Technische Universität München
Fakultät für Mathematik
Lehrstuhl für Optimalsteuerung

Numerical analysis of parabolic time-optimal control problems

Lucas Simon Bonifacius

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. Karl Kunisch
	3. Prof. Dr. Daniel Wachsmuth

Die Dissertation wurde am 07.02.2018 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 19.06.2018 angenommen.

Abstract

This thesis is concerned with the analysis of finite element discretizations for time-optimal control problems subject to linear parabolic partial differential equations and constraints for the state evaluated at the free end time. Necessary and sufficient optimality conditions are provided for the regular case and the case of bang-bang controls. *A priori* discretization error estimates are proved for different control discretization strategies. Efficient algorithms for the numerical solution are discussed.

Zusammenfassung

Diese Arbeit befasst sich mit der Analyse finiter Elemente Diskretisierungen zeitoptimaler Steuerungsprobleme mit linearer, parabolischer, partieller Differentialgleichung und Zustandsrestriktionen am freien Endzeitpunkt. Notwendige und hinreichende Optimalitätsbedingungen werden für den regulären und den bang-bang Fall untersucht. *A priori* Fehlerabschätzungen für verschiedene Diskretisierungen der Kontrolle werden bewiesen. Effiziente Algorithmen für die numerische Lösung werden diskutiert.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Boris Vexler for his advice and guidance throughout the last three years. I am also grateful for his general support, the possibility to attend international conferences, and the freedom to work on additional research projects. Moreover, I would like to thank my second supervisor Prof. Dr. Karl Kunisch for providing new impulses and sharing his expertise as well as critically discussing preliminary results during my stays in Graz.

Furthermore, I would like to express my gratitude to my mentor Prof. Dr. Ira Neitzel and to Dr. Konstantin Pieper for the fruitful collaboration and their scientific advice, especially in the beginning of my PhD studies. My sincere thanks also go to my colleges from the mathematical institutes in Munich and Graz for the friendly and supporting atmosphere.

Moreover, I gratefully acknowledge support from the International Research Training Group IGDK *Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structures*, funded by the German Science Foundation (DFG) and the Austrian Science Fund (FWF). I am especially grateful to the IGDK and its members for the frequent exchange of ideas, the possibility of collaboration, and the opportunity to organize and to take part in workshops.

Finally, I am grateful to my family, friends, and girlfriend for their support in every kind of way.

Contents

1. Introduction

In many applications a certain criterion has to be met after some time, which should be chosen as short as possible; cf., e.g., [2, 132]. For example, the objective could be to steer a system close to a desired state in the fastest time possible by applying a control to the system. This class of optimization problems is therefore called *time-optimal control*. Since it is in general difficult to give an explicit solution formula, appropriate approximations of these problems are necessary to compute solutions numerically. In this thesis we analyze finite element discretizations for a class of time-optimal control problems subject to linear parabolic partial differential equations.

To set the stage, for T denoting the terminal time, u the state, and q the control, let us consider the abstract model problem:

$$\begin{aligned} &\text{Minimize} \quad j(T,q) \coloneqq T + \int_0^T L(q(t))\,\mathrm{d}t, \\ &\text{subject to} \quad \begin{cases} T > 0, \quad q \in Q_{ad}(0,T), \\ u = u(q,T), \\ \|u(T) - u_d\|_H \leq \delta_0, \end{cases} \end{aligned} \tag{P_{model}}$$

where $Q_{ad}(0,T)$ is the set of admissible controls, H is an appropriate Hilbert space, and $u(q,T)$ denotes the solution to the time-dependent partial differential equation for the time horizon $T > 0$ and the control q. In some applications it is necessary to account for control costs or smoothing terms in the objective functional, cf., e.g., [92, 93, 125]. This motivates the additional functional L in the problem formulation. Different choices for L and its implications on the solutions will be detailed below. Many results of this thesis are valid for more general terminal constraints than the one considered in (P_{model}) and we will introduce the precise assumptions in Chapter 2.

The task is to steer the system from a given initial state close to a desired state $u_d \in H$ by an appropriate choice of the control $q\colon [0,T] \to Q_{ad}$ and the time horizon T, while minimizing T plus the running cost L for the control. It is worth mentioning that both the control q and the terminal time T are optimization variables. In particular, this means that the time horizon of the state equation is not fixed. For this reason, (P_{model}) is a nonlinear optimization problem subject to control as well as state constraints which significantly complicates the analysis and numerical realization of (P_{model}) compared to a linear parabolic optimal control problem with a fixed T; see, e.g, [116, 117, 118].

The choice of the functional L will play a central role in this work and we will consider three different situations:

- **Time-optimal control problem:** If $L \equiv 0$, then we obtain the pure time-optimal problem, where we are plainly interested in steering u_0 into the H-ball centered at the desired state u_d in the fastest time possible. This is a classical choice in control theory; see, e.g., [54, 97, 122, 149, 160] and the overview given in [104, Chapter 7]. Typically

the solutions for $L \equiv 0$ are bang-bang. We call a control *bang-bang*, if the set where it does not equal the control bounds is a set of zero measure.

- **Regularized time-optimal control problem:** A choice different from zero allows for a regularization strategy of the time-optimal control problem; cf., e.g., [82, 95]. For example, we can consider
$$L(q) = \frac{\alpha}{2}\|q\|_Q^2 \quad \text{for } \alpha > 0, \tag{1.1}$$
where Q is another Hilbert space for the controls. We will frequently choose $Q = L^2(\omega)$ with ω the control domain. In this case we are interested in letting the regularization parameter α tend to zero. Note that for this particular choice of L, the optimal control $\bar{q}$ inherits regularity properties of the adjoint state by means of a projection formula that links optimal control and adjoint state. Hence, the choice of the norm for the regularization is not arbitrary, because it qualitatively changes the solutions.

- **Control costs/Smoothing:** Moreover, L can represent inherent control costs or simply when bang-bang controls are not desirable; cf., e.g., [92, 93, 125]. In the latter, L can be chosen in a way such that it has a smoothing effect, e.g. the L^2-norm of the control. Moreover, other objective functionals can be more appropriate to model the control costs in concrete applications such as the L^1-norm of the control or a linear functional in the control variable; cf., e.g., [42, 153].

In a strict definition, if $L \not\equiv 0$ the optimization problem is not "time-optimal", because we are not minimizing just T. Nevertheless, it seems that in practice, one is often not interested in steering the system into the terminal set as fast as possible at any costs. This motivates to consider the more general problem formulation introduced above with free end time and control costs in the objective. Hence, by the term time-optimal control, we always refer to the broader definition.

To deal with the variable time horizon, it is convenient to transform the linear parabolic partial differential equation onto a fixed reference time interval, which is accomplished by introducing a transformation variable $\nu\colon [0,1] \to \mathbb{R}_+$ with the relation $T = \int_0^1 \nu(\tau)\,\mathrm{d}\tau$. Considering the transformation ν as an additional control variable allows to define a control-to-state mapping $(\nu, q) \mapsto u = S(\nu, q)$ that is (infinitely often) continuously differentiable, where u denotes the solution to the transformed partial differential equation. Moreover, as all variables are defined on the same reference time horizon, different solutions (arising for example from the regularization strategy) can be directly compared with each other. For these reasons, in this thesis we will mainly work with the transformed state equation. In particular, it is the basis for the numerical method.

By means of the control-to-state mapping, we can define the reduced and transformed optimal control problem as

$$\inf_{\nu>0,\, q\in Q_{ad}(0,1)} j(\nu, q) \quad \text{subject to} \quad \|u(1) - u_d\|_H \le \delta_0,\ u = S(\nu, q). \tag{$\hat{P}_{\text{model}}$}$$

To calculate solutions to ($\hat{P}_{\text{model}}$) in practice, we have to introduce a discretized version of the state equation and replace the control space and the state space by finite dimensional spaces. We obtain the discrete version of ($\hat{P}_{\text{model}}$) as

$$\inf_{\nu>0,\, q\in Q_{ad,\sigma}(0,1)} j(\nu, q) \quad \text{subject to} \quad \|u_{kh}(1) - u_d\|_H \le \delta_0,\ u_{kh} = S_{kh}(\nu, q), \tag{$\hat{P}_{\text{model},kh}$}$$

where k and h denote the temporal and spatial discretization parameters and σ is an additional discretization parameter for the controls. A considerable part of this work is devoted

to the numerical analysis of the discretized optimal control problem ($\hat{P}_{\text{model},kh}$). On the one hand, we show a priori discretization error estimates if L is chosen as the L^2-norm of the control similar as in (1.1) for a fixed cost parameter $\alpha > 0$. Moreover, we are concerned with the case of variable α, where we investigate the behavior of the discrete solutions if α tends to zero.

This thesis is structured as follows. In Chapter 2 we discuss first order optimality conditions for the time-optimal control problem (P_{model}). Since (P_{model}) is subject to state constraints, a constraint qualification is needed to guarantee optimality conditions in qualified form. Our approach relies on the concept of weak invariance. We first generalize a characterization of weak invariance in terms of the so-called lower Hamiltonian condition which is interesting for itself. This characterization is known for optimal control of ordinary differential equations and uncontrolled partial differential equations. We show that strengthening of the lower Hamiltonian condition leads to a sufficient criterion for qualified optimality conditions. In contrast to typical constraint qualifications our condition can be checked a priori without having to know the optimal solution. Concrete examples are discussed in Section 2.4. These results have already appeared in similar form in [18].

Chapter 3 is devoted to sufficient optimality conditions. In the non-bang-bang case, we formulate second order necessary and sufficient optimality conditions employing a cone of critical directions that leads to a minimal gap between necessity and sufficiency. Additionally, the second order sufficient optimality condition is equivalent to a scalar condition that requires the solution of one linear-quadratic optimization problem. In Section 5.4, we will verify the second order sufficient optimality condition on the discrete level by numerically computing the scalar quantity. Most of these results are already contained in [17].

In the bang-bang case, it turns out that the second order sufficient optimality condition is vacuously true and it is therefore unlikely that this guarantees local optimality. For this reason, we consider a well-established structural assumption on the adjoint state that provides a sufficient optimality condition in the bang-bang case. Chapters 2 and 3 form the basis for the a priori discretization error estimates in Chapter 5.

In Chapter 4 we discuss the theoretical and practical aspects concerning the numerical solution of (P_{model}). In case of $\alpha > 0$, we consider the augmented Lagrangian method and briefly discuss its convergence properties. For the solution of the resulting subproblems, we consider a bilevel approach and a monolithic approach. For the case $\alpha = 0$, we can solve the regularized problem for a sequence of regularization parameters with $\alpha \to 0$. Additionally, we discuss an alternative approach that relies on an equivalent reformulation of the time-optimal control problem. This reformulated problem has again a bilevel structure, where we have to find a root of a certain value function in the outer loop and need to solve convex and control constrained problems in the inner loop. We consider different methods for the solution of the optimization problems occurring on each level.

In Chapter 5 we discuss the discretization of the state equation of (P_{model}) and the corresponding adjoint state equation by means of the discontinuous Galerkin method in time and the continuous Galerkin method in space. Concerning the control variable we consider different control discretization strategies. Depending on the concrete discretization, we investigate the convergence of the solutions of the corresponding discrete variants of (P_{model}) to solutions of the original problem. In Section 5.3 we prove a priori discretization error estimates for the terminal time and the control variable in L^2 in the case of non-bang-bang controls under the hypothesis that second order sufficient optimality conditions hold. We verify our theoretical findings by numerical examples and observe that the estimates are optimal with respect to

the control variable. These results are already contained [17] and have been submitted to a scientific journal recently.

Moreover, in Section 5.5 we prove a priori discretization and regularization error estimates for the terminal time and the control variable in L^1 in the bang-bang case. It is based on the structural assumption on the adjoint state. For purely time-dependent control, these estimates directly follow from standard error estimates of the state equation. However, for distributed control we require pointwise discretization error estimates for the state equation and thus error estimates for the optimal control problem are associated with further technical effort. We provide error estimates in case of distributed control for the particular situation that the control domain has a strict distance to the boundary of the spatial domain. Numerical examples indicate that the structural assumption for purely time-dependent control is satisfied which leads to optimal error estimates.

Last, in Section 5.6 we present a different approach that leads to discretization error estimates for the terminal time without a sufficient optimality condition. It relies on the construction of feasible controls and cross-wise testing. For the construction, we use a discrete version of the strengthened Hamiltonian condition from Chapter 2. It is worth noting that the strengthened Hamiltonian condition can be checked a priori in many examples.

Further auxiliary results that are needed in the main chapters are collected in the appendix. Many of them are well-known. However, in particular for the discretization error estimates, the precise asymptotic behavior of the constants is required. Hence, we either provide references, where the constants are explicitly stated, or give independent proofs.

2. First order optimality conditions

This chapter is devoted to first order optimality conditions for a class of time-optimal control problems governed by a linear parabolic equation. It is essentially based on the paper [18] with Konstantin Pieper. For T denoting the terminal time, u the state, and q the control, we introduce the problem:

$$\begin{aligned} &\text{Minimize} \quad j(T,q) \coloneqq T + \int_0^T L(q(t))\,\mathrm{d}t, \\ &\text{subject to} \quad \begin{cases} T > 0, \\ \partial_t u(t) + Au(t) = Bq(t), \quad t \in (0,T), \\ u(0) = u_0, \\ u(T) \in U, \\ q(t) \in Q_{ad}, \quad t \in (0,T). \end{cases} \end{aligned} \tag{P}$$

Here, $A\colon V \to V^*$ is a linear, weakly coercive operator for a Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$, $Q_{ad} \subset Q$ for a Hilbert space Q a closed and bounded set of admissible controls, and B a linear and bounded control operator mapping Q into a subspace of V^*. The precise assumptions will be introduced in Section 2.1. A concrete example of a convection-diffusion equation satisfying the abstract assumptions will be discussed in Section 3.1.2.

The purpose of this chapter is the derivation of first order optimality conditions for (P) that can be stated as follows: For any optimal solution $(T, \bar{q}, \bar{u})$, there exists a nontrivial $\bar{\mu} \in N_U(\bar{u}(T))$ the normal cone to U at $\bar{u}(T)$, a corresponding adjoint state $\bar{z}$ with

$$-\partial_t \bar{z}(t) + A^* \bar{z}(t) = 0, \quad t \in (0,T), \quad \bar{z}(T) = \bar{\mu}, \tag{2.1}$$

and a $\bar{\mu}_0 \in \{0, 1\}$, such that

$$0 = \langle B\bar{q}(t) - A\bar{u}(t), \bar{z}(t)\rangle + \bar{\mu}_0[1 + L(\bar{q}(t))], \qquad t \in (0,T), \tag{2.2}$$

$$\bar{q}(t) = \operatorname*{argmin}_{q \in Q_{ad}} \left[\langle Bq, \bar{z}(t)\rangle + \bar{\mu}_0 L(q)\right], \qquad t \in (0,T). \tag{2.3}$$

This general form is fulfilled in any optimum of (P) if, e.g., the target set U is of finite co-dimension in H. We give an independent proof of the general form of the optimality conditions for (P) in Theorem 2.26; cf. [40, 104, 134]. In the case that $\bar{\mu}_0 = 1$, the optimality conditions are called *qualified*.

In order to verify that qualified optimality conditions hold, we rely on the concept of strong stability. Strong stability (also known as calmness [25, 136] or weak calmness [21]) quantifies the dependency of the optimal value function of (P) on small perturbations of the constraint. Roughly speaking strong stability means that the optimal value function of (P) (i.e. the minimal value of $j(\cdot,\cdot)$) depends Lipschitz continuously on perturbations of the target set U of the form $U_\delta = U + \mathcal{B}_\delta(0)$ with $\delta \geq 0$, where $\mathcal{B}_\delta(0)$ denotes ball in H of radius δ centered at zero. We refer to Section 2.3.1 for a precise definition.

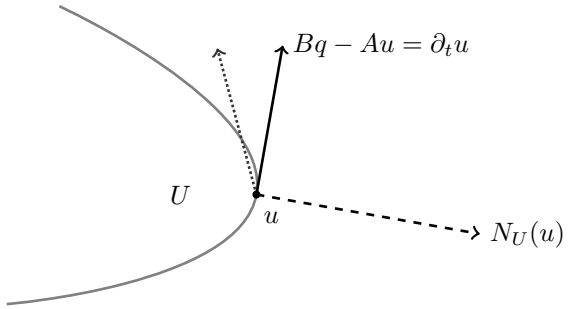

Figure 2.1.: Geometric interpretation of the lower Hamiltonian condition (2.4) with strengthened condition (2.5) (dotted).

Assuming strong stability, the qualified form holds; see Theorem 2.25; cf. also [134, Remark 2.2]. More specifically, strong stability implies the existence of an exact penalty function, which in turn allows to derive qualified optimality conditions, where we use the approach due to Clarke [39]. We emphasize that Theorem 2.25 does not require any structural assumptions on U, such as finite co-dimension; see, e.g., [104, Definition 2.1.32]. Moreover, the multiplier $\bar{\mu}$ satisfies an a priori estimate which is of independent interest. Although it is generally well-known that "almost all" problems are strongly stable, it remains a difficult task to verify strong stability of a particular problem; cf. [19, Section 3]. The main objective of this chapter is to derive conditions on the triple (A, U, BQ_{ad}) which guarantee that (P) is strongly stable for all optimal solutions.

Mathematically, our approach relies on weak invariance of the terminal set. The set U is called *weakly invariant* under (A, BQ_{ad}) if for any initial state $u_0 \in U$ there is a control such that the corresponding trajectory with initial value u_0 remains in U. The precise meaning of weak invariance used in this work is given in Definition 2.1. One of the main contributions of this chapter is the characterization of weak invariance by the conditions that the minimizing projection onto U in H denoted P_U is stable in V, i.e. $P_U(V) \subseteq V$, and

$$h(u, \zeta) \coloneqq \min_{q \in Q_{ad}} \langle Bq - Au, \zeta \rangle \leq 0 \quad \text{for all } u \in U \cap V,\ \zeta \in N_U(u) \cap V, \tag{2.4}$$

where $h \colon V \times V \to \mathbb{R}$ is the *lower Hamiltonian*; see Theorem 2.9. This extends known results for invariance under semigroups, i.e. uncontrolled systems (see, e.g., [127, Section 2.1]), and results for optimal control of ordinary differential equations (see, e.g., [40, Section 12.1]).

Precisely, our main result can now be stated as follows: Assume that the projection P_U is stable in V and that the *strengthened Hamiltonian condition*,

$$h(u, \zeta) \leq -h_0 \|\zeta\|_H \quad \text{for all } u \in U \cap V,\ \zeta \in N_U(u) \cap V, \tag{2.5}$$

holds for some $h_0 > 0$ (independent of u and ζ). Then, strong stability is satisfied for *all solutions* of the time-optimal problem; see Theorem 2.18. As already mentioned, strong stability guarantees that qualified optimality conditions hold. On top of this, condition (2.5) enables to derive Lipschitz continuity results of the value function for a variety of perturbations of the problem (P), not only in the target set. Note, that this corresponds to an estimate for the optimal time for the pure time-optimal problem, which is of independent interest.

To the best of our knowledge, several of the applications of the sufficient conditions derived in this work yield new results for concrete problems. In particular, these conditions allow to derive qualified optimality conditions for several interesting scenarios, such as the control of

the heat equation into L^2-balls around certain target sets. We will discuss these applications in Section 2.4. In the case of steering the system into a single point, i.e. $U = \{ u_d \}$, we can compare the results to those of Barbu [10, Section 5.3], who derived the maximum principle for a nonlinear monotone equation using a quadratic penalty method; cf. also [11] for the Navier-Stokes equation or [95] for the linear wave equation. Note that the qualifying condition on the target state in [10, Theorem 5.3.1] is essentially the same as the one obtained from (2.5) in the case $U = \{ u_d \}$; see Section 2.4.1. However, this condition holds in concrete applications only for controls which are acting everywhere in space. A different approach, which is based on controllability, has been proposed by Wang and Zuazua [160]. Here, the equivalence between time- and norm-optimality (see also [54]) is used in an essential way. In particular, the conditions (2.1) and (2.3) (which are independent of $\bar{\mu}_0$ in this case) are obtained for the problem of steering the heat equation into zero with pointwise bounded controls restricted to an arbitrary subset of the underlying domain. In this case, the multiplier is obtained in a space of distributions, larger than L^2. However, this technique seems to be restricted to the case $L \equiv 0$ and yields a different condition instead of (2.2) to characterize the optimality of the time variable.

To further assess the applicability of the strengthened Hamiltonian condition (2.5) in the context of concrete examples, in Section 2.4 we discuss several cases when A is given by a general convection-diffusion operator on a bounded domain Ω. On the one hand, we find that (2.5) always holds for the control of, say, the heat equation into a $L^2(\Omega)$-ball centered at a sufficiently small u_d, assuming only that the zero control is admissible. We emphasize that this already includes the classical setting $u_d = 0$ considered in, e.g., [149, 159, 162], without further assumptions. On the other hand, we find that it is fulfilled for more restrictive target sets or more general convection-diffusion operators only under additional assumptions on the form of the control operator and the admissible set. We compare these requirements to established controllability assumptions (see, e.g., [167]) and find that our conditions are stronger, in general. This can be connected to the fact that the cost of the controls resulting from controllability conditions (see [56]) grows exponentially if the length of the control horizon is decreased towards zero. However, for general A, we also give an example of a special target set where (2.5) follows directly from an established stabilizability assumption, based on the Fattorini criterion, which can be fulfilled even with finite-dimensional controls.

Clearly, as (2.5) implies (2.4), we implicitly only consider systems that are weakly invariant. This can also be justified from a practical point of view. Note first, that we only require the state to be inside the target set at the final time T in the mathematical formulation of the time optimal problem (P). However, in practice, time continues to advance afterwards and in many cases we are interested to remain inside of the target set. Therefore, it seems to be reasonable to restrict attention to systems where this is always possible. Otherwise, the optimal control might achieve $u(T) \in U$ with small cost, but every trajectory continuing from $u(T)$ might be forced to leave the target set again (possibly immediately).

We appreciate that (2.5) might not be fulfilled in all practically relevant cases. However, we anticipate that it is useful in many situations, where the objective is to steer the system "sufficiently close" to a weakly invariant, or even asymptotically stable state u_d; cf., e.g., [2, 43, 92]. Here, it could also help to guide the choice of appropriate target sets U, which guarantee both that the terminal state will be close to u_d, and that the resulting control problem will be strongly stable. We also note that, if the optimal trajectory $\bar{u}$ is assumed to be known and U has finite co-dimension with regular normal cone, condition (2.5) can be weakened to

$$h(\bar{u}(T), \zeta) \leq -h_0 \|\zeta\|_H \quad \text{for all } \zeta \in N_U(\bar{u}(T)), \tag{2.6}$$

while still implying the qualified form of the optimality conditions; see Proposition 2.30. Furthermore, if the normal cone contains only one element, this condition is already equivalent to the qualified optimality conditions (see Proposition 2.31), which further clarifies the role of the strengthened Hamiltonian condition.

Viewing (P) as an abstract constrained nonconvex optimization problem, one could also require a *constraint qualification* (CQ) to guarantee the qualified form of the optimality conditions. However, the concrete form of the standard CQs does not only depend on the parametrization of the constraint, but also on objects such as gradients, which require a proper (but in some sense arbitrary) parametrization of the time variable T; see Section 2.3.2. Therefore, strong stability appears to be the more straightforward tool in this context. Comparing CQs to the strengthened Hamiltonian condition (2.5), we remark that the latter qualifies all optimal solutions at once, whereas the other considers only one specific, but a priori unknown solution, similar to (2.6).

This chapter is organized as follows: In Section 2.1 we introduce some notation and state the main assumptions. Weak invariance is characterized in Section 2.2. The concept of strong stability is introduced in Section 2.3, where we discuss the time-optimal control problem and derive optimality conditions. Moreover, we show that strengthening of weak invariance implies strong stability as well as further perturbation results. Last, Section 2.4 is devoted to applications. The text will be accompanied by the illustrative example $U = \{ u_d \}$ with fixed $u_d \in H$, to make ideas visible to the reader. However, we emphasize that it does not represent the main application.

2.1. Notation and main assumptions

For any two Banach spaces X and Y we use $Y \hookrightarrow X$ to denote the continuous embedding and $Y \hookrightarrow_c X$ for the continuous and compact embedding. The domain of a linear (possibly unbounded) operator A on X is denoted by $\mathcal{D}_X(A)$. Let V and H be real Hilbert spaces such that $V \hookrightarrow_c H \cong H^* \hookrightarrow V^*$ form a Gelfand triple. Without restriction suppose $\|v\|_V \geq \|v\|_H$ for all $v \in V$. In general, we abbreviate the duality pairing and the inner product and norm in H by

$$\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{V^*,V}, \quad (\cdot, \cdot) = (\cdot, \cdot)_H, \quad \|\cdot\| = \|\cdot\|_H.$$

Assumption 2.1. Let $a \colon V \times V \to \mathbb{R}$ be a continuous bilinear form, which satisfies the Gårding inequality (which is also referred to as weak coercivity): we assume there are constants $\alpha_0 > 0$ and $\omega_0 \geq 0$ such that

$$a(u,u) + \omega_0 \|u\|^2 \geq \alpha_0 \|u\|_V^2, \quad u \in V. \tag{2.7}$$

We denote by $A \colon V \subset V^* \to V^*$ the unique linear operator with

$$\langle Au, v \rangle = a(u,v) \quad \text{for all } v \in V.$$

It holds $\mathcal{D}_{V^*}(A) = V$; see, e.g., [80, Theorem 3.4]. Due to the Gårding inequality, the operator $-(A + \omega_0)$ generates an analytic semigroup on V^*; see, e.g., [127, Section 1.4]. We abbreviate $\omega_0 \operatorname{Id}$ by ω_0, where Id is the identity operator on V^*, to simplify the presentation. Due to (2.7), we can define fractional powers in the sense of [128, Section 2.6]. For fixed $\theta \geq 0$, we abbreviate $X_\theta = \mathcal{D}_{V^*}((A + \omega_0)^\theta)$ and introduce the norm on X_θ as

$$\|\cdot\|_{X_\theta} \coloneqq \|(A + \omega_0)^\theta \cdot\|_{V^*}.$$

As usual, $(V^*, V)_{\theta,s}$, respectively $[V^*, V]_\theta$, stand for the real, respectively complex interpolation couple with $\theta \in (0,1)$ and $s \in (1,\infty)$. Since V is a Hilbert space (and thus V^* as well), the operator $(A + \omega_0)$ has bounded imaginary powers and it holds for $\theta \in (0,1)$ that $X_\theta = [V^*, V]_\theta = (V^*, V)_{\theta,2}$; see, e.g., [146, Section 1.15.3]. In particular, $X_{1/2} = H$; see, e.g., [108, Section 1.2.4]. Moreover,

$$X_\theta^* = [V^*, V]_\theta^* = [V^*, V]_{1-\theta} = X_{1-\theta};$$

see, e.g., [146, Theorems 1.9.3 b), 1.11.3]. Furthermore, using [146, Theorems 1.9.3 b), 1.11.3 and 1.15.3] we find

$$X_{1-\theta} = [V^*, V]_{1-\theta} = [[V^*, V]_{1/2}, V]_{1-2\theta} = [H, V]_{1-2\theta}. \tag{2.8}$$

For any set $S \subset Y$ in a Banach space Y, let $d_S^Y(\cdot)$ denote the distance function

$$d_S^Y(y) \coloneqq \inf_{y' \in S} \|y - y'\|_Y.$$

Furthermore, if Y is a Hilbert space and S is closed and convex, we denote by $P_S^Y : Y \to S$ the *minimizing projection* to S. Note that P_S^Y is Lipschitz continuous in Y (with Lipschitz constant one); see, e.g., [12, Proposition 4.8]. We denote by

$$N_S^Y(y) \coloneqq \{\, v \in Y^* \colon \langle v, y' - y \rangle_{Y^*,Y} \leq 0 \ \text{ for all } y' \in S \,\}$$

the *normal cone* to S at the point $y \in S$. In the case $Y = H$ and $S = U$ (or if no ambiguity arises), we simply write $d_U(\cdot)$, P_U, and $N_U(\cdot)$.

Concerning the problem (P), the terminal set $U \subset H$ is assumed to be nonempty, closed, and convex and the initial state satisfies $u_0 \in H$.

Assumption 2.2. Let Q be a Hilbert space, and Q_{ad} be a closed convex subset. We assume the control operator $B \colon Q \to X_{\theta_0} \hookrightarrow V^*$ for some $\theta_0 \in (0, 1/2]$ to be linear and continuous. In addition, we assume Q_{ad} to be bounded in Q, and define $C_{Q_{ad}} = \max_{q \in Q_{ad}} \|q\|_Q$. Furthermore, the functional $L \colon Q \to \mathbb{R}_+$ is Lipschitz continuous on Q_{ad} and convex.

In addition, for $T > 0$ we define $Q(0,T) \coloneqq L^2((0,T); Q)$ and

$$Q_{ad}(0,T) = \{\, q \in Q(0,T) \colon q(t) \in Q_{ad} \ \text{ a.e. } t \in (0,T) \,\} \subset L^\infty((0,T); Q).$$

Moreover, for $T > 0$ we use the symbol $W(0,T)$ to abbreviate $H^1((0,T); V^*) \cap L^2((0,T); V)$, endowed with the canonical norm and inner product. The symbol $i_T \colon W(0,T) \to H$ denotes the continuous trace mapping $i_T u = u(T)$.

2.2. Weak invariance

We first introduce the notion of weak invariance.

Definition 2.1. The set $U \subset H$ is said to be *weakly invariant* under (A, BQ_{ad}), if for every $u_0 \in U$ there exists a control $q \colon [0,\infty) \to Q_{ad}$ such that the solution u to

$$\partial_t u + Au = Bq, \quad u(0) = u_0,$$

satisfies $u(t) \in U$ for all $t \geq 0$. If ambiguity is not to be expected, we simply say U is weakly invariant.

Remark 2.2. Different terms for weak invariance are being used in the literature, such as *holdability* or *viability*; cf. [141] and [41, Section 1].

The structure of this section is as follows: We first discuss stability of the minimizing projection P_U in V. This is then needed to characterize weak invariance in terms of the lower Hamiltonian.

2.2.1. Stability of the projection to the target set

We call the minimizing projection P_U in H onto U stable in V, if $P_U(V) \subset V$. In general, stability of P_U in V is a non-trivial assumption. However, in the uncontrolled case, it is known that invariance of U under A (i.e., the property $\mathrm{e}^{-tA}U \subset U$ for all $t \geq 0$, with e^{-tA} the semigroup generated by $-A$) implies the stability of P_U in V; see, e.g., [127, Theorem 2.2], cf. also [4, Section II.6.3] for the nonautonomous case. In the following we generalize this known sufficient condition for stability of P_U in V to controlled systems. This will be a prerequisite for the characterization of weak invariance of U under (A, BQ_{ad}).

Example 2.3. As an illustrative example, we consider the set $U = \{ u_d \}$. The projection P_U is given by $P_U(u) = u_d$. Clearly, P_U is stable in V if and only if $u_d \in V$. Now, suppose that U is weakly invariant under (A, BQ_{ad}). Then there is a control $q\colon [0,\infty) \to Q_{ad}$ such that the corresponding state u satisfies $u(t) = u_d$ for all $t \geq 0$, i.e. u is the steady state solution. Thus, $0 = \partial_t u(t) = Bq(t) - Au_d$, which in turn leads to $u_d \in V$, in accordance with the results of this section. Additionally, we infer that invariance of U under the uncontrolled system (corresponding to weak invariance with the trivial choice $Q_{ad} = \{ 0 \}$) holds only for $Au_d = 0$; see [127, Theorem 2.2], cf. also Theorem 2.9.

The proof is divided into two steps. Roughly speaking, we first prove that for a weakly invariant set U, the scaled resolvent of A does not map points in U too far outside of U. We define for any $u \in H$

$$E_\lambda u \coloneqq \lambda(\lambda + A)^{-1}u = (1 + A/\lambda)^{-1}u.$$

Provided that $\lambda \geq \omega_0$, where ω_0 was defined in (2.7), we find that $E_\lambda u \in X_1 = V$ is well defined for any $u \in X_0 = V^*$. Additionally, using a resolvent identity and the interpolation inequality, there holds the estimate $\|E_\lambda u - u\|_{V^*} = \lambda^{-1}\|AE_\lambda u\|_{V^*} \leq c\,\lambda^{-1/2}\|u\|$ for all $u \in H = X_{1/2}$. For $u \in U$, an improved estimate for the distance of $E_\lambda u$ to U can be obtained under weak invariance.

Proposition 2.4. *Suppose that U is weakly invariant under (A, BQ_{ad}) and let θ_0 be the constant from Assumption 2.2. Then, for all $u \in U$ and $\gamma \in [0, 1/2]$ it holds*

$$d_U^{X_\gamma}(E_\lambda u) \leq c\,\lambda^{-1+(\gamma-\theta_0)^+}, \quad \lambda \geq \omega_0,$$

where $(\cdot)^+ = \max\{ \cdot, 0 \}$ denotes the positive part, and the constant c depends only on γ, θ_0, A, and Q_{ad}.

Proof. By assumption, there is a control such that the state $\breve{u}$ with initial value u stays in U for all $t \geq 0$. Now, we can estimate the distance of $\mathrm{e}^{-tA}u$ to U in X_γ by the distance of $\breve{u}(t)$, and obtain

$$d_U^{X_\gamma}(\mathrm{e}^{-tA}u) \leq \|\mathrm{e}^{-tA}u - \breve{u}(t)\|_{X_\gamma} \leq c\,t^{1-(\gamma-\theta_0)^+},$$

where the last inequality is an application of Proposition A.18 (iii) with $\theta = \min\{ \gamma, \theta_0 \}$. Indeed, the variable $w(t) = \mathrm{e}^{-tA}u - \breve{u}(t)$ solves a parabolic equation with right-hand side in

$L^\infty(0,\infty;X_\theta)$ and $w(0)=0$. Since the resolvent is the Laplace transform of the semigroup it holds

$$E_\lambda u = \lambda(\lambda+A)^{-1}u = \int_0^\infty \lambda \mathrm{e}^{-\lambda t}\mathrm{e}^{-tA}u\,\mathrm{d}t.$$

Note, that due to $u\in U\subset H = X_{1/2}$ and $\lambda\geq\omega_0$, the integral is defined with values in X_γ for all $\gamma\leq 1/2$; cf. [128, Sect. 1.7]. Finally, we apply the distance function on both sides of the equation, and we derive

$$\begin{aligned} d_U^{X_\gamma}(E_\lambda u) &\leq \int_0^\infty \lambda \mathrm{e}^{-\lambda t} d_U^{X_\gamma}(\mathrm{e}^{-tA}u)\,\mathrm{d}t \\ &\leq c\int_0^\infty \lambda \mathrm{e}^{-\lambda t} t^{1-(\gamma-\theta_0)^+}\,\mathrm{d}t = c\,\Gamma(2+\theta-\gamma)\,\lambda^{-1+(\gamma-\theta_0)^+}, \end{aligned}$$

with $\int_0^\infty \lambda \mathrm{e}^{-\lambda t}=1$, convexity of the distance function, and a generalized Jensen's inequality; see, e.g., [129, Theorem 3.10 (ii)]. □

Remark 2.5. Note that for the result of Proposition 2.4, we only used the assumption that BQ_{ad} is a bounded set in X_{θ_0} (using Assumption 2.2). All the results from this section remain valid under this modified assumption.

Lemma 2.6. *If U is weakly invariant under (A, BQ_{ad}), then the projection P_U is stable in V, i.e. $P_U(V)\subseteq V$.*

Proof. Let $v\in V$ be fix and set $u=P_U(v)\in H$. We first prove that $u\in X_{(n-1)/n}$ with $n=2^m$ for all $m\geq 1$. Since $u\in H = X_{1/2}$, the assertion holds for $m=1$. Proceeding by induction, we assume it holds for all $1\leq m'\leq m$ and show it for $2n=2^{m+1}$. Since $AE_\lambda u = \lambda(u-E_\lambda u)$, we compute

$$\begin{aligned} \langle AE_\lambda u, E_\lambda u\rangle &= \langle AE_\lambda u, E_\lambda u - u\rangle + \langle AE_\lambda u, u\rangle \\ &= \lambda(u-E_\lambda u, E_\lambda u - u) + \langle AE_\lambda u, u\rangle = -\lambda\|u-E_\lambda u\|^2 + \langle AE_\lambda u, u\rangle. \end{aligned}$$

Now, we take for any λ a $u'_\lambda\in U$ with $\|u'_\lambda - E_\lambda u\|_{X_{1/n}} \leq 2\, d_U^{X_{1/n}}(E_\lambda u)$. Moreover, since $X_\theta^* = X_{1-\theta}\hookrightarrow V^*$, it holds $\langle\varphi,\psi\rangle\leq \|\varphi\|_{[V^*,V]_{1-\theta}}\|\psi\|_{[V^*,V]_\theta}$ for $\varphi\in X_{1-\theta}$ and $\psi\in V$. Thus, for $v\in V$ with $u=P_U(v)$ from the beginning of the proof it holds

$$\begin{aligned} \langle AE_\lambda u, E_\lambda u\rangle + \lambda\|u-E_\lambda u\|^2 &= \langle AE_\lambda u, u-v\rangle + \langle AE_\lambda u, v\rangle \\ &= \lambda(u-u'_\lambda, u-v) + \lambda(u'_\lambda - E_\lambda u, u-v) + \langle AE_\lambda u, v\rangle \\ &\leq 0 + \lambda\|u'_\lambda - E_\lambda u\|_{X_{1/n}}\|u-v\|_{X_{(n-1)/n}} + c\|E_\lambda u\|_V\|v\|_V \\ &\leq c\,\lambda^{(1/n-\theta_0)^+}\|u-v\|_{X_{(n-1)/n}} + c\|E_\lambda u\|_V\|v\|_V, \end{aligned} \tag{2.9}$$

where we have used $(u-u'_\lambda, u-v) = (u-u'_\lambda, P_U(v)-v)\leq 0$, the estimate $d_U^{X_{1/n}}(E_\lambda u)\leq c\lambda^{-1+(1/n-\theta_0)^+}$ (from Proposition 2.4 with $\gamma=1/n$), and the continuity of A. Consequently, with Young's inequality, we arrive at

$$\langle AE_\lambda u, E_\lambda u\rangle + \lambda\|u-E_\lambda u\|^2 \leq c\,\lambda^{(1/n-\theta_0)^+}\|u-v\|_{X_{(n-1)/n}} + \frac{\alpha_0}{2}\|E_\lambda u\|_V^2 + c\|v\|_V^2,$$

and the Gårding inequality (2.7) yields

$$\frac{\alpha_0}{2}\|E_\lambda u\|_V^2 + \lambda\|u-E_\lambda u\|^2 \leq c\,\lambda^{(1/n-\theta_0)^+}\|u-v\|_{X_{(n-1)/n}} + c\|v\|_V^2 + \omega_0\|E_\lambda u\|^2.$$

With $\|E_\lambda u\| \le c\|u\| \le c\|v\|$ we obtain constants c_1 and c_2 (depending on the norms of $v \in V$ and $u \in X_{(n-1)/n}$, by the induction hypothesis) such that for all $\lambda \ge \omega_0$ it holds

$$\|E_\lambda u\|_V + \lambda^{1/2}\|u - E_\lambda u\| \le c_1 \lambda^{(1/n-\theta_0)^+/2} + c_2.$$

Note that the constants c_1 and c_2 depend on n. However, for the proof we only require finitely many steps which can be estimated a priori by $1/n \le \theta_0$. Recall the functional of the K-method of real interpolation, see, e.g., [146, Section 1.3],

$$K(u,t,V,H) = \inf_{\tilde{u}\in V} \left[\|\tilde{u}\|_V + t\|u-\tilde{u}\|\right].$$

By inserting for each $t \ge t_{\min} \coloneqq \max\{1, \sqrt{\omega_0}\}$ the values $\tilde{u} = E_\lambda u$ for $\lambda = t^2$, we obtain the estimate $K(u,t,V,H) \le c_1 t^{(1/n-\theta_0)^+} + c_2$. Moreover, inserting $\tilde{u} = 0$ yields $K(u,t,V,H) \le t\|u\| \le ct$. Thereby, we obtain

$$\begin{aligned}\|u\|^2_{(V,H)_{1/n,2}} &= \int_0^\infty \left(t^{-1/n}K(u,t,V,H)\right)^2 t^{-1}\,\mathrm{d}t \\ &\le c\int_0^{t_{\min}} t^{1-2/n}\,\mathrm{d}t + \int_{t_{\min}}^\infty \left(c_1 \max\{t^{-1/n}, t^{-\theta_0}\} + c_2 t^{-1/n}\right)^2 t^{-1}\,\mathrm{d}t < \infty.\end{aligned}$$

As in (2.8) with $\theta = 1/(2n)$, we find $(V,H)_{1/n,2} = X_{1-1/(2n)}$. Therefore, $u \in X_{(2n-1)/2n}$ and we have shown the assertion for $2n = 2^{m+1}$.

Finally, let $n \in \mathbb{N}$ such that $1/n \le \theta_0$. Then, in the last step of (2.9) we obtain that

$$\|E_\lambda u\|_V^2 \le c\|u-v\|_{X_{1-(n-1)/n}} + c\|v\|_V^2.$$

Thus, $E_\lambda u$ is uniformly bounded in V. As $E_\lambda u \to u$ in H, we conclude $u \in V$. □

Corollary 2.7. *Under the assumptions of Lemma 2.6, there exist constants $c_1, c_2 > 0$, such that*

$$\|P_U(v)\|_V \le c_1 + c_2\|v\|_V. \tag{2.10}$$

Proof. Let $v \in V$. Then, $u = P_U(v) \in U \cap V$ due to Lemma 2.6. As in the last step of Lemma 2.6, we derive

$$\|E_\lambda u\|_V^2 \le c_1 \|u-v\|_V + c_2 \|E_\lambda u\|_V \|v\|_V \le c_1 (\|u\|_V + \|v\|_V) + c_2 \|E_\lambda u\|_V \|v\|_V.$$

Recall that $E_\lambda u = \lambda(\lambda + A)^{-1}u$. Since $u \in V$, it holds $E_\lambda u \to u$ in V. Passing to the limit in the inequality above yields

$$\|u\|_V^2 \le c_1 (\|u\|_V + \|v\|_V) + c_2 \|u\|_V \|v\|_V.$$

Dividing by $\|u\|_V$, we conclude $\|P_U(v)\|_V \le \max\{1, c_1(1 + \|v\|_V) + c_2\|v\|_V\}$ and the assertion follows for appropriately modified constants c_1, c_2. □

2.2.2. Characterization of invariance

Using the result on the stability of the projection, weak invariance can be characterized by conditions involving either the projection or the normal cone. In the following, we will make repeated use of the following basic identification.

Proposition 2.8 (see [12, Proposition 6.46])**.** *Let $u \in U$. Then*

$$N_U(u) = \{\, v - u \colon v \in H \text{ with } P_U(v) = u \,\}.$$

In particular, it holds $v - P_U(v) \in N_U(P_U(v))$ for all $v \in H$, and $P_U(u+\zeta) = u$ for all $u \in U$ and $\zeta \in N_U(u)$.

Following [40, Section 12.1], we define the *lower Hamiltonian* as

$$h(u,\zeta) = \min_{q \in Q_{ad}} \langle Bq - Au, \zeta \rangle \quad \text{for } u \in V,\ \zeta \in V.$$

Analogous to the corresponding theory for ordinary differential equations, we can now characterize weak invariance in terms of the lower Hamiltonian.

Theorem 2.9. *The following conditions are equivalent:*

(i) U is weakly invariant,

(ii) P_U is stable in V and $h(u,\zeta) \le 0$ for all $u \in U \cap V$ and $\zeta \in N_U(u) \cap V$,

(iii) P_U is stable in V and $h(P_U(v), v - P_U(v)) \le 0$ for all $v \in V$.

For the proof of Theorem 2.9 we need an estimate of the distance to the target set for the controlled system, which is given next. For later use, we prove it in a more general form, including both the strengthened condition (2.5) as well as the weaker condition (2.4) (which is the special case for $h_0 = 0$).

Lemma 2.10. *Suppose that P_U is stable in V and that there is $h_0 \ge 0$ such that for all $v \in V$ we have*

$$h(u,\zeta) \le -h_0 \|\zeta\|, \quad \text{where } u = P_U(v),\ \zeta = v - u. \tag{2.11}$$

Then, for each $u_0 \in H$ with $d_U(u_0)\,\omega_0 \le h_0$ there exists a control $q \colon [0,\infty) \to Q_{ad}$ such that the solution u to

$$\partial_t u + Au = Bq, \quad u(0) = u_0,$$

satisfies

$$d_U(u(t)) \le \max\{\, 0,\ d_U(u_0) + (d_U(u_0)\,\omega_0 - h_0)\, t \,\} \quad \text{for } t \ge 0.$$

To prove this result, we construct a sequence of feedback controls which have approximately the desired property, and then we go to the limit. We start with an auxiliary result.

Proposition 2.11. *The squared distance function $d_U^2 \colon H \to \mathbb{R}$ is differentiable with*

$$\nabla d_U^2(u) = 2(u - P_U(u)).$$

Moreover, if P_U is stable in V, then ∇d_U^2 is continuous from V to $X_{1-\theta_0}$.

Proof. Differentiability of the squared distance function is proved in [12, Corollary 12.30]. Using the expression of the derivative, we infer that ∇d_U^2 is Lipschitz continuous on H with Lipschitz constant two, and stable on V due to stability of P_U in V; see Corollary 2.7. The interpolation inequality [146, Theorem 1.9.3 f)] yields

$$\begin{aligned}\frac{1}{2}\|\nabla d_U^2(u) - \nabla d_U^2(v)\|_{[H,V]_{1-2\theta_0}} &\le \frac{1}{2}\|\nabla d_U^2(u) - \nabla d_U^2(v)\|_V^{1-2\theta_0} \|\nabla d_U^2(u) - \nabla d_U^2(v)\|^{2\theta_0} \\ &\le [2c_1 + (1+c_2)(\|v\|_V + \|u\|_V)]^{1-2\theta_0}\, \|u - v\|^{2\theta_0},\end{aligned}$$

where c_1, c_2 are from estimate (2.10). Hence, ∇d_U^2 is continuous from V to $[H,V]_{1-2\theta_0} = X_{1-\theta_0}$; see (2.8). □

We now construct the desired sequence of approximate feedback controls.

Proposition 2.12. *Let $u_0 \in H$, $\gamma > 0$ and $T > 0$. Then the equation*

$$\begin{aligned} \partial_t u_\gamma + A u_\gamma &= B q_\gamma, \\ q_\gamma &= P_{Q_{ad}}\left(-\gamma^{-1} B^*(u_\gamma - P_U(u_\gamma))\right), \\ u_\gamma(0) &= u_0, \end{aligned} \tag{2.12}$$

possesses a solution $u_\gamma \in W(0,T) \cap C((0,T);V) \cap C^1((0,T);V^*)$ and $q_\gamma \in C((0,T);Q)$.

Proof. Consider the mapping $\mathcal{F}\colon Q(0,T) \to Q(0,T)$ defined by

$$\mathcal{F}(q) \coloneqq P_{Q_{ad}}\left(-(2\gamma)^{-1} B^*\left[\nabla d_U^2\left(\mathcal{S}(u_0, Bq)\right)\right]\right),$$

where $\mathcal{S}\colon H \times L^2((0,T);X_{\theta_0}) \to W(0,T)$ denotes the solution operator of the parabolic equation with initial value u_0 and right-hand side Bq. According to Proposition 2.11, the function ∇d_U^2 is continuous from V into $X_{1-\theta_0}$. Moreover, since $X_{\theta_0}^* = X_{1-\theta_0}$, and B is supposed to be continuous from Q to X_{θ_0}, we infer continuity of B^* from $X_{1-\theta_0}$ to $Q^* = Q$. Continuity of $P_{Q_{ad}}$ on Q leads to continuity of $\mathcal{F}$ from $Q(0,T)$ into itself. Using compactness of $q \mapsto \mathcal{S}(u_0, Bq)$ into $L^2((0,T);V)$ according to Proposition A.19, we deduce that $\mathcal{F}(Q_{ad}(0,T))$ is contained in a compact subset of $Q(0,T)$.

Finally, Schauder's fixed point theorem (see, e.g., [163, Theorem 2.A]) yields the existence of a fixed point $\mathcal{F}(q_\gamma) = q_\gamma$. Setting $u_\gamma = \mathcal{S}(u_0, q_\gamma)$ proves the existence of a solution to (2.12). According to Proposition A.18, u_γ is continuous on $(0,T]$ with values in V. Now, the continuity of the projection $P_{Q_{ad}}$ on Q yields the improved regularity of q_γ. Furthermore, from $\partial_t u_\gamma = Bq_\gamma - Au_\gamma$ we deduce that u_γ is continuously differentiable on $(0,T)$ with values in V^*. □

Next, we observe that the feedback control q_γ is close to the minimizing argument of the lower Hamiltonian.

Proposition 2.13. *For any $\zeta, u \in V$ and $q_\gamma = P_{Q_{ad}}\left(-\gamma^{-1} B^* \zeta\right)$ it holds*

$$\langle Bq_\gamma - Au, \zeta\rangle \le h(u,\zeta) + c\gamma, \tag{2.13}$$

where c solely depends on Q_{ad}.

Proof. Consider for $\gamma \ge 0$ the family of functions defined by

$$h_\gamma(u,\zeta) = \min_{q \in Q_{ad}} \left[\langle Bq - Au, \zeta\rangle + \frac{\gamma}{2}\|q\|_Q^2\right]. \tag{2.14}$$

Clearly, h_0 is the lower Hamiltonian h. Denote the minimizers of (2.14) by q_γ. Then, we estimate

$$\langle Bq_\gamma - Au, \zeta\rangle \le h_\gamma(u,\zeta) \le \langle Bq_0 - Au, \zeta\rangle + \frac{\gamma}{2}\|q_0\|_Q^2 \le h_0(u,\zeta) + \frac{\gamma}{2} C_{Q_{ad}}^2.$$

Furthermore, for $\gamma > 0$, from the optimality conditions for (2.14) we infer that the minimizer q_γ is given by $q_\gamma = P_{Q_{ad}}\left(-\gamma^{-1} B^* \zeta\right)$. □

Now we prove the main result of this section.

Proof of Lemma 2.10. Clearly, it suffices to show the result for $t \in (0,T)$ for some arbitrary but fixed $T > 0$. Let $u_0 \in H$ be given, let u_γ for $\gamma > 0$ denote the corresponding solution to (2.12), and define $d_\gamma(t) = d_U(u_\gamma(t))$. Then, for any $0 < t < T$ we infer

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} d_\gamma^2(t) &= \langle \partial_t u_\gamma(t), \nabla d_U^2(u_\gamma(t))\rangle = \langle Bq_\gamma(t) - Au_\gamma(t), \nabla d_U^2(u_\gamma(t))\rangle \\ &= \langle Bq_\gamma(t) - AP_U(u_\gamma(t)), \nabla d_U^2(u_\gamma(t))\rangle + \langle AP_U(u_\gamma(t)) - Au_\gamma(t), \nabla d_U^2(u_\gamma(t))\rangle,\end{aligned}$$

where we have used (2.12). For the last term, the Gårding inequality yields

$$\begin{aligned}\langle AP_U(u_\gamma(t)) - Au_\gamma(t), \nabla d_U^2(u_\gamma(t))\rangle &= -\frac{1}{2}\langle A\nabla d_U^2(u_\gamma(t)), \nabla d_U^2(u_\gamma(t))\rangle \\ &\le \frac{\omega_0}{2}\|\nabla d_U^2(u_\gamma(t))\|^2 - \frac{\alpha_0}{2}\|\nabla d_U^2(u_\gamma(t))\|_V^2 \le \frac{\omega_0}{2}\|\nabla d_U^2(u_\gamma(t))\|^2.\end{aligned}$$

Employing (2.13), the Hamiltonian condition (2.11), and

$$\|\nabla d_U^2(u_\gamma(t))\| = 2d_U(u_\gamma(t)) = 2d_\gamma(t),$$

we infer

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} d_\gamma^2(t) &\le \frac{1}{2} h(P_U(u_\gamma(t)), \nabla d_U^2(u_\gamma(t))) + c\gamma + \frac{\omega_0}{4}\|\nabla d_U^2(u_\gamma(t))\|^2 \\ &\le -h_0 d_\gamma(t) + c\gamma + \omega_0 d_\gamma^2(t).\end{aligned} \tag{2.15}$$

Using the fact that $\frac{\mathrm{d}}{\mathrm{d}t} d_\gamma^2(t) = 2d_\gamma'(t)d_\gamma(t)$, we obtain from (2.15) that

$$d_\gamma'(t) \le \omega_0 d_\gamma(t) + c\gamma/d_\gamma(t) - h_0 \quad \text{on } \{\, t\colon d_\gamma(t) > 0 \,\}.$$

According to Proposition A.25 the differential inequality implies

$$d_\gamma(t) \le \max\{\, \sqrt{\gamma}, (d_U(u_0) + \sqrt{\gamma})\mathrm{e}^{\omega_0 t} + (c\sqrt{\gamma} - h_0)\phi(t) \,\} =: D_\gamma(t), \tag{2.16}$$

where $\phi(t) = \omega_0^{-1}(\mathrm{e}^{\omega_0 t} - 1)$, if $\omega_0 > 0$, and $\phi(t) = t$ otherwise.

For $\gamma \to 0$ we now choose suitable subsequences such that $q_\gamma \rightharpoonup q$ in $Q(0,T)$ and $u_\gamma \rightharpoonup u$ in $W(0,T)$. Clearly, the weak limits satisfy

$$\partial_t u + Au = Bq, \quad u(0) = u_0.$$

Thus, with $W(0,T) \hookrightarrow C([0,T];H)$ we have $u_\gamma(t) \rightharpoonup u(t)$ in H for all $t \in [0,T]$. Using weak lower semicontinuity of the distance function $d_U(\cdot)$ and (2.16), we obtain

$$d_U(u(t)) \le \liminf_{\gamma\to 0} d_U(u_\gamma(t)) \le \lim_{\gamma\to 0} D_\gamma(t) = \max\Big\{\, 0,\ d_U(u_0)\mathrm{e}^{\omega_0 t} - h_0\phi(t) \,\Big\}.$$

Now, using the supposition $d_U(u_0)\omega_0 \le h_0$, the definition of ϕ, and the fact that $\phi(t) \ge t$, we obtain

$$d_U(u(t)) \le (d_U(u_0) + (\omega_0 d_U(u_0) - h_0)\phi(t))^+ \le (d_U(u_0) + (\omega_0 d_U(u_0) - h_0)\, t)^+$$

concluding the proof. □

Finally, we show the charactization of weak invariance by means of the lower Hamiltonian.

Proof of Theorem 2.9. We separately prove three implications.

(i) ⇒ (ii). The stability of P_U in V follows with Lemma 2.6. For the second property, let $u_0 \in U \cap V$ be arbitrary. Then, with weak invariance, there is a control $q \in Q_{ad}(0,\infty)$ such that the corresponding state satisfies $u(0) = u_0$ and $u(t) \in U$ for all $t \geq 0$. Additionally, $u(t) \in V$ for all $t \geq 0$ follows by Proposition A.18 (i). Let further $\zeta \in N_U(u_0) \cap V$. It holds $\partial_t u = Bq - Au$ in $L^2((0,s);V^*)$ for any $s > 0$, and we have

$$0 \geq \frac{1}{s}\langle u(s) - u_0, \zeta\rangle = \left\langle \frac{1}{s}\int_0^s [Bq(t) - Au(t)]\,\mathrm{d}t, \zeta \right\rangle. \tag{2.17}$$

Define the temporal averages $\bar{q}_s = (1/s)\int_0^s q(t)\,\mathrm{d}t$ and $\bar{u}_s = (1/s)\int_0^s u(t)\,\mathrm{d}t$. Due to $u \in C([0,1];V)$, it holds $\bar{u}_s \to u_0$ in V for $s \to 0$. Furthermore, with $q(t) \in Q_{ad}$ for all t, it follows $\bar{q}_s \in Q_{ad}$ (see, e.g., [40, Exercise 2.44]) and we can select a sequence $s_n \to 0$ and a $q_0 \in Q$ such that $\bar{q}_{s_n} \rightharpoonup q_0$ in Q for $n \to \infty$. By weak closedness of Q_{ad} we have $q_0 \in Q_{ad}$. Going to the limit in (2.17), we obtain

$$0 \geq \langle Bq_0 - Au_0, \zeta\rangle \geq h(u_0, \zeta),$$

using boundedness of $B\colon Q \to V^*$ and $A\colon V \to V^*$. Since u_0 and ζ were arbitrary, we finish the proof.

