \tilde{p}_\delta)) \quad \text{on } M_1 \cup M_2,$$

and $\tilde{u}_{\delta'}$ satisfies an analogous formula. Moreover, the projection operator is Lipschitz with constant 1. Therefore, on M_1 we obtain pointwise

$$\begin{aligned} |\delta u| &= \left| P_{[u_a,\infty]}(-\frac{1}{\psi_3}(-\delta_1 + \psi_5 + \psi_2 y_\delta + \tilde{p}_\delta)) - P_{[u_a,u_b]}(-\frac{1}{\psi_3}(-\delta_1' + \psi_5 + \psi_2 \tilde{y}_{\delta'} + \tilde{p}_{\delta'}))\right| \\ &\le \frac{1}{\beta_0}\{|\delta_1' - \delta_1| + |\psi_2||\delta y| + |\delta p|\}, \end{aligned}$$

where we used the Legendre-Clebsch condition from Assumption 4.2.1. An analogous calculation on M_2 implies

$$||\delta u||_{L^{2+s}(M_1\cup M_2)} \le \frac{1}{\beta_0}(||\delta_1 - \delta_1'||_{\infty,\infty} + ||\psi_2||_{\infty,\infty}||\delta y||_{L^{2+s}(Q)} + ||\delta p||_{L^{2+s}(Q)}).$$

Estimating the norms as before, we obtain

$$||\delta u||_{L^{2+s}(Q)} \le c||\delta||_{L^\infty(Q)^3}. \tag{4.4.44}$$

From (4.4.40)-(4.4.42) we then obtain analogous estimates for the state, adjoint state, and the Lagrange multipliers.

Repeating these arguments finitely many times, we obtain estimates in $L^r(Q)$, $r > n/2+1$, which then imply an $L^\infty(Q)$-estimate for all variables. The assertion is obtained with an appropriate L_∞. □

Now, we transfer these results to the original perturbed problem $(\mathbb{P}_\delta)$.

Theorem 4.4.12. *Let $\tau > 0$ be given as in Assumption 4.4.5. There exists a constant $L > 0$ such that for any δ, δ' small enough, the unique solutions $\bar{u}_\delta, \bar{u}_{\delta'}$ of $(\mathbb{P}_\delta)$ with associated states $\bar{y}_\delta, \bar{y}_{\delta'}$, adjoint state $\bar{p}_\delta, \bar{p}_{\delta'}$, and Lagrange multipliers $\bar{\mu}_a^\delta, \bar{\mu}_b^\delta$ and $\bar{\mu}_a^{\delta'}, \bar{\mu}_b^{\delta'}$ satisfy*

$$||\bar{u}_\delta - \bar{u}_{\delta'}||_{\infty,\infty} + ||\bar{y}_\delta - \bar{y}_{\delta'}||_{\infty,\infty} + ||\bar{p}_\delta - \bar{p}_{\delta'}||_{\infty,\infty} + ||\bar{\mu}_a^\delta - \bar{\mu}_a^{\delta'}||_{\infty,\infty} + ||\bar{\mu}_b^\delta - \bar{\mu}_b^{\delta'}||_{\infty,\infty} \le L||\delta - \delta'||_{L^\infty(Q)^3}.$$

Proof. We will prove that the optimal solution $\tilde{u}_\delta$ of $(\mathbb{P}_\delta^{\text{aux}})$ is the unique optimal solution of $(\mathbb{P}_\delta)$ with associated state $\bar{y}_\delta = \tilde{y}_\delta$, and unique adjoint state $\bar{p}_\delta = \tilde{p}_\delta$ and multipliers $\bar{\mu}_a^\delta = \tilde{\mu}_a^\delta$, and $\bar{\mu}_b^\delta = \tilde{\mu}_b^\delta$. Since the solutions of the auxiliary and the original problem coincide, we can then apply the Lipschitz stability result for $(\mathbb{P}_\delta^{\text{aux}})$ to (4.4.6).

We first prove the feasibility of $\tilde{u}_\delta$ for $(\mathbb{P}_\delta)$. The control $\tilde{u}_\delta$ is feasible for $(\mathbb{P}_\delta^{\text{aux}})$, hence it remains to show

$$\tilde{u}_\delta \geq u_a \text{ on } Q \setminus M_1, \ \tilde{u}_\delta \leq u_b \text{ on } Q \setminus M_2,$$
$$\lambda\tilde{u}_\delta + \tilde{y}_\delta \geq y_a + \delta_2 \text{ on } Q \setminus M_3, \ \lambda\tilde{u}_\delta + \tilde{y}_\delta \leq y_b - \delta_3 \text{ on } Q \setminus M_4.$$

It is clear that in the unperturbed setting with $\delta = 0$ both Problems $(\mathbb{P}_\delta)$ and $(\mathbb{P}_\delta^{\text{aux}})$ admit for the same solution $\tilde{u}_0$. Then we know that $\tilde{u}_0 \geq u_a + \tau$ a.e. on $Q \setminus M_1$. Hence, we have

$$\tilde{u}_\delta = \tilde{u}_0 + \tilde{u}_\delta - \tilde{u}_0 \geq \tilde{u}_0 - \|\tilde{u}_\delta - \tilde{u}_0\|_{\infty,\infty} \geq u_a + \tau - L_\infty \|\delta\|_{L^\infty(Q)^3} \geq u_a$$

almost everywhere on $Q \setminus M_1$ for $\|\delta\|_{L^\infty(Q)^3}$ sufficiently small, since we can apply the Lipschitz stability result from Theorem 4.4.11. The upper bound u_b is treated similarly. For the mixed control-state constraints, we know

$$\lambda\tilde{u}_0 + \tilde{y}_0 \geq y_a + \tau \quad \text{a.e. on } M_3.$$

Consequently,

$$\begin{aligned}
\lambda\tilde{u}_\delta + \tilde{y}_\delta - \delta_2 &= \lambda\tilde{u}_0 + \tilde{y}_0 + \lambda\tilde{u}_\delta + \tilde{y}_\delta - \delta_2 - \lambda\tilde{u}_0 - \tilde{y}_0 \\
&\geq y_a + \tau - \|\delta_2\|_{\infty,\infty} - \lambda\|\tilde{u}_\delta - \tilde{u}_0\|_{\infty,\infty} - \|\tilde{y}_\delta - \tilde{y}_0\|_{\infty,\infty} \\
&\geq y_a + \tau - \|\delta_2\|_{\infty,\infty} - \lambda L_\infty \|\delta\|_{L^\infty(Q)^3} - L_\infty \|\delta\|_{L^\infty(Q)^3} \\
&\geq y_a + \tau - c\|\delta\|_{L^\infty(Q)^3} \\
&\geq y_a
\end{aligned}$$

almost everywhere on $Q \setminus M_3$ for $\|\delta\|_{L^\infty(Q)^3}$ sufficiently small. The upper bound is treated analogously.

This proves feasibility of $\tilde{u}_\delta$ for Problem (4.4.6) and we deduce $\tilde{u}_\delta = \bar{u}_\delta$. To prove that $\tilde{p}_\delta$ as well as $\tilde{\mu}_a^\delta$ and $\tilde{\mu}_b^\delta$ are the unique dual variables for Problem (4.4.6) we show additionally that the almost active sets associated with $\bar{u}_\delta$ do not intersect. We consider a point (t^*, x^*) where the lower mixed control-state constraints for $\bar{u}_\delta$ are almost active, i.e.

$$\lambda\bar{u}_\delta(t^*, x^*) + \bar{y}_\delta(t^*, x^*) \leq y_a + \delta_2 + \tau/2.$$

We know

$$\begin{aligned}
\lambda\bar{u}_0(t^*, x^*) + \bar{y}_0(t^*, x^*) = {} & \lambda\bar{u}_0(t^*, x^*) + \bar{y}_0(t^*, x^*) - \lambda\bar{u}_\delta(t^*, x^*) - \bar{y}_\delta(t^*, x^*) && (4.4.45)\\
& + \lambda\bar{u}_\delta(t^*, x^*) + \bar{y}_\delta(t^*, x^*) && (4.4.46)\\
\leq {} & \lambda\|\bar{u}_0 - \bar{u}_\delta\|_{\infty,\infty} + \|\bar{y}_0 - \bar{y}_\delta\|_{\infty,\infty} + \|\delta_2\|_{\infty,\infty} + y_a + \tau/2 && (4.4.47)\\
\leq {} & c\|\delta\|_{L^\infty(Q)^3} + y_a \leq \tau + y_a, && (4.4.48)
\end{aligned}$$

if again $\|\delta\|_{L^\infty(Q)^3}$ is chosen sufficiently small. Hence, $(t^*, x^*) \in M_3$. Because the almost active sets do not intersect for $\delta = 0$ we find e.g.

$$\bar{u}_\delta(t^*, x^*) \geq \bar{u}_0(t^*, x^*) - \|\bar{u}_\delta - \bar{u}_0\|_{\infty,\infty} \geq \bar{u}_0(t^*, x^*) - c\|\delta\|_{L^\infty(Q)^3} \geq u_a + \tau - c\|\delta\|_{L^\infty(Q)^3} \geq u_a + \tau/2$$

for $\|\delta\|_{L^\infty(Q)}$ small enough. Applying analogous arguments to the other constraints, we obtain that the $\tau/2$-active sets according to Definition 4.4.4 do not intersect. Then, conditions (4.4.7)-(4.4.9) are indeed necessary and sufficient for optimality of $\bar{u}_\delta$, and in particular imply that the adjoint state $\bar{p}_\delta$ and the Lagrange multipliers $\bar{\mu}_a^\delta, \bar{\mu}_b^\delta$ are unique. It is easy to see that $\tilde{u}_\delta, \tilde{\mu}_a^\delta, \tilde{\mu}_b^\delta$ satisfy the optimality system (4.4.7)-(4.4.9). In summary, we have shown that $\bar{u}_\delta = \tilde{u}_\delta$ with associated state $\bar{y}_\delta = \tilde{y}_\delta$ is the unique solution of $(\mathbb{P}_\delta)$ with unique multipliers $\bar{\mu}_a^\delta = \tilde{\mu}_a^\delta$, $\bar{\mu}_b^\delta = \tilde{\mu}_b^\delta$ and unique adjoint state $\bar{p}_\delta = \tilde{p}_\delta$. $\square$

Let us end this section with a comment on the importance of L^∞-Lipschitz stability rather than L^2-stability, only. In [7], it has been illustrated that for L^2- stability, only, the stability results for $(\mathbb{P}_\delta^{\text{aux}})$ cannot be applied to Problem (4.4.6). In particular, while the Lagrange multipliers associated with $(\mathbb{P}_\delta^{\text{aux}})$ are alway unique due to the separated active sets, there exist L^2-perturbations such that the dual variables for (4.4.6) are not unique, consequently are not Lipschitz continuous, and finally the generalized equation (4.4.3) is not strongly regular. This is due to the fact that the active sets associated with L^2-perturbed controls may intersect even if $\|\delta\|_{L^2(Q)^3}$ is arbitrary small.

Collecting all our results, we obtain that the linearized generalized equation is uniquely solvable with solution and regular Lagrange multipliers depending Lipschitz continuously on the perturbations. With respect to the motivating example in Section 4.4.2 we can now conclude that the generalized equation (4.4.3) is strongly regular.

Corollary 4.4.13. *Let Assumption 4.4.1, cf. page 91, hold. Then the generalized equation* (4.4.3) *is strongly regular.*

We will comment on one important implication of this fact in the next section.

4.5. Local uniqueness of local solutions

From the last section, we know that the generalized optimality system associated with a locally optimal control $\bar{u}_\lambda$ of $(\mathbb{P}_\lambda)$, which fulfills the first order optimality conditions of Theorem 4.3.7 and the second order sufficient condition from Assumption 4.3.9 is strongly regular if Assumption 4.3.11 is fulfilled. We conclude:

Theorem 4.5.1. *Let Assumption 4.2.1 hold and let $\bar{u}_\lambda$ denote a locally optimal control of $(\mathbb{P}_\lambda)$, which fulfills the first order optimality conditions of Theorem 4.3.7 and the second order sufficient condition from Assumption 4.3.9. Moreover, suppose that Assumption 4.3.11 is satisfied. Then, $\bar{u}_\lambda$ is locally unique in the sense of $L^\infty(Q)$.*

Proof. With the strong regularity of (4.4.3) this is an immediate consequence of Robinson's implicit function theorem, [127]. □

In control-constrained optimal control problems, local uniqueness can be concluded directly from a second order sufficient optimality condition. Let us illustrate the main idea with the help of a finite-dimensional model problem, so that we do not have to take care of appropriate function spaces for differentiability.

Let $f\colon \mathbb{R}^m \to \mathbb{R}$ be given, satisfying all desired differentiability and regularity properties, and consider the optimization problem

$$\text{Minimize } f(x), \quad \text{subject to } x \in X_{\text{ad}}$$

where X_{ad} is a non-empty, convex and closed subset of $\mathbb{R}^m$. Let a local solution $x_0 \in X_{\text{ad}}$ that satisfies the first order necessary optimality condition

$$f'(x_0)(x - x_0) \geq 0 \quad \forall x \in X_{\text{ad}},$$

and the very strong second order sufficient condition

$$f''(x_0)[v,v] \geq \alpha|v|^2$$

for all v in $\mathbb{R}^m$ for some $\alpha > 0$ be given. Now assume that for all $\rho > 0$, there exists $\tilde{x}_0 \neq x_0$ with $|x_0 - \tilde{x}_0| \leq \rho$ such that the first order optimality condition

$$f'(\tilde{x}_0)(x - \tilde{x}_0) \geq 0, \quad \forall x \in X_{\text{ad}}$$

is satisfied. Then, we obtain

$$0 \leq \left(f'(x_0) - f'(\tilde{x}_0)\right)(\tilde{x}_0 - x_0),$$

and by Taylor expansion of f' in x_0 we have

$$0 \leq -f''(x_0)[\tilde{x}_0 - x_0, \tilde{x}_0 - x_0] - (f''(x_\xi) - f''(x_0))[\tilde{x}_0 - x_0, \tilde{x}_0 - x_0],$$

for some $x_\xi = x_0 + \xi(\tilde{x}_0 - x_0)$, $\xi \in (0,1)$. Assuming that f'' is Lipschitz continuous and making use of the coercivity condition for $f''(x_0)$, we obtain

$$0 \leq -\alpha|x_0 - \tilde{x}_0|^2 + \xi|x_0 - \tilde{x}_0||x_0 - \tilde{x}_0|^2.$$

Dividing by $|x_0 - \tilde{x}_0|^2$ yields

$$\alpha \leq \xi|x_0 - \tilde{x}_0| \leq \xi\rho.$$

Obviously, the strict positivity of α contradicts the fact that ρ can be arbitrary small, and we deduce that there exists $\rho_0 > 0$ such that x_0 is the unique stationary point in a neighborhood with radius ρ_0.

It is clear that this procedure is valid for control-constrained optimal control problems in function spaces, provided that differentiability and Lipschitz continuity properties are fulfilled and very strong SSC are satisfied. For state-constraints the situation is different, since the first order optimality conditions additionally contain terms associated with the Lagrange multipliers. Then, it is not sufficient to consider the objective function, but essentially a Lipschitz continuity result of the second derivative of the Lagrangian is needed. That includes a Lipschitz continuity result for the Lagrange multipliers. For Lavrentiev regularized problems, the situation is similarly complicated. We have already motivated in Remark 4.4.8 that a Lipschitz result for the multipliers is not easily obtained directly. Yet, an extension of this method to the virtual control concept has recently been employed successfully for elliptic control problems with tracking type objective function by Krumbiegel, Rösch, and the author in [83]. At first glance, this difference between the virtual control concept and Lavrentiev regularization seems rather surprising, since in both regularization approaches the pure pointwise state constraints are replaced by mixed control-state constraints. However, we have already pointed out some differences. We do not aim here at a complete analysis of the regularization of Problem (4.2.1) by the virtual control concept, but rather comment briefly on the prescribed constraints and the consequences on the optimality system. If $f_{vc} = f_{vc}(u_\varepsilon, v_\varepsilon, y_\varepsilon)$ denotes the regularized reduced objective function depending on the control u_ε, the additional, virtual control v_ε, and the state y_ε, we know that the optimality conditions with respect to the constraints

$$u_a \leq u_\varepsilon \leq u_b, \quad y_a \leq \varepsilon v_\varepsilon + y_\varepsilon \leq y_b.$$

yield

$$\frac{\partial f_{vc}}{\partial u_\varepsilon}(\bar{u}_\varepsilon, \bar{v}_\varepsilon, \bar{y}_\varepsilon)(u - \bar{u}_\varepsilon) \quad \forall u_a \leq u \leq u_b, \quad \frac{\partial f_{vc}}{\partial v_\varepsilon}(\bar{u}_\varepsilon, \bar{v}_\varepsilon, \bar{y}_\varepsilon) + \varepsilon(\bar{\mu}_b^\varepsilon - \bar{\mu}_a^\varepsilon) = 0.$$

The fact that the Lagrange multipliers appear in a gradient equation in the whole domain rather than in a variational inequality as in Theorem 4.3.7 can be exploited when proving Lipschitz stability results. For more details, we refer to [83].

4.6. Convergence analysis

4.6.1. Main assumption and an auxiliary problem

Up to now, we have collected and proven many desirable properties of Lavrentiev regularized problems. Let us now conclude with a convergence result, to motivate that the regularization is meaningful in the sense that locally optimal controls of the unregularized problem $(\mathbb{P})$ can be approximated by locally optimal controls of $(\mathbb{P}_\lambda)$. Eventually, we aim at proving a convergence result of somehow corresponding solutions. We impose an assumption of quadratic growth for the unregularized problem.

Assumption 4.6.1. *We assume there exist positive real numbers ε and α such that the local solution $\bar{u}$ of $(\mathbb{P})$ satisfies the quadratic growth condition*

$$f(\bar{u}) + \frac{\beta}{2}\|u - \bar{u}\|^2 \leq f(u) \tag{4.6.1}$$

for every control $u \in U_{feas}$ that satisfies $\|u - \bar{u}\|_{L^p(Q)} \leq \varepsilon$ with some $n/2 + 1 < p \leq \infty$.

We would like to point out that while it is difficult to derive a quadratic growth condition from a definiteness property of the second derivative of the Lagrangian in the parabolic setting with space dimension greater than one, it is still reasonable to assume such a condition. After all, the fact that second-order sufficient conditions are not readily available does not imply that there are no well-defined local solutions of the unregularized problem.

The proof of convergence is very much similar to the proof of Theorem 3.4.1 in Chapter 3, where convergence for the finite element discretization of state-constrained elliptic control problems was proven. In fact, in [72] a convergence result for local solutions of elliptic problems is shown that includes both Lavrentiev type regularization and discretization.

Therefore, we use similar arguments as in the discrete setting. We again follow [26] and consider the auxiliary problem

$$\text{Minimize } f(u), \quad \text{subject to } u \in U^\varepsilon_{\text{ad}}, \quad y_a \leq \lambda u + G(u) \leq y_b. \tag{$\mathbb{P}^\varepsilon_\lambda$}$$

Here, the admissible set $U^\varepsilon_{\text{ad}}$ associated with a given locally optimal control $\bar{u}$ of $(\mathbb{P})$ is given as

$$U^\varepsilon_{\text{ad}} := \{u \in U_{\text{ad}} \mid \|u - \bar{u}\|_{L^p(Q)} \leq \varepsilon\}, \quad p > n/2 + 1.$$

In addition, we define the associated auxiliary set of feasible controls,

$$U^\varepsilon_{\text{feas}} := \{u \in U^\varepsilon_{\text{ad}} \colon y_a \leq \lambda u + G(u) \leq y_b\}.$$

We will analyse this auxiliary problem in the following, and eventually conclude with the observation that solutions of $(\mathbb{P}^\varepsilon_\lambda)$ are also solutions to $(\mathbb{P}_\lambda)$. Moreover, we will demonstrate that some conditions we assumed in Section 4.3 can be proven to hold in the neighborhood of locally optimal controls $\bar{u}$ of $(\mathbb{P})$.

4.6.2. Auxiliary results

We proceed as in Section 3.3.4. We will construct an auxiliary sequence $\{u_{t(\lambda_n)}\}$ which is feasible for $(\mathbb{P}^\varepsilon_\lambda)$ for all λ_n sufficiently small, converging to $\bar{u}$, as well as a sequence $\{u_{\tau(\lambda_n)}\}$, feasible for $(\mathbb{P})$ and converging to an optimal solution $\bar{u}_\lambda^\varepsilon$ of $(\mathbb{P}_\lambda^\varepsilon)$. Of course, solvability of $(\mathbb{P}_\lambda^\varepsilon)$ has to be proven. All this can be obtained with the help of the Slater point $u_\gamma \in U_{\rm ad}$ from Assumption 4.2.7. We define

$$u_{\gamma_\varepsilon} := \bar{u} + t(u_\gamma - \bar{u}), \quad t := \min\left\{1, \frac{\varepsilon}{\|u_\gamma - \bar{u}\|_{L^p(Q)}}\right\}. \tag{4.6.2}$$

In other words, $u_{\gamma_\varepsilon} = u_\gamma$ if $u_\gamma \in U_{\rm ad}^\varepsilon$, or u_{γ_ε} is defined by moving the Slater point u_γ to the neighborhood of $\bar{u}$ by a convex combination. It is clear, cf. Remark 3.1.8, that u_{γ_ε} is a Slater point for $\bar{u}$ with Slater parameter $\gamma_\varepsilon = \min\{1, \varepsilon\}\gamma$.

The first result is the analogue of Lemma 3.3.13.

Lemma 4.6.2. *Let $\bar{u}$ be a feasible control for $(\mathbb{P})$ satisfying the linearized Slater condition of Assumption 4.2.7 and let $\{\lambda_n\}$, $\lambda_n > 0$, be a sequence converging to zero. Then there exists a sequence $\{u_{t(\lambda_n)}\}$ converging strongly in $L^\infty(Q)$ to $\bar{u}$ as $n \to \infty$ such that $u_{t(\lambda_n)}$ is feasible for $(\mathbb{P}_\lambda^\varepsilon)$ for all sufficiently large n.*

Proof. Notice that $\bar{u} \in L^\infty(Q)$. In what follows we use the short notation $u_n := u_{t(\lambda_n)}$ as well as $t_n := t(\lambda_n)$. Now choose

$$u_n = \bar{u} + t_n(u_{\gamma_\varepsilon} - \bar{u}),$$

where $\lambda_n > 0$ is given sufficiently small. It is clear that $u_n \to \bar{u}$ in $L^\infty(Q)$ as $t_n \to 0$, and $u_n \in U_{\rm ad}$ for $t_n \leq 1$. It remains to show the feasibility for $(\mathbb{P}_\lambda^\varepsilon)$. For the upper state constraint, we obtain

$$\begin{array}{rcl} \lambda_n u_n + G(u_n) & = & \lambda_n u_n + G(\bar{u} + t_n(u_{\gamma_\varepsilon} - \bar{u})) \\ & \leq & \lambda_n \|\bar{u} + t_n(u_{\gamma_\varepsilon} - \bar{u})\|_{\infty,\infty} + G(\bar{u}) + t_n G'(\bar{u})(u_{\gamma_\varepsilon} - \bar{u}) + o(t_n) \\ & \leq & c\lambda_n + (1 - t_n)G(\bar{u}) + t_n(G(\bar{u}) + G'(\bar{u})(u_{\gamma_\varepsilon} - \bar{u})) + o(t_n) \\ & \leq & c\lambda_n + (1 - t_n)y_b + t_n y_b - t_n\gamma_\varepsilon + o(t_n) \\ & = & y_b + c\lambda_n - t_n\left(\gamma_\varepsilon + \frac{o(t_n)}{t_n}\right). \end{array} \tag{4.6.3}$$

Take t_0 small enough to ensure $\gamma_\varepsilon + \frac{o(t_0)}{t_0} \geq \frac{\gamma_\varepsilon}{2}$ and define t_n by $c\lambda_n - t_n\frac{\gamma_\varepsilon}{2} = 0$. Then we have

$$t_n = t_n(\lambda_n) = \frac{2c}{\gamma_\varepsilon}\lambda_n,$$

which for λ_n sufficiently small yields $t(\lambda_n) \leq t_0$. Inserting this in (4.6.3) yields

$$\lambda_n u_n + G(u_n) \leq y_b + c\lambda_n - t_n(\gamma_\varepsilon + \frac{o(t_n)}{t_n}) \leq y_b \quad \forall 0 < \lambda_n \leq \lambda_0,$$

since $c\lambda - t_n(\gamma_\varepsilon + \frac{o(t_n)}{t_n}) \leq 0$. Analogously, we can deal with the lower state constraint. □

As for Theorem 3.3.15 on page 48, a side effect of the last lemma is that the feasible set for $(\mathbb{P}_\lambda^\varepsilon)$ is not empty for all sufficiently large n, i.e. for all sufficiently small λ_n.

Theorem 4.6.3. *Let $\bar{u} \in U_{ad}$ be a locally optimal control of Problem $(\mathbb{P})$ and let Assumption 4.2.7 hold. Then, for all sufficiently small $\lambda > 0$, Problem $(\mathbb{P}_\lambda^\varepsilon)$ admits at least one optimal solution $\bar{u}_\lambda^\varepsilon$.*

Proof. Just like Theorem 4.2.5, this follows from the discussion in [144]. □

Again, uniqueness of solutions is not a priori given, but this does not interfere with our analysis. Moreover, we can prove, and not just assume as in Assumption 4.3.6, that there exists a Slater point for Problem $(\mathbb{P}_\lambda^\varepsilon)$ if $\lambda > 0$ is sufficiently small. Therefore, as an analogue to Lemma 3.3.16 we obtain:

Lemma 4.6.4. *Let $\bar{u}_\lambda^\varepsilon$ be a global solution of Problem $(\mathbb{P}_\lambda^\varepsilon)$. If $\varepsilon > 0$ and $\lambda > 0$ are sufficiently small, then $u_\gamma^\varepsilon \in U_{ad}^\varepsilon$ satisfies the linearized Slater condition*

$$u_a + \frac{\gamma_\varepsilon}{2} \leq u_\gamma^\varepsilon \leq u_b - \frac{\gamma_\varepsilon}{2}, \quad y_a + \frac{\gamma_\varepsilon}{2} \leq \lambda u_\gamma^\varepsilon + G(\bar{u}_\lambda^\varepsilon) + G'(\bar{u}_\lambda^\varepsilon)(u_\gamma^\varepsilon - \bar{u}_\lambda^\varepsilon) \leq y_b - \frac{\gamma_\varepsilon}{2}.$$

Proof. The inequality for the control constraints is trivial. For the state constraints, we observe

$$\begin{aligned}\lambda u_{\gamma_\varepsilon} + G(\bar{u}_\lambda^\varepsilon) + G'(\bar{u}_\lambda^\varepsilon)(u_{\gamma_\varepsilon} - \bar{u}_\lambda^\varepsilon) = \lambda u_{\gamma_\varepsilon} + G(\bar{u}) + G'(\bar{u})(u_{\gamma_\varepsilon} - \bar{u}) \\ + (G(\bar{u}_\lambda^\varepsilon) - G(\bar{u})) + (G'(\bar{u}_\lambda^\varepsilon) - G'(\bar{u}))(u_{\gamma_\varepsilon} - \bar{u}_\lambda^\varepsilon) + G'(\bar{u})(\bar{u} - \bar{u}_\lambda^\varepsilon).\end{aligned}$$

Due to u_{γ_ε} being bounded, λ can be chosen small enough such that

$$\lambda u_{\gamma_\varepsilon} \leq \lambda \|u_{\gamma_\varepsilon}\|_{\infty,\infty} \leq \frac{\gamma_\varepsilon}{8}.$$

Also, if $\varepsilon > 0$ is sufficiently small, we obtain

$$G(\bar{u}_\lambda^\varepsilon) - G(\bar{u}) \leq c\|\bar{u}_\lambda^\varepsilon - \bar{u}\|_{L^p(Q)} \leq \frac{\gamma_\varepsilon}{8},$$

by estimate (4.2.6), as well as

$$(G'(\bar{u}_\lambda^\varepsilon) - G'(\bar{u}))(u_{\gamma_\varepsilon} - \bar{u}_\lambda^\varepsilon) \leq c\|\bar{u}_\lambda^\varepsilon - \bar{u}\|_{\infty,\infty}\|u_{\gamma_\varepsilon} - \bar{u}_\lambda^\varepsilon\|_{L^p(Q)} \leq 2c\varepsilon \leq \frac{\gamma_\varepsilon}{8}$$

by (4.2.8), and

$$G'(\bar{u})(\bar{u} - \bar{u}_\lambda^\varepsilon) \leq c\|\bar{u} - \bar{u}_\lambda^\varepsilon\|_{L^p(Q)} \leq \frac{\gamma_\varepsilon}{8},$$

by Proposition 2.5.2 and the boundedness of $\bar{y}$. Hence,

$$\lambda u_{\gamma_\varepsilon} + G(\bar{u}_\lambda^\varepsilon) + G'(\bar{u}_\lambda^\varepsilon)(u_{\gamma_\varepsilon} - \bar{u}_\lambda^\varepsilon) \leq G(\bar{u}) + G'(\bar{u})(u_{\gamma_\varepsilon} - \bar{u}) + \frac{\gamma_\varepsilon}{2} \leq y_b - \frac{\gamma_\varepsilon}{2},$$

by the assumption of a Slater condition for the unregularized problem. The lower inequality holds by analogous arguments. □

With the help of this Slater point, we provide an additional auxiliary sequence that is feasible for $(\mathbb{P})$. The analogue to Corollary 3.3.17 reads:

Corollary 4.6.5. *Let $\bar{u}$ denote an optimal control of Problem $(\mathbb{P})$ and let $\{\lambda_n\}_{n\in\mathbb{N}}$ denote a sequence of positive regularization parameters converging to zero. Moreover, let $\{\bar{u}_n^\varepsilon\}$ be any sequence of globally optimal controls of $(\mathbb{P}_\lambda^\varepsilon)$ for $\lambda_n \downarrow 0$, $n \to \infty$. Additionally, let $\varepsilon > 0$ be sufficiently small that Assumption 4.2.9 is satisfied. Then, there exists a sequence of feasible controls $\{v_{\tau(\lambda_n)}^\varepsilon\}$ of $(\mathbb{P})$ with $\|v_{\tau(\lambda_n)}^\varepsilon - \bar{u}\|_{\infty,\infty} \leq \varepsilon$ such that $\|v_{\tau(\lambda_n)}^\varepsilon - \bar{u}_n^\varepsilon\|_{\infty,\infty} \to 0$ as $n \to \infty$.*

Proof. Thanks to the control constraints, all $\bar{u}_n^\varepsilon$ belong to $L^\infty(Q)$. We abbreviate $v_n^\varepsilon := v_{\tau(\lambda_n)}^\varepsilon$ as well as $\tau_n := \tau(\lambda_n)$, and define

$$v_n^\varepsilon = \bar{u}_n^\varepsilon + \tau_n(u_{\gamma_\varepsilon} - \bar{u}_n^\varepsilon)$$

with $\tau_n = \frac{2c}{\hat{\gamma}}\lambda_n$ and $c = \max\{\|u_{\gamma_\varepsilon}\|_{\infty,\infty}, \|\bar{u}_n^\varepsilon\|_{\infty,\infty}\}$. Then for $\tau_n \downarrow 0$, $\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_{\infty,\infty} \to 0$ is satisfied and $\|v_n^\varepsilon - \bar{u}\|_{\infty,\infty} \leq \varepsilon$ holds for n large enough. We obtain

$$\begin{aligned}
-\lambda_n\|v_n^\varepsilon\|_{\infty,\infty} + G(v_n^\varepsilon) \leq\ & \lambda_n v_n^\varepsilon + G(v_n^\varepsilon) && (4.6.4)\\
=\ & (1-\tau_n)\lambda_n\,\bar{u}_n^\varepsilon + \tau_n\lambda_n u_{\gamma_\varepsilon} + G(\bar{u}_n + \tau_n(u_{\gamma_\varepsilon} - \bar{u}_n^\varepsilon)) && (4.6.5)\\
=\ & (1-\tau_n)\lambda_n\bar{u}_n^\varepsilon + (1-\tau_n)G(\bar{u}_n^\varepsilon) && (4.6.6)\\
& +\tau_n(\lambda_n u_{\gamma_\varepsilon} + G(\bar{u}_n^\varepsilon) + G'(\bar{u}_n^\varepsilon)(u_{\gamma_\varepsilon} - \bar{u}_n^\varepsilon)) + o(\tau_n)\\
\leq\ & (1-\tau_n)y_b + \tau_n\left(y_b - \frac{\gamma_\varepsilon}{2}\right) + o(\tau_n) \leq y_b - \tau_n\frac{\gamma_\varepsilon}{2} + o(\tau_n). && (4.6.7)
\end{aligned}$$

This implies $G(v_n^\varepsilon) \leq y_b - t\frac{\gamma_\varepsilon}{2} + \lambda_n\|v_n^\varepsilon\|_{\infty,\infty} \leq y_b$ by the definition of τ_n, hence v_n^ε satisfies the upper state constraint of $(\mathbb{P})$. Analogously, it satisfies the lower one. For $\lambda_n \downarrow 0$, τ_n tends to zero so that $0 < \tau_n < 1$ holds for sufficiently large n. Therefore v_n^ε, as the convex combination of two elements of $U_{\mathrm{ad}}^\varepsilon$, belongs to the same set. □

4.6.3. A regularization error estimate

Now we proceed to show that $\bar{u}$, the locally optimal reference control of $(\mathbb{P})$, can be approximated by optimal controls of $(\mathbb{P}_\lambda^\varepsilon)$. Moreover, we prove an estimate for the regularization error. We mention here the results in [31, 72], where the regularization error has been estimated for elliptic problems.

Theorem 4.6.6. *Let $\bar{u}$ be a locally optimal control of $(\mathbb{P})$ in the sense of $L^p(Q)$, p taken from Assumption 4.6.1, satisfying the quadratic growth condition (4.6.1) and Assumption 4.2.7 (linearized Slater condition) and fix $\varepsilon > 0$. Then, for all sufficiently small $\lambda > 0$, problem $(\mathbb{P}_\lambda^\varepsilon)$ has an optimal control. If $\{\bar{u}_n^\varepsilon\}$ is any sequence of (globally) optimal controls for $(\mathbb{P}_\lambda^\varepsilon)$, then it converges strongly in $L^q(Q)$ to $\bar{u}$, for all $2 \leq q < \infty$. Moreover, it converges in $L^2(Q)$ with rate $\sqrt{\lambda_n}$, i.e. there exists $c > 0$ such that*

$$\|\bar{u}_n^\varepsilon - \bar{u}\|_I \leq c\sqrt{\lambda_n}.$$

Proof. From the quadratic growth condition (4.6.1) we find

$$f(u) \geq f(\bar{u}) + \beta\|u - \bar{u}\|_I^2 \quad \forall u \in U_{\mathrm{feas}}^\varepsilon,$$

for ε sufficiently small. The above inequality holds especially for $u = v_n^\varepsilon$ constructed in Corollary 4.6.5, since this function is feasible for $(\mathbb{P})$. This yields

$$\begin{aligned}
f(v_n^\varepsilon) \ &\geq\ f(\bar{u}) + \beta\|v_n^\varepsilon - \bar{u}\|_I^2\\
&=\ f(\bar{u}) + \beta(\|\bar{u}_n^\varepsilon - \bar{u}\|_I^2 + 2(v_n^\varepsilon - \bar{u}_n^\varepsilon, \bar{u}_n^\varepsilon - \bar{u})_I + \|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I^2)\\
&\geq\ f(\bar{u}) + \beta(\|\bar{u}_n^\varepsilon - \bar{u}\|_I^2 - c\,\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I),
\end{aligned}$$

where the last inequality follows from $\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I^2 > 0$,

$$(v_n^\varepsilon - \bar{u}_n^\varepsilon, \bar{u}_n^\varepsilon - \bar{u})_I \geq -\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I\|\bar{u}_n^\varepsilon - \bar{u}\|_I,$$

and $\|\bar{u}_n^\varepsilon - \bar{u}\|_I \leq c$. We obtain

$$f(\bar{u}_n^\varepsilon) = f(v_n^\varepsilon) - (f(v_n^\varepsilon) - f(\bar{u}_n^\varepsilon)) \geq f(\bar{u}) + \beta(\|\bar{u}_n^\varepsilon - \bar{u}\|_I^2 - c\,\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I) - (f(v_n^\varepsilon) - f(\bar{u}_n^\varepsilon)),$$

which yields

$$\beta\|\bar{u}_n^\varepsilon - \bar{u}\|_I^2 \leq f(\bar{u}_n^\varepsilon) - f(\bar{u}) + c\beta\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I + f(v_n^\varepsilon) - f(\bar{u}_n^\varepsilon). \tag{4.6.8}$$

On the other hand, we have $f(\bar{u}_n^\varepsilon) \le f(u)$ for all feasible $u \in U_{\text{feas}}^\varepsilon$, hence

$$f(\bar{u}_n^\varepsilon) - f(\bar{u}) \le f(u) - f(\bar{u}).$$

Inserting this inequality in (4.6.8) yields, with

$$u = u_n := \bar{u} + t_n(u_{\gamma_\varepsilon} - \bar{u}) \tag{4.6.9}$$

and $\tau_n := \frac{2c}{\gamma_\varepsilon}\lambda_n$,

$$\beta\|\bar{u}_n^\varepsilon - \bar{u}\|_I^2 \le f(u_n) - f(\bar{u}) + f(v_n^\varepsilon) - f(\bar{u}_n^\varepsilon) + c\beta\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_I. \tag{4.6.10}$$

By the definition of $v_n^\varepsilon := \bar{u}_n^\varepsilon + \tau_n(u_{\gamma_\varepsilon} - \bar{u}_n^\varepsilon)$ with $\tau_n = \frac{2c}{\gamma_\varepsilon}\lambda_n$, as well as the uniform boundedness of $\bar{u}_n^\varepsilon$ we obtain that

$$\|v_n^\varepsilon - \bar{u}_n^\varepsilon\|_{\infty,\infty} \le c\lambda_n. \tag{4.6.11}$$

With (4.6.9), we further obtain

$$\|u_n - \bar{u}\|_{\infty,\infty} \le c\lambda_n. \tag{4.6.12}$$

The functional f is Lipschitz continuous with respect to the L^∞-norm, cf. Lemma 4.2.4. Then, inserting (4.6.11) and (4.6.12) into (4.6.10) yields

$$\|\bar{u}_n^\varepsilon - \bar{u}\|^2 \le \frac{c}{\beta}\lambda_n,$$

which implies that $\bar{u}_n^\varepsilon$ converges strongly in $L^2(Q)$ towards $\bar{u}$ with rate $\sqrt{\lambda_n}$. Since $\bar{u}_n^\varepsilon$ belongs to U_{ad}, this sequence is uniformly bounded, hence it converges also in $L^q(Q)$ with $q < \infty$, and the associated states converge uniformly on $\bar{Q}$. □

Now, we have to ensure that $\bar{u}_n^\varepsilon$ is actually a local solution of $(\mathbb{P}_\lambda)$, since it is not yet clear that $\bar{u}_n^\varepsilon$ does not touch the boundary of $U_{\text{ad}}^\varepsilon$. We need an additional assumption.

