Technische Universität München

Department of Mathematics

Chair of Optimal Control

Optimal Control of Parabolic Obstacle Problems

Optimality Conditions and Numerical Analysis

Dominik T. Hafemeyer

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzende:

Prof. Dr. Simone Warzel

Prüfer der Dissertation:

1. Prof. Dr. Boris Vexler
2. Prof. Dr. Christian Meyer
3. Prof. Dr. Thomas Apel

Die Dissertation wurde am 18.12.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 07.04.2020 angenommen.

Dedicated to all those who fight mental illness, proving that we can rise above the struggles.

Abstract

This thesis is concerned with the optimal control of parabolic obstacle problems. We first study the obstacle problems independently of the control aspect by regularizing with semilinear parabolic PDEs. We derive regularity results for the obstacle problems by taking appropriate limits of the solutions of the semlinear equations. We then discuss optimal control problems governed by obstacle problems or by the corresponding regularized problems. We derive existence of optimal solutions, necessary optimality conditions and second order sufficient conditions.
After that, we analyse space-time discretizations of the regularized obstacle problems and derive quasi-optimal L^∞-estimates for the error between the solutions to the obstacle problems on the continuous level and the regularized, discretized solutions. Lastly, we apply those estimates to discretized and regularized optimal control problems and derive partially optimal L^2-estimates for the errors betweens the continuous controls and states and the corresponding discretized, regularized variables.
Moreover, we give a quick introduction into our used algorithms and provide experiments that verify the sharpness of some of our estimates.

Zusammenfassung

Diese Arbeit beschäftigt sich mit der optimalen Steuerung von parabolischen Hindernisproblemen. Zunächst untersuchen wir die Hindernisprobleme losgelöst vom Optimierungskontext, indem wir mit semilinearen parabolischen PDEs regularisieren. Wir leiten Regularitätsresultate für die Hindernisprobleme her, indem wir passende Grenzwerte von Lösungen der semilinearen Gleichungen nutzen. Dann diskutieren wir Optimalsteuerungsprobleme mit Beschränkungen durch Hindernisprobleme bzw. durch die zugehörigen regularisierten Probleme. Wir zeigen Existenz von Lösungen und leiten notwendige Optimalitätsbedingungen sowie hinreichende Optimalitätsbedingungen zweiter Ordnung her.
Abschließend analyisieren wir Raum-Zeit-Diskretisierungen der regularisierten Hindernisprobleme und beweisen quasi-optimale L^∞-Abschätzungen für den Fehler zwischen den Lösungen der Hindernisprobleme auf der kontinuierlichen Ebene und den regularisierten, diskretisierten Lösungen. Zuletzt, wenden wir diese Abschätzungen auf die diskretisierten und regularisierten Optimalsteuerungsprobleme an und leiten teilweise optimale L^2-Abschätzungen für den Fehler zwischen den kontinuierlichen Steuerungen und Zuständen und den zugehörigen diskretsierten, regularisierten Größen her.
Außerdem geben wir eine kurze Einführung in die verwendeten Algorithmen und liefern Experimente, die die Optimalität manche unserer Abschätzungen belegen.

Contents

Introduction

This thesis is concerned with the study of optimal control of parabolic obstacle problems. The specific problems are of the form

$$\begin{aligned}
&\min_{(y,u)\in W(I)\times U_{ad}} j_v(y) + j_T(y(T)) + g(u) =: J(y,u),\\
&\text{such that } \begin{cases} (\partial_t y + Ay + f(y) - u, \varphi - y)_{L^2(I\times\Omega)} \geq 0 \\ \qquad\qquad \forall \varphi \in L^2(I,V) \text{ such that } \varphi \geq \Psi, \\ y(0) = y_0, \quad y|_{\Sigma_D} = 0, \quad y \geq \Psi. \end{cases}
\end{aligned} \tag{OC}$$

Here, u is the control, y is the corresponding state and Ψ is the name-giving obstacle. The operator A is symmetric on an appropriate Hilbert space V. Further, I is a time interval, Ω the spatial domain and f is a non-linearity. The cost functionals j_v, j_T and g can be of a quite generic form, though the quadratic case is sometimes considered for special results. The admissible set U_{ad} is bounded in an appropriate norm, which always includes simple box constraints. The interesting part is of course the connection between state and control. We will later make the following statement precise: essentially at every point in $I \times \Omega$ the state y solves the semi-linear equation $\partial_t y + Ay + f(y) = u$ or it touches the obstacle Ψ. Both can happen at once. A visual interpretation can be given by the image of a person stepping on a trampoline and gravity pulling the membrane towards the ground. If the person is light enough, the state (membrane) never touches the obstacle (ground), while for heavy persons the state might touch the obstacle.

Variational inequalities appear for example in the study of American options in the Black-Scholes model; for a mathematical derivation see for example [IK06] and the references to [Øks98, Sey02] therein. Another application is their relation/equivalence to free boundary problems which are for example studied in relation to ice sheet models, e.g. [CDV10], [JB12], where the obstacle problem is non-linear and elliptic or [CDD$^+$02], where a non-linear parabolic obstacle problem is used. An introduction to the modelling of membrane deformations with obstacle problems and various other applications can be found in [Rod87].

The first of the main achievements of this thesis is the derivation of second order sufficient conditions for the optimal control problems given by (OC). The techniques are based on the elliptic case studied in [KW12b]. There is limited related work. We shall mention [BM15] and [Bet19]. [BM15] is concerned with second order sufficient conditions in for VI models in static plasticity. [Bet19] is closer to our situation as it is considering second order sufficient conditions in the realm of semi-smooth, semilinear parabolic PDEs.

The most important result is the numerical analysis of regularized obstacle problems in Chapter 4. The analysis consists of two vital sub-results. One is the establishment of global L^∞-error estimates for linear parabolic problems in Section 4.4.1. Parts of the extensive and long proof were derived in cooperation with Lucas Bonifacius and are found in the appendix. The second result is the transfer of techniques from [Noc88] to the parabolic case. There it is shown that the convergence rates for the discretization of the regularized obstacle problem do not depend on the regularization parameter γ. This is confirmed experimentally. The obtained rates for the complete error, i.e. discretization and regularization error combined, are, up to a logarithmic factor, $\mathcal{O}(\gamma^{\frac{1}{\alpha}} + k + h^2)$, where $\alpha \geq 1$ depends on the choice of regularization, k is the time discretization parameter and h the spatial discretization parameter. The rates are shown to be optimal in examples in Chapter 7.

Lastly, we would like to mention the combination of the previous to results in Chapter 5 where the local growth conditions of the second order sufficient conditions are combined with the rates from Chapter 4 to obtain a convergence rate for the optimal control and the optimal state for the regularized, discretized problems. The proven rates are partially shown to be sharp via examples in Chapter 7. To our knowledge the only other work concerned with the numerical analysis of the optimal control with obstacle problems is [MT13]. There the obstacle problem is stationary and no regularization is used.

I would like to express my deepest gratitude to the following people. The list is non-exhaustive and not sorted by importance. For repeated programming advice I thank Lukas Failer and Dominik Meidner. For feedback regarding the thesis I would like to thank Sebastian Engel. For essential cooperation with Lucas Bonifacius I will be eternally thankful. For endless patience and his deep pool of knowledge I would like to thank Constantin Christof. For sacrificing their time and reviewing my work I would like to thank Thomas Apel and Christian Meyer. Last, but certainly not least, I would like to thank my supervisor Boris Vexler for always lending support and giving the freedom to follow ideas on my own terms.

Lastly, we give a short overview of the chapters to follow. Each chapter and some sections start with the collection of the standing assumptions of the chapter or section and a brief overview over related research.

Chapter 1: Definitions and Notation

This chapter is concerned with laying the foundation by restating well-known results, introducing notation and proving smaller lemmata. We require basics from measure theory, Sobolev spaces, Bochner spaces and interpolation space theory. All those topics are interlinked, but the measure theory becomes especially important in the formulation of the optimality system in Chapter 3. The interpolation space theory is vital to properly define our problems and then again appears in the finite element error estimates in Chapter 4.

Chapter 2: Obstacle Problems and their Regularization via Semilinear PDEs

Here, we introduce the considered obstacle problem in detail. We start with a short comment on the abstract framework and then give a more concrete problem formulation. We then introduce a family of semilinear equations, where we replace the obstacle constraint with a penalty term. We prove various regularity results, depending on the smoothness of the domain, and show, in particular, that those regularities do not depend on the penalty term. These proofs will rely on the fact that the penalty term satisfies a monotonicity condition, but other than that one can choose it relatively freely.

This independence from the penalty term allows us to easily prove existence and regularity of solutions to the specified obstacle problem by taking the limit. Convergence rates cannot be proven in general, but, by adapting techniques from [Noc88] from the elliptic setting to the parabolic setting, we are able to show L^∞-convergence rates for the error between the solutions to the regularized obstacle problem and the original, unregularized problem. The rate heavily depends on the structure of the regularization parameter.

Chapter 3: Optimal Control Problems and their Regularizations

In this chapter we introduce our optimal control problem, which has the form of (OC). The quantities are be specified and made precise.

We once again use regularization techniques. We consider the semilinear equations from the previous chapter and use them as constraints in (OC) instead of an obstacle problem. Additionally we modify the cost functionals slightly so that they include information about a chosen solution of the unregularized problem. The regularized control problem is smooth enough to use differentiability to obtain necessary first order optimality conditions. Again, one can bound important terms, this time in the optimality conditions, independently of the penalization term, and then take the limit in appropriate norms. By modifying the cost functionals as mentioned above the quantities obtained by taking the limit carry information about this chosen solution and characterize it. One obtains a so-called stationarity system. We also explain why standard approaches, like KKT-conditions, are not best suited for this class of problems. We derive some properties on the multipliers of the resulting system which is important in the discussion of second order sufficient conditions. It is also useful in showing the equivalence of the original problem to a state constraint problem, which we do for a special case.

After that we discuss sufficient optimality conditions for the original problem, which we mentioned earlier, and a regularized variant of it. Lastly, we give a short outlook on control problems without control constraints.

Chapter 4: Discretization and Numerical Analysis for Regularized Obstacle Problems

This chapter is, in our eyes, the most important. We start by an in-depth discussion of the discretization in the spatial variable via piecewise linear finite elements. We prove, in particular, the L^∞-stability of the Ritz projection and an L^∞-norm resolvent estimate for finite element operators. Then we introduce a discretization in the time variable by the means of piecewise constant elements.

We then can finally discretize the regularized obstacle problem from Chapter 2. The numerical analysis is heavily inspired by the elliptic ideas from [Noc88]. We already gave the relevant details in the discussion of our main results.

Chapter 5: Numerical Analysis of Discretized, Regularized Optimal Control Problems

This chapter is now a culmination of the previous two chapters: we regularize and discretize the optimal control problem by discretizing the state. We first leave the control untouched. We analyse the global solutions to those semi-discrete problems with respect to convergence and convergence rates. The solutions to the regularized, discretized problem do converge to solutions of the original problem. The same, however, cannot be said for the multipliers of the optimality systems. Convergence rates for control and state are proven by using a local quadratic growth condition.

To discuss the discretization of the controls we will first see that, under common assumptions on the cost functional and the admissible set, the controls for the semi-discrete problem are automatically piecewise constant. However, the discretization in the spatial variable is not automatically carried over.

We discretize the control piecewise linearly in space separately and repeat the discussion from the semi-discrete case and lean on it to prove a convergence rate of $\mathcal{O}(\gamma^{\frac{1}{2\alpha}} + k^{\frac{1}{2}} + h)$ for the L^2-norms of the controls and states, up to logarithmic factors, with the α once again depending on the regularization term.

Chapter 6: Solution Algorithms for Discretized, Regularized Obstacle Problems and Optimal Control Problems

In the algorithmic chapter we only quickly discuss ways and challenges in solving the regularized, discretized obstacle problem from Chapter 4 and the regularized, discretized optimization problems from Chapter 5. We prove here that the solution operator tending to the regularized obstacle problem is in fact two times differentiable, which is important for the application of Newton's method.

We use Newton's method to solve the semilinear equations. To solve the optimal control problem we use a pathfollowing strategy combined with a trust region method, which globalizes a semi-smooth Newton method.

Chapter 7: Numerical Examples

In this section we first show results that confirm that the estimates for the combined regularization/discretization error derived in Chapter 5 are sharp. We also observe the aforementioned decoupling of the regularization and discretization error.

We then construct a specific example to analyse the rate obtained for the regularized, discretized optimal control problem from Chapter 5. We will see that the error in the regularization parameter is often the dominant one, which makes the experimental error analysis in time and space difficult.

Chapter 8: Appendix

The appendix contains multiple auxiliary results that do either disturb the flow of the main thesis or are used at so many different points throughout the thesis that they do not belong to any specific chapter. For example, the existence results for non-linear PDEs together with regularity estimates, which include tracked constants, are found in Section 8.4.

We also find the aforementioned work with Lucas Bonifacius in the appendix.

Lastly, we provide a short index of symbols for the reader, which contains the most important variables and abbreviations.

1 Definitions and Notation

1.1 Measure Theory

We require a few tools from measure theory. We take the following definitions from [Rud74, Sections 6.1, 6.18] restricted to the real case. The definitions also apply to complex valued measures, which we do not need.

Definition 1.1 Let $\Omega \subset \mathbb{R}^N$ be a set and $\mathcal{B}(\Omega)$ its Borel σ-algebra. Then $M(\Omega)$ denotes the space of all real valued, regular Borel measures.

We define the measure of total variation $|\mu|$ for a $\mu \in M(\Omega)$ as

$$\begin{aligned} |\mu|\colon \mathcal{B}(\Omega) &\to \mathbb{R}, \\ A &\mapsto \sup\left\{\sum_{k=1}^{\infty} |\mu(A_k)| : A_k \in \mathcal{B}(\Omega), \bigcup_{k=1}^{\infty} A_k = A, A_k \text{ are pairwise disjoint}\right\}. \end{aligned}$$

Together with the norm

$$\begin{aligned} ||\cdot||_{M(\Omega)}\colon M(\Omega, \mathbb{R}^N) &\to \mathbb{R}_{\geq 0}, \\ \mu &\mapsto |\mu|(\Omega) \end{aligned}$$

$M(\Omega)$ is a Banach space; see [AFP00, Remark 1.7].

Remark 1.2 Let $\Omega \subset \mathbb{R}^N$ be a set. Considering the space of infinitely differentiable functions $C_c^\infty(\Omega)$ with compact support in Ω together with the sup norm

$$||v||_{L^\infty(\Omega)} = \sup_{x\in\Omega} |v(x)| \text{ for } v \in C_c^\infty(\Omega),$$

we define $C_0(\Omega)$ as the closure of $C_c^\infty(\Omega)$ under this norm. Then [AFP00, Theorem 1.54] and [ABM14, beginning of Chapter 10.1] hold true:

$$M(\Omega) \simeq C_0(\Omega)^*.$$

Note that for compact Ω we have $C_0(\Omega) = C(\bar{\Omega})$.

Thus we have for a $\mu \in M(\Omega)$

$$||\mu||_{M(\Omega)} = \sup\left\{\int_\Omega v(x)\, d\mu(x), v \in C_0(\Omega), ||v||_{L^\infty(\Omega)} \leq 1\right\}.$$

The following definition can be found in [AFP00, Definition 1.64].

Definition 1.3 Let $\Omega \subset \mathbb{R}^N$. For a $\mu \in M(\Omega), \mu \geq 0$ we define the support as

$$\operatorname{supp} \mu := \{x \in \Omega : x \in A, A \text{ open in } \Omega \implies \mu(A) > 0\}.$$

This is just the set of all points, whose open neighborhoods are never zero sets. For a general $\mu \in M(\Omega)$ we define the support

$$\operatorname{supp} \mu := \operatorname{supp} |\mu|$$

The next definition can be found in [AFP00, Definition 1.65 and the following remark].

Definition 1.4 Let $\Omega \subset \mathbb{R}^N$. For a $\mu \in M(\Omega)$ and $A \in \mathcal{B}(\Omega)$ we define the restriction $\mu|_A$ is the measure in $M(A)$ such that $\mu|_A(B) = \mu(A \cap B)$ for each $B \in \mathcal{B}(A)$.

1.2 Domains

Definition 1.5 We frequently use $C^{k,\alpha}$-domains, which are defined along the lines of [Trö09, Section 2.2.2] or [Gri11, Definition 1.2.1.1]. The idea is that the boundary of Ω can be locally written as the graph of a $C^{k,\alpha}$-function.

Let $\alpha \in [0,1]$ and $k \in \mathbb{N}_0$. We say that a bounded domain $\Omega \subset \mathbb{R}^N$, with $N \in \mathbb{N}$, is a $C^{k,\alpha}$-domain if the following holds true: there are constants $a, b > 0$, finitely many linear, local coordinate transformations $S_1, \dots, S_M \colon \mathbb{R}^N \to \mathbb{R}^N$ obtained by rotation and shifts and $C^{k,\alpha}$-functions $f_i \colon [-a,a]^{N-1} \to \mathbb{R}$ such that with the abbreviation $x = S_i^{-1} y$ we have

$$(y, y_N) \in \Omega \cap S_i((-a,a)^{N-1} \times (-b,b)) \iff f_i(x_{1,\dots,N-1}) < x_N,$$
$$(y, y_N) \in \partial\Omega \cap S_i((-a,a)^{N-1} \times (-b,b)) \iff f_i(x_{1,\dots,N-1}) = x_N.$$

Of course $\partial\Omega$ shall be covered by those transformation, i.e.

$$\partial\Omega \subset \bigcup_{i=1}^{M} S_i((-a,a)^{N-1} \times (-b,b)).$$

The whole affaire is illustrated in Figure 1.1. For $\alpha = 0$ we understand $C^0((-a,a)^{N-1}) = C([-a,a]^{N-1})$. This is not in conflict with the common notation used for uniformly continuous maps, since every continuous map on compact sets is uniformly continuous anyway. We also write C^k for $C^{k,0}$-domains.

We call $C^{0,1}$-domains Lipschitz domains.

Remark 1.6 In the light of the remarks at the beginning of [GT77, Section 6.2] or the remark after [Gri11, Definition 1.2.1.2] this definition implies that for each $x_0 \in \partial\Omega$ there is a open ball $B(x_0)$ and an open set D together with a bijective map $\Phi \in C^{k,\alpha}(B(x_0), D)$ such that

- $\Phi(B(x_0) \cap \Omega) \subset \mathbb{R}^{N-1} \times \mathbb{R}_{>0}$,
- $\Phi(B(x_0) \cap \partial\Omega) \subset \mathbb{R}^{N-1} \times \{0\}$,
- $\Phi^{-1} \in C^{k,\alpha}(D, B(x_0))$.

In [GT77, Section 6.2] and after [Gri11, Definition 1.2.1.2] it is remarked that, if $k \geq 1$, Definition 1.5 is equivalent to the formulations given in this remark.

Definition 1.7 Let $\Omega \subset \mathbb{R}^N$ and let Γ be a relatively open subset of $\partial\Omega$. The set $\Omega \cup \Gamma$ is called Gröger regular in the sense of [Grö89] if it is bounded and for any $x \in \partial\Omega$ there exist an open neighbourhood U_x of x and a bi-Lipschitz mapping $\phi_x : U_x \to \phi_x(U_x)$ such that $\phi_x(U_x \cap \Omega)$ is one of the following three sets: $B_1(0) \cap \mathbb{R}^{N-1} \times \mathbb{R}_{<0}$, $B_1(0) \cap \mathbb{R}^{N-1} \times \mathbb{R}_{\leq 0}$ or $\left\{ x \in B_1(0) \cap \mathbb{R}^{N-1} \times \mathbb{R}_{\leq 0} : x_N < 0 \text{ or } x_1 > 0 \right\}$.

The interpretation is given in [Grö89, Remark 1] and restated in our notation. Gröger regularity means, roughly speaking, that Γ and $\partial\Omega \backslash \Gamma$ are separated by a Lipschitzian hypersurface of $\partial\Omega$. It is straightforward to see that if Ω is a Lipschitz domain $\bar{\Omega}$ is Gröger regular.

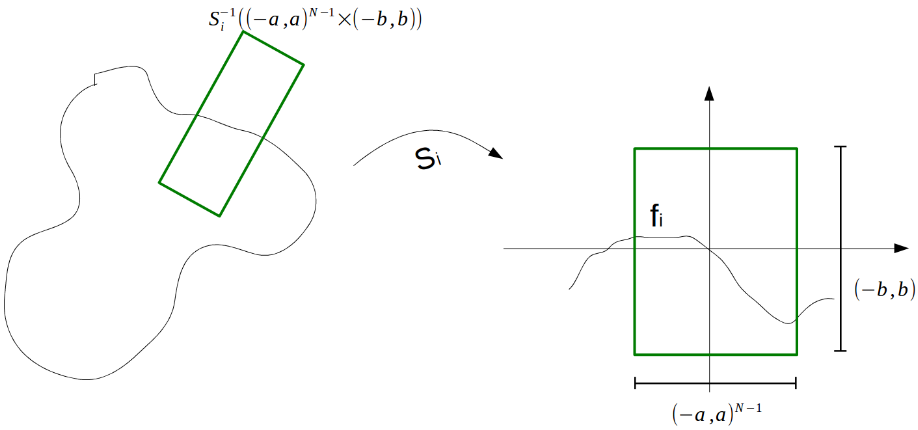


Figure 1.1: Transformation of $\partial\Omega$ onto function graphs.

Definition 1.8 Let $\Omega \subset \mathbb{R}^N$ be a domain. It is said to satisfy the cone condition, in the sense of [Tri78, Definition 4.2.3] or [Agm65, Definition 2.1], if there exists a finite, open cover $U_1, \dots, U_M$ of $\partial\Omega$ such that for each U_i there exists a cone C_i of finite length, centered at the origin such that:

$$(U_i \cap \Omega) + C_i \subset \Omega.$$

The cone condition is useful for the proofs of some embedding and extension theorems, see for example [Tri78, Remark 4.2.4.6].

In [Ada75, Remark 4.7] we find the following statement:

Theorem 1.9 *Lipschitz domains satisfy the cone condition.*

1.3 Sobolev Spaces

Throughout Section 1.3 let $\Omega \subset \mathbb{R}^N$ be a domain for $N \in \mathbb{N}$. Unless noted otherwise all the given vector spaces are considered as real vector spaces. If not specified all definitions and results for real vector spaces are found in [Trö09, Sections 2.1-2.4]. The extensions to the complex cases are usually straightforward, decomposing everything into real and imaginary parts. It is necessary to study the complex valued case as we require them in the analysis of operator resolvents later in Section 4.1.6.

Definition 1.10 Whenever we refer to the measurability of a function $f : \Omega \to \mathbb{R}$ we consider it as Lebesgue-measurability in the sense of [Rud74, Theorem 2.20, Definition 3.6]. Loosely speaking the Lebesgue-measurable sets are given by the σ-algebra obtained by completing the Borel-σ-algebra with respect to the Lebesgue-measure, cf. [Rud74, Theorem 1.36]. Also, whenever we refer to a property to hold "almost everywhere" (a.e.) in a domain or set, we mean that the property holds everywhere, except for a subset of Lebesgue-measure 0.

Let $f : \Omega \to \mathbb{R}$ be measurable. Let $p \in [1, \infty]$. We define

$$\|f\|_{L^p(\Omega)} := \left(\int_\Omega |f|^p \, dx\right)^{\frac{1}{p}} \text{ for } p < \infty \text{ and } \|f\|_{L^\infty(\Omega)} := \operatorname{esssup}_{x\in\Omega} |f(x)|$$

with possibly infinite values. We define the Lebesgue space $L^p(\Omega)$ as the vector space of equivalence classes of measurable functions $f : \Omega \to \mathbb{R}$ such that $\|f\|_{L^p(\Omega)} < \infty$. Here, two functions are considered equivalent if they differ on a set of Lebesgue measure 0. As is it customary, however, we will refer to elements $f \in L^p(\Omega)$ as functions, essentially identifying the equivalence class with one of its representatives.

For complex, measurable functions we define the analogous spaces $L^p(\Omega, \mathbb{C})$. For $f \in L^p(\Omega, \mathbb{C})$ we define $\|f\|_{L^p(\Omega,\mathbb{C})} := \||f|\|_{L^p(\Omega)}$.

For any $p \in [1, \infty]$ the spaces $L^p(\Omega)$ and $L^p(\Omega, \mathbb{C})$ are Banach spaces. The spaces $L^2(\Omega)$ and $L^2(\Omega, \mathbb{C})$ are Hilbert spaces with the inner products

$$(f, g)_{L^2(\Omega)} = \int_\Omega fg \, dx, \quad \text{respectively,} \quad (f, g)_{L^2(\Omega)} = \int_\Omega f\bar{g} \, dx.$$

Lemma 1.11 *We have the isomorphy*

$$L^p(\Omega, \mathbb{C}) \simeq L^p(\Omega) + iL^p(\Omega).$$

Where the direct sum uses an arbitrary finite dimensional norm $|\cdot|_s$ on $\mathbb{R}^2 \simeq \mathbb{C}$.

Proof. By [Rud74, Paragraph 1.9] $f\colon \Omega \to \mathbb{C}$ is measurable if and only if its real part, $\mathfrak{Re}(f)$, and imaginary part, $\mathfrak{Im}(f)$, are measurable. On the one hand we have

$$\begin{aligned}\|(\mathfrak{Re}(f), \mathfrak{Im}(f))\|_{L^p(\Omega)+iL^p(\Omega)} &= \Big|\Big(\|\mathfrak{Re}(f)\|_{L^p(\Omega)}, \|\mathfrak{Im}(f)\|_{L^p(\Omega)}\Big)\Big|_s \\ &\leq C_{s,1}\Big(\||\mathfrak{Re}(f)\|_{L^p(\Omega)} + \|\mathfrak{Im}(f)\|_{L^p(\Omega)}\Big) \leq 2C_{s,1}\|f\|_{L^p(\Omega,\mathbb{C})}.\end{aligned}$$

Here $C_{s,1}$ is an equivalency constant of $|\cdot|_s$ on $\mathbb{R}^2$ and the 1-norm. On the other hand

$$\|f\|_{L^p(\Omega,\mathbb{C})} = \||f|\|_{L^p(\Omega)} \leq \|\mathfrak{Re}(f)\|_{L^p(\Omega)} + \|\mathfrak{Im}(f)\|_{L^p(\Omega)} \leq C_{1,s}\|(\mathfrak{Re}(f), \mathfrak{Im}(f))\|_{L^p(\Omega)+iL^p(\Omega)}.$$

Here $C_{1,s}$ is another equivalency constant of the 1-norm and $|\cdot|_s$ on $\mathbb{R}^2$. □

Definition 1.12 Let $p \in [1, \infty]$ and $v \in L^p(\Omega)$. With $\partial_{x_i} v \in L^p(\Omega)$ we denote the partial weak derivative of v in the variable x_i if

$$-\int_\Omega \partial_{x_i} v\varphi \, dx = \int_\Omega v\partial_{x_i}\varphi \, dx \qquad \forall \varphi \in C_0^\infty(\Omega).$$

Schwarz' lemma holds for weak derivatives, i.e. if $\partial_{x_i}\partial_{x_j} v$ exists it is equal to $\partial_{x_j}\partial_{x_i} v$.

If the weak derivatives of appropriate order exist we write

$$\nabla v = \begin{pmatrix} \partial_{x_1} v \\ \vdots \\ \partial_{x_N} v \end{pmatrix}, \quad \nabla^2 v = \begin{pmatrix} \partial_{x_1 x_1} v & \dots & \partial_{x_1 x_N} v \\ \vdots & \vdots & \\ \partial_{x_N x_1} v & \dots & \partial_{x_N x_N} v \end{pmatrix}.$$

For $k \in \mathbb{N}$ we denote by $W^{k,p}(\Omega)$ the space of all functions in $L^p(\Omega)$ such that all the weak derivatives exist up to order k and lie in $L^p(\Omega)$. We equip them with the natural norm

$$\|v\|_{W^{k,p}(\Omega)} := \left(\sum_{\substack{\alpha_1+\dots+\alpha_N \leq k, \\ \alpha_i \in \mathbb{N}_0}} \|\partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} v\|_{L^p(\Omega)}^p \right)^{\frac{1}{p}}$$

or for $p = \infty$

$$\|v\|_{W^{k,\infty}(\Omega)} := \sup_{\substack{\alpha_1+\dots+\alpha_N \leq k, \\ \alpha_i \in \mathbb{N}_0}} \|\partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} v\|_{L^\infty(\Omega)}.$$

We can make analogous constructions for complex valued functions. By $C_c^\infty(\Omega, \mathbb{C})$ we denote the infinitely differentiable functions with values in $\mathbb{C}$ and compact support. This is not the space of holomorphic functions, but functions that are smooth from $\mathbb{R}^N$ to $\mathbb{R}^2 \simeq \mathbb{C}$. The following properties for those spaces, e.g. them being Banach spaces, follow by Lemma 1.11 and the decomposition into real and imaginary parts.

Let $v \in L^p(\Omega, \mathbb{C})$. The i-th derivative of v is defined via

$$-\int_\Omega \partial_{x_i} v \bar{\varphi}\, dx = \int_\Omega v \partial_{x_i} \bar{\varphi}\, dx \qquad \forall \varphi \in C_0^\infty(\Omega, \mathbb{C}).$$

Therefore we can define $W^{k,p}(\Omega, \mathbb{C})$ and $\|\cdot\|_{W^{k,p}(\Omega,\mathbb{C})}$ in analogy to the real case.

Lemma 1.13 *For any $p \in [1, \infty]$ and $k \in \mathbb{N}_0$ the spaces $W^{k,p}(\Omega)$ and $W^{k,p}(\Omega, \mathbb{C})$ are Banach spaces. For $p = 2$ we even have a Hilbert space, which we denote by $H^k(\Omega)$ with the inner product*

$$(y, v)_{H^k(\Omega)} = \sum_{\substack{\alpha_1+\cdots+\alpha_N \le k,\\ \alpha_i \in \mathbb{N}_0}} \left(\partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} y, \partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} v\right)_{L^2(\Omega)}.$$

The analogue can be done for the complex case. We denote $W^{k,2}(\Omega, \mathbb{C})$ by $H^k(\Omega, \mathbb{C})$. It is equipped with the inner product

$$(y, v)_{H^k(\Omega,\mathbb{C})} = \sum_{\substack{\alpha_1+\cdots+\alpha_N \le k,\\ \alpha_i \in \mathbb{N}_0}} \left(\partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} y, \partial_{x_1}^{\alpha_1} \cdots \partial_{x_N}^{\alpha_N} \bar{v}\right)_{L^2(\Omega,\mathbb{C})}.$$

Lemma 1.14 *Let $k \in \mathbb{N}_0$ and $p \in [1, \infty]$. We have*

$$W^{k,p}(\Omega, \mathbb{C}) \simeq W^{k,p}(\Omega) + iW^{k,p}(\Omega).$$

Here we use an arbitrary norm $|\cdot|_s$ on $\mathbb{R}^2$, just as in 1.11. A weak derivative $\partial_{x_l} v$, with $l \in \{1, 2, \ldots, N\}$, of a $v \in W^{k,p}(\Omega, \mathbb{C})$ is given by $\partial_{x_l} v = \partial_{x_l} \mathfrak{Re}(v) + i\partial_{x_l} \mathfrak{Im}(v)$. Higher order derivatives decompose accordingly.

Definition 1.15 Let $k \in \mathbb{N}$, $p \in [1, \infty)$. By $W_0^{k,p}(\Omega)$ we denote the closure of $C_0^\infty(\Omega)$ in $W^{k,p}(\Omega)$. In particular we write $H_0^k(\Omega)$ for $W_0^{k,2}(\Omega)$.

The second half of the following result is found in [Gri11, Corollary 1.5.1.6]. The complex valued case is again treated by decomposition.

Lemma 1.16 *Assume Ω is a bounded Lipschitz domain and $p \in [1, \infty]$. Then there exists a linear and continous map $\operatorname{tr} : W^{1,p}(\Omega) \to L^p(\partial\Omega)$, which satisfies for any $y \in C(\bar{\Omega}) \cap W^{1,p}(\Omega)$*

$$\operatorname{tr} y = y|_{\partial\Omega} \text{ a.e. on } \partial\Omega.$$

A quite well known, but seldom proven result is the following:

Lemma 1.17 *For a function $f \in L^\infty(Q)$ with $Q \subset \mathbb{R}^{N+1}$ with $N \in \mathbb{N}_0$ and $|Q| < \infty$ we have*

$$\lim_{p\to\infty} \|f\|_{L^p(Q)} = \|f\|_{L^\infty(Q)}.$$

Proof. Let $p \in (1, \infty)$, then

$$\|f\|_{L^p(Q)} \le |Q|^{\frac{1}{p}} \|f\|_{L^\infty(Q)}.$$

This upper bound converges to $\|f\|_{L^\infty(Q)}$ for $p \to \infty$. For $\delta > 0$ and a fixed representative of f we define

$$S_\delta := \left\{(t, x) \in Q : |f(t, x)| \ge \|f\|_{L^\infty(Q)} - \delta\right\}.$$

Then we have

$$\|f\|_{L^p(Q)} \ge \left(\int_{S_\delta} |f(t, x)|^p \, d(t, x)\right)^{\frac{1}{p}} \ge |S_\delta|^{\frac{1}{p}} \left(\|f\|_{L^\infty(Q)} - \delta\right).$$

So we find for any $\delta > 0$:

$$\|f\|_{L^\infty(Q)} - \delta \le \liminf_{p\to\infty} \|f\|_{L^p(Q)} \le \limsup_{p\to\infty} \|f\|_{L^p(Q)} \le \|f\|_{L^\infty(Q)}.$$

Because $\delta > 0$ and the representative of f were chosen arbitrarily this implies

$$\liminf_{p\to\infty} \|f\|_{L^p(Q)} = \limsup_{p\to\infty} \|f\|_{L^p(Q)} = \|f\|_{L^\infty(Q)}.$$

□

1.4 Bochner Spaces

Definition 1.18 Let $p \in [1, \infty]$ and X be a Banach space. By $L^p(I, X)$ we denote the equivalence classes of functions $y : I \to X$ such that y is Bochner measurable in the sense of [Emm04, Definition 7.1.8] and

$$\|y\|_{L^p(I,X)} := \left(\int_I \|y(t)\|_X^p \, dt\right)^{\frac{1}{p}} < \infty$$

or for $p = \infty$

$$\|y\|_{L^p(I,X)} := \sup_{t\in I} \|y(t)\|_X < \infty.$$

Here two functions are equivalent if they coincide almost everywhere in the sense of Lebesgue measurability. As before we identify an equivalence class with its representative for convenience.

For the closed interval $\bar{I} = [a, b]$ we define $C(\bar{I}, X)$ as the space of continuous functions $y : \bar{I} \to X$. It is equipped with the norm $\|\cdot\|_{L^\infty(I,X)}$. For $\alpha \in (0, 1]$ we define $C^\alpha(I, X)$ as the space of Hölder-continuous functions with the norm

$$|y|_{C^\alpha(I,X)} := \sup_{\substack{t,s\in I \\ t\neq s}} \frac{\|y(t) - y(s)\|_X}{|t - s|^\alpha}, \quad \|y\|_{C^\alpha(I,X)} := |y|_{C^\alpha(I,X)} + \|y\|_{L^\infty(I,X)}$$

Lastly, we introduce the following duality pairing for $y \in L^2(I, X)$ and $v \in L^2(I, X^*)$:

$$(y, v)_{L^2(I,X,X^*)} := \int_I (y(t), v(t))_{X,X^*} \, dt$$

or with the ordner of X and X^* reversed.

Depending on the situation we will sometimes drop the arguments from the integrands.

Lemma 1.19 *For any $p \in [1, \infty]$, I a real, bounded interval and any real Banach space X we have that $L^p(I, X)$ is a Banach space. $C(\bar{I}, X)$ is also a Banach space. If X is a Hilbert space and $p = 2$ we can introduce an inner product. Let $y, v \in L^2(I, X)$, then the following is an inner product*

$$(y, v)_{L^2(I,X)} := \int_I (y(t), v(t))_X \, dt.$$

Proof. The first two statements are just [Emm04, Satz 7.1.23i)] and [Emm04, Satz 7.1.1]. The last statement can be found in [Emm04, Satz 7.1.23vii)]. □

Definition 1.20 For $y \in L^p(I, X)$ with p, I and X as in Definition 1.18 we define $\partial_t y \in L^p(I, X)$, if it exists, by

$$\int_I \partial_t y(t)\varphi(t) \, dt = -\int_I y(t)\partial_t \varphi(t) \, dt \qquad \forall \varphi \in C_c^\infty(I).$$

We thus define for $k \in \mathbb{N}_0$

$$W^{k,p}(I, X) := \left\{ y \in L^p(I, X) : \partial_t^k y \in L^p(I, X) \right\}.$$

We call $V \subset H \subset V^*$ a Gelfand triple if H, V are Hilbert spaces and all embeddings are continuous and dense. V^* is purposefully not identified with V. According to [Wlo92, Theorem 17.1 and the short comment in Definition 17.1] it is sufficient for the first embedding to be continuous and dense as this implies the same for the second embedding. Given a Gelfand triple we define

$$W(I) := \left\{ y \in L^2(I, V) : \partial_t y \in L^2(I, V^*) \right\}.$$

It is a Hilbert space equipped with the inner product

$$(f, g)_{W(I)} := (f, g)_{L^2(I,V)} + (\partial_t f, \partial_t g)_{L^2(I,V^*)}$$

according to [Wlo92, Theorems 25.4, 25.5].

The following is a collection of well-known results and can be found in [Emm04, Satz 8.1.9 and Korollar 8.1.10].

Lemma 1.21 *Assume we have a Gelfand triple of Hilbert spaces H, V. Let $I = (a, b) \subset \mathbb{R}$. Then we have the continuous embedding $W(I) \hookrightarrow C(\bar{I}, H)$.*

For any $y, v \in W(I)$ we have

$$\int_I (\partial_t y(t), v(t))_{V^*,V} \, dt = (y(T), v(T))_H - (y(0), v(0))_H - \int_I (\partial_t v(t), y(t))_{V^*,V} \, dt.$$

This entails for $y \in W(I)$

$$\int_I (\partial_t y(t), y(t))_{V^*,V} \, dt = \frac{1}{2}\|y(T)\|_H^2 - \frac{1}{2}\|y(0)\|_H^2.$$

Definition 1.22 Let $p \in [1, \infty]$, $\Omega \subset \mathbb{R}^N$ a bounded domain and $I := (a, b) \subset \mathbb{R}$. We abbreviate $Q := I \times \Omega$ and define

$$W_p^{1,2}(Q) := \left\{ y \in L^p(I, W^{2,p}(\Omega)) : \partial_t y \in L^p(I, L^p(\Omega)) \right\}.$$

Equipped with the norm

$$\|y\|_{W_p^{1,2}(Q)} := \|y\|_{L^p(I,W^{2,p}(\Omega))} + \|y\|_{W^{1,p}(I,L^p(\Omega))}$$

it is a Banach space. This can be proven completely analogously to the case of real valued Sobolev spaces, see for example [Eva98, proof of Theorem 5.2.3.2].

The following result can be found in [Emm04, Satz 7.1.23].

Lemma 1.23 *Let $p \in [1, \infty)$, $I := (a, b) \subset \mathbb{R}$ and X a reflexive Banach space. Further let $q \in (1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$L^p(I, X)^* \simeq L^q(I, X^*).$$

This entails that $L^p(I, X)$ is reflexive.

The next result is just a repeated application of [Emm04, Satz 7.1.24].

Lemma 1.24 *For $p \in (1, \infty)$ we have*

$$L^p(I, L^p(\Omega)) \simeq L^p(Q).$$

The following result is usually refered to as folklore, but the author was not able to find a source and thus provides a proof.

Lemma 1.25 *For $p \in (1, \infty)$ the space $W_p^{1,2}(Q)$ is reflexive.*

Proof. We have

$$W_p^{1,2}(Q) = L^p(I, W^{2,p}(\Omega)) \cap W^{1,p}(I, L^p(\Omega)).$$

We first show that $W^{2,p}(\Omega)$ is reflexive, then we immediately get by the previous lemma that $L^p(I, W^{2,p}(\Omega))$ is reflexive.

We define the map

$$\begin{aligned} T\colon W^{2,p}(\Omega) &\to L^p(\Omega) \times L^p(\Omega)^N \times L^p(\Omega)^{N \times N} \\ y &\mapsto (y, \nabla y, \nabla^2 y). \end{aligned}$$

Choosing the natural norm for the space on the right hand side, i.e.

$$\|(a, b, c)\|_{L^p(\Omega) \times L^p(\Omega)^N \times L^p(\Omega)^{N \times N}} = \left(\|a\|^p_{L^p(\Omega)} + \|b\|^p_{L^p(\Omega)^N} + \|c\|^p_{L^p(\Omega)^{N \times N}} \right)^{\frac{1}{p}}$$

shows that T is an isometry. The image of T is thus a closed subspace of $L^p(\Omega) \times L^p(\Omega)^N \times L^p(\Omega)^{N \times N}$, which is just the Cartesian product of $1 + N + N^2$ copies of the reflexive $L^p(\Omega)$ and thus reflexive. Closed subspaces of reflexive spaces are reflexive, see for example [Alt99, Theorem 6.8].

The space $W^{1,p}(I, L^p(\Omega))$ is also a Banach space and can be treated similarly with the map

$$\begin{aligned} T\colon W^{1,p}(I, L^p(\Omega)) &\to L^p(Q) \times L^p(Q) \\ y &\mapsto (y, \partial_t y)\,. \end{aligned}$$

Arguing as before and using the previous theorem we now see that $W^{1,p}(I, L^p(\Omega))$ is also reflexive.

Now we consider the map:

$$\begin{aligned} T\colon W_p^{1,2}(Q) &\to L^p(I, W^{2,p}(\Omega)) \times W^{1,p}(I, L^p(\Omega)), \\ y &\mapsto (y, y). \end{aligned}$$

So the image of T is just the diagonal of $L^p(I, W^{2,p}(\Omega)) \times W^{1,p}(I, L^p(\Omega))$. The diagonal of a product of two Banach spaces is closed, thus the image of T is a closed set of a reflexive space and [Alt99, Theorem 6.8] again delivers the desired result. □

Lemma 1.26 *Assume Ω is a bounded Lipschitz domain. Then for $p \in (\max(1, N-1), \infty)$ the space $W_p^{1,2}(Q)$ embedds compactly into $L^p(I, W^{1,p}(\Omega))$.*

Proof. By Theorem 1.9 Ω satisfies the cone condition. This allows us to see that by [Ada75, Theorem 6.2] the space $W^{2,p}(\Omega)$ embedds compactly in $W^{1,p}(\Omega)$. Now the claim follows by [Sim85, Corollary 2]. □

It is possible to generalize the result of the density of smooth functions in Lebesgue spaces to Bochner spaces, see for example [MS17, Lemma A.1].

Theorem 1.27 *Let $I := (a, b) \subset \mathbb{R}$. $C_c^\infty(I, V)$ is dense in $L^p(I, V)$ for any $p \in [1, \infty)$ and any separable Banach space V.*

We can now improve this result by explicitly constructing the sequences of smooth functions approximating non-smooth functions.

Corollary 1.28 *Let $I := (a, b) \subset \mathbb{R}$. Let $(\eta_\epsilon)_{\epsilon>0} \subset C_c^\infty(I)$ be a family of functions satisfying for any $\epsilon > 0$*

$$\operatorname{supp}(\eta_\epsilon) \subset [-\epsilon, \epsilon], \quad \eta_\epsilon \geq 0, \quad \int_{-\infty}^{\infty} \eta_\epsilon(t)\, dt = 1.$$

Let V be a separable Banach space, $p \in [1, \infty)$ and $f \in L^p(I, V)$. Then

$$f_\epsilon := f * \eta_\epsilon = \int_{\mathbb{R}} f(s)\eta_\epsilon(\cdot - s)\, ds$$

converges to f in $L^p(I, V)$. We implicitly extend f by 0 on $\mathbb{R} \setminus I$.

Proof. Let $\delta > 0$. By Theorem 1.27 there exists a $g \in C_c^\infty(I, V)$ such that

$$\|f - g\|_{L^p(I,V)} \leq \delta.$$

Then we have

$$\begin{aligned}\|f - f_\epsilon\|_{L^p(I,V)} &\leq \|f - g\|_{L^p(I,V)} + \|g - g_\epsilon\|_{L^p(I,V)} + \|g_\epsilon - f_\epsilon\|_{L^p(I,V)}, \\ &\leq \delta + \|g - g_\epsilon\|_{L^p(I,V)} + \|g_\epsilon - f_\epsilon\|_{L^p(I,V)}.\end{aligned}$$

We have $g_\epsilon - f_\epsilon = (g - f)_\epsilon$. We abbreviate $h = g - f$ for the moment. For any $t \in \mathbb{R}$ we have for $q = \frac{p}{p-1}$, with the appropriate changes for $p = 1$ in the following,

$$\|h_\epsilon(t)\|_V \leq \int_{\mathbb{R}} \|h(s)\|_V \eta_\epsilon(t-s)\, ds \leq \left(\int_{\mathbb{R}} \|h(s)\|_V^p \eta_\epsilon(t-s)\, ds\right)^{\frac{1}{p}} \left(\int_{\mathbb{R}} \eta_\epsilon(t-s)\, ds\right)^{\frac{1}{q}}.$$

By assumption the second factor is equal to 1 and thus

$$\int_{\mathbb{R}} \|h_\epsilon(t)\|_V^p\, dt \leq \int_{\mathbb{R}} \int_{\mathbb{R}} \|h(s)\|_V^p \eta_\epsilon(t-s)\, ds\, dt = \int_{\mathbb{R}} \|h(s)\|_V^p \int_{\mathbb{R}} \eta_\epsilon(t-s)\, dt\, ds = \int_{\mathbb{R}} \|h(t)\|_V^p\, dt.$$

Thus

$$\|g_\epsilon - f_\epsilon\|_{L^p(I,V)} \leq \|g - f\|_{L^p(I,V)} \leq \delta.$$

This yields

$$\|f - f_\epsilon\|_{L^p(I,V)} \leq 2\delta + \|g - g_\epsilon\|_{L^p(I,V)}.$$

For any $t \in \mathbb{R}$ we have

$$\|g(t) - g_\epsilon(t)\|_V = \left\| \int_{\mathbb{R}} (g(s) - g(t))\eta_\epsilon(t-s)\, ds \right\|_V \leq \sup_{s \in (t-\epsilon, t+\epsilon)} \|g(s) - g(t)\|_V.$$

Since the support of g is compact, g is uniformly continuous and there exists an $\epsilon_0 > 0$, independent of t, such that

$$\sup_{s \in (t-\epsilon, t+\epsilon)} \|g(s) - g(t)\|_V \leq \delta$$

holds for any $\epsilon < \epsilon_0$. This implies for $\epsilon < \epsilon_0$

$$\|g - g_\epsilon\|_{L^\infty(I,V)} \leq \delta$$

and thus

$$\|f - f_\epsilon\|_{L^p(I,V)} \leq 2\delta + |I|^{\frac{1}{p}}\delta.$$

As $\delta > 0$ was arbitrary this concludes the proof. □

Proposition 1.29 *Let V be a separable Banach space and $I := (a,b) \subset \mathbb{R}$. Let $p \in [1,\infty), k \in \mathbb{N}_0$. Then for any $f \in W^{k,p}(I,V)$ and any $\hat{I} \subset\subset I$ we have for f_ϵ from Corollary 1.28*

$$\|f_\epsilon - f\|_{W^{k,p}(\hat{I},V)} \xrightarrow{\epsilon \to 0} 0.$$

We also have $f_\epsilon \in C^\infty(\hat{I}, V)$.

Proof. Let $(\eta_\epsilon)_{\epsilon>0}$ be as in Corollary 1.28. Let $f \in W^{1,p}(I,X)$. We may assume ϵ to be so small that $\hat{I} + [-\epsilon,\epsilon] \subset I$. This implies that $\eta_\epsilon(T-s) = \eta_\epsilon(-s) = 0$ for any $s \in \hat{I}$.

We show that $f_\epsilon := f * \eta_\epsilon$ lies in $W^{1,p}(\hat{I},X)$ with $\partial_t f_\epsilon = f * \eta'_\epsilon = (\partial_t f)_\epsilon$. Let $\varphi \in C_c^\infty(\hat{I})$, then

$$\begin{aligned} \int_I f_\epsilon(t)\varphi'(t)\, dt &= \int_{\mathbb{R}} f(s) \int_I \eta_\epsilon(t-s)\varphi'(t)\, dt\, ds \\ &= -\int_{\mathbb{R}} f(s) \int_I \eta'_\epsilon(t-s)\varphi(t)\, dt\, ds \\ &= -\int_I \int_{\mathbb{R}} f(s)\eta'_\epsilon(t-s)\, ds\varphi(t)\, dt \\ &= -\int_I \int_{\mathbb{R}} \partial_t f(t)\eta_\epsilon(t-s)\, ds\varphi(t)\, dt. \end{aligned} \tag{1.1}$$

Iterating this for $k > 1$ yields the desired regularity of f_ϵ and Corollary 1.28 yields the desired convergence.

Note that the calculations in (1.1) also show that f_ϵ lies indeed in $C^\infty(\hat{I},V)$ with the derivatives $f_\epsilon^{(k)} = f * \eta_\epsilon^{(k)}$ for $k \in \mathbb{N}$. □

Theorem 1.30 *Let V be a separable Banach space and $I := (a,b) \subset \mathbb{R}$. Let $p \in [1,\infty), k \in \mathbb{N}_0$. Then for any $f \in W^{k,p}(I,V)$ there exists a sequence $(f_\epsilon)_{\epsilon>0} \subset C^\infty(\bar{I},V) \subset C^\infty(I,V) \cap W^{k,p}(I,V)$ such that*

$$\|f_\epsilon - f\|_{W^{k,p}(I,V)} \xrightarrow{\epsilon \to 0} 0.$$

Proof. Let $(\rho_j)_{j\in\mathbb{N}}$ be a smooth partition of unity on I, cf. [Wlo92, Theorem 1.2]. We define the compact sets $\hat{I}_j := \operatorname{supp}(\rho_j)$. Let $\epsilon > 0$, by Proposition 1.29 there exists for each $\hat{I}_j$ a function $f^j_{\epsilon_j} \in C^\infty(\hat{I}_j, V)$ such that

$$\|\rho_j f - f^j_{\epsilon_j}\|_{W^{k,p}(\hat{I},V)} \le 2^{-j}\epsilon.$$

We define $f_\epsilon := \sum_{j=1}^\infty f^j_{\epsilon_j} \in C^\infty(I, V)$. Note that for each $t \in I$ only finitely many summands of this series are unequal from 0. Then

$$\|f_\epsilon - f\|_{W^{k,p}(I,V)} = \left\| \sum_{j=1}^\infty (f^j_{\epsilon_j} - f) \right\|_{W^{k,p}(I,V)} \le \sum_{j=1}^\infty \|f^j_{\epsilon_j} - \rho_j f\|_{W^{k,p}(I,V)} \le \epsilon.$$

□

Lemma 1.31 *Let Y, X separable Banach spaces such that Y is dense in X. Let $I := (a, b) \subset \mathbb{R}$ and $p \in [1, \infty)$. Then $C^\infty(\bar{I}, Y)$ is dense in $W^{1,p}(I, X) \cap L^p(I, Y)$.*

Proof. We make the same construction as in the proof of Theorem 1.30. Let $(\rho_j)_{j\in\mathbb{N}}$ be a smooth partition of unity on I, see again [Wlo92, Theorem 1.2]. We define the compact sets $\hat{I}_j := \operatorname{supp}(\rho_j) \subset I$. Let $\epsilon > 0$. By Proposition 1.29 for each $\hat{I}_j$ there exists a function $f^j_{\epsilon_j} \in C^\infty(\hat{I}_j, Y) \subset C^\infty(\hat{I}_j, X)$ such that

$$\|\rho_j f - f^j_{\epsilon_j}\|_{W^{1,p}(\hat{I},X)} + \|\rho_j f - f^j_{\epsilon_j}\|_{L^p(\hat{I},Y)} \le 2^{-j}\epsilon.$$

Now the rest follows as in the proof of Proposition 1.29. □

Remark 1.32 A special case for the situation in Lemma 1.31 is $Y = C^\infty(\bar{\Omega})$ and $X = L^p(\Omega)$. Here $\Omega \subset \mathbb{R}^N$ is a bounded domain, $k \in \mathbb{N}$, $I := (a, b) \subset \mathbb{R}$ and $p \in [1, \infty)$. Lemma 1.31 then states that

$$\overline{C^\infty(\bar{I}, C^\infty(\bar{\Omega}))}^{\|\cdot\|_{W^{1,2}_p(Q)}} = W^{1,2}_p(Q).$$

1.5 Interpolation Spaces

We shortly introduce interpolation spaces, required to consider appropriate initial conditions for our parabolic equations and inequalities. This section wholly follows [Ama95, Chapter I.2]. We later use deeper results for interpolation spaces in Section 8.6 to Section 8.8 for the numerical analysis and interpolation error estimates.

Definition 1.33 A pair (E_0, E_1) of real Banach spaces is called an interpolation couple, if there exists a locally compact space X such that both embed continuously into X.

E is called an intermediate space with respect to the interpolation couple (E_0, E_1) if we have the continuous embeddings

$$E_0 \cap E_1 \hookrightarrow E \hookrightarrow E_0 + E_1.$$

Let $x \in E_0 + E_1$, $t > 0$ and define

$$K(t, x) = K(t, x, E_0, E_1) = \inf\{\|x_0\|_{E_0} + t\|x_1\|_{E_1} : x = x_0 + x_1\}.$$

Let $\theta \in (0,1), q \in [1,\infty]$. Then we define the interpolation space

$$(E_0, E_1)_{\theta,q} := \{x_0 \in E_0 + E_1 : \|x\|_{\theta,q} < \infty\}$$

equipped with the norm

$$\|x\|_{\theta,q} = \|t^{-\theta} K(t,x)\|_{L^q_*(\mathbb{R}_{>0})}.$$

Here $L^q_*(\mathbb{R}_{>0})$ is the vector space of functions $f\colon \mathbb{R}_{>0} \to \mathbb{R}$ such that

$$\|f\|_{L^q_*(\mathbb{R}_{>0})} = \left(\int_0^\infty f(s)^q \frac{ds}{s}\right)^{\frac{1}{q}} < \infty.$$

Theorem 1.34 *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain satisfying the cone condition. For $r, p \in [1,\infty)$ with $\frac{1}{r} + \frac{N}{2p} \in (0,1)$ one has the compact embedding:*

$$W^{1,r}(I, L^p(\Omega)) \cap L^r(I, W^{2,p}(\Omega)) \hookrightarrow C^{\frac{\kappa}{2},\kappa}(Q) := C^{\frac{\kappa}{2}}(I, C^\kappa(\Omega))$$

for any $\kappa \in (0, 1 - \frac{1}{r} - \frac{N}{2p})$. In particular for any $p \in (1 + \frac{N}{2}, \infty)$ we have:

$$W^{1,2}_p(Q) \hookrightarrow C^{\frac{\kappa}{2},\kappa}(Q) \text{ compactly}$$

for any $\kappa \in (0, 1 - \frac{1}{p}(1 + \frac{N}{2}))$.

Proof. Let X, Y be function spaces, such that Y is densely embedded into X. By [Ama01, Theorem 3] and a rescaling argument we have the continuous embedding

$$W^{1,r}(I, X) \cap L^r(I, Y) \hookrightarrow C^{\kappa'}(I, (X,Y)_{\tau,1}) \tag{1.2}$$

for any $\kappa' \in (0, 1 - 1/r - N/(2p))$ and $\tau \in [N/(2p), 1 - 1/r - \kappa')$.

By [Tri78, Theorem 4.2.6] there exist norm preserving extension operators

$$S_0\colon L^p(\Omega) \to L^p(\mathbb{R}^N),$$
$$S_2\colon W^{2,p}(\Omega) \to W^{2,p}(\mathbb{R}^N).$$

By [Tri78, Theorem 1.2.4] this implies that there exists a norm preserving extension operator

$$S : \left(L^p(\Omega), W^{2,p}(\Omega)\right)_{\tau,1} \to \left(L^p(\mathbb{R}^N), W^{2,p}(\mathbb{R}^N)\right)_{\tau,1}. \tag{1.3}$$

Thus combining (1.2), (1.3) and chosing $X = L^p(\Omega)$ and $Y = W^{2,p}(\Omega)$ we have the continuous embedding

$$W^{1,r}(I, L^p(\Omega)) \cap L^r(I, W^{2,p}(\Omega)) \hookrightarrow C^{\kappa'}(I, \left(L^p(\mathbb{R}^N), W^{2,p}(\mathbb{R}^N)\right)_{\tau,1}).$$

By [BL76, Theorem 6.4.5] and [Tri78, Theorem 2.8.1c)] this implies

$$\begin{aligned} W^{1,r}(I, L^p(\Omega)) \cap L^r(I, W^{2,p}(\Omega)) &\hookrightarrow C^{\kappa'}(I, B^{2\tau}_{p,1}(\mathbb{R}^N)) \hookrightarrow C^{\kappa'}(I, C^{2\tau - \frac{N}{p}}(\mathbb{R}^N)) \\ &\hookrightarrow C^{\kappa'}(I, C^{2\tau - \frac{N}{p}}(\Omega)) \end{aligned}$$

because $\tau \geq \frac{N}{2p}$ by its earlier choice. $B^{2\tau}_{p,1}(\mathbb{R}^N)$ denotes a Besov space introduced in greater detail in [BL76, Definition 6.2.2].

We now let $\kappa \in (0, 1 - 1/r - N/(2p))$. We then choose $\kappa' = \frac{\kappa}{2} \in (0, \frac{1}{2}(1 - 1/r - N/(2p)))$ and $\tau = \frac{1}{2}(\kappa + \frac{N}{p})$. Clearly $\tau \geq N/(2p)$ and also, $\tau = \kappa + N/(2p) - \kappa/2 < 1 - 1/r - \kappa/2$. Thus we can infer

$$W^{1,r}(I, L^p(\Omega)) \cap L^r(I, W^{2,p}(\Omega)) \hookrightarrow C^{\frac{\kappa}{2}}(I, C^{\kappa}(\Omega)).$$

These embeddings are compact by the following argument: Choose $\kappa \in [0, 1 - \frac{1}{r} - \frac{N}{2p})$ and $\tilde{\kappa} \in (\kappa, 1 - \frac{1}{r} - \frac{N}{2p})$. Because Hölder spaces embedd compactly in Hölder spaces of lower exponent we have

$$W^{1,r}(I, W^{2,p}(\Omega)) \hookrightarrow C^{\frac{\tilde{\kappa}}{2}, \tilde{\kappa}}(Q) \subset\subset C^{\frac{\kappa}{2}, \kappa}(Q).$$

This concludes the main part of the proof. Choosing $p = r \in (1 + \frac{N}{2p}, \infty)$ yields the special case. $\square$

2 Obstacle Problems and their Regularization via Semilinear PDEs

2.1 Variational Inequalities in an Abstract Setting

In this section we consider parabolic variational inequalities in an abstract setting, see for example [Bar84, Chapter 4]. In later sections we give a more concrete setting and numerical analysis for that specific setting. We nevertheless quickly have a look at the abstract situation to see the underlying structures and general ideas before we study the obstacle problem specifically.

Definition 2.1 Let V, H, V^* be a Gelfand triple. With $I = (0, T)$ we denote the bounded time interval for some end time $T > 0$.

We consider a linear, continuous, symmetric operator $A : V \to V^*$ that satisfies Gårding's inequality for some $\nu_H, \nu_V > 0$:

$$(Av, v)_{V^*,V} + \nu_H \|v\|_H^2 \geq \nu_V \|v\|_V^2 \qquad \forall v \in V.$$

We introduce the following bilinear forms:

$$\begin{aligned} &a_\Omega \colon V \times V \mapsto \mathbb{R}, \quad (y, v) \mapsto (Ay, v)_{V^*,V}, \\ &a_I \colon L^2(I, V) \times L^2(I, V) \to \mathbb{R}, \quad (y, v) \mapsto \int_I (Ay(t), v(t))_{V^*,V} \, dt. \end{aligned} \tag{2.1}$$

For subsets of I we use the appropriate notation.

We also consider a convex, proper, lower semi-continuous function $\Phi : H \to \bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$. We assume that $\mathrm{dom}_V(\Phi)$ is dense in $\mathrm{dom}_H(\Phi)$. Note, $\Phi : V \to \bar{\mathbb{R}}$ is also convex and lower semi-continuous with respect to the norm topology of V. We shall assume that there is a function $\hat{0} \in \mathrm{dom}_V(\Phi)$. Note that we may not necessarily have $\hat{0} = 0$.

We keep our problem formulation close to [Bar84, Section 4.1]. Given a $y_0 \in H$ and $u \in L^2(I, V^*)$ we now look for a solution $y \in W(I)$ of

$$\begin{cases} (\partial_t y(t) + Ay(t), v - y(t))_{V^*,V} + \Phi(v) - \Phi(y(t)) \geq (u(t), v - y(t))_{V^*,V} \\ \hfill \text{a.e. in } I, \ \forall v \in V, \\ y(0) = y_0. \end{cases} \tag{VIabs}$$

Using the subdifferential, cf. [BC11, Chapter 16], $\partial\Phi : H \to \mathcal{P}(H)$ of the convex map Φ this is equivalent to writing

$$\begin{cases} \partial_t y(t) + Ay(t) + \partial\Phi(y(t)) \ni u(t) \text{ a.e. in } I, \\ y(0) = y_0. \end{cases} \tag{2.2}$$

We call a solution to (VIabs) a strong solution.

Remark 2.2 We can also define weak solutions $y \in C(\bar{I}, H) \cap L^2(I, V)$ by

$$\int_I (\partial_t v, v - y)_{V^*,V} + (Ay, v - y)_{V^*,V} + \Phi(v) - \Phi(y)\, dt + \frac{1}{2}\|v(0) - y_0\|_H^2 \\ \geq \int_I (u, v - y)_{V^*,V}\, dt \qquad \forall v \in W(I). \tag{2.3}$$

This concept of solution is obtained by integration by parts and for example used in [IK06]. We do not delve into this any further as it is clear from the structure of (2.3) that very little regularity information on the state can be gained. Since the ultimate goal of this thesis is to pursue numerical analysis, this approach is not fruitful for us.

We restate [Bar84, Theorem 4.1].

Theorem 2.3 *Let $y_0 \in \mathrm{dom}_H(\Phi)$ and $u \in L^2(I, H)$.* (VIabs) *has a unique strong solution $y := y(y_0, u) \in H^1(I, H) \cap L^2(I, V)$. It satisfies*

$$\|y(y_0^1, u^1) - y(y_0^2, u^2)\|_{L^2(I,V)\cap C(\bar{I},H)} \leq C\left(\|u^1 - u^2\|_{L^2(I,H)} + \|y_0^1 - y_0^2\|_H\right).$$

for $y_0^1, y_0^2 \in H$, $u^1, u^2 \in L^2(I, H)$.

Remark 2.4 Now we can consider an abstract version of the obstacle problem, which we make concrete in the following section. Let $\emptyset \neq K \subset H$ be a convex and closed set and $f : H \to \mathbb{R}$ be the continuous Gâteaux derivative of a convex functional $F : H \to \mathbb{R}$. We define $\Phi := F + \chi_K$, where χ_K denotes the convex indicator functional of K, i.e. $\chi_K(v) = 0$ if $v \in K$ and $\chi_K(v) = \infty$ is $v \notin K$. Φ has the desired properties because F is convex, continuous and therefore proper and K is closed and not empty making χ_K convex, closed and proper as well.

The variational inequality (VIabs) in the version of (2.2) is then equivalent to

$$\begin{cases} \partial_t y(t) + Ay(t) + f(y(t)) + \partial\chi_K(y(t)) \ni u(t) \text{ a.e. in } I, \\ y(0) = y_0. \end{cases} \tag{2.4}$$

Here we used [BC11, Corollary 16.38] and the charaterization $\partial F(y) = \{f(y)\}$, which can be found in [BC11, Proposition 17.26]. This is equivalent to

$$\begin{cases} (\partial_t y(t) + Ay(t) + f(y(t)), v - y(t))_{V^*,V} \geq (u(t), v - y(t))_{V^*,V} \qquad \text{a.e. in } I,\ \forall v \in V \cap K, \\ y(0) = y_0. \end{cases}$$

2.2 Obstacle Problems

It is possible, though, to derive strong statements about regularity for the obstacle problem when not working in the abstract setting. For the numerical analysis this will be of utmost importance. That is why we present a concrete model case for the obstacle problem and work in that setting for the rest of this thesis.

2.2.1 Standing Assumptions

For the rest of Chapter 2 the following assumptions and definitions shall apply.

Assumption 2.5 *The space dimension is denoted by* $N \in \mathbb{N}_{>0}$*. The regularity for the control is chosen as some fixed* $q_u \in (1 + N/2, \infty) \cap [2, \infty)$*. The set* $\Omega \subset \mathbb{R}^N$ *is a bounded Lipschitz domain.*

We write $H := L^2(\Omega)$*. For a boundary portion* $\Gamma_D \subset \partial\Omega$ *and* $p \in (1, \infty)$ *we define*

$$C^\infty_{\Gamma_D}(\Omega) := \left\{ v|_\Omega : v \in C^\infty_c(\mathbb{R}^N) \text{ and } \operatorname{supp} v \cap \Gamma_D = \emptyset \right\},$$
$$V := \overline{C^\infty_{\Gamma_D}(\Omega)}^{H^1(\Omega)}, \quad W^{1,p}_{\Gamma_D}(\Omega) := \overline{C^\infty_{\Gamma_D}(\Omega)}^{W^{1,p}(\Omega)}.$$

We define $\Sigma_D := I \times \Gamma_D$*,* $\Gamma_N := \partial\Omega \setminus \Gamma_D$ *and* $\Sigma_N := I \times \Gamma_N$*. The sets* Γ_D *and* Σ_D *are the Dirichlet boundary portions, where the state will vanish. The sets* Γ_N *and* Σ_N *are the corresponding Neumann boundary portions. We assume that* $\Omega \cup \Gamma_N$ *is Gröger regular according to Definition 1.7.We assume that* Γ_D *itself a* $N-1$*-dimensional Lipschitz domain. This entails that the relative boundary of* $\Gamma_D \subset \partial\Omega$ *is of* $N-1$*-Hausdorff measure* 0*. (This may or may not be implied by Gröger regularity, cf. the reference to [Grö89, Remark 1] in Definition 1.7. We do not delve deeper into this question as the focus of this thesis lies somewhere else and simply assume it.)*

This property is necessary so that we have by [Dok73, Theorem 1] that $W^{1,p}_{\Gamma_D}(\Omega)$ *is equal to the kernel of the trace operator* $\gamma_0 : W^{1,p}(\Omega) \to \Gamma_D$*. So in particular each smooth function that vanishes on* Γ_D *lies in* $W^{1,p}_{\Gamma_D}$*. For a more accessible proof in a less general situation also see [DŽ06] co-written by the same author.*

Note that V *is a closed, dense subspace of* H *and that for any* $M \geq 0$ *we have*

$$\|\max(v - M, 0)\|_V \leq C\|v\|_V \qquad v \in V \tag{2.5}$$

by Proposition 8.19.

Definition 2.6 We also introduce a nonlinearity

$$\begin{aligned} f\colon Q \times \mathbb{R} &\to \mathbb{R}, \\ (t, x, y) &\mapsto f(t, x, y). \end{aligned}$$

We often write f with only one argument $y \in \mathbb{R}$, $f(y)$ then refers to the function $f(\cdot, \cdot, y)$. Whenever we write f' it refers to the derivative in the y-component. Derivatives in time or space are not considered. The nonlinearity shall satisfy

- $f'(t, x, y) \geq 0$ and $f'(t, x, \cdot) \in C(\mathbb{R}) \qquad \forall (t, x) \in Q$,
- $f(0), f'(0) \in L^\infty(Q)$,
- $\forall M > 0$ there is an $L(M) > 0$ such that

  $$|f(t, x, y_1) - f(t, x, y_2)| + |f'(t, x, y_1) - f'(t, x, y_2)| \leq L(M)|y_1 - y_2| \tag{2.6}$$

  holds for all $(t, x) \in Q$ and all $y_1, y_2 \in B_M(0)$.

Remark 2.7 This allows for example for nonlinearities of the form $f(t, x, y) = y^3$.

Note that the assumption that f' satisfies a local Lipschitz condition can be weakened in some contexts.

Note that compared to our most general framework for the obstacle problem in Remark 2.4 the obstacle Ψ and the nonlinearity f may be time dependent. We also do not require $y_0 \in V$.

Definition 2.8 We let $A : H^1(\Omega) \to H^1(\Omega)^*$ be a symmetric operator of the form

$$(Av, y)_{V^*,V} = \int_\Omega \sum_{i,j=1}^N a_{ij}(x)\partial_{x_i} v(x)\partial_{x_j} y(x)dx$$

with $a_{ij} \in L^\infty(\Omega)$ and $a_{ij} = a_{ji}$ for any $i, j = 1, \dots, N$. We assume that there is a $\nu_{ell} > 0$ such that for almost all $x \in \bar{\Omega}$ and $\xi \in \mathbb{R}^N$ one has

$$\xi^T A(x)\xi \geq \nu_{ell}|\xi|_2^2 > 0. \tag{2.7}$$

Here we wrote $A(x)$ for the matrix $(a_{ij})_{i,j=1,\dots,N}$. It will be clear from the context whether A refers to the matrix or the operator. As $\Gamma_D = \emptyset$ can happen, (2.7) does not imply the ellipticity of $A|_V$. For readers wondering where the 0-order terms are in the definition of A: for technical reasons we include them in the non-linearity f. This includes bounded, non-negative zero order terms. So, we do not lose generality here.

We use the same bilinear forms introduced in (2.1).

For $p \in [2, \infty)$ we define

$$\begin{aligned} \operatorname{dom}_p(A) &:= \{v \in V \cap L^p(\Omega) : Av \in L^p(\Omega) \subset V^*\}, \\ \mathbb{W}_p &:= (L^p(\Omega), \operatorname{dom}_p(A))_{1-\frac{1}{p},p}. \end{aligned}$$

Definition 2.9 To fully pose the obstacle problem we also need an initial state $y_0 \in \mathbb{W}_{q_u}$. By Proposition 8.51 this implies $y_0 \in C(\bar{\Omega})$. In particular we have that y_0 is bounded on Ω, which we will use frequently. $\mathbb{W}_p$ might be difficult to interpret so we would like to simply note that $W^{2,\infty}(\Omega) \cap V$ or $\operatorname{dom}_p(A)$ embed continuously into $\mathbb{W}_p$ by Proposition 8.32. Those two spaces are both easier to imagine and the first one will play a role in our numerical analysis in the later chapters.

Lastly we of course require an obstacle $\Psi \in L^\infty(Q) \cap L^2(I, H^1(\Omega))$ with $\partial_t \Psi \in L^2(I, H^1(\Omega)^*)$. It shall satisfy $\Psi(0) \leq y_0 \in H$ and $\Psi|_{\Sigma_D} \leq 0$, so that a function $y \in W(I)$ with $y(0) = y_0$, $y|_{\Sigma_D} = 0$ and $y \geq \Psi$ exists. We further assume that $\partial_t \Psi, A\Psi \in L^{q_u}(Q)$. We define

$$K_\Psi := \left\{ v \in L^2(Q) : v \geq \Psi \text{ a.e. in } Q \right\}.$$

The regularity assumptions on Ψ are chosen in such a way that it does not destroy regularity of the state y. When a solution to an obstacle problem touches the obstacle, it clearly cannot be more regular than the obstacle on this set, which is called the active set.

Remark 2.10 If $\Psi \in W^{1,2}_{q_u}(Q)$ and A has Lipschitz continuous coefficients we automatically have $A\Psi \in L^{q_u}(Q)$. To see that first note that Rademacher's theorem, cf. [AFP00, Theorem 2.14], implies that the a_{ij} are differentiable almost everywhere in Ω for all $i, j \in \{1, \dots, N\}$. Then use the weak product rule to obtain

$$-\operatorname{div}(A\nabla\Psi) = -\sum_{i=1}^N \partial_{x_i} \left(\sum_{j=1}^N a_{ij}\partial_{x_j}\Psi \right) = -\sum_{i,j=1}^N a_{ij}\partial_{x_i}\partial_{x_j}\Psi + \partial_{x_j}\Psi\partial_{x_i}a_{ij} \in L^{q_u}(Q).$$

We also have by Theorem 1.34 that $\Psi \in W^{1,2}_{q_u}(Q)$ is Hölder continuous and in particular in $L^\infty(Q)$. Thus any $\Psi \in W^{1,2}_{q_u}(Q)$ that is smaller or equal to 0 on Γ_D and satisfies $\Psi(0) \leq y_0$ on Ω is a valid choice. In particular any constant smaller or equal to $-\|y_0\|_{L^\infty(\Omega)}$ is valid.

Definition 2.11 We now define the obstacle problem for the considered quantities, but give existence and regularity later. $y \in W(I)$ is a solution to the obstacle problem if

$$\begin{cases} y \in K_\Psi, \\ (\partial_t y + Ay, v - y)_{L^2(I,V^*,V)} + (f(y), v - y)_{L^2(I,H)} \geq (u, v - y)_{L^2(I,H)} \, \forall v \in K_\Psi \cap L^2(I,V), \\ y(0) = y_0. \end{cases} \tag{VI-OB}$$

Note that the boundary information is contained in V and thus $W(I)$.

We will note that the considered problem is very similar to the ones studied in [Fet87, IK06, IK10, Bar81, Fri87, AL02]. In contrast we allow our obstacle to be time dependent and in particular non necessarily equal to 0. We also allow for an additional non-linearity. Yet, we work in slightly less abstract spaces H and V then some of the mentioned authors since our focus lies on the regularity of solutions to the obstacle problem. The same obstacle problem is considered in [Dom82] for regularity analysis, but without any proofs. An interesting problem with obstacles from above and below and an additional non-linearity is considered in [Che03]. Then restrictions on f are, however, more strict than ours, while the right hand side is more general. A very general and abstract formulation of essentially our problem can be found in [Bar84, Chapter 4]. Most of the mentioned papers refer to various optimal control problems with a parabolic obstacle problem as a constraint. We will give a more detailed overview in the sections discussing our optimization problem.

Example 2.12 Before we continue we would like to give a simple example that satisfies all the standing assumptions. So that one may see that we do indeed not talk about an empty set of problems. Let $\Omega = B_1(0)$ and $\Gamma_D = \partial\Omega$. We further choose $\Psi = -1$. Then for $u \in L^2(Q)$ a concrete example of (VI-OB) reads

$$\begin{cases} y \geq -1 \text{ a.e. in } Q, \\ (\partial_t y - \Delta y + y^3, v - y)_{L^2(Q)} \geq (u, v - y)_{L^2(Q)} \quad \forall v \in L^2(I,V), \ v \geq -1 \text{ a.e. in } Q, \\ y(0) = 0. \end{cases}$$

2.2.2 Properties of Nonlinearities

We collect some very important continuity and differentiability properties of f.

Lemma 2.13 *As a Nemytskii-operator f is locally Lipschitz continuous from $L^\infty(Q)$ to $L^\infty(Q)$.*

Proof. This is an immediate consequence of the local Lipschitz continuity of f in its last argument. □

Corollary 2.14 *In particular $f(\Psi) \in L^\infty(Q)$.*

Lemma 2.15 *As a Nemytskii-operator f is Fréchet differentiable from $L^\infty(Q)$ to $L^\infty(Q)$.*

Proof. For almost any $(t,x) \in Q$ and $y, d \in L^\infty(Q)$ we have, with the local Lipschitz continuity of f',

$$\begin{aligned} &|f(t,x,y(t,x) + d(t,x)) - f(t,x,y(t,x)) - f'(t,x,y(t,x))d(t,x)| \\ &= \left| \int_0^1 f'(t,x,y(t,x) + sd(t,x)) - f'(t,x,y(t,x)) \, ds \; d(t,x) \right| \\ &\leq L(\|y\|_{L^\infty(Q)} + \|d\|_{L^\infty(Q)}) \int_0^1 s|d(t,x)| \, ds \, |d(t,x)| \leq \frac{L(\|y\|_{L^\infty(Q)} + \|d\|_{L^\infty(Q)})}{2} \|d\|^2_{L^\infty(Q)}. \end{aligned}$$

$L(\|y\|_{L^\infty(Q)} + \|d\|_{L^\infty(Q)})$ is the Lipschitz constant of $f'(t,x,\cdot)$ on the closed ball with radius $\|y\|_{L^\infty(Q)} + \|d\|_{L^\infty(Q)}$. By assumption it does not depend on (t,x). As we send d to 0 we may assume that the Lipschitz constant stays bounded and we obtain the claim. □

2.3 Regularization via Semilinear PDEs

We will now consider a family of regularizations of the obstacle problem established in Section 2.2. To that end we will introduce a term, that is essentially a penalization of the condition $y \geq \Psi$.

2.3.1 Regularization Terms

Definition 2.16 We fix a function $\beta \in C(\mathbb{R})$ such that

1. β is monotonically increasing and locally Lipschitz continuous,
2. $\beta|_{\mathbb{R}_{\geq 0}} = 0$,
3. $\beta(\mathbb{R}_{<0}) = (-\infty, 0)$.

We further define for $\gamma > 0$ the actual regularization term $\beta_\gamma := \frac{1}{\gamma}\beta$.

For some situations we will consider a special family of regularizations.

Proposition 2.17 *Let $\alpha \geq 1$. We define*

$$\beta\colon \mathbb{R} \to \mathbb{R},$$
$$r \mapsto \begin{cases} 0 & \text{if } r \geq 0, \\ -(-r)^\alpha & \text{if } 0 > r \geq -\alpha^{1/(1-\alpha)}, \\ r + \frac{\alpha-1}{\alpha}\alpha^{1/(1-\alpha)} & \text{if } -\alpha^{1/(1-\alpha)} > r. \end{cases}$$

For $\alpha = 1$ the expression $\alpha^{1/(1-\alpha)}$ is considered to be 0. This is motivated by $\alpha^{1/(1-\alpha)} \to 0$ for $\alpha \xrightarrow{\alpha>1} 1$. Thus the second case factually drops out.

The function β satisfies the assumptions from Definition 2.16. Even stronger we have $\beta \in C^{0,1}(\mathbb{R})$. For $\alpha > 1$ we have $\beta \in C^1(\mathbb{R})$. For $\alpha \geq 2$ we even have $\beta \in C^{1,1}(\mathbb{R})$.

Proof. We first check the continuity of β at $-\alpha^{1/(1-\alpha)}$ as at 0 the term β is clearly continuous. Coming from the left we have

$$-(-(-\alpha^{1/(1-\alpha)}))^\alpha = -\alpha^{\alpha/(1-\alpha)}$$

and coming from the right we have

$$-\alpha^{1/(1-\alpha)} + \frac{\alpha-1}{\alpha}\alpha^{1/(1-\alpha)} = -\alpha^{-1}\alpha^{1/(1-\alpha)} = -\alpha^{1/(1-\alpha)-1} = -\alpha^{\alpha/(1-\alpha)}.$$

Thus β is continuous and 2. and 3. of Definition 2.16 are clearly satisfied. The Lipschitz continuity is clear for $\alpha = 1$. For $\alpha > 1$ this will follow from the boundedness of the derivative which we prove next.

For the rest of the proof let $\alpha > 1$. Then the derivative on each section is given by

$$\beta'(r) = \begin{cases} 0 & \text{if } r \geq 0, \\ \alpha(-r)^{\alpha-1} & \text{if } 0 > r \geq -\alpha^{1/(1-\alpha)}, \\ 1 & \text{if } -\alpha^{1/(1-\alpha)} > r. \end{cases}$$

Seeing that

$$\alpha(-(-\alpha)^{1/(1-\alpha)})^{\alpha-1} = 1 \text{ and } -\alpha(-0)^{\alpha-1} = 0$$

we have that β' is continuous at the critical points. It is clearly bounded.

To see the Lipschitz continuity of the derivative for $\alpha \geq 2$ we see that we have

$$\beta''(r) = \begin{cases} 0 & \text{if } r \geq 0, \\ -\alpha(\alpha-1)(-r)^{\alpha-2} & \text{if } 0 > r \geq -\alpha^{1/(1-\alpha)}, \\ 0 & \text{if } -\alpha^{1/(1-\alpha)} > r, \end{cases}$$

in a weak sense. This is clearly bounded so a version of Rademacher's theorem, e.g. [AFP00, Proposition 2.13], delivers the Lipschitz continuity of β'.

□

The Nemytskii-operator properties for $\beta_\gamma(\cdot - \Psi)$ and their proofs are the same as for f in Section 2.2.2.

Lemma 2.18 *$\beta_\gamma(\cdot - \Psi)$ is Fréchet differentiable from $L^\infty(Q)$ to $L^\infty(Q)$ as a Nemytskii operator. In particular it is locally Lipschitz continuous and $\beta_\gamma(-\Psi) \in L^\infty(Q)$.*

Definition 2.19 We consider the following semilinear parabolic PDE for $y \in W(I)$

$$\begin{cases} \partial_t y + Ay + f(y) + \beta_\gamma(y - \Psi) = u, \\ y|_{\Sigma_D} = 0, \; y(0) = y_0, \end{cases} \tag{PDE$_\gamma$}$$

which has the weak formulation

$$\begin{cases} (\partial_t y + Ay, v)_{L^2(I,V^*,V)} + (f(y) + \beta_\gamma(y - \Psi), v)_{L^2(Q)} = (u, v)_{L^2(Q)} & \forall v \in L^2(I, V), \\ y(0) = y_0. \end{cases} \tag{2.8}$$

This definition is a relaxation of (VI-OB) by replacing the constraint $y \geq \Psi$ with a (smooth) penalization term β_γ.

This type of regularization is used frequently for the elliptic and parabolic obstacle problem. For the usage in the elliptic case we would like to point out [Noc88] specifically, as it will be the basis for later arguments. For the usage in parabolic obstacle problems see for example [AL02, Che03, Fri87]. Different regularizations are for example used in [Bar81, IK06, IK10].

Remark 2.20 A regularization of the form (PDE$_\gamma$) is motivated by the original formulation of (VI-OB) interpreted in the abstract setting of (2.4). For presentation's sake we drop the non-linearity f from the discussion in this remark. Its inclusion is straightforward.

Assume that Ψ is not time dependent. We define $\Phi = \chi_{K_\Psi}$ as above (2.4). We will show that

$$\partial\Phi \colon H \to \mathcal{P}(H), \quad y \mapsto \begin{cases} \emptyset & \text{if } y \notin K_\Psi, \\ \{g^* \in H : g|_{\{y>\Psi\}} = 0, g|_{\{y=\Psi\}} \leq 0\} & \text{else.} \end{cases}$$

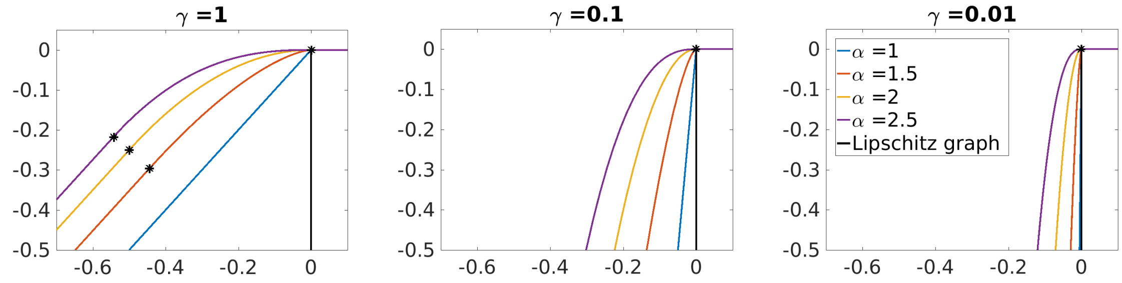


Figure 2.1: The regularization terms from Proposition 2.17 for various α and γ. The black * denote points where the function changes its definition sections.

Here, the active set $\{y = \Psi\} := \{(t,x) \in Q : y(t,x) = \Psi(t,x)\}$ has to be interpreted for fixed representatives of y and Ψ and the properties of g hold almost everywhere. The same hold true for $\{y > \Psi\}$, which is defined analogously.

To see that this is actually the correct interpretation of the subdifferential let $y \in K_\Psi$. The situation $y \notin K_\Psi$ is trivial. Let $g \in \partial\Phi(y)$, this means by definition

$$0 \geq (g, v - y)_H \qquad \forall v \in K_\Psi.$$

Let $\epsilon > 0$ and $E \subset \{y \geq \Psi + \epsilon\}$. The following statements holds almost everywhere. Then $v^\pm := y \pm \epsilon 1_E \in K_\Psi$ and $0 \geq (g, \pm 1_E)_H$. Thus $0 = \int_E g\,dx$. As $E \subset \{y \geq \Psi + \epsilon\}$ was arbitrary we have $g|_{\{y \geq \Psi + \epsilon\}} = 0$. As $\epsilon > 0$ was arbitrary this implies $g|_{\{y > \Psi\}} = 0$. Let $z \geq 0$ on $\{y = \Psi\}$ and choose $v := 1_{\{y>\Psi\}} y + 1_{\{y=\Psi\}}(z + \Psi) \in K_\Psi$, then

$$0 \geq (g, v - y)_H = (g, z)_{L^2(\{y=\Psi\})}.$$

Thus $g|_{\{y=\Psi\}} \leq 0$. It is also easy to see the any $g \in H$ with $g|_{\{y=\Psi\}} \leq 0$ and $g|_{\{y>\Psi\}} = 0$ lies in $\partial\Phi(y)$.

Defining the Lipschitz graph

$$\partial\hat{\Phi} \colon \mathbb{R} \to \mathcal{P}(\mathbb{R}),$$
$$r \mapsto \begin{cases} 0 & \text{if } r > 0, \\ (-\infty, 0] & \text{if } r = 0, \\ \emptyset & \text{if } r < 0, \end{cases}$$

The inclusion formulation in (2.4) is now equivalent to

$$\begin{cases} \partial_t y(t) + Ay(t) + \partial\hat{\Phi}(y(t) - \Psi(t)) \ni u(t) \text{ a.e. in } I, \\ y(0) = y_0. \end{cases}$$

Now it is easy to see how (2.4) and (PDE$_\gamma$) relate as problems. Visually we can see this in Figure 2.1 and Figure 2.2.

Remark 2.21 Another motivation can be derived from the elliptic case. The elliptic obstacle problem, or at least a simplified model problem, is given by

$$\begin{cases} y \geq \Psi, \\ (\nabla y, \nabla(v - y))_{L^2(\Omega)} \geq (u, v - y)_{L^2(\Omega)} \qquad \forall v \in H_0^1(\Omega), v \geq \Psi \\ y|_{\partial\Omega} = 0. \end{cases}$$

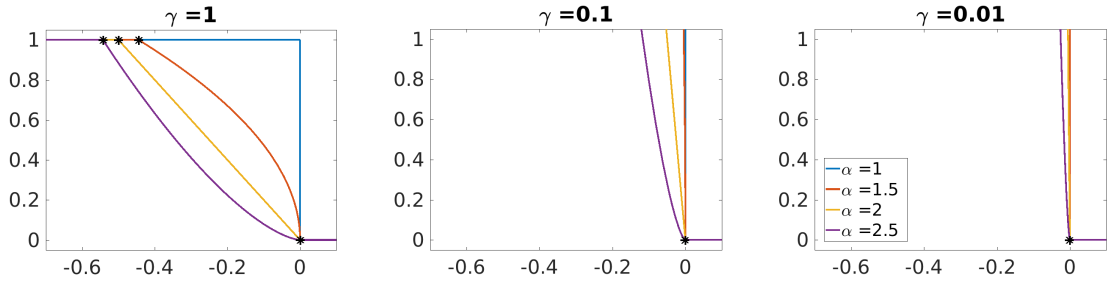


Figure 2.2: The derivatives of the regularization terms from Proposition 2.17 for various α and γ. The black * denote points where the function changes its definition sections.

This is just equivalent to y being the minimizer of

$$\min_{v\in H_0^1(\Omega),v\geq\Psi} \frac{1}{2}\|\nabla v\|^2_{L^2(\Omega)} - (u,v)_{L^2(\Omega)}.$$

Replacing the constraint $v \geq \Psi$ by a penalty term with $\gamma > 0$ results in

$$\min_{v\in H_0^1(\Omega)} \frac{1}{2}\|\nabla v\|^2_{L^2(\Omega)} - (u,v)_{L^2(\Omega)} + \frac{1}{2\gamma}\|(v-\Psi)^-\|^2_{L^2(\Omega)}.$$

The optimality condition for the optimizer y_γ of this then reads

$$(\nabla y_\gamma, \nabla v)_{L^2(\Omega)} - (u,v)_{L^2(\Omega)} + \frac{1}{\gamma}((y_\gamma - \Psi)^-, v)_{L^2(\Omega)} = 0 \qquad v \in H_0^1(\Omega).$$

This is the elliptic analogue of (PDE$_\gamma$) for a specific β_γ.

2.3.2 Existence and Regularity of Solutions to Regularized Obstacle Problems

We restate Theorem 8.17 for this situation.

Theorem 2.22 *There exists an unique weak solution $y_\gamma \in W(I)$ of*

$$\begin{cases} \partial_t y_\gamma + Ay_\gamma + f(y_\gamma) + \beta_\gamma(y_\gamma - \Psi) = u, \\ y_\gamma(0) = y_0, \quad y_\gamma|_\Sigma = 0. \end{cases}$$

Even stronger there are $\kappa^*, C_{lip} > 0$, depending only on an upper bound on the Lipschitz constants of f and β_γ on a ball with radius greater or equal to $\|y_0\|_{L^\infty(\Omega)}$, such that if $\kappa_\Omega \in [0,\kappa^*)$ and $\kappa_I \in (0,1)$ with

$$\frac{1}{q_u}\left(1+\frac{N}{2}\right) + \frac{\kappa_\Omega}{2} < 1 \text{ and } \kappa_I \in \left(0, 1 - \frac{1}{q_u}\left(1+\frac{N}{2}\right) - \frac{\kappa_\Omega}{2}\right)$$

we have

$$\begin{aligned} &\|y\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|\partial_t y\|_{L^{q_u}(Q)} + \|y\|_{L^2(I,V)} + \|Ay\|_{L^{q_u}(Q)} + \|f(y) + \beta_\gamma(y-\Psi)\|_{L^{q_u}(Q)} \\ &\quad \leq C_{lip}\left(\|u - f(0) - \beta_\gamma(-\Psi)\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}}\right). \end{aligned}$$

This regularity implies in particular that we have

$$\partial_t y_\gamma + Ay_\gamma + f(y_\gamma) + \beta_\gamma(y_\gamma - \Psi) = u \text{ a.e. in } Q.$$

We now have existence of solutions, but under certain circumstances we are able to prove that certain norms of y are actually independent of γ, allowing us to later consider the limit $\gamma \to 0$, which in turn allows rigorous analysis of (VI-OB).

Definition 2.23 Let κ_I, κ_Ω be as in Theorem 2.22. We define the solution operator

$$\begin{aligned} S_\gamma\colon L^{q_u}(Q) &\to W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)), \\ u &\mapsto S_\gamma(u) := y \text{ solution of } (\text{PDE}_\gamma). \end{aligned}$$

This solution operator is Lipschitz continuous in a relatively weak norm, but independently of γ.

Lemma 2.24 *The operator S_γ is Lipschitz continuous in the following senses: for any $u_1, u_2 \in L^{q_u}(Q)$ we have*

$$\|S_\gamma(u_1) - S_\gamma(u_2)\|_{L^2(I,V)\cap C(\bar{I},H)} \leq C\|u_1 - u_2\|_{L^2(I,V^*)}$$

and

$$\|S_\gamma(u_1) - S_\gamma(u_2)\|_{L^\infty(Q)} \leq C\|u_1 - u_2\|_{L^{q_u}(Q)}.$$

The constant $C > 0$ does not depend on u_1, u_2, β_γ or f.

Proof. To prove the first claim let $u_1, u_2 \in L^{q_u}(Q)$ and $y_i := S_\gamma(u_i)$ for $i = 1, 2$. Abbreviating $\delta u := u_1 - u_2$ and $\delta y := y_1 - y_2$ we find by the high regularity that

$$\begin{aligned} &(\partial_t \delta y + A\delta y + f(y_1) - f(y_2) + \beta_\gamma(y_1 - \Psi) - \beta_\gamma(y_2 - \Psi), \delta y)_{L^2((0,t)\times\Omega)} \\ &\quad = (\delta u, \delta y)_{L^2((0,t)\times\Omega)}. \end{aligned}$$

Here we tested with $\delta y \cdot 1_{(0,t)}$ for $t \in I$. By the monotonicity of f and β_γ and partial integration we find

$$\frac{1}{2}\|\delta y\|_H^2\big|_0^t + \int_0^t a_\Omega(\delta y, \delta y)\, dt \leq \|\delta u\|_{L^2((0,t),V^*)}\|\delta y\|_{L^2((0,t),V)}.$$

By the uniform ellipticity of the matrix $(a_{ij})_{i,j=1,\dots,N}$ with constant $\nu_{ell} > 0$ this implies

$$\frac{1}{2}\|\delta y(t)\|_H^2 + \nu_{ell}\|\nabla y\|^2_{L^2((0,t)\times\Omega))} \leq \|\delta u\|_{L^2(I,V^*)}\|\delta y\|_{L^2(I,V)}.$$

So, as t was arbitrary,

$$\sup_{t\in I}\|\delta y(t)\|_H^2 \leq 2\|\delta u\|_{L^2(I,V^*)}\|\delta y\|_{L^2(I,V)}. \tag{2.9}$$

We can finally conclude

$$\|\delta y\|^2_{L^2(I,V)} \leq \nu_{ell}^{-1}\|\delta u\|_{L^2(I,V^*)}\|\delta y\|_{L^2(I,V)} + 2T\|\delta u\|_{L^2(I,V^*)}\|\delta y\|_{L^2(I,V)}.$$

This implies

$$\|\delta y\|_{L^2(I,V)} \leq \max(\nu_{ell}^{-1}, 2T)\|\delta u\|_{L^2(I,V^*)}.$$

The $L^\infty(I,H)$-estimate now follows immediately from (2.9).

The second part is just an application of Theorem 8.22. □

Lemma 2.25 *Let $u \in L^{q_u}(Q)$ and $y_\gamma = S_\gamma(u)$. The sequence $\left(\|\beta_\gamma(y_\gamma - \Psi)\|_{L^{q_u}(Q)}\right)_{\gamma>0}$ is bounded independently of γ. In particular*

$$\|\beta_\gamma(y_\gamma - \Psi)\|_{L^{q_u}(Q)} \le \|u\|_{L^{q_u}(Q)} + \|\partial_t\Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}.$$

Proof. Let $u \in L^{q_u}(Q)$. We write $y = S_\gamma(u)$. We dropped the index from the state for presentation in the proof. By Theorem 2.22 we find

$$\begin{aligned} \|\beta_\gamma(y - \Psi)\|^{q_u}_{L^{q_u}(Q)} &= \int_Q |\beta_\gamma(y - \Psi)|^{q_u-2}\beta_\gamma(y - \Psi)\beta_\gamma(y - \Psi)\, d(t, x) \\ &= \int_Q b(y - \Psi)\,(u - \partial_t y - Ay - f(y))\; d(t, x), \end{aligned} \tag{2.10}$$

where we introduced $b(r) := |\beta_\gamma(r)|^{q_u-2}\beta_\gamma(r) \le 0$. We further define B as the antiderivative of b with $B(0) = 0$. It satisfies $B|_{[0,\infty)} = 0$ and $B \ge 0$.

The first term is easily estimated

$$\begin{aligned} &\int_Q b(y - \Psi)u\, d(t, x) \le \|b(y - \Psi)\|_{L^{\frac{q_u}{q_u-1}}(Q)}\|u\|_{L^{q_u}(Q)} \\ &= \left[\int_Q |\beta_\gamma(y - \Psi)|^{q_u}\, d(t, x)\right]^{\frac{q_u-1}{q_u}} \|u\|_{L^{q_u}(Q)} = \|\beta_\gamma(y - \Psi)\|^{q_u-1}_{L^{q_u}(Q)}\|u\|_{L^{q_u}(Q)}. \end{aligned} \tag{2.11}$$

The next term we estimate as

$$\begin{aligned} -\int_Q b(y - \Psi)\partial_t y\, d(t, x) &= -\int_Q b(y - \Psi)\partial_t(y - \Psi)\, d(t, x) - \int_Q b(y - \Psi)\partial_t\Psi\, d(t, x) \\ &= -\int_\Omega\int_I \partial_t\,(B(y - \Psi))\; dt\, dx - \int_Q b(y - \Psi)\partial_t\Psi\, d(t, x) \\ &\le -\int_\Omega B(y(T) - \Psi(T)) - B(y_0 - \Psi(0))\, dx + \|\beta_\gamma(y - \Psi)\|^{q_u-1}_{L^{q_u}(Q)}\|\partial_t\Psi\|_{L^{q_u}(Q)}. \end{aligned}$$

Here we used the chain rule for weak derivatives . Because $y_0 \ge \Psi(0)$ a.e. in Ω we have $B(y_0 - \Psi(0)) = 0$ a.e. in Ω and in general $B(y(T) - \Psi(T)) \ge 0$ a.e. in Ω by the construction of B. Therefore we find

$$-\int_Q b(y - \Psi)\partial_t y\, d(t, x) \le \|\beta_\gamma(y - \Psi)\|^{q_u-1}_{L^{q_u}(Q)}\|\partial_t\Psi\|_{L^{q_u}(Q)}. \tag{2.12}$$

The third term is estimated similarly. Note that $y - \Psi|_{\Sigma_D} = -\Psi|_{\Sigma_D} \ge 0$ so that by the definition of b we have $b(y - \Psi) \in V$ by the trace characterization of V and therefore

$$\begin{aligned} -\int_Q b(y - \Psi)Ay\, d(t, x) &= -\int_Q \sum_{i,j=1}^N a_{ij}b'(y - \Psi)\partial_{x_i}(y - \Psi)\partial_{x_j}y\, d(t, x) \\ &= -\int_Q \sum_{i,j=1}^N a_{ij}b'(y - \Psi)\partial_{x_i}(y - \Psi)\partial_{x_j}(y - \Psi)\, d(t, x) \\ &\quad -\int_Q \sum_{i,j=1}^N a_{ij}b'(y - \Psi)\partial_{x_i}(y - \Psi)\partial_{x_j}\Psi\, d(t, x). \end{aligned}$$

By the ellipticity of $(a_{ij})_{i,j=1,\ldots,N}$ and the monotonicity of b this is bounded from above by

$$-\int_Q b(y - \Psi)A\Psi\, d(t, x) \le \|\beta_\gamma(y - \Psi)\|^{q_u-1}_{L^{q_u}(Q)}\|A\Psi\|_{L^{q_u}(Q)}, \tag{2.13}$$

To see the monotonicity of b we take its weak derivative for $q_u > 2$. For $q_u = 2$ it is trivially seen.

$$\begin{aligned} b'(r) &= (q_u - 2)|\beta_\gamma(r)|^{q_u-3}\operatorname{sgn}(\beta_\gamma(r))\beta_\gamma{}'(r)\beta_\gamma(r) + |\beta_\gamma(r)|^{q_u-2}\beta_\gamma{}'(r) \\ &= (q_u - 1)|\beta_\gamma(r)|^{q_u-2}\beta_\gamma{}'(r) \geq 0, \end{aligned}$$

since β_γ is monotonely increasing.

The fourth term makes use of the monotonicity of f. The following is always to be understood in an almost everywhere sense. When we have $y \geq \Psi$ we find $b(y - \psi) = 0$ and when we have $y < \Psi$ we find $f(y) \leq f(\Psi)$. We also have $-b(y - \Psi) \geq 0$. So in total

$$-\int_Q b(y - \Psi)f(y)\, d(t,x) \leq \int_Q -b(y - \Psi)f(\Psi)\, d(t,x) \leq \|\beta_\gamma(y - \Psi)\|_{L^{q_u}(Q)}^{q_u-1}\|f(\Psi)\|_{L^{q_u}(Q)}. \tag{2.14}$$

Inserting (2.11)-(2.14) into (2.10) yields the claim. □

Lemma 2.26 *Let $u \in L^{q_u}(Q)$ and $y_\gamma = S_\gamma(u)$. The sequence $\left(\|f(y_\gamma)\|_{L^{q_u}(Q)}\right)_{\gamma>0}$ is then bounded independently of γ. In particular:*

$$\begin{aligned} \|f(y_\gamma)\|_{L^{q_u}(Q)} \leq C_{y_0}^{\frac{q_u-1}{q_u}}\|y_0\|_{L^\infty(\Omega)} + 2\|u\|_{L^{q_u}(Q)} + 2\|f(0)\|_{L^{q_u}(Q)} \\ + \|\partial_t\Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}. \end{aligned}$$

Here C_{y_0} is the Lipschitz constant of f on $L^\infty(Q)$ with respect to the ball with radius $\|y_0\|_{L^\infty(Q)}$.

Proof. We first note that by assumption $f(0) \in L^{q_u}(Q)$. Thus in (PDE$_\gamma$) we can subtract $f(0)$ from both sides and substitute u by $\tilde{u} = u - f(0)$. We thus assume $f(0) = 0$.