(ii) ⇒ (iii). This follows directly from the fact that $u = P_U(v) \in U \cap V$ and $v - P_U(v) \in N_U(u) \cap V$ for all $v \in V$ with the stability of the projection.

(iii) ⇒ (i). The last implication is consequence of Lemma 2.10 (with $h_0 = 0$). □

2.3. Time-optimal control problem

We now turn to the time-optimal control problem. In the following, we use the notation $u[q]$ for the solution of the state equation $\partial_t u + Au = Bq$ and $u(0) = u_0$ for a given control q. Let $U \subset H$ denote the terminal set that is assumed to be closed and convex. Furthermore, to exclude the trivial case with zero optimal time, we assume that $u_0 \in H \setminus U$. Problem (P) can then be restated as:

$$\inf_{T>0,\, q\in Q_{ad}(0,T)} j(T,q) \quad \text{subject to} \quad u[q](T) \in U.$$

First, we consider the question of existence of optimal controls. We show that if there exists a feasible pair $(T,q) \in \mathbb{R}_+ \times Q_{ad}(0,T)$, the problem is well-posed:

Proposition 2.14. *Suppose there exists a finite time $T > 0$ and a feasible control $q \in Q_{ad}(0,T)$ such that the corresponding state satisfies $u[q](T) \in U$. Then, problem (P) admits at least one optimal solution $(T,\bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,T)$.*

Proof. The proof is done by standard arguments (the direct method); cf., e.g., [106, Section III.17]. We use in particular the boundedness of j for bounded T due to boundedness of the admissible set Q_{ad}, $j(T,q) \geq T$ and that j is weakly lower semicontinuous in q for fixed T. Furthermore, we use the $W(0,T)$ regularity of the solution to the state equation, the continuity of the trace mapping i_T, and the convexity of U. □

Remark 2.15. In view of the preceding result, the question of existence reduces to the question of *controllability* under constraints. We exemplary state situations where feasible controls exist. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with smooth boundary and $A = -\Delta$ be the usual Laplace operator equipped with homogeneous Dirichlet boundary conditions. Moreover, for fixed $u_d \in L^2(\Omega)$ and $\delta_0 \geq 0$, suppose the terminal set to be given by $U = \{\, u \in L^2(\Omega) \colon \|u - u_d\| \leq \delta_0 \,\}$.

(i) In case of distributed control on an open subset $\omega \subset \Omega$, the state equation is known to be approximately controllable; see, e.g., [138, 167], i.e. for all $T > 0$, $u_d \in L^2(\Omega)$, and $\delta_0 > 0$, there exists a control $q \in Q(0,T)$ such that $u[q](T) \in U$. Clearly, for sufficiently large control constraints, feasible controls exist. For estimates concerning the controls, we refer to [56].

(ii) If $u_d = 0$ and $0 \in Q_{ad}(0,1)$, then for any $\delta_0 > 0$, the control $q \equiv 0$ is feasible for $T > 0$ sufficiently large, since the semigroup generated by Δ is exponentially stable in $L^2(\Omega)$; see Proposition A.21.

(iii) Furthermore, Lemma 2.10 provides a sufficient condition for existence of feasible points, under the assumption $d_U(u_0)\omega_0 < h_0$ (which is clearly true for $\omega_0 = 0$ (since $h_0 > 0$) or the initial state u_0 sufficiently close to U). Note that Lemma 2.10 generalizes the argument of (ii), since $\omega_0 = 0$ in case of homogeneous Dirichlet conditions due to the Poincaré inequality. In Section 2.4 we will explicitly verify the suppositions of Lemma 2.10 for concrete terminal constraints U.

2.3.1. Strong stability

We now introduce the strong stability condition on the objective functional with respect to small perturbations of the terminal constraint set. This will allow for exact penalization of the constraints which in turn leads to optimality conditions in qualified form. For $\delta \geq 0$, define the perturbed control problem

$$\inf_{T>0,\, q \in Q_{ad}(0,T)} j(T,q) \quad \text{subject to} \quad u[q](T) \in U_\delta, \tag{P_δ}$$

where U is replaced with $U_\delta = U + \overline{\mathcal{B}_\delta(0)} = \{\, u \in H \colon d_U(u) \leq \delta \,\}$. Evidently, (P_0) is equal to (P). We define the corresponding value function $v \colon \mathbb{R}_+ \to \mathbb{R}_+$ by

$$v(\delta) = \inf (P_\delta).$$

Clearly, v is a monotonously decreasing function with $v(d_U(u_0)) = 0$.

Definition 2.16. The problem (P_δ) is called *strongly stable (on the right)* if there exist $\varepsilon > 0$ and $\eta_0 > 0$ such that

$$v(\delta) - v(\delta') \leq \eta_0(\delta' - \delta) \quad \text{for all } \delta' \in [\delta, \delta + \varepsilon]. \tag{2.18}$$

Remark 2.17. (i) In the case that $\delta > 0$, we can also define *stability on the left* in an analogous way; cf. also Figure 2.2. In this work, we only consider stability on the right, which is meaningful also for the important case $\delta = 0$.

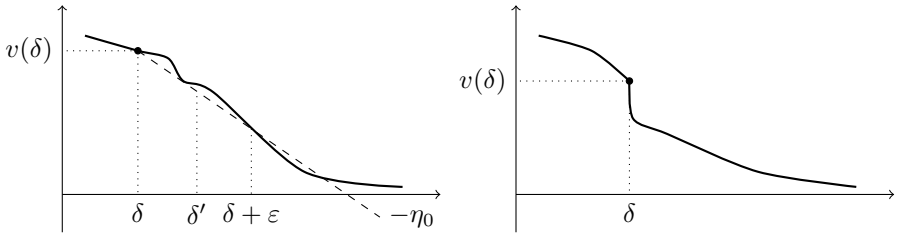

Figure 2.2.: Illustration of strong stability. The left example is strongly stable on the right at δ with radius ε and modulus η_0. The right example is not strongly stable on the right at δ. Both examples are strongly stable on the left at δ.

(ii) Strong stability is satisfied almost everywhere. Precisely, if (P) has feasible controls, then (P_δ) is strongly stable for all $\delta \in \mathbb{R}_+$ except on a set of Lebesgue measure zero; see, e.g., [19, Proposition 3.2]. This follows from monotonicity of v, because monotone functions are differentiable almost everywhere. However, since we consider the terminal set U to be a given datum, we are interested in conditions assuring strong stability on the right at $\delta = 0$.

(iii) Strong stability is also referred to as calmness, cf. [25], [136, Chapter 8.F], or weak calmness, cf. [21, Definition 3.114].

We now prove one of the main results of the chapter, which guarantees strong stability under a condition which is a direct strengthening of the necessary condition for weak invariance from Theorem 2.9. We require that there exists a $h_0 > 0$ such that

$$h(u,\zeta) \le -h_0\|\zeta\| \quad \text{for all } u \in U \cap V,\ \zeta \in N_U(u) \cap V. \tag{2.19}$$

Recall that weak invariance of (A, U, BQ_{ad}) corresponds to the same condition with $h_0 = 0$; see Theorem 2.9. In the case $h_0 > 0$, strong stability of (P_δ) holds for all small enough $\delta \ge 0$ (which includes the important case $\delta = 0$).

Theorem 2.18 (Strong stability)**.** *Let P_U be stable in V and suppose that condition* (2.19) *holds for some constant $h_0 > 0$. Then, for all $\delta \ge 0$ such that $\omega_0\delta < h_0/2$ the problem (P_δ) is strongly stable on the right with $\eta_0 \le c/h_0$, where the constant c only depends on the concrete choice of L and Q_{ad}.*

Proof. Fix $\varepsilon > 0$ such that $\omega_0(\delta+\varepsilon) \le h_0/2$. Then, let $\delta' \in [\delta, \delta+\varepsilon]$ be arbitrary and fix a solution (T', q', u') to $(P_{\delta'})$. Consider the auxiliary problem $\partial_t \breve{u} + A\breve{u} = B\breve{q}$ with initial condition $\breve{u}(0) = u'(T')$ and an auxiliary control $\breve{q}\colon [0,\infty) \to Q_{ad}$. Employing Lemma 2.10 we can choose $\breve{q}$ such that it holds

$$d_U(\breve{u}(t)) \le \max\{\,0,\ \delta' + (\delta'\omega_0 - h_0)t\,\} \quad \text{for } t \ge 0,$$

considering that $d_U(\breve{u}(0)) = d_U(u'(T')) = \delta'$. Clearly, it follows that $d_U(\breve{u}(\delta T)) \le \delta$ for the choice $\delta T = (\delta' - \delta)/(h_0 - \delta'\omega_0)$. Thus, $q \in Q_{ad}(0, T' + \delta T)$ defined by

$$q(t) = \begin{cases} q'(t) & \text{if } t \le T', \\ \breve{q}(t - T') & \text{if } t > T', \end{cases}$$

is admissible for (P_δ) and we find

$$\begin{aligned} v(\delta) = \inf(P_\delta) &\leq j(T' + \delta T, q) = j(T', q') + \int_{T'}^{T'+\delta T} \left[1 + L(\breve{q}(t - T'))\right] \mathrm{d}t \\ &\leq v(\delta') + \delta T \,(1 + L_\infty), \end{aligned}$$

where $L_\infty = \max_{q \in Q_{ad}} L(q)$. Using $\omega_0 \delta' \leq h_0/2$, we obtain that $\delta T \leq 2(\delta' - \delta)/h_0$, which results in (2.18) with a choice of $\eta_0 = 2(1 + L_\infty)/h_0$. This concludes the proof. □

2.3.2. Change of variable

In this subsection, we discuss the implications of strong stability on optimality conditions for (P_δ). To derive optimality conditions we first transform the time interval to the reference interval $(0, 1)$ (cf. Proposition 4.2 in [92], Proposition 4.1 in [134]). Consider the set of admissible scaling functions

$$N_{ad} \coloneqq \left\{ \nu \in L^\infty(0,1) \colon \operatorname*{ess\,inf}_{\tau \in (0,1)} \nu(\tau) > 0 \right\} = \{\nu \in L^\infty(0,1) \colon \nu \geq 0 \text{ and } 1/\nu \in L^\infty(0,1)\}$$

and define a family of transformations

$$T_\nu \colon [0,1] \to \mathbb{R}_+, \quad T_\nu(t) = \int_0^t \nu(\tau) \,\mathrm{d}\tau.$$

For $\nu \in N_{ad}$ and any mapping $u \colon (0,1) \to V$ we define the transformed elliptic operator

$$(\nu A u)(t) = \nu(t) A u(t),$$

and, by a change of variables, we obtain the transformed state equation

$$\partial_t u + \nu A u = \nu B q, \quad u(0) = u_0.$$

By standard results, for each right-hand side in $L^2((0,1); V^*)$ the transformed equation possesses a unique solution $u \in W(0,1)$ (see, e.g., [46, Theorem 2, Chapter XVIII, §3]). We introduce the control-to-state mapping as

$$S \colon N_{ad} \times Q_{ad}(0,1) \subset L^\infty(0,1) \times Q(0,1) \to W(0,1), \quad S(\nu, q) = u.$$

The transformed optimal control problem is then given by

$$\inf_{\nu \in N_{ad},\, q \in Q_{ad}(0,1)} j(\nu, q) \quad \text{subject to} \quad i_1 S(\nu, q) \in U, \tag{$\hat{P}$}$$

where the objective function is defined as

$$j(\nu, q) \coloneqq \int_0^1 \nu(t) \,(1 + L(q(t))) \,\mathrm{d}t.$$

Since no ambiguity arises, we do not rename variables. The definition of the set of admissible controls Q_{ad} transfers to the transformed problem, because the control constraints do not depend on time. In fact, both problems $(\hat{P})$ and (P) are equivalent in the following sense.

Proposition 2.19. *If (ν, q) is admissible for $(\hat{P})$ and $u = S(\nu, q)$, then*

$$(T_\nu(1), q \circ T_\nu, u \circ T_\nu)$$

is admissible for (P) and $j(\nu, q \circ T_\nu) = j(T, q)$. If (T, q, u) is admissible for (P), then for every $\nu \in N_{ad}$ such that $T_\nu(1) = T$,

$$(\nu, q \circ T_\nu^{-1})$$

is admissible for $(\hat{P})$ and $j(\nu, q \circ T_\nu^{-1}) = j(T, q)$.

Considering ν as an additional control variable, we obtain by standard arguments the following differentiability result.

Proposition 2.20. *The control-to-state mapping S is (infinitely often) continuously Fréchet-differentiable. In particular, $\delta u = S'(\nu, q)(\delta\nu, \delta q) \in W(0,1)$ is the unique solution to*

$$\partial_t \delta u + \nu A \delta u = \delta\nu(Bq - Au) + \nu B \delta q, \quad \delta u(0) = 0,$$

for $(\delta\nu, \delta q) \in L^\infty(0,1) \times Q(0,1)$. Moreover, $\delta\tilde{u} = S''(\nu, q)(\delta\nu_1, \delta q_1; \delta\nu_2, \delta q_2) \in W(0,1)$ is the unique solution to

$$\partial_t \delta\tilde{u} + \nu A \delta\tilde{u} = \delta\nu_1 \left(B\delta q_2 - A\delta u_2\right) + \delta\nu_2 \left(B\delta q_1 - A\delta u_1\right), \quad \delta\tilde{u}(0) = 0,$$

for $(\delta\nu_i, \delta q_i) \in L^\infty(0,1) \times L^2(I \times \omega)$ and $\delta u_i = S'(\nu, q)(\delta\nu_i, \delta q_i)$, $i = 1, 2$.

By the previous result and the continuity of the trace mapping i_1, the parameter-to-observation mapping $i_1 S(\nu, q)\colon (\nu, q) \mapsto u(1)$ is differentiable. Furthermore, for any fixed $\mu \in H$, the gradient of the functional $(\nu, q) \mapsto (i_1 S(\nu, q), \mu)$, which is given by the expression $(i_1 S'(\nu, q))^* \mu$, can be characterized by an adjoint equation.

Proposition 2.21. *Let $\nu \in N_{ad}$ and $q \in Q(0,1)$. For any $\mu \in H$ we have*

$$\left(i_1 S'(\nu, q)\right)^* \mu = \begin{pmatrix} \langle Bq - Au, z\rangle \\ \nu B^* z \end{pmatrix} \in L^1(0,1) \times L^2((0,1); Q),$$

where $z \in W(0,1)$ is the unique solution to the adjoint equation

$$-\partial_t z + \nu A^* z = 0, \quad z(1) = \mu,$$

where A^* denotes the adjoint operator of A.

Proof. Using Proposition 2.20, integration by parts, and the definition of z we observe

$$\begin{aligned}(\mu, i_1 S'(\nu, q)(\delta\nu, \delta q)) = (\delta u(1), \mu) = (\delta u(1), z(1)) - (\delta u(0), z(0)) = \int_0^1 \langle \partial_t \delta u, z\rangle + \int_0^1 \langle \partial_t z, \delta u\rangle \\ = \int_0^1 \langle \partial_t \delta u, z\rangle + \int_0^1 \langle \nu A \delta u, z\rangle = \int_0^1 \langle \delta\nu(Bq - Au) + \nu B\delta q, z\rangle,\end{aligned}$$

where $\delta u = S'(\nu, q)(\delta\nu, \delta q)$. Furthermore, we identify the partial derivative with respect to ν, i.e. $\delta\nu \mapsto \int_0^1 \delta\nu \langle Bq - Au, z\rangle$, with the function $\langle Bq - Au, z\rangle \in L^1(0,1)$. □

The transformed perturbed problems $(\hat{P}_\delta)$ for $\delta \geq 0$ are defined analogously:

$$\inf_{\nu \in N_{ad},\, q \in Q_{ad}(0,1)} j(\nu, q) \quad \text{subject to} \quad i_1 S(\nu, q) \in U_\delta. \tag{$\hat{P}_\delta$}$$

The notion of strong stability for $(\hat{P}_\delta)$ and (P_δ) are obviously equivalent, since the value function v is identical. We will derive optimality conditions by adding the terminal constraint as a penalty term to the objective functional. Under a strong stability assumption the resulting functional is exact.

Definition 2.22. Let $\delta \geq 0$ and (ν, q) be a local minimum of (P_δ). The functional

$$j_\eta(\cdot) = j(\cdot) + \eta\, d_{U_\delta}(i_1 S(\cdot))$$

is called an *exact penalty function* for (P_δ) at (ν, q), if there is $\eta \geq 0$ such that (ν, q) is a local minimizer of j_η.

Proposition 2.23. *Let $\delta \geq 0$ and $(\bar{\nu}, \bar{q})$ be a solution to (P_δ) and let (P_δ) be strongly stable on the right with constant $\eta_0 > 0$. Then, j_η is an exact penalty function for (P_δ) at $(\bar{\nu}, \bar{q})$ for any $\eta \geq \eta_0$.*

Proof. We give a proof of this well-known result for convenience of the reader: Let $\eta \geq \eta_0$ and (ν, q) be a local minimizer of j_η in a suitable small neighborhood of $(\bar{\nu}, \bar{q})$ (such that $d_{U_\delta}(i_1 S(\nu, q)) \leq \varepsilon$), and set $\delta' = d_{U_\delta}(i_1 S(\nu, q))$. Due to feasibility of $(\bar{\nu}, \bar{q})$ for (P_δ) and strong stability on the right, we obtain

$$\begin{aligned} j_\eta(\bar{\nu}, \bar{q}) = j(\bar{\nu}, \bar{q}) &\leq \inf(P_{\delta'}) + \eta(\delta' - \delta) \leq j(\nu, q) + \eta(\delta' - \delta) \\ &= j(\nu, q) + \eta\, d_{U_\delta}(i_1 S(\nu, q)) = j_\eta(\nu, q), \end{aligned}$$

where we have used optimality of (ν, q) for j_η in the last step. Whence, $(\bar{\nu}, \bar{q})$ is a local minimizer for j_η. □

Remark 2.24. The constraint in $(\hat{P})$ can be written as $g(\nu, q) = i_1 S(\nu, q) \in U_\delta$ and g is differentiable. If a constraint qualification such as Robinson's CQ holds,

$$0 \in \operatorname{int} \left\{\, g(\bar{\nu}, \bar{q}) + g'(\bar{\nu}, \bar{q})(N_{ad} - \bar{\nu}, Q_{ad}(0,1) - \bar{q}) - U_\delta \,\right\} \subset H,$$

then j_η is an exact penalty function for (P_δ); see, e.g., [21, Theorem 2.87, Proposition 3.111]. This presents an alternative approach to obtain qualified optimality conditions. We expect that the sufficient conditions from Section 2.3.4 are related to Robinson's CQ, but are unable to prove this in the general setting.

2.3.3. Optimality conditions

We define for any $\mu_0 \in \mathbb{R}_+$ the *Hamiltonian* $H_{\mu_0} \colon Q \times V \times V \to \mathbb{R}$ by

$$H_{\mu_0}(q, u, z) = \langle Bq - Au, z \rangle + \mu_0 \left[1 + L(q)\right].$$

Based on strong stability, qualified optimality conditions can be derived.

Theorem 2.25. *Let $\delta \geq 0$ and (P_δ) be strongly stable on the right (with constant $\eta > 0$). If $(\bar\nu, \bar q)$ is a solution of (P_δ) with $\bar u = S(\bar\nu, \bar q)$, then there exist $\bar\mu \in N_{U_\delta}(\bar u(1))$, $\bar\mu \neq 0$, $\|\bar\mu\| \leq \eta$, and a corresponding adjoint state $\bar z \in W(0,1)$ with*

$$-\partial_t \bar z + \bar\nu A^* \bar z = 0, \quad \bar z(1) = \bar\mu, \tag{2.20}$$

such that

$$\min_{q \in Q_{ad}} H_1(q, \bar u(t), \bar z(t)) = H_1(\bar q(t), \bar u(t), \bar z(t)) = 0, \qquad \text{a.e. } t \in (0,1). \tag{2.21}$$

The first equality in (2.21) can be equivalently expressed by

$$0 \in \partial L(\bar q(t)) + B^* \bar z(t) + N_{Q_{ad}}(\bar q(t)), \qquad \text{a.e. } t \in (0,1), \tag{2.22}$$

where ∂L denotes the convex subdifferential of L.

Proof. The proof is based on the minimization of the exact penalty function. Using Proposition 2.23, $(\bar\nu, \bar q)$ also is a minimizer of the penalty function j_η. Since $\bar\nu \in N_{ad}$, which is open, we may restrict the minimization to some neighborhood and neglect the constraints on ν in the following. We note that $j_\eta \colon L^\infty(0,1) \times Q(0,1) \to \mathbb{R}$ is locally Lipschitz continuous and derive the stationary conditions by Fermat's rule; see [40, Proposition 10.36]. We obtain

$$\begin{aligned} 0 &\in \partial_{\mathrm C} j_\eta(\bar\nu, \bar q) + N_{L^\infty(0,1) \times Q_{ad}(0,1)}(\bar\nu, \bar q) \\ &\subseteq \partial_{\mathrm C} j(\bar\nu, \bar q) + \eta\, \partial_{\mathrm C}\big[d_{U_\delta}(i_1 S(\bar\nu, \bar q))\big] + \{\,0\,\} \times N_{Q_{ad}(0,1)}(\bar q), \end{aligned} \tag{2.23}$$

where $\partial_{\mathrm C}$ denotes the generalized subdifferential due to Clarke; see, e.g., [40, Chapter 10]. Using Proposition A.24 and [40, Theorem 10.8] we find

$$\partial_{\mathrm C} j(\bar\nu, \bar q) \subseteq \{1 + L(\bar q)\} \times \bar\nu\, \partial_{\mathrm C} L(\bar q) = \{1 + L(\bar q)\} \times \bar\nu\, \partial L(\bar q),$$

because j is continuously differentiable with respect to ν and convex and Lipschitz continuous with respect to q due to the corresponding assumptions on L. Concerning the second term, we employ the chain rule [40, Theorem 10.19] and obtain

$$\partial_{\mathrm C}\left[d_{U_\delta}(i_1 S(\bar\nu, \bar q))\right] \subseteq (i_1 S'(\bar\nu, \bar q))^* \left[\partial_{\mathrm C} d_{U_\delta}(i_1 S(\bar\nu, \bar q))\right]. \tag{2.24}$$

The gradient $(i_1 S'(\bar\nu, \bar q))^*$ was computed in Proposition 2.21. Furthermore, the set $\partial_{\mathrm C} d_{U_\delta}(\cdot)$ can be identified with the ordinary convex subdifferential (see [40, Theorem 10.8]) and

$$\partial_{\mathrm C} d_{U_\delta}(v) = \partial d_{U_\delta}(v) = \{\,\mu \in N_{U_\delta}(v) \colon \|\mu\| \leq 1\,\},$$

for all $v \in U_\delta$; see, e.g., [12, Proposition 18.22]. Therefore, from (2.23) and (2.24) we obtain that there exists a $\bar\mu \in N_{U_\delta}(\bar u(1))$ with $\|\bar\mu\| \leq \eta$, a $\bar\xi \in \partial L(\bar q)$, and a $\bar\zeta \in N_{Q_{ad}(0,1)}(\bar q)$, such that

$$0 = \begin{pmatrix} 1 + L(\bar q) + \langle B\bar q - A\bar u, \bar z\rangle \\ \bar\nu(\bar\xi + B^*\bar z + \bar\zeta) \end{pmatrix},$$

where $\bar z$ solves the corresponding adjoint equation (2.20). The first component of this equation is the second equality in (2.21). Pointwise inspection of the second component for $t \in (0,1)$ and $\bar\nu(t) > 0$ implies (2.22). Now, we observe that (2.22) is the necessary and sufficient optimality condition for $\bar q(t)$ to be the solution of a convex optimization problem, namely

$$\bar q(t) = \operatorname*{argmin}_{q \in Q_{ad}} \left[L(q) + \langle Bq, \bar z(t)\rangle\right] = \operatorname*{argmin}_{q \in Q_{ad}} H_1(q, \bar u(t), \bar z(t)).$$

Finally, assume that $\bar\mu = 0$. This implies $\bar z = 0$ by unique solvability of the adjoint equation. Using the Hamiltonian condition (2.21) we infer $1 + L(\bar q) = 0$ almost everywhere in $(0,1)$. This contradicts $L \geq 0$, and we conclude $\bar\mu \neq 0$. □

Without strong stability, under a structural assumption on only the constraint set, the generalized form of the optimality conditions can be derived. To this end, we first introduce the concept of finite co-dimension; see [104, Definition 4.1.5]. Let X be a Banach space. A subset S of X is said to be of *finite co-dimension* in X, if there exists a $s_0 \in \overline{\operatorname{co}}\, S$ (the *convex hull* of S) such that $\overline{\operatorname{span}\{\, S - s_0 \,\}}$ is a finite co-dimensional subspace of X and $\overline{\operatorname{co}}\,\{\, S - s_0 \,\}$ has a nonempty interior in this subspace.

Note that if $\delta > 0$, then U_δ has finite co-dimension in H, because of $\overline{\operatorname{span}\{\, U_\delta - u' \,\}} = H$ and the fact that $\overline{\operatorname{co}}\,\{\, U_\delta - u' \,\} = U - u' + \mathcal{B}_\delta(0)$ has a non-empty interior for all $u' \in U$. In contrast, a point constraint $U = \{\, u_d \,\}$ is not of finite co-dimension, since $\overline{\operatorname{span}\{\, U - u_d \,\}} = \{\, 0 \,\}$ has an infinite co-dimension.

Theorem 2.26. *Assume that U_δ is of finite co-dimension (or $\delta > 0$); see [104, Definition 4.1.5]. Let $(\bar\nu, \bar q)$ be a solution of (P_δ), $\bar u = S(\bar\nu, \bar q)$. Then there exist $\bar\mu \in N_{U_\delta}(\bar u(1))$, $\bar\mu \neq 0$, $\bar\mu_0 \in \{\, 0, 1 \,\}$ and a corresponding adjoint state $\bar z \in W(0,1)$ which fulfills* (2.20)*, such that*

$$\min_{q \in Q_{ad}} H_{\bar\mu_0}(q, \bar u(t), \bar z(t)) = H_{\bar\mu_0}(\bar q(t), \bar u(t), \bar z(t)) = 0, \qquad \text{a.e. } t \in (0,1). \tag{2.25}$$

Proof. We only give a short outline of the proof. It combines the one of [40, Theorem 10.47] with the one of [134, Theorem 4.1]. As before, since N_{ad} is open, we may restrict the minimization to some neighborhood and neglect the constraints on ν in the following. Define the function

$$\phi^\varepsilon(\nu, q) = \sqrt{\max\{\, 0, j(\nu, q) - j(\bar\nu, \bar q) + \varepsilon \,\}^2 + d_U(i_1 S(\nu, q))^2}.$$

Ekeland's variational principle with $\lambda = \sqrt{\varepsilon}$ yields a sequence $\nu_\varepsilon \in N_{ad}$, $q_\varepsilon \in Q_{ad}(0,1)$ such that $(\nu_\varepsilon, q_\varepsilon) \to (\bar\nu, \bar q)$ for $\varepsilon \to 0$ and the function

$$\Phi^\varepsilon(\nu, q) = \phi^\varepsilon(\nu, q) + \sqrt{\varepsilon}\|\nu - \nu_\varepsilon\| + \sqrt{\varepsilon}\|q - q_\varepsilon\|$$

attains a strict (local) minimum at $(\nu_\varepsilon, q_\varepsilon)$ over $L^\infty(0,1) \times Q_{ad}(0,1)$; see, e.g., [40, Theorem 5.19]. The Lipschitz constant of Φ^ε can be bounded independently of ε, if $0 < \varepsilon \leq \varepsilon_0$ for some fixed $\varepsilon_0 > 0$. Employing [40, Theorem 10.31] there exists a constant K solely depending on the Lipschitz constant of Φ^ε, such that the mapping

$$(\nu, q) \mapsto \Phi^\varepsilon(\nu, q) + K d_{Q_{ad}(0,1)}(q)$$

has a local minimum at $(\nu_\varepsilon, q_\varepsilon)$. Nonsmooth calculus as in Theorem 2.25 yields

$$\gamma_\varepsilon \in \partial_{\mathrm{C}} \phi^\varepsilon(\nu_\varepsilon, q_\varepsilon) + \{\, 0 \,\} \times \Big(N_{Q_{ad}(0,1)}(q_\varepsilon) \cap \mathcal{B}_K(0) \Big) \tag{2.26}$$

with $\gamma_\varepsilon \to 0$ in $L^\infty(0,1)^* \times Q(0,1)$ as $\varepsilon \to 0$.

Now, we define $\lambda_\varepsilon \in \mathbb{R}^2_+$ by

$$\begin{aligned} \lambda_{\varepsilon,1} &= \max\{\, 0, j(\nu_\varepsilon, q_\varepsilon) - j(\bar\nu, \bar q) + \varepsilon \,\} / \phi^\varepsilon(\nu_\varepsilon, q_\varepsilon), \\ \lambda_{\varepsilon,2} &= d_{U_\delta}(i_1 S(\nu_\varepsilon, q_\varepsilon)) / \phi^\varepsilon(\nu_\varepsilon, q_\varepsilon). \end{aligned}$$

Clearly, it holds $\lambda_{\varepsilon,1}^2 + \lambda_{\varepsilon,2}^2 = 1$. By computing the subdifferential $\partial_{\mathrm{C}} \phi^\varepsilon$ (combining the arguments of [40, Theorem 10.47] and Theorem 2.25), we obtain sequences of $\mu_\varepsilon \in N_{U_\delta}(u_\varepsilon(1))$ with $\|\mu_\varepsilon\| \leq 1$, $\xi_\varepsilon \in \partial L(q_\varepsilon)$, $\zeta_\varepsilon \in N_{Q_{ad}(0,1)}(q_\varepsilon)$, and $\|\zeta_\varepsilon\| \leq K$ such that

$$\gamma_\varepsilon = \begin{pmatrix} \lambda_{\varepsilon,1}\,[1 + L(q_\varepsilon)] + \lambda_{\varepsilon,2} \langle B q_\varepsilon - A u_\varepsilon, z_\varepsilon \rangle \\ \nu_\varepsilon(\lambda_{\varepsilon,1} \xi_\varepsilon + \lambda_{\varepsilon,2} B^* z_\varepsilon + \zeta_\varepsilon) \end{pmatrix}, \tag{2.27}$$

where z_ε solves the corresponding adjoint equation (2.20) with terminal value μ_ε. Now, we go to the limit. Due to boundedness of the sequence $(\mu, \xi, \zeta, \lambda)_\varepsilon \in H \times Q(0,1) \times Q(0,1) \times \mathbb{R}^2$, we can go to a weak limit on a subsequence $(\mu, \xi, \zeta, \lambda)_n \rightharpoonup (\hat\mu, \hat\xi, \hat\zeta, \hat\lambda)$ for $n \to \infty$. Moreover, by combining the general result from [40, Proposition 10.10] with the continuity of the solution mapping S we can go to the limit in the inclusion (2.26) and obtain $\hat\mu \in N_{U_\delta}(\bar u(1))$ with $\|\hat\mu\| \le 1$, $\hat\xi \in \partial L(\bar q)$, $\hat\zeta \in N_{Q_{ad}(0,1)}(\bar q)$, and $\hat\lambda \in \mathbb{R}^2_+$, $\hat\lambda_1^2 + \hat\lambda_2^2 = 1$.

Now, we distinguish two cases: In the case $\hat\lambda_1 > 0$, we set $(\bar\mu, \bar\xi, \bar\zeta) = (\hat\lambda_2\hat\mu, \hat\xi, \hat\zeta)/\hat\lambda_1$, and we can derive the conditions for $\mu_0 = 1$ as in Theorem 2.25. As before, the nontriviality of $\bar\mu$ follows. Note that the case $\hat\lambda_2 = 0$ cannot occur, since from the first equation of (2.27) we would deduce $0 = 1 + L(\bar q)$.

In case $\hat\lambda_1 = 0$, it follows $\hat\lambda_2 = 1$, and we obtain the desired set of conditions with $(\bar\mu, \bar\xi, \bar\zeta) = (\hat\mu, \hat\xi, \hat\zeta)$. It remains to verify $\bar\mu \neq 0$. Since $\hat\lambda_{n,2} \to 1$, we obtain $u_n(1) = i_1 S(\nu_n, q_n) \notin U_\delta$ and $\mu_n = (u_n(1) - P_{U_\delta}(u_n(1)))/d_{U_\delta}(u_n(1))$, i.e., $\|\mu_n\| = 1$, for n sufficiently large. Moreover, as $\mu_n \in N_{U_\delta}(u_n(1))$ we find for all $u' \in U_\delta$ that

$$(\mu_n, u' - \bar u(1)) \le (\mu_n, u_n(1) - \bar u(1)) \le \|\mu_n\| \|\bar u(1) - u_n(1)\| \to 0.$$

Finally, we use the fact that U_δ has finite co-dimension with [104, Lemma 4.3.6] to conclude that $0 \neq \bar\mu = \hat\mu = \operatorname{weak}\lim_{n\to\infty} \mu_n$. □

Remark 2.27. As an example, consider the choice $L(q) = (\alpha/2)\|q\|_Q^2$ for $\alpha \ge 0$. In the qualified case, condition (2.22) reduces to the variational inequality

$$(\alpha\bar q(t) + B^*\bar z(t), q - \bar q(t)) \ge 0 \quad \text{for all } q \in Q_{ad},$$

which implies the projection formula $\bar q(t) = P_{Q_{ad}}(-(1/\alpha)B^*\bar z(t))$ for almost all $t \in (0,1)$. In contrast, in the unqualified case $\bar\mu_0 = 0$ the condition (2.25) is *independent* of the cost parameter α, and we obtain that

$$(B^*\bar z(t), q - \bar q(t)) \ge 0 \quad \text{for all } q \in Q_{ad}.$$

In this case, an unqualified stationary point for any $\alpha > 0$ corresponds to a stationary point for the pure time-optimal problem with $\alpha = 0$. Moreover, if $B^*\bar z(t) \neq 0$ for almost every $t \in (0,1)$, the control always assumes an extreme value in Q_{ad}, i.e., it is bang-bang.

2.3.4. The Hamiltonian condition and qualified optimality conditions

In this subsection we investigate connections between the strengthened Hamiltonian condition and qualified optimality conditions. We fist give the main result, which is a direct consequence of the previous results.

Corollary 2.28. *Let P_U be stable in V and suppose that the Hamiltonian condition* (2.19) *holds for some constant $h_0 > 0$. Then, the optimality conditions* (2.1)–(2.3) *hold for any optimal solution of (P) in the qualified form (with $\bar\mu_0 = 1$), and additionally $\|\bar\mu\| \le c/h_0$.*

Proof. This is a consequence of Theorem 2.18, Theorem 2.25, and the equivalence of the transformed problem $(\hat P)$ and the original problem (P). □

The Hamiltonian condition (2.19) is required to hold for all $u \in U \cap V$. Certainly, only elements of $\partial U \cap V$ are relevant; the condition is trivially fulfilled otherwise. However, if the terminal value $\bar{u}(T) \in \partial U \cap V$ of the optimal solutions to (P_δ) is assumed to be known, it appears desirable to weaken (2.19) to a local condition. In fact, at least in case of finite co-dimension of U and regular normal cones, it is sufficient to require the strengthened Hamiltonian condition only at the optimal terminal value $\bar{u}(T)$ to obtain qualified optimality conditions. We give an auxiliary lemma before the result.

Lemma 2.29. *The lower Hamiltonian $h\colon V \times V \to \mathbb{R}$ is continuous.*

Proof. We introduce the support function of Q_{ad} as $h_{Q_{ad}}(\cdot) = \sup_{q \in Q_{ad}} (q, \cdot)_Q$. Then it holds

$$h(u, \zeta) = -h_{Q_{ad}}(-B^*\zeta) - \langle Au, \zeta \rangle.$$

Employing the facts that support functions are convex and that $h_{Q_{ad}}$ is finite ($h_{Q_{ad}}(\zeta) \leq C_{Q_{ad}} \|B\|_{\mathcal{L}(Q,V^*)} \|\zeta\|_V$ for all $\zeta \in Q$), we infer that $h\colon V \times V \to \mathbb{R}$ is continuous, since convex functions are locally Lipschitz continuous; see, e.g., [40, Theorem 2.34]. □

Proposition 2.30. *Suppose that U has finite co-dimension and an optimal solution $(\bar{q}, T, \bar{u})$ of (P) is given with $N_U(\bar{u}(T)) \subset V$ and*

$$h(\bar{u}(T), \zeta) \leq -h_0 \|\zeta\| \quad \textit{for all } \zeta \in N_U(\bar{u}(T)), \tag{2.28}$$

for some constant $h_0 > 0$. Then, the optimality conditions (2.20)–(2.2) hold in the qualified form (with $\bar{\mu}_0 = 1$), and additionally $\|\bar{\mu}\| \leq c/h_0$.

Proof. We argue by contradiction. Let the conditions of Theorem 2.26 hold with $\bar{\mu}_0 = 0$. Then, $\bar{u} \in C((0,T]; V)$, $\bar{z} \in C([0,T]; V)$ according to Proposition A.18, and

$$h(\bar{u}(t), \bar{z}(t)) = \min_{q \in Q_{ad}} \langle Bq - A\bar{u}(t), \bar{z}(t) \rangle = \langle B\bar{q}(t) - A\bar{u}(t), \bar{z}(t) \rangle = 0$$

for almost all $t \in (0,T)$. However, since $t \mapsto h(\bar{u}(t), \bar{z}(t))$ is continuous on $(0,T]$ due to Lemma 2.29, this leads to a contradiction, because $h(\bar{u}(T), \bar{z}(T)) = h(\bar{u}(T), \bar{\mu}) \leq -h_0 \|\bar{\mu}\| < 0$. Thus, $\bar{\mu}_0 = 1$, and inspection of the Hamiltonian optimality condition yields

$$\begin{aligned} -h_0 \|\bar{\mu}\| &\geq h(\bar{u}(T), \bar{z}(T)) = \min_{q \in Q_{ad}} \left[H_1(q, \bar{u}(T), \bar{z}(T)) - (1 + L(q)) \right] \\ &\geq \min_{q \in Q_{ad}} H_1(q, \bar{u}(T), \bar{z}(T)) + \min_{q \in Q_{ad}} -(1 + L(q)) = -(1 + \max_{q \in Q_{ad}} L(q)) = -L_\infty, \end{aligned}$$

which implies the estimate for $\bar{\mu}$. □

Clearly, (2.28) is a weaker assumption than (2.19) (given the requirements on the terminal set U and the normal cone). Additionally, if $N_U(\bar{u}(T))$ contains just one direction, condition (2.28) is already equivalent to the qualified optimality conditions.

Proposition 2.31. *Let the qualified optimality conditions (as in Corollary 2.28) hold and assume that the normal cone $N_U(\bar{u}(T)) \subset V$ has dimension one. Then, the condition (2.28) holds with $h_0 = \|\bar{\mu}\|^{-1}$.*

Proof. First, we note that $N_U(\bar{u}(T)) = \{\, \lambda\bar{\mu} \colon \lambda \geq 0 \,\}$, since $0 \neq \bar{\mu} \in N_U(\bar{u}(T))$, and thus also $\bar{\mu} \in V$. Condition (2.21) implies

$$0 = \min_{q \in Q_{ad}} H_1(q, \bar{u}(t), \bar{z}(t)) \geq \min_{q \in Q_{ad}} \langle Bq - A\bar{u}(t), \bar{z}(t) \rangle + 1 + \min_{q \in Q_{ad}} L(q)$$

and, since $L(q) \geq 0$, we obtain

$$h(\bar{u}(t), \bar{z}(t)) = \min_{q \in Q_{ad}} \langle Bq - A\bar{u}(t), \bar{z}(t) \rangle \leq -1, \quad \text{a.e. } t \in [0, T]. \tag{2.29}$$

Recall that the lower Hamiltonian $h \colon V \times V \to \mathbb{R}$ is continuous; see Lemma 2.29. Moreover, according to Proposition A.18 with $\bar{z}(T) = \bar{\mu} \in V$ we find that $u \in C((0,T]; V)$ and $z \in C([0,T]; V)$. Thus, we can evaluate the expression (2.29) at $t = T$ and arrive at

$$h(\bar{u}(T), \bar{\mu}) = \min_{q \in Q_{ad}} \langle Bq - A\bar{u}(T), \bar{\mu} \rangle \leq -1.$$

Let $\zeta \in N_U(\bar{u}(T))$ as in (2.28). Multiplying both sides by $\lambda = \|\bar{\mu}\|^{-1}\|\zeta\| \geq 0$ in the inequality above and using the positive homogeneity of the terms on the left and right finishes the proof. □

2.3.5. Further perturbation results

Up to this point, we have studied the sensitivity of the objective functional with respect to perturbations of the terminal constraint. In this subsection, as another consequence of the theory of Section 2.2, we study perturbations with respect to the initial state u_0 (cf. [26, 62]) and the operator A (cf. [149, 162]) of problem (P). In particular, we restrict attention to the classical case $L \equiv 0$. In view of the fact that the choice $L \equiv 0$ results in $j(T, q) = T$, an estimate for the optimal value function corresponds to a perturbation estimate for the optimal time T, which is of independent interest. In the following, we introduce a perturbation parameter $\varepsilon > 0$ (to be made concrete later) and derive estimates for $T - T_\varepsilon$, where $T = T_0$ and T_ε denote the optimal times for the original and the perturbed problem, respectively. Moreover, $c > 0$ is a generic constant that may have different values at different appearances.

Perturbations of the initial state u_0

For $T > 0$, we use $u[q, u_0]$ to denote the solution to the state equation with control $q \in Q(0,T)$ and initial state $u_0 \in H$. Consider the time-optimal control problems with perturbed initial values $u_0^\varepsilon \in H$ defined as

$$\inf_{T>0,\, q \in Q_{ad}(0,T)} T \quad \text{subject to } u[q, u_0^\varepsilon](T) \in U. \tag{2.30}$$

We suppose that the initial values converge to u_0 at a rate ε, i.e. there is $c > 0$ such that

$$\|u_0^\varepsilon - u_0\| \leq c\varepsilon, \quad \varepsilon > 0. \tag{2.31}$$

Using similar arguments as in the proof of Theorem 2.18 we obtain the following perturbation result.

Theorem 2.32. *Suppose that the projection P_U is stable in V, the strengthened Hamiltonian condition* (2.11) *holds, and the perturbed initial condition fufills* (2.31)*. Then, there exists an $\varepsilon_0 > 0$ such that problem* (2.30) *has solutions for $\varepsilon \le \varepsilon_0$. Moreover, it holds*

$$|T - T_\varepsilon| \le c\varepsilon, \quad 0 < \varepsilon \le \varepsilon_0,$$

where T is the optimal time to (P) and T_ε is the optimal time to (2.30)*.*

Proof. Let $(T, \bar{q})$ be an optimal solution of (P). Since the semigroup e^{-tA} is strongly continuous, for all $T' > 0$ there is $c > 0$ such that $\|\mathrm{e}^{-tA}\|_{\mathcal{L}(H)} \le c$ for all $t \in [0, T']$. Thus, setting $\breve{u}_T = u[\bar{q}, u_0^\varepsilon](T)$ we find $c > 0$ such that

$$d_U(\breve{u}_T) \le \|u[\bar{q}, u_0^\varepsilon](T) - u[q, u_0](T)\| = \|\mathrm{e}^{-TA}(u_0^\varepsilon - u_0)\| \le c\varepsilon,$$

because $u[\bar{q}, u_0](T) \in U$. For $\varepsilon > 0$ sufficiently small, we may apply Lemma 2.10 to obtain a control $\breve{q}\colon [0, \infty) \to Q_{ad}$ such that the corresponding trajectory with initial value $\breve{u}_T = u[\bar{q}, u_0^\varepsilon](T)$ satisfies

$$d_U(u[\breve{q}, \breve{u}_T](t)) \le \max\{\, 0, d_U(u_T) + (d_U(\breve{u}_T)\omega_0 - h_0)\, t \,\} \le \max\{\, 0, c\varepsilon + (c\varepsilon\omega_0 - h_0)\, t \,\}$$

for all $t \ge 0$. Setting $\delta T = c\varepsilon/(h_0 - c\varepsilon\omega_0)$ and

$$q'(t) = \begin{cases} \bar{q}(t) & \text{if } t \le T, \\ \breve{q}(t - T) & \text{if } t > T, \end{cases}$$

the pair $(T + \delta T, q')$ is feasible for (2.30). This implies that there exists an optimal solution $(T_\varepsilon, \bar{q}_\varepsilon)$ of (2.30). Furthermore, by optimality of T_ε, we obtain

$$T_\varepsilon \le T + \delta T = T + \frac{c\varepsilon}{h_0 - c\varepsilon\omega_0} \le T + c\varepsilon.$$

In particular, this implies that T_ε is uniformly bounded. Hence, we can find an uniform estimate for $\|\mathrm{e}^{-\cdot A}\|_{\mathcal{L}(H)}$ on $[0, T_\varepsilon] \subseteq [0, T + c\varepsilon]$ and the same arguments as before (exchanging the roles of $(T, \bar{q})$ and $(T_\varepsilon, \bar{q}_\varepsilon)$) yield the estimate $T \le T_\varepsilon + c\varepsilon$. □

Note that the previous result is essentially a generalization of [26, Theorem 4.1], where a sufficient condition for the Hamiltonian condition in a specific setting is assumed to hold.

Perturbation of the operator A

Next, we consider perturbations of the operator A. Let $A_\varepsilon\colon V \to V^*$ be a family of linear operators such that for each $\varepsilon > 0$ the general assumptions from Section 2.1 are fulfilled and $A_0 = A$. Moreover, let $u_\varepsilon[q]$ denote the solution to the associated perturbed state equation for $q \in Q(0, T)$ and fixed $u_0 \in H$. We define the corresponding perturbed optimization problem as

$$\inf_{T>0,\ q \in Q_{ad}(0,T)} T \quad \text{subject to } u_\varepsilon[q](T) \in U. \tag{2.32}$$

Suppose that for every $T' > 0$ there exists $c > 0$ such that

$$\|u_\varepsilon[q](t) - u[q](t)\| \le c\varepsilon, \quad 0 \le t \le T', \quad q \in Q_{ad}(0, T'), \quad \varepsilon > 0. \tag{2.33}$$

Moreover, suppose that P_U is stable in V and the strengthened Hamiltonian condition (2.11) holds uniformly with respect to ε, i.e. there exists $h_0 > 0$ such that

$$h_\varepsilon(u,\zeta) := \min_{q\in Q_{ad}} \langle Bq - A_\varepsilon u, \zeta\rangle \le -h_0\|\zeta\|, \quad \text{where } u = P_U(v),\ \zeta = v - u, \tag{2.34}$$

for all $v \in V$ and all $\varepsilon > 0$ sufficiently small. These assumptions lead to the following error estimate.

Proposition 2.33. *Let* (2.33) *and* (2.34) *hold. Then, there exists a* $\varepsilon_0 > 0$ *such that problem* (2.32) *has solutions for* $\varepsilon \le \varepsilon_0$*. Moreover, it holds*

$$|T - T_\varepsilon| \le c\varepsilon, \quad 0 < \varepsilon \le \varepsilon_0,$$

where T *is the optimal time to* (P) *and* T_ε *is the optimal time to* (2.32)*.*

Proof. This result is shown along the lines of the proof of Theorem 2.32, where we use the supposition (2.33) instead of (2.31) as well as (2.34). □

We conclude with some comments on the assumptions of the preceding result. In particular, we show that they are always fulfilled for bounded perturbations of the operator. Concretely, assume that the perturbation is of the form:

$$A_\varepsilon = A + \delta A_\varepsilon, \quad \text{where} \quad \|\delta A_\varepsilon\|_{\mathcal{L}(H)} \le c\varepsilon. \tag{2.35}$$

We obtain the following result.

Theorem 2.34. *Let* U *be bounded in* H*. Suppose that* P_U *is stable in* V*, the strengthened Hamiltonian condition* (2.11) *holds, and the perturbed operator is of the form* (2.35)*. Then, the result of Proposition 2.33 holds true.*

Proof. We verify the conditions of Proposition 2.33: Concerning (2.34), we obtain for all $u \in U \cap V$ and $\zeta \in N_U(u) \cap V$ that

$$h_\varepsilon(u,\zeta) = \min_{q\in Q_{ad}} \langle Bq - A_\varepsilon u, \zeta\rangle = \min_{q\in Q_{ad}} \langle Bq - Au, \zeta\rangle - (\delta A_\varepsilon u, \zeta) \le -h_0\|\zeta\| + c\varepsilon\|u\|\|\zeta\|.$$

Thus, for $\varepsilon > 0$ sufficiently small, condition (2.34) holds uniformly in ε.

Concerning (2.33), consider $u^0 = u[q]$ and $u^\varepsilon = u_\varepsilon[q]$, and fix some arbitrary $T' > 0$. By straightforward calculations we verify that, for ε small enough, A_ε still satisfies the Gårding inequality (2.7) with slightly modified constants. Thus, by standard energy estimates we have the estimate $\|u^\varepsilon\|_{L^2((0,T');V)} \le c(\|u_0\| + \|q\|_{L^2((0,T');Q)})$ with a constant c independent of ε, u_0, and q, again for ε sufficiently small; see, e.g., [46, Chapter XVIII, §3]. Clearly, the perturbation $\delta u^\varepsilon = u^0 - u^\varepsilon$ solves

$$\partial_t \delta u^\varepsilon + A\,\delta u^\varepsilon = \delta A_\varepsilon\, u^\varepsilon, \quad \delta u_\varepsilon(0) = 0.$$

Hence, we obtain

$$\begin{aligned}\|\delta u^\varepsilon(t)\| \le c\|\delta A_\varepsilon u^\varepsilon\|_{L^2((0,T');V^*)} &\le c\varepsilon\|u^\varepsilon\|_{L^2((0,T');V)} \\ &\le c\varepsilon(\|u_0\| + \|q\|_{L^\infty((0,T');Q)}), \quad t \in [0,T'],\end{aligned}$$

with a constant c independent of u_0 and q. This shows (2.33). □

In Theorem 2.34, we have focused on the fundamental case of a bounded perturbation of the operator. Note that this includes the perturbation of a reaction diffusion equation in the lowest order term. In particular, this fully covers the setting considered in [162]. The uniform Hamiltonian condition (2.34) is automatically fulfilled there, since the perturbed operators are uniformly coercive ($\omega_0 = 0$), and the target set is a L^2-ball around zero; cf. Proposition 2.37.

Different scenarios are also of interest; see, e.g., [149]. Let us briefly comment on possible generalizations of Theorem 2.34. Clearly, for the verification of (2.33) it suffices that $\|\delta A_\epsilon\|_{\mathcal{L}(V,V^*)} \leq c\varepsilon$ (which is still more restrictive than [149], but allows for perturbations even in the main part of the operator). Additionally, we have to verify the uniform Hamiltonian condition (2.34). Even though it cannot simply be derived from the corresponding condition for $\varepsilon = 0$, as in the proof of Theorem 2.34, it can be done directly in concrete scenarios for the terminal set U. For instance, if the operators are uniformly coercive for small ε, the terminal set is the H ball around zero, and $0 \in Q_{ad}$, then (2.34) holds uniformly for any perturbation; cf. Proposition 2.37.

2.4. Applications

In this section we derive criteria for strong stability for different terminal sets U. It is organized as follows: First, we discuss the illustrative example $U = \{ u_d \}$ and observe that this leads to rather restrictive conditions. Significantly weaker conditions can be derived for the case of a H-ball around u_d if the operator A is coercive. In the general case, which includes unstable systems, we discuss a finite approximate controllability constraint that stabilizes the system around the zero point. The resulting conditions turn out to require at least as many controls as there are unstable modes. Finally, we only require a standard stabilizability assumption to hold, and show that there always exist target sets around zero such that the resulting optimization problem is strongly stable.

2.4.1. Point target and pointwise constraint

We first consider the example of steering the system in minimal time into a single point u_d, which has been extensively studied in the literature; see, e.g., [10, 54]. Defining U to be the singleton $U = \{ u_d \}$ with $u_d \in V$ we obtain the following result.

Proposition 2.35. *Suppose that $U = \{ u_d \}$ with $Au_d \in \operatorname{ran}(B)$ and for some $h_0 > 0$ it holds*

$$Au_d + \mathcal{B}_{h_0}(0) \subset BQ_{ad}. \tag{2.36}$$

Then (P) is strongly stable on the right for all $\delta \geq 0$.

Proof. Clearly, $P_U(u) = u_d$. Due to Proposition 2.8 it holds

$$N_U(u_d) = \{ \lambda(u' - u_d) \colon \lambda \geq 0,\, u' \in V \} = V.$$

We now take $u = u_d$ and $\zeta \in V$. Then

$$h(u, \zeta) = \min_{q \in Q_{ad}} \langle Bq - Au_d, \zeta \rangle \leq \min_{v \in Au_d + \mathcal{B}_{h_0}(0)} \langle v - Au_d, \zeta \rangle = \min_{v \in \mathcal{B}_{h_0}(0)} \langle v, \zeta \rangle = -h_0 \|\zeta\|.$$

Now, Theorem 2.18 yields the assertion. □

We point out that (2.36) is essentially the condition which is used in [10, Theorem 5.3.1] to guarantee existence of (qualified) multipliers in a similar setting; cf. also [26, Theorem 4.1] for Lipschitz continuity of the minimal time function with respect to the initial value. From an application point of view, it is rather restrictive. It is essentially only fulfilled in settings where $Q = H$, B is the identity, and Q_{ad} contains a sufficiently large H-ball. For settings with pointwise bounded control action ($BQ_{ad} \subset L^\infty(\Omega)$) for a domain $\Omega \subset \mathbb{R}^d$, controls restricted to some $\omega \subset \Omega$, or finite dimensional controls, it is not fulfilled. In this regard we also mention [160] for the pure time-optimal control (i.e. $L \equiv 0$) of the heat equation into zero with pointwise bounded controls active only on a subset of Ω. Therein, the authors obtain Lagrange multipliers in a larger space than $L^2(\Omega)$ (containing distributions) using essentially the exact null controllability of the heat equation.

Next, we turn to point-wise terminal constraints that are of independent interest in applications; cf. [97]. As an example, let $\Omega \subset \mathbb{R}^d$ be a bounded domain and assume $H = L^2(\Omega)$. We consider

$$U = \{\, u \in H \colon |u - u_d| \le u_{\max} \quad \text{a.e. in } \Omega \,\}, \tag{2.37}$$

where $u_d \in V$ and $u_{\max} \in \mathbb{R}$, $u_{\max} > 0$. For simplicity, we consider only an illustrative special case for A.

Proposition 2.36. *Let $A = -\nabla\kappa \cdot \nabla$ for a coefficient function $\kappa \in L^\infty(\Omega; \mathbb{R}^{d\times d})$ that is uniformly elliptic. Suppose that U is defined as in* (2.37) *with $Au_d \in \operatorname{ran}(B)$ and* (2.36) *holds for some $h_0 > 0$. Then (P) is strongly stable on the right for all $\delta \ge 0$.*

Proof. We will verify the supposition of Theorem 2.18. Clearly, it holds

$$P_U(v) = v - (v - u_d - u_{\max})^+ + (v + u_d - u_{\max})^-.$$

Due to Proposition 2.8 we infer

$$N_U(u) = \{\, (u' - u_d - u_{\max})^+ - (u' - u_d - u_{\max})^- \colon u' \in V,\, u = P_U(u') \,\}.$$

Take $u' \in V$ with $P_U(u') = u$ and set $\zeta = (u' - u_d - u_{\max})^+ - (u' - u_d - u_{\max})^-$. Then

$$\begin{aligned} h(u,\zeta) &= \min_{q \in Q_{ad}} \langle Bq - Au, \zeta\rangle \le \min_{v \in Au_d + \mathcal{B}_{h_0}(0)} \langle v, \zeta\rangle - \int_\Omega [\kappa \nabla P_U(u') \cdot \nabla\zeta] \\ &\le -h_0\|\zeta\| + \langle Au_d, \zeta\rangle - \int_{\{\, x\in\Omega \colon \zeta \neq 0\,\}} [\kappa \nabla u_d \cdot \nabla\zeta] = -h_0\|\zeta\|. \end{aligned}$$

Finally, Theorem 2.18 yields the assertion. □

Again we remark that (2.36) is rather restrictive. However, note that for pointwise constraints one typically searches for Lagrange multipliers in a space of regular Borel measures (cf., e.g., [134]), whereas under assumption (2.36), we obtain multipliers in $H = L^2(\Omega)$. A corresponding extension of the above theory to include multipliers in spaces of measures (under potentially weaker conditions) is outside of the scope of this chapter.