Theorem 4.6.7. *Let the assumptions of Theorem 4.6.6 be satisfied and suppose in addition that Assumption 4.6.1 is satisfied with $p < \infty$. Then, there exists a sequence of local solutions to $(\mathbb{P}_\lambda)$ that converges strongly in $L^p(Q)$ to $\bar{u}$.*

Proof. The result is a simple conclusion from the last theorem, since $\bar{u}_n^\varepsilon \to \bar{u}$ in $L^q(Q)$ for all $q < \infty$, in particular for p. Therefore, $\|\bar{u}_n^\varepsilon - \bar{u}\|_{L^p(Q)} < \varepsilon$ must hold for sufficiently large n. In this case, $\bar{u}_n^\varepsilon$ is a solution to $(\mathbb{P}_\lambda^\varepsilon)$ that is in the interior of $U_{\text{ad}}^\varepsilon$. Therefore, it is a local solution to $(\mathbb{P}_\lambda)$. □

If Assumption 4.6.1 is only satisfied for $p = \infty$, this convergence result is not applicable. However, under the strong assumption that $\bar{u}_n^\varepsilon$ converges strongly to $\bar{u}$ in $L^\infty(Q)$, Theorem 4.6.7 remains true for $p = \infty$. Moreover, under this condition, we can deduce the separation condition on active sets for $(\mathbb{P}_\lambda)$ from one imposed on $\bar{u}$ in problem $(\mathbb{P})$. Therefore, we define the notion of $\sigma-$active sets for the unregularized control simply by setting $\lambda = 0$ in Definition 4.3.4.

Definition 4.6.8. *For a control $u \in U_{feas}$ and a positive real number σ we define the σ-active sets for Problem $(\mathbb{P})$ by*

$$\begin{aligned}
M_{u,a}^\sigma(u) &:= \{(t,x) \in Q : \ u(t,x) \le u_a(t,x) + \sigma\}\\
M_{u,b}^\sigma(u) &:= \{(t,x) \in Q : \ u(t,x) \ge u_b(t,x) - \sigma\}\\
M_{y,a}^\sigma(u) &:= \{(t,x) \in Q : \ G(u)(t,x) \le y_a(t,x) + \sigma\}\\
M_{y,b}^\sigma(u) &:= \{(t,x) \in Q : \ G(u)(t,x) \ge y_b(t,x) - \sigma\}.
\end{aligned}$$

Lemma 4.6.9. *Let $\{\bar{u}_n\}$ be a sequence of locally optimal controls of $(\mathbb{P}_\lambda)$ converging strongly in $L^\infty(Q)$ to a locally optimal control $\bar{u}$ for $(\mathbb{P})$. Assume that there exists $\sigma_0 > 0$ such that the σ_0-active sets associated with $\bar{u}$ for the unregularized control problem according to Definition 4.6.8 are pairwise disjoint. Then there exists $\sigma > 0$ such that the σ-active sets for the Lavrentiev-regularized control problems according to Definition 4.3.4 are pairwise disjoint for all sufficiently small $\lambda > 0$.*

Proof. The proof is elementary. We state the main ideas for convenience of the reader. Consider for instance $(t,x) \in M_{u,a}^{\sigma_0/2}(\bar{u}_\lambda)$. Then we know

$$\bar{u}(t,x) - u_a(t,x) = \bar{u}_\lambda(t,x) - u_a(t,x) + \bar{u}(t,x) - \bar{u}_\lambda(t,x) \leq \frac{\sigma_0}{2} + \|\bar{u} - \bar{u}_\lambda\|_{\infty,\infty} \leq \sigma_0$$

for λ sufficiently small. This implies that $M_{u,a}^{\sigma_0/2}(\bar{u}_\lambda) \subset M_{u,a}^{\sigma_0}(\bar{u})$. Choosing $(t,x) \in M_{y,b}^{\sigma_0/2}(\bar{u}_\lambda)$ yields

$$\begin{aligned} \bar{y}(t,x) &= \lambda\bar{u}_\lambda(t,x) + \bar{y}_\lambda(t,x) + \bar{y}(t,x) - \bar{y}_\lambda(t,x) - \lambda\bar{u}_\lambda(t,x) \\ &\geq y_b(t,x) - \frac{\sigma_0}{2} - \|\bar{y} - \bar{y}_\lambda\|_{\infty,\infty} - \lambda\|\bar{u}_\lambda\|_{\infty,\infty} \\ &\geq y_b(t,x) - \sigma_0 \end{aligned}$$

for λ sufficiently small, since $\bar{y}_\lambda$ converges uniformly to $\bar{u}$ as λ tends to zero, and $\lambda\|\bar{u}_\lambda\|_{\infty,\infty}$ tends to zero because of the boundedness of U_{ad}. Then, we obtain $M_{y,b}^{\sigma_0/2}(\bar{u}_\lambda) \subset M_{y,b}^{\sigma_0}(\bar{u})$. The remaining sets are treated analogously. In conclusion, we obtain the $\sigma_0/2$-active sets of the regularized problem are pairwise disjoint for all sufficiently small $\lambda > 0$. $\square$

Unfortunately, we do not know a sufficient condition that guarantees the strong convergence of $\bar{u}_n$ to $\bar{u}$ in $L^\infty(Q)$.

5. A parabolic control problem with control constraints

5.1. Introduction

In the previous chapter we have demonstrated that parabolic state constrained optimal control problems can be handled well by regularization. The next step towards a complete analysis would be the discussion of a priori error estimates for a finite element discretization of such problems. In the overview in Chapter 1 we have already pointed out that the theory of a priori error estimates for parabolic control problems is far from complete. For state-constrained problems, the theoretical difficulties become obvious when pointing out that the adjoint state admits only the low regularity $\bar{p} \in L^{\tau}(I, W_0^{1,s}(\Omega))$ for some $\tau, s \in [1, 2)$, cf. also the introductory remarks in [107]. Then, standard arguments relying on interpolation error estimates are not applicable, which makes a discussion of error estimates for state-constrained parabolic problems in full generality difficult. One recent contribution to the analysis of linear-quadratic state-constrained problems with finitely many time-dependent controls has been made in [40]. For this setting, it was possible to prove uniform error estimates for the discrete states under natural regularity assumption. Error estimates for the optimal control problem were then proven with a technique that avoids the use of error estimates for the adjoint state, cf. also [74] for elliptic problems. Eventually, a rate of order $\mathcal{O}(\sqrt[4]{\ln h}|(h^{\frac{1}{2}} + k^{\frac{1}{4}})$ in two space dimensions and $\mathcal{O}(h^{\frac{1}{4}} + h^{-\frac{1}{4}}k^{\frac{1}{4}})$, in three space dimensions has been proven for the error in the $L^2(I, \mathbb{R}^m)$-norm, where h is the spatial and k is the temporal discretization parameter. Linear-quadratic problems with arbitrary control functions but state constraints that were imposed pointwise in time and averaged in space have been analyzed in [100]. In the analysis, error estimates for the state equation in the $L^\infty(I, L^2(\Omega))$ were needed and proven under regularity assumptions that can be expected for the considered problem class. Eventually, error estimates for the control in the L^2-norm of order $\mathcal{O}|\ln k|(k^{\frac{1}{2}} + h)$ have been obtained.

To our knowledge, there are no further contributions to the finite element error analysis for parabolic state-constrained optimal control problems. It is, however, known from elliptic problems that error estimates for certain Lavrentiev regularized problems, i.e. without additional control-constraints, can be obtained by transformation into a purely control constrained problem, cf. [73]. Yet, in the parabolic setting even the results for control-constrained nonconvex problems are not as complete as in the elliptic setting, where error estimates for problems with semilinear state equations have already been derived in [8]. We mention again the results by Malanowski, [92], and Meidner and Vexler [102] for linear quadratic problems and Chrysafinos, [33], for plain convergence results for problems with semilinear state equation, and refer back to the Introduction in Chapter 1 for an overview. The main goal of this chapter therefore is to provide a priori error estimates for the finite element discretization of a nonconvex control-constrained optimal control problem. As in the elliptic setting from [73], we will motivate how certain regularized state-constrained problems fit into this problem class as well. For that purpose, we will briefly discuss Lavrentiev regularization for purely state-constrained problems. The corresponding results are adapted to the nonconvex problems from the results for linear-quadratic setting that have been published by Tröltzsch and the author in [116].

The control-constrained problem is discretized by a discontinuous Galerkin scheme in time and a usual H^1-conforming finite element method in space. In [102] Meidner and Vexler analyzed the linear-quadratic counterpart of our model problem and proved error estimates that clearly separate the spatial and temporal discretization error for different types of control discretization. We demonstrate how these results extend to the nonlinear setting, where we have to deal with three challenges: first we have to take into account corresponding local solutions. We have already seen in the last two chapters how these can be handled. Moreover, a number of results on elliptic problems show how nonlinearities can be treated, cf. also the overview in Chapters 1 and 3, so that a combination of these arguments with the results from [102] seems promising. However, in the context of our model problem this involves a second challenge, since the quadratic growth condition derived from SSC dictates the spaces in which finite element error estimates have to be found. Third, we have to ensure that semidiscrete and discrete solutions to the nonlinear state equation (and adjoint equation and linearized equations) remain bounded in the $L^\infty(Q)$-norm independent of the discretization parameters.

Let us mention again that the results on error estimates to be provided here have already been published in [118] by Vexler and the author for a model problem with slightly different objective function. There, error estimates for cellwise constant control discretization have been discussed in detail and results for other types of discretization have been stated in a shorter form. In this thesis, we will discuss in detail the error for controls that are discretized piecewise constant in time and bilinear in space. This, and also the change in the objective function, are motivated by our interest in regularized state-constrained problems.

We point out that the notation in the papers [101, 102] and [118] differ essentially from the notation used in this thesis. In particular, in the three mentioned papers the control was denoted by q, the state by u, and the adjoint state by z, in contrast to our notation where u is the control, y is the state, and p is the adjoint state. Both notations are well-established, but this fact has to be kept in mind when following up on any given references.

5.2. Motivation by regularization

Before we turn to the analysis of error estimates for nonconvex control-constrained optimal control problems, let us demonstrate how certain state constrained problems fit into the context of this chapter. We will apply Lavrentiev regularization to a purely state-constrained problem, discuss convergence issues, and eventually transform the regularized problem into a control-constrained problem. The following results on regularization, especially the convergence result, have to the most part been discussed in more detail for linear quadratic problems in [116] by Tröltzsch and the author.

5.2.1. A control problem with pure pointwise state constraints

Relying on the same basic assumptions as in Chapter 4, suppose that the following purely state-constrained problem were given:

$$\text{Minimize } J(y,u) := \frac{1}{2}\int_0^T\int_\Omega \left((y-y_d)^2 + \nu u^2\right) dxdt \tag{5.2.1a}$$

subject to the PDE constraint

$$\begin{aligned} \partial_t y - \Delta y + d(t,x,y) &= u \quad \text{in } Q, \\ y(0,\cdot) &= y_0 \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Sigma. \end{aligned} \tag{5.2.1b}$$

as well as the pointwise state constraints

$$y_a \le y(t,x) \le y_b \quad \text{a.e. in } Q. \tag{5.2.1c}$$

Assumption 5.2.1. *Here, in accordance with Assumption 4.2.1, let $\nu \in \mathbb{R}$ be a positive, fixed parameter, and for simplicity let the bounds y_a, y_b be real numbers with $y_a < 0 < y_b$, $y_a \le y_0(x) \le y_b$. Moreover, let the desired state y_d be a function from $L^\infty(Q)$.*

The analysis of Problem (5.2.1) provides yet some more challenges than the problems considered in Chapter 4. In contrast to elliptic problems with mixed pointwise control-state constraints and no control bounds we will encounter some regularity issues that we will comment on. First of all, we notice that due to the missing control bounds there is no a priori reason to consider the controls in $L^\infty(Q)$. Due to the quadratic Tikhonov term

$$\iint_Q \nu u^2 \, dxdt$$

in the objective function an obvious candidate for the control space is $L^2(Q)$. Note, however, that for functions in $L^2(Q)$ and spatial domains of dimension larger than one, we only have the existence result of Theorem 2.5.5 under somewhat more severe restrictions on the nonlinearities, i.e. global boundedness and global Lipschitz continuity. Theorem 2.5.5 on page 25 provides solvability in useful spaces, but not continuity of the state – unless $n = 1$ and Theorem 2.5.6 is applicable. Consequently, the state constraints (5.2.1c) can only be formulated in the sense of $L^p(Q)$, and Slater-type arguments cannot be applied to obtain Lagrange multipliers unless we have better regularity properties of the optimal control. This cannot be shown a priori. In fact, from the objective function it can only be shown that a minimizing sequence of controls is bounded in $L^2(Q)$, and then the existence of optimal controls follows in $L^2(Q)$, only, given that the set of feasible controls in $L^2(Q)$ is not empty.

On the other hand, there exist many examples where the optimal control has higher regularity than just $L^2(Q)$ even without explicitely prescribed control bounds, or if present constraints simply do not become active. At least, numerous academic examples can easily be constructed. One such example is given in Section 5.9.5. If a control $u \in L^p(Q)$, $p \ge 2$ is - locally or globally - optimal with respect to all $L^2(Q)$ functions, it is a fortiori also optimal if the search space is restricted to

$$U := L^p(Q), \quad p > n/2 + 1.$$

Similar as before we obtain well-definedness of a control-to-state operator

$$G_p \colon L^p(Q) \to \mathcal{W}(0,T) \cap \mathcal{C}(\bar{Q}), \quad G_p(u) = y, \quad p > n/2 + 1,$$

by Theorem 2.5.5, as well as the reduced objective function

$$f_p \colon L^p(Q) \to \mathbb{R}, \quad f_p(u) := J(G_p(u), u) = \frac{1}{2} \iint_Q ((G_p(u) - y_d)^2 + \nu u^2) \, dxdt.$$

Here, the index p simply indicates that the control space is given by $L^p(Q)$. It is known, cf. [144], that G_p is of class $\mathcal{C}^2$, where the derivatives take the same form as in Proposition 4.2.3. We simply assume that a solution to the optimal control problem of the desired regularity exists.

Assumption 5.2.2. *We assume that the optimal control problem* (5.2.1) *admits at least one (globally) optimal control* $\bar{u} \in L^p(Q)$, $p > n/2 + 1$, *with associated optimal state* $\bar{y} = G_p(\bar{u})$.

Note that another limiting factor in Chapter 4 was the differentiability of the objective function. Now that we do not consider a general function $\psi(\cdot, y, u)$, but instead a somewhat simpler representative, we do not necessarily have to work with $L^\infty(Q)$. It is a well known fact, cf. also the remarks on differentiability in Chapter 2, that the objective function is differentiable in $L^2(Q) \times L^2(Q)$ into $\mathbb{R}$. Consequently, a discussion of first order optimality conditions along the lines of the last chapter is possible, but will bring little more insight. We refer to a presentation of results for linear-quadratic problems in [116]. Instead, we will use an alternative formulation of Problem (5.2.1), that will be favorable in the error estimates. Nevertheless, for the convergence analysis of regularized solutions, we rely on the following Slater-type condition.

Assumption 5.2.3. *Let* $\bar{u} \in L^p(Q)$, $p > n/2 + 1$ *denote a global solution of Problem* (5.2.1). *There exists a Slater point* $v_\gamma \in L^\infty(Q)$ *and a real number* $\gamma > 0$ *such that*

$$y_a + \gamma \leq G_p(\bar{u}) + G_p'(\bar{u})v_\gamma \leq y_b - \gamma.$$

Note the distinct difference in the roles of v_γ above and the Slater point u_γ in Assumption 4.2.7 on page 81.

5.2.2. Lavrentiev regularization and convergence

Reverting to L^2-controls and the stronger assumptions on the nonlinearity of Theorem 2.5.5, we consider a control to state operator G_2, i.e. G_p with $p = 2$, and obtain the reduced formulation

$$\text{Minimize } f_2(u) \quad \text{subject to } u \in L^2(Q), \quad y_a \leq \lambda u + G_2(u) \leq y_b \text{ a.e. in } Q. \qquad (\mathbb{P}_\lambda)$$

For fixed $\lambda > 0$, the existence of at least one global solution $\bar{u}_\lambda$ with associated optimal state $\bar{y}_\lambda = G(\bar{u}_\lambda)$ follows from standard arguments. Interestingly, the discussion of this problem without control constraints is in some aspects now significantly simpler than in the previous setting. It was already pointed out in [111] for elliptic problems that this class of Lavrentiev regularized problems can be transformed into purely control constrained model problems, whose discussion with respect to first and second-order optimality conditions is well-known. Before discussing this, though, we prove convergence for vanishing Lavrentiev parameters $\lambda \to 0$.

We have already seen twice how the convergence of local solutions can be discussed, and we will do so again when proving discretization error estimates. Therefore, let us focus here on the convergence of global solutions. We will adapt a proof from [69] for elliptic problems, that has been carried out and published for parabolic problems in [116]. In [152], the proof from [69] has been extended to a discussion of local solutions, and the same principal arguments could be applied here.

Let $\{\lambda_n\}$ be a sequence of positive real numbers converging to zero as $n \to \infty$, and let $\{\bar{u}_n\}$ denote a sequence of associated globally optimal solutions of $(\mathbb{P}_\lambda)$ with $\lambda = \lambda_n$. For now, we will assume, and later prove, that $\bar{u}_n \in L^\infty(Q)$, and hence also in $L^p(Q)$, $1 \leq p < \infty$ for all $n \in \mathbb{N}$. Then, continuity of $\bar{y}_n$ is guaranteed and we can write $\bar{y}_n = G_p(\bar{u}_n)$.

Lemma 5.2.4. *Let Assumptions 5.2.1-5.2.3 be satisfied. Then there exists* $\lambda_0 > 0$ *and a control* $u_0 \in L^2(Q)$ *such that* u_0 *is feasible for the regularized Problem* $(\mathbb{P}_\lambda)$ *for all* $\lambda \leq \lambda_0$.

Proof. Assumption 5.2.3 ensures that $\bar{y} = G_p(\bar{u})$ is a continuous function. Further, $\mathcal{C}(\bar{Q})$ is dense in $L^p(Q)$, hence there exists a sequence $\{u_k\}$ in $\mathcal{C}(\bar{Q})$ with

$$\|u_k - \bar{u}\|_{L^p(Q)} \leq \frac{1}{k},$$

i.e. $\{u_k\}$ converges strongly towards $\bar{u}$ in $L^p(Q)$. Further, there exists $c > 0$ such that

$$\|G_p'(\bar{u})(u_k - \bar{u})\|_{\mathcal{C}(\bar{Q})} \leq c\,\|u_k - \bar{u}\|_{L^p(Q)} \leq \frac{c}{k},$$

for $p > n/2 + 1$. With the help of the Slater point v_γ, we define a sequence $\{u_k^\gamma\}$ in $L^\infty(Q)$ by

$$u_k^\gamma := u_k + \frac{3c}{\gamma k} v_\gamma,$$

and obtain

$$\|u_k^\gamma - \bar{u}\|_{L^p(Q)} \leq \|u_k - \bar{u}\|_{L^p(Q)} + \frac{3c}{\gamma k}\|v_\gamma\|_{\infty,\infty} \leq \frac{1}{k}\left(1 + \frac{3c}{\gamma}\|v_\gamma\|_{\infty,\infty}\right).$$

Passing to the limit yields

$$\lim_{k\to 0} \|u_k^\gamma - \bar{u}\|_{L^p(Q)} = 0.$$

It is now easy to prove that, for k large enough, there exists $n_k \in \mathbb{N}$ such that u_k^γ is feasible for all $(\mathbb{P}_\lambda)$ with $n \geq n_k$. With $\tilde{c} := \frac{3c}{\gamma}$ and k large enough such that $\frac{1}{k} \leq \min\{1, \frac{1}{\tilde{c}}\}$ we find

$$\begin{aligned}
\lambda_n u_k^\gamma + G_p(u_k^\gamma) &= \lambda_n u_k^\gamma + G_p(\bar{u}) + G_p'(\bar{u})(u_k^\gamma - \bar{u}) + o(\|u_k^\gamma - \bar{u}\|_{L^p(Q)}) \\
&= \lambda_n u_k^\gamma + \frac{\tilde{c}}{k}(G_p(\bar{u}) + G_p'(\bar{u})(v_\gamma)) + \left(1 - \frac{\tilde{c}}{k}\right) G_p(\bar{u}) + G_p'(\bar{u})(u_k - \bar{u}) + o(\|u_k^\gamma - \bar{u}\|_{L^p(Q)}) \\
&\leq \lambda_n \|u_k^\gamma\|_{\infty,\infty} + \frac{\tilde{c}}{k}(y_b - \gamma) + \left(1 - \frac{\tilde{c}}{k}\right) y_b + \frac{c}{k} + o\left(\frac{1}{k}\right)
\end{aligned}$$

Note that k can be chosen large enough that $o(\frac{1}{k}) \leq \frac{c}{k}$, and we obtain

$$\lambda_n u_k^\gamma + G_p(u_k^\gamma) \leq y_b - 2\frac{c}{k}\gamma + o\left(\frac{1}{k}\right) + \lambda_n\|u_k^\gamma\|_{\infty,\infty} \leq y_b - \frac{c}{k} + \lambda_n \|u_k^\gamma\|_{\infty,\infty}.$$

Obviously, for λ_n sufficiently small, we have

$$\lambda_n u_k^\gamma + G(u_k^\gamma) \leq y_b.$$

The lower bound can be treated analogously, and setting $\lambda_0 = \lambda_{n_k}$ as well as $u_0 := u_k^\gamma$ for k large enough yields the assertion. □

Corollary 5.2.5. *The sequence $\{\bar{u}_n\}$ of globally optimal controls to $(\mathbb{P}_\lambda)$ is bounded in $L^2(Q)$ independently of λ.*

Proof. By the feasibility of u_0 and the optimality of $\bar{u}_n$, we know that

$$\frac{\nu}{2}\|\bar{u}_n\|_I^2 \leq f(\bar{u}_n) \leq f(u_0)$$

for all n sufficiently large, which implies the assertion. □

Now we obtain the existence of a subsequence $\bar{u}_{n_k}$ converging weakly to u^* in $L^2(Q)$.

Lemma 5.2.6. *Let $\{\bar{u}_{n_k}\}$ be the weakly converging subsequence of the sequence of globally optimal solutions $\{\bar{u}_n\}$ of $(\mathbb{P}_\lambda)$. The weak limit $u^* \in L^2(Q)$ of $\{\bar{u}_{n_k}\}$ is feasible for Problem* (5.2.1).

Proof. Since $\{\bar{u}_{n_k}\}$ is uniformly bounded in $L^2(Q)$ and $\{\lambda_n\}$ tends to zero, we can assume w.l.o.g. that $\|\lambda_n \bar{u}_{n_k}\| \to 0$ and therefore $\lambda_n \bar{u}_{n_k} \to 0$ pointwisely in Q, as $n \to \infty$. Moreover, from the feasibility of $\bar{u}_n$ for $(\mathbb{P}_\lambda)$, we obtain

$$y_a \leq \lambda \bar{u}_n + G_p(\bar{u}_n) \leq y_b. \tag{5.2.2}$$

From the compact embedding of $\mathcal{W}(0,T) \hookrightarrow L^2(Q)$ we know that the optimal regularized states $\bar{y}_n := G_p(\bar{u}_n)$ converge strongly towards $y^* := G_p(u^*)$ in $L^2(Q)$. Then, taking the limit in (5.2.2) yields the assertion, making use of the fact that $[y_a, y_b] \subset L^2(Q)$ is closed. □

Theorem 5.2.7. *Under Assumption 5.2.1-5.2.3, the sequence $\{\bar{u}_{n_k}\}$ converges strongly in $L^2(Q)$ towards u^*, and u^* is a global solution of* (5.2.1).

Proof. Since $\bar{u}_{n_k}$ is globally optimal for $(\mathbb{P}_\lambda)$, we have

$$f(\bar{u}_{n_k}) \leq f(u_k^\gamma)$$

for k large enough. Here, u_k^γ denotes the auxiliary control constructed in Lemma 5.2.4. By the lower semi-continuity of f and the weak convergence of $\bar{u}_{n_k}$ towards u^* we obtain for each k

$$f(u^*) \leq \liminf_{n\to\infty} \ f(\bar{u}_{n_k}) \leq \limsup_{n\to\infty} \ f(\bar{u}_{n_k}) \leq f(u_k^\gamma).$$

For $k \to \infty$ we arrive at

$$f(u^*) \leq \lim_{k\to\infty} f(u_k^\gamma) = f(\bar{u}) \leq f(u^*),$$

hence $f(u^*) = f(\bar{u})$, which implies optimality of u^* for (5.2.1). It remains to show that the convergence is strong in $L^2(Q)$. From the above, we know that

$$\lim_{n\to\infty} f(\bar{u}_{n_k}) = f(u^*), \tag{5.2.3}$$

and from the compactness of the embedding $\mathcal{W}(0,T) \hookrightarrow L^2(Q)$ we further deduce that

$$\lim_{n\to\infty} \|\bar{y}_{n_k} - y_d\|_I^2 = \|y^* - y_d\|_I^2, \tag{5.2.4}$$

where $\bar{y}_{n_k}$ and y^* denote the states associated with $\bar{u}_{n_k}$ and u^*, respectively. Conditions (5.2.3) and (5.2.4) imply that

$$\lim_{n\to\infty} \|\bar{u}_{n_k}\|_I^2 = \|u^*\|_I^2.$$

That, together with the weak convergence, implies strong convergence. □

5.2.3. Transformation into a control constrained problem

Following [111], we now demonstrate how Problem $(\mathbb{P}_\lambda)$ can be transformed into a control constrained problem. For a fixed $\lambda > 0$, we introduce the new control

$$w := \lambda u + y$$

and consider u, y, and w as elements of $L^2(Q)$. Introducing the set of admissible controls

$$W_{\mathrm{ad}} := \{w \in L^2(Q) \colon y_a \le w \le y_b\},$$

we immediately obtain that W_{ad} is a subspace of $W := L^\infty(Q)$ and for the regularized problem formulation with fixed $\lambda > 0$ the local Lipschitz continuity and boundedness conditions of Assumption 4.2.1 are sufficient for all following results. Then, noting that $u = \frac{1}{\lambda}(w - y)$, we express Problem $(\mathbb{P}_\lambda)$ as

$$\text{Minimize } J_w(y,w) := \frac{1}{2} \iint_Q (y - y_d)^2 + \frac{\nu}{\lambda^2}(w - y)^2 \, dx dt \tag{5.2.5a}$$

subject to the PDE constraint

$$\begin{aligned} \partial_t y - \Delta y + d(\cdot, y) + \tfrac{1}{\lambda} y &= \tfrac{1}{\lambda} w && \text{in } (0,T) \times \Omega, \\ y(0,\cdot) &= y_0 && \text{in } \Omega, \\ y &= 0 && \text{on } \Sigma, \end{aligned} \tag{5.2.5b}$$

as well as the control bounds

$$y_a \le w \le y_b \quad \text{in } Q. \tag{5.2.5c}$$

Notice that not only the PDE but also the form of the objective function has now changed. Yet, it remains differentiable from $L^2(Q) \times L^2(Q) \to \mathbb{R}$. We have thus transformed a formerly state-constrained control problem via regularization into a purely control constrained problem. Along the lines of [111], it is now possible to show the existence of regular Lagrange multipliers also for the original problem formulation $(\mathbb{P}_\lambda)$. Also, second order sufficient conditions would be much easier to discuss. However, the main goal of this chapter will be the development of a priori error estimates for a class of control-constrained parabolic problems that includes the regularized, transformed problem (5.2.5). Such a transformation into a control constrained problem has already been used by Hinze and Meyer, cf. [73], to present a priori error estimates for the variational discretization of elliptic control problems with mixed pointwise state constraints.

Let us conclude this section with a short remark on the regularity assumptions for the optimal solution. Clearly, for each fixed $\lambda > 0$ the optimal control $\bar{w}$ is bounded in $L^\infty(Q)$, which in turn implies that $\bar{u}_\lambda$ is bounded in $L^\infty(Q)$, since Theorem 2.5.6 guarantees boundedness of $\bar{y}_\lambda$ in $L^\infty(Q)$. Yet, we have no boundedness result for $\bar{u}_\lambda$ that is independent of λ. If the optimal control $\bar{u}$ is not bounded, it is likely that for λ tending to zero the L^∞-norm of $\bar{u}_\lambda$ increases, and we cannot disprove this even if $\bar{u}$ is bounded. More precisely, if control constraints u_a and u_b are present in the original unregularized problem formulation that simply do not become active, we cannot prove that $\bar{u}_\lambda$ remains within these bounds, since we have no convergence result in $L^\infty(Q)$. Consequently, it is not a priori clear whether or not a transformation into a control constrained problem is meaningful after omitting inactive control constraints . Nevertheless, we can prove a finite element error estimate for $\bar{w}$ for fixed $\lambda > 0$, and we leave it up to numerical computations to reveal whether or not $\bar{u}_\lambda$ actually remains within reasonable bounds. Moreover, the results presented next are meaningful also outside the context of regularization.

5.3. Problem formulation and analysis

We now consider a general control-constrained model problem that also includes the regularized, transformed Problem (5.2.5).

$$\text{Minimize } J(y,u) := \frac{1}{2} \iint_Q [(\psi_1 y^2 + 2\psi_2 y u + 2\psi_3 y + \nu u^2)] \, dx dt \tag{5.3.1a}$$

subject to the PDE constraint

$$\begin{aligned} \partial_t y - \Delta y + d(\cdot, y) &= u && \text{in } Q \\ y(0,\cdot) &= y_0 && \text{in } \Omega, \\ y &= 0 && \text{on } \Sigma, \end{aligned} \tag{5.3.1b}$$

as well as the control bounds

$$u_a \le u \le u_b \quad \text{in } Q. \tag{5.3.1c}$$

The terminology in the objective function is chosen similar to the linear-quadratic problem in the last chapter. We rely on the following assumption:

Assumption 5.3.1. *Let the conditions of Assumption 4.2.1 be satisfied. For simplicity, we consider real numbers ψ_1, ψ_2, ψ_3 and $\nu > 0$, as well as control bounds $u_a, u_b \in \mathbb{R}$ with $u_a < u_b$. In view of the error analysis, we will restrict the discussion to two-dimensional spatial domains, i.e. we assume $n = 2$.*

Remark 5.3.2. *All arguments in the sequel can easily be adapted to a setting with $\Psi_3 = y_d$, where y_d is a sufficiently regular desired state as in Problem $(\mathbb{P}_\lambda)$. This has been considered in [118] for a tracking type objective function, i.e. $\Psi_1 = 1$ and $\Psi_2 = 0$.*

We are eventually interested in proving a priori error estimates in the $L^2(Q)$-norm for the error between a locally optimal control $\bar{u}$ of the model problem, and a corresponding completely time-and-space-discrete locally optimal control $\bar{u}_\sigma$, and with regularity assumptions on the right-hand side that can be fulfilled for the considered optimal control problems.

Since the analysis of control-constrained model problems such as (5.3.1) is well-established, let us summarize some important results in a more concise way than we have done for the state constrained problems in Chapters 3 and 4. For simple presentation of the upcoming results we agree that

$$U := L^\infty(Q), \quad Y := \mathcal{W}(0,T),$$

as well as

$$U_{\text{ad}} := \{u \in U \colon u_a \le u \le u_b \text{ a.e. in } Q\}.$$

Moreover, we mention for convenience that with the choice of the Laplace-operator in the PDE a weak solution of the state equation is given by y fulfilling

$$\begin{aligned} \int_0^T \langle \partial_t y, \varphi \rangle_{V^*,V}\, dt + (\nabla y, \nabla \varphi)_I + (d(\cdot,y),\varphi)_I &= (u,\varphi)_I \quad \forall \varphi \in L^2(I,V), \\ y(0,\cdot) &= y_0 \qquad \text{in } \Omega. \end{aligned} \tag{5.3.2}$$

Thus, all results shown for parabolic PDEs, in particular Theorem 2.5.6 on page 26, remain valid. Let us point out that from now on we will not make use of any continuity results for state or adjoint equations, since this is not necessary for the discussion of Problem (5.3.1a) with control bounds, only. With the control-to-state mapping that we still denote by G,

$$G \colon U \to Y \cap L^\infty(Q), \quad G(u) = y,$$

we obtain the reduced objective function

$$f : U \to \mathbb{R}, \quad u \mapsto J(u, G(u))$$

and the reduced problem formulation

$$\text{Minimize } f(u) \text{ subject to } u \in U_{\text{ad}}. \tag{$\mathbb{P}$}$$

By standard methods, we obtain the existence of at least one global solution.

Theorem 5.3.3. *Let Assumption 5.3.1 be satisfied. Then Problem* $(\mathbb{P})$ *admits at least one optimal control* $\bar{u}$ *with associated state* $\bar{y} = G(\bar{u})$.

Proof. This follows from [144, Theorem 5.7]. In particular, we point out that $U_{\rm ad}$ is nonempty by assumption, and f is convex with respect to u since u only appears linearly in the term $\psi_2 yu$, and quadratically in $\nu u^2$with positive parameter ν. $\square$

We are now mainly interested in local solutions in the sense of $L^2(Q)$.

Definition 5.3.4. *A control* $\bar{u} \in U_{ad}$ *is called a local solution of* $(\mathbb{P})$ *in the sense of* $L^2(Q)$ *if there exists a constant* $\varepsilon > 0$, *such that the inequality*

$$f(u) \geq f(\bar{u})$$

is satisfied for all $u \in U_{ad}$ *with* $\|\bar{u} - u\|_I \leq \varepsilon$. *We will refer to the associated optimal state as* $\bar{y}$.

The treatment of first- and second-order optimality conditions is now very much simplified by the fact that no pointwise state constraints are given. The analysis is standard procedure and can be found in the monograph [144]. We will only collect what we need for our future analysis. In particular, we will prove a more general second order sufficient condition not restricted to L^∞-neighborhoods. We have already argued that the objective function J is Fréchet differentiable in $L^2(Q) \times L^2(Q)$, and we know that the control-to-state operator G is twice differentiable with respect to the $L^p(Q)$-norm, with $p > n/2 + 1$. In our two-dimensional setting this implies $p > 2$. By the chain rule, we obtain the following differentiability result for the reduced objective function f:

Lemma 5.3.5. *The reduced objective function* f *is also of class* $\mathcal{C}^2$. *For arbitrary* $u, v, v_1, v_2 \in L^\infty(Q)$ *with associated* $y = G(u)$, $\tilde{y} = G'(u)v$, $\tilde{y}_i = G'(u)v_i$, $i = 1, 2$, *as well as* $z := G''(u)[v_1, v_2]$, *its first and second order derivative are given by*

$$f'(u)v = \iint_Q (\psi_1 \tilde{y} y + \psi_2 \tilde{y} u + \psi_2 y v + \psi_3 \tilde{y} + \nu u v)\, dx dt$$

$$f''(u)[v_1, v_2] = \iint_Q (\psi_1 \tilde{y}_1 \tilde{y}_2 + \psi_1 z y + \psi_2 \tilde{y}_2 v_2 + \psi_2 z u + \psi_2 \tilde{y}_2 v_1 + \psi_3 z + \nu v_1 v_2)\, dx\, dt.$$

Proof. The $\mathcal{C}^2$ regularity follows from known results for Nemytskii operators, cf. Chapter 2 and the chain rule. The form of the derivatives follow by straight-forward calculations, we refer also to, e. g., [144] for details. $\square$

Now, we can formulate standard first order necessary optimality conditions with the help of a variational inequality.

Lemma 5.3.6. *Let* $\bar{u} \in U_{ad}$ *be a local solution of* $(\mathbb{P})$ *in the sense of Definition 5.3.4. Then the following variational inequality holds:*

$$f'(\bar{u})(u - \bar{u}) \geq 0 \quad \forall u \in U_{ad}. \tag{5.3.3}$$

Proof. This is a standard result in optimal control for convex admissible sets. For a proof, we refer again to [144]. $\square$

The adjoint state p for a control $u \in U$ with associated state $y = G(u) \in Y$ is defined as the solution of the adjoint equation

$$\begin{aligned} -\partial_t p - \Delta p + \partial_y d(\cdot, y)p &= \psi_1 y + \psi_2 u + \psi_3 && \text{in Q}, \\ p(T,\cdot) &= 0 && \text{in } \Omega, \\ p &= 0 && \text{on } \Sigma, \end{aligned} \tag{5.3.4}$$

whose weak formulation is given by

$$\begin{aligned} -\int_0^T \langle \partial_t p, \varphi\rangle_{V^*,V} + (\nabla\varphi, \nabla p)_I + (\varphi, \partial_y d(\cdot,y)p)_I &= (\varphi, \psi_1 y + \psi_2 u + \psi_3)_I \quad \forall \varphi \in L^2(I,V) \\ p(T,\cdot) &= 0. \end{aligned} \tag{5.3.5}$$

Note that no measures appear in (5.3.5); it is of the form (2.5.5) and therefore the existence and regularity results of Proposition 2.5.7 are applicable, yielding

$$p \in L^2(I, H^2(\Omega \cap H_0^1(\Omega))) \cap H^1(I, L^2(\Omega)) \cap L^\infty(Q) \cap \mathcal{C}(\bar{I}, H_0^1(\Omega)),$$

cf. also the summarized regularity results in Proposition 5.3.7 below. We denote by $\bar{p}$ the adjoint state associated with the locally optimal control $\bar{u}$ and its corresponding optimal state $\bar{y}$. Then, by standard calculations, cf. again [144], we can write the first order optimality conditions from Proposition 5.3.6 in the form

$$(\nu\bar{u} + \psi_2\bar{y} + \bar{p}, u - \bar{u})_I \geq 0 \quad \forall u \in U_{\text{ad}}.$$

We have seen in the previous chapters how the variational inequality can be transformed into a gradient equation with the help of Lagrange multipliers for the control constraints. Moreover, we know that a pointwise interpretation of the variational inequality yields a projection formula of $\bar{u}$ onto the set of admissible controls, cf. the discussion in [144]. Since in all what follows we need to characterize the regularity of the optimal control in a more precise way than just an L^p-setting, we formally define the pointwise projection on the admissible set by

$$P_{U_{\text{ad}}}\colon L^2(I, L^2(\Omega)) \to U_{\text{ad}}, \quad P_{U_{\text{ad}}}(r)(t,x) = \max(u_a, \in (u_b, r(t,x))).$$

We then obtain the representation

$$\bar{u} = P_{U_{\text{ad}}}\left(-\frac{1}{\nu}\bar{p} - \frac{\psi_2}{\nu}\bar{y}\right), \tag{5.3.6}$$

cf. again [144]. As in the linear quadratic setting of [102], we make use of the following regularity properties of $P_{U_{\text{ad}}}$:

$$\|\nabla(P_{U_{\text{ad}}}(v)(t))\|_{L^s(\Omega)} \leq \|\nabla v(t)\|_{L^s(\Omega)} \;\; \forall v \in L^2(I, W^{1,s}(\Omega)),\; 1 \leq s \leq \infty, \tag{5.3.7}$$

for almost all $t \in I$. For reference, we point out [80, Corollary A.6], which ensures for a projection $\tilde{P}_{[u_a,u_b]}\colon L^2(\Omega) \to L^2(\Omega)$ that the space $W^{1,s}(\Omega)$ is mapped into itself for all $1 \leq s \leq \infty$ and $\|\nabla\tilde{P}_{[u_a,u_b]}v\|_{L^s(\Omega)} \leq \|\nabla v\|_{L^2(\Omega)}$ for all $v \in W^{1,s}(\Omega)$. Applying this argument pointwise yields (5.3.7).