We write $y := S_\gamma(u)$ and again drop the index from the notation. The proof is very similar to the proof of Lemma 2.25. Introducing $\hat{f}(y) = |f(y)|^{q_u-2}f(y)$ we again have by Theorem 2.22

$$\|f(y)\|_{L^{q_u}(Q)}^{q_u} = \int_Q \hat{f}(y)\left(\tilde{u} - \partial_t y - Ay - \beta_\gamma(y - \psi)\right) d(t,x).$$

As in the proof of Lemma 2.25 we have

$$\int_Q \hat{f}(y)\tilde{u} \leq \|f(y)\|_{L^{q_u}(Q)}^{q_u-1}\|\tilde{u}\|_{L^{q_u}}. \tag{2.15}$$

We introduce $F(t,x,\cdot)$ as an antiderivative to $\hat{f}(t,x,\cdot)$ satisfying $F(t,x,0) = 0$ a.e. in Q. Because $f(t,x,\cdot)$ is monotonically increasing and $f(t,x,0) = 0$ we have $F(t,x,\cdot) \geq 0$ a.e. in Q. Using F we have

$$\begin{aligned} \int_Q -\partial_t y\hat{f}(y)\, d(t,x) &= -\int_I \partial_t\left(\int_\Omega F(y)\, dx\right) dt = -\int_\Omega F(y(T))\, dx + \int_\Omega F(y_0)\, dx \\ &\leq \int_\Omega F(y_0)\, dx. \end{aligned} \tag{2.16}$$

For almost every $(t,x) \in Q$ we have, by construction and $f(t,x,0) = 0$,

$$F(y_0) = \int_0^{y_0} |f(r)|^{q_u-2}f(r)\ dr \leq C_{y_0}^{q_u-1}\|y_0\|_{L^\infty(\Omega)}^{q_u-1}\int_0^{y_0}\ dr \leq C_{y_0}^{q_u-1}\|y_0\|_{L^\infty(\Omega)}^{q_u}.$$

Here $C_{L,\|y_0\|_{L^\infty(\Omega)}}$ is the local Lipschitz constant of f on the ball with radius $\|y_0\|_{L^\infty(\Omega)}$.

Note that because f is monotonically increasing in y, so is $\hat{f}$. We then have by the ellipticity of $(a_{ij})_{i,j=1,\dots,N}$

$$-\int_Q \hat{f}(y) Ay \, d(t,x) = -\sum_{i,j=1}^{N} \int_Q \hat{f}'(y) \, \partial_{x_i} y \, a_{ij} \, \partial_{x_j} y \leq 0. \tag{2.17}$$

The last term is estimated using Lemma 2.25:

$$\begin{aligned} -\int_Q \beta_\gamma(y-\Psi)\hat{f}(y)\, d(t,x) &\leq \|\beta_\gamma(y-\Psi)\|_{L^{q_u}(Q)} \|\hat{f}\|_{L^{\frac{q_u}{q_u-1}}(Q)} \\ &= \|\beta_\gamma(y-\Psi)\|_{L^{q_u}(Q)} \left(\int_Q \left(|f(y)|^{q_u-1} \right)^{\frac{q_u}{q_u-1}} \, d(t,x) \right)^{\frac{q_u-1}{q_u}} \\ &= \|\beta_\gamma(y-\Psi)\|_{L^{q_u}(Q)} \|f(y)\|^{q_u-1}_{L^{q_u}(Q)} \\ &\leq \left(\|\tilde{u}\|_{L^{q_u}(Q)} + \|\partial_t \Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)} \right) \|f(y)\|^{q_u-1}_{L^{q_u}(Q)}. \end{aligned} \tag{2.18}$$

Putting (2.15)-(2.18) together yields

$$\begin{aligned} \|f(y)\|^{q_u}_{L^{q_u}(Q)} &\leq C_{y_0}^{q_u-1} \|y_0\|^{q_u}_{L^\infty(Q)} \\ &\quad + \left(2\|\tilde{u}\|_{L^{q_u}(Q)} + \|\partial_t \Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)} \right) \|f(y)\|^{q_u-1}_{L^{q_u}(Q)}. \end{aligned}$$

Now Lemma 8.2 yields

$$\|f(y)\|_{L^{q_u}(Q)} \leq C_{y_0}^{\frac{q_u-1}{q_u}} \|y_0\|_{L^\infty(Q)} + 2\|\tilde{u}\|_{L^{q_u}(Q)} + \|\partial_t \Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}.$$

Recalling $\tilde{u} = u - f(0)$ concludes the proof. □

Corollary 2.27 *There exist $\kappa^*, C > 0$, independent of γ, but depending on the Lipschitz constant of $f(t,x,\cdot)$ with $(t,x) \in Q$ on the ball with radius $\|y_0\|_{L^\infty(\Omega)}$, such that the following holds: for $u \in L^{q_u}(Q)$ and $y_\gamma := S_\gamma(u)$ we have*

$$\begin{aligned} &\|y_\gamma\|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} + \|\partial_t y_\gamma\|_{L^{q_u}(Q)} + \|y_\gamma\|_{L^2(I,V)} + \|Ay_\gamma\|_{L^{q_u}(Q)} \\ &\quad \leq C \Big(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^{q_u}(Q)} \\ &\qquad + \|\partial_t \Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)} \Big) \end{aligned}$$

where $\kappa_\Omega \in [0, \kappa^*)$ and $\kappa_I \in (0,1)$ with

$$\frac{1}{q_u}\left(1 + \frac{N}{2}\right) + \frac{\kappa_\Omega}{2} < 1 \text{ and } \kappa_I \in \left(0, 1 - \frac{1}{q_u}\left(1 + \frac{N}{2}\right) - \frac{\kappa_\Omega}{2}\right).$$

Proof. We can rewrite (PDE_γ) to

$$\begin{cases} \partial_t y_\gamma + Ay_\gamma = u - f(y_\gamma) - \beta_\gamma(y_\gamma - \Psi) \text{ on } Q, \\ y_\gamma(0) = y_0, \quad y_\gamma|_{\Sigma_D} = 0. \end{cases}$$

By Theorem 8.20, Lemma 2.25 and Lemma 2.26 we obtain the desired estimate. □

We obtain the following continuity result.

Corollary 2.28 *Let $(u_n)_{n\in\mathbb{N}}$ be a sequence in $L^{q_u}(Q)$ converging weakly to some $u \in L^{q_u}(Q)$. Then we have $S_\gamma(u_n) \xrightarrow{n\to\infty} S_\gamma(u)$ uniformly.*

Proof. As weakly convergent sequences are bounded we have by Corollary 2.27 that

$$\|S_\gamma(u_n)\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|S_\gamma(u_n)\|_{W(I)} \leq C_{\not\gamma}$$

for some Hölder exponents satisfying the conditions from Corollary 2.27. By Lemma 8.8 and the reflexivity of the Hilbert space $W(I)$ we have that a subsequence of $(S_\gamma(u_n))_{n\in\mathbb{N}}$ converges uniformly and weakly in $W(I)$ to some $y \in W(I) \cap C(\bar{Q})$. Let $v \in L^2(I,V)$. Taking the limit in the weak formulation in (2.8) immediately shows that $y = S_\gamma(u)$.

To see that the whole sequence converges we make a simple calculus argument: assume that $(S_\gamma(u_n))_{n\in\mathbb{N}}$ does not converge to y uniformly. Then we have that a subsequence $(S_\gamma(u_{n_k}))_{k\in\mathbb{N}}$ satisfies $\|S_\gamma(u_{n_k}) - y\|_{L^\infty(Q)} \geq \epsilon$ for some $\epsilon > 0$. However, by the previous arguments a subsequence of $(S_\gamma(u_{n_k}))_{k\in\mathbb{N}}$ does now converge to $y = S_\gamma(u)$: a contradiction. □

2.4 Convergence of Solutions to Regularized Obstacle Problems

This section is dedicated to transfer regularity results from the solutions to the regularized problems (PDE$_\gamma$) to the solution to the unregularized problem (VI-OB). This is relatively straightforward, but not necessarily the only approach to obtain regularity results on (VI-OB). A different approach based on cutting off certain terms, which is standard in parabolic theory, is for example used, without proof, in [Dom82, Section 4].

Lemma 2.29 *Let $\kappa_\Omega, \kappa_I \in (0,1)$ satisfy the conditions from Theorem 2.22. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a zero sequence. There exist a $y \in W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$, with $\partial_t y \in L^{q_u}(Q)$ and $Ay \in L^{q_u}(Q)$, and a subsequence $(\gamma_{n_k})_{k\in\mathbb{N}}$ such that we have:*

$$S_{\gamma_{n_k}}(u) \xrightarrow{k\to\infty} y \text{ strongly in } C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \text{ and weakly in } W(I).$$

There also exists a $\lambda(u) \in L^{q_u}(Q)$ such that

$$-\beta_\gamma(S_\gamma(u) - \Psi) \xrightarrow{\gamma\to 0} \lambda(u) \text{ weakly in } L^{q_u}(Q).$$

Proof. It is possible to find some $\kappa'_I \in (\kappa_I, 1)$, $\kappa'_\Omega \in (\kappa_\Omega, 1)$ such that $\kappa'_I, \kappa'_\Omega$ still satisfy the assumptions from Corollary 2.27.

By Corollary 2.27 we have

$$\|y_\gamma\|_{C^{\kappa'_I}(I,C^{\kappa'_\Omega}(\Omega))\cap W(I)} \leq C_{\not\gamma}.$$

By the compact embedding of Hölder spaces into Hölder spaces of lower order, see Lemma 8.8, there exists a subsequence of $(\gamma_n)_{n\in\mathbb{N}}$, which we again denote by $(\gamma_n)_{n\in\mathbb{N}}$, and a $y \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ such that

$$y_{\gamma_n} \xrightarrow{n\to\infty} y \text{ strongly in } C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)).$$

By the reflexivity of the Hilbert space $W(I)$ there exists a subsequence of $(\gamma_n)_{n\in\mathbb{N}}$, which we also denote by $(\gamma_n)_{n\in\mathbb{N}}$, and a $\tilde{y} \in W(I)$ such that

$$y_{\gamma_n} \xrightarrow{n\to\infty} \tilde{y} \text{ weakly in } W(I).$$

We now show $\tilde{y} = y$. By the convergence in $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ we have

$$\|\tilde{y} - y\|^2_{L^2(Q)} = \lim_{n\to\infty} (\tilde{y} - y_n, \tilde{y} - y)_{L^2(Q)} .$$

Since $(\cdot, \tilde{y} - y)_{L^2(Q)}$ is a continuous linear functional on $W(I)$ we have by the weak convergence in $W(I)$

$$\|\tilde{y} - y\|^2_{L^2(Q)} = \lim_{n\to\infty} (\tilde{y} - y_n, \tilde{y} - y)_{L^2(Q)} = 0.$$

By Lemma 2.25 we have $\|\beta_{\gamma_n}(y_{\gamma_n} - \Psi)\|_{L^{q_u}(Q)} \leq C_{\gamma}$. This implies the weak convergence (of some subsequence) to some limit $\lambda(u) \in L^{q_u}(Q)$ by the reflexivity of $L^{q_u}(Q)$. □

Theorem 2.30 *Let $C, \kappa_\Omega, \kappa_I \in (0,1)$ be as in Corollary 2.27; in particular C is independent of γ. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a zero sequence. The y from Lemma 2.29 is the unique solution to* (VI-OB) *and the whole sequence of states $(S_{\gamma_n}(u))_{n\in\mathbb{N}}$ converges to it. Its multiplier $\lambda(u)$ from Lemma 2.29 is also unique. We also have the estimates*

$$\begin{aligned}&\|y\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|y\|_{W(I)} \\ &\quad \leq C \Big(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^{q_u}(Q)} \\ &\qquad + \|\partial_t \Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)} \Big)\end{aligned}$$

and

$$\|\lambda(u)\|_{L^{q_u}(Q)} \leq \|u\|_{L^{q_u}(Q)} + \|\partial_t \Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}.$$

The state y and the corresponding $\lambda(u)$ satisfy:

$$\begin{cases} (\partial_t y + Ay + f(y), v)_{L^2(I,V^*,V)} = (\lambda + u, v)_{L^2(Q)} & \forall v \in L^2(Q), \\ y(0) = y_0, \quad y|_{\Sigma_D} = 0, \\ \lambda(u) \geq 0, \quad y - \Psi \geq 0, \quad (\lambda(u), y - \Psi)_{L^2(Q)} = 0. \end{cases}$$

Remark 2.31 In the situation of Theorem 2.30 it is easy to see that $(\lambda(u), y - \Psi)_{L^2(Q)} = 0$ implies $\operatorname{supp}(\lambda(u)) \subset \{y = \Psi\} \cup Z$, where Z is a set of Lebesgue measure 0. Since $\lambda(u) \geq 0$ and $y - \Psi \geq 0$ a.e. in Q we also have that their product $\lambda(u)(y - \Psi) \geq 0$ a.e. in Q. Thus the complementarity is equivalent to $0 = \|\lambda(u)(y - \Psi)\|_{L^1(Q)}$ and therefore $\lambda(u)(y - \Psi) = 0$ a.e. in Q. That means only on a set of measure 0 can we expect $\lambda(u) > 0$ and $y - \Psi > 0$ both at once.

Proof. We will show that y is the unique solution to (VI-OB). Then a standard calculus argument shows that if we have for each subsequence a subsubsequence that converges to y, that y is the limit of the whole sequence. So, we may assume that $(S_{\gamma_n}(u))_{n\in\mathbb{N}}$ converges to y in the sense of Lemma 2.29. (See for example the argument at the of the proof of Corollary 2.28.)

By Lemma 2.29 and its uniform convergence in $C(\bar{Q})$ we obviously have $y(0) = y_0$. By the weak convergence in $W(I)$ we have $y|_{\Sigma_D} = 0$.

Assume there is a set $O \subset Q$ with $|O| > 0$ and an $\epsilon > 0$ such that $y \leq \Psi - \epsilon$. Now by the uniform convergence from Lemma 2.29 there is an $n_\epsilon \in \mathbb{N}$ such that $S_{\gamma_n}(u) \leq \Psi - \frac{\epsilon}{2}$ on O for any $n \geq n_\epsilon$. Then we have

$$\beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi) \leq \beta_{\gamma_n}(-\epsilon/2) = \frac{1}{\gamma_n}\beta(-\epsilon/2)$$

on O for $n \geq n_\epsilon$. Then for $n \geq n_\epsilon$ we have

$$\int_Q \beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi)\, d(t,x) \leq \int_O \beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi)\, d(t,x) \leq \frac{1}{\gamma_n}|O|\beta(-\epsilon/2) \xrightarrow{\gamma_n \to 0} -\infty.$$

This contradicts Lemma 2.25. Thus for any set $O \subset Q$ with non-zero measure and any $\epsilon > 0$ we have

$$y > \Psi - \epsilon \text{ on } O.$$

Thus $y \geq \Psi$.

We can consider the weak formulation of (PDE$_\gamma$) and test it with $v - S_{\gamma_n}(u)$ for some $v \in K_\Psi \cap L^2(I, V)$. Then we have

$$\begin{aligned}&(\partial_t S_{\gamma_n}(u) + AS_{\gamma_n}(u) + f(S_{\gamma_n}(u)), v - S_{\gamma_n}(u))_{L^2(Q)} \\ &\quad = (-\beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi), v - S_{\gamma_n}(u))_{L^2(Q)} + (u, v - S_{\gamma_n}(u))_{L^2(Q)}.\end{aligned} \tag{2.19}$$

We have

$$\begin{aligned}(-\beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi), v - S_{\gamma_n}(u))_{L^2(Q)} &= (\beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi), S_{\gamma_n}(u) - \Psi + \Psi - v)_{L^2(Q)} \\ &\geq (\beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi), \Psi - v)_{L^2(Q)}.\end{aligned}$$

Here we used the monotonicity of β_{γ_n}. Now (2.19) implies

$$(\partial_t S_{\gamma_n}(u) + AS_{\gamma_n}(u) + f(S_{\gamma_n}(u)), v - S_{\gamma_n}(u))_{L^2(Q)} \geq (u, v - S_{\gamma_n}(u))_{L^2(Q)}.$$

Now we take the limit. The convergence to (VI-OB) is clear by Lemma 2.29 and the fact that f is continuous from $L^\infty(Q)$ to $L^\infty(Q)$ by Corollary 2.14. This shows that y indeed solves (VI-OB).

To see uniqueness of the solutions let $y_1, y_2 \in W(I) \cap K_\Psi$ be solutions to (VI-OB). Then we can test the formulation for $y_1 1_{(0,\hat{t})}$ with $y_2 1_{(0,\hat{t})}$ for some $\hat{t} \in (0,T)$ and vice versa to obtain:

$$\begin{aligned}(\partial_t y_1 + Ay_1 + f(y_1) - u, y_2 - y_1)_{L^2((0,\hat{t})\times\Omega)} \geq 0, \\ (\partial_t y_2 + Ay_2 + f(y_2) - u, y_1 - y_2)_{L^2((0,\hat{t})\times\Omega)} \geq 0.\end{aligned}$$

Adding those two lines yields

$$(\partial_t(y_1 - y_2) + A(y_1 - y_2) + f(y_1) - f(y_2), y_2 - y_1)_{L^2((0,\hat{t})\times\Omega)} \geq 0.$$

By partial integration, monotonicity of f and the ellipticity of the coefficients of A we conclude

$$0 \geq \frac{1}{2}\|y_1(\hat{t}) - y_2(\hat{t})\|_H^2 + \nu_{ell}\|\nabla y_1 - \nabla y_2\|_H^2.$$

Thus in particular for any $\hat{t} \in (0,T)$ we have $y_1(\hat{t}) = y_2(\hat{t})$ almost everywhere in Ω and thus $y_1 = y_2$.

The norm estimates are just a consequence of taking the limit in Lemma 2.25 and Corollary 2.27.

To obtain the complementarity system we again consider the weak formulation of (PDE$_\gamma$) for u and $S_{\gamma_n}(u)$ and take the limit using Lemma 2.29. We thus have for any $v \in L^2(I, V)$

$$(\partial_t y + Ay + f(y), v)_{L^2(Q)} = (\lambda(u) + u, v)_{L^2(Q)} .$$

Using the density of $L^2(I, V)$ in $L^2(Q)$ we find the multiplier formulation. The density can bee seen by $C_c^\infty(Q) \subset L^2(I, V)$ and [Rud74, Theorem 3.14].

Because $\{v \in L^{q_u}(Q) : v \geq 0\}$ is closed and convex we have $\lambda(u) \geq 0$, because it is the weak limit of $-\beta_{\gamma_n}(S_{\gamma_n}(u) - \psi) \geq 0$.

The complementarity is also easily checked: We already know $y \geq \Psi$ and $\lambda(u) \geq 0$. Thus

$$(\lambda(u), y - \Psi)_{L^2(Q)} \geq 0.$$

But we also know, by the previous convergences and the monotonicity of β_γ,

$$(\lambda(u), y - \Psi)_{L^2(Q)} = \lim_{n\to\infty} (-\beta_{\gamma_n}(S_{\gamma_n}(u) - \Psi), S_{\gamma_n}(u) - \Psi)_{L^2(Q)} \leq 0.$$

The complementarity system also entails that

$$\lambda(u) = \partial_t y + Ay + f(y) - u$$

is unique. □

Definition 2.32 Based on Lemma 2.29 and Theorem 2.30 we define the solution operator to (VI-OB):

$$\begin{aligned} S\colon L^{q_u}(Q) &\to W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)), \\ u &\mapsto \lim_{\gamma\to 0} S_\gamma(u). \end{aligned}$$

Here $\kappa_I, \kappa_\Omega \in (0,1)$ are subject to the conditions from Corollary 2.27 and the convergence is to be understood in the sense of Lemma 2.29. We also define

$$\lambda_\gamma(u) := -\beta_\gamma(S_\gamma(u) - \Psi)$$

for $\gamma > 0$.

The following theorem states that the assumptions on the multiplier λ are not only necessary, but also sufficient for y to be a solution to (VI-OB).

Theorem 2.33 *Assume there is a $y \in W(I)$ and a $\lambda \in L^{q_u}(Q)$ satisfying*

$$\begin{cases} (\partial_t y + Ay + f(y), v)_{L^2(Q)} = (\lambda + u, v)_{L^2(Q)} & \forall v \in L^2(Q), \\ y(0) = y_0, \\ \lambda \geq 0, \quad y - \Psi \geq 0, \quad (\lambda, y - \Psi)_{L^2(Q)} = 0. \end{cases}$$

Then we have $y = S(u)$ and $\lambda = \lambda(u)$.

Proof. Let $v \in K_\Psi$. We then have by the complementarity and $\lambda \geq 0$

$$\begin{aligned} (\partial_t y + Ay + f(y), v - y)_{L^2(Q)} = (\lambda + u, v - y)_{L^2(Q)} = (\lambda, v - \Psi)_{L^2(Q)} + (u, v - y)_{L^2(Q)} \\ \geq (u, v - y)_{L^2(Q)} . \end{aligned}$$

□

We can now extend the previous results and show continuity of the mapping $(\gamma, u) \mapsto S_\gamma(u)$, in a certain sense.

Theorem 2.34 *Let $(u_n)_{n\in\mathbb{N}} \subset L^{q_u}(Q)$ be a weakly convergent series, converging to $u \in L^{q_u}(Q)$, and $(\gamma_n)_{n\in\mathbb{N}}$ a zero sequence, with $\gamma_n \geq 0$. We then have for any $\kappa_I, \kappa_\Omega \in (0,1)$ satisfying the conditions from Corollary 2.27*

- $S_{\gamma_n}(u_n) \xrightarrow{n\to\infty} S(u)$ *strongly in* $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ *and weakly in* $W(I)$,
- $\lambda_{\gamma_n}(u_n) \xrightarrow{n\to\infty} \lambda(u)$ *weakly in* $L^{q_u}(Q)$.

Proof. We abbreviate $y_n := S_{\gamma_n}(u_n)$. Because $(u_n)_{n\in\mathbb{N}}$ is uniformly bounded in $L^{q_u}(Q)$ we have that

$$\|y_n\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))\cap W(I)}, \|\lambda(u_n)\|_{L^{q_u}(Q)} \leq C \text{ independent of } n$$

by Corollary 2.27 (for $\gamma_n > 0$) and Theorem 2.30 (for $\gamma_n = 0$).

As in the proof of Lemma 2.29 there exists a $y \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \cap W(I)$ such that

$$y_n \xrightarrow{n\to\infty} y \text{ strongly in } C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \text{ and weakly in } W(I)$$

and $\lambda \in L^{q_u}(Q)$ such that

$$\lambda_{\gamma_n}(u_n) \xrightarrow{n\to\infty} \lambda \text{ weakly in } L^{q_u}(Q).$$

The rest of the proof is now essentially the same as the one of Theorem 2.30. The weak convergence of $(u_n)_{n\in\mathbb{N}}$ does not pose a problem as it is multiplied with a strongly convergent sequence, yielding convergence of the product. □

From Lemma 2.24 we get an easy corollary due to the uniform convergence of $(S_\gamma(u))_{\gamma>0}$.

Corollary 2.35 *The operator S is Lipschitz continuous in the following senses: there is a $C > 0$ such that for any $u_1, u_2 \in L^{q_u}(Q)$ we have*

$$\|S(u_1) - S(u_2)\|_{L^2(I,V)\cap C(\bar{I},H)} \leq C\|u_1 - u_2\|_{L^2(I,V^*)}$$

and

$$\|S(u_1) - S(u_2)\|_{L^\infty(Q)} \leq C\|u_1 - u_2\|_{L^{q_u}(Q)}.$$

In the $L^\infty(Q)$-norm we can explicitly give a convergence rate for $\gamma \to 0$. The following statements are the extension of [Noc88, Theorem 2.1] to the parabolic case. It will be later useful when deriving finite element error estimates for (PDE$_\gamma$). Note that in contrast to [Noc88, Theorem 2.1] we do not require the operator A to be elliptic.

Proposition 2.36 *Let $u \in L^\infty(Q)$ and Ψ with $\partial_t\Psi, A\Psi \in L^\infty(Q)$. Write $y_\gamma = S_\gamma(u)$. Then we have for any $\gamma > 0$:*

$$\begin{aligned}\|\beta_\gamma(y_\gamma - \Psi)\|_{L^\infty(Q)} &\leq \|u\|_{L^\infty(Q)} + \|\partial_t\Psi\|_{L^\infty(Q)} + \|A\Psi\|_{L^\infty(Q)} + \|f(\Psi)\|_{L^\infty(Q)},\\ \|f(y)\|_{L^\infty(Q)} &\leq C_{y_0}\|y_0\|_{L^\infty(\Omega)} + 2\|u\|_{L^\infty(Q)} + 2\|f(0)\|_{L^\infty(Q)}\\ &\quad + \|f(\Psi)\|_{L^\infty(Q)} + \|\partial_t\Psi\|_{L^\infty(Q)} + \|A\Psi\|_{L^\infty(Q)}.\end{aligned}$$

Here C_{y_0} is the Lipschitz constant of f with respect to its last component on a ball with radius $\|y_0\|_{L^\infty(\Omega)}$.

Proof. By Corollary 2.14 we have $f(\Psi) \in L^\infty(Q)$, so everything on the right hand sides is well-defined.

By Lemma 2.25 and Lemma 2.26 we have for any $p \in (1 + N/2, \infty) \cap [2, \infty)$

$$\begin{aligned}\|\beta_\gamma(y-\Psi)\|_{L^p(Q)} &\leq \|u\|_{L^p(Q)} + \|\partial_t\Psi\|_{L^p(Q)} + \|A\Psi\|_{L^p(Q)} + \|f(\Psi)\|_{L^p(Q)},\\ \|f(y)\|_{L^p(Q)} &\leq C_{y_0}^{\frac{p-1}{p}}\|y_0\|_{L^\infty(\Omega)} + 2\|u\|_{L^p(Q)} + 2\|f(0)\|_{L^p(Q)}\\ &\quad + \|\partial_t\Psi\|_{L^p(Q)} + \|A\Psi\|_{L^p(Q)} + \|f(\Psi)\|_{L^p(Q)}.\end{aligned}$$

We now send p to ∞ and receive the desired result by Lemma 1.17. □

Theorem 2.37 *Assume β has the form given in Proposition 2.17 for some $\alpha \geq 1$. $u \in L^\infty(Q)$ and assume Ψ with $\partial_t\Psi, A\Psi \in L^\infty(Q)$. Further, assume $\gamma \in (0, c^{-1}\alpha^{\alpha/(1-\alpha)})$ (with $\alpha^{\alpha/(1-\alpha)} = e^{-1}$ for $\alpha = 1$) where we define $c := \|u\|_{L^\infty(Q)} + \|\partial_t\Psi\|_{L^\infty(Q)} + \|A\Psi\|_{L^\infty(Q)} + \|f(\Psi)\|_{L^\infty(Q)}$. Then we have*

$$\|S(u) - S_\gamma(u)\|_{L^\infty(Q)} \leq (c\gamma)^{1/\alpha}.$$

Proof. From Proposition 2.36 we have

$$\|\lambda_\gamma(u)\|_{L^\infty(Q)} \leq \|u\|_{L^\infty(Q)} + \|\partial_t\Psi\|_{L^\infty(Q)} + \|A\Psi\|_{L^\infty(Q)} + \|f(\Psi)\|_{L^\infty(Q)}. \tag{2.20}$$

By the same proof one can obtain

$$\|\lambda(u)\|_{L^\infty(Q)} \leq \|u\|_{L^\infty(Q)} + \|\partial_t\Psi\|_{L^\infty(Q)} + \|A\Psi\|_{L^\infty(Q)} + \|f(\Psi)\|_{L^\infty(Q)} \tag{2.21}$$

from Theorem 2.34.

By definition we have for any integer $p \in (1 + N/2, \infty)$ that $S(u), S_\gamma(u) \in W(I) \cap L^\infty(Q)$. Therefore we find

$$(S(u) - S_\gamma(u))^{2p+1} =: e^{2p+1} \in L^2(I, V).$$

Now we can test the equations for $S(u)$ and $S_\gamma(u)$ with $e^{2p+1}1_{(0,\hat{t})}$ for some $\hat{t} \in I$ and take the difference, resulting in

$$\Big(\partial_t e + Ae + (f(S(u)) - f(S_\gamma(u))), e^{2p+1}\Big)_{L^2((0,\hat{t})\times\Omega)} = \Big(\lambda(u) - \lambda_\gamma(u), e^{2p+1}\Big)_{L^2((0,\hat{t})\times\Omega)}. \tag{2.22}$$

We estimate this term by term. We start with the A-term:

$$\begin{aligned}(Ae, e^{2p+1})_{L^2((0,\hat{t})\times\Omega)} &= \sum_{i,j=1}^N \int_Q a_{ij}\,\partial_{x_i}e\,\partial_{x_j}(e^{2p+1})\,d(t,x)\\ &= (2p+1)\sum_{i,j=1}^N \int_{(0,\hat{t})\times\Omega} a_{ij}\,e^{2p}\,\partial_{x_i}e\,\partial_{x_j}e\,d(t,x)\\ &= \frac{2p+1}{(p+1)^2}\sum_{i,j=1}^N \int_{(0,\hat{t})\times\Omega} a_{ij}\,\partial_{x_i}(e^{p+1})\,\partial_{x_j}(e^{p+1})\,d(t,x) \geq 0.\end{aligned} \tag{2.23}$$

Here we used the ellipticity of the matrix $(a_{ij})_{i,j=1,\ldots,N}$ in the last inequality. The terms in (2.22) with f are simply estimated by the monotonicity of f:

$$\Big(f(S(u)) - f(S_\gamma(u)), e^{2p+1}\Big)_{L^2((0,\hat{t})\times\Omega)} \geq 0. \tag{2.24}$$

Now we turn to the term with the time derivative in (2.22) and see

$$\left(\partial_t e, e^{2p+1}\right)_{L^2((0,\hat{t})\times\Omega)} = \frac{1}{2p+2}\int_0^{\hat{t}} \partial_t \left(\int_\Omega e^{2p+2}\, dx\right)\, dt = \frac{1}{2p+2}\|e(\hat{t})\|^{2p+2}_{L^{2p+2}(\Omega)}.$$

This, (2.23) and (2.24) inserted into (2.22) results in

$$\frac{1}{2p+2}\|e(\hat{t})\|^{2p+2}_{L^{2p+2}(\Omega)} \le \left(\lambda(u) - \lambda_\gamma(u), e^{2p+1}\right)_{L^2((0,\hat{t})\times\Omega)}.$$

Thus we have shown that for any $\hat{t} \in I$ there holds

$$\|e(\hat{t})\|_{L^{2p+2}(\Omega)} \le (2p+2)^{\frac{1}{2p+2}} \left(\left(\lambda(u) - \lambda_\gamma(u), e^{2p+1}\right)_{L^2((0,\hat{t})\times\Omega)}\right)^{\frac{1}{2p+2}}. \tag{2.25}$$

The following statements hold true for almost every $(t, x) \in Q$.

- If $S(u)(t, x) > \Psi(t, x)$ and $S_\gamma(u)(t, x) \ge \Psi(t, x)$ we have

  $$\lambda(u)(t, x) = -\beta_\gamma(S_\gamma(u) - \Psi)(t, x) = 0,$$

  which implies

  $$\left[(\lambda(u) + \beta_\gamma(S_\gamma(u) - \Psi))e_\gamma^{2p+1}\right](t, x) = 0.$$

- If $S(u)(t, x) > \Psi(t, x)$ and $S_\gamma(u)(t, x) < \Psi(t, x)$ we have

  $$\lambda(u)(t, x) = 0, e_\gamma(t, x) > 0 \text{ and } \beta_\gamma(S_\gamma(u) - \Psi)(t, x) \le 0.$$

  This in turn implies

  $$\left[(\lambda(u) + \beta_\gamma(S_\gamma(u) - \Psi))e_\gamma^{2p+1}\right](t, x) \le 0.$$

- If $S(u)(t, x) = \Psi(t, x)$ and $S_\gamma(u)(t, x) \ge \Psi$ we have

  $$\lambda(u)(t, x) \ge 0, e_\gamma(t, x) \le 0 \text{ and } \beta_\gamma(S_\gamma(u) - \Psi)(t, x) = 0,$$

  resulting in

  $$\left[(\lambda(u) + \beta_\gamma(S_\gamma(u) - \Psi))e_\gamma^{2p+1}\right](t, x) = (\lambda(u), e_\gamma^{2p+1})(t, x) \le 0.$$

- If $S(u)(t, x) = \Psi(t, x)$ and $S_\gamma(u)(t, x) < \Psi$ we have

  $$\begin{aligned}\left[(\lambda(u) + \beta_\gamma(S_\gamma(u) - \Psi))e_\gamma^{2p+1}\right](t, x) &= \left[(\lambda(u) + \beta_\gamma(S_\gamma(u) - \Psi))(\Psi - S_\gamma(u))^{2p+1}\right](t, x), \\ &\le (\lambda(u)(\Psi - S_\gamma(u))^{2p+1})(t, x).\end{aligned} \tag{2.26}$$

  We claim that $S_\gamma(u)(t, x) - \Psi(t, x) \ge -(c\gamma)^{1/\alpha}$. By assumption $\gamma \in (0, c^{-1}\alpha^{\alpha/(1-\alpha)})$. Assume $S_\gamma(u)(t, x) - \Psi(t, x) := r < -\alpha^{1/(1-\alpha)}$. Then

  $$\beta_\gamma(r) = \frac{1}{\gamma}\left(r + \frac{\alpha - 1}{\alpha}\alpha^{1/(1-\alpha)}\right) \le \frac{1}{\gamma}\frac{-1}{\alpha}\alpha^{1/(1-\alpha)} < -c\alpha^{-\alpha/(1-\alpha)}\frac{1}{\alpha}\alpha^{1/(1-\alpha)} = -c.$$

A contradiction to (2.20). Thus $r \in [-\alpha^{1/(1-\alpha)}, 0]$. In turn this and the form of β_γ imply

$$-c \leq -\frac{1}{\gamma}|r|^\alpha \text{ and thus } \implies c\gamma \geq |r|^\alpha.$$

By the definition of r we thus have

$$(c\gamma)^{1/\alpha} \geq |S_\gamma(u)(t,x) - \Psi(t,x)|.$$

Together with (2.26) we find by Lemma 1.17

$$\left[\lambda(u) - \lambda_\gamma(u), e^{2p+1}\right](t,x) \leq c\left((c\gamma)^{\frac{1}{\alpha}}\right)^{2p+1}$$

Those four cases inserted into (2.25) yields for any $\hat{t} \in I$

$$\|e(\hat{t})\|_{L^{2p+2}(\Omega)} \leq ((2p+2)c)^{\frac{1}{2p+2}}\left((c\gamma)^{\frac{1}{\alpha}}\right)^{\frac{2p+1}{2p+2}}.$$

Sending p to ∞ we find

$$\|e(\hat{t})\|_{L^\infty(\Omega)} \leq (c\gamma)^{\frac{1}{\alpha}}.$$

As e is continuous and $\hat{t}$ was arbitrary this implies the desired estimate. □

2.4.1 Improved Regularity Results for Obstacle Problems on Smooth Domains

In smooth domains we can obtain higher regularity of solutions to (PDE$_\gamma$).

Lemma 2.38 *Assume Ω is a $C^{1,1}$ domain and that $\Gamma_D = \partial\Omega$. There exist $C > 0$ and κ_I, κ_Ω, satisfying the conditions of Corollary 2.27 such that the following holds: for $u \in L^{q_u}(Q)$ and $y_\gamma = S_\gamma(u)$ for $\gamma > 0$ we have*

$$\|y_\gamma\|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} + \|y_\gamma\|_{W^{1,2}_{q_u}(Q)} \leq C\Big(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^{q_u}(Q)} + \|\partial_t\Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}\Big).$$

Proof. This follows from Corollary 2.27 in combination with basic higher elliptic regularity, since $\Gamma_D = \partial\Omega$, see Theorem 8.23. □

Remark 2.39 If we did not use Hölder regularity obtained by the results from [DtER15] in Theorem 8.20 and theorems based on it, Theorem 1.34 would give us Hölder regularity based on the $W^{1,2}_{q_u}(Q)$ regularity.

Now by those improved estimates one can obtain an improved version of Theorem 2.34.

Theorem 2.40 *Assume Ω is a $C^{1,1}$ domain and that $\Gamma_D = \partial\Omega$. There exist $C > 0$ and κ_I, κ_Ω, satisfying the conditions of Corollary 2.27 such that the following holds: let $(u_n)_{n\in\mathbb{N}} \subset L^{q_u}(Q)$ a weakly convergent series, converging to $u \in L^{q_u}(Q)$, and $(\gamma_n)_{n\in\mathbb{N}}$ a zero sequence with $\gamma_n \geq 0$ for all $n \in \mathbb{N}$. We then have*

- $S_{\gamma_n}(u_n) \xrightarrow{n\to\infty} S(u)$ *strongly in* $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ *and weakly in* $W^{1,2}_{q_u}(Q)$,
- $\lambda_{\gamma_n}(u_n) \xrightarrow{n\to\infty} \lambda(u)$ *weakly in* $L^{q_u}(Q)$.

This in particular implies for $y = S(u)$

$$\|y\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|y\|_{W^{1,2}_{q_u}(Q)} \leq C\Big(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^{q_u}(Q)} + \|\partial_t\Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}\Big).$$

Proof. The proof for the convergences is the same as for Theorem 2.34. We just use the reflexivity of $W^{1,2}_{q_u}(Q)$ from Lemma 1.25 instead of the reflexivity of $W(I)$.

The norm estimates follow from Lemma 2.38. □

2.4.2 Improved Regularity Results for Obstacle Problems on Polygonal Domains

As for smooth domains, one can obtain improved regularity results for polygonal domains. The strategies are the same as in Section 2.4.1, just using the elliptic regularity result Corollary 8.27 instead of Theorem 8.23.

Lemma 2.41 *Assume* $A = -\Delta$. *Let* $\Omega \subset \mathbb{R}^2$ *a polygonal domain. That means we can decompose its boundary in* $\Gamma_1, \Gamma_2, \dots, \bar{\Gamma}_M$ *edges. By* $\omega_j \in (0, 2\pi)$ *we denote the angles between* Γ_j *and* Γ_{j+1} *with* $\Gamma_{M+1} = \Gamma_1$. *We assume that the Dirichlet boundary* Γ_D *is a union of edges of* $\partial\Omega$ *and* $\Gamma_D \neq \emptyset$.

We define

$$\Phi_j := \begin{cases} 0 \text{ if } \Gamma_j \not\subset \Gamma_D, \\ \frac{\pi}{2} \text{ if } \Gamma_j \subset \Gamma_D, \end{cases}$$

for $j = 1, \dots, M$ *with* $\Phi_{M+1} := \Phi_1$. *We further define for* $j = 1, 2, \dots, M$

$$\omega_{lim,j} := \begin{cases} \frac{\pi}{2} \text{ if } \Phi_j = \Phi_{j+1}, \\ \frac{\pi}{4} \text{ if } \Phi_j \neq \Phi_{j+1}. \end{cases}$$

For each $j \in \{1, 2, \dots, M\}$ *we assume* $\omega_j \leq \omega_{lim,j}$ *or* $q_u < \frac{\omega_j}{\omega_j - \omega_{lim,j}}$.

There exist $C > 0$ *and* κ_I, κ_Ω, *satisfying the conditions of Corollary 2.27 such that the following holds: for* $u \in L^{q_u}(Q)$ *and* $y_\gamma = S_\gamma(u)$ *for* $\gamma > 0$ *we have*

$$\|y_\gamma\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|y_\gamma\|_{W^{1,2}_{q_u}(Q)} \leq C\Big(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^{q_u}(Q)} + \|\partial_t\Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}\Big).$$

Theorem 2.42 *Assume* A *and* Ω *satisfy the same properties as in Lemma 2.41.*

There exist $C > 0$ *and* κ_I, κ_Ω, *satisfying the conditions of Corollary 2.27 such that the following holds: let* $(u_n)_{n\in\mathbb{N}} \subset L^{q_u}(Q)$ *a weakly convergent series, converging to* $u \in L^{q_u}(Q)$, *and* $(\gamma_n)_{n\in\mathbb{N}}$ *a zero sequence with* $\gamma_n \geq 0$ *for all* $n \in \mathbb{N}$. *We then have*

- $S_{\gamma_n}(u_n) \xrightarrow{n\to\infty} S(u)$ *strongly in* $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ *and weakly in* $W^{1,2}_{q_u}(Q)$,
- $\lambda_{\gamma_n}(u_n) \xrightarrow{n\to\infty} \lambda(u)$ *weakly in* $L^{q_u}(Q)$.

This in particular implies for $y = S(u)$

$$\|y\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|y\|_{W^{1,2}_{q_u}(Q)} \leq C\Big(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^{q_u}(Q)} + \|\partial_t\Psi\|_{L^{q_u}(Q)} + \|A\Psi\|_{L^{q_u}(Q)} + \|f(\Psi)\|_{L^{q_u}(Q)}\Big).$$

3 Optimal Control Problems and their Regularizations

This chapter is dedicated to the derivation of the optimal control problems we shall analyze for the rest of the thesis. We study regularized and unregularized problems, which are related quite closely. We will discuss the problems with respect to the existence of solutions, necessary first order optimality conditions and second order sufficient conditions. At the beginning of each section we give a short overview over the current state of research, our motivations and strategies.

Assumption 3.1 *For the whole of Chapter 3 we obviously assume the definitions and standing assumptions from Section 2.2.1 to hold true. Therefore we may apply most results from Chapter 2 freely. Whenever mentioned, the regularization term β shall satisfy the properties from Definition 2.16.*

3.1 Optimal Control of Obstacle Problems with Control Constraints

Our problem will use a smooth cost functional of a simple structure together with distributed controls. Similar types of problems were for example also studied in [Bar81, Che03, BZ99, Fri87] and [Bar84, Chapter 5]. For the elliptic version of our problem we shall representatively mention [MP84]. We go into more depth on the relevant references at the beginning of the respective sections.

Related problems which also sometimes fall under the umbrella of "Optimal control with parabolic VIs" will be shortly listed here. For example, in [BT11] optimal control problems with parabolic VIs of second kind are analysed, see [BT11, Section 2] for the differences to the obstacle problem. The authors are able to show that their highly non-quadratic problem has a unique solution under certain, strong assumptions. Another related problem is the control of the diffusion coefficients of A studied for example in [IK10]. There a regularization approach, different from ours, is used. Problems that look similar to ours, on first glance at least, but are very different, are problems were the obstacle itself is the control. See for example [BL04, AL02].

Definition 3.2 We consider the optimal control problem

$$\begin{aligned} &\min_{(y,u)\in W(I)\times L^2(Q)} j_v(y) + j_T(y(T)) + g(u) =: J(y,u), \\ &\text{such that } S(u) = y \text{ and } u \in U_{ad}, \end{aligned} \tag{OC}$$

where

- $j_v : L^2(Q) \to \mathbb{R}$ is continuously Fréchet differentiable,
- $j_T : L^2(\Omega) \to \mathbb{R}$ is continuously Fréchet differentiable,
- $g : L^2(Q) \to \mathbb{R}$ is continuously Fréchet differentiable and convex.

The derivatives can be identified with corresponding L^2-functions according to the Riesz representation theorem. We further demand that there exists a constant $c \in \mathbb{R}$ such that $J(y, u) \geq c$ holds for any admissible (y, u).

The admissible set U_{ad} is a convex, bounded and closed subset of $L^{q_u}(Q)$. By these assumptions U_{ad} is also weakly closed. Note that the assumptions on g imply that it is weakly lower semi-continuous, cf. [BP12, Proposition 2.10].

Theorem 3.3 (OC) *has at least one solution* $(\bar{y}, \bar{u})$.

Proof. By assumption we have that J is bounded from below and thus there exists an infimizing sequence $(S(u_n), u_n)_{n\in\mathbb{N}}$ of (OC). By the boundedness of U_{ad} we have that $(\|u_n\|_{L^{q_u}(Q)})_{n\in\mathbb{N}}$ is bounded and thus there exists a weak limit $\bar{u}$ in $L^{q_u}(Q)$ of a subsequence. We denote the subsequence by the same indices. Because U_{ad} is weakly closed we have $\bar{u} \in U_{ad}$. By Lemma 2.29 and Definition 2.32 we do have that $(S(u_n))_{n\in\mathbb{N}}$ converges to $S(\bar{u})$ in the sense given in Lemma 2.29, in particularly uniformly. Thus we have

$$\begin{aligned}\inf_{u\in U_{ad}} J(S(u), u) &= \lim_{n\to\infty} J(S(u_n), u_n) = \liminf_{n\to\infty} \left(j_v(S(u_n)) + j_T(S(u_n)(T)) + g(u_n)\right) \\ &\geq \liminf_{n\to\infty} j_v(S(u_n)) + \liminf_{n\to\infty} j_T(S(u_n)(T)) + \liminf_{n\to\infty} g(u_n).\end{aligned}$$

As $(S(u_n))_{n\in\mathbb{N}}$ converges uniformly and as g is weakly lower semi-continuous we arrive at

$$\inf_{u\in U_{ad}} J(S(u), u) \geq j_v(S(\bar{u})) + j_T(S(\bar{u})(T)) + g(\bar{u}) = J(S(\bar{u}), \bar{u}).$$

□

Remark 3.4 Note that the differentiability assumptions on j_T, j_v and g were not necessary to prove existence of a solution of (OC). It is enough that j_v, j_T and g are weakly lower semi-continuous.

It is also easy to see that solutions to (OC) are not unique in general as S is clearly not injective. Consider $\Omega = (0, 1)$, $I = (0, 1)$ and the constant obstacle $\Psi = 0$. The initial condition is chosen as $y_0 = 0$ and the variational inequality is given by

$$\begin{cases} (\partial_t y - \Delta y, v - y)_{L^2(Q)} \geq (u, v - y)_{L^2(Q)} & \forall v \in L^2(I, V) \\ y|_{I\times\partial\Omega} = 0, \quad y(0) = 0. \end{cases}$$

We have $S(u) = 0$ for any $u \in L^{q_u}(Q)$ that satisfies $u \leq 0$ almost everywhere in Q. This can be easily seen by checking that for any such u the state $y = 0$ satisfies

$$(\partial_t 0 - \Delta 0 - u, v - 0)_{L^2(Q)} = (-u, v)_{L^2(Q)} \geq 0$$

for any admissible $v \geq \Psi = 0$.

We choose the admissible set as

$$U_{ad} = \left\{ u \in L^{q_u}(Q) : u \leq 0 \text{ almost everywhere in } Q, \int_Q u \, d(t, x) = -1 \right\}.$$

Further we choose $j_v(y) = \frac{1}{2}\|y\|^2_{L^2(Q)}$, $j_T = 0$ and $g(u) = \|u\|_{L^1(Q)}$. Thus for any $u \in U_{ad}$ we have

$$j_v(S(u)) + j_T(S(u)(T)) + g(u) = \frac{1}{2}\|0\|^2_{L^2(Q)} + 0 + 1 = 1.$$

Therefore all $u \in U_{ad}$ are optimal solutions.

Remark 3.5 It is possible to show uniqueness by utilizing the equivalence of (OC) to a convex, state constrained problem, at least under certain assumptions, see Section 3.4.3.

3.2 Control Problems Utilizing Regularized Obstacle Problems

For this section we let $(\bar{y}, \bar{u})$ be a global solution of (OC). We then consider a family of regularizations

$$\min_{(y,u)\in W(I)\times L^2(Q)} j_v(y) + j_T(y(T)) + g(u) + \frac{1}{2}\|u-\bar{u}\|^2_{L^2(Q)} =: J_{\bar{u}}(y,u), \qquad (\mathrm{OC}_{\gamma,\bar{u}})$$
$$\text{such that } S_\gamma(u) = y \text{ and } u \in U_{ad}.$$

Regularizations like these, relying on a chosen solution $(\bar{y}, \bar{u})$, often appear in problems without a unique solution. In case of optimal control involving variational inequalities see for example [Bar81, MP84, BL04, AL02] and in particular [Bar84, Chapter 5]. Note that these problems are obviously only of theoretical interest as they require the a priori knowledge of a solution $(\bar{y}, \bar{u})$ to (OC). We later study regularized and discretized problems, where this a priori knowledge is not required. The analysis of those problems will be similar.

Lemma 3.6 $(\mathrm{OC}_{\gamma,\bar{u}})$ *has a solution* $(\bar{y}_\gamma, \bar{u}_\gamma)$.

Proof. The same arguments and convergences used for the proof of Theorem 3.3 can be transfered to $(\mathrm{OC}_{\gamma,\bar{u}})$ as well. ☐

Theorem 3.7 *Let* $\kappa_I, \kappa_\Omega \in (0,1)$ *satisfy the properties from Theorem 2.22. For any sequence of solutions* $((\bar{u}_\gamma, y_\gamma))_{\gamma>0}$ *of* $(\mathrm{OC}_{\gamma,\bar{u}})$ *we have*

- $\bar{u}_\gamma \xrightarrow{\gamma\to 0} \bar{u}$ *strongly in* $L^2(Q)$ *and weakly in* $L^{q_u}(Q)$,
- $\bar{y}_\gamma \xrightarrow{\gamma\to 0} \bar{y}$ *strongly in* $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ *and weakly in* $W(I)$,
- $\lambda(\bar{u}_\gamma) \xrightarrow{\gamma\to 0} \lambda(\bar{u})$ *weakly in* $L^{q_u}(Q)$.

Remark 3.8 If we assume higher regularity of the domain, e.g. $C^{1,1}$ domains or polygonal domains with appropriate conditions on interior angles and q_u, we can improve this convergence result to convergence in the senses of Theorems 2.40 and 2.42.

Proof of Theorem 3.7. By the boundedness of U_{ad} we have that $(\|\bar{u}_\gamma\|_{L^{q_u}(Q)})_{\gamma>0}$ is bounded. Thus there exists a weak limit $\tilde{u} \in L^{q_u}(Q)$ along an appropriate zero sequence $(\gamma_n)_{n\in\mathbb{N}}$. By Theorem 2.34 we have for $\tilde{y} := S(\tilde{u})$ that

$$S_{\gamma_n}(\bar{u}_{\gamma_n}) \xrightarrow{n\to\infty} \tilde{y} \text{ strongly in } C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \text{ and weakly in } W(I)$$

and

$$\lambda_{\gamma_n}(u_{\gamma_n}) \xrightarrow{n\to\infty} \lambda(\tilde{u}) \text{ weakly in } L^{q_u}(Q).$$

We then have, by the lower semi-continuity of g,

$$J(\tilde{y}, \tilde{u}) \le \liminf_{n\to\infty} J(\bar{y}_{\gamma_n}, \bar{u}_{\gamma_n}) \le \limsup_{n\to\infty} J_{\bar{u}}(\bar{y}_{\gamma_n}, \bar{u}_{\gamma_n}) \le \limsup_{n\to\infty} J_{\bar{u}}(\bar{y}, \bar{u}) = J(\bar{y}, \bar{u}).$$

As $(\bar{y}, \bar{u})$ is a minimizer of J we have that $(\tilde{y}, \tilde{u})$ is also a minimizer of (OC) and

$$\lim_{n\to\infty} J(\bar{y}_{\gamma_n}, \bar{u}_{\gamma_n}) = \lim_{n\to\infty} J_{\bar{u}}(\bar{y}_{\gamma_n}, \bar{u}_{\gamma_n}) = J(\bar{y}, \bar{u}).$$

This implies

$$\lim_{n\to\infty} \frac{1}{2}\|\bar{u}_{\gamma_n} - \bar{u}\|^2_{L^2(Q)} = 0.$$

So we find $\tilde{u} = \bar{u}$ and $\tilde{y} = \bar{y}$ and the strong convergence of $(\bar{u}_{\gamma_n})_{n\in\mathbb{N}}$ in $L^2(Q)$. Because the limit is unique the whole sequence converges and not just the studied subsequence.

□

3.3 Optimality Conditions for Regularized Optimal Control Problems

We now use the regularized problems to deduce optimality conditions for the unregularized problem. We will use the differentiability of S_γ to derive optimality conditions for $(\mathrm{OC}_{\gamma,\bar{u}})$ and then in the next section pass to the limit.

Close to our approach is the work of [Bar81], respectively [Bar84]. In [Bar81] the author started with a graph-inclusion reformulation of (VIabs) and regularizes from there. Nevertheless, we still give a full discussion of the optimality conditions as understanding them fully is tantamount in applying them to the discussion of the numerical analysis of the optimal control problem itself. We also include a non-linearity in our discussion which is absent in the discussion in [Bar81].

In [Fri87] an approach similar to our was used. There the author used this approach to consider C^2-domains and specific obstacles to study bang-bang problems and boundary controls.

In [Che03] a very general approach was taken for a problem, similar to ours, posed in general metric spaces. There the author deduced Pontryagin's maximum principle in a very general setting. As we plan to use smoothness and high regularity to deduce and work with even second order conditions we abstain from pursuing this further. Another approach working with Pontryagin's maximum principle is found in [BZ99].

Throughout Sections 3.3 to 3.7 we make the following assumption on β_γ, unless noted otherwise.

Assumption 3.9 *Assume that β_γ lies in $C^{1,1}_{loc}(\mathbb{R})$.*

3.3.1 Regularity and Differentiability of Solution Operators Belonging to Regularized Obstacle Problems

We start by studying the regularity of S_γ.

Proposition 3.10 *Assume that β_γ lies in $C^{0,1}_{loc}(\mathbb{R})$. The regularized solution operator*

$$S_\gamma\colon L^{q_u}(Q) \to W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$$

is locally Lipschitz continuous. The Lipschitz constant does depend on γ.

Proof. Let $u_1, u_2 \in L^{q_u}(Q)$ and $y_i := S_\gamma(u_i) \in W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$. The regularity is given by Theorem 2.22. Taking the differences of the equation defining the y_i we have

$$\begin{cases} \partial_t(y_1 - y_2) + A(y_1 - y_2) + (f(y_1) - f(y_2)) + (\beta_\gamma(y_1 - \Psi) - \beta_\gamma(y_2 - \Psi)) = u_1 - u_2, \\ (y_1 - y_2)(0) = 0, \quad (y_1 - y_2)|_{\Sigma_D} = 0. \end{cases} \tag{3.1}$$

By Theorem 2.22 we have

$$\|y_i\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} \leq C_\gamma \left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} + \|f(0)\|_{L^\infty(Q)} + \|\beta_\gamma(-\Psi)\|_{L^\infty(Q)} \right).$$

Thus, by the local Lipschitz continuity of f we have almost everywhere in Q

$$|f(t, x, y_1(t, x)) - f(t, x, y_2(t, x))| \leq L|y_1(t, x) - y_2(t, x)| \tag{3.2}$$

with an L depending only on an upper bound to $\|u_1\|_{L^{q_u}(Q)}$ and $\|u_2\|_{L^{q_u}(Q)}$ but not on (t, x).

Thus by Rademacher's Theorem, cf. [AFP00, Proposition 2.13], and the fundamental theorem of calculus for Sobolev functions, cf. [Bré11, Theorem 8.2], we have almost everywhere in Q that

$$\begin{aligned} &f(t, x, y_1(t, x)) - f(t, x, y_2(t, x)) \\ &= \int_0^1 f'(t, x, y_2(t, x) + s(y_1(t, x) - y_2(t, x)) \, ds \cdot (y_1(t, x) - y_2(t, x)). \end{aligned}$$

Here, due to [AFP00, Proposition 2.13] and (3.2), we have that the weak derivative f' in the y-component, is bounded by L independently of $(t, x) \in Q$. The same arguments can be made for $\beta_\gamma(\cdot - \Psi)$. We denote the constant for β_γ by L as well.

If we abbreviate $\delta y := y_1 - y_2$ and $\delta u := u_1 - u_2$ we can write (3.1) as

$$\begin{cases} \partial_t \delta y + A\delta y + \int_0^1 f'(y_2 + s(y_1 - y_2)) + \beta_\gamma{}'(y_2 + s(y_1 - y_2)) \, ds \; \delta y = \delta u, \\ \delta y(0) = 0, \quad \delta y|_{\Sigma_D} = 0. \end{cases}$$

We define

$$\int_0^1 f'(y_2 + s(y_1 - y_2)) + \beta_\gamma{}'(y_2 + s(y_1 - y_2)) \, ds =: a_{y_1,y_2}.$$

By our earlier arguments a_{y_1,y_2} is bounded by $2L$ and by the monotonicity of f and β_γ it is non-negative.

So in total δy solves a parabolic PDE, with the elliptic operator A and the "non-linearity" $y \mapsto a_{y_1,y_2} y$ which has the Lipschitz constant $2L$, which does not depend on y_1 or y_2. Thus we can apply Theorem 8.17 to get

$$\|\delta y\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} \leq C\|\delta u\|_{L^{q_u}(Q)},$$

which is just the sought local Lipschitz continuity. □

Theorem 3.11 *The operator*

$$S_\gamma \colon L^{q_u}(Q) \to W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$$

is Fréchet differentiable. Its derivative at $u \in L^{q_u}(Q)$ in direction $d \in L^{q_u}(Q)$ is given by $S_\gamma{}'(u)d = z \in W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ as the weak solution of

$$\begin{cases} \partial_t z + Az + f'(S_\gamma(u))z + \beta_\gamma{}'(S_\gamma(u) - \Psi)z = d, \\ z(0) = 0, \quad z|_{\Sigma_D} = 0. \end{cases} \tag{3.3}$$

Proof. Let $u, d \in L^{q_u}(Q)$ and z the solution to (3.3). We also define $y := S_\gamma(u)$ and $y_d := S_\gamma(u+d)$. Subtracting the equation for y and z from the one for y_d yields

$$\begin{aligned} \partial_t(y_d - y - z) + A(y_d - y - z) + \big(f(y_d) - f(y) - f'(y)z\big) \\ + \big(\beta_\gamma(y_d - \Psi) - \beta_\gamma(y - \Psi) - {\beta_\gamma}'(y - \Psi)d\big) = 0 \end{aligned} \tag{3.4}$$

in a weak sense, with appropriate boundary conditions.

The mapping f is Fréchet differentiable from $L^\infty(Q)$ to $L^\infty(Q)$ by Lemma 2.15. So

$$f(y_d) - f(y) - f'(y)(y_h - y) = R_f \in L^\infty(Q)$$

with R_f satisfying

$$\|R_f\|_{L^\infty(Q)} = o(\|y_d - y\|_{L^\infty(Q)}). \tag{3.5}$$

The same holds true for β_γ with some R_{β_γ}. Introducing $y_d - y - z =: r_d$ (3.4) reads

$$\begin{cases} \partial_t r_d + A r_d + f'(y) r_d + {\beta_\gamma}'(y - \Psi) r_d = -R_f - R_{\beta_\gamma}, \\ r_d|_{\Sigma_D} = 0, \quad r_d(0) = 0. \end{cases}$$

Now, as in the last part of the proof of Proposition 3.10 Theorem 8.17 delivers,

$$\|r_d\|_{W(I)\cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} \le C \|R_f + R_{\beta_\gamma}\|_{L^{q_u}(Q)}$$

This implies, using the local Lipschitz continuity from the previous theorem and (3.5)

$$\frac{\|r_d\|_{W(I)\cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)}}{\|d\|_{L^{q_u}(Q)}} \le C \frac{\|R_f + R_{\beta_\gamma}\|_{L^\infty(Q)}}{\|y_d - y\|_{L^\infty(Q)}} \frac{\|y_d - y\|_{L^\infty(Q)}}{\|d\|_{L^{q_u}(Q)}} \xrightarrow{\|d\|_{L^{q_u}(Q)}} 0.$$

□

Remark 3.12 We now prove, if β_γ is non-smooth but has a special structure, that S_γ is still differentiable, provided that in the differentiation point u there holds that $\{S_\gamma(u) = \Psi\}$, defined below, is of measure 0. From experience this assumption is satisfied for any "naturally occurring" right hand side or control. We prove this theorem to indicate that some, or rather all, of the first order optimality discussion in Section 3.3.2 can be done for non-smooth β_γ as well as smooth β_γ. But to avoid the constant mentioning of the assumption that $\{S(u) = \Psi\}$ is a set of measure zero, we shall only discuss the cases of smooth β_γ and as the transfer to the non-smooth situation is straight forward.

The following theorem is not formulated in its utmost generality. It is possible to consider even less differentiable non-smooth regularization terms and non-linearities, cf. [Bet19]. We limit ourselves to a self contained and transparent example, which we will use later in our experiments in Chapter 7.

Theorem 3.13 *Assume that β_γ is of the form from Proposition 2.17 for $\alpha = 1$. Then*

$$S_\gamma \colon L^{q_u}(Q) \to W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$$

is Fréchet differentiable for each $u \in L^{q_u}(Q)$ such that

$$\{S_\gamma(u) = \Psi\} := \{(t,x) \in Q : S_\gamma(u)(t,x) = \Psi(t,x)\}$$

is of Lebesgue measure 0. As Ψ represents an equivalence class this set would be different for a different representative of the same equivalence class. However, sets for two different *representatives only differ on a set of of measure 0, so the statement is independent of the specific representative.*

The derivative at $u \in L^{q_u}(Q)$ in direction $d \in L^{q_u}(Q)$ is given by $S_\gamma{}'(u)d = z \in W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ as the weak solution of

$$\begin{cases} \partial_t z + Az + f'(S_\gamma(u))z + \beta_\gamma{}'(S_\gamma(u) - \Psi)z = d, \\ z(0) = 0, \quad z|_{\Sigma_D} = 0. \end{cases} \tag{3.6}$$

Here we use the notation $\beta_\gamma{}'(r) = \frac{1}{\gamma}1_{(-\infty,0)}(r)$, keeping in mind that the problematic case $r = 0$ is not relevant under our assumptions.

Proof. Let the notation be as in the proof of Theorem 3.11. By the same first steps we have, with $r_d := y_d - y - z$,

$$\begin{cases} \partial_t r_d + Ar_d + f'(y)r_d + (\beta_\gamma(y_d - \Psi) - \beta_\gamma(y - \Psi) - \beta_\gamma{}'(y - \Psi)z) = -R_f, \\ r_d(0) = 0, \quad r_d|_{\Sigma_D} = 0. \end{cases}$$

Rearranging this further results in

$$\begin{aligned} &\partial_t r_d + Ar_d + f'(y)r_d + \beta_\gamma{}'(y - \Psi)r_d \\ &= -R_f + \beta_\gamma{}'(y - \Psi)(y_d - y) + \beta_\gamma(y - \Psi) - \beta_\gamma(y_d - \Psi). \end{aligned}$$

Again, Theorem 8.17 yields

$$\begin{aligned} &\|r_d\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} \\ &\leq C(\|R_f\|_{L^{q_u}(Q)} + \|\beta_\gamma{}'(y - \Psi)(y_d - y) + \beta_\gamma(y - \Psi) - \beta_\gamma(y_h - \Psi)\|_{L^{q_u}(Q)}). \end{aligned} \tag{3.7}$$

We have to show that the right hand side is of order $o(\|d\|_{L^{q_u}(Q)})$. The term R_f is of order $o(\|d\|_{L^{q_u}(Q)})$, as in the proof of Theorem 3.11. We continue to inspect the terms with β_γ. Let $\epsilon > 0$ be arbitrary. By Proposition 3.10 we find for d sufficiently small that $y(t,x) - \Psi(t,x) < \epsilon$ implies $y_d(t,x) - \Psi(t,x) < \frac{\epsilon}{2}$. Thus, by the definition of β_γ, we find for those small d

$$\begin{aligned} &\|\beta_\gamma{}'(y - \Psi)(y_d - y) + \beta_\gamma(y - \Psi) - \beta_\gamma(y_d - \Psi)\|_{L^{q_u}(Q)} \\ &= \|\beta_\gamma{}'(y - \Psi)(y_d - y) + \beta_\gamma(y - \Psi) - \beta_\gamma(y_d - \Psi)\|_{L^{q_u}(\{|y-\Psi|<\epsilon\})} \\ &\leq \left(\frac{1}{\gamma}\|y_d - y\|_{L^\infty(Q)} + \|\beta_\gamma(y - \Psi) - \beta_\gamma(y_d - \Psi)\|_{L^\infty(\{|y-\Psi|<\epsilon\})}\right) |\{|y - \Psi| < \epsilon\}|^{\frac{1}{q_u}}. \end{aligned}$$

By the Lipschitz continuity of β_γ with constant γ^{-1} and Proposition 3.10 this implies

$$\|\beta_\gamma{}'(y - \Psi)(y_d - y) + \beta_\gamma(y - \Psi) - \beta_\gamma(y_d - \Psi)\|_{L^{q_u}(Q)} \leq \frac{2}{\gamma}\|d\|_{L^{q_u}(Q)} |\{|y - \Psi| < \epsilon\}|^{\frac{1}{q_u}}.$$

This entails

$$\frac{\|r_d\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}}{\|d\|_{L^{q_u}(Q)}} \leq \frac{o(\|d\|_{L^{q_u}(Q)})}{\|d\|_{L^{q_u}(Q)}} + C|\{|y - \Psi| < \epsilon\}|^{\frac{1}{q_u}}.$$

Sending $\|d\|_{L^{q_u}(Q)}$ to 0 yields

$$\limsup_{\|d\|_{L^{q_u}(Q)}\to 0} \frac{\|r_d\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}}{\|d\|_{L^{q_u}(Q)}} \leq C|\{|y - \Psi| < \epsilon\}|^{\frac{1}{q_u}}.$$

As $\epsilon > 0$ was arbitrary σ-continuity of the Lebesgue measure, e.g. [BK15, Proposition 3.1.], implies

$$\limsup_{\|d\|_{L^{q_u}(Q)}\to 0} \frac{\|r_d\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}}{\|d\|_{L^{q_u}(Q)}} \leq C|\{y - \Psi = 0\}|^{\frac{1}{q_u}}.$$

By assumption the right hand side is equal to 0 and we can conclude the proof here. □

3.3.2 First Order Optimality Conditions for Regularized Optimal Control Problems

Now we get an easy optimality condition by differentiating into the proper directions.

Corollary 3.14 *If* $(\bar{y}_\gamma, \bar{u}_\gamma)$ *is optimal for* $(\mathrm{OC}_{\gamma,\bar{u}})$ *then*

$$\left(j_v'(\bar{y}_\gamma), S'(\bar{y}_\gamma)(u-\bar{u}_\gamma)\right)_{L^2(Q)} + \left(j_T'(\bar{y}_\gamma(T)), (S'(\bar{y}_\gamma)(u-\bar{u}_\gamma))(T)\right)_{L^2(\Omega)} \\ + \left(g'(\bar{u}_\gamma) + \gamma(\bar{u}_\gamma - \bar{u}), u - \bar{u}_\gamma\right)_{L^2(Q)} \geq 0 \qquad \forall u \in U_{ad}.$$

We can now introduce an adjoint state to simplify this system.

Definition 3.15 For a given pair $(y,u) = (S_\gamma(u), u)$ with $u \in L^{q_u}(Q)$ we define $p_\gamma = p_\gamma(y,u) \in W(I)$ as the unique weak solution of

$$\begin{cases} -\partial_t p + Ap + f'(y)p + {\beta_\gamma}'(y-\Psi)p = j_v'(y), \\ p(T) = j_T(y(T)), \quad p|_{\Sigma_D} = 0. \end{cases} \tag{3.8}$$

Recall that A is by assumption self-adjoint. A solution to this linear problem exists by [Wlo92, Theorem 26.1]. There we also find the regularity $\bar{p}_\gamma \in W(I)$.

For an optimal solution $(\bar{y}_\gamma, \bar{u}_\gamma)$ of $(\mathrm{OC}_{\gamma,\bar{u}})$ we call $(\bar{u}_\gamma, \bar{y}_\gamma, \bar{p}_\gamma) := (\bar{u}_\gamma, \bar{y}_\gamma, p(\bar{y}_\gamma, \bar{u}_\gamma))$ an optimal triple of $(\mathrm{OC}_{\gamma,\bar{u}})$.

Using $\bar{p}_\gamma$ we can reformulate Corollary 3.14 to:

Corollary 3.16 *If* $(\bar{y}_\gamma, \bar{u}_\gamma)$ *is optimal then* $\bar{p}_\gamma = p_\gamma(\bar{y}_\gamma, \bar{u}_\gamma)$ *and* $\bar{u}_\gamma$ *satisfy*

$$\left(\bar{p}_\gamma + g'(\bar{u}_\gamma) + \gamma(\bar{u}_\gamma - \bar{u}), u - \bar{u}_\gamma\right)_{L^2(Q)} \geq 0 \qquad \forall u \in U_{ad}.$$

Proof. We test (3.8) with $\bar{p}_\gamma$ with $z := S_\gamma'(\bar{y}_\gamma)(u - \bar{u}_\gamma)$ for some $u \in U_{ad}$ and see that

$$\begin{aligned} \left(j_v'(\bar{y}_\gamma), z\right)_{L^2(Q)} &= \left(-\partial_t \bar{p}_\gamma + A\bar{p}_\gamma, z\right)_{L^2(I,V^*,V)} + \left(f'(\bar{y}_\gamma)\bar{p}_\gamma + {\beta_\gamma}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma, z\right)_{L^2(Q)} \\ &= \left(\bar{p}_\gamma, \partial_t z + Az\right)_{L^2(I,V,V^*)} + \left(f'(\bar{y}_\gamma)z + {\beta_\gamma}'(\bar{y}_\gamma - \Psi)z\right)_{L^2(Q)} \\ &\quad - \left(\bar{p}_\gamma(T), z(T)\right)_{L^2(\Omega)} + \left(\bar{p}_\gamma(0), z(0)\right)_{L^2(\Omega)} \\ &= \left(\bar{p}_\gamma, u - \bar{u}_\gamma\right)_{L^2(Q)} - \left(j_T'(\bar{y}_\gamma(T)), z(T)\right)_{L^2(\Omega)}. \end{aligned}$$

Applying this to the result of Corollary 3.14 yields the desired optimality condition. □

Lemma 3.17 *Let* $(\bar{u}_\gamma, \bar{y}_\gamma, \bar{p}_\gamma)$ *be an optimal triple of* $(\mathrm{OC}_{\gamma,\bar{u}})$. *Then*

$$\|\bar{p}_\gamma\|_{C(\bar{I},H)\cap L^2(I,V)} \leq C\left(\|j_T'(\bar{y}_\gamma(T))\|_{L^2(\Omega)} + \|j_v'(\bar{y}_\gamma)\|_{L^2(Q)}\right).$$

The constant $C > 0$ is independent of γ and the terms on the right are also bounded independently of γ.

Proof. Testing (3.8) with $\bar{p}_\gamma \cdot 1_{(t,T)}$ for some $t \in I$ yields

$$\begin{aligned} \frac{1}{2}\|\bar{p}_\gamma(t)\|_H^2 - \frac{1}{2}\|\bar{p}_\gamma(T)\|_H^2 + \nu_{\mathrm{ell}}\|\nabla \bar{p}_\gamma\|^2_{L^2((t,T)\times\Omega)} &\leq \left(-\partial_t \bar{p}_\gamma + A\bar{p}_\gamma, \bar{p}_\gamma\right)_{L^2((t,T)\times\Omega)} \\ &= \left(-f'(\bar{y}_\gamma)\bar{p}_\gamma, \bar{p}_\gamma\right)_{L^2((t,T)\times\Omega)} + \left(-{\beta_\gamma}'(\bar{y}_\gamma)\bar{p}_\gamma, \bar{p}_\gamma\right)_{L^2((t,T)\times\Omega)} + \left(j_v'(\bar{y}_\gamma), \bar{p}_\gamma\right)_{L^2((t,T)\times\Omega)} \\ &\leq \|j_v'(\bar{y}_\gamma)\|_{L^2(Q)}\|\bar{p}_\gamma\|_{L^2(Q)}. \end{aligned}$$

Here we used partial integration and the positivity of f' and ${\beta_\gamma}'$. As $t \in I$ was arbitrary this entails

$$\sup_{t\in I}\|\bar p_\gamma(t)\|_H^2 \le 2\|j_v'(\bar y_\gamma)\|_{L^2(Q)}\|\bar p_\gamma\|_{L^2(Q)} + \|j_T'(\bar y_\gamma(T))\|_H^2. \tag{3.9}$$

Thus

$$\|\bar p_\gamma\|_{L^2(I,V)}^2 \le \sqrt{T}\sup_{t\in I}\|\bar p_\gamma(t)\|_H^2 + \|\nabla \bar p_\gamma\|_{L^2(Q)}^2 \le C\left(\|j_v'(\bar y)\|_{L^2(Q)}\|\bar p_\gamma\|_{L^2(Q)} + \|j_T'(\bar y_\gamma(T))\|_H^2\right).$$

By Young's inequality we can kick back the term with $\bar p_\gamma$ on the right and obtain

$$\|\bar p_\gamma\|_{L^2(I,V)}^2 \le \sqrt{T}\sup_{t\in I}\|\bar p_\gamma(t)\|_H^2 + \|\nabla \bar p_\gamma\|_{L^2(Q)}^2 \le C\left(\|j_v'(\bar y)\|_{L^2(Q)}^2 + \|j_T'(\bar y_\gamma(T))\|_H^2\right).$$

Taking the root yields the estimate for $\|\bar p_\gamma\|_{L^2(I,V)}$. Inserting this into (3.9) yields the claim for $\|\bar p_\gamma\|_{C(\bar I,H)}$ as well. By Theorem 2.30, the boundedness of U_{ad} and the continuity of j_T' and j_v' the bounded stays bounded independently of γ. $\square$

Before we analyse its behaviour for $\gamma \to 0$ we also bound the multiplier associated with $\bar p_\gamma$.

Lemma 3.18 *Let $(\bar u_\gamma, \bar y_\gamma, \bar p_\gamma)$ be an optimal triple of $(\mathrm{OC}_{\gamma,\bar u})$. We have*

$$\|{\beta_\gamma}'(\bar y_\gamma - \Psi)\bar p_\gamma\|_{L^1(Q)} \le C(\|j_T'(\bar y_\gamma(T))\|_H + \|j_v'(\bar y_\gamma)\|_{L^2(Q)})$$

with $C > 0$ independent of $\gamma > 0$. The right hand side is bounded independently of γ.

Proof. For $\delta > 0$, $x \in \mathbb{R}$ we define $\mathrm{abs}_\delta(x) = \sqrt{\delta + x^2}$ and $\mathrm{sgn}_\delta(x) = \frac{x}{\sqrt{\delta + x^2}} = \mathrm{abs}'(x)$.

Now we test (3.8) with $\mathrm{sgn}_\delta(\bar p_\gamma)$ to receive

$$\begin{aligned}&({\beta_\gamma}'(\bar y_\gamma - \Psi)\bar p_\gamma, \mathrm{sgn}_\delta(\bar p_\gamma))_{L^2(Q)}\\ &\quad= (\partial_t \bar p_\gamma - A\bar p_\gamma, \mathrm{sgn}_\delta(\bar p_\gamma))_{L^2(I,V^*,V)} + (-f'(\bar y_\gamma)\bar p_\gamma + j_v'(\bar y_\gamma), \mathrm{sgn}_\delta(\bar p_\gamma))_{L^2(Q)}.\end{aligned}$$

We estimate the left side term by term:

$$\begin{aligned}(\partial_t \bar p_\gamma, \mathrm{sgn}_\delta\, \bar p_\gamma)_{L^2(I,V^*,V)} &= \int_\Omega \int_0^T \partial_t(\mathrm{abs}_\delta(\bar p_\gamma))\,dt\,dx = \int_\Omega \mathrm{abs}_\delta(\bar p_\gamma(T)) - \mathrm{abs}_\delta(\bar p_\gamma(0))\,dx\\ &\le \int_\Omega |\bar p_\gamma(T)| + \sqrt{\delta}\,dx \le C\|j_T'(\bar y_\gamma(T))\|_H + |\Omega|\delta.\end{aligned}$$

Here, $C > 0$ is independent of γ.

Using the ellipticity of A and the positivity of sgn' we can estimate the second term:

$$-(A\bar p_\gamma, \mathrm{sgn}_\delta\, \bar p_\gamma)_{L^2(I,V^*,V)} = -\sum_{i,j=1}^N \left(a_{ij}\partial_{x_i}\bar p_\gamma, \mathrm{sgn}_\delta'(\bar p_\gamma)\partial_{x_j}\bar p_\gamma\right)_{L^2(Q)} \le 0.$$

Because $f' \ge 0$ and $\mathrm{sgn}_\delta(s)s \ge 0$ for any s the third term satisfies

$$(-f'(\bar y_\gamma)\bar p_\gamma, \mathrm{sgn}_\delta(\bar p_\gamma))_{L^2(Q)} \le 0.$$

The fourth term is estimated directly by the boundedness of sgn_δ:

$$(j_v'(\bar y_\gamma), \mathrm{sgn}_\delta(\bar p_\gamma))_{L^2(Q)} \le \|\mathrm{sgn}_\delta(\bar p_\gamma)\|_{L^\infty(Q)}\|j_v'(\bar y_\gamma)\|_{L^1(Q)} \le C\|j_v'(\bar y_\gamma)\|_{L^2(Q)}.$$

Here, C is independent of γ. So for any $\delta > 0$ we have

$$\left(\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma, \operatorname{sgn}_\delta(\bar{p}_\gamma)\right)_{L^2(Q)} \le C(\|j_T'(\bar{y}_\gamma(T))\|_H + |\Omega|\delta + \|j_v'(\bar{y}_\gamma)\|_{L^2(Q)}). \tag{3.10}$$

Again, C is independent of γ. We now want to consider the limit $\delta \to 0$. By the boundedness of $\bar{y}_\gamma$ and Ψ the integrand on the left of (3.10) is bounded by an $L^1(Q)$-function. Thus the theorem of dominated convergence, e.g. [BK15, Proposition 5.4], yields

$$\begin{aligned} C(\|j_T'(\bar{y}_\gamma(T))\|_H + \|j_v'(\bar{y}_\gamma)\|_{L^2(Q)}) &\ge \lim_{\delta\to 0} \left(\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma, \operatorname{sgn}_\delta(\bar{p}_\gamma)\right)_{L^2(Q)} \\ &= \int_Q \lim_{\delta\to 0} \beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma \operatorname{sgn}_\delta(\bar{p}_\gamma)\, d(t,x) \\ &= \int_Q \beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma \operatorname{sgn}(\bar{p}_\gamma)\, d(t,x). \end{aligned}$$

By the positivity of $\beta_\gamma{}'$ we find

$$\|\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma\|_{L^1(Q)} \le C(\|j_T'(\bar{y}_\gamma(T))\|_H + \|j_v'(\bar{y}_\gamma)\|_{L^2(Q)}).$$

As in the last line of the proof of Lemma 3.17 this is bounded independently of γ. □

Lemma 3.19 *Let* $(\bar{u}_\gamma, \bar{y}_\gamma, \bar{p}_\gamma)$ *an optimal triple of* $(\mathrm{OC}_{\gamma,\bar{u}})$. *We have*

$$\|\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma\|_{W(I)^*} \le C\left(\|j_v'(\bar{y}_\gamma)\|_{L^2(Q)} + \|j_T'(\bar{y}_\gamma(T))\|_H + \|f'(\bar{y}_\gamma)\|_{L^\infty(Q)}\right).$$

with $C > 0$ *independent of* $\gamma > 0$. *The right hand side is bounded independently of* γ.

Proof. Let $\varphi \in W(I)$. We test (3.8) with φ and obtain

$$\begin{aligned} &(\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma, \varphi)_{L^2(Q)} \\ &\quad= (\partial_t\bar{p}_\gamma - A\bar{p}_\gamma, \operatorname{sgn}_\delta(\bar{p}_\gamma))_{L^2(I,V^*,V)} + \left(-f'(\bar{y}_\gamma)\bar{p}_\gamma + j_v'(\bar{y}_\gamma), \operatorname{sgn}_\delta(\bar{p}_\gamma)\right)_{L^2(Q)}. \\ &\quad\le \|j_v'(\bar{y}_\gamma)\|_{L^2(Q)}\|\varphi\|_{L^2(Q)} + \|A\|_{L^\infty(\Omega)}\|\bar{p}_\gamma\|_{L^2(I,V)}\|\varphi\|_{L^2(I,V)} \\ &\qquad + \|f'(\bar{y}_\gamma)\|_{L^\infty(Q)}\|\bar{p}_\gamma\|_{L^2(Q)}\|\varphi\|_{L^2(Q)} - \int_I (\bar{p}_\gamma, \partial_t\varphi)_H\, dt + (\bar{p}_\gamma(t), \varphi(t))_H\big|_0^T. \end{aligned}$$

By Lemma 3.17 we can bound these resulting in

$$\begin{aligned} (\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma, \varphi)_{L^2(Q)} &\le C\left(\|j_v'(\bar{y}_\gamma)\|_{L^2(Q)} + \|j_T'(\bar{y}_\gamma(T))\|_H + \|f'(\bar{y}_\gamma)\|_{L^\infty(Q)}\right)\|\varphi\|_{L^2(I,V)} \\ &\quad + \|\bar{p}_\gamma\|_{L^2(I,V)}\|\partial_t\varphi\|_{L^2(I,V^*)} + 2\|\bar{p}_\gamma\|_{C(\bar{I},H)}\|\varphi\|_{C(\bar{I},H)} \\ &\le C\left(\|j_v'(\bar{y}_\gamma)\|_{L^2(Q)} + \|j_T'(\bar{y}_\gamma(T))\|_H + \|f'(\bar{y}_\gamma)\|_{L^\infty(Q)}\right)\|\varphi\|_{W(I)}. \end{aligned}$$

This proves the estimate.

By Theorem 3.7 and the local Lipschitz continuity of f', which is uniform in Q, the term $\|f'(\bar{y}_\gamma)\|_{L^\infty(Q)}$ stays bounded independently of γ. The other terms stay bounded as in the last line of the proof of Lemma 3.17. □

3.4 Optimality Conditions for Unregularized Optimal Control Problems

As stated in the beginning of Section 3.3 we can now consider the limit $\gamma \to 0$ to obtain optimality conditions for the unregularized optimal control problem.

Here, we shall also discuss other approaches to establish optimality conditions for (OC). The possibly most direct approach was recently discovered in [Chr19]. It proves and uses the directional differentiability of $S : L^2(Q) \to L^2(I, V)$. Also see [Chr18] for important work on the differentiability in the elliptic case. So far the concrete form of the directional derivative of S is not known. Its existence is shown by indirect arguments. Also, to make (OC), or rather one of its regularizations, accessible to computation, a derivative with higher smoothness would be of interest, see Chapter 6. It is also not immediately clear how the analysis of the end time funtional j_T would look like when trying to use directional differentiability of S to derive optimality conditions. We thus do not delve deeper into the approach of [Chr19]. Here, we would also like to warn the reader of [JJ13]: a paper that makes the same claims. The authors, however, claim that Lipschitz continuity implies Gateaux differentiability, cf. the proof of [JJ13, (101) in Theorem 12].

Similar strategies have been discussed for related problems, for example in [JKRS03] the conical differentiability of the solution operator to a parabolic problem, involving a boundary obstacle, is analysed. Yet, we must warn that the paper contains some vagueness, which makes it impossible to transfer the results from their problem to ours. Specifically, the proof of [JKRS03, (9)] refers to an unavailable source. Personal communication with one of the authors was not productive either. Conical differentiability for elliptic variational inequalities is for example examined in [Mig76].

We will also comment shortly on why constraint qualifications and the Karush-Kuhn-Tucker conditions are unsuited for deriving optimality conditions. The appropriate way is to derive so called stationarity systems as in Theorem 3.38. There are various stationarity concepts (e.g. W-stationarity, C-stationarity, B-stationarity, S-stationarity, M-stationarity) that correspond to how much information on the involved multipliers ($(\bar{\lambda}, \bar{p}, \bar{\mu})$ in Theorem 3.38) is obtained. As we mainly use the regularized problems for computation and numerical analysis anyway, we do not delve into those concepts. For the elliptic version of (OC) see for example [Wac16b, Wac14, HW18] for stationarity concepts and their differences. The works make use of capacity theory, which is a can of worms we do not want to open within the scope of this thesis.

Remark 3.20 This is the point where we comment on the issue of constraint qualifications and why we not simply use Karush-Kuhn-Tucker conditions, cf. [UU12, Theorem 16.14]. Using Theorem 2.33 and the surrounding discussion one can reformulate (OC) to the equivalent problem

$$\min_{(y,u,\lambda)\in W(I)\times U_{ad}\times L^{q_u}(Q)} j_v(y) + j_T(y(T)) + g(u) = J(y,u),$$

$$\text{such that } \begin{cases} \partial_t y + Ay + f(y) = u + \lambda \\ \lambda \geq 0,\ y - \Psi \geq 0,\ (y - \Psi, \lambda)_{L^2(Q)} = 0. \end{cases}$$

This is now a mathematical problem with complementarity constraints (MPCC). The obstacle constraint now essentially corresponds to two equality and two inequality constraints. They are also easily differentiable, yet: one cannot expect reasonable constraint qualifications to hold on the biactive set $\{y = \Psi\} \cap \{\lambda = 0\}$. We give the following quote from [HW18, Section 4.1]: "Due to the violation of standard constraint qualifications, the classical Karush-Kuhn-Tucker conditions fail to be satisfied for some optimization problem of type (12) [a finite dimensional MPCC]; Moreover since the Mangasarian-Fromovitz condition is inevitably violated, the set of Lagrange multipliers for (12) is always unbounded (or empty)."

To see this we consider a simplistic example in two dimensions.

$$\min_{(x,y)\in\mathbb{R}^2} j(x,y) \text{ s.th. } h(x,y) := xy = 0, g_1(x,y) = -x \leq 0, g_2(x,y) = -y \leq 0 \tag{3.11}$$

The cost functional j is an arbitrary function $\mathbb{R}^2 \to \mathbb{R}$, it is not important in this example. The interesting part of this example is its biactive set: $\{(0,0)\}$. The Mangasarian-Fromovitz constraint qualification (MFCQ), cf. [UU12, Definition 16.19], is now satisfied in $(0,0)$ if $\nabla h(0,0)$ has full rank and there exists a $d \in \mathbb{R}^2$ such that $\nabla h(0,0)^T d = 0$, $\nabla g_1(0,0)^T d < 0$ and $\nabla g_2(0,0)^T d < 0$. Well, its easy to see that $\nabla h(0,0) = (0,0)^T$, which obviously does not have full rank, resulting in the consequences described in the above quote of [HW18].

Motivated by this we will derive stationarity conditions and will see that the multipliers we derive will be bounded, cf. Theorem 3.38, and possibly unique.

3.4.1 Stationarity Conditions

The goal of the following few statements is to establish so-called stationarity conditions in Theorem 3.38.

Remark 3.21 All the following analysis could be extended to less smooth β_γ than the ones in Assumption 3.9, if one makes the assumptions from Theorem 3.13 instead of $\beta_\gamma \in C^{1,1}_{loc}(\mathbb{R})$, cf. Remark 3.12. Only minor adaptions would have to be made in the following.

Lemma 3.22 *Let $(\gamma_n)_{n\in\mathbb{N}}$ be a zero sequence and $((\bar{u}_{\gamma_n}, \bar{y}_{\gamma_n}, \bar{p}_{\gamma_n}))_{n\in\mathbb{N}}$ be a sequence of corresponding optimal triples. There exists a subsequence $(\gamma_{n_k})_{k\in\mathbb{N}}$ and a $\bar{p} \in L^\infty(I,H) \cap L^2(I,V)$ such that*

$$\bar{p}_{\gamma_{n_k}} \xrightarrow{k\to\infty} \bar{p} \text{ weakly in } L^2(I,V) \text{ and weakly* in } L^\infty(I,H).$$

The adjoint $\bar{p}$ satisfies

$$\|\bar{p}\|_{L^\infty(I,H)\cap L^2(I,V)} \leq C(\|j_T'(\bar{y}(T))\|_{L^2(Q)} + \|j_v'(\bar{y})\|_{L^2(Q)}).$$

Proof. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a zero sequence. Using Lemma 3.17 we see that $\left(\|\bar{p}_{\gamma_n}\|_{L^2(I,V)}\right)_{n\in\mathbb{N}}$ is a uniformly bounded sequence in a Hilbert space, thus we find a weakly convergent subsequence, denoted by the same indices, converging weakly to some $\bar{p} \in L^2(I,V)$. By Lemma 1.23 we have that $L^1(I,H)^* = L^\infty(I,H)$. By Banach-Alaoglu, cf. [Wer11, Corollary VIII.3.12], the unit ball in $L^\infty(I,H)$ is weak* compact. $L^1(I,H)$ is separable by Theorem 1.27 in conjunction with [Emm04, Korollar 7.1.3] and thus the weak* topology is metrizable, see [Bré11, Theorem 3.28]. Therefore weak* compactness and weak* sequential compactness coincide, see for example [Bré11, Theorem 3.30]. So, in conclusion we can extract a subsequence of $(\gamma_n)_{n\in\mathbb{N}}$, denoted by the same indices, such that $(\bar{p}_{\gamma_n})_{n\in\mathbb{N}}$ has weak* limit $\hat{p}$ in $L^\infty(I,H)$. To see that this weak* limit is simply $\bar{p}$, let $\varphi \in L^2(I,H) = L^2(Q) \subset L^2(I,V)^* \cap L^1(I,H)$. Then

$$(\bar{p},\varphi)_{L^2(Q)} = \lim_{n\to\infty}(\bar{p}_{\gamma_n},\varphi)_{L^2(Q)} = (\hat{p},\varphi)_{L^2(Q)}.$$

Because $L^2(Q)$ is dense in $L^1(I,H)$ by Theorem 1.27 we can conclude $\bar{p} = \hat{p}$.

Theorem 3.7 and Lemma 3.17 together imply the statement after taking the limit on both sides of the inequality in Lemma 3.17 as the norms are weakly and weakly* lower semi-continuous. □

Definition 3.23 Given an optimal triple $(\bar{u}_\gamma, \bar{y}_\gamma, \bar{p}_\gamma)$ we abbreviate

$$\bar{\eta}_\gamma := -\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma \in L^1(Q) \cap W(I)^*.$$

In the following we consider the dual space of $C(\bar{Q})$ frequently. According to [Rud74, Theorem 6.19] we have $C(\bar{Q})^* \simeq M(\bar{Q})$. Since $C(\bar{Q})$ is separable, e.g. [Kec95, Theorem 4.19], weak* compactness is the same as weak* sequential compactness in $M(\bar{Q})$ by the same arguments as the ones used in the proof of Lemma 3.22.

Lemma 3.24 *Let $(\gamma_n)_{n\in\mathbb{N}}$ be a zero sequence and $((\bar{u}_{\gamma_n}, \bar{y}_{\gamma_n}, \bar{p}_{\gamma_n}))_{n\in\mathbb{N}}$ a sequence of corresponding optimal triples. There exists a subsequence $(\gamma_{n_k})_{k\in\mathbb{N}}$, an $\bar{\eta} \in M(\bar{Q})$ and an $\bar{\eta}_* \in W(I)^*$ such that $\bar{\eta}_{\gamma_{n_k}} \xrightarrow{k\to\infty} \bar{\eta}$ weakly* in $M(\bar{Q})$ and $\bar{\eta}_{\gamma_{n_k}} \xrightarrow{k\to\infty} \bar{\eta}_*$ weakly in $W(I)^*$. The limits satisfy*

$$\begin{aligned}\|\bar{\eta}\|_{M(\bar{Q})} &\leq C(\|j_T'(\bar{y}(T))\|_H + \|j_v'(\bar{y})\|_{L^2(Q)}),\\ \|\bar{\eta}_*\|_{W(I)^*} &\leq C(\|j_T'(\bar{y}(T))\|_H + \|j_v'(\bar{y})\|_{L^2(Q)} + \|f'(\bar{y})\|_{L^\infty(Q)})\end{aligned}$$

with some $C > 0$.

Proof. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a zero sequence. By the boundedness of $(\bar{\eta}_{\gamma_n})_{n\in\mathbb{N}}$ in $W(I)^*$ by Lemma 3.19 there exists a subsequence $(\bar{\eta}_{n_k})_{k\in\mathbb{N}}$ converging weakly to a limit $\bar{\eta}_* \in W(I)^*$ as $W(I)$ is a Hilbert space. The sequence is also bounded in $L^1(Q) \subset M(\bar{Q})$ by Lemma 3.18. We thus have by the remarks in Definition 3.23 that there exists a subsequence $(\bar{\eta}_{n_k})_{k\in\mathbb{N}}$, with the same indices, converging weakly* to a limit $\bar{\eta} \in M(\bar{Q})$.

Because norms are weakly* lower semi-continuous and because $(\bar{y}_{\gamma_{n_k}})_{k\in\mathbb{N}}$ converges uniformly by Theorem 3.7 we find the norm estimates by Lemma 3.18 and Lemma 3.19 ☐

Remark 3.25 It is not clear that $\bar{\eta}$ and $\bar{\eta}_*$ can be identified. However, for any $\varphi \in W(I)\cap C(\bar{Q})$ we find for $(\gamma_{n_k})_{k\in\mathbb{N}}$ from the proof of the previous Lemma 3.24

$$(\bar{\eta}, \varphi)_{M(\bar{Q}),C(\bar{Q})} = \lim_{k\to\infty} (\bar{\eta}_{\gamma_{n_k}}, \varphi)_{L^2(Q)} = \lim_{k\to\infty} (\bar{\eta}_{\gamma_{n_k}}, \varphi)_{W(I)^*,W(I)} = (\bar{\eta}_*, \varphi)_{W(I)^*,W(I)}.$$

It would now be possible to consider $\bar{\eta}$ and $\bar{\eta}_*$ as two seperate multipliers or to make additional assumptions on either to force a unique identification by, for example, setting $\bar{\eta}$ to 0 on certain parts of the boundary, cf. [CV19].

We refrain from doing this for the simple reason that it does add a whole lot to the resulting optimality system in Theorem 3.38, but makes the presentation a lot less clear. We shall, however, mention it from time to time. It will turn out that $\bar{\eta} \in M(\bar{Q})$ has the more interesting properties anyway.

Definition 3.26 We define the set of multipliers

$$\begin{aligned}P_{\bar{y}}^{\beta} := \Big\{&(\bar{p}, \bar{\eta}) \in (L^2(I,V) \cap L^\infty(I,H)) \times M(\bar{Q}) : \text{There exists a sequence}\\ &(\bar{p}_{\gamma_n}, \bar{\eta}_{\gamma_n})_{n\in\mathbb{N}} \text{ converging to } (\bar{p}, \bar{\eta}) \text{ weakly/weakly* } \times \text{ weakly*}\\ &\text{in } (L^2(I,V) \cap L^\infty(I,H)) \times M(\bar{Q})\Big\}.\end{aligned} \tag{3.12}$$

We say that a pair $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^{\beta}$ is generated by $(\gamma_n)_{n\in\mathbb{N}}$ if there are solutions $(\bar{y}_{\gamma_n}, \bar{u}_{\gamma_n})$ to $(\mathrm{OC}_{\gamma,\bar{u}})$ such that the corresponding pair $(\bar{p}_{\gamma_n}, \bar{\eta}_{\gamma_n})$ converges to $(\bar{p}, \bar{\eta})$ in the sense of (3.12).

Remark 3.27 Here one could extend $P_{\bar{y}}^{\beta}$ to include $\bar{\eta}_*$ from Lemma 3.24 as a third multiplier.

Definition 3.28 To establish a relation between $\bar{p}$ and $\bar{\eta}$, for $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^{\beta}$, we define

$$W_0(I) := \{v \in W(I) : v(0) = 0\}.$$

This is well-defined because $W(I) \hookrightarrow C(\bar{I}, H)$ continuously. Due to this embedding we have that $W_0(I)$ is a closed subspace of $W(I)$ and therefore a Hilbert space itself.

Lemma 3.29 *Each $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^{\beta}$ satisfies*

$$\begin{aligned} &(\bar{p}, \partial_t \varphi)_{L^2(I,V,V^*)} + a(\varphi, \bar{p}) + \left(f'(\bar{y})\bar{p}, \varphi\right)_{L^2(Q)} \\ &= (\bar{\eta}, \varphi)_{M(\bar{Q}),C(\bar{Q})} + \left(j_v'(\bar{y}), \varphi\right)_{L^2(Q)} + \left(j_T'(\bar{y}(T)), \varphi(T)\right)_H \end{aligned}$$

for any $\varphi \in W_0(I) \cap C(\bar{Q})$.

Proof. Let $\varphi \in W_0(I) \cap C(\bar{Q})$ and $\gamma > 0$. Then after partial integration (3.8) yields

$$\begin{aligned} &(\bar{p}_\gamma, \partial_t \varphi)_{L^2(I,V,V^*)} - \left(j_T'(\bar{y}_\gamma(T)), \varphi(T)\right)_H + a\left(\bar{p}_\gamma, \varphi\right)_{L^2(Q)} + \left(f'(\bar{y}_\gamma)\bar{p}_\gamma, \varphi\right)_{L^2(Q)} \\ &= (\bar{\eta}_\gamma, \varphi)_{W(I)^*,W(I)} + \left(j_v'(\bar{y}_\gamma), \varphi\right)_{L^2(Q)}. \end{aligned}$$

Taking the limit yields the desired result with Lemma 3.22 and Lemma 3.24. □

Remark 3.30 Here one could derive the same equality $\bar{\eta}_*$ from Lemma 3.24 with the strengthening assumption that the test functions φ stem only from $W_0(I)$ making the equality slightly stronger.

We can finally derive the analogue to Corollary 3.16.

Corollary 3.31 *If $(\bar{y}, \bar{u})$ is optimal, then for any $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^{\beta}$ we have*

$$\left(\bar{p} + g'(\bar{u}), u - \bar{u}\right)_{L^2(Q)} \geq 0 \quad \forall u \in U_{ad}.$$

Proof. For any $u \in U_{ad}$ we have for any $\gamma > 0$ by Corollary 3.16:

$$\left(\bar{p}_\gamma + g'(\bar{u}_\gamma) + \gamma(\bar{u}_\gamma - \bar{u}), u - \bar{u}_\gamma\right)_{L^2(Q)} \geq 0. \tag{3.13}$$

By the strong convergence of the controls from Theorem 3.7 and the continuity of g' we have

$$\left(g'(\bar{u}_\gamma) + \gamma(\bar{u}_\gamma - \bar{u}), u - \bar{u}_\gamma\right)_{L^2(Q)} \xrightarrow{\gamma \to 0} \left(g'(\bar{u}), u - \bar{u}\right)_{L^2(Q)}.$$

Let $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^{\beta}$ be generated by the zero sequence $(\gamma_n)_{n \in \mathbb{N}}$. Then we have that the weak convergence of $\left(\bar{p}_{\gamma_n}\right)_{n \in \mathbb{N}}$ together with the strong convergence of the controls yields

$$\left(\bar{p}_{\gamma_n}, u - \bar{u}_{\gamma_n}\right)_{L^2(Q)} \xrightarrow{n \to \infty} \left(\bar{p}, u - \bar{u}\right)_{L^2(Q)}.$$

Taking the limit in (3.13) yields the desired result. □

A huge part of the stationarity conditions is the complementarity of certain multipliers with certain states. The derivation of those complementarity conditions is the goal of the next few lemmas.

Lemma 3.32 *Assume $\Psi \in C(\bar{Q})$. Let $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^{\beta}$. We have $(\bar{y} - \Psi, \bar{\eta})_{C(\bar{Q}),M(\bar{Q})} = 0$. We have that $\bar{\eta}$ is concentrated on the well-defined set $\{(t,x) \in \bar{Q} : \bar{y}(t,x) = \Psi(t,x)\}$, i.e.* $\operatorname{supp}(\bar{\eta}) \subset \{(t,x) \in \bar{Q} : \bar{y}(t,x) = \Psi(t,x)\}$.

Proof. We know that $(\bar{y}_\gamma)_{\gamma>0}$ converges uniformly by Theorem 3.7 to $\bar{y}$. Thus

$$\max(\bar{y}_\gamma - \Psi, 0) \xrightarrow{\gamma\to 0} \bar{y} - \Psi \geq 0 \text{ uniformly.}$$

Let $(\bar{p}, \bar{\eta}) \in P^\beta_{\bar{y}}$ produced by $(\gamma_n)_{n\in\mathbb{N}}$. For any function $\varphi \in C(\bar{Q})$ we have, since $\beta'_{\gamma_n}(\bar{y}_{\gamma_n} - \Psi)\max(\bar{y}_\gamma - \Psi, 0) = 0$ almost everywhere in $\bar{Q}$,

$$((\bar{y} - \Psi)\varphi, \bar{\eta})_{C(\bar{Q}),M(\bar{Q})} = \lim_{n\to\infty} -\int_{\bar{Q}} \varphi \max(\bar{y}_{\gamma_n} - \Psi, 0)\beta_\gamma{}'(\bar{y}_{\gamma_n} - \Psi)\bar{p}_{\gamma_n}\, d(t,x) = 0. \quad (3.14)$$

Because $\bar{y}$ and Ψ are continuous on $\bar{Q}$ the sets $L_\epsilon := \{(t,x) \in \bar{Q} : \bar{y}(t,x) - \Psi(t,x) \geq \epsilon\}$ and $Q_0 := \{(t,x) \in \bar{Q} : \bar{y}(t,x) - \Psi(t,x) = 0\}$ are well defined, unique and closed in $\bar{Q}$.