However, it seems that in applications it is often sufficient to steer the system close to a desired state u_d. In the subsequent subsections we will derive significantly weaker conditions guaranteeing strong stability for this type of terminal constraint.

2.4.2. H-norm constraint

Let $u_d \in V$ and $\delta_0 > 0$ be given and consider the set

$$U = \{\, u \in H \colon \|u - u_d\| \leq \delta_0 \,\}.$$

We emphasize that $u_d \in V$ (instead of just $u_d \in H$, $u_d \notin V$) is required for the minimizing projection P_U to be stable in V, which is necessary for weak invariance; see Lemma 2.6.

If the operator A is coercive (i.e. $\omega_0 = 0$) we can easily verify the strengthened Hamiltonian condition assuming only the existence of one control $\breve{q} \in Q_{ad}$ such that $B\breve{q}$ is sufficiently close to Au_d in V^*. This condition can be interpreted as the requirement that u_d lies sufficiently close to an asymptotically stable state of the system with fixed control $\breve{q}$. Note that this always holds for sufficiently small $u_d \in V$ and $0 \in Q_{ad}$.

Proposition 2.37. *Let* (2.7) *hold with* $\omega_0 = 0$. *If there exists* $\breve{q} \in Q_{ad}$ *such that* $\|B\breve{q} - Au_d\|_{V^*} < \alpha_0\delta_0$, *then* (P) *is strongly stable on the right for all* $\delta \geq 0$.

Proof. Let $u \in U \cap V$. If $\|u - u_d\| < \delta_0$, we have $N_U(u) = \{\, 0 \,\}$, and nothing to show. Therefore, let $\|u - u_d\| = \delta_0$. Due to [40, Corollary 10.44] it holds

$$N_U(u) = \{\, \tau(u - u_d) \colon \tau \geq 0 \,\}.$$

Without restriction, we can therefore consider $\zeta = u - u_d$. We calculate

$$\begin{aligned} h(u,\zeta) = \min_{q \in Q_{ad}} \langle Bq - Au, \zeta \rangle &= \langle Au_d - Au, u - u_d \rangle_{V^*,V} + \min_{q \in Q_{ad}} \langle Bq - Au_d, \zeta \rangle \\ &\leq -\alpha_0 \|u - u_d\|_V^2 + \langle B\breve{q} - Au_d, \zeta \rangle \\ &\leq -\alpha_0 \|u - u_d\| \|u - u_d\|_V + \|B\breve{q} - Au_d\|_{V^*} \|\zeta\|_V \\ &= (-\alpha_0\delta_0 + \|B\breve{q} - Au_d\|_{V^*}) \, \|\zeta\|_V. \end{aligned}$$

Due to the supposition there is $h_0 > 0$ such that $h(u,\zeta) \leq -h_0\|\zeta\|_V \leq -h_0\|\zeta\|$ and we can apply Theorem 2.18 to guarantee strong stability on the right. □

However, in case $\omega_0 > 0$, the control has to counteract unstable modes of A. We will discuss this situation in the following example.

2.4.3. Finite-approximate controllability constraint

Motivated by the concept of finite-approximate controllability (see, e.g., [167]), we consider the constraint

$$U = \{\, u \in H \colon \|u\| \leq \delta_0 \text{ and } Fu = 0 \,\}. \tag{2.38}$$

Concretely, let $\{\, f_1, \ldots, f_M \,\} \subset V$ be pairwise orthonormal in H and set

$$Fu = \sum_{i=1}^{M} (f_i, u) f_i, \quad u \in H.$$

In this subsection, we will investigate weak invariance in the particular case that $\operatorname{ran} F \coloneqq \operatorname{span}\{\, f_1, \ldots, f_M \,\}$ is an invariant subspace of A^*. Concretely, we require that

$$A^* f_i \subset \operatorname{ran} F, \quad i = 1, \ldots, M. \tag{2.39}$$

A particularly interesting example is to choose the functions f_i as a basis of the unstable subspace of A^* (the real span of all eigenvalues with negative real part). A target set of the form $U = \ker F$ is then motivated by the desire to steer the system into a stable subspace; cf. [58]. From an application point of view, it might be desirable not just to steer the system into a stable subspace but also into a sufficiently small stable state. In this case, the terminal set is given by (2.38).

First, for the sake of clarity, we will investigate (2.38) with $\delta_0 = \infty$, i.e., we will consider $U = \ker(F)$. The minimizing projection onto $\ker(F)$ is given by $P_{\ker(F)} = \mathrm{Id} - F$. By virtue of Proposition 2.8 for $u \in U$ we have

$$N_U(u) = \{\, Fu' \colon u' \in H,\, u = u' - Fu' \,\}.$$

Proposition 2.38. *If $0 \in Q_{ad}$ and (2.39) holds, then $U = \ker(F)$ is weakly invariant under (A, BQ_{ad}). Moreover, if there is $h_0 > 0$ such that for all $u' \in V$ there is $\breve{q} \in Q_{ad}$ such that*

$$\langle \breve{q}, B^* F u' \rangle \leq -h_0 \|Fu'\|, \tag{2.40}$$

then (P) with $U = \ker(F)$ is strongly stable on the right for all $\delta \geq 0$.

Condition (2.40) implies that $\ker(B^*) \cap \operatorname{ran}(F) = \{\, 0 \,\}$. In particular, we require at least as many controls as $\dim \operatorname{ran}(F) = M$. Hence, this condition is in general stronger than approximate controllability (or stabilizability), where the necessary number of controls is given by the largest geometric multiplicity of the eigenvalues (resp. the unstable eigenvalues); cf. [9, Section 3.4]. We can also give examples where (2.40) holds: For instance, if the control acts in an arbitrary open subset $\omega \subset \Omega$, then (2.40) is satisfied (under certain smoothness assumptions on the coefficients of A and the domain), since the eigenfunctions of A^* restricted to ω are linearly independent; see [58, Theorem 4.1].

Proof of Proposition 2.38. Let $u' \in V$ such that $u = u' - Fu'$ and set $\zeta = Fu'$. Then

$$h(u, \zeta) = \min_{q \in Q_{ad}} \langle Bq - Au, \zeta \rangle \leq -\langle u' - Fu', A^* F u' \rangle = 0,$$

since $A^* F u' \in \operatorname{ran}(F)$. Theorem 2.9 yields the first assertion. Moreover, the strengthened Hamiltonian condition is equivalent to (2.40) due to the calculation above proving the second assertion. □

Next, we turn to the general case of (2.38) with $\delta_0 < \infty$.

Proposition 2.39. *Assume $0 \in Q_{ad}$ and let (2.39) and (2.40) hold. Moreover, suppose that $\{\, f_1, \dots, f_M \,\}$ is chosen such that for all $\varphi \in \ker(F)$ it holds $\langle A\varphi, \varphi \rangle \geq \omega_1 \|\varphi\|^2$ with $\omega_1 > 0$. Then (P) with $U = \mathcal{B}_{\delta_0}(0) \cap \ker(F)$ is strongly stable on the right for all $\delta \geq 0$.*

Proof. First, we will show the following formula for the minimizing projection P_U:

$$P_U(u) = \min\{\, 1, \delta_0 / \|u - Fu\| \,\}\ (u - Fu) =: \gamma(u)\,(u - Fu).$$

Let $u \in H$. If $\|u - Fu\| \leq \delta_0$, then for all $u' \in U$ we calculate

$$(u - P_U(u), u' - P_U(u)) = (Fu, u' - u + Fu) = (u, Fu') - (Fu, u - Fu) = 0.$$

In the other case $\|u - Fu\| > \delta_0$ set $\gamma = \gamma(u)$ and we obtain for all $v \in U$ that

$$\begin{aligned}(u - P_U(u), v - P_U(u)) &= (1-\gamma)(u, v - \gamma(u - Fu)) + \gamma(Fu, v) - \gamma^2(Fu, u - Fu)\\ &= (1-\gamma)(u - Fu, v) - (1-\gamma)\gamma\|u - Fu\|^2\\ &\le (1-\gamma)\|u - Fu\|\|v\| - (1-\gamma)\delta_0\|u - Fu\| \le 0,\end{aligned}$$

where we have used again that $(Fu, v) = (u, Fv) = 0$ and $(Fu, u - Fu) = 0$ in the second step, and $\|v\| \le \delta_0$ in the last step. By virtue of Proposition 2.8 for $u \in U$ we infer that

$$\begin{aligned}N_U(u) &= \{\, (1 - \gamma(u'))u' + \gamma(u')Fu' \colon u' \in H,\, u = P_U(u') \,\}\\ &= \{\, (1 - \gamma(u'))(u' - Fu') + Fu' \colon u' \in H,\, u = P_U(u') \,\}.\end{aligned}$$

Consider the single terms of the Hamiltonian for $u' \in V$ and set $\gamma = \gamma(u')$. We consider the case $\gamma < 1$, only; the other case is analogous to Proposition 2.38. Then for any $q \in Q$

$$\langle Bq, u' - P_U(u')\rangle = (1-\gamma)\langle Bq, u' - Fu'\rangle + \langle Bq, Fu'\rangle$$

and, since $\langle A(u' - Fu'), Fu'\rangle = 0$, we find

$$\begin{aligned}-\langle AP_U(u'), u' - P_U(u')\rangle &= -\gamma(1-\gamma)\langle A(u' - Fu'), u' - Fu'\rangle - \gamma\langle A(u' - Fu'), Fu'\rangle\\ &= -\gamma(1-\gamma)\langle A(u' - Fu'), u' - Fu'\rangle.\end{aligned}$$

Due to the supposition $\langle A\varphi, \varphi\rangle \ge \omega_1\|\varphi\|^2$ for all $\varphi \in \ker(F)$ we infer

$$-\langle AP_U(u'), u' - P_U(u')\rangle \le -\gamma(1-\gamma)\omega_1\|u' - Fu'\|^2 = -(1-\gamma)\omega_1\delta_0\|u' - Fu'\|$$

from the calculation above. Combining the previous estimates, we obtain

$$\begin{aligned}\langle Bq - AP_U(u'), u' - P_U(u')\rangle &\le (1-\gamma)\left[\langle Bq, u' - Fu'\rangle - \omega_1\delta_0\|u' - Fu'\|\right] + \langle Bq, Fu'\rangle\\ &\le (1-\gamma)(\|B\|\|q\| - \omega_1\delta_0)\|u' - Fu'\| + \langle Bq, Fu'\rangle.\end{aligned}$$

Assuming that $0 \in Q_{ad}$, choosing $q = \lambda\breve{q}$, $\lambda = \min\{\,1, (\omega_1\delta_0)/(2\|B\|\|\breve{q}\|)\,\}$, where $\breve{q}$ is the control to realize (2.40), we obtain the strengthened Hamiltonian condition (with a suitably modified constant h_0). □

2.4.4. Stabilization with finite dimensional control

We have seen that the criteria for strong stability of systems with general A and U require certain assumptions, which are somewhat restrictive. In this section, we will show that there exist neighbourhoods U of zero such that the resulting problem is strongly stable, assuming only stabilizability (controllability of the unstable modes).

Here, we suppose that the control is finite dimensional, $Q = \mathbb{R}^{N_c}$. The set of admissible controls contains a neighborhood of zero, e.g., $Q_{ad} = \{\, q \in \mathbb{R}^{N_c} \colon q \in [-K, K]^{N_c} \,\}$ for some fixed $K > 0$. We are interested to bring the system into a small neighborhood of the stationary state zero. Note that we could more generally consider weakly invariant states u_d, i.e. $\{\, u_d \,\}$ is weakly invariant under (A, BQ_{ad}). A short computation based on Theorem 2.9 reveals that $Au_d \in BQ_{ad}$. However, this case follow directly from the case $u_d = 0$ by an affine change of variables, and we omit it for simplicity of notation.

To ensure that admissible controls for (P) exist, we can employ the concept of stabilizability, which is widely accepted in the control literature. Concretely, we assume in the following

that $(-A, B)$ should be *stabilizable*, which can be verified with the Fattorini criterion; see [9] and the references therein. This means that

$$A^*\zeta = \lambda\zeta, \quad \operatorname{Re}\lambda \leq 0, \quad B^*\zeta = 0 \quad \Longrightarrow \quad \zeta = 0.$$

It is known that this implies the existence of a stabilizing feedback law, such that $\|u(t)\| \leq M_0 \exp(-\gamma_0 t)\|u_0\|$ for some $\gamma_0 > 0$, which in turn guarantees existence for (P) (given u_0 sufficiently small or Q_{ad} sufficiently large). Additionally, we will show that it is possible to choose some appropriate neighborhood U of zero, such that the criterion for strong stability (and thus weak invariance) is guaranteed.

First, we consider the infinite horizon optimization problem

$$\min_{q \in L^2((0,\infty);\mathbb{R}^{N_c})} \int_0^\infty \left[\|u[q,u'](t)\|^2 + \|q(t)\|^2_{\mathbb{R}^{N_c}} \right] \mathrm{d}t, \tag{2.41}$$

where $u[q,u']$ is the solution to the state equation on $(0,\infty)$ with control $q \in L^2((0,\infty);\mathbb{R}^{N_c})$ and initial condition $u' \in H$. This defines a linear, bounded, self-adjoint and nonnegative operator $\Pi\colon H \to H$ such that $(\Pi u', u')$ is the minimal value of (2.41) and Π satisfies the following algebraic Riccatti equation

$$-\langle A^*\Pi\varphi, \psi\rangle - \langle \Pi A\varphi, \psi\rangle + (\varphi, \psi) = (B^*\Pi\varphi, B^*\Pi\psi)_{\mathbb{R}^{N_c}}, \tag{2.42}$$

for all $\varphi, \psi \in V$; see, e.g., [101, Theorem 2.2.1 (a_2), (a_4)]. Furthermore, Π maps H into $X_{1-\theta_0}$, hence Π is compact on H; see [101, Theorem 2.2.1 (a_3)].

Define the norm $\|\cdot\|_\Pi = (\Pi\cdot, \cdot)^{1/2}$ induced by the operator Π. Let the terminal constraint be given by

$$U = \{\, u \in H\colon \|u\|_\Pi \leq \delta_0 \,\}. \tag{2.43}$$

Thus, $u \in U$ corresponds to a constraint on the optimal value function of (2.41) with initial value u. Since Π is self-adjoint, according to [40, Corollary 10.44] for all $u \in \partial U$ we have

$$N_U(u) = \{\, \lambda\Pi u\colon \lambda \geq 0 \,\} \subset V.$$

Inserting the optimal feedback law $\breve{q} = -B^*\Pi u$ we estimate

$$\begin{aligned} h(u,\zeta) = h(u,\Pi u) &= \inf_{q \in Q_{ad}} (q, B^*\Pi u)_{\mathbb{R}^{N_c}} - \frac{1}{2}\langle u, (A^*\Pi + \Pi A)u\rangle \\ &\leq -(B^*\Pi u, B^*\Pi u)_{\mathbb{R}^{N_c}} - \frac{1}{2}\langle u, (A^*\Pi + \Pi A)u\rangle. \end{aligned}$$

This is valid as long as $\breve{q} = -B^*\Pi u \in Q_{ad}$. Since

$$\|B^*\Pi u\|_{\mathbb{R}^{N_c}} \leq \|B^*\Pi^{1/2}\|_{\mathcal{L}(H,\mathbb{R}^{N_c})}\|u\|_\Pi = \|B^*\Pi^{1/2}\|_{\mathcal{L}(H,\mathbb{R}^{N_c})}\delta_0,$$

this can be achieved by a sufficiently small choice of δ_0. Now we use (2.42) to obtain

$$h(u,\Pi u) \leq -\frac{1}{2}(B^*\Pi u, B^*\Pi u)_{\mathbb{R}^{N_c}} - \frac{1}{2}\|u\|^2 \leq 0 - \frac{1}{2\|\Pi\|\|\Pi^{1/2}\|}\|\Pi u\|\|u\|_\Pi \leq -h_0\|\Pi u\|,$$

where $h_0 = \delta_0/(2\|\Pi\|^{3/2}_{\mathcal{L}(H)})$. Thus, strong stability of (P) is guaranteed by Theorem 2.18, assuming only stabilizability (approximate controllability of the unstable modes).

From a practical point of view, the choice of the target set (2.43) can be interpreted as follows: Since the norm $\|u\|_\Pi$ corresponds to the optimal value of (2.41), we have in particular the estimates $\|\breve{u}(t)\|_{L^2((0,\infty);H)} \leq \|u\|_\Pi$ and $\|\breve{q}(t)\|_{L^2((0,\infty);\mathbb{R}^{N_c})} \leq \|u\|_\Pi$ where $\breve{u}(t)$ is the trajectory starting at $\breve{u}(0) = u$ with control given by the feedback law $\breve{q}(t) = -B^*\Pi\breve{u}(t)$. Thus, we aim to enter a neighboorhood of zero that contains only states which can be stabilized at low cost. After the end of the optimization horizon T, the control can be chosen by the optimal feedback law to keep the trajectory stable.

3. Second order and sufficient optimality conditions

Since $(\hat{P})$ is a nonconvex optimization problem, first order optimality conditions are not sufficient for optimality in general. We therefore discuss second order optimality conditions and sufficient optimality conditions in the following. In particular, these results will be essential for proving a priori discretization error estimates in Chapter 5. Generally, this chapter relies on the first order optimality conditions of Chapter 2 and the problem formulation used therein. However, here we in addition suppose that the terminal set U is given as a sublevel set of some smooth function G. Moreover, we restrict to the choice

$$L(q) = \frac{\alpha}{2}\|q\|_Q^2 \quad \text{for } \alpha \geq 0.$$

The problem setting will be introduced in detail in Section 3.1, where we also recap the first order optimality condition for the specific problem. A concrete example of a convection-diffusion equation subject to mixed boundary conditions for different control scenarios will be discussed at the end of that section.

As the time-optimal control problems with $\alpha > 0$ and $\alpha = 0$ typically lead to different solutions, we have to distinguish these two cases. This is also reflected in the main structure of this chapter. In Section 3.2 we will provide second order necessary and sufficient optimality conditions for the case $\alpha > 0$. Employing a critical cone, this leads to a minimal gap between necessity and sufficiency. Here we rely on the work of Casas and Tröltzsch on second order optimality conditions; see, e.g., [30, 34, 35].

In general, it seems to be a difficult task to verify a second order condition for a given problem both theoretically and numerically. However, for the problem under consideration, we show that the second order condition on the (infinite dimensional) critical cone is equivalent to a scalar condition where we have to solve one (infinite dimensional) linear system. This condition gives rise to the verification of second order conditions on the discrete level, by calculating the scalar quantity numerically; see Section 5.4. Furthermore, the scalar condition can be interpreted in terms of the value function with respect to the time transformation ν. More specifically, we show that the second order sufficient optimality condition holds if and only if the second derivative of the value function is strictly positive. These results are already contained in [17] in similar form, however most of them without detailed proofs.

Section 3.3 is devoted to the case $\alpha = 0$, where we rely on an established structural assumption on the adjoint state; cf. Remark 3.27 and the references given there. This assumption is sufficient for optimality and leads to a growth condition in $L^1(I \times \omega)$. It is worth mentioning, that due to the particular structure of the objective functional, we do not require additional assumptions such as conditions on the second derivative of the Lagrange function. As a first application, we study the stability of solutions to the regularized time-optimal control problem. Under certain assumptions we show Lipschitz continuity of the optimal time and the optimal control in $L^1(I \times \omega)$ with respect to the regularization parameter α.

3.1. Problem formulation

In addition to the general assumptions of Section 2.1, we suppose the following problem setup throughout this chapter. Let (ω, ϱ) be a measure space and set $Q = L^2(\omega, \varrho)$. This notation allows for the unified treatment of different control situations. For example, in case of a distributed control, we take $\omega \subset \Omega$ equipped with the usual Lebesgue measure, where Ω denotes the spatial domain of the parabolic equation. An example of a reaction-diffusion equation with different control scenarios will be discussed in Section 3.1.2. Since no ambiguity arises, we simply write $L^2(\omega)$ instead of $L^2(\omega, \varrho)$ in the following. If we write almost everywhere in ω, then this always refers to the respective measure. The space of admissible controls is defined as

$$Q_{ad} \coloneqq \left\{ q \in L^2(\omega) \colon q_a \le q \le q_b \text{ a.e. in } \omega \right\} \subset L^\infty(\omega)$$

for $q_a, q_b \in L^\infty(\omega)$ with $\operatorname{ess\,inf}_{x\in\omega}(q_b(x) - q_a(x)) > 0$. The set $I \times \omega$ is equipped with the completion of the corresponding product measure. Recall that $Q(0,1) \coloneqq L^2((0,1); L^2(\omega))$ and

$$Q_{ad}(0,1) \coloneqq \left\{ q \in L^2(I \times \omega) \colon q(t) \in Q_{ad} \text{ a.a. } t \in (0,1) \right\} \subset L^\infty(I \times \omega).$$

Concerning the regularization or cost term L, we suppose that

$$L(q) = \frac{\alpha}{2} \|q\|^2_{L^2(\omega)}$$

for $\alpha \ge 0$. Moreover, in place of the general terminal set U, we assume that U is given as a sublevel set of a continuously differentiable function $G \colon H \to \mathbb{R}$, precisely,

$$U = \{ u \in H \colon G(u) \le 0 \}.$$

In this and the next chapters we work again with the state equation transformed to the reference interval $(0,1)$ by means of a transformation ν. Here, we restrict to $\nu \in \mathbb{R}_+$ for the following reason: Recall that in the previous chapter we took $\nu \in L^\infty((0,1))$ such that $\operatorname{ess\,inf}_{\tau\in(0,1)} \nu(\tau) > 0$. With this choice, the transformed and the untransformed problems are equivalent with the relation $T = \int_0^1 \nu(\tau)\,\mathrm{d}\tau$. In particular, given $T > 0$ there is a sequence $\nu_n \in L^\infty((0,1))$ such that $T = \int_0^1 \nu_n(\tau)\,\mathrm{d}\tau$, $\nu_n \not\equiv T$, and $\nu_n \to \nu \coloneqq T$ in $L^\infty((0,1))$. Hence, the time transformation ν is not locally unique. However, typically second order sufficient optimality conditions imply local uniqueness and this is also valid in our setting; see Theorem 3.25. Therefore, we have to take $\nu \in \mathbb{R}_+$ for the time transformation, otherwise a local solution cannot be locally unique. The time-optimal control problem reads as

$$\inf_{\substack{\nu\in\mathbb{R}_+ \\ q\in Q_{ad}(0,1)}} \quad j(\nu, q) \quad \text{subject to} \quad g(\nu, q) \le 0,$$

where the objective function is given by

$$j(\nu, q) = \nu \left(1 + \int_0^1 \frac{\alpha}{2} \|q\|^2_{L^2(\omega)} \right),$$

and the reduced terminal constraint is defined as

$$g(\nu, q) \coloneqq G(i_1 S(\nu, q)), \quad (\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1).$$

Note that due to continuity of $u \colon [0,1] \to H$, the optimal solution must fulfill the terminal constraint with equality (otherwise, a control with a shorter time is still admissible, while having a smaller objective value). Hence, we can equivalently use $g(\nu, q) \le 0$ or $g(\nu, q) = 0$ in the problem formulation above. Furthermore, to avoid confusion with the spatial gradient ∇, we denote the gradients of g and G by $g'(\cdot)^*$ and $G'(\cdot)^*$ in the following.

3.1.1. First order optimality conditions

The first order optimality conditions from Chapter 2 still hold if we take $\nu \in \mathbb{R}_+$, except for the constancy of the Hamiltonian condition (2.21) where the 'almost everywhere' is substituted by the integral over the time interval. For convenience we summarize the first order optimality conditions for the setting considered in this chapter. We require the following *linearized Slater* condition.

Assumption 3.1. We assume that

$$\bar{\eta} \coloneqq -\partial_\nu g(\bar{\nu}, \bar{q}) > 0. \tag{3.1}$$

Note that by Assumption 3.1 and $g(\bar{\nu}, \bar{q}) = 0$, the point $\breve{\chi}^\rho = (\bar{\nu} + \rho, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ defined for $\rho > 0$ fulfills

$$g(\bar{\chi}) + g'(\bar{\chi})(\breve{\chi}^\rho - \bar{\chi}) = -\bar{\eta}\rho < 0, \tag{3.2}$$

where $\bar{\chi} = (\bar{\nu}, \bar{q})$, which corresponds to a more familiar presentation of the linearized Slater condition. We will see that for the particular problem it is not more general to suppose that first order optimality conditions hold in qualified form than to assume that the linearized Slater condition in the form of (3.1) is valid (or any other constraint qualification).

In order to state optimality conditions, we introduce the Lagrange function as

$$\mathcal{L}\colon \mathbb{R}_+ \times Q(0,1) \times \mathbb{R} \to \mathbb{R}, \quad \mathcal{L}(\nu, q, \mu) \coloneqq j(\nu, q) + \mu\, g(\nu, q).$$

Now, optimality conditions for $(\hat{P})$ in qualified form can be stated as follows: For given $\bar{\nu} > 0$ and $\bar{q} \in Q_{ad}(0,1)$ with $g(\bar{\nu}, \bar{q}) = 0$ there exists a $\bar{\mu} \geq 0$, such that

$$\partial_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})(\delta\nu, q - \bar{q}) \geq 0 \quad \text{for all } (\delta\nu, q) \in \mathbb{R} \times Q_{ad}(0,1). \tag{3.3}$$

With Assumption 3.1, a multiplier always exists and, due to the special structure, it is always positive. We summarize this in the next result.

Lemma 3.1. *Let $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be a solution of $(\hat{P})$ with associated state $\bar{u} = S(\bar{\nu}, \bar{q})$ and the linearized Slater condition* (3.1) *holds. Then there exist $\bar{\mu} \in \mathbb{R}$ and $\bar{z} \in W(0,1)$ such that*

$$\bar{\mu} > 0, \tag{3.4}$$

$$\int_0^1 1 + \frac{\alpha}{2}\|\bar{q}(t)\|^2_{L^2(\omega)} + \langle B\bar{q}(t) - A\bar{u}(t), \bar{z}(t)\rangle \,\mathrm{d}t = 0, \tag{3.5}$$

$$\int_0^1 \langle \alpha\bar{q}(t) + B^*\bar{z}(t), q(t) - \bar{q}(t)\rangle \,\mathrm{d}t \geq 0, \qquad q \in Q_{ad}(0,1), \tag{3.6}$$

$$G(\bar{u}(1)) = 0, \tag{3.7}$$

where $\bar{z}$ is the adjoint state *determined by*

$$-\partial_t \bar{z} + \bar{\nu} A^* \bar{z} = 0, \quad \bar{z}(1) = G'(\bar{u}(1))^* \bar{\mu}.$$

Proof. We first note that the linearized Slater condition allows for exact penalization of $(\hat{P})$; see [21, Theorem 2.87, Proposition 3.111]. The optimality conditions now follow as in the proof of Theorem 2.25. Since in our setting the multiplier $\bar{\mu}$ belongs to the normal cone $N_{(-\infty,0]}(0) \subset \mathbb{R}$, we in addition infer $\bar{\mu} \geq 0$. If $\bar{\mu} = 0$, then $\bar{z} = 0$, which is a contradiction to the Hamiltonian condition (3.5). Thus, $\bar{\mu} > 0$. □

The condition (3.5) is equivalent to $\partial_\nu \mathcal{L}(\bar\nu, \bar q, \bar\mu) = 0$ and (3.6) arises from (3.3) for $\delta\nu = 0$. Note that the optimality conditions of Lemma 3.1 are consistent with the theory of Chapter 2, because of the characterization

$$N_U(\bar u(1)) = \{\, G'(\bar u(1))^* \lambda \colon \lambda \geq 0 \,\}$$

according to [40, Corollary 10.44]. As in Proposition 2.21, for $\nu \in \mathbb{R}_+$, $q \in Q(0,1)$, and $\mu \in \mathbb{R}$ we have the representation

$$g'(\nu, q)^* \mu = \begin{pmatrix} \int_0^1 \langle Bq - Au, z\rangle \\ \nu B^* z \end{pmatrix}, \tag{3.8}$$

where $z \in W(0,1)$ is the unique solution to

$$-\partial_t z + \nu A z = 0, \quad z(1) = G'(i_1 S(\nu, q))^* \mu. \tag{3.9}$$

Constancy of the Hamiltonian (3.5) allows to prove equivalence of qualified optimality conditions and condition (3.1).

Proposition 3.2. *The first order optimality conditions of Lemma 3.1 hold in qualified form if and only if the linearized Slater condition* (3.1) *is valid.*

Proof. According to (3.8) and (3.5) we have

$$\begin{aligned} \langle \partial_\nu g(\bar\nu, \bar q)\rho, \bar\mu\rangle &= G'(i_1 S(\bar\nu, \bar q)) i_1 S'(\bar\nu, \bar q)(\rho, 0)\bar\mu = \int_0^1 \rho \langle B\bar q(t) - A\bar u(t), \bar z(t)\rangle \,\mathrm{d}t \\ &= -\int_0^1 \rho \left(1 + \frac{\alpha}{2}\|\bar q(t)\|^2\right) \mathrm{d}t \quad < 0 \end{aligned}$$

for any $\rho > 0$. Hence, with the choice $\varepsilon = -\partial_\nu g(\bar\nu, \bar q) > 0$, condition (3.1) holds. The remaining implication is the assertion of Lemma 3.1. □

Furthermore, as in the linear parabolic case, see, e.g., [147, Section 3.6], the following projection formula holds

$$\bar q = P_{Q_{ad}}\left(-\frac{1}{\alpha} B^* \bar z\right), \tag{3.10}$$

where $P_{Q_{ad}}(\cdot)$ denotes the pointwise projection onto the set $Q_{ad}(0,1)$, defined by

$$P_{Q_{ad}} \colon L^2(I \times \omega) \to Q_{ad}(0,1), \quad P_{Q_{ad}}(r)(t,x) = \max\{q_a(x), \min\{q_b(x), r(t,x)\}\}.$$

3.1.2. Example of a convection-diffusion equation

We conclude the introduction of this chapter by an example of a convection-diffusion equation on a bounded domain subject to mixed boundary conditions that satisfies the abstract assumptions. First, we introduce corresponding function spaces and the operator A. Concrete examples for the measure space (ω, ϱ) will be given at the end of this subsection.

Let $\Omega \subset \mathbb{R}^d$ with $d \in \{2, 3, \dots\}$ be a bounded domain with boundary $\partial\Omega$. We use Γ_N for the relatively open subset of $\partial\Omega$ denoting the Neumann boundary part and $\Gamma_D = \partial\Omega \setminus \Gamma_N$ the Dirichlet boundary part. We assume that $\Omega \cup \Gamma_N$ is regular in the sense of Gröger; see [69, Definition 2]. In addition, we suppose that each mapping ϕ_x in the definition of Gröger regularity is volume-preserving.

Remark 3.3. The notion of Gröger regular has been introduced in [69] and is meanwhile widely used in the regularity theory for partial differential equations. For clarity, we summarize well-known properties and elaborate on its relation to Lipschitz domains and domains with Lipschitz boundary:

(i) If $\Omega \cup \Gamma_N$ is regular in the sense of Gröger, then Ω is a Lipschitz domain; see [72, Theorem 5.1]. Conversely, if Ω is a Lipschitz domain, then Ω and $\Omega \cup \partial\Omega$ are Gröger regular; cf. Definition 1.2.1.2 in [68].

(ii) For simplified characterizations of regular sets we refer to [72, Theorems 5.2 and 5.4].

(iii) Note that the cases $\Gamma_D = \emptyset$ or $\Gamma_D = \partial\Omega$ are not excluded.

(iv) The additional requirement of volume-preserving bi-Lipschitz transformations is satisfied in many practical situations. For example, in spatial dimension three, two crossing beams allow for a volume-preserving bi-Lipschitz transformation; see Section 7.3 in [73]. Moreover, this is true for domains with Lipschitz boundary; see Remark 3.3 in [73].

As usual, for $\theta \in (0,1]$ and $p \in (1,\infty)$, we define the space $H_D^{\theta,p}(\Omega)$ as the closure of

$$C_D^\infty(\Omega) = \left\{ \psi|_\Omega \colon \psi \in C^\infty(\mathbb{R}^d),\, \operatorname{supp}(\psi) \cap \Gamma_D = \emptyset \right\}$$

in the Bessel-potential space $H^{\theta,p}(\Omega)$, i.e.

$$H_D^{\theta,p}(\Omega) = \overline{C_D^\infty(\Omega)}^{H^{\theta,p}(\Omega)}.$$

If $\theta = 1$, then the space $H_D^{\theta,p}(\Omega)$ coincides with the usual Sobolev space that we denote by $W_D^{1,p}(\Omega)$. Of course, if $\Gamma_N = \emptyset$, then $H_D^{\theta,p}(\Omega) = W_0^{\theta,p}(\Omega)$, and if $\Gamma_N = \partial\Omega$, then $H_D^{\theta,p}(\Omega) = W^{\theta,p}(\Omega)$. The corresponding dual space of $H_D^{\theta,p}(\Omega)$ is denoted by $H_D^{-\theta,p'}(\Omega)$, where p' denotes the Hölder conjugate $1 = 1/p + 1/p'$. In addition, if $\Gamma_N = \emptyset$, we write $W^{-1,p'}(\Omega) = W_D^{-1,p'}$ for the dual space of $W_D^{1,p}$. These function spaces have the following properties.

Proposition 3.4. *Let $\theta \in (0,1]$ and $p \in (1,\infty)$.*

(i) If $p \geq d$, then $W_D^{1,p}(\Omega) \hookrightarrow_c L^r(\Omega)$ for $1 \leq r \leq dp/(p-d)$.

(ii) If $1 - d/p \geq \theta - d/r$, then $W_D^{1,p}(\Omega) \hookrightarrow H_D^{\theta,r}(\Omega)$.

Proof. Both injections follow by first extending the functions to $\mathcal{B} \subset \mathbb{R}^d$ an open ball containing $\overline{\Omega}$, using the corresponding results for smooth domains, and finally restricting to Ω again. According to [8, Lemma 3.2 with Remark 3.3 (i)] there is a continuous extension operator $E \colon W_D^{k,p}(\Omega) \to W^{k,p}(\mathcal{B})$ for all $p \in [1,\infty]$ and $k = 0, 1$. Thus, [146, Theorem 4.6.1] yields

$$\|u\|_{H_D^{\theta,r}(\Omega)} = \|Eu\|_{H_D^{\theta,r}(\Omega)} \leq \|Eu\|_{H_D^{\theta,r}(\mathcal{B})} \leq \|Eu\|_{W^{1,p}(\mathcal{B})} \leq \|u\|_{W_D^{1,p}(\Omega)}.$$

Compactness of (i) follows from the arguments above and [114, Theorem 1.4.6.2]. □

Note that the space $H_D^{-\theta,p}(\Omega)$ allows for distributional objects such as surface charge densities or thermal sources concentrated on hypersurfaces due to the following trace theorem.

Proposition 3.5 ([73, Theorem 3.6]). *For all $p \in (1,\infty)$ and $\theta \in (1/p, 1)$ there exists a continuous trace operator*

$$\mathrm{Tr}\colon H_D^{\theta,p}(\Omega) \to L^p(\Gamma_N),$$

and by duality its adjoint satisfies

$$\mathrm{Tr}^*\colon L^{p'}(\Gamma_N) \to H_D^{-\theta,p'}(\Omega).$$

The function spaces introduced above now allow for the definition of the operator A. It is given by the bilinear form

$$a(u,\varphi) = \int_\Omega [\kappa \nabla u \cdot \nabla \varphi + b \cdot \nabla u \varphi + c_0 u \varphi]\,\mathrm{d}x + \int_{\Gamma_N} c_1 u \varphi \,\mathrm{d}s, \quad u, \varphi \in H_D^1(\Omega),$$

for $\kappa\colon \Omega \to \mathbb{R}^{d\times d}$ a coefficient function satisfying the usual uniform ellipticity condition

$$\kappa_\bullet \|z\|^2 \le \sum_{i,j=1}^d \kappa_{ij}(x) z_j z_i \quad \text{for all } z \in \mathbb{R}^d \text{ and a.a. } x \in \Omega,$$

and $\|\kappa_{ij}\|_{L^\infty(\Omega)} \le \kappa^\bullet$, $i,j = 1,2,\dots,d$, for constants $\kappa^\bullet, \kappa_\bullet > 0$. Moreover, $b \in L^\infty(\Omega;\mathbb{R}^d)$, $c_0 \in L^\infty(\Omega)$, and $c_1 \in L^\infty(\Gamma_N)$. We note that the first term is the weak formulation of a convection-diffusion-reaction equation and the second term allows for either the Robin boundary conditions ($\kappa \partial u/\partial n + c_1 u = 0$, where n is the outer normal to Ω) or homogeneous Neumann boundary conditions by setting $c_1 = 0$. According to [114, Corollary 1.4.7.2] there is $c > 0$ such that

$$\|\mathrm{Tr}\, u\|_{L^2(\Gamma_N)}^2 \le c \|u\|_{L^2(\Omega)} \|u\|_{H_D^1(\Omega)}, \quad u \in H_D^1(\Omega).$$

Thus, using the assumptions on the coefficients and Young's inequality, we immediately infer

$$a(u,u) \ge \frac{\kappa_\bullet}{4}\|u\|_{H_D^1}^2 - \left(\frac{\kappa_\bullet}{2} + \frac{\|b\|_{L^\infty}^2}{2\kappa_\bullet} + \|c_0\|_{L^\infty} + \frac{c^2\|c_1\|_{L^\infty}^2}{\kappa_\bullet}\right)\|u\|_{L^2}^2.$$

For these reasons, taking $V = H_D^1(\Omega)$, $H = L^2(\Omega)$, and $V^* = H_D^{-1}(\Omega)$, the Gårding inequality (2.7) holds for $\alpha_0 = \kappa_\bullet/4$ and $\omega_0 = \kappa_\bullet/2 + \|b\|_{L^\infty}^2/(2\kappa_\bullet) + \|c_0\|_{L^\infty} + c^2\|c_1\|_{L^\infty}^2/\kappa_\bullet$. Next, we turn to the control operator B.

In Chapter 2 we have essentially worked with domains of fractional powers of A. Under the assumptions of this subsection, we can provide a convenient characterization in terms of the Bessel-potential spaces. Due to [65, Theorem 3.5] we have

$$H_D^{2\theta-1,2}(\Omega) = [H_D^{-1}(\Omega), H_D^1(\Omega)]_\theta, \quad \frac{1}{4} \ne \theta \ne \frac{3}{4}.$$

Since both H_D^1 and H_D^{-1} are Hilbert spaces, A exhibits bounded imaginary powers. Thus, from [146, Theorem 1.15.3] we conclude

$$X_\theta = \mathcal{D}_{V^*}(A^\theta) = [V^*, \mathcal{D}_{V^*}(A)]_\theta = [H_D^{-1}(\Omega), H_D^1(\Omega)]_\theta = H_D^{2\theta-1,2}(\Omega).$$

Finally, we give three examples of concrete control scenarios that satisfy the assumptions of Chapters 2 and 3.

Example 3.6 (Distributed control). In case of a distributed control on a subset $\omega \subset \Omega$ of the spatial domain, we simply take B as the extension by zero operator. Clearly, B is linear and continuous from $L^2(\omega)$ into $L^2(\Omega) = X_{\theta_0}$ with $\theta_0 = 1/2$.

Example 3.7 (Neumann boundary control). For Neumann boundary control we set $\omega = \Gamma_N$ and $B = \mathrm{Tr}^*$. Since the adjoint of the trace operator Tr^* is continuous from $L^2(\Gamma_N)$ into $H_D^{2\theta_0-1,2}(\Omega)$ for any $\theta_0 \in (0, 1/4)$, we infer $B\colon L^2(\Gamma_N) \to X_{\theta_0}$. Note that this holds true independently of the spatial dimension.

Example 3.8 (Purely time-dependent controls). Last, we consider the case of purely time-dependent controls that is of independent interest in theory as well as applications. For $\theta_0 \in (0, 1/4)$ let $e_1, \dots, e_{N_c} \in H_D^{2\theta_0-1,2}$ be given form functions. Define the control operator as

$$B\colon \mathbb{R}^{N_c} \to H_D^{2\theta_0-1,2}, \quad Bq = \sum_{i=1}^{N_c} q_i e_i.$$

The measure space (ω, ϱ) is defined as $\omega = \{1, 2, \dots, N_c\}$ equipped with the counting measure. Hence, the control space and the space of admissible controls, respectively, are given by

$$Q = L^2(\omega) \cong \mathbb{R}^{N_c}, \quad Q_{ad} = \{q \in Q\colon q_a \le q \le q_b\},$$

where $q_a, q_b \in \mathbb{R}^{N_c}$ and the inequality is to be understood componentwise.

3.2. Second order optimality conditions ($\alpha > 0$)

We first consider the case with cost term, i.e. $\alpha > 0$. These results are already contained in [17], however most of them without detailed proofs. We require the following regularity assumption concerning the terminal constraint.

Assumption 3.2. The function $G\colon H \to \mathbb{R}$ is twice continuously Fréchet-differentiable. In addition, the mapping $\eta \mapsto G''(u)\eta^2$ is weakly lower semicontinuous for all $u \in H$.

Moreover, the product space $\mathbb{R} \times L^2(I \times \omega)$ is endowed with the canonical inner product and we abbreviate its norm as

$$\|(\delta\nu, \delta q)\| = \left(|\delta\nu|^2 + \|\delta q\|^2_{L^2(I\times\omega)}\right)^{1/2}.$$

By means of Proposition 2.20, the reduced constraint mapping $g\colon \mathbb{R}_+ \times Q_{ad}(0,1) \to \mathbb{R}$ is twice continuously Fréchet-differentiable. Moreover, recalling

$$g''(\nu, q)[\delta\nu, \delta q]^2 = G''(u(1)) \left[i_1 S'(\nu, q)(\delta\nu, \delta q)\right]^2 + G'(u(1)) i_1 S''(\nu, q))[\delta\nu, \delta q]^2,$$

where $u = S(\nu, q)$, we have: If $\delta\nu_n \to \delta\nu$ and $\delta q_n \rightharpoonup \delta q$ weakly in $L^2(I \times \omega)$, then

$$\begin{aligned} S'(\nu, q)(\delta\nu_n, \delta q_n) &\rightharpoonup S'(\nu, q)(\delta\nu, \delta q) \quad \text{in } W(0,1), \\ S''(\nu, q)[\delta\nu_n, \delta q_n]^2 &\rightharpoonup S''(\nu, q)[\delta\nu, \delta q]^2 \quad \text{in } W(0,1), \end{aligned}$$

thanks to the bilinear structure. Hence, using weak lower semicontinuity of G'' and that $G'(i_1 S(\nu, q))$ is a linear bounded functional, we infer the following result.

Corollary 3.9. *Let $(\nu, q) \in \mathbb{R}_+ \times Q(0,1)$. If $\delta\nu_n \to \delta\nu$ in $\mathbb{R}$ and $\delta q_n \rightharpoonup \delta q$ weakly in $L^2(I\times\omega)$, then*

$$g''(\nu, q)[\delta\nu, \delta q]^2 \le \liminf_{n\to\infty} g''(\nu, q)[\delta\nu_n, \delta q_n]^2.$$

We introduce a *cone of critical directions* (or simply called *critical cone*) as

$$C_{(\bar\nu,\bar q)} = \left\{ (\delta\nu,\delta q) \in \mathbb{R}\times L^2(I\times\omega) \;\middle|\; \begin{array}{r}\delta q \text{ satisfies the sign condition (3.11), and}\\ g'(\bar\nu,\bar q)(\delta\nu,\delta q) = 0\end{array} \right\},$$

where the *sign condition* is given by

$$\delta q(t,x)\begin{cases} \le 0 \text{ if } \bar q(t,x) = q_b(x)\\ \ge 0 \text{ if } \bar q(t,x) = q_a(x)\\ = 0 \text{ if } \alpha\bar q(t,x) + B^*\bar z(t,x) \neq 0\end{cases} \quad \text{a.e. in } I\times\omega. \tag{3.11}$$

3.2.1. Second order necessary optimality conditions

With this definition, we can formulate second order necessary conditions, which hold in any locally optimal stationary point.

Theorem 3.10. *Let $(\bar\nu,\bar q)\in\mathbb{R}_+\times Q_{ad}(0,1)$ be a local minimum of $(\hat P)$ satisfying first order optimality conditions of Lemma 3.1 and $\bar\mu>0$. Then the following inequality holds*

$$\partial^2_{(\nu,q)}\mathcal{L}(\bar\nu,\bar q,\bar\mu)[\delta\nu,\delta q]^2 \ge 0 \quad \textit{for all } (\delta\nu,\delta q)\in C_{(\bar\nu,\bar q)}. \tag{3.12}$$

In general, for second order necessary conditions one needs a further constraint qualification. It is worth mentioning that – in our setting – we may conclude the regularity assumption from the first order optimality conditions. According to the linearized Slater condition (3.1), we have $g'(\bar\nu,\bar q)(\rho,0)\neq 0$ for any $\rho\neq 0$. Hence, there is a $\breve\nu\in\mathbb{R}$ such that $g'(\bar\nu,\bar q)(\breve\nu,0)=1$; cf. the regularity assumption (2.1) in [35]. For the proof, we require the following auxiliary result concerning admissible perturbations.

Proposition 3.11. *Let $(\nu,q)\in\mathbb{R}\times Q(0,1)$ satisfy $g(\nu,q)=0$. Moreover, let $\delta\nu\in\mathbb{R}$ and $\delta q\in Q(0,1)$ such that $g'(\nu,q)(\delta\nu,\delta q)=0$. Additionally, suppose $\partial_\nu g(\nu,q)\neq 0$. Then there are $\varepsilon>0$ and a function $\gamma\colon(-\varepsilon,\varepsilon)\to\mathbb{R}$ of class C^2 satisfying*

$$g(\nu(\theta),q(\theta))=0,\quad \theta\in(-\varepsilon,\varepsilon),\quad \gamma(0)=\gamma'(0)=0,$$

where $\nu(\theta)=\nu+\theta\delta\nu+\gamma(\theta)$ and $q(\theta)=q+\theta\delta q$.

Proof. Due to the supposition, there exists $\breve\nu\in\mathbb{R}$ such that $g'(\nu,q)(\breve\nu,0)=1$. We define

$$f(\theta,\rho)=g(\nu+\theta\delta\nu+\rho\breve\nu, q+\theta\delta q).$$

According to Assumption 3.2 and Proposition 2.20 f is of class C^2 and we have

$$\begin{aligned}\partial_\theta f(0,0) &= g'(\nu,q)(\delta\nu,\delta q)=0,\\ \partial_\rho f(0,0) &= g'(\nu,q)(\breve\nu,0)=1.\end{aligned}$$

By the implicit function theorem there exist $\varepsilon>0$ and a C^2-function $\widetilde\gamma\colon(-\varepsilon,\varepsilon)\to\mathbb{R}$ such that

$$f(\theta,\widetilde\gamma(\theta))=f(0,0)=0,\quad \theta\in(-\varepsilon,\varepsilon),\quad \widetilde\gamma(0)=0.$$

Moreover, differentiating the identity above we infer

$$\partial_\theta f(0,0)+\partial_\rho f(0,0)\widetilde\gamma'(0)=0,$$

so we conclude $\widetilde\gamma'(0)=0$. Last, we set $\gamma=\widetilde\gamma\breve\nu$. □

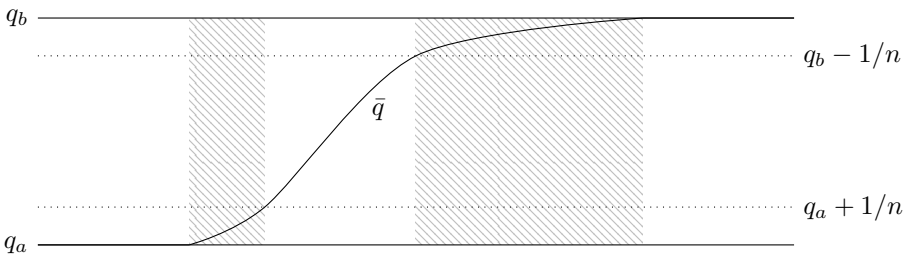

Figure 3.1.: Truncation procedure in the proofs of Theorem 3.10 and Proposition 3.12. Hachured parts are set to zero, the remaining parts are truncated.

Proof of Theorem 3.10. The assertion can be proved similarly as in [35] with an additional truncation procedure. Let $(\delta\nu, \delta q) \in C_{(\bar{\nu},\bar{q})}$. For $n \in \mathbb{N}$, we introduce the following truncation

$$\delta q_n = \begin{cases} 0 & \text{if } q_a < \bar{q} < q_a + 1/n \text{ or } q_b - 1/n < \bar{q} < q_b, \\ \max\{-n, \min\{n, \delta q\}\} & \text{else,} \end{cases}$$

almost everywhere in $I \times \omega$; see also Figure 3.1. Moreover, we set

$$\delta\nu_n = -\partial_q g(\bar{\nu}, \bar{q})\delta q_n / \partial_\nu g(\bar{\nu}, \bar{q}),$$

which is justified because of $\partial_\nu g(\bar{\nu}, \bar{q}) \neq 0$. By construction we have $g'(\bar{\nu}, \bar{q})(\delta\nu_n, \delta q_n) = 0$. According to Proposition 3.11 there is $\varepsilon > 0$ and a function $\gamma\colon (-\varepsilon, \varepsilon) \to \mathbb{R}$ such that the state constraints remain active for the pair $\nu(\theta) = \bar{\nu} + \theta\delta\nu_n + \gamma(\theta)$ and $q(\theta) = \bar{q} + \theta\delta q_n$. Due to the sign condition (3.11), we have $\bar{q} + \theta\delta q_n \in Q_{ad}(0,1)$ for $0 \leq \theta \leq \min\{\, 1/n^2, \theta_0/n \,\}$, where $\theta_0 := \operatorname{ess\,inf}_{x\in\omega}(q_b(x) - q_a(x)) > 0$. Moreover, $\delta\nu(\theta) > 0$ for all θ sufficiently small. Thus, the function

$$\phi\colon [0, \varepsilon) \to \mathbb{R}, \quad \phi(\theta) = \mathcal{L}(\nu(\theta), q(\theta), \bar{\mu}),$$

has a local minimum at $\theta = 0$. Since $\gamma'(0) = 0$ (see Proposition 3.11) we have

$$\phi'(0) = \partial_{(\nu,q)}\mathcal{L}(\nu(0), q(0), \bar{\mu})\,(\delta\nu_n + \gamma'(0), \delta q_n) = \partial_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})\,(\delta\nu_n, \delta q_n)\,.$$

Using the first order necessary optimality condition

$$\partial_\nu \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu}) = 0 \tag{3.13}$$

as well as condition (3.11) implying $\delta q_n = 0$ whenever $\alpha\bar{q} + B^*\bar{z} \neq 0$, we find

$$\partial_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})\,(\delta\nu_n, \delta q_n) = \partial_q \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})\delta q_n = \int_0^1 \bar{\nu}\langle \alpha\bar{q} + B^*\bar{z}, \delta q_n\rangle = 0.$$

Hence, $\phi'(0) = 0$. Therefore, the second order optimality condition has to hold, i.e.

$$\begin{aligned} 0 \leq \phi''(0) &= \partial^2_{(\nu,q)}\mathcal{L}(\nu(0), q(0), \bar{\mu})[\delta\nu_n, \delta q_n]^2 + \partial_{(\nu,q)}\mathcal{L}(\nu(0), q(0), \bar{\mu})\,(\gamma''(0), 0) \\ &= \partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu_n, \delta q_n]^2 + \gamma''(0)\partial_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})\,(1, 0) \\ &= \partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu_n, \delta q_n]^2, \end{aligned}$$

where we have used again (3.13). Since δq_n converges pointwise almost everywhere in $I \times \omega$ and $|\delta q_n| \leq |\delta q|$, the dominated convergence theorem implies $\delta q_n \to \delta q$ in $L^2(I \times \omega)$. Thus, $\delta\nu_n \to \delta\nu$ due to $g'(\bar{\nu}, \bar{q})(\delta\nu, \delta q) = 0$. Hence,

$$0 \leq \lim_{n\to\infty} \partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu_n, \delta q_n]^2 = \partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu, \delta q]^2,$$

where we have used continuity of $(\delta\nu, \delta q) \mapsto \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu, \delta q]^2$ on $\mathbb{R} \times L^2(I \times \omega)$, completing the proof. □

The proof of the second order necessary conditions relied on the construction of feasible points. Using very similar arguments, we obtain the following result on the existence of feasible controls for perturbed time transformations. This result is needed for later reference in Section 3.2.3. In the absence of control constraints, it directly follows from the linearized Slater condition and the implicit function theorem. However, for the problem with control constraints, we have to argue as for the second order necessary conditions. We say that $q \in Q_{ad}(0,1)$ satisfies the non-triviality condition, if

$$|\{\, (t,x) \in I \times \omega \colon q_a(x) < q(t,x) < q_b(x),\, B^*z(t,x) \neq 0 \,\}| > 0, \tag{3.14}$$

where z is the solution to the adjoint state equation with terminal value $G'(i_1 S(\nu,q))^*$. We will see that (3.14) is satisfied in many situations; cf. Assumption 3.3 and Remark 3.17.

Proposition 3.12. *Let $(\nu,q) \in \mathbb{R} \times Q_{ad}(0,1)$. Suppose that Assumption 3.1 and (3.14) hold. Then there exists $\varepsilon > 0$ such that for all $\nu' \in (\nu-\varepsilon, \nu+\varepsilon)$ there is an admissible control $q(\nu') \in Q_{ad}(0,1)$ satisfying $g(\nu, q(\nu)) = 0$. Moreover, we have*

$$\|q(\nu') - q\|_{L^2(I\times\omega)} = \mathcal{O}(|\nu' - \nu|) \quad \text{as } \nu' \to \nu.$$

Proof. Due to the non-triviality condition (3.14), there exists a subset $\Lambda \subset I \times \omega$ with non-trivial measure such that the control constraints are (strictly) not active and $\mathbb{1}_\Lambda B^*z \neq 0$. Taking $\delta q = \mathbb{1}_\Lambda B^*z$ yields

$$\partial_q g(\nu,q)\delta q = \nu (B^*z, \delta q)_{L^2(I\times\omega)} = \nu \|B^*z\|^2_{L^2(\Lambda)} > 0$$

using the adjoint state representation of g'. Moreover, $\alpha q + B^*z = 0$ on Λ. Hence, δq satisfies the sign condition on Λ. It also satisfies the sign condition on the complement of Λ, because there it is identical zero. Then we define $\delta\nu = -\partial_q g(\nu,q)\delta q / \partial_\nu g(\nu,q)$, which is well-defined since $\partial_\nu g(\nu,q) \neq 0$ due to the supposition.

As in the proof of Theorem 3.10, we define a truncation of δq by

$$\delta q_N = \begin{cases} 0 & \text{if } q_a < q < q_a + 1/N \text{ or } q_b - 1/N < q < q_b, \\ \max\{-N, \min\{N, \delta q\}\} & \text{else,} \end{cases}$$

almost everywhere in $I \times \omega$; cf. also Figure 3.1. Moreover, we set

$$\delta\nu_N = -\partial_q g(\nu,q)\delta q_N / \partial_\nu g(\nu,q).$$

Due to the choice of δq, we have $\delta\nu_N \neq 0$ for N sufficiently large as well as $g'(\nu,q)(\delta\nu_N, \delta q_N) = 0$. According to Proposition 3.11 there are $\varepsilon > 0$ and a C^2-function $\gamma\colon (-\varepsilon,\varepsilon) \to \mathbb{R}$ such that for all $\theta \in (-\varepsilon,\varepsilon)$ we have $g(\nu(\theta), q(\theta)) = 0$, where $\nu(\theta) = \nu + \theta\delta\nu_N + \gamma(\theta)$ and $q(\theta) = q + \theta\delta q_N$. Furthermore, since δq is nonvanishing only where the control is strictly nonactive, we infer $q(\theta) \in Q_{ad}(0,1)$ for θ sufficiently small. Using that $\gamma(0) = \gamma'(0) = 0$, we obtain

$$\nu(\theta) = \nu(0) + \nu'(0)\theta + \mathcal{O}(\theta^2) = \nu + \delta\nu_N\theta + \mathcal{O}(\theta^2).$$

Moreover, from the equality above, we deduce that

$$|\theta| = |\nu(\theta) - \nu||\delta\nu_N + \mathcal{O}(\theta^2)/\theta|^{-1} \leq c|\nu(\theta) - \nu|, \quad \theta \in (\nu-\varepsilon, \nu+\varepsilon).$$

Hence,

$$\|q(\theta) - q\| = \|\theta\delta q_N\| \leq c|\nu(\theta) - \nu|.$$

Taking θ close to zero, yields the assertion for $\nu' \coloneqq \nu(\theta)$ and $q(\nu') \coloneqq q(\theta)$. □

3.2.2. Second order sufficient optimality conditions

Next, we postulate "minimal-gap" second order sufficient conditions, which result from replacing the inequality in (3.10) by a strict inequality.