Proposition 5.3.7. *Let $\bar{u}$ be a local solution of the optimization problem $(\mathbb{P})$, and $\bar{p}$ denote the corresponding adjoint state. Then*

$$\begin{gathered} \bar{y}, \bar{p} \in L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(I, L^2(\Omega)) \cap L^\infty(Q) \cap \mathcal{C}(\bar{I}, H_0^1(\Omega)), \\ \bar{u} \in L^2(I, W^{1,s}(\Omega)) \cap H^1(I, L^2(\Omega)) \cap L^\infty(Q) \end{gathered}$$

holds for any $s < \infty$.

Proof. The regularity of the state and adjoint state is a consequence of Theorem 2.5.6 and Proposition 2.5.7 due to the boundedness of U_{ad}. Then, the regularity of $\bar{u}$ follows from the projection formula (5.3.6) and the properties of the projection operator. □

Remark 5.3.8. *Note that the set of admissible controls U_{ad} is bounded in $L^\infty(Q)$ by $|u_a| + |u_b|$. Then, Theorem 2.5.6 on page 26 guarantees boundedness of all states $y = G(u)$ in the $L^\infty(Q)$ norm independent from $u \in U_{ad}$. Consequently, the (local) Lipschitz continuity of e.g. d and its derivatives can be used with constants C independent of y in all proofs in the sequel, since we will only consider controls $u \in U_{ad}$. Similar arguments apply to linearized and adjoint equations, using the boundedness results from Propositions 2.5.2 and 2.5.7.*

We proceed with the discussion of second order sufficient conditions.

Assumption 5.3.9. *Let $\bar{u} \in U_{ad}$ fulfill the first-order necessary optimality conditions* (5.3.3). *We assume that there exists a constant $\alpha > 0$ such that*

$$f''(\bar{u})[v,v] \geq \alpha \|v\|_I^2 \quad \forall\, v \in L^\infty(Q).$$

Notice again the difference to a setting with pointwise state constraints, where the second derivative of the Lagrangian had to be considered. For this problem, second order sufficient conditions are well understood. We refer to the exposition in [144] for a more general objective function, where optimality of $\bar{u}$ in the sense of $L^\infty(Q)$ has been deduced from such a condition. In view of the finite element error estimates, it will be convenient to avoid the use of the space $L^\infty(Q)$ in order to avoid the need for discretization error estimates in the L^∞-norm. We can make use of the special structure of the objective function and prove that the second derivative of f is even Lipschitz continuous in $L^2(Q)$. Eventually, this will allow to prove a quadratic growth condition in an L^2-neighborhood. Before, let us mention that $f''[v,v]$ can be represented by

$$f''(u)[v,v] = \iint\limits_Q \left(\psi_1 \tilde{y}^2 + 2\psi_2 \tilde{y} v + \nu v^2 - p\partial_{yy} d(\cdot, y)\tilde{y}^2\right)\, dxdt, \tag{5.3.8}$$

where p fulfills (5.3.5) and $\tilde{y} := G'(u)v$. To prove Lipschitz continuity of f'' we will rely on the already proven Lipschitz results for the state and linearized state from Lemma 4.2.4 on page 79. For the adjoint state, we also need a Lipschitz stability result in the L^2-norm, which is stated next.

Lemma 5.3.10. *Let $u_1, u_2 \in U_{ad}$ be given, and let p_1, p_2 denote the associated adjoint states fulfilling* (5.3.5). *Then there exists a constant $C > 0$ such that the estimate*

$$\|p_1 - p_2\|_I \leq C\,\|u_1 - u_2\|_I \tag{5.3.9}$$

is fulfilled.

Proof. We set $y_i = G(u_i)$ and consider the difference $p := p_1 - p_2$, which fulfills the equation

$$\begin{aligned} -\partial_t p + \mathcal{A}p + \partial_y d(\cdot, y_1)p &= ((\partial_y d(\cdot, y_2) - \partial_y d(\cdot, y_1))p_2 + \psi_1(y_1 - y_2) && \text{in } Q,\\ p(T,\cdot) &= 0 && \text{in } \Omega,\\ p &= 0 && \text{on } \Sigma. \end{aligned}$$

The stability result for adjoint equations from Proposition 2.5.7, the Lipschitz properties of y due to Lemma 4.2.4, the Lipschitz continuity of $\partial_y d$ guarantee

$$\|p\|_I \leq c(\|(\partial_y d(\cdot, y_2) - \partial_y d(\cdot, y_1))p_2\|_I + \|y_1 - y_2\|_I) \leq c(\|p_2\|_{\infty,\infty} + 1)\|y_1 - y_2\|_I \leq c\,\|u_1 - u_2\|_I,$$

since U_{ad} is bounded. □

We can now deduce the desired Lipschitz result for f''. Since this is one of the central arguments in the following convergence proofs, we present the proof, though obtained by standard methods, in detail.

Lemma 5.3.11. *There exists a constant C, such that for all $u_1, u_2 \in U_{ad}$ and all $v \in L^\infty(Q)$*

$$|f''(u_1)[v,v] - f''(u_2)[v,v]| \le C\,\|u_1 - u_2\|_I \|v\|_I^2$$

is satisfied.

Proof. Consider the auxiliary functions

$$y_i := G(u_i), \quad \tilde{y}_i := G'(u_i)v,$$

and let p_i be the adjoint state associated with u_i. Notice again that y_1, y_2 are uniformly bounded in $L^\infty(Q)$ due to the boundedness of U_{ad}. Direct calculations show that

$$\begin{aligned}
|f''(u_1)[v,v] - f''(u_2)[v,v]| \le \; & \iint\limits_Q \left|p_2 \partial_{yy} d(\cdot, y_2)\tilde{y}_2^2 - p_1 \partial_{yy} d(\cdot, y_1)\tilde{y}_1^2\right| \,dxdt \\
& + \iint\limits_Q \left|\psi_1(\tilde{y}_1^2 - \tilde{y}_2^2) + 2\psi_2 v(\tilde{y}_1 - \tilde{y}_2)\right| dxdt \\
\le \; & \iint\limits_Q |(\psi_1(\tilde{y}_1 + \tilde{y}_2) + 2\psi_2 v)(\tilde{y}_1 - \tilde{y}_2)| \,dxdt \\
& + \iint\limits_Q \left|(p_2 - p_1)\partial_{yy} d(\cdot, y_2)\tilde{y}_2^2\right| \,dxdt \\
& + \iint\limits_Q |p_1 \partial_{yy} d(\cdot, y_2)(\tilde{y}_2 - \tilde{y}_1)(\tilde{y}_1 + \tilde{y}_2)| \,dxdt \\
& + \iint\limits_Q \left|p_1(\partial_{yy} d(\cdot, y_2) - \partial_{yy} d(\cdot, y_1))\tilde{y}_1^2\right| \,dxdt.
\end{aligned}$$

Estimating the integrals by Hölder's inequality yields

$$\begin{aligned}
|f''(u_1)[v,v] - f''(u_2)[v,v]| \le \; & c(\|\tilde{y}_1\|_I + \|\tilde{y}_2\|_I + \|v\|_I)\|\tilde{y}_1 - \tilde{y}_2\|_I \\
& + c\,\|\partial_{yy} d(\cdot, y_1)\|_{\infty,\infty}\|p_1 - p_2\|_I \|\tilde{y}_2\|^2_{L^4(Q))} \\
& + c\,\|p_1\|_{\infty,\infty}\|\partial_{yy} d(\cdot, y_2)\|_{\infty,\infty}(\|\tilde{y}_1\|_I + \|\tilde{y}_2\|_I)\|\tilde{y}_1 - \tilde{y}_2\|_I \\
& + c\|p_1\|_{\infty,\infty}\|y_1 - y_2\|_I \|\tilde{y}_1\|^2_{L^4(Q)}.
\end{aligned}$$

In the last line, we have used the Lipschitz continuity of $\partial_{yy} d$. By the embedding

$$L^\infty(I, H_0^1(\Omega)) \hookrightarrow L^4(Q)$$

and Proposition 2.5.2 we obtain

$$\|\tilde{y}_i\|_{L^4(Q)} \le c\,\|v\|_I, \quad i = 1, 2.$$

Then, the boundedness of U_{ad} and $\partial_{yy} d$, Proposition 2.5.2 as well as the Lipschitz estimates in Lemma 4.2.4 and 5.3.10 yield the assertion. ☐

These results allow to formulate a quadratic growth condition in an L^2-neighborhood of a local solution $\bar{u}$.

Theorem 5.3.12. *Let Assumption 4.2.1 hold and let additionally $\bar{u} \in U_{ad}$ fulfill Assumption 5.3.9. Then there exist constants $\varepsilon, \beta > 0$ such that the quadratic growth condition*

$$f(u) \geq f(\bar{u}) + \beta \|u - \bar{u}\|_I^2$$

is satisfied for all $u \in U_{ad}$ with $\|u - \bar{u}\|_I \leq \varepsilon$.

Proof. We proceed by Taylor expansion of f in $\bar{u}$. For $u \in U_{\mathrm{ad}}$ we obtain

$$f(u) = f(\bar{u}) + f'(\bar{u})(u - \bar{u}) + \frac{1}{2} f''(u^\xi)[u - \bar{u}, u - \bar{u}]$$

with $u^\xi = \bar{u} + \xi(u - \bar{u})$ for a $\xi \in (0,1)$ due to the differentiability of G in $L^\infty(Q)$. With the variational inequality (5.3.3), Assumption 5.3.9, as well as Lemma 5.3.11 we obtain

$$\begin{aligned} f(u) =& f(\bar{u}) + f'(\bar{u})(u - \bar{u}) + \frac{1}{2} f''(\bar{u})[u - \bar{u}, u - \bar{u}] + \frac{1}{2}(f''(u^\xi) - f''(\bar{u}))[u - \bar{u}, u - \bar{u}] \\ \geq& f(\bar{u}) + \frac{\alpha}{2}\|u - \bar{u}\|_I^2 - c\,\|u^\xi - \bar{u}\|_I \|u - \bar{u}\|_I^2 \end{aligned}$$

The assertion follows noting that $\|u^\xi - \bar{u}\|_I = \xi \|u - \bar{u}\|_I$. □

This theorem implies that a local analysis of discretized problems is useful in L^2-neighborhoods. Consequently, associated error estimates have to be supplied primarily in the L^2-norm. Following [8] or [26], we will make use of coercivity properties of the second derivative of the objective function on the different levels of discretization. For future reference, let us therefore prove that the coercivity of f'' transfers to an L^2-neighborhood of a locally optimal control.

Corollary 5.3.13. *Let $\bar{u}$ satisfy Assumption 5.3.9. There exists $\varepsilon > 0$ such that*

$$f''(u)[v, v] \geq \frac{\alpha}{2}\|v\|_I^2$$

for all $v \in L^\infty(Q)$ and all $u \in U_{ad}$ with $\|u - \bar{u}\|_I \leq \varepsilon$.

Proof. This follows from Assumption 5.3.9 and Lemma 5.3.11 noting

$$f''(u)[v,v] = f''(\bar{u})[v,v] + (f''(u)[v,v] - f''(\bar{u})[v,v]) \geq \alpha\|v\|_I^2 - c\,\|u - \bar{u}\|_I \|v\|_I^2.$$

If $\varepsilon > 0$ is chosen sufficiently small, the assertion is obtained. □

5.4. Semi-discretization of the state equation in time

In this section, we begin our discussion of discretized versions of $(\mathbb{P})$. We first focus on a semi-discretization in time of the state equation (5.3.2). In Section 5.5, we will then provide results on the spatial discretization. Only in Section 5.6 will we start taking into account different aspects of control discretization, so that we can provide error estimates for a completely discretized problem. Our estimates in time and space are derived on the basis of separate Galerkin finite element methods along the lines of [102]. This allows to derive first- and second order optimality conditions in a way very close to the problem formulation on the continuous level.

5.4.1. The discontinuous Galerkin method

We present the discretization of the state equation in time by a discontinuous Galerkin scheme of degree zero. This method is explained in detail for instance in [44]. We will present the main ideas, keeping as close to the notation in [102] or [118] as possible. However, we point out again that in this thesis we use the notation y, u, p for the state, control and adjoint state, as opposed to u, q, z.

Along the lines of [101, 102] and [118], we consider a discretization of the time interval $\bar{I} = [0, T]$ with time points

$$0 = t_0 < t_1 < \ldots < t_{M-1} < t_M = T.$$

The interval $\bar{I}$ is then divided into subintervals $I_m = (t_{m-1}, t_m]$ of length $k_m := t_m - t_{m-1}$ such that

$$\bar{I} = \{0\} \cup I_1 \cup I_2 \cup \ldots \cup I_M.$$

The discretization parameter $k\colon I \to \mathbb{R}$ is defined as a piecewise constant function by setting $k_{|I_m} = k_m$ for $m = 1, 2, \ldots, M$. We will refer to k also as the maximal size of the time steps, i. e.,

$$k = \max k_m,$$

which is also called the diameter of the partitioning. We denote by $\mathcal{P}_0(I_m, V)$ the space of polynomials of degree zero defined on I_m with values in $V = H_0^1(\Omega)$, and define the semidiscrete trial and test space

$$Y_k^0 = \{v_k \in L^2(I, V) | v_{k|I_m} \in \mathcal{P}_0(I_m, V),\ m = 1, 2, \ldots, M\}.$$

Note that

$$Y_k^0 \not\subset Y,$$

since $Y = \mathcal{W}(0, T)$, but functions from Y_k^0 are not continuous. The control space $U = L^\infty(Q)$ and the set of admissible controls U_{ad} remains unchanged, since no control discretization is considered at this point. We will refer to this discretization as dG(0) scheme. Note that higher order Galerkin methods can be defined utilizing higher order polynomials $\mathcal{P}_r(I_m, V)$. However, the constant polynomials already fit well to the regularity results that we can make use of.

For functions $v_k \in Y_k^0$ we define a limit "from above", $v_{k,m}^+$, "from below", $v_{k,m}^-$, and the "jump" $[v_k]_m$ by

$$\begin{aligned} v_{k,m}^+ &:= \lim_{t \to 0^+} v_k(t_m + t) = v_k(t_{m+1}) =: v_{k,m+1}, \\ v_{k,m}^- &:= \lim_{t \to 0^+} v_k(t_m - t) = v_k(t_m) \quad =: v_{k,m}, \end{aligned}$$

$$[v_k]_m := v_{k,m}^+ - v_{k,m}^- = v_{k,m+1} - v_{k,m},$$

as in [44] which are illustrated in Figure 5.1. Moreover, we introduce the short notation

$$(v, w)_{I_m} \quad := \quad (v, w)_{L^2(I_m, L^2(\Omega))}, \quad \|v\|_{I_m} \quad := \quad \|v\|_{L^2(I_m, L^2(\Omega))}$$

as well as a semidiscrete bilinear form $B(\cdot,\cdot)\colon Y_k^0 \times Y_k^0 \to \mathbb{R}$ by

$$\begin{aligned} B(y_k, \varphi) &:= (\nabla y_k, \nabla \varphi)_I + \sum_{m=2}^{M} ([y_k]_{m-1}, \varphi_{m-1}^+) + (y_{k,0}^+, \varphi_0^+) \\ &= (\nabla y_k, \nabla \varphi)_I + \sum_{m=2}^{M} (y_{k,m} - y_{k,m-1}, \varphi_m) + (y_{k,1}, \varphi_1). \end{aligned} \tag{5.4.1}$$

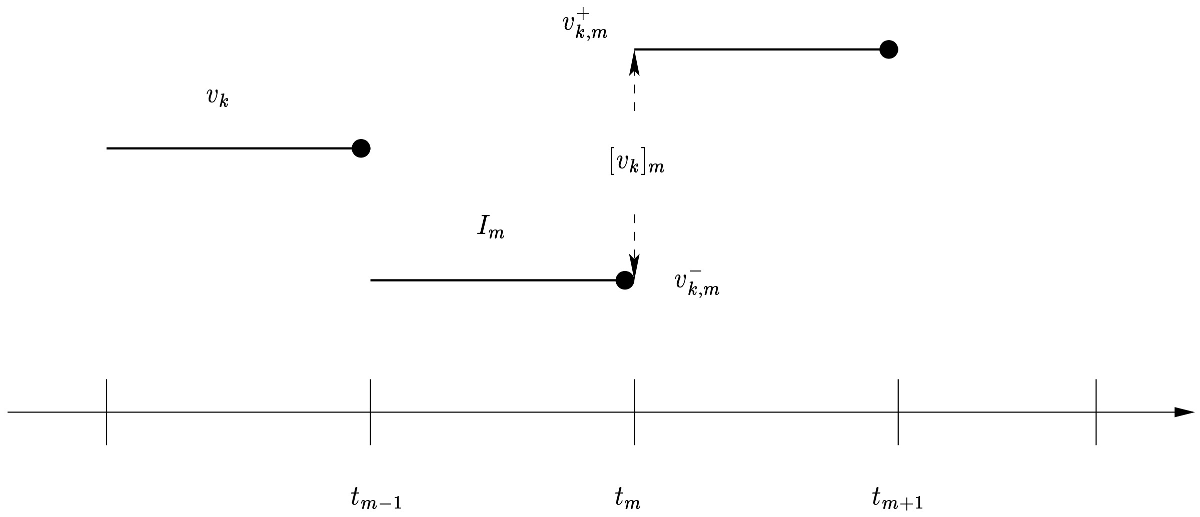


Figure 5.1.: Notation for the dG(0) method

This can also be expressed as

$$B(y_k, \varphi) = (\nabla y_k, \nabla \varphi)_I + \sum_{m=1}^{M-1} ([y_k]_m, \varphi_m^+) + (y_{k,1}, \varphi_1).$$

No value of φ is required at $t = 0$. We refer to [47] for further details as well as an equivalent formulation of the $dG(0)$ scheme. The $dG(0)$ semidiscretization of the state equation (5.3.2) for fixed control $u \in U$ now reads as follows:

Find a state $y_k = y_k(u) \in Y_k^0$ such that

$$B(y_k, \varphi) + (d(\cdot, y_k), \varphi)_I = (u, \varphi)_I + (y_0, \varphi_1) \quad \forall \varphi \in Y_k^0. \tag{5.4.2}$$

Note that we do not require y_k to fulfill the initial condition exactly. To explain this, consider the case $M = 1$, i.e. a setting with only one time interval $(t_0, t_1] = (0, T]$. Then, we obtain the semidiscrete state equation

$$(\nabla y_k, \nabla \varphi)_I + (y_k, \varphi) + (d(\cdot, y_k), \varphi)_I = (u, \varphi)_I + (y_0, \varphi) \quad \forall \varphi \in \mathcal{P}_0(I, V), \tag{5.4.3}$$

which is equivalent to

$$(\nabla y_k, \nabla \varphi)_I + (y_k - y_0, \varphi) + (d(\cdot, y_k), \varphi)_I = (u, \varphi)_I \quad \forall \varphi \in \mathcal{P}_0(I, V). \tag{5.4.4}$$

Since both the state space and the test space are the same, i.e. constant polynomials in time with values in $V = H_0^1(\Omega)$, y_k cannot be expected to fulfill the initial condition

$$y_k(0, x) = y_0(x),$$

exactly, as this would lead to an overdetermined system. Instead, the term

$$(y_k - y_0, \varphi)$$

in (5.4.4) imposes the initial condition in a variational sense. For $M > 1$, successive application of this method on the time intervals I_m, $m = 1, \ldots, M$ explains the jump terms in the definition of B. We refer again to [44].

For later use of the bilinear form for the adjoint equation we point out that for $\varphi, v_k \in Y_k^0$, B fulfills

$$
\begin{aligned}
B(\varphi, v_k) = \;& (\nabla\varphi, \nabla v_k)_I + \sum_{m=1}^{M-1} ([\varphi_m], v_{k,m}^+) + (\varphi_1, v_{k,1}) \\
= \;& (\nabla\varphi, \nabla v_k)_I + \sum_{m=1}^{M-1} (\varphi_{m+1} - \varphi_m, v_{k,m+1}) + (\varphi_1, v_{k,1}) \\
= \;& (\nabla\varphi, \nabla v_k)_I - \sum_{m=1}^{M-1} (\varphi_m, v_{k,m+1} - v_{k,m}) + \sum_{m=1}^{M-1} (\varphi_{m+1}, v_{k,m+1}) \\
& - \sum_{m=1}^{M-1} (\varphi_m, v_{k,m}) + (\varphi_1, v_{k,1}) \\
= \;& (\nabla\varphi, \nabla v_k)_I - \sum_{m=1}^{M-1} (\varphi_m^-, [v_k]_m) + (\varphi_M^-, v_{k,M}^-) \\
= \;& (\nabla\varphi, \nabla v_k)_I - \sum_{m=1}^{M-1} (\varphi_m^-, [v_k]_m) + (\varphi_M, v_{k,M}),
\end{aligned}
\tag{5.4.5}
$$

cf. [101].

For all what follows, we point out again that the spatial dimension $n = 2$ is considered. The reason for this lies in regularity results for the space-time-discretized state, but we will adapt all further discussion to this setting. In principal, we will proceed with the same or similar steps as for the continuous problem. We will first discuss existence, regularity, and stability estimates for the semi-discrete state equation, where we cannot easily make use of available results. Once these are available, optimality conditions of the semi-discrete control problem can be discussed, along with a discussion of linearized and adjoint equations. Then, a discretization error of uncontrolled equations is derived, that will eventually determine the temporal error in the optimal controls.

5.4.2. Existence results and stability estimates for the semi-discrete state equation

The properties of the semi-discrete state equation essentially influence the analysis of the semidiscrete optimal control problem. We will therefore very carefully discuss the semi-discrete equations. Once results comparable to the continuous level are available we can proceed with the analysis of the semidiscrete optimal control problem. Most stability estimates needed can be obtained extending results for linear equations from [101], cf. also [45, 46]. We begin with plain existence and uniqueness, as well as boundedness in $L^\infty(Q)$ independently of the discretization. In some sense, this is the central argument in the following proofs.

Theorem 5.4.1. *Let Assumption 5.3.1 be satisfied. Then, for every fixed control $u \in U$ and initial state $y_0 \in V \cap L^\infty(Q)$, the semidiscrete state equation* (5.4.2) *admits a unique semidiscrete solution $y_k \in Y_k^0 \cap L^\infty(Q)$.*

Proof. The existence of a solution follows by applying standard arguments from elliptic theory to the system of semilinear elliptic PDEs for each time interval obtained after semidiscretization in time. It is sufficient to prove the assertion on the interval I_1, cf. also [138]. We obtain the formulation

$$k_1(\nabla y_{k,1}, \nabla \varphi_1) + (y_{k,1}, \varphi_1) + (d(t,x,y_{k,1}), \varphi_1)_{I_1} = (u, \varphi_1)_{I_1} + (y_0, \varphi_1) \quad \forall \varphi_1 \in V,$$

where u and y_0 are given. With the auxiliary functions

$$\begin{aligned} \tilde{d}(x, y_{k,1}) &:= \int_{I_1} d(t,x,y_{k,1}(x))dt, \\ \tilde{u}(x) &:= \int_{I_1} u(t,x)dt, \end{aligned}$$

this yields

$$k_1(\nabla y_{k,1}, \nabla \varphi_1) + (y_{k,1}, \varphi_1) + (\tilde{d}(\cdot, y_{k,1}), \varphi_1) = (\tilde{u}, \varphi_1) + (y_0, \varphi_1) \quad \forall \varphi_1 \in V. \tag{5.4.6}$$

Clearly, the monotonicity and Lipschitz properties of d remain valid for $\tilde{d}$, and $\tilde{u}$ is a function from $L^\infty(\Omega)$. We can apply Theorem 2.4.6 to this elliptic boundary value problem and for each $y_0, \tilde{u} \in L^2(\Omega)$ we obtain a unique solution $y_{k,1} \in H_0^1(\Omega) \cap L^\infty(\Omega)$. This argumentation can successively be applied to all time intervals, and we obtain the existence of a semidiscrete solution $y_k \in Y_k^0 \cap L^\infty(Q)$ to (5.4.2). □

The existence of a unique solution $y_k \in L^\infty(Q)$ is the starting point for our analysis, but not sufficient for later results. For the stability and error estimates to come, we will need to rely on boundedness estimates that do not depend on the discretization parameter k. Yet, applying elliptic stability arguments to (5.4.6) will lead to an estimate which is proportional to $1/k$.

In elliptic problems, boundedness of the discrete states y_h in $L^\infty(\Omega)$ can be obtained with the help of an inverse estimate between the norms on $L^\infty(\Omega)$ and $L^2(\Omega)$ for cellwise polynomials, once an error estimate in the L^2-norm depending on the spatial discretization parameter h is established under strong assumptions on the appearing nonlinearities. These strong assumptions can be weakened after boundedness of the discrete states is obtained, cf. [23] for a detailed discussion. This procedure is not easily transferable to parabolic problems, since an estimate in the $L^\infty(Q)$ norm on the whole space-time-domain is needed. In [46], error estimates in the $L^\infty(I \times \Omega)$-norm have been shown for linear problems, but the regularity assumptions are too strong for our optimal control problem. In order to obtain a plain boundedness result independent of k, we will continue differently and apply Stampacchia's method to the semidiscrete state equation (5.4.2).

Theorem 5.4.2. *Let the conditions from Theorem 5.4.1 hold and note in particular that the spatial dimension $n = 2$. Then, the solution y_k satisfies the boundedness result*

$$\|y_k\|_{\infty,\infty} \le C\left(\|u\|_{L^p(Q)} + \|d(\cdot,0)\|_{L^p(Q)} + \|y_0\|_\infty\right)$$

for every $p > 2$ and a constant C independent of the discretization parameter k.

Proof. We follow closely the proof of [144, Theorem 4.5] for elliptic equations. Let therefore $b \ge \|y_0\|_\infty$ be a real number and choose a test-function $v_k \in Y_k^0$ such that

$$v_{k,m} = \begin{cases} y_{k,m} - b & \text{on} \quad \Omega_{k,m}^+(b) := \{x \in \Omega \colon y_{k,m}(x) > b\} \\ y_{k,m} + b & \text{on} \quad \Omega_{k,m}^-(b) := \{x \in \Omega \colon y_{k,m}(x) < -b\} \\ 0 & \text{on} \quad \Omega \setminus (\Omega_{k,m}^+(b) \cup \Omega_{k,m}^-(b)). \end{cases}$$

For brevity, we will write $\Omega^+_{k,m}$ instead of $\Omega^+_{k,m}(b)$ as well as $\Omega^-_{k,m}$ instead of $\Omega^-_{k,m}(b)$.

In a first step, we estimate $B(v_k, v_k) + (d(\cdot, v_k), v_k)_I$ against $B(y_k, v_k) + (d(\cdot, y_k), v_k)_I$. It is easily verified that

$$(d(\cdot, y_k) - d(\cdot, 0), v_k)_{I_m} \geq 0 \quad \forall m = 1, \ldots, M \tag{5.4.7}$$

as well as

$$\|\nabla v_{k,m}\|^2 = (\nabla v_{k,m}, \nabla y_{k,m}) \quad \forall m = 1, \ldots, M \tag{5.4.8}$$

hold. Note for instance that on $\Omega^+_{k,m}$, where $y_{k,m} - b > 0$ is satisfied, we have

$$\begin{aligned}\int_{I_m}\int_{\Omega^+_{k,m}} (d(\cdot, y_k) - d(\cdot, 0)) v_k \, dx dt &= \int_{I_m}\int_{\Omega^+_{k,m}} (d(\cdot, y_k - b + b) - d(\cdot, 0))(y_k - b) \, dx dt \\ &\geq \int_{I_m}\int_{\Omega^+_{k,m}} (d(\cdot, y_k - b) - d(\cdot, 0))(y_k - b) \, dx dt \geq 0\end{aligned}$$

by the monotonicity of d with respect to y. Statement (5.4.8) follows directly from the definition of v_k. To estimate the jump terms observe that for $m = 2, \ldots, M$, we have

$$(y_{k,m} - y_{k,m-1}, v_{k,m}) = (v_{k,m} + b - y_{k,m-1}, v_{k,m})_{\Omega^+_{k,m}} + (v_{k,m} - b + y_{k,m-1}, v_{k,m})_{\Omega^-_{k,m}},$$

where the index $\Omega^+_{k,m}$ or $\Omega^-_{k,m}$ indicates L^2 inner products on the respective set. On $\Omega^+_{k,m}$, we observe further:

- $$(v_{k,m} + b - y_{k,m-1}, v_{k,m})_{\Omega^+_{k,m} \cap \Omega^+_{k,m-1}} = (v_{k,m} - v_{k,m-1}, v_{k,m})_{\Omega^+_{k,m} \cap \Omega^+_{k,m-1}}, \tag{5.4.9}$$

  since $b - y_{k,m-1} = v_{k,m-1}$ on $\Omega^+_{k,m-1}$ by definition of v_k

- $$(v_{k,m} + b - y_{k,m-1}, v_{k,m})_{\Omega^+_{k,m} \cap \Omega^-_{k,m-1}} \geq (v_{k,m} - v_{k,m-1}, v_{k,m})_{\Omega^+_{k,m} \cap \Omega^-_{k,m-1}}, \tag{5.4.10}$$

  since $b - y_{k,m-1} = -v_{k,m-1} + 2b > -v_{k,m-1}$ on $\Omega^-_{k,m-1}$ and $v_{k,m} > 0$ on $\Omega^+_{k,m}$

- $$\begin{aligned}(v_{k,m} + b - y_{k,m-1}, v_{k,m})_{\Omega^+_{k,m} \setminus (\Omega^-_{k,m-1} \cup \Omega^+_{k,m-1})} &\geq (v_{k,m}, v_{k,m})_{\Omega^+_{k,m} \setminus (\Omega^-_{k,m-1} \cup \Omega^+_{k,m-1})}, \\ &= (v_{k,m} - v_{k,m-1}, v_{k,m})_{\Omega^+_{k,m} \setminus (\Omega^-_{k,m-1} \cup \Omega^+_{k,m-1})}\end{aligned} \tag{5.4.11}$$

  since $|y_{k,m-1}| \leq b$ and $v_{k,m-1} = 0$ on $\Omega^+_{k,m} \setminus (\Omega^-_{k,m-1} \cup \Omega^+_{k,m-1})$.

Combining (5.4.9)–(5.4.11) yields

$$(y_{k,m} - y_{k,m-1}, v_{k,m})_{\Omega^+_{k,m}} \geq (v_{k,m} - v_{k,m-1}, v_{k,m})_{\Omega^+_{k,m}}, \tag{5.4.12}$$

and with analogous calculations on $\Omega^-_{k,m}$, we arrive at

$$(y_{k,m} - y_{k,m-1}, v_{k,m}) \geq (v_{k,m} - v_{k,m-1}, v_{k,m}) \quad \forall\, m = 2, \ldots, M. \tag{5.4.13}$$

For $m = 1$, we proceed similarly, taking into account the term $(y_0, v_{k,1})$ from the right-hand side of the state equation (5.4.2). On $\Omega^+_{k,1}$, we have

$$(y_{k,1} - y_0, v_{k,1})_{\Omega^+_{k,1}} = (v_{k,1} + b - y_0, v_{k,1})_{\Omega^+_{k,1}} \geq \|v_{k,1}\|^2_{\Omega^+_{k,1}},$$

where the last inequality follows from the fact that $b \geq \|y_0\|_\infty$ and $v_{k,1} > 0$ on $\Omega^+_{k,1}$. Together with analogous calculations on $\Omega^-_{k,1}$, this yields

$$(y_{k,1} - y_0, v_{k,1}) \geq \|v_{k,1}\|^2. \tag{5.4.14}$$

From (5.4.7),(5.4.8),(5.4.13), and (5.4.14) we obtain

$$B(v_k, v_k) \leq B(y_k, v_k) - (y_0, v_{k,1}) \leq (u - d(\cdot, 0), v_k)_I,$$

which in particular implies

$$\|v_k\|^2_{L^\infty(I,L^2(\Omega))\cap L^2(I,H^1_0(\Omega))} \leq (u - d(\cdot,0), v_k)_I, \tag{5.4.15}$$

as we will discuss in more detail in the proof of Theorem 5.4.5 below.

In the second step of the proof, we estimate (5.4.15) further. Note that we have the embeddings

$$H^1_0(\Omega) \hookrightarrow L^\sigma(\Omega), \quad \forall\, 1 < \sigma < \infty$$

as well as

$$L^\infty(I, L^2(\Omega)) \hookrightarrow L^\sigma(I, L^2(\Omega)), \quad \forall\, 1 < \sigma < \infty$$

since Ω is two dimensional, cf. [1]. Then, known interpolation error estimates, cf. [139], yield

$$L^2(I, H^1_0(\Omega)) \cap L^\sigma(I, L^2(\Omega)) \hookrightarrow L^{r_s}(I, [H^1_0(\Omega), L^2(\Omega)]_s) \hookrightarrow L^{r_s}(I, L^{q_s}(\Omega))$$

with

$$\frac{1}{r_s} = \frac{1-s}{2} + \frac{s}{\sigma}, \quad \frac{1}{q_s} = \frac{1-s}{\sigma} + \frac{s}{2}, \quad s \in (0,1).$$

Choosing $s = \frac{1}{2}$ yields

$$r_s = q_s = \frac{4\sigma}{2+\sigma} < 4,$$

which is monotonically increasing in σ. Hence, we deduce that

$$L^\infty(I, L^2(\Omega)) \cap L^2(I, H^1_0(\Omega)) \hookrightarrow L^\tau(Q) \tag{5.4.16}$$

is satisfied for any positive real number $\tau < 4$. To eventually be able to apply Stampacchia's Lemma we point out that for all $p > 2$ there exist $\lambda > 1$ and $p' > 0$ such that $\frac{1}{p} + \frac{1}{p'} = 1$ and $p' = \frac{\tau}{2\lambda}$ is satisfied. With the embedding (5.4.16) as well as estimate (5.4.15) we obtain

$$\|v_k\|^2_{L^\tau(Q)} \leq c\|v_k\|^2_{L^\infty(I,L^2(\Omega))\cap L^2(I,H^1_0(\Omega))} \leq c\,\|u - d(\cdot,0)\|_{L^p(Q)}\|v_k\|_{L^{p'}(Q)} \tag{5.4.17}$$

where we used Hölder's inequality. Defining $\mathcal{I}(b) \subset Q$ as

$$\mathcal{I}(b) := \{(t,x) \in Q \mid |y_k(t,x)| > b\},$$

and denoting by $\chi_{\mathcal{I}(b)}$ its characteristic function we obtain from (5.4.17)

$$\begin{aligned}
\|v_k\|^2_{L^\tau(Q)} &\leq c\,\|u - d(\cdot,0)\|_{L^p(Q)}\|v_k\|_{L^{p'}(Q)} \\
&= c\,\|u - d(\cdot,0)\|_{L^p(Q)} \cdot \left(\iint_Q \chi_{\mathcal{I}(b)}\, |v_k|^{p'}\, dxdt\right)^{\frac{1}{p'}} \\
&\leq c\,\|u - d(\cdot,0)\|_{L^p(Q)}\|v_k\|_{L^{2p'}(Q)}|\mathcal{I}(b)|^{\frac{1}{2p'}},
\end{aligned}$$

where we applied Hölder's inequality with Hölder exponent 2. Estimating this further using $2p' = \frac{\tau}{\lambda} < \tau$ since $\lambda > 1$ leads to

$$\begin{aligned} \|v_k\|^2_{L^\tau(Q)} &\leq c\,\|u - d(\cdot,0)\|_{L^p(Q)}\|v_k\|_{L^{\frac{\tau}{\lambda}}(Q)}|\mathcal{I}(b)|^{\frac{\lambda}{\tau}} \\ &\leq c\,\|u - d(\cdot,0)\|_{L^p(Q)}\|v_k\|_{L^\tau(Q)}|\mathcal{I}(b)|^{\frac{\lambda}{\tau}}. \end{aligned}$$

With Young's inequality this transforms to

$$\|v_k\|^2_{L^\tau(Q)} \leq c_{1/\varepsilon}\,\|u - d(\cdot,0)\|^2_{L^p(Q)}|\mathcal{I}(b)|^{2\frac{\lambda}{\tau}} + c_\varepsilon\|v_k\|^2_{L^\tau(Q)}.$$

Choosing c_ε sufficiently small, we obtain

$$\|v_k\|^2_{L^\tau(Q)} \leq c\,\|u - d(\cdot,0)\|^2_{L^p(Q)}|\mathcal{I}(b)|^{\frac{2}{\tau}\lambda}, \tag{5.4.18}$$

where c might be large but finite, depending on the choice of c_ε.