We now show $\bar{\eta}|_{Q\setminus Q_0} = 0$. Let $\varphi \in C_c(\bar{Q} \setminus Q_0)$, where $C_c(\bar{Q} \setminus Q_0)$ is the set of continuous functions with compact support in $\bar{Q} \setminus Q_0$. It is dense in $C_0(\bar{Q} \setminus Q_0)$, e.g. [Rud74, Section 6.18]. On the compact set $\operatorname{supp}(\varphi)$ the continuous map $\bar{y} - \Psi \geq 0$ attains a minimum $\epsilon \geq 0$ at some $(t_\epsilon, x_\epsilon) \in \operatorname{supp}(\varphi) \subset \bar{Q} \setminus Q_0$. We claim that $\epsilon > 0$. If $\bar{y}(t_\epsilon, x_\epsilon) - \Psi(t_\epsilon, x_\epsilon) = 0$, we would have $(t_\epsilon, x_\epsilon) \in Q_0$, a contradiction.

Then $\operatorname{supp}(\varphi) \in C(L_\epsilon)$ and $\frac{\varphi}{\bar{y}-\Psi} \in C(\bar{Q})$. Therefore (3.14), testing it with $\frac{\varphi}{\bar{y}-\Psi}$, yields

$$0 = (\varphi, \bar{\eta})_{C(\bar{Q}),M(\bar{Q})} = \int_{\bar{Q}\setminus Q_0} \varphi\, d\bar{\eta} = \int_{\bar{Q}\setminus Q_0} \varphi\, d\bar{\eta}|_{\bar{Q}\setminus Q_0}.$$

As φ was arbitrary in $C_c(\bar{Q} \setminus Q_0)$ and by the aforementioned density and [Rud74, Theorem 2.19] once again we find $0 = \bar{\eta}|_{\bar{Q}\setminus Q_0} \in M(\bar{Q} \setminus Q_0) \simeq C_0(\bar{Q} \setminus Q_0)^*$. Proposition 8.1 now entails the claim on the support.

This also entails the complementarity condition. $\square$

We will now state a growth assumption on β_γ. As we plan to take the limit $\gamma \to 0$, to go from (PDE$_\gamma$) to (VI-OB), this is not a meaningful restriction. In Proposition 2.17 we have a concrete example for a β_γ satisfying Assumption 3.33 as we will prove in Proposition 3.34.

Assumption 3.33 *Assume that there exist $c_1^\beta, c_2^\beta > 0$ with*

$$c_1^\beta|\beta(r)| \leq |\beta'(r)r| \leq c_2^\beta|\beta(r)| \qquad \forall r \in \mathbb{R}.$$

Proposition 3.34 *Assume β has the form from Proposition 2.17 for some $\alpha > 1$. Then β satisfies Assumption 3.33.*

Remark 3.35 This proposition still holds true for $\alpha = 1$, as the only non-differentiable spot of β is $r = 0$, where we might assign an arbitrary value as it is weighted by $r = 0$.

Proof. Clearly for $r \geq 0$ there is nothing to show. Let $0 > r \geq -\alpha^{1/(1-\alpha)}$. Then we have

$$|\beta(r)| = |(-r)^\alpha| < \alpha|(-r)^\alpha| = |\beta'(r)r| = \alpha|\beta(r)|.$$

Now let $r < -\alpha^{1/(1-\alpha)}$. Then we have

$$|\beta(r)| = -r - \frac{\alpha - 1}{\alpha}\alpha^{1/(1-\alpha)} < -r = |\beta'(r)r|.$$

We show $|\beta'(r)r| \leq \alpha|\beta(r)|$ by showing this is equivalent to a true inequality:

$$\begin{aligned} |\beta'(r)r| \leq \alpha|\beta(r)| &\iff -r \leq -\alpha r - (\alpha-1)\alpha^{1/(1-\alpha)} \\ &\iff (\alpha-1)r \leq -(\alpha-1)\alpha^{1/(1-\alpha)} \\ &\iff r \leq -\alpha^{1/(1-\alpha)}. \end{aligned}$$

The last line is true by the assumption we made on r in this case. □

The following is a "correction" of [Bar81, Lemma 7] following the ideas of [Bar84] by the same author. Unfortunately in [Bar81, Lemma 7] an application of Arzela-Ascoli's theorem was not possible.

Lemma 3.36 *Let $(\bar{p},\bar{\eta}) \in P^{\beta}_{\bar{y}}$ be generated by $(\gamma_n)_{n\in\mathbb{N}}$. We then have for any $s > \frac{N}{2}$ that $\bar{p} \in BV(\bar{I},(H^s(\Omega)\cap V)^*)$ and*

$$\bar{p}_{\gamma_n} \xrightarrow{n\to\infty} \bar{p} \text{ strongly in } L^2(Q).$$

Here $H^s(\Omega)$ is a fractional Sobolev space, see Section 8.8, and $BV(\bar{I},(H^s(\Omega)\cap V)^*)$ is defined as the set of functions $v : \bar{I} \to (H^s(\Omega)\cap V)^*$ such that

$$\operatorname{Var}(v,(H^s(\Omega)\cap V)^*) := \sup_{n\in\mathbb{N}, 0=t_0<t_1<\cdots<t_n=T} \sum_{j=1}^{n} \|v(t_{j-1}) - v(t_j)\|_{(H^s(\Omega)\cap V)^*} < \infty.$$

see for example [BP86, Chapter 1, Section 3.2.]. In this definition no equivalence classes are used.

We could obviously use this to make the definition of $P^{\beta}_{\bar{y}}$ sharper without loss of generality. However, it is not necessary and the presentation is clearer without it.

Proof. We only give a sketch of the proof as a relatively detailed proof is given above [Bar84, (5.34), p. 181]. The arguments there are essentially that $\|\partial_t \bar{p}_{\gamma_n}\|_{L^1(I,(H^s(\Omega)\cap V)^*)}$ is bounded independently of γ_n. We have $-\partial_t \bar{p}_{\gamma_n} = -A\bar{p}_{\gamma_n} - f'(\bar{y}_{\gamma_n})\bar{p}_{\gamma_n} + \bar{\eta}_{\gamma_n} + j_v'(\bar{y}_{\gamma_n}(T))$. The only difference to [Bar84] is that we have an additional term stemming from the non-linearity. All the terms are bounded in the given norm by the arguments in [Bar84, p. 181], except $f'(\bar{y}_{\gamma_n})\bar{p}_{\gamma_n}$. This term is bounded by the local Lipschitz continuity of f' in its third argument, Theorem 3.7 and Lemma 3.17. Now an application of Helly's theorem and the arguments in [Bar84, p. 181] yield the claim. □

Based on ideas of [Bar81, Theorem 2] we find:

Lemma 3.37 *Let $(\bar{p},\bar{\eta}) \in P^{\beta}_{\bar{y}}$. Assume that β_γ satisfies Assumption 3.33. We have*

$$\left(\bar{\lambda},\bar{p}\right)_{L^2(Q)} = 0.$$

Proof. For $\gamma > 0$ we have by Assumption 3.33

$$\begin{aligned} |(\bar{\lambda}_\gamma,\bar{p}_\gamma)_{L^2(Q)}| &\leq \int_Q |\beta_\gamma(\bar{y}_\gamma - \Psi)||\bar{p}_\gamma|\, d(t,x) \leq C\int_{\{\bar{y}_\gamma-\Psi<0\}} |\beta_\gamma{}'(\bar{y}_\gamma-\Psi)(\bar{y}_\gamma-\Psi)||\bar{p}_\gamma|\, d(t,x) \\ &= C\int_{\{\bar{y}_\gamma-\Psi<0\}} |\bar{\eta}_\gamma||\bar{y}_\gamma-\Psi|\, d(t,x) \leq C\|\bar{\eta}_\gamma\|_{L^1(Q)}\|\max(0,\bar{y}_\gamma-\Psi)\|_{L^\infty(Q)}. \end{aligned}$$

By Lemma 3.18, Theorem 3.7 and $\bar{y} \geq \Psi$ this upper bound converges to 0.

By Theorem 3.7 $(\lambda_\gamma)_{\gamma>0}$ converges weakly to $\bar\lambda$ in $L^2(Q)$. Along the sequence $(\gamma_n)_{n\in\mathbb{N}}$ that generates $(\bar p,\bar\eta)$ we have that $(\bar p_{\gamma_n})_{n\in\mathbb{N}}$ converges strongly in $L^2(Q)$ by Lemma 3.36. We thus find

$$(\bar\lambda,\bar p)_{L^2(Q)} = \lim_{n\to\infty}\left(\bar p_{\gamma_n},\bar\lambda_{\gamma_n}\right)_{L^2(Q)} = 0.$$

□

We now put all the previous results together and arrive at our first order optimality conditions.

Theorem 3.38 *Let $(\bar y,\bar u)$ be a globally optimal solution to* (OC)*. Assume $\Psi\in C(\bar Q)$ and let $s>\frac{N}{2}$. Then there exists a $(\bar p,\bar\eta)\in(L^2(I,V)\cap L^\infty(I,H)\cap BV(\bar I,(H^s(\Omega)\cap V)^*)\times M(\bar Q)$ such that*

- $S(\bar u)=\bar y$,
- *For all $\varphi\in W_0(I)\cap C(\bar Q)$ we have*

$$\begin{aligned}&(\bar p,\partial_t\varphi)_{L^2(I,V,V^*)} + a_I(\bar p,\varphi) + (f'(\bar y)\bar p,\varphi)_{L^2(Q)}\\ =&(\bar\eta,\varphi)_{M(\bar Q),C(\bar Q)} + (j_v'(\bar y),\varphi)_{L^2(Q)} + (j_T'(\bar y(T)),\varphi(T))_H,\end{aligned} \tag{3.15}$$

- *For all $u\in U_{ad}$ we have*

$$(\bar p+g'(\bar u),u-\bar u)_{L^2(Q)}\ge 0, \tag{3.16}$$

- $\operatorname{supp}(\bar\eta)\subset\{(t,x)\in\bar Q:\bar y(t,x)=\Psi(t,x)\}$ *and in particular* $(\bar y-\Psi,\bar\eta)_{C(\bar Q),M(\bar Q)}=0$,
- $(\bar\lambda,\bar p)_{L^2(Q)}=0$ *where* $\bar\lambda=\bar\lambda(\bar u)$.

There even exists a $C>0$, depending only on $(\bar y,\bar u)$ such that $\|\bar p\|_{L^2(I,V)\cap L^\infty(I,H)}\le C$ and $\|\bar\eta\|_{M(\bar Q)}\le C$.

Proof. This theorem is just the collection of Lemmas 3.29, 3.32, 3.36 and 3.37 and Corollary 3.31 for the first five points. Lemma 3.37 is applicable by choosing β according to Proposition 2.17 so that Proposition 3.34 implies Assumption 3.33. To see the norm bounds simply consider Lemma 3.22 and Lemma 3.24. □

Remark 3.39 One could extend this system to include $\bar\eta_*$ with the results mentioned in Remark 3.30 and the norm estimate from Lemma 3.24, but one can now see that the additional knowledge is limited as things like the complementarity conditions to not hold for $\bar\eta_*$.

Example 3.40 We provide an example where we can clearly see that the adjoint is in general discontinuous in time, in particular it is not in $W(I)$, and not unique.

For presentation's sake we choose $I:=(-1,1)$. A simple shift of $+1$ makes the example fit exactly into our setting. We also choose $N=2$ and $\Omega:=B_1(0)$ with the Dirichlet boundary $\Gamma_D=\partial B_1(0)$. As admissible set we choose $U_{ad}:=\{u\in L^\infty(Q): 0\le u\le 1 \text{ a.e. in } Q\}$. We choose the non-linearity $f=0$ and the operator $A:=-\Delta$. The control constraints have to be active on a non-zero set, as we will later see that there is some form of uniqueness of the adjoint and its multiplier on the set where the control constraints are inactive, cf. Theorem 3.46.

We choose the control $\bar u:=0\in U_{ad}$, the obstacle $\Psi:=0$ and state $\bar y:=S(0)=0$. Here the multiplier $\bar\lambda$ is equal to 0. As functionals we choose $j_T=0$, $j_v(y):=\frac{1}{2}\|y-y_Q\|^2_{L^2(Q)}$ and

$g(u) := \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}$ with $\alpha_g > 0$. Here $y_Q := -1 < 0$. As non-linearity we simply choose $f = 0$.

We now show that $(\bar{y}, \bar{u})$ are indeed optimal. For any control $0 \neq u \in U_{ad}$ we of course have that $\frac{1}{2}\|u\|_{L^2(Q)} > 0$. Because $y_Q \leq 0$ and $S(u) \geq \Psi = 0$, we also have $\frac{1}{2}\|S(u) - y_Q\|^2_{L^2(Q)} \geq \frac{1}{2}\|0 - y_Q\|^2_{L^2(Q)}$. Thus we find for any $u \in U_{ad}$

$$\frac{\alpha_g}{2}\|0\|^2_{L^2(Q)} + \frac{1}{2}\|0 - y_Q\|^2_{L^2(Q)} = J(\bar{u}, \bar{y}) < J(u, S(u)).$$

Hence, $(\bar{y}, \bar{u})$ is optimal.

To define the adjoint we let $q \in (1 + N/2, \infty)$ be arbitrary and introduce p_1 as the solution to the linear PDE

$$\begin{cases} -\partial_t p_1 - \Delta p_1 = 0 \text{ on } (-1,0) \times \Omega \\ p_1(0) = p_0, \quad p_1|_{\Sigma_D} = 0. \end{cases} \tag{3.17}$$

Here $0 \geq p_0 \in \mathbb{W}_q$ is a starting value, which we specify later. We also define p_2 as the solution to

$$\begin{cases} -\partial_t p_2 - \Delta p_2 = -y_Q \text{ on } Q \\ p_2(1) = 0, \quad p_2|_{\Sigma_D} = 0. \end{cases} \tag{3.18}$$

By Theorem 8.17, after changing the sign in the time component, we have $p_1, p_2 \in W^{1,2}_q(Q)$.

We now define the adjoint as $\bar{p} := p_1 \cdot 1_{(-1,0)} + p_2$. It is easy to see that $\bar{p} \in L^2(I, V) \cap L^\infty(I, H)$. To see the $BV(\bar{I}, (H^s(\Omega) \cap V)^*)$-regularity, for any $s > \frac{N}{2}$, we show that $\operatorname{Var}(\bar{p}, L^1(\Omega)) < \infty$. By [DNPV12, Theorem 8.2] we have $H^s(\Omega) \subset L^\infty(\Omega)$, so $\operatorname{Var}(\bar{p}, (H^s(\Omega) \cap V)^*)) \leq C \operatorname{Var}(\bar{p}, L^1(\Omega)) < \infty$. Note that we have to prove this for any fixed representative of $\bar{p}$, which we shall also call $\bar{p}$. Its easy to see from the definition of Var that

$$\operatorname{Var}(\bar{p}, L^1(\Omega)) \leq \operatorname{Var}(p_2, L^1(\Omega)) + \operatorname{Var}(p_1 \cdot 1_{(-1,0)}, L^1(\Omega)).$$

By the $W^{1,2}_q(Q)$-regularity of p_2 we immediately find $\operatorname{Var}(p_2, L^1(\Omega)) \leq C\|p_2\|_{W^{1,2}_q(Q)}$. To estimate $\operatorname{Var}(p_1 \cdot 1_{(-1,0)}, L^1(\Omega))$ let $-1 = t_0 < t_1 < \cdots < t_M = 1$. Let $\hat{j} \in \{0, 1, \ldots, M-1\}$ such that $t_{\hat{j}} < 0 < t_{\hat{j}+1}$. Using the $W^{1,2}_q(Q)$ regularity of p_1 we find

$$\begin{aligned}
&\sum_{j=1}^{M} \|p_1(t_j) \cdot 1_{(-1,0)}(t_j) - p_1(t_{j-1}) \cdot 1_{(-1,0)}(t_{j-1})\|_{L^1(\Omega)} \\
&= \sum_{j=1}^{\hat{j}} \|p_1(t_j) - p_1(t_{j-1})\|_{L^1(\Omega)} + \|p_1(t_{\hat{k}})\|_{L^1(\Omega)} \\
&\leq \sum_{j=1}^{\hat{j}} \int_{t_{j-1}}^{t_j} \|\partial_t p_1\|_{L^1(\Omega)}\, dt + \int_{t_{\hat{j}}}^{0} \|\partial_t p_1\|_{L^1(\Omega)}\, dt + \|p_1(0)\|_{L^1(\Omega)} \\
&\leq \|\partial_t p_1\|_{L^1(Q)} + \|p_0\|_{L^1(\Omega)}.
\end{aligned}$$

Thus $\operatorname{Var}(p_1, L^1(\Omega)) \leq C\|p_1\|_{W^{1,2}_q(Q)} + \|p_0\|_{L^1(\Omega)}$.

We define the corresponding multiplier $\bar{\eta} := p_0 \delta_0 \in M(\bar{Q})$, where $\delta_0 = \delta_0(t)$ is the Dirac measure in time at the point 0 and p_0 has to be interpreted as a density of the Lebesgue measure.

Now we show that all conditions from Theorem 3.38 are satisfied. All points, except the second and third one, are seen immediately. We now show that

$$(\bar{p}, \partial_t \varphi)_{L^2(I,V,V^*)} + a_I(\bar{p}, \varphi) = (\bar{\eta}, \varphi)_{M(\bar{Q}),C(\bar{Q})} + (j'_v(\bar{y}), \varphi)_{L^2(Q)} \tag{3.19}$$

holds for any $\varphi \in W(I) \cap C(\bar{Q})$ with $\varphi(-1) = 0$. Let φ be of this form and let us compute:

$$\begin{aligned}
(\bar{p}, \partial_t \varphi)_{L^2(I,V,V^*)} + a_I(\bar{p}, \varphi) &= \int_{-1}^{0} (p_1, \partial_t \varphi)_{V,V^*} + a_\Omega(p_1, \varphi)\, dt \\
&\quad + \int_{-1}^{1} (p_2, \partial_t \varphi)_{V,V^*} + a_\Omega(p_2, \varphi)\, dt \\
&= \int_{-1}^{0} -(\partial_t p_1, \varphi)_{V,V^*} + a_\Omega(p_1, \varphi)\, dt + (p_1, \varphi)_H|_{-1}^{0} \\
&\quad + \int_{-1}^{1} -(\partial_t p_2, \varphi)_{V,V^*} + a_\Omega(p_2, \varphi)\, dt + (p_2, \varphi)_H|_{-1}^{1} \\
&= (p_0, \varphi(0))_H + (-y_Q, \varphi)_{L^2(Q)}.
\end{aligned}$$

By construction we have $j_v'(\bar{y}) = 0 - y_Q = -y_Q$ and thus (3.19) is satisfied.

Lastly we have to show that the third point of Theorem 3.38 is satisfied. Because of the structures of g and U_{ad} this condition is equivalent to

$$\bar{u} = P_{[0,1]}\left(-\frac{1}{\alpha_g}\bar{p}\right).$$

To see that, inspect the proof of the later Corollary 3.41. Thus, if we can show that $\bar{p} \geq 0$, we have proven that all conditions of Theorem 3.38 are met. If $t \in (0,1)$ we have $\bar{p}(t) = p_2(t)$. By Theorem 8.15 we have $p_2(t) \geq 0$ and thus $\bar{p}(t) \geq 0$ almost everywhere in Ω. Note that $\bar{p}$ satisfies

$$\begin{cases} -\partial_t \bar{p} + A\bar{p} = -y_Q \text{ in } (-1,0) \times \Omega \\ \bar{p}(0) = p_0 + p_2(0), \quad \bar{p}|_{\Sigma_D} = 0. \end{cases}$$

By Theorem 8.15 we have $p_2(0) \geq 0$ and, since $y_Q = -1 < 0$, we also have $p_2(0) \neq 0$. We will show $p_2(0) \in \mathbb{W}_q$ and choose $p_0 \in -p_2(0) \cdot (0,1) \leq 0$ to see that $p_0 + p_2(0) \geq 0$. Thus Theorem 8.15 again delivers $\bar{p} \geq 0$ on $(-1,0) \times \Omega$. To see $p_2(0) \in \mathbb{W}_q$ one can use the trace method for interpolation spaces, see Section 8.6, and immediately deduce via the definition that

$$\|p_2(0)\|_{\mathbb{W}_q} \leq C \|p_2\|_{W_q^{1,2}(Q)}.$$

Thus we have shown $\bar{p} \geq 0$ and conclude the proof. We also see here that p_0 and thus $\bar{p}$ and $\bar{\eta} = p_0 \delta_0$ are not uniquely determined by the properties of Theorem 3.38.

3.4.2 Properties of Adjoints and Multipliers

We can now deduce a few properties from this stationarity system. Note that this system was obtained via regularization, but its final form in Theorem 3.38 has nothing to do with the regularization anymore. The following statements thus hold true without any knowledge about the regularization.

A famous consequence is the relation of control and adjoint in the case of box constraints. For this corollary the assumption $\Psi \in C(\bar{Q})$ is not needed.

Corollary 3.41 *Let $\bar{u}$ and $\bar{p}$ satisfy* (3.16) *and assume g is given by*

$$g(u) = \int_Q \varphi(u)\, d(t,x) \tag{3.20}$$

with φ strictly convex and differentiable. Also assume U_{ad} is given as a box, bounded from below and above by $u_l, u_u \in L^{q_u}(Q)$. Then we have

$$\bar{u} = P_{[u_l,u_u]}[-(\varphi')^{-1}(\bar{p})].$$

Proof. The proof is standard, e.g. [Trö09, Lemma 2.26], so we only give a sketch. By (3.16) we get almost everywhere in Q

$$\bar{u}(t,x) \begin{cases} = u_l(t,x) & \text{if } \bar{p}(t,x) + \varphi'(\bar{u}(t,x)) > 0, \\ \in [u_l(t,x), u_u(t,x)] & \text{if } \bar{p}(t,x) + \varphi'(\bar{u}(t,x)) = 0, \\ = u_u(t,x) & \text{if } \bar{p}(t,x) + \varphi'(\bar{u}(t,x)) < 0. \end{cases} \tag{3.21}$$

All the following arguments are to be interpreted to hold almost everywhere in Q.

The function φ is strictly convex and thus its derivative is strictly monotonically increasing and invertible. Assume we have at some point in Q

$$(\varphi')^{-1}(-\bar{p}) > u_u \geq \bar{u}.$$

Applying φ' to both sides implies

$$\bar{p} + \varphi'(\bar{u}) < 0.$$

Thus by (3.21) we have here $\bar{u} = u_u$.

The same line of arguing implies $\bar{u} = u_l$ whenever $(\varphi')^{-1}(-\bar{p}) < u_l$. For the remaining case, the middle portion of (3.21) implies the projection formula directly by rearranging terms. □

Remark 3.42 Under the assumptions on g by Corollary 3.41 we also can deduce some minimum principles. The formulation in (3.21) implies the weak minimum principle directly. Almost everywhere in Q we have

$$\min_{v\in[u_l,u_u]} (\bar{p} + \varphi'(\bar{u}))\, v = (\bar{p} + \varphi'(\bar{u}))\, \bar{u}.$$

Analogously to the proof of [Trö09, Theorem 2.27] we obtain the minimum principle:

$$\min_{v\in[u_l,u_u]} \bar{p}v + \varphi(v) = \bar{p}\bar{u} + \varphi(\bar{u}).$$

Remark 3.43 An important consequence of the projection formula from Corollary 3.41 is the fact that the regularity of $\bar{p}$ transfers to $\bar{u}$, provided u_l, u_u are regular enough. If for example $u_l, u_u \in L^{q_u}(Q) \cap L^2(I,V)$ we have $\bar{u} \in L^2(I,V)$. Taking the positive part, and thus projection onto sufficiently regular box constraints, preserves V-regularity, see Proposition 8.19.

As we have hinted at, here and there, the multipliers are not unique. Yet, we can show that a given adjoint $\bar{p}$ determines its multiplier $\bar{\eta}$ and vice versa, to an extent.

Theorem 3.44 *For any $(\bar{p},\bar{\eta}), (\tilde{p},\tilde{\eta}) \in L^2(I,V) \times M(\bar{Q})$ satisfying (3.15) we have*

- *$\bar{p} = \tilde{p}$ in $L^2(I,V)$ implies $\bar{\eta} = \tilde{\eta}$ in $M(Q \cup \Sigma_N) \subset M(\bar{Q})$,*
- *$\bar{\eta} = \tilde{\eta}$ in $M(\bar{Q})$ implies $\bar{p} = \tilde{p}$ in $L^2(Q)$.*

Proof. Let $(\bar{p}, \bar{\eta}), (\tilde{p}, \tilde{\eta})$ be as above.

First assume that $\bar{p} = \tilde{p}$. Let $\varphi \in C_c^\infty(Q \cup \Sigma_N)$. Then by (3.15) have:

$$(\bar{\eta} - \tilde{\eta}, \varphi)_{M(Q\cup\Sigma_N),C(Q\cup\Sigma_N)} = (\bar{\eta} - \tilde{\eta}, \varphi)_{M(\bar{Q}),C(\bar{Q})} = 0.$$

By density of $C_c^\infty(Q \cup \Sigma_N)$ in $C_0(Q \cup \Sigma_N)$ and the representation theorem, see for example [Rud74, Theorem 6.17], this implies $\bar{\eta} = \tilde{\eta}$ in $M(Q \cup \Sigma_N)$.

Now assume $\bar{\eta} = \tilde{\eta}$ in $M(\bar{Q})$. Then for any $\varphi \in W_0(I) \cap C(\bar{Q})$ we have

$$(\bar{p} - \tilde{p}, \partial_t \varphi)_{L^2(I,V,V^*)} + (A(\bar{p} - \tilde{p}), \varphi)_{L^2(I,V^*,V)} + (f'(\bar{y})(\bar{p} - \tilde{p}), \varphi)_{L^2(Q)} = 0.$$

This implies

$$(\bar{p} - \tilde{p}, \partial_t \varphi + A\varphi)_{L^2(I,V,V^*)} + \left(\bar{p} - \tilde{p}, f'(\bar{y})\varphi\right)_{L^2(Q)} = 0.$$

Now let $v \in C_c^\infty(Q)$ be arbitrary. Let φ be the solution to

$$\begin{cases} \partial_t \varphi + A\varphi + f'(\bar{y}) = v, \\ \varphi|_{\Sigma_D} = 0, \quad \varphi(0) = 0. \end{cases} \tag{3.22}$$

By Theorem 8.20 this solution satisfies $\varphi \in W(I) \cap C(\bar{Q})$. Thus $\varphi_0 \in W_0(I) \cap C(\bar{Q})$. We therefore have for any $v \in C_c^\infty(Q)$

$$(\bar{p} - \tilde{p}, v)_{L^2(Q)} = 0.$$

Using the density of $C_c^\infty(Q)$ in $L^2(Q)$ we have $\bar{p} = \tilde{p}$. □

Remark 3.45 We see that in the first case of Theorem 3.44 the measures do not necessarily agree on $\bar{Q} \setminus (Q \cup \Sigma_N) = \Sigma_D$. One could force uniqueness by for example forcing any $\bar{\eta}$ appearing as a component in $P_{\bar{y}}^\beta$ to be 0 on Σ_D, cf. [CV19].

An immediate consequence of Corollary 3.41 is the following result.

Theorem 3.46 *Let $U_{ad} = [u_l, u_u]$ with $u_l, u_u \in L^{q_u}(Q)$. Assume g has the form of* (3.20) *with φ being strictly convex and differentiable. Let $Q_l := \left\{(t,x) \in \bar{Q} : \bar{u}(t,x) = u_l(t,x)\right\}$. It is defined up to a set of Lebesgue measure zero. Technically the set has to be defined for two fixed representatives of $\bar{u}$ and u_l. Analogously we define Q_u for u_u.*

For any $(\bar{p}, \bar{\eta}), (\tilde{p}, \tilde{\eta}) \in L^2(I, V) \times M(\bar{Q})$ satisfying (3.15) *we have $\bar{p} = \tilde{p}$ a.e. on $\bar{Q} \setminus (Q_l \cup Q_u)$. We also have $\bar{\eta}|_{\bar{Q}\setminus(Q_l\cup Q_u)} = \tilde{\eta}|_{\bar{Q}\setminus(Q_l\cup Q_u)}$.*

So in a certain sense the multipliers are “unique” where the control $\bar{u}$ is “interesting” and not just touching the box constraints.

Proof. The following statements are understood to be true almost everywhere. Let $(t,x) \in \bar{Q} \setminus (Q_l \cup Q_u)$. By Corollary 3.41 we have

$$-(\varphi')^{-1}(\tilde{p}(t,x)) = \bar{u}(t,x) = -(\varphi')^{-1}(\bar{p}(t,x)).$$

By the invertibility of φ' this yields the desired equality $\tilde{p}(t,x) = \bar{p}(t,x)$.

Now we have by (3.15) for any $v \in C_c^\infty(\bar{Q} \setminus (Q_l \cup Q_u))$:

$$\begin{aligned}(\bar{\eta}, v)_{M(\bar{Q}),C(\bar{Q})} &= (\bar{p}, \partial_t v + Av + f'(\bar{y})\bar{p} - j'(\bar{y}), v)_{L^2(Q)} - (j_T'(\bar{y}(T)), v(T))_H \\ &= (\tilde{p}, \partial_t v + Av + f'(\bar{y})\tilde{p} - j'(\bar{y}), v)_{L^2(Q)} - (j_T'(\bar{y}(T)), v(T))_H = (\tilde{\eta}, v)_{M(\bar{Q}),C(\bar{Q})}.\end{aligned}$$

By the density of $C_c^\infty(\bar{Q} \setminus (Q_l \cup Q_u))$ in $C_0(\bar{Q} \setminus (Q_l \cup Q_u))$ we have $\bar{\eta} = \tilde{\eta}$ as functionals on $C_0(\bar{Q} \setminus (Q_l \cup Q_u))$. The Riesz representation theorem, e.g. [Rud74, Theorem 6.19], now delivers the desired equality as measures. □

Under relatively mild assumptions, we can prove that $\bar{p}$ is bounded in $L^\infty(Q)$. Under these additional assumptions one could strengthen the stationarity system from Theorem 3.38 to include the regularity $\bar{p} \in L^\infty(Q)$ as a condition.

Theorem 3.47 *Let $\tilde{q_u} \in (1 + N/2, \infty)$. Assume j_v' is continuous as an operator $L^{\tilde{q_u}}(Q) \to L^{\tilde{q_u}}(Q)$. Also let $j_T = 0$ and g be of the form from Corollary 3.41.*

Let $(\bar{p}, \bar{\eta}) \in P_{\bar{y}}^\beta$ be generated by $(\gamma_n)_{n\in\mathbb{N}}$. We then have

$$\bar{p}_{\gamma_n} \xrightarrow{n\to\infty} \bar{p} \text{ weak* in } L^\infty(Q),$$
$$\bar{u}_{\gamma_n} \xrightarrow{n\to\infty} \bar{u} \text{ weak* in } L^\infty(Q)$$

and the estimates

$$\|\bar{p}\|_{L^\infty(Q)}, \|\bar{u}\|_{L^\infty(Q)} \le C\|j_v'(\bar{y})\|_{L^{\tilde{q_u}}(Q)}.$$

The constant $C > 0$ does not depend on U_{ad}.

Proof. By Theorem 8.22 we find for $\tilde{q_u} > 1 + N/2$ that

$$\|\bar{p}_{\gamma_n}\|_{L^\infty(Q)} \le C\|j_v'(\bar{y}_{\gamma_n})\|_{L^{\tilde{q_u}}(Q)} \tag{3.23}$$

where C does not depend on f, $\beta_\gamma{}'$ or U_{ad}.

Since $(\bar{y}_{\gamma_n})_{n\in\mathbb{N}}$ converges uniformly to $\bar{y}$ by Theorem 3.7 the right hand side stays bounded and even converges to $\|j_v'(\bar{y})\|_{L^{\tilde{q_u}}(Q)}$. Therefore there is a $C > 0$ independent of n such that

$$\|\bar{p}_{\gamma_n}\|_{L^\infty(Q)} \le C.$$

Because $L^1(Q)^* \simeq L^\infty(Q)$ we have that a subsequence of $(\bar{p}_{\gamma_n})_{n\in\mathbb{N}}$ converges weakly* in $L^\infty(Q)$ to some limit, cf. the arguments in the proof of Lemma 3.22. The sequence $(\bar{p}_{\gamma_n})_{n\in\mathbb{N}}$ also converges weakly in $L^2(Q) \supset L^2(I, V)$ to $\bar{p}$ by assumption. This entails that this subsequence converges weakly* in $L^\infty(Q)$ to $\bar{p}$. Because the limit is unique basic calculus arguments prove that the whole sequence converges to $\bar{p}$ weakly* in $L^\infty(Q)$. Because norms are weakly* lower semi-continuous we can conclude the claimed norm bound from (3.23).

The norm bounds for $\bar{u}$ immediately follow from Corollary 3.41 and those for $\bar{p}$. The weak* convergence of the controls follows from the norm bound, as for $(\bar{p}_{\gamma_n})_{n\in\mathbb{N}}$, and their $L^2(Q)$-convergence from Theorem 3.7. □

Theorem 3.48 *Assume the obstacle satisfies $\partial_t\Psi + A\Psi + f(\Psi) \le 0$ and that $\bar{y}, \Psi \in L^2(I, H^2(\Omega))$. Then*

$$\bar{u} \le 0 \text{ almost everywhere on } \{(t, x) \in Q : \bar{y}(t, x) = \Psi(t, x)\}.$$

The given set has to interpreted for a fixed representative of Ψ. For two different representatives of Ψ the corresponding sets differ only on a set of Lebesgue measure 0.

If additionally the control cost term g and U_{ad} are of the form of Corollary 3.41 with φ satisfying $\varphi'(0) = 0$ and U_{ad} satisfying $0 < u_u$ almost everywhere, we have the following: let $\bar{p} \in L^2(I, V)$ satisfy (3.16). Then

$$\bar{p} \geq 0 \text{ almost everywhere on } \{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}.$$

Remark 3.49 Technically $\partial_t \Psi + A\Psi + f(\Psi) \leq 0$ and $0 < u_u$ could be just assumed to hold on $\{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}$. This would be a very technical and unnatural condition.

The regularity assumptions on $\bar{y}$ and Ψ are satisfied on $C^{1,1}$-domains and sufficiently nice polygonal domains by Definition 2.9, Theorem 2.40 and Theorem 2.42, provided they satisfy proper boundary conditions.

The other conditions of Theorem 3.48 are satisfied, e.g. if

- $U_{ad} = [-a, a]$ for some $a > 0$,
- $g(u) = \frac{\alpha_g}{2} \|u\|^2_{L^2(Q)}$ for some $\alpha_g > 0$,
- f satisfies the required properties and additionally $f(0) = 0$,
- $\Psi = -c$ for some constant $c > \|y_0\|_{L^\infty(\Omega)}$.

More generally one could also consider functions $0 \geq \Psi_{source} \in L^{q_u}(Q)$, $0 \geq \Psi_0 \in \mathbb{W}_{q_u}$ and the obstacle Ψ as the solution of

$$\begin{cases} \partial_t \Psi + A\Psi + f(\Psi) = \Psi_{source}, \\ \Psi(0) = \Psi_0, \quad \Psi|_{\Sigma_D} = 0. \end{cases}$$

By the maximum principle, cf. Theorem 8.15, we find $\Psi \leq 0$. Depending on the regularity of the domain the obstacle Ψ has different regularity. Checking the results in Section 8.4 we can see that Ψ satisfies Assumption 2.5.

Proof of Theorem 3.48. All of the following arguments are assumed to hold almost everywhere. On $\{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}$ we have by Lemma 8.14 that

$$\bar{u} \leq \bar{u} + \lambda(\bar{u}) = \partial_t \bar{y} + A\bar{y} + f(\bar{y}) = \partial_t \Psi + A\Psi + f(\Psi) \leq 0, \tag{3.24}$$

as $\bar{\lambda}(\bar{u})$ is non-negative.

Under the additional assumptions we have by Corollary 3.41 that on $\{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}$ there holds

$$u_u > 0 \geq \bar{u} = P_{u_l,u_u}\left[(\varphi')^{-1}(-\bar{p})\right].$$

Thus we have $0 \geq (\varphi')^{-1}(-\bar{p})$ on $\{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}$. This implies by the monotonicity of φ'

$$-\varphi'(0) \leq \bar{p}$$

on the set in question. As we assumed $\varphi'(0) = 0$ we conclude the proof. □

Remark 3.50 We now obtain the following, intuitive observation: if Ψ evolves negatively, i.e. it becomes drop further and further below $y_0 \geq \Psi(0)$, and if $\bar{u} > 0$, i.e. the state constantly moves upwards from y_0, obstacle and state never touch. Therefore in this situation the active set $\{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}$ is empty.

In a nutshell: if, in the situation of Theorem 3.48, we have $u_l > 0$, we immediately find $\{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\} = \emptyset$.

We also make the following observation which is akin to [CW19, Theorem 5.1].

Theorem 3.51 *Assume the obstacle satisfies* $\partial_t\Psi + A\Psi + f(\Psi) = 0$ *and that* $\bar{y}, \Psi \in L^2(I, H^2(\Omega))$. *Assume that* $g(u) = \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}$ *for some* $\alpha_g > 0$. *Then*

$$\bar{u} = 0 \text{ almost everywhere on } \{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\} \text{ and } \lambda = 0.$$

The given set has to interpreted for a fixed representative of Ψ. *For two different representatives of* Ψ *the corresponding sets differ on a set of Lebesgue measure* 0.

If the admissible set U_{ad} *is of the form of Corollary 3.41 with* $u_l < 0 < u_u$ *almost everywhere, then the following holds: let* $\bar{p} \in L^2(I, V)$ *satisfy* (3.16). *Then*

$$\bar{p} = 0 \text{ almost everywhere on } \{(t,x) \in Q : \bar{y}(t,x) = \Psi(t,x)\}.$$

Proof. By (3.24) of the previous Theorem 3.48 we find $\bar{u} + \lambda(\bar{u}) = 0$ on $\{\bar{y} = \Psi\}$. Here we abbreviated the active set. Now consider

$$\hat{u} := \begin{cases} \bar{u} & \text{on } \{\bar{y} > \Psi\}, \\ 0 & \text{on } \{\bar{y} = \Psi\}, \end{cases} \quad \text{and } \hat{\lambda} = 0.$$

We see that

$$\partial_t \bar{y} + A\bar{y} + f(\bar{y}) = \bar{u} + \bar{\lambda}(\bar{u}) = \hat{u} + \hat{\lambda}. \tag{3.25}$$

Here we used that $\lambda(\bar{u}) = 0$ a.e. in $\{\bar{y} > \Psi\}$ by Remark 2.31 and $\bar{u} + \lambda(\bar{u}) = 0$ a.e. in $\{\bar{y} = \Psi\}$. Hence $S(\hat{y}) = \bar{y}$ with the multiplier $\hat{u} = \lambda(\hat{u}) = 0$ by Theorem 2.33. By construction and Theorem 3.48 we have $\bar{u} \leq \hat{u}$ and find

$$J(\bar{y}, \bar{u}) \geq J(\bar{y}, \hat{u}) \geq J(\bar{y}, \bar{u}),$$

which entails $g(\bar{u}) = g(\hat{u})$ and therefore $\|\bar{u}\|_{L^2(Q)} = \|\hat{u}\|_{L^2(Q)}$. Thus, we have by construction $\bar{u} = \hat{u}$. In particular we have $\bar{u} = 0$ on $\{\bar{y} = \Psi\}$.

The claim on $\lambda(\bar{u})$ follows from the arguments below Equation (3.25).

Under the additional assumptions the projection formula from Corollary 3.41 implies the claim on $\bar{p}$ similar to the arguments in Theorem 3.48. ☐

3.4.3 Equivalence to State Constrained Problems

In this section we will only scratch the surface of showing the equivalence between (OC) and a state constrained problem. The developments in this area are pretty recent, see for example [CW19] for the discussion of this equivalence in the elliptic case and [CV19] for a similar discussion in the parabolic case. The usefulness of this approach cannot be understated, as (OC) is a non-linear problem (even for $f = 0$), while state constrained problems are (for $f = 0$ at least) typically convex. This allows us to characterize solutions to either type of problem in two ways, yielding stronger optimality conditions. Extending those results and possibly including the non-linearity f will be subject to future research.

As this discussion has been too recent to include fully in this thesis, we give a sketch for a special case, that will be used in our numerical examples, see Section 7.3.1. The articles are very general and technical, so for simplicity we give a self contained proof for a special case based on the general ideas of the aforementioned articles. We will use constant obstacles, while the works mentioned above use conditions like $-\Delta\Psi \leq 0$, respectively, $\partial_t\Psi - \Delta\Psi \leq 0$, cf. Theorem 3.48.

Theorem 3.52 *Assume Ψ is a negative constant smaller or equal to $\|y_0\|_{L^\infty(\Omega)}$, $f = 0$, and that Ω is a $C^{1,1}$-domain with $\Gamma_D = \partial\Omega$. Further assume that $g(u) = \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}$ for some $\alpha_g > 0$. Then $\tilde{u} \in U_{ad}$ is a solution to*

$$\min_{u \in U_{ad}} J(y,u) \ s.th. \ y = S(u) \tag{OC}$$

iff it is a solution to

$$\min_{u \in U_{ad}} J(y,u) \ s.th. \ y = S_{lin}(u) \ and \ y \geq \Psi. \tag{3.26}$$

Here S_{lin} is the solution operator to the linear PDE

$$\begin{cases} \partial_t y + Ay = u \ in \ Q, \\ y(0) = y_0, \ y|_{\Sigma_D} = 0. \end{cases}$$

For the well-definedness of S_{lin} see Theorem 8.17.

Proof. By Theorem 3.51 we immediately find that

$$\min_{u \in U_{ad}} J(y,u) \text{ s.th. } y = S(u)$$

is equivalent to

$$\min_{u \in U_{ad}} J(y,u) \text{ s.th. } y = S(u), \ u = 0 \text{ on } \{y = \Psi\}. \tag{3.27}$$

Here, $\{y = \Psi\}$ is an abbreviation of the set appearing in Theorem 3.51. Theorem 3.51 is applicable, because $S(u) \in L^2(I, H^2(\Omega))$ by the regularity of Ω, see Theorem 2.40.

We claim that $(y,u) \in S(U_{ad}) \times U_{ad}$ satisfies

$$y = S(u) \text{ and } u = 0 \text{ on } \{y = \Psi\} \tag{3.28}$$

iff

$$y = S_{lin}(u) \text{ and } y \geq \Psi. \tag{3.29}$$

Let (3.28) be satisfied. Then the second condition of (3.29) is clearly satisfied. We also have by Remark 2.31 that $\partial_t y + Ay = u$ almost everywhere in $\{y > \Psi\} := Q \setminus \{y = \Psi\}$. Lastly, since Ψ is constant, we find $u = 0 = \partial_t y + Ay$ on $\{y = \Psi\}$ by Lemma 8.14. The boundary and initial conditions are obviously satisfied. So in total, indeed, $S_{lin}(u) = y$.

To see that (3.29) implies (3.28) we define $\lambda = 0$ and see that

$$\partial_t y + Ay = u + \lambda \text{ a.e. in } Q$$

and

$$(\lambda, y - \Psi)_{L^2(Q)} = 0.$$

Now Theorem 2.33 implies that $y = S(u)$. To see the claim on the support of u we again use that Ψ is constant and find $u = \partial_t y + Ay = 0$ on $\{y = \Psi\}$.

Now, (3.26) and (3.27) have the same constraints, yielding the equivalence. □

Corollary 3.53 *Let the situation be as in Theorem 3.52. Additionally assume that j_v and j_T are convex. Assume that $(\hat{y}, \hat{u}) \in S(U_{ad}) \times U_{ad}$ satisfies*

$$(\hat{p} + \alpha_g \hat{u}, u - \hat{u})_{L^2(Q)} \geq 0 \qquad \forall u \in \{u \in U_{ad} : S_{lin}(u) \geq \Psi\}. \tag{3.30}$$

Here $\hat{p}$ is the solution to

$$\begin{cases} -\partial_t \hat{p} + A\hat{p} = j_v'(\hat{y}) \text{ in } Q, \\ \hat{p}(T) = j_T(\hat{y}(T)), \quad \hat{p}|_{\Sigma_D} = 0. \end{cases}$$

This is well-defined by [Wlo92, Theorem 26.1].

Then $(\hat{y}, \hat{u})$ is an optimal solution to (OC).

Proof. Because S_{lin} is linear (3.30) is just the necessary and sufficient optimality condition for the convex state constrained problem (3.26) in Theorem 3.52, see for example the calculations in [Trö09, Chapter 3.6]. Thus, by Theorem 3.52 we find that $(\hat{y}, \hat{u})$ is optimal for (OC). □

3.5 Second Order Sufficient Conditions for Unregularized Control Problems

We now transfer ideas and results for the elliptic case from [KW12b] to our parabolic situation. We see that with those arguments it is possible, under specific assumptions, to prove local quadratic growth of the functional for optimizers. This can be done without stumbling upon famous norm gap problems, where the growth of the functional happens in a norm different from the one that the local neighborhood is determined in, e.g. [Trö09, Chapters 4,5]. Interestingly the two norm gap seems unavoidable when discussing a certain type of semi-linear, semi-smooth parabolic PDE, see for example [Bet19]. Also see [BM15] where second order assumptions are made in the field of plasticity modelling with VIs.

Assumption 3.54 *Let $(\bar{y}, \bar{u})$ be an admissible pair for* (OC).

1. *$j_v : L^2(Q) \to \mathbb{R}$ is twice continuously Frechét differentiable,*
2. *$j_T : L^2(\Omega) \to \mathbb{R}$ is twice continuously Frechét differentiable,*
3. *$g : L^2(Q) \to \mathbb{R}$ is twice continuously Frechét differentiable.*

Additionally the following properties shall be satisfied:

4. *There exists a $\nu_g > 0$ such that $g''(\bar{u})$ is elliptic with ellipticity constant ν_g.*
5. *We have one of the following:*
   5.1 *f is linear and $j_v''(\bar{y})v^2 + j_T''(\bar{y}(T))v(T)^2 \geq 0$ for all $v \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \cap W_0(I)$,*
   5.2 *$N \leq 3$, $f''(t, x, \cdot)$ exists, is uniformly continous for all $(x, y) \in Q$, satisfies $f''(\bar{y}) \in L^\infty(Q)$ and*
   $$j_v''(\bar{y})v^2 + j_T''(\bar{y}(T))v(T)^2 - \int_Q f''(\bar{y})\bar{p}v^2 \, d(t, x) \geq 0 \qquad \forall v \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \cap W_0(I).$$

6. *Assume $\Psi \in C(\bar{Q})$. Let $(\bar{\lambda}, \bar{p}, \bar{\eta})$ be a triple of adjoint and multipliers satisfying the necessary optimality conditions from Theorem 3.38 together with $(\bar{y}, \bar{u})$. As additional assumptions we assume there exists a $\tau > 0$ such that*

$$\bar{p} \geq 0 \text{ on } \{(t,x) \in Q : \Psi(t,x) \leq \bar{y}(t,x) < \Psi(t,x) + \tau\} \text{ and } \bar{\eta} \leq 0 \text{ as a measure on } \bar{Q}.$$

Remark 3.55 Assumption 3.54.1-Assumption 3.54.5 are satisfied for quadratic cost functionals, which may or may not have an additional linear part, provided f is linear.

By Theorem 3.48 we already know that under common conditions we have $\bar{p} \geq 0$ on the active set. So, Assumption 3.54.6 can be seen as a tightening of that condition.

Under assumptions we can even get $\bar{p} \geq 0$ globally. The following idea is a modification of [Mig76, Théorème 4.1] and [KW12b, Remark 2.13] to the parabolic case. The following lemma is applicable for the case of quadratic functionals of the form

$$\frac{\alpha_Q}{2}\|y - y_Q\|^2_{L^2(Q)}, \qquad \frac{\alpha_\Omega}{2}\|y(T) - y_T\|^2_{L^2(\Omega)} \tag{3.31}$$

if the desired states $y_Q \in L^2(Q)$, $y_T \in L^2(\Omega)$ satisfy $y_Q \leq \Psi$, $y_T \leq \Psi(T)$.

Lemma 3.56 *Assume that $(\bar{p}, \bar{\eta}) \in P^{\beta}_{\bar{y}}$ is generated by $(\gamma_n)_{n\in\mathbb{N}}$ and $j'_v(\bar{y}), j'_T(\bar{y}(T)) \geq 0$ almost everywhere as $L^2(Q)$-/$L^2(\Omega)$-functions. Then we have $\bar{p} \geq 0$ and $\bar{\eta} \leq 0$ as a measure on Q.*

Under these assumption one could thus strengthen the optimality system in Theorem 3.38 to include those sign conditions as restrictions on $(\bar{p}, \bar{\eta})$.

Proof. Let $((\bar{p}_{\gamma_n}, \bar{\eta}_{\gamma_n}))_{n\in\mathbb{N}}$ be the sequence that converges to $(\bar{p}, \bar{\eta})$ in the sense of Definition 3.26. We first prove the sign of $\bar{p}$. We decompose $\bar{p}_{\gamma_n} = \bar{p}^+_{\gamma_n} - \bar{p}^-_{\gamma_n}$ into its positive and negative parts. Since taking those parts preserves V-regularity, cf. Proposition 8.19, we can test the regularized adjoint equation with $-\bar{p}^-_{\gamma_n}$ to get

$$\begin{aligned}&\left(j'_v(\bar{y}_{\gamma_n}), -\bar{p}^-_{\gamma_n}\right)_{L^2(Q)} \\ &\quad = \left(-\partial_t \bar{p}_{\gamma_n}, \bar{p}^-_{\gamma_n}\right)_{L^2(Q)} + a(\bar{p}_{\gamma_n}, -\bar{p}^-_{\gamma_n}) + \left(\left(f'(\bar{y}_{\gamma_n}) + \beta'_{\gamma_n}(\bar{y}_{\gamma_n} - \Psi)\right)\bar{p}_{\gamma_n}, -\bar{p}^-_{\gamma_n}\right)_{L^2(Q)}.\end{aligned}$$

By Proposition 8.19 the supports of $\nabla \bar{p}^-_{\gamma_n}$ and $\nabla \bar{p}^+_{\gamma_n}$ are disjoint. This implies

$$a(\bar{p}_{\gamma_n}, -\bar{p}^-_{\gamma_n}) = a(-\bar{p}^-_{\gamma_n}, -\bar{p}^-_{\gamma_n}) \geq \nu_{\text{ell}}\|\bar{p}^-_{\gamma_n}\|^2_{L^2(I,V)}. \tag{3.32}$$

By assumption $f' \geq 0$ and $\beta'_{\gamma_n} \geq 0$, thus

$$\left(\left(f'(\bar{y}_{\gamma_n}) + \beta'_{\gamma_n}(\bar{y}_{\gamma_n} - \Psi)\right)\bar{p}_{\gamma_n}, -\bar{p}^-_{\gamma_n}\right)_{L^2(Q)} = \left(\left(f'(\bar{y}_{\gamma_n}) + \beta'_{\gamma_n}(\bar{y}_{\gamma_n} - \Psi)\right)\bar{p}^-_{\gamma_n}, \bar{p}^-_{\gamma_n}\right)_{L^2(Q)} \geq 0.$$

[Wac16a, Lemma 3.3], partial integration of Bochner functions, also applies to $W(I)$. Here we need to consider that the there mentioned [Rou13, Lemma 7.2] also applies to V and not only $H^1(\Omega)$. This means

$$\left(-\partial_t \bar{p}_{\gamma_n}, -\bar{p}^-_{\gamma_n}\right)_{L^2(Q)} = -\frac{1}{2}\|\bar{p}^-_{\gamma_n}(T)\|^2_H + \frac{1}{2}\|\bar{p}^-_{\gamma_n}(0)\|^2_H \geq -\frac{1}{2}\|\bar{p}^-_{\gamma_n}(T)\|^2_H = -\frac{1}{2}\|j'_T(\bar{y}_{\gamma_n}(T))^-\|^2_H.$$

So in total we have

$$\left(j'(\bar{y}_{\gamma_n}), -\bar{p}_{\gamma_n}^-\right)_{L^2(Q)} + \frac{1}{2}\|j_T'(\bar{y}_{\gamma_n}(T))^-\|_H^2 \geq \nu_{\text{ell}}\|\bar{p}_{\gamma_n}^-\|_{L^2(Q)}^2. \tag{3.33}$$

Since $\bar{p}_{\gamma_n} \xrightarrow{n\to\infty} \bar{p}$ in $L^2(Q)$, from Lemma 3.36, implies $\bar{p}_{\gamma_n}^{\pm} \xrightarrow{n\to\infty} \bar{p}^{\pm}$ in $L^2(Q)$ we have:

$$\left(j_v'(\bar{y}), -\bar{p}^-\right)_{L^2(Q)} + \frac{1}{2}\|j_T'(\bar{y}(T))^-\|_H^2 \geq \nu_{\text{ell}}\|\bar{p}^-\|_{L^2(Q)}^2.$$

By the positivity assumptions on j_v' and j_T' this implies $0 \geq \|\bar{p}^-\|_{L^2(Q)}$. Thus $\bar{p} = \bar{p}^+ \geq 0$.

We now turn to the sign of $\bar{\eta}$. Let $\varphi \in C_c^\infty(Q)$ with $\varphi \geq 0$. Note that $S_{\gamma_n}'(\bar{u}_{\gamma_n})$ is a positive operator by Theorem 3.11 and Theorem 8.15. Thus

$$C(\bar{Q}) \ni S_{\gamma_n}'(\bar{u}_{\gamma_n})(\beta_{\gamma_n}'(\bar{y}_{\gamma_n} - \Psi)\varphi) =: w_{\gamma_n} \geq 0. \tag{3.34}$$

For the regularity see again Theorem 3.11.

By the definition of $\bar{p}_{\gamma_n}$ and $\bar{\eta}_{\gamma_n}$ we find

$$\left(j_v'(\bar{y}_{\gamma_n}), w_{\gamma_n}\right)_{L^2(Q)} + \left(j_T'(\bar{y}_{\gamma_n}(T)), w_{\gamma_n}(T)\right)_H = \left(-\bar{\eta}_{\gamma_n}, \varphi\right)_{L^2(Q)}. \tag{3.35}$$

By standard arguments, e.g. the ones used for the adjoint in the proof of Lemma 3.17, we have that $\|w_{\gamma_n}\|_{C(\bar{I},H)\cap L^2(I,V)}$ is bounded independently of γ_n.

So in particular we have that $(w_{\gamma_n})_{n\in\mathbb{N}}$ is bounded in $L^2(Q)$ and thus an appropriate subsequence has a weak limit $w \in L^2(Q)$. Recall that each $w_{\gamma_n} \geq 0$ by (3.34). As the set of non-negative functions in $L^2(Q)$ is closed and convex, it is weakly closed and we have $w \geq 0$ almost everywhere. The exact same way it can be shown that a subsequence of $(w_{\gamma_n}(T))_{n\in\mathbb{N}}$ converges weakly to some $0 \leq w_T \in L^2(\Omega)$.

By the uniform convergence of the states, see Theorem 3.7, we have

$$\|j_v'(\bar{y}) - j_v'(\bar{y}_{\gamma_n})\|_{L^2(Q)}, \|j_T'(\bar{y}(T)) - j_T'(\bar{y}_{\gamma_n}(T))\|_H \xrightarrow{n\to\infty} 0.$$

Thus (3.35) yields, after taking the limit along the appropriate subsequence,

$$-(\bar{\eta}, \varphi)_{M(\bar{Q}),C(\bar{Q})} = (j_v'(\bar{y}), w)_{L^2(Q)} + (j_T'(\bar{y}(T)), w_T)_H \geq 0.$$

Thus we have shown for any $\varphi \in C_c^\infty(\bar{Q})$ that is non-negative:

$$(\bar{\eta}, \varphi)_{M(\bar{Q}),C(\bar{Q})} \leq 0.$$

It is well-known that the non-negative functions in $C_c^\infty(\bar{Q})$ are dense in the non-negative functions in $C_0(\bar{Q})$ (in the standard mollification argument non-negative mollifiers are used) with respect to $\|\cdot\|_{L^\infty(Q)}$ and thus $\bar{\eta} \leq 0$ as a measure in $M(\bar{Q}) \simeq C(\bar{Q})^*$. □

We now finally prove local quadratic growth under Assumption 3.33. This theorem and the proof are a modification of the elliptic case in [KW12b].

Theorem 3.57 *Let Assumption 3.54 hold for some $(\bar{u}, \bar{y}, \bar{\lambda}, \bar{p}, \bar{\eta})$. Then there exist $r, \delta > 0$ such that*

$$\|u - \bar{u}\|_{L^2(Q)} < r \implies J(S(u), u) \geq J(\bar{y}, \bar{u}) + \delta\|u - \bar{u}\|_{L^2(Q)}^2.$$

Proof. Assume there exists a sequence $(u_k)_{k\in\mathbb{N}} \subset L^2(Q)$ converging to $\bar{u}$ strongly in $L^2(Q)$ such that

$$J(y_k, u_k) - J(\bar{y}, \bar{u}) < \frac{\rho_k^2}{k},$$

where we abbreviated $\rho_k := \|u_k - \bar{u}\|_{L^2(Q)}$ and $y_k = S(u_k)$. We also abbreviate $\lambda_k = \lambda(u_k)$, $\bar{\lambda} = \lambda(\bar{u})$.

Then, using the mean value theorem and differentiability, we have for appropriate ξ_k^Q lying between y_k and $\bar{y}$, ξ_k^T lying between $y_k(T)$ and $\bar{y}(T)$ and ξ_k^u lying between u_k and $\bar{u}$ that

$$\begin{aligned} J(y_k, u_k) - J(\bar{y}, \bar{u}) = {} & j_v'(\bar{y})(y_k - \bar{y}) + \frac{1}{2} j_v''(\xi_k^Q)(y_k - \bar{y})^2 \\ & + j_T'(\bar{y}(T))(y_k(T) - \bar{y}(T)) + \frac{1}{2} j_T''(\xi_k^T)(y_k(T) - \bar{y}(T))^2 \\ & + g'(\bar{u})(u_k - \bar{u}) + \frac{1}{2} g''(\xi_k^u)(u_k - \bar{u})^2 \\ & - \Big((\partial_t + A)(y_k - \bar{y}) - (\lambda_k - \bar{\lambda}) + (f(y_k) - f(\bar{y})) - (u_k - \bar{u}), \bar{p} \Big)_{L^2(Q)} . \end{aligned} \tag{3.36}$$

The last line is just an added 0. We know by Corollary 3.31 that

$$\left(\bar{p} + g'(\bar{u}), u_k - \bar{u}\right)_{L^2(Q)} \geq 0$$

so the whole expression in (3.36) is bounded from below by

$$\begin{aligned} & j_v'(\bar{y})(y_k - \bar{y}) + \frac{1}{2} j_v''(\xi_k^Q)(y_k - \bar{y})^2 \\ & + j_T'(\bar{y}(T))(y_k(T) - \bar{y}(T)) + \frac{1}{2} j_T''(\xi_k^T)(y_k(T) - \bar{y}(T))^2 \\ & + \frac{1}{2} g''(\xi_k^u)(u_k - \bar{u})^2 \\ & - \Big((\partial_t + A)(y_k - \bar{y}) - (\lambda_k - \bar{\lambda}) + (f(y_k) - f(\bar{y})), \bar{p} \Big)_{L^2(Q)} . \end{aligned} \tag{3.37}$$

By (3.15) from Theorem 3.38 and the fact that $y_k - \bar{y} \in W_0(I) \cap C(\bar{Q})$ we have

$$\begin{aligned} - \left((\partial_t + A)(y_k - \bar{y}), \bar{p} \right)_{L^2(Q)} + \left(j_v'(\bar{y}), y_k - \bar{y} \right)_{L^2(Q)} + \left(j_T'(\bar{y}(T))(y_k(T) - \bar{y}(T)) \right)_H \\ = - \left(\bar{\eta}, y_k - \bar{y} \right)_{M(\bar{Q}), C(\bar{Q})} + \left(f'(\bar{y})\bar{p}, y_k - \bar{y} \right)_{L^2(Q)} . \end{aligned}$$

Inserting this into (3.37) yields the lower bound

$$\begin{aligned} & \frac{1}{2} j_v''(\xi_k^Q)(y_k - \bar{y})^2 + \frac{1}{2} j_T''(\xi_k^T)(y_k(T) - \bar{y}(T))^2 + \frac{1}{2} g''(\xi_k^u)(u_k - \bar{u})^2 \\ & - \Big(-(\lambda_k - \bar{\lambda}) + (f(y_k) - f(\bar{y})) - f'(\bar{y})(y_k - \bar{y}), \bar{p} \Big)_{L^2(Q)} \\ & - \left(\bar{\eta}, y_k - \bar{y} \right)_{M(\bar{Q}), C(\bar{Q})} . \end{aligned} \tag{3.38}$$

Because U_{ad} is bounded and $L^{q_u}(Q)$ is reflexive we may assume, without loss of generality, that $(u_k)_{k\in\mathbb{N}}$ converges weakly to $\bar{u}$ in $L^{q_u}(Q)$. We now know by Theorem 2.34 that $(y_k)_{k\in\mathbb{N}}$ converges uniformly to $\bar{y}$. Thus for k large enough we have $\|y_k - \bar{y}\|_{L^\infty(Q)} < \tau$ and thus, by Remark 2.31,

$$\operatorname{supp} \lambda_k \subset \{y_k = \Psi\} \subset \{\Psi \leq \bar{y} < \Psi + \tau\}.$$

The sets are defined as the corresponding set in Theorem 3.51 and only defined up to sets of measure zero, see Remark 2.31. So for large k we have together with $(\bar{p}, \bar{\lambda})_{L^2(Q)} = 0$, the positivity assumption on $\bar{p}$ and the positivity of λ_k:

$$\left(\lambda_k - \bar{\lambda}, \bar{p}\right)_{L^2(Q)} = \int_{\{\Psi \le \bar{y} < \Psi + \tau\}} \lambda_k \bar{p} \, d(t,x) \ge 0 \tag{3.39}$$

Because $\bar{\eta}$ is "orthogonal" to $\bar{y} - \Psi$ by Lemma 3.32 we have

$$-\left(\bar{\eta}, y_k - \bar{y}\right)_{M(\bar{Q}),C(\bar{Q})} = -\left(\bar{\eta}, y_k - \Psi\right)_{M(\bar{Q}),C(\bar{Q})} \ge 0, \tag{3.40}$$

because $-\bar{\eta} \ge 0$ by assumption and $y_k \ge \Psi$ anyway.

So (3.36), (3.38), together with (3.39) and (3.40) yields for large k

$$\begin{aligned} \frac{\rho_k^2}{k} > {} & \frac{1}{2} j_v''(\xi_k^Q)(y_k - \bar{y})^2 + \frac{1}{2} j_T''(\xi_k^T)(y_k(T) - \bar{y}(T))^2 + \frac{1}{2} g''(\xi_k^u)(u_k - \bar{u})^2 \\ & - \left(f(y_k) - f(\bar{y}) - f'(\bar{y})(y_k - \bar{y}), \bar{p}\right)_{L^2(Q)} . \end{aligned}$$

By Assumption 3.54.5 we arrive at

$$\begin{aligned} \frac{\rho_k^2}{k} > {} & \frac{1}{2}\left(j_v''(\xi_k^Q) - j_v''(\bar{y})\right)(y_k - \bar{y})^2 + \frac{1}{2}\left(j_T''(\xi_k^T) - j_T''(\bar{y}(T))\right)(y_k(T) - \bar{y}(T))^2 \\ & + \frac{1}{2} g''(\xi_k^u)(u_k - \bar{u})^2 - \left(f(y_k) - f(\bar{y}) - f'(\bar{y})(y_k - \bar{y}) - \frac{1}{2} f''(\bar{y})(y_k - \bar{y})^2, \bar{p}\right)_{L^2(Q)} . \end{aligned} \tag{3.41}$$

We now study the term involving f and its derivatives. If f is linear it simply vanishes. We thus can assume that $N \le 3$ and all the other properties for f of Assumption 3.54.5.2 are satisfied. We will show that

$$\left| \left(f(y_k) - f(\bar{y}) - f'(\bar{y})(y_k - \bar{y}) - \frac{1}{2} f''(\bar{y})(y_k - \bar{y})^2, \bar{p}\right)_{L^2(Q)} \right| = o(\rho_k^2). \tag{3.42}$$

Let $\epsilon > 0$ be arbitrary and $\delta_f > 0$ such that $|a - b| < \delta_f$ implies $|f''(t,x,a) - f''(t,x,b)| < \epsilon$ for all $(t,x) \in Q$ by the uniform continuity assumption on f''. For k large enough we have by the uniform convergence of the states that $\|y_k - \bar{y}\|_{L^\infty(Q)} < \delta_f$. Thus (3.42) is bounded from above by

$$\left| \left(\int_0^1 \int_0^1 (f''(\bar{y} + st(y_k - \bar{y})) - f''(\bar{y})) s \, dt \, ds (y_k - \bar{y})^2, \bar{p} \right)_{L^2(Q)} \right| \le \int_Q \frac{1}{2} \epsilon |y_k - \bar{y}|^2 |\bar{p}| \, d(t,x).$$

By Proposition 8.11 this bounded from above by

$$C\epsilon \|\bar{p}\|_{L^2(I,V) \cap C(\bar{I},H)} \|y_k - \bar{y}\|^2_{L^2(I,V) \cap C(\bar{I},H)}.$$

With the Lipschitz continuity of S, Corollary 2.35, we arrive at

$$\left| \left(f(y_k) - f(\bar{y}) - f'(\bar{y})(y_k - \bar{y}) - \frac{1}{2} f''(\bar{y})(y_k - \bar{y})^2, \bar{p}\right)_{L^2(Q)} \right| \le C\epsilon\rho_k^2 \tag{3.43}$$

for large enough k. Thus

$$\limsup_{k \to \infty} \rho_k^{-2} \left| \left(f(y_k) - f(\bar{y}) - f'(\bar{y})(y_k - \bar{y}) - \frac{1}{2} f''(\bar{y})(y_k - \bar{y})^2, \bar{p}\right)_{L^2(Q)} \right| \le C\epsilon.$$

As $\epsilon > 0$ was arbitrary this implies

$$\left| \left(f(y_k) - f(\bar{y}) - f'(\bar{y})(y_k - \bar{y}) - \frac{1}{2} f''(\bar{y})(y_k - \bar{y})^2, \bar{p} \right)_{L^2(Q)} \right| = o(\rho_k^2). \tag{3.44}$$

Rearranging (3.41) now delivers

$$\left(j_v''(\xi_k^Q) - j_v''(\bar{y}) \right) \left(\frac{y_k - \bar{y}}{\rho_k} \right)^2 + \left(j_T''(\xi_k^T) - j_T''(\bar{y}(T)) \right) \left(\frac{y_k(T) - \bar{y}(T)}{\rho_k} \right)^2 + g''(\xi_k^u) \left(\frac{u_k - \bar{u}}{\rho_k} \right)^2 \leq \frac{2}{k} + \frac{o(\rho_k^2)}{\rho_k^2}. \tag{3.45}$$

On to the final stretch: By ellipticity of $g''(\bar{u})$ and (3.45) we have

$$\begin{aligned} 0 < \nu_g \leq g''(\bar{u}) \left(\frac{u_k - \bar{u}}{\rho_k} \right)^2 &= (g''(\bar{u}) - g''(\xi_k^u)) \left(\frac{u_k - \bar{u}}{\rho_k} \right)^2 + g''(\xi_k^u) \left(\frac{u_k - \bar{u}}{\rho_k} \right)^2 \\ \leq (g''(\bar{u}) - g''(\xi_k^u)) \left(\frac{u_k - \bar{u}}{\rho_k} \right)^2 + \frac{2}{k} + \frac{o(\rho_k^2)}{\rho_k^2} \\ - (j_v''(\bar{y}) - j_v''(\xi_k^Q)) \left(\frac{y_k - \bar{y}}{\rho_k} \right)^2 - (j_T''(\bar{y}) - j_v''(\xi_k^T)) \left(\frac{y_k(T) - \bar{y}(T)}{\rho_k} \right)^2 . \end{aligned} \tag{3.46}$$

We have assumed g'' to be continuous, therefore

$$(g''(\bar{u}) - g''(u_k)) \left(\frac{u_k - \bar{u}}{\rho_k} \right)^2 \leq \| g''(\bar{u}) - g''(u_k) \|_{Bil(L^2(Q), L^2(Q))} \left\| \frac{u_k - \bar{u}}{\rho_k} \right\|_{L^2(Q)}^2 \xrightarrow{k \to \infty} 0. \tag{3.47}$$

Here $\|\cdot\|_{Bil(L^2(Q),L^2(Q))}$ is the norm of a linear, continuous operator $L^2(Q) \to L^2(Q)^*$. Analogously the terms involving $(j_v''(\bar{y}) - j_v''(\xi_k^Q))$ and $(j_T''(\bar{y}) - j_v''(\xi_k^T))$ vanish for $k \to \infty$. Thus (3.46) implies after taking the limit that

$$0 < \nu_g \leq 0. \tag{3.48}$$

This is a contradiction and thus the quadratic growth condition has to be satisfied. □

Remark 3.58 It may be possible to improve the previous result by weakening the conditions from Assumption 3.54 by, for example, employing the directional derivative of S from [Chr19].

Remark 3.59 We remark here that the previous Example 3.40 also satisfies second order sufficient conditions. This shows that even second order conditions do not necessarily entail that the set of multipliers in the first order conditions are unique.

Assumption 3.33.1 - Assumption 3.33.5 are clearly satisfied by Example 3.40.

By construction $(\bar{u}, \bar{y}, \bar{\lambda}, \bar{p}, \bar{\eta})$ satisfies the necessary optimality conditions from Theorem 3.38 and by the arguments in Example 3.40 we have $\bar{p} \geq 0$ globally on Q. Also by construction $\bar{\eta} = p_0 \cdot \delta_0 \leq 0$ as a measure. Thus Assumption 3.33.6 is satisfied.

3.6 Second Order Sufficient Conditions for Regularized Control Problems

We consider the following regularized control problem for comparison with (OC):

$$\min_{(y,u)\in W(I)\times L^2(Q)} j_v(y) + j_T(y(T)) + g(u),\quad \text{such that } S_\gamma(u) = y \text{ and } u \in U_{ad}. \tag{OC$_\gamma$}$$

Note that this differs from $(\mathrm{OC}_{\gamma,\bar{u}})$ in so far that we do not require any optimal solution $(\bar{u}, \bar{y})$ to define it. We collect the most important statements about (OC_γ) in the following theorem:

Theorem 3.60 (OC_γ) *has at least one optimal solution* $(\bar{y}_\gamma, \bar{u}_\gamma)$*.*

Assume that $\beta_\gamma \in C^{1,1}_{loc}(\mathbb{R})$ *or that* β_γ *is of the form from Proposition 2.17 for* $\alpha_\beta = 1$ *and that* $|\{\bar{y}_\gamma = \Psi\}| = 0$*. Then* $(\bar{y}_\gamma, \bar{u}_\gamma)$ *satisfies*

$$(\bar{p}_\gamma + g'(\bar{u}_\gamma), u - \bar{u}_\gamma)_{L^2(Q)} \geq 0 \qquad \forall u \in U_{ad}.$$

Here $\bar{p}_\gamma \in W(I)$ *solves the linear, parabolic PDE*

$$\begin{cases} -\partial_t \bar{p}_\gamma + A\bar{p}_\gamma + {\beta_\gamma}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma + f'(\bar{y}_\gamma)\bar{p}_\gamma = j_v'(\bar{y}_\gamma) \\ \bar{p}_\gamma(T) = j_T'(\bar{y}_\gamma(T)), \quad \bar{p}_\gamma|_{\Sigma_D} = 0. \end{cases}$$

Here we again have ${\beta_\gamma}'(r) = \gamma^{-1}1_{(-\infty,0)}(r)$ *if* $\alpha_\beta = 1$*.*

Proof. The existence proof is essentially the same as the one for the unregularized problem (OC) in Theorem 3.3.

The optimality conditions are just basic optimality conditions obtained by differentiating the problem in proper directions, e.g. [Trö09, Lemma 2.21]. For the differentiability see Theorem 3.11 or Theorem 3.13. Introducing the adjoint just works as in Corollary 3.31. □

Theorem 3.61 *Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a zero sequence and* $((\bar{y}_{\gamma_n}, \bar{u}_{\gamma_n}))_{n\in\mathbb{N}}$ *a sequence of solutions to* (OC_γ) *for* $\gamma = \gamma_n$*. There exists a subsequence* $(\gamma_{n_k})_{k\in\mathbb{N}}$ *and* $(\bar{y}, \bar{u}) \in W(I) \times U_{ad}$ *such that*

- $\bar{u}_{\gamma_{n_k}} \xrightarrow{k\to\infty} \bar{u}$ *weakly in* $L^{q_u}(Q)$*,*
- $\bar{y}_{\gamma_{n_k}} \xrightarrow{k\to\infty} \bar{y}$ *strongly in* $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ *and weakly in* $W(I)$

and $(\bar{y}, \bar{u})$ *is a solution to* (OC)*.*

Proof. The existence of limits and the indicated convergence types are proven as in Theorem 3.7. To see that the limit $(\bar{y}, \bar{u})$ is indeed optimal, we have by the weakly lower semicontinuity of g:

$$J(\bar{y}, \bar{u}) \leq \liminf_{k\to\infty} J(\bar{y}_{\gamma_{n_k}}, \bar{u}_{\gamma_{n_k}}).$$

By the optimality of $(\bar{y}_{\gamma_{n_k}}, \bar{u}_{\gamma_{n_k}})$ we have for any $u \in U_{ad}$

$$J(\bar{y}, \bar{u}) \leq \lim_{k\to\infty} J(S_{\gamma_{n_k}}(u), u) = J(S(u), u). \tag{3.49}$$

Here we used that by definition, cf. Definition 2.32, $(S_{\gamma_{n_k}}(u))_{k\in\mathbb{N}}$ converges uniformly to $S(u)$. □

Remark 3.62 One could now retrace a lot of the theorems and deductions from Section 3.3 and obtain boundedness and convergence of the controls, states, adjoints and the corresponding multipliers. This would lead to the analysis of accumulation points of the respective quantities and thus to a link to the unregularized problem. We do not, however, reiterate all the theorems, because that would produce little new information in our eyes. We therefore limit ourselves to the analysis of second order sufficient conditions, because this is also interesting within the numerical analysis of (OC), respectively (OC_γ), see Chapter 5 and in particular Remark 5.8.

The only thing that may not be obvious is the question whether the regularized controls converge to $\bar{u}$ strongly in $L^2(Q)$. For $u = \bar{u}$ (3.49) implies $g(\bar{u}_k) \xrightarrow{k\to\infty} g(\bar{u})$. If g is for example given by $\frac{\nu_g}{2}\|\cdot\|^2_{L^2(Q)}$, this convergence implies that $\|u_k\|_{L^2(Q)} \xrightarrow{k\to\infty} \|u\|_{L^2(Q)}$. Together with the weak convergence of the controls, this implies the strong $L^2(Q)$-convergence. Thus the rest of the statements is now indeed retraceable.

Before we prove an analogue of Theorem 3.57 for the regularized control problem (OC_γ). We note that it is not straightforward to transfer the result of Theorem 3.57 to the regularized setting as the following one-dimensional example shows.

Example 3.63 The unique minimizer of $d : [-1, 1] \to \mathbb{R}$, $d(x) = x^2$ is clearly 0 and there clearly holds a quadratic growth condition in 0 with constant 1. Yet, consider the following family of functions:

$$d_\epsilon \colon [-1,1] \to \mathbb{R},$$
$$x \mapsto \begin{cases} x^2 & \text{if } x \le 0, \\ \epsilon(x-\epsilon)^2 - \epsilon^3 & \text{if } 0 < x \le \epsilon, \\ (x-\epsilon)^2 - \epsilon^3 & \text{if } \epsilon < x, \end{cases}$$

with $\epsilon \in (0, 1)$. We see that ϵ is the minimizer with minimal value $-\epsilon^3$. The minimizers also satisfy local growth conditions, more specifically

$$d(x) \ge d(\epsilon) + \epsilon(x-\epsilon)^2 \qquad \forall x \in [-1, 1].$$

We also see that for any $c > \epsilon$ that

$$d(x) = \epsilon(x-\epsilon)^2 - \epsilon^3 < d(\epsilon) + c(x-\epsilon)^2 \qquad \forall x \in (0, \epsilon).$$

Thus ϵ is the maximal constant for the local quadratic growth and not just a lower bound.

We also see that $(d_\epsilon)_{\epsilon\in(0,1)}$ converges uniformly to d:

$$\begin{aligned} \sup_{x\in[-1,1]} |d_\epsilon(x) - d(x)| &= \max\left(\sup_{x\in(0,\epsilon]} x^2 - \epsilon(x-\epsilon)^2 + \epsilon^3, \sup_{x\in(\epsilon,1]} x^2 - (x-\epsilon)^2 + \epsilon^3 \right) \\ &\le \max\left(\sup_{x\in(0,\epsilon]} \epsilon^2 - \epsilon x^2 + 2x\epsilon^2, \sup_{x\in(\epsilon,1]} 2x\epsilon - \epsilon^2 + \epsilon^3 \right) \le 2(\epsilon + \epsilon^2 + \epsilon^3). \end{aligned}$$

Concluding, we have uniform convergence of functions $(d_\epsilon)_{\epsilon\in(0,1)}$ to the function d. Their unique minimizers $(\epsilon)_{\epsilon\in(0,1)}$ converge to the unique minimizer 0 of d. The minimizers of d_ϵ each satisfy a quadratic growth condition with the maximal growth constant $\epsilon > 0$. Yet, the minimizer 0 of d satisfies a growth condition with constant 1.

That means that one cannot deduce any behaviour on the quadratic growth in the minimizer of d_ϵ from the two facts that: d, the limit, satisfies a growth condition in its minimizer, and d_ϵ itself satisfies a growth condition in its minimizer.

Even with this possible weakness in mind we will prove quadratic growth conditions for (OC_γ) where the growth constants and radii will depend on γ. This is still interesting for the numerical analysis, see the related Remark 5.8.

Assumption 3.64 *Let $(\bar{u}_\gamma, \bar{y}_\gamma, \bar{p}_\gamma)$ be such that they satisfy the necessary first order optimality conditions of (OC_γ) from Theorem 3.60. This assumes in particular that S_γ is Frechét differentiable in $\bar{u}_\gamma$. This triple and the functionals j_v, j_T, g shall satisfy the following points*

1. *$j_v : L^2(Q) \to \mathbb{R}$ is twice continuously Frechét differentiable,*
2. *$j_T : L^2(\Omega) \to \mathbb{R}$ is twice continuously Frechét differentiable,*
3. *$g : L^2(Q) \to \mathbb{R}$ is twice continuously Frechét differentiable.*

Additionally the following properties shall be satisfied:

4. *There exists a $\nu_g > 0$ such that $g''(\bar{u}_\gamma)$ is elliptic with ellipticity constant ν_g,*
5. *Assume $N \leq 3$, $\beta_\gamma \in C^2(\mathbb{R})$, $f''(t, x, \cdot)$ exists, is uniformly continous for all $(x, y) \in Q$, satisfies $f''(\bar{y}_\gamma) \in L^\infty(Q)$ and*

$$j_v''(\bar{y}_\gamma)v^2 + j_T''(\bar{y}_\gamma(T))v(T)^2 - \int_Q \bar{p}_\gamma({\beta_\gamma}''(\bar{y}_\gamma - \Psi) + f''(\bar{y}_\gamma))v^2 \, d(t,x) \geq 0 \qquad (3.50)$$

   for all $v \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \cap W_0(I)$.

Remark 3.65 We shall once again comment on these assumptions and their validity. Assumption 3.64.1-Assumption 3.64.4 are again quite obvious and relatively benign assumptions.

Assumption 3.64.5, however, ist complicated. It essentially is the analogue/replacement of Assumption 3.33.5 - Assumption 3.33.6. This is not immediately obvious, but becomes clearer throughout the proof of local quadratic growth in Theorem 3.67. We can at least provide a heuristic motivation. Assume for the moment that the sign of $\bar{p}$ from the unregularized case in Assumption 3.33 is the same as the sign of $\bar{p}_\gamma$, for sufficiently small γ. This is possible under a restrictive, but simple assumption, cf. Remark 3.69. So, if $\bar{p}_\gamma \geq 0$ on $\{\Psi \leq \bar{y} < \Psi + \tau\}$, we find by uniform convergence of $(\bar{y}_{\gamma_n})_{n\in\mathbb{N}}$ to $\bar{y}$, cf. Theorem 3.61, that $\bar{p}_\gamma \geq 0$ on $\{\bar{y}_\gamma \leq \Psi\}$ for γ small enough. Therefore we see that

$$-(\bar{p}_\gamma {\beta_\gamma}''(\bar{y}_\gamma - \Psi), d^2)_{L^2(Q)} \geq 0.$$

So, if the analogue of Assumption 3.33.5 holds for $\bar{y}_\gamma$ this "implies" Assumption 3.64.5.

Our analysis is closely related to the study of second order sufficient conditions of semilinear, semi-smooth parabolic PDEs. The literature is limited and we are only aware of [Bet19]. There assumption (4.22) is the analogue to our Assumption 3.64.5. The author of [Bet19] comes to the conclusion, at the end of section 4, that the there given second order conditions are comparatively sharp. We therefore conclude that there are no obvious ways to avoid this type of condition.

The author of [Bet19] cannot avoid a two-norm gap, while we can, using the special structure of our problem and fact that Assumption 3.64.5 is structurally similar, but stronger, than (4.22) in [Bet19]. We comment later how to weaken Assumption 3.64.5 by accepting a two-norm gap.

Proposition 3.66 *Assume that $\beta \in C^2(\mathbb{R})$. Then for any $y \in L^\infty(Q)$ we have*

$$\|\beta''(y + \delta y) - \beta''(y)\|_{L^\infty(Q)} \xrightarrow{\|\delta y\|_{L^\infty(Q)} \to 0} 0.$$

Proof. We abbreviate $r_y := \|y\|_{L^\infty(Q)} + 1$. As β'' is continuous and $[-r_y, r_y]$ is compact we have that $\beta''|_{[-r_y,r_y]}$ is uniformly continuous. So any $\epsilon > 0$ there exists a $\delta > 0$ such that for any $r, r' \in [-r_y, r_y]$ with $|r - r'| < \delta$ we have $|\beta''(r) - \beta''(r')| < \epsilon$. Thus we can conclude

$$\limsup_{\|\delta y\|_{L^\infty(Q)} \to 0} \|\beta''(y + \delta y) - \beta''(y)\|_{L^\infty(Q)} < \epsilon.$$

As $\epsilon > 0$ was arbitrary we conclude the proof. □

Theorem 3.67 *Assume that Assumption 3.64 is satisfied for some* $(\bar{u}_\gamma, \bar{y}_\gamma, \bar{p}_\gamma)$. *Then there exist* $r_\gamma, \delta_\gamma > 0$ *such that*

$$\|u - \bar{u}_\gamma\|_{L^2(Q)} < r_\gamma \implies J(S_\gamma(u), u) \geq J(\bar{y}_\gamma, \bar{u}_\gamma) + \delta_\gamma \|u - \bar{u}_\gamma\|^2_{L^2(Q)}.$$

Proof. Copying the proof of Theorem 3.57 until (3.38) verbatim yields

$$\begin{aligned} \frac{\rho_k^2}{k} >& \frac{1}{2} j''(\xi_\gamma^{Q,k})(y_\gamma^k - \bar{y}_\gamma)^2 + \frac{1}{2} j''(\xi_\gamma^{T,k})(y_\gamma^k(T) - \bar{y}_\gamma(T))^2 + \frac{1}{2} g''(\xi_\gamma^{u,k})(u_\gamma^k - \bar{u}_\gamma)^2 \\ &- \Big(-(\lambda_\gamma^k - \bar{\lambda}_\gamma) + f(y_\gamma^k) - f(\bar{y}_\gamma) - f'(\bar{y}_\gamma)(y_k - \bar{y}_\gamma), \bar{p}_\gamma\Big)_{L^2(Q)} \\ &- \Big(\bar{\eta}_\gamma, y_\gamma^k - \bar{y}_\gamma\Big)_{L^1(Q), L^\infty(Q)}. \end{aligned} \tag{3.51}$$

Here we have $\lambda_\gamma^k := -\beta_\gamma(\bar{y}_\gamma^k - \Psi)$, $\bar{\lambda}_\gamma = -\beta_\gamma(\bar{y}_\gamma - \Psi)$ and $\bar{\eta}_\gamma = -\beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma$. As in the proof of Theorem 3.57 we may still assume that $(u_\gamma^k)_{k \in \mathbb{N}}$ converges weakly to $\bar{u}_\gamma$ in $L^{q_u}(Q)$. By Corollary 2.28 we therefore have that $(y_\gamma^k)_{k \in \mathbb{N}}$ converges uniformly to $\bar{y}_\gamma$.

We now analyse the terms in (3.51) involving the regularization term:

$$\begin{aligned} &- \Big(-(\lambda_\gamma^k - \bar{\lambda}_\gamma), \bar{p}_\gamma\Big)_{L^2(Q)} - \Big(\bar{\eta}_\gamma, y_\gamma^k - \bar{y}_\gamma\Big)_{L^1(Q), L^\infty(Q)} \\ &= \int_Q -\beta_\gamma(\bar{y}_\gamma^k - \Psi)\bar{p}_\gamma + \beta_\gamma(\bar{y}_\gamma - \Psi)\bar{p}_\gamma + \beta_\gamma{}'(\bar{y}_\gamma - \Psi)\bar{p}_\gamma(\bar{y}_\gamma^k - \bar{y}_\gamma)\, d(t,x) \\ &= -\int_Q \left(\int_0^1 \int_0^1 \beta_\gamma{}''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) s\, dt\, ds\right)(y_\gamma^k - \bar{y}_\gamma)^2 \bar{p}_\gamma\, d(t,x). \end{aligned} \tag{3.52}$$

We now compare this to $(-\frac{1}{2}\beta_\gamma{}''(\bar{y}_\gamma - \Psi)\bar{p}_\gamma, (y_\gamma^k - \bar{y}_\gamma)^2)_{L^2(Q)}$. The arguments will be similar to those for f in the proof of the unregularized case, cf. Theorem 3.57. We find

$$\begin{aligned} &\left|\int_Q \left(\int_0^1 \int_0^1 \beta_\gamma{}''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) s\, dt\, ds - \frac{1}{2}\beta_\gamma{}''(\bar{y}_\gamma - \Psi)\right)(y_\gamma^k - \bar{y}_\gamma)^2 \bar{p}_\gamma\, d(t,x)\right| \\ &\leq \|\bar{p}_\gamma\|_{L^3(Q)} \|y_\gamma^k - \bar{y}_\gamma\|^2_{L^3(\Omega)} \\ &\quad \cdot \left\|\int_0^1 \int_0^1 [\beta_\gamma{}''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) - \beta_\gamma{}''(\bar{y}_\gamma - \Psi)] s\, dt\, ds\right\|_{L^\infty(\Omega)} dt. \end{aligned}$$

We again have by Proposition 8.11 and Corollary 2.35 that

$$\begin{aligned} &\left|\int_Q \left(\int_0^1 \int_0^1 \beta_\gamma{}''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) s\, dt\, ds - \frac{1}{2}\beta_\gamma{}''(\bar{y}_\gamma - \Psi)\right)(y_\gamma^k - \bar{y}_\gamma)^2 \bar{p}_\gamma\, d(t,x)\right| \\ &\leq C\rho_k^2 \int_0^1 \int_0^1 \|\beta_\gamma{}''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) - \beta_\gamma{}''(\bar{y}_\gamma - \Psi)\|_{L^\infty(Q)} s\, dt\, ds. \end{aligned} \tag{3.53}$$

By our assumptions Proposition 3.66 applies. Thus for fixed $s, t \in (0,1)$ we have

$$\|\beta_\gamma''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) - \beta_\gamma''(\bar{y}_\gamma - \Psi)\|_{L^\infty(Q)} \xrightarrow{k\to\infty} 0.$$

By the theorem of dominated convergence, e.g. [BK15, Proposition 5.4], and the boundedness of β_γ'' on the compact set $[-\sup_{k\in\mathbb{N}}\|y_\gamma^k - \Psi\|_{L^\infty(Q)}, \sup_{k\in\mathbb{N}}\|y_\gamma^k - \Psi\|_{L^\infty(Q)}]$ we can conclude that

$$\int_0^1\int_0^1 \|\beta_\gamma''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma)) - \beta_\gamma''(\bar{y}_\gamma - \Psi)\|_{L^\infty(Q)} s\,dt\,ds \xrightarrow{k\to\infty} 0.$$

Thus (3.53) entails

$$\left|\int_Q \left(\int_0^1\int_0^1 \beta_\gamma''(\bar{y}_\gamma - \Psi + st(y_\gamma^k - \bar{y}_\gamma))s\,dt\,ds - \frac{1}{2}\beta_\gamma''(\bar{y}_\gamma - \Psi)\right)(y_\gamma^k - \bar{y}_\gamma)^2\bar{p}_\gamma\,d(t,x)\right| = o(\rho_k^2).$$

This and (3.52) therefore entail

$$\begin{aligned} &-\left(-(\lambda_\gamma^k - \bar{\lambda}_\gamma), \bar{p}_\gamma\right)_{L^2(Q)} - \left(\bar{\eta}_\gamma, y_\gamma^k - \bar{y}_\gamma\right)_{L^1(Q),L^\infty(Q)} \\ &= -\int_Q \frac{1}{2}\beta_\gamma''(\bar{y}_\gamma - \Psi)\bar{p}_\gamma(y_\gamma^k - \bar{y}_\gamma)^2\,d(t,x) + o(\rho_k^2). \end{aligned} \tag{3.54}$$

As in the proof of (3.44) in Theorem 3.57 we have

$$\left|(f(y_\gamma^k) - f(\bar{y}_\gamma) - f'(\bar{y}_\gamma)(y_k - \bar{y}_\gamma) - \frac{1}{2}f''(\bar{y}_\gamma)(y_k - \bar{y}_\gamma)^2, \bar{p}_\gamma)_{L^2(Q)}\right| = o(\rho_k^2). \tag{3.55}$$

As in (3.47) we find

$$g''(\xi_\gamma^{u,k})(u_\gamma^k - \bar{u}_\gamma)^2 = g''(\bar{u}_\gamma)(u_\gamma^k - \bar{u}_\gamma)^2 + o(\rho_k^2).$$

Again the same holds true for j_v and j_T so that together with (3.54) and (3.55) we conclude from (3.51):

$$\begin{aligned} \frac{2\rho_k^2}{k} + o(\rho_k^2) >& j_v''(\bar{y}_\gamma)(y_\gamma^k - \bar{y}_\gamma)^2 + j_T''(\bar{y}_\gamma(T))(y_\gamma^k(T) - \bar{y}_\gamma(T))^2 + g''(\bar{u}_\gamma)(u_\gamma^k - \bar{u}_\gamma)^2 \\ &- \int_Q \bar{p}_\gamma(\beta_\gamma''(\bar{y}_\gamma - \Psi) + f''(\bar{y}_\gamma))(y_\gamma^k - \bar{y}_\gamma)^2\,d(t,x). \end{aligned} \tag{3.56}$$

By (3.50), keeping in mind that $y_\gamma^k - \bar{y}_\gamma \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \cap W_0(I)$, and Assumption 3.64.4 we find

$$\frac{2\rho_k^2}{k} + o(\rho_k^2) > \nu_g.$$

Sending $k \to \infty$ now yields the desired contradiction. □

Remark 3.68 As promised at the end of Remark 3.65 we shall comment on a variant of Theorem 3.67 involving a two norm gap. The following thoughts are essentially [Bet19, Theorem 4.13]. Presume that in the proof of Theorem 3.67 we assume $\rho_k = \|u_\gamma^k - \bar{u}_\gamma\|_{L^{q_u}(Q)} \xrightarrow{k\to\infty} 0$, which is stronger than the situation in Theorem 3.67. One arrives at (3.56) just the same.

Dividing this by ρ_k and using the differentiability of $S_\gamma : L^{q_u}(Q) \to C(\bar{Q})$ from Theorem 3.13 yields

$$\begin{aligned} 0 \geq \frac{1}{2} j_v''(\bar{y}_\gamma)[S_\gamma{}'(\bar{u}_\gamma)d]^2 + \frac{1}{2} j_T''(\bar{y}_\gamma(T))[S_\gamma{}'(\bar{u}_\gamma)d](T)^2 + \frac{1}{2}\nu_g \\ - \left(\frac{1}{2}\bar{p}_\gamma \beta_\gamma{}''(\bar{y}_\gamma - \Psi) + \frac{1}{2} f''(\bar{y}_\gamma), [S_\gamma{}'(\bar{u}_\gamma)d]^2\right)_{L^2(Q)}. \end{aligned} \tag{3.57}$$

Now (3.50) again yields a contradiction. However, it is sufficient to assume a weaker version of (3.50), namely

$$\begin{aligned} j_v''(\bar{y}_\gamma)(S_\gamma{}'(\bar{u}_\gamma)d)^2 + j_T''(\bar{y}_\gamma(T))(S_\gamma{}'(\bar{u}_\gamma)d)(T)^2 \\ - \int_Q \bar{p}_\gamma(\beta_\gamma{}''(\bar{y}_\gamma - \Psi) + f''(\bar{y}_\gamma))(S_\gamma{}'(\bar{u}_\gamma)d)^2 \, d(t,x) \geq 0 \end{aligned}$$

for all $d \in L^{q_u}(Q)$.

Remark 3.69 In Theorem 3.67 all the constants $r_\gamma, \delta_\gamma > 0$ depend on γ. But in a very special situation, we can easily obtain a quadratic growth condition uniform in γ. This is more or less just an academic situation but it gives us an indication that the quadratic growth of (OC_γ) might be uniform in γ after all. This is reserved for future research.

We study the case that $j_T = 0$, j_v is quadratic of the form (3.31) and $y_Q \leq \Psi - \epsilon$ on Q for some $\epsilon > 0$. We also consider $g(u) := \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}$ for some $\alpha_g > 0$. For presentations' sake we also choose $f = 0$. Lastly, we of course presume $\beta \in C^2(\mathbb{R})$.

By the uniform convergence of $(\bar{y}_\gamma)_{\gamma>0}$ to some $\bar{y}$ by Theorem 3.61, after consideration of subsequences, we have $\bar{y}_\gamma \geq y_Q$ for all sufficiently small γ. By the maximum principle and the definition of $\bar{p}_\gamma$ we have $\bar{p}_\gamma \geq 0$ as in the proof of Lemma 3.56. Hence, we see that the reduced cost functional

$$j_\gamma(u) := \frac{\alpha_Q}{2}\|S_\gamma(u) - y_Q\|^2_{L^2(Q)} + \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}$$

satisfies for any $d \in L^{q_u}(Q)$

$$\begin{aligned} j_\gamma''(\bar{u}_\gamma)d^2 &= (S'(\bar{u}_\gamma)d, S'(\bar{u}_\gamma)d)_{L^2(Q)} + (S(\bar{u}_\gamma) - y_Q, S''(\bar{u}_\gamma)d^2)_{L^2(Q)} + \alpha_g(d,d)_{L^2(Q)} \\ &\geq \alpha_g\|d\|^2_{L^2(Q)} + (S(\bar{u}_\gamma) - y_Q, S''(\bar{u}_\gamma)d^2)_{L^2(Q)} \\ &= \alpha_g\|d\|^2_{L^2(Q)} + ((-\partial_t + A + \beta_\gamma{}'(S(\bar{u}_\gamma) - \Psi))\bar{p}_\gamma, S''(\bar{u}_\gamma)d^2)_{L^2(Q)}. \end{aligned}$$

S'' will be shown to be well-defined later in Theorem 6.3 and Theorem 6.5. Using the definition of $S''(\bar{u}_\gamma)d^2$ and under the assumption that $\{\bar{y}_\gamma = \Psi\}$ is a zero set, we arrive at

$$j_\gamma''(\bar{u}_\gamma)d^2 \geq \alpha_g\|u_\gamma\|^2_{L^2(Q)} + (\bar{p}_\gamma, -\beta_\gamma{}''(\bar{y}_\gamma - \Psi)[S'(\bar{u}_\gamma)d]^2)_{L^2(Q)}.$$

As we argued before we have $\bar{p}_\gamma \geq 0$. We also know that $\beta_\gamma{}'' \leq 0$ and thus $j_\gamma''(\bar{u}_\gamma)d^2 \geq \alpha_g\|d\|^2_{L^2(Q)}$. We thus see that in this special case a quadratic growth condition holds independently of γ. Sending γ to 0 obviously entails that the same growth condition holds for the unregularized problem.

3.7 Optimal Control of Obstacle Problems Without Control Constraints

We now shortly comment on the control problem when U_{ad} is unbounded. The trick is reducing it to a constrained control problem.

Definition 3.70 In Section 3.7 we consider the control problem

$$\min_{(y,u)\in W(I)\times L^\infty(Q)} j_v(y) + j_T(y(T)) + g(u) =: \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}, \quad \text{such that } S(u) = y. \tag{OC$_{unbd}$}$$

In particular, we have $j_T = 0$ and $\alpha_g > 0$ is some arbitrary constant. Note that this problem admits only bounded controls, but is not a control constrained problem in and of itself.

As a tool we also introduce the constrained control problems for $a > 0$:

$$\min_{(y,u)\in W(I)\times L^\infty(Q)} \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}, \quad \text{such that } S(u) = y \text{ and } u \in [-a, a]. \tag{OC^a}$$

We immediately see that (OC^a) satisfies all the assumptions from Definition 3.2. Thus all results from Sections 3.1 through 3.5 hold for any $a > 0$. We will show that for a sufficiently large a (OC^a) is equal to (OC$_{unbd}$) and thus can easily transfer the results from the previous sections to (OC$_{unbd}$).

Theorem 3.71 *Assume $N \leq 3$. Let $\tilde{q_u} \in (1+N/2, 3]$ and assume that $j_v(y) = \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)}$ with $y_Q \in L^{\tilde{q_u}}(Q)$. Assume g is of the form from Corollary 3.41. Then there exists a $a_0 > 0$ such that (OC^a) is equivalent to (OC$_{unbd}$) for all $a \geq a_0$.*

Proof. As mentioned before it is clear that (OC^a) satisfies the assumptions from Definition 3.2 and thus a solution $(\bar{y}_a, \bar{u}_a)$ exists. By Theorem 3.47 we have

$$\|\bar{u}_a\|_{L^\infty(Q)} \leq C_{\not{a}}\|j_v'(\bar{y})\|_{L^{\tilde{q_u}}(Q)} \leq C_{\not{a}}(\|\bar{y}\|_{L^3(Q)} + \|y_Q\|_{L^{\tilde{q_u}}(Q)}). \tag{3.58}$$

By Proposition 8.11 and the assumption $N \leq 3$ there is a $\theta \in (0, 1)$ such that

$$\|\bar{y}\|_{L^3(Q)} \leq C\|\bar{y}\|^{1-\theta}_{C(\bar{I},H)}\|\bar{y}\|^\theta_{L^2(I,V)}. \tag{3.59}$$

$C > 0$ does not depend on a. We will show at the end that

$$\|\bar{y}\|_{L^2(I,V)\cap C(\bar{I},H)} \leq C \tag{3.60}$$

where C does not depend on a. Then (3.59) entails

$$\|\bar{y}\|_{L^3(Q)} \leq C_{\not{a}}$$

and then (3.58)

$$\|\bar{u}_a\|_{L^\infty(Q)} \leq C_{\not{a}}.$$

We choose a_0 as the right hand side and let $a \geq a_0$. We show that any solution to (OC^a) solves (OC$_{unbd}$) and vice versa. Let $\bar{u}$ be a solution to (OC$_{unbd}$). Thus

$$J(S(\bar{u}), \bar{u}) = \min_{u\in L^\infty(Q)} J(S(u), u) \leq \min_{u\in[-\|\bar{u}\|_{L^\infty(Q)}, \|\bar{u}\|_{L^\infty(Q)}]} J(S(u), u) = J(S(\bar{u}), \bar{u}). \tag{3.61}$$

Hence $\bar{u}$ is, unsurprisingly, a solution to

$$\min_{u\in[-\|\bar{u}\|_{L^\infty(Q)}, \|\bar{u}\|_{L^\infty(Q)}]} J(S(u), u).$$

By the earlier discussion this entails $\|\bar{u}\|_{L^\infty(Q)} \leq a_0$. By the same line of arguing as in (3.61) we have that $\bar{u}$ solves (OC^a). This also automatically entails

$$\min_{u\in L^\infty(Q)} J(S(u),u) = \min_{u\in[-a,a]} J(S(u),u).$$

In turn, this entails that every solution to (OC^a) also solves (OC$_{unbd}$).