Theorem 3.13. *Suppose $(\bar\nu, \bar q, \bar\mu) \in \mathbb{R}_+ \times Q_{ad}(0,1) \times \mathbb{R}_+$ satisfies the first order necessary conditions of Lemma 3.1 and the second order sufficient condition*

$$\partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)[\delta\nu, \delta q]^2 > 0 \quad \text{for all } (\delta\nu, \delta q) \in C_{(\bar\nu,\bar q)} \setminus \{(0,0)\}. \tag{3.15}$$

Then there exist $\varepsilon > 0$ and $c > 0$ such that for every admissible pair $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ the quadratic growth condition

$$j(\bar\nu, \bar q) + \frac{c}{2}|\nu - \bar\nu|^2 + \frac{c}{2}\|q - \bar q\|^2_{L^2(I\times\omega)} \le j(\nu, q), \tag{3.16}$$

is satisfied if $|\nu - \bar\nu| + \|q - \bar q\|_{L^2(I\times\omega)} \le \varepsilon$.

Proof. We closely follow the ideas of [34, Theorem 4.13] for the semilinear heat equation; cf. also [30, Section 4] for pointwise state constraints. Assume by contradiction that for all integer n there exist a time transformation $\nu_n \in \mathbb{R}_+$ and a control $q_n \in Q_{ad}(0,1)$ such that the corresponding state $S(\nu_n, q_n)$ is feasible and

$$\|(\nu_n - \bar\nu, q_n - \bar q)\| < \frac{1}{n}, \qquad j(\bar\nu, \bar q) + \frac{1}{2n}\|(\nu_n - \bar\nu, q_n - \bar q)\|^2 > j(\nu_n, q_n). \tag{3.17}$$

Set $\bar\chi = (\bar\nu, \bar q)$ and $\chi_n = (\nu_n, q_n)$. Define $\rho_n = \|(\nu_n - \bar\nu, q_n - \bar q)\|$ and

$$v_n = (v_n^\nu, v_n^q) = \frac{1}{\rho_n}(\chi_n - \bar\chi).$$

We may assume w.l.o.g. that $v_n^\nu \to v^\nu$ in $\mathbb{R}$ and $v_n^q \rightharpoonup v^q$ in $L^2(I\times\omega)$.

Step 1: $\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v = 0$. Due to the variational inequality (3.3) we have

$$\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v = \lim_{n\to\infty} \partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v_n \ge 0.$$

For the reverse inequality, we observe

$$\begin{aligned}\mathcal{L}(\bar\chi, \bar\mu) + \frac{1}{2n}\|(\nu_n - \bar\nu, q_n - \bar q)\|^2 &= j(\bar\chi) + \frac{1}{2n}\|(\nu_n - \bar\nu, q_n - \bar q)\|^2 \\ &> j(\chi_n) \ge \mathcal{L}(\chi_n, \bar\mu),\end{aligned} \tag{3.18}$$

since χ_n is feasible and $\bar\mu > 0$. The Taylor expansion yields

$$\mathcal{L}(\chi_n, \bar\mu) = \mathcal{L}(\bar\chi, \bar\mu) + \rho_n \partial_\chi \mathcal{L}(\check\chi_n, \bar\mu)v_n, \tag{3.19}$$

for some appropriate $\check\chi_n \in \mathbb{R} \times L^2(I\times\omega)$. Moreover, $\check\chi_n \to \bar\chi$ as $n \to \infty$. Combining (3.18) and (3.19) we arrive at

$$\partial_\chi \mathcal{L}(\check\chi_n, \bar\mu)v_n \le \frac{1}{2n\rho_n}\|(\nu_n - \bar\nu, q_n - \bar q)\|^2 = \frac{1}{2n}\|(\nu_n - \bar\nu, q_n - \bar q)\| < \frac{1}{2n^2}.$$

Using continuity of $\partial_\chi \mathcal{L}$ we infer

$$\begin{aligned}\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v &= \lim_{n\to\infty} \partial_\chi \mathcal{L}(\chi,\bar\mu)v_n \\ &\le \limsup_{n\to\infty} \partial_\chi \mathcal{L}(\check\chi_n,\bar\mu)v_n + \limsup_{n\to\infty} \left[\partial_\chi \mathcal{L}(\bar\chi,\bar\mu) - \partial_\chi \mathcal{L}(\check\chi_n,\bar\mu)\right] v_n \\ &\le \limsup_{n\to\infty} \frac{1}{2n^2} \le 0.\end{aligned}$$

Step 2: $v \in C_{(\bar\nu,\bar q)}$. First, because the set

$$\left\{ \delta q \in L^2(I\times\omega) \;\middle|\; \begin{array}{l} \delta q \le 0 \text{ if } \bar q(t,x) = q_b(x) \\ \delta q \ge 0 \text{ if } \bar q(t,x) = q_a(x) \end{array} \right\},$$

is closed and convex, it is in particular weakly closed, hence the weak limit satisfies $v^q(t,x) \le 0$, if $\bar q(t,x) = q_b(x)$, and $v^q(t,x) \ge 0$, if $\bar q(t,x) = q_a(x)$. For this reason, (3.6) implies

$$\int_0^1 \int_\omega \bar\nu(\alpha\bar q + B^*\bar z)v^q \,\mathrm{d}x\,\mathrm{d}t = \int_0^1 \int_\omega \bar\nu|(\alpha\bar q + B^*\bar z)v^q| \,\mathrm{d}x\,\mathrm{d}t.$$

Moreover, due to $\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v = 0$ and the first order necessary condition $\partial_\nu \mathcal{L}(\bar\chi,\bar\mu) = 0$ we have the equality

$$0 = \partial_q \mathcal{L}(\bar\chi,\bar\mu)v^q = \int_0^1 \bar\nu(\alpha\bar q + B^*\bar z, v^q)_{L^2(\omega)} \,\mathrm{d}t = \int_0^1 \int_\omega \bar\nu|(\alpha\bar q + B^*\bar z)v^q| \,\mathrm{d}x\,\mathrm{d}t.$$

Hence, $v^q = 0$, if $\alpha\bar q + B^*\bar z \ne 0$ almost everywhere in $I\times\omega$, and v^q satisfies the sign condition (3.11) as well. Furthermore, according to the assumption on G we have

$$g'(\bar\chi)v = \lim_{n\to\infty} \frac{1}{\rho_n}\left(g(\bar\chi + \rho_n v_n) - g(\bar\chi)\right) = \lim_{n\to\infty} \frac{1}{\rho_n} g(\chi_n) \le 0,$$

since $g(\bar\chi) = 0$ and $g(\chi_n) \le 0$. Similarly, using (3.17), we find

$$j'(\bar\chi)v = \lim_{n\to\infty} \frac{1}{\rho_n}\left(j(\bar\chi + \rho_n v_n) - j(\bar\chi)\right) \le \lim_{n\to\infty} \frac{1}{2n^2} = 0.$$

Due to $\bar\mu > 0$ and $\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v = j'(\bar\chi)v + \bar\mu\, g'(\bar\chi)v = 0$ we conclude $g'(\bar\chi)v = 0$. In summary, we have proved that $v \in C_{(\bar\nu,\bar q)}$.

Step 3: $v = 0$. Using again Taylor expansion we have

$$\mathcal{L}(\chi_n,\bar\mu) = \mathcal{L}(\bar\chi,\bar\mu) + \rho_n \partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v_n + \frac{\rho_n^2}{2}\partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2, \tag{3.20}$$

with intermediate points $\check\chi_n \in \mathbb{R}\times L^2(I\times\omega)$. Plugging (3.20) into (3.18) and dividing by ρ_n^2 we obtain

$$\frac{1}{\rho_n}\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v_n + \frac{1}{2}\partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2 \le \frac{1}{n}.$$

Hence, using Corollary 3.9 and weak lower semicontinuity of $j''(\bar\chi)$, it follows that

$$\begin{aligned}\partial_\chi^2 \mathcal{L}(\bar\chi,\bar\mu)v^2 &\le \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi,\bar\mu)v_n^2 \\ &\le \limsup_{n\to\infty} \partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2 + \limsup_{n\to\infty} \left[\partial_\chi^2 \mathcal{L}(\bar\chi,\bar\mu) - \partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)\right] v_n^2 \\ &\le \limsup_{n\to\infty} \left[\frac{2}{\rho_n}\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v_n + \partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2\right] \\ &\quad + \limsup_{n\to\infty} \frac{-2}{\rho_n}\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v_n \le 0,\end{aligned} \tag{3.21}$$

where we have used that $\partial_\chi \mathcal{L}(\bar\chi, \bar\mu) v_n \geq 0$ for all $n \in \mathbb{N}$ in the last step. According to the supposition, this is only possible if $v = 0$.

Step 4: Final contradiction. Since $\|(v_n^\nu, v_n^q)\| = 1$ and $v_n^\nu \to 0$ we finally obtain

$$0 < \alpha\bar\nu = \alpha\bar\nu \liminf_{n\to\infty} \|(v_n^\nu, v_n^q)\|^2 = \liminf_{n\to\infty} \alpha\bar\nu \int_0^1 \|v_n^q(t)\|^2_{L^2(\omega)} \,\mathrm{d}t.$$

Using the specific structure of j'', we see that

$$\liminf_{n\to\infty} \alpha\bar\nu \int_0^1 \|v_n^q(t)\|^2_{L^2(\omega)} \,\mathrm{d}t = \liminf_{n\to\infty} j''(\bar\chi) v_n^2.$$

Hence, employing Corollary 3.9, we conclude that

$$\begin{aligned} 0 &< \liminf_{n\to\infty} j''(\bar\chi) v_n^2 \leq \liminf_{n\to\infty} j''(\bar\chi) v_n^2 + \bar\mu \liminf_{n\to\infty} g''(\bar\chi) v_n^2 \\ &\leq \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi, \bar\mu) v_n^2 \leq 0, \end{aligned}$$

where we have used again (3.21) in the last inequality, yielding a contradiction. □

The second order sufficient condition (3.15) and the quadratic growth condition (3.16) will form the basis of the a priori discretization error estimates in Chapter 5. Last, we note that for the given objective functional coercivity of $\partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)$ is equivalent to the seemingly weaker positivity condition, as already observed for semilinear parabolic PDEs in [34].

Theorem 3.14. *Let* $(\bar\nu, \bar q, \bar\mu) \in \mathbb{R}_+ \times Q_{ad}(0,1) \times \mathbb{R}_+$. *The* positivity condition (3.15) *is equivalent to the* coercivity condition

$$\exists \gamma > 0 \colon \partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)[\delta\nu, \delta q]^2 \geq \gamma \left(|\delta\nu|^2 + \|\delta q\|^2_{L^2(I\times\omega)} \right) \quad \text{for all } (\delta\nu, \delta q) \in C_{(\bar\nu,\bar q)}. \tag{3.22}$$

Proof. This result can be proved along the lines of the proof of [34, Theorem 4.11], where we in particular use that $\delta\nu$ is from a finite dimensional space. The proof is given for convenience. Obviously, the condition of coercivity implies the positivity condition. To prove the reverse implication, we set

$$\gamma := \inf \left\{ \partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)[\delta\nu, \delta q]^2 \colon (\delta\nu, \delta q) \in C_{(\bar\nu,\bar q)}, \|(\delta\nu, \delta q)\| = 1 \right\}.$$

Due to the assumptions on j and g, the infimum exists and is nonnegative. Let $(\delta\nu_n, \delta q_n) \in C_{(\bar\nu,\bar q)}$ be a minimizing sequence. Without restriction we may assume $\delta\nu_n \to \delta\nu$ in $\mathbb{R}$ and $\delta q_n \rightharpoonup \delta q$ in $L^2(I\times\omega)$. Now, we distinguish two cases:

Case: $(\delta\nu, \delta q) = (0,0)$. From $\delta\nu_n \to 0$ in $\mathbb{R}$ and $\delta q_n \rightharpoonup 0$ in $L^2(I\times\omega)$, we conclude

$$\liminf_{n\to\infty} j''(\bar\nu, \bar q)[\delta\nu_n, \delta q_n]^2 = \alpha\bar\nu \liminf_{n\to\infty} \int_0^1 \|\delta q_n\|^2_{L^2(\omega)} = \alpha\bar\nu \liminf_{n\to\infty} \|(\delta\nu_n, \delta q_n)\|^2 = \alpha\bar\nu,$$

where we have used $\|(\delta\nu_n, \delta q_n)\| = 1$ in the last step. Moreover, due to Corollary 3.9, we have

$$\liminf_{n\to\infty} g''(\bar\nu, \bar q)(\delta\nu_n, \delta q_n) \geq 0.$$

Since $\bar\mu > 0$, we conclude that

$$\gamma = \liminf_{n\to\infty} \partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)[\delta\nu_n, \delta q_n]^2 \geq \liminf_{n\to\infty} \alpha \int_0^1 \bar\nu \|\delta q_n\|^2_{L^2(\omega)} = \alpha\bar\nu > 0.$$

Case: $(\delta\nu, \delta q) \neq (0,0)$. Using the same arguments as before for j'' and Corollary 3.9 we find

$$\gamma = \liminf_{n\to\infty} \partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)[\delta\nu_n, \delta q_n]^2 \geq \partial^2_{(\nu,q)} \mathcal{L}(\bar\nu, \bar q, \bar\mu)[\delta\nu, \delta q]^2 > 0,$$

due to the positivity condition. In both cases we proved that $\gamma > 0$. Thus, the coercivity condition holds. □

3.2.3. Reduction to a scalar condition

In general it seems to be difficult to verify whether a second order sufficient optimality condition is satisfied for a given a problem – both theoretically and numerically. However, for the problem under consideration, we will provide a scalar condition that is equivalent to the second order sufficient optimality condition of Theorem 3.13; cf. [82] for a similar approach for time-optimal control of ODEs. The idea is based on an infinite dimensional version of the Schur complement. We suppose that $G''(u)[\cdot,\cdot]$ is positive semi-definite for all $u \in H$. Recall that positive semi-definiteness implies weak lower semicontinuity, see, e.g, [61, Proposition 3.2], so this a strengthening of Assumption 3.2.

In order to keep the presentation of this section simple, we impose additional assumptions, which will turn out to be fulfilled in most situations. First, if the critical cone is trivial, i.e. $C_{(\bar\nu,\bar q)} = \{0\}$, the condition (3.15) is vacuously true. Note that this case typically corresponds to a bang-bang control. Similarly, to avoid other degenerate cases, we impose the additional assumption:

Assumption 3.3. We assume that the *strict complementarity condition*

$$|\{(t,x) \in I \times \omega \colon \bar q(t,x) \in \{q_a(x), q_b(x)\},\ \alpha\bar q(t,x) + B^*\bar z(t,x) = 0\}| = 0, \tag{3.23}$$

and the *non-triviality condition*

$$|\{(t,x) \in I \times \omega \colon q_a(x) < \bar q(t,x) < q_b(x),\ B^*\bar z(t,x) \neq 0\}| > 0, \tag{3.24}$$

hold, where $|\cdot|$ denotes the product-measure associated with $I \times \omega$.

Under Assumption 3.3 the critical cone $C_{(\bar\nu,\bar q)}$ is a linear space, which contains elements of the form $(\delta\nu, \delta q)$ with $\delta\nu \neq 0$. In the following we show that if the strict complementarity condition (3.23) holds, then the non-triviality condition (3.24) is equivalent to $C_{(\bar\nu,\bar q)} \neq \{0\}$. Moreover, we prove that under an appropriate approximate controllability assumption on the pair (A, B) strict complementarity holds.

Proposition 3.15. *Suppose that the strict complementarity condition* (3.23) *holds. Let* $(\bar\nu, \bar q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ *satisfy the qualified first order optimality conditions of Lemma 3.1. Then the non-triviality condition* (3.24) *is equivalent to* $C_{(\bar\nu,\bar q)} \neq \{0\}$.

Proof. Clearly, if the non-triviality condition (3.24) holds, then $C_{(\bar\nu,\bar q)} \neq \{0\}$. Hence, we only have to show the reverse implication. Suppose that (3.24) is violated. Let $(\delta\nu, \delta q) \in C_{(\bar\nu,\bar q)}$. Then we have $(B^*\bar z, \delta q)_{L^2(I\times\omega)} = 0$, because $\delta q = 0$ if $\alpha\bar q + B^*\bar z \neq 0$ (due to the sign condition (3.11)), $B^*\bar z = 0$ if $q_a < \bar q < q_b$ (because (3.24) is violated), and the remaining case $\bar q \in \{q_a, q_b\}$ and $\alpha\bar q + B^*\bar z = 0$ has zero measure (due to the strict complementarity condition (3.23)). Hence, using the condition $g'(\bar\nu, \bar q)(\delta\nu, \delta q) = 0$ from the critical cone and $\partial_\nu g(\bar\nu, \bar q) \neq 0$ (see Assumption 3.1 and Proposition 3.2), we infer that $\delta\nu = 0$. Thus, we conclude that $C_{(\bar\nu,\bar q)} = \{0\}$. □

In the case that $\bar{q}$ is bang-bang, the non-triviality condition (3.24) is clearly violated. Hence, in view of Proposition 3.15, if strict complementarity and qualified first order optimality conditions hold, and $\bar{q}$ is bang-bang, then the critical cone is trivial.

Proposition 3.16. *Consider the case of purely time-dependent controls, i.e.* $B\colon \mathbb{R}^{N_c} \to V^*$, $Bq = \sum_{i=1}^{N_c} q_i e_i$. *Suppose that the solution* $z \in W(0,1)$ *to the adjoint state equation with terminal state* $z_1 \in H$ *and time transformation* $\nu > 0$, *i.e.*

$$-\partial_t z + \nu A^* z = 0, \quad z(1) = z_1,$$

satisfies a backwards uniqueness property, i.e.

$$B^* z \equiv 0 \text{ on } I_0 \times \omega_0 \quad \Rightarrow \quad z = 0 \quad \text{for all } I_0 \times \omega_0 \subseteq I \times \omega. \tag{3.25}$$

Moreover, suppose that

$$t \mapsto z(t) \text{ constant} \quad \Rightarrow \quad z = 0. \tag{3.26}$$

Let $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ *satisfy the qualified first order optimality conditions of Lemma 3.1. Then the strict complementarity condition* (3.23) *holds.*

Proof. If (3.23) is violated, then $B_i^* \bar{z} = -\alpha q_{a,i}$ or $B_i^* \bar{z} = -\alpha q_{b,i}$ on a subset $I_0 \subseteq I$ such that $|I_0| \neq 0$ for some $i \in \{1, 2, \dots, N_c\}$. Without restriction suppose that $B_i^* \bar{z} = -\alpha q_{a,i}$ on I_0. Hence, the mapping $t \mapsto \phi(t) := B_i^* \bar{z}(t)$ is constant on I_0. Moreover, analyticity of the semigroup generated by $-A^*$ implies that ϕ is also analytic. Thus, ϕ is constant on I, which means that $\phi' = \partial_t B_i^* \bar{z}$ vanishes on I. We set $v = \partial_t \bar{z}$. Since $\bar{z} \in C^1(I; V)$, we observe that v also solves a backwards parabolic equation on $(0, 1-\varepsilon)$ with terminal value $v(1-\varepsilon) = \partial_t \bar{z}(1-\varepsilon) \in V$ for any small $\varepsilon > 0$. Because of $B_i^* v = \partial_t B_i^* \bar{z} = 0$ on I, from the backwards uniqueness property (3.25), we deduce that $v = 0$ on the time interval $(0, 1-\varepsilon)$, which in turn implies that the adjoint state $\bar{z}$ is constant on $(0, 1-\varepsilon)$. Letting $\varepsilon \to 0$, we see that $\bar{z}$ is constant on I. Thus, the second supposition (3.26) implies $\bar{z} = 0$. However, this contradicts the optimality conditions from Lemma 3.1. □

Remark 3.17. We comment on situations, where the suppositions (3.25) and (3.26) of the preceding proposition are guaranteed to hold:

(i) The backwards uniqueness property (3.25) is equivalent to the assumption that for each $B_i\colon \mathbb{R} \to V^*$, $B_i q = q e_i$, the pair (A, B_i) is approximately controllable; see [150, Definition 11.1.1]. Note that in the context of optimal control for ordinary differential equations this property would correspond to normality of the pair (A, B); see, e.g., [74, Section II.16] or [112, Section III.3]. Employing the fact that $t \mapsto B^* z(t)$ is analytic and [150, Theorem 11.2.1, Definition 6.1.1], we infer that (3.25) holds.

(ii) If the Gårding inequality (2.7) holds with $\omega_0 = 0$ (e.g. if $A = -\Delta$ with homogeneous Dirichlet boundary conditions), then the semigroup generated by $-A$ is uniformly exponentially stable in H, i.e. $\|\mathrm{e}^{-tA}\|_{\mathcal{L}(H)} \leq \mathrm{e}^{-\rho t}$ for some $\rho > 0$ and all $t > 0$; see Proposition A.21. Additionally, according to [128, Corollary 1.10.6] we have $\mathrm{e}^{-\cdot A^*} = (\mathrm{e}^{-\cdot A})^*$. Considering the canonical extension of z to $(-\infty, 0)$, exponential stability yields $\lim_{t \to -\infty} z(t) = 0$. Thus, if $t \mapsto z(t)$ is constant, we conclude that $z = 0$ due to analyticity of $t \mapsto z(t)$, i.e. (3.26) holds.

(iii) In certain situations, we can dispense with the controllability assumption. Let the control bounds satisfy $q_a < 0 < q_b$ and Gårding's inequality (2.7) hold with $\omega_0 = 0$. Suppose that the strict complementarity condition (3.23) is violated. Hence, there is an index

$i \in \{1, 2, \dots, N_c\}$ such that $(-1/\alpha)B_i^*\bar{z}(t) = \bar{q}_i(t)$ equals one of the control bounds on a subset of I. First, we have $|B_i^*\bar{z}(t)| \le \|e^i\|_{V^*}\|\bar{z}(t)\|_V$. Moreover, exponential stability as before yields

$$\|\bar{z}(t)\|_V \le \|\mathrm{e}^{-(1-t-\varepsilon)A}A^{1/2}\|_{\mathcal{L}(H)}\|\mathrm{e}^{-\varepsilon A}\bar{z}(1)\|_H \to 0$$

as $t \to -\infty$ (for some small $\varepsilon > 0$), where we have considered the canonical extension of $\bar{z}$ to $(-\infty, 0)$. Thus, analyticity implies $B_i^*\bar{z}(t) = 0$ on I, which contradicts the supposition $q_a < 0 < q_b$. Hence, the strict complementarity condition (3.23) is satisfied. Note that we have used the coupling between $\bar{q}$ and $\bar{z}$ on I only, not on $(-\infty, 0)$.

If Assumption 3.3 holds, the critical cone $C_{(\bar{\nu},\bar{q})}$ is a linear space. It consists exactly of the elements $(\delta\nu, \delta q)$ with $\delta\nu \in \mathbb{R}$, $\delta q \in C_{\bar{q}}$, and $\partial_q g(\bar{\nu}, \bar{q})\delta q + \partial_\nu g(\bar{\nu}, \bar{q})\delta\nu = 0$, where

$$C_{\bar{q}} := \{\, \delta q \in L^2(I \times \omega) \colon \delta q(t,x) = 0 \text{ if } \alpha\bar{q}(t,x) + B^*\bar{z}(t,x) \neq 0 \,\}.$$

For ease of presentation, we sometimes abbreviate the arguments $(\bar{\nu}, \bar{q})$ and simply write $\bar{\chi}$ in the following. Under Assumption 3.3 we now prove that the second order sufficient optimality condition is equivalent to a scalar condition.

Lemma 3.18. *Let $(\bar{\nu}, \bar{q}, \bar{\mu}) \in \mathbb{R}_+ \times Q_{ad}(0,1) \times \mathbb{R}_+$. Assume that Assumption 3.3 holds and that $G''(u)[\cdot,\cdot]$ is positive semi-definite for all $u \in H$. Then, the second order sufficient optimality condition of Theorem 3.13 is equivalent to*

$$\bar{\gamma} := \partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[1, \delta\bar{q}]^2 > 0, \tag{3.27}$$

where $(\delta\bar{q}, \delta\bar{\mu}) \in C_{\bar{q}} \times \mathbb{R}$ is the unique solution of the linear system

$$\begin{aligned} \partial_q^2\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\bar{q}, \delta q] + \delta\bar{\mu}\,\partial_q g(\bar{\nu}, \bar{q})\delta q &= -\partial_\nu\partial_q\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[1, \delta q], \quad \delta q \in C_{\bar{q}}, \\ \partial_q g(\bar{\nu}, \bar{q})\delta\bar{q} &= -\partial_\nu g(\bar{\nu}, \bar{q}). \end{aligned} \tag{3.28}$$

Proof. We start by proving that (3.27) implies the second order sufficient optimality condition. Let $(\delta\nu, \delta q) \in C_{(\bar{\nu},\bar{q})}$. We distinguish two cases for $\delta\nu$: If $\delta\nu = 0$, we use the fact that the second derivative of g with respect to q has the form $\partial_q^2 g(\bar{\chi})[\delta q]^2 = G''(\bar{u}(1))[i_1\partial_q S(\bar{\chi})\delta q]^2$ to obtain

$$\partial_q^2\mathcal{L}(\bar{\chi}, \bar{\mu})[\delta q]^2 \ge \partial_q^2 j(\bar{\chi})[\delta q]^2 = \alpha\bar{\nu}\|\delta q\|^2_{L^2(I\times\omega)}, \tag{3.29}$$

which immediately implies (3.15).

Now, consider the case $\delta\nu \neq 0$. Since the expression on the left in (3.15) is bi-linear in $\delta\nu$, and the critical cone $C_{(\bar{\nu},\bar{q})}$ is linear, it suffices to consider the case $\delta\nu = 1$. By minimizing the expression on the left for admissible δq (such that $(1, \delta q) \in C_{(\bar{\nu},\bar{q})}$), writing out the second derivative in terms of the partial derivatives and dropping constant terms, we arrive at the following minimization problem:

$$\inf_{\delta q \in C_{\bar{q}}} \frac{1}{2}\partial_q^2\mathcal{L}(\bar{\chi}, \bar{\mu})[\delta q]^2 + \partial_\nu\partial_q\mathcal{L}(\bar{\chi}, \bar{\mu})[1, \delta q] \quad \text{subject to} \quad \partial_q g(\bar{\chi})\delta q = -\partial_\nu g(\bar{\chi}). \tag{3.30}$$

Since $(1, \delta q) \in C_{(\bar{\nu},\bar{q})}$, we have $\partial_q g(\bar{\chi})\delta q = -\partial_\nu g(\bar{\chi})$. Hence, problem (3.30) has admissible points, and we easily verify existence of a minimizer using the direct method. Moreover, due to Assumption 3.3 (or using the first order optimality condition $\partial_\nu g(\bar{\chi}) \neq 0$ and linearity of $C_{\bar{q}}$), we have $\partial_q g(\bar{\chi})C_{\bar{q}} = \mathbb{R}$, which means that a constraint qualification condition (see, e.g., [166]) is fulfilled. Thus, we obtain the necessary and sufficient optimality conditions of

the convex problem (3.30) in the form (3.28). Hence, for the positivity condition (3.15) we only have to require that $\bar{\gamma} > 0$, which guarantees

$$\partial^2_{(\nu,q)}\mathcal{L}(\bar{\chi},\bar{\mu})[1,\delta q]^2 \geq \partial^2_{(\nu,q)}\mathcal{L}(\bar{\chi},\bar{\mu})[1,\delta\bar{q}]^2 = \bar{\gamma} > 0, \tag{3.31}$$

for any δq with $(1,\delta q) \in C_{(\bar{\nu},\bar{q})}$, where $\delta\bar{q}$ is the solution to (3.28).

Last, we prove that the second order sufficient optimality condition implies (3.27). As already observed, (3.28) possesses a unique solution $(\delta\bar{q},\delta\bar{\mu})$ and $(1,\delta\bar{q}) \in C_{(\bar{\nu},\bar{q})}$. From the second order sufficient optimality condition we obtain $\bar{\gamma} > 0$. □

Remark 3.19. If the non-triviality condition (3.24) is violated, then $C_{(\bar{\nu},\bar{q})}$ contains elements of the form $(0,\delta q)$, only. Hence, the second order sufficient condition is always satisfied for positive semi-definite $G''(u)[\cdot,\cdot]$. Note that in this degenerate case, the system (3.28) does not posses a solution, because the second equality cannot be satisfied for $\delta\bar{q} \in C_{\bar{q}}$.

While $\bar{\gamma} > 0$ implies the second order sufficient conditions from Theorem 3.13, it does not represent the coercivity constant for the second derivative of the Lagrange function as obtained in Theorem 3.14. Instead, we can derive a lower bound on the coercivity constant in terms of $\bar{\gamma}$, which also depends explicitly on $\alpha > 0$.

Proposition 3.20. *Adapt the assumptions of Lemma 3.18 and suppose that $\bar{\gamma} > 0$. Then, the coercivity constant γ from Theorem 3.14 is bounded from below by*

$$\gamma \geq (\bar{\gamma}/3)\min\{\,\alpha\bar{\nu}/(\bar{\gamma}+c_1),1\,\},$$

where c_1 depends on the optimal solution and on α.

Proof. By replacing δq with $\delta q/\delta\nu$ in (3.31) and using linearity we directly obtain

$$\partial^2_{(\nu,q)}\mathcal{L}(\bar{\chi},\bar{\mu})[\delta\nu,\delta q]^2 \geq \bar{\gamma}|\delta\nu|^2 \quad \text{for all } (\delta\nu,\delta q) \in C_{(\bar{\nu},\bar{q})}.$$

Furthermore, by using the coercivity of $\partial^2_q\mathcal{L}(\bar{\chi},\bar{\mu})$ with constant $\alpha\bar{\nu}$ and straightforward estimates (using Young's inequality), we can derive that

$$\partial^2_{(\nu,q)}\mathcal{L}(\bar{\chi},\bar{\mu})[\delta\nu,\delta q]^2 \geq \frac{\alpha\bar{\nu}}{2}\|\delta q\|^2_{L^2(I\times\omega)} - c_1|\delta\nu|^2 \quad \text{for all } (\delta\nu,\delta q),$$

where $c_1 = \Big(|\partial^2_\nu\mathcal{L}(\bar{\chi},\bar{\mu})| + 2\|\partial_\nu\partial_q\mathcal{L}(\bar{\chi},\bar{\mu})\|^2/(\alpha\bar{\nu})\Big)$. By taking a convex combination of $(1-\theta)$ times the former and θ times the latter estimate, where $\theta = (2/3)(\bar{\gamma}/(\bar{\gamma}+c_1))$, we arrive at

$$\begin{aligned}\partial^2_{(\nu,q)}\mathcal{L}(\bar{\chi},\bar{\mu})[\delta\nu,\delta q]^2 &\geq \left((1-\theta)\bar{\gamma} - \theta c_1\right)|\delta\nu|^2 + \frac{\theta\alpha}{2}\|\delta q\|^2_{L^2(I\times\omega)} \\ &\geq \frac{\bar{\gamma}}{3}\left(|\delta\nu|^2 + \frac{\alpha\bar{\nu}}{\bar{\gamma}+c_1}\|\delta q\|^2_{L^2(I\times\omega)}\right)\end{aligned}$$

for all $(\delta\nu,\delta q) \in C_{(\bar{\nu},\bar{q})}$. □

The scalar condition (3.27) still involves the solution of an infinite dimensional problem. However, the same calculation holds true for the discrete problem, which means that we can verify the SSC on the discrete level, by computing numerically the constant $\bar{\gamma}$ defined in Lemma 3.18; see Section 5.4. In this regard we also mention [137] on the numerical verification of second order optimality conditions.

We can also give a different interpretation of the scalar condition (3.27) in terms of the curvature of a certain value function with respect to ν. To understand the condition (3.27), we introduce the value function

$$V(\nu) = \min_{\substack{q \in Q_{ad}(0,1) \\ g(\nu,q)=0}} j(\nu, q) = \mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu)), \tag{3.32}$$

which is obtained by fixing an arbitrary time ν and resolving the resulting constraint optimization problem for the controls. For well posedness of $V(\cdot)$ in a neighborhood of $\bar{\nu}$ we have to argue that feasible points exist. However, this is the assertion of Proposition 3.12, provided that Assumption 3.3 holds.

The aim is to show that the value function V is differentiable. For the proof we rely on established arguments where we refer to [21, Section 5.1] and the review article [20]; cf. also [67, Proposition 3.16] and [81, Chapter 2]. We first study the stability of the optimal solution $\bar{q}(\nu)$ and the Lagrange multiplier $\bar{\mu}(\nu)$ with respect to ν associated with the minimization problem (3.32).

Proposition 3.21. *Let Assumption 3.3 hold and assume that $G''(u)[\cdot,\cdot]$ is positive semi-definite for all $u \in H$. There exists $\delta > 0$ such that for all $\nu_1, \nu_2 \in \mathbb{R}_+$ with $|\nu_i - \bar{\nu}| \leq \delta$, $i = 1, 2$, we have*

$$\|\bar{q}(\nu_1) - \bar{q}(\nu_2)\|_{L^2(I\times\omega)} + |\bar{\mu}(\nu_1) - \bar{\mu}(\nu_2)| = \mathcal{O}(|\nu_1 - \nu_2|) \quad \text{as } |\nu_1 - \nu_2| \to 0.$$

Proof. Step 1: Continuity. To begin with, consider the case $\nu_1 = \nu$ and $\nu_2 = \bar{\nu}$ for some given $\nu \in \mathbb{R}_+$. Let $\mathcal{L}(\nu, q, \mu) = j(\nu, q) + \mu g(\nu, q)$ denote the Lagrange function associated with the minimization problem (3.32). Since the mapping $q \mapsto i_1 S(\nu, q)$ is (affine) linear, positive semi-definiteness of $G''[\cdot,\cdot]$ yields $\partial_{qq} g(\nu, q)\delta q^2 \geq 0$. Therefore, we have

$$\alpha\bar{\nu}\|q - \bar{q}\|^2_{L^2(I\times\omega)} \leq \partial_{qq}\mathcal{L}(\bar{\nu}, q_\xi, \bar{\mu})[q - \bar{q}]^2$$

for any $q_\xi \in L^2(I \times \omega)$. Hence, Taylor expansion of $\mathcal{L}$ at $(\bar{\nu}, \bar{q}, \bar{\mu})$ implies

$$\begin{aligned}\frac{\alpha\bar{\nu}}{2}\|q - \bar{q}\|^2_{L^2(I\times\omega)} &\leq \mathcal{L}(\bar{\nu}, q, \bar{\mu}) - \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu}) - \partial_q\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})(q - \bar{q}) \\ &\leq j(\bar{\nu}, q) - j(\bar{\nu}, \bar{q}),\end{aligned}$$

for all $q \in Q_{ad}(0, 1)$ such that $g(\bar{\nu}, q) = 0$. Plugging the admissible control $q(\nu)$ from Proposition 3.12 into the inequality above and using Lipschitz continuity of j, we obtain

$$j(\nu, \bar{q}(\nu)) - j(\bar{\nu}, \bar{q}) \leq j(\nu, q(\nu)) - j(\bar{\nu}, \bar{q}) = \mathcal{O}(|\nu - \bar{\nu}|),$$

i.e. $\bar{q}(\nu)$ is an ε-optimal solution of order $\mathcal{O}(|\nu - \bar{\nu}|)$. Thus, [21, Proposition 4.41] with the quadratic growth condition of the unperturbed problem implies

$$\|\bar{q}(\nu) - \bar{q}\|_{L^2(I\times\omega)} = \mathcal{O}(|\nu - \bar{\nu}|^{1/2}) \quad \text{as } \nu \to \bar{\nu}.$$

Step 2: Uniqueness of $\bar{\mu}(\nu)$. Due to the non-triviality condition, the control $\bar{q} = \bar{q}(\bar{\nu})$ is not bang-bang and there exists a subset $\Lambda \subset I \times \omega$ such that $\bar{q} = (-1/\alpha)B^*\bar{z} \neq 0$ on Λ. Since $\nu \mapsto \bar{q}(\nu)$ is (Hölder) continuous at $\bar{\nu}$ in $L^2(I \times \omega)$, we have $\bar{q}(\nu) = -(1/\alpha)B^*\bar{z}(\nu) \neq 0$ on a (possibly smaller set) Λ for all ν close to $\bar{\nu}$ using the Theorem of Egorov. If μ is a different multiplier, then

$$\bar{q}(\nu) = (-1/\alpha)\mu B^* z = (-1/\alpha)\bar{\mu}(z)B^* z,$$

where z is the adjoint state with terminal value $G'(i_1 S(\nu, \bar{q}(\nu)))^*$. Hence, $\mu = \bar{\mu}(\nu)$.

Step 3: Hölder stability. Arguments similar as above yield

$$\begin{aligned}\frac{\alpha\nu_2}{2}&\|\bar{q}(\nu_1) - \bar{q}(\nu_2)\|^2_{L^2(I\times\omega)} \\ &\leq \mathcal{L}(\nu_2, \bar{q}(\nu_1), \bar{\mu}(\nu_2)) - \mathcal{L}(\nu_2, \bar{q}(\nu_2), \bar{\mu}(\nu_2)) - \partial_q \mathcal{L}(\nu_2, \bar{q}(\nu_2), \bar{\mu}(\nu_2))(\bar{q}(\nu_1) - \bar{q}(\nu_2)) \\ &\leq \mathcal{L}(\nu_2, \bar{q}(\nu_1), \bar{\mu}(\nu_2)) - \mathcal{L}(\nu_1, \bar{q}(\nu_1), \bar{\mu}(\nu_2)) + j(\nu_1, \bar{q}(\nu_1)) - j(\nu_2, \bar{q}(\nu_2)).\end{aligned}$$

Lipschitz continuity of j and g as well as boundedness of $\bar{\mu}(\nu_2)$ imply

$$\mathcal{L}(\nu_2, \bar{q}(\nu_1), \bar{\mu}(\nu_2)) - \mathcal{L}(\nu_1, \bar{q}(\nu_1), \bar{\mu}(\nu_2)) \leq c|\nu_1 - \nu_2|.$$

Moreover, since $\nu \mapsto \bar{q}(\nu)$ is continuous, the non-triviality condition (3.14) for $\bar{q}(\nu_2)$ used in Proposition 3.12 is satisfied for ν_2 sufficiently close to $\bar{\nu}$. Using the admissible control $q(\nu_1)$ from Proposition 3.12, we get

$$j(\nu_1, \bar{q}(\nu_1)) - j(\nu_2, \bar{q}(\nu_2)) \leq j(\nu_1, q(\nu_1)) - j(\nu_2, \bar{q}(\nu_2)) = \mathcal{O}(|\nu_1 - \nu_2|).$$

Due to [21, Proposition 4.41], we have

$$\|\bar{q}(\nu_1) - \bar{q}(\nu_2)\|_{L^2(I\times\omega)} = \mathcal{O}(|\nu_1 - \nu_2|^{1/2}).$$

Step 4: Lipschitz stability. Finally, we would like to apply [21, Theorem 4.51] with $\mathcal{G}(q, \nu) \coloneqq (g(\nu, q), q)$ and constraint $\mathcal{K} \coloneqq \{0\} \times Q_{ad}(0, 1)$, and verify its assumptions. The non-triviality condition yields $\partial_q g(\nu, \bar{q}(\nu))Q(0, 1) = \mathbb{R}$. Hence, Robinson's constraint qualification holds at the tuple $(\bar{q}(\nu), \nu)$. Moreover, upper Lipschitz continuity of the multifunction defined in [21, (4.116)], follows from surjectivity of $D\mathcal{G}(\bar{q}(\nu), \nu)$ and [21, Remark 4.45 (i)]. Last, the second order condition holds on $Q(0, 1)$ (not only on the approximate critical cone). Since the Lagrange multiplier $\bar{\mu}(\nu)$ is unique, the inequality [21, (4.127)] yields the assertion. □

In order to show that the value function is differentiable, we require the notion of polyhedricity that we will introduce next; see, e.g., [21, Definition 3.51].

Definition 3.22. Let K be a closed convex set of a Banach space Y. K is called *polyhedric* at $\bar{x} \in K$, if for any $v \in N_K(\bar{x})$ the identity

$$T_K(\bar{x}) \cap \ker v = \overline{R_K(\bar{x}) \cap \ker v}$$

holds, where N_K denotes the normal cone, T_K the tangent cone, and

$$R_K(\bar{x}) = \{\, \delta x \in Y \colon \exists t > 0 \text{ such that } \bar{x} + t\delta x \in K \,\}$$

the *radial cone* with $T_K(\bar{x}) = \overline{R_K(\bar{x})}$. It is called polyhedric, if it is polyhedric for all $\bar{x} \in K$.

The radial cone is also referred to as the cone of feasible directions; cf. [33]. In the case of pointwise control constraints, the set $Q_{ad}(0, 1)$ is polyhedric.

Proposition 3.23. *Let (ω, ϱ) be a finite measure space. If*

$$Q_{ad} \coloneqq \left\{ q \in L^2(\omega) \colon q_a \leq q \leq q_b \text{ a.e. in } \omega \right\} \subset L^\infty(\omega)$$

for $q_a, q_b \in L^\infty(\omega)$ with $q_a < q_b$ almost everywhere, then $Q_{ad}(0, 1)$ is polyhedric in $L^2(I \times \omega)$.

Proof. Due to $T_{Q_{ad}(0,1)}(\bar{q}) = \overline{R_{Q_{ad}(0,1)}(\bar{q})}$, the right-hand side in Definition 3.22 is automatically contained in the left-hand side. Let $v \in N_{Q_{ad}(0,1)}(\bar{q})$ and $\delta q \in T_{Q_{ad}(0,1)}(\bar{q}) \cap \ker v$. We define almost everywhere

$$\delta q_n = \begin{cases} 0 & \text{if } q_a < \bar{q} < q_a + 1/n \text{ or } q_b - 1/n < \bar{q} < q_b, \\ \max\{-n, \min\{n, \delta q\}\} & \text{else.} \end{cases}$$

Then $q_n \coloneqq \bar{q} + \rho_n \delta q_n$ for $\rho_n = \min\{1/n^2, (q_b - q_a)/n\}$ satisfies $q_a \le q_n \le q_b$ almost everywhere in $I \times \omega$. Hence, $\delta q_n \in R_{Q_{ad}(0,1)}(\bar{q})$. Let $\Lambda_+ \subseteq I \times \omega$ denote the subset where $v > 0$ and let $\Lambda_- \subseteq I \times \omega$ denote the subset where $v < 0$. Inspection of the variational inequality from the definition of the normal cone, cf., e.g., [147, Lemma 2.26], yields $\bar{q} = q_b$ on Λ_+ and $\bar{q} = q_a$ on Λ_- almost everywhere. Therefore, we have $\delta q_n \le 0$ on Λ_+ and $\delta q_n \ge 0$ on Λ_-, and the same holds for δq. Using the definition of δq_n, we infer that $\delta q_n \le \delta q$ a.e. on Λ_+ and $-\delta q \le -\delta q_n$ a.e. on Λ_-. Since $v \in N_{Q_{ad}(0,1)}(\bar{q})$ and $\delta q_n \in R_{Q_{ad}(0,1)}(\bar{q})$, we deduce

$$0 \le -(v, \delta q_n)_{L^2(I\times\omega)} = -(v, \delta q_n)_{L^2(\Lambda_+)} - (v, \delta q_n)_{L^2(\Lambda_-)} \le -(v, \delta q)_{L^2(I\times\omega)} = 0,$$

i.e. $\delta q_n \in \ker v$. Since δq_n converges pointwise almost everywhere and $|\delta q_n| \le |\delta q|$, the dominated convergence theorem implies $\delta q_n \to \delta q$ in $L^2(I \times \omega)$. □

With these preparation, we will verify that the value function V is two times differentiable. Furthermore, the scalar condition (3.27) can be identified with the value of the second derivative of V at the optimal time $\bar{\nu}$, i.e. the scalar condition (3.27) describes the curvature of the value function.

Proposition 3.24. *Let Assumption 3.3 hold and assume that $G''(u)[\cdot,\cdot]$ is positive semi-definite for all $u \in H$. Then the value function V is two times differentiable in a neighborhood of $\bar{\nu}$ and the expression*

$$V''(\bar{\nu}) = \partial^2_{(\nu,q)} \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[1, \delta\bar{q}]^2$$

holds, where $\delta\bar{q}$ is the solution to (3.28).

Proof. Let $\nu \in \mathbb{R}_+$ and $\tau_n \in \mathbb{R}$ such that $\tau_n \to 0$. Set $\nu_n = \nu + \tau_n$, $\bar{q}_n = \bar{q}(\nu_n)$, and $\bar{\mu}_n = \bar{\mu}(\nu_n)$. Using Proposition 3.21 we infer that the quotients $\delta q_n = \tau_n^{-1}(\bar{q}_n - \bar{q}(\nu))$ and $\delta\mu_n = \tau_n^{-1}(\bar{\mu}_n - \bar{\mu}(\nu))$ are bounded. Thus, by taking a subsequence if necessary, $\delta q_n \rightharpoonup \delta q(\nu)$ in $L^2(I \times \omega)$ and $\delta\mu_n \to \delta\mu(\nu)$.

We would like to apply [21, Theorem 5.10] again with $\mathcal{G}(q, \nu) \coloneqq (g(\nu, q), q)$ and constraint $\mathcal{K} \coloneqq \{0\} \times Q_{ad}(0,1)$, and verify its assumptions. First, the non-triviality condition yields $\partial_q g(\nu, \bar{q}(\nu))Q(0,1) = \mathbb{R}$. Therefore, upper Lipschitz continuity of the solution mapping follows from surjectivity of $D\mathcal{G}(\bar{q}(\nu), \nu)$ and [21, Remark 4.45 (i)]. According to [21, Proposition 3.76], the mapping $\partial^2_{(q,\mu)}\mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu))$ is a Legendre form. Furthermore, Robinson's constraint qualification holds at $\bar{q}(\nu)$, since $\partial_q g(\nu, \bar{q}(\nu))Q(0,1) = \mathbb{R}$. According to [21, Theorem 5.10], the weak limit $\delta q(\nu)$ is in fact a strong limit and $(\delta q(\nu), \delta\mu(\nu))$ satisfies a so-called linearized generalized equality. More specifically, there exists a triple $(\delta q(\nu), \delta\mu(\nu), \delta\xi(\nu))$ such that

$$\begin{aligned} \partial_{qq}\mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu))[\delta q(\nu), \cdot] + \delta\mu(\nu)\partial_q g(\nu, \bar{q}(\nu))^* + \partial_{\nu q}\mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu))[1, \cdot] &= -\delta\xi(\nu), \\ \partial_q g(\nu, \bar{q}(\nu))\delta q(\nu) + \partial_\nu g(\nu, \bar{q}(\nu)) &= 0, \\ \delta q(\nu) &\in T_{Q_{ad}(0,1)}(\bar{q}(\nu)), \\ (\xi(\nu), \delta q(\nu))_{Q(0,1)} &= 0, \\ (\delta\xi(\nu), \delta q)_{Q(0,1)} \le 0 \quad \forall \delta q \in T_{Q_{ad}(0,1)}(\bar{q}(\nu)) \text{ with } (\xi(\nu), \delta q)_{Q(0,1)} &= 0, \end{aligned}$$

where $\xi(\nu)$ is the Lagrange multiplier corresponding to the constraint $\bar{q}(\nu) \in Q_{ad}(0,1)$. Note first that for $\nu = \bar{\nu}$ we have

$$C_{\bar{q}} = \{\, \delta q \in T_{Q_{ad}(0,1)}(\bar{q}) \colon (\xi(\bar{\nu}), \delta q)_{Q(0,1)} = 0 \,\}\,.$$

Since the critical cone $C_{\bar{q}}$ is a subspace due to strict complementarity, the last condition in the system above is an equality for $\nu = \bar{\nu}$. Hence, this is exactly the linear system (3.28). Therefore, we deduce

$$\begin{aligned} V'(\nu) = \lim_{n\to\infty} \tau_n^{-1} \left(V(\nu_n) - V(\nu)\right) &= \lim_{n\to\infty} \tau_n^{-1} \left(\mathcal{L}(\nu_n, \bar{q}_n, \bar{\mu}_n) - \mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu))\right) \\ &= \partial\mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu))[1, \delta q(\nu), \delta\mu(\nu)] \\ &= \partial_\nu \mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu)). \end{aligned}$$

As $\nu \mapsto \bar{q}(\nu)$ and $\nu \mapsto \bar{\mu}(\nu)$ are continuous, we infer that V is continuously differentiable. Turning to the second derivative, the chain rule yields

$$V''(\nu) = \partial_{\nu\nu}\mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu)) + \partial_{\nu q}\mathcal{L}(\nu, \bar{q}(\nu), \bar{\mu}(\nu))[1, \delta q(\nu)] + \delta\mu(\nu)\partial_\nu g(\nu, \bar{q}(\nu)).$$

Therefore, setting $\nu = \bar{\nu}$, we obtain the expression

$$V''(\bar{\nu}) = \partial_{\nu\nu}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu}) + \partial_{\nu q}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[1, \delta\bar{q}] + \delta\bar{\mu}\partial_\nu g(\bar{\nu}, \bar{q}) = \partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[1, \delta\bar{q}]^2,$$

proving the assertion. □

We would like to verify (3.27) numerically, to get at least an indicator, whether a second order sufficient optimality condition holds. Due to the dimension of the linear system (3.28), iterative solvers seem to be appropriate for its solution. Hence, we have to efficiently calculate products of the system (3.28) times $(\delta q, \delta\mu)$. From the definition of the critical cone and employing (3.8) we find the condition

$$0 = \bar{\mu} g'(\bar{\nu}, \bar{q})(\delta\nu, \delta q) = \int_0^1 \langle \delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q, \bar{z}\rangle.$$

Thus, $\partial_\nu g(\bar{\nu}, \bar{q}) = \int_0^1 \langle \delta\nu(B\bar{q} - A\bar{u}), \bar{z}\rangle$ and $\partial_q g(\bar{\nu}, \bar{q})^* = \bar{\nu}B^*\bar{z}$. For the second derivative of the Lagrange function we obtain

$$\partial^2_{(\nu,q)}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu, \delta q]^2 = \alpha\bar{\nu}\|\delta q\|^2_{L^2(I\times\omega)} + 2\alpha\int_0^1 \delta\nu(\delta q, \bar{q}) + \bar{\mu}g''(\bar{\nu}, \bar{q})[\delta\nu, \delta q]^2,$$

with

$$\bar{\mu}g''(\bar{\nu}, \bar{q})[\delta\nu, \delta q]^2 = \int_0^1 \delta\nu(2B\delta q - A\delta u, \bar{z}) + \int_0^1 (\delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q, \delta\bar{z}),$$

where $\delta u = S'(\bar{\nu}, \bar{q})(\delta\nu, \delta q)$ and $\delta\bar{z}$ is a second adjoint state solving

$$-\partial_t\delta\bar{z} + \bar{\nu}A^*\delta\bar{z} = -\delta\nu A^*\bar{z}, \quad \delta\bar{z}(1) = \bar{\mu}G''(\bar{u}(1))[\delta u(1), \cdot\,].$$

Considering the splitting $\delta\bar{z} = \delta\nu\hat{z} + \delta\bar{z}_2$, where $\hat{z}$ solves

$$-\partial_t\hat{z} + \bar{\nu}A^*\hat{z} = -A^*\bar{z}, \quad \hat{z}(1) = 0,$$

we obtain

$$-\int_0^1 \delta\nu(A\delta u, \bar{z}) = \delta\nu\int_0^1 (\delta u, (-\partial_t + \bar{\nu}A^*)\,\hat{z}) = \delta\nu\int_0^1 (\delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q, \hat{z}),$$

and

$$\int_0^1 (\delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q, \delta\bar{z}) = \bar{\mu}G''(\bar{u}(1))[\delta u(1)]^2 + \delta\nu \int_0^1 (\delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q, \hat{z}).$$

Let $S = (\partial_t + \bar{\nu}A)^{-1}$ be the solution operator to the state equation with homogeneous initial condition. Moreover, let $G''(\bar{u}(1))\colon H \to H$ denote the operator representation of $G''(\bar{u}(1))[\cdot,\cdot]$ and set $E = (i_1 S)^* G''(\bar{u}(1))(i_1 S)$. Then

$$\begin{aligned} G''(\bar{u}(1))[\delta u(1)]^2 &= (E(\delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q), \delta\nu(B\bar{q} - A\bar{u}) + \bar{\nu}B\delta q)_H \\ &= \delta\nu^2 (E(B\bar{q} - A\bar{u}), B\bar{q} - A\bar{u})_H \\ &\quad + 2\delta\nu(E(B\bar{q} - A\bar{u}), \bar{\nu}B\delta q)_H + \bar{\nu}^2 (EB\delta q, B\delta q)_H. \end{aligned}$$

This in summary leads to

$$\begin{aligned} \partial_{qq}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu}) &= \alpha\bar{\nu}\,\mathrm{Id} + \bar{\mu}\bar{\nu}^2 B^* EB \\ \partial_{\nu q}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})^* &= \alpha\bar{q} + B^*\bar{z} + \bar{\nu}B^*(\hat{z} + \bar{\mu}E(B\bar{q} - A\bar{u})) \\ \partial_{\nu\nu}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu}) &= \bar{\mu}(E(B\bar{q} - A\bar{u}), B\bar{q} - A\bar{u})_H + 2\int_0^1 (B\bar{q} - A\bar{u}, \hat{z}). \end{aligned}$$

For these reasons, the application of $\partial_{qq}\mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})$ to δq can be calculated by solving two partial differential equations. Given the adjoint state $\bar{z}$, the application of $\partial_q g(\bar{\nu}, \bar{q})$ to δq can be easily calculated by solving the resulting integral. These expressions directly transfer to the discrete problems discussed in Chapter 5. Using an iterative solver the corresponding discrete set of equations to (3.28) can be efficiently evaluated numerically without building the matrix $\partial_{qq}\mathcal{L}_{kh}(\bar{\nu}_{kh}, \bar{q}_{kh}, \bar{\mu}_{kh})$. Based on this approach, in Section 5.4 we will eventually verify the second order sufficient optimality condition on the discrete level.

3.2.4. Local uniqueness of local solutions

In related context it is known that second order sufficient optimality conditions imply local uniqueness of local solutions. Using similar arguments as in the proof of Theorem 3.13 we obtain local uniqueness in $\mathbb{R} \times L^2(I \times \omega)$, if the second order sufficient optimality conditions for $(\hat{P})$ hold. Note that this does not automatically follow from Taylor's expansion and the coercivity condition (3.22), since we formulated the second order condition employing a cone of critical directions.

Theorem 3.25. *Let $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be a local solution of $(\hat{P})$ such that the (qualified) first order optimality conditions of Lemma 3.1 and the second order sufficient optimality conditions of Theorem 3.13 hold. Then $(\bar{\nu}, \bar{q})$ is locally unique in the sense of $\mathbb{R} \times L^2(I \times \omega)$.*

Proof. Suppose that $(\bar{\nu}, \bar{q})$ not locally unique, i.e. there exist locally optimal solutions $\bar{\chi}_n = (\bar{\nu}_n, \bar{q}_n) \in \mathbb{R}_+ \times Q_{ad}(0,1)$, $n \in \mathbb{N}$, such that $(\bar{\nu}_n, \bar{q}_n) \to (\bar{\nu}, \bar{q}) = \bar{\chi}$ in $\mathbb{R} \times Q(0,1)$.

Define $\rho_n = \|(\bar{\nu}_n - \bar{\nu}, \bar{q}_n - \bar{q})\|$ and

$$v_n = (v_n^\nu, v_n^q) = \frac{1}{\rho_n}(\bar{\chi}_n - \bar{\chi}).$$

We may assume w.l.o.g. that $v_n^\nu \to v^\nu$ in $\mathbb{R}$ and $v_n^q \rightharpoonup v^q$ in $L^2(I \times \omega)$.