The third and last step of the proof involves rewriting (5.4.18) such that Stampacchia's Lemma, cf. [80] can be applied. By the definition of v_k, (5.4.18) implies

$$\left(\int\limits_{\mathcal{I}(b)} (|y_k| - b)^\tau dx\,dt\right)^{\frac{2}{\tau}} \leq c\,\|u - d(\cdot,0)\|^2_{L^p(Q)}|\mathcal{I}(b)|^{\frac{2}{\tau}\lambda}. \tag{5.4.19}$$

For every $\tilde{b} > b$ we have $|\mathcal{I}(\tilde{b})| \leq |\mathcal{I}(b)|$, and we can estimate

$$\left(\int\limits_{\mathcal{I}(b)} (|y_k| - b)^\tau dx\,dt\right)^{\frac{2}{\tau}} \geq \left(\int\limits_{\mathcal{I}(\tilde{b})} (\tilde{b} - b)^\tau\,dxdt\right)^{\frac{2}{\tau}} \geq (\tilde{b} - b)^2|\mathcal{I}(\tilde{b})|^{\frac{2}{\tau}}.$$

Combining this with (5.4.19) finally yields

$$(\tilde{b} - b)^2|\mathcal{I}(\tilde{b})|^{\frac{2}{\tau}} \leq c\,\|u - d(\cdot,0)\|^2_{L^p(Q)}|\mathcal{I}(b)|^{\frac{2}{\tau}\lambda}.$$

Applying Stampacchia's Lemma, which is given in Lemma 7.5 from [144] in a form close to our notation, yields that $|\mathcal{I}(b)| = 0$ for b large enough, i.e. $y_k(t,x) \leq b$ almost everywhere in Q. The desired boundedness estimate follows from the same lemma. □

Corollary 5.4.3. *As an immediate consequence of the last theorem we obtain that for all controls $u \in U_{ad}$ the associated states $y_k(u)$ are uniformly bounded in $L^\infty(Q)$ independently of k and u by the boundedness of U_{ad}.*

As an analogue to Remark 5.3.8 on page 123 we obtain:

Remark 5.4.4. *The set of admissible controls U_{ad} is still the same as on the continuous level and therefore it is bounded in $L^\infty(Q)$ by $|u_a| + |u_b|$. Then, Theorem 5.4.2 guarantees boundedness of all semidiscrete states $y_k = y_k(u)$ in the $L^\infty(Q)$ norm independent from $u \in U_{ad}$ as well as independent from the discretization parameter k. Consequently, we again observe that the (local) Lipschitz continuity of e.g. d and its derivatives can be used with constants C independent of y_k as well as k in all proofs in the sequel, since we will only consider controls $u \in U_{ad}$.*

Now, we complete the semidiscrete analogue to Theorem 2.5.6 by a stability estimate that we will essentially rely on when proving our error estimates. This is the extension of an analogous result in [101] for linear equations, cf. also [45, 46] for similar estimates. Let us point out two helpful and straight-forward identities. Every function $v_k \in Y_k^0$ satisfies

$$([v_k]_{m-1}, v_{k,m}) = \frac{1}{2}\left(\|v_{k,m}\|^2 + \|[v_k]_{m-1}\|^2 - \|v_{k,m-1}\|^2\right), \tag{5.4.20}$$

as well as

$$([\nabla v_k]_{m-1}, \nabla v_{k,m}) = \frac{1}{2}\left(\|\nabla v_{k,m}\|^2 + \|[\nabla v_k]_{m-1}\|^2 - \|\nabla v_{k,m-1}\|^2\right). \tag{5.4.21}$$

Theorem 5.4.5. *For the solution $y_k \in Y_k^0$ of the dG(0) semidiscretized state equation* (5.4.2) *with right-hand side $u \in U$ and initial condition $y_0 \in H_0^1(\Omega)$, the stability estimate*

$$\|y_k\|_I^2 + \|y_k\|_{\infty,2}^2 + \|\nabla y_k\|_I^2 + \|\nabla y_k\|_{\infty,2}^2 + \|\Delta y_k\|_I^2 + \sum_{m=1}^{M} k_m^{-1}\|[y_k]_{m-1}\|^2$$
$$\leq C\left(\|u\|_I^2 + \|d(\cdot,0)\|_I^2 + \|\nabla y_0\|^2\right)$$

holds. The constant C depends only on the domain Ω. The jump term $[y_k]_0$ at $t = 0$ is defined as $y_{k,1} - y_0$.

Proof. For all $\varphi \in Y_k^0$, the solution $y_k \in Y_k^0$ of (5.4.2) satisfies the following system of equations:

$$(\nabla y_k, \nabla\varphi)_{I_m} + (y_{k,m} - y_{k,m-1}, \varphi_m) + (d(\cdot, y_k), \varphi)_{I_m} = (u, \varphi)_{I_m}, \quad m = 1, \ldots, M. \tag{5.4.22}$$

The proof consists of three steps. First, we test (5.4.22) with the semidiscrete solution y_k to obtain an estimate for $\|\nabla y_k\|_I$ and $\|y_k\|_I$ in principle using the monotonicity of d, where it is helpful that $y_{k,m} \in H_0^1(\Omega)$, since this allows to use Poincaré's inequality. Testing (5.4.22) with $y_k \in Y_k^0$ yields

$$(u, y_k)_{I_m} = \|\nabla y_k\|_{I_m}^2 + (d(\cdot, y_k), y_k)_{I_m} + \frac{1}{2}(\|y_{k,m}\|^2 + \|[y_k]_{m-1}\|^2 - \|y_{k,m-1}\|^2), \tag{5.4.23}$$

for all $m = 1, \ldots, M$, where we used (5.4.20). By summation over all $m = 1, \ldots, M$, we obtain

$$\|\nabla y_k\|_I^2 + (d(\cdot, y_k), y_k)_I + \frac{1}{2}\|y_{k,M}\|^2 + \sum_{i=1}^{M}\|[y_k]_{m-1}\|^2 = (u, y_k)_I + \frac{1}{2}\|y_0\|^2. \tag{5.4.24}$$

This implies in particular that

$$\|\nabla y_k\|_I^2 + (d(\cdot, y_k) - d(\cdot, 0), y_k)_I \leq \frac{1}{2}\|y_0\|^2 + \|u - d(\cdot, 0)\|_I\|y_k\|_I. \tag{5.4.25}$$

By the monotonicity of d and Young's and Poincarè's inequality, (5.4.25) leads to

$$\|\nabla y_k\|_I^2 \leq c(\|u\|_I^2 + \|d(\cdot, 0)\|_I^2 + \|\nabla y_0\|^2), \tag{5.4.26}$$

since $y_k \in H_0^1(\Omega)$. Note that then Poincaré's inequality also implies

$$\|y_k\|_I^2 \leq c(\|u\|_I^2 + \|d(\cdot, 0)\|_I^2 + \|\nabla y_0\|^2). \tag{5.4.27}$$

In the second step, we estimate the terms $\|\Delta y_k\|_I^2$ and $\sum_{m=1}^{M} k_m^{-1}\|[y_k]_{m-1}\|^2$ by applying Theorem 4.1 from [101] to the linear equation

$$B(y_k,\varphi) = (u - d(\cdot,y_k),\varphi)_I + (y_0,\varphi_1), \quad \forall \varphi \in Y_k^0. \tag{5.4.28}$$

Then, we know that

$$\|\Delta y_k\|_I^2 + \sum_{m=1}^{M} k_m^{-1}\|[y_k]_{m-1}\|^2 \leq c\left(\|u - d(\cdot,y_k)\|_I^2 + \|\nabla y_0\|^2\right), \tag{5.4.29}$$

which follows in principle from testing (5.4.2) with $\varphi = [y_k]_{m-1} \in \mathcal{P}_0(I, H_0^1(\Omega))$ as well as integrating (5.4.2) by parts in space to be able to test with $\varphi = -\Delta y_k \in \mathcal{P}_0(I, L^2(\Omega))$. For more details, we refer to [101]. The right-hand side of (5.4.29) can be estimated by

$$\|u - d(\cdot,y_k)\|_I^2 \leq c(\|u\|_I^2 + \|d(\cdot,0)\|_I^2 + \|d(\cdot,0) - d(\cdot,y_k)\|_I^2) \leq c\left(\|u\|_I^2 + \|d(\cdot,0)\|_I^2 + \|y_k\|_I^2\right) \tag{5.4.30}$$

by the local Lipschitz continuity of d and the fact that y_k is uniformly bounded for all $u \in U_{\text{ad}}$ due to Theorem 5.4.2. Inserting (5.4.27) yields the proposed estimate for $\|\Delta y_k\|_I$ and the jump terms.

In a third step, we need to estimate $\|y_k\|_{\infty,2}$ and $\|\nabla y_k\|_{\infty,2}$. In fact, the proof in [101, Theorem 4.1] already provides this (for linear equations). Consider again the linear equation (5.4.28). Partial integration in space and testing with $\varphi = -\Delta y_k$ yields

$$\|\Delta y_k\|_{I_m}^2 + ([\nabla y_k]_{m-1}, \nabla y_{k,m}) = (u - d(\cdot,0) + d(\cdot,0) - d(\cdot,y_k), -\Delta y_k)_{I_m} \tag{5.4.31}$$

for all $m = 1,\dots,M$. We refer to [101] for further details. Using (5.4.21) and the Lipschitz property of d and summation over all m yield

$$\frac{1}{2}\|\nabla y_{k,M}\|^2 + \frac{1}{2}\sum_{m=1}^{M} \|[\nabla y_k]_{m-1}\|^2 + \|\Delta y_k\|_I^2 \leq \|u - d(\cdot,0)\|_I\|\Delta y_k\|_I + c\,\|y_k\|_I\|\Delta y_k\|_I + \frac{1}{2}\|\nabla y_0\|_I^2.$$

With Young's inequality and (5.4.27) this implies

$$\|\nabla y_{k,M}\|^2 + \sum_{m=1}^{M} \|[\nabla y_k]_{m-1}\|^2 \leq c\left(\|u\|_I^2 + \|\nabla y_0\|^2 + 1\right).$$

Since now each of the jump terms and therefore also their maximum is bounded, this yields the desired estimate for $\|y_k\|_{\infty,2}$ and $\|\nabla y_k\|_{\infty,2}$ after again applying Poincaré's inequality. □

Remark 5.4.6. *Note that elliptic regularity and the fact that Ω is convex and polygonal imply*

$$\|\nabla^2 y_k\|_I^2 \leq c\|\Delta y_k\|_I^2 \leq C\left(\|u\|_I^2 + \|d(\cdot,0)\|_I^2 + \|\nabla y_0\|^2\right).$$

5.4.3. The optimal control problem with semidiscrete state equation

The existence and regularity results for the semidiscrete state equation allow to treat a time-discrete version of the model problem by the same means as on the continuous level. We start with the definition of a control-to-state operator and a reduced objective function.

Definition 5.4.7. *The mapping*

$$G_k : U \to Y_k^0, \quad u \mapsto y_k = G_k(u),$$

where y_k solves (5.4.2)*, is called the semidiscrete control-to-state-operator. In addition, we introduce the semi-discrete reduced objective functional*

$$f_k : U \to \mathbb{R}, \quad f_k(u) := J(G_k(u), u).$$

The resulting reduced semidiscrete problem formulation reads

$$\text{Minimize } f_k(u) \quad u \in U_{\text{ad}}. \tag{$\mathbb{P}_k$}$$

We point out that the index k in f_k only indicates that the control belongs to a semidiscretized state equation. The controls themselves remain undiscretized at this point. Yet, to be able to distinguish between the optimal controls on the different levels of discretization, we denote an optimal control belonging to the semidiscrete state equation by an index k.

Theorem 5.4.8. *The semidiscrete optimal control problem* ($\mathbb{P}_k$) *admits at least one (globally) optimal control $\bar{u}_k \in U_{ad}$ with associated optimal semidiscrete state $\bar{y}_k = G_k(\bar{u}_k)$.*

Proof. We again refer to the technique of proof presented in [144]. We particularly mention again that the set of admissible controls U_{ad} is the same as for Problem ($\mathbb{P}$) and therefore not empty. □

A semidiscrete local solution is defined in essentially the same way as in the last section.

Definition 5.4.9. *A control $\bar{u}_k \in U_{ad}$ is called a semidiscrete local solution to* ($\mathbb{P}_k$) *in the sense of $L^2(Q)$ if there exists an $\varepsilon > 0$ such that for all $u \in U_{ad}$ with $\|u - \bar{u}_k\|_I \leq \varepsilon$*

$$f_k(u) \geq f_k(\bar{u}_k)$$

is satisfied.

Next, we summarize differentiability results.

Proposition 5.4.10. *The control-to-state-operator G_k is of class $\mathcal{C}^2$. Let $u, v, v_1, v_2 \in U$ and $y_k = G_k(u)$ be given. Then, its first and second order derivatives are given by $\tilde{y}_k := G_k'(u)v$ and $\tilde{z}_k := G_k''(u)v_1v_2$ being the solutions of*

$$B(\tilde{y}_k, \varphi) + (\partial_y d(\cdot, y_k)\tilde{y}_k, \varphi)_I \quad = \quad (v, \varphi)_I \quad \forall \varphi \in Y_k^0, \tag{5.4.32}$$

as well as

$$B(\tilde{z}_k, \varphi) + (\partial_y d(\cdot, y_k)\tilde{z}_k, \varphi)_I \quad = \quad (-\partial_{yy} d(\cdot, y_k)\tilde{y}_{k,1}\tilde{y}_{k,2}, \varphi)_I \quad \forall \varphi \in Y_k^0, \tag{5.4.33}$$

where $y_k = G_k(u)$,$\tilde{y}_{k,1} = G_k'(u)v_1$ and $\tilde{y}_{k,2} = G_k'(u)v_2$. For the reduced objective function, we obtain

$$f_k'(u)v = \iint\limits_Q (\psi_1 \tilde{y}_k y_k + \psi_2 \tilde{y}_k u + \psi_2 y_k v + \psi_3 \tilde{y}_k + \nu u v)\, dxdt$$

$$f_k''(u)[v_1, v_2] = \iint\limits_Q (\psi_1 \tilde{y}_{k,1}\tilde{y}_{k,2} + \psi_1 z_k y_k + \psi_2 \tilde{y}_{k,2} v_2 + \psi_2 z_k u + \psi_2 \tilde{y}_{k,2} v_1 + \psi_3 z_k + \nu v_1 v_2)\, dx\, dt.$$

Proof. The proof follows as in Theorems 5.9, 5.15, and 5.16 in [144] for the control-to-state operator. By the differentiability of J in $L^2(Q) \times L^2(Q)$ and the chain rule, we know that f is of class $\mathcal{C}^2$. Straight-forward calculations yield the precise formulation of the derivatives. □

It is clear that the stability estimates of Theorems 5.4.2 and 5.4.5 are also valid for linearized semidiscrete state equations:

Lemma 5.4.11. *Let $u \in U_{ad}$ and $v \in U$ be given and denote by $\tilde{y}_k$ the semidiscrete linearized state $\tilde{y}_k = G_k'(u)v$. Then there exists a constant $c > 0$ independent of k and $u \in U_{ad}$ such that*

$$\begin{aligned} \|\tilde{y}_k\|_I &\leq c\,\|v\|_I && (5.4.34)\\ \|\tilde{y}_k\|_{\infty,\infty} &\leq c\,\|v\|_{L^p(Q)} && (5.4.35)\\ \|\tilde{y}_k\|_{\infty,2} + \|\nabla\tilde{y}_k\|_{\infty,2} &\leq c\,\|v\|_I && (5.4.36) \end{aligned}$$

holds for any $p > 2$.

Note in particular that the constant c are also independent of the term $\partial_y d(\cdot, y_k)$ appearing in the linearized equations, since y_k is uniformly bounded for all $u \in U_{\mathrm{ad}}$, independent of k.

For $u \in U$ and $y_k := G_k(u)$ we define the semidiscrete adjoint state $p_k = p_k(u) \in Y_k^0$ as the solution of the semidiscrete adjoint equation

$$B(\varphi, p_k) + (\varphi, \partial_y d(\cdot, y_k)p_k) = (\varphi, \psi_1 y_k + \psi_2 u + \psi_3)_I \quad \forall \varphi \in Y_k^0. \tag{5.4.37}$$

We denote by $\bar{p}_k$ the adjoint state associated with the locally optimal control $\bar{u}_k$ and its corresponding optimal state $\bar{y}_k$. For a more general right-hand side $g \in L^\infty(Q)$ and terminal condition $p_T \in H_0^1(\Omega)$ we consider the semidiscretized dual equation

$$B(\varphi, p_k) + (\varphi, \partial_y d(\cdot, y_k(u))p_k)_I = (\varphi, g)_I + (\varphi_M, p_T) \quad \forall\, \varphi \in Y_k^0. \tag{5.4.38}$$

and obtain the following result applicable to (5.4.37).

Corollary 5.4.12. *For each right-hand side $g \in L^\infty(Q)$ and terminal condition $p_T \in H_0^1(\Omega) \cap L^\infty(\Omega)$, there exists a unique solution*

$$p_k \in Y_k^0 \cap L^2(I, H^2(\Omega) \cap H_0^1(\Omega)) \cap L^\infty(Q)$$

of the semidiscrete dual equation (5.4.38) and a real number $p > 2$ such that the estimates

$$\|p_k\|_{\infty,\infty} \leq C(\|g\|_{L^p(Q)} + \|p_T\|_\infty)$$

$$\|p_k\|_I^2 + \|\nabla p_k\|_I^2 + \|p_k\|_{\infty,2}^2 + \|\nabla p_k\|_{\infty,2}^2 + \|\Delta p_k\|_I^2 + \sum_{m=1}^{M} k_m^{-1}\|[p_k]_m\|^2 \leq C(\|g\|_I^2 + \|\nabla p_T\|^2)$$

hold with a constant $C > 0$. Here, the jump term $[p_k]_M$ at $t = T$ is defined as $p_T - p_{k,M}$.

First order optimality conditions are stated next. Let us mention that the derivative of f_k at $\bar{u}_k$ in the direction $v = u - \bar{u}_k$ can be expressed as

$$f_k'(\bar{u}_k)(u - \bar{u}_k) = (\nu\bar{u}_k + \psi_2\bar{y}_k + \bar{p}_k, u - \bar{u}_k)_I.$$

Then we obtain:

Lemma 5.4.13. *Let $\bar{u}_k \in U_{ad}$ be a local solution of $(\mathbb{P}_k)$ in the sense of Definition 5.4.9. Then the following semi-discrete variational inequality holds:*

$$f_k'(\bar{u}_k)(u - \bar{u}_k) \geq 0 \quad \forall u \in U_{ad}. \tag{5.4.39}$$

Equivalently, $\bar{u}_k$ satisfies the projection formula

$$\bar{u}_k = P_{U_{ad}}\left(-\frac{1}{\nu}\bar{p}_k + \frac{\psi_2}{\nu}\bar{y}_k\right). \tag{5.4.40}$$

Proof. The proof is obtained by standard methods. □

Remark 5.4.14. *The projection formula* (5.4.40) *implies in particular that any locally optimal control $\bar{u}_k$ is piecewise constant in time.*

The following Lipschitz results are obtained as in the continuous case making use of the appropriate semidiscrete existence, boundedness, and stability results for the state from Theorem 5.4.2 and Theorem 5.4.5, the corresponding estimates for linearized and adjoint equations from Lemma 5.4.11 and Corollary 5.4.12, respectively, as well as the boundedness of U_{ad}. We obtain:

Lemma 5.4.15. *There exists a constant $C > 0$ independent of k such that*

$$\begin{aligned}||G_k(u_1) - G_k(u_2)||_I &\leq C\,||u_1 - u_2||_I\\ ||G_k'(u_1)v - G_k'(u_2)v||_I &\leq C\,||u_1 - u_2||_I||v||_I\end{aligned}$$

is fulfilled for all $u_1, u_2 \in U_{ad}$ and $v \in U$. Moreover, if $p_{k,1}, p_{k,2}$ denote the associated semidiscrete adjoint states solving (5.4.37) *we have the estimate*

$$||p_{k,1} - p_{k,2}||_I \leq C\,||u_1 - u_2||_I.$$

For completeness and later use, we can now supply a Lipschitz stability result f_k''.

Lemma 5.4.16. *There exists a constant $C > 0$ depending only on the data ψ_3 from the objective function and the nonlinearity d, such that for all $u_1, u_2 \in U_{ad}$ and all $v \in U$*

$$|f_k''(u_1)[v,v] - f_k''(u_2)[v,v]| \leq C\,||u_1 - u_2||_I||v||_I^2$$

is satisfied.

Proof. This follows exactly as in Lemma 5.3.11, utilizing the results for semidiscrete linearized and adjoint equations from Lemma 5.4.11 and Corollary 5.4.12, as well as the semidiscrete Lipschitz properties summarized in Lemma 5.4.15. □

5.4.4. Error estimates for semi-discrete uncontrolled equations

As a basis for our error estimates for the control in Section 5.7 we will now provide error estimates for the uncontrolled semidiscrete state equation, associated linearized equations, as well as the adjoint equation in the L^2-norm. Let therefore $y \in Y$ be the solution of the state equation (5.3.2) for a fixed

$u \in U$, and let $y_k \in Y_k^0$ be the solution of the corresponding semidiscretized equation (5.4.2). For our further analysis we emphasize that

$$\partial_t y, \nabla^2 y, \nabla^2 y_k \in L^2(Q),$$

is guaranteed by Proposition 2.5.6 and Theorem 5.4.5 in combination with Remark 5.4.6. We will estimate the temporal discretization error

$$e_k := y - y_k.$$

As a useful tool, we define the semidiscrete projection

$$\pi_k \colon \mathcal{C}(\bar{I}, H_0^1(\Omega)) \to Y_k^0 \quad \text{for } m = 1, 2, \dots, M$$

by

$$\pi_k y_{|I_m} \in \mathcal{P}_0(I_m, H_0^1(\Omega)), \quad \pi_k y(t_m) = y(t_m),$$

cf. [138]. Since Theorem 2.5.6 guarantees $y \in \mathcal{C}(\bar{I}, H_0^1(\Omega))$ we can apply the projection to the continuous state y. For short notation in the following proofs, we introduce the abbreviations

$$\eta_k := y - \pi_k y, \quad \xi_k := \pi_k y - y_k$$

and split the error

$$e_k = \eta_k + \xi_k$$

into a projection error η_k and a remainder ξ_k. We summarize some known auxiliary results that will be helpful in the error analysis.

Lemma 5.4.17. *Let $u \in U_{ad}$ and the states $y = G(u)$ as well as $y_k = G_k(u)$ be given. The error $e_k = y - y_k$ satisfies*

$$B(e_k, \varphi) = -(d(\cdot, y) - d(\cdot, y_k), \varphi)_I \quad \forall \varphi \in Y_k^0. \tag{5.4.41}$$

Proof. This is a standard analogue to Galerkin orthogonality for linear equations. By a density argument, it is clear that the exact solution $y = y(u) \in Y$ satisfies

$$B(y, \varphi) + (d(\cdot, y), \varphi)_I = (u, \varphi)_I + (y_0, \varphi_1) \quad \forall \varphi \in Y_k^0.$$

By definition, y_k fulfills

$$B(y_k, \varphi) + (d(\cdot, y_k), \varphi)_I = (u, \varphi)_I + (y_0, \varphi_1) \quad \forall \varphi \in Y_k^0,$$

and the assertion follows by subtraction. □

We will make use of (5.4.41) quite frequently in the sequel.

Lemma 5.4.18 ([99, 101]). *For the projection error η_k, the identity*

$$B(\eta_k, \varphi) = (\nabla \eta_k, \nabla \varphi)_I$$

holds for all $\varphi \in Y_k^0$. Moreover, the estimate

$$\|\eta_k\|_{I_m} \le c k_m \|\partial_t y\|_{I_m}$$

is satisfied for a constant $c > 0$ not depending on the discretization parameter k_m.

Proof. The first identity follows from straight-forward calculations utilizing the properties of π_k. The projection error estimate is a conclusion from the Bramble-Hilbert Lemma, cf. [138]. □

Lemma 5.4.19. *For all $v \in Y_k^0$, the following estimate is satisfied:*

$$\|v\|_{\infty,2}^2 + \|\nabla v\|_I^2 \leq B(v,v).$$

Proof. We refer to the proof of Lemma 5.7 in [101], where an estimate for the term $\|\nabla v\|_I^2$ was shown, which was sufficient for the linear setting. We follow these steps and add

$$B(v,v) = (\nabla v, \nabla v)_I + \sum_{m=1}^{M-1} ([v]_m, v_m^+) + (v_0^+, v_0^+) \tag{5.4.42}$$

$$B(v,v) = (\nabla v, \nabla v)_I + \sum_{m=1}^{M-1} (-v_m^-, [v]_m) + (v_M^-, v_M^-) \tag{5.4.43}$$

which is valid for $v \in Y_k^0$, cf. (5.4.5) on page 128. We obtain

$$B(v,v) = \|\nabla v\|_I^2 + \frac{1}{2}\left(\sum_{m=1}^{M-1} \|[v]_m\|^2 + \|v_0^+\|^2 + \|v_M^-\|^2\right),$$

since $v_m^+ - v_m^- = [v_m]$. Now, the assertion is obtained similar to the proof of Theorem 5.4.5. □

The error estimate due to time-discretization of the state equation is now obtained following ideas that are known for elliptic problems, cf. [23].

Theorem 5.4.20. *For the error $e_k := y - y_k$ between the continuous solution $y \in Y$ of* (5.3.2) *and the dG(0) semidiscretized solution $y_k \in Y_k^0$ of* (5.4.2) *the estimate*

$$\|e_k\|_I \leq Ck\|\partial_t y\|_I$$

holds with a constant C that is independent of the size of the time steps k.

Proof. As in [23] for elliptic problems, we introduce a function $\tilde{d}$ defined by

$$\tilde{d}(t,x) = \begin{cases} \frac{d(t,x,y(t,x)) - d(t,x,y_k(t,x))}{y(t,x) - y_k(t,x)} & \text{if } y(t,x) \neq y_k(t,x) \\ 0 & \text{else.} \end{cases}$$

Note that $\|\tilde{d}\|_{\infty,\infty} \leq c$ for a $c > 0$ due to the boundedness of U_{ad}, the uniform boundedness of the states y and y_k, and the Lipschitz continuity of d. Then, we use known duality arguments and define $\tilde{p}_k \in Y_k^0$ to be the solution of the auxiliary dual equation

$$B(\varphi, \tilde{p}_k) + (\varphi, \tilde{d}\tilde{p}_k)_I = (\varphi, e_k)_I \quad \forall\, \varphi \in Y_k^0. \tag{5.4.44}$$

Testing (5.4.44) with $\varphi = \xi_k \in Y_k^0$ yields

$$(\xi_k, e_k)_I = B(\xi_k, \tilde{p}_k) + (\xi_k, \tilde{d}\tilde{p}_k)_I = B(e_k, \tilde{p}_k) - B(\eta_k, \tilde{p}_k) + (e_k - \eta_k, \tilde{d}\tilde{p}_k)_I. \tag{5.4.45}$$

To this, we apply (5.4.41) from Lemma 5.4.17 and arrive at

$$(\xi_k, e_k)_I = -B(\eta_k, \tilde{p}_k) - (d(\cdot, y) - d(\cdot, y_k), \tilde{p}_k)_I + (e_k, \tilde{d}\tilde{p}_k)_I - (\eta_k, \tilde{d}\tilde{p}_k)_I. \tag{5.4.46}$$

Since we have $\tilde{p}_k \in Y_k^0$ we can apply Lemma 5.4.18 to (5.4.46) and obtain

$$(\xi_k, e_k)_I = -(\nabla \eta_k, \nabla \tilde{p}_k)_I - (d(\cdot, y) - d(\cdot, y_k), \tilde{p}_k)_I + (e_k, \tilde{d}\tilde{p}_k)_I - (\eta_k, \tilde{d}\tilde{p}_k)_I. \tag{5.4.47}$$

With the definition of $\tilde{d}$ this simplifies further to

$$(\xi_k, e_k)_I = -(\nabla \eta_k, \nabla \tilde{p}_k) - (\eta_k, \tilde{d}\tilde{p}_k)_I = (\eta_k, \Delta \tilde{p}_k)_I - (\eta_k, \tilde{d}\tilde{p}_k)_I, \tag{5.4.48}$$

where the last equality follows by integration by parts in space since $\eta_k = y - \pi_k y$ fulfills homogeneous Dirichlet boundary conditions. Then, we obtain

$$\|e_k\|_I^2 = (\xi_k, e_k)_I + (\eta_k, e_k)_I \le \|\eta_k\|_I \|\Delta \tilde{p}_k\|_I + \|\tilde{d}\|_{\infty,\infty} \|\eta_k\|_I \|\tilde{p}_k\|_I + \|\eta_k\|_I \|e_k\|_I. \tag{5.4.49}$$

The stability estimate for semidiscrete adjoint equations from Corollary 5.4.12 and the well-known projection error estimate from Lemma 5.4.18 guarantee

$$\|e_k\|_I^2 \le c\|\eta_k\|_I \|e_k\|_I \le ck\|\partial_t y\|_I \|e_k\|_I, \tag{5.4.50}$$

and the assertion is obtained after dividing by $\|e_k\|_I$. □

We prove a similar estimate for linearized equations and adjoint equations. We have to take into account that the derivative of d with respect to y is evaluated at $y = G(u)$ on the continuous level and at $y_k = G_k(u)$ at the semi-discrete level. We have already encountered a similar situation for the semi-infinite elliptic problem in Chapter 3, cf. the proof of Proposition 3.3.10.

Lemma 5.4.21. *For fixed $u \in U_{ad}$ let $y = G(u) \in Y$ be given by the solutions of the state equation* (5.3.2) *and let $\tilde{u} := G'(u)v$ for $v \in U$. Moreover, let $y_k \in Y_k^0$ be determined as the solution of the semidiscrete state equation* (5.4.2)*, and let $\tilde{y}_k \in Y_k^0$ denote the solution $\tilde{y}_k := G_k'(u)v$. Then the error estimate*

$$\|\tilde{y} - \tilde{y}_k\|_I \le Ck\|v\|_I$$

is satisfied for a constant $C > 0$ independent of the time discretization parameter k and the control $u \in U_{ad}$.

Proof. On the continuous level, we define the auxiliary equations

$$\int_0^T \langle \partial_t \tilde{z}, \varphi \rangle_{V^*,V}\, dt + (\nabla \tilde{z}, \nabla \varphi)_I + (\partial_y d(\cdot, y_k)\tilde{z}, \varphi)_I = (v, \varphi)_I \quad \forall \varphi \in Y, \quad \tilde{z}(0,\cdot) = 0, \tag{5.4.51}$$

as well as

$$-\int_0^T \langle \varphi, \partial_t \tilde{p} \rangle_{V,V^*}\, dt + (\nabla \varphi, \nabla \tilde{p})_I + (\varphi, \partial_y d(\cdot, y_k)\tilde{p})_I = (\varphi, \tilde{y} - \tilde{z})_I \quad \forall \varphi \in Y, \quad \tilde{p}(T,\cdot) = 0 \tag{5.4.52}$$

with solutions $\tilde{z}, \tilde{p} \in Y$, respectively. Now we split the error $\|\tilde{y} - \tilde{y}_k\|_I$ into

$$\|\tilde{y} - \tilde{y}_k\|_I \le \|\tilde{y} - \tilde{z}\|_I + \|\tilde{z} - \tilde{y}_k\|_I, \tag{5.4.53}$$

where the first term indicates the error due to linearization at different states y and y_k, and the second term is a pure discretization error. Since $\tilde{y} = G'(u)v$ fulfills

$$\int_0^T \langle \partial_t \tilde{y}, \varphi \rangle_{V^*,V}\, dt + (\nabla \tilde{y}, \nabla \varphi)_I + (\partial_y d(\cdot, y)\tilde{y}, \varphi)_I = (v, \varphi)_I \quad \forall \varphi \in Y, \quad \tilde{y}(0,\cdot) = 0 \tag{5.4.54}$$

with $y = G(u)$ we obtain for the first term in (5.4.53) that

$$\begin{aligned} \int_0^T \langle \partial_t(\tilde{y}-\tilde{z}),\varphi\rangle_{V^*,V} + (\nabla(\tilde{y}-\tilde{z}),\nabla\varphi)_I + (\partial_y d(\cdot,y)(\tilde{y}-\tilde{z}),\varphi)_I &= ((\partial_y d(\cdot,y_k)-\partial_y d(\cdot,y))\tilde{z},\varphi)_I, \\ (\tilde{y}-\tilde{z})(0,\cdot) &= 0 \end{aligned} \tag{5.4.55}$$

is satisfied for all $\varphi \in Y$. Testing (5.4.52) with $\varphi = \tilde{y} - \tilde{z}$ and integration by parts in time yields

$$\begin{aligned} \|\tilde{y}-\tilde{z}\|_I^2 &= -\int_0^T \langle \tilde{y}-\tilde{z},\partial_t\tilde{p}\rangle_{V,V^*}\,dt + (\nabla(\tilde{y}-\tilde{z}),\nabla\tilde{p})_I + (\tilde{y}-\tilde{z},\partial_y d(\cdot,y_k)\tilde{p})_I \\ &= \int_0^T \langle \partial_t(\tilde{y}-\tilde{z}),\tilde{p}\rangle_{V^*,V}\,dt + (\nabla(\tilde{y}-\tilde{z}),\nabla\tilde{p})_I + (\tilde{y}-\tilde{z},\partial_y d(\cdot,y_k)\tilde{p})_I. \end{aligned} \tag{5.4.56}$$

On the other hand, testing (5.4.55) with $\varphi = \tilde{p}$ leads to

$$\int_0^T \langle \partial_t(\tilde{y}-\tilde{z}),\tilde{p}\rangle_{V^*,V}\,dt + (\nabla(\tilde{y}-\tilde{z}),\nabla\tilde{p})_I + (\partial_y d(\cdot,y)(\tilde{y}-\tilde{z}),\tilde{p})_I = ((\partial_y d(\cdot,y_k)-\partial_y d(\cdot,y))\tilde{z},\tilde{p})_I. \tag{5.4.57}$$

Combining (5.4.56) and (5.4.57) yields

$$\|\tilde{y}-\tilde{z}\|_I^2 = -(\partial_y d(\cdot,y)(\tilde{y}-\tilde{z}),\tilde{p})_I + ((\partial_y d(\cdot,y_k)-\partial_y d(\cdot,y))\tilde{z},\tilde{p})_I + (\tilde{y}-\tilde{z},\partial_y d(\cdot,y_k)\tilde{p})_I, \tag{5.4.58}$$

which simplifies to

$$\|\tilde{y}-\tilde{z}\|_I^2 = ((\partial_y d(\cdot,y_k)-\partial_y d(\cdot,y))\,\tilde{y},\tilde{p})_I. \tag{5.4.59}$$

This can be estimated as

$$\|\tilde{y}-\tilde{z}\|_I^2 \le \|\partial_y d(\cdot,y_k)-\partial_y d(\cdot,y)\|_I\|\tilde{y}\|_{L^4(Q)}\|\tilde{p}\|_{L^4(Q)}, \tag{5.4.60}$$

and by the embedding $L^\infty(I,V) \hookrightarrow L^4(Q)$ and the Lipschitz continuity of $\partial_y d$ as well as the boundedness of U_{ad} we arrive at

$$\|\tilde{y}-\tilde{z}\|_I^2 \le c\,\|y-y_k\|_I\|\tilde{y}\|_{L^\infty(I,V)}\|\tilde{p}\|_{L^\infty(I,V)}.$$

The error estimate for the states from Theorem 5.4.20 as well as the stability estimates for linearized and adjoint equations from Propositions 2.5.2 and 2.5.7 deliver the estimate

$$\|\tilde{y}-\tilde{z}\|_I \le ck\|\partial_t y\|_I\|v\|_I,$$

and using Theorem 2.5.6 to estimate $\|\partial_t y\|_I$ we obtain

$$\|\tilde{y}-\tilde{z}\|_I \le ck\|v\|_I$$

by the boundedness of U_{ad}. To estimate the second term in (5.4.53), we can apply the discretization error estimates from [101] with only minor adaptation due to the term $\partial_y d(\cdot,y_k(u))$, where again y_k is uniformly bounded independent of the discretization parameter. Alternatively, the discretization error estimate for semilinear equations from Theorem 5.4.20 can be applied. □

It remains to estimate the error for the adjoint equation.

Lemma 5.4.22. *For fixed $u \in U_{ad}$ let $y \in Y$ and $p \in Y$ be given by the solutions of the state equation* (5.3.2) *and the adjoint equation* (5.3.5)*, respectively. Moreover, let $y_k \in Y_k^0$ and $p_k \in Y_k^0$ be determined as solutions of the semidiscrete state equation* (5.4.2) *and adjoint equation* (5.4.37)*. Then the error estimate*

$$\|p-p_k\|_I \le Ck\,(\|\partial_t y\|_I + \|\partial_t p\|_I)$$

is satisfied for a constant $C > 0$ independent of the time discretization parameter k and independent of $u \in U_{ad}$.

Proof. The proof is similar to the one of Lemma 5.4.21, introducing an auxiliary adjoint state $\hat{p} \in Y$, defined as the solution of

$$\begin{aligned} -\int_0^T \langle \varphi, \partial_t \hat{p} \rangle_{V,V^*} \, dt + (\nabla \varphi, \nabla \hat{p})_I + (\varphi, \partial_y d(\cdot, y_k(u))\hat{p})_I &= (\varphi, \psi_1 y_k + \psi_2 u + \psi_3)_I \quad \forall \varphi \in Y, \\ \hat{p}(T, \cdot) &= 0. \end{aligned}$$

We split the error into

$$\|p - p_k\|_I \le \|p - \hat{p}\|_I + \|\hat{p} - p_k\|_I.$$

The term $\|p - \hat{p}\|_I$ can be estimated with similar steps as in the proof of Lemma 5.4.21, making use of Theorem 5.4.20, Proposition 2.5.7, and the boundedness of U_{ad}. We also refer to [118] for further details. The second term to estimate is again a pure discretization error and accounts for the term $\|\partial_t p\|$. This is obtained with arguments similar to [101]. □

5.4.5. Semidiscrete coercivity

Before discussing the spatial discretization of the state equation we state a semidiscrete analogue to Corollary 5.3.13. We prove a coercivity result for f_k'' in the neighborhood of $\bar{u}$, which will later be extended to a coercivity result in the neighborhood of a local semidiscrete solution $\bar{u}_k$ approximating $\bar{u}$.

Lemma 5.4.23. *Let $\bar{u}$ be a local solution of* $(\mathbb{P})$ *and let Assumptions 5.3.1 and 5.3.9 be valid. Then there exists an $\varepsilon > 0$, such that for all $u \in U_{ad}$ with $\|u - \bar{u}\|_I \le \varepsilon$ and all $v \in L^\infty(Q)$*

$$f_k''(u)[v, v] \ge \frac{\alpha}{4}\|v\|_I^2$$

holds for k sufficiently small.