It remains to prove (3.60). This is not difficult, but lengthy. We test the multiplier formulation, see Theorem 2.30, with $\bar{y}\cdot 1_{(0,t)}$ for $t \in I$ and use the monotonicity of f to see that

$$\begin{aligned}\left(\|\lambda(\bar{u})\|_{L^2(Q)} + \|\bar{u}\|_{L^2(Q)}\right)\|\bar{y}\|_{L^2(Q)} &\geq \int_0^t (\partial_t\bar{y} + f(\bar{y}), \bar{y}_\gamma)_H + a_\Omega(\bar{y},\bar{y})\,ds \\ &\geq \frac{1}{2}\|\bar{y}(t)\|_H^2 - \frac{1}{2}\|y_0\|_H^2 + \nu_{ell}\|\nabla\bar{y}\|^2_{L^2((0,t)\times\Omega)} - \|f(0)\|_{L^2(Q)}\|\bar{y}\|_{L^2(Q)}.\end{aligned}$$

As $t \in I$ was arbitrary we conclude

$$\left(\|\bar{u}\|_{L^2(Q)} + \|f(0)\|_{L^2(Q)} + \|\lambda(\bar{u})\|_{L^2(Q)}\right)\|\bar{y}\|_{L^2(Q)} + \frac{1}{2}\|y_0\|_H^2 \geq \frac{1}{2}\|\bar{y}\|^2_{L^\infty(I,H)}.$$

As in the proof of Lemma 3.17, cf. the arguments below (3.9), we get that

$$\|\bar{y}\|_{L^2(I,V)\cap C(\bar{I},H)} \leq C\left(\|\bar{u}\|_{L^2(Q)} + \|f(0)\|_{L^2(Q)} + \|\lambda(\bar{u})\|_{L^2(Q)} + \|y_0\|_H\right).$$

Here C does only depend on T and ν_{ell}. By the optimality of $\bar{u}$ and the structure of g we find

$$\|\bar{u}\|_{L^2(Q)} \leq \sqrt{\alpha_g^{-1} J(S(0),0)}. \tag{3.62}$$

Thus we have for a C independent of a or $\bar{u}_a$ that

$$\|\bar{y}\|_{L^2(I,V)\cap C(\bar{I},H)} \leq C\left(1 + \|\lambda(\bar{u})\|_{L^2(Q)}\right). \tag{3.63}$$

It remains to estimate $\|\lambda(\bar{u})\|_{L^2(Q)}$.

Using the same techniques as in Lemma 2.25 one can show that

$$\|\lambda_\gamma(\bar{u})\|_{L^2(Q)} \leq \|\bar{u}\|_{L^2(Q)} + \|\partial_t\Psi\|_{L^2(Q)} + \|A\Psi\|_{L^2(Q)} + \|f(\Psi)\|_{L^2(Q)}$$

holds for any $\gamma > 0$. Since by Theorem 2.30 we have $\lambda_\gamma(\bar{u}) \xrightarrow{\gamma\to 0} \lambda(\bar{u})$ weakly in $L^{q_u}(Q)$ and therefore weakly in $L^2(Q)$ we deduce by the weakly lower semi-continuity of the norms that

$$\|\lambda(\bar{u})\|_{L^2(Q)} \leq \|\bar{u}\|_{L^2(Q)} + \|\partial_t\Psi\|_{L^2(Q)} + \|A\Psi\|_{L^2(Q)} + \|f(\Psi)\|_{L^2(Q)}.$$

This together with (3.62) and (3.63) we have shown (3.60). □

Remark 3.72 Now the important results of the previous sections can be transferred to (OC$_{unbd}$) by means of (OC^a). We do not go over every little lemma, but most importantly the necessary optimality conditions from Theorem 3.38 hold true. The most interesting part here is the relation between control and adjoint. Let $(\bar{y},\bar{u})$ be an optimal solution to (OC$_{unbd}$), a_0 from Theorem 3.71 and $a > 0$. Then we know, by Theorem 3.38, or rather by Corollary 3.41, that

$$\bar{u} = P_{[-a,a]}[-\alpha_g^{-1}\bar{p}].$$

As $a \geq a_0$ was arbitrary this show that $\bar{u} = -\alpha_g\bar{p}$. Exactly what one would hope for an unbounded problem. This does entail that the adjoint $\bar{p}$ and the multiplier $\bar{\eta}$ are unique by Theorem 3.44.

The sufficient conditions from Section 3.5 obviously also hold true under the appropriate assumptions.

4 Discretization and Numerical Analysis for Regularized Obstacle Problems

Throughout Chapter 4 we assume the definitions and assumptions of Section 2.2.1 to apply. We also assume that $N \geq 2$, even though up to and including Section 4.1.3 $N = 1$ is possible by the same proofs and references. However, the later sections often make reference to works where $N \geq 2$ is assumed, which makes it easier to parse by ignoring the special case $N = 1$.

The structure of this chapter is as follows: we first start with the discussion of the discretization of $V = H_0^1(\Omega)$ by linear elements. This discussion includes the derivation of $L^\infty(\Omega)$-quasi-stability results of the Ritz projection by techniques of [SW82] and resolvent estimates for the operator A by techniques of [BTW03].

After that we discretize the time dependencies by piecewise constant functions and combine this discretization with the spatial discretization to obtain $L^\infty(Q)$-error estimates for linear parabolic PDEs. We will then finally turn to the discretization of the regularized obstacle problem and its numerical analysis, which requires the theory for linear parabolic PDEs.

4.1 Spatial Discretization

4.1.1 Definitions

Definition 4.1 By $\mathcal{K}_h$ we denote a triangulation of Ω. That means that $\mathcal{K}_h$ is a set of open, convex, and tetrahedral cells $K \subset \Omega$ with diameters h_K. The mesh size is denoted by $h := \max_{K\in\mathcal{K}_h} h_K$. We define $\Omega_h := \operatorname{int} \bigcup_{K\in\mathcal{K}_h} \bar{K} \subset \Omega$. The set Ω_h is polygonal/polyhedral. We do not allow hanging nodes and the closures of two different cells, which intersect, shall always intersect in a common facet. The set of nodes is denoted by $\mathcal{N}_h$. We will use the terms "mesh" and "triangulation" interchangeably throughout this thesis.

The reference element $\hat{K}$ is defined as the unit simplex, i.e. $\hat{K} := \operatorname{span}(0, e_1, \ldots, e_N)$. By $J_K : \hat{K} \to K$ we denote the linear mapping from the reference element to the general element. We always assume it to be bijective, which implicitly entails, for example, that K cannot be a line.

As usual we can not work on arbitrary cells. We require several regularity and quasi-uniformity assumptions:

Definition 4.2 We call a family $(\mathcal{K}_h)_{h>0}$ of meshes quasi-uniform if there is a constant $C > 0$ such that

$$h_K \leq h \leq C|K|^{\frac{1}{N}} \qquad \forall K \in \mathcal{K}_h, h > 0.$$

This immediately implies for some $C > 0$, independent of K or h

$$h_K \leq h \leq Ch_K.$$

This property is also sometimes referred to as quasi-uniformity.

We call a family of meshes $(\mathcal{K}_h)_{h>0}$ shape regular if there exists a $C > 0$ such that

$$\frac{h_K}{\rho_K} \leq C \qquad \forall K \in \mathcal{K}_h. \tag{4.1}$$

Here ρ_K denotes the diameter of the largest circle/ball inscribed in K.

Proposition 4.3 *Let $(\mathcal{K}_h)_{h>0}$ be a family of shape regular meshes. Then there exists a $c > 0$ such that for all $h > 0$, $K \in \mathcal{K}_h$ and all edges E of K we have $|E| \geq ch_K$.*

If $(\mathcal{K}_h)_{h>0}$ is additionally quasi-uniform this entails $|E| \geq ch$.

Proof. It is obvious that the diameter of the incircle of a tetrahedron is always smaller than the shortest edge. Thus any edge $E \subset K$ satisfies $|E| \geq \rho_K \geq ch_K$, where we used the shape regularity.

It is clear that the first claim implies the second by the quasi-uniformity assumption. □

Definition 4.4 Note that for the rest of the thesis we assume $V := H_0^1(\Omega)$. This is stressed again later. Whenever we refer to A^{-1} or similar operators, we assume that A is equipped with homogenous boundary conditions. This is clear by the definition of $\text{dom}_p(A)$ for $p \in (1, \infty)$ but it is important to keep in mind nevertheless. The space of piecewise linear functions in V subordinate to a triangulation $\mathcal{K}_h$ is denoted by V_h, i.e.

$$V_h := \{v_h \in C_0(\Omega_h) : v_h|_K \text{ is linear } \forall K \in \mathcal{K}_h\}.$$

By linear we technically mean affine linear.

We assume $\Omega_h \subset \Omega$ and thus can extend each $v_h \in V_h$ by 0 to a function in $H_0^1(\Omega) \cap C(\bar{\Omega})$. This is the reason we consider $V = H_0^1(\Omega)$: extending the finite element function with something else than 0 poses its own challenges. So, we have $V_h \subset V$. This is very important so that all the following definitions are well-defined and intuitive. In general it is of course possible to consider non-conforming case $\Omega_h \not\subset \Omega$. Then one would have to extend $H_0^1(\Omega)$ by 0 onto $H_0^1(\Omega \cup \Omega_h)$ and V_h in the same fashion. Then one could redefine the bilinear forms a_Ω/a_I on the space $H_0^1(\Omega \cup \Omega_h)$, respectively $L^2(I, H_0^1(\Omega \cup \Omega_h))$, and work from there. We do not do this as we think it does not add much to the present analysis in terms of new results. From time to time we will comment on it nevertheless.

For $k \in \mathbb{N}, p \in [1, \infty]$ we define:

$$W^{k,p,h}(\Omega_h) := \left\{ v_h \in L^1(\Omega_h) : \|v\|^p_{W^{k,p,h}(\Omega_h)} := \sum_{K \in \mathcal{K}_h} \|v_h\|^p_{W^{k,p}(K)} < \infty \right\}.$$

We can clearly use this definition for any subset of Ω_h consisting of a union of cells.

Note that in this definition $\|v_h\|_{W^{2,p,h}(\Omega_h)} = \|v_h\|_{W^{1,p}(\Omega_h)}$ for any $v_h \in V_h$. This notation is frequently used in our references.

We of course require the notions of interpolation and projection, which we will introduce here. They are well-known, we therefore see the following paragraph as an introduction of notation.

Definition 4.5 Let $\mathcal{K}_h$ be a triangulation of Ω and V_h the corresponding space of affine elements. Then we define the interpolation operator

$$\begin{aligned} I_h \colon C_0(\bar{\Omega}) &\to V_h, \\ v &\mapsto I_h v, \end{aligned}$$

where $I_h v$ is the unique affine function that satisfies $v(\hat{x}) = I_h v(\hat{x})$ for any node $\hat{x} \in \mathcal{N}_h$.

We also define the L^2-projection via

$$\begin{aligned} P_h \colon L^2(\Omega) &\to V_h, \\ v &\mapsto \text{ the unique } v_h \text{ such that } (v - v_h, \varphi_h)_{L^2(\Omega)} = 0 \qquad \forall \varphi_h \in V_h. \end{aligned}$$

Lastly we define the Ritz-projection via

$$\begin{aligned} R_h \colon H_0^1(\Omega) &\to V_h, \\ v &\mapsto \text{ the unique } v_h \text{ such that } a_\Omega(v - v_h, \varphi_h) = 0 \qquad \forall \varphi_h \in V_h \end{aligned}$$

and the discrete version of the operator A via

$$A_h \colon V_h \to V_h^*, \quad v_h \mapsto \int_\Omega \nabla v_h^T A \nabla \cdot \, dx.$$

The symbol A in the integral refers to the coefficient matrix of the operator.

4.1.2 Preliminary Results About Meshes

In this section we collect minor results on meshes that mostly verify assumptions made by authors we cite. A common assumption on the distance of $\partial\Omega$ and $\partial\Omega_h$ is implied under regularity of Ω, see for example [Ran17, proof of Theorem 3.3].

Proposition 4.6 *Assume $N = 2$ and that Ω is a convex C^2 domain. Let $(\mathcal{K}_h)_{h>0}$ be a family of meshes such that the boundary nodes of each Ω_h lie on $\partial\Omega$, then*

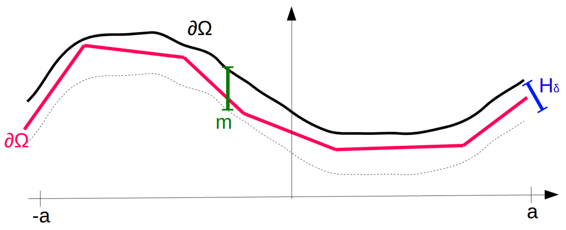
$$\max_{x \in \partial\Omega_h} \operatorname{dist}(x, \partial\Omega) = \mathcal{O}(h^2).$$

The following proposition proves a uniform Lipschitz property of the meshes that is for example used in [SW82] and other papers by the same authors. There it is, however, not verified but assumed. We verify it in a case we also use in programming.

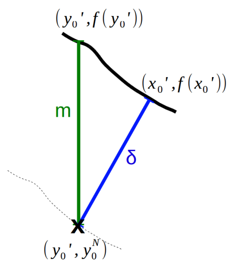
Proposition 4.7 *Let $N = 2$. Let $(\mathcal{K}_h)_{h>0}$ a family of quasi-uniform and shape regular meshes. Assume $\delta := \max_{x \in \partial\Omega_h} \operatorname{dist}(x, \partial\Omega) \leq C_\delta h^2$; the $\max$ is attained, because $\partial\Omega_h$ is compact. Finally assume the smallest angle in $\partial\Omega_h$ is bounded from below by $\alpha_0 \in (0, \pi)$ independently of h.*

Then the $(\Omega_h)_{h>0}$ and Ω are uniform Lipschitz domains for sufficiently small h. More precisely there is a $h_0 > 0$ such that for $h \in (0, h_0)$ one has the following: for the quantities a, b, S_i, h_i, f_i from the definition of the Lipschitz property of Ω one has that there are Lipschitz functions f_i^h such that after transforming with S_i we have for all $(y, y_N) \in (-a, a)^{N-1} \times (-b, b)$:

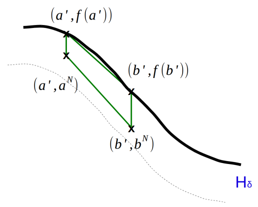
$$\begin{aligned} (y, y_N) \in \Omega_h &\iff f_i^h(y) > y_i > f_i^h(y) - b, \\ (y, y_N) \in \partial\Omega_h &\iff f_i^h(y) = y_N. \end{aligned}$$



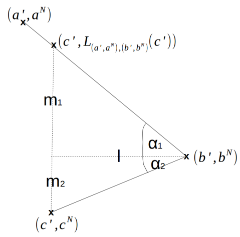
(a) The "tube" H_δ around $\partial\Omega$ of size δ. The blue line has the length δ.



(b) A sketch illustrating the points used in the estimate of m. It is basically a zoomed in part of Figure 4.1a.



(c) The points used to estimate the maximum slope of the line connecting (a', a^N) and (b', b^N).



(d) The points used to prove that $\partial\Omega_h$ can be written as a graph for h sufficiently small.

Here, we implicitly used Ω_h as the symbol for the original Ω_h transformed by S_i^{-1}, as it is customary.

Also all the $S_1((-a,a)^{N-1} \times (-b,b)), \dots, S_M((-a,a)^{N-1} \times (-b,b))$ cover $\partial\Omega_h$ for each $h > 0$. Furthermore if we let $L_\Omega := \max_{i=1,\dots,\#f_i} Lip_{f_i}$ we have

$$\max_{i=1,\dots,\#f_i} Lip_{f_i^h} \leq L_\Omega + \frac{(2 + L_\Omega) C_\delta h}{c_e}.$$

Here $c_e > 0$ is the h-independent constant from Proposition 4.3. We also have

$$\max_{i=1,\dots,\#f_i} \|f_i - f_i^h\|_{L^\infty([-a,a]^{N-1})} \leq (2 + L_\Omega)\delta \leq C_\delta (2 + L_\Omega) h^2.$$

In words: almost the same transformations, that locally represent $\partial\Omega$ as a graph, can be used to represent $\partial\Omega_h$ locally as a graph.

Proof. We do everything for one coordinate system, i.e. after transformation by one S_i, and drop the index i. We first analyse what the maximum distance of $\partial\Omega_h$ to $\partial\Omega$ in x_N direction could be. Looking at Figure 4.1a one can see that $\partial\Omega_h$ is contained in the "tube" H_δ of thickness δ around $\partial\Omega$. Formally

$$\partial\Omega_h \cap ((-a,a)^{N-1} \times \mathbb{R}) \subset H_\delta := \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) \leq \delta\} \cap ((-a,a)^{N-1} \times \mathbb{R}).$$

We analyse H_δ to obtain some behaviour of $\partial\Omega_h$.

We denote the maximum vertical expansion of H_δ by m. In formulae:

$$m := \sup_{(y_0', y_0^N) \in H_\delta} |(y_0', y_0^N) - (y_0', f(y_0'))| = \sup_{(y_0', y_0^N) \in H_\delta} |y_0^N - f(y_0')|.$$

Let $(y_0', y_0^N) \in H_\delta$. By definition of H_δ there exists a $(x_0', f(x_0')) \in (-a,a)^{N-1} \times \mathbb{R}$ such that $|(y_0', y_0^N) - (x_0', f(x_0'))| \leq \delta$. Then we have

$$\begin{aligned} |y_0^N - f(y_0')| = |(y_0', y_0^N) - (y_0', f(y_0')| &\leq \delta + |(x_0', f(x_0')) - (y_0', f(y_0'))| \\ &\leq \delta + |x_0' - y_0'| + L_\Omega |x_0' - y_0'| \leq (2 + L_\Omega)\delta. \end{aligned}$$

Thus

$$m \leq (2 + L_\Omega)\delta. \tag{4.2}$$

So for $h < \sqrt{\frac{b}{C_\delta(2+L_\Omega)}}$ we have $m < b$ and thus the whole tube around $\partial\Omega$ is contained in our local coordinate system and thus in particular $\partial\Omega_h$ is contained there as well. Yet, this does not imply that $\partial\Omega_h$ can be represented as a graph.

We now compute the maximum slope a line segment of at least length $c_e h$, i.e. a possible facet of $\partial\Omega_h$ contained in H_δ, can have. Here $c_e > 0$ is the constant from Proposition 4.3 obtained by quasi-uniformity and shape regularity. Let $a = (a', a^N), b = (b', b^N) \in H_\delta$ such that $|(a', a^N) - (b', b^N)| \geq c'h$. Without loss of generality we may assume $a^N \geq b^N$, cf. Figure 4.1c. Then we define the linear map connecting a and b:

$$L_{(a',a^N),(b',b^N)}(x') = a^N + \frac{(x' - a')^T (b' - a')}{|b' - a'|^2} (b^N - a^N).$$

Its gradient is given by

$$\nabla L_{(a',a^N),(b',b^N)}(x') = \frac{b' - a'}{|b' - a'|^2} (b^N - a^N).$$

Thus the slope of the line segment between a and b is bounded by $\frac{|b^N - a^N|}{|b' - a'|}$. By the assumption $a^N \geq b^N$ we have the following estimate on the upper bound

$$\frac{a^N - b^N}{|a' - b'|} \leq \frac{f(a') - f(b')}{|a' - b'|} + \frac{f(b') - b^N}{|a' - b'|} \leq L_\Omega + \frac{m}{c_e h}.$$

By (4.2) we find the following bound on the slope

$$|\nabla L_{(a',a^N),(b',b^N)}| \leq L_\Omega + \frac{(2 + L_\Omega)}{c_e} \frac{\delta}{h} =: L_\Omega^{disc}. \tag{4.3}$$

We now show that $\partial\Omega_h$ can indeed be written locally as a graph. We have already shown that the open sets $S_1((-a,a)^{N-1} \times (-b,b)), \ldots, S_M((-a,a)^{N-1} \times (-b,b))$ cover $\partial\Omega_h$. Thus by the Lebesgue's number lemma, e.g. [Mun75, Chapter 3, Lemma 7.2], there exists a $h_0 > 0$ so small that for each $h \leq h_0$ we have that each subset of $\partial\Omega_h$ of diameter at most $2h$ is contained in one of the covering sets. Then each two connected edges are contained fully in one of the systems. Now consider any two connected edges of $\partial\Omega_h$. We will now show that the situation of Figure 4.1d cannot happen for sufficiently small h. That means that the two edges cannot run into opposite directions. Let everything be transformed into a coordinate system that completely contains those two edges, which exists by our earlier observations.

In the following $(a', a^N), (b', b^N), (c', c^N), (c', L_{(a',a^N),(b',b^N)}(c'))$ are chosen as in Figure 4.1d. $[a, b]$ and $[b, c]$ are edges and thus are longer than $c_e h$ by assumption. We estimate l from Figure 4.1d. We see by (4.3) that

$$\begin{aligned} c_e h \leq |c' - b'| + |c^N - b^N| &\leq |c' - b'| + |L_{(c',c^N),(b',b^N)}(c') - L_{(c',c^N),(b',b^N)}(b')| \\ &\leq (1 + L_\Omega^{disc})|c' - b'| = (1 + L_\Omega^{disc})l. \end{aligned}$$

Thus $l \geq \frac{c_e h}{1+L_\Omega^{disc}}$. Now we can estimate $\sin(\alpha_1+\alpha_2) > 0$:

$$\sin(\alpha_1+\alpha_2) = \sin(\alpha_1)\cos(\alpha_2) + \sin(\alpha_2)\cos(\alpha_1) = \frac{m_1}{|d-b|}\frac{l}{|d-b|} + \frac{m_2}{|c-b|}\frac{l}{|c-b|}.$$

Since $|d-b| \geq l$ and $|c-b| \geq l$ we have

$$\sin(\alpha_1+\alpha_2) \leq \frac{2m}{l} \leq \frac{2m(1+L_\Omega^{disc})}{c_e h}.$$

Applying (4.2) and using the definition of α_0 yields

$$0 < \sin(\alpha_0) \leq \sin(\alpha_1+\alpha_2) \leq \frac{2(2+L_\Omega^{disc})^2}{c}\frac{\delta}{h} \leq \frac{2(2+L_\Omega^{disc})^2 C_\delta}{c}h.$$

For h sufficiently small this yields a contradiction.

Since $\Omega_h \subset \Omega$ those coordinate systems and transformations cover the whole of $\partial\Omega_h$ for those small enough h. □

Remark 4.8 We note here that the angle condition from Proposition 4.7 is also known as Zlámal condition, cf. [BKK11], which implies that triangular meshes are shape regular. Consider a triangle with edges of length a, b, c and the corresponding angles α, β, γ. Without loss of generality assume $a \leq b \leq c$. Then by the law of sines we see that $\alpha_0 \leq \alpha \leq \beta \leq \gamma$. By school knowledge the formula for the incircle ρ is given as

$$\rho = \frac{1}{2}\sqrt{\frac{(b+c-a)(a+c-b)(a+b-c)}{a+b+c}} \geq \frac{1}{2}\sqrt{\frac{a}{3c}(a^2-(c-b)^2)}.$$

By the law of sines and basic identities for the sine we have

$$c = \frac{\sin(\gamma)}{\sin(\alpha)}a \leq \frac{1}{\sin(\alpha)}a.$$

Thus we deduce

$$\begin{aligned}
\rho &\geq \frac{1}{2}\sqrt{\frac{\sin(\alpha)}{3}}\sqrt{a^2\left(1-\left(\frac{\sin(\gamma)}{\sin(\alpha)}-\frac{\sin(\beta)}{\sin(\alpha)}\right)^2\right)} \\
&\geq \frac{1}{2}\sqrt{\frac{\sin(\alpha)}{3}}\frac{a}{\sin(\alpha)}\sqrt{\sin^2(\alpha)-(\sin(\gamma)-\sin(\beta))^2} \\
&\geq \frac{1}{2}\sqrt{\frac{\sin(\alpha_0)}{3}}c\sqrt{\sin^2(\alpha)-(\sin(\gamma)-\sin(\beta))^2}.
\end{aligned}$$

We know that $\gamma = \pi - (\alpha+\beta)$ and therefore find

$$\begin{aligned}
\rho &\geq \frac{c}{2}\sqrt{\frac{\sin(\alpha_0)}{3}}\sqrt{\sin^2(\alpha)-(\sin(\alpha+\beta)-\sin(\beta))^2} \\
&= \frac{c}{2}\sqrt{\frac{\sin(\alpha_0)}{3}}\sqrt{\sin^2(\alpha)-(\sin(\beta)(\cos(\alpha)-1)+\sin(\alpha)\cos(\beta))^2} \\
&= \frac{c}{2}\sqrt{\frac{\sin(\alpha_0)}{3}} \\
&\quad \sqrt{\sin^2(\alpha)(1-\cos^2(\beta))-\sin^2(\beta)(1-\cos(\alpha))^2-2\sin(\beta)\cos(\beta)\sin(\alpha)(\cos(\alpha)-1)}.
\end{aligned}$$

Clearly $0 < \alpha \le \beta < \frac{\pi}{2}$ so that the last summand in the root is larger than 0 resulting in

$$\begin{aligned}\rho &\ge \frac{c}{2}\sqrt{\frac{\sin(\alpha_0)}{3}}\sqrt{(1-\cos^2(\alpha)-(1-\cos(\alpha))^2)\sin^2(\beta)}\\ &\ge \frac{c}{2}\sqrt{\frac{\sin(\alpha_0)}{3}}\sin(\alpha)\sqrt{2\cos(\alpha)-2\cos^2(\alpha)}.\end{aligned}$$

As the smallest angle α can not be larger than $\pi/3$ we get by the monotonitcity of the cosine

$$\rho \ge c\sqrt{\frac{\sin(\alpha_0)}{6}}\sin(\alpha)\sqrt{\cos(\pi/3)}\sqrt{1-\cos(\alpha)} \ge \frac{c}{2}\sqrt{\frac{\sin^3(\alpha_0)}{12}}\sqrt{1-\cos(\alpha_0)}.$$

This shows the shape regularity.

An important consequence of having uniform Lipschitz domains is the following

Proposition 4.9 *Let $(\mathcal{K}_h)_{h>0}$ be a family of meshes such that $(\Omega)_{h>0}$ and Ω are uniform Lipschitz domains. For the quantities a, b, S_i, h_i, f_i from the definition of the Lipschitz property of Ω and f_i^h satisfying the properties listed in Proposition 4.7 we have in each coordinate system S_i*

$$\|f_i - f_i^h\|_{L^\infty([-a,a]^{N-1})} \le C_{\not h}\delta.$$

Proof. This is just proven as (4.2), whose proof does not require any regularity assumptions on $(\mathcal{K}_h)_{h>0}$, N or any interior angles. □

4.1.3 Preparatory Steps to Proving L^∞-stability of the Ritz Projection

In this section we prove and verify statements often used as assumptions made by the authors of [SW82]. That way we can assure applicability of their theorems and our deductions. We start with the well-known interpolation inequality.

Lemma 4.10 *Let $(\mathcal{K}_h)_{h>0}$ be a family of quasi-uniform and shape regular meshes. Let $p \in [1,\infty]$ and $m \in \{0,1,2\}$, satisfying $m > \frac{N}{p}$ for $p > 1$ and $m \ge N$ for $p = 1$. Then we have for any $k \in \{0,\dots,m\}$, $K \in \mathcal{K}_h$ and $v \in W^{m,p}(K)$:*

$$|v - I_h v|_{W^{k,p}(K)} \le C h^{m-k}|v|_{W^{m,p}(K)}.$$

The constant $C > 0$ does not depend on K, h, k, p.

Proof. This standard estimate can be found in [BS08, Theorem (4.4.24)]. □

The next proposition corresponds to assumption [SW82, (A.2)] and is essentially a trace theorem where the dependency on the domain has been tracked.

Proposition 4.11 *Assume that $(\mathcal{K}_h)_{h>0}$ is quasi-uniform and shape regular. Let $K \in \mathcal{K}_h$ and let $v \in W^{1,1}(K)$. Then there exist a $C > 0$ independent of h, v and K such that:*

$$\int_{\partial K} |v|\, dS \le C\left(h^{-1}\|v\|_{L^1(K)} + |v|_{W^{1,1}(K)}\right).$$

Proof. Defining $\hat{v} := v \circ J_K$ we have $\nabla \hat{v} = DJ_K^T \nabla v \circ J_K$ and $\nabla v = D(J_K^{-T}) \nabla \hat{v} \circ J_K^{-1}$. For each edge/surface $E \subset K$ and $\hat{E} := J_K^{-1}(E)$ we have

$$\int_E |v| \, dS = \int_{J_K(\hat{E})} |v| \, dS = \int_{\hat{E}} |v \circ J_K| \, |\det DJ_K| \, dS \leq C \int_{\hat{E}} |\hat{v}| h^{N-1} \, dS$$

where we used the Jacobi transformation formula, see for example [BK15, Chapter 10]. We only get an h^{N-1} instead of h^N from Proposition 8.28 as we consider an $N-1$-dimensional subset, i.e. $\hat{E}$, of $\hat{K}$ and thus a restriction of J to $\hat{E}$.

By the trace theorem for Sobolev spaces, see for example [Gri11, Theorem 1.5.1.3], we can bound this further by $Ch^{N-1}\|v\|_{W^{1,1}(K)}$. Using the transformation again this bound is equal to

$$Ch^{N-1} \int_K \left(\left| \hat{v} \circ J_K^{-1} \right| + \left| \nabla \hat{v} \circ J_K^{-1} \right| \right) \left| \det D \left(J_K^{-1} \right) \right| \, dx.$$

Using the definition of v and $\hat{v}$ and the bound on the Jacobi determinant from Proposition 8.28 this is in turn bounded by

$$Ch^{-1} \int_K |v| + \left| \left(D(J_K^{-1}) \right)^{-1} \nabla v \right| \, dx \leq Ch^{-1} \int_K |v| + \left| \left(DJ_K \circ J_K^{-1} \right) \nabla v \right| \, dx.$$

Using the norm estimate for the Jacobian from Proposition 8.28 and the fact that the edge was arbitrary we conclude

$$\int_{\partial K} |v| \, dS \leq Ch^{-1} \int_K |v| + h|\nabla v| \, dx.$$

□

The following proposition is a generalization of the famous inverse inequalities frequently encountered in numerical analysis.

Proposition 4.12 *Let $m, k \in \mathbb{N}_0$ with $0 \leq m \leq k$ and $p, q \in [1, \infty]$ with $q \leq p$. Assume that $(\mathcal{K}_h)_{h \in (0,1]}$ is quasi-uniform and shape regular. Then there exist $\rho, C > 0$, independent of h, K, p and q such that for any linear $v_h : K \to \mathbb{R}$*

$$\|v_h\|_{W^{k,p}(K)} \leq C h_K^{m-k-N(\frac{1}{q}-\frac{1}{p})} \|v_h\|_{W^{m,q}(K_{\rho,h})} \quad \forall K \in \mathcal{K}_h$$

where $K_{\rho,h} := \{x \in K : \operatorname{dist}(x, \partial K) > \rho h\}$.

Proof. The statement can be proven by changing the proof for the well known inverse inequality, see for example [EG04, Lemma 1.138] or [BS08, Lemma (4.5.3)]. The core of all the proofs is to establish

$$\|\hat{v}_h\|_{W^{k,p}(\hat{K})} \leq C \|\hat{v}_h\|_{L^1(\hat{K})}$$

by the equivalence of norms on the finite dimensional space of linear polynomials on $\hat{K}$. We instead use

$$\|\hat{v}_h\|_{W^{k,p}(\hat{K})} \leq C \|\hat{v}_h\|_{L^1(\rho\hat{K})}.$$

for some $\rho \in (0,1)$. We now assume that $\hat{K}$ is centered at 0 such that $J_K(0) = 0 \in K$, i.e. J_K is linear. This obviously does not change the norms. Now the exact same arguments of [BS08, Lemma (4.5.3)] yield

$$\|v_h\|_{W^{k,p}(K)} \leq Ch^{-k+\frac{N}{p}-\frac{N}{q}} \|v_h\|_{L^q(J_K(\rho\hat{K}))} = Ch^{-k+\frac{N}{p}-\frac{N}{q}} \|v_h\|_{L^q(\rho K)}.$$

The estimate for $m > 0$ also follows by the same arguments as in [BS08, Lemma (4.5.3)]. □

The next proposition is a useful result which shows that the L^2-projection onto V_h behaves nicely with any norm and not just the L^2-norm.

Proposition 4.13 *Assume $(\mathcal{K}_h)_{h>0}$ is a family of quasi-uniform and shape regular meshes. Let $p \in [1,\infty]$. Then there is a $C > 0$ independent of h such that*

$$\|P_h u\|_{L^p(\Omega_h)} \leq C\|u\|_{L^p(\Omega_h)} \qquad \forall u \in L^\infty(\Omega).$$

Proof. The main theorem of [DDW75] can be applied to the polygonal domain Ω_h, as the constants there do not depend on Ω. The assumptions in [DDW75] are satisfied for affine triangular elements by the remarks at the end of the paper. □

Assumption 4.14 *We now state a collection of mesh regularity assumptions that will be required by the following theorems. Its requirement will be stated explicitly. In cases where they could be relaxed we will give note. The assumptions are strong enough such that all the previous results from Section 4.1.3 until this point are valid.*

Assume $N \in \{2,3\}$ and that Ω is a $C^{2,\alpha}$-domain for some $\alpha > 0$.

Assume that $(\mathcal{K}_h)_{h\in(0,1]}$ is a family of quasi-uniform and shape regular meshes. Further assume

$$\delta := \max_{x\in\partial\Omega_h} \operatorname{dist}(x,\partial\Omega) \leq Ch^2. \tag{4.4}$$

Assume that $(\Omega_h)_{h>0}$ and Ω are uniform Lipschitz domains. That means that there exist finitely many local coordinate systems represented by the affine transformations $S_1,\dots,S_m$ and local transformations $f_1,\dots,f_m$, $f_1^h,\dots,f_m^h$ for $h > 0$, as in Definition 1.5 for $k = 0$, $\alpha = 1$, such that the the boundaries of each Ω_h and Ω can be represented as a Lipschitz graph as in Definition 1.5. Also see the results in Proposition 4.7 for a rigorous formulation. If Ω is assumed to more regular than a Lipschitz $f_1,\dots,f_m$ are assumed to be of the appropriate, higher regularity. There also exists a $C > 0$ such that all the Lipschitz constants of the f_i^h are bounded by C.

The restriction $N \in \{2,3\}$ and the domain regularity are not necessary for all of the following statements, for example, Proposition 4.15 and Proposition 4.20 do not require it, while Proposition 4.18 and all results based on it do. Corollary 8.31 also requires $N \in \{2,3\}$, which is proven in the appendix and used throughout Section 4.1.6.

The following is a result about local approximation, which, together with all the following propositions, allows us to estimate the usually difficult to handle L^∞-errors in Section 4.1.5. It corresponds to [SW82, Assumption A.4], a paper we will cite frequently.

Proposition 4.15 *Let Assumption 4.14 be satisfied. There exist $C, c > 0$ independent of Ω such that for any $v \in W^{2,\infty}(\Omega) \cap C_0(\Omega)$ and any $h \in (0,1]$ there is a $v_h \in V_h$ with the following properties: let $B := B_d(y)$ and $B' := B_{2d}(y)$ two balls with $y \in \Omega$, $d \geq ch$. We set $D_h := B \cap \Omega_h$, $D' := B' \cap \Omega$. Then*

$$h^{-1}\|v - v_h\|_{L^\infty(D_h)} + \|v - v_h\|_{W^{1,\infty}(D_h)} \leq Ch\|v\|_{W^{2,\infty}(D')} + Ch^{-1}\delta\|v\|_{W^{1,\infty}(D')}. \tag{4.5}$$

Remark 4.16 Proposition 4.15 can be generalized to higher order finite elements, cf. [SW82, Remarks after A.4].

Proof. The proof is close to the ideas in the remark after [SW82, Assumption A.4]. Let $v \in W^{2,\infty}(\Omega)$ with $v|_{\partial\Omega} = 0$. The idea is to use the nodal interpolant on the interior of Ω_h and a slight modification on the boundary of Ω_h to obtain v_h.

We define $L_h := \left\{K \in \mathcal{K}_h : \bar{K} \cap \partial\Omega_h \neq \emptyset\right\}$. This is the boundary layer of Ω_h that needs special consideration. By $\mathcal{N}_h^0 := \mathcal{N}_h \cap \partial\Omega_h$ we denote the boundary nodes of Ω_h. We define v_h on the nodes $\hat{x} \in \mathcal{N}_h$ via

$$v_h(\hat{x}) := \begin{cases} v(\hat{x}) & \text{if } \hat{x} \notin \partial\Omega_h, \\ 0 & \text{if } \hat{x} \in \partial\Omega_h. \end{cases}$$

We can clearly extend v_h to a function in V_h. Note that v_h is the nodal interpolant of v far enough away from $\partial\Omega_h$.

We first consider $K \notin L_h$. Then $v_h|_K = I_h v|_K$ and therefore, by shape regularity and a standard estimate for the interpolant, see Lemma 4.10,

$$h^{-1}\|v - v_h\|_{L^\infty(K)} + \|v - v_h\|_{W^{1,\infty}(K)} \leq Ch\|v\|_{W^{2,\infty}(K)}. \tag{4.6}$$

We now check what happens at the boundary. For any $K \in L_h$ the inverse property of Proposition 4.12 gives us

$$h^{-1}\|v_h - I_h v\|_{L^\infty(K)} + \|v_h - I_h v\|_{W^{1,\infty}(K)} \leq Ch^{-1}\|v_h - I_h v\|_{L^\infty(K)}. \tag{4.7}$$

We will show that

$$h^{-1}\|v_h - I_h v\|_{L^\infty(K)} \leq Ch^{-1}\delta\|v\|_{W^{1,\infty}(S_h)}, \tag{4.8}$$

where we clarify S_h later. We then have for any K in L_h

$$\begin{aligned} &h^{-1}\|v - v_h\|_{L^\infty(K)} + \|v - v_h\|_{W^{1,\infty}(K)} \\ &\leq h^{-1}\|v - I_h v\|_{L^\infty(K)} + \|v - I_h v\|_{W^{1,\infty}(K)} + h^{-1}\|I_h v - v_h\|_{L^\infty(K)} + \|I_h v - v_h\|_{W^{1,\infty}(K)} \\ &\leq Ch\|v\|_{W^{2,\infty}(K)} + Ch^{-1}\delta\|v\|_{W^{1,\infty}(S_h)}. \end{aligned} \tag{4.9}$$

Here we used the same interpolation error estimate as before and (4.8).

We now establish (4.8). Because for each cell $K \in L_h$ the nodes $\hat{x} \in K \setminus \partial\Omega_h$ satisfy $v_h(\hat{x}) = I_h v(\hat{x})$ and we have, by linearity,

$$\sup_{x \in K} |v_h(x) - I_h v(x)| = \sup_{x \in \partial K \cap \mathcal{N}_h} |v_h(x) - I_h v(x)| = \sup_{x \in \partial K \cap \mathcal{N}_h^0} |I_h v(x)|. \tag{4.10}$$

Therefore (4.7) entails

$$h^{-1}\|v_h - I_h v\|_{L^\infty(K)} + \|v_h - I_h v\|_{W^{1,\infty}(K)} \leq Ch^{-1}\|I_h v\|_{L^\infty(\partial K \cap \mathcal{N}_h^0)}.$$

Because $I_h v$ is the nodal interpolant of v we have by the mean value theorem, since $v|_{\partial\Omega} = 0$,

$$h^{-1}\|v_h - I_h v\|_{L^\infty(K)} + \|v_h - I_h v\|_{W^{1,\infty}(K)} \leq Ch^{-1}\delta\|\nabla v\|_{L^\infty(S_h)}. \tag{4.11}$$

Here $S_h := \cup_{\hat{x} \in \partial K \cap \mathcal{N}_h^0} \bar{B}_\delta(\hat{x})$ so that in particular a straight line from each $\hat{x} \in \partial K \cap \mathcal{N}_h^0$ to $\partial\Omega$ is contained in S_h.

Its remains to translate these elementwise estimates to the stated balls. We define $c := 1 + C_\delta$ where $C_\delta > 0$ is such that $\delta \le C_\delta h^2 \le C_\delta h$. Let $y \in \Omega$ and $d \ge ch$. We define

$$B_{\mathcal{K}_h} := \operatorname{int} \bigcup_{\substack{K \in \mathcal{K}_h, \\ K \cap B \neq \emptyset}} \bar{K},$$

the "discrete neighborhood" of B. We then have for any $x \in B_{\mathcal{K}_h}$, since $|x - y| \le d + h \le 2d$,

$$D_h \subset B_{\mathcal{K}_h} \subset D'.$$

Now let $K \subset B_{\mathcal{K}_h}$. If $K \notin L_h$ we have by (4.6) the estimate

$$h^{-1}\|v - v_h\|_{L^\infty(K)} + \|v - v_h\|_{W^{1,\infty}(K)} \le Ch\|v\|_{W^{2,\infty}(D')} + Ch^{-1}\delta\|v\|_{W^{1,\infty}(D')}. \tag{4.12}$$

If $K \in L_h$ we have by (4.9)

$$h^{-1}\|v - v_h\|_{L^\infty(K)} + \|v - v_h\|_{W^{1,\infty}(K)} \le Ch\|v\|_{W^{2,\infty}(D')} + Ch^{-1}\delta\|v\|_{W^{1,\infty}(S_h)}.$$

If we can show that $S_h = S_h(K) \subset D'$ holds, this inequality and (4.12) conclude the proof.

For any $\hat{y} \in S_h$ there exists by definition an $\hat{x} \in \partial K$ such that $\hat{y} \in \bar{B}_\delta(\hat{x})$. Because $K \in B_{\mathcal{K}_h}$ there is an $x_B \in B \cap \bar{K}$. Then we have

$$|y - \hat{y}| \le |y - x_B| + |x_B - \hat{x}| + |\hat{x} - \hat{y}| \le d + h + \delta \le d + (1 + C_\delta)h \le 2d.$$

This implies the inclusion $S_h \subset D'$ and concludes the proof. □

Proposition 4.17 *Let Assumption 4.14 be satisfied. By the uniform Lipschitz domain property we have $a, b > 0$, affine transformations $S_1, \dots, S_m$ and mappings $f_1, \dots, f_m, f_1^h, \dots, f_m^h : (-a, a)^{N-1} \to (-b, b)$ such that these quantities satisfy Definition 1.5, the local representations of boundaries via the f_i and f_i^h, for Ω and Ω_h simualtenously.*

Then we have the following: there exists an $\epsilon > 0$ such that

$$\bar{\Omega} \setminus \operatorname{int}(\Omega_h) \subset \cup_{i=1}^m S_i([-a + \epsilon, a - \epsilon]^{N-1} \times [-b + \epsilon, b - \epsilon]).$$

There also exists an $\epsilon' > 0$ such that for each $i \in \{1, 2, \dots, m\}$ and all $x_a \in [-a + \epsilon, a - \epsilon]^{N-1}$ we have $f_i(x_a) \ge -b + \epsilon'$.

Proof. We first show $\bar{\Omega} \setminus \operatorname{int} \Omega_h \subset \cup_{i=1}^m S_i((-a, a)^{N-1} \times (-b, b)) =: \cup_{i=1}^m R_i$. Assume there was an $x \in \bar{\Omega} \setminus \operatorname{int}(\Omega_h)$ such that $x \notin \cup_{i=1}^m R_i$. Then, since Ω_h is compact, there has to exist some $x_h \in \partial\Omega_h$ such that $\operatorname{dist}(x, \Omega_h) = |x - x_h|$. We now take note of the set $\Omega_h \cap \cup_{i=1}^m R_i$, the area indicated green in Figure 4.2. By the placement of x outside of $\cup_{i=1}^m R_i$ it is easy to see that

$$\operatorname{dist}(x, \Omega_h) \le \operatorname{dist}(x, \Omega_h \cap \cup_{i=1}^m R_i) < \operatorname{dist}(x, \partial\Omega_h \cap \cup_{i=1}^m R_i).$$

Therefore $x_h \notin \partial\Omega_h \cap \cup_{i=1}^m R_i$. But by assumption we do have $\partial\Omega_h \cap \cup_{i=1}^m R_i = \partial\Omega_h$, which is now a contradiction to $x_h \in \partial\Omega_h$. We therefore can now operate with the knowledge that $\bar{\Omega} \setminus \operatorname{int}(\Omega_h) \subset \cup_{i=1}^m R_i$.

We now argue that there exists a $\epsilon > 0$ such that

$$\bar{\Omega} \setminus \operatorname{int}(\Omega_h) \subset \cup_{i=1}^m S_i([-a + \epsilon, a - \epsilon]^{N-1} \times [-b + \epsilon, b - \epsilon]) =: \cup_{i=1}^m R_i^\epsilon.$$

Assume for all $n \in \mathbb{N}$ there was an $x^n \in \bar{\Omega} \setminus \operatorname{int}(\Omega_h)$ such that $x^n \notin \cup_{i=1}^m R_i^{\frac{1}{n}}$. By the compactness of $\bar{\Omega} \setminus \operatorname{int}(\Omega_h)$ we may assume, after going to a subsequence with the same name,

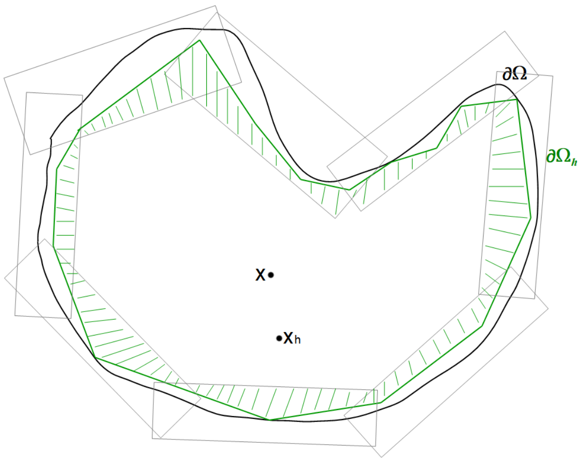


Figure 4.2: The boundaries of $\partial\Omega$ and $\partial\Omega_h$. The grey rectangles represent the open cuboids $R_i = S_i((-a,a)^{N-1} \times (-b,b))$. The area indicated green is $\Omega_h \cap \cup_{i=1}^m R_i$. The points x and x_h indicate the situation of the proof that $\Omega \setminus \Omega_h \subset \cup_{i=1}^m R_i$.

that there exists an $x \in \bar{\Omega} \setminus \operatorname{int}(\Omega_h)$ such that $\lim_{n\to\infty} x_n = x$. We have $x \in \bar{\Omega} \setminus \operatorname{int}(\Omega_h) \subset \cup_{i=1}^m S_i((-a,a)^{N-1} \times (-b,b))$. The sets on the right hand side are open and the S_i are affine transformations. Thus there exists a $\eta > 0$ such that $x \in \cup_{i=1}^m S_i((-a+\eta, a-\eta)^{N-1} \times (-b+\eta, b-\eta))$. By the convergence of $(x^n)_{n\in\mathbb{N}}$ to x we have for n large enough that $x^n \in \cup_{i=1}^m S_i((-a+\eta, a-\eta)^{N-1} \times (-b+\eta, b-\eta)) \subset \cup_{i=1}^m R_i^\eta$. This is a contradiction to the choice of x^n for n large enough.

The claim of the f_i follows from the fact that $[-a+\epsilon, a-\epsilon]^{N-1}$ is compact and thus each continuous $f_i : [-a+\epsilon, a-\epsilon]^{N-1} \to (-b,b)$ takes a mimimum $f(\hat{x}_a^i) > -b$ at some $\hat{x}_a^i \in [-a+\epsilon, a-\epsilon]^{N-1}$. We can thus choose $\epsilon' := \min_{i=1,2,\dots,m}(f(\hat{x}_a^i)) + b > 0$. □

This next statement about lower order global approximation corresponds to [SW82, Assumption A.5].

Proposition 4.18 *Let Assumption 4.14 be satisfied. Let $h \in (0,1]$.*

Then there exists $C > 0$ such that for $v \in H^2(\Omega) \cap H_0^1(\Omega)$ there exists $v_h \in V_h$ such that

$$h^{-1}\|v - v_h\|_{L^2(\Omega)} + \|v - v_h\|_{H^1(\Omega)} + h\|v - v_h\|_{H^{2,h}(\Omega_h)} \le Ch\|v\|_{H^2(\Omega)}. \tag{4.13}$$

Here $H^{2,h}(\Omega_h) := W^{2,2,h}(\Omega_h)$ as defined in Definition 4.4.

Proof. We argue for smooth v since the general statement will follow via density from Proposition 8.9. Let $v \in C^{2,\alpha}(\bar{\Omega}) \cap H_0^1(\Omega)$ and define v_h as in Proposition 4.15. *In the first step we will show that the difference $v - v_h$ is quite easily estimated on Ω_h by comparing v_h with the nodal interpolant $I_h v$.* Because $h \le 1$ we see that

$$\begin{aligned}
&h^{-1}\|v - v_h\|_{L^2(\Omega)} + \|v - v_h\|_{H^1(\Omega)} + h\|v - v_h\|_{H^{2,h}(\Omega_h)} \\
&\le h^{-1}\|v - v_h\|_{L^2(\Omega)} + |v - v_h|_{H^1(\Omega)} + h|v - v_h|_{H^{2,h}(\Omega_h)} \\
&\le h^{-1}\|v - I_h v\|_{L^2(\Omega)} + |v - I_h v|_{H^1(\Omega)} + h|v|_{H^{2,h}(\Omega_h)} \\
&\quad + h^{-1}\|I_h v - v_h\|_{L^2(\Omega_h)} + |I_h v - v_h|_{H^1(\Omega_h)}.
\end{aligned}$$

Here $|\cdot|_{H^1(\Omega)}$ and $|\cdot|_{H^{2,h}(\Omega)}$ refer to the seminorms on their respective spaces and only involve the highest order derivatives.

By Lemma 4.10 and Proposition 4.12, note that $h \leq Ch_K$ by quasi-uniformity, we find

$$\begin{aligned} &h^{-1}\|v - v_h\|_{L^2(\Omega)} + \|v - v_h\|_{H^1(\Omega)} + h\|v - v_h\|_{H^{2,h}(\Omega_h)} \\ &\leq h^{-1}\|v\|_{L^2(\Omega\setminus\Omega_h)} + |v|_{H^1(\Omega\setminus\Omega_h)} + h\|v\|_{H^2(\Omega)} + Ch^{-1}\|I_h v - v_h\|_{L^2(\Omega_h)}. \end{aligned} \tag{4.14}$$

As in the proof of (4.10) we see that

$$\|I_h v - v_h\|_{L^2(\Omega)} \leq Ch^{\frac{N}{2}} \max_{x \in \partial K \cap \mathcal{N}_h^0} |v(x)|.$$

Because $v \in H^2(\Omega) \subset W^{1,4}(\Omega)$ by $N \leq 3$ and standard Sobolev embeddings, we have, by Morrey's inequality, see [Ada75, Chapter V, Theorem 5.4 Part II], that

$$\|v\|_{C^{0,1-N/4}(\Omega)} \leq C\|v\|_{H^2(\Omega)}.$$

Thus, we find by the 0-boundary values of v that

$$\|I_h v - v_h\|_{L^2(\Omega)} \leq Ch^{\frac{N}{2}}\delta^{1-\frac{N}{4}}\|v\|_{H^2(\Omega)} \leq Ch^{\frac{N}{2}+2-\frac{N}{2}}\|v\|_{H^2(\Omega)} = Ch^2\|v\|_{H^2(\Omega)}. \tag{4.15}$$

We proceed to make local arguments, similar to those in [LMWZ10, Lemma 2.1], but more rigorous. We use the same notation for the local boundary representation as in Proposition 4.17, cf. Definition 1.5, and abbreviate $S_i((-a,a)^{N-1} \times (-b,b)) =: R_i$ and $S_i([-a+\epsilon, a-\epsilon]^{N-1} \times [-b+\epsilon, b-\epsilon]) =: R_i^\epsilon$. As a reminder: the S_i are the affine transformations into the local coordinate systems. Thus by Proposition 4.17 there is an $\epsilon > 0$ be such that $\bar{\Omega} \setminus \mathrm{int}(\Omega_h) \subset \cup_{i=1}^m R_i^\epsilon \subset \cup_{i=1}^m R_i$. We therefore find

$$h^{-1}\|v\|_{L^2(\Omega\setminus\Omega_h)} + |v|_{H^1(\Omega\setminus\Omega_h)} \leq \sum_{i=1}^m (h^{-1}\|v\|_{L^2(R_i\cap(\Omega\setminus\Omega_h))} + |v|_{H^1(R_i^\epsilon\cap(\Omega\setminus\Omega_h))}). \tag{4.16}$$

Now, let $i \in \{1, \ldots, m\}$ be fixed. Transforming everything into the local system and using the Jacobi transformation formula, e.g. [BK15, Chapter 10], we have, with $\tilde{v} = v \circ S_i$,

$$\|v\|^2_{L^2(R_i\cap(\Omega\setminus\Omega_h))} = |\det(DS_i)| \int_{(-a,a)^{N-1}} \int_{f_i^h(x_a)}^{f_i(x_a)} \tilde{v}(x_a, r)^2 \, dr \, dx_a.$$

Recall that $f_i, f_i^h : (-a,a)^{N-1} \to (-b,b)$ are the functions locally representing $\partial\Omega$ and $\partial\Omega_h$. As $v = 0$ a.e on $\partial\Omega$ we can use the fundamental theorem of calculus, keeping $v \in C^2(\bar{\Omega})$ in mind, to find

$$\|v\|^2_{L^2(R_i\cap(\Omega\setminus\Omega_h))} \leq C \int_{(-a,a)^{N-1}} \int_{f_i^h(x_a)}^{f_i(x_a)} \left(-\int_r^{f_i(x_a)} \partial_{x_N}\tilde{v}(x_a, s) \, ds \right)^2 dr \, dx_a.$$

Continuing with Proposition 4.9 and the Cauchy-Schwarz inequality we find

$$\begin{aligned} \|v\|^2_{L^2(R_i\cap(\Omega\setminus\Omega_h))} &\leq C \int_{(-a,a)^{N-1}} \delta \left(\int_{f_i^h(x_a)}^{f_i(x_a)} |\partial_{x_N}\tilde{v}(x_a, s)| \, ds \right)^2 dx_a \\ &\leq C\delta \int_{(-a,a)^{N-1}} \|\nabla\tilde{v}(x_a, \cdot)\|^2_{L^2((f_i^h(x_a), f_i(x_a)))} \|1\|^2_{L^2((f_i^h(x_a), f_i(x_a)))} \, dx_a \\ &\leq C\delta^2 \int_{(-a,a)^{N-1}} \|\nabla\tilde{v}(x_a, \cdot)\|^2_{L^2((f_i^h(x_a), f_i(x_a)))} \, dx_a. \end{aligned}$$

Transforming everything back, S_i is again only entering in form of Jacobi matrices, yields $\|v\|_{L^2(R_i\cap(\Omega\setminus\Omega_h))} \leq C\delta\|v\|_{H^1(R_i\cap(\Omega\setminus\Omega_h))}$. By assumption $\delta \leq Ch^2$ and we arrive at

$$\|v\|_{L^2(R_i\cap(\Omega\setminus\Omega_h))} \leq Ch^2\|v\|_{H^2(\Omega)}. \tag{4.17}$$

For the second group of summands in (4.16) *we again transform everything* to see

$$|v|^2_{H^1(R_i^\epsilon\cap(\Omega\setminus\Omega_h))} \leq C\int_{(-a+\epsilon,a-\epsilon)^{N-1}}\int_{f_i^h(x_a)}^{f_i(x_a)} |\nabla\tilde{v}(x_a,r)|^2\,dr\,dx_a.$$

Applying Proposition 8.4, a specific embedding statement, and Proposition 4.9 we find

$$\begin{aligned}|v|^2_{H^1(R_i^\epsilon\cap(\Omega\setminus\Omega_h))} &\leq C\delta\int_{(-a+\epsilon,a-\epsilon)^{N-1}} \|\nabla\tilde{v}(x_a,\cdot)\|^2_{L^\infty((-b,f_i(x_a)))}\,dx_a\\ &\leq C\delta\int_{(-a,a)^{N-1}} \max((f_i(x_a)+b),(f_i(x_a)+b)^{-1})\|\nabla\tilde{v}(x_a,\cdot)\|^2_{H^1((-b,f_i(x_a)))}\,dx_a\end{aligned}$$

From Proposition 4.17 we have the existence of an $\epsilon' > 0$ such that $f_i(x_a) \geq -b+\epsilon'$ for all $x_a \in [-a+\epsilon, a-\epsilon]^{N-1}$. So in particular $\max((f_i(x_a)+b),(f_i(x_a)+b)^{-1}) \leq 2b + (\epsilon')^{-1}$ for all $x_a \in [-a+\epsilon, a-\epsilon]^{N-1}$. Thus we find the following upper bound

$$|v|^2_{H^1(R_i\cap(\Omega\setminus\Omega_h))} \leq C\delta\int_{(-a,a)^{N-1}} \|\nabla\tilde{v}(x_a,\cdot)\|^2_{H^1((-b,f_i(x_a)))}\,dx_a$$

Transforming everything back we get

$$|v|_{H^1(R_i\cap(\Omega\setminus\Omega_h))} \leq Ch\|v\|_{H^2(R_i\cap\Omega)}.$$

This and (4.17) inserted into (4.16) shows $h^{-1}\|v\|_{L^2(\Omega\setminus\Omega_h)} + |v|_{H^1(\Omega\setminus\Omega_h)} \leq Ch^2\|v\|_{H^2(\Omega)}$. This in conjunction with (4.15) and (4.14) yields (4.13) for $v \in C^{2,\alpha}(\bar{\Omega}) \cap H_0^1(\Omega)$.

To see the general statement let $v \in H^2(\Omega) \cap H_0^1(\Omega)$ and $(v_n)_{n\in\mathbb{N}} \subset C^{2,\alpha}(\bar{\Omega}) \cap H_0^1(\Omega)$ be a sequence converging to it in $H^2(\Omega)$ by Proposition 8.9. Let $(v_n^h)_{n\in\mathbb{N}} \subset V_h$ be the sequence of functions satisfying (4.13) for their corresponding v_n. Then it is clear that $(\|v_n^h\|_{H^1(\Omega)})_{n\in\mathbb{N}}$ is bounded. Since V_h is finite dimensional, we may assume (after possibly going to a subsequence) that $v_n^h \xrightarrow{n\to\infty} v^h$ for some $v^h \in V_h$. Thus taking the limit in

$$h^{-1}\|v_n - v_n^h\|_{L^2(\Omega)} + \|v_n - v_n^h\|_{H^1(\Omega)} + h\|v_n\|_{H^{2,h}(\Omega_h)} \leq Ch\|v_n\|_{H^2(\Omega)}$$

yields the claim, since $\|v - v^h\|_{H^{2,h}(\Omega_h)} = \|v\|_{H^{2,h}(\Omega_h)}$. □

Remark 4.19 In the previous proposition it is not possible to extend the result to higher dimension by the same techniques. For $N = 4$ one would still have the embedding $H^2(\Omega) \subset W^{1,4}(\Omega)$, but Morrey's inequality would no longer be applicable, since $1 \cdot 4 = N$.

The following property is called superapproximation and corresponds to [SW82, Assumption A.6].

Proposition 4.20 *Assume Assumption 4.14 holds true. There exist $C, c > 0$ such that the following properties are satisfied. Let $d \geq ch$ and $D_1 \subset \mathbb{R}^N$ be an open set. We define $D_{i+1} := \{x \in \mathbb{R}^N : \operatorname{dist}(x, D_i) \leq d\}$ for $i = 1, 2, 3$. We define $D_i^h := D_i \cap \Omega_h$.*

Let $\omega \in C_c^\infty(D_3)$ with

$$\|\omega\|_{W^{k,\infty}(D_3)} \leq Ld^{-k} \text{ for } k = 0, 1, 2 \text{ and } \omega = 1 \text{ on } D_2.$$

Then for any $v_h \in V_h$ there is a $\chi_h \in V_h$ with $\operatorname{supp}\chi_h \subset D_4^h$ such that

$$\|\omega^2 v_h - \chi_h\|_{H^1(D_4^h)} \le CLh\left(d^{-2}\|v_h\|_{L^2(D_4^h\setminus D_1)} + d^{-1}\|v_h\|_{H^1(D_4^h\setminus D_1)}\right).$$

The constant $C > 0$ does not depend on Ω.

Proof. We choose $\chi_h = I_h(\omega^2 v_h)$. By Lemma 4.10 we have for any $K \in \mathcal{K}_h$

$$|\omega^2 v_h - I_h(\omega^2 v_h)|^2_{H^1(K)} \le Ch^2|\omega^2 v_h|^2_{H^2(K)}.$$

By $\operatorname{supp}\omega \subset D_3$ we see that $\operatorname{supp}(I_h(\omega^2 v_h)) \subset \bigcup_{K\in\mathcal{K}_h, K\cap D_3^h\neq\emptyset} K$. Thus we deduce

$$\begin{aligned}|\omega^2 v_h - I_h(\omega^2 v_h)|^2_{H^1(D_4^h)} &\le Ch^2 \sum_{\substack{K\in\mathcal{K}_h\\ K\cap D_3^h\neq\emptyset}} |\omega^2 v_h|^2_{H^2(K)}\\ &\le Ch^2 \sum_{\substack{K\in\mathcal{K}_h\\ K\cap D_3^h\neq\emptyset}} \|\nabla^2(\omega^2)v_h + \nabla(\omega^2)\nabla v_h^T\|^2_{L^2(K)}.\end{aligned}$$

Since $\nabla\omega = 0$ on D_2 we deduce for $d \ge h$

$$\begin{aligned}&\|\omega^2 v_h - I_h(\omega^2 v_h)\|^2_{H^1(D_4^h)}\\ &\quad\le Ch^2 \sum_{\substack{K\in\mathcal{K}_h\\ K\cap D_3^h\neq\emptyset\\ K\cap D_1^h=\emptyset}} \|\nabla^2(\omega^2)\|^2_{L^\infty(K)}\|v_h\|^2_{L^2(K)} + \|\nabla(\omega^2)\|^2_{L^\infty(K)}\|\nabla v_h\|^2_{L^2(K)}\\ &\quad\le Ch^2 \sum_{\substack{K\in\mathcal{K}_h\\ K\cap D_3^h\neq\emptyset\\ K\cap D_1^h=\emptyset}} L^2 d^{-4}\|v_h\|^2_{L^2(K)} + L^2 d^{-2}\|\nabla v_h\|^2_{L^2(K)}.\end{aligned}$$

Estimating this from above yields, since $d \ge h$,

$$\|\omega^2 v_h - I_h(\omega^2 v_h)\|^2_{H^1(D_4^h)} \le CL^2h^2\left(d^{-4}\|v_h\|^2_{L^2(D_4^h\setminus D_1)} + d^{-2}\|\nabla v_h\|^2_{L^2(D_4^h\setminus D_1)}\right).$$

Taking the root yields the claim. □

We lastly state a generalization of [SW82, Theorem 4.1]. This generalization is required in [BTW03], while the special case $D = B_d(y) \cap \Omega_h$ and $D_d = B_{2d}(y) \cap \Omega_h$ is used in [SW82, Theorem 4.1].

Proposition 4.21 *Let Assumption 4.14 hold. Let $h \in (0, \frac{1}{2})$. Let A have $C^{0,1}(\Omega)$-coefficients. Let $D \subset \Omega_h$ be open. For $d > 0$ we define $D_d := \{x \in \Omega_h : \operatorname{dist}(x, D) \le d\}$. Then there exist $C, c > 0$ such that for $d \ge ch$ and any $v \in H_0^1(\Omega)$ and $v_h \in V_h$ satisfying*

$$(A_h(v - v_h), \varphi_h)_{V_h^*, V_h} = 0 \text{ for } \varphi_h \in V_h \text{ with } \operatorname{supp} v_h \subset D_d$$

we have

$$\|\nabla(v_h - v)\|_{L^2(D)} \le C \inf_{\varphi_h\in V_h}(\|\nabla(v-\varphi_h)\|_{L^2(D_d)} + d^{-1}\|v-\varphi_h\|_{L^2(D_d)} + d^{-1}\|v_h - v\|_{L^2(D_d)}).$$

Proof. The proof was done in [SW82, Theorem 4.1, Remark 4.1] for balls, C^∞-domains and the Laplacian. From the operator one only needs ellipticity, boundedness and higher elliptic $H^2(\Omega)$-regularity, which is provided by our A and Ω; see Theorem 8.23. Checking the proof one can see that our assumptions on Ω and A are sufficient, as Proposition 4.12, Proposition 4.18 and Proposition 4.20 still hold. The fact that balls are used in the proof is basically incidental and any open set will do the job. □

4.1.4 Green's Functions

As we will have to dig deep into the nature of the numerical analysis of elliptic equations, we obviously have to provide some basic results on Green's functions. The following statement is the main result of [Sol71]. To the best of our knowledge it is the theorem with the least regularity assumptions on Ω and A that still provides estimates for sufficiently many derivatives of the Green's function for our purposes.

Proposition 4.22 *Assume $N \geq 2$. Assume Ω is a $C^{3,\alpha}$-domain and A is symmetric, uniformly elliptic and has coefficients in $C^{1,\alpha}(\bar{\Omega})$ for some $\alpha \in (0,1)$. Then there is a Green's function $G(x,y)$, i.e. for any $y \in \Omega$ the function $G(\cdot,y)$ solves*

$$\begin{cases} AG(\cdot,y) & = \delta_y \text{ in } \Omega, \\ G(\cdot,y)|_{\partial\Omega} & = 0, \end{cases}$$

in a distributional sense. It satisfies for any $k,l \in \mathbb{N}_0$, $k \leq 2$, $l \leq 1$

$$|\nabla_x^k \nabla_y^l G(x,y)| \leq C \begin{cases} |\ln|x-y|| & \text{if } 2-N-k-l=0, \\ |x-y|^{2-N-k-l} & \text{else.} \end{cases}$$

Remark 4.23 For very smooth domains and coefficients see also the results of [Kra67, GW82, Väh12]. For similar statements for convex domains see [Fro93].

4.1.5 L^∞-stability of Ritz Projections

This section is devoted to generalizing [SW82, Theorem 5.1] in Theorem 4.29. [SW82, Theorem 5.1] was proven for C^∞ domains and the Laplace operator. Yet, with the previous results from Sections 4.1.3 and 4.1.4 we can generalize their statements.

Proposition 4.24 *Let $(\mathcal{K}_h)_{h\in(0,1]}$ be a family of meshes satisfying Assumption 4.14. Then there exists a $c' > 0$ such that the following holds: let $u \in C(\bar{\Omega}) \cap H_0^1(\Omega)$ and let u_h be its Ritz projection. Let $x_0 \in \bar{\Omega}_h$ with*

$$\|u - u_h\|_{L^\infty(\Omega_h)} = |u(x_0) - u_h(x_0)|.$$

If $\operatorname{dist}(x_0, \partial\Omega_h) \leq c'h$ *then*

$$\|u - u_h\|_{L^\infty(\Omega_h)} \leq 2\|u\|_{L^\infty(\Omega_h)}.$$

Proof. This is just [SW82, Lemma 5.1]. Their proof does apply without any changes. □

Notation 4.25 For the rest of Section 4.1.5 we use the following notation. Let $u \in C(\bar{\Omega}) \cap H_0^1(\Omega)$ and u_h its Ritz projection. The point $x_0 \in \bar{\Omega}_h$ is chosen such that: $\|u - u_h\|_{L^\infty(\Omega_h)} = |u(x_0) - u_h(x_0)|$. Let c' be from Proposition 4.24 and K_0 denote a cell containing x_0. We define

$$K_0' := \{x \in K : \operatorname{dist}(x, \partial\Omega_h) \geq c'h\}$$

and assume $x_0 \in K_0'$. We may assume c' to be small enough so that Proposition 4.12 applies to a subset of $K_0' \subset K_0$.

Theorem 4.29, the main result of this section, is formulated without use of this notation and can be cited freely without confusion. The notation is discarded after the proof of Theorem 4.29.

Proposition 4.26 *Let $(\mathcal{K}_h)_{h\in(0,1]}$ a family of meshes satisfying Assumption 4.14. In the situation of Notation 4.25 we have*

$$\|u-u_h\|_{L^\infty(\Omega_h)} \le C\|u\|_{L^\infty(\Omega_h)} + Ch^{-\frac{N}{2}}\|u-u_h\|_{L^2(K_0')}.$$

Proof. The proof is given in [SW82, (5.4)]. □

It remains to estimate $\|u-u_h\|_{L^2(K_0')}$. If we can do this, Proposition 4.24 and Proposition 4.26 give an estimate of $\|u-u_h\|_{L^\infty(\Omega_h)}$. The following statement can be found in [SW82, Lemma 5.2] for the Laplacian and a smooth domain.

Proposition 4.27 *Assume Ω is a $C^{3,\alpha}$-domain for some $\alpha\in(0,1)$ and A is symmetric and has coefficients in $C^{1,1}(\Omega)$. Let $(\mathcal{K}_h)_{h\in(0,1/2)}$ be a family of meshes such that Assumption 4.14 is satisfied. Additionally assume that $(\Omega_h)_{h\in(0,1]}$ and Ω are uniform Lipschitz domains.*

Let $\varphi\in C_c^\infty(K_0')$ with $\|\varphi\|_{L^2(K_0')}=1$. Let v be defined by

$$\begin{cases} Av=\varphi \text{ in } \Omega, \\ v|_{\partial\Omega}=0. \end{cases} \tag{4.18}$$

Then we have

$$\int_{\partial\Omega_h} |\partial_{\nu_A} v|\,dS \le Ch^{\frac{N}{2}}, \quad \int_{\Omega\setminus\Omega_h} |\nabla v|\,dS \le C\delta h^{\frac{N}{2}}.$$

Here $\partial_{\nu_A}v$ is the co-normal derivative $\nu^T A\nabla v$, where ν is the outer normal to $\partial\Omega$ and A the coefficient matrix of the operator with the same symbol.

Proof. The proof of [SW82, Lemma 5.2] is immediately applicable provided we have pointwise estimates for the Green's function up to second order derivatives in Proposition 4.22. Our assumptions are more than enough for Proposition 4.22 to apply. □

Proposition 4.28 *Assume Ω is a $C^{3,\alpha}$-domain for some $\alpha\in(0,1)$ and A is symmetric and has coefficients in $C^{1,1}(\bar\Omega)$. Let $(\mathcal{K}_h)_{h\in(0,1)}$ satisfy Assumption 4.14.*

Let $\varphi\in C_0^\infty(K_0')$ with $\|\varphi\|_{L^2(K_0')}=1$. Let v be defined by

$$\begin{cases} Av &= \varphi \text{ in } \Omega, \\ v|_{\partial\Omega} &= 0. \end{cases} \tag{4.19}$$

Then

$$\|\nabla(v-v_h)\|_{W^{1,1,h}(\Omega_h)} + h^{-1}\|\nabla(v-v_h)\|_{L^1(\Omega_h)} \le C|\ln h|h^{\frac{N}{2}}.$$

Proof. The proof of [SW82, Lemma 5.3] immediately applies. It is based "only" on Propositions 4.12, 4.15, 4.18, 4.20, 4.21, 4.26 and 4.27 and the estimates for the Green's function for one and two derivatives in the same variable, see Proposition 4.22. All their prerequisites are satisfied by Assumption 4.14. □

The next theorem is now finally the culmination of the previous small lemmas and propositions. It is a generalization of [SW82, Theorem 5.1] with respect to domain and operator regularity.

Theorem 4.29 *Assume Ω is a $C^{3,\alpha}$-domain for some $\alpha \in (0,1)$ and A has $C^{1,1}(\bar{\Omega})$ coefficients. Let Assumption 4.14 be satisfied for a family of meshes $(\mathcal{K}_h)_{h\in(0,1/2)}$. Additionally assume that $(\Omega_h)_{h\in(0,1/2)}$ and Ω are uniform Lipschitz domains.*

For any $u \in H_0^1(\Omega) \cap C(\bar{\Omega})$ and its Ritz-projection u_h we have

$$\|u - u_h\|_{L^\infty(\Omega_h)} \le C|\ln h| \inf_{\varphi_h \in V_h} \|u - \varphi_h\|_{L^\infty(\Omega_h)}.$$

Proof. Let $x_0 \in \bar{\Omega}_h$ such that $|u(x_0) - u_h(x_0)| = \|u - u_h\|_{L^\infty(\Omega_h)}$. If $\operatorname{dist}(x_0, \partial\Omega) \le c'h$ with the c' from Notation 4.25 we have by Proposition 4.24 $\|u - u_h\|_{L^\infty(\Omega_h)} \le 2\|u\|_{L^\infty(\Omega_h)}$. If $\operatorname{dist}(x_0, \partial\Omega) \ge c'h$ we have by Proposition 4.26

$$\|u - u_h\|_{L^\infty(\Omega_h)} \le C\|u\|_{L^\infty(\Omega_h)} + Ch^{-\frac{N}{2}}\|u - u_h\|_{L^2(K_0')} \tag{4.20}$$

with the K_0' from Notation 4.25. Now let $\varphi \in C_c^\infty(K_0')$ with $\|\varphi\|_{L^2(K_0')} = 1$ and $v \in H^2(\Omega)$ defined as in (4.18). We see that

$$\begin{aligned}\int_{K_0'} (u - u_h)\varphi\, dx &= \int_{\Omega_h} (u - u_h)Av\, dx,\\ &= \int_{\partial\Omega_h} (u - u_h)\partial_{\nu_A} v\, dS - a_{\Omega_h}(u - u_h, v)\\ &= \int_{\partial\Omega_h} u\, \partial_{\nu_A} v\, dS - a_{\Omega_h}(u - u_h, v).\end{aligned} \tag{4.21}$$

We define v_h as the Ritz projection of v and have via Galerkin-orthogonality and extension by 0 that

$$\begin{aligned} a_{\Omega_h}(u - u_h, v_h) &= a_\Omega(u - u_h, v_h) = 0,\\ a_{\Omega_h}(u_h, v - v_h) &= a_\Omega(u_h, v - v_h) = 0.\end{aligned}$$

Thus (4.21) yields

$$\int_{K_0'} (u - u_h)\varphi\, dx = \int_{\partial\Omega_h} u\, \partial_{\nu_A} v\, dS - a_{\Omega_h}(u, v - v_h). \tag{4.22}$$

The first term is estimated using Proposition 4.27:

$$\left|\int_{\partial\Omega_h} u\, \partial_{\nu_A} v\, dS\right| \le \|u\|_{L^\infty(\Omega_h)} \int_{\partial\Omega_h} |\partial_{\nu_A} v|\, dS \le Ch^{\frac{N}{2}}\|u\|_{L^\infty(\Omega_h)}. \tag{4.23}$$

The second term in (4.22) is more difficult:

$$\begin{aligned} a_{\Omega_h}(u, v - v_h) &= \sum_{K\in\mathcal{K}_h} a_K(u, v - v_h)\\ &= \sum_{K\in\mathcal{K}_h} \int_K uAv\, dx - \int_K uAv_h\, dx - \int_{\partial K} u\, \partial_{\nu_A}(v - v_h)\, dx := T_1 + T_2 + T_3.\end{aligned} \tag{4.24}$$

The term T_2 simply vanishes as $Av_h = 0$ on each cell $K \in \mathcal{K}_h$.

The term T_1 in (4.24) is treated easily by realizing that $Av = \varphi$ and $\operatorname{supp}(\varphi) \subset K_0'$ which shows that

$$\left|\sum_{K\in\mathcal{K}_h} \int_K uAv\, dx\right| \le \int_{K_0'} |u\varphi|\, dx \le \|u\|_{L^\infty(\Omega)}\|\varphi\|_{L^1(K_0)} \le h^{\frac{N}{2}}\|u\|_{L^\infty(\Omega)}.$$

Here we used $\|\varphi\|_{L^2(K_0')} = 1$ and Hölder's inequality.

The term T_3 is estimated by using Proposition 4.11 applied to $\nabla(v - v_h) \in W^{1,1,h}(\Omega_h)$ and an application of Proposition 4.28:

$$\begin{aligned}\left|\sum_{K\in\mathcal{K}_h} \int_{\partial K} u\, \partial_{\nu_A}(v - v_h)\, dx\right| &\leq C\|u\|_{L^\infty(\Omega_h)} \sum_{K\in\mathcal{K}_h} \Big(h^{-1}\|\nabla(v - v_h)\|_{L^1(K)} + \|\nabla(v - v_h)\|_{W^{1,1}(K)}\Big) \\ &\leq C\|u\|_{L^\infty(\Omega)} |\ln h| h^{\frac{N}{2}}.\end{aligned}$$

We have now shown that all three terms, T_1, T_2 and T_3, are bounded by $C|\ln h|h^{\frac{N}{2}}\|u\|_{L^\infty(\Omega)}$. This, (4.24) and (4.23) inserted into (4.21) results in

$$\int_{K_0'} (u - u_h)\varphi\, dx \leq C|\ln h|h^{\frac{N}{2}}\|u\|_{L^\infty(\Omega)}.$$

As $\varphi \in C_c^\infty(K_0')$ was arbitrary with $\|\varphi\|_{L^2(K_0')} = 1$ this entails

$$\|u - u_h\|_{L^2(K_0')} \leq C|\ln h|h^{\frac{N}{2}}\|u\|_{L^\infty(\Omega)}.$$

(4.20) now finally shows

$$\|u - u_h\|_{L^\infty(\Omega_h)} \leq C|\ln h|\|u\|_{L^\infty(\Omega)}.$$

For any $\varphi_h \in V_h$ the Ritz projection of $(u - \varphi_h)$ is given by $(u_h - \varphi_h)$. Thus we have

$$\|u - u_h\|_{L^\infty(\Omega_h)} = \|(u - \varphi_h) - (u_h - \varphi_h)\|_{L^\infty(\Omega_h)} \leq C\|u - \varphi_h\|_{L^\infty(\Omega_h)}.$$

□

Remark 4.30 We would like to reiterate comments made in [SW82] regarding the case of non-conforming finite elements. So let all the assumptions of Theorem 4.29 be satisfied except $\Omega_h \subset \Omega$ and Ω being a $C^{3,\alpha}$-domain. Now consider a domain Ω_δ with δ from Assumption 4.14 that is $C^{3,\alpha}$-smooth for some $\alpha \in (0,1)$ and that satisfies $\Omega_\delta \supset \Omega \cup \Omega_h$ and

$$\max_{x\in\Omega_h} \operatorname{dist}(x, \partial\Omega_\delta) \leq C\delta \qquad \text{and} \qquad \max_{x\in\Omega} \operatorname{dist}(x, \partial\Omega_\delta) \leq C\delta.$$

For the sake of presentation we only consider $A := -\Delta$, one would have to take additional care of the coefficients of A in the general case.

Let u be given as the solution of

$$\begin{cases} -\Delta u = f \text{ in } \Omega, \\ u|_{\partial\Omega} = 0. \end{cases}$$

We now extend f by 0 onto Ω_δ and solve

$$\begin{cases} -\Delta u^\delta = f \text{ in } \Omega_\delta, \\ u^\delta|_{\partial\Omega_\delta} = 0. \end{cases}$$

Now the Ritz projection u_h^δ of u^δ satisfies:

$$\|u^\delta - u_h^\delta\|_{L^\infty(\Omega_h)} \leq C(\Omega^\delta)|\ln h| \inf_{\varphi_h\in V_h} \|u^\delta - \varphi_h\|_{L^\infty(\Omega^\delta)}.$$

The Ritz-projection of u is just $u_h^\delta|_{\Omega_h}$. To see this let $\varphi_h \in V_h$ and observe that

$$\begin{aligned}(f,\varphi_h)_{L^2(\Omega)} &= a_\Omega(u,\varphi_h) = a_{\Omega\cup\Omega_h}(u,\varphi_h),\\ &= (f,\varphi_h)_{L^2(\Omega^\delta)} = a_{\Omega^\delta}(u_h^\delta,\varphi_h) = a_{\Omega\cup\Omega_h}(u_h^\delta,\varphi_h).\end{aligned}$$

Thus

$$a_{\Omega\cup\Omega_h}(u-u_h^\delta,\varphi_h)=0 \qquad \forall \varphi_h \in V_h.$$

So we can write $u_h = u_h^\delta$ and get

$$\|u-u_h\|_{L^\infty(\Omega_h\cap\Omega)} \le \|u-u^\delta\|_{L^\infty(\Omega)} + C(\Omega^\delta)|\ln h| \inf_{\varphi_h\in V_h} \|u^\delta-\varphi_h\|_{L^\infty(\Omega_h)}.$$

By the maximum principle, e.g. [GT01, Theorem 8.1], and $-\Delta(u-u^\delta)=0$ on Ω we find:

$$\|u-u_h\|_{L^\infty(\Omega_h\cap\Omega)} \le \|u^\delta\|_{L^\infty(\partial\Omega)} + C(\Omega^\delta)|\ln h| \inf_{\varphi_h\in V_h} \|u^\delta-\varphi_h\|_{L^\infty(\Omega_h)}.$$

If now one exploits higher elliptic regularity for sufficiently nice f one can derive further estimates. If $f\in L^\infty(\Omega)$ we have by higher elliptic regularity, e.g. Theorem 8.23,

$$\|u^\delta\|_{W^{2,p}(\Omega)} \le C_{\not p}(\Omega^\delta)p\|f\|_{L^p(\Omega_\delta)} = C_{\not p}(\Omega^\delta)p\|f\|_{L^p(\Omega)}.$$

By standard interpolation theory we have

$$\inf_{\varphi_h\in V_h}\|u^\delta-\varphi_h\|_{L^\infty(\Omega_h)} \le C(\Omega^\delta)h^{2-\frac{N}{p}}p\|f\|_{L^p(\Omega)}.$$

Choosing $p=|\ln h|$ yields

$$\inf_{\varphi_h\in V_h}\|u^\delta-\varphi_h\|_{L^\infty(\Omega_h)} \le C(\Omega^\delta)|\ln h|h^2\|f\|_{L^\infty(\Omega)}.$$

For $p>N+1$ we also have

$$\|u^\delta\|_{C^1(\Omega^\delta)} \le C(\Omega^\delta)\|u^\delta\|_{W^{2,N+1}(\Omega^\delta)} \le C(\Omega^\delta)\|f\|_{L^\infty(\Omega)}.$$

Now the mean value theorem, $u^\delta|_{\partial\Omega^\delta}=0$ and $\max_{x\in\partial\Omega}\operatorname{dist}(x,\partial\Omega^\delta)\le Ch^2$ imply

$$\|u^\delta\|_{L^\infty(\partial\Omega)} \le C(\Omega^\delta)h^2\|f\|_{L^\infty(\Omega)}.$$

So in total

$$\|u-u_h\|_{L^\infty(\Omega_h\cap\Omega)} \le C(\Omega^\delta)|\ln h|h^2\|f\|_{L^\infty(\Omega)}.$$

So if one shows or presumes that $C(\Omega^\delta)$ stays bounded for $\delta\to 0$, we have the same convergence rate for the non-conforming case as for the conforming case. The behaviour of the constant $C(\Omega^\delta)$, with respect to Ω and Ω^δ, could theoretically be tracked throughout our proofs. The claim is made, but, to the authors knowledge, not proven below [SW82, (1.6)].

4.1.6 L^∞-norm Resolvent Estimates for Finite Element Operators

This section is devoted to generalize the results from [BTW03]. There all statements are given for convex C^∞-domains and $A=-\Delta$, which is more restrictive than we appreciate. We therefore trace the domain regularity and the properties of A throughout [BTW03] and note where changes are necessary.

In the following we need the notion of complex valued partial differential equations. They are defined by the same weak formulations except that the test functions and the solution are chosen from complex valued spaces. For example for $z \in \mathbb{C}$ the PDE

$$\begin{cases} Av + zv = \chi, \\ v|_{\partial\Omega} = 0, \end{cases}$$

has the weak formulation

$$\int_\Omega \nabla v^T A \nabla \bar{\varphi} + zv\bar{\varphi}\, dx = \int_\Omega \chi\bar{\varphi}\, dx \quad \forall \varphi \in H_0^1(\Omega, \mathbb{C}).$$

If $z \in \mathbb{R}$ and χ is real valued, we realize, by the linearity of the problem, that the solution v solves the real valued version of the PDE; just test with any $\varphi \in H_0^1(\Omega) \subset H_0^1(\Omega, \mathbb{C})$. So in fact introducing complex numbers into partial differential equations extends our used framework naturally.

First we collect some minor estimates which correspond to [BTW03, (2.1)-(2.6)], so that the reader may easily transfer their results to our situation.

Proposition 4.31 *Let $(\mathcal{K}_h)_{h\in(0,1]}$ be a family of meshes satisfying Assumption 4.14. We define $\tilde{V}_h := V_h + iV_h$. Then we have:*

1. *There is a $C > 0$, independent of h, such that for any $v_h \in \tilde{V}_h$, $p, q \in [1, \infty]$, $0 \le l \le k$ and $K \in \mathcal{K}_h$*

$$\|v_h\|_{W^{k,p}(K,\mathbb{C})} \le Ch^{l-k-N(\frac{1}{q}-\frac{1}{p})}\|v_h\|_{W^{l,q}(K,\mathbb{C})}$$

2. *The L^2-projection P_h of $L^2(\Omega, \mathbb{C})$ onto $\tilde{V}_h$ satisfies with a C independent of h*

$$\|P_h v\|_{L^\infty(\Omega,\mathbb{C})} \le C\|v\|_{L^\infty(\Omega_h,\mathbb{C})} \quad \forall v \in L^\infty(\Omega, \mathbb{C}),$$

3. *The L^2-projection P_h of $L^2(\Omega, \mathbb{C})$ onto $\tilde{V}_h$ satisfies with a C independent of h*

$$\|P_h v\|_{W^{1,\infty}(\Omega,\mathbb{C})} \le C\|v\|_{W^{1,\infty}(\Omega_h,\mathbb{C})} \quad \forall v \in W^{1,\infty}(\Omega, \mathbb{C}) \cap C_0(\bar{\Omega}, \mathbb{C}),$$

4. *Additionally assume we have for some $\alpha \in (0, 1)$ that Ω is a $C^{3,\alpha}$-domain and that A has $C^{1,\alpha}(\Omega)$-coefficients. Then there is a $C > 0$, independent of h, such that for any $u \in W^{1,\infty}(\Omega, \mathbb{C}) \cap C_0(\bar{\Omega}, \mathbb{C})$ and its complex Ritz-projection $u_h \in \tilde{V}_h$, defined via*

$$\int_\Omega (\nabla u - \nabla u_h)^T A \nabla \bar{v}_h\, dx = 0 \quad \forall v_h \in \tilde{V}_h,$$

we have

$$\|u_h\|_{W^{1,\infty}(\Omega,\mathbb{C})} \le C\left(\|u\|_{W^{1,\infty}(\Omega_h,\mathbb{C})} + h\|u\|_{W^{1,\infty}(\Omega\setminus\Omega_h)}\right).$$

Proof. 1. is just a consequence of Proposition 4.12 applied to real and imaginary part of $v_h \in \tilde{V}_h$.

2. is just the application of Proposition 4.13 onto real and imaginary part of $v \in L^\infty(\Omega, \mathbb{C})$.

3. can be proven by using the nodal interpolant $I_h v$ of $v \in W^{1,\infty}(\Omega, \mathbb{C})$. We have

$$\|P_h v - I_h v\|_{L^\infty(\Omega,\mathbb{C})} = \|P_h(v - I_h v)\|_{L^\infty(\Omega,\mathbb{C})} \le C\|v - I_h v\|_{L^\infty(\Omega,\mathbb{C})}.$$

Here we used the $L^\infty(\Omega,\mathbb{C})$-stability of the projection from *2.* By a standard interpolation error estimate applied to real and imaginary part, cf. Lemma 4.10, we thus have

$$\|P_h v - I_h v\|_{L^\infty(\Omega,\mathbb{C})} \le Ch\|v\|_{W^{1,\infty}(\Omega,\mathbb{C})}.$$

By the inverse inequality from *1.* we therefore have

$$\|P_h v - I_h v\|_{W^{1,\infty}(\Omega,\mathbb{C})} \le C\|v\|_{W^{1,\infty}(\Omega,\mathbb{C})}.$$

Clearly $\|I_h v\|_{W^{1,\infty}(\Omega,\mathbb{C})} \le \|v\|_{W^{1,\infty}(\Omega,\mathbb{C})}$ by Lemma 4.10 applied to real and imaginary part. Therefore

$$\|P_h v\|_{W^{1,\infty}(\Omega,\mathbb{C})} \le \|P_h v - I_h v\|_{W^{1,\infty}(\Omega,\mathbb{C})} + \|I_h v\|_{W^{1,\infty}(\Omega,\mathbb{C})} \le C\|v\|_{W^{1,\infty}(\Omega,\mathbb{C})}.$$

4. is just an application of Corollary 8.31 to real and imaginary part. □

Now we state a generalization of [BTW03, Theorem 2.1].

Proposition 4.32 *Let Ω be a C^2-domain and let A have $C^{0,1}(\Omega)$-coefficients. For any $\rho_0 \in (0, \frac{\pi}{2})$ there is a $C > 0$ such that*

$$\|(\lambda \operatorname{Id} - A)^{-1} v\|_{W^{j,\infty}(\Omega,\mathbb{C})} \le C(1+|\lambda|)^{-1+\frac{j}{2}}\|v\|_{L^\infty(\Omega,\mathbb{C})} \tag{4.25}$$

for $\lambda \notin \Sigma_{\rho_0} := \{z \in \mathbb{C} : |\arg z| \le \rho_0\}$, $j = 0, 1$, $v \in C(\bar{\Omega}, \mathbb{C})$.

Proof. As stated in the first remark in the proof of [BTW03, Theorem 2.1] the results of [Ste74, Theorem 1] hold true if we consider any elliptic operator A with no lower order terms and thus get, as in [BTW03, Theorem 2.1], that (4.25) holds true for any $\rho_0 \in (0, \frac{\pi}{2})$ and λ with $|\lambda| \ge R$ for some $R > 0$. [Ste74] does not require higher regularity than C^2 for Ω or more than uniform continuity of the coefficients of A.

The remaining case that $|\lambda| < R$ is treated in the rest of the proof of [BTW03, Theorem 2.1]. There only higher $W^{2,p}$-regularity for elliptic problems is needed, which is true by Theorem 8.23. □

Inspecting the proof of [BTW03, Theorem 2.2], we immediately see that it is enough to assume that Ω is a C^2-domain. The domain regularity is used to construct a C^2 function, whose regularity has to be preserved by the boundary regularity under extension.

Proposition 4.33 *Let Ω be a C^2-domain and let A have coefficients in $C^{0,1}(\Omega)$. Then for any $\rho_0 \in (0, \frac{\pi}{2})$ there is a $C > 0$ such that*

$$\|A(\lambda \operatorname{Id} - A)^{-1} v\|_{L^\infty(\Omega,\mathbb{C})} \le C(1+|\lambda|)^{-\frac{1}{2}}\|v\|_{W^{1,\infty}(\Omega,\mathbb{C})} \quad \lambda \notin \Sigma_{\rho_0}, v \in W^{1,\infty}(\Omega,\mathbb{C}) \cap C_0(\Omega,\mathbb{C}).$$

The following proposition is one of the main reasons, next to the usage of Green's functions, why we need higher domain regularity than the usual $C^{1,1}$-regularity.

Proposition 4.34 *Let Ω be a $C^{2,1}$-domain. Let A be an operator with uniformly elliptic, symmetric coefficients in $C^{1,1}(\Omega)$. Then we have that for any $\alpha \in (0,1)$ there is a $C > 0$ such that for any $f \in C^{0,\alpha}(\Omega,\mathbb{C})$ the solution g of*

$$\begin{cases} Ag &= f \text{ in } \Omega, \\ g|_{\partial\Omega} &= 0 \text{ on } \partial\Omega, \end{cases} \tag{4.26}$$

satisfies

$$\|g\|_{C^{2,\alpha}(\Omega)} \le C\|f\|_{C^{0,\alpha}(\Omega)}.$$

Proof. We first discuss the real-valued case. Let $f \in C^{0,\alpha}(\Omega)$ and g the solution to (4.26). By [Gri11, Lemma 2.4.2.1 and the remarks thereafter, Theorem 2.4.2.5] we have that for any $p \in (1,\infty)$ there is $C_p > 0$, independent of f and g, such that

$$\|g\|_{W^{2,p}(\Omega)} \leq C_p \|f\|_{L^p(\Omega)} \leq C_p \|f\|_{C^{0,\alpha}(\Omega)}. \tag{4.27}$$

For p large enough we have by the Sobolev embeddings, e.g. [Eva98, Theorem 5.6],

$$\|g\|_{C^{1,\alpha}(\Omega)} \leq C_{p(\alpha)} \|f\|_{C^{0,\alpha}(\Omega)}.$$

[Gri11, Theorem 6.3.2.1] yields $g \in C^{2,\alpha}(\Omega)$, but no estimate. But now [Gri11, Theorem 6.3.1.4] is also applicable and yields, together with (4.27),

$$\|g\|_{C^{2,\alpha}(\Omega)} \leq C_\alpha \left(\|f\|_{C^{0,\alpha}(\Omega)} + \|g\|_{C^{1,\alpha}(\Omega)} \right) \leq C_\alpha \|f\|_{C^{0,\alpha}(\Omega)}.$$

The complex valued statement now follows from separation into real and imaginary part. □

This is a generalization of [BTW03, Theorem 2.3].

Proposition 4.35 *Let Ω be a $C^{2,1}$-domain and let A be a symmetric operator with coefficients in $C^2(\bar{\Omega})$. Then for any $\rho_0 \in (0, \frac{\pi}{2})$ and $\alpha \in (0,1)$ there is a $C > 0$ such that*

$$\|(\lambda \operatorname{Id} - A)^{-1} v\|_{C^{1,\alpha}(\bar{\Omega})} \leq C \left(1 + |\lambda|\right)^{-1+\frac{\alpha}{2}} \|v\|_{W^{1,\infty}(\Omega)} \qquad \forall \lambda \notin \Sigma_{\rho_0}, v \in W^{1,\infty}(\Omega) \cap C_0(\Omega).$$

Proof. First we note that [Tri78, Theorem 4.5.2] applies to $C^{2,1}$-domains and not only C^∞-domains. It implies in particular that for any $\alpha \in (0,1)$ one has

$$\left(C^\alpha(\Omega), C^{2,\alpha}(\Omega) \right)_{\frac{1}{2},\infty} = C^{1,\alpha}(\Omega).$$

Here $(\cdot,\cdot)_{\frac{1}{2},\infty}$ denotes an interpolation space. The reason [Tri78, Theorem 4.5.2] is still applicable for less smooth domains is that it is only necessary to extend and retract functions $v \in C^{2,\alpha}(\Omega)$ to and from functions in $C^{2,\alpha}(\mathbb{R}^N)$. This is possible by [Gri11, Theorem 6.2.4].

This observation and the Schauder-type estimate from Proposition 4.34 for less regular domains can now be inserted into the proof of [BTW03, Theorem 2.3], generalizing it. □

We can now state a generalization of [BTW03, Theorem 1.1].

Theorem 4.36 *Let Assumption 4.14 be satisfied. Let Ω be a $C^{3,\alpha}$-domain for some $\alpha \in (0,1)$. Let $h \in (0, \frac{1}{2})$. Assume A has $C^2(\bar{\Omega})$ coefficients. Then for any $\rho_0 \in (0, \frac{\pi}{2})$ there exists a $C > 0$, independent of h, such that*

$$\|(\lambda \operatorname{Id} - A_h)^{-1} v_h\|_{L^\infty(\Omega)} \leq C (1 + |\lambda|)^{-1} \|v_h\|_{L^\infty(\Omega)} \quad \forall \lambda \notin \Sigma_{\rho_0}, v_h \in \tilde{V}_h.$$

Proof. The proofs of [BTW03, Theorem 1.1] in [BTW03, Section 3] are based on the inequalities given by Propositions 4.31 to 4.33 and 4.35 and do not require any more boundary regularity or regularity of the operator. □

Corollary 4.37 *Let Assumption 4.14 be satisfied. Let Ω be a $C^{3,\alpha}$-domain for some $\alpha \in (0,1)$. Let $h \in (0, \frac{1}{2})$. Assume A has $C^2(\bar{\Omega})$ coefficients. Then for any $\rho_0 \in (0, \frac{\pi}{2})$ there exists a $C > 0$, independent of h, such that for any $p \in [1,\infty]$ one has*

$$\|(\lambda \operatorname{Id} - A_h)^{-1} v\|_{L^p(\Omega)} \leq C (1 + |\lambda|)^{-1} \|v\|_{L^p(\Omega)} \quad \forall \lambda \notin \Sigma_{\rho_0}, \quad v \in L^p(\Omega, \mathbb{C}). \tag{4.28}$$

Proof. Let $h \in (0, \frac{1}{2})$ and $\lambda \in \mathbb{C} \setminus \Sigma_{\rho_0}$ for some $\rho_0 \in (0, \frac{\pi}{2})$.

We first show the estimate for $p = 2$. Let $v \in L^2(\Omega, \mathbb{C})$. We define $w_h := -(\lambda \operatorname{Id} - A_h)^{-1} v \in V_h$ that means

$$a_I(w_h, \bar{\varphi}_h) - (\lambda w_h, \bar{\varphi}_h)_{L^2(Q)} = (v, \bar{\varphi}_h)_{L^2(Q)} \quad \forall \varphi_h \in \tilde{V}_h.$$

Choosing $\varphi_h = w_h$ and splitting into real and imaginary part yields

$$\begin{aligned} c_{ell} \|\nabla w_h\|^2_{L^2(\Omega,\mathbb{C}^N)} - \mathfrak{Re}(\lambda) \|w_h\|^2_{L^2(\Omega,\mathbb{C})} &\leq \mathfrak{Re}((v, \bar{w}_h)_{L^2(\Omega)}) \\ -\mathfrak{Im}(\lambda) \|w_h\|^2_{L^2(\Omega,\mathbb{C})} &= \mathfrak{Im}((v, \bar{w}_h)_{L^2(\Omega)}). \end{aligned} \tag{4.29}$$

If $\mathfrak{Re}(\lambda) \leq 0$ we take the absolute value on both sides and get

$$\begin{aligned} 2\|v\|_{L^2(\Omega,\mathbb{C})} \|w_h\|_{L^2(\Omega,\mathbb{C})} &\geq \mathfrak{Re}((v, \bar{w}_h)_{L^2(\Omega)}) + |\mathfrak{Im}((v, \bar{w}_h)_{L^2(\Omega)})| \\ &\geq c_{ell} \|\nabla w_h\|^2_{L^2(\Omega,\mathbb{C}^N)} + |\mathfrak{Re}(\lambda)| \|w_h\|^2_{L^2(\Omega,\mathbb{C})} + |\mathfrak{Im}(\lambda)| \|w_h\|^2_{L^2(\Omega,\mathbb{C})}. \end{aligned}$$

Using the Poincaré inequality and the triangle inequality we can bound this further from below by

$$c_1 \|w_h\|^2_{L^2(\Omega,\mathbb{C})} + c_2 |\lambda| \|w_h\|^2_{L^2(\Omega,\mathbb{C})} \geq c(1 + |\lambda|) \|w_h\|^2_{L^2(\Omega,\mathbb{C})}.$$

Thus

$$\|w_h\|_{L^2(\Omega,\mathbb{C})} \leq C(1 + |\lambda|)^{-1} \|v\|_{L^2(\Omega,\mathbb{C})}.$$

Now assume $\mathfrak{Re}(\lambda) > 0$ and, due to symmetry, without loss of generality $\mathfrak{Im}(\lambda) > 0$. Because $\lambda \notin \Sigma_{\rho_0}$ we cannot have $\mathfrak{Re}(\lambda) > 0$ and $\mathfrak{Im}(\lambda) = 0$. In this situation we have

$$\rho_0 \leq \arg \lambda = \arctan\left(\frac{\mathfrak{Im}(\lambda)}{\mathfrak{Re}(\lambda)}\right) \implies \tan(\rho_0) \leq \frac{\mathfrak{Im}(\lambda)}{\mathfrak{Re}(\lambda)}.$$

Thus $\mathfrak{Re}(\lambda) \leq (\tan(\rho_0))^{-1} \mathfrak{Im}(\lambda)$. Then by (4.29)

$$\begin{aligned} c_{ell} \|\nabla w_h\|^2_{L^2(\Omega,\mathbb{C}^N)} &\leq \mathfrak{Re}(\lambda) \|w_h\|^2_{L^2(\Omega,\mathbb{C})} + \mathfrak{Re}((v, \bar{w}_h)_{L^2(\Omega)}) \\ &\leq \frac{1}{\tan(\rho_0)} \mathfrak{Im}(\lambda) \|w_h\|^2_{L^2(\Omega,\mathbb{C})} + \mathfrak{Re}((v, \bar{w}_h)_{L^2(\Omega)}) \\ &= -\frac{1}{\tan(\rho_0)} \mathfrak{Im}((v, \bar{w}_h)_{L^2(\Omega)}) + \mathfrak{Re}((v, \bar{w}_h)_{L^2(\Omega)}) \\ &\leq \left(1 + \frac{1}{\tan(\rho_0)}\right) \|v\|_{L^2(\Omega,\mathbb{C})} \|w_h\|_{L^2(\Omega,\mathbb{C})}. \end{aligned}$$

Here we used $\mathfrak{Im}(\lambda) \|w_h\|^2_{L^2(\Omega)} = -|\mathfrak{Im}((v, \bar{w}_h)_{L^2(\Omega)}|$ and the Cauchy Schwarz inequality. Now we can argue analogously to the earlier case and obtain (4.28) for $p = 2$.