Step 0: Preparation. Since $\partial_\nu g(\bar\chi) < 0$ according to (3.1), the convergence $\bar\chi_n \to \bar\chi$, and since $\partial_\nu g$ is continuous, there exists $g_0 > 0$ such that $\partial_\nu g(\bar\chi_n) < -g_0$ for all n sufficiently large, which is a constraint qualification. Therefore, qualified first order optimality conditions hold for $\bar\chi_n$, i.e. there exist multipliers $\bar\mu_n > 0$ such that

$$0 = \partial_\nu \mathcal{L}(\bar\chi_n, \bar\mu_n) = \partial_\nu j(\bar\chi_n) + \bar\mu_n \partial_\nu g(\bar\chi_n).$$

Clearly, we also have $0 = \partial_\nu \mathcal{L}(\bar\chi, \bar\mu) = \partial_\nu j(\bar\chi) + \bar\mu \partial_\nu g(\bar\chi)$. Adding both equalities implies

$$\begin{aligned} |\bar\mu - \bar\mu_n| &\le |\partial_\nu g(\bar\chi)|^{-1} |\partial_\nu j(\bar\chi) - \partial_\nu j(\bar\chi_n)| + \frac{|\partial_\nu g(\bar\chi) - \partial_\nu g(\bar\chi_n)|}{|\partial_\nu g(\bar\chi)\partial_\nu g(\bar\chi_n)|} \partial_\nu j(\bar\chi_n) \\ &\le c \|\bar\chi - \bar\chi_n\|, \end{aligned} \tag{3.33}$$

where we have used that $\partial_\nu j(\chi) = \int (1 + \frac{\alpha}{2}\|q\|^2)$.

Step 1: $\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v = 0$. Clearly, since $\bar q_n \in Q_{ad}(0,1)$, it holds $\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v \ge 0$. To show the reverse inequality, from the first order optimality conditions

$$\partial_\chi \mathcal{L}(\bar\chi_n, \bar\mu_n)(\chi - \bar\chi_n) \ge 0, \quad \chi \in \mathbb{R} \times Q_{ad}(0,1),$$

for $\chi = \bar\chi$ we obtain

$$\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v = \partial_\chi \left[\mathcal{L}(\bar\chi, \bar\mu) - \mathcal{L}(\bar\chi_n, \bar\mu_n)\right] v + \partial_\chi \mathcal{L}(\bar\chi_n, \bar\mu_n)v \le c\left(\|\bar\chi - \bar\chi_n\| + |\bar\mu - \bar\mu_n|\right),$$

which tends to zero as $n \to \infty$ due to the estimate (3.33).

Step 2: $v \in C_{(\bar\nu, \bar q)}$. As both $\bar\chi_n$ and $\bar\chi$ are (locally) optimal, we have

$$g'(\bar\chi)v = \lim_{n\to\infty} \frac{1}{\rho_n}\left[g(\bar\chi + \rho_n v_n) - g(\bar\chi)\right] = \lim_{n\to\infty} \frac{1}{\rho_n}\left[g(\bar\chi_n) - g(\bar\chi)\right] = 0.$$

Moreover, because the set

$$\left\{ \delta q \in L^2(I\times\omega) \;\middle|\; \begin{array}{l} \delta q \le 0 \text{ if } \bar q(t,x) = q_b(x) \\ \delta q \ge 0 \text{ if } \bar q(t,x) = q_a(x) \end{array} \right\},$$

is closed and convex, it is in particular weakly closed. Moreover, due to feasibility of q_n every $(q_n - \bar q)/\rho_n$ belongs to the set above, so does the weak limit. Thus, v satisfies $v^q \le 0$, if $\bar q(t,x) = q_b(x)$, and $v^q \ge 0$, if $\bar q(t,x) = q_a(x)$. For this reason, (3.6) implies

$$\int_0^1 \int_\omega \bar\nu(\alpha\bar q + B^*\bar z)v^q \,\mathrm{d}x\,\mathrm{d}t = \int_0^1 \int_\omega \bar\nu|(\alpha\bar q + B^*\bar z)v^q| \,\mathrm{d}x\,\mathrm{d}t.$$

Moreover, due to $\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)v = 0$ and the first order necessary condition $\partial_\nu \mathcal{L}(\bar\chi, \bar\mu) = 0$ we have the equality

$$0 = \partial_q \mathcal{L}(\bar\chi, \bar\mu)v^q = \int_0^1 \bar\nu(\alpha\bar q + B^*\bar z, v^q)_{L^2(\omega)} \,\mathrm{d}t = \int_0^1 \int_\omega \bar\nu|(\alpha\bar q + B^*\bar z)v^q| \,\mathrm{d}x\,\mathrm{d}t.$$

Hence, $v^q = 0$, if $\alpha\bar q(t,x) + B^*\bar z(t,x) \ne 0$, and v^q satisfies the sign condition (3.11) as well.

Step 3: $v = 0$. Employing Taylor expansion of $\mathcal{L}$ we find

$$0 = \partial_\chi \left[\mathcal{L}(\bar\chi_n, \bar\mu) - \mathcal{L}(\bar\chi, \bar\mu)\right](\bar\chi_n - \bar\chi) - \partial_\chi^2 \mathcal{L}(\breve\chi_n, \bar\mu)[\bar\chi_n - \bar\chi]^2$$

with some appropriate $\check{\chi}_n$ satisfying $\check{\chi}_n \to \bar{\chi}$ as $n \to \infty$. Using the optimality conditions for $\bar{\chi}_n$ and $\bar{\chi}$, i.e.

$$\partial_\chi \mathcal{L}(\bar{\chi}, \bar{\mu})(\bar{\chi}_n - \bar{\chi}) \geq 0, \quad \partial_\chi \mathcal{L}(\bar{\chi}_n, \bar{\mu}_n)(\bar{\chi} - \bar{\chi}_n) \geq 0,$$

we find

$$0 \leq \partial_\chi \left[\mathcal{L}(\bar{\chi}_n, \bar{\mu}) - \mathcal{L}(\bar{\chi}_n, \bar{\mu}_n)\right](\bar{\chi}_n - \bar{\chi}) - \partial_\chi^2 \mathcal{L}(\check{\chi}_n, \bar{\mu})[\bar{\chi}_n - \bar{\chi}]^2$$

Thus, adding $\partial_\chi^2 \mathcal{L}(\bar{\chi}, \bar{\mu})$ and dividing by ρ_n^2 we arrive at

$$\partial_\chi^2 \mathcal{L}(\bar{\chi}, \bar{\mu}) v_n^2 \leq \rho_n^{-1} \partial_\chi \left[\mathcal{L}(\bar{\chi}_n, \bar{\mu}) - \mathcal{L}(\bar{\chi}_n, \bar{\mu}_n)\right] v_n + \partial_\chi^2 \left[\mathcal{L}(\bar{\chi}, \bar{\mu}) - \mathcal{L}(\check{\chi}_n, \bar{\mu})\right] v_n^2. \tag{3.34}$$

For the first term of the right-hand side of (3.34) we have

$$\begin{aligned}\rho_n^{-1} \partial_\chi \left[\mathcal{L}(\bar{\chi}_n, \bar{\mu}) - \mathcal{L}(\bar{\chi}_n, \bar{\mu}_n)\right] v_n &= \rho_n^{-1}(\bar{\mu} - \bar{\mu}_n) g'(\bar{\chi}_n) v_n \\ &= \rho_n^{-1}(\bar{\mu} - \bar{\mu}_n) \left[g'(\bar{\chi}) + (g'(\bar{\chi}_n) - g'(\bar{\chi}))\right] v_n \to 0,\end{aligned}$$

using again the estimate (3.33), the fact that $g'(\bar{\chi})v = 0$ (step 1), and convergence $\bar{\chi}_n \to \bar{\chi}$ with continuity of g'. In summary, inequality (3.34), weak lower semicontinuity of $\partial_\chi^2 \mathcal{L}(\bar{\chi}, \bar{\mu})$, and continuity of $\partial_\chi^2 \mathcal{L}$ yield

$$\partial_\chi^2 \mathcal{L}(\bar{\chi}, \bar{\mu}) v^2 \leq \liminf_{n \to \infty} \partial_\chi^2 \mathcal{L}(\bar{\chi}, \bar{\mu}) v_n^2 \leq 0.$$

The second order sufficient optimality condition implies $v = 0$.

Step 4: Final contradiction. Using $\|(v_n^\nu, v_n^q)\| = 1$ and $v^\nu \to 0$ we obtain

$$0 < \alpha\bar{\nu} = \alpha\bar{\nu} \liminf_{n \to \infty} \|(v_n^\nu, v_n^q)\|^2 = \alpha\bar{\nu} \liminf_{n \to \infty} \|v_n^q\|_{L^2(I \times \omega)}^2 = \liminf_{n \to \infty} \alpha \int_0^1 \bar{\nu} \|v_n^q(t)\|_{L^2(\omega)}^2 \, \mathrm{d}t.$$

Using the specific structure of j'', we see that

$$\liminf_{n \to \infty} \alpha \int_0^1 \bar{\nu} \|v_n^q(t)\|_{L^2(\omega)}^2 \, \mathrm{d}t = \liminf_{n \to \infty} j''(\bar{\chi})[v_n^\nu, v_n^q]^2.$$

Thus, employing Corollary 3.9 we conclude that

$$\begin{aligned}0 &< \liminf_{n \to \infty} j''(\bar{\chi})[v_n^\nu, v_n^q]^2 \leq \liminf_{n \to \infty} j''(\bar{\chi})[v_n^\nu, v_n^q]^2 + \bar{\mu} \liminf_{n \to \infty} g''(\bar{\chi})[v_n^\nu, v_n^q]^2 \\ &\leq \liminf_{n \to \infty} \partial_\chi^2 \mathcal{L}(\bar{\chi}, \bar{\mu})[v_n^\nu, v_n^q]^2 \leq 0,\end{aligned}$$

where we have used again (3.34) in the last inequality. □

3.3. Sufficient optimality conditions for bang-bang controls ($\alpha = 0$)

After the discussion of second order optimality conditions for a fixed cost parameter $\alpha > 0$, we now turn to the case of variable α that typically leads to bang-bang controls in the limit case $\alpha \to 0$. For $\alpha \geq 0$ we introduce the regularized and transformed problem

$$\inf_{\substack{\nu \in \mathbb{R}_+ \\ q \in Q_{ad}(0,1)}} j_\alpha(\nu, q) \quad \text{subject to} \quad g(\nu, q) \leq 0, \tag{$\hat{P}_\alpha$}$$

where the objective function is given by

$$j_\alpha(\nu, q) = \nu \left(1 + \int_0^1 \frac{\alpha}{2} \|q\|_{L^2(\omega)}^2\right).$$

As before $\mathcal{L}(\nu, q, \mu) \coloneqq j_\alpha(\nu, q) + \mu\, g(\nu, q)$ denotes the Lagrange function associated to $(\hat{P}_\alpha)$. In order to simplify the notation, we neglect the α-dependence in the symbol $\mathcal{L}$.

Let $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be a locally optimal solution for $(\hat{P}_\alpha)$ with $\alpha = 0$. Then, from the optimality conditions of Lemma 3.1 we infer

$$\bar{q}(t,x) = \begin{cases} q_a(x) & \text{if } (B^*\bar{z})(t,x) > 0, \\ q_b(x) & \text{if } (B^*\bar{z})(t,x) < 0, \end{cases}$$

and

$$(B^*\bar{z})(t,x) \begin{cases} \geq 0 & \text{if } \bar{q}(t,x) = q_a(x), \\ \leq 0 & \text{if } \bar{q}(t,x) = q_b(x), \\ = 0 & \text{if } q_a(x) < \bar{q}(t,x) < q_b(x). \end{cases} \tag{3.35}$$

In this section we are interested in the case when $\bar{q}$ is a bang-bang control. If

$$|\{\, (t,x) \in I \times \omega \colon (B^*\bar{z})(t,x) = 0 \,\}| = 0, \tag{3.36}$$

where $|\cdot|$ denotes the measure associated with $I \times \omega$, then from the first order necessary optimality condition (3.35) we conclude that the control $\bar{q}$ is bang-bang. The condition (3.36) on the set of zeros can be deduced from a backwards uniqueness property. For example, if Ω is a bounded domain, $A = -\Delta$ equipped with homogeneous Dirichlet boundary conditions, and $B \colon L^2(\omega) \to L^2(\Omega)$ for $\omega \subset \Omega$ open is the extension by zero operator, then the backwards uniqueness property is valid; see Holmgren's uniqueness theorem [79, Theorem 5.3.3] or [105]. The bang-bang property for time-optimal control problems subject to parabolic partial differential equations has been extensively studied; see, e.g., [88, 96, 157].

Note that if (3.36) holds, then the critical cone used in the formulation of the second order necessary and sufficient optimality conditions in the preceding section is trivial, i.e. $C_{(\bar{\nu},\bar{q})} = \{\, (0,0) \,\}$. This immediately follows from Proposition 3.15, because (3.36) and the bang-bang property imply the strict complementarity condition (3.23). Therefore, the second order sufficient optimality condition from Theorem 3.13 is vacuously true and does not provide any additional information.

However, it should be noted that global uniqueness of a solution can still be guaranteed. First, in view of $j(\nu, q) = \nu$ due to $\alpha = 0$, the optimal time $T = \bar{\nu}$ is unique. Concerning the control variable, we can state the following criterion.

Proposition 3.26. *Let G be convex and $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be a global solution to $(\hat{P}_0)$. Suppose that $C_{(\bar{\nu},\bar{q})} = \{\, 0 \,\}$. Then $(\bar{\nu}, \bar{q})$ is globally unique.*

Proof. As already noted $\bar{\nu}$ is uniquely determined because of $\alpha = 0$. Let $(\bar{\nu}, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be another global solution to $(\hat{P}_0)$. We set $q_\lambda = \lambda q + (1-\lambda)\bar{q}$ for $\lambda \in [0,1]$. Convexity of G and linearity of the control-to-state mapping (for fixed $\bar{\nu}$) imply

$$g(\bar{\nu}, q_\lambda) \leq \lambda g(\bar{\nu}, q) + (1-\lambda) g(\bar{\nu}, \bar{q}) = 0.$$

Hence, q_λ is also feasible for $(\hat{P}_0)$. Moreover, a simple contradiction argument shows that q_λ is also optimal for $(\hat{P}_0)$. In particular, $g(\bar{\nu}, q_\lambda) = 0$. Therefore,

$$\partial_q \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})(q - \bar{q}) = \lim_{\substack{\lambda \in (0,1] \\ \lambda \to 0}} \frac{1}{\lambda} \left[\mathcal{L}(\bar{\nu}, q_\lambda, \bar{\mu}) - \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu}) \right] = 0.$$

Hence, $q - \bar{q}$ satisfies the sign condition (3.11), because $q \in Q_{ad}(0,1)$. Moreover,

$$\partial_q g(\bar{\nu}, \bar{q})(q - \bar{q}) = \lim_{\substack{\lambda \in (0,1] \\ \lambda \to 0}} \frac{1}{\lambda} \left[g(\bar{\nu}, q_\lambda) - g(\bar{\nu}, \bar{q}) \right] = 0.$$

Thus, $g'(\bar{\nu}, \bar{q})(0, q - \bar{q}) = 0$ and we conclude that $(0, q - \bar{q}) \in C_{(\bar{\nu}, \bar{q})}$. Since the critical cone is supposed to be trivial, this implies $q = \bar{q}$. □

Alternatively, if the terminal value of the adjoint state equation is unique, then from (3.36) we deduce uniqueness of the control variable. Uniqueness of the terminal value for the adjoint state equation can be shown for certain problems, e.g., employing a dual problem such as in [160, Theorem 3.2].

However, to quantify local uniqueness we require an additional condition. To this end, in this section we assume that there are constants $C > 0$ and $\kappa > 0$ such that the adjoint state $\bar{z}$ satisfies

$$|\{ (t,x) \in I \times \omega \colon -\varepsilon \leq (B^* \bar{z})(t,x) \leq \varepsilon \}| \leq C \varepsilon^\kappa \quad \text{for all } \varepsilon > 0. \tag{3.37}$$

Note that the structural assumption (3.37) (that is sometimes also referred to as measure condition, see, e.g., [44, Assumption 7]) is a strengthening of the condition (3.36) that ensures the bang-bang property. We collect situations where (3.37) is guaranteed to hold and relate it to similar conditions from the literature.

Remark 3.27. (i) Similar assumptions on the adjoint state as in (3.37) have been used in related contexts; see, e.g., [36, 37, 44, 47, 145, 152, 155, 156] for PDE-constrained optimization problems. In the context of optimal control problems with ODEs, one typically assumes that the differentiable switching function $\sigma \colon I \to \mathbb{R}$ has only finitely many zeros with nonvanishing first derivatives; see, e.g., [55, 113]. Condition (3.37) can be considered to be a generalization to the distributed control case. Furthermore, (3.37) is a strengthened complementarity condition.

(ii) If $B^* \bar{z} \in C^1(\overline{I \times \omega})$ and if there exists a constant $c > 0$ such that

$$|\nabla_{(t,x)} B^* \bar{z}(t,x)| \geq c$$

for all $(t,x) \in I \times \omega$ such that $B^* \bar{z}(t,x) = 0$, then the condition (3.37) holds with $\kappa = 1$; see [47, Lemma 3.2].

(iii) Condition (3.37) is also compatible with purely time-dependent controls; see Example 3.8. In this case ω would be a discrete set equipped with the counting measure and the control operator is defined by $Bq = \sum_{i=1}^{N_c} q_i e_i$, where $e_i \in V^*$ are given form functions. The adjoint operator of B is $(B^* \varphi)_i = \langle e_i, \varphi \rangle$ for $i = 1, 2 \ldots, N_c$. Hence, the measure condition (3.37) can be written as

$$\sum_{i=1}^{N_c} |\{ t \in I \colon |(B^* \bar{z}(t))_i| \leq \varepsilon \}| \leq C \varepsilon^\kappa.$$

(iv) Consider the case of purely time-dependent controls and suppose $B^* \bar{z} \neq 0$ (otherwise (3.36) is clearly violated). Since $-A$ generates an analytic semigroup $\mathrm{e}^{-\cdot A}$, the function $t \mapsto (B^* \bar{z}(t))_i$ can have only finitely many zeros on the interval $(0, 1-\varepsilon)$ for all $\varepsilon \in (0,1)$. However, these zeros may accumulate at $t = 1$. If $\mathrm{e}^{-\cdot A}$ is an analytic group, then $(B^* \bar{z}(t))_i$ is analytic on $\mathbb{R}$. Consequently it can have only finitely many zeros and not all derivatives vanish. Therefore, there is $\kappa > 0$ such that (3.37) is satisfied; cf. also [88, Theorem 1.1].

(v) For a particular situation we can verify (3.37) also in the case that $-A$ is the infinitesimal generator of a semigroup. Suppose that $\bar{z}(1) = \sum_{j=1}^{n} v_j$ with $Av_j = \lambda_j v_j$, $v_j \in V$, and $\lambda_j \in \mathbb{R}$. Employing [128, Lemma 2.2.2], for purely time-dependent controls as considered in Example 3.8 we obtain

$$(B^*\bar{z}(t))_i = \sum_{j=1}^{n} \mathrm{e}^{\lambda_j t}\langle e_i, v_j\rangle, \quad i = 1, 2, \dots, N_c.$$

Since $t \mapsto (B^*\bar{z}(t))_i$ is analytic, it can have only finitely many zeros and not all derivatives vanish. Thus, there is $\kappa > 0$ such that (3.37) is satisfied.

3.3.1. Sufficient optimality conditions

We will show that (3.37) is sufficient for optimality of $(\bar{\nu}, \bar{q})$. Throughout this section, we suppose the following assumption to hold.

Assumption 3.4. The function $G\colon H \to \mathbb{R}$ is twice continuously Fréchet-differentiable. In addition, we assume that

$$G''(u)\delta u^2 \geq 0 \quad \text{for all } u, \delta u \in H.$$

The proof of sufficiency of the structural assumption for a pair $(\bar{\nu}, \bar{q})$ to be locally optimal will rely on the following observation.

Proposition 3.28. *Let $(\bar{\nu}, \bar{q}, \bar{\mu}) \in \mathbb{R}_+ \times Q_{ad}(0,1) \times \mathbb{R}_+$ and (3.37) hold. Then there is a constant $c_0 > 0$ such that*

$$\partial_q \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})(q - \bar{q}) \geq c_0 \bar{\nu} \|q - \bar{q}\|^{1+1/\kappa}_{L^1(I\times\omega)} \quad \textit{for all } q \in Q_{ad}(0,1). \tag{3.38}$$

Proof. The proof is along the lines of [37, Proposition 2.7] and we give it for convenience of the reader. For $q \in Q_{ad}(0,1)$, set

$$\varepsilon := \left(2\|q_b - q_a\|_{L^\infty(\omega)} C\right)^{-1/\kappa} \|q - \bar{q}\|^{1/\kappa}_{L^1(I\times\omega)}$$

and $E_\varepsilon := \{\, (t,x) \in I \times \omega \colon |B^*\bar{z}(t,x)| \geq \varepsilon \,\}$. Then due to (3.35) we see that

$$\begin{aligned}\partial_q \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})(q - \bar{q}) &= \bar{\nu} \int_0^1 \int_\omega (q - \bar{q}) B^*\bar{z} = \bar{\nu} \int_0^1 \int_\omega |q - \bar{q}||B^*\bar{z}| \geq \bar{\nu} \int_{E_\varepsilon} |q - \bar{q}||B^*\bar{z}| \\ &\geq \varepsilon\bar{\nu}\|q - \bar{q}\|_{L^1(E_\varepsilon)} = \varepsilon\bar{\nu}\left(\|q - \bar{q}\|_{L^1(I\times\omega)} - \|q - \bar{q}\|_{L^1((I\times\omega)\setminus E_\varepsilon)}\right).\end{aligned}$$

According to (3.37) we have $|(I \times \omega) \setminus E_\varepsilon| \leq C\varepsilon^\kappa$. Hence,

$$\|q - \bar{q}\|_{L^1((I\times\omega)\setminus E_\varepsilon)} \leq \|q_b - q_a\|_{L^\infty(\omega)} C\varepsilon^\kappa.$$

Therefore, we arrive at

$$\begin{aligned}\partial_q \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})(q - \bar{q}) &\geq \varepsilon\bar{\nu}\left(\|q - \bar{q}\|_{L^1(I\times\omega)} - \|q - \bar{q}\|_{L^1((I\times\omega)\setminus E_\varepsilon)}\right) \\ &\geq \varepsilon\bar{\nu}\left(\|q - \bar{q}\|_{L^1(I\times\omega)} - \|q_b - q_a\|_{L^\infty(\omega)} C\varepsilon^\kappa\right) \\ &= \frac{\varepsilon\bar{\nu}}{2}\|q - \bar{q}\|_{L^1(I\times\omega)} = c_0\bar{\nu}\|q - \bar{q}\|^{1+1/\kappa}_{L^1(I\times\omega)}\end{aligned}$$

with $c_0 = \frac{1}{2}\left(2\|q_b - q_a\|_{L^\infty(\omega)} C\right)^{-1/\kappa}$. □

Definition 3.29. The tuple $(\bar\nu, \bar q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ is called a *local solution in the sense of* L^1 with radius $\varepsilon > 0$ for $(\hat P_0)$, if the inequality

$$\bar\nu \le \nu$$

holds for all admissible tuple $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ with $|\nu - \bar\nu| + \|q - \bar q\|_{L^1(I\times\omega)} \le \varepsilon$.

The structural assumption of the adjoint state allows to prove the following growth condition without two norm discrepancy. We emphasize that due to the particular objective functional we do not require any additional assumption such as a condition on the second derivative of the Lagrange function; cf. [28, Theorem 2.2] and [37, Theorem 2.8].

Theorem 3.30. *Let $(\bar\nu, \bar q, \bar\mu) \in \mathbb{R}_+ \times Q_{ad}(0,1) \times \mathbb{R}_+$ satisfy the first order necessary optimality conditions of Lemma 3.1. Assume that the associated adjoint state satisfies the structural assumption* (3.37)*. Then there are constants $\varepsilon > 0$ and $c > 0$ such that the growth condition*

$$c\|q - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)} \le \nu - \bar\nu \quad [= j_0(\nu, q) - j_0(\bar\nu, \bar q)] \tag{3.39}$$

holds for all admissible $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ with $|\nu - \bar\nu| \le \varepsilon$.

Note that a localization with respect to q is implicitly contained in Theorem 3.30, due to

$$c\|q - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)} \le \nu - \bar\nu \le \varepsilon.$$

To prove the result, we first observe that under Assumption 3.4 the second derivative of the Lagrange function can be bounded from below as follows.

Proposition 3.31. *Let $(\bar\nu, \bar q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$, $\bar\mu > 0$, and $0 < \nu_{\min} < \nu_{\max}$. There is $c > 0$ such that*

$$\partial^2_{(\nu,q)}\mathcal{L}(\nu_\xi, q_\xi, \bar\mu)[\nu - \bar\nu, q - \bar q]^2 \ge -c|\nu - \bar\nu|^2 - c|\nu - \bar\nu|\|q - \bar q\|_{L^2(I\times\omega)}$$

for all $\nu, \nu_\xi \in \mathbb{R}_+$, $q, q_\xi \in Q_{ad}(0,1)$ with $\nu_{\min} \le \nu, \nu_\xi \le \nu_{\max}$.

Proof. Set $\delta\nu = \nu - \bar\nu$ and $\delta q = q - \bar q$. Define $u = S(\nu_\xi, q_\xi)$, $\delta u = S'(\nu_\xi, q_\xi)(\delta\nu, \delta q)$, and $\delta\tilde u = S''(\nu_\xi, q_\xi)[\delta\nu, \delta q]^2$. Moreover, let z_ξ be the corresponding adjoint state with terminal value $\bar\mu G'(u_\xi(1))^*$. Then we observe

$$\begin{aligned}
\bar\mu G'(u(1))\delta\tilde u(1) &= (z_\xi(1), \delta\tilde u(1)) - (z_\xi(0), \delta\tilde u(0)) \\
&= \int_0^1 \langle \partial_t \delta\tilde u, z_\xi\rangle + \int_0^1 \langle \partial_t z_\xi, \delta\tilde u\rangle = \int_0^1 \langle \partial_t\delta\tilde u, z_\xi\rangle + \int_0^1 \langle \bar\nu A\delta\tilde u, z_\xi\rangle \\
&= 2\delta\nu \int_0^1 \langle B\delta q - A\delta u, z_\xi\rangle \,\mathrm{d}t.
\end{aligned}$$

Thus, because $\alpha = 0$, and using Assumption 3.4 on G'', we find

$$\begin{aligned}
\partial^2_{(\nu,q)}\mathcal{L}(\nu_\xi, q_\xi, \bar\mu)[\delta\nu, \delta q]^2 &= \bar\mu G''(u_\xi(1))[\delta u(1)]^2 + 2\delta\nu \int_0^1 \langle B\delta q - A\delta u, z_\xi\rangle \,\mathrm{d}t \\
&\ge -2|\delta\nu| \int_0^1 |\langle B\delta q - A\delta u, z_\xi\rangle| \,\mathrm{d}t.
\end{aligned}$$

The Cauchy-Schwarz inequality and the stability estimates for $u, \delta u$, and z, see Proposition A.26, further imply

$$\begin{aligned}\partial^2_{(\nu,q)}\mathcal{L}(\nu_\xi, q_\xi, \bar\mu)[\delta\nu, \delta q]^2 &\geq -|\delta\nu| \left(\|B\delta q\|_{L^2(I;V^*)} + \|\delta u\|_{L^2(I;V)} \right) \|z_\xi\|_{L^2(I;V)} \\ &\geq -c|\delta\nu| \left(\|B\delta q\|_{L^2(I;V^*)} + \frac{|\delta\nu|}{\nu_\xi} \left(\|Bq_\xi\|_{L^2(I;V^*)} + \|u_\xi\|_{L^2(I;V)} \right) \right) \frac{\bar\mu}{\sqrt{\nu_\xi}} \|G'(u_\xi(1))^*\|_H.\end{aligned}$$

Since q_ξ is uniformly bounded due to boundedness of $Q_{ad}(0,1)$ as well as ν_ξ is uniformly bounded from below and from above, there exists a constant $c > 0$ such that

$$\partial^2_{(\nu,q)}\mathcal{L}(\nu_\xi, q_\xi, \bar\mu)[\delta\nu, \delta q]^2 \geq -c|\delta\nu|^2 - c|\delta\nu| \|\delta q\|_{L^2(I\times\omega)}$$

proving the assertion. □

Proof of Theorem 3.30. Let $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be admissible with $|\nu - \bar\nu| \leq \bar\nu/2$. Set $\delta\nu = \nu - \bar\nu$ and $\delta q = q - \bar q$. Using feasibility of (ν, q), the fact that $\bar\mu > 0$ from the first order necessary optimality conditions for $(\bar\nu, \bar q)$, as well as Taylor expansion we find

$$\begin{aligned}\nu - \bar\nu = j_0(\nu, q) - j_0(\bar\nu, \bar q) &\geq j_0(\nu, q) + \bar\mu g(\nu, q) - (j_0(\bar\nu, \bar q) + \bar\mu g(\bar\nu, \bar q)) \\ &= \mathcal{L}(\nu, q, \bar\mu) - \mathcal{L}(\bar\nu, \bar q, \bar\mu) \\ &= \partial_{(\nu,q)}\mathcal{L}(\bar\nu, \bar q, \bar\mu)(\delta\nu, \delta q) + \frac{1}{2}\partial^2_{(\nu,q)}\mathcal{L}(\nu_\xi, q_\xi, \bar\mu)[\delta\nu, \delta q]^2,\end{aligned}$$

with appropriate $\nu_\xi = \bar\nu + \xi_\nu(\nu - \bar\nu)$ and $q_\xi = \bar q + \xi_q(q - \bar q)$ for $0 \leq \xi_\nu, \xi_q \leq 1$. Thus, according to Proposition 3.31 there is $c_1 > 0$ such that

$$\nu - \bar\nu \geq \partial_{(\nu,q)}\mathcal{L}(\bar\nu, \bar q, \bar\mu)(\delta\nu, \delta q) - c_1|\delta\nu|^2 - c_1|\delta\nu| \|\delta q\|_{L^2(I\times\omega)}.$$

Since $\partial_\nu\mathcal{L}(\bar\nu, \bar q, \bar\mu) = 0$ and using Proposition 3.28, this further implies

$$\nu - \bar\nu \geq c_0\bar\nu \|q - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)} - c_1|\delta\nu|^2 - c_1|\delta\nu| \|\delta q\|_{L^2(I\times\omega)}.$$

Applying Young's inequality to the last term with $p = 2 + 2/\kappa$ and $p' = p/(p-1)$ yields

$$|\delta\nu| \|\delta q\|_{L^2(I\times\omega)} \leq \frac{1}{p'} \left(\frac{|\delta\nu|}{\varepsilon} \right)^{p'} + \frac{\varepsilon^p}{p} \|\delta q\|^{2+2/\kappa}_{L^2(I\times\omega)}$$

for any $\varepsilon > 0$. Clearly, we have

$$\|\delta q\|_{L^2(I\times\omega)} \leq \|\delta q\|^{1/2}_{L^\infty(I\times\omega)} \|\delta q\|^{1/2}_{L^1(I\times\omega)} \leq \|q_b - q_a\|^{1/2}_{L^\infty(\omega)} \|\delta q\|^{1/2}_{L^1(I\times\omega)}.$$

Choosing $\varepsilon = (c_0\bar\nu p/(2c_1))^{1/p} \|q_b - q_a\|^{-1/2}_{L^\infty(\omega)}$ we obtain

$$c_1|\delta\nu| \|\delta q\|_{L^2(I\times\omega)} \leq c_2|\delta\nu|^{p'} + \frac{c_0\bar\nu}{2} \|\delta q\|^{1+1/\kappa}_{L^1(I\times\omega)},$$

where $c_2 > 0$ is a new constant only depending on the quantities c_0, c_1, κ, q_a, and q_b. Hence, it follows that

$$\delta\nu + c_2|\delta\nu|^{p'} + c_1|\delta\nu|^2 \geq \frac{c_0\bar\nu}{2} \|q - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)}.$$

For $|\delta\nu| \leq \min\{ (3c_1)^{-1}, (3c_2)^{-1/(p'-1)} \}$ we deduce that

$$\delta\nu + \frac{2}{3}|\delta\nu| \geq \frac{c_0\bar\nu}{2} \|q - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)},$$

which in particular implies that $\nu - \bar\nu = \delta\nu \geq 0$ and we conclude the growth condition with $\varepsilon = \min\{ \bar\nu/2, (3c_1)^{-1}, (3c_2)^{-1/(p'-1)} \}$ and $c = 3c_0\bar\nu/10$. □

3.3.2. Stability analysis with respect to α

Last, we discuss the stability of $(\hat{P}_0)$ with respect to the regularization parameter α. Let a locally optimal solution $(\bar{\nu}, \bar{q})$ of $(\hat{P}_0)$, i.e. with $\alpha = 0$ be given. We would like to approximate this solution by locally optimal solutions $(\bar{\nu}_\alpha, \bar{q}_\alpha)$ for $(\hat{P}_\alpha)$ with regularization parameter $\alpha > 0$. Of course, we are interested in estimating the order of convergence.

Proposition 3.32. *Let $(\bar{\nu}, \bar{q})$ be a globally optimal solution to $(\hat{P}_0)$ and $(\bar{\nu}_\alpha, \bar{q}_\alpha)$ be a globally optimal solution to $(\hat{P}_\alpha)$ for some $\alpha > 0$. Then*

$$0 \leq \bar{\nu}_\alpha - \bar{\nu} \leq \frac{\bar{\nu}}{2} C_{Q_{ad}}\, \alpha, \tag{3.40}$$

where $C_{Q_{ad}} = \max_{q \in Q_{ad}} \|q\|^2_{L^2(\omega)}$.

Proof. Since $\bar{\nu}$ is globally optimal for $(\hat{P}_0)$, we infer $\bar{\nu} \leq \bar{\nu}_\alpha$. Similarly, as $(\bar{\nu}_\alpha, \bar{q}_\alpha)$ is globally optimal for $(\hat{P}_\alpha)$, we have

$$\bar{\nu}_\alpha \leq j_\alpha(\bar{\nu}_\alpha, \bar{q}_\alpha) \leq j_\alpha(\bar{\nu}, \bar{q}) = \bar{\nu} \left(1 + \frac{\alpha}{2} \int_0^1 \|\bar{q}\|^2_{L^2(\omega)}\right).$$

Combining both estimates yields

$$\bar{\nu}_\alpha \leq \bar{\nu} \left(1 + \frac{\alpha}{2} \int_0^1 \|\bar{q}\|^2_{L^2(\omega)}\right) \leq \bar{\nu} \left(1 + \frac{\alpha}{2} C_{Q_{ad}}\right),$$

where $C_{Q_{ad}} = \max_{q \in Q_{ad}} \|q\|^2_{L^2(\omega)}$, from which we conclude (3.40). □

Proposition 3.33. *Let $\{(\bar{\nu}_\alpha, \bar{q}_\alpha)\}_{\alpha > 0}$ be a sequence of global solutions of $(\hat{P}_\alpha)$. Then $\bar{\nu}_\alpha \to \bar{\nu}$ in $\mathbb{R}_+$ and $\bar{q}_\alpha \to q^*$ in $L^r(I \times \omega)$ as $\alpha \to 0$ for some $q^* \in Q_{ad}(0,1)$ and any $r \in [1, \infty)$. Moreover, the pair $(\bar{\nu}, q^*)$ is a global solution of $(\hat{P}_0)$.*

Proof. From Proposition 3.32 we immediately infer $\bar{\nu}_\alpha \to \bar{\nu}$. Moreover, due to boundedness of $\bar{q}_\alpha$ in $Q_{ad}(0,1)$, there is a subsequence, denoted in the same way for simplicity, such that $\bar{q}_\alpha \rightharpoonup q^*$ in $L^s(I \times \omega)$ with some fixed $s > 2$ and $q^* \in Q_{ad}(0,1)$. In the last step we have used sequentially compactness of $Q_{ad}(0,1)$ with respect to the weak star topology.

Employing feasibility of $(\bar{\nu}_\alpha, \bar{q}_\alpha)$ for $(\hat{P}_\alpha)$, i.e. $g(\bar{\nu}_\alpha, \bar{q}_\alpha) \leq 0$, we find

$$g(\bar{\nu}, q^*) \leq g(\bar{\nu}_\alpha, \bar{q}_\alpha) + |g(\bar{\nu}, q^*) - g(\bar{\nu}_\alpha, \bar{q}_\alpha)| = |g(\bar{\nu}, q^*) - g(\bar{\nu}_\alpha, \bar{q}_\alpha)|.$$

Hence, complete continuity of the mapping $(\nu, q) \mapsto i_1 S(\nu, q)$, see Proposition A.20, and passing to the limit $\alpha \to 0$ imply feasibility of the pair $(\bar{\nu}, q^*)$ for $(\hat{P}_0)$. In summary, $(\bar{\nu}, q^*)$ is a global solution of $(\hat{P}_0)$ and it remains to verify the convergence.

Since $(\bar{\nu}, q^*)$ is also feasible for $(\hat{P}_\alpha)$, we infer

$$\bar{\nu}_\alpha \left(1 + \frac{\alpha}{2} \int_0^1 \|\bar{q}_\alpha\|^2_{L^2(\omega)}\right) = j_\alpha(\bar{\nu}_\alpha, \bar{q}_\alpha) \leq j_\alpha(\bar{\nu}, q^*) = \bar{\nu} \left(1 + \frac{\alpha}{2} \int_0^1 \|q^*\|^2_{L^2(\omega)}\right).$$

Because $\bar{\nu} \leq \bar{\nu}_\alpha$, the above estimate implies

$$\int_0^1 \|\bar{q}_\alpha\|^2_{L^2(\omega)} \leq \int_0^1 \|q^*\|^2_{L^2(\omega)}.$$

Using weak lower semicontinuity we obtain

$$\int_0^1 \|q^*\|^2_{L^2(\omega)} \le \liminf_{\alpha\to 0} \int_0^1 \|\bar q_\alpha\|^2_{L^2(\omega)} \le \limsup_{\alpha\to 0} \int_0^1 \|\bar q_\alpha\|^2_{L^2(\omega)} \le \int_0^1 \|q^*\|^2_{L^2(\omega)}.$$

This gives $\int_0^1 \|\bar q_\alpha\|^2_{L^2(\omega)} \to \int_0^1 \|q^*\|^2_{L^2(\omega)}$ as $\alpha \to 0$, which implies $\bar q_\alpha \to q^*$ in $L^2(I\times\omega)$ due to weak convergence. The convergence result in $L^r(I\times\omega)$ for $r \in [1,\infty)$ follows from Hölder's inequality and the control constraints. □

Theorem 3.34. *Let $(\bar\nu, \bar q)$ be a local solution to $(\hat P_0)$ and suppose that there exist constants $\varepsilon > 0$, $c > 0$, and $\kappa > 0$ such that the growth condition*

$$c\|q - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)} \le \nu - \bar\nu, \tag{3.41}$$

holds for all admissible $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ with $|\nu - \bar\nu| + \|q - \bar q\|_{L^1(I\times\omega)} \le \varepsilon$. Then there are constants $\alpha_0, c > 0$, and a sequence of local solutions $\{(\bar\nu_\alpha, \bar q_\alpha)\}_{\alpha>0}$ of $(\hat P_\alpha)$ such that

$$0 \le \bar\nu_\alpha - \bar\nu \le c\alpha \quad \text{and} \quad \|\bar q_\alpha - \bar q\|_{L^1(I\times\omega)} \le c\alpha^\kappa$$

for all $0 < \alpha \le \alpha_0$.

Proof. We apply a localization argument, cf. [32], and introduce the auxiliary problem

$$\inf_{\substack{\nu_\alpha\in\mathbb{R}_+\\ q_\alpha\in Q_{ad}(0,1)}} j_\alpha(\nu_\alpha, q_\alpha) \quad \text{subject to} \quad \begin{cases} g(\nu_\alpha, q_\alpha) \le 0, \\ |\nu_\alpha - \bar\nu| + \|q_\alpha - \bar q\|_{L^1(I\times\omega)} \le \rho, \end{cases} \tag{3.42}$$

where $\rho = \varepsilon > 0$ is from the growth condition (3.41). Noting that $g(\bar\nu, \bar q) = 0$, i.e. the admissible set for (3.42) is nonempty, existence of at least one solution $(\bar\nu^\rho_\alpha, \bar q^\rho_\alpha)$ to (3.42) follows by standard arguments. Moreover, similar as in the proofs of Propositions 3.32 and 3.33 one can verify that

$$0 \le \bar\nu^\rho_\alpha - \bar\nu \le c\alpha,$$

and $\bar q^\rho_\alpha \to q^*$ in $L^1(I\times\omega)$ for some $q^* \in Q_{ad}(0,1)$. Hence, the growth condition (3.41) implies

$$\begin{aligned} \|q^* - \bar q\|_{L^1(I\times\omega)} &\le \|q^* - \bar q^\rho_\alpha\|_{L^1(I\times\omega)} + \|\bar q^\rho_\alpha - \bar q\|_{L^1(I\times\omega)} \\ &\le \|q^* - \bar q^\rho_\alpha\|_{L^1(I\times\omega)} + c\,(\bar\nu^\rho_\alpha - \bar\nu)^{\kappa/(1+\kappa)} \\ &\le \|q^* - \bar q^\rho_\alpha\|_{L^1(I\times\omega)} + c\alpha^{\kappa/(1+\kappa)} \to 0 \end{aligned}$$

as $\alpha \to 0$. This clearly forces $q^* = \bar q$. Therefore, for $\alpha > 0$ sufficiently small the auxiliary constraint in (3.42) is not active and $(\bar\nu^\rho_\alpha, \bar q^\rho_\alpha)$ is a local solution to $(\hat P_\alpha)$, so that we will omit the additional super-index ρ in the following.

Using again the growth condition and feasibility of $(\bar\nu, \bar q)$ for $(\hat P_\alpha)$, we find

$$\begin{aligned} c\|\bar q_\alpha - \bar q\|^{1+1/\kappa}_{L^1(I\times\omega)} + \frac{\alpha}{2}\bar\nu_\alpha \int_0^1 \|\bar q_\alpha\|^2_{L^2(\omega)} &\le \bar\nu_\alpha - \bar\nu + \frac{\alpha}{2}\bar\nu_\alpha \int_0^1 \|\bar q_\alpha\|^2_{L^2(\omega)} = j_\alpha(\bar\nu_\alpha, \bar q_\alpha) - \bar\nu \\ &\le j_\alpha(\bar\nu, \bar q) - \bar\nu = \frac{\alpha}{2}\bar\nu \int_0^1 \|\bar q\|^2_{L^2(\omega)}. \end{aligned}$$

Thus, since $\bar{\nu} \leq \bar{\nu}_\alpha$, we obtain

$$\begin{aligned} c\|\bar{q}_\alpha - \bar{q}\|^{1+1/\kappa}_{L^1(I\times\omega)} &\leq \frac{\alpha}{2}\left(\bar{\nu}\int_0^1 \|\bar{q}\|^2_{L^2(\omega)} - \bar{\nu}_\alpha \int_0^1 \|\bar{q}_\alpha\|^2_{L^2(\omega)}\right) \\ &\leq \frac{\alpha}{2}\bar{\nu}_\alpha\left(\int_0^1 \|\bar{q}\|^2_{L^2(\omega)} - \int_0^1 \|\bar{q}_\alpha\|^2_{L^2(\omega)}\right) \\ &= \frac{\alpha}{2}\bar{\nu}_\alpha \int_0^1 (\bar{q} + \bar{q}_\alpha, \bar{q} - \bar{q}_\alpha) \leq c\alpha\|\bar{q}_\alpha - \bar{q}\|_{L^1(I\times\omega)}, \end{aligned}$$

where we have used Hölder's inequality and that $\bar{q}_\alpha, \bar{q} \in Q_{ad}(0,1) \subset L^\infty(I\times\omega)$ with uniform bound independent of α in the last inequality. □

4. Optimization algorithms

This chapter is devoted to the theoretical and practical aspects concerning the numerical solution of the time-optimal control problem (P). We consider the general formulation of Chapter 2, where we in addition restrict ourselves to the choice

$$L(q) = \frac{\alpha}{2}\|q\|_Q^2 \quad \text{for } \alpha \geq 0.$$

In the case $\alpha > 0$, the resulting problems can be solved by standard methods. Therefore, in Section 4.1 we will only discuss one method, namely the augmented Lagrangian method to deal with the state constraint. For the particular case that U is the sublevel set of a smooth function G, i.e.

$$U = \{\, u \in H \colon G(u) \leq 0 \,\},$$

we will prove convergence of the augmented Lagrangian method under certain assumptions. Here, we essentially follow the presentation from [81, Chapter 3]. For the subproblems arising in the augmented Lagrangian method we will briefly discuss a bilevel optimization and a monolithic approach; cf. [92, 93]. Since all algorithms will be analyzed in a function space setting, we expect that appropriate realizations of the algorithms will show mesh independence, i.e. the number of iterations is essentially independent of the number of degrees of freedom of a concrete discretization. The discretization of the state and adjoint state equations by means of the Galerkin method will be discussed in detail in Chapter 5.

In order to solve the time-optimal control problem in the case $\alpha = 0$, a straightforward approach consists in solving the regularized problems for a monotonically decreasing sequence of regularization parameters $\alpha_1 > \alpha_2 > \ldots > 0$ such that $\lim_{n\to\infty} \alpha_n = 0$. In view of the stability results from Section 3.3.2, the corresponding solutions to the regularized problems converge to a solution of the original problem. However, with decreasing values of α_n the associated problems become computationally very expensive. To this end, in Section 4.2 we will discuss a different approach that relies on a certain equivalence of minimal time and minimal distance controls. This allows for a reformulation of (P), where we can separate the nonconvex influence of T and the convex structure of the remaining problem. Leading again to a bilevel optimization problem, we will discuss different methods for the numerical treatment of the outer and inner minimization problems. Numerical examples indicate that the resulting algorithm is capable to solve the problem up to high precision in reasonable time. Thus, it seems to be at least competitive with the regularization strategy. The equivalence of minimal time and minimal distance controls can be related to the equivalence of minimal time and minimal norm controls that is well-known in the literature; see, e.g, [54, 62, 89, 160]. In [160, Remark 3.3] the latter equivalence has been proposed for the algorithmic treatment of time-optimal control problems. Inspired by [160], an algorithm based on the bisection method for the numerical solution of time-optimal control problems subject to ordinary differential equations has been discussed in [109]. However, to the best of the authors knowledge an algorithm for the setting subject to partial differential equations has not been studied so far. We will compare these different approaches concerning the bang-bang case at the end of the chapter.

4.1. Optimization algorithms for $\alpha > 0$

In this section we discuss the algorithmic solution of (P) in the case $\alpha > 0$ by means of the augmented Lagrangian method. For the following considerations, we suppose that the terminal set U can be expressed as a sublevel set of a two times continuously differentiable function $G\colon H \to \mathbb{R}$, precisely,

$$U = \{\, u \in H \colon G(u) \leq 0 \,\}.$$

As before, we use $g(\nu, q) = G(i_1 S(\nu, q))$ for $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ to denote the reduced terminal constraint. Furthermore, we generally suppose that $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ is a locally optimal solution to $(\hat{P})$ that satisfies the qualified optimality conditions of Lemma 3.1.

4.1.1. Augmented Lagrangian method

The augmented Lagrangian method has been introduced by Hestenes [75] and Powell [133]. It can be seen as a hybrid method combining the multiplier and the penalty method. For any $c > 0$, we define the augmented Lagrangian as

$$\mathcal{L}_c(\nu, q, \mu) = j(\nu, q) + \mu g(\nu, q) + \frac{c}{2} |g(\nu, q)|^2.$$

Let $c_\bullet \geq 0$ be fixed. Consider a sequence $(c_n)_{n \in \mathbb{N}}$ of nondecreasing penalty parameters to be specified later such that $c_n \geq c_\bullet$. In each iteration, for μ_{n-1} from the previous iteration, we determine (ν_n, q_n) as the solution to

$$\min \mathcal{L}_{c_n}(\nu, q, \mu_{n-1}) \quad \text{subject to} \quad (\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1). \tag{4.1}$$

Clearly, for the choice $\mu_{n-1} = 0$ we obtain the quadratic penalty method and for $c_n = 0$ the multiplier method. The advantage of the augmented Lagrangian method over penalty methods is that it avoids the necessity of increasing the penalization parameter c_n to infinity, which typically leads to ill-conditioning of the involved problems. The Lagrange multiplier is updated by the rule

$$\mu_n = \mu_{n-1} + (c_n - c_\bullet) g(\nu_n, q_n).$$

We summarize the resulting method in Algorithm 1.

Algorithm 1: Augmented Lagrangian method

Choose $\mu_0 > 0$ and set $n = 1$;
do
  Find a solution (ν_n, q_n) to

$$\min \mathcal{L}_{c_n}(\nu, q, \mu_{n-1}) \quad \text{subject to} \quad (\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$$

  Update $\mu_n = \mu_{n-1} + (c_n - c_\bullet) g(\nu_n, q_n)$ and set $n = n + 1$;
while $|g(\nu_n, q_n)| > \varepsilon_{tol}$;

The augmented Lagrangian method has been extensively studied in the context of finite dimensional problems; see in addition to the references given above [14, 15] and [57, Chapter 1]. For the infinite dimensional case we can refer to [131, 142] as well as the monographs [57, Chapter 3] and [81, Chapter 3]. We have the following convergence result.

Proposition 4.1. *Suppose there are $c_\bullet \geq 0$, $\varepsilon > 0$, and $\delta > 0$ such that the quadratic growth condition*

$$j(\bar{\nu}, \bar{q}) + \frac{\delta}{2}|\nu - \bar{\nu}|^2 + \frac{\delta}{2}\|q - \bar{q}\|^2_{Q(0,1)} \leq \mathcal{L}_{c_\bullet}(\nu, q, \bar{\mu})$$

holds for all $(\nu, q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ such that $|\nu - \bar{\nu}|^2 + \|q - \bar{q}\|^2_{Q(0,1)} \leq \varepsilon$. Consider a sequence of penalty parameters with $c_n > c_\bullet$. Let (ν_n, q_n) and μ_n be defined by the augmented Lagrangian method and suppose that $|\nu_n - \bar{\nu}|^2 + \|q_n - \bar{q}\|^2_{Q(0,1)} \leq \varepsilon$. Then for any $n \geq 1$ and $\sigma_n = c_n - c_\bullet$ the estimate

$$\delta|\nu_n - \bar{\nu}|^2 + \delta\|q_n - \bar{q}\|^2_{Q(0,1)} + \frac{1}{\sigma_n}|\mu_n - \bar{\mu}|^2 \leq \frac{1}{\sigma_n}|\mu_{n-1} - \bar{\mu}|^2 \tag{4.2}$$

holds. In particular, this implies

$$|\nu_n - \bar{\nu}|^2 + \|q_n - \bar{q}\|^2_{Q(0,1)} \leq \frac{1}{\sigma_n\delta}|\mu_{n-1} - \bar{\mu}|^2 \leq \frac{1}{\sigma_n\delta}|\mu_0 - \bar{\mu}|^2 \tag{4.3}$$

and

$$\sum_{n=1}^{\infty} \sigma_n \left(|\nu_n - \bar{\nu}|^2 + \|q_n - \bar{q}\|^2_{Q(0,1)}\right) \leq \frac{1}{\delta}|\mu_0 - \bar{\mu}|^2. \tag{4.4}$$

Proof. The result can be shown as in [81, Theorem 3.8]. Since the proof is short and instructive, we give it for the convenience of the reader. First, we have

$$\begin{aligned}
\mathcal{L}_{c_n}(\nu_n, q_n, \mu_{n-1}) &= j(\nu_n, q_n) + \mu_{n-1}g(\nu_n, q_n) + \frac{c_n}{2}|g(\nu_n, q_n)|^2 \\
&= \mathcal{L}_{c_\bullet}(\nu_n, q_n, \bar{\mu}) + (\mu_{n-1} - \bar{\mu})g(\nu_n, q_n) + \frac{c_n - c_\bullet}{2}|g(\nu_n, q_n)|^2 \\
&= \mathcal{L}_{c_\bullet}(\nu_n, q_n, \bar{\mu}) + (\mu_{n-1} - \bar{\mu})g(\nu_n, q_n) + \frac{1}{2}(\mu_n - \mu_{n-1})g(\nu_n, q_n) \\
&= \mathcal{L}_{c_\bullet}(\nu_n, q_n, \bar{\mu}) + \frac{1}{2\sigma_n}\left(\mu_{n-1} + \mu_n - 2\bar{\mu}\right)\left(\mu_n - \mu_{n-1}\right) \\
&= \mathcal{L}_{c_\bullet}(\nu_n, q_n, \bar{\mu}) + \frac{1}{2\sigma_n}\left(|\mu_n - \bar{\mu}|^2 - |\mu_{n-1} - \bar{\mu}|^2\right).
\end{aligned}$$

Since $\mathcal{L}_{c_n}(\nu_n, q_n, \mu_{n-1}) \leq \mathcal{L}_{c_n}(\bar{\nu}, \bar{q}, \mu_{n-1}) = j(\bar{\nu}, \bar{q})$, we arrive at

$$\frac{1}{2\sigma_n}|\mu_n - \bar{\mu}|^2 \leq j(\bar{\nu}, \bar{q}) - \mathcal{L}_{c_\bullet}(\nu_n, q_n, \bar{\mu}) + \frac{1}{2\sigma_n}|\mu_{n-1} - \bar{\mu}|^2.$$

Finally, the growth condition yields the first estimate (4.2). The last estimates (4.3) and (4.4) follow from the first. □

Remark 4.2. The quadratic growth condition of Proposition 4.1 can be deduced from a strong second order sufficient optimality condition. It would be desirable to prove the supposition of Proposition 4.1 assuming only that the second order sufficient condition of Theorem 3.13 holds.

Proposition 4.3. *Adapt the assumptions of Proposition 4.1. There is $c > 0$ independent of n such that*

$$|\mu_n - \bar{\mu}| \leq \frac{c}{\sqrt{\sigma_n\delta}}|\mu_{n-1} - \bar{\mu}|$$

for all $n \geq 1$.

Proof. Since the constraint for ν is not active, we can still argue as in the proof of [81, Theorem 3.10], even though the problem with controls in a convex and closed set $Q_{ad}(0,1)$ is not included in the setting considered in [81, Chapter 3]. From the optimality conditions for (4.1) and the optimality conditions for the original problem we infer that

$$\begin{aligned}\partial_\nu j(\nu_n, q_n) + (\mu_{n-1} + c_n g(\nu_n, q_n))\, \partial_\nu g(\nu_n, q_n) = 0,\\ \partial_\nu j(\bar\nu, \bar q) + \bar\mu \partial_\nu g(\bar\nu, \bar q) = 0.\end{aligned}$$

Abbreviating $\tilde\mu_n = \mu_{n-1} + c_n g(\nu_n, q_n)$ and summing both equalities imply

$$(\tilde\mu_n - \bar\mu)\partial_\nu g(\bar\nu, \bar q) = \frac{\alpha}{2}\left(\|\bar q\|^2_{Q(0,1)} - \|q_n\|^2_{Q(0,1)}\right) + (\partial_\nu g(\bar\nu, \bar q) - \partial_\nu g(\nu_n, q_n))\,\tilde\mu_n. \tag{4.5}$$

Moreover, $\tilde\mu_n - \mu_n = c_\bullet g(\nu_n, q_n)$. Hence, $\tilde\mu_n$ is uniformly bounded due to boundedness of (ν_n, q_n, μ_n) that is guaranteed by Proposition 4.1. Since

$$\mu_n - \bar\mu = (\tilde\mu_n - \bar\mu) + c_\bullet\,(g(\nu_n, q_n) - g(\bar\nu, \bar q))$$

the desired estimate follows from (4.5), Lipschitz continuity of g and g', and (4.3). Here we have used that qualified optimality conditions hold at $(\bar\nu, \bar q)$, i.e. $\partial_\nu g(\bar\nu, \bar q) \neq 0$. □

In view of Proposition 4.1 for c_n sufficiently large, the iterates (ν_n, q_n) converge to $(\bar\nu, \bar q)$ at least at the same rate as μ_n converges to $\bar\mu$. If $c_n \to c^\bullet$ for $n \to \infty$ with some $c^\bullet < \infty$ and $c_n \geq c_\bullet$ sufficiently large, then from Proposition 4.3 we deduce that μ_n converges q-linearly to $\bar\mu$. If $c^\bullet = \infty$, then we even obtain q-superlinear convergence of μ_n.

It seems that for the choice of the sequence of penalty parameters c_n there is no general rule. In the numerical experiments we therefore take the heuristic from [14, p. 405]. If the constraint violation measured in $|g(\nu_n, q_n)|$ is not decreased by a certain factor, then the penalty parameter is multiplied by a factor (say $2 - 10$).