Proof. With the definition of f'' and f_k'' we obtain

$$\begin{aligned} |f_k''(u)[v,v] - f''(u)[v,v]| \le & \iint_Q \left| p \partial_{yy} d(\cdot, y)\tilde{y}^2 - p_k \partial_{yy} d(\cdot, y_k)\tilde{y}_k^2 \right| \, dxdt \\ & + \iint_Q \left| \psi_1(\tilde{y}_k^2 - \tilde{y}^2) + 2\psi_2 v(\tilde{y}_k - \tilde{y}) \right| dxdt \\ \le & \iint_Q |(\psi_1(\tilde{y}_k + \tilde{y}) + 2\psi_2 v)(\tilde{y}_k - \tilde{y})| \, dxdt \\ & + \iint_Q \left| (p - p_k)\partial_{yy} d(\cdot, y)\tilde{y}^2 \right| \, dxdt \\ & + \iint_Q |p_k \partial_{yy} d(\cdot, y)(\tilde{y} - \tilde{y}_k)(\tilde{y}_k + \tilde{y})| \, dxdt \\ & + \iint_Q \left| p_k(\partial_{yy} d(\cdot, y) - \partial_{yy} d(\cdot, y_k))\tilde{y}_k^2 \right| \, dxdt, \end{aligned}$$

similar to Lemma 5.3.11. The following steps are completely analogous to the proof of Lemma 5.4.16, making use of semidiscrete stability and discretization error estimates. Estimating the integrals by Hölder's inequality yields

$$\begin{aligned}|f_k''(u)[v,v] - f''(u)[v,v]| \leq\ & c(\|\tilde{y}_k\|_I + \|\tilde{y}\|_I + \|v\|_I)\|\tilde{y}_k - \tilde{y}\|_I \\ &+c\,\|\partial_{yy}d(\cdot,y_k)\|_{\infty,\infty}\|p_k - p\|_I\|\tilde{y}\|^2_{L^4(Q))} \\ &+c\,\|p_k\|_{\infty,\infty}\|\partial_{yy}d(\cdot,y)\|_{\infty,\infty}(\|\tilde{y}_k\|_I + \|\tilde{y}\|_I)\|\tilde{y}_k - \tilde{y}\|_I \\ &+c\|p_k\|_{\infty,\infty}\|y_k - y\|_I\|\tilde{y}_k\|^2_{L^4(Q)}.\end{aligned}$$

In the last line, we have used the uniform boundedness of y as well as y_k and p_k independent of k and u as well as the Lipschitz continuity of $\partial_{yy}d$. By the embedding

$$L^\infty(I, H_0^1(\Omega)) \hookrightarrow L^4(Q),$$

Proposition 2.5.2 and estimates (5.4.34) and (5.4.36) yield

$$\|\tilde{y}\|_{L^4(Q)} \leq c\,\|v\|_I, \quad \|\tilde{y}_k\|_{L^4(Q)} \leq c\,\|v\|_I,$$

as well as

$$\|\tilde{y}\|_I \leq c\,\|v\|_I, \quad \|\tilde{y}_k\|_I \leq c\,\|v\|_I.$$

Then, the boundedness of U_{ad} and $\partial_{yy}d$ as well as $y_k, p_k, \tilde{y}_k$ yield

$$|f_k''(u)[v,v] - f''(u)[v,v]| \leq c\|\tilde{y}_k - \tilde{y}\|_I\|v\|_I + c\,\|p_k - p\|_I\|v\|_I^2 + c\|y_k - y\|_I\|v\|_I^2. \tag{5.4.61}$$

Now, the discretization error estimates

$$\|y - y_k\|_I + \|p - p_k\|_I \leq c\,k, \quad \|\tilde{y} - \tilde{y}_k\|_I \leq c\,k\|v\|_I$$

from Theorem 5.4.20 and Lemmas 5.4.21 and 5.4.22 lead to

$$|f_k''(u)[v,v] - f''(u)[v,v]| \leq c\,k\|v\|_I^2.$$

Since this tends to zero as k tends to zero, the assertion is obtained by Corollary 5.3.13, observing that

$$f_k''(u)[v,v] = f''(u)[v,v] + f_k''(u)[v,v] - f''(u)[v,v] \geq \frac{\alpha}{2}\|v\|_I^2 - ck\|v\|_I^2.$$

□

5.5. Discretization of the state equation in space

On the second level of discretization we discretize the state equation in space. Here, we use usual conforming finite elements. We proceed similar to the last section. After a brief introduction to the discretization, we will basically discuss the same steps as before. Special emphasis will be placed on a boundedness result for discrete solutions of the state equation, and on error estimates for uncontrolled equations due to the spatial discretization.

5.5.1. Preliminaries

Following the lines of [102], we consider here two-dimensional shape regular meshes consisting of nonoverlapping quadrilateral cells K. We refer to e.g., [34] for the definition of regular meshes. We denote the mesh by $\mathcal{T}_h = \{K\}$ and define the discretization parameter $h\colon \Omega \to \mathbb{R}$ as a cellwise constant function by setting $h_{|_K} = h_K$ with the diameter h_K of the cell K. We use the symbol h also for the maximal cell size, i.e.,

$$h = \max h_K.$$

We benefit from the fact that Ω is polygonally bounded.

On the mesh $\mathcal{T}_h$ we construct a conforming finite element space $V_h \subset H_0^1(\Omega)$ in a standard way. Note that in contrast to Chapter 3 the use of quadrilateral cells instead of triangles makes it necessary to consider bilinear functions. We define the discrete space

$$V_h = \{v \in H_0^1(\Omega) | v_{|_K} \in \mathcal{Q}_1(K) \text{ for } K \in \mathcal{T}_h\},$$

where $\mathcal{Q}_1(K)$ consists of shape functions obtained via bilinear transformations of polynomials in $\widehat{\mathcal{Q}_1}(\hat{K})$ defined on the reference cell $\hat{K} = (0,1)^2$; cf. also Section 3.2 in [101]. Then, we utilize the space-time finite element space

$$Y_{k,h}^{0,1} = \{v_{kh} \in L^2(I, V_h) | v_{kh}{}_{|_{I_m}} \in \mathcal{P}_0(I_m, V_h),\ m = 1, 2, \ldots, M\} \subset Y_k^0.$$

The so-called cG(1)dG(0) discretization of the state equation for given control $u \in U$ has the following form:

Find a state $y_{kh} = y_{kh}(u) \in Y_{k,h}^{0,1}$ such that

$$B(y_{kh}, \varphi) + (d(\cdot, y_{kh}), \varphi)_I = (u, \varphi)_I + (y_0, \varphi_1) \quad \forall \varphi \in Y_{k,h}^{0,1}. \tag{5.5.1}$$

5.5.2. Existence results and stability estimates for the discrete state equation

We are first concerned with the existence of a solution y_{kh} of the fully discrete state equation for all $u \in U$ and $y_0 \in H_0^1(\Omega)$.

Theorem 5.5.1. *Let Assumption 5.2.1 be satisfied. Then, for each $u \in U$ and $y_0 \in H_0^1(\Omega) \cap L^\infty(\Omega)$, there exists a unique solution $y_{kh} \in Y_{k,h}^{0,1} \cap L^\infty(Q)$ of equation* (5.5.1).

Proof. The existence of y_{kh} is a consequence of Brouwer's fixed point theorem using the monotonicity and Lipschitz continuity of d; we refer to Appendix A.1 for more details. The uniqueness follows from the monotonicity of d with respect to y by assumption of the contrary. □

In a second step, we will again need to supply a boundedness result for y_{kh} that is independent of the discretization parameters. Noting that error estimates for the semidiscrete equations in the $L^\infty(I, L^2(\Omega))$-norm appeared in a quite natural way, it is now promising to obtain boundedness of the discrete solutions by an inverse estimate between the norms on $L^\infty(\Omega)$ and $L^2(\Omega)$. In fact, a direct extension of Stampacchia's method to spatially discrete problems is not possible, since test functions $v_{kh} = y_{kh} - b$ are not generally functions in $Y_{k,h}^{0,1}$. By means of the spatial L^2-projection

$$\Pi_h\colon H_0^1(\Omega) \to V_h$$

we define the projection

$$\pi_h : Y_k^0 \to Y_{k,h}^{0,1}$$

pointwise in time as

$$(\pi_h y_k)(t) = \Pi_h y_k(t).$$

The projection π_h is applicable to the semidiscrete state $y_k = G_k(u)$. Therefore, we define $\eta_h \in Y_k^0$ as well as $\xi_h \in Y_{k,h}^{0,1}$ by

$$\eta_h := y_k - \pi_h y_k, \quad \xi_h := \pi_h y_k - y_{kh}. \tag{5.5.2}$$

We emphasize that

$$\nabla y_k \in L^\infty(I, L^2(\Omega)), \quad \nabla^2 y_k \in L^2(I, L^2(\Omega))$$

by Theorem 5.4.5 and Remark 5.4.6. Note that well known error estimates for the spatial L^2-projection allow to estimate

$$\begin{aligned} \|v - \Pi_h v\| + h\|\nabla(v - \Pi_h v)\| &\le ch^2\|\nabla^2 v\| \\ \|v - \Pi_h v\| &\le ch\|\nabla v\| \end{aligned}$$

for all $v \in H^2(\Omega)$, cf. for instance [34, Chapter 3]. Then, the estimates

$$\|\eta_h\|_I + h\|\nabla \eta_h\|_I \le ch^2\|\nabla^2 y_k\|_I, \tag{5.5.3}$$

$$\|\eta_h\|_{\infty,2} \le ch\|\nabla y_k\|_{\infty,2} \tag{5.5.4}$$

are fulfilled. We point out two additional helpful result. The first one is the inverse estimate

$$\|v_h\|_{L^\infty(\Omega)} \le ch^{-1}\|v_h\|_{L^2(\Omega)} \tag{5.5.5}$$

which follows from a more general result in [34] for functions $v_h \in V_h$. We will apply it to functions $v \in Y_{k,h}^{0,1}$ separately on each time interval I_m. The second helpful property is

$$B(y_k - y_{kh}, \varphi) + (d(\cdot, y_k) - d(\cdot, y_{kh}), \varphi) = 0 \quad \forall \varphi \in Y_{k,h}^{0,1}, \tag{5.5.6}$$

which is obtained as its semidiscrete analogue (5.4.41) from Lemma 5.4.17.

With these auxiliary results, we now prove a boundedness result for the discrete states y_{kh} that holds independent from k and h.

Theorem 5.5.2. *For the solution y_{kh} of (5.5.1), we have the following stability estimates: There is a constant $C > 0$ independent of k and h such that*

$$\|y_{kh}\|_{\infty,\infty} \le C\left(\|u\|_{L^p(Q)} + \|d(\cdot,0)\|_{L^p(Q)} + \|\nabla y_0\| + \|y_0\|_\infty\right)$$

is satisfied for $p > 2$.

Proof. We prove uniform boundedness following an idea that Vexler suggested in [118]. We first prove that $\pi_h y_k$ is uniformly bounded. We consider the pointwise in time interpolant

$$(i_h y_k)(t) = I_h y_k(t),$$

where I_h denotes a quasi-interpolant that is stable with respect to the L^∞-norm. Consider for instance the Clément interpolant, cf. [35]. Then, by the boundedness of y_k in $L^\infty(Q)$ as well as $\nabla y_k \in L^\infty(I, L^2(\Omega))$ due to Theorem 5.4.5, we obtain

$$\|\pi_h y_k\|_{\infty,\infty} \le \|i_h y_k\|_{\infty,\infty} + \|\pi_h y_k - i_h y_k\|_{\infty,\infty} \le \|y_k\|_{\infty,\infty} + ch^{-1}\|i_h y_k - \pi_h y_k\|_{\infty,2}, \tag{5.5.7}$$

where we used the inverse (5.5.5), applied to $v_h = (i_h y_k - \pi_h y_k)_{|I_m}$. Now, we estimate

$$\|i_h y_k - \pi_h y_k\|_{\infty,2} \le \|i_h y_k - y_k\|_{\infty,2} + \|\pi_h y_k - y_k\|_{\infty,2} \le ch\|\nabla y_k\|_{\infty,2} \tag{5.5.8}$$

where we used estimate (5.5.4), that also holds for i_h. Inserting (5.5.8) into (5.5.7) yields

$$\|\pi_h y_k\|_{\infty,\infty} \le c(\|y_k\|_{\infty,\infty} + \|\nabla y_k\|_{\infty,2}). \tag{5.5.9}$$

Now, using (5.5.6), we consider

$$\begin{aligned} B(\xi_h, \xi_h) &= B(y_k - y_{kh}, \xi_h) - B(\eta_h, \xi_h) \\ &= -B(\eta_h, \xi_h) - (d(\cdot, y_k) - d(\cdot, y_{kh}), \xi_h)_I \\ &= -B(\eta_h, \xi_h) - (d(\cdot, y_k) - d(\cdot, \pi_h y_k), \xi_h)_I - (d(\cdot, \pi_h y_k) - d(\cdot, y_{kh}), \xi_h)_I. \end{aligned}$$

Using Lemma 5.4.18 and the monotonicity and Lipschitz continuity of d, this leads to

$$\begin{aligned} B(\xi_h, \xi_h) &\le -(\nabla \eta_h, \nabla \xi_h)_I - (d(\cdot, y_k) - d(\cdot, \pi_h y_k), \xi_h)_I \\ &\le \|\nabla \eta_h\|_I \|\nabla \xi_h\|_I + c\|\eta_h\|_I \|\xi_h\|_I, \end{aligned}$$

where we used in particular the boundedness of y_k from Theorem 5.4.2 and the boundedness of $\pi_h y_k$ from (5.5.9). Combining Lemma 5.4.19 with the previous estimate yields

$$\|\xi_h\|_{\infty,2}^2 + \|\nabla \xi_h\|_I^2 \le c\,\|\nabla \eta_h\|_I^2 + c\,\|\eta_h\|_I^2 \tag{5.5.10}$$

by Young's inequality, and the fact that $\|\xi_h\|_I \le c\,\|\nabla \xi_h\|_I$. The error estimate (5.5.3) yields

$$\|\xi_h\|_{\infty,2}^2 + \|\nabla \xi_h\|_I^2 \le ch^2\|\nabla^2 y_k\|_I + ch^4\|\nabla^2 y_k\|_I \le ch^2\|\nabla^2 y_k\|_I.$$

Applying again the inverse estimate (5.5.5), we obtain

$$\|\xi_h\|_{\infty,\infty} \le c\|\nabla^2 y_k\|_I,$$

which together with (5.5.9) implies

$$\|y_{kh}\|_{\infty,\infty} = \|y_{kh} - \pi_h y_k + \pi_h y_k\|_{\infty,\infty} \le \|\xi_h\|_{\infty,\infty} + \|\pi_h y_k\|_{\infty,\infty} \le c(\|y_k\|_{\infty,\infty} + \|\nabla y_k\|_{\infty,2} + \|\nabla^2 y_k\|_I).$$

Theorems 5.4.5 and 5.4.2 then imply the assertion. □

Remark 5.5.3. *In the previous proof, estimates* (5.5.7) *and* (5.5.8) *show the reason for the restrictions on the spatial dimension. For space dimension* $n > 2$, h^{-1} *is not sufficient to estimate the* $L^\infty(Q)$ *norm against the* $L^\infty(I, L^2(\Omega))$ *norm. Then, higher order error estimates for* $\|i_h y_k - \pi_h y_k\|_{\infty,2}$ *would have to be used, which in turn requires higher order spatial derivatives of* y_k *to be bounded in* $L^\infty(I, L^2(\Omega))$. *This is not guaranteed by the stability estimates from Theorem 5.5.5.*

Remark 5.5.4. *Analogously to Remark 5.4.4, note that Theorem 5.5.2 implies in particular that all discrete states* y_{kh} *belonging to controls* $u \in U_{ad}$ *are uniformly bounded in* $L^\infty(Q)$ *by a constant* $C > 0$ *independent of* k, h, *and* u, *since* U_{ad} *is bounded.*

Before we continue, let us define the discrete Laplace operator by

$$(\Delta_h y, \varphi) = -(\nabla y, \nabla \varphi) \quad \forall \varphi \in V_h.$$

Theorem 5.5.5. *Let the conditions of Theorem 5.5.2 be satisfied. The discrete solution y_{kh} satisfies the stability estimate*

$$\|y_{kh}\|_I^2 + \|y_{kh}\|_{\infty,2}^2 + \|\nabla y_{kh}\|_I^2 + \|\nabla y_{kh}\|_{\infty,2}^2 + \|\Delta_h y_{kh}\|_I^2 + \sum_{m=1}^{M} k_m^{-1} \|[y_{kh}]_{m-1}\|^2 \\ \leq C \left(\|u\|_I^2 + \|d(\cdot,0)\|_I^2 + \|\nabla \Pi_h y_0\|^2 \right)$$

for a constant $c > 0$ that does not depend on the discretization parameters k and h.

Proof. With the uniform boundedness result of Theorem 5.5.2, the assertion follows as in the semi-discrete setting of Theorem 5.4.5. □

5.5.3. The optimal control problem with discrete state equation

We now proceed as in the semidiscrete setting and summarize all necessary results for the optimal control problem with discrete state equation. For brevity, we will sometimes speak of a discrete problem or even a discrete solution, but we point out that at this time the controls are still not discretized.

Definition 5.5.6. *The mapping*

$$G_{kh} \colon U \to \cap Y_{k,h}^{0,1}, \quad u \mapsto y_{kh} = G_{kh}(u),$$

where y_{kh} solves (5.5.1)*, is called the discrete control-to-state-operator associated with Problem* (5.3.1)*. In addition, we introduce the discrete reduced objective functional*

$$f_{kh} \colon U \to \mathbb{R}, \quad f_{kh}(u) := J(G_{kh}(u), u).$$

The discrete problem formulation now reads

$$\text{Minimize } f_{kh}(u) \quad u \in U_{\text{ad}}. \tag{$\mathbb{P}_{kh}$}$$

We agree to denote a (local) solution of $(\mathbb{P}_{kh})$ by $\bar{u}_{kh}$ to indicate the discretization of the state equation and to distinguish it from the optimal solutions on other levels of discretization. The existence of at least one optimal solution $\bar{u}_{kh} \in U_{\text{ad}}$ is again easily obtained, since U_{ad} is not empty.

Theorem 5.5.7. *The discrete optimal control problem $(\mathbb{P}_{kh})$ admits at least one (globally) optimal control $\bar{u}_{kh} \in U_{ad}$ with associated optimal discrete state $\bar{y}_{kh} = G_{kh}(\bar{u}_{kh})$.*

Proof. We again refer to [144] for the technique of proof. □

To be complete, we define a discrete local solution.

Definition 5.5.8. *A control $\bar{u}_{kh} \in U_{ad}$ is called a discrete local solution to $(\mathbb{P}_{kh})$ in the sense of $L^2(Q)$, if there exists an $\varepsilon > 0$ such that for all $u \in U_{ad}$ with $\|u - \bar{u}_{kh}\|_I \leq \varepsilon$*

$$f_{kh}(u) \geq f_{kh}(\bar{u}_{kh})$$

is satisfied.

Proposition 5.5.9. *The control-to-state-operator G_{kh} is of class $\mathcal{C}^2$. Let $v, v_1, v_2 \in U$ and $y_{kh} = G_{kh}(u)$ be given. Then, its first and second order derivatives are given by $\tilde{y}_{kh} := G'_{kh}(u)v$ and $\tilde{z}_{kh} := G''_{kh}(u)v_1v_2$ being the solutions of* (5.4.32) *and* (5.4.33) *for test functions in $Y^{0,1}_{k,h}$. For the reduced objective function, we obtain*

$$f'_{kh}(u)v = \iint_Q (\psi_1 \tilde{y}_{kh} y_{kh} + \psi_2 \tilde{y}_{kh} u + \psi_2 y_{kh} v + \psi_3 \tilde{y}_{kh} + \nu u v)\, dx dt$$

$$f''_{kh}(u)[v_1, v_2] = \iint_Q (\psi_1 \tilde{y}_{kh,1} \tilde{y}_{kh,2} + \psi_1 z_{kh} y_{kh} + \psi_2 \tilde{y}_{kh,2} v_2 + \psi_2 z_{kh} u + \psi_2 \tilde{y}_{kh,2} v_1 + \psi_3 z_{kh} + \nu v_1 v_2)\, dx\, dt,$$

Proof. The proof follows as in Theorems 5.9, 5.15, and 5.16 in [144] for the control-to-state operator. □

Again, it is quite obvious that the stability estimates of Theorems 5.5.2 and 5.5.5 are also valid for linearized semidiscrete state equations.

Lemma 5.5.10. *Let $u \in U_{ad}$ and $v \in U$ be given and denote by $\tilde{y}_{kh}$ the discrete linearized state $\tilde{y}_{kh} = G'_{kh}(u)v$. Then there exists a constant $c > 0$ independent of k and h as well as $u \in U_{ad}$ such that*

$$\|\tilde{y}_{kh}\|_I \le c\, \|v\|_I \tag{5.5.11}$$

$$\|\tilde{y}_{kh}\|_{\infty,\infty} \le c\, \|v\|_{L^p(Q)} \tag{5.5.12}$$

$$\|\tilde{y}_{kh}\|_{\infty,2} + \|\nabla \tilde{y}_{kh}\|_{\infty,2} \le c\, \|v\|_I \tag{5.5.13}$$

holds for any $p > 2$.

For $u \in U$ and $y_{kh} = G_{kh}(u)$, the discrete adjoint state $p_{kh} = p_{kh}(u) \in Y^{0,1}_{k,h}$ is the solution of the discrete adjoint equation

$$B(\varphi, p_{kh}) + (\varphi, \partial_y d(\cdot, y_{kh}) p_{kh}) = (\varphi, \psi_1 y_{kh} + \psi_2 u)_I \quad \forall \varphi \in Y^{0,1}_{k,h}. \tag{5.5.14}$$

We can show existence as well as stability estimates for the adjoint equation. We denote by $\bar{p}_{kh}$ the adjoint state associated with the locally optimal control $\bar{u}_{kh}$ and its corresponding optimal state $\bar{y}_{kh}$. For a more general right-hand side $g \in L^\infty(Q)$ and terminal condition $p_T \in H^1_0(\Omega)$ we consider the discrete dual equation

$$B(\varphi, p_{kh}) + (\varphi, \partial_y d(\cdot, y_{kh}) p_{kh})_I = (\varphi, g)_I + (\varphi_M, p_T) \quad \forall\, \varphi \in Y^{0,1}_{k,h}. \tag{5.5.15}$$

and obtain the following result applicable to (5.5.14).

Corollary 5.5.11. *For the solution $p_{kh} \in Y^{0,1}_{k,h}$ of the discrete dual equation* (5.5.14) *with right-hand side $g \in L^2(I, H)$ and terminal condition $p_T \in H^1_0(\Omega)$, the estimate*

$$\|p_{kh}\|_I^2 + \|\nabla p_{kh}\|_I^2 + \|\Delta_h p_{kh}\|_I^2 + \sum_{m=1}^{M} k_m^{-1} \|[p_{kh}]_m\|^2 \le C\{\|g\|_I^2 + \|\Pi_h \nabla p_T\|^2\}$$

holds. Here, the jump term $[p_{kh}]_M$ at $t = T$ is defined as $p_T - p^-_{kh,M}$.

The first order necessary optimality conditions for a discrete locally optimal solution are now obtained in the form we are already familiar with from the continuous and semidiscrete levels.

Lemma 5.5.12. *Let $\bar{u}_{kh} \in U_{ad}$ be a local solution of $(\mathbb{P}_{kh})$ in the sense of Definition 5.5.8. Then the following discrete variational inequality holds:*

$$f'_{kh}(\bar{u}_{kh})(u - \bar{u}_{kh}) \geq 0 \quad \forall u \in U_{ad}, \tag{5.5.16}$$

Equivalently, $\bar{u}_{kh}$ satisfies the projection formula

$$\bar{u}_{kh} = P_{U_{ad}}\left(-\frac{1}{\nu}\bar{p}_{kh} + \frac{\psi_2}{\nu}\bar{y}_{kh}\right). \tag{5.5.17}$$

Proof. The proof is standard, we refer, again e. g., to [144]. □

We infer that $\bar{u}_{kh}$ is piecewise constant in time, but not necessarily cellwise bilinear in space.

For future reference, let us also state a Lipschitz continuity result for G_{kh} and its derivatives, which can be shown as in the semidiscrete setting of Lemma 5.4.15.

Lemma 5.5.13. *Let $u_1, u_2 \in U_{ad}$ and $v \in U$ be given. Then there exists a constant $C > 0$ independent of k and h such that*

$$\begin{aligned} \|G_{kh}(u_1) - G_{kh}(u_2)\|_I &\leq C\, \|u_1 - u_2\|_I \\ \|G'_{kh}(u_1)v - G'_{kh}(u_2)v\|_I &\leq C\, \|u_1 - u_2\|_I \|v\|_I \end{aligned}$$

is fulfilled.

The following Lipschitz result for the discrete objective function is equally straight-forward to prove.

Lemma 5.5.14. *The first derivative of f_{kh} fulfills a Lipschitz condition, i. e. there exists a constant $C > 0$ such that for all $u_1, u_2 \in U_{ad}$ and all $v \in U$*

$$|f'_{kh}(u_1)v - f'_{kh}(u_2)v| \leq C\, \|u_1 - u_2\|_I \|v\|_I$$

is satisfied.

Proof. This follows in a straight-forward manner, making use of Lipschitz results for the state and linearized state equation from Lemma 5.5.13. □

5.5.4. Error estimates for discrete uncontrolled equations

We now prove error estimates for the uncontrolled discrete state and adjoint equations in the L^2-norm. Analogously to the semidiscrete level, we define the error e_h due to spatial discretization by

$$e_h := y_k - y_{kh},$$

and again decompose it into a projection error and a remainder, term, i.e.

$$e_h = \eta_h + \xi_h,$$

where η_h and ξ_h are define in (5.5.2) on page 145. Again, we follow [23]. We define $\hat{p}_k \in Y_k^0$ as the solution of the auxiliary semidiscrete dual equation

$$B(\varphi, \hat{p}_k) + (\varphi, \hat{d}\hat{p}_k)_I = (\varphi, e_h)_I \quad \forall\, \varphi \in Y_k^0, \tag{5.5.18}$$

with

$$\hat{d}(t,x) = \begin{cases} \frac{d(t,x,y_k(t,x)) - d(t,x,y_{kh}(t,x))}{y_k(t,x) - y_{kh}(t,x)} & \text{if } y_k(t,x) \neq y_{kh}(t,x) \\ 0 & \text{else.} \end{cases}$$

In addition to the abbreviations η_h and ξ_h we introduce the short notation

$$\eta_h^* := \hat{p}_k - \pi_h \hat{p}_k.$$

and make use of the statements

$$\begin{aligned} B(\xi_h, \eta_h^*) &= (\nabla \xi_h, \nabla \eta_h^*), && (5.5.19)\\ B(\eta_h, \eta_h^*) &\leq \|\nabla \eta_h\|_I \|\nabla \eta_h^*\|_I + c\, \|\eta_h\|_I \|e_h\|_I, && (5.5.20) \end{aligned}$$

that have been proven in [101, Lemmas 5.7, 5.8] by direct calculations. Our main error estimate for the discrete state equation is now easily obtained.

Theorem 5.5.15. *For the error $e_h := y_k - y_{kh}$ between the dG(0) semidiscretized solution $y_k \in Y_k^0$ of (5.4.2) and the fully cG(1)dG(0) discretized solution $y_{kh} \in Y_{k,h}^{0,1}$ of (5.5.1), we have the error estimate*

$$\|e_h\|_I \leq C h^2 \|\nabla^2 y_k\|_I,$$

where the constant C is independent of the mesh size h and the size of the time steps k.

Proof. Testing (5.5.18) with $\varphi = e_h \in Y_k^0$ yields

$$\begin{aligned} \|e_h\|_I^2 &= B(e_h, \hat{p}_k) + (e_h, \hat{d}\hat{p}_k)_I \\ &= B(e_h, \hat{p}_k - \pi_h \hat{p}_k) + B(e_h, \pi_h \hat{p}_k) + (e_h, \hat{d}(\hat{p}_k - \pi_h \hat{p}_k))_I + (e_h, \hat{d}\pi_h \hat{p}_k) \\ &= B(e_h, \eta_h^*) + B(e_h, \pi_h \hat{p}_k) + (e_h, \hat{d}\eta_h^*) + (e_h, \hat{d}\pi_h \hat{p}_k) \\ &= B(e_h, \eta_h^*) - (d(\cdot, y_k) - d(\cdot, y_{kh}), \pi_h \hat{p}_k) + (e_h, \hat{d}\eta_h^*) + (e_h, \hat{d}\pi_h \hat{p}_k), \end{aligned}$$

where the last equality follows from (5.5.6) by the fact that $\pi_h \hat{p}_k \in Y_{k,h}^{0,1}$. With the definition of $\hat{d}$, this simplifies further to

$$\|e_h\|_I^2 = B(e_h, \eta_h^*) + (e_h, \hat{d}\eta_h^*)_I = B(\xi_h, \eta_h^*) + B(\eta_h, \eta_h^*) + (e_h, \hat{d}\eta_h^*)_I. \tag{5.5.21}$$

With (5.5.19) and (5.5.20) we estimate (5.5.21) by

$$\|e_h\|_I^2 \leq \|\nabla \xi_h\|_I \|\nabla \eta_h^*\|_I + \|\nabla \eta_h\|_I \|\nabla \eta_h^*\|_I + c\, \|\eta_h\|_I \|e_h\|_I + c\, \|\xi_h\|_I \|\eta_h^*\|_I + c\, \|\eta_h\|_I \|\eta_h^*\|_I \tag{5.5.22}$$

since y_k, y_{kh}, and $\hat{d}$ are bounded in $L^\infty(Q)$. From Lemma 5.4.19 we also know that

$$\|\nabla \xi_h\|_I^2 + \|\xi_h\|_I^2 \leq B(\xi_h, \xi_h).$$

In the proof of Theorem 5.5.2, we have estimated $B(\xi_h, \xi_h)$ against projection error estimates, cf. (5.5.10), and we obtain

$$\|\nabla \xi_h\|_I^2 + \|\xi_h\|_I^2 \leq c \left(\|\nabla \eta_h\|_I^2 + \|\eta_h\|_I^2 \right).$$

Inserting this in (5.5.22), we obtain

$$\|e_h\|_I^2 \le c(\|\nabla\eta_h\|_I\|\nabla\eta_h^*\|_I + \|\eta_h\|_I\|\nabla\eta_h^*\|_I + \|\nabla\eta_h\|_I\|\eta_h^*\|_I + \|\eta_h\|_I\|\eta_h^*\|_I + \|\eta_h\|_I\|e_h\|_I). \tag{5.5.23}$$

Now observe that analogously to the error estimates (5.5.3) we have

$$\|\eta_h^*\|_I \le ch^2\|\nabla^2\hat{p}_k\|_I, \quad \|\nabla\eta_h^*\|_I \le ch\|\nabla^2\hat{p}_k\|_I.$$

Applying this and the error estimate (5.5.3) to (5.5.23) leads to

$$\|e_h\|_I^2 \le ch^2\|\nabla^2\hat{p}_k\|_I\|\nabla^2 y_k\|_I + ch^2\|\nabla^2 y_k\|_I\|e_h\|_I.$$

Now, with the same arguments as in Remark 5.4.6, we have

$$\|\nabla^2\hat{p}_k\|_I \le c\,\|\Delta\hat{p}_k\|_I \le c\,\|e_h\|_I,$$

due to Corollary 5.4.12, which inserted into (5.5.23) yields the assertion. □

As on the semidiscrete discretization level in Section 5.4.4 we also obtain error estimates for linearized states and adjoint states. We state them without proof.

Lemma 5.5.16. *For $u \in U_{ad}$ let $y_k \in Y_k^0$ be given by the solutions of the semidiscrete state equation* (5.4.2) *and and let $\tilde{y}_k := G_k'(u)v \in Y_k^0$ be the solution of the linearized state equation, respectively. Moreover, let $y_{kh} \in Y_{k,h}^{0,1}$ and $\tilde{y}_{kh} := G_{kh}'(u)v \in Y_{k,h}^{0,1}$ denote the solutions of the discrete state equation* (5.5.1) *and a corresponding linearized equation. Then the error estimate*

$$\|\tilde{y}_k - \tilde{y}_{kh}\|_I \le Ch^2\|v\|_I$$

is satisfied for a constant $C > 0$ independent of the space discretization.

Lemma 5.5.17. *For $u \in U_{ad}$ let $y_k \in Y_k^0$ and $p_k \in Y_k^0$ be given by the solutions of the semidiscrete state equation* (5.4.2) *and adjoint equation* (5.4.37)*, respectively. Moreover, let $y_{kh} \in Y_{k,h}^{0,1}$ and $p_{kh} \in Y_{k,h}^{0,1}$ denote the solutions of the discrete state equation* (5.5.1) *and adjoint equation* (5.5.14)*. Then the error estimate*

$$\|p_k - p_{kh}\|_I \le Ch^2\left(\|\nabla^2 y_k\|_I + \|\nabla^2 p_k\|_I\right)$$

is satisfied for a constant $C > 0$ independent of the space discretization.

Finally, we prove an auxiliary discretization error estimate for the objective function.

Lemma 5.5.18. *Let $u \in U_{ad}$ and $v \in U$ be given. Then the estimate*

$$|f_k'(u)v - f_{kh}'(u)v| \le Ch^2\|v\|_I$$

holds with a constant $C > 0$ independent of h.

Proof. With $y_k := G_k(u)$, $y_{kh} := G_{kh}(u)$, as well as $\tilde{y}_k := G_k'(u)v$ and $\tilde{y}_{kh} := G_{kh}'(u)v$, direct calculations imply

$$\begin{aligned}|f_k'(u)v - f_{kh}'(u)v| =& |(\psi_1(y_k - y_{kh}), \tilde{y}_k)_I + (\psi_1(\tilde{y}_k - \tilde{y}_{kh}), y_{kh})_I + (\psi_2(\tilde{y}_k - \tilde{y}_{kh}), u)_I + (\psi_2(y_k - y_{kh}), v)_I| \\ \le& c\|\tilde{y}_{kh} - \tilde{y}_k\|_I + c\,\|y_{kh} - y_k\|_I\end{aligned}$$

by the boundedness of U_{ad}, the boundedness results of Theorem 5.5.2 and estimate (5.4.35). The error estimates from Theorem 5.5.15 and Lemma 5.5.16 yield the assertion. □

We will also need coercivity results of the second derivative of f_{kh}, but only in the neighborhood of a semi-discrete optimal solution $\bar{u}_k$ that has to correspond to a local solution $\bar{u}$ on the continuous level. We therefore postpone the desired estimates until we prove error estimates for locally optimal controls.

5.6. Control discretization

The third level of discretization involves the discretization of the controls, if desired. We introduce a subspace $U_\delta \subset U$, where the subscript δ indicates control discretization in one of several possible ways to be discussed below. A discrete version of $(\mathbb{P})$ then reads

$$\text{Minimize } f_{kh}(u_\delta), \quad \text{subject to } u_\delta \in U_{\text{ad}} \cap U_\delta, \tag{$\mathbb{P}_\sigma$}$$

where the subscript δ indicates discretization of the control (in space and time) and the index kh indicates the discretization of the state equation in space and time. Collecting all discretization parameters k, h, δ into a single abstract parameter σ simplifies the notation, and we will denote locally optimal controls of $(\mathbb{P}_\sigma)$ by $\bar{u}_\sigma$. The question of existence will be discussed in the further analysis for a specific choice of U_δ, but can generally be expected if the set $U_\delta \cap U_{\text{ad}}$ is not empty.

In [102], different discretization concepts for the control have been discussed in the linear-quadratic setting. We only mention the results that correspond to our two-dimensional setting, but point out that estimates in three space dimensions were also proven in [102]. First of all, for controls that are cellwise constant with respect to space and time an error estimate of the order $\mathcal{O}(k+h)$ was shown, if the same temporal and spatial meshes as for the state equation are used. Choosing an H^1-conforming finite element discretization in space and piecewise constant discretization in time, again on the same meshes as the state equation, an error estimate of the order $\mathcal{O}(k + h^{\frac{3}{2}-\varepsilon})$, $\varepsilon > 0$ was proven. Under certain regularity assumptions, this could even be improved to the order $\mathcal{O}(k + h^{\frac{3}{2}})$. The variational discretization concept from [71] for elliptic problems, without explicit control discretization, led to an optimal error estimate of order $\mathcal{O}(k+h^2)$. Last, the postprocessing approach originally from [109], where cellwise constant control discretization is combined with a postprocessing step utilizing the projection formula for optimal controls has also been discussed and an error estimate of order $\mathcal{O}(k + h^{2-\varepsilon})$, $\varepsilon > 0$ has been derived, that was improved to $\mathcal{O}(k + h^2)$ under certain assumptions.

In [118], the first strategy has been analyzed in detail for problems with semilinear equation, and it has been discussed why the other error estimates are also valid for the semilinear setting. In view of the motivating examples with regularized pointwise state constraints from Section 5.2 we will discuss here the second strategy in detail. That is, we discretize the controls piecewise constant in time and cellwise bilinear in space, using the same meshes as for the state equation.

5.7. Error estimates for the optimal control problem

We are eventually interested in the error $\|\bar{u} - \bar{u}_\sigma\|_I$ of somehow corresponding local solutions on the undiscretized and fully discretized levels. The error estimate will be obtained in two main steps. The first step involves the estimate between $\bar{u}$ and an auxiliary semidiscrete problem. In the second step the error due to the spatial discretization of the PDE as well as the controls is estimated. Since any local solution of this semidiscrete problem is itself already piecewise constant in time, cf. Remark 5.4.14, no additional discretization of the controls in time has to be considered. We formally split the error into

$$\|\bar{u} - \bar{u}_\sigma\|_I \leq \|\bar{u} - \bar{u}_k\|_I + \|\bar{u}_k - \bar{u}_\sigma\|_I, \tag{5.7.1}$$

but auxiliary problems that guarantee closeness of the different solutions have to be taken into account. We point out that all auxiliary problems to be considered in the following fulfill first order necessary optimality conditions in form of variational inequalities like (5.3.3), (5.4.39), or (5.5.16), respectively, if the admissible sets are chosen appropriately.

5.7.1. Error estimates for the semi-discrete optimal control problem

As in Chapters 3 and 4, we adapt the idea from [26] of considering auxiliary problems, and discuss the error between corresponding local solutions of the continuous problem ($\mathbb{P}$) and the time discrete problem ($\mathbb{P}_k$).