Now we discuss $p = \infty$. Let $v \in L^\infty(\Omega, \mathbb{C})$. Theorem 4.36 almost gives us the desired (4.28), but $v \notin V_h$. Yet, Theorem 4.36 in conjunction with Proposition 4.31.2 does deliver

$$\begin{aligned} &\|(\lambda \operatorname{Id} - A_h)^{-1} v\|_{L^\infty(\Omega,\mathbb{C})} = \|(\lambda \operatorname{Id} - A_h)^{-1} P_h v\|_{L^\infty(\Omega,\mathbb{C})} \\ &\leq C(1 + |\lambda|)^{-1} \|P_h v\|_{L^\infty(\Omega,\mathbb{C})} \leq C(1 + |\lambda|)^{-1} \|v\|_{L^\infty(\Omega,\mathbb{C})}. \end{aligned}$$

We now make an interpolation argument. We may assume that the constant C in (4.28) is the same for $p = 2$ and $p = \infty$. Let $p \in (2, \infty)$. By [BL76, Theorem 5.1.1] we have

$$L^p(\Omega, \mathbb{C}) = \left[L^2(\Omega, \mathbb{C}), L^\infty(\Omega, \mathbb{C})\right]_{1-\frac{2}{p}}.$$

All the spaces are complex and $[\cdot,\cdot]_{1-\frac{2}{p}}$ denotes the complex interpolation functor. It is used and introduced for example in [BL76, Theorem 4] or [Tri78, Section 1.9]. We do not need any details.

By [Tri78, Theorem 1.9.3] the interpolation is exact of type $1-\frac{2}{p}$, that means for any linear continuous operator $T : L^\infty(\Omega,\mathbb{C}) \to L^\infty(\Omega,\mathbb{C})$ we have

$$\|T\|_{L^p(\Omega,\mathbb{C}),L^p(\Omega,\mathbb{C})} \leq \|T\|^{\frac{2}{p}}_{L^2(\Omega,\mathbb{C}),L^2(\Omega,\mathbb{C})}\|T\|^{1-\frac{2}{p}}_{L^\infty(\Omega,\mathbb{C}),L^\infty(\Omega,\mathbb{C})}.$$

Thus we have in particular

$$\|(\lambda \operatorname{Id} - A_h)^{-1}\|_{L^p(\Omega,\mathbb{C}),L^p(\Omega,\mathbb{C})} \leq C(1+|\lambda|)^{-1}.$$

Here C is the same C as for $p=2$ and $p=\infty$.

By duality arguments we now argue for $p=1$. Let $v_h \in V_h$ and $\Phi \in L^\infty(\Omega,\mathbb{C})$ with $\|\Phi\|_{L^\infty(\Omega,\mathbb{C})} \leq 1$. Now

$$\begin{aligned}\left((\lambda \operatorname{Id} - A_h)^{-1}v_h, \bar{\Phi}\right)_{L^2(\Omega)} &= \left(v_h, \left(\bar{\lambda}\operatorname{Id} - A_h\right)^{-1}\bar{\Phi}\right)_{L^2(\Omega)} \\ &\leq \|v_h\|_{L^1(\Omega,\mathbb{C})}\left\|\left(\bar{\lambda}\operatorname{Id} - A_h\right)^{-1}\bar{\Phi}\right\|_{L^\infty(\Omega,\mathbb{C})}.\end{aligned}$$

Now, because $\lambda \notin \Sigma_{\rho_0} \iff \bar{\lambda} \notin \Sigma_{\rho_0}$, we can use the already stated estimate for $p=\infty$ and the fact that $\|\Phi\|_{L^\infty(\Omega,\mathbb{C})} \leq 1$ to obtain

$$\left((\lambda \operatorname{Id} - A_h)^{-1}v_h, \bar{\Phi}\right)_{L^2(\Omega)} \leq \|v_h\|_{L^1(\Omega,\mathbb{C})}C(1+|\lambda|)^{-1}\|\Phi\|_{L^\infty(\Omega,\mathbb{C})} \leq C(1+|\lambda|)^{-1}\|v_h\|_{L^1(\Omega,\mathbb{C})}.$$

As $\Phi \in L^\infty(\Omega,\mathbb{C})$ with $\|\Phi\|_{L^\infty(\Omega,\mathbb{C})} \leq 1$ was arbitrary this yields the desired $L^1(\Omega,\mathbb{C})$-estimate. The constant is the same as for $p=\infty$.

By the same interpolation arguments as before the case for $p \in (1,2)$ *follows.* ☐

4.2 Time Discretization

In the following definitions we stick closely to the notation used in [LV17b, Section 3], but also adapt it to smooth domains.

Definition 4.38 We divide $I=[0,T]$ into M subintervals of the form $(t_{m-1},t_m] =: I_m$ with $0=t_0<t_1<\cdots<t_{M-1}<t_M=T$. The interval size is abbreviated by $k_m := t_m - t_{m-1} > 0$. We also introduce $k := \max_{m=1,\ldots,M} k_m$.

The semidiscrete space X_k^0 denotes the functions which are piecewise constant in time:

$$X_k^0 = \left\{v_k \in L^2(I,V) : v_k|_{I_m} \in \mathcal{P}_0(I_m,V), m=1,\ldots,M\right\}.$$

Here $\mathcal{P}_0(I_m,V)$ is the space of constant functions with values in the Hilbert space $V=H_0^1(\Omega)$.

We use the following notation for limits and jumps at t_m of $v_k \in X_k^0$:

$$v_k(t_m^+) := \lim_{\epsilon\to 0^+} v_k(t_m+\epsilon), \qquad v_k(t_m^-) := \lim_{\epsilon\to 0^-} v_k(t_m+\epsilon), \qquad [v_k]_m := v_k(t_m^+) - v_k(t_m^-).$$

Note that the end time value for $v_k \in X_k^0$ is well-defined as $T \in I_M$.

We also define the nodal interpolant $I_k : C(\bar{I}, V) \to X_k^0$. Given a $v \in C(\bar{I}, V)$ the interpolant $I_k v$ is the unique function in X_k^0 such that $v(t_m) = I_k v(t_m)$ holds for all $m = 1, 2, \ldots, M$. The same construction obviously works if one uses a different space than V.

Given a mesh $\mathcal{K}_h$ we also introduce the fully discrete spaces

$$X_{k,h}^{0,1} := \left\{ v_{kh} \in X_k^0 : v_{kh}|_{I_m} \in \mathcal{P}_0(I_m, V_h), m = 1, \ldots, M \right\}$$

and the bilinear form B on $X_k^0 \cup W(I)$.

$$\begin{aligned} B_k(v, \varphi) &:= \sum_{m=1}^{M} (\partial_t v, \varphi)_{L^2(I_m, V^*, V)} + a_I(v, \varphi) + \sum_{m=2}^{M} \Big([v]_{m-1}, \varphi_{m-1}^+ \Big)_H + \Big(v(0^+), \varphi(0^+) \Big)_H \\ &= -\sum_{m=1}^{M} (v, \partial_t \varphi)_{L^2(I_m, V, V^*)} + a_I(v, \varphi) - \sum_{m=1}^{M-1} (v_m^-, [\varphi]_m)_H + (v(T), \varphi(T))_H \,. \end{aligned}$$

Remark 4.39 For $v \in W(I)$ the expression $B(v, \varphi)$ has a simple structure, because $v \in C(\bar{I}, H)$ and thus the jumps vanish:

$$\begin{aligned} B(v, \varphi) &= (\partial_t v, \varphi)_{L^2(I, V^*, V)} + a_I(v, \varphi) + \Big(v(0), \varphi(0^+) \Big)_H \\ &= -\sum_{m=1}^{M} (v, \partial_t \varphi)_{I_m \times \Omega} + a_I(v, \varphi) + \sum_{m=1}^{M-1} (v(t_m), [\varphi]_m)_H + (v(T), \varphi(T))_H \,. \end{aligned}$$

For $v \in X_k^0$ the expression also looks simpler:

$$B(v, \varphi) = a_I(v, \varphi) + \sum_{m=2}^{M} \Big([v]_{m-1}, \varphi_{m-1}^+ \Big)_{L^2(\Omega)} + \Big(v(0^+), \varphi(0^+) \Big)_{L^2(\Omega)} \,.$$

Lemma 4.40 *B_k is a positive definite bilinear form on X_k^0. In particular we have $B(v_k, v_k) \geq a_I(v_k, v_k) + \|v(T)\|_H^2 + \|v(0^+)\|_H^2$ for all $v_k \in X_k^0$.*

Proof. The bilinearity is clear. To show positive definiteness let $v_k \in X_k^0$. Then, by the two formulations of B and the piecewise constantness of v_k we have:

$$\begin{aligned} 2B(v_k, v_k) &= \sum_{m=1}^{M} \Big((\partial_t v_k, v_k)_{L^2(I_m, V^*, V)} - (v_k, \partial_t v_k)_{L^2(I_m, V, V^*)} \Big) + 2a_I(v_k, v_k) \\ &\quad + \sum_{m=1}^{M-1} \Big(\Big([v_k]_m, v(t_m^+) \Big)_H - (v(t_m^-), [v_k]_m)_H \Big) + \Big(v(0^+), v(0^+) \Big)_H + (v(T), v(T))_H \\ &= \sum_{m}^{M-1} \Big([v_k]_m, v(t_m^+) - v(t_m^-) \Big) + \Big(v(0^+), v(0^+) \Big)_H + (v(T), v(T))_H + 2a_I(v_k, v_k) \\ &= \sum_{m}^{M-1} ([v_k]_m, [v_k]_m)_H + \Big(v(0^+), v(0^+) \Big)_H + (v(T), v(T))_H + 2a_I(v_k, v_k) \geq 2a_I(v_k, v_k). \end{aligned}$$

If $B(v_k, v_k) = 0$ we have by the ellipticity of a_I that $\|v_k\|_{L^2(I,V)} = 0$, yielding the positivity. □

4.3 Discretization of Regularized Obstacle Problems

Definition 4.41 Let $\mathcal{K}_h$ be a triangulation of Ω and I be discretized by intervals of size at most k. Let $\gamma > 0$ and $u \in L^2(Q)$ and $y_0 \in L^2(\Omega)$. The regularized, discretized state $y_{\gamma kh} = y_{\gamma kh}(u) \in X_{k,h}^{0,1}$ is defined as the solution to

$$B_k(y_{kh}, \varphi_{kh}) + (f(y_{kh}) + \beta_\gamma(y_{kh} - \Psi), \varphi_{kh})_{L^2(Q)} = (u, \varphi_{kh})_{L^2(Q)} + \Big(y_0, \varphi_{kh}(0^+)\Big)_H \quad (\mathrm{R}_{\gamma kh})$$

for all $\varphi_{kh} \in X_{k,h}^{0,1}$. We introduce the solution operator

$$\begin{aligned} S_{\gamma kh} \colon L^2(Q) &\to X_{k,h}^{0,1}, \\ u &\mapsto y_{\gamma kh}(u). \end{aligned}$$

[MV17, Theorem 3.1] gives the well-definedness of $S_{\gamma kh}$. Its proof immediately applies to both $C^{1,1}$-domains and convex polygonal/polyhedral domains:

Theorem 4.42 $(\mathrm{R}_{\gamma kh})$ *has a unique solution. In particular $S_{\gamma kh}$ is well-defined.*

Lemma 4.43 *$S_{\gamma kh}$ is Lipschitz continuous in the following sense: there exists a $C > 0$, independent of γ, k, h, such that for all $u_1, u_2 \in L^2(Q)$ we have*

$$\|S_{\gamma kh}(u_1) - S_{\gamma kh}(u_2)\|_{L^2(I,V)} + \|S_{\gamma kh}(u_1)(T) - S_{\gamma kh}(u_2)(T)\|_H \le C\|u_1 - u_2\|_{L^2(I,V^*)}.$$

By the equivalence of norms in the finite dimensional space $X_{k,h}^{0,1}$ this implies that $S_{\gamma kh}$ is Lipschitz continuous with respect to any norm in $X_{k,h}^{0,1}$. The Lipschitz constant does then depend on k and h.

Proof. We abbreviate $y_i := S_{\gamma kh}(u_i) \in X_{k,h}^{0,1}$ for $i = 1, 2$. Testing $(\mathrm{R}_{\gamma kh})$ with $y_1 - y_2$ for u_1 and u_2 and taking the difference yields

$$\begin{aligned} &B_k(y_1 - y_2, y_1 - y_2) + (f(y_1) - f(y_2) + \beta_\gamma(y_1 - \Psi) - \beta_\gamma(y_2 - \Psi), y_1 - y_2)_{L^2(Q)} \\ &\qquad = (u_1 - u_2, y_1 - y_2)_{L^2(Q)}. \end{aligned}$$

We abbreviate $\delta y := y_1 - y_2$ and $\delta u := u_1 - u_2$. By the monotonicity of f and β_γ and Lemma 4.40 we have

$$a_I(\delta y, \delta y) + \|\delta y(T)\|_H^2 \le \|\delta u\|_{L^2(I,V^*)}\|\delta y\|_{L^2(I,V)}.$$

Because $V = H_0^1(\Omega)$ we have that $a_I : L^2(I, V) \times L^2(I, V) \to \mathbb{R}$ is actually elliptic by the Poincaré inequality, e.g. [Eva98, Theorem 5.6.3], so that

$$c\|\delta y\|_{L^2(I,V)}^2 + \|\delta y(T)\|_H^2 \le \|\delta u\|_{L^2(I,V^*)}\|\delta y\|_{L^2(I,V)} \le \frac{1}{2c}\|\delta u\|_{L^2(I,V^*)}^2 + \frac{c}{2}\|\delta y\|_{L^2(I,V)}^2.$$

Here we used Young's inequality in the last estimate. Subtracting $\frac{c}{2}\|\delta y\|_{L^2(I,V)}^2$ from both sides and taking the root yields the claim. □

We shortly discuss the regularity of the mapping $S_{\gamma kh} \colon L^2(Q) \to X_{k,h}^{0,1}$. The proof is very similar to the proof of Theorem 3.11.

Theorem 4.44 *Assume that $\beta_\gamma \in C^1(\mathbb{R})$ and $f(t,x,\cdot) \in C^{1,1}_{loc}(\mathbb{R})$ for any $(t,x) \in Q$. Here the Lipschitz constants of $f(t,x,\cdot)$ or $f'(t,x,\cdot)$ may not depend on $(t,x) \in Q$. Then*

$$S_{\gamma kh} \colon L^2(Q) \to X^{0,1}_{k,h}$$

is Fréchet differentiable. For $u, d \in L^2(Q)$ its derivative $S'_{\gamma kh}(u)d =: z_{\gamma kh}(u,d)$ solves

$$\begin{aligned} B_k(z_{\gamma kh}(u,d), \varphi_{kh}) + &\left(({\beta_\gamma}'(S_{\gamma kh}(u) - \Psi) + f'(S_{\gamma kh}(u)))\, z_{\gamma kh}(u,d), \varphi_{kh} \right)_{L^2(Q)} \\ &= (d, \varphi_{kh})_{L^2(Q)} \qquad \varphi_{kh} \in X^{0,1}_{k,h}. \end{aligned}$$

Proof. Let $u \in L^2(Q)$ and $d \in L^2(Q)$. We define $y_{\gamma kh}(u) := S_{\gamma kh}(u)$, $y_{\gamma kh}(u+d) := S_{\gamma kh}(u+d)$ and $z_{\gamma kh}(u,d)$ as above. They satisfy

$$\begin{aligned} &B_k(y_{\gamma kh}(u), \varphi_{kh}) + (\beta_\gamma(y_{\gamma kh}(u) - \Psi) + f(y_{\gamma kh}(u)), \varphi_{kh})_{L^2(Q)} \\ &\qquad = (u, \varphi_{kh})_{L^2(Q)} + \left(y_0, \varphi_{kh}(0^+) \right)_H, \\ &B_k(y_{\gamma kh}(u+d), \varphi_{kh}) + (\beta_\gamma(y_{\gamma kh}(u+d) - \Psi) + f(y_{\gamma kh}(u+d)), \varphi_{kh})_{L^2(Q)} \\ &\qquad = (u+d, \varphi_{kh})_{L^2(Q)} + \left(y_0, \varphi_{kh}(0^+) \right)_H, \\ &B_k(z_{\gamma kh}(u,d), \varphi_{kh}) + \left(({\beta_\gamma}'(y_{\gamma kh}(u) - \Psi) + f'(y_{\gamma kh}(u)))\, z(u,d), \varphi_{kh} \right)_{L^2(Q)} \\ &\qquad = (d, \varphi_{kh})_{L^2(Q)}. \end{aligned}$$

Those equations hold for all $\varphi_{kh} \in X^{0,1}_{k,h}$. Subtracting the lines from each other and introducing

$$r_{\gamma kh}(u,d) := y_{\gamma kh}(u+d) - y_{\gamma kh}(u) - z_{\gamma kh}(u,d)$$

yields for any $\varphi_{kh} \in X^{0,1}_{k,h}$:

$$\begin{aligned} B_k\left(r_{\gamma kh}(u,d), \varphi_{kh} \right) + (&\beta_\gamma(y_{\gamma kh}(u+d) - \Psi) - \beta_\gamma(y_{\gamma kh}(u) - \Psi) - {\beta_\gamma}'(y_{\gamma kh}(u) - \Psi) z_{\gamma kh}(u,d) \\ &+ f(y_{\gamma kh}(u+d)) - f(y_{\gamma kh}(u)) - f'(y_{\gamma kh}(u)) z_{\gamma kh}(u,d), \varphi_{kh})_{L^2(Q)} = 0. \end{aligned} \tag{4.30}$$

Using the fundamental theorem of calculus we have almost everywhere in Q:

$$\begin{aligned} &f(y_{\gamma kh}(u+d)) - f(y_{\gamma kh}(u)) - f'(y_{\gamma kh}(u)) z_{\gamma kh}(u,d) \\ &= \int_0^1 f'(y_{\gamma kh}(u) + s(y_{\gamma kh}(u+d) - y_{\gamma kh}(u)))(y_{\gamma kh}(u+d) - y_{\gamma kh}(u))\, ds \\ &\quad - \int_0^1 f'(y_{\gamma kh}(u)) z_{\gamma kh}(u,d)\, ds \\ &= \int_0^1 f'(y_{\gamma kh}(u) + s(y_{\gamma kh}(u+d) - y_{\gamma kh}(u))) r_{\gamma kh}(u,d)\, ds \\ &\quad + \int_0^1 \left(f'(y_{\gamma kh}(u) + s(y_{\gamma kh}(u+d) - y_{\gamma kh}(u))) - f'(y_{\gamma kh}(u)) \right) z_{\gamma kh}(u,d)\, ds. \end{aligned}$$

We abbreviate $\Theta_{f,s} := f'(y_{\gamma kh}(u) + s(y_{\gamma kh}(u+d) - y_{\gamma kh}(u)))$ to obtain

$$\begin{aligned} &f(y_{\gamma kh}(u+d)) - f(y_{\gamma kh}(u)) - f'(y_{\gamma kh}(u)) z_{\gamma kh}(u,h) \\ &= \int_0^1 \Theta_{f,s} r_{\gamma kh}(u,d)\, ds + \int_0^1 (\Theta_{f,s} - f'(y_{\gamma kh}(u))) z_{\gamma kh}(u,d)\, ds. \end{aligned} \tag{4.31}$$

Analogously we obtain

$$\begin{aligned}
&\beta_\gamma(y_{\gamma kh}(u+d)-\Psi)-\beta_\gamma(y_{\gamma kh}(u)-\Psi)-\beta_\gamma{}'(y_{\gamma kh}(u)-\Psi)z_{\gamma kh}(u,d)\\
&=\int_0^1 \Theta_{\beta_\gamma,s} r_{\gamma kh}(u,d)\;ds+\int_0^1(\Theta_{\beta_\gamma,s}-\beta_\gamma{}'(y_{\gamma kh}(u)-\Psi))z_{\gamma kh}(u,d)\,ds.
\end{aligned}\tag{4.32}$$

Inserting (4.31) and (4.32) into (4.30) thus yields

$$\begin{aligned}
&B_k(r_{\gamma kh}(u,d),\varphi_{kh})+\Big((\Theta_{f,s}+\Theta_{\beta_\gamma,s})r_{\gamma kh},\varphi_{kh}\Big)_{L^2(Q)}\\
&=-\Big(\int_0^1\Theta_{\beta_\gamma,s}-\beta_\gamma{}'(y_{\gamma kh}(u)-\Psi)\,ds\cdot z_{\gamma kh}(u,d),\varphi_{kh}\Big)_{L^2(Q)}\\
&\quad-\Big(\int_0^1\Theta_{f,s}-f'(y_{\gamma kh}(u))\,ds\cdot z_{\gamma kh}(u,d),\varphi_{kh}\Big)_{L^2(Q)}.
\end{aligned}$$

Choosing $\varphi_{kh}=r_{\gamma kh}(u,d)$ and using Lemma 4.40 together with the ellipticity of a_I, note that Poincaré inequality, e.g. [Eva98, Theorem 5.6.3], is applicable, and $\Theta_{f,s},\Theta_{\beta_\gamma,s}\geq 0$ yields

$$\begin{aligned}
c\|r_{\gamma kh}(u,d)\|^2_{L^2(I,V)}\leq{}&\left\|\int_0^1\Theta_{\beta_\gamma,s}-\beta_\gamma{}'(y_{\gamma kh}(u)-\Psi)\,ds\cdot z_{\gamma kh}(u,d)\right\|_{L^2(Q)}\|r_{\gamma kh}(u,d)\|_{L^2(Q)}\\
&+\left\|\int_0^1\Theta_{f,s}-f'(y_{\gamma kh}(u))\,ds\cdot z_{\gamma kh}(u,d)\right\|_{L^2(Q)}\|r_{\gamma kh}(u,d)\|_{L^2(Q)}.
\end{aligned}$$

Thus

$$\begin{aligned}
c\|r_{\gamma kh}(u,h_u)\|_{L^2(I,V)}\leq{}&\left\|\int_0^1\Theta_{\beta_\gamma,s}-\beta_\gamma{}'(y_{\gamma kh}(u)-\Psi)\,ds\cdot z_{\gamma kh}(u,d)\right\|_{L^2(Q)}\\
&+\left\|\int_0^1\Theta_{f,s}-f'(y_{\gamma kh}(u))\,ds\cdot z_{\gamma kh}(u,d)\right\|_{L^2(Q)}.
\end{aligned}\tag{4.33}$$

We now may assume $\|d\|_{L^2(Q)}\leq 1$. By the Lipschitz continuity of $S_{\gamma kh}$ and the equivalency of norms in $X^{0,1}_{k,h}$ we have

$$\|y_{\gamma kh}(u)\|_{L^\infty(Q)},\|y_{\gamma kh}(u+d)\|_{L^\infty(Q)}\leq R_0 \text{ independent of } d.\tag{4.34}$$

By the continuity of $\beta_\gamma{}'$ it is also uniformly continuous on the compact set $[-R_0,R_0]$. Let $\delta>0$. By this uniform continuity and the Lipschitz continuity of $S_{\gamma kh}$ then there exists a $\epsilon>0$ such that $\|d\|_{L^2(Q)}<\epsilon$ implies $|\beta_\gamma(s(y_{\gamma kh}(u+d)-y_{\gamma kh}(u))+y_{\gamma kh}(u)-\Psi)-\beta_\gamma(y_{\gamma kh}(u)-\Psi)|<\delta$, independently of s. Thus the first term on the right in (4.33) is bounded by $\delta\|z_{\gamma kh}(u,d)\|_{L^2(Q)}$ for those small d.

We also have by the local Lipschitz continuity of f' and (4.34) that

$$\left|\int_0^1\Theta_f-f'(y_{\gamma kh}(u))\,ds\right|\leq C|y_{\gamma kh}(u+d)-y_{\gamma kh}(u)|.$$

Thus (4.33) entails

$$\begin{aligned}
&\limsup_{\|d\|_{L^2(Q)}\to 0}\|r_{\gamma kh}(u,d)\|_{L^2(I,V)}\\
&\leq C\limsup_{\|d\|_{L^2(Q)}\to 0}(\delta\|z_{\gamma kh}(u,d)\|_{L^2(Q)}+\||y_{\gamma kh}(u)-y_{\gamma kh}(u+d)|\cdot|z_{\gamma kh}(u,d)|\|_{L^2(Q)}).
\end{aligned}$$

By the usual arguments, the ellipticity of B_k (Lemma 4.40), and the positivity of $\beta_\gamma{}'$ and f', we get the estimate $\|z_{\gamma kh}(u,d)\|_{L^2(Q)} \le C\|d\|_{L^2(Q)}$. With this and the Lipschitz continuity from Lemma 4.43 we find

$$\limsup_{\|d\|_{L^2(Q)}\to 0} \frac{\|r_{\gamma kh}(u,d)\|_{L^2(I,V)}}{\|d\|_{L^2(Q)}} \le C \limsup_{\|d\|_{L^2(Q)}\to 0} (\delta + \|d\|_{L^2(Q)}) = C\delta.$$

As $\delta > 0$ was arbitrary we conclude the proof.

□

Remark 4.45 As with Theorem 3.11 and Theorem 3.13 we can deviate from the assumption $\beta_\gamma \in C^1(\mathbb{R})$ a little bit, by considering a β_γ of the form from Proposition 2.17 for $\alpha = 1$ and assuming $\{S_{\gamma kh}(u) = \Psi\}$ is of Lebesgue measure 0.

4.4 Convergence Rates for Solutions to Discretized, Regularized Obstacle Problems

Before we continue we would like to quickly address the question why we use first order finite elements in space and not possibly higher order elements. Especially in view of the high regularity of solutions to our variational inequalities and their regularizations. One reason is that affine elements are easier to handle from an implementation perspective. But, more importantly, we would like to give the following quote from [EG04, page 67]: "*In particular, if a domain with curved boundary is approximated geometrically with affine meshes, using finite elements of degree larger than one is not asymptotically more accurate than using first-order finite elements.*" So, to see an improvement in our situation by using higher order elements, one would have to use non-isoparametric elements and non-affine meshes. This, however, is out of the scope of this thesis.

We again stress that the assumptions and properties from Section 2.2.1 are assumed to hold throughout Chapter 4. The following assumptions shall also hold for the rest of Chapter 4. For some intermediate results the assumptions might be stronger than necessary. But for presentations' sake it is convenient to consider all assumptions from the beginning, as in the end we all need them anyway.

We also note that we will use results from Sections 8.6 to 8.8. They are formulated for $I = (0,1)$, but the results suited for $I = (0,T)$ can be all obtained by rescaling with the factors T and $\frac{1}{T}$. Very importantly we would like to stress that Assumption 4.47 is strong enough such that Lemma 8.53 applies, cf. Remark 8.54. This in particular entails that results about interpolation spaces from Section 8.7 are freely applicable.

Assumption 4.46 *We assume there exist constants $c_1, c_2 > 0$ such that:*

$$k^{c_2} \le c_1 \min_{m=1,\dots,M} k_m.$$

We further assume that there exists a $\kappa > 0$ such that

$$\kappa^{-1} \le \frac{k_m}{k_{m-1}} \le \kappa \quad \forall m = 1,\dots,M-1.$$

We also assume $k \le k_0 < \min(\frac{T}{4}, e^{-(N+1)})$ for some constant k_0.

Assumption 4.47 *We assume $(\mathcal{K}_h)_{h>0}$ to be quasi-uniform and shape regular.*

Assume $\Omega_h \subset \Omega$.

Assume that $(\Omega_h)_{h>0}$ and Ω are uniform Lipschitz domains. That means that there exist finitely many local coordinate systems $S_1, \ldots, S_m$ and local transformations $f_1, \ldots, f_m$, $f_1^h, \ldots, f_m^h$ for $h > 0$, as in Definition 1.5 for $k = 0$, $\alpha = 1$, such that the the boundaries of each Ω_h and Ω can be represented as a Lipschitz graph as in Definition 1.5. Also see the results in Proposition 4.7 for a rigorous formulation. If Ω is assumed to more regular than a Lipschitz $f_1, \ldots, f_m$ are assumed to be of the appropriate, higher regularity. There also exists a $C > 0$ such that all the Lipschitz constants of the f_i^h are bounded by C.

We also assume that

$$\max_{x \in \partial\Omega_h} \operatorname{dist}(x, \partial\Omega) = \mathcal{O}(h^2).$$

We shall also assume that $h \leq h_0 < e^{-(N+2)}$ for some constant h_0.

Assumption 4.47 is essentially a sharpening of Assumption 4.14 so that results from Chapter 8 are freely applicable.

Assumption 4.48 *Assume $N \in \{2, 3\}$. Assume Ω to be a $C^{3,\alpha}$-domain. Here $\alpha \in (0, 1)$ is some arbitrary but fixed constant. We also assume A has coefficients in $C^2(\bar{\Omega})$.*

Remark 4.49 The limiting factors on the regularity on Ω and A are Theorem 4.29 and Proposition 4.35. The regularity is required for pointwise estimates of the Green's function.

The numerical analysis for smooth domains is different to the numerical analysis for polygonal domains insofar that the solutions to the undiscretized problem are very regular globally, but we have to consider the error introduced by approximating the smooth Ω by a polygonal Ω_h.

4.4.1 An L^∞-error Estimate for Linear Parabolic PDEs

Our goal in the following is to derive $L^\infty(Q)$-error esimates for the linear parabolic problem and for the regularization (PDE$_\gamma$) of (VI-OB). The linear parabolic problem has been analysed, for example, in [STW98, STW80]. There, however, a semi-discretization of the linear parabolic problem was considered, i.e. the time was not discretized. Also the results have been derived under very strong regularity assumptions on solutions of

$$\begin{cases} \partial_t y - \Delta y = u \text{ in } Q, \\ y(0) = 0, \; y|_{\Sigma_D} = 0. \end{cases}$$

It was assumed that $y \in W^{1,2}_\infty(Q)$, which is stronger than one could expect, even for $u \in L^\infty(Q)$. For a semi-discretization for linear parabolic problems see also [Nit79].

So, to our knowledge the global $L^\infty(Q)$-error estimate for the linear parabolic problem in smooth domains is novel. In [Bon18, Section 5.3.3] interior estimates for polygonal domains were considered and the strategies used there and here were co-developed by the author of [Bon18], Lucas Bonifacius, and myself; see Sections 8.6 to 8.8.

The structure of this section is as follows: we combine a resolvent estimate for the operator A and with the $L^\infty(\Omega)$-quasi optimality of the Ritz projection to derive an appropriate quasi best approximation result for the Galerkin approximation in the time dependent case. This is the same strategy as in [LV16, LV17a, LV17b], where it was used for polygonal domains. Then we can use an interpolation error estimate, see Section 8.6 through Section 8.8, to get an error estimate for the Galerkin approximation.

Note, even though [LV17a] and the other mentioned works were written for polygonal domains the proofs we will cite work for our situation as well, since they only depend on the resolvent estimates for the discrete operators. We have proven those in Corollary 4.37. Since the proofs are so similar we only restate the most important lemmas required to prove Theorem 4.54 via the strategies of [LV16].

Remark 4.50 The following estimate can be generalized to different norms, e.g. weighted $L^2(\Omega)$-norms. See, for example, [LV16, Theorem 4.1.] in the case that Ω is convex and polygonal/polyhedral.

The proof of the following is essentially the combination of [LV17a, Theorem 1 and Corollary 1], which are the semi-discrete analogoues of this statement. The fully discrete version is stated in [LV17a, Theorem 10].

Lemma 4.51 *Let Assumption 4.47 and Assumption 4.48 be satisfied. Let $p \in [1, \infty]$. Let $y_0 \in L^p(\Omega)$ and y_{kh} defined by*

$$B(y_{kh}, \varphi_{kh}) = \left(y_0, \varphi_{kh,0}^+\right)_{L^2(Q)} \qquad \forall \varphi_{kh} \in X_{k,h}^{0,1}. \tag{4.35}$$

Then there exists a $C > 0$ independent of k, h and p such that

$$\sup_{t \in I_m} \|A_h y_{kh}(t)\|_{L^p(\Omega_h)} + k_m^{-1} \|[y_{kh}]_{m-1}\|_{L^p(\Omega_h)} \leq \frac{C}{t_m} \|P_h y_0\|_{L^p(\Omega_h)}.$$

Proof. The proof of [EJL98, Theorem 5.1] can be applied to any norm and not only the $L^2(\Omega)$-norm used there since Corollary 4.37 holds for any $p \in [1, \infty]$. It yields

$$\sup_{t \in I_m} \|A_h y_{kh}(t)\|_{L^p(\Omega_h)} \leq \frac{C}{t_m} \|P_h y_0\|_{L^p(\Omega_h)} \tag{4.36}$$

for any $m = 1, \ldots, M$, see the proof of [EJL98, (5.7)].

Formulating (4.35) as an Euler method we can prove the second part. This can be done by spelling out $B(y_{kh}, \varphi_{kh})$ for $\varphi_{kh} \in X_{k,h}^{0,1}$:

$$\begin{aligned} B(y_{kh}, \varphi_{kh}) = \sum_{m=2}^{M} \left([y_{kh}]_{m-1}, \varphi_{kh}(t_{m-1}^+)\right)_{L^2(\Omega)} + \left(y_{kh}(0^+), \varphi_{kh}(0^+)\right)_{L^2(\Omega)} \\ + k_m \left(A_h y_{kh}(t_{m-1}^+), \varphi_{kh}(t_{m-1}^+)\right)_{L^2(\Omega)}. \end{aligned}$$

So for any $\varphi_h \in V_h$ we have, by choosing $\varphi_{kh}(t,x) := \varphi_h(x) 1_{(t_{m-1}, t_m)}(t)$ for $m > 0$,

$$([y_{kh}]_{m-1}, \varphi_h)_{L^2(\Omega)} + k_m \left(A_h y_{kh}(t_{m-1}^+), \varphi_h\right)_{L^2(\Omega)} = 0.$$

As $\dim(V_h) < \infty$ and $\varphi_h \in V_h$ is arbitrary this implies $k_m^{-1}[y_{kh}]_{m-1} = -A_h y_{kh}(t_{m-1}^+)$. Now the claim follows from (4.36). □

The following lemma is an easy consequence of Lemma 4.51. The proof is given in [LV16, Lemma 12].

Lemma 4.52 *Let p, y, y_{kh} as in Lemma 4.51. Then there exists a constant $C > 0$ independent of k and h such that*

$$\sum_{m=1}^{M}\Big(\|A_h y_{kh}\|_{L^p(\Omega_h)} + k_m\|[y_{kh}]_{m-1}\|_{L^p(\Omega_h)}\Big) \le C \ln\frac{T}{k}\|P_h y_0\|_{L^p(\Omega_h)}.$$

For $m = 1$ the jump term is understood in the following sense:

$$[y_{kh}]_0 = y_{kh}(0^+) - P_h y_0.$$

The following is our version of [LV16, Lemma 13].

Lemma 4.53 *Let Assumption 4.47 and Assumption 4.48 be satisfied. Let $p \in [1,\infty]$. Let $f \in L^p(Q)$ and y_{kh} defined by*

$$B(y_{kh}, \varphi_{kh}) = (f, \varphi_{kh})_{L^2(Q)} \qquad \forall \varphi_{kh} \in X^{0,1}_{k,h}. \tag{4.37}$$

Let $s \in [1,\infty]$. Then there exists a $C > 0$ independent of k, h, p and s such that

$$\|A_h y_{kh}\|_{L^s(I,L^p(\Omega))} + \left(\sum_{m=1}^{M} k_m\|k_m^{-1}[y_{kh}]_{m-1}]\|^s_{L^p(\Omega)}\right)^{\frac{1}{s}} \le C\ln\frac{T}{k}\|P_h f\|_{L^s(I,L^p(\Omega))}$$

respectively

$$\|A_h y_{kh}\|_{L^\infty(I,L^p(\Omega))} + \max_{m\in\{1,\dots,M\}}\|k_m^{-1}[y_{kh}]_{m-1}]\|_{L^p(\Omega)} \le C\ln\frac{T}{k}\|P_h f\|_{L^\infty(I,L^p(\Omega))}.$$

Here P_h denotes the L^2-Projection of f onto V_h in each time point.

Proof. This is proven as [LV17a, Theorem 11], which refers to the semi-discrete case for the proof. The estimates for $\|A_h y_{kh}\|_{L^s(I,L^p(\Omega))}$ are proven as in [LV17a, Theorem 2], which is the aforementioned semi-discrete case. The estimates for the jump points follows, similarly to the proof of Lemma 4.51, by writing the discrete scheme as an Euler's method. This is done in detail in [LV17a, Corollary 2]. □

The following theorem is the analogue of [LV16, Theorem 1] for smooth domains.

Theorem 4.54 *Let $y \in W(I) \cap C(\bar{Q})$ and $y_{kh} \in X^{0,1}_{k,h}$ such that*

$$B(y - y_{kh}, \varphi_{kh}) = 0 \quad \varphi_{kh} \in X^{0,1}_{k,h}.$$

Then

$$\|y - y_{kh}\|_{L^\infty(Q)} \le C|\ln k||\ln h| \inf_{\varphi_{kh}\in X^{0,1}_{k,h}}\|y - \varphi_{kh}\|_{L^\infty(Q)} + \|y\|_{L^\infty(Q\backslash Q_h)}.$$

The constant $C > 0$ does not depend on k, h or y.

Proof. We first note that

$$\|y - y_{kh}\|_{L^\infty(Q)} = \max(\|y - y_{kh}\|_{L^\infty(Q_h)}, \|y - y_{kh}\|_{L^\infty(Q\backslash Q_h)}).$$

By definition of V_h we have

$$\|y - y_{kh}\|_{L^\infty(Q\backslash Q_h)} = \|y\|_{L^\infty(Q\backslash Q_h)}.$$

The proof of [LV16, Theorem 1] can be applied to $\|y - y_{kh}\|_{L^\infty(Q_h)}$ based on Lemma 4.53 and Lemma 4.51 for $p = 1$, Proposition 4.13 and Theorem 4.29. This yields the desired estimate. □

We can now state our main theorem for the Galerkin approximation of linear parabolic problems.

Theorem 4.55 *Assume $f \in L^\infty(Q), y_0 \in W^{2,\infty}(\Omega) \cap V$. Let $y \in W(I) \cap C(\bar{Q})$ be the weak solution of*

$$\begin{cases} \partial_t y + Ay = f \text{ on } Q, \\ y(0) = y_0, \; y|_{\Sigma_D} = 0. \end{cases}$$

Let y_{kh} be the Ritz projection of y, i.e.

$$B(y - y_{kh}, \varphi_{kh}) = 0 \quad \forall \varphi_{kh} \in X^{0,1}_{k,h}.$$

Then we have

$$\|y - y_{kh}\|_{L^\infty(Q)} \le C|\ln k|^2 |\ln h|^4 (k + h^2) \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right).$$

Proof. By Theorem 4.54 and the regularity from Theorem 8.17 the previous Theorem 4.54 applies and we have

$$\|y - y_{kh}\|_{L^\infty(Q)} \le C|\ln k||\ln h| \inf_{\varphi_{kh} \in X^{0,1}_{k,h}} \|y - \varphi_{kh}\|_{L^\infty(Q)} + \|y\|_{L^\infty(Q \setminus Q_h)}.$$

Choosing $\varphi_{kh} = I_k I_h y$ we find the estimate

$$\begin{aligned} \|y - y_{kh}\|_{L^\infty(Q)} &\le C|\ln k||\ln h| (\|y - I_k y\|_{L^\infty(Q)} + \|I_k(y - I_h y)\|_{L^\infty(Q)}) + \|y\|_{L^\infty(Q\setminus Q_h)} \\ &\le C|\ln k||\ln h| (\|y - I_k y\|_{L^\infty(Q)} + \|y - I_h y\|_{L^\infty(Q)}) + \|y\|_{L^\infty(Q\setminus Q_h)}. \end{aligned} \tag{4.38}$$

We first estimate the last term as it is the easiest. We give an estimate of the Hölder norm of the spatial part of y. By (8.34), a consequence of Proposition 8.39, and the following remark we have

$$W^{1,2}_p(Q) \hookrightarrow C\left(\bar{I}, \left(L^p(\Omega), W^{2,p}(\Omega)\right)_{1-1/p,p}\right)$$

with an embedding constant independent of p. Thus by Corollary 8.24 we have

$$\|y\|_{C(\bar{I},(L^p(\Omega),W^{2,p}(\Omega))_{1-1/p,p})} \le C\|y\|_{W^{1,2}_p(Q)} \le Cp^2 \left(\|f\|_{L^p(Q)} + \|y_0\|_{\mathbb{W}_p} \right).$$

By Proposition 8.32 and Remark 8.33 we have:

$$\|y\|_{C(\bar{I},(L^p(\Omega),W^{2,p}(\Omega))_{1-1/p,p})} \le Cp^2 \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right). \tag{4.39}$$

By Proposition 8.57 this implies for $p \ge N + 2 > 2$

$$\|y\|_{C(\bar{I}, W^{2(1-1/p),p}(\Omega))} \le C_p p^2 \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right).$$

By the same lemma C_p has the behaviour $\sim p$ for $p \to \infty$. By Lemma 8.59 this implies in turn for $p > N + 2$.

$$\|y\|_{C(\bar{I}, C^{1,1-(N+2)/p}(\Omega))} \le Cp^3 \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right). \tag{4.40}$$

The Hölder exponent is $2(1 - 1/p) - 1 - N/p = 1 - 2/p - N/p = 1 - (N + 2)/p$. This implies in particular, that $y(t, \cdot)$ is Lipschitz continuous in space for ever $t \in \bar{I}$.

Now let $(t,x) \in Q \setminus Q_h$. By $y|_{\Gamma_D} = 0$ and (4.40) we have for $p > N+2$ and for some $\hat{x} \in \partial\Omega$ with $\operatorname{dist}(x, \partial\Omega) = |\hat{x} - x|_2$ that

$$|y(t,x)| = |y(t,x) - y(t,\hat{x})| \leq \|y\|_{C(\bar{I}, C^{1,1-(N+2)/p}(\Omega))} |x - \hat{x}| \\ \leq Cp^3 |x - \hat{x}| \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right).$$

By assumption we have $|\hat{x} - x| \leq Ch^2$. Choosing $p = N+3 > N+2$ therefore yields

$$|y(t,x)| \leq Ch^2 \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right).$$

Thus the third term of (4.38) satisfies the given bound.

The second term of (4.38) *is treated similarly.* Let $t \in \bar{I}$ and $x \in \bar{\Omega}_h$ such that

$$|y(t,x) - I_h y(t,x)| = \|y(t,\cdot) - I_h y(t,\cdot)\|_{L^\infty(\Omega_h)}.$$

By the definition of Ω_h there has to exist a cell $K \in \mathcal{K}_h$ such that $x \in \bar{K}$. Further, there has to exist a subset $\operatorname{conv}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_l) =: E \subset K$ such that $\hat{x} \in \operatorname{relint}(E)$. Here, $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_l$ are corners of K. That basically means that the maximizer x has to lie in the relative interior of one of the facets of K. If we had $l = 1$, i.e. x is a corner of K, we could infer

$$0 = |y(t,x) - I_h y(t,x)| = \|y(t,\cdot) - I_h y(t,\cdot)\|_{L^\infty(\Omega_h)}.$$

This is a trivial case that obviously satisfies the given estimate. We may therefore assume $l \geq 2$. We define the one dimensional function

$$f \colon (0, 1+\epsilon) \to \mathbb{R}, \\ s \mapsto y(t, \hat{x}_1 + s(x - \hat{x}_1)) - I_h y(t, \hat{x}_1 + s(x - \hat{x}_1)).$$

Here, $\epsilon > 0$ is so small such that $\hat{x}_1 + s(x - \hat{x}_1) \in \bar{K}$ for all $s \in (0, 1+\epsilon)$. This is well-defined due to x lying in the relative interior of E. Obviously f has an extremum at $s = 1$ with $|f(1)| = \|y(t,\cdot) - I_h y(t,\cdot)\|_{L^\infty(\Omega_h)}$. If $E \subset \partial K$ have by Proposition 4.56 that f is continuously differentiable. If $E = K$ the differentiability of f is clear. So in either case

$$0 = f'(1) = \nabla(y(t,x) - I_h y(t,x))^T (x - \hat{x}_1).$$

We therefore find for some $\xi_s \in (0,1)$

$$|y(t,x) - I_h y(t,x)| = |f(1) - f(0)| = |f'(\xi_s)| = |f'(\xi_s) - f'(1)| \\ = |\nabla((y(t,\xi) - I_h y(t,\xi)) - (y(t,x) - I_h y(t,x)))^T (x - \hat{x}_1)| \\ \leq |\nabla(y(t,\xi) - y(t,x))| |x - \hat{x}_1|.$$

Here, $\xi = \hat{x}_1 + s(x - \hat{x}_1) \in \operatorname{relint}(E)$. By the Hölder continuity of ∇y from (4.40) we can deduce for any $p > N+2$

$$|y(t,x) - I_h y(t,x)| \leq Ch^{2-(N+2)/p} p^3 \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right).$$

Choosing $p = |\ln h|$ we deduce

$$\|y - I_h y\|_{L^\infty(Q)} \leq C |\ln h|^3 h^2 \left(\|f\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)} \right).$$

Thus the second term in (4.38) satisfies the claimed bound as well.

All that remains to estimate is the first term on the right hand side of (4.38). By (8.35), a consequence of Proposition 8.39, we have the following embedding for $s \in [2, \infty)$, $p \in [2, \infty)$ and $\tau \in (0, 1)$

$$W^{1,s}(I, L^p(\Omega)) \cap L^s(I, \mathrm{dom}_p(A)) \hookrightarrow C^\alpha(I, (L^p(\Omega), \mathrm{dom}_p(A))_{\tau,1})$$

for any $\alpha \in [0, 1 - 1/s - \tau)$. The embedding constant does not depend on s, p, τ or α. For $\tau = N/p$ we have by Proposition 8.51

$$W^{1,s}(I, L^p(\Omega)) \cap L^s(I, \mathrm{dom}_p(A)) \hookrightarrow C^\alpha(I, C(\bar{\Omega}))$$

where the embedding constant is bounded by

$$C \frac{\Gamma(\frac{N}{2p})}{\Gamma(\frac{N}{p})}$$

where Γ is the Γ-function. This quotient is bounded independently of p by Proposition 8.6. Now Corollary 8.50 yields

$$\begin{aligned} \|y\|_{C^\alpha(I,C(\bar{\Omega}))} &\le C\|y\|_{W^{1,s}(I,L^p(\Omega))\cap L^s(I,\mathrm{dom}_p(A))} \\ &\le C\frac{s^2}{s-1}(\|u\|_{L^s(I,L^p(\Omega))} + \|y_0\|_{(L^p(\Omega),\mathrm{dom}_p(A))_{1-1/s,s}}). \end{aligned}$$

By Proposition 8.32 and Remark 8.33 we can conclude for $s > N + 1$

$$\|y\|_{C^\alpha(I,C(\bar{\Omega}))} \le Cs(\|u\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)})$$

for any $\alpha \in [0, 1 - 1/s - N/p)$. Choosing $s = p \ge |\ln k_0| > N + 1$ and $\alpha = 1 - |\ln k_0|/s$ results in

$$\|y - I_k y\|_{L^\infty(Q)} \le Ck^{1-(N+2)/s} s(\|u\|_{L^\infty(Q)} + \|y_0\|_{W^{2,\infty}(\Omega)}).$$

Choosing $s = |\ln k| \ge |\ln k_0|$ yields the desired estimate for the first term of (4.38). □

Proposition 4.56 *Let $D \subset \mathbb{R}^N$ be a polygonal domain. Let* $\mathrm{conv}(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_l) = E \subset \partial D$. *Let $g \in C^1(D) \cap C(\bar{D})$ such that ∇g can be extended to a continuous function on $\bar{D}$. Lastly, let $a, b \in$* $\mathrm{relint}(E)$. *Then the function*

$$\begin{aligned} f\colon (0,1) &\to \mathbb{R}, \\ s &\mapsto g(a + s(b - a)) \end{aligned}$$

lies in $C^1((0,1))$. Its derivative is given by $g'(s) = \nabla f(a + s(b - a))^T(b - a)$.

Proof. The idea will be to, more or less, shift the analysis into the open set D, where standard definitions and results apply. Let all the quantities be as above and $s \in (0, 1)$ be fixed. We abbreviate $c_s := a + s(b - a)$. Since a polygonal domain is also a Lipschitz domain, it satisfies the cone condition from Definition 1.8 via Theorem 1.9. The cone condition implies that there exists an open set $U \subset \mathbb{R}^N$ containing c_s and a cone C such that for each $x \in U \cap \partial\Omega$ we have $x + C \subset D$. Let ν be the axis of the cone, then for each $x \in U \cap E$ we have for any $\epsilon \in (0, 1)$ that $x + \epsilon\nu \in D$. Lastly, since U is open we can choose a $d_0 > 0$ small enough such that $B_{d_0}(c_s) \subset U$ and $B_{d_0}(c_s) \cap E \subset \mathrm{relint}(E)$.

We then have for any $d \in \mathbb{R}$, with $|d|$ sufficiently small, that

$$\begin{aligned} &|f(s + d) - f(s) - \nabla g(c_s)^T(b - a)d| \\ &\le |g(c_{s+d}) - g(c_{s+d} - \epsilon\nu) - g(c_s) + g(c_s - \epsilon\nu) - \nabla g(c_s)^T(b - a)d + \nabla g(c_s - \epsilon\nu)^T(b - a)d| \\ &\quad + |g(c_{s+d} - \epsilon\nu) - g(c_s - \epsilon\nu) - \nabla g(c_s - \epsilon\nu)^T(b - a)d|. \end{aligned} \tag{4.41}$$

We analyse the second term on the right hand side. It is equal to

$$\left| \int_0^1 \nabla g(c_s - \epsilon\nu + r(c_{s+d} - c_s))^T (b-a) - \nabla g(c_s - \epsilon\nu)^T (b-a)\, dr \right| |d|$$

and further bounded by

$$\int_0^1 |\nabla g(c_s - \epsilon\nu + r(c_{s+d} - c_s)) - \nabla g(c_s - \epsilon\nu)|\, dr |b-a||d|.$$

By the theorem of dominated convergence, e.g. [BK15, Proposition 5.4], this converges for $\epsilon \to 0$ to the following

$$\int_0^1 |\nabla g(c_s + r(c_{s+d} - c_s)) - \nabla g(c_s)|\, dr |b-a||d|.$$

Thus, after taking the limit $\epsilon \to 0$ in (4.41), we find with the above and the continuity of g and ∇g that

$$|g(s+d) - g(s) - \nabla f(c_s)^T (b-a)d| \leq \int_0^1 |\nabla g(c_s + r(c_{s+d} - c_s)) - \nabla g(c_s)|\, dr |b-a||d|.$$

Using the theorem of dominated convergence again, we see that the right hand side is of order $o(|d|)$, which concludes the proof.

□

4.4.2 An L^∞-error Estimate for the Discretization of Regularized Obstacle Problems

Remark 4.57 Before we prove our estimate:

$$\|S_\gamma(u) - S_{\gamma kh}(u)\|_{L^\infty(Q)} \leq C |\ln h|^2 |\ln k|^2 (k + h^2)$$

with C independent of $\gamma > 0$ we would like to adress the work of [Fet87]. In [Fet87] it is proven that under strong regularity assumptions an appropriate discretization of (VI-OB), using the same spaces for the discretization we use, directly leads to an $L^\infty(Q)$-error estimate of the order $C(\epsilon)ck^{-\epsilon}h^{-\epsilon}(k+h^2)$. The assumptions are that Ω is a two-dimensional, smooth convex domain and $A = -\Delta$. Further, the solution $y = S(u)$ of (VI-OB) satisfies

$$\begin{aligned} y \in L^\infty(I, W^{2,p}(\Omega)), \quad \forall p \in [1,\infty), \\ \partial_t y \in L^\infty(I, L^\infty(\Omega)), \\ \Delta y \in L^\infty(I, L^\infty(\Omega)). \end{aligned} \tag{4.42}$$

It shall also satisfy

$$\partial_{tt} y \in L^2(Q). \tag{4.43}$$

Further the triangulation has no angle that exceeds $\frac{\pi}{2} - c_1$ for some fixed $c_1 > 0$ and that there exists a $c_2 > 0$ such that $k \geq c_2 h^2$.

The regularities in (4.42) are satisfied under the assumptions $u \in C(I, L^\infty(\Omega))$, $\partial_t u \in L^1(I, L^\infty(\Omega))$, $y_0 \in W^{2,\infty}(\Omega) \cap V$ and $y_0 \geq \Psi$; see [Fet87, remark before (1.3)] and the there mentioned [Bré72]. Under appropriate conditions one can show that (4.42) and (4.43) both hold true for the one phase Stefan problem; see [FK75] and the corresponding [Fet87, remark above the first theorem].

In [YWG14] the authors prove a rate of $\mathcal{O}(h + k^{\frac{1}{2}})$ for a very similar numerical scheme and a simplified version of our obstacle problem. This result implicitly uses a regularization approach that is, however, different from ours. The difficulty in [YWG14] is the initial condition which only lies in V, while our initial condition will lie in $W^{2,\infty}(\Omega) \cap V$.

The main result of [GM19] is a convergence rate of $\mathcal{O}(h + k^{\frac{3}{4}})$ times some logarithmic factors in a specific, pointwise discrete norm. They do not use any form of regularization, but make high regularity assumptions on the appearing quantities and use conforming finite elements.

[OS16] discusses a full discretization of the parabolic obstacle problem involving fractional operators which is note-worthy, but not close to our question. Statements about semi-discretizations of semilinear parabolic problems and their convergence rates under minimal regularity assumptions can for example be found in [CH02].

The next proofs and ideas are based on the elliptic case studied in [Noc88]. We start by introducing some form of a regularized parabolic Green's function.

Definition 4.58 For $0 \leq b \in L^\infty(Q)$, and $\tilde{\delta} \in L^\infty(Q)$ with $\|\tilde{\delta}\|_{L^1(Q)} \leq 1$ we define $G \in W(I)$ as the solution of

$$\begin{cases} -\partial_t G + AG + bG = \tilde{\delta}, \\ G(T) = 0, \ G|_{\Sigma_D} = 0. \end{cases} \tag{4.44}$$

See Theorem 8.17 for existence and regularity of G.

Its Galerkin approximation is denoted by $G_{kh} \in X_{k,h}^{0,1}$ and is defined via

$$B_k(\varphi_{kh}, G - G_{kh}) = 0 \qquad \forall \varphi_{kh} \in X_{k,h}^{0,1}.$$

Note that b does not play a role in the definition of G_{kh}.

Later b will be a sum of terms involving β_γ and f. In [Noc88] the $\tilde{\delta}$ is some form of regularization of the delta distribution centered at an appropriate point satisfying $\|\tilde{\delta}\|_{L^1(Q)} = 1$. Our argument is subtly different, as it utilizes duality where this interpretation is not necessary, see [HKP19, Section 4], where we first used this approach.

The general idea is to estimate the $L^\infty(Q)$-error of the states by the $L^1(Q)$-error of G and G_{kh}. This $L^1(Q)$-error in turn will be estimated by the $L^\infty(Q)$-error for a linear PDE. While this is our motivation the actual arguments are made in reverse; starting with the estimation of the $L^1(Q)$-error of G and G_{kh}.

Lemma 4.59 *For the G from Definition 4.58 we have*

$$\|bG\|_{L^1(Q)} \leq 1 \ \textit{and} \ \|-\partial_t G + AG\|_{L^1(Q)} \leq 2.$$

Proof. For $\epsilon > 0$ we define $\operatorname{abs}_\epsilon(x) := \sqrt{x^2 + \epsilon}$ and $\operatorname{sgn}_\epsilon(x) := \operatorname{abs}'_\epsilon(x) = \frac{x}{\sqrt{x^2+\epsilon}}$. We now test (4.44) with $\operatorname{sgn}_\epsilon(G)$ to obtain

$$(-\partial_t G + AG + bG, \operatorname{sgn}_\epsilon(G))_{L^2(Q)} = \left(\tilde{\delta}, \operatorname{sgn}_\epsilon(G)\right)_{L^2(Q)} \leq 1.$$

We now estimate the terms without b starting with

$$\begin{aligned} -(\partial_t G, \operatorname{sgn}_\epsilon(G))_{L^2(Q)} &= -\int_I \partial_t \left(\int_\Omega \operatorname{abs}_\epsilon(G)\, dx \right) dt \\ &= -\int_\Omega \operatorname{abs}_\epsilon(G)(T) - \operatorname{abs}_\epsilon(G)(0)\, dx = \int_\Omega \operatorname{abs}_\epsilon(G)(0)\, dx \geq 0. \end{aligned}$$

Now we treat the elliptic term

$$(AG, \operatorname{sgn}_\epsilon(G))_{L^2(Q)} = \sum_{i,j=1}^{N} \int_Q a_{ij}\partial_{x_i} G \operatorname{sgn}'_\epsilon(G)\partial_{x_j} G \, d(t,x) \geq 0.$$

Here we used the a.e. positive semi-definiteness of the matrix $(a_{ij})_{i,j=1,\ldots,N}$ and the fact that $\operatorname{sgn}_\epsilon$ is monotonically increasing.

Thus, we have proven so far

$$1 \geq \int_Q b \frac{G^2}{\sqrt{G^2+\epsilon}} \, d(t,x).$$

The integrand on the right-hand side is bounded by $\|b\|_{L^\infty(Q)} \cdot |G| \in L^1(Q)$, so that the theorem of dominated convergence, e.g. [BK15, Proposition 5.4], yields

$$1 \geq \lim_{\epsilon\to 0} \int_Q b \frac{G^2}{\sqrt{G^2+\epsilon}} \, d(t,x) = \int_Q b|G| \, d(t,x) = \|bG\|_{L^1(Q)}.$$

Here we used that $b \geq 0$.

Using the regularity from Theorem 8.17 we can write a.e. in Q that $-\partial_t G + AG = -bG + \tilde{\delta}$. Therefore we can easily conclude

$$\|-\partial_t G + AG\|_{L^1(Q)} \leq \|bG\|_{L^1(Q)} + \|\tilde{\delta}\|_{L^1(Q)} \leq 2.$$

□

Lemma 4.60 *Let G be as in Definition 4.58. For $\Phi_Q \in L^\infty(Q), \Phi_T \in W^{2,\infty}(\Omega) \cap V$ we have*

$$(G - G_{kh}, \Phi_Q)_{L^2(Q)} + \left(G(0) - G_0^+, \Phi_0\right)_{L^2(\Omega)} \leq 2\|\zeta - \zeta_{kh}\|_{L^\infty(Q)}$$

where ζ is defined as the solution to

$$\begin{cases} \partial_t \zeta + A\zeta = \Phi_Q, \\ \zeta(0) = \Phi_0, \ \ \Phi|_{\Sigma_D} = 0, \end{cases} \tag{4.45}$$

and ζ_{kh} as its Galerkin approximation, i.e. ζ_{kh} satisfies

$$B_k(\zeta - \zeta_{kh}, \varphi_{kh}) = 0 \qquad \forall \varphi \in X^{0,1}_{k,h}.$$

Proof. Using the bilinear form B_k equation (4.45) is equivalent to

$$B_k(\zeta, \varphi) = \left(\Phi_0, \varphi_0^+\right)_H + (\Phi_Q, \varphi)_{L^2(Q)} \qquad \forall \varphi \in X_k^0.$$

Thus

$$(G - G_{kh}, \Phi_Q)_{L^\infty(Q)} + \left(G(0) - G_0^+, \Phi_0\right)_{L^2(\Omega)} = B_k(\zeta, G - G_{kh}) = B_k(\zeta - \zeta_{kh}, G - G_{kh}).$$

Here we used the definition of G_{kh}. Using the same idea for ζ_{kh} the whole expression equals

$$B_k(\zeta - \zeta_{kh}, G) = -\sum_{m=1}^{M} (\zeta - \zeta_{kh}, \partial_t G)_{L^2(I_m, V, V^*)} + a_I(\zeta - \zeta_{kh}, G) = (-\partial_t G + AG, \zeta - \zeta_{kh})_{L^2(Q)}$$

where we used the fact that $G \in W_2^{1,2}(Q) \subset W(I)$ and the secondary formulation of B_k. Using the previous Lemma 4.59 we can derive the estimate immediately. □

Theorem 4.61 *Let $\hat{f} : Q\times\mathbb{R} \to \mathbb{R}$ that satisfies all the properties of f from Definition 2.6. In addition we assume there exists a $C_\infty > 0$ such that $|\hat{f}(t,x,y)| \leq C_\infty$ for all $(t,x,y) \in Q\times\mathbb{R}$. Let $u \in L^\infty(Q)$ and $y_0 \in W^{2,\infty}(\Omega)\cap V$. Let y be the solution to*

$$\begin{cases} \partial_t y + Ay + \hat{f}(y) = u, \\ y|_\Sigma = 0, \;\; y(0) = y_0, \end{cases} \tag{4.46}$$

and let y_{kh} be its Galerkin approximation for some $k, h > 0$. Then there exists a $C > 0$, independent of y, y_{kh}, k, h and $\hat{f}$ such that

$$\begin{aligned} &\|y - y_{kh}\|_{L^\infty(Q)} + \|y(T) - y_{kh}(T)\|_{L^\infty(\Omega)} \\ &\qquad \leq C|\ln k|^2|\ln h|^2(k+h^2)\left(\|u\|_{L^\infty(Q)} + C_\infty + \|y_0\|_{W^{2,\infty}(\Omega)}\right). \end{aligned}$$

Proof. We define

$$b := \begin{cases} \frac{\hat{f}(y)-\hat{f}(y_{kh})}{y-y_{kh}} & \text{if } y - y_{kh} \neq 0, \\ 0 & \text{else.} \end{cases}$$

By the monotonicity of $\hat{f}$ this is non-negative. By the local Lipschitz continuity of $\hat{f}$ the b stays bounded as well. Now let $\tilde{\delta} \in L^\infty(Q)$ with $\|\tilde{\delta}\|_{L^1(Q)} \leq 1$. We consider G with this specific b and $\tilde{\delta}$. We also define the error $e_{kh} := y - y_{kh}$. Then we have

$$\begin{aligned} \left(e_{kh}, \tilde{\delta}\right)_{L^2(Q)} &= (bG, e_{kh})_{L^2(Q)} + B_k(e_{kh}, G) \\ &= \left(\hat{f}(y) - \hat{f}(y_{kh}), G\right)_{L^2(Q)} + B_k(y - y_{kh}, G) \\ &= \left(\hat{f}(y) - \hat{f}(y_{kh}), G - G_{kh}\right)_{L^2(Q)} + B_k(y - y_{kh}, G - G_{kh}). \end{aligned}$$

Here we used the Galerkin orthogonality of $y - y_{kh}$. We decompose this into the parts involving y and those involving y_{kh}:

$$\begin{aligned} &\left(\hat{f}(y) + \partial_t y + Ay, G - G_{kh}\right)_{L^2(Q)} + \left(y_0, G(0) - G_0^+\right)_H \\ &\qquad = (u, G - G_{kh})_{L^2(Q)} + \left(y_0, G(0) - G_0^+\right)_H. \end{aligned}$$

The parts involving y_{kh} are easily estimated, by the Galerkin orthogonality of G and G_{kh}

$$\left(\hat{f}(y_{kh}), G - G_{kh}\right)_{L^2(Q)} + B_k(y_{kh}, G - G_{kh}) = \left(\hat{f}(y_{kh}), G - G_{kh}\right)_{L^2(Q)}.$$

Thus we have

$$\left(e_{kh}, \tilde{\delta}\right)_{L^2(Q)} = (u, G - G_{kh})_{L^2(Q)} + \left(y_0, G(0) - G_0^+\right)_H - \left(\hat{f}(y_{kh}), G - G_{kh}\right)_{L^2(Q)}. \tag{4.47}$$

Applying Lemma 4.60 and Theorem 4.55 to $\Phi_Q = u - \hat{f}(y_{kh})$ and $\Phi_0 = y_0$ we obtain

$$\left(e_{kh}, \tilde{\delta}\right)_{L^2(Q)} \leq C|\ln k|^2|\ln h|^2(k+h^2)\left(\|u\|_{L^\infty(Q)} + C_\infty + \|y_0\|_{W^{2,\infty}(\Omega)}\right). \tag{4.48}$$

The constant C is from Theorem 4.55 and thus does not depend on $\hat{f}$ or any of the other indicated quantities. Since $C_c^\infty(Q)$ is dense in $L^1(Q)$ one can easily deduce that

$$\left\{\varphi \in C_c^\infty(Q) : \|\varphi\|_{L^1(Q)} \leq 1\right\}$$

is dense in

$$\left\{\varphi \in L^1(Q) : \|\varphi\|_{L^1(Q)} \leq 1\right\}.$$

Thus

$$\begin{aligned}
\|e_{kh}\|_{L^\infty(Q)} &= \sup_{\tilde{\delta} \in \left\{\varphi \in C_c^\infty(Q) : \|\varphi\|_{L^1(Q)} \leq 1\right\}} \left(e_{kh}, \tilde{\delta}\right)_{L^2(Q)} \\
&\leq C |\ln k|^2 |\ln h|^2 (k + h^2) \left(\|u\|_{L^\infty(Q)} + C_\infty + \|y_0\|_{W^{2,\infty}(\Omega)}\right).
\end{aligned}$$

To see the end time estimate note that for each k there is an intervall $I_{M-1} = (t_{M-1}, T] \subset I$ such that y_{kh} is constant on. Thus

$$\|y(T) - y_{kh}(T)\|_{L^\infty(\Omega)} \leq \|y - y_{kh}\|_{L^\infty(I_{M-1} \times \Omega)} \leq \|y - y_{kh}\|_{L^\infty(Q)}.$$

□

Remark 4.62 To apply the previous result to our regularized obstacle problem we have to discuss one last hurdle: β_γ and, in general, f are not bounded by some $C_\infty > 0$ which is required to obtain (4.48) from (4.47), which is a crucial step in the proof of Theorem 4.61. There might be the option to prove that $|\beta_\gamma(y_{\gamma kh}) + f(y_{\gamma kh})|$ is bounded independently of k, h or γ, but this is, as of now, future research, which would have a lot of interesting implications. We go a different route and show that we can replace β_γ and f by bounded functions without changing the regularizing PDE.

Assume $u, \partial_t \Psi, A\Psi \in L^\infty(Q)$. Then, by Proposition 2.36, we have an upper bound $C_\infty > 0$ independent of $\gamma > 0$ such that

$$\|\beta_\gamma(y_\gamma - \Psi)\|_{L^\infty(Q)}, \|f(y_\gamma)\|_{L^\infty(Q)} \leq C_\infty - 1 \text{ for any } \gamma > 0.$$

The bound C_∞ depends only on an upper bound of $\|u\|_{L^\infty(Q)}$, $\|f(0)\|_{L^\infty(Q)}$, $\|f(\Psi)\|_{L^\infty(Q)}$, $\|\partial_t \Psi\|_{L^\infty(Q)}$, $\|A\Psi\|_{L^\infty(Q)}$ and $\|y_0\|_{L^\infty(\Omega)}$.

By the monotonicity of β_γ, the continuity of β_γ and the basic assumption that $\beta_\gamma(r) \xrightarrow{r \to -\infty} -\infty$ there is a largest $R_\infty \in (0, \infty)$ such that $\beta_\gamma(-R_\infty) = -C_\infty + \frac{1}{2}$. Thus we can consider a function $\hat{\beta}_\gamma$ that satisfies:

- $\hat{\beta}_\gamma \in C(\mathbb{R})$, $\hat{\beta}_\gamma \in C^1(\mathbb{R})$ provided $\beta_\gamma \in C^1(\mathbb{R})$,
- $\hat{\beta}_\gamma|_{[-R_\infty, \infty)} = \beta_\gamma|_{[-R_\infty, \infty)}$,
- $\|\hat{\beta}_\gamma\|_{L^\infty(\mathbb{R})} \leq C_\infty$.

Since $\beta_\gamma(y_\gamma - \Psi) > -C_\infty + 1 > -C_\infty + \frac{1}{2}$ we have $y_\gamma - \Psi > -R_\infty$. Thus $\beta_\gamma(y_\gamma - \Psi) = \hat{\beta}_\gamma(y_\gamma - \Psi)$ and we can replace β_γ by $\hat{\beta}_\gamma$ in (PDE$_\gamma$) and essentially change nothing. So we may assume $\|\beta_\gamma\|_{L^\infty(\mathbb{R})} \leq C_\infty$ and still solve the same equation. Importantly, note that the truncation point R_∞ does depend on $\gamma > 0$, while the cut-off height C_∞ does not.

In the case that $\beta_\gamma \in C(\mathbb{R})$ we can simply choose

$$\hat{\beta}_\gamma(r) := \max(-C_\infty, \beta_\gamma(r)) \in C(\mathbb{R}). \tag{4.49}$$

This obviously satisfies the desired properties. Clearly the same construction cannot be used for the case when one wants to keep differentiability. This can be for example achieved by defining for $\beta_\gamma \in C^1(\mathbb{R})$

$$\hat{\beta_\gamma}(t) = \begin{cases} -C_\infty & \text{if } t < -R_\infty - \frac{1}{\beta_\gamma'(-R_\infty)}, \\ -C_\infty + \frac{\beta_\gamma'(-R_\infty)^2}{2}\left(t + R_\infty + \frac{1}{\beta_\gamma'(-R_\infty)}\right)^2 & \text{if } -R_\infty - \frac{1}{\beta_\gamma'(-R_\infty)} \leq t < -R_\infty, \\ \beta_\gamma(t) & \text{if } -R_\infty \leq t. \end{cases} \tag{4.50}$$

Higher smoothness preservation is obviously possible by using more sophisticated formulae or smoothing kernels.

The analogous line of arguing yields that we may assume $\|f\|_{L^\infty(I\times\Omega\times\mathbb{R})} \leq C_\infty$ by Proposition 2.36. Note that here C_∞ additionally depends on the Lipschitz constant of f on a ball with radius $\|y_0\|_{L^\infty(\Omega)}$.

Whenever we refer to this truncation strategy we implicitly use (4.49) for $\beta_\gamma \in C(\mathbb{R}) \setminus C^1(\mathbb{R})$ and (4.50) for $\beta_\gamma \in C^1(\mathbb{R})$.

Remark 4.63 Note that these truncated functions cannot be used in Chapter 2, *because* they are bounded and the unboundedness of β_γ is used in proving $\lambda(S(u)) = \lim_{\gamma\to\infty} -\beta_\gamma(S_\gamma(u))$; see the middle portion of the proof of Theorem 2.30 with its contradiction argument.

We obtain an immediate, but important corollary:

Theorem 4.64 *Assume that $u, \partial_t\Psi, A\Psi \in L^\infty(Q)$, $y_0 \in W^{2,\infty}(\Omega) \cap V$ and that the results from Remark 4.62 hold true, i.e. β_γ and f are truncated at height C_∞. Then*

$$\begin{aligned} &\|y_\gamma - y_{\gamma kh}\|_{L^\infty(Q)} + \|y_\gamma(T) - y_{\gamma kh}(T)\|_{L^\infty(\Omega)} \\ &\quad \leq C|\ln k|^2|\ln h|^4(k + h^2)\left(\|u\|_{L^\infty(Q)} + C_\infty + \|y_0\|_{W^{2,\infty}(\Omega)}\right) \end{aligned}$$

for $y = S(u)$ and $y_{\gamma kh} = S_{\gamma kh}(u)$ for $\gamma > 0$.

The constants C and C_∞ do not depend on γ.

Proof. This follows from Theorem 4.61 with $\hat{f}(t,x,y) = \beta_\gamma(y - \Psi(t,x)) + f(t,x,y)$ as both, β_γ and f are truncated. □

We can now give a complete error estimate for an essentially computable approximation of $y = S(u)$. We say "essentially", because on still needs to discuss how to evaluate all the appearing integrals.

Theorem 4.65 *Assume that $u, \partial_t\Psi, A\Psi \in L^\infty(Q)$, $y_0 \in W^{2,\infty}(\Omega) \cap V$ and that the results from Remark 4.62 hold true, i.e. β_γ and f are truncated at height C_∞. Assume that β has the form of Proposition 2.17 for some $\alpha \geq 1$. Then*

$$\begin{aligned} &\|y - y_{\gamma kh}\|_{L^\infty(Q)} + \|y(T) - y_{\gamma kh}(T)\|_{L^\infty(\Omega)} \\ &\quad \leq C\left(\gamma^{\frac{1}{\alpha}} + |\ln k|^2|\ln h|^2(k + h^2)\right)\left(\|u\|_{L^\infty(Q)} + C_\infty + \|y_0\|_{W^{2,\infty}(\Omega)}\right) \end{aligned}$$

for $y = S(u)$ and $y_{\gamma kh} = S_{\gamma kh}(u)$ for $\gamma > 0$.

The constants C and C_∞ do not depend on γ, k or h.

Proof. This is a direct consequence of Theorem 2.37 and Theorem 4.64. □

Remark 4.66 Concluding this chapter we would like to shortly comment on the difficulties one has when trying to transfer our strategies to a polygonal domain Ω. One is obviously the lack of regularity at the corners of the polygon. While we still have Hölder regularity according to Theorem 2.42, it is not clear how exactly the exponents behave as one can also see in Theorem 2.42 that the space-time-derivative regularity, which we used heavily for our interpolation error estimates in Theorem 4.55, is limited by the polygonality of the domain.

One way around this, is the use of interior estimates, where one picks smooth subdomains of Ω and derives the estimates on those. This is possible, e.g. [Bon18, Section 5.3.3], but yields difficulties when trying to emulate the duality arguments from [Noc88].

Another approach was made in [HKP19], where $L^2(\Omega)$-estimates for the regularized, elliptic obstacle problem were derived, by 'cutting out' the corners in the estimates. Whether this approach is transferable to the parabolic situation, is, as of this moment, unknown.

5 Numerical Analysis of Discretized, Regularized Optimal Control Problems

5.1 Semi-discrete, Regularized Control Problems

Before we go into the details we would like to comment on related approaches of discretizing an optimal control problem involving the obstacle problem. We focus on a priori error estimates.

One great result, using an elliptic variational inequality and not a parabolic one, is found in [MT13]. Under the assumption that a discrete maximum principle is satisfied, the authors derive a (quasi optimal) convergence order for the controls in the $L^2(\Omega)$-norm of $o(h^{1-\epsilon})$ and of order $o(h^{2-\epsilon})$ for the $L^2(\Omega)$-norm of the states, without the use of regularization. We do not follow this approach as we would like to avoid the use of a discrete maximum principle in this thesis. In possible future research it might still be interesting to see how the arguments may transfer to the parabolic case.

Not concerned with a priori error estimates, but a posteriori error estimates, is the work of [GHHL14] where an elliptic obstacle problem is considered. Combining a posteriori estimates and adaptivity, again in the elliptic case, is also done in [MRW15, CH15].

A related topic is the discretization error analysis for optimal control problems with semilinear parabolic equations. The "problem" with those papers is, however, that the constants in those works do depend on the non-linearity. Thus applying them to a regularized problem of the form (OC_γ) is possible, but the constants would then depend on the regularization parameter γ. Therefore one would lose the nice additive structure of the error from Theorem 4.65. We shall nevertheless mention the following showcase examples: [NV12, FR11, TC13]

To the author's knowledge there is no work concerning the numerical analysis, in particular a priori error estimates, of optimal control problems with parabolic variational inequalities as constraint.

The assumptions are the same as in the end of Chapter 4. To be precise: the assumptions and properties from Section 2.2.1 are assumed to apply. Assumption 4.46, Assumption 4.47 and Assumption 4.48 are also considered valid throughout Chapter 5.

We start with considering the following regularization and discretization of (OC):

$$\begin{aligned} &\min_{(y_{\gamma kh},u)\in X^{0,1}_{k,h}\times L^2(Q)} j_v(y_{\gamma kh}) + j_T(y_{\gamma kh}(T)) + g(u) = J(y_{\gamma kh},u), \\ &\text{such that } S_{\gamma kh}(u) = y_{\gamma kh} \text{ and } u\in U_{ad}. \end{aligned} \tag{$\mathrm{SOC}_{\gamma kh}$}$$

Note, the control is not discretized, yet. The control will turn out to be discretized implicitly in time in Section 5.2.

All the appearing quantities have the properties from (OC). In addition, we assume that U_{ad} is bounded in $L^\infty(Q)$ by some constant $R_{U_{ad}} > 0$. We also define $g(u) = \frac{\alpha_g}{2}\|u\|^2_{L^2(Q)}$ with $\alpha_g > 0$. For transparency and simplicity we assume β to be of the form of Proposition 2.17, for some $\alpha_\beta > 1$. It and f are truncated in the sense of Remark 4.62. For Remark 4.62 to hold we assume $\partial_t\Psi, A\Psi, f(\Psi) \in L^\infty(Q)$.

As we want to use Theorem 4.65 we also assume $y_0 \in W^{2,\infty}(\Omega) \cap V$.

To make our solution operators differentiable, i.e. to apply Theorem 4.44, let $f(t,x,\cdot) \in C^{1,1}_{loc}(\mathbb{R})$ for any $(t,x) \in Q$. Here the Lipschitz constants of $f(t,x,\cdot)$ and $f'(t,x,\cdot)$ must not depend on $(t,x) \in Q$.

Throughout Chapter 5 $((\gamma_n, k_n, h_n))_{n\in\mathbb{N}} \subset \mathbb{R}^3_{>0}$ shall refer to a zero sequence satisfying

$$|\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2) \xrightarrow{n\to\infty} 0.$$

This is necessary for the convergence of the right hand side to 0 in Theorem 4.65.

Remark 5.1 The following results hold true if we consider a β_γ as in Proposition 2.17 for $\alpha_\beta = 1$ and assume that the boundary of the active set is of Lebesgue measure 0. This is enough to keep $S_{\gamma kh}$ differentiable, cf. Remark 4.45.

5.1.1 Existence of Global Solutions and Convergence Rates

Theorem 5.2 (SOC$_{\gamma kh}$) *has at least one solution* $(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) \in X^{0,1}_{k,h} \times U_{ad}$.

Proof. The proof is basically the same as the one in the continuous case in Theorem 3.3. For the continuity of $S_{\gamma kh}$ see Lemma 4.43. □

Theorem 5.3 *Let* $((\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n}))_{n\in\mathbb{N}}$ *be a sequence of solutions to* (SOC$_{\gamma_n k_n h_n}$). *There exists a subsequence and a global solution* $(\bar{y}, \bar{u})$ *of* (OC) *such that*

$$\begin{aligned} \bar{y}_{\gamma_{n_l},k_{n_l},h_{n_l}} &\xrightarrow{l\to\infty} \bar{y} \text{ strongly in } L^\infty(Q),\\ \bar{y}_{\gamma_{n_l},k_{n_l},h_{n_l}}(T) &\xrightarrow{l\to\infty} \bar{y}(T) \text{ strongly in } L^\infty(\Omega),\\ \bar{u}_{\gamma_{n_l},k_{n_l},h_{n_l}} &\xrightarrow{l\to\infty} \bar{u} \text{ weakly* in } L^\infty(Q) \text{ and strongly in } L^2(Q). \end{aligned}$$

Proof. By the boundedness of U_{ad} in $L^\infty(Q)$ we have that there exists a subsequence $(n_l)_{l\in\mathbb{N}}$ and a $\bar{u} \in L^\infty(Q)$ such that

$$\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}} \xrightarrow{l\to\infty} \bar{u} \text{ weakly* in } L^\infty(Q) \text{ and weakly in } L^{q_u}(Q).$$

We have to show that the states converge as well. Define $\bar{y} := S(\bar{u})$. We have

$$\begin{aligned} &\|\bar{y} - \bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|_{L^\infty(Q)}\\ &\quad \le \|S(\bar{u}) - S(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}})\|_{L^\infty(Q)} + \|S(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}) - S_{\gamma_{n_l} k_{n_l} h_{n_l}}(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}})\|_{L^\infty(Q)}. \end{aligned}$$

Theorem 4.65 now implies

$$\|S(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}) - S_{\gamma_{n_l} k_{n_l} h_{n_l}}(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}})\|_{L^\infty(Q)} \le C(\gamma_{n_l}^{\frac{1}{\alpha_\beta}} + |\ln k_{n_l}|^2 |\ln h_{n_l}|^4 (k_{n_l} + h_{n_l}^2)) R_{U_{ad}}.$$

By Theorem 2.34 we have

$$\|S(\bar{u}) - S(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}})\|_{L^\infty(Q)} \xrightarrow{l\to\infty} 0.$$

Hence, by our assumption of $(\gamma_n, k_n, h_n)_{n\in\mathbb{N}}$,

$$\|\bar{y} - \bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|_{L^\infty(Q)} \xrightarrow{l\to\infty} 0.$$

This implies that the end time values converge as well.

Let us now show that $(\bar{y}, \bar{u})$ is an optimal solution to (OC). Note that weak* convergence in $L^\infty(Q)$ implies weak convergence in $L^2(Q)$. Let $u \in U_{ad}$. Weakly lower semi-continuity of g in $L^2(Q)$ and the continuity of j_v and j_T imply

$$\begin{aligned} J(\bar{y}, \bar{u}) &= j_v(\bar{y}) + j_T(\bar{y}) + g(\bar{u}) \\ &\le \liminf_{l\to\infty} j_v(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}) + j_T(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T)) + g(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}) \\ &\le \liminf_{l\to\infty} j_v(S_{\gamma_{n_l} k_{n_l} h_{n_l}}(u)) + j_T(S_{\gamma_{n_l} k_{n_l} h_{n_l}}(u)(T)) + g(u). \end{aligned} \tag{5.1}$$

Similar arguments as before lead to

$$\|S_{\gamma_{n_l} k_{n_l} h_{n_l}}(u) - S(u)\|_{L^\infty(Q)} \xrightarrow{l\to\infty} 0, \qquad \|S_{\gamma_{n_l} k_{n_l} h_{n_l}}(u)(T) - S(u)(T)\|_{L^\infty(\Omega)} \xrightarrow{l\to\infty} 0,$$

and thus

$$J(\bar{y}, \bar{u}) \le \liminf_{l\to\infty} j_v(S_{\gamma_{n_l} k_{n_l} h_{n_l}}(u)) + j_T(S_{\gamma_{n_l} k_{n_l} h_{n_l}}(u)(T)) + g(u) = J(S(u), u).$$

This shows that $(\bar{y}, \bar{u})$ is a solution to (OC).

It remains to show the strong $L^2(Q)$ convergence of the controls. By the arguments in (5.1) we have the following chain of inequalities:

$$J(\bar{y}, \bar{u}) = \liminf_{l\to\infty} J(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}, \bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}) \le \liminf_{l\to\infty} J(S_{\gamma_{n_l} k_{n_l} h_{n_l}}(\bar{u}), \bar{u}) = J(\bar{y}, \bar{u}). \tag{5.2}$$

This implies $\liminf_{l\to\infty} g(\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}) = g(\bar{u})$. Taking a subsequence, denoted by the same indices, we have

$$\frac{\alpha_g}{2}\|\bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|^2_{L^2(Q)} \xrightarrow{l\to\infty} \frac{\alpha_g}{2}\|\bar{u}\|_{L^2(Q)}.$$

Together with the weak convergence in $L^2(Q)$ this implies strong convergence in $L^2(Q)$. □

We now continue to show a priori error estimates for the functional values. Estimates for the controls and the states are presented later.

Theorem 5.4 *Let* $(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh})$ *be a global solution to* $(\mathrm{SOC}_{\gamma kh})$ *and* $(\bar{y}, \bar{u})$ *be a global solution to* (OC). *We then have*

$$|J(\bar{y}, \bar{u}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh})| \le C\left(\gamma^{\frac{1}{\alpha_\beta}} + |\ln k|^2 |\ln h|^4 (k + h^2)\right).$$

Proof. By optimality of $(\bar{y}, \bar{u})$ we have

$$\begin{aligned} J(\bar{y}, \bar{u}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) &\le J(S(\bar{u}_{\gamma kh}), \bar{u}_{\gamma kh}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) \\ &= j_v(S(\bar{u}_{\gamma kh})) - j_v(S_{\gamma kh}(\bar{u}_{\gamma kh})) + j_T(S(\bar{u}_{\gamma kh})(T)) - j_T(S_{\gamma kh}(\bar{u}_{\gamma kh})(T)). \end{aligned} \tag{5.3}$$

The functionals j_v, j_T are Fréchet differentiable and thus locally Lipschitz continuous. By the boundedness of U_{ad} and the Lipschitz continuity of S, see Corollary 2.35, we have

$$\|S(\bar{u}_{\gamma kh})\|_{L^2(Q)\cap L^\infty(I,H)} \le C$$

with a $C > 0$ independent of (γ, k, h). We will show that

$$\|S_{\gamma kh}(\bar{u}_{\gamma kh})\|_{L^2(Q)}, \|S_{\gamma kh}(\bar{u}_{\gamma kh})(T)\|_H \leq C \text{ independent of } \bar{u}_{\gamma kh}, \gamma, k, h. \tag{5.4}$$

Admitting this for the moment we can deduce

$$\begin{aligned} J(\bar{y}, \bar{u}) - & J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) \\ & \leq C(\|S(\bar{u}_{\gamma kh}) - S_{\gamma kh}(\bar{u}_{\gamma kh})\|_{L^2(Q)} + \|S(\bar{u}_{\gamma kh})(T) - S_{\gamma kh}(\bar{u}_{\gamma kh})(T)\|_H). \end{aligned} \tag{5.5}$$

By Theorem 4.65 and the boundedness of U_{ad} we have

$$J(\bar{y}, \bar{u}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) \leq C \left(\gamma^{\frac{1}{\alpha_\beta}} + |\ln k|^2 |\ln h|^4 (k + h^2) \right).$$

The estimate

$$J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) - J(\bar{y}, \bar{u}) \leq \left(\gamma^{\frac{1}{\alpha_\beta}} + |\ln k|^2 |\ln h|^4 (k + h^2) \right).$$

is obtained analogously.

It remains to show (5.4). Testing the defining equation of $\bar{y}_{\gamma kh}$ with $\bar{y}_{\gamma kh}$ itself yields

$$B_k(\bar{y}_{\gamma kh}, \bar{y}_{\gamma kh}) + (f(\bar{y}_{\gamma kh}) + \beta_\gamma(\bar{y}_{\gamma kh} - \Psi), \bar{y}_{\gamma kh})_{L^2(Q)} = (u, \bar{y}_{\gamma kh})_{L^2(Q)} + (y_0, \bar{y}_{\gamma kh}(0^+))_H.$$

By the fact that β_γ a nnd f are truncated in the sense of Remark 4.62 we can conclude

$$B_k(\bar{y}_{\gamma kh}, \bar{y}_{\gamma kh}) - 2C_\infty \|\bar{y}_{\gamma kh}\|_{L^2(I,V)} \leq \|u\|_{L^\infty(Q)} \|\bar{y}_{\gamma kh}\|_{L^2(I,V)} + \|y_0\|_H \|\bar{y}_{\gamma kh}(0^+)\|_H.$$

By the positive definiteness of Lemma 4.40, the Poincaré inequality, e.g. [Eva98, Theorem 5.6.3], and Young's inequality we find

$$\begin{aligned} & c\|\bar{y}_{\gamma kh}\|^2_{L^2(I,V)} + \|\bar{y}_{\gamma kh}(0^+)\|^2_H + \|\bar{y}_{\gamma kh}(T)\|^2_H \\ & \quad \leq \frac{4}{c} C_\infty^2 + \frac{c}{4} \|\bar{y}_{\gamma kh}\|^2_{L^2(I,V)} + \frac{1}{c} \|u\|^2_{L^\infty(Q)} + \frac{c}{4} \|\bar{y}_{\gamma kh}\|^2_{L^2(I,V)} + \frac{1}{2} \|y_0\|^2_H + \frac{1}{2} \|\bar{y}_{\gamma kh}(0^+)\|^2_H. \end{aligned}$$

Ordering terms results in

$$\|\bar{y}_{\gamma kh}\|^2_{L^2(I,V)} + \|\bar{y}_{\gamma kh}(0^+)\|^2_H + \|\bar{y}_{\gamma kh}(T)\|^2_H \leq C \left(C_\infty^2 + \|u\|^2_{L^\infty(Q)} + \|y_0\|^2_H \right).$$

By the boundedness of U_{ad} the right hand side is bounded independently of $\bar{u}_{\gamma kh}$ and the other quantities. □

5.1.2 L^2-Convergence Rates for Local Solutions

So far we have not paid attention to local solutions of (OC). We rectify this now, as not all solution algorithms applicable to ($\text{SOC}_{\gamma kh}$) or, later, ($\text{FOC}_{\gamma kh}$) provide global solutions. As the finite dimensional problem ($\text{FOC}_{\gamma kh}$) is accessible to computation we consider local solutions as something of note.

Lemma 5.5 *Let $(\bar{y}, \bar{u})$ be a strict local minimum to* (OC)*. That means there exists a $r > 0$ such that*

$$J(\bar{y}, \bar{u}) < J(S(u), u) \qquad \forall u \in \bar{B}_r(\bar{u}).$$

Here $\bar{B}_r(\bar{u})$ refers to the closed unit ball in $L^2(Q)$.

Then there is a sequence of local solutions of ($\text{SOC}_{\gamma kh}$) *converging to $(\bar{y}, \bar{u})$ in the sense of Theorem 5.3. They are locally optimal on balls closed in $L^2(Q)$ with the radii $(r_n)_{n\in\mathbb{N}}$, which satisfy $r_n \xrightarrow{n\to\infty} r$.*

Proof. Let $(\bar{y}, \bar{u})$ be as above and $\gamma, k, h > 0$. We consider the auxiliary problem

$$\min_{(y,u)\in X^{0,1}_{k,h}\times L^2(Q)} J(y,u), \qquad \text{such that } S_{\gamma kh}(u) = y \text{ and } u \in U_{ad} \cap \bar{B}_r(\bar{u}). \tag{AUX}$$

Theorem 5.2 is also applicable to (AUX) as $U_{ad} \cap \bar{B}_r(\bar{u})$ is convex, closed in $L^{q_u}(Q)$ and bounded in $L^\infty(Q)$. Thus there exists a solution $(\bar{y}^r_{\gamma kh}, \bar{u}^r_{\gamma kh})$ to (AUX). Just as in the proof of Theorem 5.3 there is a subsequence of $(\gamma_n, k_n, h_n)_{n\in\mathbb{N}}$, denoted by the same indices, such that $(\bar{y}^r_{\gamma_n k_n h_n}, \bar{u}^r_{\gamma_n k_n h_n})_{n\in\mathbb{N}}$ converges to some $(\hat{y}, \hat{u})$. It convergences strongly in $L^\infty(Q)$ in the states, strongly in $L^\infty(\Omega)$ in the end times, weakly* in $L^\infty(Q)$ and weakly in $L^2(Q)$ in the controls. We have $\hat{u} \in \bar{B}_r(\bar{u})$, because $\bar{B}_r(\bar{u})$ is weakly closed, and see that

$$J(\hat{y}, \hat{u}) \le \liminf_{n\to\infty} J(y_{\gamma_n k_n h_n}, u_{\gamma_n k_n h_n}) \le \liminf_{n\to\infty} J(S_{\gamma_n k_n h_n}(\bar{u}), \bar{u}) = J(\bar{y}, \bar{u}).$$

By the strict optimality of $(\bar{y}, \bar{u})$ this implies $(\hat{y}, \hat{u}) = (\bar{y}, \bar{u})$. This line also entails, as in the proof of Theorem 5.3, see (5.2), $g(\bar{u}^r_{\gamma_n k_n h_n}) \xrightarrow{n\to\infty} g(u)$ and thus the strong $L^2(Q)$ convergence of the controls.

Therefore we have for n large enough $\bar{u}^r_{\gamma_n k_n h_n} \in B_r(\bar{u}) \subsetneq \bar{B}_r(\bar{u})$. Thus choosing $r_n := r - \|\bar{u}^r_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} > 0$, for n large enough, we have $\bar{B}_{r_n}(\bar{u}^r_{\gamma_n k_n h_n}) \subset \bar{B}_\delta(\bar{u})$. This entails

$$J(\bar{y}^r_{\gamma_n k_n h_n}, \bar{u}^r_{\gamma_n k_n h_n}) = \min_{u\in U_{ad}\cap \bar{B}_r(\bar{u})} J(S_{\gamma_n k_n h_n}(u), u) \le \min_{u\in U_{ad}\cap \bar{B}_{r_n}(\bar{u}^r_{\gamma_n k_n h_n})} J(S_{\gamma_n k_n h_n}(u), u).$$

Thus for n large enough the $(\bar{y}^r_{\gamma_n k_n h_n}, \bar{u}^r_{\gamma_n k_n h_n})$ are indeed local solutions to $(\mathrm{SOC}_{\gamma kh})$. □

Theorem 5.6 *Let $(\bar{y}, \bar{u})$ be a strict local minimum of* (OC) *in the sense of Lemma 5.5. Let $((\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n}))_{n\in\mathbb{N}}$ be the sequence of local solutions to* $(\mathrm{SOC}_{\gamma kh})$ *that converges to $(\bar{y}, \bar{u})$ by Lemma 5.5. Then we have*

$$|J(\bar{y}, \bar{u}) - J(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n})| \le C\left(\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2)\right)$$

for $n \in \mathbb{N}$ sufficiently large.

We slightly abuse notation here. The fixed sequence $(\gamma_n, k_n, h_n)_{n\in\mathbb{N}}$ would require us to use subsequences in Theorem 5.6, i.e. $((\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}, \bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}))_{l\in\mathbb{N}}$. This would however lead to a massive influx in the use of indices, which we avoid here and in similar situations.

Proof. The proof is almost the same as for Theorem 5.4 applied to the local auxiliary problems (AUX).

By Lemma 5.5 we have $\|\bar{u}_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} < r$ for n sufficiently large and therefore $J(\bar{y}, \bar{u}) < J(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n})$. Thus

$$|J(\bar{y}, \bar{u}) - J(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n})| = J(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n}) - J(\bar{y}, \bar{u}).$$

Using the $(r_n)_{n\in\mathbb{N}}$ from Lemma 5.5, it follows for n sufficiently large, that $r_n > \frac{r}{2}$ and $\|\bar{u}_{\gamma kh} - \bar{u}\|_{L^2(Q)} < \frac{r}{2}$ so that we have $\bar{u} \in B_{r_n}(\bar{u}_{\gamma_n k_n h_n})$ and by the local optimality of $(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n})$ that

$$J(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n}) - J(\bar{y}, \bar{u}) \le J(S_{\gamma_n k_n h_n}(\bar{u}), \bar{u}) - J(S(\bar{u}), \bar{u}).$$

This is now estimated as the terms in (5.3) from the proof of Theorem 5.4. □

Theorem 5.7 *Let $(\bar{y}, \bar{u})$ be a local solution to* (OC) *such that a local quadratic growth condition of the form of Theorem 3.57 holds, i.e. there are $r, \delta > 0$ such that*

$$\|u - \bar{u}\|_{L^2(\Omega)} < r \implies J(S(u), u) \geq J(\bar{y}, \bar{u}) + \delta \|u - \bar{u}\|^2_{L^2(\Omega)}.$$

Let $((\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n}))_{n \in \mathbb{N}}$ be the sequence of local solutions to $(\mathrm{SOC}_{\gamma kh})$ *that converges to $(\bar{y}, \bar{u})$ by Lemma 5.5. Then, for n so large that $\|\bar{u}_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} < r$, we obtain*

$$\|\bar{u}_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} \leq C \sqrt{\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2)},$$
$$\|\bar{y}_{\gamma_n k_n h_n} - \bar{y}\|_{L^\infty(I,H)} \leq C \sqrt{\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2)}.$$

Proof. The local quadratic growth condition obviously entails that $(\bar{y}, \bar{u})$ is a strict local minimum. Thus Lemma 5.5 actually applies. The quadratic growth condition implies, for n so large such that $\|\bar{u}_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} < r$, that

$$\begin{aligned}
&\delta \|\bar{u}_{\gamma_n k_n h_n} - \bar{u}\|^2_{L^2(Q)} \\
&\leq J(S(\bar{u}_{\gamma_n k_n h_n}), \bar{u}_{\gamma_n k_n h_n}) - J(\bar{y}, \bar{u}) \\
&= J(S(\bar{u}_{\gamma_n k_n h_n}), \bar{u}_{\gamma_n k_n h_n}) - J(S_{\gamma_n k_n h_n}(\bar{u}_{\gamma_n k_n h_n}), \bar{u}_{\gamma_n k_n h_n}) + J(\bar{y}_{\gamma_n k_n h_n}, \bar{u}_{\gamma_n k_n h_n}) - J(\bar{y}, \bar{u}) \\
&\leq j_v(S(\bar{u}_{\gamma_n k_n h_n})) - j_v(S_{\gamma_n k_n h_n}(\bar{u}_{\gamma_n k_n h_n})) + j_T(S(\bar{u}_{\gamma_n k_n h_n})(T)) - j_T(S_{\gamma_n k_n h_n}(\bar{u}_{\gamma_n k_n h_n})(T)) \\
&\quad + C(\gamma_n + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2)).
\end{aligned}$$

Here we used Theorem 5.6 in the last inequality. The last two terms are estimated as in the proof of Theorem 5.4. We thus obtain

$$\|\bar{u}_{\gamma_n k_n h_n} - \bar{u}\|^2_{L^2(Q)} \leq C \left(\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2) \right).$$

Taking the root yields the desired estimate for the controls.

To get the estimate for the states we split the error:

$$\begin{aligned}
&\|\bar{y}_{\gamma_n k_n h_n} - \bar{y}\|_{L^\infty(I,H)} \\
&\qquad \leq \|S_{\gamma_n k_n h_n}(\bar{u}_{\gamma_n k_n h_n}) - S(\bar{u}_{\gamma_n k_n h_n})\|_{L^\infty(I,H)} + \|S(\bar{u}_{\gamma_n k_n h_n}) - S(\bar{u})\|_{L^\infty(I,H)}.
\end{aligned}$$

The first term converges with the stated rate by Theorem 4.65. The second term is a consequence of the estimates for the controls and the Lipschitz continuity of S, cf. Corollary 2.35. □

Remark 5.8 We compare these rates to the ones obtained in [NV12]. The article is concerned with the optimal control of semilinear equations under state constraints. Our regularized optimization problem (OC_γ) fall essentially into this category, as we could simply add empty state constraints. [NV12, Theorem 5.1] establishes a convergence rate of the controls of order $\mathcal{O}(k)$. Note, however, that in their semi-discrete analysis the space was not discretized. Essentially the authors obtain twice as good an order as we have. Why is that? We had to take great precautions to make the constant in front of the terms in Theorem 5.7 independent of γ, i.e. independent of the non-linearity in the equation. Thus citing [NV12] directly would result in convergence rates where the constants would depend on the regularization parameter.

For the fully discretized problem [NV12, Proposition 5.1] obtains a rate of $\mathcal{O}(k + h^{\frac{3}{2}-\epsilon}$ for arbitrary $\epsilon > 0$ and piecewise bilinear elements. We will obtain a worse rate in Theorem 5.28, while having the benefit that our rate does explicitly not depend on the regularization parameter.

It would be of great interest to see if the techniques of [NV12] or related work could be used to obtain better convergence rates in k and h, under certain assumptions at least, and keep them independent of the regularization parameter.

5.1.3 Optimality Conditions

Theorem 5.9 *Each locally optimal control $\bar{u}_{\gamma kh}$ of the reduced problem*

$$\min_{u \in U_{ad}} J(S_{\gamma kh}(u), u) = J(S_{\gamma kh}(\bar{u}_{\gamma kh}), \bar{u}_{\gamma kh})$$

satisfies for any $u \in U_{ad}$ with $\bar{y}_{\gamma kh} := S_{\gamma kh}(\bar{u}_{\gamma kh})$

$$\begin{aligned} 0 \le & \Big(j_v'(\bar{y}_{\gamma kh}), S_{\gamma kh}'(\bar{u}_{\gamma kh})(u - \bar{u}_{\gamma kh}) \Big)_{L^2(Q)} \\ & + \Big(j_T'(\bar{y}_{\gamma kh}(T)), [S_{\gamma kh}'(\bar{u}_{\gamma kh})(u - \bar{u}_{\gamma kh})](T) \Big)_H \\ & + \alpha_g \left(\bar{u}_{\gamma kh}, u - \bar{u}_{\gamma kh} \right)_{L^2(Q)} . \end{aligned}$$

Here "locally" is to be understood in the $L^2(Q)$-sense, see Lemma 5.5.