4.1.2. Bilevel optimization

In order to implement Algorithm 1, we have to determine a solution to (4.1) for given μ_{n-1} and c_n. A bilevel approach consists of splitting the optimization in two steps, where we optimize for ν in the outer loop and for q in the inner loop. Whence, we obtain

$$\min_{\nu\in\mathbb{R}_+} \min_{q\in Q_{ad}(0,1)} \mathcal{L}_{c_n}(\nu, q, \mu_{n-1}). \tag{4.6}$$

Clearly, the optimization problems (4.1) and (4.6) are equivalent. In the context of time-optimal control problems a bilevel approach has also been proposed in [93, 94].

Let us introduce the value function of the subproblem as

$$V(\nu) = \min_{q\in Q_{ad}(0,1)} \mathcal{L}_{c_n}(\nu, q, \mu_{n-1}).$$

We denote by $\bar q(\nu)$ an optimal solution to the minimization problem above. For the solution of the outer problem we are interested in continuity and differentiability properties of V. To this end we require the notion of polyhedricity; see Definition 3.22. Recall from Proposition 3.23 that in case of box constraints for the control the set of admissible controls is polyhedric.

Proposition 4.4. *Let $\nu \in \mathbb{R}_+$ and $\mu_{n-1} \in \mathbb{R}_+$. Suppose that $Q_{ad}(0,1)$ is polyhedric and that $G''(u)[\cdot,\cdot]$ is positive semi-definite for all $u \in H$. Moreover, let $\mu_{n-1} + c_n g(\nu, \bar{q}(\nu)) \geq 0$. Then the value function V is differentiable with locally Lipschitz continuous derivative and we have the expression*

$$V'(\nu) = 1 + \frac{\alpha}{2}\|\bar{q}(\nu)\|^2_{Q(0,1)} + (\mu_{n-1} + c_n g(\nu, \bar{q}(\nu))) \int_0^1 \langle B\bar{q}(\nu) - Au, z\rangle, \tag{4.7}$$

where $u = S(\nu, \bar{q}(\nu))$ and $z \in W(0,1)$ satisfies

$$-\partial_t z + \nu A^* z = 0, \quad z(1) = G'(u(1))^*. \tag{4.8}$$

Moreover, V admits a second order directional derivative

$$\begin{aligned} V''(\nu)\delta\nu = (\mu_{n-1} + c_n g(\nu, \bar{q}(\nu))) \left[\int_0^1 \langle B\bar{q}(\nu) - Au, \delta z\rangle - \langle A\delta u, z\rangle\right] \\ + c_n \int_0^1 \langle B\bar{q}(\nu) - Au, z\rangle \int_0^1 \langle B\bar{q}(\nu) - Au + \nu B\delta q(\delta\nu), z\rangle, \end{aligned}$$

where $\delta q(\delta\nu)$ is the directional derivative of $\bar{q}(\nu)$ in direction $\delta\nu$ that is determined by a variational inequality (given in the proof below), $\delta u = S'(\nu, \bar{q}(\nu))(\delta\nu, \delta q(\delta\nu))$, and δz is the solution to

$$-\partial_t \delta z + \nu A^* \delta z = -\delta\nu A^* z, \quad \delta z(1) = G''(u(1))[\delta u(1), \cdot]^*. \tag{4.9}$$

Proof. The proof relies on established arguments where we refer to [21, Section 5.1]; cf. also [67, Proposition 3.16] and [81, Chapter 2]. For convenience we abbreviate

$$f(\nu, q) \coloneqq \mathcal{L}_{c_n}(\nu, q, \mu_{n-1}).$$

The chain rule yields

$$\partial_q f(\nu, q)\delta q = \alpha\nu(q, \delta q)_{L^2(I\times\omega)} + (\mu_{n-1} + c_n g(\nu, q))\, \partial_q g(\nu, q)\delta q.$$

Since $q \mapsto i_1 S(\nu, q)$ is affine linear and $G''(u)$ is positive semi-definite we obtain

$$\begin{aligned} \partial_{qq} f(\nu, \bar{q}(\nu))\delta q^2 &= \alpha\nu\|\delta q\|^2_{Q(0,1)} + c_n\left(\partial_q g(\nu, \bar{q}(\nu))\right)^2 + (\mu_{n-1} + c_n g(\nu, \bar{q}(\nu)))\, \partial_{qq} g(\nu, \bar{q}(\nu))\delta q^2 \\ &\geq \alpha\nu\|\delta q\|^2_{Q(0,1)} \end{aligned}$$

for all $\delta q \in Q(0,1)$. Hence, f satisfies a strong second order sufficient optimality condition. Therefore, according to [21, Proposition 5.2 (ii)], the mapping $\nu \mapsto \bar{q}(\nu)$ is locally Lipschitz continuous.

Let $\nu \in \mathbb{R}_+$ and $\tau_j \in \mathbb{R}$ such that $\tau_j \to 0$. Set $q_j = \bar{q}(\nu + \tau_j)$ and $q = \bar{q}(\nu)$. Employing local Lipschitz continuity of $\nu \mapsto \bar{q}(\nu)$, we conclude that the quotient $\tau_j^{-1}(q_j - q)$ converges weakly to some $\delta q \in Q(0,1)$. In addition, since $\partial_{qq} f(\nu, q)$ is elliptic, it defines a Legendre form; see [21, Proposition 3.76]. According to [21, Theorem 5.5], the weak limit δq is in fact a strong limit and satisfies a so-called linearized variational inequality: Find $\delta q \in C$ such that

$$F(\delta\nu, \delta q)(\xi - \delta q) \geq 0 \quad \text{for all } \xi \in C_{\bar{q}(\nu)},$$

where C_q denotes the critical cone

$$C_q \coloneqq T_{Q_{ad}(0,1)} \cap \{\, \delta q \in L^2(I\times\omega) \colon \partial_q f(\nu, q)\delta q = 0 \,\},$$

and the functional F is defined by

$$\begin{aligned} F(\delta\nu,\delta q)(\cdot) = {} & \alpha(\delta\nu\bar{q}(\nu)+\nu\delta q,\cdot)_{L^2(I\times\omega)} \\ & + (\mu_{n-1}+c_n g(\nu,\bar{q}(\nu)))\,(\partial_{\nu q}g(\nu,\bar{q}(\nu))(\delta\nu,\cdot)+\partial_{qq}g(\nu,\bar{q}(\nu))[\delta q,\cdot]) \\ & + c_n\,(\partial_\nu g(\nu,\bar{q}(\nu))\delta\nu+\partial_q g(\nu,\bar{q}(\nu))\delta q)\,\partial_q g(\nu,\bar{q}(\nu))(\cdot). \end{aligned}$$

Since $\partial_q f(\nu,q)\delta q = 0$, we finally obtain

$$\begin{aligned} \tau_j^{-1}\left[V(\nu+\tau_j)-V(\nu)\right] &= \tau_j^{-1}\left[f(\nu+\tau_j,q_j)-f(\nu,q_j)+f(\nu,q_j)-f(\nu,q)\right] \\ &\to \partial_\nu f(\nu,q)+\partial_q f(\nu,q)\delta q = \partial_\nu f(\nu,q). \end{aligned}$$

The concrete expression (4.7) for $\partial_\nu f(\nu,q)$ follows as in Proposition 2.21. Moreover, from (4.7), local Lipschitz continuity of $\nu\mapsto\bar{q}(\nu)$, and Lipschitz stability of the solution to the state and adjoint state equation, we further deduce that $\nu\mapsto V'(\nu)$ is locally Lipschitz continuous.

The formula for the second derivative follows by total directional differentiation of the expression for $V'(\nu)$ and using again $\partial_q f(\nu,q)\delta q = 0$. □

Note that if $\mu_{n-1}>0$ the additional assumption $\mu_{n-1}+c_n g(\nu,\bar{q}(\nu))\geq 0$ in Proposition 4.4 is satisfied at least close to an optimal solution $(\bar{\nu},\bar{q})$.

In view of Proposition 4.4 in order to solve the outer loop of the bilevel optimization problem, we have to determine $\bar{\nu}\in\mathbb{R}_+$ such that $V'(\bar{\nu})=0$. Thus, the second derivative of V allows for a semismooth Newton method to solve the optimization problem; cf. also [94, Algorithm 2]. Alternatively, one can employ a derivative free optimization method such as a bisection type method to find a minimum of the value function V. This avoids the calculation of the directional derivative $\delta q(\delta\nu)$.

4.1.3. Monolithic optimization

We next discuss an alternative approach to the bilevel optimization for the solution of the subproblem (4.1) arising in the augmented Lagrangian method. In contrast to the previous subsection, we consider the joint optimization for the free terminal time and the control in one combined optimization variable $(\nu,q)\in\mathbb{R}_+\times Q_{ad}(0,1)$. We therefore refer to this approach as the monolithic optimization; cf. also [92].

The resulting optimization problem is nonlinear and subject to control constraints. Using the differentiability results for the control-to-state mapping Proposition 2.20 we infer that the reduced objective functional is twice continuously differentiable. Hence, one can use standard optimization methods for the solution of (4.1). To efficiently deal with the control constraints, we employ the semismooth Newton method proposed in [92] for time-optimal control of the monodomain equations. It is based on an equivalent reformulation of the problem by means of the normal map due to Robinson [135].

The method from [92] can be directly applied to our setting with slight modifications. Instead of the terminal tracking formulation of [92, Section 3.2], we take the augmented Lagrangian as the objective functional. We sketch the main steps in the derivation of the method and define the "normal map" for our purposes as

$$F(\nu,\psi) = \partial_{(\nu,q)}\mathcal{L}_{c_n}(\nu,P_{Q_{ad}}(\psi),\mu_{n-1})^* + \begin{pmatrix} 0 \\ \alpha\nu(\psi-P_{Q_{ad}}(\psi)) \end{pmatrix}, \quad (\nu,\psi)\in\mathbb{R}_+\times Q(0,1).$$

The crucial observation is that if $(\nu_n, q_n) \in \mathbb{R} \times Q_{ad}(0,1)$ is a local solution to (4.1), then there exists $\psi_n \in Q(0,1)$ such that $q_n = P_{Q_{ad}}(\psi_n)$ and $F(\nu_n, \psi_n) = 0$; cf. [130, Proposition 3.5] or [92, Proposition 5.6]. Hence, in order to solve the minimization problem (4.1), we have to determine zeros of F.

In the following we abbreviate $q = P_{Q_{ad}}(\psi)$ and $\tilde{\mu}_n = \mu_{n-1} + c_n g(\nu, q)$. Since the term $\alpha\nu P_{Q_{ad}}(\psi)$ in the second component of F cancels out, we have the expression

$$F(\nu, \psi) = \begin{pmatrix} 1 + \frac{\alpha}{2}\|q\|^2_{Q(0,1)} + \tilde{\mu}_n \int_0^1 \langle Bq - Au, z\rangle \\ \alpha\nu\psi + \nu\tilde{\mu}_n B^* z \end{pmatrix},$$

where $u = S(\nu, q)$ denotes the state and z is the solution to the adjoint state equation (4.8). To apply a Newton type method to $F(\nu, \psi) = 0$, we require the linearization of F. In the case of box constraints for Q_{ad} defined on a measure space (ω, ϱ), the generalized differential of $P_{Q_{ad}}$ can be given as $\mathrm{D}P_{Q_{ad}}(\psi)\,\delta\psi = \mathbb{1}_{\mathcal{I}}\delta\psi$, where $\mathbb{1}_{\mathcal{I}}$ denotes the indicator function associated to the set of inactive constraints

$$\mathcal{I} \coloneqq \{\, (t,x) \in I \times \omega \colon q_a(x) \le \psi(t,x) \le q_b(x) \,\}.$$

With a chain rule for nonsmooth operators, we calculate the generalized derivative of F at the point (ν, ψ) as

$$\mathrm{D}F(\nu,\psi)(\delta\nu,\delta\psi) = \begin{pmatrix} \alpha(q,\delta q)_{Q(0,1)} + \tilde{\mu}_n \int_0^1 \left[\langle Bq - Au, \delta z\rangle + \langle B\delta q - A\delta u, z\rangle\right] \mathrm{d}t \\ \quad + c_n \left[\int_0^1 \langle Bq - Au, z\rangle \,\mathrm{d}t\right] \int_0^1 \left[\langle Bq - Au, z\rangle\delta\nu + \nu(B^*z, \delta q)_Q\right] \mathrm{d}t \\ \alpha\left[\delta\nu\psi + \nu\delta\psi\right] + \tilde{\mu}_n\left[\delta\nu B^* z + \nu B^*\delta z\right] \\ \quad + \nu c_n \int_0^1 \left[\langle Bq - Au, z\rangle\delta\nu + (B^*z, \delta q)_Q\right] \mathrm{d}t\, B^* z \end{pmatrix},$$

where $\delta q = \mathbb{1}_{\mathcal{I}}\delta\psi$, $\delta u = S'(\nu, q)(\delta\nu, \delta q)$ denotes the linearized state, and δz is the solution to the second adjoint state equation (4.9). For the convergence analysis of this semismooth Newton method and globalization approaches we refer to [130, Chapter 3].

4.2. An algorithmic approach for bang-bang controls ($\alpha = 0$)

After the discussion of algorithms for the time-optimal control problems with a fixed cost parameter $\alpha > 0$, we now consider an algorithmic approach for the case of bang-bang controls, i.e. $\alpha = 0$. First of all we will prove the equivalence of minimal time and minimal distance controls. This gives rise to a reformulation of the time-optimal control problem (P) which allows to separate the minimization for the terminal time T and for the control q.

4.2.1. Equivalence of time and distance optimal controls

For given $\delta \ge 0$ consider the perturbed time-optimal control problem

$$\inf_{\substack{T>0 \\ q \in Q_{ad}(0,T)}} \quad T \quad \text{subject to} \quad u[q](T) \in U_\delta, \tag{P_δ}$$

where $U_\delta = U + \overline{\mathcal{B}_\delta(0)} = \{\, u \in H \colon d_U(u) \le \delta \,\}$. Here, $u[q] \in W(0,T)$ denotes the solution to the state equation defined above.

Moreover, for fixed $T > 0$ we consider the *minimal distance control problem*

$$\inf_{q \in Q_{ad}(0,T)} d_U(u[q](T)), \tag{δ_T}$$

where $d_U(\cdot)$ denotes the distance function

$$d_U(u) \coloneqq \inf_{u' \in U} \|u - u'\|_H.$$

We define the value functions $T\colon [0,\infty) \to [0,\infty]$ and $\delta\colon [0,\infty) \to [0,\infty)$ as

$$T(\delta) = \inf(P_\delta) \quad \text{and} \quad \delta(T) = \inf(\delta_T).$$

From boundedness of Q_{ad}, linearity of the control-to-state mapping (for fixed $T > 0$), and weak lower semicontinuity of the distance function, we immediately infer that the value function $\delta(\cdot)$ is well-defined. However, to verify well-posedness of $T(\cdot)$ we require an additional assumption; cf. also Proposition 2.14 and Remark 2.15.

Proposition 4.5. *Let $\delta \geq 0$. If (P_δ) has a feasible point, then $T(\cdot)$ is well-defined on $[\delta, \infty)$.*

Proof. This result follows by standard arguments using the direct method. □

Throughout the rest of this chapter we assume that there exists a feasible point for $\delta = 0$.

Proposition 4.6. *Set $\delta^\bullet = d_U(u_0)$. The function $T\colon [0,\delta^\bullet] \to [0,\infty)$ is strictly monotonically decreasing and right-continuous.*

Proof. Step 1: T strictly decreasing. Clearly, T is monotonically decreasing. To show strict monotonicity, let $\delta_1 > \delta_2 \geq 0$. We have to show $T(\delta_1) < T(\delta_2)$. Suppose $T(\delta_1) = T(\delta_2)$ and let $(T(\delta_i), q_i) \in \mathbb{R}_+ \times Q_{ad}(0, T(\delta_i))$ be optimal solutions to (P_{δ_i}), $i = 1, 2$. Since

$$d_U(u[q_2](T(\delta_2))) = \delta_2 < \delta_1,$$

we infer that $(T(\delta_2), q_2)$ is also feasible for (P_{δ_1}). Note that in the problem formulation we can equivalently use $d_U(u[q](T)) \leq \delta$ and $d_U(u[q](T)) = \delta$. From continuity of $u[q_2]\colon [0, T(\delta_2)] \to H$ and $T(\delta_1) = T(\delta_2)$ we deduce that $(T(\delta_1), q_1)$ cannot be optimal for the time-optimal problem (P_{δ_1}). This contradicts the assumption and we conclude $T(\delta_1) < T(\delta_2)$.

Step 2: T is right-continuous. Consider a sequence $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_n \to \delta \in [0, \delta^\bullet)$. We have to show $\lim_{n\to\infty} T(\delta_n) = T(\delta)$. Assume that $\lim_{n\to\infty} T(\delta_n) \neq T(\delta)$. Then, due to monotonicity of T, there is $\varepsilon > 0$ such that

$$\lim_{n\to\infty} T(\delta_n) = T(\delta) - \varepsilon.$$

Let $q_n = q_n(\delta_n, T(\delta_n)) \in Q_{ad}(0, T(\delta_n))$ denote an optimal control to (P_{δ_n}). We can extend each q_n to the time-interval $(0, T(\delta))$ so that $q_n \in Q_{ad}(0, T(\delta))$ for all $n \in \mathbb{N}$. Due to boundedness of Q_{ad}, there is a subsequence denoted in the same way such that $q_n \rightharpoonup q$ in $Q(0, T(\delta))$ with $q \in Q_{ad}(0, T(\delta))$. Now, Lipschitz continuity of $d_U(\cdot)$ implies

$$\begin{aligned}
\lim_{n\to\infty} d_U(u[q_n](T(\delta_n))) &\geq \lim_{n\to\infty} d_U(u[q](T(\delta_n))) - c \lim_{n\to\infty} \|u[q](T(\delta_n)) - u[q_n](T(\delta_n))\| \\
&\geq \lim_{n\to\infty} d_U(u[q](T(\delta_n))) - c \lim_{n\to\infty} \sup_{t\in[0,T(\delta)]} \|u[q](t) - u[q_n](t)\| \\
&= \lim_{n\to\infty} d_U(u[q](T(\delta_n))),
\end{aligned}$$

where in the last step we have used compactness of the control-to-state mapping from $Q_{ad}(0, T(\delta))$ to $C([0, T(\delta)]; H)$; see Proposition A.20. Therefore, continuity of $u[q]\colon [0, T(\delta)] \to H$ yields

$$\delta = \lim_{n\to\infty} \delta_n = \lim_{n\to\infty} d_U(u[q_n](T(\delta_n))) \geq d_U(u[q](T(\delta) - \varepsilon)).$$

Thus, $(T(\delta) - \varepsilon, q)$ is admissible for (P_δ), contradicting optimality of $T(\delta)$. □

Proposition 4.7. *Let $T(\cdot)$ be left-continuous. Then $\delta(\cdot)$ is continuous and strictly monotonically decreasing. Moreover,*

$$T(\delta(T')) = T' \quad \text{for all } T' \in [0, T(0)] \tag{4.10}$$

and

$$\delta(T(\delta')) = \delta' \quad \text{for all } \delta' \in [0, \delta^\bullet]. \tag{4.11}$$

Proof. First, since T is strictly decreasing, its inverse T^{-1} is continuous. Moreover, as T is continuous, T^{-1} is defined everywhere on $[0, T(0)]$; see, e.g., [5, Theorem III.5.7].

Let $T > 0$. Then there exists $q \in Q_{ad}(0, T)$ such that $d_U(u[q](T)) = \delta(T)$. Hence, it holds $T(\delta(T)) \leq T$. Suppose that $T(\delta(T)) < T$. Then by continuity of T there exists $\delta' < \delta(T)$ such that $T(\delta') = T$. Let $q' \in Q_{ad}(0, T)$ be an optimal control to $(P_{\delta'})$. Then

$$\delta' < \delta(T) \leq d_U(u[q'](T)) \leq \delta',$$

a contradiction, which proves (4.10).

Moreover, (4.10) implies that $T(\delta(T(\delta'))) = T(\delta')$ for all $\delta' \in [0, \delta^\bullet]$. Strict monotonicity of T therefore yields (4.11). For these reasons, $\delta = T^{-1}$ and we conclude that δ is continuous and strictly monotonically decreasing. □

After this preparation we can now prove equivalence of time and distance optimal controls.

Lemma 4.8. *Let $T(\cdot)$ be left-continuous. If $T > 0$ and $q \in Q_{ad}(0, T)$ is distance-optimal for (δ_T), then (T, q) is also time-optimal for $(P_{\delta(T)})$. Conversely, if $\delta \geq 0$ and $(T, q) \in \mathbb{R}_+ \times Q_{ad}(0, T)$ is time-optimal for (P_δ), then q is also distance-optimal for (δ_T).*

Proof. Let $T > 0$ and $q \in Q_{ad}(0, T)$ be distance-optimal for (δ_T), i.e. $\delta(T) = d_U(u[q](T))$. Due to (4.10) we have $T(\delta(T)) = T$. Thus, (T, q) is also time-optimal for $(P_{\delta(T)})$.

Conversely, let $\delta \geq 0$ and $(T, q) \in \mathbb{R}_+ \times Q_{ad}(0, T)$ be time-optimal for (P_δ). In particular, this gives $d_U(u[q](T)) = \delta$. Using (4.11) we infer that

$$\delta(T(\delta)) = \delta = d_U(u[q](T)),$$

i.e. q is also distance-optimal for (δ_T). □

Remark 4.9. The assumption that the value function $T(\cdot)$ is left-continuous used in Proposition 4.7 and Lemma 4.8 generally depends on the state equation, the set of admissible controls, and the terminal constraint. Recall that in Theorem 2.18 we derived a sufficient condition for Lipschitz continuity of $T(\cdot)$ from the right. By similar techniques, we can derive a condition that guarantees Lipschitz continuity of $T(\cdot)$ from the left. As seen in Section 2.4, such conditions can be verified explicitly for many examples.

If (P_δ) is strongly stable on the right at $\delta = 0$, we immediately infer an estimate for the optimal times in terms of the minimal distances. Recall that strong stability on the right is defined as

$$T(0) - T(\delta') \leq \eta_0\, \delta', \quad \delta' \in [0, \varepsilon],$$

for constants $\eta_0 > 0$ and $\varepsilon > 0$; see Definition 2.16. Since $T(\cdot)$ and $\delta(\cdot)$ are inverse to each other, this implies the following estimate.

Corollary 4.10. *If (P_δ) is strongly stable on the right at $\delta = 0$, then*

$$0 \leq T - T' \leq \eta_0\, \delta(T')$$

for all $(T - \varepsilon\eta_0)^+ \leq T' \leq T$, where $T = T(0)$ is the optimal time for $\delta = 0$.

4.2.2. Regularization of the minimal distance problem

We will suppose throughout the rest of this section that $T(\cdot)$ is left continuous. In view of Lemma 4.8, we are interested in finding a root of the value function $\delta(\cdot)$ in order to solve the time-optimal control problem (P). This will generally lead to a bi-level optimization problem: In the outer loop we optimize for T and the inner loop determines for each given T a control with minimal distance to the target set. For the outer optimization we will discuss a bisection and a Newton method. Concerning the inner optimization, we will consider the conditional gradient method and the primal-dual active set strategy. Before we turn to the optimization methods, we will first introduce a regularized version of (δ_T).

Since $d_U(u[q](T)) = 0$ if and only if $d_U^2(u[q](T)) = 0$, we can alternatively minimize the squared distance function. For fixed $T > 0$ and $\alpha \geq 0$ we consider the regularized minimum squared distance control problem

$$\inf_{q \in Q_{ad}(0,T)} \frac{1}{2} d_U^2(u[q](T)) + \frac{\alpha}{2} \|q\|_{Q(0,T)}^2. \tag{4.12}$$

We emphasize that the unregularized case $\alpha = 0$ is not excluded. Again, we transform the problem (4.12) onto the reference interval $I = (0, 1)$ and obtain the optimization problem

$$\inf_{q \in Q_{ad}(0,1)} \frac{1}{2} d_U^2(i_1 S(\nu, q)) + \frac{\alpha}{2} \nu \|q\|_{Q(0,1)}^2. \tag{4.13}$$

As in Proposition 2.19 we find that the problems (4.12) and (4.13) are equivalent. For given $\alpha \geq 0$ and $\nu \in \mathbb{R}_+$, we define $\bar{q}_\alpha(\nu)$ as

$$\bar{q}_\alpha(\nu) \in \operatorname*{argmin}_{q \in Q_{ad}(0,1)} \left[d_U^2(i_1 S(\nu, q)) + \alpha\nu \|q\|_{Q(0,1)}^2 \right], \tag{4.14}$$

i.e. $\bar{q}_\alpha(\nu)$ is a solution to (4.13). Since the distance function is convex, the objective functional in (4.13) is strictly convex in case $\alpha > 0$ and whence the optimal solution $\bar{q}_\alpha(\nu)$ is uniquely determined. In case that $\alpha = 0$, there might be several minimizers $\bar{q}_\alpha(\nu)$. We consider the value function $V_\alpha \colon \mathbb{R}_+ \to \mathbb{R}$ associated with (4.13) defined by

$$V_\alpha(\nu) = \frac{1}{2} d_U^2(i_1 S(\nu, \bar{q}_\alpha(\nu))) + \frac{\alpha}{2} \nu \|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2.$$

Then we easily obtain the following a priori regularization error estimate.

Proposition 4.11. *Let $\nu \in \mathbb{R}_+$. If $\bar{q}(\nu) \in Q_{ad}(0,1)$ is a solution to (4.13) with $\alpha = 0$, then*

$$0 \leq \frac{1}{2} d_U^2(i_1 S(\nu, \bar{q}_\alpha(\nu))) - \frac{1}{2} d_U^2(i_1 S(\nu, \bar{q}(\nu))) \leq V_\alpha(\nu) - V_0(\nu) \leq \alpha \frac{\nu}{2} \|\bar{q}(\nu)\|_{Q(0,1)}^2$$

for all $\alpha > 0$.

Proof. This follows as in Proposition 3.32 using optimality of $\bar{q}(\nu)$ and $\bar{q}_\alpha(\nu)$. □

The minimizers of (4.14) satisfy the following necessary optimality conditions.

Lemma 4.12. *Let $\nu \in \mathbb{R}_+$ and $\alpha \geq 0$. Then $\bar{q}_\alpha(\nu) \in Q_{ad}(0,1)$ is locally optimal for (4.13) if and only if*

$$\int_0^1 (\alpha \bar{q}_\alpha(\nu) + B^* z_\alpha, q - \bar{q}_\alpha(\nu)) \geq 0 \quad \text{for all } q \in Q_{ad}(0,1),$$

where the associated adjoint state $\bar{z}_\alpha \in W(0,1)$ is the solution to

$$-\partial_t z_\alpha + \nu A^* z_\alpha = 0, \quad z_\alpha(1) = i_1 S(\nu, \bar{q}_\alpha(\nu)) - P_U(i_1 S(\nu, \bar{q}_\alpha(\nu))).$$

Proof. Since the squared distance function $d_U^2(\cdot)$ is convex and Fréchet differentiable with $\nabla d_U^2(u) = 2\,(u - P_U(u))$, see Proposition 2.11, the result follows by standard arguments; see, e.g., [147, Lemma 2.21]. □

Depending on the choice of the optimization methods for the inner and outer loops, we might require $\alpha > 0$. To determine $V_0(\nu)$ we are therefore interested in continuity and differentiability properties of $\bar{q}_\alpha(\nu)$ and $V_\alpha(\nu)$ with respect to α. In particular, these results will be used for a path-following approach for the solution of the inner problem by means of the primal-dual active set strategy.

Proposition 4.13. *Let $\nu > 0$ and $\alpha_1, \alpha_2 \in \mathbb{R}_+$. Then*

$$\|\bar{q}_{\alpha_1}(\nu) - \bar{q}_{\alpha_2}(\nu)\|_{Q(0,1)} \leq \frac{|\alpha_1 - \alpha_2|}{\alpha_1} \|\bar{q}_{\alpha_2}(\nu)\|_{Q(0,1)}.$$

Proof. This follows along the lines of the proof of [130, Proposition 2.31]. In particular, we use the optimality conditions of Lemma 4.12 and the fact that $\nabla d_U^2(i_1 S(\nu, q))$ is monotone since $d_U^2(i_1 S(\nu, q))$ is convex; see [12, Proposition 17.10]. □

In related contexts, it is well-known that the value function is differentiable and concave with respect to the regularization parameter α; see, e.g., [130, Section 2.5.2] and the references given therein. This is also valid in our situation. More specifically, we obtain

Proposition 4.14. *The function $\mathbb{R}_+ \ni \alpha \mapsto V_\alpha(\nu)$ is continuously differentiable and concave. Additionally, the expression*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} V_\alpha(\nu) = \frac{\nu}{2} \|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2, \quad \nu \in \mathbb{R}_+, \tag{4.15}$$

holds for all $\alpha \in (0, \infty)$.

Proof. Let $\nu \in \mathbb{R}_+$. We first show that $\alpha \mapsto V_\alpha(\nu)$ is concave. Let $\theta \in (0,1)$, $\alpha_0, \alpha_1 > 0$ be given and set $\alpha_\theta = \theta\alpha_0 + (1-\theta)\alpha_1$. Using optimality of $\bar{q}_{\alpha_0}$ and $\bar{q}_{\alpha_1}$ yields

$$\begin{aligned}\theta V_{\alpha_0}(\nu) + (1-\theta)V_{\alpha_1}(\nu) &\leq \frac{\theta}{2}\left(d_U^2(u_{\alpha_\theta}(1)) + \alpha_0\nu\|\bar{q}_{\alpha_\theta}(\nu)\|_{Q(0,1)}^2\right)\\ &\quad + \frac{1-\theta}{2}\left(d_U^2(u_{\alpha_\theta}(1)) + \alpha_1\nu\|\bar{q}_{\alpha_\theta}(\nu)\|_{Q(0,1)}^2\right)\\ &= V_{\alpha_\theta}(\nu),\end{aligned}$$

where we have set $u_{\alpha_\theta} = S(\nu, \bar{q}_{\alpha_\theta}(\nu))$. Concavity of the mapping $\alpha \mapsto V_\alpha(\nu)$ implies that it is locally Lipschitz continuous; see, e.g., [12, Proposition 8.28]. Hence, using Rademacher's theorem we infer that it is differentiable almost everywhere; see, e.g., [53, Section 3.1.2].

To verify the expression for the first derivative, from (4.15), we observe that for $\varepsilon > 0$ the difference quotient is bounded from below and above by

$$\begin{aligned}\frac{1}{\varepsilon}\left(V_{\alpha+\varepsilon}(\nu) - V_\alpha(\nu)\right) &\leq \frac{1}{2\varepsilon}\Big(d_U^2(u_\alpha(1)) + (\alpha+\varepsilon)\nu\|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2\\ &\quad - d_U^2(u_\alpha(1)) - \alpha\nu\|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2\Big)\\ &= \frac{\nu}{2}\|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2 \leq \frac{1}{\varepsilon}\left(V_\alpha(\nu) - V_{\alpha-\varepsilon}(\nu)\right).\end{aligned}$$

Hence, for $\varepsilon \to 0$ we find that

$$\mathrm{d}_\alpha^+ V_\alpha(\nu) \leq \frac{\nu}{2}\|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2 \leq \mathrm{d}_\alpha^- V_\alpha(\nu),$$

where d_α^+ and d_α^- denote the directional derivatives with respect to α in positive and negative direction. Thus, we conclude (4.15). Finally, Proposition 4.13 yields that $\alpha \mapsto \|\bar{q}_\alpha\|_{Q(0,1)}$ is continuous completing the proof. □

Proposition 4.15. *The function $\mathbb{R}_+ \ni \alpha \mapsto V_\alpha(\nu)$ is two times differentiable almost everywhere. Furthermore, the estimate*

$$0 \leq -\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}V_\alpha(\nu) \leq \frac{\nu}{\alpha}\|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2$$

holds for almost all $\alpha \in \mathbb{R}_+$ and all $\nu \in \mathbb{R}_+$.

Proof. Since $-V_\alpha$ is convex with respect to α due to Proposition 4.14, existence of the second derivative almost everywhere is consequence of Alexandrov's theorem; see, e.g., [53, Section 6.4]. Additionally, from convexity of $-V_\alpha$, we infer that its derivative is monotone; see [12, Proposition 17.10]. Thus

$$\left(\frac{\mathrm{d}}{\mathrm{d}\alpha}V_{\alpha+\tau}(\nu) - \frac{\mathrm{d}}{\mathrm{d}\alpha}V_\alpha(\nu)\right)\tau \leq 0, \quad \tau \in \mathbb{R}, \quad \alpha + \tau > 0.$$

Dividing by τ^2 and letting $\tau \to 0$, yields $\frac{\mathrm{d}^2}{\mathrm{d}\alpha^2}V_\alpha(\nu) \leq 0$. For the remaining estimate, using Proposition 4.13, for any $\tau \in \mathbb{R}$ such that $\alpha + \tau > 0$ we obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\alpha}V_{\alpha+\tau}(\nu) - \frac{\mathrm{d}}{\mathrm{d}\alpha}V_\alpha(\nu) &= \frac{\nu}{2}(\bar{q}_{\alpha+\tau}(\nu) - \bar{q}_\alpha(\nu), \bar{q}_{\alpha+\tau}(\nu) + \bar{q}_\alpha(\nu))\\ &\leq \frac{\nu}{2}\|\bar{q}_{\alpha+\tau}(\nu) - \bar{q}_\alpha(\nu)\|_{Q(0,1)}\|\bar{q}_{\alpha+\tau}(\nu) + \bar{q}_\alpha(\nu)\|_{Q(0,1)}\\ &\leq \frac{\nu|\tau|}{2\alpha}\left(\|\bar{q}_{\alpha+\tau}(\nu)\|_{Q(0,1)}\|\bar{q}_\alpha(\nu)\|_{Q(0,1)} + \|\bar{q}_\alpha(\nu)\|_{Q(0,1)}^2\right).\end{aligned}$$

Finally, dividing by $|\tau|$ and letting $\tau \to 0$ yields the remaining estimate. □

4.2.3. Bisection method for the outer optimization

In view of Lemma 4.8, in the limit case $\alpha = 0$, we find that $V_0(\nu) = 0$ if and only if $\nu = T$ is time-optimal for (P_δ) with $\delta = 0$. Therefore, to solve the time-optimal control problem $(\hat{P})$, we can equivalently determine the root of $V_0(\cdot)$. Since $\delta(\cdot)$ is strictly monotonically decreasing, see Proposition 4.7, so is $V_0(\cdot)$. Hence, a first approach would be to use the bisection method to iteratively find a root of $V_0(\cdot)$. For simplicity, we suppose that U is weakly invariant under (A, BQ_{ad}); see Definition 2.1. This automatically implies that $V_0(\bar{\nu} + \tau) = 0$ for all $\tau > 0$. Note that the standard bisection method is applied to functions that have a zero with nonvanishing first derivative at that point. Since V_0 equals zero for times larger than $\bar{\nu}$, this leads to the modified bisection algorithm sketched in Algorithm 2.

Due to the fact the time interval is halved in each iteration, its accuracy can be controlled by the number of iterations denoted $n_{\max}$. More specifically, to reach the tolerance ε_{tol}, we require

$$n = \frac{\log(\nu_b - \nu_a) - \log \varepsilon_{\text{tol}}}{\log 2}$$

number of iterations. Under strong stability, we immediately obtain q-linear convergence for the value function.

Algorithm 2: Bisection method for solution of minimal distance problem (outer loop)

```
Choose ν_a < ν_b;
Calculate d_a = V_0(ν_a) and d_b = V_0(ν_b);
if d_a = 0 or d_b ≠ 0 then
    Error: Optimal time is not contained in [ν_a, ν_b];
end
Set ν_0 = (ν_a + ν_b)/2;
for n = 0 to n_max do
    Calculate d_n = V_0(ν_n);
    if d_n = 0 then
        Set ν_b = ν_n;
    else
        Set ν_a = ν_n;
    end
    Set ν_{n+1} = (ν_a + ν_b)/2;
end
```

Proposition 4.16. *Let $\bar{\nu} \in \mathbb{R}_+$ be the optimal time for $(\hat{P})$ with $\alpha = 0$ and suppose that U is weakly invariant under (A, BQ_{ad}). Moreover, let $\nu_a < \nu_b$ be such that $\nu_a < \bar{\nu} < \nu_b$. If (P_δ) is strongly stable on the right at $\delta = 0$, then for $|\nu_a - \bar{\nu}|$ sufficiently small we have*

$$0 \leq V_0(\nu_n) - V_0(\bar{\nu}) \leq \frac{1}{2} \left(\frac{\nu_b - \nu_a}{\eta_0} \right)^2 2^{-2n},$$

where ν_n denote the iterates generated by the bisection method; see Algorithm 2.

Proof. Due to strong stability on the right, we have

$$\delta(T') - \delta(T) \leq \eta_0^{-1}(T - T').$$

for all $(T - \varepsilon\eta_0)^+ \leq T' \leq T$. Moreover, by definition of Algorithm 2, the iterates satisfy $|\nu_n - \bar{\nu}| \leq 2^{-n}(\nu_b - \nu_a)$. Hence, if $\nu_n \leq \bar{\nu}$, it holds

$$0 \leq V_0(\nu_n) - V_0(\bar{\nu}) \leq \eta_0^{-2}(\nu_b - \nu_a)^2 2^{-2n+1}$$

because of $V_0(\bar{\nu}) = 0$. For the remaining case $\nu_n > \bar{\nu}$ we have $V_0(\nu_n) - V_0(\bar{\nu}) = 0$, due to weak invariance. □

Hence, we have to determine $\bar{q}_0(\nu)$, in order to iteratively solve the time-optimal control problem. We will come back to this question in the following. Let us first discuss a Newton method for the outer loop that leads to improved order of convergence under appropriate assumptions.

4.2.4. Newton method for the outer optimization

In order to apply the Newton method to efficiently compute a root of the value function V_α, we require differentiability of V_α with respect to ν. For the following considerations, we in addition suppose that the terminal constraint U is given by

$$U = \{\, u \in H \colon \|u - u_d\|_H \leq \delta_0 \,\}$$

for some $\delta_0 > 0$ and $u_d \in H$. Then, we easily find a simple reformulation of (4.13). Instead of minimizing the squared distance function, we can equivalently consider the minimization of the squared norm, i.e.

$$\inf_{q \in Q_{ad}(0,1)} \frac{1}{2}\|i_1 S(\nu, q) - u_d\|^2 - \frac{\delta_0^2}{2} + \frac{\alpha}{2}\nu\|q\|^2_{Q(0,1)}. \tag{4.16}$$

By an abuse of notation for given $\alpha \geq 0$ and $\nu \in \mathbb{R}_+$, we define $\bar{q}_\alpha(\nu)$ as

$$\bar{q}_\alpha(\nu) \in \operatorname*{argmin}_{q \in Q_{ad}(0,1)} \left[\|i_1 S(\nu, q) - u_d\|^2 + \alpha\nu\|q\|^2_{Q(0,1)}\right].$$

Furthermore, we consider the associated value function $V_\alpha \colon \mathbb{R}_+ \to \mathbb{R}$ defined by

$$V_\alpha(\nu) = \frac{1}{2}\|i_1 S(\nu, \bar{q}(\nu)) - u_d\|^2 - \frac{\delta_0^2}{2} + \frac{\alpha}{2}\nu\|\bar{q}_\alpha(\nu)\|^2_{Q(0,1)}.$$

In this case, the necessary and sufficient optimality conditions of (4.16) are given by

$$\int_0^1 (\alpha\bar{q}_\alpha(\nu) + B^* z_\alpha, q - \bar{q}_\alpha(\nu))\,\mathrm{d}t \geq 0 \quad \text{for all } q \in Q_{ad}(0,1),$$

where the associated adjoint state $z_\alpha \in W(0,1)$ is the solution to

$$-\partial_t z_\alpha + \nu A^* z_\alpha = 0, \quad z_\alpha(1) = u_\alpha(1) - u_d, \tag{4.17}$$

with $u_\alpha = i_1 S(\nu, \bar{q}_\alpha(\nu))$. As before, we obtain the following a priori regularization error estimate

$$0 \leq \|i_1 S(\nu, \bar{q}_\alpha(\nu)) - u_d\|^2 - \|i_1 S(\nu, \bar{q}(\nu)) - u_d\|^2 \leq \alpha\nu\|\bar{q}(\nu)\|^2_{Q(0,1)}; \tag{4.18}$$

cf. Proposition 4.11. Moreover, for differentiability of the value function, we require the notion of polyhedricity; see Definition 3.22. Recall from Proposition 3.23 that in case of box constraints for the controls the set of admissible controls is polyhedric.

Proposition 4.17. *Let $\alpha \in \mathbb{R}_+$ and $\nu \in \mathbb{R}_+$. Suppose that $Q_{ad}(0,1)$ is polyhedric and that $U = \mathcal{B}_{\delta_0}(u_d)$. Then the value function V_α is differentiable with locally Lipschitz continuous derivative and the expression*

$$V'_\alpha(\nu) = \int_0^1 \langle B\bar{q}_\alpha(\nu) - Au_\alpha, z_\alpha\rangle + \frac{\alpha}{2}\|\bar{q}_\alpha(\nu)\|^2_{Q(0,1)}\,\mathrm{d}t \tag{4.19}$$

holds, where $z_\alpha \in W(0,1)$ satisfies

$$-\partial_t z_\alpha + \nu A^* z_\alpha = 0, \quad z_\alpha(1) = u_\alpha(1) - u_d,$$

and $u_\alpha = S(\nu, \bar{q}_\alpha(\nu))$.

Proof. This follows as in Proposition 4.4. We give the proof for the convenience of the reader. Set

$$d(\nu, q) := \frac{1}{2}\|i_1 S(\nu, q) - u_d\|^2_H \quad \text{and} \quad f(\nu, q) := d(\nu, q) + \frac{\alpha}{2}\nu\|q\|^2_{Q(0,1)}.$$

Since $q \mapsto i_1 S(\nu, q)$ is affine linear, we immediately infer that

$$\partial_{qq} f(\nu, q)\delta q^2 \geq \alpha\nu\|\delta q\|^2_{Q(0,1)} \quad \text{for all } q, \delta q \in Q(0,1)$$

and f satisfies a strong second order sufficient optimality condition. Therefore, according to [21, Proposition 5.2 (ii)], the mapping $\nu \mapsto \bar{q}_\alpha(\nu)$ is locally Lipschitz continuous.

Let $\nu \in \mathbb{R}_+$ and $\tau_n \in \mathbb{R}$ such that $\tau_n \to 0$. Set $q_n = \bar{q}_\alpha(\nu + \tau_n)$ and $q = \bar{q}_\alpha(\nu)$. Employing local Lipschitz continuity of $\nu \mapsto \bar{q}_\alpha(\nu)$, we conclude that the quotient $\tau_n^{-1}(q_n - q)$ converges weakly to some $\delta q \in Q(0,1)$. In addition, since $\partial_{qq} f(\nu, q)$ is elliptic, it defines a Legendre form; see [21, Proposition 3.76]. According to [21, Theorem 5.5], the weak limit δq is in fact a strong limit and satisfies a so-called linearized variational inequality. The latter in particular implies that $\partial_q f(\nu, q)\delta q = 0$, because δq belongs to the critical cone.

For these reasons, we finally obtain

$$\begin{aligned}\tau_n^{-1}\left[V_\alpha(\nu + \tau_n) - V_\alpha(\nu)\right] &= \tau_n^{-1}\left[f(\nu + \tau_n, q_n) - f(\nu, q_n) + f(\nu, q_n) - f(\nu, q)\right] \\ &\to \partial_\nu f(\nu, q) + \partial_q f(\nu, q)\delta q = \partial_\nu f(\nu, q).\end{aligned}$$

The concrete expression (4.19) for $\partial_\nu f(\nu, q)$ follows as in Proposition 2.21. Moreover, from (4.19), local Lipschitz continuity of $\nu \mapsto \bar{q}_\alpha(\nu)$, and Lipschitz stability of the solution to the state and adjoint state equation, we further deduce that $\nu \mapsto V'_\alpha(\nu)$ is locally Lipschitz continuous. □

Remark 4.18. It would be desirable to prove differentiability of V_α for $\alpha = 0$ under a structural assumption on the adjoint state such as the one used in Section 3.3. This would probably lead to a setting in L^1. However, [21, Theorem 5.5] relied on reflexivity of the underlying space.

We emphasize that given the solution $\bar{q}_\alpha(\nu)$, it is computationally very cheap to evaluate the derivative $V'_\alpha(\nu)$. Indeed, using for instance the primal-dual active set strategy to solve (4.16), then all variables required for the computation of $V'_\alpha(\nu)$ have already been computed for the inner loop and we simply have to calculate the inner product and the norm in (4.19). For this reason, one step of the Newton method has approximately the same computational costs as one step of the bisection method; see Algorithm 2.

Basically without additional costs, we can use the following Newton method to find a root of $V_\alpha(\cdot)$. Given $\nu_n > 0$, we calculate the next iterate ν_{n+1} by the formula

$$\nu_{n+1} = \nu_n - \frac{V_\alpha(\nu_n)}{V'_\alpha(\nu_n)}.$$

We have to argue that the Newton iterates are well-posed, i.e. we have to show that V'_α is uniformly bounded away from zero in some neighborhood of $\bar{\nu}$. The following proposition provides a sufficient condition for well-posedness, under the assumption that qualified optimality conditions hold for all solutions to the original problem.

Proposition 4.19. *Let $(\bar{\nu}, \bar{q})$ denote an optimal solution to $(\hat{P})$. Suppose that qualified optimality conditions hold for all optimal solutions to $(\hat{P})$ with $\alpha = 0$. Then for $\alpha > 0$ sufficiently small, we have $V'_\alpha(\bar{\nu}) < 0$.*

Proof. Consider a sequence of optimal controls $\bar{q}_\alpha(\bar{\nu})$ with $\alpha \to 0$. Due to boundedness of $Q_{ad}(0,1)$, there is a subsequence denoted in the same way converging weakly to some $q \in Q_{ad}(0,1)$. Hence, the corresponding states $u_\alpha = S(\bar{\nu}, \bar{q}_\alpha(\bar{\nu}))$ satisfy $u_\alpha \rightharpoonup u$ in $W(0,1)$ and $u_\alpha(1) \to u(1)$ in H due to compactness of the control-to-observation mapping; see Proposition A.20. The latter implies $z_\alpha \to z$ in $W(0,1)$. Moreover, employing (4.18) we have

$$\begin{aligned}\|u(1) - u_d\| &\le \|u_\alpha(1) - u_d\| - \|\bar{u}(1) - u_d\| + \delta_0 + \|u(1) - u_\alpha(1)\| \\ &\le c\sqrt{\alpha} + \delta_0 + c\|u(1) - u_\alpha(1)\| \to \delta_0.\end{aligned}$$

For this reason, we infer that $(\bar{\nu}, q)$ is feasible for (P). Due to optimality of $\bar{\nu}$ and $\alpha = 0$, we further deduce that the tuple $(\bar{\nu}, q)$ is an optimal solution to $(\hat{P}_0)$. Let $\mu \in N_U(u(1))$ be an associated Lagrange multiplier. According to Proposition 2.8 and [40, Corollary 10.44], since $U = \mathcal{B}_{\delta_0}(u_d)$, the normal cone is given by

$$N_U(u) = \{\, \lambda(u - u_d) \colon \lambda \ge 0 \,\} = \{\, v - u \colon v \in H \text{ with } P_U(v) = u \,\}$$

for all $u \in H$ with $\|u - u_d\| = \delta_0$. Thus, there is $\mu_0 > 0$ such that $\mu = \mu_0(u(1) - u_d)$. In summary, we get

$$\begin{aligned}V'_\alpha(\bar{\nu}) &= \int_0^1 \langle B\bar{q}_\alpha(\bar{\nu}) - Au_\alpha, z_\alpha\rangle + \frac{\alpha}{2}\|\bar{q}_\alpha(\bar{\nu})\|^2_{Q(0,1)} \\ &\le -\mu_0 + \int_0^1 \langle B(\bar{q}_\alpha(\bar{\nu}) - q) - A(u_\alpha - u), z\rangle + \int_0^1 \langle B\bar{q}_\alpha(\bar{\nu}) - Au_\alpha, z_\alpha - z\rangle + \frac{\alpha}{2}C_{Q_{ad}}\end{aligned}$$

due to the qualified optimality conditions. Weak convergence of $\bar{q}_\alpha(\bar{\nu}) \rightharpoonup q$ and $u_\alpha \rightharpoonup u$ as well as convergence of $z_\alpha \to z$ imply the existence of $\alpha_0 > 0$ such that $V'_\alpha(\bar{\nu}) < 0$ for all $\alpha < \alpha_0$. □

We summarize the Newton method for the outer loop of the optimization in Algorithm 3. By means of Proposition 4.17 and well-known properties of the Newton method, see, e.g., [126, Theorem 11.2], we infer fast local convergence of Algorithm 3.

Proposition 4.20. *Adopt the assumptions of Proposition 4.17 and suppose $V'_\alpha(\bar{\nu}) \neq 0$. Then the sequence ν_n generated by Algorithm 3 converges locally q-quadratically to $\bar{\nu}$.*

In order to implement this method in practice, as for the bisection algorithm one has to efficiently calculate $\bar{q}_\alpha(\nu)$. We will discuss two approaches in the sequel.

Algorithm 3: Newton method for solution of minimal distance problem (outer loop)

Let $\alpha > 0$ be given. Choose $\nu_0 > 0$;
for $n = 0$ **to** $n_{\max}$ **do**
  Calculate $q_n = \bar{q}_\alpha(\nu_n)$ and $u_n = S(\nu_n, q_n)$;
  if $V_\alpha(\nu_n) < \varepsilon_{tol}$ **then**
    return;
  end
  Evaluate $V'_\alpha(\nu_n)$ using (4.19);
  Set $\nu_{n+1} = \nu_n - V_\alpha(\nu_n)V'_\alpha(\nu_n)^{-1}$;
end

4.2.5. Conditional gradient method for the inner optimization

For both the bisection and the Newton method we have to determine the solution to (4.14) for a sequence of ν. Following the presentation from [50], we introduce the conditional gradient method (cG) as follows. For convenience we abbreviate

$$f(q) = \frac{1}{2} d_U^2(i_1 S(\nu, q)) + \frac{\alpha}{2} \nu \|q\|^2_{Q(0,1)}$$

neglecting the ν and α dependence for a moment. Clearly, we are interested in minimizing f over $Q_{ad}(0,1)$. Let $\bar{q}$ denote an optimal control. We emphasize that all statements also hold for the case $\alpha = 0$ that is of particular interest in this section. By means of Propositions 2.11, 2.20 and 2.21, we infer that $f\colon Q(0,1) \to \mathbb{R}$ is continuously differentiable and its gradient can be expressed as

$$f'(q)^* = \nu \left(\alpha q + B^* z\right),$$

where $z \in W(0,1)$ satisfies

$$-\partial_t z + \nu A^* z = 0, \quad z(1) = u(1) - P_U(u(1)),$$

and $u = S(\nu, q)$. Given $q_n \in Q_{ad}(0,1)$, we take

$$q_{n+1/2} \in \{\, q \in Q_{ad}(0,1) \colon f'(q_n)(q - q_n) = \inf_{v \in Q_{ad}(0,1)} f'(q_n)(v - q_n) \,\}. \tag{4.20}$$

In many cases $q_{n+1/2}$ is directly given by a simple formula. For example, if $Q = L^2(\omega)$ with (ω, ϱ) a finite measure space and box constraints

$$Q_{ad} = \{\, q \in L^2(\omega) \colon q_a \le q(x) \le q_b \text{ a.e. } x \in \omega \,\},$$

where $q_a, q_b \in \mathbb{R}$, $q_a < q_b$, we obtain the explicit expression

$$q_{n+1/2} = \begin{cases} q_a, & \text{if } \alpha q_n + B^* z_n > 0, \\ q_b, & \text{if } \alpha q_n + B^* z_n < 0, \end{cases}$$

almost everywhere. Moreover, we determine the optimal convex combination of q_n and $q_{n+1/2}$ as

$$\lambda_n = \operatorname*{argmin}_{0 \le \lambda \le 1} f((1-\lambda) q_n + \lambda q_{n+1/2}). \tag{4.21}$$

For example, in the case that $U = \mathcal{B}_{\delta_0}(u_d)$, the expression can be analytically determined, employing the fact that $q \mapsto S(\nu, q)$ is affine linear. Finally, the next iterate is defined by the minimizing argument from (4.21), i.e.

$$q_{n+1} = (1 - \lambda_n)q_n + \lambda_n q_{n+1/2}.$$

Using convexity of f and the definition of $q_{n+1/2}$, we immediately obtain the following a posteriori error estimator

$$0 \leq f(q_n) - f(\bar{q}) \leq f'(q_n)(q_n - \bar{q}) \leq \max_{q \in Q_{ad}(0,1)} f'(q_n)(q_n - q) = f'(q_n)(q_n - q_{n+1/2}).$$

The expression on the right-hand side can be efficiently evaluated using the adjoint representation and serves as a termination criterion for the conditional gradient method. The algorithm is summarized in Algorithm 4.

The conditional gradient method has the following convergence properties.

Proposition 4.21. *Let $(q_n)_n$ be a sequence generated by the conditional gradient method. Then $f(q_n)$ decreases monotonically and*

$$0 \leq f(q_n) - f(\bar{q}) \leq \frac{f(q_0) - f(\bar{q})}{1 + cn}, \quad n \geq 0,$$

with a constant c exclusively depending on the Lipschitz constant of f' on $Q_{ad}(0,1)$, the initial residuum, and $Q_{ad}(0,1)$.

Proof. This follows from [50, Theorem 3.1 (i)], since both f and $Q_{ad}(0,1)$ are convex. □

Under additional assumptions, improved order of convergence can be shown. To this end, we assume that $Q = L^2(\omega)$ for a finite measure space (ω, ϱ). If the control operator B defines a linear bounded operator from $L^1(\omega)$ to H, then under a structural assumption on the adjoint state, the objective values converges q-linearly. Recall that a similar assumption has been used for sufficient optimality conditions in case of bang-bang controls; cf. Section 3.3. We will encounter this condition again in Section 5.5 in the context of finite element discretization error estimates for bang-bang controls.

Proposition 4.22. *Let $Q = L^2(\omega)$, $\alpha = 0$, and $B\colon L^1(\omega) \to H$. Moreover, let $\bar{z}$ denote the adjoint state associated to $\bar{q}$. Suppose that there is $C > 0$ such that*

$$|\{\, (t,x) \in I \times \omega\colon\ -\varepsilon \leq (B^*\bar{z})(t,x) \leq \varepsilon \,\}| \leq C\varepsilon$$

for all $\varepsilon > 0$. Then there is $\lambda \in [1/2, 1)$ such that

$$0 \leq f(q_n) - f(\bar{q}) \leq [f(q_0) - f(\bar{q})]\, \lambda^n, \quad n \geq 0.$$

The constant λ exclusively depends on C, q_a, q_b, ω, and the Lipschitz constant of f' on $Q_{ad}(0,1)$. Moreover, for a constant $c > 0$ we have

$$\|q_n - \bar{q}\|_{L^1(I\times\omega)} \leq c\lambda^{n/2}, \quad n \geq 0.$$

Algorithm 4: Conditional gradient method for solution of (4.13)

Let $\alpha > 0$ and $\nu > 0$ be given. Choose $q_0 \in Q_{ad}(0,1)$;
for $n = 0$ **to** $n_{\max}$ **do**
  Calculate $u_n = S(\nu, q_n)$ and z_n;
  Choose $q_{n+1/2}$ as in (4.20);
  if $f'(q_n)(q_n - q_{n+1/2}) < \varepsilon_{tol}$ **then**
    return;
  end
  Calculate λ_n by (4.21);
  Set $q_{n+1} = (1 - \lambda_n)q_n + \lambda_n q_{n+1/2}$;
end

Proof. Since $B\colon L^1(\omega) \to H$, the variation of constants formula implies that the control-to-state mapping is linear and continuous from $L^1(I \times \omega)$ to $C([0,1]; H)$. Hence, f as a mapping defined on $L^1(I \times \omega)$ is (infinitely often) continuously differentiable. Furthermore, as in the proof of Proposition 3.28, we find that

$$f'(\bar{q})(q - \bar{q}) \geq c_0 \|q - \bar{q}\|^2_{L^1(I\times\omega)}, \quad q \in Q_{ad}(0,1),$$

for some constant $c_0 > 0$. Therefore, the assertion follows from [50, Theorem 3.1 (iii)]. □

In practice, the desired tolerance ε_{tol} for the inner loop can be heuristically chosen based on the current iterate. If $V_\alpha(\nu_n) \gg 0$, then for the outer loop it is sufficient to solve the inner optimization only up to a coarser tolerance. This suggests the heuristic

$$\varepsilon_{\text{tol}} = \max\{\varepsilon_{\text{target}}, \beta f(q_n)\}$$

with a suitable chosen $\beta \in (0,1)$ and $\varepsilon_{\text{target}}$ denoting the target tolerance at the optimum. We observe good results in our numerical examples with $\beta = 10^{-3}$.