Assumption 5.7.1. *Let Assumptions 5.3.1 and 5.3.9 be satisfied. Moreover, let $\varepsilon > 0$ be small enough, that Lemma 5.4.23 is satisfied.*

We consider a semidiscrete problem with controls in the neighborhood of $\bar{u}$.

$$\text{Minimize} \quad f_k(u), \text{ subject to } u \in U_{\text{ad}}^{\varepsilon}, \tag{$\mathbb{P}_k^{\varepsilon}$}$$

where $U_{\text{ad}}^{\varepsilon}$ is defined as

$$U_{\text{ad}}^{\varepsilon} := \{u \in U_{\text{ad}} \colon \ \|\bar{u} - u\|_I \le \varepsilon\}.$$

Lemma 5.7.2. *Let Assumption 5.7.1 hold. For k sufficiently small, the auxiliary problem ($\mathbb{P}_k^{\varepsilon}$) admits a unique global solution $\bar{u}_k^{\varepsilon}$.*

Proof. The existence of a solution is clear noting that $U_{\text{ad}}^{\varepsilon}$ is not empty, since we have at least $\bar{u} \in U_{\text{ad}}^{\varepsilon}$. Uniqueness of $\bar{u}_k^{\varepsilon}$ is a standard conclusion from the coercivity of f_k''. Let us assume that there exist two global minima $\bar{u}_k^{\varepsilon}$, $\bar{r}_k^{\varepsilon}$ of ($\mathbb{P}_k^{\varepsilon}$) with $\bar{u}_k^{\varepsilon} \ne \bar{r}_k^{\varepsilon}$ and $f_k(\bar{u}_k^{\varepsilon}) = f_k(\bar{r}_k^{\varepsilon})$. For some $\xi \in (0,1)$, we obtain

$$\begin{aligned} f_k(\bar{r}_k^{\varepsilon}) =& f_k(\bar{u}_k^{\varepsilon}) + f_k'(\bar{u}_k^{\varepsilon})(\bar{r}_k^{\varepsilon} - \bar{u}_k^{\varepsilon}) + \frac{1}{2} f_k''(\bar{u}_k^{\varepsilon} + \xi(\bar{r}_k^{\varepsilon} - \bar{u}_k^{\varepsilon}))[\bar{r}_k^{\varepsilon} - \bar{u}_k^{\varepsilon}, \bar{r}_k^{\varepsilon} - \bar{u}_k^{\varepsilon}] \\ \ge& f_k(\bar{u}_k^{\varepsilon}) + \frac{\alpha}{8}\|\bar{r}_k^{\varepsilon} - \bar{u}_k^{\varepsilon}\|_I^2 > f_k(\bar{u}_k^{\varepsilon}) \end{aligned}$$

for k sufficiently small by Taylor expansion of f_k at $\bar{u}_k^{\varepsilon}$, first order optimality conditions for $\bar{u}_k^{\varepsilon}$, and Lemma 5.4.23, that is applicable to $u^{\xi} = \bar{u}_k^{\varepsilon} + \xi(\bar{r}_k^{\varepsilon} - \bar{u}_k^{\varepsilon})$ due to the convexity of $U_{\text{ad}}^{\varepsilon}$. This contradiction yields uniqueness of $\bar{u}_k^{\varepsilon}$. □

Theorem 5.7.3. *Let Assumption 5.7.1 hold and let moreover $k > 0$ be small enough that Lemma 5.7.2 is satisfied. Then there exists a constant $C > 0$ such that*

$$\|\bar{u} - \bar{u}_k^{\varepsilon}\|_I \le Ck.$$

Proof. We make use of the variational inequalities for Problems ($\mathbb{P}$) and ($\mathbb{P}_k^{\varepsilon}$). Here we can exploit the fact that both controls $\bar{u}$ and $\bar{u}_k^{\varepsilon}$ are feasible for each problem. We observe that

$$0 \le f'(\bar{u})(\bar{u}_k^{\varepsilon} - \bar{u})$$

as well as

$$0 \le f_k'(\bar{u}_k^{\varepsilon})(\bar{u} - \bar{u}_k^{\varepsilon})$$

is satisfied. Summation of both inequalities yields

$$\begin{aligned} 0 \ &\le \ (f'(\bar{u}) - f_k'(\bar{u}_k^{\varepsilon}))(\bar{u}_k^{\varepsilon} - \bar{u}) \\ &= \ (f'(\bar{u}) - f_k'(\bar{u}))(\bar{u}_k^{\varepsilon} - \bar{u}) + (f_k'(\bar{u}) - f_k'(\bar{u}_k^{\varepsilon}))(\bar{u}_k^{\varepsilon} - \bar{u}) \\ &= \ (\psi_2(\bar{y} - \bar{y}_k) + p(\bar{u}) - p_k(\bar{u}), \bar{u}_k^{\varepsilon} - \bar{u})_I - f_k''(u_{\xi})[\bar{u}_k^{\varepsilon} - \bar{u}, \bar{u}_k^{\varepsilon} - \bar{u}], \end{aligned} \tag{5.7.2}$$

with $u_{\xi} = \bar{u} + \xi(\bar{u}_k^{\varepsilon} - \bar{u})$ for some $0 < \xi < 1$ after Taylor expansion of f_k at $\bar{u}$. We know that $u_{\xi} \in U_{\text{ad}}^{\varepsilon}$, because

$$\|u_{\xi} - \bar{u}\|_I \le \xi\|\bar{u}_k^{\varepsilon} - \bar{u}\|_I \le \|\bar{u}_k^{\varepsilon} - \bar{u}\|_I \le \varepsilon.$$

With Lemma 5.4.23, (5.7.2) leads to

$$\frac{\alpha}{8}\|\bar{u}_k^\varepsilon - \bar{u}\|_I^2 \leq c\left(\|\bar{y} - \bar{y}_k\|_I + \|p(\bar{u}) - p_k(\bar{u})\|_I\right)\|\bar{u}_k^\varepsilon - \bar{u}\|_I$$

from which we obtain

$$\|\bar{u}_k^\varepsilon - \bar{u}\|_I \leq ck(\|\partial_t p(\bar{u})\|_I + \|\partial_t y(\bar{u})\|_I)$$

with Theorem 5.7.4 and Lemma 5.4.22, and the assertion is obtained with the help of the stability estimates from Theorem 5.4.5 and Corollary 5.4.12. □

The error estimate for the auxiliary problem can now easily be transferred to the semidiscrete problem $(\mathbb{P}_k)$.

Theorem 5.7.4. *Let $\bar{u}$ be an optimal control of Problem $(\mathbb{P})$ satisfying the first order optimality conditions* (5.3.3) *as well as the second order sufficient conditions from Assumption 5.3.9. Moreover, let Assumption 5.7.1 be satisfied. Then there exists a sequence of locally optimal controls $\bar{u}_k$ to problem $(\mathbb{P}_k)$ converging strongly in $L^2(Q)$ to $\bar{u}$ as k tends to zero, and the discretization error estimate*

$$\|\bar{u} - \bar{u}_k\|_I \leq Ck$$

is satisfied for a constant $C > 0$, which is independent of k.

Proof. It is sufficient to mention that by Theorem 5.7.3 $\bar{u}_k^\varepsilon$ converges to $\bar{u}$ in $L^2(Q)$ as k tends to zero. Hence, for k sufficiently small, we will have $\|\bar{u}_k^\varepsilon - \bar{u}\| < \varepsilon$, and therefore we find that $\bar{u}_k^\varepsilon =: \bar{u}_k$ is a local solution to Problem $(\mathbb{P}_k)$, for which the error estimate from Theorem 5.7.3 remains valid. □

In preparation of error estimates due to spatial discretization, we prove coercivity results of f_k'' and f_{kh}'' in the neighborhood of $\bar{u}_k$.

Lemma 5.7.5. *Let Assumption 5.7.1 be satisfied and let $\bar{u}_k^\varepsilon \in U_{ad}$ denote the solution of Problem $(\mathbb{P}_k^\varepsilon)$. There exists a constant ε_k such that*

$$f_k''(u)[v,v] \geq \frac{\alpha}{8}\|v\|_I^2$$

for all $\|u - \bar{u}_k^\varepsilon\|_I \leq \varepsilon_k$ and all $v \in U$.

Proof. This follows directly from the Lipschitz continuity result for f_k'' in Lemma 5.4.16 and the coercivity result for f_k'' from Lemma 5.4.23 considering

$$f_k''(u)[v,v] = f_k''(\bar{u}_k^\varepsilon)[v,v] + \left(f_k''(u)[v,v] - f_k''(\bar{u}_k^\varepsilon)[v,v]\right),$$

and noting that $\|\bar{u}_k^\varepsilon - \bar{u}\|_I \leq \varepsilon$, so that Lemma 5.4.23 is applicable. Then we observe

$$f_k''(u)[v,v] \geq \frac{\alpha}{4}\|v\|_I^2 - c\|u - \bar{u}_k^\varepsilon\|_I\|v\|_I^2,$$

where the constant c is in particular independent of k, cf. Lemma 5.4.16. □

The index k in ε_k merely serves as a means to distinguish between ε_k and ε from Lemma 5.4.23.

Lemma 5.7.6. *Let Assumption 5.7.7 hold. Then, for all $u \in U_{ad}$ with $\|u - \bar{u}_k\|_I \leq \varepsilon_k$ and all $u \in U$, the inequality*

$$f''_{kh}(u)[v, v] \geq \frac{\alpha}{16}\|v\|_I^2$$

is satisfied for all h sufficiently small.

Proof. With Lemma 5.7.5, the proof follows as the one for Lemma 5.4.23, making use of Theorems 5.5.2 and 5.5.15, Lemmas 5.5.16 and 5.5.17, as well as the stability estimates (5.5.11)-(5.5.13). □

5.7.2. Error estimates for the discrete optimal control problem

Let us now come to the last step of our error estimate, i.e. the spatial error estimate in the control with cellwise linear control discretization with the same discretization parameter h as is used for the PDE. Let us point mention that we give more details than in [118], where the focus has been on cellwise constant controls. We follow closely the arguments in [21], where semilinear elliptic problems have been discussed, and [102], where a corresponding estimate has been shown for linear-quadratic parabolic problems. To minimize the notational effort, we agree on the following assumption.

Assumption 5.7.7. *Within this section, we denote by $\bar{u}_k \in B_\varepsilon(\bar{u})$ a locally optimal control of $(\mathbb{P}_k)$ from Theorem 5.7.4 that approximates $\bar{u}$. Moreover, let ε_k be small enough that Lemmas 5.7.5 and 5.7.6 are satisfied.*

With the discrete space

$$U_h = \{v \in \mathcal{C}(\bar{\Omega}) | v_{|_K} \in \mathcal{Q}_1(K) \text{ for } K \in \mathcal{T}_h\},$$

we define the discrete control space $U_\delta \supset Y_{k,h}^{0,1}$ as

$$U_\delta = \{u \in U | u_{|_{I_m}} \in \mathcal{P}_0(I_m, U_h)\},$$

as well as the set of discrete admissible controls

$$U_{\text{ad}}^\delta := U_\delta \cap U_{\text{ad}}.$$

Let us mention that the discrete state space $Y_{k,h}^{0,1}$ is only a subspace of U_δ, since both spaces differ in the boundary conditions. For convenience, we formulate again the completely discretized problem $(\mathbb{P}_\sigma)$,

$$\text{Minimize } f_{kh}(u_\delta) \quad \text{subject to } u_\delta \in U_{\text{ad}}^\delta, \tag{$\mathbb{P}_\sigma$}$$

and introduce an auxiliary problem to discuss local solutions of $(\mathbb{P}_\sigma)$ that correspond to $\bar{u}_k$ and hence to $\bar{u}$. This reads

$$\text{Minimize } f_{kh}(u) \quad \text{subject to } u \in U_{\text{ad}}^{\delta,\varepsilon_k}, \tag{$\mathbb{P}_\sigma^{\varepsilon_k}$}$$

wherethe auxiliary set $U_{\text{ad}}^{\delta,\varepsilon_k}$ is defined by

$$U_{\text{ad}}^{\delta,\varepsilon_k} := \{u \in U_{\text{ad}}^\delta \colon \ \|\bar{u}_k - u\|_I \leq \varepsilon_k\}.$$

As a link between the semidiscrete and the discrete level, we make use of the usual nodal interpolation operator

$$I_\delta \colon \mathcal{C}(\Omega) \to U_h, \quad I_\delta g(x_i) = g(x_i) \quad \text{for each node } x_i \text{ of } \mathcal{T}_h.$$

It is well known, cf. for instance [34], that I_δ satisfies

$$\|g - I_\delta g\| \leq ch\|\nabla g\| \tag{5.7.3}$$

for all $g \in H^1(\Omega) \cap \mathcal{C}(\Omega)$. For time-dependent functions, we simply set

$$(I_\delta g)(t) = I_\delta g(t).$$

We will apply the operator to the semidiscrete control $\bar{u}_k$ separately on each time interval. Note that (5.7.3) implies that $\bar{u}_k \in U_{\text{ad}}^{\delta,\varepsilon_k}$ for h sufficiently small, since $\bar{u}_k$ is already piecewise constant in time. Therefore, the set of feasible controls for the auxiliary problem $(\mathbb{P}_\sigma^{\varepsilon_k})$ is not empty.

Lemma 5.7.8. *Let Assumption 5.7.7 be satisfied. For all $\delta > 0$ sufficiently small, the auxiliary problem $(\mathbb{P}_\sigma^{\varepsilon_k})$ admits a unique global solution $\bar{u}_\sigma^\varepsilon$.*

Proof. The proof of existence follows by standard arguments noting that $I_\delta \bar{u}_k \in U_{\text{ad}}^{\delta,\varepsilon_k}$ for all h sufficiently small. The proof of uniqueness then follows similarly to Lemma 5.7.2, making use of the coercivity of f''_{kh} from Lemma 5.7.6. □

In the following analysis, we make an assumption on the structure of the active sets as in [130] or [109] for elliptic problems, and for linear-quadratic parabolic problems in [102]. In [130], this was used to obtain the order $h^{\frac{3}{2}}$ for linearly discretized controls for the first time. For each time interval I_m, we group the cells K of the mesh $\mathcal{T}_h$ depending on the value of $\bar{u}_k$ on K into three sets

$$\mathcal{T}_h = \mathcal{T}_{h,m}^1 \cup \mathcal{T}_{h,m}^2 \cup \mathcal{T}_{h,m}^3$$

with $\mathcal{T}_{h,m}^i \cap \mathcal{T}_{h,m}^j = \emptyset$ for $i \neq j$, where the three sets are chosen as follows:

$$\begin{aligned}
\mathcal{T}_{h,m}^1 &= \{K \in \mathcal{T}_h \,|\, \bar{u}_k(t_m, x) = u_a \text{ or } \bar{u}_k(t_m, x) = u_b \quad \forall x \in K\}, \\
\mathcal{T}_{h,m}^2 &= \{K \in \mathcal{T}_h \,|\, u_a < \bar{u}_k(t_m, x) < u_b \quad \forall x \in K\}, \\
\mathcal{T}_{h,m}^3 &= \mathcal{T}_h \setminus (\mathcal{T}_{h,m}^1 \cup \mathcal{T}_{h,m}^2).
\end{aligned}$$

In the sequel, we will see that both the sets $\mathcal{T}_{h,m}^1$ and $\mathcal{T}_{h,m}^2$ can be handled easily in the error analysis, since the optimal control $\bar{u}_k$ is either constant or fulfills a gradient equation. Both properties can be used to estimate $\|\bar{u}_k - \mathcal{I}_d \bar{u}_k\|_I$ conveniently. The cells contained in $\mathcal{T}_{h,m}^3$ contain part of the free boundary between the active and the inactive sets for the time interval I_m. This will need special attention.

Assumption 5.7.9. *We assume that there exists a positive constant C independent of k, h, and m such that*

$$\sum_{K \in \mathcal{T}_{h,m}^3} |K| \leq Ch$$

separately for all $m = 1, 2, \ldots, M$.

Theorem 5.7.10. *Let Assumption 5.7.7 hold and let $h > 0$ be small enough that Lemmas 5.7.6 and 5.7.8 are satisfied. Then there exists a constant $C > 0$ such that*

$$\|\bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I \leq Ch^{\frac{3}{2}-\rho}$$

is fulfilled for every $\rho > 0$ and a constant $C > 0$ independent of k and h.

Proof. The proof is analogous to the one presented for linear-quadratic problems in [102], noting that the coercivity result from Lemma 5.7.6 is fulfilled. Cf. also the semilinear elliptic setting in [21]. We give the complete proof for convenience of the reader. We point out that $\bar{y}_k, \bar{p}_k \in L^2(I, H^2(\Omega))$. Therefore, by a Sobolev embedding we have $\bar{y}_k, \bar{p}_k \in L^2(I, W^{1,s}(\Omega))$ for all $s < \infty$, since Ω is twodimensional. Consequently, we can make use of the boundedness of $\|\nabla \bar{y}_k\|_{L^s(\Omega)}$ and $\|\nabla \bar{p}_k\|_{L^s(\Omega)}$.

- In a first step, note again that $I_\delta \bar{u}_k \in U_{\mathrm{ad}}^{\delta,\varepsilon_k}$ if h is sufficiently small. Then, by Taylor expansion of f_{kh} in $\bar{u}_\sigma^\varepsilon$ we obtain

$$f'_{kh}(I_\delta \bar{u}_k)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) - f'_{kh}(\bar{u}_\sigma^\varepsilon)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) = f''_{kh}(u_\xi)[I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon, I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon] \tag{5.7.4}$$

with $u_\xi = \bar{u}_\sigma^\varepsilon + \xi(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon)$ for a $\xi \in (0,1)$. To apply the coercivity result of Lemma 5.7.6 to the right-hand side of (5.7.4), note that

$$\|u_\xi - \bar{u}_k\|_I \le (1-\xi)\|\bar{u}_\sigma^\varepsilon - \bar{u}_k\|_I + \xi \|I_\delta \bar{u}_k - \bar{u}_k\|_I \le \varepsilon_k.$$

Therefore, Lemma 5.7.6 and representation (5.7.4) lead to

$$\frac{\alpha}{8}\|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I^2 \le f'_{kh}(I_\delta \bar{u}_k)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) - f'_{kh}(\bar{u}_\sigma^\varepsilon)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon). \tag{5.7.5}$$

Moreover, from the optimality conditions for problems $(\mathbb{P}_k)$ and $(\mathbb{P}_\sigma^{\varepsilon_k})$ we deduce

$$\begin{aligned} -f'_{kh}(\bar{u}_\sigma^\varepsilon)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) &\le 0 \\ &\le -f'_k(\bar{u}_k)(\bar{u}_k - \bar{u}_\sigma^\varepsilon) \\ &= -f'_k(\bar{u}_k)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) - f'_k(\bar{u}_k)(\bar{u}_k - I_\delta \bar{u}_k), \end{aligned} \tag{5.7.6}$$

and combining (5.7.5) and (5.7.6) yields

$$\begin{aligned} \frac{\alpha}{8}\|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I^2 \le \quad & f'_{kh}(I_\delta \bar{u}_k)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) - f'_k(\bar{u}_k)(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) \\ & -f'_k(\bar{u}_k)(\bar{u}_k - I_\delta \bar{u}_k) \\ = \quad & (f'_{kh}(I_\delta \bar{u}_k) - f'_{kh}(\bar{u}_k))(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) + (f'_{kh}(\bar{u}_k) - f'_k(\bar{u}_k))(I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon) \\ & -f'_k(\bar{u}_k)(\bar{u}_k - I_\delta \bar{u}_k) \\ \le \quad & c\,\|\bar{u}_k - I_\delta \bar{u}_k\|_I \|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I + ch^2 \|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I \\ & +(\nu \bar{u}_k + \psi_2 \bar{y}_k + \bar{p}_k, I_\delta \bar{u}_k - \bar{u}_k)_I, \end{aligned} \tag{5.7.7}$$

similar to the proof of Theorem 5.5 in [102], where we used the Lipschitz stability result error estimate from Lemmas 5.5.14 and 5.5.18, respectively. This yields

$$\|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I^2 \le c\,\|\bar{u}_k - I_\delta \bar{u}_k\|_I \|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I + ch^2 \|I_\delta \bar{u}_k - \bar{u}_\sigma^\varepsilon\|_I + (\nu \bar{u}_k + \psi_2 \bar{y}_k + \bar{p}_k, I_\delta \bar{u}_k - \bar{u}_k)_I. \tag{5.7.8}$$

- We proceed by estimating $\|I_\delta \bar{u}_k - \bar{u}_k\|_I$. Following the proof of Lemma 5.7 in [102], which needs only minor adaptation, we observe

$$\|I_\delta \bar{u}_k - \bar{u}_k\|_I^2 = \sum_{i=1}^M \int_{I_m} \|I_\delta \bar{u}_k(t) - \bar{u}_k(t)\|^2\,dt = \sum_{i=1}^M k_m \|I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m)\|^2, \tag{5.7.9}$$

since $\bar{u}_k$ and consequently $\mathcal{I}_d \bar{u}_k$ are piecewise constant in time. For each $m = 1, \dots, M$, we split

$$\|I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m)\|^2 = \sum_{K \in \mathcal{T}_{h,m}^2} \|I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m)\|_{L^2(K)}^2 + \sum_{K \in \mathcal{T}_{h,m}^3} \|I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m)\|_{L^2(K)}^2, \tag{5.7.10}$$

since $I_\delta \bar{u}_k = \bar{u}_k$ on $\mathcal{T}^1_{h,m}$. We know that

$$\bar{u}_k(t_m) = -\frac{1}{\nu}\bar{p}_k(t_m) - \frac{\psi_2}{\nu}\bar{y}_k(t_m) \quad \text{on } \mathcal{T}^2_{h,m},$$

and with (5.7.3) this allows to estimate

$$\begin{aligned} \sum_{K\in\mathcal{T}^2_{h,m}} \|I_\delta\bar{u}_k(t_m) - \bar{u}_k(t_m)\|^2_{L^2(K)} &\leq ch^4 \sum_{K\in\mathcal{T}^2_{h,m}} \|\nabla^2\bar{u}_k(t_m)\|_{L^2(K)} \\ &\leq ch^4\left(\|\nabla^2\bar{y}_k\|^2_{L^2(K)} + \|\nabla^2\bar{p}_k\|^2_{L^2(K)}\right). \end{aligned} \tag{5.7.11}$$

For the second term in (5.7.10), we make use of Assumption 5.7.9 and estimate (5.7.3) and obtain

$$\begin{aligned} \sum_{K\in\mathcal{T}^3_{h,m}} \|I_\delta\bar{u}_k(t_m) - \bar{u}_k(t_m)\|^2_{L^2(K)} \leq& \sum_{K\in\mathcal{T}^3_{h,m}} |K|^{1-\frac{2}{s}}\|I_\delta\bar{u}_k(t_m) - \bar{u}_k(t_m)\|^2_{L^s(K)} \\ &\leq ch^2 \sum_{K\in\mathcal{T}^3_{h,m}} |K|^{1-\frac{2}{s}}\|\nabla\bar{u}_k(t_m)\|^2_{L^s(K)} \\ &\leq ch^2 \left(\sum_{K\in\mathcal{T}^3_{h,m}} |K|\right)^{1-\frac{2}{s}} \|\nabla\bar{u}_k(t_m)\|^2_{L^s(\Omega)} \\ &\leq ch^{3-\frac{2}{s}}\left(\|\nabla\bar{y}_k(t_m)\|^2_{L^s(\Omega)} + \|\nabla\bar{p}_k(t_m)\|^2_{L^s(\Omega)}\right). \end{aligned}$$

Combining this with (5.7.9)-(5.7.11) yields

$$\|I_\delta\bar{u}_k(t_m) - \bar{u}_k(t_m)\| \leq ch^{\frac{3}{2}-\frac{1}{s}} \tag{5.7.12}$$

for all $p < \infty$.

- Now, as in Lemma 5.6 [102], we observe

$$\begin{aligned} (\nu\bar{u}_k + \psi_2\bar{y}_k + \bar{p}_k, \bar{u}_k - I_\delta\bar{u}_k)_I &= \sum_{i=1}^{M}\int_{I_m}(\nu\bar{u}_k(t) + \psi_2\bar{y}_k(t) + \bar{p}_k(t), \bar{u}_k(t) - I_\delta\bar{u}_k(t))\,dt \\ &= \sum_{i=1}^{M} k_m(\nu\bar{u}_k(t_m) + \psi_2\bar{y}_k(t_m) + \bar{p}_k(t_m), \bar{u}_k(t_m) - I_\delta\bar{u}_k(t_m)), \end{aligned}$$

since all appearing functions are piecewise constant. Setting

$$g_m := \nu\bar{u}_k(t_m,\cdot) + \psi_2\bar{y}_k(t_m,\cdot) + \bar{p}_k(t_m,\cdot)$$

delivers

$$\begin{aligned} (\nu\bar{u}_k + \psi_2\bar{y}_k + \bar{p}_k, \bar{u}_k - I_\delta\bar{u}_k)_I &= \sum_{i=1}^{M} k_m \sum_{K\in\mathcal{T}_{h,m}} (g_m, \bar{u}_k(t_m) - I_\delta\bar{u}_k(t_m))_{L^2(K)} \\ &= \sum_{i=1}^{M} k_m \sum_{K\in\mathcal{T}^3_{h,m}} (g_m, \bar{u}_k(t_m) - I_\delta\bar{u}_k(t_m))_{L^2(K)}. \end{aligned} \tag{5.7.13}$$

Here, we used that $I_\delta\bar{u}_k(t_m) = \bar{u}_k(t_m)$ on $\mathcal{T}^1_{h,m}$ by construction, as well as

$$g_m = 0$$

on $\mathcal{T}^2_{h,m}$. Now, by definition of $\mathcal{T}^3_{h,m}$, in every cell $K \in \mathcal{T}^3_{h,m}$ there is a point x_K with

$$g_m(x_K) = 0.$$

With (5.7.3) this implies

$$\begin{aligned} |(g_m, I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m))_{L^2(K)}| &\le |K|^{1-\frac{2}{s}} \|g_m\|_{L^s(K)} \|I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m)\|_{L^s(K)} \\ &= |K|^{1-\frac{2}{s}} \|g_m - g_m(x_K)\|_{L^s(K)} \|I_\delta \bar{u}_k(t_m) - \bar{u}_k(t_m)\|_{L^s(K)} \\ &\le ch^2 |K|^{1-\frac{2}{s}} \|\nabla g_m\|_{L^s(K)} \|\nabla \bar{u}_k(t_m)\|_{L^s(K)} \end{aligned}$$

for any $2 < s < \infty$. Note that an estimate like (5.7.3) can be applied to $\|g_m - g_m(x_K)\|$, where $g_m(x_k)$ is constant.

Inserting this into (5.7.13) yields

$$\begin{aligned} &(\nu \bar{u}_k + \psi_2 \bar{y}_k + \bar{p}_k, \bar{u}_k - I_\delta \bar{u}_k)_I \\ &\quad \le \sum_{i=1}^{M} c k_m h^2 \sum_{K \in \mathcal{T}^3_{h,m}} |K|^{1-\frac{2}{s}} \|\nabla g_m\|_{L^s(K)} \|\nabla \bar{u}_k(t_m)\|_{L^s(K)} \\ &\quad \le \sum_{i=1}^{M} c k_m h^2 (\sum_{K \in \mathcal{T}^3_{h,m}} |K|)^{1-\frac{2}{s}} \|\nabla g_m\|_{L^s(\Omega)} \|\nabla \bar{u}_k(t_m)\|_{L^s(\Omega)} \\ &\quad \le \sum_{i=1}^{M} c k_m h^{3-\frac{2}{s}} \|\nabla g_m\|_{L^s(\Omega)} \|\nabla \bar{u}_k(t_m)\|_{L^s(\Omega)} \\ &\quad \le \sum_{i=1}^{M} c k_m h^{3-\frac{2}{s}} (\|\nabla \bar{y}_k(t_m)\|_{L^2(I,L^s(\Omega))} + \|\nabla \bar{p}_k(t_m)\|_{L^2(I,L^s(\Omega))} + 1) \|\nabla \bar{u}_k(t_m)\|_{L^s(\Omega)} \\ &\quad \le c h^{3-\frac{2}{s}} (\|\nabla \bar{y}_k\|_{L^2(I,L^s(\Omega))} + \|\nabla \bar{p}_k\|_{L^2(I,L^s(\Omega))} + 1) \|\nabla \bar{u}_k(t_m)\|_{L^2(I,L^s(\Omega))} \end{aligned} \tag{5.7.14}$$

which implies

$$(\nu \bar{u}_k + \psi_2 \bar{y}_k + \bar{p}_k, \bar{u}_k - I_\delta \bar{u}_k)_I \le c h^{\frac{3}{2} - \frac{1}{s}}. \tag{5.7.15}$$

- Finally, inserting (5.7.12) and (5.7.15) into (5.7.8) and applying Young's inequality yields the assertion, noting that

$$\|\bar{u}_k - \bar{u}^\varepsilon_\sigma\|_I \le \|I_\delta \bar{u}_k - \bar{u}_k\|_I + \|I_\delta \bar{u}_k - \bar{u}^\varepsilon_\sigma\|_I.$$

□

Again, we obtain that $\bar{u}^\varepsilon_\sigma$ converges to $\bar{u}_k$ in $L^2(Q)$ as h tends to zero. Therefore, $\bar{u}^\varepsilon_\sigma =: \bar{u}_\sigma$ is a local solution of Problem $(\mathbb{P}_\sigma)$ for all h sufficiently small, and we directly obtain

Theorem 5.7.11. *Let $\bar{u}$ be an optimal solution of Problem $(\mathbb{P})$, let $\bar{u}_k \in B_\varepsilon(\bar{u})$ be an optimal control of Problem $(\mathbb{P}_k)$ and let Assumptions 5.7.1-5.7.9 be satisfied. Then there exists a sequence of locally optimal controls to problem $(\mathbb{P}_{kh})$ converging strongly in $L^2(Q)$ to $\bar{u}_k$ as h tends to zero, and the discretization error estimate*

$$\|\bar{u}_k - \bar{u}_\sigma\|_I \le C(k + h^{\frac{3}{2} - \rho})$$

is fulfilled for every $\rho > 0$ and a constant $C > 0$ independent of k and h.

Proof. This is a direct conclusion of Theorem 5.7.10, noting that the sequence $\{\bar{u}_\sigma^\varepsilon\}$ tends to $\bar{u}_k$ in $L^2(Q)$, which makes $\bar{u}_\sigma^\varepsilon$ a local solution of $(\mathbb{P}_\sigma)$. $\square$

From the last line of (5.7.14) in the proof of Theorem 5.7.10 it becomes obvious that if we have the regularity $\bar{y}_k, \bar{p}_k \in L^2(I, W^{1,\infty}(\Omega))$ we obtain an error estimate of order $\mathcal{O}(k + h^{\frac{3}{2}})$. Cf. also [102], where only $\|\nabla \bar{p}_k\|_{L^s(\Omega)}$ appeared in the gradient equation due to the different objective functional.

5.8. Main result

Combining the Theorems 5.7.4 and 5.7.11, we arrive at the main result of this chapter, an a priori error estimate between a completely discrete solution $\bar{u}_\sigma$ and the corresponding continuous solution $\bar{u}$.

Theorem 5.8.1. *Let the general Assumption 5.3.1 hold. Let $\bar{u}$ be an optimal control of Problem $(\mathbb{P})$ satisfying the first order optimality conditions* (5.3.3) *as well as second order sufficient condition from Assumption 5.3.9. Moreover, let a corresponding semidiscrete solution $\bar{u}_k$ satisfy the structural Assumption 5.7.9. Then there exists a sequence of locally optimal controls $\{\bar{u}_\sigma\}$ to problem $(\mathbb{P}_\sigma)$ converging strongly in $L^2(Q)$ to $\bar{u}$ as k,h tend to zero, and the discretization error estimate*

$$\|\bar{u} - \bar{u}_\sigma\|_I \leq C(k + h^{\frac{3}{2}-\rho})$$

is fulfilled for all $\rho > 0$ and a constant $C > 0$ independent of k and h.

Proof. This is a direct consequence of Theorems 5.7.3 and 5.7.10. $\square$

Remark 5.8.2. *We point out again the results by Vexler and the author presented in [118], where an error estimate of order $\mathcal{O}(k+h)$ has been proven in detail for cellwise constant control approximations. It has also been mentioned there that in case of variational discretization only the error between $\bar{u}$ and $\bar{u}_{kh}$ needs to be analyzed, leading to an error estimate of order $\mathcal{O}(k+h^2)$. For the postprocessing approach, we expect an order of $\mathcal{O}(k+h^{2-\rho})$ or even $\mathcal{O}(k+h^2)$ under certain regularity conditions, since the coercivity results for the second derivative of the objective functional is the key ingredient in transferring arguments from the linear quadratic setting to a problem with semilinear state equation.*

We will conclude this section with a numerical example that illustrates the proven order of convergence.

5.9. Numerical results

We are now going to validate the proven error estimates with the help of a numerical example. We are interested in evaluating the temporal and spatial discretization error for the control with the help of a control-constrained test example with known exact solution.

5.9.1. The SQP primal-dual active set strategy

We will solve the non-convex control problem with an (outer) SQP method, where the inequality constrained quadratic subproblems are solved with an (inner) primal-dual active set strategy. Starting from an initial guess (u^0, y^0, p^0) for the control, state, and adjoint state, the SQP method generates a sequence $(u^n, y^n, p^n)_{n\in\mathbb{N}}$. If (u^n, y^n, p^n) is already computed, the quadratic subproblem to be solved in the next iteration is given by

$$\text{Minimize } \tilde{J}(y,u) := J'(y^n,u^n)(y-y^n, u-u^n) + \frac{1}{2}\mathcal{L}''(y^n,u^n,p^n)[y-y^n, u-u^n]^2 \tag{5.9.1a}$$

subject to

$$\begin{aligned} \partial_t y - \Delta y + d(t,x,y^n) + \partial_y d(t,x,y^n)(y-y^n) &= u \quad \text{in } Q, \\ y(0,\cdot) &= y_0 \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \Sigma, \end{aligned} \tag{5.9.1b}$$

as well as the inequality constraints

$$u_a \le u(t,x) \le u_b \quad \text{a.e. in } Q. \tag{5.9.1c}$$

Here, the Lagrangian $\mathcal{L}$ is formally defined by

$$\mathcal{L} = J - \iint\limits_Q (\partial_t y - \Delta y - d(t,x,y) - u)p\, dxdt - \iint\limits_\Sigma yp\, dsdt,$$

and J' as well as $\mathcal{L}''$ denote the first and second derivative of the objective function and the Lagrangian, respectively, with respect to y and u. We refer to [144] for a more detailed overview, and to [140] for a convergence analysis. For our model problem (5.3.1) $\tilde{J}$ takes the form

$$\begin{aligned} \tilde{J}(y,u) = &\iint\limits_Q \left(\begin{bmatrix} \Psi_1 y^n + \Psi_2 u^n + \Psi_3 \\ \Psi_2 y^n + \nu u^n \end{bmatrix} \cdot \begin{bmatrix} y-y^n \\ u-u^n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} y-y^n \\ u-u^n \end{bmatrix}^T \begin{bmatrix} \Psi_1 & \Psi_2 \\ \Psi_2 & \nu \end{bmatrix} \begin{bmatrix} y-y^n \\ u-u^n \end{bmatrix} \right) dxdt \\ &- \frac{1}{2} \iint\limits_Q p^n \partial_{yy} d(t,x,y^n)(y-y^n)^2\, dxdt. \end{aligned}$$

The optimal control of this subproblem, together with its associated state and adjoint state will form the next iterate $(u^{n+1}, y^{n+1}, p^{n+1})$. Therefore, we denote the optimal control of (5.9.1) by u^{n+1} and the associated state by y^{n+1}. The adjoint state p^{n+1} is given by the solution p of the adjoint equation

$$\begin{aligned} -\partial_t p - \Delta p + \partial_y d(t,x,y^n)p &= \Psi_1 y^{n+1} + \Psi_2 u^{n+1} + \Psi_3 - p^n \partial_{yy} d(t,x,y^n)(y^{n+1} - y^n) \quad \text{in } Q, \\ p(T,\cdot) &= 0 \quad \text{in } \Omega, \\ p &= 0 \quad \text{on } \Sigma. \end{aligned} \tag{5.9.2}$$

and u^{n+1} satisfies the projection formula

$$u^{n+1} = P_{U_{\text{ad}}}\left(-\frac{1}{\nu} p^{n+1} - \frac{\Psi_2}{\nu} y^{n+1} \right).$$

It is clear, cf. for instance [144], that Lagrange multipliers $\eta_a^{n+1}, \eta_b^{n+1} \in L^\infty(Q)$ for the control constraints can be determined by

$$\begin{aligned} \eta_a^{n+1} &= (\nu u^{n+1} + \Psi_2 y^{n+1} + p^{n+1})_+, \\ \eta_b^{n+1} &= (\nu u^{n+1} + \Psi_2 y^{n+1} + p^{n+1})_-, \end{aligned}$$

which we combine into a function

$$\eta^{n+1} := \eta_a^{n+1} - \eta_b^{n+1}.$$

To determine $(u^{n+1}, y^{n+1}, p^{n+1})$ as well as η^{n+1} an interior active set loop is used. We drop the superscripts n and $n+1$, and introduce a subscript k, or $k+1$, respectively, to indicate the iterates of the primal-dual active set strategy, $(u_k, y_k, p_k)_{k\in\mathbb{N}}$ as well as $(\eta_k)_{k\in\mathbb{N}}$, starting from $u_0 = u^n, y_0 = y^n, p_0 = p^n$ and $\eta_0 = \eta^n$. Known convergence results, cf. [52] guarantee that (u_k, y_k, p_k) tends to $(u^{n+1}, y^{n+1}, p^{n+1})$. In each active set loop the active sets

$$\begin{aligned}
\mathcal{A}_{k+1}^a &= ((t,x) \in Q \colon u_k(t,x) + \eta_k(t,x) \le u_a) \\
\mathcal{A}_{k+1}^b &= ((t,x) \in Q \colon u_k(t,x) + \eta_k(t,x) \ge u_b)
\end{aligned}$$

are determined and the inequality constraints

$$u_a \le u \le u_b$$

are replaced by equality constraints of the form

$$u_{k+1} = u_a \quad \text{on } \mathcal{A}_{k+1}^a, \qquad u_{k+1} = u_b \quad \text{on } \mathcal{A}_{k+1}^b.$$

Noting additionally that u_{k+1} fulfills

$$u_{k+1} = -\frac{1}{\nu} p_{k+1} - \frac{\Psi_2}{\nu} y_{k+1} \quad \text{on } \mathcal{I}_{k+1} := Q \setminus (\mathcal{A}_{k+1}^a \cup \mathcal{A}_{k+1}^b)$$

we obtain the concise representation

$$u_{k+1} := \chi_{\mathcal{I}_{k+1}} \cdot (-\frac{1}{\nu} p_{k+1} - \frac{\Psi_2}{\nu} y_{k+1}) + \chi_{\mathcal{A}_{k+1}^a} \cdot u_a + \chi_{\mathcal{A}_{k+1}^b} \cdot u_b,$$

where $\chi_{\mathcal{I}_{k+1}}$, $\chi_{\mathcal{A}_{k+1}^a}$, and $\chi_{\mathcal{A}_{k+1}^b}$ denote the characteristic functions of the indexed sets. We refer for instance to [108] for a more detailed description of the primal dual active set strategy for elliptic problems. After obtaining the solution triple $(u_{k+1}, y_{k+1}, p_{k+1})$, we update the multiplier

$$\eta_{k+1} = \nu u_{k+1} + \Psi_2 y_{k+1} + p_{k+1},$$

and determine the new active sets. The interior active set loop is terminated when the active sets in two consecutive loops do not change.

5.9.2. Construction of a control-constrained test example

We consider a model problem in one space dimension with tracking type objective function, i.e. $\Psi_1 = 1$, $\Psi_2 = 0$, as well as $\Psi_3 = y_d$, where y_d is a function to be specified rather than a fixed real number. The spatial domain Ω and the time interval I are given by $\Omega = (0,1)$ as well as $I = (0,T)$.

$$\text{Minimize } \frac{1}{2} \int_0^T \int_\Omega ((y - y_d)^2 + u^2)\, dx dt \tag{E}$$

subject to

$$\begin{aligned}
\partial_t y - \Delta y + y^3 &= u + f(t,x) && \text{in } Q \\
y(0,\cdot) &= 0 && \text{in } \Omega, \\
y &= 0 && \text{on } \Sigma,
\end{aligned}$$

as well as

$$u_a \le u \le u_b,$$

with the specific choice

$$y_d(t,x) = -\pi\cos(\pi t)\sin(\pi x) + \pi^2\sin(\pi t)\sin(\pi x) + 3\sin^3(\pi t)\sin^3(\pi x) + \sin(\pi t)\sin(\pi x),$$
$$f(t,x) = \pi\cos(\pi t)\sin(\pi x) + \pi^2\sin(\pi t)\sin(\pi x) + \sin^3(\pi t)\sin^3(\pi x) - \min\left(\frac{1}{2}, \sin(\pi t)\sin(\pi x)\right),$$

and $u_a = -10$ and $u_b = \frac{1}{2}$.