Proof. This is an immediate consequence of Theorem 4.44 and the usual optimality conditions for control constrained problems, e.g. the proof of [Trö09, Lemma 2.21]. □

Lemma 5.10 *Given $u \in L^2(Q)$ we define $p_{\gamma kh}(u) \in X_{k,h}^{0,1}$ as the solution to*

$$\begin{aligned} & B_k(\varphi_{kh}, p_{\gamma kh}(u)) + ([\beta_\gamma{}'(S_{\gamma kh}(u) - \Psi) + f'(S_{\gamma kh}(u))] p_{\gamma kh}(u), \varphi_{kh})_{L^2(Q)} \\ & = (j_T'(S_{\gamma kh}(u)(T)), \varphi_{kh}(T))_H + (j_v'(S_{\gamma kh}(u)), \varphi_{kh})_{L^2(Q)} \qquad \forall \varphi_{kh} \in X_{k,h}^{0,1}. \end{aligned}$$

Then for all $u, v \in L^2(Q)$ we have

$$\begin{aligned} & \Big(j_T'(S_{\gamma kh}(u)(T)), [S_{\gamma kh}'(u)v](T) \Big)_H + \Big(j_v'(S_{\gamma kh}(u)), S_{\gamma kh}'(u)v \Big)_{L^2(Q)} \\ & = (p_{\gamma kh}(u), v)_{L^2(Q)} . \end{aligned}$$

Proof. Let $u, v \in L^2(Q)$. The definition of $p_{\gamma kh}(u)$ is tested with $z_{\gamma kh} := S_{\gamma kh}'(u)v$ to obtain

$$\begin{aligned} & (j_T'(S_{\gamma kh}(u)(T)), z_{\gamma kh}(T))_H + (j_v'(S_{\gamma kh}(u)), z_{\gamma kh})_{L^2(Q)} \\ & = B_k(z_{\gamma kh}, p_{\gamma kh}(u)) + \left([\beta_\gamma{}'(S_{\gamma kh}(u) - \Psi) + f'(S_{\gamma kh}(u))]\, p_{\gamma kh}(u), z_{\gamma kh} \right)_{L^2(Q)} . \end{aligned}$$

Using the definition of $z_{\gamma kh} = S_{\gamma kh}'(u)v$ by Theorem 4.44 yields that the last line is equal to

$$(p_{\gamma kh}(u), v)_{L^2(Q)} .$$

□

Combining Theorem 5.9 and Lemma 5.10 immediately yields

Corollary 5.11 *Each locally optimal control $\bar{u}_{\gamma kh}$ of the reduced problem*

$$\min_{u \in U_{ad}} J(S_{\gamma kh}(u), u) = J(S_{\gamma kh}(\bar{u}_{\gamma kh}), \bar{u}_{\gamma kh})$$

satisfies

$$(p_{\gamma kh}(\bar{u}_{\gamma kh}) + \alpha_g \bar{u}_{\gamma kh}, u - \bar{u}_{\gamma kh})_{L^2(Q)} \geq 0 \qquad \forall u \in U_{ad}.$$

Here "locally" has to be understood in the $L^2(Q)$-sense.

Definition 5.12 Given an optimal solution $(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh})$ to $(\mathrm{SOC}_{\gamma kh})$ we define $\bar{p}_{\gamma kh} := p_{\gamma kh}(\bar{u}_{\gamma kh})$. We call $(\bar{u}_{\gamma kh}, \bar{y}_{\gamma kh}, \bar{p}_{\gamma kh})$ an optimal triple.

5.1.4 Convergence of Adjoints and Multipliers

We now discuss how taking the limit $(\gamma, k, h) \to 0$ relates to the analogous discussion for $\gamma \to 0$ in Section 3.4.

Lemma 5.13 *The Ritz projection $R_h : V \to V_h$ satisfies*

$$\|R_h u - u\|_H \leq C\|u\|_V h \qquad \forall u \in V.$$

The constant $C > 0$ does not depend on h or u.

Proof. This is a standard result and is implied by [BS08, Theorem (5.7.6)]. The prerequisites are clearly satisfied because we have $V = H_0^1(\Omega)$, with Ω being smooth, and the results from Section 4.1.3. □

We now discuss how the adjoints and the multipliers converge. Throughout Section 5.1.4 let $(\bar{u}_{\gamma kh}, \bar{y}_{\gamma kh}, \bar{p}_{\gamma kh})$ be an globally optimal triple to $(\mathrm{SOC}_{\gamma kh})$ for $(\gamma, k, h) > 0$. We start with some abstract, preparatory propositions.

Proposition 5.14 *For $n \in \mathbb{N}$ let $v_n \in X^{0,1}_{k_n,h_n}$, here (γ_n, k_n, h_n) is the chosen sequence from the beginning of this chapter. The sequence $(v_n)_{n\in\mathbb{N}}$ shall converge to a $v \in C(\bar{I}, H) \cap L^2(I, V)$ weakly in $L^2(I, V)$ and strongly in $L^\infty(I, H)$. Then we have*

$$B_k(v_n, I_{k_n} R_{h_n} \varphi) \xrightarrow{n\to\infty} a_I(v, \varphi) + (v(T), \varphi(T))_H - (v, \partial_t \varphi)_{L^2(I,H)} \qquad \forall \varphi \in H^1(I, V).$$

Proof. Let $\varphi \in H^1(I, V)$ be fixed. We first note that

$$\begin{aligned}
\|I_k\varphi - \varphi\|^2_{L^2(I,V)} &= \sum_{m=1}^{M} \|\varphi(t_m) - \varphi(\cdot)\|^2_{L^2(I_m,V)} = \sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} \left\| \int_t^{t_m} \partial_t \varphi(s)\, ds \right\|_V^2 dt \\
&\leq \sum_{m=1}^{M} \int_{t_{m-1}}^{t_m} \left(\int_t^{t_m} \|\partial_t\varphi(s)\|_V \, ds \right)^2 dt \\
&\leq \sum_{m=1}^{M} k_m \left(\int_{t_{m-1}}^{t_m} \|\partial_t\varphi(s)\|_V \, ds \right)^2 \\
&\leq \sum_{m=1}^{M} k_m^2 \int_{t_{m-1}}^{t_m} \|\partial_t\varphi(s)\|_V^2 \, ds\, dt \leq k^2 \|\partial_t\varphi\|^2_{L^2(I,V)}.
\end{aligned}$$

Thus $I_k\varphi \xrightarrow{k\to 0} \varphi$ in $L^2(I,V)$. By $a_I(v_n, I_{k_n}R_{h_n}\varphi) = a_I(v_n, R_{h_n}I_{k_n}\varphi) = a_I(v_n, I_{k_n}\varphi)$ it follows that

$$\begin{aligned}|a_I(v_n, I_{k_n}R_{h_n}\varphi) - a_I(v,\varphi)| &\le |a_I(v_n, I_{k_n}\varphi - \varphi)| + |a_I(v_n - v,\varphi)| \\ &\le C\|v_n\|_{L^2(I,V)}\|I_{k_n}\varphi - \varphi\|_{L^2(I,V)} + |a_I(v_n - v,\varphi)|.\end{aligned}$$

By the weak convergence of $(v_n)_{n\in\mathbb{N}}$ in $L^2(I,V)$ and the strong convergence of $(I_{k_n}\varphi)_{n\in\mathbb{N}}$ in $L^2(I,V)$ we find

$$|a_I(v_n, I_{k_n}R_{h_n}\varphi) - a_I(v,\varphi)| \xrightarrow{n\to\infty} 0. \tag{5.6}$$

Furthermore, we have

$$\begin{aligned}&-\sum_{m=1}^{M-1}\left(v_n(t_m^-), [I_{k_n}R_{h_n}\varphi]_m\right)_H + (v_n(T), I_{k_n}R_{h_n}\varphi(T))_H \\ &= -\sum_{m=1}^{M-1}(v(t_m^-), I_{k_n}R_{h_n}\varphi(t_m^+) - I_{k_n}R_{h_n}\varphi(t_m^-))_H + (v_n(T), I_{k_n}R_{h_n}\varphi(T))_H \\ &= -\sum_{m=1}^{M-1}(v(t_m^-), R_{h_n}\varphi(t_{m+1}) - R_{h_n}\varphi(t_m))_H + (v_n(T), R_{h_n}\varphi(T))_H.\end{aligned}$$

Lemma 5.13 gives us

$$\|R_{h_n}\varphi(T) - \varphi(T)\|_H \le Ch_n\|\varphi(T)\|_V \le Ch_n\|\varphi\|_{H^1(I,V)} \xrightarrow{n\to\infty} 0. \tag{5.7}$$

We now treat the jump terms:

$$\begin{aligned}&\left|\sum_{m=1}^{M-1}(v(t_m^-), R_{h_n}\varphi(t_{m+1}) - R_{h_n}\varphi(t_m))_H - (v,\partial_t\varphi)_{L^2((0,t_1),H)}\right| \\ &= \left|\sum_{m=1}^{M-1}\int_{t_m}^{t_{m+1}}(v_n(s), \partial_t(R_{h_n}\varphi)(s))_H - (v(s),\partial_t\varphi(s))_H\,ds\right| \\ &\le \sum_{m=1}^{M-1}\int_{t_m}^{t_{m+1}} |(v_n(s) - v(s), \partial_t(R_{h_n}\varphi)(s))_H| + |(v(s), \partial_t(R_{h_n}\varphi)(s) - \partial_t\varphi(s))_H|\,ds \\ &\le \|v_n - v\|_{L^\infty(I,H)}\|\partial_t(R_h\varphi)\|_{L^1(I,H)} + \|v\|_{L^\infty(I,H)}\|\partial_t(R_{h_n}\varphi) - \partial_t\varphi\|_{L^1(I,H)}.\end{aligned}$$

By Proposition 8.12 we have almost everywhere in I that $\partial_t R_h\varphi = R_h\partial_t\varphi$. Lemma 5.13 implies almost everywhere in I that

$$\|\partial_t R_{h_n}\varphi - \partial_t\varphi\|_H = \|R_{h_n}\partial_t\varphi - \partial_t\varphi\|_H \le Ch_n\|\partial_t\varphi\|_V.$$

Thus we have shown

$$\begin{aligned}&\left|\sum_{m=1}^{M-1}(v_n(t_m^-), R_{h_n}\varphi(t_m^+) - R_{h_n}\varphi(t_{m-1}^+))_H - (v,\partial_t\varphi)_{L^2((t_1,T),H)}\right| \\ &\le C\left(\|v_n - v\|_{L^\infty(I,H)}\|\partial_t(R_h\varphi)\|_{L^1(I,H)} + h\|\partial_t\varphi\|_V\right).\end{aligned}$$

Using Proposition 8.12 again and recalling that φ is fixed we deduce:

$$\begin{aligned}&\left|\sum_{m=1}^{M-1}(v_n(t_m^-), R_{h_n}\varphi(t_m^+) - R_{h_n}\varphi(t_{m-1}^+))_H - (v,\partial_t\varphi)_{L^2((t_1,T),H)}\right| \\ &\le C\left(\|v_n - v\|_{L^\infty(I,H)} + h\right).\end{aligned} \tag{5.8}$$

It remains to show that

$$|(v,\partial_t\varphi)_{L^2((0,t_1),H)}| \xrightarrow{n\to\infty} 0. \tag{5.9}$$

This follows from the theorem of dominated convergence, e.g. [BK15, Proposition 5.4]:

$$|(v,\partial_t\varphi)_{L^2((0,t_1),H)}| \le \int_0^T \|v(s)\|_H \|\partial_t\varphi(s)\|_H \cdot 1_{(0,t_1)}\, ds \le \|v\|_{L^2(I,H)}\|\varphi\|_{L^2(I,H)}$$

which is independent of t_1. Thus the theorem of dominated convergence is applicable and yields

$$\lim_{n\to\infty}\int_0^T \|v(s)\|_H \|\partial_t\varphi(s)\|_H \cdot 1_{(0,t_1)}\, ds = \int_0^T \lim_{n\to\infty} \|v(s)\|_H \|\partial_t\varphi(s)\|_H \cdot 1_{(0,t_1)}\, ds = 0$$

as $t_1 \to T$ for $n\to\infty$. Hence, this implies (5.9). Combining (5.6), (5.7), (5.8) and (5.9) yields

$$|B_k(v_n, I_{k_n}R_{h_n}\varphi) - a_I(v,\varphi) - (v(T),\varphi(T))_H - (v,\partial_t\varphi)_{L^2(I,H)}| \xrightarrow{n\to\infty} 0.$$

□

Proposition 5.15 *For $n\in\mathbb{N}$ let $v_n \in X^{0,1}_{k_n,h_n}$. The sequence $(v_n)_{n\in\mathbb{N}}$ shall converge to a $v\in L^2(I,V)$ weakly in $L^2(I,V)$. Let $I_k\colon H^1(I,V)\to L^\infty(I,V)$ be the nodal interpolant in time and $R_h\colon V\to V_h$ the usual Ritz projection along a_Ω. Then we have*

$$B_k(I_kR_h\varphi, v_n) \xrightarrow{n\to\infty} a_I(v,\varphi) + (\partial_t\varphi, v)_{L^2(I,H)} \qquad \forall\varphi\in H^1(I,V)\cap W_0(I).$$

Proof. Let $\varphi\in H^1(I,V)\cap W_0(I)$ be fixed.

As in the proof of (5.6) in the previous Proposition 5.14 we find

$$a_I(I_{k_n}R_{h_n}\varphi, v_n) = a_I(v_n, I_{k_n}\varphi) \xrightarrow{n\to\infty} a_I(v,\varphi).$$

Furthermore, it holds

$$\begin{aligned}\sum_{m=2}^{M}\Big([I_kR_h\varphi]_{m-1}, v_n(t^+_{m-1})\Big)_H &= \sum_{m=2}^{M}\Big(I_kR_h\varphi(t^+_{m-1}) - I_kR_h\varphi(t^-_{m-1}), v_n(t^+_{m-1})\Big)_H \\ &= \sum_{m=2}^{M}\Big(R_h\varphi(t_m) - R_h\varphi(t_{m-1}), v_n(t^+_{m-1})\Big)_H .\end{aligned}$$

We now treat the jump terms once again:

$$\begin{aligned}&\sum_{m=2}^{M}\Big(I_kR_h\varphi(t^+_{m-1}) - I_kR_h\varphi(t^-_{m-1}), v_n(t^+_{m-1})\Big)_H - (\partial_t\varphi, v)_{L^2((0,t_{M-1}),H)} \\ &= \sum_{m=2}^{M}\int_{t_{m-1}}^{t_m}\Big(\partial_t(R_{h_n}\varphi)(s), v_n(t^+_{m-1})\Big)_H - (\partial_t\varphi(s), v(s))_H\, ds \\ &= \sum_{m=2}^{M}\int_{t_{m-1}}^{t_m}\Big(\partial_t(R_{h_n}\varphi)(s), v_n(t^+_{m-1}) - v(s)\Big)_H + (\partial_t\varphi(s) - \partial_t(R_{h_n}\varphi)(s), v(s))_H\, ds \\ &= \int_Q \partial_t(R_{h_n}\varphi)1_{(t_1,T)}(v_n - v)\, d(t,x) + \sum_{m=2}^{M}\int_{t_{m-1}}^{t_m}(\partial_t\varphi(s) - \partial_t(R_{h_n}\varphi)(s), v(s))_H\, ds.\end{aligned} \tag{5.10}$$

The first term converges to 0, because $R_{h_n}(\partial_t\varphi)1_{(t_1,T)}$ converges strongly to $\partial_t\varphi$ in $L^2(Q)$ by Lemma 5.13 and $v_n \xrightarrow{n\to\infty} v$ weakly in $L^2(I,V) \supset L^2(Q)$.

By Proposition 8.12 we have almost everywhere in I that $\partial_t R_h\varphi = R_h\partial_t\varphi$. Thus using Lemma 5.13 again implies almost everywhere in I:

$$\|\partial_t R_{h_n}\varphi - \partial_t\varphi\|_H = \|R_{h_n}\partial_t\varphi - \partial_t\varphi\|_H \leq Ch_n\|\partial_t\varphi\|_V.$$

Thus the second term in (5.10) converges to 0 as well, cf. the arguments in Proposition 5.14.

The remaining part $(\partial_t\varphi, v)_{L^2((t_{M-1},T),H)}$ converges to 0, which can be proven by the theorem of dominated convergence as in the proof of (5.9). □

Proposition 5.16 *Let $\varphi \in C(\bar{I},V)\cap H^1(I,H)$. Then we have*

$$\|I_kR_h\varphi - \varphi\|_{L^2(Q)} \leq C(h\|\varphi\|_{L^\infty(I,V)} + k\|\partial_t\varphi\|_{L^2(Q)}),$$
$$\|I_kR_h\varphi(T) - \varphi(T)\|_H \leq Ch\|\varphi\|_{L^\infty(I,V)}.$$

Proof. By the triangle inequality and Hölder's inequality we immediately find

$$\begin{aligned}\|I_kR_h\varphi - \varphi\|_{L^2(Q)} &\leq \|I_k(R_h\varphi-\varphi)\|_{L^2(Q)} + \|I_k\varphi - \varphi\|_{L^2(Q)}\\ &\leq \|I_k(R_h\varphi-\varphi)\|_{L^\infty(I,H)} + \left(\sum_{j=1}^M \int_{I_j}\|\varphi(t_j)-\varphi(t)\|_H^2\,dt\right)^{\frac{1}{2}}.\end{aligned}$$

By the stability of the nodal interpolant and a standard estimate for the Ritz projection, cf. Lemma 5.13, we bound this from above by

$$\begin{aligned}\|I_kR_h\varphi - \varphi\|_{L^2(Q)} &\leq \|R_h\varphi-\varphi\|_{L^\infty(I,H)} + \left(\sum_{j=1}^M \int_{I_j}\left(\int_t^{t_j}\|\partial_t\varphi(s)\|_H\,ds\right)^2 dt\right)^{\frac{1}{2}}\\ &\leq Ch\|\varphi\|_{L^\infty(I,V)} + k^{\frac{1}{2}}\left(\sum_{j=1}^M\left(\int_{I_j}\|\partial_t\varphi(s)\|_H\,ds\right)^2\right)^{\frac{1}{2}}\\ &\leq Ch\|\varphi\|_{L^\infty(I,V)} + k\left(\sum_{j=1}^M\int_{I_j}\|\partial_t\varphi(s)\|_H^2\,ds\right)^{\frac{1}{2}}.\end{aligned}$$

This shows the first estimate. The second estimate follows readily from Lemma 5.13:

$$\|I_kR_h\varphi(T) - \varphi(T)\|_H = \|R_h\varphi(T) - \varphi(T)\|_H \leq Ch\|\varphi(T)\|_V.$$

□

Theorem 5.17 *There exists a subsequence of $(\gamma_n,k_n,h_n)_{n\in\mathbb{N}}$ and a 5-tuple*

$$(\bar{u},\bar{y},\bar{\lambda},\bar{p},\bar{\eta}) \in U_{ad}\times \cap \bigcap_{p\in[q_u,\infty)} W_p^{1,2}(Q)\times L^\infty(Q)\times L^2(I,V)\times W_0(I)^*$$

such that

- *$(\bar{u},\bar{y})$ is optimal for* (OC) *and a sequence of optimal solutions $(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}}, \bar{u}_{\gamma_{n_l}k_{n_l}h_{n_l}})_{l\in\mathbb{N}}$ to the discretized, regularized problems converges to $(\bar{y},\bar{u})$ in the sense of Theorem 5.3,*
- *$\bar{\lambda}_{\gamma_{n_l}k_{n_l}h_{n_l}} := -\beta_\gamma(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}} - \Psi) \xrightarrow{l\to\infty} \bar{\lambda}$ weakly* in $L^\infty(Q)$ such that $(\bar{u},\bar{y},\bar{\lambda})$ satisfy the conditions from Theorem 2.33,*

- $\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}} \xrightarrow{l\to\infty} \bar{p}$ *weakly in* $L^2(I,V)$,
- $\bar{\eta}_{\gamma_{n_l} k_{n_l} h_{n_l}} := -R_{h_n}^* I_{k_n}^* (\beta_{\gamma_{n_l}}(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}} - \Psi)\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}) \xrightarrow{l\to\infty} \bar{\eta}$ *weakly in* $(H^1(I,V) \cap W_0(I))^*$. *Here* $R_{h_n}^*$ *is the adjoint operator taken with respect to the* H*-norm and* $I_{k_n}^*$ *is the adjoint operator taken with respect to the* $L^2(Q)$*-norm.*

The pair $(\bar{p}, \bar{\eta})$ *satisfies for any* $\varphi \in W_0(I)$:

$$
\begin{aligned}
&(\bar{p}, \partial_t \varphi)_{L^2(I,V,V^*)} + a_I(\bar{p}, \varphi) + (f'(\bar{y})\bar{p}, \varphi)_{L^2(Q)} \\
=&(\bar{\eta}, \varphi)_{W_0(I)^*, W_0(I)} + (j_v'(\bar{y}), \varphi)_{L^2(Q)} + (j_T'(\bar{y}(T)), \varphi(T))_H,
\end{aligned}
$$

Proof. The first claim follows immediately from Theorem 5.3, while the regularity of $\bar{y}$ follows from Lemma 2.38.

Since by construction, i.e. being truncated as in Remark 4.62, the $\bar{\lambda}_{\gamma_{n_l} k_{n_l} h_{n_l}}$ are bounded uniformly in $L^\infty(Q)$, we have a subsequence converging weakly* to some $\bar{\lambda}$ in $L^\infty(Q)$.

Let us show that $(\bar{u}, \bar{y}, \bar{\lambda})$ satisfy the conditions from Theorem 2.33. For $v \in C_c^\infty(Q)$ we have by Proposition 5.14 and Proposition 5.16 and the already established convergence behaviour that the following equality

$$
\begin{aligned}
B_k(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}, I_{k_n} R_{h_n} v) &+ \Big(f(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}), I_{k_n} R_{h_n} v\Big)_{L^2(Q)} \\
&= \Big(\bar{\lambda}_{\gamma_{n_l} k_{n_l} h_{n_l}} + \bar{u}_{\gamma_{n_l} k_{n_l} h_{n_l}}, I_{k_n} R_{h_n} v\Big)_{L^2(Q)}
\end{aligned}
$$

converges to

$$
-(\bar{y}, \partial_t v)_{L^2(Q)} + a_I(\bar{y}, v) + (f(\bar{y}), v)_{L^2(Q)} = (\bar{\lambda} + \bar{u}, v)_{L^2(Q)}.
$$

By the high regularity of $\bar{y}$ we can use partial integration in space and time to arrive at

$$
(\partial_t \bar{y} + A\bar{y} + f(\bar{y}), v)_{L^2(Q)} = (\bar{\lambda} + \bar{u}, v)_{L^2(Q)}.
$$

By the density of $C_c^\infty(Q)$ in $L^2(Q)$, we conclude that the first condition of $\bar{\lambda}$ of Theorem 2.33 is satisfied. To see the complementarity condition note that we already have, by the monotonicity of β_γ, that $\bar{\lambda} \geq 0$. Thus by the definition of $\bar{\lambda}_{\gamma_{n_l} k_{n_l} h_{n_l}}$ and β_γ there holds

$$
\begin{aligned}
0 \leq (\bar{\lambda}, \bar{y} - \Psi)_{L^2(Q)} &= \lim_{l\to\infty} (\bar{\lambda}_{\gamma_{n_l} k_{n_l} h_{n_l}}, -(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}} - \Psi)_-)_{L^2(Q)} \\
&\leq \limsup_{l\to\infty} C_\infty \|(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}} - \Psi)_-\|_{L^1(Q)} = 0.
\end{aligned}
$$

Testing the defining equation for $\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}$ with $\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}$ itself and using the ellipticity of B_k from Lemma 4.40 yields

$$
\begin{aligned}
&c\|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|_{L^2(I,V)}^2 + \|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T)\|_H^2 \\
&\leq \Big(j_v'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}) - [\beta_\gamma{}'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}} - \Psi) + f'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}})]\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}, \bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}\Big)_{L^2(Q)} \\
&\quad + (j_T'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T)), \bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T))_H \\
&\leq \|j_v'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}})\|_{L^2(I,V^*)} \|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|_{L^2(I,V)} + \|j_T'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T))\|_H \|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T)\|_H.
\end{aligned}
\tag{5.11}
$$

Here we used the monotonicity of β_γ and f. Note that $c > 0$ does not depend on γ_{n_l},k_{n_l} or h_{n_l}. Using Young's inequality we arrive at

$$
\begin{aligned}
&c\|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|_{L^2(I,V)}^2 + \|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T)\|_H^2 \\
&\leq \frac{1}{2c}\|j_v'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}})\|_{L^2(I,V^*)}^2 + \frac{c}{2}\|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}\|_{L^2(I,V)}^2 + \frac{1}{2}\|j_T'(\bar{y}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T))\|_H^2 + \frac{1}{2}\|\bar{p}_{\gamma_{n_l} k_{n_l} h_{n_l}}(T)\|_H^2.
\end{aligned}
$$

Thus

$$\begin{aligned}\frac{c}{2}\|\bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}}\|^2_{L^2(I,V)} &\leq \frac{c}{2}\|\bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}}\|^2_{L^2(I,V)} + \frac{1}{2}\|\bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}}(T)\|^2_H \\ &\leq \frac{1}{2c}\|j'_v(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}})\|^2_{L^2(I,V^*)} + \frac{1}{2}\|j'_T(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}}(T))\|^2_H.\end{aligned} \tag{5.12}$$

Taking the square root yields the desired result. By the uniform convergence of the states the upper bound actually stays bounded and thus extracting an appropriate subsequence we find a weak $L^2(I,V)$-limit $\bar{p}$ of $(\bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}})_{l\in\mathbb{N}}$. Here, we denoted the subsequence by the same sign.

The analysis of the $(\bar{\eta}_{\gamma_{n_l}k_{n_l}h_{n_l}})_{l\in\mathbb{N}}$ is more involved. Let $\varphi \in H(I,V)\cap W_0(I)$. We have

$$\begin{aligned}(\bar{\eta}_{\gamma_{n_l}k_{n_l}h_{n_l}}, \varphi)_{L^2(Q)} &= -(\beta'_{\gamma_{n_l}}(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}} - \Psi)\bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}}, I_{k_{n_l}}R_{k_{n_l}}\varphi)_{L^2(Q)} \\ &= B_k(I_{k_{n_l}}R_{k_{n_l}}\varphi, \bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}}) + (f'(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}})\bar{p}_{\gamma_{n_l}k_{n_l}h_{n_l}}, I_{k_{n_l}}R_{k_{n_l}}\varphi)_{L^2(Q)} \\ &\quad - (j'_v(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}}), I_{k_{n_l}}R_{k_{n_l}}\varphi)_{L^2(Q)} - (j'_T(\bar{y}_{\gamma_{n_l}k_{n_l}h_{n_l}}(T)), I_{k_{n_l}}R_{k_{n_l}}\varphi(T))_H.\end{aligned}$$

By the strong convergence of the states in $L^\infty(Q)$, the weak convergences of the adjoints, Proposition 5.15 and Proposition 5.16 we conclude that the right hand side converges to

$$\begin{aligned}&a_I(\varphi,\bar{p}) + (\bar{p}, \partial_t\varphi)_{L^2(I,H)} + (f'(\bar{y})\bar{p},\varphi)_{L^2(Q)} - (j'_v(\bar{y}),\varphi)_{L^2(Q)} - (j'_T(\bar{y}(T)),\varphi(T))_H \\ &\quad =: (\bar{\eta},\varphi)_{H^1(I,V)^*,H^1(I,V)}.\end{aligned}$$

The operator $\bar{\eta}$ is, however, not only well-defined on $H^1(I,V)\cap W_0(I)$. The expression on the left is, in fact, continuous with respect to the $W(I)$-norm. Because $H^1(I,V)\cap W_0(I)$ is dense in $W_0(I)$ by Proposition 8.13, $\bar{\eta}$ can now be extended to an operator on $W_0(I)$ such that

$$\begin{aligned}&(\bar{\eta},\varphi)_{W_0(I)^*,W_0(I)} \\ &\quad = a_I(\varphi,\bar{p}) + (\bar{p},\partial_t\varphi)_{L^2(I,H)} + (f'(\bar{y})\bar{p},\varphi)_{L^2(Q)} - (j'_v(\bar{y}),\varphi)_{L^2(Q)} - (j'_T(\bar{y}(T)),\varphi(T))_H\end{aligned}$$

holds for all $\varphi \in W_0(I)$. □

Before we continue to interpret those optimality conditions, let us prove the following corollary:

Corollary 5.18 *The control $\bar{u}$ and the adjoint $\bar{p}$ from Theorem 5.17 satisfy*

$$(\bar{p} + g'(\bar{u}), u - \bar{u})_{L^2(Q)} \geq 0 \qquad \forall u \in U_{ad}.$$

Proof. This is an immediate consequence of the same inequality for the discrete solutions in Corollary 5.11, the strong convergence of the controls in $L^2(Q)$, cf. Theorem 5.3, and the weak convergence in $L^2(Q)\subset L^2(I,V)$ of the adjoints to $\bar{p}$ by Theorem 5.17. □

Remark 5.19 We now compare the properties we just derived, to the stationarity conditions from Theorem 3.38, to see how our "new" multipliers from Theorem 5.17 behave, when compared to the continuous derivation of those multipliers.

The first condition in Theorem 3.38 is clearly satisfied, as we have shown $S(\bar{u}) = \bar{y}$ in the beginning of this chapter. The second condition is actually weaker than what we obtain for $\bar{\eta}$ in Theorem 5.17. In fact the $\bar{\eta}$ from Theorem 5.17 satisfies the same condition as $\bar{\eta}_*$ from Remark 3.30, see also Remark 3.39. The third condition is also satisfied by Corollary 5.18. The fourth condition does not have an equivalent, since we cannot interpret $\bar{\eta}$ as a measure in Theorem 5.17.

The complementarity conditions from Theorem 3.38 are not satisfied, or rather: we were not able to prove them. The difficulty is to interpret $\bar{\eta}_{\gamma kh}$ as a measure by finding an $L^1(Q)$-bound for it. Without a $L^1(Q)$-bound it is essentially impossible to prove the fourth condition of Theorem 3.38, which in turn was used in Lemma 3.37 to show the fifth condition in Theorem 3.38. To establish an $L^1(Q)$-bound on $\bar{\eta}_{\gamma kh}$ or a similar multiplier might be of interest to future research as finding $L^1(Q)$-bounds of (partially) discretized quantities might be related to finding $L^\infty(Q)$-bounds of discretized quantities, which is of great interest, cf. the truncation in $L^\infty(Q)$ of the multiplier in Remark 4.62 and Remark 5.21.

5.2 Implicit Time Discretization of the Controls in the Semi-discrete Case

We will now show that, although we have just discretized the state, we automatically have a temporal discretization of the control, at least under certain assumptions. This is important as optimal controls of (OC) lack regularity in general. This can be seen by inspecting Corollary 3.41: the control $\bar{u}$ has the regularity of the adjoint state $\bar{p}$. But the adjoint state only solves

$$\begin{cases} -\partial_t \bar{p} + A\bar{p} + f'(\bar{y})\bar{p} = j_v'(\bar{y}) + \bar{\eta}, \\ \bar{p}|_{\Sigma_D} = 0, \quad \bar{p}(T) = j_T'(\bar{y}(T)) \end{cases}$$

in the weak sense of Lemma 3.29. That means in general $\bar{p}$ does not posses any form of useful differentiability regularity in time.

Corollary 5.20 *Assume that U_{ad} is given by box constraints $u_l < u_u \in \mathbb{R}$. Then each locally optimal control $\hat{u}_{\gamma kh}$ to* (SOC$_{\gamma kh}$) *satisfies*

$$\bar{u}_{\gamma kh} = P_{[u_l,u_u]}(-\alpha_g^{-1}\bar{p}_{\gamma kh}). \tag{5.13}$$

In particular $\bar{u}_{\gamma kh} \in X_k^0$ with

$$\|\bar{u}_{\gamma kh}\|_{L^2(I,V)} \le C\left(\|j_v'(\bar{y}_{\gamma kh})\|_{L^2(I,V^*)} + \|j_T'(\bar{y}_{\gamma kh}(T))\|_H\right).$$

C does not depend on γ, k or h.

Proof. The proof for the projection formula is the same as for Corollary 3.41 using Corollary 5.11 and the structure of $g(u) = \int_Q \frac{\alpha_g}{2} u^2 \, d(t,x)$. The fact that $\bar{u}_{\gamma kh}$ is piecewise constant in time is now an consequence of the fact that projecting a function from $X_{k,h}^{0,1}$ unto constant constraints keep it piecewise constant on the same intervalls.

By (5.13) and Proposition 8.19 the following holds for almost every $t \in I$

$$\|\bar{u}_{\gamma kh}(t)\|_V \le \alpha_g^{-1}\|\bar{p}_{\gamma kh}(t)\|_V$$

and thus

$$\|\bar{u}_{\gamma kh}\|_{L^2(I,V)} \le \alpha_g^{-1}\|\bar{p}_{\gamma kh}\|_{L^2(I,V)}.$$

We already have bounded $\|\bar{p}_{\gamma kh}\|_{L^2(I,V)}$ by the desired quantities in (5.12). □

Note that we can now state, under the assumptions of Corollary 5.20, that the following equality is true:

$$\min_{u \in U_{ad}} J(S_{\gamma kh}(u), u) = \min_{u \in X_k^0 \cap U_{ad}} J(S_{\gamma kh}(u), u). \tag{5.14}$$

Remark 5.21 Motivated by Section 3.7, the equivalence of the unbounded control problem and the bounded control problem, one can ask the question, wether the unbounded and bounded control problem are also equivalent in the discrete case? That means, essentially, finding a $C > 0$, independent of γ, k or h, such that

$$\|\bar{u}_{\gamma kh}\|_{L^\infty(Q)} \leq C.$$

This bound is essential in proving the convergence rates in this section, e.g. in Theorem 5.3 and all proofs based on it. From our experience in numerical experiments this $C > 0$ appears to exist, see the explanations below (7.2) in Section 7.3.1, but an actual proof would be of great interest for future research.

Lastly, if we happen to have $-\alpha\bar{p}_{\gamma kh} \in U_{ad}$, for all possible adjoints, Equation (5.13) delivers $\bar{u}_{\gamma kh} = -\alpha\bar{p}_{\gamma kh} \in X^{0,1}_{k,h}$, for all those adjoints and controls. Thus, in this case

$$\min_{u \in U_{ad}} J(S_{\gamma kh}(u), u) = \min_{u \in X^{0,1}_{k,h} \cap U_{ad}} J(S_{\gamma kh}(u), u).$$

Hence, the semi-discrete problem is now equivalent to the fully discrete problem and the controls are discretized in time and space automatically. From our experience this happens when U_{ad} is given by sufficiently large box constraints, see for example Section 7.3.1.

5.3 Fully Discrete, Regularized Control Problems

In this section we will finally discuss the fully discrete problem:

$$\begin{aligned} &\min_{(y_{\gamma kh}, u_{kh}) \in X^{0,1}_{k,h} \times X^{0,1}_{k,h}} j_v(y_{\gamma kh}) + j_T(y_{\gamma kh}(T)) + g(u_{kh}) = J(y_{\gamma kh}, u_{kh}), \\ &\text{s.th. } S_{\gamma kh}(u_{kh}) = y_{\gamma kh} \text{ and } u_{kh} \in U_{ad}. \end{aligned} \tag{FOC$_{\gamma kh}$}$$

The problem (FOC$_{\gamma kh}$) is completely finite dimensional and thus accessible to computation, cf. Chapter 6.

Theorem 5.22 *The problem* (FOC$_{\gamma kh}$) *has at least one solution* $(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh}) \in X^{0,1}_{k,h} \times X^{0,1}_{k,h}$.

Proof. The proof is basically the same as the one in the continuous case in Theorem 3.3. For the continuity of $S_{\gamma kh}$ see Lemma 4.43. □

Theorem 5.23 *Each locally optimal control* $\hat{u}_{\gamma kh}$ *of the reduced problem*

$$\min_{u \in U_{ad} \cap X^{0,1}_{k,h}} J(S_{\gamma kh}(u), u) = J(S_{\gamma kh}(\hat{u}_{\gamma kh}), \hat{u}_{\gamma kh})$$

satisfies

$$(p_{\gamma kh}(\hat{u}_{\gamma kh}) + \alpha_g \hat{u}_{\gamma kh}, u_{kh} - \hat{u}_{\gamma kh})_{L^2(Q)} \geq 0 \qquad \forall u_{kh} \in U_{ad} \cap X^{0,1}_{k,h}. \tag{5.15}$$

Let us define $\hat{p}_{\gamma kh} := p_{\gamma kh}(\hat{u}_{\gamma kh})$. *In particular, we find*

$$\hat{u}_{\gamma kh} = P_{U_{ad} \cap X^{0,1}_{k,h}}[-\alpha_g^{-1}\hat{p}_{\gamma kh}].$$

Proof. The optimality condition follows similarly to the semi-discrete case from Theorem 5.9, Lemma 5.10 and Corollary 5.11. The relation in (5.15) just defines $\hat{u}_{\gamma kh}$ as the $L^2(Q)$-projection of $\hat{p}_{\gamma kh}$ on $U_{ad} \cap X^{0,1}_{k,h}$. □

Proposition 5.24 *Let $u_l, u_u \in \mathbb{R}$ with $u_l \leq u_u$ and define $\tilde{U}_{ad} := \{v \in H : u_l \leq v \leq u_u$ a.e. in $\Omega\}$. Then there exist $C > 0$ such that for all $h \in (0, 1]$ and all $u \in V \cap \tilde{U}_{ad}$ there exists a $u_h \in V_h \cap \tilde{U}_{ad}$ such that*

$$\|u_h - u\|_{V^*} \leq Ch^2\|u\|_V \text{ and } \|u_h - u\|_H \leq Ch\|u\|_V.$$

Proof. The desired result is almost just a combination of [dlRMV08, Lemmas 4.2-4.5]. In that paper, however, the authors assume that $\Omega_h = \bar{\Omega}$, which is not the case for our smooth domains. [dlRMV08] treats curved boundaries differently. We also consider homogenous Dirichlet boundary conditions and therefore have to check if their ideas are applicable to our situation.

The authors of [dlRMV08] make the following construction, which goes back to [Car99]: given a nodal hat function $\varphi_{\hat{x}}$, centered at a node $\hat{x} \in \mathcal{N}_h$, they define

$$\pi_{\hat{x}}(u) := \frac{\int_{\operatorname{supp}(\varphi_i)} u\varphi_{\hat{x}}\,dx}{\int_{\operatorname{supp}(\varphi_{\hat{x}})} \varphi_{\hat{x}}\,dx} \qquad \text{and} \qquad \Pi_h u := \sum_{\hat{x}\in\mathcal{N}_h} \pi_{\hat{x}}(u)\varphi_{\hat{x}}.$$

This does, however, not necessarily preserve boundary values. We thus consider the set of interior nodes $\mathcal{N}_h^{int}$, i.e. those $\hat{x} \in \mathcal{N}_h$ such that $\hat{x} \notin \partial\Omega_h$. We define

$$u_h := \sum_{\hat{x}\in\mathcal{N}_h^{int}} \pi_{\hat{x}}(u)\varphi_{\hat{x}} \in V_h.$$

First of all: it easy easy to see that $u_h \in V_h \cap U_{ad}$ by the definition of the $\pi_{\hat{x}}(u)$, cf. [dlRMV08, Lemma 4.5].

[dlRMV08, Lemma 4.2] is applicable to our situation and it delivers for each $\hat{x} \in \mathcal{N}_h$

$$\|u - \pi_{\hat{x}}(u)\|_{L^2(\operatorname{supp}(\varphi_{\hat{x}}))} \leq Ch\|\nabla u\|_{L^2(\operatorname{supp}(\varphi_{\hat{x}}))}.$$

We define the domain

$$\Omega'_h := \bigcup_{\substack{K\in\mathcal{K}_h,\\ K\cap\partial\Omega_h=\emptyset}} K.$$

This is the union of those cells that have positive distance to the discrete boundary. This is where our u_h and the $\Pi_h u$ from [dlRMV08] coincide. The proof of [dlRMV08, Lemma 4.3] is immediately applicable to Ω'_h and delivers

$$\|u - u_h\|_{L^2(\Omega'_h)} \leq Ch\|\nabla u\|_{L^2(\Omega'_h)} \tag{5.16}$$

because we have $\sum_{\hat{x}\in\mathcal{N}_h} \varphi_{\hat{x}} = 1$ on Ω'_h, which is a key part of the proof.

The rest of the domain has to be considered as well, however. We have by [LMWZ10, Lemma 2.1]

$$\|u - u_h\|_{L^2(\Omega\setminus\Omega'_h)} \leq Ch\|\nabla(u - u_h)\|_{L^2(\Omega\setminus\Omega'_h)}.$$

This is possible because $\Omega \setminus \Omega'_h$ is contained in a tube around $\partial\Omega$ with width proportional to h, cf. the arguments in the proof of Proposition 4.18. By Proposition 4.12 we find

$$\|u - u_h\|_{L^2(\Omega\setminus\Omega'_h)} \leq Ch\|\nabla u\|_{L^2(\Omega\setminus\Omega'_h)} + C\|u_h\|_{L^2(\Omega_h\setminus\Omega'_h)} \tag{5.17}$$

Let $K \in \mathcal{K}_h$ be such that $K \subset \Omega_h \setminus \Omega_h'$. Then we have

$$\begin{aligned}
\|u_h\|_{L^2(K)} &\leq \sum_{\hat{x}\in\mathcal{N}_h^{int}\cap K} \|\pi_{\hat{x}}(u)\varphi_{\hat{x}}\|_{L^2(K)} \leq \sum_{\hat{x}\in\mathcal{N}_h^{int}\cap K} \|\varphi_{\hat{x}}\|_{L^2(K)}|\pi_{\hat{x}}(u)| \\
&\leq \sum_{\hat{x}\in\mathcal{N}_h^{int}\cap K} Ch^{\frac{N}{2}} \frac{\int_{\operatorname{supp}(\varphi_{\hat{x}})} |u|\varphi_{\hat{x}}\, dx}{\int_{\operatorname{supp}(\varphi_{\hat{x}})} \varphi_{\hat{x}}\, dx} \\
&\leq Ch^{\frac{N}{2}} \sum_{\hat{x}\in\mathcal{N}_h^{int}\cap K} \|u\|_{L^2(\operatorname{supp}(\varphi_{\hat{x}}))}\|\varphi_{\hat{x}}\|_{L^2(K)}\|\varphi_{\hat{x}}\|_{L^1(K)}^{-1}.
\end{aligned}$$

By Proposition 4.12 we know that $Ch^{-N}\|\varphi_{\hat{x}}\|_{L^1(K)} \geq \|\varphi_{\hat{x}}\|_{L^\infty(K)} = 1$. Thus

$$\|u_h\|_{L^2(K)} \leq C \sum_{\hat{x}\in\mathcal{N}_h^{int}\cap K} \|u\|_{L^2(\operatorname{supp}(\varphi_{\hat{x}}))} \leq C(N+1)\|u\|_{L^2(\bigcup_{\tilde{K}\in\mathcal{K}_h,\tilde{K}\cap K\neq\emptyset}\tilde{K})}.$$

Thus we find

$$\|u_h\|^2_{L^2(\Omega_h\setminus\Omega_h')} \leq C \sum_{\substack{K\in\mathcal{K}_h\\ K\cap\partial\Omega_h\neq\emptyset}} \|u\|^2_{L^2(\bigcup_{\tilde{K}\in\mathcal{K}_h,\tilde{K}\cap K\neq\emptyset}\tilde{K})}.$$

The cells $\tilde{K}$ in the unions are those cells that touch another cell K that touches a boundary. Due to our regularity assumptions each cell has at most L neighbours, cf. [dlRMV08, Remark 3.4 in reference to Assumption 3.3], so that we can conclude

$$\|u_h\|^2_{L^2(\Omega_h\setminus\Omega_h')} \leq CL \sum_{\substack{K\in\mathcal{K}_h:\\ \exists K_1\in\mathcal{K}_h: K_1\cap\partial\Omega_h\neq\emptyset\\ \text{and } K\cap K_1\neq\emptyset}} \|u\|^2_{L^2(K)} \leq C\|u\|^2_{L^2(\{x\in\Omega:\operatorname{dist}(x,\partial\Omega)\leq 2h\})}.$$

Using [LMWZ10, Lemma 2.1] once again we find

$$\|u_h\|_{L^2(\Omega_h\setminus\Omega_h')} \leq Ch\|\nabla u\|_{L^2(\Omega)}.$$

Thus (5.17) delivers

$$\|u - u_h\|_{L^2(\Omega\setminus\Omega_h')} \leq Ch\|\nabla u\|_{L^2(\Omega)}.$$

Together with (5.16) we find

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch\|\nabla u\|_{L^2(\Omega)}.$$

The estimate

$$\|u - u_h\|_{V^*} \leq Ch^2\|\nabla u\|_{L^2(\Omega)}$$

follows now from the proof of [dlRMV08, Lemma 4.4]. □

Theorem 5.25 *Assume U_{ad} is given by box constraints $u_l < u_u$. Let $(\hat{y}_{\gamma_n k_n h_n}, \hat{u}_{\gamma_n k_n h_n})_{n\in\mathbb{N}}$ be a sequence of solutions to* (FOC$_{\gamma kh}$)*. There exists a subsequence and a global solution $(\bar{y}, \bar{u})$ of* (OC) *such that*

$$\begin{aligned}
\hat{y}_{\gamma_{n_l},k_{n_l},h_{n_l}} &\xrightarrow{l\to\infty} \bar{y} \text{ strongly in } L^\infty(Q),\\
\hat{y}_{\gamma_{n_l},k_{n_l},h_{n_l}}(T) &\xrightarrow{l\to\infty} \bar{y}(T) \text{ strongly in } L^\infty(\Omega),\\
\hat{u}_{\gamma_{n_l},k_{n_l},h_{n_l}} &\xrightarrow{l\to\infty} \bar{u} \text{ weakly* in } L^\infty(Q) \text{ and strongly in } L^2(Q).
\end{aligned}$$

Proof. The proof is in line with the semi-discrete case in Theorem 5.3. □

Lemma 5.26 *Let* $(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh})$ *be a global solution to* (FOC$_{\gamma kh}$) *and* $(\bar{y}, \bar{u})$ *be a global solution to* (OC). *We then have*

$$|J(\bar{y}, \bar{u}) - J(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh})| \leq C\left(\gamma^{\frac{1}{\alpha_\beta}} + |\ln k|^2 |\ln h|^4 (k + h^2)\right).$$

Proof. As in the proof of Theorem 5.4 we have

$$J(\bar{y}, \bar{u}) - J(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh}) \leq C\left(\gamma^{\frac{1}{\alpha_\beta}} + |\ln k|^2 |\ln h|^4 (k + h^2)\right).$$

However, to see the reverse estimate we cannot simply do the same as $\bar{u} \notin X_{k,h}^{0,1} \cap U_{ad}$. We insert a solution $\bar{u}_{\gamma kh}$ to (SOC$_{\gamma kh}$), which implies by Corollary 5.20 that $\bar{u}_{\gamma kh} \in X_k^0 \cap U_{ad}$, and see

$$J(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh}) - J(\bar{y}, \bar{u}) = J(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) + J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) - J(\bar{y}, \bar{u}).$$

The last two terms are bounded by the claimed estimate via Theorem 5.4. It remains to estimate

$$J(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}).$$

Thus consider $u_{kh} \in X_{k,h}^{0,1}$, where for almost every $t \in I$ $u_{kh}(t)$ is the function obtained by Proposition 5.24 applied to $\bar{u}_{\gamma kh}(t)$. We then find

$$\begin{aligned}
&J(\hat{y}_{\gamma kh}, \hat{u}_{\gamma kh}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) \\
&\leq J(S_{\gamma kh}(u_{kh}), u_{kh}) - J(\bar{y}_{\gamma kh}, \bar{u}_{\gamma kh}) \\
&= j_v(S_{\gamma kh}(u_{kh})) - j_v(S(\bar{u}_{\gamma kh})) + j_T(S_{\gamma kh}(u_{kh})(T)) - j_T(S(\bar{u}_{\gamma kh})(T)) + g(u_{kh}) - g(\bar{u}_{\gamma kh}).
\end{aligned}$$

As in the proof of (5.5) in the proof of Theorem 5.4 it follows

$$\begin{aligned}
&j_v(S_{\gamma kh}(u_{kh})) - j_v(S(\bar{u})) + j_T(S_{\gamma kh}(u_{kh})(T)) - j_T(S(u)(T)) \\
&\leq C\|S_{\gamma kh}(u_{kh}) - S_{\gamma kh}(\bar{u}_{\gamma kh})\|_{L^2(Q)} + \|S_{\gamma kh}(u_{kh}) - S_{\gamma kh}(\bar{u}_{\gamma kh})(T)\|_H.
\end{aligned}$$

By Lemma 4.43 and Proposition 5.24 this is bounded from above by

$$C\|u_{kh} - \bar{u}_{\gamma kh}\|_{L^2(I,V^*)} \leq Ch^2 \|\bar{u}_{\gamma kh}\|_{L^2(I,V)}. \tag{5.18}$$

By Corollary 5.20 this is bounded from above by

$$Ch^2\left(\|j_v'(\bar{y}_{\gamma kh})\|_{L^2(I,V^*)} + \|j_T'(\bar{y}_{\gamma kh}(T))\|_H\right)$$

which behaves like Ch^2 as $\|\bar{y}_{\gamma kh}\|_{L^\infty(Q)}$ and $\|\bar{y}_{\gamma kh}(T)\|_H$ stay bounded for $(\gamma, k, h) \to 0$ by Theorem 4.65 and the boundedness of U_{ad} in $L^\infty(Q)$, see the proof of (5.4).

It remains to estimate $g(u_{kh}) - g(\bar{u}_{\gamma kh})$. This is straightforward using its definition, (5.18) and the subsequent estimates

$$\begin{aligned}
g(u_{kh}) - g(\bar{u}_{\gamma kh}) &= \alpha \int_Q (u_{kh} + \bar{u}_{\gamma kh})(u_{kh} - \bar{u}_{\gamma kh})\, d(t, x) \\
&\leq \|u_{kh} + \bar{u}_{\gamma kh}\|_{L^2(I,V)} \|u_{kh} - \bar{u}_{\gamma kh}\|_{L^2(I,V^*)} \leq Ch^2.
\end{aligned}$$

This concludes the proof.

□

Just the same way as in the semi-discrete case one can obtain Lemma 5.5 and Theorem 5.6. We combine the result into the following theorem:

Theorem 5.27 *Let* $(\bar{y}, \bar{u})$ *be a strict local minimum to* (OC) *in the sense of Lemma 5.5. Then there exists a sequence of local solutions* $(\hat{y}_{\gamma_n k_n h_n}, \hat{u}_{\gamma_n k_n h_n})_{n\in\mathbb{N}}$ *to* $(\mathrm{FOC}_{\gamma kh})$ *converging to* $(\bar{y}, \bar{u})$ *in the sense of Theorem 5.25. They satisfy*

$$|J(\bar{y}, \bar{u}) - J(\hat{y}_{\gamma_n k_n h_n}, \hat{u}_{\gamma_n k_n h_n})| \leq C(\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2) + h_n).$$

Lastly we obtain the analogoue of Theorem 5.7 by the same proof:

Theorem 5.28 *Let* $(\bar{y}, \bar{u})$ *be a local solution to* (OC) *such that a local quadratic growth condition of the form of Theorem 3.57 holds, i.e. there are* $r, \delta > 0$ *such that*

$$\|u - \bar{u}\|_{L^2(\Omega)} < r \implies J(S(u), u) \geq J(\bar{y}, \bar{u}) + \delta \|u - \bar{u}\|^2_{L^2(\Omega)}.$$

Let $(\hat{y}_{\gamma_n k_n h_n}, \hat{u}_{\gamma_n k_n h_n})$ *be the sequence of local solutions to* $(\mathrm{SOC}_{\gamma kh})$ *that converges to* $(\bar{y}, \bar{u})$ *by Theorem 5.27. Then we have, for* n *so large that* $\|\hat{u}_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} < r$*,*

$$\|\hat{u}_{\gamma_n k_n h_n} - \bar{u}\|_{L^2(Q)} \leq C\sqrt{\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2)},$$
$$\|\hat{y}_{\gamma_n k_n h_n} - \bar{y}\|_{L^\infty(I,H)} \leq C\sqrt{\gamma_n^{\frac{1}{\alpha_\beta}} + |\ln k_n|^2 |\ln h_n|^4 (k_n + h_n^2)}.$$

We decided to skip an in depth discussion of the convergence of multipliers as in Section 5.1.4 as we already saw in Section 5.1.4 that it is difficult to derive a stronger stationarity system than Theorem 3.38 using discrete quantities, cf. Remark 5.19.

6 Solution Algorithms for Discretized, Regularized Obstacle Problems and Optimal Control Problems

We only shortly present the algorithms used to solve ($\mathrm{R}_{\gamma kh}$) and ($\mathrm{FOC}_{\gamma kh}$). We will essentially use a semi-smooth Newton method to solve the regularized, discretized obstacle problem and a Newton method or a trust region method to solve the optimization problem. We then combine those two with a pathfollowing method for the regularization parameter. The focus in this thesis does not lie on the algorithms and their theory. Hence, the following introduction will be short and not focused on theoretical results.

For more extensive discussions on how to solve obstacle problems and related variational inequalities we would like to direct the reader to [IK90, IK03, KKT03], where elliptic variational inequalities are solved using Lagrangian methods, in conjunction with semi-smooth Newton methods and active set strategies. A similar approach using Lagrange multipliers on parabolic variational inequalities is used in [IK06].

The solution of optimal control problems with elliptic variational inequalities are discussed in [IK00] by the means of Lagrange multipliers and active set strategies. The same type of problem is considered in [KW12a] where the Moreau-Yosida regularization and a semi-smooth Newton method in conjunction with a pathfollowing method is used.

6.1 Solving Regularized, Discretized Obstacle Problems

6.1.1 An Implicit Euler Scheme

It is well-known, or at least easily provable, that a discontinuous Galerkin method as introduced in Definition 4.41 is equivalent to an implicit Euler scheme in the finite dimensional space V_h, see also the proof of Lemma 4.51. That means that the equation in V_h^* in ($\mathrm{R}_{\gamma kh}$) is equivalent to the system of equations

$$\begin{aligned} &y_{\gamma kh}^1 = y_0 \in V_h, \\ &y_{\gamma kh}^{j+1} = k_{j+1}(-A_h y_{\gamma kh}^{j+1} + u^{j+1} - f(y_{\gamma kh}^{j+1}) - \beta_\gamma(y_{\gamma kh}^{j+1} - \Psi)) + y_{\gamma kh}^j \in V_h^* \text{ for } j = 0, \dots, M-1. \end{aligned}$$

Here $y_{\gamma kh}^j := y_{\gamma kh}(t_j^+)$. Thus we have to solve $M-1$ non-linear equations in the finite dimensional spaces V_h, V_h^*. For each j we determine $y_{\gamma kh}^{j+1}$ by solving the equation given by the previous iterate $y_{\gamma kh}^j$. We solve the system

$$y_{\gamma kh}^{j+1} + k_{j+1} A_h y_{\gamma kh}^{j+1} + k_{j+1} f(y_{\gamma kh}^{j+1}) + k_{j+1}\beta_\gamma(y_{\gamma kh}^{j+1} - \Psi) = y_{\gamma kh}^j + k_{j+1} u^{j+1}$$

by a semi-smooth Newton method as for example presented in [Ulb11, Chapter 3]. Note that all β_γ-terms of the form of Proposition 2.17 are semi-smooth of order 1, according to [Ulb11, Definition 2.13]. The proof is immediate and thus skipped. To solve the appearing linear systems a multigrid solver with an ILU decomposition as smoother is used, see for example [BB00].

Remark 6.1 It is well-known that Newton's method for finding a root of $F\colon \mathbb{R}^n \to \mathbb{R}$ converges locally quadratic if its derivative F' is Lipschitz continuous, see for example [UU12]. There can also be seen that the constant in the local error estimates is proportional to the Lipschitz constant of F'. In our case we can clearly see that the derivative of

$$\begin{aligned} F\colon V_h &\to V_h^*, \\ y_h &\mapsto y_h + k_{j+1}A_h y_h + k_{j+1} f(y_h) + \beta_\gamma(y_h - \Psi) - y^j_{\gamma kh} + k_{j+1}u^{j+1} \end{aligned}$$

is proportional to γ^{-1} due to the presence of β_γ. This means that the smaller γ, i.e. the more non-linear the F, the worse the convergence behaviour of the Newton's method will be. This is not surprising, but of course unfortunate for us as we of course intend to decrease γ to 0. This cannot be compensated, but one has to keep in mind the errors caused by (k, h). By Theorem 4.65 we have, under appropriate circumstances,

$$\|y - y_{\gamma kh}\|_{L^\infty(Q)} \leq C(\gamma^{\frac{1}{\alpha}} + |\ln k|^2 |\ln h|^4 (k + h^2)).$$

Thus choosing $\gamma \ll (k + h^2)^\alpha$, ignoring the log-terms, has no benefit to the approximation error. Therefore balancing the errors caused by (γ, k, h) might make "larger" γ relatively "unproblematic". Another strategy is a pathfollowing method proposed in Section 6.3.2.

6.1.2 Quadrature Errors in the Numerical Solution of Semilinear Equations

In Section 4.4.2, in particular Theorem 4.61 and Theorem 4.64, we have shown that the Galerkin approximation of a semilinear parabolic equation satisfies an a priori error estimate that is independent of the involved non-linearity. Yet, the numerical intergration of the non-linear term poses certain challenges that could interfere with those estimates. Illustratively we consider the following PDE, with zero initial and boundary conditions:

$$\partial_t y + Ay + \beta_\gamma(y + 1) = u$$

for β_γ being of the form Proposition 2.17 for $\alpha = 1$. This means the obstacle is constant and equal to -1. We also recall its Galerkin approximation

$$B(y_{kh}, \varphi_{kh}) + (\beta_\gamma(y_{kh} - \Psi), \varphi_{kh})_{L^2(Q)} = (u, \varphi_{kh})_{L^2(Q)} \qquad \forall \varphi_{kh} \in X^{0,1}_{k,h}.$$

To compute the solution of this equation we need to evaluate

$$(\beta_\gamma(y_{kh} - \Psi), \varphi_{\hat{x}_i})_{L^2(I_m,\Omega)} = \sum_{T \in \mathcal{T}} k_m \left(\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i}\right)_{L^2(K)}$$

for the nodal basis functions $\varphi_{\hat{x}_i}$ and each time interval I_m for $m \in \{1, \ldots, M\}$. Note that the portion $\Omega \setminus \Omega_h$ can be ignored since y_{kh} and $\varphi_{\hat{x}_i}$ vanish there.

If we have $y_{kh}|_{I_m} \geq \Psi$ on K, i.e. the discrete state is locally admissible, we have

$$\left(\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i}\right)_{L^2(K)} = 0.$$

This integral is very easy to compute numerically.

The other simple case is that $y_{kh}|_{I_m} < \Psi$ on K. Then we have

$$\left(\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i}\right)_{L^2(K)} = \frac{1}{\gamma} \left(y_{kh}|_{I_m} - \Psi, \varphi_{\hat{x}_i}\right)_{L^2(K)}.$$

Due to the smoothness of Ψ and the simple structures of y_{kh} and $\varphi_{\hat{x}_i}$ this is integrable up to a high order error or, in the case $\Psi = -1$, even exactly. Thus causing no quadrature error.

The highest complexity arises when K intersects the boundary of the discrete, regularized active set. Then we have that β_γ is no longer linear on the set $(y_{kh} - \Psi)(K)$, since

$$\beta_\gamma(y_{kh}|_{I_m} - \Psi) = \begin{cases} \frac{1}{\gamma}(y_{kh}|_{I_m} - \Psi) & \text{if } y_{kh}|_{I_m} - \Psi < 0, \\ 0 & \text{if } y_{kh}|_{I_m} - \Psi \geq 0. \end{cases}$$

This is a non-smooth term. Thus usual quadrature strategies can potentially cause a significant error.

There are ways in which we mitigate this effect. Firstly, under many circumstances the boundary of the active set is a $N-1$-dimensional set. Thus by all the convergences one can expect that the regularized, active boundary $\partial\{y_{kh} = \Psi\}$ behaves similarly. Thus we expect that $\partial\{y_{kh} = \Psi\}$ intersects a number of elements proportional to h^{-N+1}. The quadrature error for any quadrature rule based on point evaluations is bounded from above like this:

$$\begin{aligned} &\left|(\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i})_{L^2(\Omega)} - (\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i})_{L^2(\Omega),quad}\right| \\ &\leq C \sum_{\substack{K \in \mathcal{K}_h, \\ K \cap \partial\{y_{kh} = \Psi\} \neq \emptyset}} \left|(\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i})_{L^2(K)} - (\beta_\gamma(y_{kh}|_{I_m} - \Psi), \varphi_{\hat{x}_i})_{L^2(K),quad}\right| \\ &\leq C \sum_{\substack{K \in \mathcal{K}_h, \\ K \cap \partial\{y_{kh} = \Psi\} \neq \emptyset}} 2C_\infty |K| \leq C h^{-N+1} C_\infty h^N \leq C h C_\infty. \end{aligned} \tag{6.1}$$

Here C_∞ is the usual γ-independent truncation of β_γ according to Remark 4.62. This, however, is a large over estimate.

Alternatively one can make a similar argument for smooth β_γ. Assume for the moment that $\beta_\gamma \in C^2(\Omega)$, then clearly on each cell K we have $\beta_\gamma \varphi_{\hat{x}_i}$ for each nodal function. Then the quadrature error for the trapezoidal rule is bounded from above by

$$\begin{aligned} &C \|\nabla^2(\beta_\gamma(y_{kh}|_{I_m} - \Psi)\varphi_{\hat{x}_i})\|_{L^\infty(K)} h^2 \\ &\quad \leq C(\|\nabla^2 \beta_\gamma(y_{kh}|_{I_m} - \Psi)\|_{L^\infty(\Omega)} + \|\nabla(y_{kh}|_{I_m} - \Psi)\beta_\gamma\|_{L^\infty(\Omega)} \|\nabla \varphi_{\hat{x}_i}\|_{L^\infty(\Omega)}) h^2. \end{aligned}$$

The derivatives of β_γ are proportional to $1/\gamma$ and thus the global quadrature error is then of order $\mathcal{O}(\gamma^{-1} h^2 h^{-N+1} h^{\frac{N}{2}}) = \mathcal{O}(\gamma^{-1} h^{3-\frac{N}{2}})$. This follows from the same calculations as in (6.1). So, if one keeps γ fixed, or for a sequence on experiments at least bounded from below, this gives a rate of $\mathcal{O}(h^2)$ for $N = 2$.

Secondly, we do use a quadrature formula on a finer mesh than than the one used for Ω_h. In our examples we subdivide each cell into 16 subcells on which we integrate $\beta_\gamma(y_{kh}|_{I_m} - \Psi) \cdot \varphi_{\hat{x}_i}$. This makes the error relatively small by essentially reducing the constants appearing in the given estimates.

Lastly, in our experiments we do observe not any influence of γ in the discussed regard anyway, see the experiments in Section 7.1.

6.2 Second Order Fréchet derivatives for the Solution Operators of Regularized Obstacle Problems

The purpose of this section is to compile some second order differentiability information on S_γ, as we plan to use a semi-smooth Newton method. The arguments are mostly standard, except where the non-differentiability of β_γ at 0 comes into play. For this reason we give a self-contained proof in this situation. For the fully smooth situation one can consult standard text books, for example [Trö09].

Theorem 6.2 *Assume that β_γ and $f(t,x,\cdot)$ are in $C^{1,1}_{loc}(\mathbb{R})$ for any $(t,x) \in Q$. Here the Lipschitz constants of $f(t,x,\cdot)$ and $f'(t,x,\cdot)$ may not depend on $(t,x) \in Q$. Then*

$$S''_\gamma \colon L^{q_u}(Q) \to \mathrm{Lin}(L^{q_u}(Q), W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)))$$

is locally Lipschitz continuous.

Proof. As β_γ and f have the same behaviour, we integrate β_γ into f for the purposes of this proof. Let $u, d, v \in L^{q_u}(Q)$ and define $w := S_\gamma{}'(u)v$, $w_d := S_\gamma{}'(u+d)v$. We have to show $\|w_d - w\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} \leq C\|v\|_{L^{q_u}(Q)}\|d\|_{L^{q_u}(Q)}$ for sufficiently small d. We see that $\delta w := w_d - w$ solves

$$\begin{cases} \partial_t \delta w + A\delta w + f'(S_\gamma(u+h))w_d - f'(S_\gamma(u))w = 0, \\ \delta w(0) = 0, \quad \delta w|_{\Sigma_D} = 0. \end{cases}$$

Rearranging this equation shows

$$\partial_t \delta w + A\delta w + f'(S_\gamma(u))\delta w = [f'(S_\gamma(u)) - f'(S_\gamma(u+d))]w_d.$$

Theorem 8.17 and the local Lipschitz continuity of f entail

$$\begin{aligned} \|\delta w\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} &\leq \|[f'(S_\gamma(u)) - f'(S_\gamma(u+d))]w_d\|_{L^{q_u}(Q)} \\ &\leq C\|S_\gamma(u) - S_\gamma(u+d)\|_{L^\infty(Q)}\|w_d\|_{L^{q_u}(Q)}. \end{aligned}$$

Here $C > 0$ depends on an upper bound to $\|S_\gamma(u)\|_{L^\infty(Q)}$ and $\|S_\gamma(u+d)\|_{L^\infty(Q)}$. By the Lipschitz continuity from Theorem 8.22 we can conclude

$$\|\delta w\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} \leq C\|d\|_{L^{q_u}(Q)}\|v\|_{L^{q_u}(Q)}$$

with a $C > 0$ depending on a upper bound to $\|u\|_{L^\infty(Q)}$ and $\|d\|_{L^\infty(Q)}$. This shows the local Lipschitz continuity. □

Theorem 6.3 *Assume $f(t,x,\cdot) \in C^{2,1}_{loc}(\mathbb{R})$ for any $(t,x) \in Q$, with derivatives being bounded independently of (t,x). The Lipschitz constant of $f''(t,x,\cdot)$ may also not depend on (t,x). Let β_γ be of the form Proposition 2.17 for $\alpha = 2$ and assume γ to be sufficiently small. Then*

$$S_\gamma{}' \colon L^{q_u}(Q) \to \mathrm{Lin}(L^{q_u}(Q), W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)))$$

is Fréchet differentiable in each u such that $\{S_\gamma(u) = \Psi\}$ is of Lebesgue measure 0. For $u, d_1, d_2 \in L^{q_u}(Q)$ the derivative $S_\gamma{}''(u)[d_1, d_2] =: z$ satisfies

$$\begin{cases} \partial_t z + Az + f'(S_\gamma(u))z + \beta_\gamma{}'(S_\gamma(u) - \Psi)z \\ \qquad = -f''(S_\gamma(u))[S_\gamma{}'(u)d_1\, S_\gamma{}'(u)d_2] - \beta_\gamma{}''(S_\gamma(u) - \Psi)[S_\gamma{}'(u)d_1\, S_\gamma{}'(u)d_2] \\ z(0) = 0, \quad z|_{\Sigma_D} = 0. \end{cases}$$

Remark 6.4 The assumption that γ is small may seem counterintuitive, but is basically only done to avoid multiple case distinctions. It is possible to prove more general results, but this would not enhance our understanding of the appearing equations any further.

Proof. Note that we drop γ from S_γ for the purposes of this proof to reduce visual cluttering. We do not, in fact, prove that the unregularized operator is twice differentiable.

Let $u, d_1, d_2 \in L^{q_u}(Q)$. With the z from above we have to show

$$\|S'(u+d_1)d_2 - S'(u)d_2 - z\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} \leq o(\|d_1\|_{L^{q_u}(Q)})\|d_2\|_{L^{q_u}(Q)}$$

as this is equivalent to

$$\|S'(u+d_1)-S'(u)-S''(u)d_1\|_{\mathrm{Lin}(L^{q_u}(Q),W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega)))}\leq o(\|d_1\|_{L^{q_u}(Q)}).$$

Here we already presupposed the existence of S'' in this step. The step is only illustrative and thus this is no circular argument.

We begin by arguing that only the first two cases of the definition of β_γ are relevant. For γ sufficiently small we have by Theorem 2.37 and some trivial calculations that $S(u)-\Psi\geq -c\gamma^{\frac{1}{2}}$. Thus let γ be small enough that $S(u)-\Psi\geq-\frac{1}{4}$. Because $S: L^{q_u}(Q)\to L^\infty(Q)$ is Lipschitz continuous by Theorem 8.22 we have for $\|d_1\|_{L^{q_u}(Q)}$ sufficiently small that $S(u+d_1)-\Psi>-\frac{1}{2}$. This implies the claim with respect to the cases of β_γ.

We define $w_1:=S'(u+d_1)d_2$ and $w_2:=S'(u)d_2$. They and z satisfy

$$\begin{aligned}(\partial_t+A+f'(S(u))+\beta_\gamma{}'(S(u)-\Psi))w_1&=d_2+[f'(S(u))-f'(S(u+d_1))\\&\quad+\beta_\gamma{}'(S(u)-\Psi)-\beta_\gamma{}'(S(u+d_1)-\Psi)]w_1,\\(\partial_t+A+f'(S(u))+\beta_\gamma{}'(S(u)-\Psi))w_2&=d_2,\\(\partial_t+A+f'(S(u))+\beta_\gamma{}'(S(u)-\Psi))z&=-f''(S(u))[S'(u)d_1\,S'(u)d_2]\\&\quad-\beta_\gamma{}''(S(u)-\Psi)[S'(u)d_1\,S'(u)d_2].\end{aligned}\tag{6.2}$$

Defining $r:=w_1-w_2-z$ we have to show $\|r\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}=o(\|d_1\|_{L^{q_u}(Q)})\|d_2\|_{L^{q_u}(Q)}$. Adding and subtracting the equations in (6.2) and using Theorem 8.17 we see that

$$\begin{aligned}&\|r\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}\\&\leq C\|[f'(S(u))-f'(S(u+d_1))+\beta_\gamma{}'(S(u)-\Psi)-\beta_\gamma{}'(S(u+d_1)-\Psi)]w_1\\&\quad+f''(S(u))[S'(u)d_1,S'(u)d_2]+\beta_\gamma{}''(S(u)-\Psi)[S'(u)d_1,S'(u)d_2]\|_{L^{q_u}(Q)}\\&\leq C\|[f'(S(u+d_1))w_1-f'(S(u))w_1-f''(S(u))[S'(u)d_1,S'(u)d_2]\|_{L^{q_u}(Q)}\\&\quad+C\|\beta_\gamma{}'(S(u+d_1)-\Psi)w_1-\beta_\gamma{}'(S(u)-\Psi)w_1-\beta_\gamma{}''(S(u)-\Psi)[S'(u)d_1,S'(u)d_2]\|_{L^{q_u}(Q)}.\end{aligned}\tag{6.3}$$

The terms involving f are treated similarly to the terms with β_γ. We thus only analyse the terms involving β_γ. We have the following estimates by integration and the triangle inequality:

$$\begin{aligned}&\|\beta_\gamma{}'(S(u+d_1)-\Psi)w_1-\beta_\gamma{}'(S(u)-\Psi)w_1-\beta_\gamma{}''(S(u)-\Psi)[S'(u)d_1,S'(u)d_2]\|_{L^{q_u}(Q)}\\&\leq\left\|\int_0^1\beta_\gamma{}''(S(u+d_1)-\Psi+s(S(u)-S(u+d_1)))\,ds\cdot\right.\\&\qquad\cdot[(S(u)-S(u+d_1))w_1-S'(u)d_1,S'(u)d_2]\|_{L^{q_u}(Q)}\\&+\left\|\int_0^1\beta_\gamma{}''(S(u+d_1)-\Psi+s(S(u)-S(u+d_1)))-\beta_\gamma{}''(S(u)-\Psi)\,ds\cdot\right.\\&\qquad\left.\cdot[S'(u)d_1,S'(u)d_2]\right\|_{L^{q_u}(Q)}.\end{aligned}\tag{6.4}$$

Using basic Hölder estimates we bound this from above by

$$\begin{aligned}&\frac{2}{\gamma}\|(S(u)-S(u+d_1))w_1-S'(u)d_1S'(u)d_2\|_{L^{q_u}(Q)}\\&+\int_0^1\|\beta_\gamma{}''(S(u+d_1)-\Psi+s(S(u)-S(u+d_1)))-\beta_\gamma{}''(S(u)-\Psi)\|_{L^{q_u}(Q)}\,ds\\&\qquad\cdot\|S'(u)d_1\|_{L^\infty(Q)}\|S'(u)d_2\|_{L^\infty(Q)}.\end{aligned}\tag{6.5}$$

Recalling $w_1 = S'(u+d_1)d_2$ and the Fréchet differentiability of S, cf. Theorem 3.11, the first term in (6.5) is bounded by

$$\begin{aligned}&\frac{C}{\gamma}(\|(S(u) - S(u+d_1) - S'(u)d_1)w_1\|_{L^{q_u}(Q)} + \|S'(u)d_1(S'(u+d_1)d_2 - S'(u)d_2)\|_{L^{q_u}(Q)})\\ &\leq \frac{C}{\gamma}(o(\|d_1\|_{L^{q_u}(Q)})\|S'(u+d_1)d_2 - S'(u)d_2\|_{L^\infty(Q)} + o(\|d_1\|_{L^{q_u}(Q)})\|S'(u)d_2\|_{L^\infty(Q)}\\ &\quad + \|S'(u)d_1\|_{L^\infty(Q)}\|S'(u+d_1)d_2 - S'(u)d_2\|_{L^{q_u}(Q)})\end{aligned}$$

By the Lipschitz continuity of S from Theorem 6.2 and Theorem 8.17 we have that for sufficiently small d_1, independent of d_2, this is bounded from above by

$$\begin{aligned}&\frac{C}{\gamma}(o(\|d_1\|_{L^{q_u}(Q)})(\|d_1\|_{L^{q_u}(Q)} + 1)\|d_2\|_{L^{q_u}(Q)} + \|d_1\|^2_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)})\\ &= \frac{C}{\gamma}o(\|d_1\|_{L^{q_u}(Q)})\|d_2\|_{L^{q_u}(Q)}.\end{aligned} \tag{6.6}$$

The second term in (6.5) is treated similarly as in Theorem 3.13. Let $\epsilon > 0$. By the Lipschitz continuity of $S : L^{q_u}(Q) \to L^\infty(Q)$ from Theorem 8.22 we have for d_1 sufficiently small that $|S(u) - \Psi| > \epsilon$ implies $|S(u+d_1) - \Psi| > \frac{\epsilon}{2}$, cf. the arguments above (6.2). Thus $S(u) - \Psi$ and $S(u+d_1) - \Psi$ share the same sign on $\{|S(u) - \Psi| > \epsilon\}$. By the form of the second derivative of β_γ'' and Theorem 8.17 the second term of (6.5) is thus bounded from above by

$$\frac{C}{\gamma}\|1\|_{L^{q_u}(\{|S(u)-\Psi|\leq\epsilon\})}\|d_1\|_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)} = \frac{C}{\gamma}|\{|S(u) - \Psi| \leq \epsilon\}|^{\frac{1}{q_u}}\|d_1\|_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)}.$$

This and (6.6) inserted into (6.4) thus yields

$$\begin{aligned}&\|\beta_\gamma'(S(u+d_1) - \Psi)w_1 - \beta_\gamma'(S(u) - \Psi)w_1 - \beta_\gamma''(S(u) - \Psi)[S'(u)d_1\, S'(u)d_2]\|_{L^{q_u}(Q)}\\ &\leq \frac{C}{\gamma}\left(o(\|d_1\|_{L^{q_u}(Q)})\|d_2\|_{L^{q_u}(Q)} + |\{|S(u) - \Psi| \leq \epsilon\}|^{\frac{1}{q_u}}\|d_1\|_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)}\right).\end{aligned}$$

As f'' is locally Lipschitz continuous in the third component the same arguments yield

$$\begin{aligned}&\|[f'(S(u+d_1))w_1 - f'(S(u))w_1 - f''(S(u))[S'(u)d_1\, S'(u)d_2]\|_{L^{q_u}(Q)}\\ &\leq o(\|d_1\|_{L^{q_u}(Q)})\|d_2\|_{L^{q_u}(Q)}.\end{aligned}$$

Thus by (6.3) we find

$$\frac{\|r\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}}{\|d_1\|_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)}} \leq \frac{C}{\gamma}\left(\frac{o(\|d_1\|_{L^{q_u}(Q)})}{\|d_1\|_{L^{q_u}(Q)}} + |\{|S(u) - \Psi| \leq \epsilon\}|^{\frac{1}{q_u}}\right)$$

and can conclude

$$\limsup_{\|d_1\|_{L^{q_u}(Q)}\to 0} \frac{\|r\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}}{\|d_1\|_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)}} \leq |\{|S(u) - \Psi| \leq \epsilon\}|^{\frac{1}{q_u}}.$$

As $\epsilon > 0$ was arbitrary, we can send it to 0 and see that

$$\limsup_{\|d_1\|_{L^{q_u}(Q)}\to 0} \frac{\|r\|_{W(I)\cap C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))}}{\|d_1\|_{L^{q_u}(Q)}\|d_2\|_{L^{q_u}(Q)}} \leq |\{S(u) = \Psi\}|^{\frac{1}{q_u}}.$$

By assumption the right hand side is 0, concluding the proof.

□

For the case that one would consider a smooth β_γ the differentiability can obviously also be obtained. Remark 6.4 applies here as well.

Theorem 6.5 *Assume $f(t,x,\cdot) \in C^{2,1}_{loc}(\mathbb{R})$ for any $(t,x) \in Q$, with derivatives being bounded independently of (t,x). The Lipschitz constant of $f''(t,x,\cdot)$ may also not depend on (t,x). Let β_γ be of the form from Proposition 2.17 for $\alpha > 2$ and γ sufficiently small or $\beta_\gamma \in C^{2,1}_{loc}(\mathbb{R})$. Then*

$$S'_\gamma \colon L^{q_u}(Q) \to \mathrm{Lin}(L^{q_u}(Q), W(I) \cap C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)))$$

is Fréchet differentiable. For $u, d_1, d_2 \in L^{q_u}(Q)$ the derivative $S''(u)[d_1, d_2] =: z$ satisfies

$$\begin{cases} \partial_t z + Az + f'(S(u))z + \beta_\gamma{}'(S(u) - \Psi)z \\ \qquad = -f''(S(u))[S'(u)d_1\, S'(u)d_2] - \beta_\gamma{}''(S(u) - \Psi)[S'(u)d_1\, S'(u)d_2] \\ z(0) = 0, \quad z|_\Sigma = 0. \end{cases}$$

Proof. If β_γ is of the form from Proposition 2.17 for $\alpha > 2$ we have by the same arguments from the previous theorem, that only the first two cases in β_γ are active. Thus in either case we basically have $\beta_\gamma \in C^{2,1}_{loc}(\mathbb{R})$. Thus retracing the steps of the previous theorem or a standard text book, e.g. [Trö09], yield the desired result. □

6.3 Solving Regularized, Fully Discretized Optimal Control Problems

6.3.1 Newton's Method

We solve the fully discretized optimal control problem given by ($\mathrm{FOC}_{\gamma kh}$). It is a finite dimensional optimization problem and thus its solution is relatively straight forward. Let us assume that β_γ is smooth, i.e. at least in $C^{1,1}(\mathbb{R})$, so that Theorem 4.44 applies and $S_{\gamma kh}$ is differentiable. Considering the same strategies used for S_γ in Section 6.2 we can obtain the following theorem analogous to Theorem 6.3:

Theorem 6.6 *Assume $f(t,x,\cdot) \in C^{2,1}_{loc}(\mathbb{R})$ for any $(t,x) \in Q$, with derivatives being bounded independently of (t,x). Let β_γ be of the form Proposition 2.17 for $\alpha = 2$ and assume (γ, k, h) to be sufficiently small. Then*

$$S'_{\gamma kh} \colon L^{q_u}(Q) \to \mathrm{Lin}(L^{q_u}(Q), X^{0,1}_{k,h})$$

is Fréchet differentiable in each u such that $\{S_{\gamma kh}(u) = \Psi\}$ is of Lebesgue measure 0. For $u, d_1, d_2 \in L^{q_u}(Q)$ the derivative $S''_{\gamma kh}(u)[d_1, d_2] =: z$ satisfies

$$\begin{aligned} &B(z, \varphi_{kh}) + (f'(S_{\gamma kh}(u))z + \beta_\gamma{}'(S_{\gamma kh}(u) - \Psi)z, \varphi_{kh})_{L^2(Q)} \\ &= (-f''(S(u))[S'_{\gamma kh}(u)d_1, S'_{\gamma kh}(u)d_2] - \beta_\gamma{}''(S_{\gamma kh}(u) - \Psi)[S'_{\gamma kh}(u)d_1, S'_{\gamma kh}(u)d_2])_{L^2(Q)} \end{aligned}$$

for any $\varphi_{kh} \in X^{0,1}_{k,h}$.

One could also prove the corresponding variant of Theorem 6.5. We leave this out as we only actively use the stated version in the numerical experiments.

Under the assumptions of the previous theorem, we have that that $J(S_{\gamma kh}(u), u) =: j_{\gamma kh}(u)$ is smooth. Thus we can apply a Newton's method to $j_{\gamma kh} \colon X^{0,1}_{k,h} \to \mathbb{R}$. Note that the Lipschitz continuity of second derivative is responsible for the locally quadratic convergence behaviour of the method. The derivatives of $j_{\gamma kh}$ are of the form

$$\begin{aligned} j'_{\gamma kh}(u)d_u &= J_y(S_{\gamma kh}(u),u)S'_{\gamma kh}(u)d_u + J_u(S_{\gamma kh}(u),u)d_u, \\ j''_{\gamma kh}(u)[d_1,d_2] &= J_{yy}(S_{\gamma kh}(u),u)[S'_{\gamma kh}(u)d_1, S'_{\gamma kh}(u)d_2] + J_y(S_{\gamma kh}(u),u)S''_{\gamma kh}(u)[d_1,d_2]+ \\ &\quad + J_{yu}(S_{\gamma kh}(u),u)[S'_{\gamma kh}(u)d_1,d_2] + J_{uy}(S_{\gamma kh}(u),u)[S'_{\gamma kh}(u)d_2,d_1] \\ &\quad + J_{uu}(S_{\gamma kh}(u),u)[d_1,d_2]. \end{aligned}$$

Remark 6.7 We see that the regularities of $S_{\gamma kh}$ play a vital role in the smoothness of $j''_{\gamma kh}$. As in Remark 6.1 this means that smaller γ cause worse convergence behaviour of the applied Newton method.

Note that a simple Newton's method does not necessarily yield convergence or even admissibility of the computed solution. The admissibility can be obtained by using a primal dual active set strategy, cf. [UU12, Algorithmus 20.1] for the schematic overview in the context of quadratic optimization. The global convergence will not be proven here, but we use a trust region Method in conjunction with Newton's method, which is known to have nice, globalizing properties. The method can be found in [UU12]. Any appearing symmetric operators are inverted using a conjugate gradient method.

6.3.2 A Pathfollowing Strategy

As stated in Remark 6.1 and Remark 6.7 small regularization parameters are adverse for the convergence behaviour of both, the solution of the semilinear equation and the optimal control problem itself. Thus it is advisable to use a pathfollowing strategy. We therefore first solve (FOC$_{\gamma kh}$) for a large $\gamma_0 \gg 0$ with an arbitrary starting control $u_0 \in U_{ad} \cap X^{0,1}_{k,h}$. We obtain an optimal solution $\bar{u}_{\gamma_0 kh} \in U_{ad} \cap X^{0,1}_{k,h}$. Then reduce γ_0 by some factor $\eta_\gamma \in (0,1)$ to obtain a $\gamma_1 < \gamma_0$. Then we solve (FOC$_{\gamma kh}$) for γ_1 using $\bar{u}_{\gamma_0 kh}$ as the initial guess. This should reduce the number of iterations required as $\bar{u}_{\gamma_0 kh}$ and $\bar{u}_{\gamma_1 kh}$ are expected to be very close. Iterating this we obtain a standard pathfollowing strategy, presented in pseudo code in Algorithm 2.

We do not give any convergence proof or any deeper results as this is not the focus of this thesis. Considering that in each iteration one does usually not obtain $\bar{u}_{\gamma_j kh}$ but only an approximation, makes it clear that this analysis is non-trivial. A really in depth discussion of pathfollowing strategies and their convergence in the context of optimal control of state constrained problems, which are related to control problems with obstacle problem constraints, cf. [CW19, CV19], can be found in [Kru14].

Algorithm 2: A simple pathfollowing strategy in the regularization parameter.

```
Data: γ_0, γ > 0, η_γ ∈ (0,1) such that γ_0 · η_γ^K = γ for some K ∈ ℕ, u_0 ∈ U_ad ∩ X^{0,1}_{k,h}
Result: ū_{γkh}
Set l = 1, γ_l = γ_0;
while γ_l ≥ γ do
  | Solve (FOC_{γkh}) for (γ_l, k, h) and the initial guess u_0 and obtain ū_{γ_l kh} via
  |   Section 6.3.1; Set γ_l = η_γ · γ_l, u_0 = ū_{γ_l kh};
end
```

7 Numerical Examples

7.1 Sharpness of Convergence Rates of the Discretization of Regularized Obstacle Problems

We consider a numerical example that indicates that the error estimate from Theorem 4.64 is indeed sharp, apart from maybe the powers in logarithmic factors. We also see that the constant in the estimate does indeed not depend on the regularization parameter γ. We work on the time-space cylinder $Q = (0,1) \times B_1(0)$. We have a small gap to the theory here, as we use bilinear elements. . That means all cells $K \in \mathcal{K}_h$ are quadrilaterals that are obtained by a bilinear transformation J_T from the reference element $[0,1]^2$. Each $v_h \in V_h$ satisfies that $v_h|_T = \tilde{v}_h \circ J_T$ where $\tilde{v}_h : [0,1]^2 \to \mathbb{R}$ is bilinear, see for example [Kat08, Chapter 13]. Another small gap is that all integrals and functionals are evaluated only on $I \times \Omega_h$ instead of Q. Since the boundaries of $\partial\Omega_h$ and $\partial\Omega$ have a distance proportional to h^2, since the assumptions of Proposition 4.6 will always be satisfied, this error is considered negligible.

As parameters for the algorithms described in the previous chapter we choose a relative tolerance of 10^{-15} for the semi-smooth Newton method to solve the semilinear equations, a global tolerance of 10^{-13}, a dampening factor of 0.01 and a tolerance of 10^{-6} for the appearing multigrid solver for the linear system.

For the actual problem we consider the bounded, but non-smooth function

$$u(t,x) = \begin{cases} -5 & \text{if } t \leq 0.8, \\ 5 & \text{if } t > 0.8. \end{cases}$$

Furthermore, consider the initial state $y_0 = 0$ and the nonlinearity $f = 0$. A is given as the negative Laplacian. As regularization term β we choose the one from Proposition 2.17 for $\alpha_\beta = 1$ and we truncate it, in the sense of Remark 4.62, liberally at -1000. We will observe than this does not negatively influence the convergence order at all and was mainly done to avoid slow down of the code as β has to be defined via case distinctions and having one less case happening turned out to be positive for the computation speed. All appearing semilinear equations are solved by a semi-smooth Newton method, cf. Section 6.1.

For all experiments in Section 7.1 we use a reference solution $y_\gamma^{ref} \in X_{k,h}^{0,1}$ as approximation to the real solution $y_\gamma = S_\gamma(u)$. We choose $y_\gamma^{ref} = S_{\gamma,k_{ref},h_{ref}}(u)$ with $h_{ref} \approx 1.2 \cdot 10^{-2}$ and $k_{ref} \approx 9.8 \cdot 10^{-4}$. The parameter γ will vary throughout the experiments.

In the first experiment we send k to zero and observe the behaviour of

$$\|S_\gamma(u) - S_{\gamma kh}(u)\|_{L^\infty(Q)} \approx \|S_{\gamma,k_{ref},h_{ref}}(u) - S_{\gamma kh}(u)\|_{L^\infty(Q)}.$$

Here we choose and keep h and γ fixed. Looking at Figure 7.1 one can see that for large h the error in h dominates, leading to a saturation behaviour. For the smaller h one can however see that the convergence behaviour is roughly of $\mathcal{O}(k)$. This indicates that the error estimate from Theorem 4.64 is indeed sharp in k and has the additive structure.

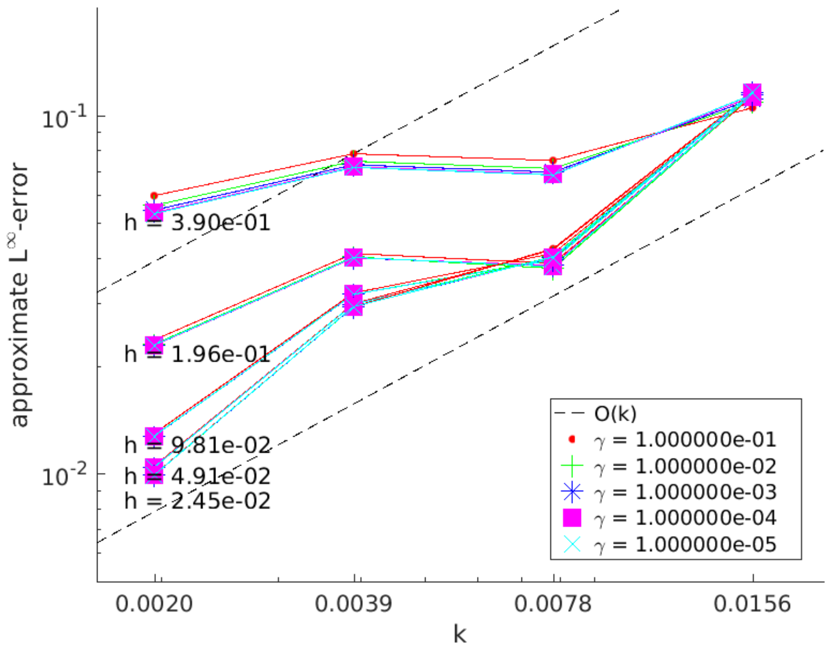


Figure 7.1: The approximate convergence behaviour of $\|S_\gamma(u) - S_{\gamma kh}(u)\|_{L^\infty(Q)}$ for $k \to 0$ for various, fixed h and γ.

The most interesting part, however, is the fact that even for different γ the curves overlap quite heavily. That means γ does not play a part in the the $L^\infty(Q)$-finite element error, exactly as predicted in Theorem 4.64.

In the second experiment we now keep k fixed and send h to zero and observe the error behaviour again. Looking at Figure 7.2 one can see a similar, but more pronounced error saturation as before. Indeed only for the smallest $k = k_{ref}$ an unobstructed convergence order of roughly $\mathcal{O}(h^2)$ be observed. This is due to the fact that the linear convergence in k causes a much greater error in the time discretization than in the spatial discretization.

Once again, even for different γ the curves overlap quite heavily. That means γ does not play a part in the convergence behaviour of the $L^\infty(Q)$-estimate.

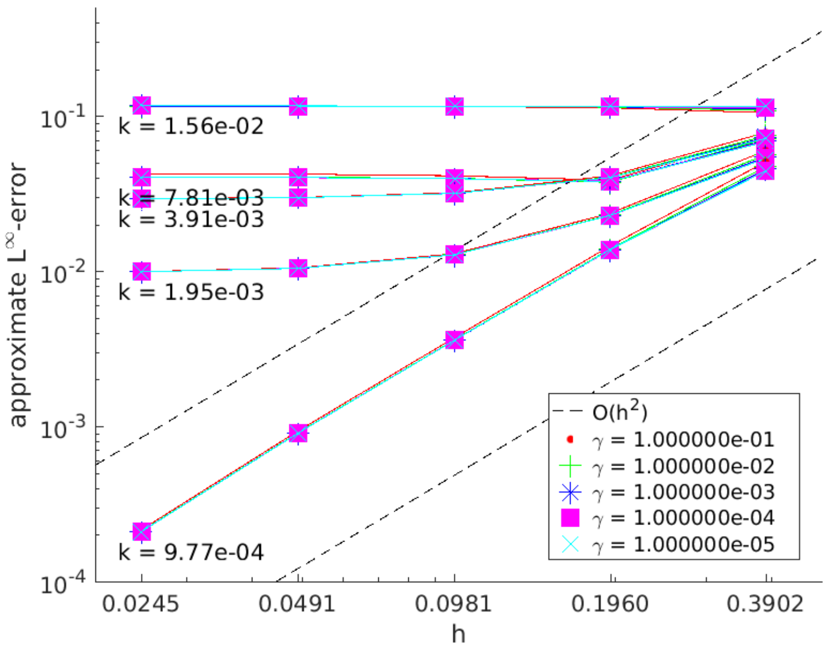


Figure 7.2: The approximate convergence behaviour of $\|S_\gamma(u) - S_{\gamma kh}(u)\|_{L^\infty(Q)}$ for $h \to 0$ for various, fixed k and γ.

7.2 Sharpness of the Regularization Error Estimate of Regularized Obstacle Problems

We consider the same quantities as in Section 7.1, except we use different parameters. Here $y^{ref} = S_{\gamma_{ref},k_{ref},h_{ref}}(u)$ with $\gamma_{ref} = 10^{-4}$, $h_{ref} \approx 6.14 \cdot 10^{-3}$ and $k_{ref} \approx 4.88 \cdot 10^{-4}$. We compute the error $\|S_{\gamma kh}(u) - y^{ref}\|_{L^\infty(Q)} \approx \|S_{\gamma kh}(u) - S(u)\|_{L^\infty(Q)}$ for various, fixed k, h and send γ to 0 and observe the convergence behaviour.

In Figure 7.3 one can see that for small h and k the error in γ dominates in a linear fashion. Showing that Theorem 2.37 is a sharp estimate. One can also clearly see that the error in k is the dominant error, as the curves $+$ and $*$ are very close, yet the h varies by a factor of ≈ 4 between the two curves.

Remark 7.1 [Noc88] also argues for sharpness of the regularization error in the elliptic case, by explicitly stating an example, see [Noc88, Remark 2.2].

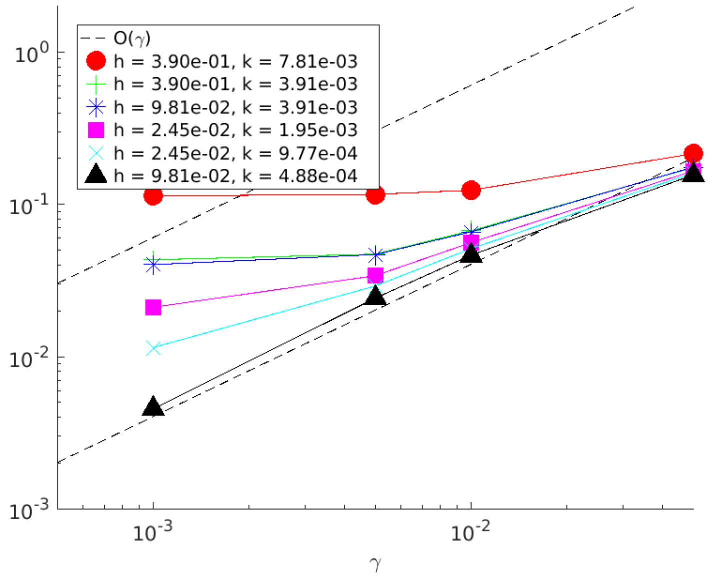


Figure 7.3: The approximate convergence behaviour of $\|S(u) - S_{\gamma kh}(u)\|_{L^\infty(Q)}$ for $\gamma \to 0$ for various, fixed k, h.

7.3 An Optimal Control Example with Exact Solution

7.3.1 Construction of the Example

In this section we consider an example where the exact solution is known. We are mostly interested in analysing how sharp the estimates from Theorem 5.28 for the controls are. In the context of state constrained problems, which are related to the optimization of obstacle problems, see [CW19, CV19], it is well-known that the states can exhibit better convergence rates, even though the predicted rates are worse, cf. [MRV11, the end of Section 8].

The following construction is inspired by the example in [CV19]. We work with quadratic cost functionals, specifically

$$J(y, u) = \frac{1}{2}\|y - y_Q\|^2_{L^2(Q)} + \frac{10^{-3}}{2}\|u\|^2_{L^2(Q)}.$$

We will heavily use rotational symmetry. To that end we define $\Omega := B_1(0)$, $I := [0, 1]$,

$$\begin{aligned} g\colon [0,\pi] &\to \mathbb{R}, \\ r &\mapsto \cos(r) - \left(1 - \frac{1}{2}r^2\right) - \frac{3}{16\pi^2}r^4 + \left(2 - \frac{5}{16}\pi^2\right) \\ &= \cos(r) + \left(1 - \frac{5}{16}\pi^2\right) + \frac{1}{2}r^2 - \frac{3}{16\pi^2}r^4 \end{aligned} \tag{7.1}$$

and

$$\begin{aligned} \varphi\colon [0,1] &\to \mathbb{R}, \\ t &\mapsto \begin{cases} 48t^2 - 128t^3 & \text{if } t \le 0.25 \\ 1 & \text{if } 0.25 < t < 0.75 \\ 48(1-t)^2 + 128(1-t)^3 & \text{if } 0.75 \le t. \end{cases} \end{aligned}$$

Using those two functions we define the rotationally symmetric state on Q

$$\bar{y}(t,x) = \varphi(t)g(\pi|x|).$$

We will choose the control such that there is no multiplier $\bar{\lambda}$, i.e.

$$\bar{u}(t,x) := (\partial_t - \Delta)\bar{y}(t,x) = \varphi'(t)g(\pi|x|) - \varphi(t)\Delta(g(\pi|x|)). \tag{7.2}$$

By some basic, but tedious calculations, see Proposition 8.7, we see that $\|\bar{u}\|_{L^\infty(Q)} \leq 60$. This is a strong over estimate, yet, we choose U_{ad} to be given by box constraints $[-70, 70]$. The reason we do this, is that we can now avoid the usage of a primal-dual active set strategy and essentially solve an unbounded problem as it turns out that all computed, regularized and discretized controls, happen to lie comfortably in $[-70, 70]$ for any point in space-time, even without this additional strategy. This yields increased performance of the code. In the computed example the control for the "roughest" approximation, i.e. $\gamma = 1$, $k \approx 1.6 \cdot 10^{-2}$ and $h \approx 7.7 \cdot 10^{-1}$, lie in $[-47.7, 22.7]$. An approximation for $\gamma = 10^{-5}$, $k \approx 2.0 \cdot 10^{-3}$, $h \approx 4.9 \cdot 10^{-2}$, however, lies in $[-9.9, 6.4]$. All other observed controls lies in between.

By basic formulae for the Laplacian we have $\Delta g(|x|) = g''(|x|) + \frac{g'(|x|)}{|x|}$ for $x \neq 0$, which is one of the reasons we chose this particular g: the first few terms of its Taylor expansion vanish so that $\Delta g(|x|)$ is still smooth. Thus

$$\bar{u}(t,x) = \varphi'(t)g(\pi|x|) - \pi^2\varphi(t)\left(g''(\pi|x|) + \frac{g'(\pi|x|)}{\pi|x|}\right). \tag{7.3}$$

As the obstacle we choose a function, that is constant over time:

$$\Psi(t,x) := \begin{cases} g(\pi|x|) & \text{if } |x| < \frac{1}{2}, \\ g(\frac{\pi}{2}) + \pi g'(\frac{\pi}{2})(|x| - \frac{1}{2}) & \text{if } |x| \geq \frac{1}{2}. \end{cases}$$

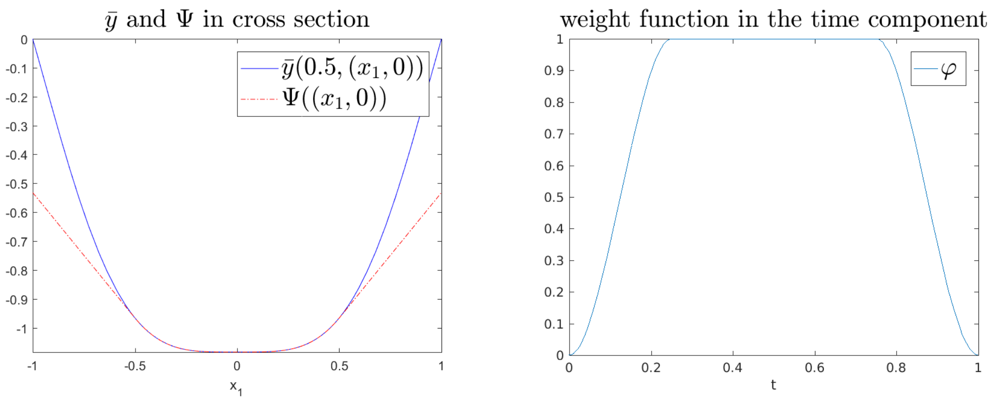


Figure 7.4: The cross sections from $\bar{y}$ and Ψ at $t = 0.5$ from our constructed example together with the function φ on the right.

With Figure 7.4 it is easy to see that the active set is $[0.25, 0.75] \times B_{\frac{1}{2}}(0) \subset Q$. A proof is a lenghty, but elementary calculation utilizing one-dimensional calculus. As we do not have active box constraints, we simply choose

$$\bar{p}(t,x) := -\alpha\bar{u}(t,x).$$

Its straight forward to check boundary conditions. Since this $\bar{p}$ is quite smooth, it should satisfy

$$(-\partial_t - \Delta)\bar{p} = \bar{y}(t,x) - y_Q(t,x) + \bar{\eta}(t,x) \text{ a.e. in } Q.$$

We choose $\bar{\eta}$ last, as its the key of this construction, and see that

$$\begin{aligned} y_Q(t,x) &:= -\alpha(\partial_t + \Delta)\bar{u} + \bar{y}(t,x) + \bar{\eta}(t,x) \\ &= -\alpha(\partial_t^2 - \Delta\Delta)\bar{y}(t,x) + \bar{y}(t,x) + \bar{\eta}(t,x) \\ &= -\alpha\varphi''(t)g(\pi|x|) + \alpha\pi^4\left(g^{(IV)}(\pi|x|) + \frac{2g'''(\pi|x|)}{\pi|x|} - \frac{g''(\pi|x|)}{\pi^2|x|^2} + \frac{g'(\pi|x|)}{\pi^3|x|^3}\right) + \bar{\eta}(t,x). \end{aligned}$$

The multiplier $\bar{\eta}$ is supposed to live on the active set $[0.25, 0.75] \times B_{\frac{1}{2}}(0) \subset Q$. Thus we simply choose $\bar{\eta} = -2 \cdot 1_{[0.25,0.75]\times B_{\frac{1}{2}}(0)}$.