4.2.6. Primal-dual active set strategy for the inner optimization

In order to solve the minimization problem (4.13), in this subsection we consider an alternative method to the first order method discussed before. We will discuss the solution of (4.14) by means of the primal-dual active set strategy (PDAS), where we essentially follow the presentation of [81, Chapter 7]. In the following, we suppose $\alpha > 0$.

As before for the Newton method for the outer optimization, we restrict to the special case when the terminal constraint U is given by

$$U = \{\, u \in H\colon \|u - u_d\|_H \leq \delta_0 \,\}$$

for some $\delta_0 > 0$ and $u_d \in H$. Moreover, we assume $Q = L^2(\omega)$ as before and box constraints for Q_{ad}. Let $C\colon L^2(I \times \omega) \to H$ denote the (affine linear) control-to-observation operator associated with the state equation for fixed ν, i.e. $Cq = i_1 S(\nu, q)$. We abbreviate

$$f(q) = \frac{1}{2}\|Cq - u_d\|^2_H + \frac{\alpha}{2}\nu\|q\|^2_{L^2(I\times\omega)}$$

neglecting the ν and α dependence for a moment. Clearly, we are interested in minimizing f over $Q_{ad}(0,1)$. The primal-dual active set strategy is introduced as follows. We choose $d > 0$

and an initial control $q_0 \in Q(0,1)$. By a slight abuse of notation, in each iteration, we compute the associated state $u_n = Cq_n$ (evaluated at the terminal time) and the corresponding adjoint state $z_n = (C')^*(u_n - u_d)$. Note that z_n is not the adjoint state as before, because it is already multiplied by B^*. Moreover, we set

$$\mu_n = -\frac{1}{\alpha} z_n - q_n,$$

and the active index sets are defined as

$$\begin{aligned} \mathcal{A}_n^a &= \{\, (t,x) \in I \times \omega \colon \mu_n(t,x) + d(q_n(t,x) - q_a(x)) < 0 \,\}, \\ \mathcal{A}_n^b &= \{\, (t,x) \in I \times \omega \colon \mu_n(t,x) + d(q_n(t,x) - q_b(x)) > 0 \,\}. \end{aligned}$$

The PDAS can be seen as a prediction strategy that predicts on the basis of (q_n, μ_n) the true active and inactive sets. Given the current iterate q_n, the new iterate q_{n+1} is determined as the solution to the linear system

$$\begin{aligned} u_{n+1} &= Cq_{n+1}, \\ z_{n+1} &= (C')^*(u_{n+1} - u_d), \end{aligned} \qquad q_{n+1} = \begin{cases} q_a & \text{on } \mathcal{A}_n^a, \\ q_b & \text{on } \mathcal{A}_n^b, \\ -\frac{1}{\alpha} z_{n+1} & \text{else.} \end{cases}$$

The system above can be equivalently written as

$$\begin{aligned} u_{n+1} - Cq_{n+1} &= 0, \\ z_{n+1} - (C')^* u_{n+1} &= -(C')^* u_d, \\ (1 - \mathbb{1}_{\mathcal{A}_n^a} - \mathbb{1}_{\mathcal{A}_n^b}) \alpha^{-1} z_{n+1} + q_{n+1} &= \mathbb{1}_{\mathcal{A}_n^a} q_a + \mathbb{1}_{\mathcal{A}_n^b} q_b, \end{aligned} \tag{4.22}$$

where $\mathbb{1}_{\mathcal{A}_n^a}$ and $\mathbb{1}_{\mathcal{A}_n^b}$ denote the characteristic functions associated with $\mathcal{A}_n^a$ and $\mathcal{A}_n^b$, respectively. Note that (4.22) can be efficiently solved numerically employing an iterative solver such as GMRES or BICGSTAB; see, e.g., [139, Sections 6.5, 7.4.2] and [52, Section 7.1].

If $\mathcal{A}_n^a = \mathcal{A}_{n-1}^a$ and $\mathcal{A}_n^b = \mathcal{A}_{n-1}^b$, then the optimal solution is found. In practice it is frequently observed, that this condition can be used as a termination criterion; see [81, Remark 7.1.1] and the reference therein. However, scattering might occur and therefore we use the norm of the indicator function of changed indices as a stopping criterion; cf. also [83, Example 5.3]. The PDAS for the solution of (4.14) is summarized in Algorithm 5.

Algorithm 5: Primal-dual active set strategy for solution of (4.13)

Let $\alpha > 0$ and $\nu > 0$ be given. Choose $q_0 \in Q_{ad}(0,1)$ and $d > 0$;
for $n = 0$ **to** $n_{\max}$ **do**
  Determine active sets $\mathcal{A}_n^a$ and $\mathcal{A}_n^b$;
  Set $r_n = (\mathcal{A}_n^a \neq \mathcal{A}_{n-1}^a) \vee (\mathcal{A}_n^b \neq \mathcal{A}_{n-1}^b)$;
  if $\|r_n\|_{L^2(I\times\omega)} < \varepsilon_{tol}$ **then**
    return;
  end
  Calculate solution q_{n+1} to (4.22);
end

Under suitable assumptions on the regularization parameter α and the parameter d, Algorithm 5 is guaranteed to converge globally, i.e. independent of the initial value q_0.

Proposition 4.23. *Let $d > 0$ as in Algorithm 5. Suppose there is $\gamma > 0$ such that*

$$\alpha + \gamma < d < \alpha - \frac{\alpha^2}{\gamma} + \frac{\alpha^2}{\|C'\|^2}.$$

Then $q_n \to \bar{q}_\alpha(\nu)$ in $L^2(I \times \omega)$ as $n \to \infty$.

Proof. Noting that the control-to-observation operator is affine linear and compact from $Q_{ad}(0,1)$ into H, see Proposition A.20, this follows as in [91, Theorem 3]. □

Choosing $\gamma = \alpha$ in Proposition 4.23, we obtain the following sufficient criterion for global convergence

$$2\|C'\|^2 < \alpha.$$

It is well-known that the PDAS can be interpreted as a semismooth Newton method; see [76]. We therefore can expect fast convergence, if the initial value is sufficiently close to the solution.

Proposition 4.24. *Let $\alpha > 0$ and $\nu \in \mathbb{R}_+$. Suppose that $B^* \colon V \to L^p(\omega)$ for some $p > 2$. The primal-dual active set strategy converges locally q-superlinearly.*

Proof. Since $q \mapsto Cq$ is affine linear, we obtain

$$f''(q)\delta q^2 \geq \alpha\nu \|\delta q\|^2_{L^2(I\times\omega)}, \quad q, \delta q \in Q(0,1),$$

which is a strong second order sufficient optimality condition. Moreover, the mapping $q \mapsto Cq$ is continuous from $L^2(I \times \omega)$ into V. In addition, $(C')^*$ is linear and continuous from V into $C([0,1]; L^p(\omega)) \hookrightarrow L^p(I \times \omega)$ for some $p > 2$ due to the supposition on B^*. In summary, $(C')^*Cq$ is continuous from $L^2(I \times \omega)$ into $L^p(I \times \omega)$. Hence, arguing as in [76, Theorem 4.1] yields the assertion; cf. also [151]. □

Remark 4.25. The regularity assumption on the adjoint of the control operator can be satisfied for all prototypical control scenarios considered in Section 3.1.2.

(i) In case of a distributed control, B^* is the restriction to ω operator. According to Proposition 3.4, we have $H^1_D(\Omega) \hookrightarrow L^p(\Omega)$ for $p > 2$. For these reasons, we obtain $B^* \colon V \to L^p(\omega)$ for some $p > 2$.

(ii) For Neumann boundary control, we take $B^* = \mathrm{Tr}$, where Tr denotes the trace operator. According to Proposition 3.5, we have $\mathrm{Tr} \colon H^{\theta,p}_D(\Omega) \to L^p(\Gamma_N)$ for $\theta \in (1/p, 1)$. Employing Proposition 3.4, we find $H^1_D(\Omega) \hookrightarrow H^{\theta,p}_D(\Omega)$ for $1 - d/2 \geq \theta - d/p$ or, equivalently, $d(\theta - 1 + d/2) \geq p$. Hence, the supposition is satisfied for, e.g., $\theta = 3/4$. Since $d \geq 2$, we have $p > 2$ for all $\theta > 0$, so in particular for $\theta = 1 - 2\theta_0$ with $\theta_0 \in (0, 1/4)$.

(iii) In case of purely time-dependent control, we have $B^* \colon X_{1-\theta_0} \to L^2(\omega)$ due to Assumption 2.2. Since $V \hookrightarrow X_{1-\theta_0}$ and $L^2(\omega) \cong \mathbb{R}^{N_c} \cong L^p(\omega)$ the supposition is clearly satisfied.

In practice, a path-following strategy with respect to the regularization parameter α is recommendable. For a systematic derivation we adapt the idea of an appropriate model function,

where we follow [77]. Motivated by the a priori estimate of Proposition 4.11, we consider an affine linear model function

$$V_\alpha(\nu) \approx m_n(\alpha) = m_{n,0} + m_{n,1}\alpha$$

for parameters $m_{n,0}$, $m_{n,1}$ to be calibrated in each iteration of the path-following strategy. For $\alpha_n > 0$, based on current data, we require

$$m_n(\alpha_n) = V_{\alpha_n}(\nu), \quad m_n'(\alpha_n) = \frac{\mathrm{d}}{\mathrm{d}\alpha} V_{\alpha_n}(\nu).$$

Hence, Proposition 4.14, implies

$$m_{n,0} = V_{\alpha_n}(\nu) - m_{n,1}\alpha_n, \quad m_{n,1} = \frac{\nu}{2}\|\bar{q}_{\alpha_n}(\nu)\|^2_{L^2(I\times\omega)}.$$

Using Proposition 4.15 and Taylor's expansion of $V_\alpha(\nu)$ at α_n immediately imply

$$V_{\alpha_0}(\nu) = m_n(\alpha_0) + \int_{\alpha_0}^{\alpha_n} \frac{\mathrm{d}^2}{\mathrm{d}\alpha^2} V_{\alpha'}(\nu)(\alpha' - \alpha_0)\,\mathrm{d}\alpha' \le m(\alpha_0).$$

for $0 < \alpha_0 < \alpha_n$. Letting $\alpha_0 \to 0$, yields $0 \le m_n(0) - V_0(\nu)$. Moreover, from Proposition 4.11 we infer that

$$m_n(0) - V_0(\nu) = V_{\alpha_n}(\nu) - m_{n,1}\alpha_n - V_0(\nu) \le \alpha_n \frac{\nu}{2}\|\bar{q}(\nu)\|^2_{L^2(I\times\omega)}.$$

In summary, we have the error estimate

$$0 \le m_n(0) - V_0(\nu) \le \alpha_n \frac{\nu}{2}\|\bar{q}(\nu)\|^2_{L^2(I\times\omega)}.$$

Using the model function we will deduce an update strategy to get the next regularization parameter α_{n+1}. Ideally for a sequence of $\tau_n \in (0,1)$ we would like to have

$$|V_{\alpha_{n+1}}(\nu) - V_0(\nu)| \le \tau_n |V_{\alpha_n}(\nu) - V_0(\nu)|.$$

Plugging the model function into the inequality for $V_0(\nu)$ and $V_{\alpha_{n+1}}(\nu)$, due to linearity of m_n, we simply obtain $\alpha_{n+1} = \tau_n \alpha_n$.

4.2.7. Numerical examples

Last, we conduct two numerical examples in order to verify our findings of the preceding subsections in practice. The discretization scheme of the state equation, the adjoint state equation, and the control variable will be discussed in detail in Chapter 5. Therefore, we will be brief here. The discretization is based on a Galerkin method. The state and adjoint state equations are discretized by piecewise constant in time and continuous and cellwise linear functions in space. Since we expect the control to be bang-bang, the control variable is discretized by temporally and spatially piecewise and cellwise constant functions.

We use both the bisection method and the Newton method for the computation of the optimal time. The inner optimization problems are solved using the conditional gradient method. In addition, we have implemented the following acceleration strategy for the conditional gradient method: Instead of minimizing the convex combination of the last iterate q_n and the new point $q_{n+1/2}$ in (4.21), we search for the best convex combination of all previous iterates plus the new point $q_{n+1/2}$. For the acceleration strategy, we use CVX to solve the arising

convex subproblems; see [63, 64]. To keep the memory requirements moderate, points that are associated with small coefficients in the convex combination are being deleted from the stock. In practice, we observe that this strategy significantly improves the convergence. This is of particular interest for problems, where the structural assumption of Proposition 4.22 is not fulfilled and, hence, q-linear convergence is not guaranteed. However, we are not able to give conditions that guarantee fast convergence of the accelerated conditional gradient method. For further details on improved convergence of variants of the conditional gradient method in finite dimensions we also refer to [98].

Moreover, we compare the conditional gradient method and the primal dual active set strategy with path following for the inner optimization. For the path following strategy we use the heuristic choice of τ_n given by

$$\tau_{n+1} = \max\{\,\tau_{\min}, \min\{\,\tau_{\max}, \|q_n\|_{L^2(I\times\omega)}/(d\alpha_n + \|q_n\|_{L^2(I\times\omega)})\,\}\,\}, \tag{4.23}$$

where d, $\tau_{\min}$, and $\tau_{\max}$ have to be calibrated manually. We obtain good numerical results for $d = 1000$, $\tau_{\min} = 0.1$, and $\tau_{\max} \in [0.8, 0.95]$.

Numerical example with purely time-dependent control

As a first example, we consider the case of purely time-dependent controls. We control the heat equation on a bounded domain $\Omega \subset \mathbb{R}^2$ with homogeneous Dirichlet boundary conditions. The precise problem data reads as

$$\begin{aligned}
&\Omega = (0,1)^2, \quad \omega_1 = (0, 0.5) \times (0,1), \quad \omega_2 = (0.5, 1) \times (0, 0.5),\\
&B\colon \mathbb{R}^2 \to L^2(\Omega), \quad Bq = q_1 \mathbb{1}_{\omega_1} + q_2 \mathbb{1}_{\omega_2},\\
&G(u) = \tfrac{1}{2}\|u\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad \delta_0 = \tfrac{1}{10},\\
&Q_{ad}(0,1) = \{q \in L^2((0,1);\mathbb{R}^2)\colon\ -1.5 \le q \le 0\}, \quad u_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2^3),
\end{aligned}$$

where $\mathbb{1}_{\omega_1}$ and $\mathbb{1}_{\omega_2}$ denote the characteristic functions on ω_1 and ω_2. The spatial mesh is chosen such that the boundaries of ω_1 and ω_2 coincide with edges of the mesh. We will revisit this example again in Sections 5.4.2 and 5.7.1 on a priori discretization error estimates. The corresponding value function is depicted in Figure 4.1.

For the outer optimization, we observe linear convergence of the bisection method and quadratic convergence of the Newton method; see Figure 4.2. This is almost in accordance with the theory, except for the fact that we do not know that the value function is differentiable with Lipschitz continuous derivative for $\alpha = 0$. Concerning the inner optimization, we observe that the conditional gradient method without and with acceleration converges faster than the primal-dual active set strategy; see Figure 4.3. However, for high accuracy the conditional gradient method with acceleration and the primal-dual active set strategy perform better than the pure conditional gradient method that shows sublinear convergence at some point.

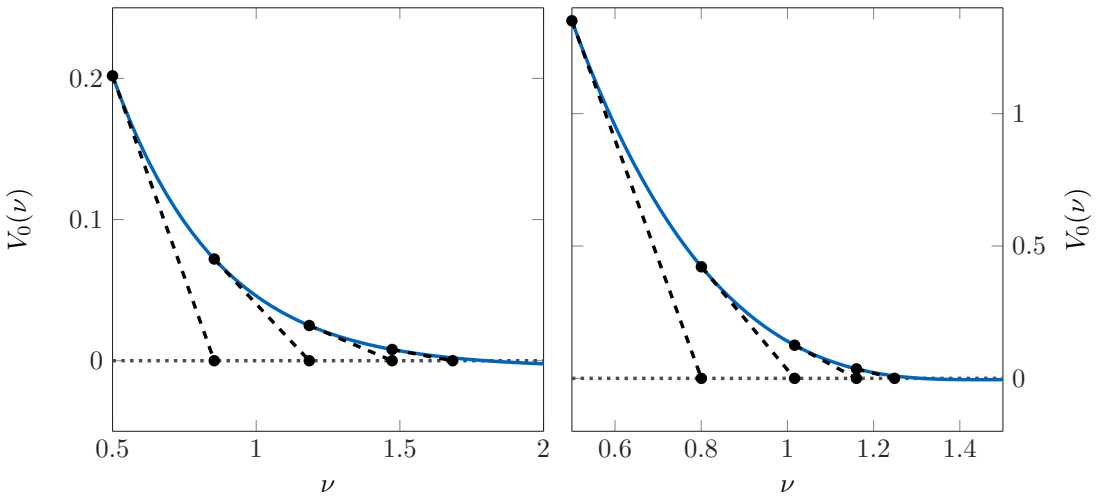

Figure 4.1.: Value function $V_0(\nu)$ for the example with purely time dependent control (left) and the example with distributed control (right). Dotted lines indicate the first Newton steps. Function values calculated by the Newton method for $\alpha = 0$ and the conditional gradient method with acceleration strategy.

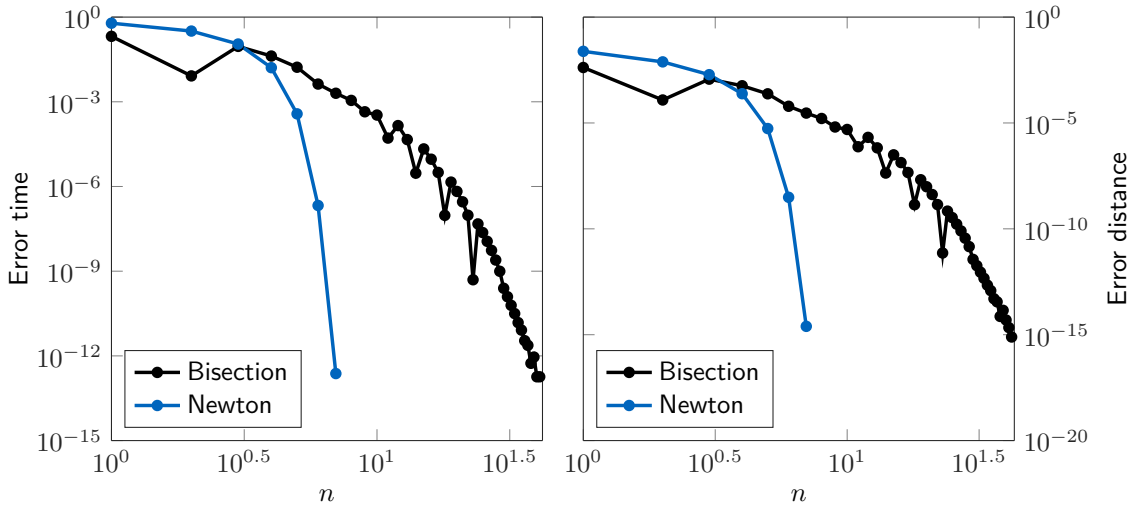

Figure 4.2.: Absolute error $|\nu_n - \bar{\nu}|$ (left) and $|V_0(\nu_n)|$ (right) for the example with purely time dependent control over the iteration number in the outer loop. For each fixed ν_n the inner problem is solved using the conditional gradient method with acceleration strategy.

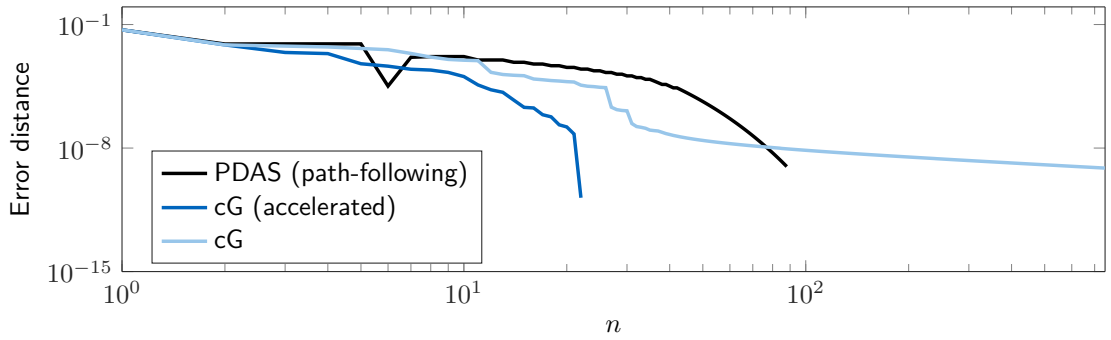

Figure 4.3.: Error $|f(q_n) - f(q)|$ for the inner loop calculated by different methods for the example with purely time-dependent control and fixed $T = \nu = 1.6$ over the iteration number. The PDAS is embedded into a path-following strategy as described in Section 4.2.6, where we take $\alpha_0 = 10^{-2}$ and the update $\alpha_{n+1} = \tau_n \alpha_n$ with τ_n determined by (4.23) and $\tau_{\max} = 0.8$. The cG method is carried out for $\alpha = 0$.

Numerical example with distributed control

As a second numerical example, we consider the distributed control on the whole domain, i.e. $\omega = \Omega$. More specifically, let

$$\begin{aligned} \Omega = (0,1)^2, \quad \omega = (0,1)^2, \quad \delta_0 = \tfrac{1}{10}, \\ G(u) = \tfrac{1}{2}\|u\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad \delta_0 = \tfrac{1}{10}, \\ Q_{ad} = \{q \in L^2(I \times \omega) \colon -1.5 \le q \le 0\}, \\ u_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2)^3. \end{aligned}$$

The corresponding value function is plotted in Figure 4.1 (right). Snapshots of the optimal control are depicted in Figure 4.6. As before, we observe linear convergence of the bisection method and quadratic convergence of the Newton method; see Figure 4.4. Moreover, we compare the different methods for the solution of the inner optimization problem. The accelerated conditional gradient method performs slightly better than the pure conditional gradient method; see Figure 4.5. However, it is difficult to solve the minimal distance problem to the same hight accuracy as in the first example.

4.2.8. Comparison to other approaches

The equivalence of time and distance optimal controls as stated in Lemma 4.8 can be related to the equivalence of time and norm optimal controls. We will give a brief overview and introduce two problems. Let $p \in [1, \infty]$ be fixed. First, for any $\rho > 0$, we introduce the *minimal time problem* as

$$\inf_{\substack{T>0 \\ q \in Q(0,T)}} T \quad \text{subject to} \quad u[q](T) \in U, \quad \|q\|_{L^p((0,T);Q)} \le \rho. \tag{4.24}$$

Here the set of admissible controls is defined by the additional constraint $\|q\|_{L^p((0,T);Q)} \le \rho$ instead of $Q_{ad}(0,1)$ used in our problem formulation. Note that in (P_δ) we studied perturbations on the terminal constraint set, whereas in (4.24) we consider perturbations in the control constraints in terms of the parameter ρ.

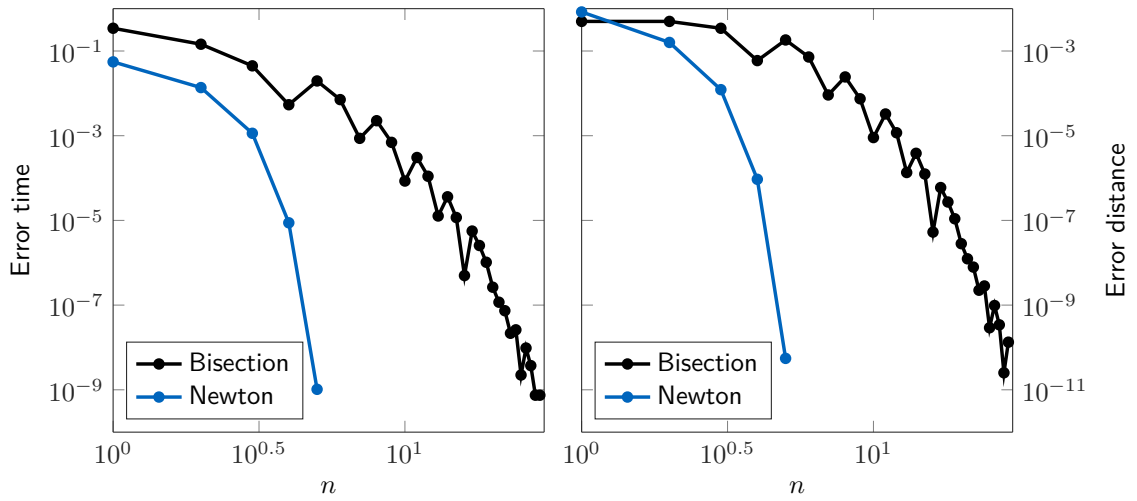

Figure 4.4.: Absolute error $|\nu_n - \bar{\nu}|$ (left) and $|V_0(\nu_n)|$ (right) for the example with distributed control over the iteration number in the outer loop. For each fixed ν_n the inner problem is solved using the conditional gradient method with acceleration strategy.

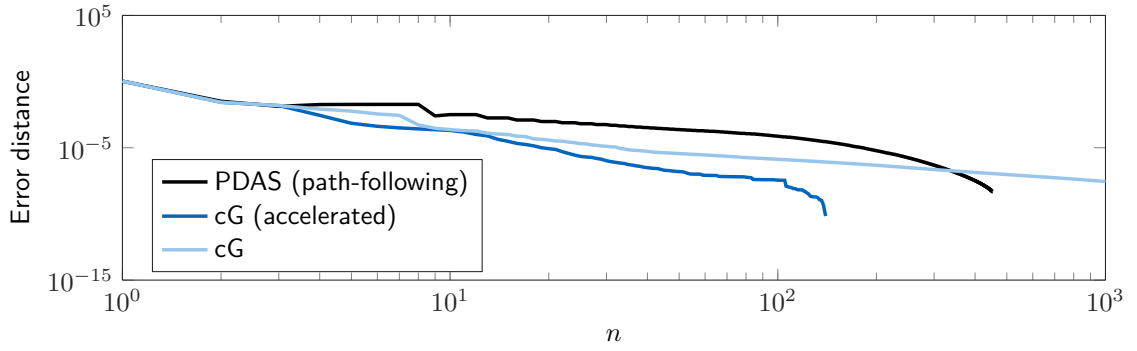

Figure 4.5.: Error $|f(q_n) - f(q)|$ for the inner loop calculated by different methods for the example with distributed control for fixed $T = \nu = 1.1$ over the iteration number. The PDAS is embedded into a path-following strategy as described in Section 4.2.6, where we take $\alpha_0 = 10^{-2}$ and the update $\alpha_{n+1} = \tau_n \alpha_n$ with τ_n determined by (4.23) and $\tau_{\max} = 0.95$. The cG method is carried out for $\alpha = 0$.

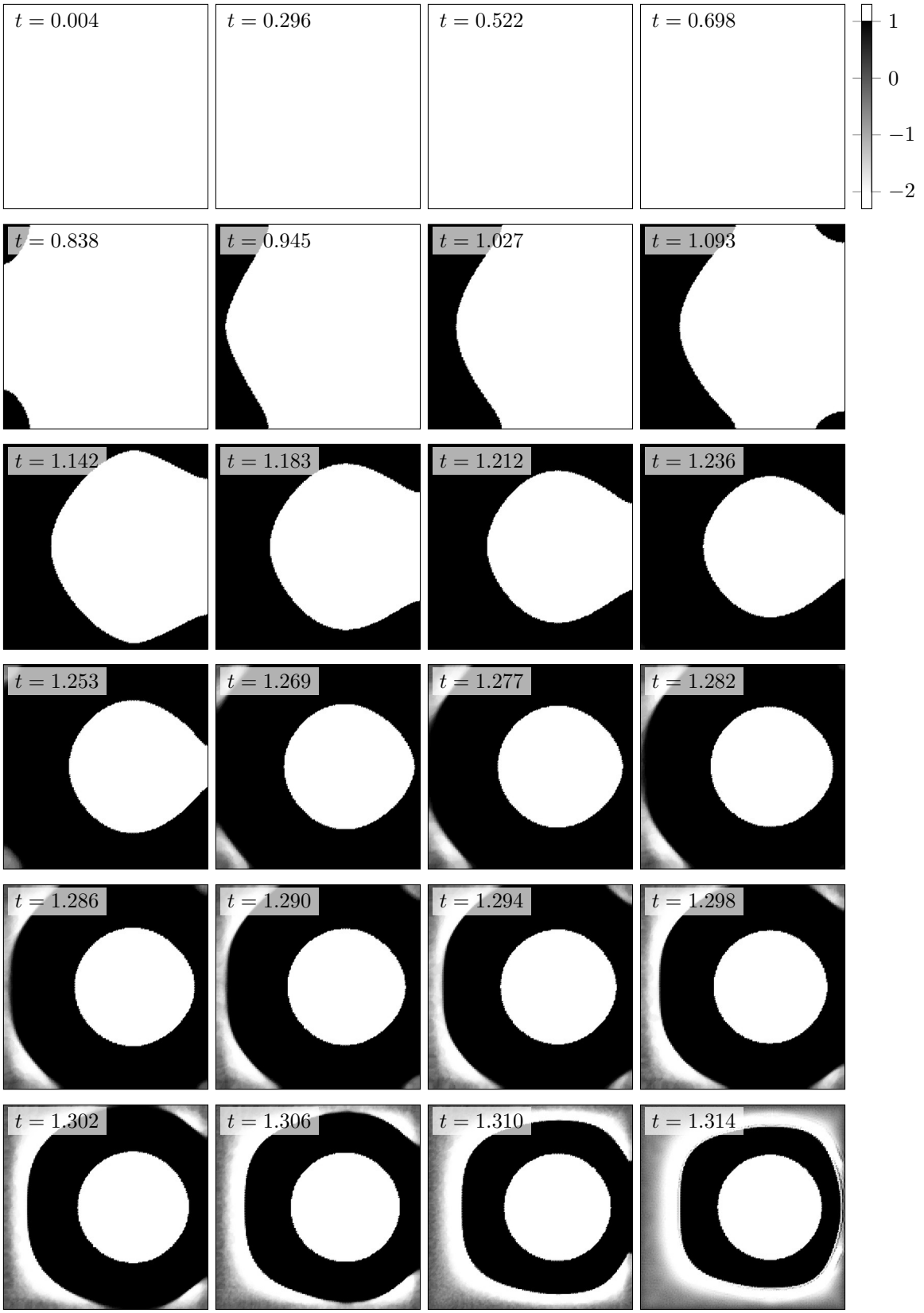

Figure 4.6.: Logarithmically spaced snapshots of control for example with distributed control. Solution calculated by the Newton method Algorithm 3 for $\alpha = 0$ and the conditional gradient method Algorithm 4 for $\varepsilon_{\text{tol}} = 10^{-8}$. White and black denote the lower and the upper control bound, respectively.

Second, for given $T > 0$ we introduce the *minimal norm problem* as

$$\inf_{q \in Q(0,T)} \|q\|_{L^p((0,T);Q)} \quad \text{subject to} \quad u[q](T) \in U. \tag{4.25}$$

For each problem, we can define corresponding value functions

$$T(\rho) = \inf (4.24) \quad \text{and} \quad M(T) = \inf (4.25).$$

The connection of (4.24) and (4.25) has been extensively investigated; see, e.g, [54, 62, 89, 97, 160, 164]. More specifically, under certain conditions, it has been shown that $T(\cdot)$ and $M(\cdot)$ are inverse to each other. For example, if $U = \{ 0 \}$ and the system is null controllable by L^p controls, then according to [62, Theorem 4.1] the value function M is implicitly defined by the relation

$$M(T(\rho)) = \rho, \quad \rho > 0.$$

Recall that a system is called null controllable by L^p controls, if for each $T > 0$ and initial condition $u_0 \in H$ there exists a control $q \in L^p((0,T);Q)$ such that $u[u_0, q](T) = 0$. Moreover, equivalence of (4.24) and (4.25) in the sense above has been shown for the heat equation with distributed control on a subset of the spatial domain with $p = \infty$ and again $U = \{ 0 \}$; see [160, Theorem 2.1].

In fact, the idea to build algorithms based on an equivalent reformulation of the time-optimal control problem is not new. Wang and Zuazua proposed in [160, Remark 3.3] to solve the minimal norm problem in order to solve the time-optimal control problem by means of the equivalence of time and norm optimal controls. Inspired by [160], a bisection method has been used to solve time-optimal control problems subject to ordinary differential equations in [109]; cf. also [164, Theorem 1.2]. However, to the best of the authors knowledge, neither theoretical results nor numerical examples have been published so far in the context of partial differential equations. An equivalence that is similar to the one of this section has been shown in [158] for the situation of delaying the activation of the control as long as possible. Moreover, a related approach has been developed in [70] for time-optimal control of a one-dimensional vibrating system with controls in a subspace of L^2 determined by certain moment equations.

We note that both approaches require different assumptions: While the equivalence of minimal time and minimal norm controls relies on exact null controllability with L^p controls, our approach requires that a certain value function is left continuous; see Lemma 4.8. It seems to be difficult to compare these assumptions with each other. Independently, they essentially rely on the state equation under consideration.

In comparison with the approach based on the equivalence of time and distance optimal controls, we observe that the minimum norm problem (4.25) is still subject to state constraints, in contrast to the minimal distance problem (δ_T) that is convex and subject to control constraints. For the latter class of optimization problems, efficient algorithms are available, while the algorithmic solution of state constrained control problems is generally more difficult.

Additionally we note that, in view of the equivalence of minimal time and minimal distance controls as well as the equivalence of minimal time and minimal norm controls, the distance optimal solution also solves the minimal norm problem. Whence, our approach also provides a solution to the minimal norm problem, which seems to be a nontrivial optimization problem itself.

In the particular case of purely time-dependent controls, an alternative algorithm can be described as follows: It directly solves the time-optimal control problem by parametrizing

the switching points by its location and optimize for the parametrization; see, e.g., [85, 86] for time-optimal control problems subject to ordinary differential equations. However, in the case of distributed control such an ansatz would require to parametrize the time-dependent switching hyperplanes which seems to cause further difficulties.

Last, the approaches based on equivalent reformulations discussed in Section 4.2 can be compared to the augmented Lagrangian method from Section 4.1 equipped with a path-following strategy in the regularization parameter α. Here, we would consider a sequence of regularization parameters $\alpha_1 > \alpha_2 > \ldots > 0$ such that $\lim_{n\to\infty} \alpha_n = 0$. For each such α_n we solve the regularized time-optimal control problem with the method discussed in Section 4.1. In view of the stability results from Section 3.3.2, the corresponding terminal times converge at the rate α and the optimal controls converge to a solution of the unregularized problem. Moreover, under additional assumptions the control variable is guaranteed to converge in $L^1(I \times \omega)$ at the same rate α. However, for small regularization parameters the resulting optimization problems become computationally very expensive. Comparing running times for the numerical examples we observe that our approach based on the equivalence of minimal time and minimal distance controls is at least competitive with the regularization approach. However, it is difficult to find a fair measure for the comparison as the total running time depends on various aspects. For example, the augmented Lagrangian approach equipped with a path-following strategy in the regularization parameter strongly depends on the choice of the initial value for the optimization. In contrast, the approach discussed in this section is not very sensitive to the initial time for the outer optimization.

For a numerical realization there are of course further error contributions besides the error due to regularization such as the discretization error and the modelling error between the model and the real problem. Clearly, one is interested in controlling the overall approximation and regularization error; see also Section 5.5 for a detailed discussion on the quantitative behavior of the discretization and regularization error. Therefore, one could argue that we do not have to consider α arbitrarily small as long as other error contributions dominate the total error. However, this is only half an argument, as we wish to have control over each error contribution.

5. A priori discretization error estimates

This chapter is devoted to a priori discretization error estimates for the time-optimal control problem. To set the stage, we consider the following model problem:

$$\text{Minimize } j(T,q) \coloneqq T + \frac{\alpha}{2}\int_0^T \|q(t)\|^2_{L^2(\omega)}\,\mathrm{d}t,$$

$$\text{subject to } \begin{cases} T > 0, & \\ \partial_t u - \Delta u = Bq, & \text{in } (0,T)\times\Omega, \\ u = 0, & \text{on } (0,T)\times\partial\Omega, \\ u(0) = u_0, & \text{in } \Omega, \\ G(u(T)) \le 0, & \\ q_a \le q(t) \le q_b, & \text{in } \omega,\ t\in(0,T). \end{cases}$$

With the notation of the preceding chapters, $A = -\Delta$ denotes the usual Laplace operator on a bounded domain $\Omega \subset \mathbb{R}^d$ equipped with homogeneous Dirichlet boundary conditions. Accordingly, we take $V = H^1_0(\Omega)$ and $H = L^2(\Omega)$. Moreover, for a finite measure space (ω, ϱ) as in Chapter 3, the control operator B maps from $L^2(I\times\omega)$ into $L^2(I\times\Omega)$. This allows to treat different control scenarios such as purely time-dependent control or distributed control with one consistent notation. The parameter $\alpha \ge 0$ models control costs or is a regularization parameter. Its implications on error estimates will be discussed in detail below.

We start by giving a brief overview on related literature. Although time-optimal control is considered to be a classical subject in control theory, to the best of our knowledge there are only a few publications concerning the numerical solutions of such problems in the context of parabolic equations. The existing contributions have in common that the terminal set is given by an L^2-ball around a desired state (often assumed to be zero), the objective functional is $j(T,q) = T$, and the state is discretized only in space by means of continuous linear finite elements. In [140] convergence of optimal times for a one dimensional heat equation is proved based on a bang-bang principle. Purely time-dependent controls acting on the boundary have been considered in [87]. For $u_0 \in H^{3/2}(\Omega)$ the author proved the error estimate $\mathcal{O}(h^{3/2-\varepsilon})$ with arbitrary small $\varepsilon > 0$ for the optimal times. Furthermore, convergence of optimal times and the controls for the terminal set the L^2-ball centered at some u_d with $u_0, u_d \in H^{1/2-\varepsilon}(\Omega)$ for boundary control has been shown in [100]. More recently, for distributed control and $u_0 \in H^1_0(\Omega)$ the error estimate $\mathcal{O}(h)$ has been proved in [159] for the linear heat equation and for a semilinear heat equation in [165]. Both articles use cellwise linear discretization for the control and the set of admissible controls is defined by $Q_{ad} \coloneqq \{\, q \in L^\infty((0,\infty); L^2(\omega)) \colon \|q(t)\|_{L^2} \le 1 \text{ a.a. } t \,\}$. Employing the variational control discretization the error estimates $\mathcal{O}(h)$ for T and $\mathcal{O}(h^{1-\varepsilon})$ for the control and the state have been shown in [60]. Convergence of optimal times and controls for a class of abstract evolution equations has been recently shown in [148] with terminal set a closed ball centered at the origin. We point out that the authors impose less regularity on the initial value as in the references before, which in our setting would correspond to the assumption $u_0 \in L^2(\Omega)$.

In contrast to the contributions mentioned above, we consider fully space-time discretization of all variables. The state and the adjoint state equation are being discretized by means of the discontinuous Galerkin scheme in time and the continuous Galerkin scheme in space. In this regard we also mention [117, 118] on a priori discretization error estimates for linear parabolic and [124] for semilinear parabolic optimal control problems. Moreover, pointwise control constraints are included in our setting compared to $L^\infty((0,T);L^2(\omega))$ constraints that are typically considered in the contributions mentioned above. Furthermore, we allow for more general terminal sets and we may deal with different control discretizations.

As in the preceding chapter, we will discuss the case of bang-bang controls and non-bang-bang controls (i.e. $\alpha > 0$) separately. The results for $\alpha > 0$ are already contained in [17] in similar form. We prove optimal convergence rates in $L^2(I \times \omega)$ for the control variable in the case $\alpha > 0$ for different control discretization strategies. For example, in case of the variational control discretization we obtain the convergence rate $k+h^2$ in all variables up to a logarithmic term with k and h denoting the temporal and spatial mesh size, respectively. The proof is done in two steps and strongly depends on the second order sufficient optimality condition of Section 3.2. First, we obtain a suboptimal convergence rate for the control variable, where we rely on a quadratic growth condition that follows from a second order sufficient optimality condition (SSC). Conceptionally the discretization error is related to differences of the objective functional for the continuous and the discrete solutions, where we have to take square roots in the end. In the context of pointwise state constraints this is often acceptable, as low regularity of the problem prevents better convergence; cf., e.g., [123]. However, the solutions of (P) exhibit improved regularity if $\alpha > 0$, so we can expect an improved rate of convergence. For the proof we adapt ideas from [31] for unconstrained problems to the constrained case. In this second step, the discretization error is conceptionally related to differences of derivatives of the Lagrange function which avoids taking square roots.

In contrast, for the discretization error estimates for bang-bang controls (i.e. $\alpha = 0$), we rely on the structural assumption on the adjoint state from Section 3.3. Here, we show convergence rates in $L^1(I \times \omega)$ for the control variable that are optimal, if the structural assumption (3.37) holds with $\kappa = 1$. For example, for purely time-dependent controls we prove the convergence rate $\alpha + k + h^2$. It is worth mentioning that all three quantities are independent of each other and which also justifies the terminology robust error estimates. It seems that there is a growing interest in the numerical analysis of optimal control problems with bang-bang solutions which is reflected in a number of articles that have appeared recently. The variational control discretization for a linear elliptic equation has been considered in [47] and for a linear parabolic equation in [152] subject to pointwise control constraints. Moreover, a bilinear optimal control has been investigated in [36]. The latter contribution is based on second order sufficient optimality conditions from [37] that use both a structural assumption (3.37) as well as a condition on the second derivative.

However, it is in general difficult to validate the structural assumption without the optimal solution. Recall that in Section 2.3.5 we proved that the value function is Lipschitz continuous with respect to certain perturbations of the terminal set, the initial state, and the operator under a condition that is a direct strengthening of the lower Hamiltonian condition. In view of $\alpha = 0$, this immediately implies an estimate for the optimal time. This motivates the derivation of a sufficient condition that allows to construct feasible points for the discretized problems. The sufficient condition is basically the strengthened Hamiltonian condition (2.19) for the discrete problems. Since $\alpha = 0$, two-way testing yields an error estimate for the optimal times; cf. [87]. However, estimates with rates for the controls, require further assumptions; cf. [47] (in the context of finite element discretizations) and [37].

This chapter is structured as follows. In Section 5.1 we recap the optimality conditions for the concrete problem that imply improved regularity of the solution in the case $\alpha > 0$. Thereafter, we introduce the discretization scheme and provide general discretization error estimates in Section 5.2. Section 5.3 is devoted to a priori discretization error estimates for the case $\alpha > 0$, where we rely on second order sufficient optimality conditions. Last, we turn to case of bang-bang controls, i.e. $\alpha = 0$. Based on the structural assumption from Section 3.3, we show discretization error estimates for the control variable in Section 5.5. Error estimates based on the strengthened Hamiltonian condition are presented in Section 5.6. Each part of this chapter is accompanied by numerical examples to validate the theoretical findings.

5.1. Assumptions and optimality conditions

We summarize the main assumptions used throughout this chapter.

Assumption 5.1. We assume $\Omega \subset \mathbb{R}^d$ with $d \in \{2, 3\}$ to be a polygonal or polyhedral and convex domain and the initial value satisfies $u_0 \in H_0^1(\Omega)$.

Concerning the control operator we consider one of the following situations:

(i) Distributed control: Let $\omega \subseteq \Omega$ be the control domain that is polygonal or polyhedral as well. The control operator $B\colon L^2(\omega) \to L^2(\Omega)$ is the extension by zero operator. Clearly, its adjoint $B^*\colon L^2(\Omega) \to L^2(\omega)$ is the restriction to ω operator.

(ii) Purely time-dependent control: Let ω be a discrete set equipped with the counting measure. The control operator is defined by $Bq = \sum_{i=1}^{N_c} q_i e_i$, where $e_i \in L^2(\Omega)$ are given form functions. Then $L^2(\omega) \cong \mathbb{R}^{N_c}$ and $B^*\colon L^2(\Omega) \to \mathbb{R}^{N_c}$ with $(B^*\varphi)_i = (e_i, \varphi)_{L^2(\Omega)}$ for $i = 1, 2, \dots, N_c$.

The space of admissible controls is defined as

$$Q_{ad} \coloneqq \Big\{ q \in L^2(\omega)\colon q_a \le q \le q_b \text{ a.e. in } \omega \Big\} \subset L^\infty(\omega)$$

for $q_a, q_b \in \mathbb{R}$ with $q_a < q_b$. Recall that $Q(0,1) \coloneqq L^2(I; L^2(\omega))$ and

$$Q_{ad}(0,1) \coloneqq \Big\{ q \in L^2(I \times \omega)\colon q(t) \in Q_{ad} \text{ a.a. } t \in (0,1) \Big\} \subset L^\infty(I \times \omega).$$

Concerning the state equation, we suppose that $A = -\Delta$ is the usual Laplace operator equipped with homogeneous Dirichlet boundary conditions. As usual, $H_0^1(\Omega)$ is the Sobolev space with zero trace and the corresponding dual space is denoted by $H^{-1}(\Omega)$. The duality pairing between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$ is denoted $\langle \cdot, \cdot \rangle$. If ambiguity is not to be expected, we drop the spatial domain Ω from the notation of the spaces. Moreover, we use $W(0,1)$ to abbreviate $H^1((0,1); H^{-1}) \cap L^2((0,1); H_0^1)$, endowed with the canonical norm and inner product. The symbol $i_1\colon W(0,1) \to H$ denotes the trace mapping $i_1 u = u(1)$. We also define $B\colon L^2(I \times \omega) \to L^2(I \times \Omega)$ by setting $(Bq)(t) = Bq(t)$ for all $t \in (0,1)$ and any control $q \in L^2(I; L^2(\omega)) \cong L^2(I \times \omega)$. Last, $\mathbb{R} \times L^2(I \times \omega)$ is endowed with the canonical inner product and we abbreviate its norm as

$$\|(\delta\nu, \delta q)\| = \Big(|\delta\nu|^2 + \|\delta q\|^2_{L^2(I\times\omega)} \Big)^{1/2}.$$

Assumption 5.2. The terminal constraint G is defined by

$$G(u) \coloneqq \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega)} - \frac{\delta_0^2}{2}.$$

for fixed $u_d \in H^1_0(\Omega)$ and $\delta_0 > 0$.

Remark 5.1. (i) The regularity assumption $u_d \in H^1_0(\Omega)$ is required for optimal order of convergence. Since $G'(u)^* = u - u_d$ defines the terminal value of the adjoint equation, this leads to improved regularity of the adjoint equation, which in turn allows to prove full order of convergence.

(ii) In addition, we would like to justify the regularity assumption $u_d \in H^1_0(\Omega)$ from a different perspective, namely that of weak invariance. Recall that the target set $U = \{\, u \in L^2(\Omega) \colon G(u) \leq 0 \,\}$ is called *weakly invariant* under the state equation if for any u_0 satisfying $G(u_0) \leq 0$ there is a admissible control $q(t) \in Q_{ad}$ such that the corresponding trajectory with initial value u_0 satisfies $G(u(t)) \leq 0$ for all times; cf. Section 2.2. Since the formulation of (P) only requires the state to be inside the target set at the final time T (but not at later times), it seems to be desirable to require the target set to be weakly invariant, since this guarantees that $G(u(t)) \leq 0$ can be maintained for $t > T$. However, this requirement already implies that the minimizing projection P_U to U in $L^2(\Omega)$ is stable in $H^1_0(\Omega)$; see Lemma 2.6. This further leads to the requirement $G'(P_U(u))^* = P_U(u) - u_d \in H^1_0(\Omega)$ for all $u \in H^1_0(\Omega)$, which implies the desired property for u_d.

(iii) The error analysis remains valid for more general terminal constraints. Precisely, we require that G is two times continuously Fréchet-differentiable, the mapping $\eta \mapsto G''(u)[\eta]^2$ is weakly lower semicontinuous, and G'' is bounded on bounded sets in $L^2(\Omega)$. Furthermore, $G'(u)^* \in H^1_0$ for any $u \in H^1_0$.

Assumption 5.3. There exist a finite time $T > 0$ and a feasible control $q \in Q_{ad}(0,T)$ such that the solution to the state equation of (P) satisfies $G(u(T)) \leq 0$. To exclude the trivial case, we additionally assume $G(u_0) > 0$.

Under Assumption 5.3 the time-optimal control problem is well-posed; cf. Proposition 2.14. Moreover, we assume that the constraint qualification from Assumption 3.1 holds, i.e.

$$\eta \coloneqq -\partial_\nu g(\bar\nu, \bar q) > 0.$$

As in Chapter 3 we define the reduced terminal constraint by $g(\nu, q) = G(i_1 S(\nu, q))$. For ν bounded uniformly from below and above, the derivatives of g can be estimated by uniform constants, which will be important in the following.

Proposition 5.2. *Let $0 < \nu_{\min} < \nu_{\max}$ be given. Then there exists $c > 0$ such that for all $\delta\nu \in \mathbb{R}$ and $\delta q \in L^2(I \times \omega)$ the stability estimates*

$$|g'(\nu, q)(\delta\nu, \delta q)| \leq c\|(\delta\nu, \delta q)\|,$$
$$|g''(\nu, q)[\delta\nu, \delta q]^2| \leq c\|(\delta\nu, \delta q)\|^2,$$

hold for all $\nu_{\min} \leq \nu \leq \nu_{\max}$ and $q \in Q_{ad}(0,1)$. Moreover,

$$|(g'(\nu_1, q_1) - g'(\nu_2, q_2))\,(\delta\nu, \delta q)| \leq c\|(\nu_1 - \nu_2, q_1 - q_2)\|\|(\delta\nu, \delta q)\|,$$

for all $\nu_{\min} \leq \nu_1, \nu_2 \leq \nu_{\max}$ and $q_1, q_2 \in Q_{ad}(0,1)$.

Proof. Since $g(\nu, q) = G(i_1 S(\nu, q))$ the result is a consequence of Proposition A.26 and the assumptions on G; see Assumption 5.2. □

Under these assumptions, the first order optimality conditions of Lemma 3.1 imply improved regularity of the optimal solution. More specifically, we infer from the optimality condition (3.6) that

$$\int_0^1 \langle \alpha \bar{q} + B^* \bar{z}, q - \bar{q} \rangle \geq 0 \quad \text{for all } q \in Q_{ad}(0,1).$$

In particular, if $\alpha > 0$ this implies (almost everywhere) in $I \times \omega$

$$\begin{cases} \bar{q}(t,x) = q_a & \text{if } \alpha \bar{q}(t,x) + B^* \bar{z}(t,x) > 0, \\ \bar{q}(t,x) = q_b & \text{if } \alpha \bar{q}(t,x) + B^* \bar{z}(t,x) < 0. \end{cases} \tag{5.1}$$

Furthermore, as in the linear parabolic case, see, e.g., [147, Section 3.6], if $\alpha > 0$, then the following projection formula

$$\bar{q} = P_{Q_{ad}} \left(-\frac{1}{\alpha} B^* \bar{z} \right) \tag{5.2}$$

holds, where $P_{Q_{ad}}(\cdot)$ denotes the pointwise projection onto the set $Q_{ad}(0,1)$, defined by

$$P_{Q_{ad}} \colon L^2(I \times \omega) \to Q_{ad}(0,1), \quad P_{Q_{ad}}(r)(t,x) = \max\{q_a, \min\{q_b, r(t,x)\}\}.$$

Proposition 5.3. *The optimal state $\bar{u}$ and the adjoint state $\bar{z}$ to $(\hat{P})$ exhibit the improved regularity*

$$\bar{u}, \bar{z} \in H^1(I; L^2) \cap L^2(I; H^2 \cap H_0^1) \hookrightarrow C([0,1]; H_0^1).$$

Additionally, in case of distributed control with $\alpha > 0$, we have

$$\bar{q} \in H^1(I; L^2(\omega)) \cap L^2(I; H^1(\omega)).$$

Moreover, if $u_d \in W_0^{1,p}$ for some $p \in [2, \infty)$, then it holds $\bar{z} \in C([0,1]; W_0^{1,p})$ and (for distributed control with $\alpha > 0$) we have $\bar{q} \in C([0,1]; W^{1,p}(\omega))$.

Proof. We first note that elliptic regularity yields $\mathcal{D}_{L^2}(-\Delta) = H^2 \cap H_0^1$, since Ω is convex; see, e.g., [68, Theorem 3.2.1.2]. Furthermore, since $-\Delta$ exhibits maximal parabolic regularity on L^2, see, e.g., [99, Theorem 1], we infer $\bar{u} \in C([0,1]; H_0^1)$ due to $\mathcal{D}_{L^2}((-\Delta)^{1/2}) = H_0^1$. According to Assumption 5.2 we have $\bar{z}(1) = G'(\bar{u}(1))^* \bar{\mu} = \bar{\mu}(\bar{u}(1) - u_d) \in H_0^1$. Hence,

$$\bar{z} \in H^1(I; L^2) \cap L^2(I; \mathcal{D}_{L^2}(-\Delta)) \hookrightarrow C([0,1]; H_0^1).$$

The projection formula (5.2) leads to $\bar{q} \in H^1(I; L^2(\omega)) \cap L^2(I; H^1(\omega))$.

According to [45, Corollary 3.12], it holds $\mathcal{D}_{W^{-1,p}}(-\Delta) = W_0^{1,p}$ for any $p \in [2, \infty)$, because Ω is a convex polyhedron. Moreover, since $-\Delta$ generates an analytic semigroup on $W^{-1,p}$, see, e.g., [8, Theorem 11.5 (i)], we have $\bar{z} \in C([0,1]; \mathcal{D}_{W^{-1,p}}(-\Delta))$ due to [128, Theorem 4.3.5 (ii)]. The projection formula (5.2) yields the last assertion. □

5.2. Finite element discretization

Consider a partitioning of the (reference) time interval $[0,1]$ given as

$$[0,1] = \{0\} \cup I_1 \cup I_2 \cup \ldots \cup I_M$$

with disjoint subintervals $I_m = (t_{m-1}, t_m]$ of size k_m defined by the time points

$$0 = t_0 < t_1 < \ldots < t_{M-1} < t_M = 1.$$

We abbreviate the time discretization by the parameter k defined as the piecewise constant function by setting $k|_{I_m} = k_m$ for all $m = 1, 2, \ldots, M$. Simultaneously, we denote by k the maximal size of the time steps, i.e. $k = \max k_m$. Moreover, we assume the following regularity conditions on the time mesh:

(i) There are constants $c, \beta > 0$ independent on k such that

$$\min_m k_m \geq c k^\beta,$$

(ii) There is a constant $k_{\text{ratio}} > 0$ independent of k such that

$$k_{\text{ratio}}^{-1} \leq \frac{k_m}{k_{m+1}} \leq k_{\text{ratio}},$$

(iii) Last, $k \leq 1/4$ holds.