Noting that the adjoint equation takes the form

$$\begin{aligned} -\partial_t p - \Delta p + 3\bar{y}^2 p &= \bar{y} - y_d && \text{in } Q \\ p(T,\cdot) &= 0 && \text{in } \Omega, \\ p &= 0 && \text{on } \Sigma, \end{aligned}$$

direct computations verify that

$$\bar{u} = \min\left(\frac{1}{2}, \sin(\pi t)\sin(\pi x)\right)$$

together with the associated state and adjoint state

$$\bar{y} = \sin(\pi t)\sin(\pi x), \quad \bar{p} = -\sin(\pi t)\sin(\pi x)$$

satisfy the first order optimality conditions of Lemma 5.3.6 or the projection formula (5.3.6). For the specific setting of Problem (E) we find that the second derivative of the reduced objective function is given by

$$f''(\bar{u})[v,v] = \iint\limits_Q \left(\tilde{y}^2 + \nu v^2 - 6\bar{y}\bar{p}\tilde{y}^2\right)\, dxdt,$$

with $\tilde{y} = G'(\bar{u})v$, cf. (5.3.8). We observe that

$$-6\bar{y}\bar{p}\tilde{y}^2 = 6\tilde{y}^2\sin^2(\pi t)\sin^2(\pi x) \ge 0.$$

Therefore, the second order sufficient condition from Assumption 5.3.9 is satisfied since

$$f''(\bar{u})[v,v] = \iint\limits_Q \left(\tilde{y}^2 + \nu v^2 - 6\bar{y}\bar{p}\tilde{y}^2\right)\, dxdt \ge \nu\|v\|_I^2$$

holds, and $\bar{u}$ is indeed an optimal solution. We will verify the error estimates for the temporal and the spatial discretization separately.

5.9.3. The error due to time discretization

We first solve the problem on a spatial grid with $h = 2^{-5}$ for temporal meshes of size $k = 2^{-1}, \dots, 2^{-11}$. We determine the experimental order of convergence by

$$EOC_k = \frac{\ln(\|\bar{u} - \bar{u}_{\sigma_1}\|) - \ln(\|\bar{u} - \bar{u}_{\sigma_2}\|)}{\ln(k_1) - \ln(k_2)},$$

where k_1 and k_2 are consecutive (temporal) mesh sizes, and σ_i denotes the discretization of the PDE and the control with fixed mesh size $h = 2^{-5}$ and consecutive temporal discretization parameters k_i.

k	♯SQP it	$\|\|\bar{u}-\bar{u}_\sigma\|\|_I$	EOC_k	$\|\|\bar{y}-\bar{y}_\sigma\|\|_I$	$\bar{p}-\bar{p}_\sigma\|\|_I$
2^{-1}	4	1.3624e-01	-	8.7468e-02	1.3854e-01
2^{-2}	4	8.1902e-02	0.73	4.8211e-02	8.6794e-02
2^{-3}	4	4.8459e-02	0.76	2.4857e-02	5.0066e-02
2^{-4}	4	2.6695e-02	0.86	1.2717e-02	2.7579e-02
2^{-5}	4	1.4245e-02	0.91	6.4683e-03	1.4622e-02
2^{-6}	4	7.3740e-03	0.95	3.3075e-03	7.5540e-03
2^{-7}	4	3.7627e-03	0.97	1.7152e-03	3.8430e-03
2^{-8}	4	1.9061e-03	0.98	9.1685e-04	1.9397e-03
2^{-9}	4	9.6566e-04	0.98	5.1803e-04	9.7770e-04
2^{-10}	4	4.9299e-04	0.97	3.1985e-04	4.9844e-04
2^{-11}	4	2.5781e-04	0.94	2.2211e-04	2.6735e-04

Table 5.1.: Example (E): linear rate of convergence for the time discretization for $h = 2^{-5}$

The error $||\bar{u}-\bar{u}_\sigma||_I$ and the experimental order of convergence EOC_k are displayed in 5.1. For further illustration, we include the number of SQP iterations, noting that each quadratic subproblem took at most five primal-dual active set iterations to be solved. Moreover, for completeness, we display the errors $||\bar{y}-\bar{y}_\sigma||$ and $||\bar{p}-\bar{p}_\sigma||$ for the states and adjoint states. The experiments clearly show the proven order

$$\mathcal{O}(k)$$

for the L^2-error in the controls. The convergence behavior is also depicted in the left-hand-side plot of Figure 5.2. Note that the linear convergence behavior can only be expected up to the spatial discretization error. To emphasize this, we repeat the numerical tests with a coarser space discretization of $h = 2^{-3}$. Then, due to the convergence result of Theorem 5.8.1, we expect that the spatial discretization error dominates the total error once we have roughly $k < h^2 = 2^{-6}$. In the right-hand side plot of Figure 5.2, this can be observed.

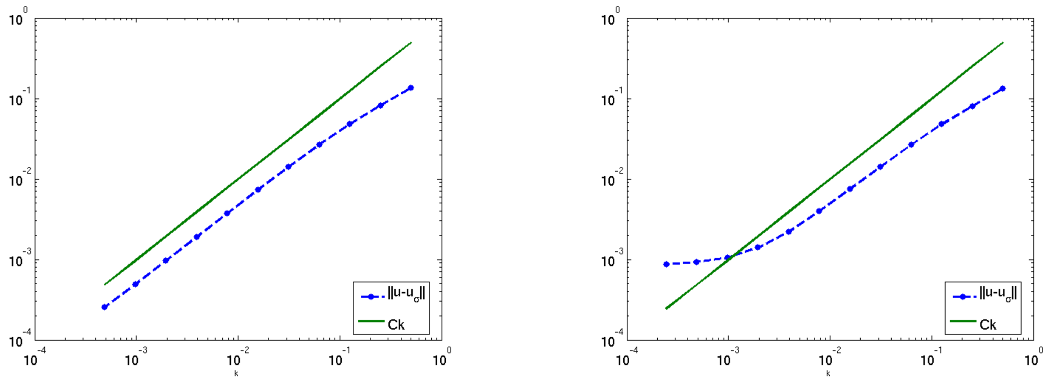


Figure 5.2.: Example (E): Experimental order of convergence for the time discretization for $h = 2^{-5}$ (left) and $h = 2^{-3}$ (right)

h	$\sharp$SQP it	$\|\bar{u} - \bar{u}_\sigma\|_I$	EOC_h	$\|\bar{y} - \bar{y}_\sigma\|_I$	$\bar{p} - \bar{p}_\sigma\|_I$
2^{-1}	5	1.4245e-02	-	5.2684e-02	2.6788e-02
2^{-2}	4	2.8440e-03	2.32	8.1590e-03	5.5432e-03
2^{-3}	4	8.8033e-04	1.69	2.2066e-03	1.7775e-03
2^{-4}	4	2.6771e-04	1.72	5.3353e-04	4.5982e-04

Table 5.2.: Example (E): Experimental order of convergence for the space discretization for $k = 2^{-12}$

5.9.4. The spatial discretization error

In order to verify the spatial discretization error estimate of order $h^{3/2}$ we proceed similarly. The proven order of convergence can only be expected up to the temporal discretization error. Due to the error estimate of Theorem 5.8.1 the temporal discretization has to be much finer than the spatial discretization. Because of that, even in one-dimensional spatial domains the number of unknowns becomes very large. We solve the problem for a temporal discretization parameter $k = 2^{-12}$ and spatial discretization parameter $h = 2^{-1}, \dots, 2^{-4}$. We determine the experimental order of convergence by

$$EOC_h = \frac{\ln(\|\bar{u} - \bar{u}_{\sigma_1}\|_I) - \ln(\|\bar{u} - \bar{u}_{\sigma_2}\|_I)}{\ln(h_1) - \ln(h_2)},$$

where h_1 and h_2 are consecutive mesh sizes, and σ_i denotes the discretization of the PDE and the control with fixed temporal discretization parameter $k = 2^{-12}$ and consecutive spatial discretization parameters h_i. The error $\|\bar{u} - \bar{u}_\sigma\|_I$ and the experimental order of convergence EOC_h are displayed in Table 5.2. We again include the number of SQP iterations as well as the errors $\|\bar{y} - \bar{y}_\sigma\|$ and $\|\bar{p} - \bar{p}_\sigma\|$ for the states and adjoint states. The order of convergence is also displayed in Figure 5.3. The computations underline the theoretical results of Theorem 5.8.1.

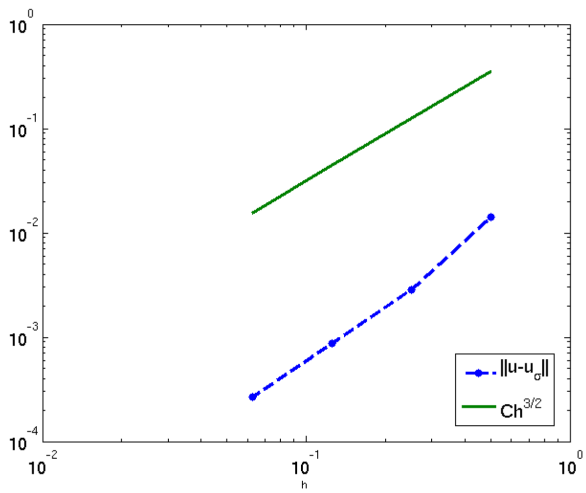


Figure 5.3.: Example (E): Experimental order of convergence for the space discretization for $k = 2^{-12}$

5.9.5. The regularization error

To conclude this section, we demonstrate how a purely state-constrained optimal control problem can be solved via regularization. We consider a special setting of Problem (5.2.1) in $Q := (0,1) \times (0,1)$, given by

$$\text{Minimize } \frac{1}{2} \int_0^1 \int_\Omega \left((y - y_d)^2 + u^2\right) dx dt \tag{E_λ}$$

subject to

$$\begin{aligned} \partial_t y - \Delta y + y^3 &= u + f(t,x) && \text{in } Q \\ y(0,\cdot) &= \sin(\pi x) && \text{in } \Omega, \\ y &= 0 && \text{on } \Sigma, \end{aligned}$$

as well as

$$-1 \leq y \leq 1,$$

with the specific choice

$$\begin{aligned} y_d(t,x) &= \sin(\pi x) + (|x - \frac{1}{2}| - \frac{1}{2}) - 3\sin^2(\pi x)(t-1)\left(|x - \frac{1}{2}| - \frac{1}{2}\right), \\ f(t,x) &= \pi^2 \sin(\pi x) + \sin^3(\pi x) + (t-1)(|x - \frac{1}{2}| - \frac{1}{2}). \end{aligned}$$

The adjoint equation takes the form

$$\begin{aligned} -\partial_t p - \Delta p + 3\bar{y}^2 p &= \bar{y} - y_d + \bar{\mu}_{b,Q} - \bar{\mu}_{a,Q} && \text{in } Q \\ p(1,\cdot) &= \bar{y} - y_d + \bar{\mu}_{b,T} - \bar{\mu}_{a,T} && \text{in } \Omega, \\ p &= 0 && \text{on } \Sigma, \end{aligned}$$

and direct computations verify that

$$\bar{u} = -(t-1)\left(\left|x - \frac{1}{2}\right| - \frac{1}{2}\right)$$

together with the associated state and adjoint state

$$\bar{y} = \sin(\pi x), \quad \bar{p} = (t-1)\left(\left|x - \frac{1}{2}\right| - \frac{1}{2}\right),$$

as well as the Lagrange multipliers $\bar{\mu}_a = 0$ and $\bar{\mu}_b = 2(t-1)\delta_{x=\frac{1}{2}}$ satisfy first order optimality conditions similar to Theorem 4.2.8, i.e. with $U_{\text{ad}} = L^\infty(Q)$. Note in particular that $\bar{u} \in L^\infty(Q)$. Moreover, the problem is constructed with very low regularity of the adjoint state due to the presence of Dirac measures in the right-hand side. Second order sufficient conditions are again verified by noting that

$$-\partial_y d(\cdot, \bar{y})\bar{p} = -3\bar{y}^2 \bar{p} > 0.$$

The problem is regularized by Lavrentiev regularization and transformed into a control-constrained problem as pointed out in Section 5.2. We then solve the regularized problem with discretization parameters $k = 2^{-10}$ and $h = 2^{-5}$ for decreasing values of $\lambda = \lambda_1, \ldots, \lambda_6$. We choose $\lambda_i = 10^{-i}$ and take the solution pertaining to λ_{i-1} as an initial guess for the SQP method for $i > 1$. For an extrapolation based strategy to initialize the solution algorithm we refer to [70] for elliptic problems. In Table 5.3, we display for each λ the number of SQP iterations and the overall number of primal-dual-active set iterations, as well as the L^2-error for the control, the state, and the adjoint state. Of

λ	$\sharp$SQP it	$\sum \sharp$PD it	$\|\bar{u} - \bar{u}_\lambda\|_I$	$\|\bar{y} - \bar{y}_\lambda\|_I$	$\bar{p} - \bar{p}_\lambda\|_I$
10^{-1}	4	17	1.1288e-01	8.4406e-03	1.1617e-01
10^{-2}	3	8	4.8357e-02	2.9222e-03	5.0328e-02
10^{-3}	3	11	6.3851e-03	2.4020e-04	6.7558e-03
10^{-4}	2	7	1.8456e-03	2.4920e-05	2.1536e-03
10^{-5}	2	5	1.7163e-03	6.7960e-06	1.7544e-03
10^{-6}	2	4	1.7162e-03	5.7849e-06	1.7200e-03

Table 5.3.: Example (E_λ): The regularization error

course, the errors displayed in Table 5.3 also contain a certain error due to discretization. Nevertheless, the results indicate that the regularization strategy produces reasonable results. It becomes apparent that the error in the control stagnates once λ becomes too small. Then, the regularization error is dominated by the discretization error. For more experiments on regularization of linear-quadratic parabolic problems that are solved with finite difference schemes we refer to [116]. There, also results on the regularization of boundary control problems can be found.

6. Conclusion and outlook

In this thesis, we have made some new contributions to the theoretical and numerical analysis of nonconvex PDE constrained optimal control problems. The focus has been placed on model problems with pointwise state constraints.

As a first model problem, we have discussed a semi-infinite optimal control problem arising from an elliptic PDE constrained model problem with finite dimensional control space and state constraints on a set of positive measure. While the theory for the model problem with respect to first and second order optimality conditions is known, the novelty of this chapter lies in the a priori finite element error estimates for the control. The available theory of error estimates for elliptic control-constrained problems is rather rich, but less is known for state-constrained optimal control problems, cf. also the references given in Chapters 1 and 3. Relying on some assumptions that are typically made in optimal control, we additionally assumed a certain structure of the active sets that is natural in semi-infinite programming. That is, we assumed that the bound on the state is only active in finitely many points. In contrast to the situation considered in [106], the location of these active points is assumed unknown and the bounds are therefore prescribed in a subset of Ω. The difference to the model problem from [106] also becomes obvious in the final results. Only under certain conditions does the error of the state equation reflect upon the convergence of the discrete optimal controls. This is due to the fact that the controls can vary more freely since the active points of the states may change as well. We provided conditions that ensure an error estimate of order $h^2|\ln h|$ and also demonstrated that in other situations the lower order estimate of order $h\sqrt{|\ln h|}$ is sharp. Throughout, we took into account that the nonconvex problem may admit multiple local solutions. These results have been published for linear-quadratic problems in [104, 105]. The extension to semilinear state equations required some additional considerations such as second order sufficient conditions and finding an adequate substitute for the superposition principle. We have shown some numerical examples to underline the theoretical results. Our main focus has been on the development of the error estimates rather than on the implications in numerical computations. It will be interesting to investigate if and how the results can be of advantage in numerical solution algorithms. Another interesting question is to consider different types of bounds. A fundamental ingredient in our convergence proofs was the fact that due to the constant bounds and the linear finite element approximation the state constraints could be prescribed in the grid points only without changing the admissible sets in the neighborhood of local solutions. For more general bounds on the state, and additional contribution to the final error estimate is to be expected, cf. also [4].

In addition to the state-constrained elliptic problem we were interested in the analysis of time-dependent, parabolic formulations, where we admit control functions that can vary arbitrarily in space and time. The challenges of pointwise state constraints for the considered problem class are well-known. In Chapter 4, we therefore focused on the analysis of a well-established regularization method. We applied Lavrentiev regularization that transforms pointwise state constraints into mixed control-state constraints, and obtained a problem formulation that behaves well with respect to first and second order optimality conditions. There are two main new contributions in this chapter: On the one hand, we provided the stability analysis of an associated linearized problem that eventually led to the strong regularity for generalized equations that represent the first order optimality conditions for the regularized nonlinear problem. This property is a fundamental condition for many results related to the analysis of numerical solution algorithms. We used it primarily to deduce a local uniqueness property of a well-defined locally optimal control, but it could also be used to deduce convergence of SQP methods. Similar results for elliptic control problems have been provided in [7] and [61, 60]. We point out, however, that the lower regularity properties of parabolic equations require special attention. A second important result of this chapter is the regularization error estimate, and the convergence result for regularization parameter λ tending to zero, that were only available for elliptic problems, cf. [31]. Most of the results have been published in [115], cf. also [114].

Finally, Chapter 5 was devoted to the finite element error estimates for certain parabolic optimal control problems. We began this chapter by showing how a problem with Lavrentiev-regularized mixed control-state constraints and no further control constraints can be transformed into a purely-control-constrained model problem, cf. also [111]. This transformation had already been used by Hinze and Meyer, cf. [73] to obtain finite element error estimates for a variational discretization of elliptic Lavrentiev regularized problems. However, the lower regularity properties of the parabolic control-to-state operator made the convergence analysis for Lavrentiev parameters tending to zero a delicate issue, even more so than in Chapter 4. We have proven convergence under certain assumptions adapting arguments from the elliptic setting, cf. [69]. For linear-quadratic problems, this has been published in [116]. We then derived a priori error estimates for the finite element discretization of a parabolic control-constrained model problem in two-dimensional spatial domains with semilinear state equation that also includes Lavrentiev regularized problems discussed at the beginning of Chapter 5. The problems are discretized by a discontinuous Galerkin scheme in time and usual conforming finite elements in space. The controls themselves were discretized piecewise constant in time and cellwise bilinear in space. We extended the results from [102] for linear quadratic problems to problems with semilinear state equation. This is a new contribution to the finite element error analysis of parabolic optimal control problems where the theory is less complete than in the elliptic setting, cf. also the introductory remarks in Chapter 1. For the considered problem class we were able to provide and use a quadratic growth condition for locally optimal controls in an L^2-neighborhood, thus avoiding the need for convergence in $L^\infty(I\times\Omega)$. However, we needed a uniform boundedness result for solutions of the semidiscrete and discrete state equation. The former was obtained by applying Stampacchia's method to the semidiscrete system of equations, whereas the latter followed essentially from appropriate error estimates between the semidiscrete and the discrete level. We eventually obtained an error estimate of order $\mathcal{O}(k + h^{\frac{3}{2}-\varepsilon})$, $\varepsilon > 0$, where k denotes the temporal discretization parameter and h denotes the spatial discretization parameter. These results have been published in [118] for a slightly different objective function, with a detailed proof of convergence for controls that are discretized by piecewise constants in time and space. This change in the objective function was mainly motivated by our interest in state constrained control problems. At the end of Chapter 5 we have also provided some numerical examples that underline the theoretical results.

A. Supplementary results

A.1. Function spaces

We give a brief overview about the function spaces used in this thesis. For further information we point out textbooks on functional analysis, e.g. [1] and partial differential equations, e.g. [48]. An overview relevant for optimal control with PDEs can also be found in [144], whose notation we adopt here and where the following definitions can be found.

A.1.1. L^p-spaces, Sobolev spaces, and spaces of continuous functions

Definition A.1.1. *By* $\{L^p(\Omega), \|\cdot\|_{L^p(\Omega)}\}$, $1 \le p \le \infty$, *we denote the Banach space of all Lebesgue measurable functions* v *defined on* Ω *that satisfy* $\int_\Omega |v(x)|^p dx < \infty$, *or in case of* $p = \infty$ *are essentially bounded, respectively, whose norms are given by*

$$\|v\|_{L^p(\Omega)} := \left(\int_\Omega |v(x)|^p dx\right)^{\frac{1}{p}}, \quad \|v\|_{L^\infty(\Omega)} := ess \sup_{x\in\Omega} |v(x)| := \inf_{|F=0|} \left(\sup_{x\in\Omega\setminus F} |v(x)|\right).$$

The case $p = 2$ is of special interest, and we note that $L^2(\Omega)$ is even a Hilbert space with the inner product

$$(v, w)_{L^2(\Omega)} := \int_\Omega vw dx.$$

Definition A.1.2. *By* $\mathcal{C}^k(\Omega)$, *we denote the linear space of all real-valued functions on* Ω *that, together with their partial derivatives up to order* k, *are continuous on* Ω. *Moreover,* $\mathcal{C}^k(\bar\Omega)$ *denotes the space of all elements of* $\mathcal{C}^k(\Omega)$ *that, together with their partial derivatives up to order* k, *can be continuously extended to* $\bar\Omega$. *We agree on the short notation* $\mathcal{C}(\bar\Omega)$ *instead of* $\mathcal{C}^0(\bar\Omega)$. *The space* $\mathcal{C}^k(\bar\Omega)$ *is a Banach space with norm*

$$\|v\|_{\mathcal{C}(\bar\Omega)} := \max_{x\in\bar\Omega} |v(x)|, \quad \|v\|_{\mathcal{C}^k(\bar\Omega)} := \sum_{|\alpha|\le k} \|D^\alpha v\|_{\mathcal{C}(\bar\Omega)},$$

where α *is a multi-index. Last, for a positive real number* ν, *we introduce the space* $\mathcal{C}^{k,\nu}(\bar\Omega)$ *of all functions* $g \in \mathcal{C}^k(\bar\Omega)$ *that, together with their partial derivatives up to order* k, *are Hölder continuous, i.e.*

$$|D^\alpha g(x_1) - D^\alpha g(x_2)| \le C|x_1 - x_2|^\nu_{\mathbb{R}^n}, \quad \forall\, |\alpha| \le k$$

for a constant $C > 0$.

Since we will look for solutions of the governing PDEs in weak sense, we introduce the Sobolev spaces $W^{k,p}(\Omega)$. We briefly recall that a weak derivative of order $|\alpha|$, $w \in L^1_{loc}(\Omega)$ of $u \in L^1_{loc}(\Omega)$ is defined by

$$\int_\Omega u(x) D^\alpha v(x) dx = (-1)^{|\alpha|} \int_\Omega w(x) v(x) dx \quad \forall\, v \in \mathcal{C}_0^\infty(\Omega),$$

where α is a multiindex, $L^1_{loc}(\Omega)$ denotes the space of locally Lebesgue integrable functions, and the subscript 0 in $\mathcal{C}_0^\infty(\Omega)$ indicates functions with compact support in Ω.

Definition A.1.3. *Let $1 \le p \le \infty$ and $k \in \mathbb{N}$ be given. We denote by $W^{k,p}(\Omega)$ the linear space of all functions $v \in L^p(\Omega)$ having weak derivatives $D^\alpha v$ in $L^p(\Omega)$ for all multi-indices α of length $|\alpha| \le k$, endowed with the norm*

$$\|v\|_{W^{k,p}(\Omega)} = \left(\sum_{|\alpha| \le k} \int_\Omega |D^\alpha v(x)|^p dx \right)^{\frac{1}{p}}, \quad \|v\|_{W^{k,\infty}(\Omega)} = \max_{|\alpha| \le k} \|D^\alpha v\|_{L^\infty(\Omega)},$$

for $p < \infty$ and $p = \infty$, respectively. We write in short $H^k(\Omega) := W^{k,2}(\Omega)$. $H^1(\Omega)$ becomes a Hilbert space with the associated inner product

$$(u, v)_{H^1(\Omega)} := \int_\Omega uv dx + \int_\Omega \nabla u \nabla v dx.$$

To incorporate the Dirichlet boundary conditions in a meaningful way, we introduce

$$H^1_0(\Omega) := W^{1,2}_0(\Omega) = \{v \in H^1(\Omega) \colon v_{|\Gamma} = 0\}, \quad \|v\|_{H^1_0(\Omega)} = \left(\int_\Omega |\nabla|^2 dx \right)^{\frac{1}{2}},$$

where $W^{k,p}_0(\Omega)$ denotes the closure of $\mathcal{C}_0^\infty(\Omega)$.

A.1.2. Spaces of vector valued functions

Spaces like $L^p(\Omega)$ and $\mathcal{C}(\Omega)$ can be extended to other domains, e.g. time-space domains Q, simply by replacing Ω by Q. A fundamental tool for treating nonstationary equations is the concept of abstract functions, i.e. mappings from an interval $[a, b] \subset \mathbb{R}$, e.g. a time interval $I := (0, T)$, into a Banach space X. We will mainly make use of the following spaces, cf. [144].

Let $\{X, \|\cdot\|_X\}$ be a real Banach space.

Definition A.1.4. *We denote the space of all abstract functions that are continuous at every $t \in [a, b]$ by $\mathcal{C}([a, b], X)$. This space is a Banach space with norm $\|v\|_{\mathcal{C}([a,b],X)} = \max_{t \in [a,b]} \|v\|_X$.*

Definition A.1.5. *We denote by $L^p(a, b; X)$, $1 \le p < \infty$, the linear space of all measurable abstract functions $v \colon [a, b] \to X$ with the property*

$$\int_a^b \|v(t)\|_X^p dt < \infty.$$

This space becomes a Banach space when endowed with respect to the norm

$$\|v\|_{L^p(a,b;X)} := \left(\int_a^b \|v(t)\|_X^p dt \right)^{\frac{1}{p}}.$$

The space $L^\infty(a,b;X)$ denotes the space of all measurable abstract functions $v\colon [a,b] \to X$ having the property that

$$\|v\|_{L^\infty(a,b;X)} := ess \sup_{[a,b]} \|v(t)\|_X < \infty.$$

If $X = L^p(\Omega)$ with $p < \infty$, the space $L^p([a,b], L^p(\Omega)$ can be identified with $L^p([a,b] \times \Omega)$. For $p = \infty$, we only obtain $L^\infty([a,b], L^\infty(\Omega)) \subset L^\infty([a,b] \times \Omega)$.

Definition A.1.6. *We denote by $H^1([a,b];X)$ the space of all abstract functions $v\colon [a,b] \to X$ with $v \in L^2(X)$ and $\partial_t v \in L^2(X)$, with norm*

$$\|v\|^2_{H^1([a,b],X)} := \|v\|^2_{L^2(a,b;X)} + \|\partial_t v\|^2_{L^2(a,b;L^2(X)}.$$

Last, let us introduce a state space with less regularity of the time derivatives.

Definition A.1.7. *Let V be a Banach space with dual space V^*. We denote by $\mathcal{W}(0,T)$ the linear space of*

$$\mathcal{W}(0,T) = \{v \in L^2(0,T,V) \mid \partial_t v \in L^2(0,T,V^*\}$$

with the norm

$$\|v\|_{\mathcal{W}(0,T)} = \left(\int_0^T (\|v(t)\|_V^2 + \|v'(t)\|^2_{V^*}) dt \right)^{\frac{1}{2}}$$

We mention here that the dual space of a Banach space $\{X, \|\cdot\|_X\}$, denoted by X^*, is the space of all continuous linear functionals on $\{X, \|\cdot\|_X\}$. For Hilbert spaces $(H, (\cdot,\cdot))$, the Riesz representation theorem guarantees that for each continuous linear functional $F \in H^*$ there exists a uniquely determined $f \in H$ such that

$$\|F\|_{H^*} = \|f\|_H, \quad F(v) = (f,v)_H \quad \forall v \in H.$$

For L^p-spaces with $1 < p < \infty$, we have

$$L^p(\Omega)^* \cong L^q(\Omega), \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Also, we have

$$L^1(\Omega)^* \cong L^\infty(\Omega),$$

but unfortunately $L^\infty(\Omega)^*$ cannot be identified with $L^1(\Omega)$. Instead, the elements of $L^\infty(\Omega)^*$ admit very low regularity, which makes $L^\infty(\Omega)$ an unsuitable state space in the context of state-constrained optimal control problems. We refer to [134] for a discussion of Lagrange multipliers in these spaces. On the other hand, the dual space of $\mathcal{C}(\bar{\Omega})$ can at least be identified with the space of regular Borel measures, $\mathcal{M}(\bar{\Omega})$, cf. [6].

A.2. Some properties of the control-to-state operators

We will provide the proof of Lemma 3.1.4, which states the Lipschitz-continuity of the control-to-state-operator and the objective function and their derivatives for the elliptic semi-infinite problem, and its parabolic counterpart, Lemma 4.2.4. The techniques are similar, yet in the elliptic case we only deal with finitely many control parameters, whereas the functions from the parabolic setting require some caution with respect to the appearing norms.

A.2.1. Proof of Lemma 3.1.4

Proof. For u_1, u_2 and v given consider $y_i = G(u_i)$, $\tilde{y}_i = G'(u_i)v$ and $\hat{y}_i = G''(u_i)[v,v]$. For the first estimate, it will be convenient to use the weak formulations of the appearing PDEs. We first prove (3.1.6). The difference $z := y_1 - y_2$ fulfills

$$\mathbf{a}(z,\varphi) \quad = \quad (d(\cdot,y_2,u_2) - d(\cdot,y_1,u_2),\varphi) + (d(\cdot,y_1,u_2) - d(\cdot,y_1,u_1),\varphi) \quad \forall\,\varphi \in V. \tag{A.2.1}$$

Testing (A.2.1) with $\varphi = z$ yields

$$c\,\|z\|^2_{H^1_0(\Omega)} \leq (d(\cdot,y_2,u_2) - d(\cdot,y_1,u_2),z) + (d(\cdot,y_1,u_2) - d(\cdot,y_1,u_1),z) \tag{A.2.2}$$

by the $H^1_0(\Omega)$-ellipticity of $\mathbf{a}$. The monotonicity of d with respect to y implies

$$(d(\cdot,y_2,u_2) - d(\cdot,y_1,u_2),z) \leq 0,$$

and the local Lipschitz continuity with respect to u gives rise to

$$(d(\cdot,y_1,u_2) - d(\cdot,y_1,u_1),z) \leq c\|d(\cdot,y_1,u_2) - d(\cdot,y_1,u_1)\|\|z\| \leq c|u_1 - u_2|\|z\|,$$

where c can be taken independently of u_1, u_2 by the boundedness of U_{ad}. Insertion of both estimates in (A.2.2) yields

$$\|z\|^2_{H^1_0(\Omega)} \leq c|u_1 - u_2|\|z\| \leq c|u_1 - u_2|\|z\|_{H^1_0(\Omega)},$$

due to Friedrich's inequality applied to the left hand side. After dividing by $\|z\|_{H^1_0(\Omega)}$ we obtain Lipschitz stability in $H^1(\Omega)$ and thus also in $L^2(\Omega)$. The estimate in the norm $\mathcal{C}(\bar{\Omega})$ follows by applying Proposition 2.4.2 to the linear equation

$$\begin{aligned} \mathcal{A}z &= d(\cdot,y_2,u_2) - d(\cdot,y_1,u_1) && \text{in } \Omega, \\ z &= 0 && \text{on } \Gamma, \end{aligned}$$

which due to the Lipschitz properties of d and the fact that $\|y_1 - y_2\|$ can be estimated as just proven yields

$$\|z\|_{\mathcal{C}(\bar{\Omega})} \leq c\,(\|y_1 - y_2\| + |u_1 - u_2|) \leq c|u_1 - u_2|.$$

Now, to prove (3.1.7), consider the difference $\tilde{z} = \tilde{y}_1 - \tilde{y}_2$, which fulfills

$$\begin{aligned} \mathcal{A}\tilde{z} + \partial_y d(\cdot,y_1,u_1)\tilde{z} &= (\partial_y d(\cdot,y_2,u_2) - \partial_y d(\cdot,y_1,u_1))\tilde{y}_2 + (\partial_u d(\cdot,y_2,u_2) - \partial_u d(\cdot,y_1,u_1))v && \text{in } \Omega \\ \tilde{z} &= 0 && \text{on } \Gamma \end{aligned}$$

By the Lipschitz continuity of $\partial_u d$ and $\partial_y d$, we obtain

$$\|\tilde{z}\|_{L^2(\Omega)\cap H^1(\Omega)\cap\mathcal{C}(\bar{\Omega})} \leq c(\|y_1 - y_2\|_\infty + |u_1 - u_2|)(\|\tilde{y}_2\| + |v|) \leq c|u_1 - u_2||v|,$$

where we applied Proposition 2.4.2, and the Lipschitz estimate (3.1.6), as well as the boundedness of U_{ad}. Estimate (3.1.8), i.e. the Lipschitz stability for the second derivative G'' is proven in a similar way, taking into account the Lipschitz continuity of d''. To demonstrate this consider $\hat{z} := \hat{y}_1 - \hat{y}_2$, which fulfills

$$\begin{aligned} \mathcal{A}\hat{z} + \partial_y d(\cdot, y_1, u_1)\hat{z} = \quad & (\partial_y d(\cdot, y_2, u_2) - \partial_y d(\cdot, y_1, u_1))\hat{y}_2 + (\tilde{y}, 0^T)d''(\cdot, y_2, u_2)(\tilde{y}_2, v^T)^T \\ & +(\tilde{y}_1, v^T)(d''(\cdot, y_2, u_2) - d''(\cdot, y_1, u_1))(\tilde{y}_2, v^T)^T \\ & +(\tilde{y}_1, v^T)d''(\cdot, y_1, u_1))(\tilde{y}, 0^T)^T && \text{in } \Omega \\ \hat{z} = \quad & 0 && \text{on } \Gamma. \end{aligned}$$

Then, Proposition 2.4.2 combined with the Lipschitz properties of d and its derivatives, as well as the boundedness of U_{ad} yields

$$\begin{aligned} \|z\|_{L^2(\Omega)\cap H^1(\Omega)\cap \mathcal{C}(\Omega)} \leq & c(\|y_1 - y_2\|_\infty + |u_1 - u_2|)\|\hat{y}_2\| + c(\|\tilde{y}\|_\infty(\|\tilde{y}_2\|_\infty + |v|) \\ & + c(\|\tilde{y}_1\|_\infty + |v|)(\|y_1 - y_2\|_\infty + |u_1 - u_2|)(\|\tilde{y}_2\|_\infty + |v|) + c\|\tilde{y}\|_\infty(\|\tilde{y}_1\|_\infty + |v|). \end{aligned}$$

Note that some parts of the right-hand-side could have been estimated in weaker norms. Nevertheless, noting that

$$\|y_1 - y_2\|_\infty \leq c|u_1 - u_2|$$

by (3.1.6), as well as

$$\|\tilde{y}\|_\infty + \|\tilde{y}_1\|_\infty + \|\tilde{y}_2\|_\infty \leq c|v|$$

by Proposition 2.4.2, which then implies

$$\|z_2\| \leq c|v|^2,$$

we finally obtain

$$\|\hat{y}\|_{L^2(\Omega)\cap H^1(\Omega)\cap \mathcal{C}(\Omega)} \leq c|u_1 - u_2||v|^2.$$

For the objective function, we proceed in a similar way. Estimate (3.1.9) is a simple conclusion of the Lipschitz properties of Ψ and (3.1.6), since

$$|f(u_1) - f(u_2)| \leq \int_\Omega |\Psi(x, y_1(x), u_1) - \Psi(x, y_2(x), u_2)|\, dx \leq c(\|y_1 - y_2\| + |u_1 - u_2|) \leq c\,|u_1 - u_2|$$

is satisfied. To prove (3.1.10), not that the first derivative satisfies

$$\begin{aligned} |f'(u_1)v - f'(u_2)v| \leq & \int_\Omega |\partial_u \Psi(x, y_1, u_1)v - \partial_u \Psi(x, y_2, u_2)v + \partial_y \Psi(x, y_1, u_1)\tilde{y}_1 - \partial_y \Psi(x, y_2, u_2)\tilde{y}_2|\, dx \\ \leq & \int_\Omega |\,(\partial_u \Psi(x, y_1, u_1) - \partial_u \Psi(x, y_2, u_2))\, v + \partial_y \Psi(x, y_1, u_1)(\tilde{y}_1 - \tilde{y}_2) \\ & \quad + (\partial_y \Psi(x, y_1, u_1) - \partial_y(x, y_2, u_2))\, \tilde{y}_2|\, dx \\ \leq & c \quad (\|y_1 - y_2\|_\infty + |u_1 - u_2|)\, |v| + c|u_1 - u_2||v| + c\,(\|y_1 - y_2\|_\infty + |u_1 - u_2|)\, |v| \\ \leq & c \quad |u_1 - u_2||v| \end{aligned}$$

by (3.1.9) as well as Proposition 2.4.2. Estimate (3.1.11) for the second derivative, we obtain by straightforward calculations:

$$\begin{aligned} |f''(u_1)[v, v] - f''(u_2)[v, v]| \leq \int_\Omega & |(\tilde{y}, 0^T)\Psi''(\cdot, y_2, u_2)(\tilde{y}_2, v^T)^T \\ & + (\tilde{y}_1, v^T)(\Psi''(\cdot, y_2, u_2) - \Psi''(\cdot, y_1, u_1))(\tilde{y}_2, v^T)^T \\ & + ((\tilde{y}_1, v^T)\Psi''(\cdot, y_1, u_1))(\tilde{y}, 0^T)^T|\, dx, \end{aligned}$$

The right-hand-side can be estimated completely analogously to the previous estimates. □

A.2.2. Proof of Lemma 4.2.4

Proof. We begin with proving the Lipschitz continuity result for the states, as in [144, Theorem 5.8]. Denote in short $y_1 := G(u_1)$ and $y_2 := G(u_2)$. Utilizing the properties of d, it is easily verified that the difference $y := y_1 - y_2$ fulfills

$$\begin{aligned} \partial_t y + \mathcal{A} y + d_\xi(t,x) y &= u_1 - u_2 && \text{in } Q \\ y(0,\cdot) &= 0 && \text{in } \Omega \\ y &= 0 && \text{on } \Sigma, \end{aligned}$$

with

$$d_\xi(t,x) = \int_0^1 \partial_y d(t,x,y_1(t,x) + \xi(y_2(t,x) - y_1(t,x)))d\xi$$

The function d_ξ is bounded in $L^\infty(Q)$ independently from y_1 and y_2 by the boundedness of U_{ad}. The first two assertions (4.2.5) and (4.2.6) then follow from Proposition 2.5.2 and the boundedness of U_{ad}.