All in all, the given data satisfy the necessary first order optimality conditions from Theorem 3.38. It is also easy to see that all the assumptions for quadratic growth, cf. Assumption 3.54, are satisfied, except $\bar{p} \geq 0$ on a neighborhood of the active set. We, however, still get that $(\bar{y}, \bar{u})$ is the unique, global minimum our optimization problem by Theorem 3.52 and Corollary 3.53. Let S_{lin} be defined as in Theorem 3.52. It is self-adjoint because A is self-adjoint. For any $u \in \{u \in U_{ad} : S(u) \geq \Psi\}$ we find

$$\begin{aligned} (S_{lin}^*(\bar{y} - y_Q) + \alpha_g\bar{u}, u - \bar{u})_{L^2(Q)} &= (S_{lin}^*(\bar{y} - y_Q) - S_{lin}^*(\bar{y} - y_Q + \bar{\eta}), u - \bar{u})_{L^2(Q)} \\ &= (S_{lin}^*(-\bar{\eta}), u - \bar{u})_{L^2(Q)} = (-\bar{\eta}, S(u) - \bar{y})_{L^2(Q)}. \end{aligned}$$

Here we used in particular that $\bar{\lambda} = 0$. By the definition of $\bar{\eta}$ and $\bar{u}$ we find

$$(S_{lin}^*(\bar{y} - y_Q) + \alpha_g\bar{u}, u - \bar{u})_{L^2(Q)} = 2\int_{\{\bar{y}=\Psi\}} S(u) - \Psi \, d(t,x) \geq 0.$$

This is the necessary condition from Corollary 3.53 and shows the optimality of $(\bar{y}, \bar{u})$.

Remark 7.2 With respect to the implementation it is important to treat the factors $|x|^{-1}$, $|x|^{-2}$, $|x|^{-3}$ in the previous example properly, when integrating numerically. They seemingly explode for $|x| \to 0$. All quantities in the example were chosen in such a way that this does not happen. This can be seen by considering the Taylor expansion of the appearing terms. We have for $|x| = r \neq 0$

$$\begin{aligned} \Delta(g(|x|)) &= g''(r) + \frac{g'(r)}{r} \\ &= -\frac{\sin(r)}{r} - \cos(r) - \frac{3}{\pi^2}r^2 + 2 \\ &= -\left(1 - \frac{1}{6}r^2 + \mathcal{O}(r^4)\right) - \left(1 - \frac{1}{2}r^2 + \mathcal{O}(r^4)\right) - \frac{3}{\pi^2}r^2 + 2 \\ &= \left(\frac{2}{3} - \frac{3}{\pi^2}\right)r^2 + \mathcal{O}(r^4) \end{aligned}$$

and

$$\begin{aligned} \Delta\Delta(g(|x|)) &= \cos(r) + \frac{2\sin(r)}{r} + \frac{\cos(r)}{r^2} - \frac{\sin(r)}{r^3} - \frac{12}{\pi^2} \\ &= \dots \\ &= \left(\frac{7}{3} - \frac{12}{\pi^2}\right) - \frac{47}{60}r^2 + \mathcal{O}(r^4). \end{aligned}$$

We recommend using the derived Taylor approximations of order $\mathcal{O}(r^2)$ to compute the functionals for small r, thus avoiding integration errors.

7.3.2 Estimated Convergence Rates

In this section all computations were done with a regularization term of the form of Proposition 2.17 for $\alpha_\beta = 2$, making it smooth and thus accessible to computation, see Section 6.3.

To solve the appearing semilinear equations we choose the same parameters as in Section 7.1, i.e. we choose a relative tolerance of 10^{-15} for the semi-smooth Newton method, a global tolerance of 10^{-13}, a damping factor of 0.01 and a tolerance of 10^{-6} for the appearing multigrid solver for the linear system.

To solve the optimization problems we used a semi-smooth Newton method with global and relative tolerances of 10^{-5} and at most 50 steps. As globalization we used a trust region method a maximum of 10 line search steps, a damping factor of 0.8, a starting trust region radius of 1, a maximum trust region radius of 10. For the CG-method we used at most 20 iterations and global and relative tolerances of 10^{-5}

The predicted rate in for the regularization error is $\mathcal{O}(\gamma^{\frac{1}{4}})$ for small enough γ, as long as the error caused by h or k is not dominant, see Theorem 5.28. In Figure 7.5 one can see that the rates in the regularization term are sharp for the controls, but not for the states, which appear to have the convergence order of $\mathcal{O}(\gamma^{\frac{1}{2}})$.

We will see that this rate, as sharp as it may be, causes a major issue. Looking at Figure 7.6 we can see that the convergence behaviour for $h \to 0$ for the states seems fine. The rate looks similar to $\mathcal{O}(h^{1.5})$, which is at least worse than the rate from the solution operator itself, c.f. Theorem 4.65. But for the controls it is not possible to see anything similar, as one can easily see that the error curve flattens out basically immediately. As we can see that two curves for different k, but the same γ, almost overlap completely, we can conclude that γ is the culprit for the flattening out, i.e. $\mathcal{O}(\gamma^{\frac{1}{4}})$ is the dominant error. We also note that the curves for $\gamma = 10^{-6}$ and $\gamma = 10^{-7}$ are also close, indicating that the error in k is also comparatively large.

The effect is less pronounced for the error with respect to k. While the error in Figure 7.7 for the states flattens out quickly and gives no real hint about possible convergence rates, for the controls we get at least the hint that the error in the times is of order $\mathcal{O}(k^{0.8})$. This rate is better than the predicted rate of $\mathcal{O}(k^{0.5})$ from Theorem 5.28, keeping Remark 5.1 in mind. But it is also worse than the rate of the solution operator which is $\mathcal{O}(k)$, up to logarithmic factor, c.f. Theorem 4.65.

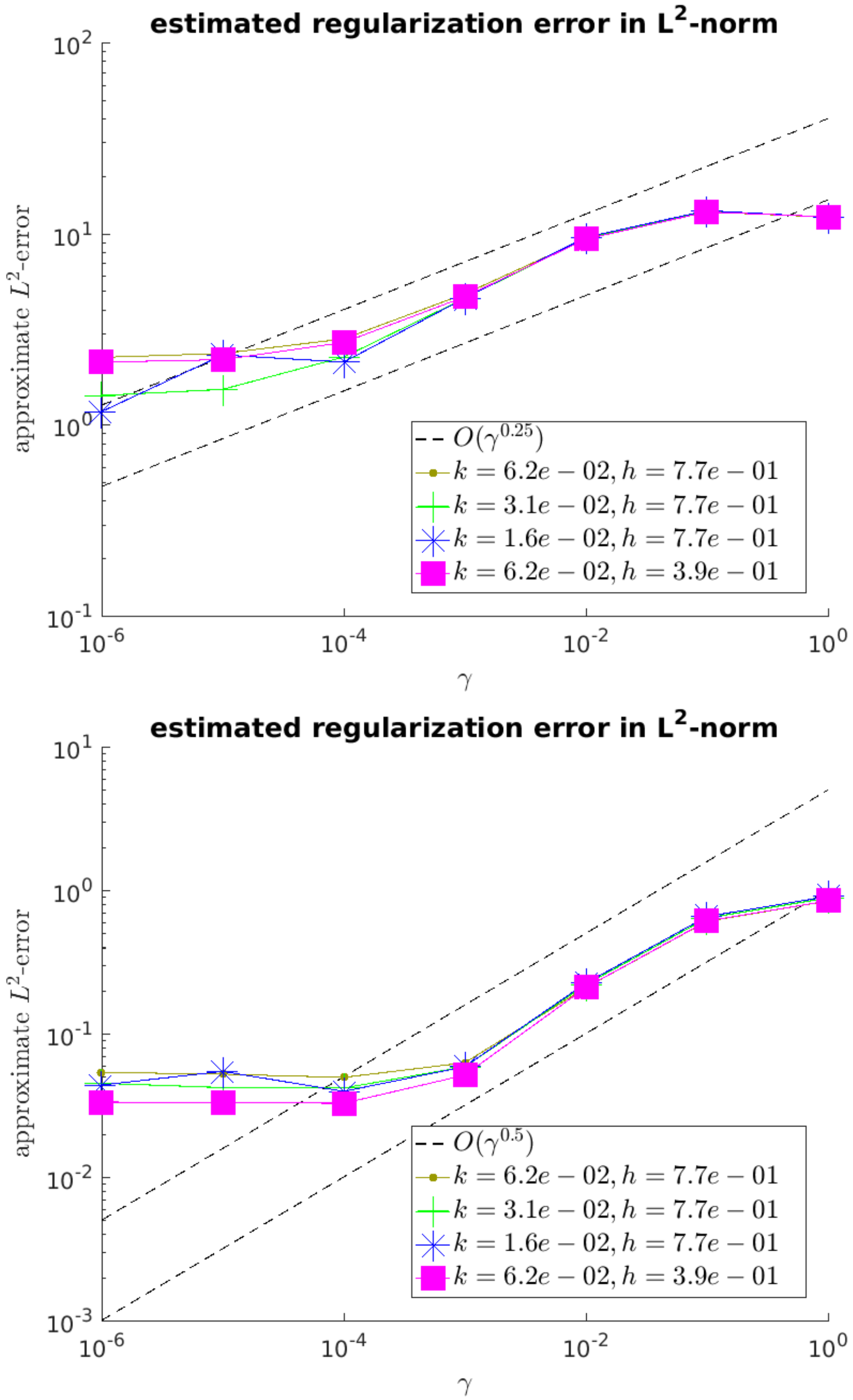


Figure 7.5: The behaviour of the errors in L^2-norm for the control (top) and the state (bottom) for decreasing regularization parameters.

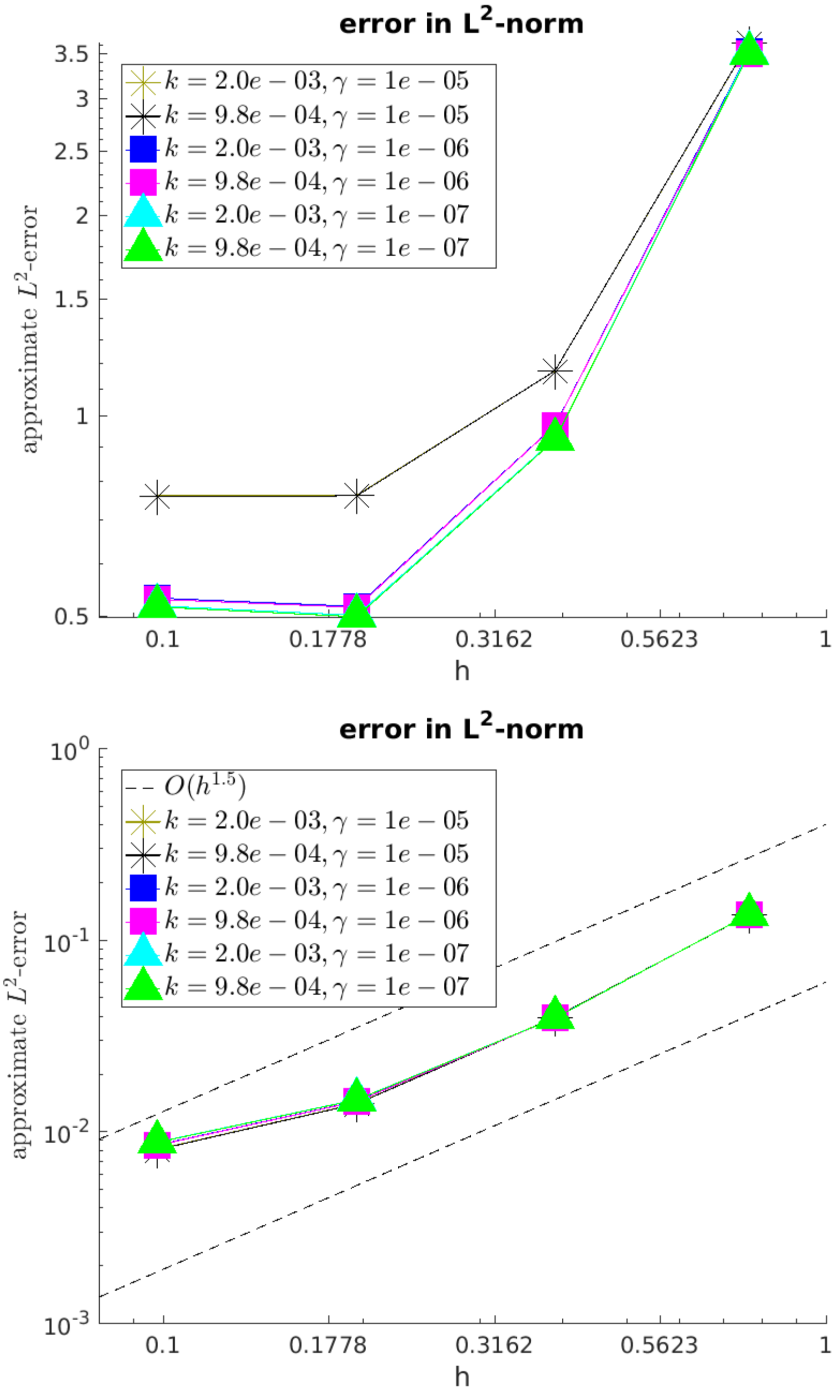


Figure 7.6: The behaviour of the errors in L^2-norm for the control (top) and the state (bottom) for decreasing h.

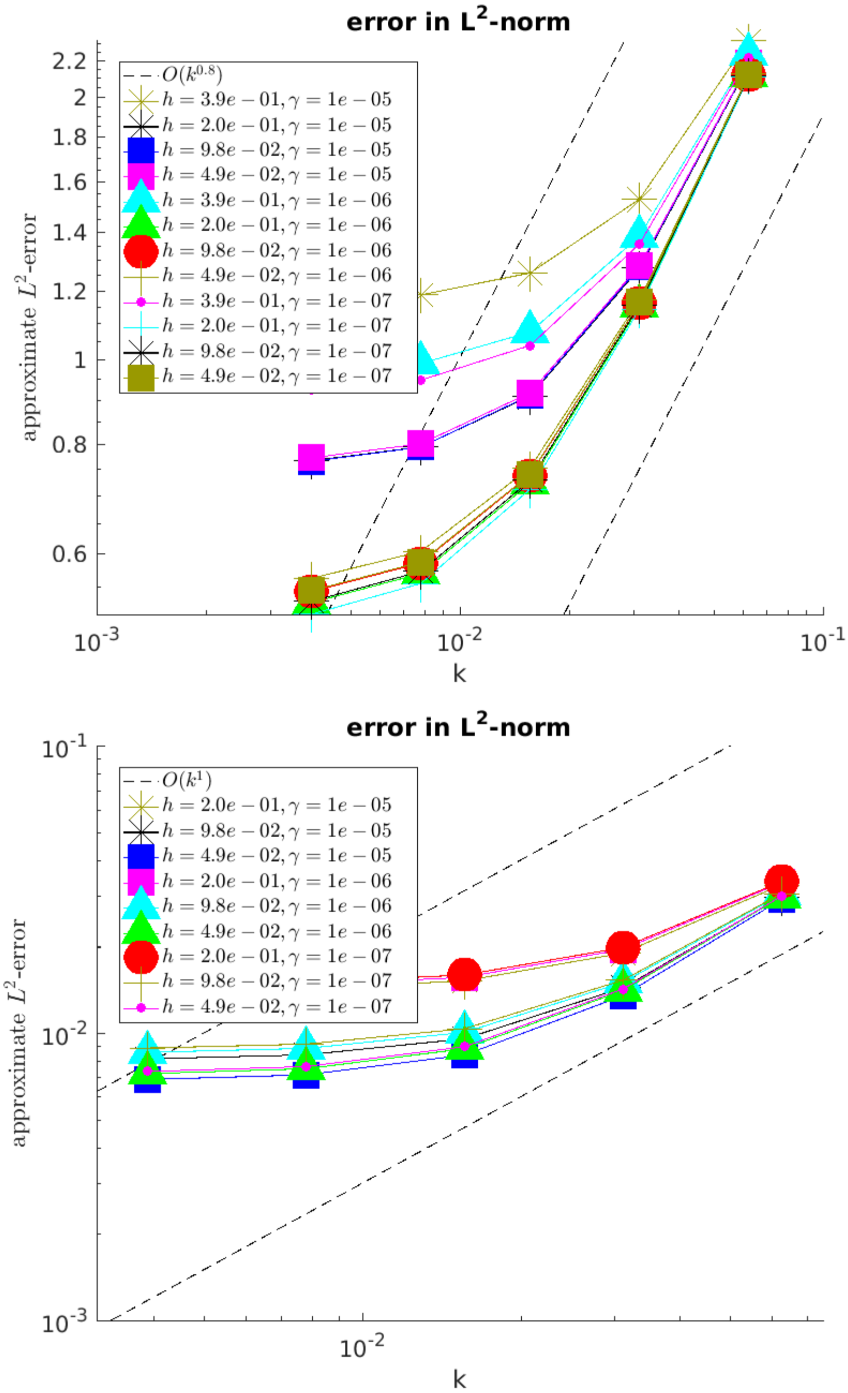


Figure 7.7: The behaviour of the errors in L^2-norm for the control (top) and the state (bottom) for decreasing k.

8 Appendix

8.1 Miscellaneous Auxiliary Results

8.1.1 Supports of Measures

This section just contains a minor additional result on measure spaces, but does not really fit in the preceeding sections.

Proposition 8.1 *Let $\Omega \subset \mathbb{R}^N$ be bounded and Borel-measurable, $A \subset \Omega$ be open in Ω and $\mu \in M(\Omega)$. Then $\mu|_A = 0$ iff $\operatorname{supp}(\mu) \subset \Omega \setminus A$.*

Proof. First let $\mu|_A = 0$ and $x \in A$. This immediately implies $|\mu|_A|(A) = 0$ as this is the norm on $M(A)$. Assume $x \in \operatorname{supp}(\mu)$. Then by the openness of A, there must be a $\epsilon > 0$ such that $B_\epsilon(x) \subset A$. Then we have $0 < |\mu|(B_\epsilon(x)) = |\mu|_A|(B_\epsilon(x)) = 0$. A contradiction.

Now let $A \subset \Omega \setminus \operatorname{supp}(\mu)$ be open in Ω. This implies $A \in \mathcal{B}(\Omega)$. We define $A_n := \{x \in A : \operatorname{dist}(x, \partial A) \geq \frac{1}{n}\}$. Because the distance function is continuous those sets are closed, and by the boundedness of Ω, even compact.

Let $n \in \mathbb{N}$. It is easy to check that $A_n \cap \mathbb{Q}^N$ is dense in A_n. We write this countable set as $\{x_1, x_2, \dots\}$. Because none of these points lie in $\operatorname{supp}(\mu)$ we find for each $j \in \mathbb{N}$ and open set $x \in U_j \subset \Omega$ such that $|\mu|(U_j) = 0$. By the compactness of A_n there exists a finite subcover $\cup_{j=1}^M U_j$ of the open covering $\cup_{j=1}^\infty U_j \supset A_n$. Here we obviously rearranged the indices for the sake of presentation. From this we conclude by the subadditivity of measures

$$|\mu|(A_n) \leq \sum_{j=1}^M |\mu|(U_j) = 0.$$

By the definition of regular measures, cf. [Rud74, Defintion 2.15], we have

$$|\mu|(A) = \sup_{\substack{K \subset A, \\ K \text{ is compact}}} |\mu|(K).$$

Its now very is easy to check that every compact $K \subset A$ lies in some A_n, because A is open. Therefore $|\mu|(K) \leq |\mu|(A_n) = 0$. Therefore the regularity implies $|\mu|(A) = 0$ which entails the claim. $\square$

8.1.2 Inequalities

This section just contains some technical inequalities which do not fit anywhere else.

Lemma 8.2 *Let $a, b, c \geq 0$ and $p > 1$. Assume $a^p \leq b a^{p-1} + c^p$. Then*

$$a \leq b + c.$$

Proof. Assume we have the converse $a > b + c$. Then we have

$$ba^{p-1} + c^p \geq a^p > a^{p-1}b + a^{p-1}c.$$

This implies $c^{p-1} > a^{p-1}$ and thus $c > a > b + c \geq c$. This is a contradiction. □

Lemma 8.3 *Let $C, T > 0$ and $f \in C([0,T])$ such that*

$$f(t)^2 \leq C\|f\|_{L^2(0,t)} + f(0)^2 \qquad \forall t \in [0,T]. \tag{8.1}$$

Then

$$\|f\|_{C([0,T])} \leq C\sqrt{T} + \frac{1+\sqrt{3}}{2}|f(0)|.$$

Proof. Integrating over both sides in (8.1) from 0 to $t \in (0,T)$ yields

$$\|f\|^2_{L^2(0,t)} \leq Ct\|f\|_{L^2(0,t)} + tf(0)^2.$$

The previous lemma now entails:

$$\|f\|_{L^2(0,t)} \leq Ct + \sqrt{t}|f(0)|.$$

Using this in (8.1) yields

$$f(t)^2 \leq C^2 t + C\sqrt{t}|f(0)| + |f(0)|^2 = \left(\sqrt{t}C + \frac{1}{2}|f(0)|\right)^2 + \frac{3}{4}|f(0)|^2.$$

Taking the square root on both sides and using the subadditivity of the root yields

$$|f(t)| \leq \sqrt{t}C + \frac{1}{2}|f(0)| + \frac{\sqrt{3}}{2}|f(0)|.$$

□

The following is an embedding where the constant in the estimate has been tracked. Note that the behaviour of the constant is sharp with respect to the inverse interval length. To see that consider constant functions.

Proposition 8.4 *Let $a < b \in \mathbb{R}$ and $f \in H^1((a,b))$. Then we have*

$$\|f\|^2_{L^\infty((a,b))} \leq 2\max((b-a),(b-a)^{-1})\|f\|^2_{H^1((a,b))}.$$

Proof. Let $f \in C^1((a,b))$. The general statement follows from density, e.g. [Ada75, Theorem 3.16], and the embedding $H^1((a,b)) \hookrightarrow L^\infty((a,b))$, e.g. [Ada75, Theorem 5.4].

Let $r \in (a,b)$. Then we find

$$|f(r)|^2 = (b-a)^{-1}\int_a^b |f(r)|^2\,ds = (b-a)^{-1}\int_a^b \left|\int_s^r f'(t)\,dt + f(s)\right|^2 ds.$$

Pulling the absolute further inside yields

$$\begin{aligned}|f(r)|^2 &\leq (b-a)^{-1}\int_a^b \left(\int_s^r |f'(t)|\,dt + |f(s)|\right)^2 ds \\ &\leq (b-a)^{-1}\int_a^b \left(\|f'\|_{L^1((a,b))} + |f(s)|\right)^2 ds \leq 2(b-a)^{-1}\int_a^b \|f'\|^2_{L^1((a,b))} + |f(s)|^2\,ds.\end{aligned}$$

Integrating each summand and using Hölder's inequality we now arrive at

$$\frac{1}{2}|f(r)|^2 \leq \|f'\|^2_{L^1((a,b))} + (b-a)^{-1}\|f\|^2_{L^2((a,b))} \leq (b-a)\|f'\|^2_{L^2((a,b))} + (b-a)^{-1}\|f\|^2_{L^2((a,b))}.$$

This yields the claim. □

Proposition 8.5 *Let $d > 0$. Then for any $N \in \mathbb{N}_{\geq 2}$ and $h > 0$ we have*

$$\int_0^d \frac{r^{N-2}}{r^N + h^N}\,dr \leq \begin{cases} \frac{2}{N}\arctan((hd)^{\frac{N}{2}})h^{-1} & \text{if } hd \leq 1, \\ N^{-1}(\frac{\pi}{2} + (hd)^{N-1} - 1)h^{-1} & \text{if } hd > 1. \end{cases} \tag{8.2}$$

In particular:

$$\int_0^d \frac{r^{N-2}}{r^N + h^N}\,dr \leq Ch^{-1} \tag{8.3}$$

where $C > 0$ only depends on an upper bound on hd and N.

Proof. We make the coordinate transformation $z = h^{-1}r$ so that we obtain:

$$\int_0^d \frac{r^{N-2}}{r^N + h^N}\,dr = \int_0^{hd} h^{-2}\frac{z^{N-2}}{z^N + 1}h\,dz = h^{-1}\int_0^{hd} \frac{z^{N-2}}{z^N + 1}\,dz.$$

If $dh \leq 1$ we have

$$\int_0^d \frac{r^{N-2}}{r^N + h^N}\,dr \leq h^{-1}\int_0^{hd} \frac{z^{\frac{N}{2}-1}}{z^N + 1}\,dz = h^{-1}\frac{2}{N}\arctan\left(z^{\frac{N}{2}}\right)|_0^{hd} = \frac{2}{N}\arctan((hd)^{\frac{N}{2}})h^{-1}.$$

If $dh > 1$ we have similarly

$$\begin{aligned}\int_0^d \frac{r^{N-2}}{r^N + h^N}\,dr &\leq h^{-1}\int_0^1 \frac{z^{\frac{N}{2}-1}}{z^N + 1}\,dz + h^{-1}\int_1^{hd} \frac{z^{N-2}}{z^N + 1}\,dz \\ &\leq \frac{2}{N}\arctan(1)h^{-1} + h^{-1}\int_1^{hd} z^{N-2}\,dz = \frac{\pi}{2N}h^{-1} + \frac{1}{N-1}((hd)^{N-1} - 1)h^{-1}.\end{aligned}$$

This implies the claimed bound in (8.2).

(8.3) is now an immediate consequence of (8.2). □

Proposition 8.6 *Let $z > 0$ then the Γ-function satisfies*

$$\Gamma(z) \leq 2e\Gamma(2z).$$

Proof. Let $z > 0$. The Γ-function is defined and estimated as follows:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt \leq \int_0^1 t^{z-1}\,dt + \int_1^\infty t^{2z-1}e^{-t}\,dt \leq z^{-1} + 2e\int_1^\infty t^{2z-1}e^{-t}\,dt.$$

We have

$$\int_0^1 t^{2z-1}e^{-t}\,dt \geq e^{-1}\int_0^1 t^{2z-1}\,dt = e^{-1}(2z)^{-1}.$$

Thus we conclude

$$\Gamma(z) \leq 2e\int_0^1 t^{2z-1}e^{-t} + 2e\int_1^\infty t^{2z-1}e^{-t}\,dt. = 2e\Gamma(2z).$$

□

Proposition 8.7 *The function $\bar{u} : (0,1) \times B_1(0) \to \mathbb{R}$, defined in (7.2) satisfies*

$$\|\bar{u}\|_{L^\infty((0,1)\times B_1(0))} \leq 60.$$

Proof. By the reformulation of $\bar{u}$ in (7.3) and abbreviating $r := \pi|x|$ we find for any $(t,x) \in (0,1) \times B_1(0)$ that

$$|\bar{u}(t,x)| \leq |\varphi'(t)||g(r)| + \pi^2|\varphi(t)| \left| g''(r) + r^{-1}g'(r) \right|.$$

We first estimate the terms with φ. We see

$$\varphi'(t) = \begin{cases} 96t - 384t^2 & \text{if } t \leq 0.25, \\ 0 & \text{if } 0.25 < t < 0.75, \\ -96(1-t) - 384(1-t)^2 & \text{if } 0.75 \leq t, \end{cases}$$

and

$$\varphi''(t) = \begin{cases} 96 - 768t & \text{if } t \leq 0.25, \\ 0 & \text{if } 0.25 < t < 0.75, \\ 96 + 768(1-t) & \text{if } 0.75 \leq t. \end{cases}$$

Due to symmetry it is enough to consider the case $t \leq 0.5$. If $t \in (0, 0.25)$ we have $\varphi''(t) = 0$ iff $t = 0.125$, thus only there can φ' have a local maximum. Thus

$$|\varphi'(t)| \leq \max(|\varphi'(0)|, |\varphi'(0.125)|, |\varphi'(0.25)|) = |\varphi'(0.125)| = 6.$$

By the same ideas we also have $|\varphi(t)| \leq 1$. Thus

$$|\bar{u}(t,x)| \leq 6|g(r)| + \pi^2|g''(r) + r^{-1}g'(r)|.$$

We first estimate the polynomial part of g. To that end define

$$g_p(r) := \left(1 - \frac{5}{16}\pi^2\right) + \frac{1}{2}r^2 - \frac{3}{16\pi^2}r^4.$$

We observe

$$g_p'(r) = r - \frac{3}{4\pi^2}r^3, \;\; g_p''(r) = 1 - \frac{9}{4\pi^2}r^2.$$

The roots of g_p' are given by 0 and $\pm\frac{2\pi}{\sqrt{3}}$. There are no roots in $(0,\pi)$ meaning that

$$|g_p(r)| \leq \max(|g_p(0)|, |g_p(\pi)|) = \frac{5}{16}\pi^2 - 1. \tag{8.4}$$

We also have the trigonometric part of g and see

$$g_t(r) := \cos(r), \quad -(g_t''(r) + r^{-1}g_t'(r)) = \cos(r) + r^{-1}\sin(r).$$

Thus by (8.4) we have $|g(r)| \leq \frac{5}{16}\pi^2$ and find

$$|\bar{u}(t,x)| \leq \frac{15}{8}\pi^2 + \pi^2|g''(r) + r^{-1}g'(r)|.$$

For the polynomial part we see

$$\left|g_p''(r) + r^{-1}g_p'(r)\right| = \left|2 - \frac{3}{\pi^2}r^2\right| \leq 2.$$

For the trigonometric part we see that

$$|g_t''(r) + r^{-1}g_t'(r)| \leq 1 + \max_{s\in[0,\pi]} |s^{-1}\sin(s)|$$

Setting derivative of $s^{-1}\sin(s)$ to 0 (for $s \in (0, \pi)$) we find

$$0 = \frac{-s\cos(s) + \sin(s)}{s^2} \iff 0 = s - \tan(s).$$

Here we used that $s = \frac{\pi}{2}$ is clearly not a solution so that we may divide by $\cos(s)$. Numerically it is easy to see that this has no solutions in $(0, \pi)$. Thus

$$|g_t''(r) + r^{-1}g_t'(r)| \leq 1 + \max\left((\lim_{t\to 0} |t^{-1}\sin(t)|, \left|\frac{2}{\pi}\sin\left(\frac{\pi}{2}\right)\right|, |\pi^{-1}\sin(\pi)| \right) = 2.$$

This finally results in

$$|\bar{u}(t,x)| \leq \frac{15}{8}\pi^2 + 4\pi^2 \leq 6\pi^2 < 60.$$

□

8.1.3 Compact Embeddings of Hölder Spaces

The following statement is well-known, when the Hölder exponents κ_I, κ_Ω are the same.

Lemma 8.8 *Let $I = (0,T)$ for some $T > 0$, $\Omega \subset \mathbb{R}^N$ for some $N \geq 1$ and $\kappa_I, \kappa_\Omega \in (0,1]$. Then*

$$C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega)) \hookrightarrow\hookrightarrow C^{\kappa_I'}(I, C^{\kappa_\Omega'}(\Omega))$$

for any $\kappa_I' \in (0, \kappa_I)$, $\kappa_\Omega' \in (0, \kappa_\Omega)$.

Proof. Let $(y_n)_{n\in\mathbb{N}} \subset C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ be a sequence bounded by some $C > 0$. We see that for any $(t,x),(s,y) \in I \times \Omega$ we have

$$\begin{aligned} |y_n(t,x) - y_n(s,y)| &\leq |y_n(t,x) - y_n(s,x)| + |y_n(s,x) - y_n(s,y)| \\ &\leq \|y_n(t,\cdot) - y_n(s,\cdot)\|_{L^\infty(\Omega)} + |x-y|^{\kappa_\Omega}|y_n(s)|_{C^{\kappa_\Omega}(\Omega)} \\ &\leq |t-s|^{\kappa_I}\|y_n\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + |x-y|^{\kappa_\Omega}\|y_n(s)\|_{C^{\kappa_\Omega}(\Omega)} \\ &\leq (|t-s|^{\kappa_I} + |x-y|^{\kappa_\Omega})C. \end{aligned}$$

Hence the sequence $(y_n)_{n\in\mathbb{N}}$ is equicontinuous. Now the theorem of Arzela-Ascoli, e.g. [Ş03, Theorem 5.7.8], implies that the sequence has an accumulation point y with respect to the $L^\infty(I \times \Omega)$-norm. The converging subsequence is denoted by the same indices for simplicity.

We now show $y \in C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$. Let $s,t \in I$ with $t \neq s$ and $x,y \in \Omega$ with $x \neq y$. We see

$$\begin{aligned} &\frac{|(y(t,\cdot) - y(s,\cdot))(x) - (y(t,\cdot) - y(s,\cdot))(y)|}{|x-y|^{\kappa_\Omega}} \\ &= \lim_{n\to\infty} \frac{|(y_n(t,\cdot) - y_n(s,\cdot))(x) - (y_n(t,\cdot) - y_n(s,\cdot))(y)|}{|x-y|^{\kappa_\Omega}} \\ &\leq \limsup_{n\to\infty} |y_n(t,\cdot) - y_n(s,\cdot)|_{C^{\kappa_\Omega}(\Omega)}. \end{aligned}$$

Thus, taking the supremum over all $x \neq y$ and using the uniform convergence we find

$$\|y(t,\cdot) - y(s,\cdot)\|_{C^{\kappa_\Omega}(\Omega)} \le \limsup_{n\to\infty} \|y_n(t,\cdot) - y_n(s,\cdot)\|_{C^{\kappa_\Omega}(\Omega)}.$$

Dividing both sides by $|t-s|^{\kappa_I}$ yields

$$\begin{aligned}\frac{\|y(t,\cdot) - y(s,\cdot)\|_{C^{\kappa_\Omega}(\Omega)}}{|t-s|^{\kappa_I}} &\le \limsup_{n\to\infty} \frac{\|y_n(t,\cdot) - y_n(s,\cdot)\|_{C^{\kappa_\Omega}(\Omega)}}{|t-s|^{\kappa_I}} \\ &\le \limsup_{n\to\infty} |y_n|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} \le C.\end{aligned}$$

Taking the supremum over all $t \neq s$ yields

$$|y|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} \le C.$$

It remains to estimate $\|y\|_{L^\infty(I, C^{\kappa_\Omega}(\Omega))}$. This can be done completely analogously. So, in total we have $\|y\|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} \le C$.

It remains to show $y_n \xrightarrow{n\to\infty} y$ in $C^{\kappa_I'}(I, C^{\kappa_\Omega'}(\Omega))$. Let $\epsilon \in (0,1)$, $\kappa_I' = (1-\epsilon)\kappa_I$ and $\kappa_\Omega' = (1-\epsilon)\kappa_\Omega$. We argue the general case at the end. Due to the just proven regularity of y we may subtract it from $(y_n)_{n\in\mathbb{N}}$ and assume without loss of generality $y = 0$. Then we have for any $t, s \in I$ that

$$\begin{aligned}|y_n(t,\cdot) - y_n(s,\cdot)|_{C^{\kappa_\Omega'}(\Omega)} &= \sup_{\substack{x,y\in\Omega,\\ x\neq y}} \frac{|(y_n(t,\cdot) - y_n(s,\cdot))(x) - (y_n(t,\cdot) - y_n(s,\cdot))(y)|}{|x-y|^{\kappa_\Omega'}} \\ &= \sup_{\substack{x,y\in\Omega,\\ x\neq y}} \left(\frac{|(y_n(t,\cdot) - y_n(s,\cdot))(x) - (y_n(t,\cdot) - y_n(s,\cdot))(y)|}{|x-y|^{\kappa_\Omega}} \right)^{\frac{\kappa_\Omega'}{\kappa_\Omega}} \\ &\qquad \cdot |(y_n(t,\cdot) - y_n(s,\cdot))(x) - (y_n(t,\cdot) - y_n(s,\cdot))(y)|^{1-\frac{\kappa_\Omega'}{\kappa_\Omega}} \\ &\le 4 |y_n(t,\cdot) - y_n(s,\cdot)|_{C^{\kappa_\Omega}(\Omega)}^{\frac{\kappa_\Omega'}{\kappa_\Omega}} \|y_n\|_{L^\infty(I\times\Omega)}^{1-\frac{\kappa_\Omega'}{\kappa_\Omega}}.\end{aligned} \tag{8.5}$$

By assumption $(y_n)_{n\in\mathbb{N}}$ is bounded in $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ by C. Thus, we can conclude

$$\begin{aligned}|y_n(t,\cdot) - y_n(s,\cdot)|_{C^{\kappa_\Omega'}(\Omega)} &\le 4 \|y_n\|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))}^{\frac{\kappa_\Omega'}{\kappa_\Omega}} \left(|t-s|^{\kappa_I}\right)^{\frac{\kappa_\Omega'}{\kappa_\Omega}} \|y_n\|_{L^\infty(I\times\Omega)}^{1-\frac{\kappa_\Omega'}{\kappa_\Omega}} \\ &\le 4(C+1)|t-s|^{\kappa_I'} \|y_n\|_{L^\infty(I\times\Omega)}^{\epsilon}.\end{aligned}$$

Here we used $\frac{\kappa_I}{\kappa_\Omega} - \kappa_\Omega' = \kappa_I'$ by their specific definitions. Also

$$\begin{aligned}\frac{\|y_n(t,\cdot) - y_n(s,\cdot)\|_{L^\infty(\Omega)}}{|t-s|^{\kappa_I'}} &= \left(\frac{\|y_n(t,\cdot) - y_n(s,\cdot)\|_{L^\infty(\Omega)}}{|t-s|^{\kappa_I}} \right)^{\frac{\kappa_I'}{\kappa_I}} \|y_n(t,\cdot) - y_n(s,\cdot)\|_{L^\infty(\Omega)}^{1-\frac{\kappa_I'}{\kappa_I}} \\ &\le 2(C+1)\|y_n\|_{L^\infty(I\times\Omega)}^{\epsilon}.\end{aligned}$$

Combining these two estimates yields

$$\frac{\|y_n(t,\cdot) - y_n(s,\cdot)\|_{C^{\kappa_\Omega'}(\Omega)}}{|t-s|^{\kappa_I'}} \le 6(C+1)\|y_n\|_{L^\infty(I\times\Omega)}^{\epsilon}.$$

Thus $|y_n|_{C^{\kappa_I'}(I, C^{\kappa_\Omega'}(\Omega))} \le 6(C+1)\|y_n\|_{L^\infty(I\times\Omega)}^{\epsilon}$. Just as in (8.5) one can show the estimate $\|y_n\|_{L^\infty(I, C^{\kappa_\Omega}(\Omega))} \le 2(C+1)\|y_n\|_{L^\infty(I\times\Omega)}^{\epsilon}$. Thus

$$\|y_n\|_{C^{\kappa_I'}(I, C^{\kappa_\Omega'}(\Omega))} \le 8(C+1)\|y_n\|_{L^\infty(I\times\Omega)}^{\epsilon}. \tag{8.6}$$

This now finally converges to 0 as $\epsilon > 0$.

Now, let $\kappa_I' \in (0, \kappa_I)$ and $\kappa_\Omega' \in (0, \kappa_\Omega)$. We define $\epsilon := \min\left(1 - \frac{\kappa_I'}{\kappa_I}, 1 - \frac{\kappa_\Omega'}{\kappa_\Omega}\right) > 0$. We then have $(1-\epsilon)\kappa_I \in [\kappa_I', \kappa_I)$ and $(1-\epsilon)\kappa_\Omega \in [\kappa_\Omega', \kappa_\Omega)$ and can immediately conclude

$$\begin{aligned}\|y_n\|_{C^{\kappa_I'}(I, C^{\kappa_\Omega'}(\Omega))} &\leq \operatorname{diam}(\Omega)^{(1-\epsilon)\kappa_\Omega - \kappa_\Omega'} \|y_n\|_{C^{\kappa_I'}(I, C^{(1-\epsilon)\kappa_\Omega}(\Omega))} \\ &\leq (\operatorname{diam}(\Omega) + 1) \|y_n\|_{C^{\kappa_I'}(I, C^{(1-\epsilon)\kappa_\Omega}(\Omega))}.\end{aligned}$$

Doing the same for the time variable and using (8.6) we can finally conclude

$$\begin{aligned}\|y_n\|_{C^{\kappa_I'}(I, C^{\kappa_\Omega'}(\Omega))} &\leq (T+1)(\operatorname{diam}(\Omega) + 1) \|y_n\|_{C^{(1-\epsilon)\kappa_I}(I, C^{(1-\epsilon)\kappa_\Omega}(\Omega))} \\ &\leq 8(T+1)(\operatorname{diam}(\Omega) + 1)(C+1) \|y_n\|_{L^\infty(I \times \Omega)}^{\epsilon}.\end{aligned}$$

Sending $n \to \infty$ yields the claim in the general case. □

8.1.4 Boundary value preserving density

The following density result is very specific, which is unfortunately necessary for the previous proof of Proposition 4.18.

Proposition 8.9 *Let $\Omega \subset \mathbb{R}^N$ be a $C^{2,\tilde{\alpha}}$-domain for some $\tilde{\alpha} > 0$. Then $C^{2,\tilde{\alpha}}(\bar{\Omega}) \cap H_0^1(\Omega)$ is dense in $H^2(\Omega) \cap H_0^1(\Omega)$ with respect to $\|\cdot\|_{H^2(\Omega)}$.*

Proof. Let $v \in H^2(\Omega) \cap H_0^1(\Omega)$ be given and $\epsilon > 0$ be arbitrary. Since $C_c^\infty(\Omega)$ is dense in $L^2(\Omega)$, e.g. [Ada75, Theorem 2.19], there exists a $f_\epsilon \in C_c^\infty(\Omega)$ such that $\|(-\Delta v) - f_\epsilon\|_{L^2(\Omega)} \leq \epsilon$. By [GT01, Theorem 6.14] there now exists a unique solution $v_\epsilon \in C^{2,\tilde{\alpha}}(\bar{\Omega}) \cap H_0^1(\Omega)$ to the PDE $-\Delta v_\epsilon = f_\epsilon$ on Ω, $v_\epsilon|_{\partial\Omega} = 0$. By higher elliptic regularity, e.g. [Gri11, Lemma 2.4.2.1 and the remarks thereafter, Theorem 2.4.2.5], we therefore find

$$\|v - v_\epsilon\|_{H^2(\Omega)} \leq C \|(-\Delta v) - f_\epsilon\|_{L^2(\Omega)} \leq C\epsilon.$$

As $\epsilon > 0$ was arbitrary this shows the claim. □

8.2 Statements on Bochner Spaces

The following proposition is used in the proof of second order sufficient conditions in Section 3.5.

Proposition 8.10 *Let $N \in \{1,2,3\}$. Let $\Omega \subset \mathbb{R}^N$ be a bounded domain satisfying the cone condition and $I := (0,T)$ for some $T > 0$. We have $L^2(I, H^1(\Omega)) \cap C(\bar{I}, L^2(\Omega)) \subset L^4(L^{\frac{4N}{N+2}}(\Omega))$. Even stronger there are $C > 0$ and $\theta \in (0,1)$ such that*

$$\|y\|_{L^4(L^{\frac{4N}{N+2}}(\Omega))} \leq C \|y\|_{C(\bar{I}, L^2(\Omega))}^{1-\theta} \|y\|_{L^2(I, H^1(\Omega))}^{\theta}$$

holds for any $y \in C(\bar{I}, L^2(\Omega)) \cap L^2(I, H^1(\Omega))$.

Proof. We show that we have the embedding

$$\Big(L^2(I,H^1(\Omega)),C(\bar{I},H)\Big)_{\theta,4} \hookrightarrow L^4(L^{\frac{4N}{N+2}}(\Omega))$$

for some $\theta \in (0,1)$. Then [Tri78, Theorem 1.3.3g)] delivers the desired norm estimate.

We only treat the case $N=3$. The other cases are similar but easier as $H^1(\Omega)$ then embedds into higher order Lebesgue spaces.

We have the Sobolev embedding $H^1(\Omega) \hookrightarrow L^{\frac{2N}{N-2}}(\Omega)$. For any $r \in [1,\infty)$ we also have $C(\bar{I},L^2(\Omega)) \hookrightarrow L^r(I,L^2(\Omega))$. Thus we find for any $\theta \in (0,1)$, checking the definition of the interpolation spaces,

$$\Big(L^2(I,H^1(\Omega)),C(\bar{I},L^2(\Omega))\Big)_{\theta,4} \hookrightarrow \Big(L^2(I,L^{\frac{2N}{N-2}}(\Omega)),L^r(I,L^2(\Omega))\Big)_{\theta,4}.$$

We now let $r \in (2,\infty)$ and $\theta_r \in (1/2,1)$ such that

$$\frac{1}{4} = \frac{1-\theta_r}{2} + \frac{\theta_r}{r}. \tag{8.7}$$

Then [Tri78, Theorem 1.18.4] delivers the isomorphism

$$\begin{aligned}\Big(L^2(I,H^1(\Omega)),C(\bar{I},L^2(\Omega))\Big)_{\theta_r,4} &\hookrightarrow \Big(L^2(I,L^{\frac{2N}{N-2}}(\Omega)),L^r(I,L^2(\Omega))\Big)_{\theta_r,4}\\ &\simeq L^4(I,(L^{\frac{2N}{N-2}}(\Omega),L^2(\Omega))_{\theta_r,4}).\end{aligned}$$

By [Tri78, Theorem 1.3.3e)] we have for any $\epsilon_r > 0$ small enough such that $\theta_r + \epsilon_r < 1$:

$$(L^{\frac{2N}{N-2}}(\Omega),L^2(\Omega))_{\theta_r,4} \hookrightarrow (L^{\frac{2N}{N-2}}(\Omega),L^2(\Omega))_{\theta_r+\epsilon_r,\frac{4N}{N+2}}.$$

Using [Tri78, Theorem 4.3.1.1 and its reference to Theorem 2.4.2.2 (10)] we have

$$(L^{\frac{2N}{N-2}}(\Omega),L^2(\Omega))_{\theta_r,4} \hookrightarrow L^{\frac{4N}{N+2}}(\Omega) \tag{8.8}$$

provided $\theta_r + \epsilon_r$ satisfies

$$\frac{1}{\left(\frac{4N}{N+2}\right)} = \frac{1-(\theta_r+\epsilon_r)}{\left(\frac{2N}{N-2}\right)} + \frac{\theta_r+\epsilon_r}{2}.$$

In turn, this is equivalent to

$$\frac{N+2}{2N} = \frac{(1-(\theta_r+\epsilon_r))(N-2)}{N} + (\theta_r+\epsilon_r) = 1 - \frac{2}{N}(1-(\theta_r+\epsilon_r)).$$

This is equivalent to

$$\begin{aligned} N+2 &= 2N - 4(1-(\theta_r+\epsilon_r))\\ \iff 6 - 4(\theta_r+\epsilon_r) &= N\\ \iff \theta_r+\epsilon_r &= \frac{6-N}{4}.\end{aligned} \tag{8.9}$$

As $(6-N)/4 = 3/4$ we can choose r large enough in (8.7) to have $\theta_r = 2.5/4$. Choosing $\epsilon_r := 0.5/4$ we have satisfied both (8.7) and (8.9) and thus have obtained (8.8). This concludes the proof. □

The following statement is very similar to Proposition 8.10 for slightly different spaces. The proof is very similar, so we only give a rough sketch. It appears for example in discussion of second order conditions in Section 3.6.

Proposition 8.11 *Let $N \in \{1,2,3\}$. Let $\Omega \subset \mathbb{R}^N$ be a bounded domain satisfying the cone condition and $I := (0,T)$ for some $T > 0$. We have $L^2(I,H^1(\Omega)) \cap C(\bar{I},L^2(\Omega)) \subset L^3(Q)$. Even stronger there are $C > 0$ and $\theta \in (0,1)$ such that*

$$\|y\|_{L^3(Q)} \leq C\|y\|_{C(\bar{I},L^2(\Omega))}^{1-\theta}\|y\|_{L^2(I,H^1(\Omega))}^{\theta}$$

holds for any $y \in C(\bar{I},L^2(\Omega)) \cap L^2(I,H^1(\Omega))$.

Proof. As in the proof of Proposition 8.10 we obtain the embedding

$$\left(L^2(I,H^1(\Omega)),C(\bar{I},L^2(\Omega))\right)_{\theta_r,4} \hookrightarrow L^3\left(I,\left(L^{\frac{2N}{N-2}}(\Omega),L^2(\Omega)\right)_{\theta_r,3}\right)$$

for any $r \in (2,\infty)$ and $\theta_r \in (0,1)$ satisfying

$$\frac{1}{3} = \frac{1-\theta_r}{2} + \frac{\theta_r}{r}. \tag{8.10}$$

We also obtain

$$\left(L^{\frac{2N}{N-2}}(\Omega),L^2(\Omega)\right)_{\theta_r,3} \hookrightarrow L^3(\Omega)$$

provided we have

$$\frac{1}{3} = \frac{1-\theta_r}{\frac{2N}{N-2}} + \frac{\theta_r}{2}. \tag{8.11}$$

We have to find $r \in (2,\infty)$ and $\theta_r \in (0,1)$ such that (8.10) and (8.11) are both satisfied. Reforming (8.11) yields the equivalent forms

$$\begin{aligned}\frac{1}{3} = \frac{N-2}{2N} + \theta_r\left(\frac{N}{2N} - \frac{N-2}{2N}\right) &\iff \frac{2N}{6N} - \frac{3N-6}{6N} = \theta_r\frac{1}{N}\\ &\iff -\frac{N}{6} + 1 = \theta_r.\end{aligned}$$

For $N \leq 3$ this is compatible with $\theta_r \in (0,1)$. Inserting this into (8.10) we have

$$\begin{aligned}\frac{1}{3} = \frac{N}{12} + \left(1 - \frac{N}{6}\right)\frac{1}{r} &\iff \frac{\frac{4}{12} - \frac{N}{12}}{1 - \frac{N}{6}} = \frac{1}{r}\\ &\iff \frac{1}{2}\frac{4-N}{6-N} = \frac{1}{r}\\ &\iff r = 2\frac{6-N}{4-N}.\end{aligned}$$

For $N \leq 3$ this is again well defined and we can conclude the proof. □

Proposition 8.12 *Let $I \subset \mathbb{R}$ be an open interval, V a Hilbert space and $T\colon V \to V$ a linear and continuous operator. Let $y \in H^1(I,V)$ then $Ty \in H^1(I,V)$ with $\partial_t(Ty) = T\partial_t y$. We also have the following estimate*

$$\|Ty\|_{H^1(I,V)} \leq \|T\|_{\mathcal{L}(V,V)}\|y\|_{H^1(I,V)}.$$

Proof. Let $\varphi \in C_c^\infty(I)$ and $v \in V$ be arbitrary test functions. Then

$$\begin{aligned}\left(\int_I T(\partial_t y)(t)\varphi(t)\,dt, v\right)_V &= \int_I \varphi(t)\left(T(\partial_t y)(t), v\right)_V\,dt\\ &= \int_I \varphi(t)\left(\partial_t y(t), T^*v\right)_V\,dt = \left(\int_I \varphi(t)\partial_t y(t)\,dt, T^*v\right)_V.\end{aligned}$$

By the definition of the Bochner derivative this is in turn equal to

$$\left(-\int_I \varphi'(t)y(t)\,dt, T^*v\right)_V = \left(-\int_I \varphi'(t)Ty(t)\,dt, v\right)_V.$$

As v was arbitrary this implies

$$\int_I T(\partial_t y)(t)\varphi(t)\,dt = -\int_I \varphi'(t)Ty(t)\,dt.$$

By the definition of the Bochner derivative this implies $\partial_t(Ty) = T\partial_t y$. □

Proposition 8.13 *Let $I \subset \mathbb{R}$ be an open interval and (V, H, V^*) a Gelfand triple of Hilbert spaces. Then the space $H^1(I, V) \cap W_0(I)$ is dense in $W_0(I)$ with respect to the $W(I)$-norm.*

Proof. Let $f \in W_0(I)$. We show that f can be approximated by some $(f_\epsilon)_{\epsilon>0} \subset W_0(I)$ such that $f|_{(0,\epsilon)} = 0$. For $\epsilon > 0$ define the following for $t \in I$

$$\rho_\epsilon(t) := \begin{cases} 0 & \text{if } t \in (0, \epsilon),\\ \epsilon^{-1}(t-\epsilon) & \text{if } t \in (\epsilon, 2\epsilon),\\ 1 & \text{if } t \in (2\epsilon, T),\end{cases} \qquad \in W^{1,\infty}(I).$$

Furthermore, define $f_\epsilon := f\,\rho_\epsilon$. It satisfies

$$\|f_\epsilon - f\|^2_{L^2(I,V)} = \int_0^{2\epsilon} \|f(t)\rho_\epsilon(t) - f(t)\|_V^2\,dt = \int_0^{2\epsilon} |\rho_\epsilon(t) - 1|^2 \|f(t)\|_V^2\,dt.$$

Because $|\rho_\epsilon(t) - 1| \in [0, 1]$ we have

$$\|f_\epsilon - f\|_{L^2(I,V)} \le \|f\|_{L^2((0,2\epsilon),V)}. \tag{8.12}$$

By the theorem of dominated convergence, e.g. [BK15, Proposition 5.4], the term converges to 0 for $\epsilon \to 0$. The derivatives are treated similarly, but are slightly more complicated.

$$\|\partial_t(f_\epsilon) - \partial_t(f)\|_{L^2(I,V^*)} \le \|\rho_\epsilon \partial_t f - \partial_t f\|_{L^2(I,V^*)} + \|f\rho_\epsilon'\|_{L^2(I,V^*)}.$$

The first term converges to 0 by the same arguments used to obtain (8.12). The second term equals by the definition of ρ_ϵ

$$\begin{aligned}\|f\rho_\epsilon'\|^2_{L^2(I,V^*)} &= \int_\epsilon^{2\epsilon} \|f(t)\|_{V^*}^2 \epsilon^{-2}\,dt = \epsilon^{-2}\int_0^{2\epsilon} \left\|\int_0^t \partial_t f(s)\,ds\right\|_{V^*}^2 dt\\ &\le \epsilon^{-2}\int_0^{2\epsilon}\left(\int_0^t \|\partial_t f(s)\|_{V^*}\,ds\right)^2 dt.\end{aligned}$$

By Hölder's inequality we get

$$\|f\rho_\epsilon'\|^2_{L^2(I,V^*)} \le \epsilon^{-2}\int_0^{2\epsilon} t\int_0^t \|\partial_t f(s)\|_{V^*}^2\,ds\,dt \le 4\int_0^{2\epsilon} \|\partial_t f(s)\|_{V^*}^2\,ds.$$

As before this converges to 0 for $\epsilon \to 0$.

Now, each f_ϵ can be approximated by smooth functions in $W(I)$. This can be done analogously to the proof of [Wlo92, Lemma 25.1]. □

The following is an almost verbatim quote of [CV19, Lemma A.1], which is still in preparation and not published as of yet, so in particular its name in the paper might change. We obtained the result via private communication and want to stress that is not our work and it is presented to not leave any gaps. The only change we made is to adapt the notation.

Lemma 8.14 (Stampacchia's Lemma for Bochner-Sobolev Functions) *Suppose that a $T > 0$, a domain $\Omega \subset \mathbb{R}^N$, $N \geq 1$, and a $v \in L^2((0,T), H^1(\Omega))$ are given. Define $v^+ := \max(0, v)$, where $\max(0,\cdot)$ acts pointwise a.e. in $(0,T) \times \Omega$. Then, the function v^+ is an element of $L^2(0,T,H^1(\Omega))$, the gradient $\nabla(v^+) \in L^2((0,T), L^2(\Omega, \mathbb{R}^N)) \simeq L^2((0,T) \times \Omega, \mathbb{R}^N)$ of v^+ satisfies*

$$\nabla(v^+) = \begin{cases} \nabla v & a.e.\ in\ \{v > 0\} \\ 0 & a.e.\ in\ \{v \leq 0\} \end{cases}, \tag{A.1}$$

and we have

$$\nabla v = 0 \ a.e.\ in\ \{v = 0\}. \tag{A.2}$$

If, further, the function v additionally possesses $H^1((0,T), L^2(\Omega))$-regularity, then v^+ is also an element of $H^1((0,T), L^2(\Omega))$ and it holds

$$\partial_t(v^+) = \begin{cases} \partial_t v & a.e.\ in\ \{v > 0\} \\ 0 & a.e.\ in\ \{v \leq 0\} \end{cases}, \tag{A.3}$$

as well as

$$\partial_t v = 0 \ a.e.\ in\ \{v = 0\}, \tag{A.4}$$

and if v is even in $L^2(0,T; H^2(\Omega))$, then the Hessian $\nabla^2 v \in L^2((0,T), L^2(\Omega, \mathbb{R}^{N\times N})) \simeq L^2((0,T) \times \Omega, \mathbb{R}^{N\times N})$ of v satisfies

$$\nabla^2 v = 0 \ a.e.\ in\ \{v = 0\}. \tag{A.5}$$

As always $\{v > 0\} := \{(t,x) \in Q : v(t,x) > 0\}$ and similar sets are defined for a fixed, but arbitrary representative of v.

Proof. The $L^2((0,T), H^1(\Omega))$-regularity of v^+ follows straightforwardly from the results in [ABM14, Section 5.8] and [HKST15, Section 3], the formula (A.1) can be established completely analogously to [ABM14, Theorem 5.8.2], and to obtain (A.2), it suffices to note that (A.1) and the lineartiy of the operator ∇ yield

$$0 = \nabla v - \nabla(v^+ + v^-) = \nabla v - \nabla(v^+) - \nabla(v^-) = \begin{cases} \nabla v & \text{a.e. in } \{v = 0\} \\ 0 & \text{a.e. in } \{v \neq 0\} \end{cases},$$

where v^- is short for $\min(0, v) = -\max(0, -v)$. This proves the first part of the lemma. Let us assume now that v is also an element of $H^1((0,T), L^2(\Omega))$. Then, the $H^1((0,T), L^2(\Omega))$-regularity of v^+ and the formula (A.3) follow straightforwardly from [Woo07, Corollary 2.3, Equation (2)], and the derivation (A.4) is completely along the lines of that of (A.2). It remains to establish (A.5). To this end, we first note that, for every $v \in L^2((0,T), H^2(\Omega))$, we have (due to (A.2) and since $\partial_n v \in L^2((0,T), H^1(\Omega))$ holds for all spatial partial derivatives $\partial_n v$, $n = 1, \ldots, N$)

$$(\partial_n v) 1_{\{v=0\}} = 0 \in L^2((0,T) \times \Omega) \qquad \forall n = 1, \ldots, N$$

and

$$(\partial_m \partial_n v) 1_{\{\partial_n v = 0\}} = 0 \in L^2((0,T) \times \Omega) \qquad \forall m, n = 1, \dots, N.$$

The above implies in particular that

$$1_{\{v=0\}} = 1_{\{\partial_n v = 0\}} 1_{\{v=0\}} \in L^2((0,T) \times \Omega) \qquad \forall n = 1, \dots, N$$

This establishes (A.5) and completes the proof. □

8.3 A Parabolic Maximum Principle

The following theorem and its proof are a generalization of [USL88, Theorem III.7.2]. A difficulty in the proof is that $y \in W(I)$ does not imply that $y^+ \notin W(I)$.

Theorem 8.15 *Let $\Omega \subset \mathbb{R}^N$ be open and bounded, $T > 0$, $I := (0,T)$ and $Q := I \times \Omega$. Let $\Gamma_D \subset \partial\Omega$ and $\Sigma_D := I \times \Gamma_D$ be the Dirichlet boundary portions. Let V, H and $W(I)$ be as in Assumption 2.5. Let $u, f \in L^2(Q)$ with $u, f \geq 0$ a.e. in Q. Let $y_0 \in L^2(\Omega)$ with $y_0 \geq 0$ a.e. in Ω. Let $y \in W(I)$ be a solution to*

$$\begin{cases} \partial_t y + Ay + fy = u, \\ y(0) = y_0, \quad y|_{\Sigma_D} = 0. \end{cases}$$

Then we have $y \geq 0$ a.e. in Q.

Proof. We define $y^- := \min(0, y) \in L^2(I, V) \cap C(\bar{I}, H)$. Note that y^- has this regularity by Proposition 8.19 and $y \in W(I) \subset C(\bar{I}, H)$. Then we test the equation for y with $y^- \cdot 1_{(0,t)}$ for $t \in I$ to obtain

$$\int_0^t \left(\partial_t y, y^-\right)_{V^*,V} \, dt + a_{(0,t)}(y, y^-) = \left(u - fy, y^-\right)_{L^2((0,t)\times\Omega)}. \tag{8.13}$$

By Proposition 8.19 we also have

$$a_{(0,t)}(y, y^-) \geq \nu_{ell} \|y^-\|^2_{L^2((0,t),V)} \geq 0. \tag{8.14}$$

Note that the reference to [Rou13, Lemma 7.2] in the proof of [Wac16a, Lemma 3.3] also applies to our situation where V is not necessarily equal to $H^1(\Omega)$. So [Wac16a, Lemma 3.3] is applicable and we have

$$\int_0^t \left(\partial_t y, y^-\right)_{V^*,V} \, dt = \frac{1}{2}\|y^-(t)\|^2_H - \frac{1}{2}\|y^-(0)\|^2_H. \tag{8.15}$$

We also have

$$\left(u - fy, y^-\right)_{L^2((0,t)\times\Omega)} = \left(u - fy^-, y^-\right)_{L^2((0,t)\times\Omega)} \leq 0. \tag{8.16}$$

Inserting (8.14), (8.15) and (8.16) into (8.13) yields

$$\frac{1}{2}\|y^-(t)\|^2_H - \frac{1}{2}\|y^-(0)\|^2_H \leq 0.$$

But since $y_0 \geq 0$ we have $y^-(0) = y_0^- = 0$ and thus $\|y^-(t)\|^2_H \leq 0$. Therefore $y^-(t) = 0$ a.e. in Ω. In turn, as $t \in I$ was arbitrary, we find $y^- = 0$ a.e. in Q. □

8.4 Existence Results and Estimates for Semilinear Parabolic PDEs

This section contains a general existence result for solutions of non-linear PDEs. The following assumption shall hold throughout Section 8.4. We will refer to it explicitly.

Assumption 8.16 *In Section 8.4 let I, Ω, Q, Γ_D, Σ_D, V, q_u and A have the same properties as in Section 2.2.1. Note that in particular $\mathbb{W}_{q_u} \subset C(\bar{\Omega})$ by Proposition 8.51. Let $\tilde{f}(t,x,y) : I \times \Omega \times \mathbb{R} \to \mathbb{R}$ be a function satisfying:*

- *$\tilde{f}(\cdot,\cdot,y)$ is measurable for each $y \in \mathbb{R}$,*
- *$\tilde{f}$ is locally Lipschitz continuous in y: For all $M > 0$ there is a $L_M > 0$ such that*

$$|\tilde{f}(t,x,y_1) - \tilde{f}(t,x,y_2)| \le L_M |y_1 - y_2| \qquad \forall y_1, y_2 \in B_M(0),$$

- *$\tilde{f}(t,x,\cdot)$ is monotonically increasing for a.e. $(t,x) \in Q$,*
- *$\tilde{f}(0) \in L^{q_u}(Q)$.*

Theorem 8.17 *Assume Assumption 8.16 holds. Let $y_0 \in \mathbb{W}_{q_u}$, $u \in L^{q_u}(Q)$. There exists a unique solution y of*

$$\begin{cases} \partial_t y + Ay + \tilde{f}(y) = u, \\ y(0) = y_0, \quad y|_{\Sigma_D} = 0. \end{cases} \tag{8.17}$$

There is a κ^* such for any $\kappa_\Omega \in [0,\kappa^*)$ and $\kappa_I \in (0,1)$ with

$$\frac{1}{q_u}(1 + N/2) + \frac{\kappa_\Omega}{2} < 1 \text{ and } \kappa_I \in \left(0, 1 - \frac{1}{q_u}(1+N/2) - \frac{\kappa_\Omega}{2}\right)$$

there is a $C > 0$ such that

$$\begin{aligned} &\|y\|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} + \|\partial_t y\|_{L^{q_u}(Q)} + \|y\|_{L^2(I,V)} + \|Ay\|_{L^{q_u}(Q)} + \|\tilde{f}(y)\|_{L^{q_u}(Q)} \\ &\quad \le C\left(\|u - \tilde{f}(0)\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}}\right). \end{aligned}$$

The proof is split in multiple parts and consists of various existence and regularity results spread over various articles and books. We collect the results throughout Section 8.4 and start the proof of Theorem 8.17 on page 184.

Proposition 8.18 *Under the assumptions of Theorem 8.17 and the additional assumption $|\tilde{f}(t,x,y)| \le C$ for all $(t,x,y) \in Q \times \mathbb{R}$ the parabolic equation* (8.17) *has a unique solution $y \in W(I)$ which satisfies*

$$\|y\|_{C(\bar{I},H) \cap L^2(I,V)} \le C_{\tilde{f}}\left(\|u - \tilde{f}(0)\|_{L^2(Q)} + \|y_0\|_{L^2(\Omega)}\right).$$

For $M \ge 0$ we also have for $y^M := \max(y - M, 0)$

$$\|y^M\|_{C(\bar{I},H) \cap L^2(I,V)} \le C_{\tilde{f},\not{M}}\left(\|u - \tilde{f}(0)\|_{L^2(Q)} + \|y_0\|_{L^2(\Omega)}\right).$$

Proof. Note that [Trö09, Lemma 5.3] was proven for different boundary conditions, nevertheless the proof still applies with minor modifications and yields the existence of a solution $y \in W(I)$ to (8.17).

Now we can test (8.17) with $y \cdot 1_{(0,T')}$ for $T' \in I$ and get

$$\int_0^{T'} (\partial_t y, y)_{V^*,V} + (\tilde{f}(y) - \tilde{f}(0), y)_H \, dt + a_{(0,T')}(y,y) = \int_0^{T'} (u - \tilde{f}(0), y)_H \, dt.$$

By partial integration, the monotonicity of $\tilde{f}$ and the ellipticity of A we get

$$\frac{1}{2}\|y(T')\|_H^2 - \frac{1}{2}\|y_0\|_H^2 + \nu_{ell}\|\nabla y\|^2_{L^2((0,T'),H)} \leq \|u - \tilde{f}(0)\|_{L^2(I,H)}\|y\|_{L^2(0,T',H)}. \tag{8.18}$$

This implies

$$\|y(T')\|_H^2 \leq 2\|u - \tilde{f}(0)\|_{L^2(Q)}\|y\|_{L^2(0,T',H)} + \|y_0\|_H^2.$$

By Lemma 8.3 this yields

$$\|y\|_{C(\bar{I},H)} \leq 2\sqrt{T}\|u - \tilde{f}(0)\|_{L^2(I,H)} + \frac{1+\sqrt{3}}{2}\|y_0\|_H \leq 2\sqrt{T}\|u - \tilde{f}(0)\|_{L^2(I,H)} + 2\|y_0\|_H$$

and this implies

$$\|y\|_{L^2(I,H)} \leq 2T\|u - \tilde{f}(0)\|_{L^2(I,H)} + 2\sqrt{T}\|y_0\|_H.$$

It remains to estimate $\|\nabla y\|_{L^2(I,H)}$. (8.18) implies

$$\begin{aligned}\nu_{ell}\|\nabla y\|^2_{L^2(I,H)} &\leq \|u - \tilde{f}(0)\|_{L^2(I,H)}\left(2T\|u - \tilde{f}(0)\|_{L^2(I,H)} + 2\sqrt{T}\|y_0\|_H\right) + \frac{1}{2}\|y_0\|_H^2 \\ &= \left(\sqrt{2T}\|u - \tilde{f}(0)\|_{L^2(I,H)} + \frac{1}{\sqrt{2}}\|y_0\|_H\right)^2.\end{aligned}$$

Taking the root yields the desired estimate for ∇y. Note that all the constants do not depend on $\tilde{f}$.

Now let $M \geq 0$. By Proposition 8.19 we have $y^M \in L^2(I,V)$ and $\|y^M\|_{L^2(I,V)} \leq C_{\not M}\|y\|_{L^2(I,V)}$. It is also clear that $\|y^M\|_{C(\bar{I},H)} \leq \|y\|_{C(\bar{I},H)}$. Thus the second estimate follows from the first.

□

The following proposition can be seen as a variant of [KS80, II Proposition 5.3], which was also used in the context of variational inequalities.

Proposition 8.19 *Let $\Omega \subset \mathbb{R}^N$ be bounded and open. Let $v \in H^1(\Omega)$ and $M \geq 0$, then $v^M := \max(v - M, 0)$ lies in $H^1(\Omega)$. Let V be as in Assumption 8.16. If $v \in V$, then $v^M \in V$. We also have*

$$\nabla v^M = \begin{cases} 0 & \text{if } v \leq M, \\ \nabla v & \text{if } v > M. \end{cases}$$

Proof. Let $v \in H^1(\Omega)$ by [KS80, Section II, Theorem A.1] we have $v^M \in H^1(\Omega)$ and the given form of the gradient. To see that the boundary data are preserved let $v \in V$. Let $(v_\epsilon)_{\epsilon>0} \subset C^\infty_{\Gamma_D}(\Omega)$ be a sequence approximating $v \in V$ in the $H^1(\Omega)$-norm. Such a sequence exists by the definition of V. Then by [KS80, Theorem A.1] we again have $v_\epsilon^M \in H^1(\Omega)$, but also $v_\epsilon^M \in C(\bar{\Omega})$ by construction. Thus for any $x \in \Gamma_D \subset \partial\Omega$ we find

$$v_\epsilon(x) - M = -M \leq 0 \quad \Longrightarrow \quad \max(v_\epsilon(x) - M, 0) = 0.$$

Therefore $v_\epsilon^M \in V$ by the characterization of V via the trace operator, see the discussion above (2.5). We now show $v_{\epsilon_n}^M \xrightarrow{n\to\infty} v^M$ weakly $H^1(\Omega)$, for an appropriate zero sequence. Because V is closed and convex, by definition, it is weakly closed and thus $v_M \in V$.

We have by the $H^1(\Omega)$-convergence of $(v_\epsilon)_{\epsilon>0}$

$$\|v_\epsilon^M\|_{H^1(\Omega)} \le \|v_\epsilon - M\|_{H^1(\Omega)} \le \|v_\epsilon\|_{H^1(\Omega)} + M|\Omega|^{\frac{1}{2}} \le C.$$

Thus for any zero sequence $(\epsilon_n)_{n\in\mathbb{N}}$ there exists a subsequence, denoted by the same name, such that $(v_{\epsilon_n})_{n\in\mathbb{N}}$ converges weakly in $H^1(\Omega)$, a Hilbert space, to some $\tilde{v}$. We show that $\tilde{v} = v^M$. The squared norm $\|\cdot\|_H^2$ is convex and continuous with respect to the $H^1(\Omega)$-norm therefore it is weakly lower semi-continuous. This allows us to estimate

$$\|\tilde{v} - v^M\|_H^2 \le \liminf_{n\to\infty} \|v_{\epsilon_n}^M - v^M\|_H^2 \le \liminf_{n\to\infty} \|v_{\epsilon_n} - v\|_H^2 = 0.$$

Here we used the fact that $\max(\cdot, 0)$ is a Lipschitz continuous mapping with Lipschitz constant 1 and the strong convergence $v_{\epsilon_n} \xrightarrow{n\to\infty} v$ in $H^1(\Omega)$. We have shown $\tilde{v} = v^M$, where $\tilde{v}$ is the weak $H^1(\Omega)$-limit of $(v_{\epsilon_n})_{n\in\mathbb{N}}$. This concludes the proof as stated earlier. □

The following result is almost [DtER15, Theorem 3.1]. A similar statement was also proven in [USL88, Chapter V, Sections 1 and 2].

Theorem 8.20 *Assume Assumption 8.16 holds. There exists a $\kappa^* > 0$ such that the following holds: if $\kappa_\Omega \in [0, \kappa^*)$ and $\kappa_I \in (0,1)$ satisfy*

$$\frac{1}{q_u}(1 + N/2) + \frac{\kappa_\Omega}{2} < 1 \text{ and } \kappa_I \in \left(0, 1 - \frac{1}{q_u}(1+N/2) - \frac{\kappa_\Omega}{2}\right) \tag{8.19}$$

then there is a $C > 0$ such that the following holds true: Let $u \in L^{q_u}(Q)$ and $y_0 \in \mathbb{W}_{q_u}$. Then there exists a solution $y \in W(I)$ of

$$\begin{cases} \partial_t y + Ay = u, \\ y(0) = y_0, \quad y|_{\Sigma_D} = 0. \end{cases} \tag{8.20}$$

It lies in $C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))$ with

$$\|y\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|y\|_{L^2(I,V)} + \|\partial_t y\|_{L^{q_u}(Q)} + \|Ay\|_{L^{q_u}(Q)} \le C\left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}}\right).$$

Proof. By [BN18, Propositions A.2, A.3] Ω satisfies the regularity assumptions [DtER15, Assumptions 2.3, 2.4, 2.5]. Thus [DtER15, Theorem 2.9b)] is applicable and gives that $A\colon \operatorname{dom}_{q_u}(A) \to L^{q_u}(\Omega)$ has maximal parabolic $L^{q_u}(I, L^{q_u}(\Omega))$-regularity. By [Ama04, Proposition 3.1] this implies that there exists a unique solution to (8.20) satisfying:

$$\|\partial_t y\|_{L^{q_u}(Q)} + \|Ay\|_{L^{q_u}(Q)} \le C\left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}}\right).$$

Note that we cheated a little bit: [DtER15] was written for $N \ge 2$, while would like to admit $N = 1$ as well. The arguments so far do not really depend on the dimension as a one-dimensional domain, an interval, has arbitrary regularity.

[DtER15, Theorem 3.1] contains the Hölder estimate of the given form for $N \ge 2$. For $N = 1$, however, we see the regularity directly by Theorem 1.34 with $\kappa^* = 1 - {q_u}^{-1}(1 + N/2)$. □

Remark 8.21 The results of [DtER15] are far more general that we let on. It would be possible in fact, to use more general domains than Lipschitz domains, for example Gröger regular domains that were introduced in [GR89]. To see that those domains are more general see [HDR09, Section 7.3].

It also be possible to use right hand sides that are in $L^s(I, L^p(\Omega))$ for different s and p, but this theorem suffices for our presentation. We are mostly interested in the application of those regularity results in special domains, together with high regularity of the control u provided by control constraints, c.f. Proposition 2.36 and Remark 4.62.

Proof of Theorem 8.17. By considering $u - \tilde{f}(0)$ instead of u we may assume $\tilde{f}(0) = 0$. Then for $k \geq 0$ consider

$$\tilde{f}_k(t,x,y) := \begin{cases} \tilde{f}(t,x,k) & \text{if } k \leq y, \\ \tilde{f}(t,x,y) & \text{if } -k < y < k, \\ \tilde{f}(t,x,-k) & \text{if } y \leq -k. \end{cases}$$

By Proposition 8.18 the problem

$$\begin{cases} \partial_t y_k + A y_k + \tilde{f}_k(y_k) = u, \\ y_k(0) = y_0, \quad y_k|_{\Sigma_D} = 0 \end{cases} \tag{8.21}$$

has a unique solution y_k satisfying

$$\|y_k\|_{C(\bar{I},H)} + \|y\|_{L^2(I,V)} \leq C_{\not{k}} \left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{L^2(\Omega)} \right).$$

Just as in the proof of Lemma 2.26 one can show:

$$\|\tilde{f}_k(y_k)\|_{L^{q_u}(Q)} \leq C_{Lip,\not{k}} \left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{L^\infty(\Omega)} \right). \tag{8.22}$$

The C does only depend on the Lipschitz constant of $\tilde{f}_k$ on the Ball $B_{\|y_0\|_{L^\infty(\Omega)}}$. This is obviously the same Lipschitz constant as for $\tilde{f}$, on the ball $B_{\|y_0\|_{L^\infty(\Omega)}}$. Therefore y_k is the unique solution of

$$\begin{cases} \partial_t y_k + A y_k = u - \tilde{f}_k(y_k) \text{ in Q}, \\ y_k(0) = y_0, \qquad y_k|_{\Sigma_D} = 0. \end{cases} \tag{8.23}$$

By Theorem 8.20 there are $\kappa^*, C > 0$ such that for all $\kappa_\Omega \in (0, \kappa^*]$ and $\kappa_I \in (0,1)$ satisfying (8.19) we have

$$\begin{aligned} &\|y_k\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|y_k\|_{L^2(I,V)} + \|\partial_t y_k\|_{L^{q_u}(Q)} + \|A y_k\|_{L^{q_u}(Q)} \\ &\leq C \left(\|u - \tilde{f}_k(y_k)\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right). \end{aligned} \tag{8.24}$$

Here no constant depends on k or u. By (8.22) and $\mathbb{W}_{q_u} \subset C(\bar{\Omega})$ from Proposition 8.51 we find

$$\begin{aligned} &\|y_k\|_{C^{\kappa_I}(I,C^{\kappa_\Omega}(\Omega))} + \|\partial_t y_k\|_{L^{q_u}(Q)} + \|y_k\|_{L^2(I,V)} + \|A y_k\|_{L^{q_u}(Q)} \\ &\leq C_{\not{k}} \left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right). \end{aligned}$$

Note that the right hand side does not depend on k. So choosing

$$k > C_{\not{k}} \left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right) \geq \|y_k\|_{L^\infty(Q)}$$

we have $\tilde{f}_k(y_k) = \tilde{f}(y_k)$ and y_k is the unique solution to (8.17). □

Theorem 8.22 *Let everything be as in Assumption 8.16. The solution operator belonging to* (8.17) *is Lipschitz continuous from* $L^{q_u}(Q)$ *to* $L^\infty(Q)$ *with a Lipschitz constant independent of the non-linearity* $\tilde{f}$.

If one could prove that taking the positive part of a $y_0 \in \mathbb{W}_{q_u}$ implies $y_0^+ \in \mathbb{W}_{q_u}$, it would be straight forward to extend the result to Lipschitz continuity of the operator from $L^{q_u}(Q) \times \mathbb{W}_{q_u} \to L^\infty(Q)$.

Proof. Let $u_1, u_2 \in L^{q_u}(Q)$, $y_0 \in \mathbb{W}_{q_u}$ and y_1, y_2 the corresponding solutions to (8.17). Subtracting the equations for the states yields, abbreviating $\delta u := u_1 - u_2$ and $\delta y := y_1 - y_2$,

$$\delta u = \partial_t \delta y + A\delta y + \tilde{f}(y_1) - \tilde{f}(y_2) = \partial \delta y + A\delta y + \int_0^1 \tilde{f}'(y_2 + s\delta y)\, ds\; \delta y \text{ a.e. in } Q$$

by the regularities of y_1 and y_2 from Theorem 8.17. We abbreviate $a := \int_0^1 \tilde{f}'(y_2 + s\delta y)\, ds$. By the monotonicity of $\tilde{f}$ we have $a \geq 0$. We denote the positive and negative part of δu by $\delta u^\pm$, respectively. We then denote by $\delta y^{p/m}$ the solutions to

$$\begin{cases} \partial_t \delta y^p + A\delta y^p + a y^p = \delta u^+ \text{ in Q}, \\ \delta y^p(0) = 0, \quad \delta y^p|_\Sigma = 0. \end{cases} \quad \text{and} \quad \begin{cases} \partial_t \delta y^m + A\delta y^m + a y^m = \delta u^- \text{ in Q}, \\ \delta y^m(0) = 0, \quad \delta y^m|_\Sigma = 0. \end{cases}$$

They exist by Theorem 8.17. By the maximum principle from Theorem 8.15 we find $\delta y^p \geq 0$ and $\delta y^m \leq 0$. We also have the usual Lebesgue decomposition of δy into positive and negative parts: $\delta y = \delta y^+ + \delta y^-$. Let $(t,x) \in Q$ with $\delta y(t,x) \geq 0$. Then $\delta y^+(t,x) = \delta y(t,x) = \delta y^p(t,x) + \delta y^m(t,x)$. Thus $\delta y^p(t,x) \geq \delta y^+(t,x)$. We can therefore conclude $\delta y^p \geq \delta y^+$ on Q. Analogously we can deduce $\delta y^m \leq \delta y^-$.

We define $\hat{y}$ as the solution to

$$\begin{cases} \partial_t \hat{y} + A\hat{y} = \delta u^+, \\ \hat{y}(0) = 0, \quad \hat{y}|_\Sigma = 0. \end{cases}$$

It exists and exhibits high regularity by Theorem 8.17. We can compare this to δy^p and see

$$\partial_t \delta y^p + A\delta y^p = \delta u^+ - a y^p \leq \delta u^+ = \partial_t \hat{y} + A\hat{y} \text{ a.e. in } Q.$$

By Theorem 8.15 we find

$$0 \leq \delta y^+ \leq \delta y^p \leq \hat{y} \leq \|\hat{y}\|_{L^\infty(Q)}.$$

By Theorem 8.17 applied to $\hat{y}$ we find thus

$$0 \leq \delta y^+ \leq C\|\delta u^+\|_{L^{q_u}(Q)} \leq C\|\delta u\|_{L^{q_u}(Q)}.$$

Note that this C does not depend on a or $\tilde{f}$ by the definition of $\hat{y}$.

Arguing analogously for δy^- yields the claim. □

The following statements are just a corollary/restating of the well-known results of [Gri11] on elliptic regularity.

Theorem 8.23 *In addition to the assumptions in Assumption 8.16 let Ω be a $C^{1,1}$-domain, let A have $C^{0,1}(\Omega)$-coefficients and assume $V = H_0^1(\Omega)$. Let $p \in [2, \infty)$. Then there is a $C_{\not p} > 0$ such that for every $u \in L^p(\Omega)$ and the solution y of*

$$\begin{cases} Ay = u \text{ on } \Omega, \\ y|_{\partial\Omega} = 0 \end{cases}$$

one has

$$\|y\|_{W^{2,p}(\Omega)} \le C_{\not p} \frac{p^2}{p-1} \|u\|_{L^p(\Omega)}.$$

Proof. This is just the well-known result about higher elliptic regularity, see for example [Gri11, Lemma 2.4.2.1, Theorem 2.4.2.5] or [GT01, Theorem 9.15, Lemma 9.17]. The behaviour of the constant can be obtained by tracking it throughout the proofs of [GT01]. □

Corollary 8.24 *In addition to the assumptions in Assumption 8.16 let Ω be a $C^{1,1}$-domain, let A have $C^{0,1}(\Omega)$-coefficients and assume $V = H_0^1(\Omega)$. Then we have*

$$\|y\|_{W^{1,2}_{q_u}(Q)} \le C_{q_u} \left(\|u - \tilde{f}(0)\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right).$$

Here $C_{q_u} > 0$ does not depend on y or u, but only on the same quantities as in Theorem 8.17.

If we additionally have that Ω is a C^2-domain we have

$$\|y\|_{W^{1,2}_{q_u}(Q)} \le C_{\not q_u} {q_u}^2 \left(\|u - \tilde{f}(0)\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right).$$

Here $C_{\not q_u} > 0$ does not depend on y, u or q_u, but only on the same quantities as in Theorem 8.17.

Proof. The first claim is a simple combination of Theorem 8.23 with Theorem 8.17.

Under the higher regularity assumptions Lemma 8.53 and Remark 8.54 apply. Therefore Corollary 8.50, rescaled to $(0, T) = I$, together with Theorem 8.17 imply that (8.24), in the proof of Theorem 8.17, can be improved to

$$\begin{aligned} &\|y_k\|_{C^{\kappa_I}(I, C^{\kappa_\Omega}(\Omega))} + \|y_k\|_{L^2(I,V)} + \|\partial_t y_k\|_{L^{q_u}(Q)} + \|Ay_k\|_{L^{q_u}(Q)} \\ &\le C_{\not q_u} \frac{{q_u}^2}{q_u - 1} \left(\|u - \tilde{f}_k(y_k)\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right). \end{aligned}$$

Using this new knowledge about the behaviour of the constant the proof of Theorem 8.17 entails in particular

$$\|\partial_t y\|_{L^{q_u}(Q)} + \|Ay\|_{L^{q_u}(Q)} \le C_{\not q_u} \frac{{q_u}^2}{q_u - 1} \left(\|u\|_{L^{q_u}(Q)} + \|y_0\|_{\mathbb{W}_{q_u}} \right).$$

This together with the higher elliptic regularity from Theorem 8.23 yields the claim. □

Lemma 8.25 *Assume $A = -\Delta$ and that $\Omega \subset \mathbb{R}^2$ is a polygonal domain. That means we can decompose its boundary in $\Gamma_1, \Gamma_2 \ldots, \Gamma_M$ edges. By $\omega_j \in (0, 2\pi)$ we denote the angles between Γ_j and Γ_{j+1} with $\Gamma_{M+1} = \Gamma_1$. We assume that the Dirichlet boundary Γ_D is a union of edges of $\partial\Omega$.*

We define

$$\Phi_j := \begin{cases} 0 \text{ if } \Gamma_j \not\subset \Gamma_D, \\ \frac{\pi}{2} \text{ if } \Gamma_j \subset \Gamma_D, \end{cases}$$

for $j = 1, \dots, M$ with $\Phi_{M+1} := \Phi_1$. We assume $\Gamma_D \neq \emptyset$.

Let $p \in (2, \infty)$. We assume that following assumptions hold: For any $j = 1, \dots, M$ we have

$$\frac{1}{\pi}\left(\Phi_{j+1} - \Phi_j - \frac{2}{q}\omega_j\right) \notin \mathbb{Z}.$$

Here $q := \frac{p}{p-1}$. For all $k \in \mathbb{Z}$ one shall have

$$\frac{\Phi_{j+1} - \Phi_j + k\pi}{\omega_j} =: \lambda_{j,k} \leq -\frac{2}{q} \quad \text{or } \lambda_{j,k} \geq 0 \quad \text{or } \lambda_{j,k} = -1.$$

Then each solution $y \in H^1(\Omega)$ to

$$\begin{cases} -\Delta y = u \text{ in } \Omega, \\ y|_{\Gamma_D} = 0, \quad \partial_\nu y|_{\partial\Omega \setminus \Gamma_D} = 0, \end{cases} \tag{8.25}$$

with $u \in L^p(\Omega)$ satisfies

$$\|y\|_{W^{2,p}(\Omega)} \leq C\|u\|_{L^p(\Omega)}$$

for a constant $C > 0$ depending on p, but not on y or u.

It is possible to show higher regularity results for y even if $\Gamma_D = \emptyset$ by the same chapters of [Gri11]. We do not develop this further as the focus of this thesis lies mostly on smooth domains.

Proof. Let $p \in (2, \infty)$. By [Gri11, Theorem 4.4.3.7] there is a unique solution $y \in W^{2,p}(\Omega)$ to (8.25). By [Gri11, Theorem 4.3.2.4] and [Gri11, Remark 4.3.2.5] we have

$$\|y\|_{W^{2,p}(\Omega)} \leq C(\|u\|_{L^p(\Omega)} + \|y\|_{L^p(\Omega)}). \tag{8.26}$$

We have $N = 2$ so we find by standard Sobolev embeddings $\|y\|_{L^p(\Omega)} \leq C\|y\|_{H^1(\Omega)}$. By standard ellipticity estimates we also have $\|y\|_{H^1(\Omega)} \leq C\|u\|_{L^p(\Omega)}$ since the Poincaré inequality, e.g. [Eva98, Theorem 5.6.3], yields the ellipticity of A by the assumption $\Gamma_D \neq \emptyset$. Using these two estimates in (8.26) yields the claim. □

Proposition 8.26 *Assume that $\Omega \subset \mathbb{R}^2$ is a polygonal domain. That means we can decompose its boundary in $\Gamma_1, \Gamma_2, \dots, \Gamma_M$ edges. By $\omega_j \in (0, 2\pi)$ we denote the angles between Γ_j and Γ_{j+1} with $\Gamma_{M+1} = \Gamma_1$. We assume that the Dirichlet boundary Γ_D is a union of edges of $\partial\Omega$.*

We define

$$\Phi_j := \begin{cases} 0 \text{ if } \Gamma_j \not\subset \Gamma_D, \\ \frac{\pi}{2} \text{ if } \Gamma_j \subset \Gamma_D, \end{cases}$$

for $j = 1, \dots, M$ with $\Phi_{M+1} := \Phi_1$. We further define for $j = 1, 2, \dots, M$

$$\omega_{lim,j} := \begin{cases} \frac{\pi}{2} \text{ if } \Phi_j = \Phi_{j+1}, \\ \frac{\pi}{4} \text{ if } \Phi_j \neq \Phi_{j+1}. \end{cases}$$

Let $p \in (1,\infty)$ and $j \in \{1,2,\dots,M\}$. If $\omega_j \le \omega_{lim,j}$ or $p < \frac{\omega_j}{\omega_j - \omega_{lim,j}}$ we have

$$\frac{1}{\pi}\left(\Phi_{j+1} - \Phi_j - \frac{2}{q}\omega_j\right) \notin \mathbb{Z}. \tag{8.27}$$

Here $q := \frac{p}{p-1}$. We then also have

$$\frac{\Phi_{j+1} - \Phi_j + k\pi}{\omega_j} =: \lambda_{j,k} \le -\frac{2}{q} \quad \text{or } \lambda_{j,k} \ge 0. \tag{8.28}$$

Proof. Let all the quantities be as above. To make the proofs more readable, we drop the index j from ω_j and $\omega_{lim,j}$.

We first consider the case $\Phi_{j+1} = \Phi_j$. We first show (8.28) and then (8.27). If $0 \le k \in \mathbb{Z}$, we immediately see $\lambda_{j,k} \ge 0$. If $0 > k \in \mathbb{Z}$ the first inequality in (8.28) is equivalent to

$$k \le -\frac{2}{\pi}\frac{1}{q}\omega.$$

By assumption we have either $\omega \le \omega_{lim}$ or $q = \frac{p}{p-1} > \frac{\omega}{\omega_{lim}}$. In the first case we find

$$0 > -\frac{2}{\pi}\frac{1}{q}\omega \ge -\frac{2}{\pi}\frac{1}{q}\omega_{lim} = -\frac{1}{q} > -1 \ge k. \tag{8.29}$$

In the second case we find

$$0 > -\frac{2}{\pi}\frac{1}{q}\omega > -\frac{2}{\pi}\omega_{lim} = -1 \ge k. \tag{8.30}$$

This shows (8.28). Now (8.29) and (8.30) immediately imply (8.27).

We now consider the case $\Phi_{j+1} = \Phi_j + \frac{\pi}{2}$. The last case $\Phi_{j+1} = \Phi_j - \frac{\pi}{2}$ follows by shifting k in the following arguments. Again, if $0 \le k \in \mathbb{Z}$, we immediately see $\lambda_{j,k} \ge 0$. If $0 > k \in \mathbb{Z}$ the first inequality in (8.28) is equivalent to

$$\frac{1}{2} + k \le -\frac{2}{\pi}\frac{1}{q}\omega.$$

If $\omega \le \omega_{lim}$ we find

$$0 > -\frac{2}{\pi}\frac{1}{q}\omega \ge -\frac{2}{\pi}\frac{1}{q}\omega_{lim} = -\frac{1}{2q} > -\frac{1}{2} \ge k + \frac{1}{2}.$$

If $q > \frac{\omega}{\omega_{lim}}$ we find

$$0 > -\frac{2}{\pi}\frac{1}{q}\omega > -\frac{2}{\pi}\omega_{lim} = -\frac{1}{2} \ge k + \frac{1}{2}.$$

Now (8.29) and (8.30) once again imply (8.27). □

Combining Lemma 8.25 and Proposition 8.26 immediately implies the following result.

Corollary 8.27 *Assume $A = -\Delta$ and that $\Omega \subset \mathbb{R}^2$ is a polygonal domain. That means we can decompose its boundary in $\Gamma_1, \Gamma_2, \dots, \Gamma_M$ edges. By $\omega_j \in (0, 2\pi)$ we denote the angles between Γ_j and Γ_{j+1} with $\Gamma_{M+1} = \Gamma_1$. We assume that the Dirichlet boundary Γ_D is a union of edges of $\partial\Omega$. We assume $\Gamma_D \neq \emptyset$. Let $p \in (2,\infty)$.*

We define

$$\Phi_j := \begin{cases} 0 \text{ if } \Gamma_j \not\subset \Gamma_D, \\ \frac{\pi}{2} \text{ if } \Gamma_j \subset \Gamma_D, \end{cases}$$

for $j = 1, \ldots, M$ *with* $\Phi_{M+1} := \Phi_1$. *We further define for* $j = 1, 2, \ldots, M$

$$\omega_{lim,j} := \begin{cases} \frac{\pi}{2} \text{ if } \Phi_j = \Phi_{j+1}, \\ \frac{\pi}{4} \text{ if } \Phi_j \neq \Phi_{j+1}. \end{cases}$$

For each $j \in \{1, 2, \ldots, M\}$ *we assume* $\omega_j \leq \omega_{lim,j}$ *or* $p < \frac{\omega_j}{\omega_j - \omega_{lim,j}}$.

Then each solution $y \in H^1(\Omega)$ *to*

$$\begin{cases} -\Delta y = u \text{ in } \Omega, \\ y|_{\Gamma_D} = 0, \quad \partial_\nu y|_{\partial\Omega \setminus D} = 0, \end{cases} \tag{8.31}$$

with $u \in L^p(\Omega)$ *satisfies*

$$\|y\|_{W^{2,p}(\Omega)} \leq C\|u\|_{L^p(\Omega)}$$

for a constant $C > 0$ *depending on* p*, but not on* y *or* u.

8.5 Minor Results on Finite Element Spaces

Throughout Section 8.5 we assume that the definitions and implicit assumptions of Section 4.1.1 apply. This includes the assumptions in Section 2.2.1 as they hold for the whole of Chapter 4.

The following estimates on the Jacobian of the element transformation J_K, see Definition 4.1, are well-known, but for simple referencing we restate it.

Proposition 8.28 *Let* $(\mathcal{K}_h)_{h>0}$ *be a family of quasi-uniform meshes. We have for some* $C_1, C_2 > 0$

$$C_1 h^N \leq |\det(DJ_K(x))| \leq C_2 h^N \qquad \forall K \in \mathcal{K}_h, x \in K.$$

If $(\mathcal{K}_h)_{h>0}$ *is additionally shape regular we have for some* $C_1, C_2 > 0$ *and all* $x \in K$

$$\|DJ_K(x)\| \leq Ch \text{ and } \|D(J_K^{-1})(x)\| \leq Ch^{-1}.$$

Proof. In [EG04, Lemma 1.100] the well-known estimates

$$|\det(DJ_K)| = \frac{|K|}{|\hat{K}|}, \quad \|DJ_K\| \leq \frac{h_K}{\rho_{\hat{K}}}, \quad \|D(J_K^{-1})\| \leq \frac{h_{\hat{K}}}{\rho_K}$$

are proven. The claims are now an immediate estimate of the quasi-uniformity and, when required, shape regularity.

□

The next two results are devoted to a short recap and generalization of parts of [BTW03]. The results given there are made for “sufficiently smooth domains” and the Laplacian. We were interested in the least regularity required to obtain the same results by essentially the same proofs. The following proposition is a generalization of [BTW03, Theorem A.1]. Note that there is a small typo in the statement of [BTW03, Theorem A.1]: $W^{1,\infty}(\Omega)$ instead of $W^{1,\infty}(\Omega) \cap C_0(\Omega)$.

Proposition 8.29 *Assume Ω is a $C^{3,\alpha}$-domain and A has $C^{1,\alpha}(\Omega)$-coefficients for some $\alpha \in (0,1)$. Let $(\mathcal{K}_h)_{h\in(0,1]}$ be a family of meshes satisfying Assumption 4.14.*

Define $\rho_h^x(y) := \left(|x-y|^2 + h^2\right)^{\frac{1}{2}}$ for $x, y \in \Omega$. Then there is a $C > 0$ such that for any $W^{1,\infty}(\Omega) \cap C_0(\bar{\Omega})$, its Ritz projection v_h and any we have $x \in \Omega_h$

$$|\nabla v_h(x)| \le C\|\nabla v\|_{L^\infty(\Omega_h)} + \|(\rho_h^x(y))^{-N} v\|_{L^1(\partial\Omega_h)}.$$

Proof. The proof of [BTW03, Theorem A.1] is immediately applicable. It only requires the existence of a Green's function $G(x,y)$ which satisfies the bounds

$$\begin{aligned} |\nabla_x G(x,y)| &\le C|x-y|^{1-N}, \\ |\nabla_x^2 G(x,y)| &\le C|x-y|^{-N}, \\ |\nabla_x^2 \nabla_y G(x,y)| &\le C|x-y|^{-1-N}. \end{aligned}$$

This is true by Proposition 4.22. The fact that we use a general symmetric, uniformly elliptic operator with Lipschitz continuous coefficients does not change the proof at all. Just note that $H^2(\Omega) \hookrightarrow C(\bar{\Omega})$ for $N \le 3$, that is why we can use the nodal interpolant of $H^2(\Omega)$ functions freely and do not require the use of a different kind of interpolant, see the proof of [BTW03, (A.15)]. □

Remark 8.30 It is possible to generalize the result to dimensions greater than 3. But then one can longer use the nodal interpolant for $v \in H^2(\Omega)$. In fact the authors of [BTW03] use Scott-Zhang type interpolants, see for example [SZ90]. However, their estimates for Scott-Zhang type interpolants require convexity of the domain, which we would like to avoid.

Corollary 8.31 *Assume Ω is a $C^{3,\alpha}$-domain and that A has $C^{1,\alpha}(\Omega)$-coefficients for some $\alpha \in (0,1)$. Let $(\mathcal{K}_h)_{h\in(0,1]}$ a family of meshes such that Assumption 4.14 holds.*

Then there is a $C > 0$ independent of h such that for any $W^{1,\infty}(\Omega) \cap C_0(\Omega)$ and its Ritz-projection $v_h \in \tilde{V}_h$ we have

$$\|v_h\|_{W^{1,\infty}(\Omega)} \le C\|v\|_{W^{1,\infty}(\Omega_h)} + Ch\|v\|_{W^{1,\infty}(\Omega\setminus\Omega_h)}.$$

Proof. By the remark under [BTW03, Theorem A.1] Proposition 8.29 implies the desired result. □

8.6 Interpolation Spaces

The following Sections 8.6 to 8.8 were all created in very close collaboration with Lucas Bonifacius[1] and were also used in his dissertation ([Bon18, Section 5.3.3, Section A.1, Section A.6]).

[1] Lucas Bonifacius, formerly: TU München, Lehrstuhl für Optimalsteuerung, Department of Mathematics, Boltzmannstr. 3, 85748 Garching b. München, Germany (`lucas.bonifacius@ma.tum.de`, `https://www-m17.ma.tum.de/Lehrstuhl/LucasBonifacius`).

We collect several well-known properties of interpolation spaces. For further information we refer to the monographs [BL76, Tri78, Lun09]. To facilitate access to the individual topics, this appendix is rendered as self-contained as possible. Furthermore, since for the pointwise discretization error estimate we require the precise dependencies of the constants, we will state them explicitly.

Let X and Y be real or complex Banach spaces. The couple $\{ X, Y \}$ is called an interpolation couple, if both X and Y are continuously embedded into a linear Hausdorff space $\mathcal{V}$. Then the space $X \cap Y$ equipped with the norm

$$\|u\|_{X\cap Y} = \max \{ \|u\|_X, \|u\|_Y \}$$

is a linear subspace of $\mathcal{V}$. Moreover, the space $X + Y$ with the norm

$$\|u\|_{X+Y} = \inf_{\substack{x\in X, y\in Y\\ u=x+y}} \|x\|_X + \|y\|_Y$$

is also a linear subspace of $\mathcal{V}$. The interpolation theory is concerned with intermediate spaces, i.e. is any Banach space E such that

$$X \cap Y \hookrightarrow E \hookrightarrow X + Y.$$

An intermediate space E is called *interpolation space*, if for every linear operator $T \in \mathcal{L}(X+Y)$ whose restriction to X belongs to $\mathcal{L}(X)$ and whose restriction to Y belongs to $\mathcal{L}(Y)$, the restriction of T to E belongs to $\mathcal{L}(E)$.

In the following we will introduce the K-method and the trace method that lead to the so-called real interpolation spaces. Thereafter, we will discuss the connection of real interpolation spaces and domains of fractional powers of sectorial operators.

Given a Banach space X, let $L^s_*(\mathbb{R}_+; X)$ denote the space of s integrable functions with values in X with respect to the measure $\mathrm{d}t/t$. Moreover, we set $L^\infty_*(\mathbb{R}_+; X) = L^\infty(\mathbb{R}_+; X)$. For $X = \mathbb{R}$ and any s we write $L^s(\mathbb{R}_+; \mathbb{R}) =: L^s(\mathbb{R}_+)$.