To prove the estimate (4.2.7) for the first derivative of G, we set $\tilde{y}_1 := G'(u_1)v$ and $\tilde{y}_2 := G'(u_2)v$, and note that the difference $\tilde{y} := \tilde{y}_1 - \tilde{y}_2$ fulfills

$$\begin{aligned} \partial_t \tilde{y} + \mathcal{A}\tilde{y} + \partial_y d(\cdot, y_1)\tilde{y} &= -(\partial_y d(\cdot,y_1) - \partial_y d(\cdot,y_2))\tilde{y}_2 && \text{in } Q \\ \tilde{y}(0,\cdot) &= 0 && \text{in } \Omega \\ \tilde{y} &= 0 && \text{on } \Sigma. \end{aligned}$$

Hence from the Lipschitz continuity of $\partial_y d$, Proposition 2.5.2 and the proven Lipschitz continuity of the control-to-state operator, we obtain

$$\begin{aligned} \|\tilde{y}\|_I \le c\,\|(\partial_y d(\cdot,y_1) - \partial_y d(\cdot,y_2))\tilde{y}_2\|_I &\le c\,\|y_1 - y_2\|_{L^4(Q)}\|\tilde{y}_2\|_{L^4(Q)} \\ &\le c\,\|y_1 - y_2\|_{L^\infty(I,V)}\|\tilde{y}_2\|_{L^\infty(I,V)} \\ &\le c\,\|u_1 - u_2\|_I\|v\|_I \end{aligned}$$

by the embedding

$$L^\infty(I,V) \hookrightarrow L^4(Q).$$

The estimate (4.2.8) in the $L^\infty(Q)$ norm follows by estimation of the $L^\infty(Q)$-norm of the right-hand-side.

The estimate (4.2.9) for G'' is an immediate consequence of the Lipschitz continuity of G, G', and the boundedness of U_{ad} in conjunction with the second order Lipschitz and boundedness properties of d. We observe

$$\begin{aligned} G''(u_1)[v,v] - G''(u_2)[v,v] = \; & (G'(u_2) - G'(u_1))\partial_{yy} d(\cdot,y_2)[\tilde{y}_2,\tilde{y}_2] \\ & + G'(u_1)\partial_{yy}(d(\cdot,y_2) - d(\cdot,y_1))[\tilde{y}_2,\tilde{y}_2] \\ & + G'(u_1)\partial_{yy} d(\cdot,y_1))([\tilde{y}_2,\tilde{y}_2] - [\tilde{y}_1,\tilde{y}_1]), \end{aligned}$$

hence Proposition 2.5.2 allows to estimate

$$\|G''(u_1)[v,v] - G''(u_2)[v,v]\|_I \le c\,\|u_1 - u_2\|_I\|\tilde{y}_2\|^2_{L^4(Q)} + c\,\|y_1 - y_2\|_I\|\tilde{y}_2\|^2_{L^4(Q)} + (\|\tilde{y}_1\|_I + \|\tilde{y}_2\|_I)\|\tilde{y}_1 - \tilde{y}_2\|_I.$$

The assertion is then obtained by the embedding $L^\infty(I,V) \hookrightarrow L^4(Q)$ and the stability estimates from Proposition 2.5.2 and Theorem 2.5.6 applied to the right-hand-side.

The first estimate (4.2.10) for the objective function is trivial. Direct calculations show that

$$
\begin{aligned}
&|f'(u_1)v - f'(u_2)v| \\
&\quad \le \iint_Q |\partial_y \Psi(t,x,y_1,u_1)\tilde{y}_1 - \partial_y \Psi(t,x,y_2,u_2)\tilde{y}_2 + \partial_u \Psi(t,x,y_1,u_1)v - \partial_u \Psi(t,x,y_2,u_2)v|\, dxdt \\
&\quad \le \iint_Q |\partial_y \Psi(t,x,y_1,u_1)(\tilde{y}_1 - \tilde{y}_2) + (\partial_y \Psi(t,x,y_1,u_1) - \partial_y \Psi(t,x,y_2,u_2))\tilde{y}_2 \\
&\qquad\qquad + (\partial_u \Psi(t,x,y_1,u_1) - \partial_u \Psi(t,x,y_2,u_2))v|\, dxdt
\end{aligned}
$$

Then, by the Lipschitz properties of Ψ and its derivatives, we can estimate

$$
|f'(u_1)v - f'(u_2)v| \le c\|\tilde{y}_1 - \tilde{y}_2\|_I + c(\|y_1 - y_2\|_I + \|u_1 - u_2\|_I)(\|\tilde{y}_2\|_I + \|v\|_I).
$$

Applying Lipschitz results for G and G', and Proposition 2.5.2 to the right-hand-side yields the desired estimate (4.2.11) for the first derivative. Moreover, we have

$$
\begin{aligned}
&|f''\ (u_1)[v,v] - f''(u_2)(v,v)| \\
&\le \iint_Q \Bigg| \left(\begin{bmatrix} \tilde{y} \\ 0 \end{bmatrix}^T \begin{bmatrix} \partial_{yy}\Psi(t,x,y_1,u_1) & \partial_{yu}\Psi(t,x,y_1,u_1) \\ \partial_{uy}\Psi(t,x,y_1,u_1) & \partial_{uu}\Psi(t,x,y_1,u_1) \end{bmatrix} \begin{bmatrix} \tilde{y}_1 \\ v \end{bmatrix} \right. \\
&\qquad + \begin{bmatrix} \tilde{y}_2 \\ v \end{bmatrix}^T \left(\begin{bmatrix} \partial_{yy}\Psi(\cdot,y_1,u_1) & \partial_{yu}\Psi(\cdot,y_1,u_1) \\ \partial_{uy}\Psi(\cdot,y_1,u_1) & \partial_{uu}\Psi(\cdot,y_1,u_1) \end{bmatrix} - \begin{bmatrix} \partial_{yy}\Psi(\cdot,y_2,u_2) & \partial_{yu}\Psi(\cdot,y_2,u_2) \\ \partial_{uy}\Psi(\cdot,y_2,u_2) & \partial_{uu}\Psi(\cdot,y_2,u_2) \end{bmatrix} \right) \begin{bmatrix} \tilde{y}_1 \\ v \end{bmatrix} \\
&\qquad \left. + \begin{bmatrix} \tilde{y}_2 \\ v \end{bmatrix}^T \begin{bmatrix} \partial_{yy}\Psi(\cdot,y_2,u_2) & \partial_{yu}\Psi(\cdot,y_2,u_2) \\ \partial_{uy}\Psi(\cdot,y_2,u_2) & \partial_{uu}\Psi(\cdot,y_2,u_2) \end{bmatrix} \begin{bmatrix} \tilde{y} \\ 0 \end{bmatrix} \right) \Bigg|\, dxdt \\
&\quad + \iint_Q (\partial_y \Psi(t,x,y_1,u_1)(\hat{y}) + (\partial_y \Psi(t,x,y_1,u_1) - \partial_y \Psi(t,x,y_2,u_2)\hat{y}_2))\ dxdt.
\end{aligned}
$$

Then we estimate

$$
\begin{aligned}
|f''(u_1)[v,v] - f''(u_2)[v,v]| \le\ & c(\|\tilde{y}_1 - \tilde{y}_2\|_I(\|\tilde{y}_1\|_I + \|v\|_I) \\
& +(\|\tilde{y}_2\|_I + \|v\|_I(\|y\|_{\infty,\infty} + \|u_1 - u_2\|_{\infty,\infty})(\|\tilde{y}_1\|_I + \|v\|_I) \\
& \ (\|\tilde{y}_2\|_I + \|v\|_I)\|\tilde{y}_1 - \tilde{y}_2\|_I + \|\hat{y}\|_I + (\|y\|_I + \|u_1 - u_2\|_I)\|\hat{y}_2\|_I).
\end{aligned}
$$

The boundedness of U_{ad} and Ψ'', Proposition 2.5.2 as well as the Lipschitz continuity results for G, G' and G'' yield the last estimate (4.2.12). □

A.3. Proof of Theorem 5.5.1

Proof. Let k and h be fixed. We discuss only the first time interval I_1. We obtain the formulation

$$
k_1(\nabla y_{kh,1}, \nabla\varphi_1) + (y_{kh,1}, \varphi_1) + (d(t,x,y_{kh,1}), \varphi_1)_{I_1} = (u, \varphi_1)_{I_1} + (y_0, \varphi_1) \quad \forall \varphi_1 \in V_h,
$$

where u and y_0 are given. For better readability, we omit the index 1 in k_1, $y_{kh,1}$, and φ_1 and simply write k, y_{kh}, and φ instead. With the auxiliary functions

$$\begin{aligned}\tilde{d}(x, y_{kh}) &:= \int_{I_1} d(t, x, y_{kh}(x))dt, \\ \tilde{u}(x) &:= \int_{I_1} u(t, x)dt,\end{aligned}$$

this yields

$$k(\nabla y_{kh}, \nabla\varphi) + (y_{kh}, \varphi_1) + (\tilde{d}(\cdot, y_{kh}), \varphi_1) = (\tilde{u}, \varphi_1) + (y_0, \varphi_1) \quad \forall\varphi \in V_h, \tag{A.3.1}$$

cf. the proof of Theorem 5.4.8. This is a standard finite element discretization of an elliptic equation, where existence theory is available. We refer to [96] for more details, but give the proof for convenience of the reader.

Let n_h denote the dimension of V_h and let $\{\Phi_i\}$ denote a basis of V_h. Writing y_{kh} as a linear combination of the basis functions Φ_i, $i = 1, \ldots, n_h$, tacitly identifying y_{kh} with its coefficient vector, and testing (A.3.1) with each basis function yields

$$kSy_{kh} + My_{kh} = -D(y_{kh}) + \hat{y}_0, \tag{A.3.2}$$

where $S \in \mathbb{R}^{n_h \times n_h}$ is a stiffness matrix with entries $S_{ij} = (\nabla\Phi_i, \nabla\Phi_j)$, $M \in \mathbb{R}^{n_h \times n_h}$ is a mass matrix with entries $M_{ij} = (\Phi_i, \Phi_j)$, $\hat{y}_0 \in \mathbb{R}^{n_h}$ accommodates y_0, and $D \in \mathbb{R}^{n_h}$ takes into account the nonlinear terms

$$D_i = \left(\tilde{d}(x, \sum_{j=1}^{n_h} y_{kh}^j \Phi_j), \Phi_i\right),$$

and remains locally Lipschitz continuous. For clarity of presentation, assume that $D(0) = 0$. We apply a truncation argument of the form

$$D_b = \begin{cases} D(y_{kh}) & \text{if } |D(y_{kh})| \le b \\ b\frac{D(y_{kh})}{y_{kh}} & \text{if } |D(y_{kh})| \ge b \end{cases}$$

For every $z \in \mathbb{R}^{n_h}$, the equation

$$(kS + M)y_z = -D_b(z) + \hat{y}_0$$

admits a unique solution

$$y_z = (kS + M)^{-1}(-D_b(z) + \hat{y}_0).$$

The norm of y_z can be estimated with the help of the smallest eigenvalue λ of $kS + M$,

$$|y_z| \le \lambda^{-1}(b + |y_0|). \tag{A.3.3}$$

Brouwer's fixed point theorem yields the existence of a solution y_{kh}^b of

$$(kS + M)y_{kh}^b = -D_b(y_{kh}^b) + \hat{y}_0. \tag{A.3.4}$$

Multiplying (A.3.4) with y_{kh}^b yields

$$\lambda|y_{kh}^b|^2 \le \langle y_{kh}^b, (kS + M)y_{kh}^b\rangle = -\langle y_{kh}^b, D_b(y_{kh}^b)\rangle + \langle y_{kh}^b, \hat{y}_0\rangle \le \langle y_{kh}^b, \hat{y}_0\rangle \le |y_{kh}^b||\hat{y}_0|$$

by the monotonicity of D and therefore also D_b. Then we deduce

$$|y_{kh}^b| \le \frac{|\hat{y}_0|}{\lambda},$$

which in particular implies boundedness of y_{kh}^b independent of the truncation bound b. Since D is Lipschitz continuous on the closed ball $\bar{B}(0, \frac{\hat{y}_0}{\lambda})$, we have

$$|D(y_{kh}^b)| \le \hat{y}_0 \frac{|\hat{y}_0|}{\lambda} \quad \forall b > 0.$$

If we choose $b > \hat{y}_0 \lambda^{-1}|b|$, we obtain $D(y_{kh}^b) = D_b(y_{kh}^b)$, and we have guaranteed the existence of a solution y_{kh} to (A.3.1). Uniqueness is an immediate consequence of the monotonicity of d by assumption of the contrary. $\square$

B. Notation

We give a brief overview about some frequently appearing symbols and abbreviations.

Symbols, sets, and spaces

∇	Gradient
Δ	Laplace operator
∇^2	Hessian
$\sharp A$	cardinality of the set A
Ω	spatial domain
$I = (0,T)$	time interval
$Q = I \times \Omega$	time-space domain
$V = H_0^1(\Omega)$	
$I_m = (t_{m-1}, t_m]$	subinterval of I

Inner products and norms

$$
\begin{aligned}
\langle \cdot, \cdot \rangle &:= \langle \cdot, \cdot \rangle_{\mathbb{R}^n}, \\
(\cdot, \cdot) &:= (\cdot, \cdot)_{L^2(\Omega)}, \\
(\cdot, \cdot)_I &:= (\cdot, \cdot)_{L^2(Q)}, \\
\|\cdot\|_{I_m} &:= \|\cdot\|_{L^2(I_m, L^2(\Omega))} \\
|\cdot| &:= |\cdot|_{\mathbb{R}^n} \\
\|\cdot\| &:= \|\cdot\|_{L^2(\Omega)} \\
\|\cdot\|_\infty &:= \|\cdot\|_{L^\infty(\Omega)} \\
\|\cdot\|_I &:= \|\cdot\|_{L^2(Q)} \\
\|\cdot\|_{\infty,2} &:= \|\cdot\|_{L^\infty(I, L^2(\Omega))}, \\
\|\cdot\|_{\infty,\infty} &:= \|\cdot\|_{L^\infty(Q)} \\
(\cdot, \cdot)_{I_m} &:= (\cdot, \cdot)_{L^2(I_m, L^2(\Omega))}, \\
\|\cdot\|_{\infty,2} &:= \|\cdot\|_{L^\infty(I, L^2(\Omega))}
\end{aligned}
$$

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] 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.

[3] H. W. Alt. *Lineare Funktionalanalysis.* Springer-Verlag, Berlin, 1999.

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

[5] W. Alt. The Lagrange-Newton method for infinite dimensional optimization problems. *Numerical Functional Analysis and Optimization*, 11:201–224, 1990.

[6] W. Alt. *Nichtlineare Optimierung.* Vieweg, Braunschweig, 2002.

[7] W. Alt, R. Griesse, N. Metla, and A. Rösch. Lipschitz stability for elliptic optimal control problems with mixed control-state constraints. *Optimization*, 59:833–849, 2010.

[8] 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.

[9] N. Arada and J. P. Raymond. Dirichlet boundary control of semilinear parabolic equations. Part 1: Problems with no state constraints. *Appl. Math. Optim.*, 45:125–143, 2002.

[10] M. Bergounioux, M. Haddou, M. Hintermüller, and K. Kunisch. A comparison of a Moreau-Yosida-based active set strategy and interior point methods for constrained optimal control problems. *SIAM J. Optimization*, 11:495–521, 2000.

[11] M. Bergounioux and K. Kunisch. On the structure of the Lagrange multipliers for state-constrained optimal control problems. *Systems and Control Letters*, 48:16–176, 2002.

[12] M. Bergounioux and K. Kunisch. Primal-dual active set strategy for state-constrained optimal control problems. *Computational Optimization and Applications*, 22:193–224, 2002.

[13] M. Bergounioux and F. Tröltzsch. Optimal control of linear bottleneck problems. *ESAIM COCV*, 3:235–250, 1998.

[14] M. Bergounioux and F. Tröltzsch. Optimal control of semilinear parabolic equations with state-constraints of bottleneck type. *ESAIM COCV*, 4:595–608, 1999.

[15] J. T. Betts and S. L. Campbell. Discretize then Optimize. In D. R. Ferguson and T. J. Peters, editors, *Mathematics in Industry: Challenges and Frontiers A Process View: Practice and Theory.* SIAM Publications, Philadelphia, 2005.

[16] F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems.* Springer, New York, 2000.

[17] E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM J. Control and Optimization*, 4:1309–1322, 1986.

[18] E. Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM J. Control and Optimization*, 31:993–1006, 1993.

[19] E. Casas. Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control and Optimization*, 35:1297–1327, 1997.

[20] E. Casas. Error estimates for the numerical approximation of semilinear elliptic control problems with finitely many state constraints. *ESAIM: Control, Optimization and Calculus of Variations*, 31:345–374, 2002.

[21] E. Casas. Using piecewise linear functions in the numerical approximation of semilinear elliptic control problems. *Advances in Computational Mathematics*, 26:137–153, 2007.

[22] E. Casas, J. de los Reyes, and F. Tröltzsch. Sufficient second-order optimality conditions for semilinear control problems with pointwise state constraints. *SIAM J. on Optimization*, 19(2):616–643, 2008.

[23] E. Casas and M. Mateos. Uniform convergence of the FEM. Applications to state constrained control problems. *J. of Computational and Applied Mathematics*, 21:67–100, 2002.

[24] E. Casas and M. Mateos. Error estimates for the numerical approximation of boundary semilinear elliptic control problems. Continuous piecewise linear approximations. In *Systems, control, modeling and optimization*, volume 202 of *IFIP Int. Fed. Inf. Process.*, pages 91–101. Springer, New York, 2006.

[25] 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.

[26] E. Casas and F. Tröltzsch. Error estimates for the finite-element approximation of a semilinear elliptic control problem. *Control and Cybernetics*, 31:695–712, 2002.

[27] E. Casas and F. Tröltzsch. Error estimates for linear-quadratic elliptic control problems. In *Analysis and Optimization of Differential Systems.* Kluwer Academic Publishers, 2003.

[28] E. Casas, F. Tröltzsch, and A. Unger. Second order sufficient optimality conditions for a nonlinear elliptic control problem. *J. for Analysis and its Applications*, 15:687–707, 1996.

[29] E. Casas, F. Tröltzsch, and A. Unger. Second order sufficient optimality conditions for some state-constrained control problems of semilinear elliptic equations. *SIAM Control Optim.*, 38:1369–1391, 2000.

[30] S. Cherednichenko, K. Krumbiegel, and A. Rösch. Error estimates for the Lavrentiev regularization of elliptic optimal control problems. *Inverse Problems*, 24(5):055003, 2008.

[31] S. Cherednichenko and A. Rösch. Error estimates for the regularization of optimal control problems with pointwise control and state constraints. *ZAA*, 27(2):195–212, 2008.

[32] S. Cherednichenko and A. Rösch. Error estimates for the discretization of elliptic control problems with pointwise control and state constraints. *COAP*, 44:27–55, 2009.

[33] K. Chrysafinos. Convergence of discontinuous Galerkin approximations of an optimal control problem associated to semilinear parabolic pde's. *ESAIM: M2AN*, 44:189–206, 2010.

[34] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[35] P. Clément. Approximation by finite element functions using local regularization. *Revue Franc. Automat. Inform. Rech. Operat.*, 9(2):77–84, 1975.

[36] J. C. de los Reyes, P. Merino, J. Rehberg, and F. Tröltzsch. Optimality conditions for state-constrained PDE control problems with finite-dimensional control space. *Control Cybern.*, 37:5–38, 2008.

[37] 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 Control Optim.*, 48(4):2798–2819, 2009.

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

[39] K. Deckelnick and M. Hinze. Numerical analysis of a control and state constrained elliptic control problem with piecewise constant control approximations. In K. Kunisch, G. Of, and O. Steinbach, editors, *Proceedings of ENUMATH 2007, the 7th European Conference on Numerical Mathematics and Advanced Applications*, Heidelberg, September 2007 2008. Springer.

[40] K. Deckelnick and M. Hinze. Variational discretization of parabolic control problems in the presence of pointwise state constraints. *J. Comp. Math.*, 29:1–16, 2011.

[41] P. Deuflhardt, M. Seebass, D. Stalling, R. Beck, and H.-C. Hege. Hyperthermia treatment planning in clinical cancer therapy: Modelling, simulation, and visualization. In A. Sydow, editor, *Computational Physics, Chemistry and Biology*, pages 9–17. Wissenschaft und Technik- Verlag, 1997.

[42] A. L. Dontchev. Local analysis of a Newton–type method based on partial linearization. In J. Renegar, M. Shub, and S. Smale, editors, *Proceedings of the AMS–SIAM Summer Seminar in Applied Mathematics on Mathematics of Numerical Analysis: Real Number Algorithms*, volume 32 of *AMS Lectures in Appl. Math.*, pages 295–306, 1996.

[43] K. Eppler and F. Tröltzsch. Fast optimization methods in the selective cooling of steel. In M. Grötschel, S. O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems*, pages 185–204. Springer, 2001.

[44] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. *Computational Differential Equations*. Cambridge Univ. Press, UK, 1996.

[45] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems I: A linear model problem. *SIAM J. Numer. Anal.*, 28(1):43–77, 1991.

[46] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems II: Optimal error estimats in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM J. Numer. Anal.*, 32(3):706–740, 1995.

[47] K. Eriksson, C. Johnson, and V. Thomée. Time discretization of parabolic problems by the discontinuous Galerkin method. *M2AN Math. Model. Numer. Anal.*, 19:611–643, 1985.

[48] L. C. Evans. *Partial Differential Equations*. American Math. Society, Providence, Rhode Island, 1998.

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

[50] T. Geveci. On the approximation of the solution of an optimal control problem governed by an elliptic equation. *R.A.I.R.O. Analyse numérique/ Numerical Analysis*, 13:313–328, 1979.

[51] D. Gilbarg and N. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer, 1977.

[52] H. Goldberg and F. Tröltzsch. Second order sufficient optimality conditions for a class of control problems governed by non-linear integral equations with application to parabolic boundary control. *Optimization*, 20:671–698, 1989.

[53] H. Goldberg and F. Tröltzsch. Second order sufficient optimality conditions for a class of non-linear parabolic boundary control problems. *SIAM Control Optim.*, 31:1007–1025, 1993.

[54] H. Goldberg and F. Tröltzsch. On a Lagrange-Newton method for a nonlinear parabolic boundary control problem. *Optimization Methods and Software*, 8(3):225–247, 1998.

[55] G. Gramlich, R. Hettich, and E. Sachs. Local convergence of SQP methods in semi-infinite programming. *SIAM J. Optim.*, 5:641–658, 1995.

[56] J. A. Griepentrog. Linear elliptic boundary value problems with non-smooth data: Campanato spaces of functionals. *Math. Nachr.*, 243:19–42, 2002.

[57] J. A. Griepentrog. Maximal regularity for nonsmooth parabolic problems in Sobolev-Morrey spaces. *Adv. Differential Eequations*, 12:1031–1078, 2007.

[58] J. A. Griepentrog. Sobolev-Morrey spaces associated with evolution equations. *Adv. Differential Equations*, 12:781–840, 2007.

[59] R. Griesse. Lipschitz stability of solutions to some state-constrained elliptic optimal control problems. *Zeitschrift für Analysis und ihre Anwendungen*, 25(4):435–455, 2006.

[60] R. Griesse, N. Metla, and A. Rösch. Convergence analysis of the SQP method for nonlinear mixed-constrained elliptic optimal control problems. *ZAMM*, 88(10):776–792, 2008.

[61] R. Griesse, N. Metla, and A. Rösch. Local quadratic convergence of SQP for elliptic optimal control problems with mixed control-state constraints. *Control Cybern.*, 39(3):717–738, 2010.

[62] P. Grisvard. *Elliptic Problems in Nonsmooth Domains.* Pitman, Boston, 1985.

[63] A. Günter and M. Hinze. Elliptic control problems with gradient constraints - variational discrete versus piecewise constant controls. *Comput. Optim. Appl.*, appeared online:DOI: 10.1007/s10589–009–9308–8, 2009.

[64] M. D. Gunzburger. *Perspectives in flow control and optimization.* SIAM, 2003.

[65] M. Hintermüller and M. Hinze. Moreau-yosida regularization in state constrained elliptic control problems: Error estimates and parameter adjustment. *SIAM J. on Numerical Analysis*, 47:1666–1683, 2009.

[66] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth Newton method. *SIAM J. Optim.*, 13:865–888, 2003.

[67] M. Hintermüller and K. Kunisch. Path-following methods for a class of constrained minimization problems in function space. *SIAM J. on Optimization*, 17:159–187, 2006.

[68] M. Hintermüller and K. Kunisch. PDE-constrained optimization subject to pointwise constraints on the control, the state, and its derivative. *SIAM J. Optim.*, 20(3):1133–1156, 2009.

[69] M. Hintermüller, F. Tröltzsch, and I. Yousept. Mesh-independence of semismooth Newton methods for Lavrentiev-regularized state constrained nonlinear optimal control problems. *Numerische Mathematik*, 108(4):571–603, 2008.

[70] M. Hintermüller and I. Yousept. A sensitivity-based extrapolation technique for the numerical solution of state-constrained optimal control problems. *ESAIM COCV*, 16(3):503–522, 2010.

[71] M. Hinze. A variational discretization concept in control constrained optimization: the linear-quadratic case. *J. Computational Optimization and Applications*, 30:45–63, 2005.

[72] M. Hinze and C. Meyer. Stability of semilinear elliptic optimal control problems with pointwise state constraints. Technical Report 1236, WIAS, 2007.

[73] M. Hinze and C. Meyer. Variational discretization of Lavrentiev-regularized state constrained elliptic optimal control problems. *Computational Optimization and Applications*, 46:487–510, 2010.

[74] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with pde constraints. Mathematical Modelling: Theory and Applications.* Springer, 2008.

[75] M. Hinze and A. Schiela. Discretization of interior point methods for state constrained elliptic optimal control problems: Optimal error estimates and parameter adjustment. *COAP, published online, DOI: 10.1007/s10589-009-9278-x*, 2009.

[76] D. Hömberg. A mathematical model for induction hardening including mechanical effects. *Nonlinear Anal. Real World Appl.*, 5:55–90, 2004.

[77] M. Huth and R. Tichatschke. A hybrid method for semi-infinite programming problems. *Operations research, Proc. 14th Symp. Ulm/FRG 1989,Methods Oper. Res.*, 62:79–90, 1990.

[78] K. Ito and K. Kunisch. Semi-smooth Newton methods for state-constrained optimal control problems. *Systems and Control Letters*, 50:221–228, 2003.

[79] N. Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, WI, 1979.

[80] D. Kinderleher and G. Stampacchia. *An introduction to variational inequalities and their applications.* Pure and applied mathematics: a series of monographs and textbooks. Academic Press, New York – San Francisco – London, 1980.

[81] M. A. Krasnoselskii, P. P. Zabreiko, E. I. Pustylnik, and P. E. Sbolevskii. *Integral operators in spaces of summable functions.* Noordhoff International Publishing, Leyden, 1976.

[82] A. Kröner and B. Vexler. A priori error estimates for elliptic optimal control problems with bilinear state equation. *J. Comp. Appl. Math.*, 230(2):781–802, 2009.

[83] K. Krumbiegel, I. Neitzel, and A. Rösch. Regularization for semilinear elliptic optimal control problems with pointwise state and control constraints. *Comput. Optim. Appl.*, online first(6):DOI: 10.1007/s10589–010–9357–z, 2010.

[84] K. Krumbiegel, I. Neitzel, and A. Rösch. Sufficient optimality for the moreau–yosida-type regularization concept applied to semilinear elliptic optimal control problems with pointwise state constraints. *Annals of the Academy of Romanian Scientists, Series on Mathematics and its Applications*, 2(2):222–246, 2010.

[85] K. Krumbiegel and A. Rösch. A virtual control concept for state constrained optimal control problems. *COAP*, 43(2):213–233, 2009.

[86] O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Ural'ceva. *Linear and Quasilinear Equations of Parabolic Type.* American Math. Society, Providence, R.I., 1968.

[87] I. Lasiecka and K. Malanowski. On discrete-time Ritz-Galerkin approximation of control constrained optimal control problems for parabolic systems. *Control Cybern.*, 7:21–36, 1978.

[88] M. M. Lavrentiev. *Some improperly posed problems of mathematical physis.* Springer, 1967.

[89] J. L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires.* Dunod Gauthier–Villars, 1969.

[90] J. L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations.* Springer-Verlag, Berlin, 1971.

[91] D. Luenberger. *Optimization by Vector Space Methods.* Wiley, New York, 1969.

[92] K. Malanowski. Convergence of approximations vs. regularity of solutions for convex, control–constrained optimal control problems. *Applied Math. Optimization*, 8:69–95, 1981.

[93] K. Malanowski. Stability and sensitivity analysis for optimal control problems with control–state constraints. *Dissertationes Mathematicae (Rozprawy Matematyczne)*, 394:51 pp, 2001.

[94] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control for parabolic equations. *Journal of Analysis and its Applications*, 18(2):469–489, 1999.

[95] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control of elliptic equations. *Control Cybern.*, 29:237–256, 2000.

[96] M. Mateos. *Problemas de Control Óptimo Gobernados por Ecuaciones Semilineales con Restricciones de Tipo Integral sobre el Gradiente del Estado.* PhD thesis, Universidad de Cantabria, 2000.

[97] The Mathworks. *Optimization Toolbox User's guide.* www.mathworks.com/help/toolbox/optim/.

[98] R. S. McNight and W. E. Bosarge, Jr. The Ritz-Galerkin procedure for parabolic control problems. *SIAM J. Control*, 11:510–524, 1973.

[99] D. Meidner. *Adaptive Space-Time Finite Element Methods for Optimization Problems Governed by Nonlinear Parabolic Systems.* PhD thesis, Ruprecht-Karls-Universität Heidelberg, 2008.

[100] D. Meidner, R. Rannacher, and B. Vexler. A priori error estimates for finite element discretizations of parabolic optimization problems with pointwise state constraints in time. Technical Report 099, SPP1253, 2010.

[101] D. Meidner and B. Vexler. A priori error estimates for space-time finite element discretization of parabolic optimal control problems Part I: Problems without control constraints. *SIAM Control Optim.*, 47(3):1150–1177, 2008.

[102] D. Meidner and B. Vexler. A priori error estimates for space-time finite element discretization of parabolic optimal control problems Part II: Problems with control constraints. *SIAM Control Optim.*, 47(3):1301–1329, 2008.

[103] D. Meidner and B. Vexler. A priori error analysis of the petrov galerkin crank nicolson scheme for parabolic optimal control problems. Technical Report 109, SPP1253, 2010.

[104] P. Merino, I. Neitzel, and F. Tröltzsch. Error estimates for the finite element discretization of semi-infinite elliptic optimal control problems. *Discussiones Mathematicae Differential Inclusions, Control and Optimization*, 30(2):221–236, 2010.

[105] P. Merino, I. Neitzel, and F. Tröltzsch. On linear-quadratic elliptic optimal control problems of semi-infinite type. *Applicable Analysis.*, published online:DOI 10.1080/00036811.2010.489187, 2010.

[106] P. Merino, F. Tröltzsch, and B. Vexler. Error Estimates for the Finite Element Approximation of a Semilinear Elliptic Control Problem with State Constraints and Finite Dimensional Control Space. *ESAIM:Mathematical Modelling and Numerical Analysis*, 44:167–188, 2010.

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

[108] C. Meyer, U. Prüfert, and F. Tröltzsch. On two numerical methods for state-constrained elliptic control problems. *Optimization Methods and Software*, 22(6):871–899, 2007.

[109] C. Meyer and A. Rösch. Superconvergence properties of optimal control problems. *SIAM J. Control and Optimization*, 43:970–985, 2004.

[110] C. Meyer, A. Rösch, and F. Tröltzsch. Optimal control of PDEs with regularized pointwise state constraints. *Computational Optimization and Applications*, 33(2003-14):209–228, 2006.

[111] C. Meyer and F. Tröltzsch. On an elliptic optimal control problem with pointwise mixed control-state constraints. In A. Seeger, editor, *Recent Advances in Optimization. Proceedings of the 12th French-German-Spanish Conference on Optimization held in Avignon, September 20-24, 2004*, Lectures Notes in Economics and Mathematical Systems. Springer-Verlag, 2005.

[112] C. Meyer and I. Yousept. Regularization of state-constrained elliptic optimal control problems with nonlocal radiation interface conditions. *Computational Optimization and Applications.*, 44:734–755, 2009.

[113] J. Nečas. *Les Méthodes Directes en Théorie des Equationes Elliptiques.* Academia, Praque, 1967.

[114] I. Neitzel and F. Tröltzsch. On convergence of Lavrentiev regularization for semilinear parabolic optimal control problems with pointwise control and state constraints. *Proc. Appl. Math. Mech.*, 8(1):10865–10866, 2008.

[115] I. Neitzel and F. Tröltzsch. On convergence of regularization methods for nonlinear parabolic optimal control problems with control and state constraints. *Control and Cybernetics*, 37(4):1013–1043, 2008.

[116] I. Neitzel and F. Tröltzsch. On regularization methods for the numerical solution of parabolic control problems with pointwise state constraints. *ESAIM Control, Optimisation and Calculus of Variations*, 15(2):426–453, 2009.

[117] I. Neitzel and F. Tröltzsch. Numerical analysis of state-constrained optimal control problems for PDEs. to appear in SPP1253 theme issue, 2010.

[118] I. Neitzel and B. Vexler. A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems. Technical Report 107, SPP1253, 2010. submitted to Numerische Mathematik.

[119] C. Ortner and W. Wollner. A priori error estimates for optimal control problems with pointwise constraints on the gradient of the state. Technical Report SPP1253-071, SPP1253, 2009.

[120] U. Prüfert and F. Tröltzsch. An interior point method for a parabolic optimal control problem with regularized pointwise state constraints. *ZAMM*, 87(8–9):564–589, 2007.

[121] U. Prüfert, F. Tröltzsch, and M. Weiser. The convergence of an interior point method for an elliptic control problem with mixed control-state constraints. *Comput. Optim. Appl.*, 39(2):183–218, March 2008.

[122] R. Rannacher and B. Vexler. A priori error estimates for the finite element discretization of

elliptic parameter identification problems with pointwise measurements. *SIAM Control Optim.*, 44:1844–1863, 2005.

[123] J. Raymond and F. Tröltzsch. Second order sufficient optimality conditions for nonlinear parabolic control problems with state constraints. *Discrete Contin. Dyn. Syst.*, 6:431–450, 2000.

[124] J.-P. Raymond and H. Zidani. Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls. *SIAM J. Control and Optimization*, 36:1853–1879, 1998.

[125] 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.

[126] R. Reemtsen and J.-J. Rückmann (Eds.). *Semi-Infinite Programming.* Kluwer Academic Publishers, Boston, 1998.

[127] S. Robinson. Strongly regular generalized equations. *Mathematics of Operation Research*, 5:43–62, 1980.

[128] A. Rösch. Error estimates for linear-quadratic control problems with control constraints. Technical Report 740-2002, TU Berlin, 2002.

[129] A. Rösch. Error estimates for parabolic optimal control problems with control constraints. *Z. für Analysis und ihre Anwendungen (ZAA)*, 23:353–376, 2004.

[130] A. Rösch. Error estimates for linear-quadratic control problems with control constraints. *Optimization Methods and Software*, 21, No. 1:121–134, 2006.

[131] A. Rösch and F. Tröltzsch. Sufficient second order optimality conditions for a parabolic optimal control problem with pointwise state constraints. *SIAM Control Optim.*, 42(1):138–154, 2003.

[132] A. Rösch and F. Tröltzsch. Existence of regular Lagrange multipliers for a nonlinear elliptic optimal control problem with pointwise control-state constraints. *SIAM J. Control and Optimization*, 45:548–564, 2006.

[133] A. Rösch and F. Tröltzsch. Sufficient second-order optimality conditions for an elliptic optimal control problem with pointwise control-state constraints. *SIAM J. on Optimization*, 17(3):776–794, 2006.

[134] A. Rösch and F. Tröltzsch. On regularity of solutions and Lagrange multipliers of optimal control problems for semilinear equations with mixed pointwise control-state constraints. *SIAM J. Control and Optimization*, 46(3):1098–1115, 2007.

[135] A. Schiela. Barrier methods for optimal control problems with state constraints. *SIAM J. on Optimization*, 20:1002–1031, 2009.

[136] G. Still. Discretization in semi-infinite programming: the rate of convergence. *Mathematical Programming. A Publication of the Mathematical Programming Society*, 91:53–69, 2001.

[137] G. Still. Generalized semi-infinite programming: Numerical aspects. *Optimization*, 49:223–242, 2001.

[138] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*, volume 1054 of *Lecture Notes in Mathematics*. Springer, 1984.

[139] H. Triebel. *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland, Amsterdam, New York, Oxford, 1978.

[140] F. Tröltzsch. On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. *SIAM Control Optim.*, 38:294–312, 1999.

[141] F. Tröltzsch. Lipschitz stability of solutions of linear-quadratic parabolic control problems with respect to perturbations. *Dyn. Contin. Discrete Impulsive Syst.*, 7:289–306, 2000.

[142] F. Tröltzsch. Regular Lagrange multipliers for problems with pointwise mixed control-state constraints. *SIAM J. Optimization*, 15(2):616–634, 2005.

[143] F. Tröltzsch. On Finite Element error estimates for optimal control problems with elliptic PDEs. In I. Lirkov, S. Margenov, and J. Wasniewski, editors, *Large Scale Scientific Computing*, volume 5910 of *Lecture Notes in Computer Science*, pages 40–53. Springer, 2010.

[144] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*. AMS, Providence, 2010.

[145] F. Tröltzsch and I. Yousept. A regularization method for the numerical solution of elliptic boundary control problems with pointwise state constraints. *COAP*, 42(1):43–66, 2009.

[146] F. Tröltzsch and I. Yousept. Source representation strategy for optimal boundary control problems with state constraints. *Journal of Analysis and its Applications*, 28(2):189–203, 2009.

[147] M. Ulbrich and S. Ulbrich. Primal-dual interior-point methods for PDE-constrained optimization. *Math. Program.*, 117(1-2, Ser. B):435–485, 2009.

[148] A. Unger. *Sufficient optimality conditions of second order and convergence of the SQP method for semilinear elliptic boundary control problems. (Hinreichende Optimalitätsbedingungen 2. Ordnung und Konvergenz des SQP-Verfahrens für semilineare elliptische Randsteuerprobleme.)*. PhD thesis, Chemnitz: Technische Universität Chemnitz, Fakultät Mathematik, 126 S. , 1997.

[149] F. G. Vázquez, J.-J. Rückmann, O. Stein, and G. Still. Generalized semi-infinite programming: a tutorial. *J. Computational and Applied Mathematics*, 217:394–419, 2008.

[150] R. Winther. Error Estimates for a Galerkin approximation of a parabolic control problem. *Ann. Math. Pura Appl.*, 117(4):173–206, 1978.

[151] J. Wloka. *Partielle Differentialgleichungen*. Teubner-Verlag, Leipzig, 1982.

[152] I. Yousept. *Optimal control of partial differential equations involving pointwise state constraints*. PhD thesis, Technische Universität Berlin, 2008.

[153] J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optimization*, 5:49–62, 1979.