The K-method

Let $\{X, Y\}$ be an interpolation couple. For $t \in (0, \infty)$ and $u \in \mathcal{V}$ the K-functional is defined as

$$K(t, u, X, Y) = \inf_{x\in X, u-x\in Y} [\|x\|_X + t\|u - x\|_Y] .$$

For $\tau \in (0, 1)$ and $1 \le s \le \infty$ we define the real interpolation space

$$(X, Y)_{\tau,s} := \{ u \in X + Y : t \mapsto t^{-\tau} K(t, u, X, Y) \in L^s_*(\mathbb{R}_+) \}$$

equipped with the norm

$$\|u\|_{\tau,s} = \|t^{-\tau} K(t, u, X, Y)\|_{L^s_*(\mathbb{R}_+)};$$

see, e.g. [Lun09, Section 1.1]. If ambiguity is not to be expected, we simply write $K(t, u)$ instead of $K(t, u, Y, X)$. In this thesis the norm of the real interpolation space is always defined by the K-functional as above, if not indicated otherwise.

Proposition 8.32 *Let $\tau \in (0, 1)$, $1 \le s \le \infty$, and $\{X, Y\}$ an interpolation couple such that $Y \hookrightarrow X$ with embedding constant C. Then for any $u \in (X, Y)_{\tau,s}$*

$$\|u\|_{\tau,s} \le \left(\frac{s}{(s-\tau)\tau}\right)^{1/s} C^{1-\tau/s} \|u\|_Y$$

if $s < \infty$ and

$$\|u\|_{\tau,\infty} \le C^{1-\tau} \|u\|_Y .$$

Remark 8.33 A common choice for us will be $s \in (s_0, \infty)$ for some $s_0 > 1$ and $\tau := 1 - 1/s$. Then the constant from Proposition 8.32 stays bounded for large s, since

$$\left(\frac{s}{(s-\tau)\tau}\right)^{1/s} = \left(\frac{s}{(s-1+\frac{1}{s})(1-\frac{1}{s})}\right)^{1/s} \le \left(\frac{s^2}{(s-1)^2}\right)^{1/s} \le \frac{1}{(s_0-1)^2}\left(s^{1/s}\right)^2.$$

Proof of Proposition 8.32. Let $\tau \in (0,1)$, $1 \le s \le \infty$, and $u \in (X,Y)_{\tau,s}$. Then by the definition of the K-functional we obtain

$$K(t,u,X,Y) \le \min(t\|u\|_Y, \|u\|_X) \le \min(t,C)\|u\|_Y.$$

For $s = \infty$ we now immediately see

$$\|u\|_{\tau,\infty} \le \sup_{t\in(0,\infty)} t^{-\tau}\min(t,C)\|u\|_Y \le C^{1-\tau}\|u\|_Y.$$

For $s < \infty$ we split the integral in the definition of the norm and obtain

$$\begin{aligned}\|u\|_{\tau,s}^s &\le \int_0^C t^{-\tau}t^s\|u\|_Y^s\frac{\mathrm{d}t}{t} + \int_C^\infty t^{-\tau}C^s\|u\|_Y^s\frac{\mathrm{d}t}{t}\\ &= \left(\frac{1}{s-\tau}t^{s-\tau}|_0^C - \frac{1}{\tau}t^{-\tau}|_C^\infty C^s\right)\|u\|_Y^s = \left(\frac{1}{s-\tau}+\frac{1}{\tau}\right)C^{s-\tau}\|u\|_Y^s.\end{aligned}$$

Taking the s-th root yields the claim. □

Proposition 8.34 *Let $\tau \in (0,1)$, $1 \le s_1 \le s_2 \le \infty$. Then*

$$(X,Y)_{\tau,s_1} \hookrightarrow (X,Y)_{\tau,s_2}$$

with embedding constant bounded by $c(\tau,s_1,s_2) = [s_1 \min\{\tau, 1-\tau\}]^{1/s_1-1/s_2}$. For the case $s_1 = s_2 = \infty$ we obviously have $c(\tau,s_1,s_2) = 1$.

Proof. See proof of [Lun09, Proposition 1.1.3]. □

Proposition 8.35 *Suppose $Y \hookrightarrow X$. If $0 < \tau_1 < \tau_2 < 1$, then*

$$(X,Y)_{\tau_2,\infty} \hookrightarrow (X,Y)_{\tau_1,1}$$

with embedding constant bounded by $c(\tau_1,\tau_2) = (\tau_2-\tau_1)^{-1} + \tau_1^{-1}$.

Proof. See proof of [Lun09, Proposition 1.1.4]. □

Combination of Propositions 8.34 and 8.35 immediately implies the following embedding; see also [Tri78, Theorem 1.3.3 e)].

Proposition 8.36 *Suppose $Y \hookrightarrow X$. If $0 < \tau_1 < \tau_2 < 1$ and $1 \le s_1, s_2 \le \infty$, then*

$$(X,Y)_{\tau_2,s_1} \hookrightarrow (X,Y)_{\tau_1,s_2}$$

with embedding constant bounded by $c(\tau_1,\tau_2,s_1,s_2) = c(\tau_2,s_1,\infty)c(\tau_1,\tau_2)c(\tau_1,1,s_2)$.

Remark 8.37 For the particular choice $\tau_1 = 1-2/r$, $\tau_2 = 1-1/r$, $s_1 = r$, and $s_2 = p$ for any $r > 2$ and $p > 1$, the embedding constant of Proposition 8.36 is bounded by

$$\begin{aligned}c(1-2/r,1-1/r,r,p) &= \left[r\min\left\{1-\frac{1}{r},\frac{1}{r}\right\}\right]^{1/r}\left(r+\left(1-\frac{1}{r}\right)^{-1}\right)\left[p\min\left\{1-\frac{2}{r},\frac{2}{r}\right\}\right]^{1-1/p}\\ &\le r^{1/r}(r+1)p^{1-1/p} \le cr.\end{aligned}$$

The trace method

Let γ_0 denote the trace mapping, i.e. $\gamma_0 u = u(0)$. Moreover, for $\tau \in (0,1)$ set

$$v_{0,1-\tau}(t) = t^{1-\tau} v(t) \quad \text{and} \quad v_{1,1-\tau}(t) = t^{1-\tau}\partial_t v(t)$$

and introduce the trace space as

$$V(s, 1-\tau, Y, X) := \{\, \gamma_0 v \colon v_{0,1-\tau} \in L^s_*(\mathbb{R}_+; Y), v_{1,1-\tau} \in L^s_*(\mathbb{R}_+; X) \,\},$$

equipped with the norm

$$\|u\|^{\mathrm{Tr}}_{\tau,s} = \inf \{\, \|v_{0,1-\tau}\|_{L^s_*(\mathbb{R}_+;Y)} + \|v_{1,1-\tau}\|_{L^s_*(\mathbb{R}_+;X)} \colon \gamma_0 v = u \,\}.$$

It is well-known that the trace method is equivalent to the K-method and thus leads to the same interpolation spaces. More specifically, it holds

Proposition 8.38 *Let $\{X, Y\}$ be an interpolation couple, $\tau \in (0,1)$, $1 \le s \le \infty$. Then*

$$V(s, 1-\tau, Y, X) = (X, Y)_{\tau,s}$$

with equivalent norms. Precisely, it holds

$$\|u\|_{\tau,s} \le \frac{1}{\tau}\|u\|^{\mathrm{Tr}}_{\tau,s} \le \frac{2}{\tau}\left(2 + \frac{1}{\tau}\right)\|u\|_{\tau,s}.$$

Proof. See [Lun09, Proposition 1.2.2], where also constants are given explicitly in the proof. □

The trace method yields an important embedding result for spaces of maximal parabolic regularity.

Proposition 8.39 *Let $T > 0$ and X, Y be Banach spaces that satisfy the assumptions from Lemma 1.31. If $s \in (1, \infty)$, then*

$$W^{1,s}((0,T); X) \cap L^s((0,T); Y) \hookrightarrow C([0,T]; (X,Y)_{1-1/s,s}). \tag{8.32}$$

If $\tau \in (0, 1 - \frac{1}{s})$, then

$$W^{1,s}((0,T); X) \cap L^s(I; Y) \hookrightarrow C^\alpha((0,T); (X,Y)_{\tau,1}), \qquad 0 \le \alpha < 1 - \frac{1}{s} - \tau. \tag{8.33}$$

Moreover, the embedding constants are bounded by

$$c_{(8.32)}(s) = \frac{cs}{s-1} \quad \text{and} \quad c_{(8.33)}(\tau, s) = 2\left(c_{(8.32)}(s)\right)^{\tau/(1-1/s)}.$$

Proof. The embedding constant for (8.33) is explicitly verified in the proof of [DtER15, Lemma 3.4 b)]. Precisely, the constant for (8.33) is bounded by $2c^\lambda$ with $\lambda = \tau/(1-1/s)$ and c from (8.32). For these reasons, it remains to verify the dependencies of (8.32), where we follow the ideas of [Ama95, Theorem III.4.10.2].

For the particular choice $\tau = 1 - 1/s$, the trace space becomes

$$V(s, 1/s, Y, X) := \{\, \gamma_0 v \colon v \in W^{1,s}(\mathbb{R}_+; X) \cap L^s(\mathbb{R}_+; Y) \,\},$$

equipped with the norm

$$\|u\|^{\mathrm{Tr}}_{1-1/s,s} = \inf\{\, \|v\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \colon \gamma_0 v = u \,\}.$$

Clearly, the trace mapping $\gamma_0 \colon W^{1,s}(\mathbb{R}_+;X) \cap L^s(\mathbb{R}_+;Y) \to V(s,1/s,Y,X)$ is linear and continuous with norm less than or equal to one.

Let λ_t denote the semigroup of left translations, i.e. $\lambda_t u(t') = u(t+t')$ for all $t,t' \geq 0$. It is easily verified that λ_t is a contraction semigroup on $W^{1,s}(\mathbb{R}_+;X) \cap L^s(\mathbb{R}_+;Y)$. Moreover, λ_t is strongly continuous; cf. [Ama95, Lemma III.4.10.1 (i)]. Noting that $\gamma_0\lambda_t u = u(t)$, we infer

$$\|u(t)\|^{\mathrm{Tr}}_{1-1/s,s} \leq \|\lambda_t u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \leq \|u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)}, \quad t \geq 0.$$

Furthermore, if $0 \leq t < t' < \infty$, we have

$$\begin{aligned} \|u(t') - u(t)\|^{\mathrm{Tr}}_{1-1/s,s} &\leq \|\lambda_t(\lambda_{t'-t} - 1)u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \\ &\leq \|(\lambda_{t'-t} - 1)u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \end{aligned}$$

for all $u \in W^{1,s}(\mathbb{R}_+;X) \cap L^s(\mathbb{R}_+;Y)$. Employing strong continuity of λ_t, we deduce that $u \colon \mathbb{R}_+ \to V(s,1/s,Y,X)$ is continuous. In summary,

$$W^{1,s}(\mathbb{R}_+;X) \cap L^s(\mathbb{R}_+;Y) \hookrightarrow C(\mathbb{R}_+;V(s,1/s,Y,X))$$

with embedding constant less than or equal to one.

To prove (8.32), we use the result on $\mathbb{R}_+$ combined with a retraction/coretraction argument. Let $u \in \mathcal{D}([0,T);Y)$, where $\mathcal{D}([0,T);Y)$ denotes the space of Y valued C^∞-functions on $[0,T)$ with compact supports. We define the reflection of u as

$$\hat{u}(t) = \begin{cases} u(t), & \text{if } 0 \leq t \leq T, \\ u(2T-t), & \text{if } T < t \leq 2T. \end{cases}$$

Let $\eta \in C^\infty(\mathbb{R}_+)$ be a smooth cut-off function such that η equals one on $[0,(4/3)T]$ and vanishes on $[(5/3)T,\infty)$. Then we define the extension of u by $Eu = \eta\hat{u}$. Since $\mathcal{D}([0,T);Y)$ is dense in $W^{1,s}((0,T);X) \cap L^s((0,T);Y)$ (see for example Lemma 1.31), we obtain

$$\|Eu\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \leq 2\|\eta\|_{C^1(\mathbb{R}_+)}\|u\|_{W^{1,s}((0,T);X)\cap L^s((0,T);Y)}$$

for all $u \in W^{1,s}((0,T);X) \cap L^s((0,T);Y)$. Thus, for any $t \in [0,T)$,

$$\begin{aligned} \|u(t)\|^{\mathrm{Tr}}_{1-1/s,s} = \|(Eu)(t)\|^{\mathrm{Tr}}_{1-1/s,s} &\leq \|Eu\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \\ &\leq c\|u\|_{W^{1,s}((0,T);X)\cap L^s((0,T);Y)}, \end{aligned}$$

with $c = 2\|\eta\|_{C^1(\mathbb{R}_+)}$, which is independent of s. Finally, according to Proposition 8.38 it holds $V(s,1/s,Y,X) = (X,Y)_{1-1/s,s}$ and

$$\|u\|_{1-1/s,s} \leq \frac{s}{s-1}\|u\|^{\mathrm{Tr}}_{1-1/s,s},$$

which yields (8.32). □

We will frequently use the following embeddings for spaces of maximal parabolic regularity; see Proposition 8.39. Let X and Y be Banach spaces such that they satisfy the assumptions from Lemma 1.31 and $s \in (1,\infty)$. Then

$$W^{1,s}(I;X) \cap L^s(I;Y) \hookrightarrow C([0,T];(X,Y)_{1-1/s,s}). \tag{8.34}$$

If $\tau \in (0, 1-\frac{1}{s})$, then

$$W^{1,s}(I;X) \cap L^s(I;Y) \hookrightarrow C^\alpha(I;(X,Y)_{\tau,1}), \qquad 0 \leq \alpha < 1 - \frac{1}{s} - \tau. \tag{8.35}$$

Furthermore, the embedding constants can be chosen uniformly for $s \in [2,\infty)$ and $\tau \in (0,1)$.

Intermediate spaces and the reiteration theorem

Let $\{X, Y\}$ be an interpolation couple, $0 \leq \theta \leq 1$, and E be an intermediate space, i.e. $X \cap Y \hookrightarrow E \hookrightarrow X + Y$. The space E is said to belong to the class $J_\theta(X, Y)$ between X and Y if there is $c > 0$ such that

$$\|x\|_E \leq c\|x\|_X^{1-\theta}\|x\|_Y^{\theta}, \quad x \in X \cap Y.$$

We write $E \in J_\theta(X, Y)$ for short. It holds the following reiteration theorem.

Proposition 8.40 *Let $0 \leq \theta_0 < \theta_1 \leq 1$ and $\tau \in (0,1)$. If $E_i \in J_{\theta_i}(X, Y)$, $i = 0, 1$, then*

$$(X, Y)_{(1-\tau)\theta_0+\tau\theta_1,s} \hookrightarrow (E_0, E_1)_{\tau,s}, \quad s \in [1, \infty].$$

Moreover, the embedding constant is bounded by

$$2(\theta_1 - \theta_0)^{-1-1/s}(1 + 3\theta^{-1})(c_0 + c_1(1-\tau)^{-1})\tau^{-2}\left(2 + \frac{1}{\tau}\right)$$

where c_i denotes the constant from the definition of the class $J_{\theta_i}(X, Y)$ and $\theta := (1-\tau)\theta_0 + \tau\theta_1$.

Before we give a proof of Proposition 8.40, we have to trace the constants mentioned in [Lun09, Remark 1.2.4].

Proposition 8.41 *For each $v \in V(p, 1-\theta, Y, X)$, with $\theta \in (0,1)$, we have that the mean*

$$w(t) := \frac{1}{t}\int_0^t v(s)\,ds, \quad t > 0,$$

satisfies the estimate

$$\begin{aligned} &\|t^{1-\theta}w\|_{L^p_*(\mathbb{R}_+;Y)} + \|t^{2-\theta}w'\|_{L^p_*(\mathbb{R}_+;Y)} + \|t^{1-\theta}w'\|_{L^p_*(\mathbb{R}_+;X)} \\ &\qquad\leq \left(1 + \frac{3}{\theta}\right)\left(\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)} + \|v_{1,1-\theta}\|_{L^p_*(\mathbb{R}_+;X)}\right). \end{aligned} \tag{8.36}$$

We also have $\gamma_0 w = \gamma_0 v$.

Proof. Let v and w be as above. First note that the derivative of w is given by

$$\begin{aligned} w'(t) &= -\frac{1}{t^2}\int_0^t v(s)\,ds + \frac{1}{t}v(t) \\ &= \frac{1}{t}\left(-w(t) + v(t)\right) = \frac{1}{t^2}\int_0^t -v(s) + v(t)\,ds. \end{aligned} \tag{8.37}$$

We estimate the first summand in (8.36), using [Lun09, Corollary A.3.1]

$$\|t^{1-\theta}w\|_{L^p_*(\mathbb{R}_+;Y)} \leq \frac{1}{\theta}\|t^{1-\theta}v\|_{L^p_*(\mathbb{R}_+;Y)} = \frac{1}{\theta}\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)}. \tag{8.38}$$

As a consequence, the second summand in (8.36) can now be estimated as

$$\begin{aligned} \|t^{2-\theta}w'\|_{L^p_*(\mathbb{R}_+;Y)} &= \|t^{1-\theta}(-w(t) + v(t))\|_{L^p_*(\mathbb{R}_+;Y)} \\ &\leq \|t^{1-\theta}w(t)\|_{L^p_*(\mathbb{R}_+;Y)} + \|t^{1-\theta}v(t)\|_{L^p_*(\mathbb{R}_+;Y)} \\ &\leq \left(1 + \frac{1}{\theta}\right)\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)}. \end{aligned} \tag{8.39}$$

The third summand in (8.36) can be estimated employing the last expression of (8.37). Thus,

$$\begin{aligned}\|w'(t)\|_X &\le \|\frac{1}{t^2}\int_0^t\int_s^t v'(\sigma)\,d\sigma\,ds\|_X\\ &\le \frac{1}{t^2}\int_0^t\int_0^t \|v'(\sigma)\|_X\,d\sigma\,ds \le \frac{1}{t}\int_0^t \|v'(\sigma)\|_X\,d\sigma.\end{aligned}$$

Now we have

$$\|t^{1-\theta}w'\|^p_{L^p_*(\mathbb{R}_+;X)} = \int_0^\infty t^{(1-\theta)p}\|w'\|^p_X\frac{\mathrm{d}t}{t} \le \int_0^\infty t^{-\theta p}\left(\int_0^t \sigma\|v'(\sigma)\|_X\frac{\mathrm{d}\sigma}{\sigma}\right)^p\frac{\mathrm{d}t}{t}.$$

Now the Hardy-Young inequality, see, e.g. [Lun09, Equation (A.3.1)], leads to

$$\begin{aligned}\|t^{1-\theta}w'\|^p_{L^p_*(\mathbb{R}_+;X)} &\le \theta^{-p}\int_0^\infty s^{-\theta p}\left(s\|v'(s)\|_X\right)^p\frac{\mathrm{d}s}{s}\\ &= \theta^{-p}\int_0^\infty s^{(1-\theta)p}\|v'(s)\|^p_X\frac{\mathrm{d}s}{s} = \theta^{-p}\|v_{1,1-\theta}\|^p_{L^p_*(\mathbb{R}_+;X)}.\end{aligned}\tag{8.40}$$

Thus the desired inequality follows by adding (8.38), (8.39), and the p-th root of (8.40).

The last statement directly follows from continuity of v: For $t>0$, we have

$$|\frac{1}{t}\int_0^t v(s)\,ds - \gamma_0 v| = |\frac{1}{t}\int_0^t v(s)-v(0)\,ds| \le \sup_{s\in[0,t]}|v(s)-v(0)|.$$

Continuity of v on $[0,\infty)$ and going to the limit $t\to 0$ in the inequality above yields

$$\gamma_0 w = \lim_{t\to 0} w(t) = \lim_{t\to 0}\frac{1}{t}\int_0^t v(s)\,ds = \gamma_0 v,$$

concluding the proof. □

Proof of Proposition 8.40. This is a standard result in interpolation theory. To trace the constants, we follow the proof of [Lun09, Theorem 1.3.5] that relies on the trace method. Set $\theta = (1-\tau)\theta_0 + \tau\theta_1$ and let $u \in (X,Y)_{\theta,s}$. Then there exists $v \in W^{1,s}(I;X)\cap L^s(I;Y)$ such that u is the trace of v at $t=0$, i.e. $\gamma_0 v = u$. Defining w by the mean of v as in Proposition 8.41 we obtain

$$\|t^{1-\theta}w'(t)\|_{L^s_*(\mathbb{R}_+;X)} + \|t^{2-\theta}w'(t)\|_{L^s_*(\mathbb{R}_+;Y)} \le c(\theta,v),$$

where $c(\theta,v) := \left(1+3\theta^{-1}\right)\left(\|v_{0,1-\theta}\|_{L^s_*(\mathbb{R}_+;Y)} + \|v_{1,1-\theta}\|_{L^s_*(\mathbb{R}_+;X)}\right)$ and $v_{0,1-\theta}$ and $v_{1,1-\theta}$ are defined as in the trace method. We have to verify that

$$g(t) = w(t^{1/(\theta_1-\theta_0)}),\quad t>0,$$

belongs to $V(s,1-\tau,E_0,E_1)$. This will imply that $u=\gamma_0 v = \gamma_0 w = \gamma_0 g$ belongs to the interpolation space $(E_0,E_1)_{\tau,s}$. Let c_i be such that

$$\|x\|_{E_i} \le c_i\|x\|_X^{1-\theta_i}\|x\|_Y^{\theta_i},\quad x\in X\cap Y.$$

Clearly, it holds

$$\|w'(t)\|_{E_i} \le \frac{c_i}{t^{\theta_i+1-\tau}}\|t^{1-\tau}w'(t)\|_X^{1-\theta_i}\|t^{2-\tau}w'(t)\|_Y^{\theta_i},\quad i=0,1.$$

Whence, from the equalities

$$\theta_0+1-\theta = 1-\tau(\theta_1-\theta_0),\quad \theta_1+1-\theta = 1+(1-\tau)(\theta_1-\theta_0),$$

we infer

$$\|t^{1-\tau(\theta_1-\theta_0)}w'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} \le c_0 c(\theta,v), \tag{8.41}$$

$$\|t^{1+(1-\tau)(\theta_1-\theta_0)}w'(t)\|_{L^s_*(\mathbb{R}_+;E_1)} \le c_1 c(\theta,v). \tag{8.42}$$

Substitution in the integral yields

$$\|t^{1-\tau}g(t)\|_{L^s_*(\mathbb{R}_+;E_1)} = (\theta_1-\theta_0)^{-1/s}\|t^{(1-\tau)(\theta_1-\theta_0)}w(t)\|_{L^s_*(\mathbb{R}_+;E_1)}.$$

Furthermore, using $w(t) = -\int_t^\infty w'(\sigma)\,d\sigma$, inequality (8.42), and the Hardy-Young inequality, we get

$$\|t^{(1-\tau)(\theta_1-\theta_0)}w(t)\|_{L^s_*(\mathbb{R}_+;E_1)} \le \frac{c_1 c(\tau,v)}{(1-\tau)(\theta_1-\theta_0)},$$

and thus

$$\|t^{1-\tau}g(t)\|_{L^s_*(\mathbb{R}_+;E_1)} \le (\theta_1-\theta_0)^{-1/s}\frac{c_1 c(\tau,v)}{(1-\tau)(\theta_1-\theta_0)}.$$

Moreover, since

$$g'(t) = (\theta_1-\theta_0)^{-1}t^{-1+1/(\theta_1-\theta_0)}w'(t^{1/(\theta_1-\theta_0)}),$$

we obtain, by (8.41),

$$\begin{aligned}\|t^{1-\tau}g'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} &= (\theta_1-\theta_0)^{-1-1/s}\|t^{1-\tau(\theta_1-\theta_0)}w'(t)\|_{L^s_*(\mathbb{R}_+;E_0)}\\ &\le (\theta_1-\theta_0)^{-1-1/s}c_0 c(\theta,v).\end{aligned}$$

This and (8.6) yields the estimate

$$\|t^{1-\tau}g'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} + \|t^{1-\tau}g(t)\|_{L^s_*(\mathbb{R}_+;E_1)} \le (\theta_1-\theta_0)^{-1-1/s}(c_0 + c_1(1-\tau)^{-1})c(\theta,v).$$

This implies, by the definition of the trace norm (note $\gamma_0 g = u$) and its equivalence to the K-method, see Proposition 8.38, that

$$\begin{aligned}\|u\|_{(E_0,E_1)_{\tau,s}} &\le \tau^{-1}\|u\|^{\mathrm{Tr}}_{(E_0,E_1)_{\tau,s}}\\ &\le \tau^{-1}\left(\|t^{1-\tau}g'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} + \|t^{1-\tau}g(t)\|_{L^s_*(\mathbb{R}_+;E_1)}\right),\\ &\le \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0 + c_1(1-\tau)^{-1})c(\theta,v),\\ &= \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0 + c_1(1-\tau)^{-1})\\ &\quad\left(1+\frac{3}{\theta}\right)\left(\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)} + \|v_{1,1-\theta}\|_{L^p_*(\mathbb{R}_+;X)}\right).\end{aligned}$$

Finally, taking the infimum over all v with $\gamma_0 v = u$, we find

$$\begin{aligned}\|u\|_{(E_0,E_1)_{\tau,s}} &\le \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0 + c_1(1-\tau)^{-1})\left(1+3\theta^{-1}\right)\|u\|^{\mathrm{Tr}}_{(X,Y)_{\tau,s}}\\ &\le \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0 + c_1(1-\tau)^{-1})\left(1+3\theta^{-1}\right)\frac{2}{\tau}\left(2+\frac{1}{\tau}\right)\|u\|_{(X,Y)_{\tau,s}}\end{aligned}$$

concluding the proof. □

8.7 The Real Interpolation Method and Domains of Fractional Operators

In this paragraph we consider linear operators A on a Banach space X with $\rho(A) \supset (-\infty, 0)$. Suppose there exists $M > 0$ such that

$$\|zR(z,A)\|_{\mathcal{L}(X)} \le M, \quad z < 0.$$

The real interpolation space between X and the domain of A can be characterized as follows.

Proposition 8.42 *Let $\tau \in (0,1)$ and $1 \le s \le \infty$. Then*

$$(X, \operatorname{dom}_X(A))_{\tau,s} = \Big\{x \in X \colon t \mapsto x_\tau(t) := t^\tau \|AR(-t,A)x\|_X \in L^s_*(\mathbb{R}_+)\Big\},$$

and the norms $\|\cdot\|_{\tau,s}$ and

$$\|x\|^*_{\tau,s} := \|x\|_X + \|x_\tau\|_{L^s_*(\mathbb{R}_+)}.$$

are equivalent. Precisely,

$$\|x\|_{\tau,s} \le \Big(2 + M((1-\tau)s)^{-1/s}\Big)\|x\|^*_{\tau,s}, \quad \|x\|^*_{\tau,s} \le (M+1)\|x\|_{\tau,s}.$$

Proof. This follows from the proof of [Lun09, Proposition 3.1.1]. □

Definition 8.43 A linear operator A on a Banach space is called *sectorial*, if there exists $M > 0$ such that $\rho(A) \supset (-\infty, 0)$ and

$$\|R(z,A)\|_{\mathcal{L}(X)} \le \frac{M}{1+|z|}, \quad z \le 0.$$

This allows to define fractional powers of A by means of the Dunford-Taylor integral; see, e.g. [Tri78, Section 1.15] and [Lun09, Chapter 4]. The theory of interpolation spaces is closely related to domains of fractional operators. We summarize some of these properties in the sequel.

Proposition 8.44 *Let $z_1, z_2 \in \mathbb{C}$ such that $\mathfrak{Re} z_1 < \mathfrak{Re} z_2$. Then*

$$\operatorname{dom}_X(A^{z_2}) \hookrightarrow \operatorname{dom}_X(A^{z_1})$$

and the embedding constant is bounded by $\max\{1, \|A^{z_1-z_2}\|_{\mathcal{L}(X)}\}$. Moreover, the mapping $z \mapsto A^z \in \mathcal{L}(X)$ is holomorphic in the half plane $\mathfrak{Re} z < 0$ and $A^0 = \mathrm{Id}$.

Proof. From the proof of [Lun09, Theorem 4.1.6] we have

$$\|A^{z_1}x\|_X \le \|A^{z_1-z_2}\|_{\mathcal{L}(X)}\|A^{z_2}x\|_X.$$

The remaining properties immediately follow from the definition; see also the text after [Lun09, Definition 4.1.3]. □

Proposition 8.45 *Let A be a sectorial operator on a Banach space X. Then*

$$(X, \operatorname{dom}_X(A))_{\tau,1} \hookrightarrow \operatorname{dom}_X(A^\tau), \quad \tau \in (0,1),$$

where the embedding constant is bounded by $(M+1)\max\{1, (\Gamma(\tau)\Gamma(1-\tau))^{-1}\}$.

Remark 8.46 For $\tau \in (0,1)$ we have $\Gamma(\tau) \ge 1 - e^{-1}$ and thus $\max(1, (\Gamma(\tau)\Gamma(1-\tau))^{-1}) \le \text{const.}$, independently of τ. This can be seen from the definition of the gamma function

$$\Gamma(\tau) = \int_0^\infty t^{\tau-1}e^{-t}\,dt \ge \int_0^1 t^{\tau-1}e^{-t}\,dt$$
$$\ge \int_0^1 e^{-t}\,dt = 1 - e^{-1} > 0.$$

Proof of Proposition 8.45. We closely follow the proof of [Lun09, Proposition 4.1.7]. Let $x \in (X, \mathrm{dom}_X(A))_{\tau,1}$. Due to Proposition 8.42, the mapping $t \mapsto t^\tau \|AR(-t,A)x\|_X$ belongs to $L^1_*(\mathbb{R}_+)$. Using the representation formula for A^τ, see, e.g. [Lun09, Equation (4.1.7)], we obtain

$$\|A^\tau x\|_X \le \frac{1}{\Gamma(\tau)\Gamma(1-\tau)} \int_0^\infty t^\tau \|AR(-t,A)x\|_X \frac{\mathrm{d}t}{t} \le \frac{1}{\Gamma(\tau)\Gamma(1-\tau)} \|x_\tau\|^*_{L^1_*(\mathbb{R}_+)}.$$

Hence, using again Proposition 8.42 we obtain,

$$\begin{aligned}\|x\|_X + \|A^\tau x\|_X &\le \max\{1, (\Gamma(\tau)\Gamma(1-\tau))^{-1}\} \|x\|^*_{\tau,1} \\ &\le (M+1)\max\{1, (\Gamma(\tau)\Gamma(1-\tau))^{-1}\} \|x\|_{\tau,1},\end{aligned}$$

concluding the proof. □

Proposition 8.47 *Let A be a sectorial operator on a Banach space X. Then*

$$\mathrm{dom}_X(A^\tau) \hookrightarrow (X, \mathrm{dom}_X(A))_{\tau,\infty}, \quad \tau \in (0,1),$$

where the embedding constant is bounded by

$$\frac{(2+M)M(M+1)^2}{\Gamma(1-\tau)\Gamma(1+\tau)} \left(\frac{1}{1-\tau} + \frac{1}{\tau}\right).$$

Proof. We closely follow the proof of [Lun09, Proposition 4.1.7]. Let $x \in \mathrm{dom}_X(A^\tau)$. According to Proposition 8.42 we have

$$\|x\|_{\tau,s} \le (2+M) \sup_{t>0} t^\tau \|AR(-t,A)x\|_X.$$

Using the representation formula

$$A^{-\tau-1}x = \frac{1}{\Gamma(1-\tau)\Gamma(1+\tau)} \int_0^\infty z^{-\tau} R(-z,A)^2 x\, dz,$$

see [Lun09, Equation (4.1.8)], we obtain

$$AR(-t,A)x = \frac{A^2R(-t,A)}{\Gamma(1-\tau)\Gamma(1+\tau)} \int_0^\infty z^{-\tau} R(-z,A)^2 A^\tau x\, dz.$$

Moreover, for any $t > 0$ we estimate

$$\|AR(-t,A)x\|_X \le \frac{M}{1+t} \int_0^t z^{-\alpha}(M+1)^2 \|A^\tau x\|_X + (M+1) \int_t^\infty z^{-\alpha} \frac{M(M+1)}{1+z} \|A^\tau x\|_X.$$

Hence,

$$t^\tau \|AR(-t,A)x\|_X \le \frac{M(M+1)^2}{\Gamma(1-\tau)\Gamma(1+\tau)} \left(\frac{t}{1+t}\frac{1}{1-\tau} + \frac{1}{\tau}\right) \|A^\tau x\|_X$$

concluding the proof. □

In the Hilbert space case, we can give the following embedding.

Proposition 8.48 *Let A be a sectorial, self-adjoint operator on a Hilbert space H. Then*

$$\mathrm{dom}_H(A^\tau) \hookrightarrow (H, \mathrm{dom}_H(A))_{\tau,2},$$

where the embedding constant is bounded by

$$1 + (-2\cos(\pi\tau)\Gamma(-2\tau))^{1/2}.$$

Proof. Following the proof of [Tri78, Theorem 1.18.10], the constant c in step 2 is given by

$$\int_0^\infty \frac{|e^{it\mu}-1|^2}{(t\mu)^{2\tau}}\frac{\mathrm{d}t}{t} = -2\cos(\pi\tau)\Gamma(-2\tau).$$

Taking square roots yields the bound. □

The following two results were derived independently of Lucas Bonifacius. Its proof is partly based on [Lun95, Proposition 2.2.2].

Theorem 8.49 *Let X be a Banach space, A be a sectorial operator on X and $p \in (1,\infty)$. Then we have for each $x \in (X, \mathrm{dom}_X(A))_{1-\frac{1}{p},p}$ that*

$$\|Ae^{-tA}x\|_{L^p(0,1;X)} \le C\left(\frac{p}{p-1}\right)^2 \|x\|_{1-\frac{1}{p},p}.$$

C is independent of p or X and depends only on the M in the sectoriality definition of A. $(e^{-tA})_{t>0}$ denotes the analytic semigroup generated by A, see for example [Lun95, Definition 2.0.2].

Proof. Let $t \in (0,\infty)$ and the other quantities as above. We will use the trace method and Proposition 8.38 as key components. To that end let $v \in L^p_*(\mathbb{R}_+;\mathrm{dom}_X(A))$ with $v' \in L^p(\mathbb{R}_+, X)$ and $\gamma_0 v = x$. We see that

$$\begin{aligned}\|Ae^{-tA}x\|_X &\le \|Ae^{-tA}v(t)\|_X + \|Ae^{-tA}\int_0^t v'(s)\,ds\|_X \\ &\le \|e^{-tA}\|_{\mathrm{Lin}(X,X)}\|Av(t)\|_X + \|tAe^{-tA}\|_{\mathrm{Lin}(X,X)}\left\|\frac{1}{t}\int_0^t v'(s)\,ds\right\|_X.\end{aligned} \tag{8.43}$$

By [Lun95, Proposition 2.1.1.(iii)] there exists a $C > 0$ depending only on M such that

$$\|e^{-tA}\|_{\mathrm{Lin}(X,X)}, \|tAe^{-tA}\|_{\mathrm{Lin}(X,X)} \le C.$$

Integrating (8.43) on both sides from 0 to 1 yields

$$\|Ae^{-tA}x\|_{L^p(0,1;X)} \le C(\|v\|_{L^p(0,\infty;\mathrm{dom}_X(A))} + \|\left\|\frac{1}{t}\int_0^t v'(s)\,ds\right\|_X\|_{L^p(0,\infty)}).$$

By [Lun95, Corollary 1.2.9] with $\theta = \frac{1}{p}$ we find $\|\left\|\frac{1}{t}\int_0^t v'(s)\,ds\right\|_X\|_{L^p(0,\infty;X)} \le \frac{p}{p-1}\|v'(t)\|_{L^p(0,\infty;X)}$. Thus

$$\begin{aligned}\|Ae^{-tA}x\|_{L^p(0,1;X)} &\le C\frac{p}{p-1}(\|v\|_{L^p(0,\infty;\mathrm{dom}_X(A))} + \|v'(t)\|_{L^p(0,\infty;X)}) \\ &\le C\frac{p}{p-1}\left(\left\|t^{\frac{1}{p}}v\right\|_{L^p_*(\mathbb{R}_+;\mathrm{dom}_X(A))} + \left\|t^{\frac{1}{p}}v'(t)\right\|_{L^p_*(\mathbb{R}_+;X)}\right).\end{aligned}$$

As v was arbitrary with $\gamma_0 v = x$ we have by the definition of $\|x\|^{Tr}_{1-\frac{1}{p},p}$

$$\|Ae^{-tA}x\|_{L^p(0,1;X)} \le C\frac{p}{p-1}\|x\|^{Tr}_{1-\frac{1}{p},p}.$$

By Proposition 8.38 this implies

$$\|Ae^{-tA}x\|_{L^p(0,1;X)} \le C\frac{p}{p-1}\left(1+\frac{1}{1-\frac{1}{p}}\right)\|x\|_{1-\frac{1}{p},p}.$$

This implies the claim. □

The following is an important estimate for maximal parabolic regularity with tracked constants. We require it in our numerical analysis.

Corollary 8.50 *Let X be a Banach space and A be a sectorial operator on X. Assume that for a $p_0 \in (1,\infty)$ the following holds: for each $f \in L^{p_0}(0,1;X)$ there exists a $y \in L^{p_0}(0,1;X)$ such that*

$$\begin{cases} \partial_t y + Ay = f \text{ a.e. in } (0,1) \\ y(0) = 0. \end{cases}$$

Further, there needs to exists a $C_{p_0} > 0$, independent of X, such that

$$\|y\|_{L^{p_0}(0,1;X)} \leq C_{p_0} \|f\|_{L^{p_0}(0,1;X)}.$$

This is called well-posedness in the sense of [AS94, Definition 3.1].

Then we have the following for each $p \in (1,\infty)$: for all $y_0 \in (X, \mathrm{dom}_X(A))_{1-\frac{1}{p},p}$ and $f \in L^p(0,1;X)$ there exist a $y \in L^p(0,1;X)$ such that

$$\begin{cases} \partial_t y + Ay = f \text{ a.e. in } (0,1) \\ y(0) = y_0. \end{cases}$$

Further y satisfies

$$\|\partial_t y\|_{L^p(0,1;X)} + \|Ay\|_{L^p(0,1;X)} \leq C\left(\frac{p^2}{p-1}\|f\|_{L^p(0,1;X)} + \left(\frac{p}{p-1}\right)^2 \|y_0\|_{(X,\mathrm{dom}_X(A))_{1-\frac{1}{p},p}}\right).$$

C depends only p_0 and M, but not on X or p.

Proof. Let $p \in (1,\infty)$ and f, y_0 as above. By linearity we may consider the following two equations seperately:

$$\begin{cases} \partial_t y_f + Ay_f = f \text{ a.e. in } (0,1) \\ y_f(0) = 0. \end{cases} \qquad \begin{cases} \partial_t y_{y_0} + Ay_{y_0} = 0 \text{ a.e. in } (0,1) \\ y_{y_0}(0) = y_0. \end{cases}$$

So, for the moment assume $y_0 = 0$. By [AS94, Theorem 3.2] there exists a solution y_f, thanks to our assumption, and it satisfies

$$\|y_f\|_{L^p(0,1;X)} \leq C_{p_0} \frac{p^2}{p-1} \|f\|_{L^p(0,1;X)}.$$

Now we consider the case $f = 0$. By [Lun95, Proposition 2.1.1] we find $y_{y_0} = e^{-tA} y_0$. Therefore, by Theorem 8.49,

$$\|\partial_t y_{y_0}\|_{L^p(0,1;X)} = \|Ay_{y_0}\|_{L^p(0,1;X)} \leq C\left(\frac{p}{p-1}\right)^2 \|y_0\|_{(X,\mathrm{dom}_X(A))_{1-\frac{1}{p},p}}.$$

Adding y_f and y_{y_0} yields the solution with the the desired estimates. □

Proposition 8.51 *For all $p \in (1,\infty)$ and $\tau \in (0,1)$ such that $d/(2p) < \tau$ and that A is sectorial on $L^p(\Omega)$ it holds*

$$(L^p(\Omega), \mathrm{dom}_p(A))_{\tau,1} \hookrightarrow C(\overline{\Omega}).$$

Moreover, the embedding constant is bounded by

$$\frac{c(M+1)^2 \Gamma(\tau - N/(2p))}{\Gamma(\tau)}$$

with $c > 0$ is independent of τ or p. M is the sectoriality constant of A from Definition 8.43.

Proof. According to Proposition 8.45 we have

$$(L^p(\Omega), \mathrm{dom}_p(A))_{\tau,1} \hookrightarrow \mathrm{dom}_p(A^\tau).$$

Note that the embedding constant can be bounded independently of τ or p; see Remark 8.46. As in the proof of [DtER15, Theorem 2.10 c)], for $\omega > 0$ to be specified later, we use the integral representation of the fractional operator

$$(A+\omega+1)^{-\tau} = \frac{1}{\Gamma(\tau)} \int_0^\infty t^{\tau-1} e^{-t(A+\omega+1)}\, dt;$$

see, e.g. [Paz83, Equation (6.9), Chapter 2]. Employing [DtER15, Theorem 2.10 b)], there are $c > 0$ and $\omega > 0$ such that for $\kappa > 0$ sufficiently small we find

$$\|u\|_{C^\kappa(\Omega)} \leq \frac{cM^2}{\Gamma(\tau)} \int_0^\infty t^{\tau-1} t^{-N/(2p)-\kappa/2} e^{-t} \|(A+\omega+1)^\tau u\|_{L^p(\Omega)}\, dt,$$

where the constants c and ω are independent of κ, p, and τ. For the integral we have the expression

$$\int_0^\infty t^{\tau-1-N/(2p)-\kappa/2} e^{-t}\, dt = \Gamma(\tau - N/(2p) - \kappa/2)$$

and thus arrive at

$$\|u\|_{C^\kappa(\Omega)} \leq \frac{c\Gamma(\tau - N/(2p) - \kappa/2)}{\Gamma(\tau)} \|(A+\omega+1)^\tau u\|_{L^p(\Omega)}.$$

c is still independent of κ,p and τ.

We will show that

$$\|(A+\omega+1)^\tau u\|_{L^p(\Omega)} \leq (M+1)^2(\omega+1)\|u\|_{\mathrm{dom}_p(A^\tau)}, \tag{8.44}$$

and thus show

$$\|u\|_{C(\bar{\Omega})} \leq \frac{c(M+1)^2\Gamma(\tau - N/(2p) - \kappa/2)}{\Gamma(\tau)} \|\mathrm{dom}_p(A^\tau)\|_{L^p(\Omega)}.$$

Finally, going to the limit $\kappa \to 0$ yields the bound on the embedding constant as specified in the proposition.

It remains to show (8.44). Tracking the constant through the proof of [Lun09, Lemma 4.1.11] shows that we have

$$\|A^\tau u - (A+\omega+1)^\tau u\|_X \leq \frac{\sin(\pi\tau)}{\pi} \left(\frac{M^2}{1-\tau} + \frac{(1+M)}{\tau} \right) (\omega+1)^\tau \|u\|_X.$$

By elementary calculus von has $|\sin(\pi\tau)/\tau|, |\sin(\pi\tau)/(1-\tau)| \leq \pi$ and thus

$$\|A^\tau u - (A+\omega+1)^\tau u\|_X \leq (M+1)^2(\omega+1)\|u\|_X.$$

This finally implies (8.44):

$$\|(A+\omega+1)^\tau u\|_X \leq (M+1)^2(\omega+1)\|u\|_X + \|A^\tau u\|_X \leq (M+1)^2(\omega+1)(\|u\|_X + \|A^\tau x\|_X).$$

□

Remark 8.52 It is worth mentioning that Proposition 8.51 holds for fairly general domains and divergence form operators even with mixed boundary conditions. In case of homogeneous Dirichlet conditions [DtER15, Assumptions 2.3 and 2.5] are vacuously true. Moreover, [DtER15, Assumptions 2.4] requires the Dirichlet boundary part to be a $(N-1)$-set; see [JW84, Chapter II]. Since Ω is a Lipschitz domain and there is no Neumann boundary part, from [MR16, Theorem 4.3] we conclude that $\partial\Omega$ is a $(N-1)$-set. Furthermore, [DtER15] deals with operators of the form $A = -\nabla \cdot \mu\nabla$, where μ is a uniformly elliptic and essentially bounded coefficient function that is clearly satisfied in our setting.

The following result is our replacement for [Bon18, Remark A.17]. While [Bon18] works with the Laplacian, there the domain restrictions are weak. We, however, allow for a more general operator, while posing restrictions on the domain. This is not a problem for our application as the required regularity has to be assumed anyway.

Lemma 8.53 *Let Ω be a C^2-domain and let A be an elliptic, symmetric operator in divergence form with coefficients in $C^{0,1}(\Omega)$, equipped with homogenous Dirichlet boundary conditions. Then for any $\gamma_0 \in (0, \frac{\pi}{2})$ there is a $M > 0$ such that for any $p \in [1, \infty]$ one has*

$$\|(\lambda \operatorname{Id} - A)^{-1} v\|_{L^p(\Omega,\mathbb{C})} \leq \frac{M}{1+|\lambda|}\|v\|_{L^p(\Omega,\mathbb{C})}$$

for any $\lambda \notin \Sigma_{y_0} := \{z \in \mathbb{C} : |\arg z| \leq \gamma_0\}$ and $v \in L^p(\Omega, \mathbb{C})$.

Proof. The proof for $p = 2$ is the same as in the discrete case, see Corollary 4.37. We will now show the case $p = \infty$. Then the rest of the estimates for $p \neq 2, \infty$ follows by interpolation and duality by the same arguments as in Corollary 4.37.

The statement holds true for $p = \infty$, provided $v \in C(\bar{\Omega}, \mathbb{C})$, by Proposition 4.32. Now let $v \in L^\infty(\Omega, \mathbb{C})$ be arbitrary, then we approximate v by a sequence of smooth functions. We introduce a family of mollifiers $(\eta_\epsilon)_{\epsilon>0} \subset C_c^\infty(\Omega)$ satisfying for any $\epsilon > 0$

$$\operatorname{supp}(\eta_\epsilon) \subset \bar{B}_\epsilon(0), \quad \eta_\epsilon \geq 0, \quad \int_\Omega \eta_\epsilon(t)\ dt = 1.$$

Such a family is well-known to exist. It is also well-known that for each $\epsilon > 0$ we have

$$v_\epsilon(y) := (v * \eta_\epsilon)(y) = \int_\Omega v(x)\eta_\epsilon(y-x)\ dy \in C_c^\infty(\mathbb{R}^N)$$

where we implicitly extend v by 0 onto $\mathbb{R}^N \setminus \Omega$. It is well-known that $v_\epsilon \xrightarrow{\epsilon \to 0} v$ in $L^2(\Omega)$ and we can also immediately see that $\|v_\epsilon\|_{L^\infty(\Omega,\mathbb{C})} \leq \|v\|_{L^\infty(\Omega,\mathbb{C})}$. Thus we have by Proposition 4.32:

$$\|(\lambda \operatorname{Id} - A)^{-1} v_\epsilon\|_{L^\infty(\Omega,\mathbb{C})} \leq \frac{M}{1+|\lambda|}\|v\|_{L^\infty(\Omega,\mathbb{C})}$$

for any $\lambda \notin \Sigma_{y_0} := \{z \in \mathbb{C} : |\arg z| \leq \gamma_0\}$. Now fix $\lambda \notin \Sigma_{y_0}$. By Banach-Alaoglu there exists a sequence $y_{\epsilon_n} := (\lambda \operatorname{Id} - A)^{-1} v_{\epsilon_n}$ converging weakly* in $L^\infty(\Omega, \mathbb{C})$ to some limit $y \in L^\infty(\Omega, \mathbb{C})$. For details see the proof of Lemma 3.22. Then we have by the weakly lower semi-continuity of the norm that

$$\|y\|_{L^\infty(\Omega,\mathbb{C})} \leq \frac{M}{1+|\lambda|}\|v\|_{L^\infty(\Omega,\mathbb{C})}.$$

We will show $y = (\lambda \operatorname{Id} - A)^{-1} v$. Then the desired statement follows. Just as in the proof of Corollary 4.37 one can show

$$\|y_{\epsilon_n}\|_{H^1(\Omega,\mathbb{C})} \leq C\|v_{\epsilon_n}\|_{L^2(\Omega,\mathbb{C})}.$$

By the L^2-convergence of the $(v_\epsilon)_{\epsilon>0}$ the right hand side stays bounded and thus a subsequence of $(y_{\epsilon_n})_{n\in\mathbb{N}}$, denoted by the same indices, converges weakly to y in $H_0^1(\Omega,\mathbb{C})$, implying in particular $y \in H_0^1(\Omega,\mathbb{C})$. Thus for any $\varphi \in H_0^1(\Omega,\mathbb{C}) \subset L^1(\Omega,\mathbb{C})$ we have that

$$0 = a_\Omega(y_{\epsilon_n},\bar{\varphi}) - \lambda\,(y_{\epsilon_n},\bar{\varphi})_{L^2(\Omega)} - (-v_{\epsilon_n},\bar{\varphi})_{L^2(\Omega)}$$

converges to

$$0 = a_\Omega(y,\bar{\varphi}) - \lambda\,(y,\bar{\varphi})_{L^2(\Omega)} - (-v,\varphi)_{L^2(\Omega)}$$

for $n \to \infty$. Thus $y = (A - \lambda\,\mathrm{Id})^{-1}(-v) = (\lambda\,\mathrm{Id} - A)^{-1}v$. □

Remark 8.54 The very important and preceeding Lemma 8.53 entails that for those operators and domains the results of Section 8.7 are applicable for any $X = L^p(\Omega)$ with embedding constants that do not depend on p.

8.8 Fractional Sobolev spaces

We summarize well-known properties of fractional Sobolev spaces that are also called Sobolev-Slobodeckij spaces. For more details, we refer to the monograph [AF03, Chapter 7]; see also [DNPV12] for an introduction to this topic.

Let $\Omega \subseteq \mathbb{R}^N$ be an open set. For $\theta \in (0,1)$ and $p \in [1,\infty)$ we define

$$[f]_{\theta,p,\Omega} := \left(\int_\Omega\int_\Omega \frac{|f(x)-f(y)|^p}{|x-y|^{d+\theta p}}\,dx\,dy\right)^{1/p}$$

the Gagliardo (semi)norm of f and define the norm of the fractional Sobolev space on Ω denoted $W^{\theta,p}(\Omega)$ by

$$\|f\|_{W^{\theta,p}(\Omega)} := \left(\|f\|^p_{L^p(\Omega)} + [f]^p_{\theta,p,\Omega}\right)^{1/p}.$$

If $\theta > 1$ and θ is not an integer, then write $\theta = m + \sigma$ with $m \in \mathbb{N}$ and $\sigma \in (0,1)$. In this case the norm of $W^{\theta,p}(\Omega)$ is given by

$$\|f\|_{W^{\theta,p}(\Omega)} := \left(\|f\|^p_{W^{m,p}(\Omega)} + \sum_{|\alpha|=m}[D^\alpha f]^p_{\sigma,p,\Omega}\right)^{1/p}.$$

Here, α denotes the multindex and $|\alpha| = \sum \alpha_j$. We emphasize that the fractional Sobolev norm does not reproduce the (classical) Sobolev norm in the limit cases $\theta \to k$ with $k \in \mathbb{N}$; cf. [BBM01, Remark 5].

For the point-wise error estimates we require the embedding of the real interpolation space between Sobolev spaces into the fractional Sobolev space. To clearly see the dependencies of the constants, we give an independent proof that relies on elementary arguments. Note that in the following even equality holds (up to equivalent norms), but we only need one injection.

Proposition 8.55 *For any $p \in [1,\infty)$ and $\theta \in (0,1)$ one has*

$$(W^{m,p}(\mathbb{R}^N), W^{m+1,p}(\mathbb{R}^N))_{\theta,p} \hookrightarrow W^{m+\theta,p}(\mathbb{R}^d), \quad m \in \mathbb{N}, \tag{8.45}$$

where the embedding constant is bounded by

$$\left(\min\{\,\theta, 1-\theta\,\}\,p + 2^{2p}c_N\right)^{1/p},$$

and c_N exclusively depends on the spatial dimension N.

Proof. We follow the proof of [Lun09, Example 1.1.8]; cf. also [AF03, Theorem 7.47].

Step 1: $m = 0$. Let $u \in (L^p(\mathbb{R}^N), W^{1,p}(\mathbb{R}^N))_{\theta,p}$. Consider a splitting $u = v + w$ with $v \in L^p(\mathbb{R}^N)$ and $w \in W^{1,p}(\mathbb{R}^N)$. Recall that

$$\int_{\mathbb{R}^N} |w(x+h) - w(x)|^p \, dx \le |h|^p \|\nabla w\|_{L^p}^p.$$

Therefore, using Jensen's inequality twice,

$$\begin{aligned}
[u]_{\theta,p}^p &\le 2^{p-1} \int_{\mathbb{R}^d} \int_{\mathbb{R}^N} \frac{|v(x+h) - v(x)|^p}{|h|^{d+\theta p}} \, dx \, dh + 2^{p-1} \int_{\mathbb{R}^d} \int_{\mathbb{R}^N} \frac{|w(x+h) - w(x)|^p}{|h|^{d+\theta p}} \, dx \, dh \\
&\le \int_{\mathbb{R}^N} |h|^{-N-\theta p} \left(2^{2p-2} \|v\|_{L^p}^p + 2^{p-1} |h|^p \|w\|_{W^{1,p}}^p \right) dh \\
&\le 2^{2p-2} \int_{\mathbb{R}^N} |h|^{-N-\theta p} \left(\|v\|_{L^p} + |h| \|w\|_{W^{1,p}} \right)^p \, dh.
\end{aligned}$$

Hence, by means of the definition of the K-functional, we obtain

$$\begin{aligned}
[u]_{\theta,p}^p &\le 2^{2p-2} \int_{\mathbb{R}^N} |h|^{-N-\theta p} K(|h|, u)^p \, dh \\
&\le 2^{2p-2} \int_0^\infty t^{-1-\theta p} K(t,u)^p \, dt \int_{\partial B_1(0)} d\sigma = 2^{2p-2} c_d \|u\|_{\theta,p}^p,
\end{aligned}$$

where the constant c_N exclusively depends on the spatial dimension N. Furthermore, we have

$$\|u\|_{L^p} \le \|u\|_{L^p + W^{1,p}} = K(1, u, L^p, W^{1,p}) \le \|u\|_{\theta,\infty} \le (p \min\{\theta, 1-\theta\})^{1/p} \|u\|_{\theta,p},$$

due to Proposition 8.34. Hence,

$$\|u\|_{W^{\theta,p}} = \left(\|u\|_{L^p}^p + [u]_{\theta,p}^p \right)^{1/p} \le \left(p \min\{\theta, 1-\theta\} + 2^{2p} c_d \right)^{1/p} \|u\|_{\theta,p}.$$

Step 2: $m \ge 1$. The general case $m \ge 1$ follows by analogous arguments as above, where we simply replace the space L^p by $W^{m,p}$ and $W^{1,p}$ by $W^{m+1,p}$. Moreover, we estimate the seminorm $[D^\alpha u]_{\theta,p}$ instead of $[u]_{\theta,p}$. □

Lemma 8.56 *For any $p \in [1,\infty)$ and $\theta \in (0,1) \setminus \{\frac{1}{2}\}$ one has*

$$(L^p(\mathbb{R}^N), W^{2,p}(\mathbb{R}^N))_{\theta,p} \hookrightarrow W^{2\theta,p}(\mathbb{R}^d).$$

Furthermore, the embedding constant $c(\tau)$ is uniform in $p \in [1,\infty)$ and satisfies $c(\tau) \sim (1-\theta)^{-1}$ as $\theta \to 1$ and $c(\tau) \sim |1/2 - \theta|^{-1}$ as $\theta \to 1/2$.

Proof. According to [Maz85, Corollary 1.4.7.1] it holds

$$\|\nabla u\|_{L^p} \le c_p \|u\|_{W^{2,p}}^{1/2} \|u\|_{L^p}^{1/2}, \quad \text{for all } u \in W^{2,p}(\mathbb{R}^d),$$

where $c_p \le c K_N^{1/p}$ and K_d denotes the volume of the N dimensional unit ball. Thus,

$$\|u\|_{W^{1,p}} \le (1 + c K_N^{1/p}) \|u\|_{W^{2,p}}^{1/2} \|u\|_{L^p}^{1/2}, \quad u \in W^{2,p}(\mathbb{R}^N).$$

Whence, the space $W^{1,p}(\mathbb{R}^N)$ belongs to the class $J_{1/2}(L^p(\mathbb{R}^d), W^{2,p}(\mathbb{R}^N))$. For these reasons, if $\theta > 1/2$, then the reiteration theorem Proposition 8.40 (with $\theta_0 = 1/2$ and $\theta_1 = 1$) and the embedding (8.45) imply

$$(L^p, W^{2,p})_{\theta,p} \hookrightarrow (W^{1,p}, W^{2,p})_{2\theta-1,p} \hookrightarrow W^{2\theta,p}.$$

Similarly, if $\theta < 1/2$, then the reiteration theorem (with $\theta_0 = 0$ and $\theta_1 = 1/2$) yields

$$(L^p, W^{2,p})_{\theta,p} \hookrightarrow (L^p, W^{1,p})_{2\theta,p} \hookrightarrow W^{2\theta,p}.$$

Moreover, the embedding constants from the reiteration theorem are bounded by

$$(2\theta-1)^{-1}2^{1+1/p}(c_0 + c_1(2-2\theta)^{-1})(1+3\theta^{-1})2(1-2\theta)^{-1}\left(2+(1-2\theta)^{-1}\right)$$

in the first case, and by

$$(2\theta)^{-1}2^{1+1/p})(c_0 + c_1(1-2\theta)^{-1})(1+3\theta)\theta^{-1}(2+(2\theta)^{-1})$$

in the second case. With the constant from Proposition 8.55 we obtain the asymptotic behavior of the embedding constant as stated in the proposition. □

Proposition 8.57 *Let $\omega' \subset \mathbb{R}^N$ be a domain with a Lipschitz boundary. For all $\theta \in (0,1) \setminus \{\frac{1}{2}\}$ and $p \in (1,\infty)$ it holds*

$$(L^p(\omega'), W^{2,p}(\omega'))_{\theta,p} \hookrightarrow W^{2\theta,p}(\omega').$$

Furthermore, the embedding constant $c(\theta)$ is uniform in $p \in [1,\infty)$ and satisfies $c(\theta) \sim (1-\theta)^{-1}$ as $\theta \to 1$ and $c(\theta) \sim |1/2-\theta|^{-1}$ as $\theta \to 1/2$.

Before we give the proof we need to restate the Stein extension theorem.

Lemma 8.58 *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary and $m \in \mathbb{N}$. Then there exists an extension operator E mapping $W^{k,p}(\Omega)$ into $W^{k,p}(\mathbb{R}^N)$ for all $k = 0, \dots, m$ and $p \in [1,\infty)$. Moreoever there is a $c > 0$ independent of p, k, f such that*

$$\|Ef\|_{W^{k,p}(\mathbb{R}^N)} \le c\|f\|_{W^{k,p}(\Omega)} \qquad \forall f \in W^{k,p}(\Omega).$$

Proof. The result is proved in [Ste70, Theorem VI.3.5]. The bound on the norm of E as stated above can be found in [Ste70, Chapter VI.3, Equation (32)]. □

proof of Proposition 8.57. The proof is based on the corresponding result on $\mathbb{R}^N$ by first extending the functions from ω' to $\mathbb{R}^d$ and retraction afterwards; cf., e.g. [BS08, Equation 14.2.4].

According to Lemma 8.58 there exists an extension operator $E\colon W^{k,p}(\omega') \to W^{k,p}(\mathbb{R}^N)$ for all $k = 0, 1, 2$ and its norm is independent of p. Hence

$$\begin{aligned}\|f\|_{W^{2\tau,p}(\omega')} = \|Ef\|_{W^{2\tau,p}(\omega')} &\le \|Ef\|_{W^{2\tau,p}(\mathbb{R}^N)}\\ &\le c(\tau)\|Ef\|_{(L^p(\mathbb{R}^N),W^{2,p}(\mathbb{R}^N))_{\tau,p}}\\ &\le c(\tau)\|f\|_{(L^p(\mathbb{R}^N),W^{2,p}(\mathbb{R}^N))_{\tau,p}},\end{aligned}$$

where we have used the interpolation result Lemma 8.56 on $\mathbb{R}^N$ in the second inequality and a general interpolation principle for linear operators, see, e.g. [Tri78, Section 1.2.2], in the last inequality. Note that for the above estimate it is essential that the extension operator E is the same for $k = 0$ and $k = 2$ in order to interpolate operators. □

The following result was derived independently of Lucas Bonifacius, but we clearly use related techniques and ideas.

Lemma 8.59 *Let Ω be a Lipschitz domain. For $p \in [1, \infty)$, $s \in (0, 1)$ with $s > \frac{N}{p}$ we have the embedding*

$$W^{1+s,p}(\Omega) \hookrightarrow C^{1,\alpha}(\Omega)$$

with embedding constant uniform in s and p and $\alpha = s - \frac{N}{p}$.

Proof. The proof consist of two parts. It requires the use of Campanato spaces, but we will not require any prior knowledge of those. An extensive discussion can be found in [Gia83] and [Mor66]. For $p, \lambda \in [1, \infty)$ and $f \in L^p(\Omega)$ we define

$$|f|^p_{L^{p,\lambda}(\Omega)} := \sup_{a \in \Omega, 0 < r < \Omega} r^{-\lambda} \int_{B_r(a) \cap \Omega} |f(x) - f_{a,r}|^p \, dx.$$

Here we use the mean value

$$f_{a,r} := \frac{1}{B_r(a) \cap \Omega} \int_{B_r(a) \cap \Omega} f(x) \, dx.$$

We now show that $|f|_{L^{p,sp}(\Omega)} \leq C_{\not p, \not s} \|f\|_{W^{s,p}(\Omega)}$. After that we will use $|f|_{L^{p,sp}(\Omega)}$ to estimate the Hölder norm of f. The first part of the proof is very close to [DNPV12, Theorem 8.2]. We introduce the abbreviation $\Omega_r(a) := B_r(a) \cap \Omega$ for $r > 0$ and $a \in \Omega$. For any $\xi \in \mathbb{R}$ we have after an application of Hölder's inequality

$$|\xi - (f)_{a,r}| = \frac{1}{|\Omega_r(a)|^p} \left| \int_{\Omega_r(a)} \xi - f(y) \, dy \right|^p \leq \frac{1}{|\Omega_r(a)|} \int_{\Omega_r(a)} |\xi - f(y)|^p \, dy.$$

Setting $\xi = f(x)$ and integrating yields

$$\begin{aligned} \int_{\Omega_r(a)} |f(x) - (f)_{a,r}|^p \, dx &\leq \frac{1}{|\Omega_r(a)|} \int_{\Omega_r(a)} \int_{\Omega_r(a)} |f(x) - f(y)|^p \, dy \, dx \\ &\leq \frac{(2r)^{N+sp}}{|\Omega_r(a)|} \int_{\Omega_r(a)} \int_{\Omega_r(a)} \frac{|f(x) - f(y)|^p}{|x - y|^{N+sp}} \, dy \, dx. \end{aligned} \tag{8.46}$$

Because Ω is a Lipschitz domain there is a $c > 0$ independent of a and r (and of course s and p) such that

$$|\Omega_r(a)| \geq c r^N; \tag{8.47}$$

see for example the remark after [Gia83, Chapter III, Definition 1.3]. Thus (8.46) yields

$$\int_{\Omega_r(a)} |f(x) - (f)_{a,r}|^p \, dx \leq \frac{2^{N+sp} r^{sp}}{c} \|f\|^p_{W^{s,p}(\Omega)}.$$

This implies

$$|f|_{L^{p,sp}(\Omega)} \leq \frac{1}{c^{\frac{1}{p}}} 2^{\frac{N}{p}+s} \|f\|_{W^{s,p}(\Omega)} \leq \frac{1}{\min(c, 1)} 2^{\frac{N}{p}+s} \|f\|_{W^{s,p}(\Omega)}.$$

Because $2^{\frac{N}{p}+s} \leq 2^{N+1}$ we have

$$|f|_{L^{p,sp}(\Omega)} \leq C_{\not s, \not p} \|f\|_{W^{s,p}(\Omega)}. \tag{8.48}$$

We now continue very closely to [Gia83, Chapter III, Theorem 1.2] and track the appearing constants carefully. For $0 < r < R$ we have

$$|(f)_{a,r} - (f)_{a,R}|^p \leq 2^{p-1} \left(|f(x) - (f)_{a,r}|^p + |f(x) - (f)_{a,R}|^p \right).$$

Using (8.47) again yields after integration over $\Omega_r(a)$

$$\begin{aligned} cr^N |(f)_{a,r} - (f)_{a,R}|^p &\le 2^{p-1} \left(\int_{\Omega_r(a)} |f(x) - (f)_{a,r}|^p \, dx + \int_{\Omega_R(a)} |f(x) - (f)_{a,R}|^p \, dx \right) \\ &\le 2^{p-1} \left(r^{sp} |f|^p_{L^{p,sp}(\Omega)} + R^{sp} |f|^p_{L^{p,sp}(\Omega)} \right) \le 2^p R^{sp} |f|^p_{L^{p,sp}(\Omega)}. \end{aligned}$$

Dividing by cr^N and taking the p-th root on both sides yields

$$|(f)_{a,r} - (f)_{a,R}| \le \frac{2}{c^{\frac{1}{p}}} r^{-\frac{N}{p}} R^s |f|_{L^{p,sp}(\Omega)} \le \frac{2}{\min(c,1)} r^{-\frac{N}{p}} R^s |f|_{L^{p,sp}(\Omega)} = C_{\not{s},\not{p}} r^{-\frac{N}{p}} R^s |f|_{L^{p,sp}(\Omega)}.$$

Setting $R_j := 2^{-j} R$ for $j \in \mathbb{N}$ implies for $j \le k \in \mathbb{N}$.

$$\begin{aligned} |(f)_{a,R_j} - (f)_{a,R_k}| &\le C_{\not{s},\not{p}}\, R_k^{-\frac{N}{p}} R_j^s |f|_{L^{p,sp}(\Omega)} = C_{\not{s},\not{p}}\, R_j^{s-\frac{N}{p}} 2^{-\frac{N}{p}(k-j)} |f|_{L^{p,sp}(\Omega)} \\ &\le C_{\not{s},\not{p}}\, R_j^{s-\frac{N}{p}} |f|_{L^{p,sp}(\Omega)}. \end{aligned} \tag{8.49}$$

Thus $((f)_{a,R_k})_{k\in\mathbb{N}}$ is a Cauchy sequence, since we assume $s > N/p$, and we therefore can define the limit

$$\tilde{f}(a) := \lim_{k\to\infty} (f)_{a,R_k}.$$

By the Lebesgue differentiation theorem, cf. [GM09, Theorem 2.16], we have $f(a) = \tilde{f}(a)$ a.e. in Ω. Therefore taking the limit $k \to \infty$ in (8.49) and setting $j = 0$ yields

$$|(f)_{a,R} - f(a)| \le C_{\not{s},\not{p}} R^{s-\frac{N}{p}} |f|_{L^{p,sp}(\Omega)}. \tag{8.50}$$

We can now show Hölder continuity of f. For $x, y \in \Omega$ and $R := |x - y|$ we have

$$|f(x) - f(y)| \le |(f)_{x,2R} - f(x)| + |(f)_{x,2R} - (f)_{y,2R}| + |(f)_{y,2R} - f(y)|. \tag{8.51}$$

The first and third term are estimated by (8.50). For the middle term we integrate over $\Omega_{2R}(x) \cap \Omega_{2R}(y)$ to get

$$\begin{aligned} \int_{\Omega_{2R}(x)\cap\Omega_{2R}(y)} |(f)_{x,2R} - (f)_{y,2R}| \, dz &\le \int_{\Omega_{2R}(x)\cap\Omega_{2R}(y)} |(f)_{x,2R} - f(z)| + |f(z) - (f)_{y,2R}| \, dz \\ &\le \int_{\Omega_{2R}(x)} |(f)_{x,2R} - f(z)| \, dz + \int_{\Omega_{2R}(y)} |f(z) - (f)_{y,2R}| \, dz \\ &\le |\Omega_{2R}(x)|^{1-\frac{1}{p}} \left(\int_{\Omega_{2R}(x)} |(f)_{x,2R} - f(z)|^p \, dz \right)^{\frac{1}{p}} \\ &\quad + |\Omega_{2R}(y)|^{1-\frac{1}{p}} \left(\int_{\Omega_{2R}(y)} |(f)_{y,2R} - f(z)|^p \, dz \right)^{\frac{1}{p}} \\ &\le 2(2R)^{N-\frac{N}{p}} (2R)^s |f|_{L^{p,sp}(\Omega)} \le 2^{N+2} R^{N-\frac{N}{p}+s} |f|_{L^{p,sp}(\Omega)}. \end{aligned}$$

This implies

$$|(f)_{x,2R} - (f)_{y,2R}| \le |\Omega_{2R}(x) \cap \Omega_{2R}(y)|^{-1} 2^{N+2} R^{N-\frac{N}{p}+s} |f|_{L^{p,sp}(\Omega)}.$$

By construction $\Omega_R(x) \subset \Omega_{2R}(x) \cap \Omega_{2R}(y)$ and thus by (8.47) we have

$$|\Omega_{2R}(x) \cap \Omega_{2R}(y)|^{-1} \le c^{-1} R^{-N}$$

and therefore

$$|(f)_{x,2R} - (f)_{y,2R}| \leq C_{\not{s},\not{p}} R^{s-\frac{N}{p}} |f|_{L^{p,sp}(\Omega)}. \tag{8.52}$$

Inserting (8.52) and (8.50) into (8.51) yields

$$|f(x) - f(y)| \leq C_{\not{s},\not{p}}\, R^{s-\frac{N}{p}} |f|_{L^{p,sp}(\Omega)}$$

Thus with $\alpha := s - \frac{N}{p} > 0$, the earlier choice of $R = |x-y|$ and (8.48) we have shown:

$$|f|_{C^{0,\alpha}(\Omega)} \leq C_{\not{s},\not{p}} \|f\|_{W^{s,p}(\Omega)}.$$

It remains to restimate $\|f\|_{L^\infty(\Omega)}$. Let $x \in \Omega$ arbitrary. Let $y \in \Omega$ such that $f(y) = |\Omega|^{-1} \int_\Omega f(z)\, dz$. This is possible because f is continuous and Ω is connected. Then

$$|f(x)| \leq |f(x) - f(y)| + |f(y)| \leq C_{\not{s},\not{p}} \|f\|_{W^{s,p}(\Omega)} + |\Omega|^{-1} \|f\|_{L^1(\Omega)}.$$

Because $\|f\|_{L^1(\Omega)} \leq C\|f\|_{W^{s,p}(\Omega)}$, where C can be chosen independently of s and p, we have shown:

$$\|f\|_{L^\infty(\Omega)} \leq C_{\not{s},\not{p}} \|f\|_{W^{s,p}(\Omega)}.$$

□

Symbols

General

$\mathbb{N}$	natural numbers; starts at 1
$\mathbb{N}_0$	natural numbers including 0
a.e.	abbreviation for 'almost everywhere'
N	Spatial dimension; at least 1; sometimes restricted to 2 or 3
$\lvert\cdot\rvert, \lvert\cdot\rvert_p$	the euclidean and the p-Norm on $\mathbb{R}^N$; with one exception: $\lvert\cdot\rvert$ can also refer to the Lebesgue measure of a set

Domains, Space and Time

supp	the support of a measure or function	p. 11
Ω	a domain in $\mathbb{R}^N$; usually exhibits higher regularity	
Γ_D, Γ_N	refers to the Dirichlet boundary part, respectively, the Neumann boundary parts of $\partial\Omega$	p. 27
T	a positive, real number; the endtime	
I	time intervall $I := (0, T)$	
Σ_D, Σ_N	time-space Dirichlet and Neumann boundaries; $\Sigma_D := I \times \Gamma_D$, $\Sigma_N := I \times \Gamma_N$	p. 27

Function Spaces

$C^k(\Omega)$	for $k \in \mathbb{N}$, functions that are k-times differentiable on Ω	
$C^\alpha(\Omega)$	for $\alpha \in (0, 1]$, functions that α-Hölder continuous on Ω	
$C^k(\bar{\Omega})$	for $k \in \mathbb{N}$, functions that are k-times differentiable on Ω, where all derivatives can be extended to $\bar{\Omega}$	
$C^{k,\alpha}(\Omega)$	functions that are k-times differentiable on Ω; The k-th derivative has α-Hölder regularity; here $k \in \mathbb{N}_0$ and $\alpha \in (0, 1]$; all derivatives can be extended to $\bar{\Omega}$, i.e. this set coincides with the set $C^{k,\alpha}(\bar{\Omega})$	
$C^{\alpha,\beta}(Q)$	continuous function on $\bar{Q}$ that are Hölder continuous with exponent $\alpha \in (0, 1)$ in the first component and Hölder continuous with exponent $\beta \in (0, 1)$ in all but the first component	p. 22
$C_c^\infty(\Omega)$	infinitely differentiable functions with compact support in a domain Ω	
$C_0(\Omega)$	continuous functions with 0 boundary value	

$L^p(\Omega), W^{1,p}(\Omega)$	real Lebesgue and Sobolev spaces for $p \in [1,\infty]$ and a domain $\Omega \subset \mathbb{R}^N$	p. 13, 14
$L^p(\Omega,\mathbb{C})$	complex Lebesgue spaces for $p \in [1,\infty]$ and a domain $\Omega \subset \mathbb{R}^N$	p. 13
$W^{1,p}(\Omega,\mathbb{C})$	complex Sobolev spaces for $p \in [1,\infty]$ and a domain $\Omega \subset \mathbb{R}^N$	p. 14
$\|\cdot\|_{L^p(\Omega)}$	real $L^p(\Omega)$ norm	p. 13
$\|\cdot\|_{L^p(\Omega,\mathbb{C})}$	complex $L^p(\Omega)$ norm	p. 13
H	a Hilbert space; mostly utilized as an abbreviation of the commonly used $L^2(\Omega) = H$	p. 25, 27
V	a Hilbert space contained in a larger Hilbert space H; mostly utilized as an abbreviation of the commonly used $H^1(\Omega)$ equipped with Dirichlet data	p. 25, 27
$W^{1,p}_{\Gamma_D}(\Omega)$	a Sobolev space equipped with Dirichlet data	p. 27
$L^p(I,X)$	Bochner space for $p \in [1,\infty]$ with values in a Banach space X	p. 16
$C(\bar{I},X)$	space of continous functions with values in a Banach space X	p. 16
$C^\alpha(I,X)$	spaces of α-Hölder continuous functions with values in a Banach space X for $\alpha \in (0,1]$	p. 16
$W^{k,p}(I,X)$	Bochner-Sobolev space for $p \in [1,\infty]$, $k \in \mathbb{N}_0$ with values in a Banach space X	p. 17
$W(I)$	a special Hilbert space: $W(I) = L^2(I,V) \cap W^{1,2}(I,V^*)$ with V a Hilbert space	p. 17
$W_0(I)$	a subspace of $W(I)$; $W_0(I) = \{v \in W(I) : v(0) = 0\}$	p. 60
$\mathbb{W}_p$	a special interpolation space for initial conditions	p. 28

Discretization

$K, \mathcal{K}_h$	cell of a mesh; all cells of a mesh	p. 87
h	mesh size	p. 87
$\mathcal{N}_h$	nodes of a mesh	p. 87
V_h	subspace of V of piecewise linear elements	p. 87
$\tilde{V}_h$	the complex valued version of V_h	p. 107
$H^{k,h}(\Omega_h)$, $W^{k,p,h}(\Omega_h)$	spaces of finite element functions v_h such that $\|v\|^p_{W^{k,p,h}(\Omega_h)} := \sum_{K\in\mathcal{K}_h} \|v\|^p_{W^{k,p}(K)}$ is finite; $H^{k,h}(\Omega_h) := W^{k,2,h}(\Omega_h)$	p. 88
X^0_k	space of functions that are piecewise constant in time	p. 111
$X^{0,1}_{k,h}$	space of functions that are piecewise constant in time and piecewise linear in space	p. 111
R_h, P_h, I_h	the Ritz projection onto V_h along the bilinear form a_Ω; the $L^2(\Omega)$-projection onto V_h and the interpolation in V_h	p. 89
A_h	the discretization of the operator A	p. 89
I_k	the nodal interpolant in the time component of a continuous Bochner function	p. 111

Obstacle Problem

Ψ	the obstacle; its regularity varies	p. 27
K_Ψ	set of admissible functions; i.e. functions greator or equal to Ψ almost everywhere	p. 27
$P^{\beta}_{\bar{y}}$	a set of multipliers obtained by special limits used in the establishment of optimality conditions	p. 59

Constants and Embeddings

$C, c > 0$	generic constants whose meaning may change from line to line; usually independent of certain quantities of interest; often equipped with an additional index
$C_{\not f}, C_{\not q}$	examples for constants that do not depend on f or q respectively
$A \subset B$	the standard inclusion; $A = B$ is allowed
$A \subset\subset B$	A, B are topological spaces and A is a compact subset of B
$X \hookrightarrow Y$	the normed vector space X is continuously embedded into the normed vector space Y
$X \overset{C}{\hookrightarrow} Y$	the normed vector space X is continuously embedded into the normed vector space Y with embedding constant C
$X \hookrightarrow\hookrightarrow Y$	the normed vector space X is compactly embedded into the normed vector space Y

Functions and Operators

1_A	indicator function of a set A	
$\mathrm{dom}_p(A)$	the domain of an operator $A\colon \mathrm{dom}_p(A) \subset L^p(\Omega) \to L^p(\Omega)$, for $p \in [1, \infty]$	
$\mathrm{dom}_X(A)$	the domain of an operator $A\colon \mathrm{dom}_X(A) \subset X \to X$	
$(\cdot)'$	this refers to the derivative of a one dimensional function; or in the case of functions $f\colon Q \times \mathbb{R} \to \mathbb{R}, (t, x, y) \mapsto f(t, x, y)$ to the derivative in the third, singled out component	
$\partial_t, \partial_{x_j}$	this refers to the derivative in time or j-th space component; it can refer to the strong, weak or distributional derivative, depending on the context	
∇, ∇^2	the (weak) gradient and hessian in the spatial coordinates	
a_Ω	bilinear form in spatial coordinates	p. 28
$a_I, a_{(0,T)}$	bilinear form integrated over a given time interval	p. 28
$(\cdot, \cdot)_{X,X^*}$	the dual pairing of a Banach space X and its dual; the order in the index does not matter	
$(\cdot, \cdot)_H$	an inner product on a Hilbert space H	

$(\cdot,\cdot)_{L^2(I,X,X^*)}$	a special duality pairing; let X be a Banach space, $f \in L^2(I,X)$ and $g \in L^2(I,X^*)$ we have $(f,g)_{L^2(I,X,X^*)} := \int_I (f(t),g(t))_{X,X^*}\, dt$	p. 16

Bibliography

[ABM14] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2. edition, 2014.

[Ada75] R.A. Adams. *Sobolev Spaces*. Academic Press, 1975.

[AF03] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[Agm65] S. Agmon. *Lectures on Elliptic Boundary Value Problems*. AMS, 1965.

[AL02] D. R. Adams and S. Lenhart. Optimal control of the obstacle for a parabolic variational inequality. *Journal of Mathematical Analysis and Applications*, 268:602–214, 2002.

[Alt99] H.W. Alt. *Lineare Funktionalanalysis*. Springer, 3 edition, 1999.

[Ama95] H. Amann. *Linear and quasilinear parabolic problems. Vol. I*, volume 89 of *Monographs in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1995. Abstract linear theory.

[Ama01] H. Amann. Linear parabolic problems involving measures. *RACSAM. Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat.*, 95(1):85–119, 2001.

[Ama04] H. Amann. Maximal regularity for nonautonomous evolution equations. *Adv. Nonlinear Stud.*, 4(4):417–430, 2004.

[AS94] A. Ashyralyev and P.E. Sobolevskii. *Well-Posedness of Parabolic Difference Equations*. Springer, 1994.

[Bar81] V. Barbu. Necessary conditions for distributed control problems governed by parabolic variational inequalities. *SIAM Journal on Control and Optimization*, 19(1):64.86, 1981.

[Bar84] V. Barbu. *Optimal control of variational inequalities*, volume 100 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.

[BB00] R. Becker and M. Braack. Multigrid techniques for finite elements on locally refined meshes. *Numer. Linear Algebra Appl.*, 7(6):363–379, 2000. Numerical linear algebra methods for computational fluid flow problems.

[BBM01] J. Bourgain, H. Brézis, and P. Mironescu. Another look at Sobolev spaces. In *Optimal control and partial differential equations*, pages 439–455. IOS, Amsterdam, 2001.

[BC11] H. Bauschke and P. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, 2011.

[Bet19] L. M. Betz. Second-order sufficient optimality conditions for optimal control of nonsmooth, semilinear parabolic equations. *SIAM J. Control Optim.*, 57(6):4033–4062, 2019.

[BK15] M. Brokate and G. Kersting. *Measure und Integral*. Compact Textbooks in Mathematics. Birkhäuser/Springer Basel AG, 2015.

[BKK11] J. Brandts, S. Korotov, and M. Křížek. Generalization of the Zlámal condition for simplicial finite elements in $\mathbb{R}^d$. *Appl. Math.*, 56(4):417–424, 2011.

[BL76] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*. Springer-Verlag, Berlin-New York, 1976. Grundlehren der Mathematischen Wissenschaften, No. 223.

[BL04] M. Bergounioux and S. Lenhart. Optimal control of bilateral obstacle problems. *SIAM J. Control Optim.*, 43(1):240–255, 2004.

[BM15] T. Betz and C. Meyer. Second-order sufficient optimality conditions for optimal control of static elastoplasticity with hardening. *ESAIM: COCV*, 21(1):271–300, 2015.

[BN18] L. Bonifacius and I. Neitzel. Second order optimality conditions for optimal control of quasilinear parabolic equations. *Math. Control Relat. Fields*, 8(1):1–34, 2018.

[Bon18] L. Bonifacius. *Numerical analysis of parabolic time-optimal control problems*. PhD thesis, Technische Universität München, 2018.

[BP86] V. Barbu and Th. Precupanu. *Convexity and optimization in Banach spaces*. Mathematics and its Applications (East European Series). D. Reidel Publishing Co., Dordrecht; Editura Academiei Republicii Socialiste România, Bucharest, romanian edition, 1986.

[BP12] V. Barbu and T. Precupanu. *Convexity and optimization in Banach spaces*. Springer Monographs in Mathematics. Springer, Dordrecht, fourth edition, 2012.

[Bré72] H. Brézis. Problèmes unilatéraux. *J. Math. Pures Appl. (9)*, 51:1–168, 1972.

[Bré11] H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[BS08] S.C. Brenner and L.R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[BT11] M. Boukrouche and D. A. Tarzia. Existence, uniqueness, and convergence of optimal control problems associated with parabolic variational inequalities of the second kind. *Nonlinear Anal. Real World Appl.*, 12(4):2211–2224, 2011.

[BTW03] N. Y. Bakaev, V. Thomée, and L. B. Wahlbin. Maximum-norm estimates for resolvents of elliptic finite element operators. *Math. Comp.*, 72(244):1597–1610, 2003.

[BZ99] M. Bergounioux and H. Zidani. Pontryagin maximum principle for optimal control of variational inequalities. *SIAM J. Control Optim.*, 37(4):1273–1290, 1999.

[Car99] C. Carstensen. Quasi-interpolation and a posteriori error analysis in finite element methods. *M2AN Math. Model. Numer. Anal.*, 33(6):1187–1202, 1999.

[CDD+02] N. Calvo, J. I. Díaz, J. Durany, E. Schiavi, and C. Vázquez. On a doubly nonlinear parabolic obstacle problem modelling ice sheet dynamics. *SIAM J. Appl. Math.*, 63(2):683–707, 2002.

[CDV10] N. Calvo, J. Durany, and C. Vázquez. A new fully nonisothermal coupled model for the simulation of ice sheet flow. *Phys. D*, 239(5):248–257, 2010.

[CH02] K. Chrysafinos and L. S. Hou. Error estimates for semidiscrete finite element approximations of linear and semilinear parabolic equations under minimal regularity assumptions. *SIAM J. Numer. Anal.*, 40(1):282–306, 2002.

[CH15] C. Carstensen and J. Hu. An optimal adaptive finite element method for an obstacle problem. *Comput. Methods Appl. Math.*, 15(3):259–277, 2015.

[Che03] Q. Chen. Optimal control for semilinear evolutionary variational bilateral problem. *Journal of Mathematical Analysis And Applications*, 277:303–323, 2003.

[Chr18] C. Christof. *Sensitivity Analysis of Elliptic Variational Inequalities of the First and the Second Kind.* Dissertation, Technische Universität Dortmund, 2018.

[Chr19] C. Christof. Sensitivity analysis and optimal control of obstacle-type evolution variational inequalities. *SIAM J. Control Optim.*, 57(1):192–218, 2019.

[Ş03] E. S. Şuhubi. *Functional analysis.* Kluwer Academic Publishers, Dordrecht, 2003.

[CV19] C. Christof and B. Vexler. New regularity results and finite element error estimates for a class of parabolic optimal control problems with pointwise state constraints. *in preparation*, 2019, last version in private communication from 08.03.19.

[CW19] C. Christof and G. Wachsmuth. On second-order optimality conditions for optimal control problems governed by the obstacle problem. *Preprint*, 2019, last access 28.11.19.

[DDW75] J. Douglas, Jr., T. Dupont, and L. Wahlbin. The stability in L^q of the L^2-projection into finite element function spaces. *Numer. Math.*, 23:193–197, 1974/75.

[dlRMV08] J.C. de los Reyes, C. Meyer, and B. Vexler. Finite element error analysis for state-constrained optimal control of the Stokes equations. *Control Cybernet.*, 37(2):251–284, 2008.

[DNPV12] E. Di Nezza, G. Palatucci, and E. Valdinoci. Hitchhiker's guide to the fractional Sobolev spaces. *Bull. Sci. Math.*, 136(5):521–573, 2012.

[Dok73] P. Doktor. On the density of smooth functions in certain subspaces of Sobolev space. *Comment. Math. Univ. Carolinae*, 14:609–622, 1973.

[Dom82] A. Domarkas. Regularity of solutions of unilateral problems for quasilinear parabolic problems. *Lithuanian Mathematical Journal*, 22:275–284, 1982.

[DtER15] K. Disser, A. F. M. ter Elst, and J. Rehberg. Hölder estimates for parabolic operators on domains with rough boundary. *Ann. Sc. Norm. Sup. Pisa*, 2015.

[DŽ06] P. Doktor and A. Ženíšek. The density of infinitely differentiable functions in Sobolev spaces with mixed boundary conditions. *Appl. Math.*, 51(5):517–547, 2006.

[EG04] A. Ern and J. Guermond. *Theory and practice of finite elements*, volume 159 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[EJL98] K. Eriksson, C. Johnson, and S. Larsson. Adaptive finite element methods for parabolic problems vi: Analytic semigroups. *SIAM J. Numer. Anal.*, 35:1315–1325, 1998.

[Emm04] E. Emmrich. *Gewöhnliche und Operator-Differentialgleichungen*. Vieweg, 2004.

[Eva98] L. Evans. *Partial Differential Equations*. American Mathematical Society, second edition, 1998.

[Fet87] A. Fetter. L_∞-error estimate for an approximation of a parabolic variational inequality. *Numerische Mathematik*, 50:557–565, 1987.

[FK75] A. Friedman and D. Kinderlehrer. A one phase stefan problem. *Indiana Univ. Math. J.*, 24:1005–1035, 1975.

[FR11] H. Fu and H. Rui. Finite element approximation of semilinear parabolic optimal control problems. *Numer. Math. Theory Methods Appl.*, 4(4):489–504, 2011.

[Fri87] A. Friedman. Optimal control for parabolic variational inequalities. *SIAM J. Control Optim.*, 25(2):482–497, 1987.

[Fro93] S. J. Fromm. Potential space estimates for Green potentials in convex domains. *Proc. Amer. Math. Soc.*, 119(1):225–233, 1993.

[GHHL14] A. Gaevskaya, M. Hintermüller, R. H. W. Hoppe, and C. Löbhard. Adaptive finite elements for optimally controlled elliptic variational inequalities of obstacle type. In *Optimization with PDE constraints*, volume 101 of *Lect. Notes Comput. Sci. Eng.*, pages 95–150. Springer, Cham, 2014.

[Gia83] M. Giaquinta. *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, volume 105 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1983.

[GM09] M. Giaquinta and G. Modica. *Mathematical Analysis An Introduction to Functions of Several Variables*. Birkhäuser, 2009.

[GM19] T. Gudi and P. Majumder. Convergence analysis of finite element method for a parabolic obstacle problem. *J. Comput. Appl. Math.*, 357:85–102, 2019.

[GR89] K. Gröger and J. Rehberg. Resolvent estimates in $W^{-1,p}$ for second order elliptic differential operators in case of mixed boundary conditions. *Math. Ann.*, 285(1):105–113, 1989.

[Gri11] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Society for Industrial and Applied Mathematics, reissue edition, 2011.

[Grö89] K. Gröger. A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Math. Ann.*, 283(4):679–687, 1989.

[GT77] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin-New York, 1977. Grundlehren der Mathematischen Wissenschaften, Vol. 224.

[GT01] D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*, volume Reprint of the 1998 Edition. Springer, 2001.

[GW82] M. Grüter and K-O. Widman. The Green function for uniformly elliptic equations. *Manuscripta Math.*, 37(3):303–342, 1982.

[HDR09] R. Haller-Dintelmann and J. Rehberg. Maximal parabolic regularity for divergence operators including mixed boundary conditions. *J. Differential Equations*, 247(5):1354–1396, 2009.

[HKP19] D. Hafemeyer, C. Kahle, and J. Pfefferer. Finite element error estimates in L^2 for regularized discrete approximations to the obstacle problem. *Numerische Mathematik*, 2019.

[HKST15] J. Heinonen, P. Koskela, N. Shanmugalingam, and J. T. Tyson. *Sobolev spaces on metric measure spaces*, volume 27 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2015. An approach based on upper gradients.

[HW18] F. Harder and Gerd Wachsmuth. Comparison of optimality systems for the optimal control of the obstacle problem. *GAMM-Mitteilungen*, 40(4):312–338, 2018.

[IK90] K. Ito and K. Kunisch. An augmented Lagrangian technique for variational inequalities. *Appl. Math. Optim.*, 21(3):223–241, 1990.

[IK00] K. Ito and K. Kunisch. Optimal control of elliptic variational inequalities. *Appl. Math. Optim.*, 41(3):343–364, 2000.

[IK03] K. Ito and K. Kunisch. Semi-smooth Newton methods for variational inequalities of the first kind. *M2AN Math. Model. Numer. Anal.*, 37(1):41–62, 2003.

[IK06] K. Ito and K. Kunisch. Parabolic variational inequalities: the Lagrange multiplier approach. *J. Math. Pures Appl. (9)*, 85(3):415–449, 2006.

[IK10] K. Ito and K. Kunisch. Optimal control of parabolic variational inequalities. *J. Math. Pures Appl.*, 93:329–360, 2010.

[JB12] G. Jouvet and E. Bueler. Steady, shallow ice sheets as obstacle problems: well-posedness and finite element approximation. *SIAM J. Appl. Math.*, 72(4):1292–1314, 2012.

[JJ13] E. Y. Ju and J.-M. Jeong. Optimal control problems for nonlinear variational evolution inequalities. *Abstr. Appl. Anal.*, pages Art. ID 724190, 10, 2013.

[JKRS03] J. Jarušek, M. Krbec, M. Rao, and J. Sokołowski. Conical differentiability for evolution variational inequalities. *J. Differential Equations*, 193(1):131–146, 2003.

[JW84] A. Jonsson and H. Wallin. *Function spaces on subsets of* $\mathbb{R}^n$. Chur u.a., Harwood, 1984.

[Kat08] P. I. Kattan. *MATLAB Guide to Finite Elements: An Interactive Approach.* Springer Berlin Heidelberg, 2008.

[Kec95] A. S. Kechris. *Classical descriptive set theory*, volume 156 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[KKT03] T. Kärkkäinen, K. Kunisch, and P. Tarvainen. Augmented Lagrangian active set methods for obstacle problems. *J. Optim. Theory Appl.*, 119(3):499–533, 2003.

[Kra67] J. P. Krasovskiĭ. Isolation of singularities of the green's function. *Mathematics of the USSR-Izvestiya*, 1(5):935–966, 1967.

[Kru14] F. Kruse. *Interior point methods for optimal control problems with pointwise state constraints*. Dissertation, Technische Universität München, München, 2014. The author is now known as Florian Mannel.

[KS80] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications*. Academic Press, Inc., 1980.

[KW12a] K. Kunisch and D. Wachsmuth. Path-following for optimal control of stationary variational inequalities. *Comput. Optim. Appl.*, 51(3):1345–1373, 2012.

[KW12b] K. Kunisch and D. Wachsmuth. Sufficient optimality conditions and semi-smooth Newton methods for optimal control of stationary variational inequalities. *ESAIM Control Optim. Calc. Var.*, 18(2):520–547, 2012.

[LMWZ10] J. Li, J.M. Melenk, B. Wohlmuth, and J. Zou. Optimal a priori estimates for higher order finite elements for elliptic interface problems. *Appl. Numer. Math.*, 60(1-2):19–37, 2010.

[Lun95] A. Lunardi. *Analytic semigroups and optimal regularity in parabolic problems*. Modern Birkhäuser Classics. Birkhäuser/Springer Basel AG, Basel, 1995. [2013 reprint of the 1995 original] [MR1329547].

[Lun09] A. Lunardi. *Interpolation theory*. Appunti. Scuola Normale Superiore di Pisa (Nuova Serie). [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)]. Edizioni della Normale, Pisa, second edition, 2009.

[LV16] D. Leykekhman and B. Vexler. Pointwise best approximation results for galerkin finite element solutions of parabolic problems. *SIAM Journal on Control and Optimization*, 2016.

[LV17a] D. Leykekhman and B. Vexler. Discrete maximal parabolic regularity for galerkin finite element methods. *Numer. Math.*, 135(3):923–952, 2017.

[LV17b] D. Leykekhman and B. Vexler. Optimal a priori error estimates of parabolic optimal control problems with a moving point control. *Submitted*, 2017.

[Maz85] V. G. Maz'ja. *Sobolev spaces*. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1985. Translated from the Russian by T. O. Shaposhnikova.

[Mig76] F. Mignot. Contrôle dans les inéquations variationelles elliptiques. *J. Functional Analysis*, 22(2):130–185, 1976.

[Mor66] C. B. Morrey, Jr. *Multiple integrals in the calculus of variations*. Die Grundlehren der mathematischen Wissenschaften, Band 130. Springer-Verlag New York, Inc., New York, 1966.

[MP84] F. Mignot and J.P. Puel. Optimal control in some variational inequalities. *SIAM Journal on Control and Optimization*, 22:466–476, 1984.

[MR16] H. Meinlschmidt and J. Rehberg. Hölder-estimates for non-autonomous parabolic problems with rough data. *Evolution Equations and Control Theory*, 5(1):147–184, 2016.

[MRV11] 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. *SIAM J. Control Optim.*, 49(5):1961–1997, 2011.

[MRW15] C. Meyer, A. Rademacher, and W. Wollner. Adaptive optimal control of the obstacle problem. *SIAM J. Sci. Comput.*, 37(2):A918–A945, 2015.

[MS17] C. Meyer and L. M. Susu. Optimal control of nonsmooth, semilinear parabolic equations. *SIAM J. Control Optim.*, 55(4):2206–2234, 2017.

[MT13] C. Meyer and O. Thoma. A priori finite element error analysis for optimal control of the obstacle problem. *SIAM J. Numer. Anal.*, 51(1):605–628, 2013.

[Mun75] J. R. Munkres. *Topology: a first course.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

[MV17] D. Meidner and B. Vexler. Optimal error estimates for fully discrete galerkin approximations of semilinear parabolic equations. *Submitted*, 2017.

[Nit79] J. Nitsche. L_∞-convergence of finite element Galerkin approximations for parabolic problems. *RAIRO Anal. Numér.*, 13(1):31–54, i, 1979.

[Noc88] R. Nochetto. Sharp L^∞-error estimates for semilinear elliptic problems with free boundaries. *Numer. Math.*, 54(3):243–255, 1988.

[NV12] I. Neitzel and B. Vexler. A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems. *Numer. Math.*, 120(2):345–386, 2012.

[Øks98] B. Øksendal. *Stochastic differential equations.* Universitext. Springer-Verlag, Berlin, fifth edition, 1998. An introduction with applications.

[OS16] E. Otárola and A. J. Salgado. Finite element approximation of the parabolic fractional obstacle problem. *SIAM J. Numer. Anal.*, 54(4):2619–2639, 2016.

[Paz83] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences.* Springer-Verlag, New York, 1983.

[Ran17] R. Rannacher. *Numerik 2: Numerik partieller Differentialgleichungen.* Heidelberg University Publishing, 2017.

[Rod87] J-F. Rodrigues. *Obstacle problems in mathematical physics*, volume 134 of *North-Holland Mathematics Studies.* North-Holland Publishing Co., Amsterdam, 1987. Notas de Matemática [Mathematical Notes], 114.

[Rou13] T. Roubíček. *Nonlinear Partial Differential Equations with Applications.* Birkhäuser, 2013.

[Rud74] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, second edition, 1974.

[Sey02] R. Seydel. *Tools for computational finance.* Universitext. Springer-Verlag, Berlin, 2002.

[Sim85] J. Simon. Compact sets in the space $L^p(0,T;B)$. *Annali di Matematica Pura ed Applica*, 1985.

[Sol71] V. A. Solonnikov. The Green's matrices for elliptic boundary value problems. II. *Trudy Mat. Inst. Steklov.*, 116:181–216, 237, 1971. Boundary value problems of mathematical physics, 7.

[Ste70] E. M. Stein. *Singular integrals and differentiability properties of functions.* Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970.

[Ste74] H.B. Stewart. Generation of analytic semigroups by strongly elliptic operators. *Transactions of the American Mathematical Society*, 199:141–161, 1974.

[STW80] A. H. Schatz, V. C. Thomée, and L. B. Wahlbin. Maximum norm stability and error estimates in parabolic finite element equations. *Comm. Pure Appl. Math.*, 33(3):265–304, 1980.

[STW98] A. H. Schatz, V. C. Thomée, and L. B. Wahlbin. Stability, analyticity, and almost best approximation in maximum norm for parabolic finite element equations. *Comm. Pure Appl. Math.*, 51(11-12):1349–1385, 1998.

[SW82] A. H. Schatz and L. B. Wahlbin. On the quasi-optimality in L_∞ of the $\dot{H}^1$-projection into finite element spaces. *Mathematics of Computation*, 38(157), 1982.

[SZ90] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[TC13] Y. Tang and Y. Chen. Superconvergence analysis of fully discrete finite element methods for semilinear parabolic optimal control problems. *Front. Math. China*, 8(2):443–464, 2013.

[Tri78] H. Triebel. *Interpolation theory, function spaces, differential operators*, volume 18 of *North-Holland Mathematical Library.* North-Holland Publishing Co., Amsterdam-New York, 1978.

[Trö09] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen.* Vieweg + Teubner Verlag, 2. auflage edition, 2009.

[Ulb11] M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, volume 11 of *MOS-SIAM Series on Optimization.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2011.

[USL88] N.N. Ural'ceva, V.A. Solonnikov, and O.A. Ladyženskaja. *Linear and Quasilinear Equations of Parabolic Type*, volume 23. AMS, 1988.

[UU12] M. Ulbrich and S. Ulbrich. *Nichtlineare Optimierung.* Birkhäuser, 2012.

[Väh12] A. V. Vähäkangas. On regularity and extension of Green's operator on bounded smooth domains. *Potential Anal.*, 37(1):57–77, 2012.

[Wac14] G. Wachsmuth. Strong stationarity for optimal control of the obstacle problem with control constraints. *SIAM J. Optim.*, 24(4):1914–1932, 2014.

[Wac16a] D. Wachsmuth. The regularity of the positive part of functions in $L^2(I;H^1(\Omega))\cap H^1(I;H^1(\Omega)^*)$ with applications to parabolic equations. *Comment. Math. Univ. Carolin.*, 57(3):327–332, 2016.

[Wac16b] G. Wachsmuth. Towards M-stationarity for optimal control of the obstacle problem with control constraints. *SIAM J. Control Optim.*, 54(2):964–986, 2016.

[Wer11] D. Werner. *Funktionalanalysis.* Springer, 7. edition, 2011.

[Wlo92] J. Wloka. *Partial differential equations.* Cambridge University Press, Cambridge, 1992. Translated from the German by C. B. Thomas and M. J. Thomas.

[Woo07] I. Wood. Maximal L^p-regularity for the Laplacian on Lipschitz domains. *Math. Z.*, 255(4):855–875, 2007.

[YWG14] X. Yang, G. Wang, and X. Gu. Numerical solution for a parabolic obstacle problem with nonsmooth initial data. *Numer. Methods Partial Differential Equations*, 30(5):1740–1754, 2014.