Concerning the spatial discretization, we consider a discretization consisting of triangular or tetrahedral cells K that constitute a non-overlapping cover of the domain Ω. We define the discretization parameter h as the cellwise constant function $h|_K = h_K$ with diameter h_K of the cell K. Moreover, we set $h = \max h_K$. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$. We suppose throughout that $\mathcal{T}_h$ is regular; see Definition A.31. Let $V_h \subset H_0^1$ denote the subspace of continuous and cellwise linear functions associated with $\mathcal{T}_h$. We define the spatial L^2-projection $\Pi_h \colon L^2 \to V_h$ by

$$(u - \Pi_h u, \varphi)_{L^2} = 0 \quad \text{for all } \varphi \in V_h.$$

The corresponding space-time finite element space is constructed in a standard way by

$$X_{k,h} = \left\{ v_{kh} \in L^2(I; V_h) \colon v_{kh}|_{I_m} \in \mathcal{P}_0(I_m; V_h),\ m = 1, 2, \ldots, M \right\},$$

where $\mathcal{P}_0(I_m; V_h)$ denotes the space of constant functions on the time interval I_m with values in V_h. For any function $\varphi_k \in X_{k,h}$ we set $\varphi_{k,m} \coloneqq \varphi_k(t_m)$ with $m = 1, 2, \ldots, M$, as well as

$$[\varphi_k]_m \coloneqq \varphi_{k,m+1} - \varphi_{k,m}, \quad m = 1, 2, \ldots, M-1.$$

Now, we define the trilinear form $\mathrm{B} \colon \mathbb{R} \times X_{k,h} \times X_{k,h} \to \mathbb{R}$ as

$$\begin{aligned} \mathrm{B}(\nu, u_{kh}, \varphi_{kh}) \coloneqq \sum_{m=1}^{M} & \langle \partial_t u_{kh}, \varphi_{kh} \rangle_{L^2(I_m; L^2)} \\ & + \nu (\nabla u_{kh}, \nabla \varphi_{kh})_{L^2(I; L^2)} + \sum_{m=2}^{M} ([u_{kh}]_{m-1}, \varphi_{kh,m}) + (u_{kh,1}, \varphi_{kh,1}). \end{aligned} \tag{5.3}$$

Note that the definition of B above can be directly extended on the larger space $X_{k,h}+W(0,1)$, which allows to formulate Galerkin orthogonality. Given $\nu \in \mathbb{R}_+$ and $q \in Q(0,1)$ the discrete state equation reads as follows: Find a state $u_{kh} \in X_{k,h}$ satisfying

$$\mathrm{B}(\nu, u_{kh}, \varphi_{kh}) = \nu(Bq, \varphi_{kh})_{L^2(I;L^2)} + (u_0, \varphi_{kh,1})_{L^2} \quad \text{for all } \varphi_{kh} \in X_{k,h}. \tag{5.4}$$

We define the discrete control-to-state mapping $S_{kh}\colon \mathbb{R}_+ \times Q(0,1) \to X_{k,h}$, $S_{kh}(\nu, q) = u_{kh}$, where u_{kh} is the solution to (5.4). In addition, for $\nu \in \mathbb{R}_+$ and $q \in L^2(I \times \omega)$ we introduce the discrete version of the reduced constraint mapping as

$$g_{kh}(\nu, q) \coloneqq G(i_1 S_{kh}(\nu, q)).$$

In the following we verify that S_{kh} is well-defined and prove stability estimates as well as differentiability properties. This will be imported for discretization error estimates for the reduced constraint mapping.

5.2.1. Stability estimates for the PDE

We introduce the discrete analogue $-\Delta_h\colon V_h \to V_h$ to the operator $-\Delta$ as

$$-(\Delta_h u_h, \varphi_h)_{L^2} = (\nabla u_h, \nabla \varphi_h)_{L^2}, \quad \varphi_h \in V_h.$$

For the discretization error estimates we require stability estimates for the state, linearized state, and adjoint state.

Proposition 5.4. *For every tuple $(\nu, q) \in \mathbb{R}_+ \times Q(0,1)$ there exists a unique solution $u_{kh} \in X_{k,h}$ to the discrete state equation. Moreover, the stability estimates*

$$\|u_{kh}(1)\|_{L^2}^2 + \nu \|u_{kh}\|_{L^2(I;H_0^1)}^2 \le c \left(\nu \|Bq\|_{L^2(I;H^{-1})}^2 + \|\Pi_h u_0\|_{L^2(I;L^2)}^2 \right), \tag{5.5}$$

$$\nu \|\Delta_h u_{kh}\|_{L^2}^2 \le c \left(\nu \|Bq\|_{L^2(I;L^2)}^2 + \|\Pi_h u_0\|_{H^1}^2 \right), \tag{5.6}$$

$$\|\nabla u_{kh}(1)\|_{L^2}^2 \le c \left(\nu \|Bq\|_{L^2(I;L^2)}^2 + \frac{1}{\nu} \|\Pi_h u_0\|_{L^2}^2 \right), \tag{5.7}$$

hold with a constant $c > 0$ that is independent of k, h, ν, q, u_0, and u_{kh}.

Proof. We proceed as in [117]. Setting $u_{kh,0} = \Pi_h u_0$ the equation (5.4) can be written as

$$\nu(\nabla u_{kh}, \nabla \varphi_{kh})_{L^2(I_m;L^2)} + ([u_{kh}]_{m-1}, \varphi_{kh,m})_{L^2} = \nu(Bq, \varphi_{kh})_{L^2(I_m;L^2)} \tag{5.8}$$

for all $m = 1, 2, \dots, M$ and all $\varphi_{kh} \in X_{k,h}$. Hence, existence of a solution for each time interval follows by the lemma of Lax-Milgram. Concatenation of the interval-wise defined solutions yields $u_{kh} \in X_{kh}$. Concerning the stability estimates, first testing with $\varphi = u_{kh}$ implies for all $m = 1, 2, \dots, M$

$$\nu \|\nabla u_{kh}\|_{L^2(I_m;L^2)}^2 + \frac{1}{2} \|u_{kh,m}\|_{L^2}^2 - \frac{1}{2} \|u_{kh,m-1}\|_{L^2}^2 \le \nu(Bq, u_{kh})_{L^2(I_m;L^2)},$$

where we have used that

$$\frac{1}{2} \left(\|\varphi_m\|_{L^2}^2 + \|[\varphi]_{m-1}\|_{L^2}^2 - \|\varphi_{m-1}\|_{L^2}^2 \right) = ([\varphi]_{m-1}, \varphi_m)_{L^2}. \tag{5.9}$$

Summation over all $m = 1, 2, \dots, M$, using Poincaré's and Young's inequalities, yields (5.5).

Concerning the second stability estimate (5.6), testing in (5.8) with $\varphi = -\Delta_h u_{kh}$, using the definition of $-\Delta_h$, the identity (5.9), and summation over all $m = 1, 2, \dots, M$ lead to

$$\frac{1}{2}\|\nabla u_{kh}(1)\|_{L^2}^2 + \frac{1}{2}\sum_{m=1}^{M}\|[\nabla u_{kh}]_{m-1}\|_{L^2}^2 + \nu\|\Delta_h u_{kh}\|_{L^2(I;L^2)}^2 \\ \leq \nu(Bq, -\Delta_h u_{kh})_{L^2(I;L^2)} + \|\nabla \Pi_h u_0\|_{L^2}^2.$$

Whence, from Young's inequality we conclude

$$\|\nabla u_{kh}(1)\|_{L^2}^2 + \nu\|\Delta_h u_{kh}\|_{L^2(I;L^2)}^2 \leq c\left(\nu\|Bq\|_{L^2(I;L^2)}^2 + \|\Pi_h u_0\|_{H^1}^2\right).$$

This proves (5.6). Moreover, if $u_0 = 0$, then the estimate above immediately yields (5.7). Proceeding by superposition, it thus remains the case $q = 0$. We argue as in [116, Theorem 4.5]. Testing in (5.8) with $\varphi = -t_m \Delta_h u_{kh}$ gives

$$\nu t_m \|\Delta_h u_{kh}\|_{L^2(I_m;L^2)}^2 + t_m([\nabla u_{kh}]_{m-1}, \nabla u_{kh,m})_{L^2} = 0.$$

Then (5.9) with the relations $t_m = t_{m-1} + k_m$ and $k_m \leq k_{\text{ratio}} k_{m-1}$ implies

$$t_m\|\nabla u_{kh,m}\|_{L^2}^2 + 2\nu t_m\|\Delta_h u_{kh}\|_{L^2(I_m;L^2)}^2 = t_m\|\nabla u_{kh,m-1}\|_{L^2}^2 - t_m\|[\nabla u_{kh}]_{m-1}\|_{L^2}^2 \\ \leq t_{m-1}\|\nabla u_{kh,m-1}\|_{L^2}^2 + k_{\text{ratio}} k_{m-1}\|\nabla u_{kh,m-1}\|_{L^2}^2.$$

Summations yields

$$\|\nabla u_{kh}(1)\|_{L^2}^2 + 2\nu\sum_{m=2}^{M} t_m\|\Delta_h u_{kh}\|_{L^2(I_m;L^2)}^2 \leq k_1\|\nabla u_{kh,1}\|_{L^2}^2 + k_{\text{ratio}}\sum_{m=2}^{M}\|\nabla u_{kh}\|_{L^2(I_m;L^2)}^2 \\ \leq (1 + k_{\text{ratio}})\|\nabla u_{kh}\|_{L^2(I\times\Omega)}^2.$$

Finally, (5.5) and superposition of the result for $u_0 = 0$ proves (5.7). □

Corollary 5.5. *Let $u_{kh} \in X_{k,h}$ be the state corresponding to $(\nu, q) \in \mathbb{R}_+ \times Q(0,1)$. For all $(\delta\nu, \delta q) \in \mathbb{R} \times Q(0,1)$ there are unique solutions $\delta u_{kh} \in X_{k,h}$ and $\delta\tilde{u}_{kh} \in X_{k,h}$ to the discrete linearized and second linearized state equation, i.e.*

$$\mathrm{B}(\nu, \delta u_{kh}, \varphi_{kh}) = (\delta\nu(Bq + \Delta_h u_{kh}) + \nu B\delta q, \varphi_{kh})_{L^2(I;L^2)},$$
$$\mathrm{B}(\nu, \delta\tilde{u}_{kh}, \varphi_{kh}) = 2(\delta\nu(B\delta q + \Delta_h \delta u_{kh}), \varphi_{kh})_{L^2(I;L^2)},$$

for all $\varphi_{kh} \in X_{k,h}$. Moreover, the estimates

$$\|\delta u_{kh}(1)\|_{L^2}^2 \leq c\left(|\delta\nu|^2(\|Bq\|_{L^2(I;H^{-1})}^2 + \frac{1}{\nu}\|\Pi_h u_0\|_{L^2}^2) + \nu\|B\delta q\|_{L^2(I;L^2)}^2\right),$$
$$\|\delta\tilde{u}_{kh}(1)\|_{L^2}^2 \leq c|\delta\nu|^2\left(\|Bq\|_{L^2(I;H^{-1})}^2 + \|\delta u_{kh}\|_{L^2(I;H^1)}^2\right),$$

hold. The constant $c > 0$ is independent of k, h, $\delta\nu$, ν, δq, q, δu_{kh}, and $\delta\tilde{u}_{kh}$.

Similarly, we obtain for the auxiliary adjoint equation the following stability result.

Proposition 5.6. *For every triple $(\nu, f, z_1) \in \mathbb{R}_+ \times L^2(I;L^2) \times H_0^1$ there exists a unique solution $\tilde{z}_{kh} \in X_{k,h}$ to*

$$\mathrm{B}(\nu, \varphi_{kh}, \tilde{z}_{kh}) = \nu(f, \varphi_{kh})_{L^2(I;L^2)} + (z_1, \varphi_{kh}(1)) \quad \text{for all } \varphi_{kh} \in X_{k,h}.$$

Moreover,

$$\|\tilde{z}_{kh}\|_{L^2(I;H_0^1)} \leq c\left(\|f\|_{L^2(I;L^2)} + \frac{1}{\sqrt{\nu}}\|\Pi_h z_1\|_{L^2}\right), \tag{5.10}$$

$$\|\Delta_h \tilde{z}_{kh}\|_{L^2(I;L^2)} \leq c\left(\|f\|_{L^2(I;L^2)} + \frac{1}{\sqrt{\nu}}\|\Pi_h z_1\|_{H^1}\right), \tag{5.11}$$

and the constant $c > 0$ is independent of k, h, ν, f, z_1, and $\tilde{z}_{kh}$.

Proof. Existence of a solution and the stability estimates follow as in Proposition 5.4. □

As in the continuous case we obtain a discrete analogue to Proposition 5.2 using the stability estimates of Proposition 5.4 and Corollary 5.5 for the discrete states.

Proposition 5.7. *Let $0 < \nu_{\min} < \nu_{\max}$ be given. Then there exists $c > 0$ independent of k and h such that for all $\delta\nu \in \mathbb{R}$ and $\delta q \in L^2(I \times \omega)$ we have*

$$|g'_{kh}(\nu, q)(\delta\nu, \delta q)| \leq c\|(\delta\nu, \delta q)\|, \tag{5.12}$$

$$|g''_{kh}(\nu, q)[\delta\nu, \delta q]^2| \leq c\|(\delta\nu, \delta q)\|^2, \tag{5.13}$$

for all $\nu_{\min} \leq \nu \leq \nu_{\max}$ and $q \in Q_{ad}(0,1)$. Moreover, g_{kh} and g'_{kh} are Lipschitz continuous on bounded sets.

5.2.2. Discretization error estimates for the terminal constraint

Based on the discretization error estimates for the state that are collected in Appendix A.7, we establish discretization error estimates concerning the reduced terminal constraint.

Proposition 5.8. *Let $0 < \nu_{\min} < \nu_{\max}$ and $(\nu, q) \in [\nu_{\min}, \nu_{\max}] \times Q_{ad}(0,1)$. For the adjoint state z defined in (3.9) associated with $u = u(\nu, q)$ and the discrete adjoint state z_{kh} associated with $u_{kh} = u_{kh}(\nu, q)$, i.e. z_{kh} satisfies*

$$\mathrm{B}(\nu, \varphi_{kh}, z_{kh}) = \mu(u_{kh}(1) - u_d, \varphi_{kh}(1)) \quad \text{for all } \varphi_{kh} \in X_{k,h},$$

the estimate

$$\|z - z_{kh}\|_{L^2(I;L^2)} \leq c|\log k|(k + h^2)|\mu|\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{L^2}\right) \tag{5.14}$$

holds. If additionally Π_h is stable in H^1, then

$$\|\nabla z - \nabla z_{kh}\|_{L^2(I;L^2)} \leq c|\log k|(k^{1/2} + h)|\mu|\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{L^2}\right). \tag{5.15}$$

The constant $c > 0$ is independent of k, h, ν, μ, q, u_0, z, and z_{kh}.

Proof. We consider the splitting

$$z - z_{kh} = z - \tilde{z} + \tilde{z} - z_{kh}, \tag{5.16}$$

where $\tilde{z}$ denotes the solution to

$$-\partial_t \tilde{z} - \nu\Delta\tilde{z} = 0, \quad \tilde{z}(1) = \mu(u_{kh}(1) - u_d).$$

By means of the stability estimates of Proposition A.26 for u and Proposition 5.4 for u_{kh} as well as boundedness of $q \in Q_{ad}(0,1)$ and $\nu \in [\nu_{\min}, \nu_{\max}]$ we find that $u(1)$ and $u_{kh}(1)$ are uniformly bounded in L^2. Employing a stability result similar as Proposition A.26 and Lipschitz continuity of G' on bounded sets in L^2 we infer

$$\begin{aligned}\|z - \tilde{z}\|_{L^2(I;H^1)} &\le c\frac{|\mu|}{\sqrt{\nu}}\|u(1) - u_{kh}(1)\|_{L^2} \\ &\le c(\nu_{\min}, \nu_{\max})|\mu||\log k|(k+h^2)\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{L^2}\right), \end{aligned} \tag{5.17}$$

where we have used the discretization error estimate (A.42) in the last step. The second term in (5.16) is a pure discretization error, therefore,

$$\begin{aligned}\|\tilde{z} - z_{kh}\|_{L^2(I;L^2)} &\le c(k+h^2)|\mu|\|u_{kh}(1) - u_d\|_{H^1}, \\ \|\nabla\tilde{z} - \nabla z_{kh}\|_{L^2(I;L^2)} &\le c(k^{1/2}+h)|\mu|\|u_{kh}(1) - u_d\|_{H^1};\end{aligned}$$

cf. (A.31) and (A.32). The assertion follows from (5.17), the two preceding estimates and the stability estimates (5.5) and (5.7) applied for u_{kh}. □

Proposition 5.9. *Let $0 < \nu_{\min} < \nu_{\max}$ be fixed. Consider $(\nu, q) \in [\nu_{\min}, \nu_{\max}] \times Q_{ad}(0,1)$ and $(\delta\nu, \delta q) \in \mathbb{R} \times Q(0,1)$. Then*

$$|g(\nu,q) - g_{kh}(\nu,q)| \le c|\log k|(k+h^2)\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{L^2}\right). \tag{5.18}$$

If additionally Π_h is stable in H^1, then

$$|(g'(\nu,q) - g'_{kh}(\nu,q))(\delta\nu,\delta q)| \le c|\log k|(k+h^2)\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{H^1}\right)\|(\delta\nu,\delta q)\|. \tag{5.19}$$

The constant $c > 0$ is independent of k, h, $\delta\nu$, ν, δq, q, and u_0.

Proof. From the discretization error estimate (A.42) and Lipschitz continuity of G on bounded sets in L^2 we conclude

$$\begin{aligned}|g(\nu,q) - g_{kh}(\nu,q)| &\le c(\nu_{\min},\nu_{\max})\|u(1) - u_{kh}(1)\|_{L^2} \\ &\le c(\nu_{\min},\nu_{\max})|\log k|(k+h^2)\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{L^2}\right).\end{aligned}$$

To prove (5.19), we use the adjoint representation (3.8) and its discrete analogue. Let $\mu \in \mathbb{R}$, then

$$[g'(\nu,q) - g'_{kh}(\nu,q)]^*\mu = \begin{pmatrix} \int_0^1 \langle Bq, z - z_{kh}\rangle + \langle \Delta u, z\rangle - \langle \Delta_h u_{kh}, z_{kh}\rangle \,\mathrm{d}t \\ \nu B^*(z - z_{kh}) \end{pmatrix}.$$

Clearly, the terms involving $z - z_{kh}$ can be estimated using (5.14). Concerning the remaining terms of the first component, we have

$$\langle \Delta u, z\rangle - \langle \Delta_h u_{kh}, z_{kh}\rangle = -\langle u_{kh} - u, \Delta z\rangle + \langle \nabla u_{kh} - \nabla u, \nabla z_{kh} - \nabla z\rangle - \langle \Delta u, z_{kh} - z\rangle.$$

Since $\Delta u, \Delta z \in L^2(I;L^2)$, we conclude

$$\begin{aligned}|\langle \Delta_h u_{kh}, z_{kh}\rangle - \langle \Delta u, z\rangle| &\le c\Big(\|u_{kh} - u\|_{L^2(I;L^2)}|\mu| + \|z_{kh} - z\|_{L^2(I;L^2)} \\ &\qquad + \|\nabla u_{kh} - \nabla u\|_{L^2(I;L^2)}\|\nabla z_{kh} - \nabla z\|_{L^2(I;L^2)}\Big) \\ &\le c(\nu_{\min},\nu_{\max})|\log k|(k+h^2)|\mu|\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{H^1}\right)\end{aligned}$$

according to (A.31), (5.14), (A.32), and (5.15). Thus,

$$\|[g'(\nu,q) - g'_{kh}(\nu,q)]^*\mu\| \le c(\nu_{\min},\nu_{\max})|\log k|(k+h^2)|\mu|\left(\|Bq\|_{L^\infty(I;L^2)} + \|u_0\|_{H^1}\right),$$

which implies (5.19) due to linearity of $[g'(\nu,q) - g'_{kh}(\nu,q)]^*$. □

5.3. Error estimates for controls ($\alpha > 0$)

In this section we establish a priori discretization error estimates in the case that $\alpha > 0$. The results are already contained in [17] in similar form. Throughout this section we suppose that the general regularity conditions concerning the temporal and spatial mesh from Section 5.2 are satisfied. Moreover, we assume that the projection Π_h onto V_h is stable in H^1. This is satisfied if, e.g., the mesh is quasi-uniform but weaker conditions are known; cf. [23].

To consider different control discretizations at the same time, we introduce the operator I_σ onto the (possibly discrete) control space $Q_\sigma(0,1) \subset L^2(I \times \omega)$ with an abstract parameter σ for the control discretization. In case of distributed control, we additionally assume that a subset denoted $\mathcal{T}_h^\omega$ of the mesh $\mathcal{T}_h$ is a non-overlapping cover of ω to simplify the discussion. We use the symbol $\sigma(k,h)$ to denote the error due to control discretization, i.e.

$$\|\bar{q} - \mathrm{I}_\sigma \bar{q}\|_{L^2(I\times\omega)} \leq \sigma(k,h)\|\bar{q}\|_\sigma, \tag{5.20}$$

where $\|\cdot\|_\sigma$ stands for a potentially different norm on $Q(0,1)$. We suppose $\sigma(k,h) \to 0$ as $k,h \to 0$ and $\mathrm{I}_\sigma Q_{ad}(0,1) \subset Q_{ad}(0,1)$. Moreover, we assume $\|\bar{q}\|_\sigma < \infty$ and $\|\bar{q}\|_{L^2(I\times\omega)} \leq \|\bar{q}\|_\sigma$. For notational simplicity we write $\mathrm{I}_\sigma(\nu,q) = (\nu, \mathrm{I}_\sigma q)$ using the same symbol. Concrete discretization strategies for the control will be discussed in Section 5.3.3. For convenience we define $Q_{ad,\sigma}(0,1) = Q_\sigma(0,1) \cap Q_{ad}(0,1)$.

The discrete optimal control problem now reads as follows:

$$\inf_{\substack{\nu_{kh}\in\mathbb{R}_+ \\ q_{kh}\in Q_{ad,\sigma}(0,1)}} j(\nu_{kh}, q_{kh}) \quad \text{subject to} \quad g_{kh}(\nu_{kh}, q_{kh}) \leq 0. \tag{$\hat{P}_{kh}$}$$

At this point, the well-posedness of $(\hat{P}_{kh})$ is not clear. In the following, as a by-product of the error analysis, we will show existence of feasible points (for k and h sufficiently small), local uniqueness, and optimality conditions. Note that the linearized Slater condition (3.1) is sufficient to ensure existence (for k, h sufficiently small), whereas the SSC is essential for the local uniqueness and rates of convergence of the optimization variables.

5.3.1. Construction of feasible controls

In order to deal with local solutions, we apply a standard localization argument, cf. [32]. For a given locally optimal control $(\bar{\nu}, \bar{q})$ of $(\hat{P})$ in $Q_{ad} \cap \overline{\mathcal{B}_\rho(\bar{\nu}, \bar{q})}$ with $\rho > 0$ sufficiently small satisfying first-order optimality conditions, we introduce the auxiliary problem

$$\inf_{\substack{\nu_{kh}\in\mathbb{R}_+ \\ q_{kh}\in Q_{ad,\sigma}(0,1)}} j(\nu_{kh}, q_{kh}) \quad \text{subject to} \quad \begin{cases} g_{kh}(\nu_{kh}, q_{kh}) \leq 0, \\ \|(\nu_{kh} - \bar{\nu}, q_{kh} - \bar{q})\| \leq \rho. \end{cases} \tag{$\hat{P}_{kh}^\rho$}$$

We first construct a sequence of tuples $\{(\nu_\gamma, q_\gamma)\}_{\gamma>0}$ converging to $(\bar{\nu}, \bar{q})$ as $\gamma \to 0$ that is feasible for the localized problem. In particular, this implies existence of solutions to $(\hat{P}_{kh}^\rho)$. Thereafter we construct a sequence $\{(\nu_\tau, q_\tau)\}_{\tau>0}$ converging to $(\bar{\nu}_{kh}^\rho, \bar{q}_{kh}^\rho)$ as $\tau \to 0$ that is feasible for $(\hat{P})$. Feasibility of the τ-sequence for $(\hat{P})$ with the quadratic growth condition (3.16) yields convergence of discrete solutions to $(\bar{\nu}, \bar{q})$ at a suboptimal rate. The convergence result will later be the basis for the improved convergence rate in Section 5.3.3.

Proposition 5.10. *Let $(\bar\nu, \bar q)$ be a locally optimal control of problem $(\hat P)$. There exists a sequence $\{(\nu_\gamma, q_\gamma)\}_{\gamma>0}$ of controls with $\gamma = \gamma(k,h)$ that are feasible for $(\hat P^\rho_{kh})$ for k, h, ρ sufficiently small. Moreover,*

$$|\nu_\gamma - \bar\nu| + \|q_\gamma - \bar q\|_{L^2(I\times\omega)} \le c(\sigma(k,h) + |\log k|(k+h^2)).$$

Proof. The proof follows the one of [123, Lemma 4.2]. We abbreviate $\bar\chi = (\bar\nu, \bar q)$. Moreover, for $\gamma > 0$ to be determined in the course of the proof we set

$$\chi_\gamma \coloneqq \mathrm{I}_\sigma \breve\chi^\gamma = (\bar\nu + \gamma, \mathrm{I}_\sigma \bar q).$$

Employing the supposition on I_σ, see (5.20), we obtain

$$\|\chi_\gamma - \bar\chi\| \le \gamma + \sigma(k,h)\|\bar q\|_\sigma. \tag{5.21}$$

Moreover, using Taylor expansion of g_{kh} at $\mathrm{I}_\sigma\bar\chi$ we find

$$g_{kh}(\chi_\gamma) = g_{kh}(\mathrm{I}_\sigma\bar\chi) + \gamma g'_{kh}(\mathrm{I}_\sigma\bar\chi)(1,0) + \frac{\gamma^2}{2} g''_{kh}(\chi_\zeta)[1,0]^2.$$

Using the triangle inequality we estimate the first term by

$$\begin{aligned} g_{kh}(\mathrm{I}_\sigma\bar\chi) &\le g(\bar\chi) + |g(\bar\chi) - g_{kh}(\bar\chi)| + c\|\mathrm{I}_\sigma\bar\chi - \bar\chi\| \\ &\le c_1(|\log k|(k+h^2) + \sigma(k,h)) =: \delta_1(k,h) \end{aligned} \tag{5.22}$$

with Lipschitz continuity of g_{kh} and Proposition 5.9. For the second term, we estimate similarly

$$g'_{kh}(\mathrm{I}_\sigma\bar\chi)(1,0) \le g'(\bar\chi)(1,0) + c_2\left(|\log k|(k+h^2) + \sigma(k,h)\right) \le -\bar\eta + \delta_2(k,h). \tag{5.23}$$

using Assumption 3.1 and $g'(\bar\chi)(1,0) = \partial_\nu g(\bar\chi) = -\bar\eta < 0$. Finally, for the third term, we find due to (5.13) that

$$g''_{kh}(\chi_\zeta)[\gamma,0]^2 \le c_3\gamma^2.$$

Collecting all estimates, we have

$$g_{kh}(\chi_\gamma) \le c_1\delta_1(k,h) - \gamma\left(\bar\eta - c_2\delta_2(k,h) - c_3\gamma\right).$$

Note that the first component of χ_γ is bounded below by $\bar\nu$ and bounded above by $\bar\nu + 1$, so that all constants of Propositions 5.7 and 5.9 can be chosen to be independent of χ_γ. Taking

$$\gamma = \frac{3c_1\delta_1(k,h)}{\bar\eta} \le \frac{\bar\eta}{3c_3} \quad \text{and} \quad c_2\delta_2(k,h) \le \frac{\bar\eta}{3}$$

for k, h sufficiently small, we obtain $g_{kh}(\chi_\gamma) \le 0$. From the definition of γ we further deduce

$$\gamma = \gamma(k,h) = \mathcal{O}(\sigma(k,h) + |\log k|(k+h^2)).$$

Moreover, it holds $\|\chi_\gamma - \bar\chi\| \le \rho$ for γ, k, h sufficiently small due to (5.21). In summary, we have that the sequence χ_γ is feasible for $(\hat P^\rho_{kh})$. □

In particular, Proposition 5.10 guarantees that for h, k, and ρ sufficiently small, the set of admissible controls of the discrete problem $(\hat P^\rho_{kh})$ is nonempty. Hence, by standard arguments we obtain that the localized discrete problem is well-posed.

Corollary 5.11. *Let h, k, and ρ be sufficiently small. Then there exists a solution $\bar{\chi}^\rho_{kh} = (\bar{\nu}^\rho_{kh}, \bar{q}^\rho_{kh}) \in \mathbb{R}_+ \times Q_{ad,\sigma}(0,1)$ to $(\hat{P}^\rho_{kh})$.*

In order to ensure that the constants in the following arguments are independent of $\bar{\nu}^\rho_{kh}$, we have to guarantee that $\bar{\nu}^\rho_{kh}$ is strictly and uniformly bounded from zero; cf., e.g., Propositions 5.7 and 5.9 and Appendix A.7. To this end, we always assume in the following that $\rho \le \bar{\nu}/2$, which implies $\bar{\nu}/2 \le \bar{\nu}^\rho_{kh} \le (3/2)\bar{\nu}$ by the localization in $(\hat{P}^\rho_{kh})$.

If k, h, and ρ are sufficiently small, then we easily verify that the linearized Slater condition holds at $\bar{\chi}^\rho_{kh}$ for the discrete problem.

Proposition 5.12. *For k, h, and ρ sufficiently small we have*

$$\partial_\nu g_{kh}(\bar{\chi}^\rho_{kh}) \le -\bar{\eta}/2 < 0.$$

Proof. This follows with Assumption 3.1 and

$$\partial_\nu g_{kh}(\bar{\chi}^\rho_{kh}) \le \partial_\nu g(\bar{\chi}) + |\partial_\nu g_{kh}(\bar{\chi}^\rho_{kh}) - \partial_\nu g(\bar{\chi}^\rho_{kh})| + |\partial_\nu g(\bar{\chi}^\rho_{kh}) - \partial_\nu g(\bar{\chi})|,$$

using the error estimate (5.19), the Lipschitz-continuity of $\partial_\nu g$ from Proposition 5.2, and the fact that $\|\bar{\chi}^\rho_{kh} - \bar{\chi}\| \le \rho$ by the construction of $(\hat{P}^\rho_{kh})$. □

Last, we construct a sequence that is feasible for $(\hat{P})$ and its distance to $(\bar{\nu}^\rho_{kh}, \bar{q}^\rho_{kh})$ converges at the rate $|\log k|(k + h^2)$.

Proposition 5.13. *Let k, h, and ρ be sufficiently small. Moreover, let $(\bar{\nu}, \bar{q})$ be a locally optimal solution of $(\hat{P})$ and let $(\bar{\nu}^\rho_{kh}, \bar{q}^\rho_{kh})$ be any globally optimal control of $(\hat{P}^\rho_{kh})$. Then there exists a sequence $\{\nu_\tau\}_{\tau>0}$ with $\tau = \tau(k,h)$ such that $(\nu_\tau, \bar{q}^\rho_{kh})$ is feasible for $(\hat{P})$ and that fulfill*

$$|\nu_\tau - \bar{\nu}^\rho_{kh}| \le c|\log k|(k + h^2).$$

Proof. We set

$$\chi_\tau = (\nu_\tau, q_\tau) = (\bar{\nu}^\rho_{kh} + \tau, \bar{q}^\rho_{kh}).$$

for some $\tau \in (0,1]$ to be determined later. Now, the proof proceeds along the lines of the proof of Proposition 5.10, interchanging the roles of $\bar{\chi}$ and $\bar{\chi}_{kh}$ and g and g_{kh} and using the result of Proposition 5.12 instead of Assumption 3.1. Clearly, we have

$$\|\chi_\tau - \bar{\chi}^\rho_{kh}\| \le \tau.$$

Moreover, using Taylor expansion of g at $\bar{\chi}^\rho_{kh}$ we find

$$g(\chi_\tau) = g(\bar{\chi}^\rho_{kh}) + \tau\, g'(\bar{\chi}^\rho_{kh})(1,0) + \frac{\tau^2}{2}\, g''(\chi_\zeta)[1,0]^2.$$

Using the triangle inequality, we estimate the first term by

$$g(\bar{\chi}^\rho_{kh}) \le g_{kh}(\bar{\chi}^\rho_{kh}) + |g_{kh}(\bar{\chi}^\rho_{kh}) - g(\bar{\chi}^\rho_{kh})| \le c_1\, |\log k|(k + h^2)$$

with Proposition 5.9. For the second term, we estimate similarly

$$g'(\bar{\chi}^\rho_{kh})(1,0) \le g'_{kh}(\bar{\chi}^\rho_{kh})(1,0) + c_2\, |\log k|(k + h^2).$$

Due to Proposition 5.12, it holds $g'_{kh}(\bar{\chi}^\rho_{kh})(1,0) = \partial_\nu g_{kh}(\bar{\chi}^\rho_{kh}) = -\bar{\eta}/2 < 0$. Finally, for the third term, we find that $g''(\chi_\zeta)[\tau,0]^2 \leq c_3\tau^2$. Collecting all estimates, we have

$$g(\chi_\tau) \leq c_1 \left|\log k\right|(k+h^2) - \tau\left(\bar{\eta}/2 - c_2\left|\log k\right|(k+h^2) - c_3\tau\right).$$

Note that the first component of χ_τ is bounded below by $\bar{\nu}^\rho_{kh} \geq \bar{\nu}/2$ and bounded above by $\bar{\nu}^\rho_{kh} + 1 \leq (3/2)\bar{\nu} + 1$, so that all constants of Proposition 5.9 can be chosen to be independent of χ_τ. Taking

$$\tau = \frac{6c_1\left|\log k\right|(k+h^2)}{\bar{\eta}} \leq \frac{\bar{\eta}}{6c_3} \quad \text{and} \quad c_2\left|\log k\right|(k+h^2) \leq \frac{\bar{\eta}}{6}$$

for k and h sufficiently small, we obtain $g(\chi_\tau) \leq 0$. From the definition of τ we further deduce $\tau = \tau(k,h) = \mathcal{O}(|\log k|(k+h^2))$. In summary, we have that the sequence χ_τ is feasible for $(\hat{P})$. □

5.3.2. Suboptimal error estimates for controls

Two-way insertion of the auxiliary sequences constructed in the preceding subsections with the quadratic growth condition yields a first convergence result.

Proposition 5.14. *Let $(\bar{\nu},\bar{q})$ be a local solution to $(\hat{P})$. Moreover, let $\{(k,h)\}$ be a sequence of positive mesh sizes converging to zero and $\{(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh})\}_{k,h>0}$ be a sequence of globally optimal solutions to $(\hat{P}^\rho_{kh})$ for $\rho > 0$ sufficiently small such that the quadratic growth condition* (3.16) *as well as Propositions 5.10 and 5.13 hold. Then $(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh})$ converges to $(\bar{\nu},\bar{q})$ and*

$$|\bar{\nu} - \bar{\nu}^\rho_{kh}| + \|\bar{q} - \bar{q}^\rho_{kh}\|_{L^2(I\times\omega)} \leq c\left(\sigma(k,h)^{1/2} + |\log k|^{1/2}(k^{1/2}+h)\right).$$

Proof. Because the tuple $(\nu_\tau,\bar{q}^\rho_{kh})$ from Proposition 5.13 is feasible for $(\hat{P})$, we may use the quadratic growth condition (3.16) to estimate

$$\begin{aligned}\frac{\delta}{2}\|(\bar{\nu}-\nu_\tau,\bar{q}-\bar{q}^\rho_{kh})\|^2 &\leq j(\nu_\tau,\bar{q}^\rho_{kh}) - j(\bar{\nu},\bar{q})\\ &\leq j(\nu_\tau,\bar{q}^\rho_{kh}) - j(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh}) + j(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh}) - j(\nu_\gamma,q_\gamma) + j(\nu_\gamma,q_\gamma) - j(\bar{\nu},\bar{q})\\ &\leq j(\nu_\tau,\bar{q}^\rho_{kh}) - j(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh}) + j(\nu_\gamma,q_\gamma) - j(\bar{\nu},\bar{q}),\end{aligned}$$

where the last inequality follows from optimality of the pair $(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh})$ and feasibility of (ν_γ,q_γ) for $(\hat{P}^\rho_{kh})$. Then, we observe

$$\begin{aligned}j(\nu_\tau,\bar{q}^\rho_{kh}) - j(\bar{\nu}^\rho_{kh},\bar{q}^\rho_{kh}) &= (\nu_\tau - \bar{\nu}^\rho_{kh})\left(1 + \frac{\alpha}{2}\|\bar{q}^\rho_{kh}\|^2_{L^2(I\times\omega)}\right)\\ &\leq c\left(1+\frac{\alpha}{2}\right)|\log k|(k+h^2)\end{aligned}$$

due to Proposition 5.13 and boundedness of $\bar{q}^\rho_{kh}$. Similarly,

$$\begin{aligned}j(\nu_\gamma,q_\gamma) - j(\bar{\nu},\bar{q}) &= (\nu_\gamma - \bar{\nu})\left(1 + \frac{\alpha}{2}\|q_\gamma\|^2_{L^2(I\times\omega)}\right)\\ &\quad + \bar{\nu}\frac{\alpha}{2}\|q_\gamma + \bar{q}\|_{L^2(I\times\omega)}\|q_\gamma - \bar{q}\|_{L^2(I\times\omega)}\\ &\leq c\left(1+\frac{\alpha}{2}\right)(\sigma(k,h) + |\log k|(k+h^2))\end{aligned}$$

employing Proposition 5.10. Taking square roots yields the assertion. □

Lemma 5.15. *Let $(\bar\nu, \bar q)$ be a local solution to $(\hat P)$ satisfying the quadratic growth condition (3.16) and $\{(k,h)\}$ be a sequence of positive mesh sizes converging to zero. There is a sequence $\{(\bar\nu_{kh}, \bar q_{kh})\}_{k,h>0}$ of local solutions to problem $(\hat P_{kh})$ such that*

$$|\bar\nu - \bar\nu_{kh}| + \|\bar q - \bar q_{kh}\|_{L^2(I\times\omega)} \le c\left(\sigma(k,h)^{1/2} + |\log k|^{1/2}(k^{1/2}+h)\right), \tag{5.24}$$

where $c > 0$ is independent of k, h, $\bar\nu_{kh}$, and $\bar q_{kh}$. Moreover, there exists a Lagrange multiplier $\bar\mu_{kh}$ such that the following optimality system is satisfied:

$$\bar\mu_{kh} > 0, \tag{5.25}$$

$$\int_0^1 1 + \frac{\alpha}{2}\|\bar q_{kh}(t)\|^2_{L^2(\omega)} + \langle B\bar q_{kh}(t) + \Delta_h \bar u_{kh}(t), \bar z_{kh}(t)\rangle \,\mathrm{d}t = 0, \tag{5.26}$$

$$\int_0^1 \bar\nu_{kh}\langle \alpha\bar q_{kh} + B^*\bar z_{kh}, q - \bar q_{kh}\rangle \ge 0, \quad q \in Q_{ad,\sigma}(0,1), \tag{5.27}$$

$$G(\bar u_{kh}(1)) = 0, \tag{5.28}$$

where $\bar u_{kh} = S_{kh}(\bar\nu_{kh}, \bar q_{kh})$ and $\bar z_{kh} \in X_{k,h}$ is the solution to the discrete adjoint equation

$$\mathrm{B}(\bar\nu_{kh}, \varphi_{kh}, \bar z_{kh}) = \bar\mu_{kh}(\bar u_{kh}(1) - u_d, \varphi_{kh}(1)), \quad \varphi_{kh} \in X_{k,h}.$$

Proof. The assertion follows from Proposition 5.14 noting that global solutions of $(\hat P^\rho_{kh})$ are local solutions of $(\hat P_{kh})$, since the constraint $\|(\nu_{kh} - \bar\nu, q_{kh} - \bar q)\| \le \rho$ is not active for sufficiently small $k, h > 0$ due to the convergence result of Proposition 5.14. Furthermore, Proposition 5.12 guarantees the existence of KKT multipliers satisfying the optimality system stated above. □

Proposition 5.16. *Adopt the assumptions of Lemma 5.15. Then it holds*

$$|\bar\mu - \bar\mu_{kh}| \le c\left(|\log k|(k+h^2) + \|(\bar\nu - \bar\nu_{kh}, \bar q - \bar q_{kh})\|\right), \tag{5.29}$$

with a constant $c > 0$ independent of k, h, $\bar\nu_{kh}$, $\bar q_{kh}$, and $\bar\mu_{kh}$.

Proof. We abbreviate $\bar\chi = (\bar\nu, \bar q)$ and $\bar\chi_{kh} = (\bar\nu_{kh}, \bar q_{kh})$. Combining the optimality conditions for $(\hat P)$ and $(\hat P_{kh})$ we obtain

$$\bar\mu - \bar\mu_{kh} = \partial_\nu g(\bar\chi)^{-1}\partial_\nu j(\bar\chi) - \partial_\nu g_{kh}(\bar\chi_{kh})^{-1}\partial_\nu j(\bar\chi_{kh}).$$

Now, we may use the discretization error estimate (5.19) to infer

$$\begin{aligned}
|\bar\mu - \bar\mu_{kh}| &\le |\partial_\nu g(\bar\chi)^{-1} - \partial_\nu g_{kh}(\bar\chi)^{-1}|\partial_\nu j(\bar\chi) \\
&\quad + |\partial_\nu g_{kh}(\bar\chi)^{-1}\partial_\nu j(\bar\chi) - \partial_\nu g_{kh}(\bar\chi_{kh})^{-1}\partial_\nu j(\bar\chi_{kh})| \\
&\le \frac{|\partial_\nu g(\bar\chi) - \partial_\nu g_{kh}(\bar\chi)|}{|\partial_\nu g(\bar\chi)\partial_\nu g_{kh}(\bar\chi)|}\partial_\nu j(\bar\chi) + \frac{|\partial_\nu g_{kh}(\bar\chi) - \partial_\nu g_{kh}(\bar\chi_{kh})|}{|\partial_\nu g_{kh}(\bar\chi)\partial_\nu g_{kh}(\bar\chi_{kh})|}\partial_\nu j(\bar\chi) \\
&\quad + |\partial_\nu g_{kh}(\bar\chi_{kh})^{-1}||\partial_\nu j(\bar\chi) - \partial_\nu j(\bar\chi_{kh})| \\
&\le c|\log k|(k+h^2) + c\|\bar\chi - \bar\chi_{kh}\|,
\end{aligned}$$

where we have used that $\partial_\nu j(\chi) = \int_0^1(1 + (\alpha/2)\|q\|^2)$ and that $|\partial_\nu g_{kh}(\bar\chi)| \ge \eta/2$ for k and h small enough, using again the discretization error estimate (5.19). □

5.3.3. Optimal error estimates for controls

Using the convergence result of the preceding subsection, we now prove optimal order of convergence with respect to the control variable. While the previous result is based on the quadratic growth condition, we now directly rely on the second order sufficient optimality condition and thus avoid taking square roots in the end. The improved convergence result will be consequence of the following Lemma.

Lemma 5.17. *Let $(\bar\nu, \bar q)$ be a local solution to $(\hat P)$ satisfying the second order sufficient optimality condition (3.15) and let $\{(k,h)\}$ be a sequence of positive mesh sizes such that $|\log k|(k+h^2) \to 0$. Let $\{(\bar\nu_{kh}, \bar q_{kh})\}_{k,h>0}$ be a sequence of local solutions to $(\hat P_{kh})$ converging in $\mathbb{R} \times L^2(I\times\omega)$ and associated Lagrange multipliers $\bar\mu_{kh}$ converging in $\mathbb{R}$. Then there are constants $c > 0$ and $k_0, h_0 > 0$ such that*

$$\|(\bar\nu - \bar\nu_{kh}, \bar q - \bar q_{kh})\|^2 \le c\left[|\log k|^2(k+h^2)^2 + \|\bar q - q_{kh}\|^2_{L^2(I\times\omega)} + \partial_q \mathcal{L}(\bar\nu, \bar q, \bar\mu)(q_{kh} - \bar q)\right] \tag{5.30}$$

for all $q_{kh} \in Q_{ad,\sigma}(0,1)$ and all $k \le k_0$ and $h \le h_0$.

Proof. We adapt the ideas of the proof of Theorem 2.14 in [31] for optimal control problems without state constraints. Instead of working with the objective functional, we use the Lagrange function $\mathcal{L}$ and the corresponding second order sufficient optimality condition (3.15). We abbreviate $\bar\chi = (\bar\nu, \bar q)$ and $\bar\chi_{kh} = (\bar\nu_{kh}, \bar q_{kh})$ with the norm $\|\chi\| = (|\nu|^2 + \|q\|^2_{L^2(I\times\omega)})^{1/2}$ on the product space.

Step 0: Preparation. Since $(\bar\nu, \bar q)$ is optimal for $(\hat P)$, it holds

$$\partial_\chi \mathcal{L}(\bar\chi, \bar\mu)(\chi - \bar\chi) \ge 0 \tag{5.31}$$

for all $\chi \in \mathbb{R}_+ \times Q_{ad}(0,1)$, and by the same arguments for the discrete problem $(\hat P_{kh})$

$$\partial_\chi \mathcal{L}_{kh}(\bar\chi_{kh}, \bar\mu_{kh})(\chi_{kh} - \bar\chi_{kh}) \ge 0 \tag{5.32}$$

for all $\chi_{kh} \in \mathbb{R}_+ \times Q_{ad,\sigma}(0,1)$.

Using (5.31) and the fact that $Q_\sigma(0,1) \subset Q(0,1)$, we find

$$\begin{aligned}\partial_\chi\left[\mathcal{L}(\bar\chi_{kh}, \bar\mu) - \mathcal{L}(\bar\chi, \bar\mu)\right](\bar\chi_{kh} - \bar\chi) &\le \partial_\chi \mathcal{L}(\bar\chi_{kh}, \bar\mu)(\bar\chi_{kh} - \bar\chi) \\ &\le \partial_\chi\left[\mathcal{L}(\bar\chi_{kh}, \bar\mu) - \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})\right](\bar\chi_{kh} - \bar\chi) + \partial_\chi \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})(\bar\chi_{kh} - \bar\chi).\end{aligned} \tag{5.33}$$

The first term on the right-hand side of (5.33) satisfies

$$\partial_\chi\left[\mathcal{L}(\bar\chi_{kh}, \bar\mu) - \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})\right](\bar\chi_{kh} - \bar\chi) = (\bar\mu - \bar\mu_{kh}) g'(\bar\chi_{kh})(\bar\chi_{kh} - \bar\chi).$$

Concerning the second term on the right-hand side of (5.33), using (5.32) and inserting additional terms with some arbitrary $\chi_{kh} \in \mathbb{R}_+ \times Q_{ad,\sigma}(0,1)$ yield

$$\begin{aligned}\partial_\chi \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})(\bar\chi_{kh} - \bar\chi) &\le \partial_\chi \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})(\bar\chi_{kh} - \bar\chi) + \partial_\chi \mathcal{L}_{kh}(\bar\chi_{kh}, \bar\mu_{kh})(\chi_{kh} - \bar\chi_{kh}) \\ &= \partial_\chi\left[\mathcal{L}_{kh}(\bar\chi_{kh}, \bar\mu_{kh}) - \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})\right](\bar\chi - \bar\chi_{kh}) + \partial_\chi \mathcal{L}_{kh}(\bar\chi_{kh}, \bar\mu_{kh})(\chi_{kh} - \bar\chi) \\ &= \partial_\chi\left[\mathcal{L}_{kh}(\bar\chi_{kh}, \bar\mu_{kh}) - \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})\right](\bar\chi - \bar\chi_{kh}) \\ &\quad + \partial_\chi\left[\mathcal{L}_{kh}(\bar\chi_{kh}, \bar\mu_{kh}) - \mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh})\right](\chi_{kh} - \bar\chi) \\ &\quad + \partial_\chi\left[\mathcal{L}(\bar\chi_{kh}, \bar\mu_{kh}) - \mathcal{L}(\bar\chi, \bar\mu_{kh})\right](\chi_{kh} - \bar\chi) + \partial_\chi \mathcal{L}(\bar\chi, \bar\mu_{kh})(\chi_{kh} - \bar\chi).\end{aligned} \tag{5.34}$$

Concerning the first term on the right-hand side, we find

$$\begin{aligned}\partial_\chi \left[\mathcal{L}_{kh}(\bar\chi_{kh},\bar\mu_{kh}) - \mathcal{L}(\bar\chi_{kh},\bar\mu_{kh})\right](\bar\chi - \bar\chi_{kh}) &= \bar\mu_{kh}\left[g'_{kh}(\bar\chi_{kh}) - g'(\bar\chi_{kh})\right](\bar\chi - \bar\chi_{kh})\\ &\le c|\log k|(k+h^2)\|\bar\chi - \bar\chi_{kh}\|,\end{aligned}$$

where we have used boundedness of the Lagrange multipliers $\bar\mu_{kh}$ due to Proposition 5.16 and the estimate (5.19). Similarly for the second term of (5.34), it holds

$$\partial_\chi \left[\mathcal{L}_{kh}(\bar\chi_{kh},\bar\mu_{kh}) - \mathcal{L}(\bar\chi_{kh},\bar\mu_{kh})\right](\chi_{kh} - \bar\chi) \le c|\log k|(k+h^2)\|\chi_{kh} - \bar\chi\|.$$

The third term of (5.34) is estimated using Lipschitz continuity of $\partial_\chi\mathcal{L}$ (due to Lipschitz continuity of g' on bounded sets)

$$\partial_\chi \left[\mathcal{L}(\bar\chi_{kh},\bar\mu_{kh}) - \mathcal{L}(\bar\chi,\bar\mu_{kh})\right](\chi_{kh} - \bar\chi) \le c\|\bar\chi_{kh} - \bar\chi\|\|\chi_{kh} - \bar\chi\|.$$

Since $\mathcal{L}$ is two times continuously differentiable we find

$$\partial^2_\chi\mathcal{L}(\check\chi_{kh},\bar\mu)[\bar\chi_{kh} - \bar\chi]^2 = \partial_\chi\left[\mathcal{L}(\bar\chi_{kh},\bar\mu) - \mathcal{L}(\bar\chi,\bar\mu)\right](\bar\chi_{kh} - \bar\chi) \tag{5.35}$$

with $\check\chi_{kh}$ in between $\bar\chi$ and $\bar\chi_{kh}$. Together with the estimates above, we obtain

$$\begin{aligned}\partial^2_\chi\mathcal{L}(\check\chi_{kh},\bar\mu)[\bar\chi_{kh} - \bar\chi]^2 \le{}& c|\log k|(k+h^2)\left(\|\bar\chi - \bar\chi_{kh}\| + \|\bar\chi - \chi_{kh}\|\right)\\ &+ c\|\bar\chi_{kh} - \bar\chi\|\|\chi_{kh} - \bar\chi\| + \partial_\chi\mathcal{L}(\bar\chi,\bar\mu_{kh})(\chi_{kh} - \bar\chi)\\ &+ |\bar\mu - \bar\mu_{kh}||g'(\bar\chi_{kh})(\bar\chi_{kh} - \bar\chi)|.\end{aligned} \tag{5.36}$$

We argue by contradiction. Suppose that (5.30) is false, then there exist a subsequence of mesh sizes $\{k_n, h_n\}$ converging to zero and $(\bar\nu_n, \bar q_n) \in \mathbb{R}_+ \times Q_{ad,\sigma}(0,1)$ such that $(\bar\nu_n, \bar q_n) \to (\bar\nu, \bar q)$ with

$$\|\bar\chi_n - \bar\chi\|^2 > n\left[(|\log k_n|(k_n + h_n^2))^2 + \|q_n - \bar q\|^2_{L^2(I\times\omega)} + \partial_q\mathcal{L}(\bar\chi,\bar\mu)(q_n - \bar q)\right],$$

where we use for convenience the short notation $\bar\nu_n = \bar\nu_{k_nh_n}$ and $\mathcal{L}_n = \mathcal{L}_{k_nh_n}$ etc. Setting $\chi_n = (\bar\nu, q_n)$, the inequality is equivalent to

$$\frac{1}{n} > \frac{(|\log k_n|(k_n + h_n^2))^2}{\|\bar\chi_n - \bar\chi\|^2} + \frac{\|\chi_n - \bar\chi\|^2}{\|\bar\chi_n - \bar\chi\|^2} + \frac{\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)(\chi_n - \bar\chi)}{\|\bar\chi_n - \bar\chi\|^2}. \tag{5.37}$$

Define $\rho_n = \|(\bar\nu_n - \bar\nu, \bar q_n - \bar q)\|$ and

$$v_n = (v_n^\nu, v_n^q) = \frac{1}{\rho_n}(\bar\chi_n - \bar\chi).$$

We may assume w.l.o.g. that $v_n^\nu \to v^\nu$ in $\mathbb{R}$ and $v_n^q \rightharpoonup v^q$ in $L^2(I\times\omega)$ and set $v = (v^\nu, v^q)$.

Step 1: $\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)v = 0$. The optimality condition (3.3) implies

$$\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)v = \lim_{n\to\infty}\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)v_n \ge 0.$$

To show the reverse inequality, we consider

$$\begin{aligned}\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)v &= \lim_{n\to\infty}\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)v_n \qquad (5.38)\\ &= \lim_{n\to\infty}\partial_\chi\mathcal{L}(\bar\chi,\bar\mu_n)v_n + \partial_\chi\left[\mathcal{L}(\bar\chi,\bar\mu) - \mathcal{L}(\bar\chi,\bar\mu_n)\right]v_n\\ &= \lim_{n\to\infty}\partial_\chi\mathcal{L}_n(\bar\chi_n,\bar\mu_n)v_n\\ &\quad + \lim_{n\to\infty}\partial_\chi\left[\mathcal{L}(\bar\chi_n,\bar\mu_n) - \mathcal{L}_n(\bar\chi_n,\bar\mu_n)\right]v_n\\ &\quad + \lim_{n\to\infty}\partial_\chi\left[\mathcal{L}(\bar\chi,\bar\mu_n) - \mathcal{L}(\bar\chi_n,\bar\mu_n)\right]v_n. \qquad (5.39)\end{aligned}$$

The limit in (5.38) exists due to weak convergence of (v_n^ν, v_n^q). Concerning the second limit in (5.39) we observe

$$\lim_{n\to\infty} \left[\partial_\chi \mathcal{L}(\bar\chi_n, \bar\mu_n) - \partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n)\right] v_n \\ = \lim_{n\to\infty} \bar\mu_n \left[g'(\bar\chi_n) - g_n'(\bar\chi_n)\right] v_n \le c \lim_{n\to\infty} |\log k_n|(k_n + h_n^2) = 0,$$

where we have used boundedness of $\bar\mu_n$ and (5.19). Using Lipschitz continuity we estimate the third limit as

$$\lim_{n\to\infty} \left[\partial_\chi \mathcal{L}(\bar\chi, \bar\mu_n) - \partial_\chi \mathcal{L}(\bar\chi_n, \bar\mu_n)\right] v_n \le c \lim_{n\to\infty} \|\bar\chi - \bar\chi_n\| = 0,$$

since $\|v_n\| = 1$. Thus, the first limit in (5.39) must exist as well.

Using continuity of $\partial_\chi \mathcal{L}$ in $\mathbb{R} \times L^2(I\times\omega)$ and the optimality condition (5.32) for $\bar\chi_n = (\bar\nu_n, \bar q_n)$ with $\chi_n = (\bar\nu, q_n)$ we find

$$\begin{aligned}\partial_\chi \mathcal{L}(\bar\chi, \bar\mu) v &\le \lim_{n\to\infty} \partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n) v_n \\ &= \lim_{n\to\infty} \frac{1}{\rho_n} \left[\partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n)(0, q_n - \bar q) + \partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n)(\bar\nu_n - \bar\nu, \bar q_n - q_n)\right] \\ &\le \lim_{n\to\infty} \frac{1}{\rho_n} \partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n)(0, q_n - \bar q).\end{aligned}$$

Since for any $\varphi \in \mathbb{R} \times L^2(I\times\omega)$ it holds

$$\begin{aligned}\partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n)\varphi &\le |\partial_\chi \mathcal{L}(\bar\chi_n, \bar\mu_n)\varphi| + |[\partial_\chi \mathcal{L}_n(\bar\chi_n, \bar\mu_n) - \partial_\chi \mathcal{L}(\bar\chi_n, \bar\mu_n)]\,\varphi| \\ &\le c\left(1 + |\log k_n|(k_n + h_n^2)\right) \|(\varphi^\nu, \varphi^q)\|,\end{aligned}$$

we conclude

$$\partial_\chi \mathcal{L}(\bar\chi, \bar\mu) v \le \lim_{n\to\infty} c\left(1 + |\log k_n|(k_n + h_n^2)\right) \frac{\|q_n - \bar q\|_{L^2(I\times\omega)}}{\rho_n} = 0,$$

due to (5.37). In summary, we proved $\partial_\chi \mathcal{L}(\bar\chi, \bar\mu) v = 0$.

Step 2: $g'(\bar\chi)v = 0$. Using $g(\bar\chi) = g_n(\bar\chi_n) = 0$, (5.18), (5.37), and step 1 we infer

$$\begin{aligned}j'(\bar\chi)v &= \lim_{n\to\infty} \frac{1}{\rho_n} \left[j(\bar\chi_n) - j(\bar\chi)\right] = \lim_{n\to\infty} \frac{1}{\rho_n} \left[\mathcal{L}_n(\bar\chi_n, \bar\mu) - \mathcal{L}(\bar\chi, \bar\mu)\right] \\ &= \lim_{n\to\infty} \frac{1}{\rho_n} \left[\mathcal{L}_n(\bar\chi_n, \bar\mu) - \mathcal{L}(\bar\chi_n, \bar\mu) + \mathcal{L}(\bar\chi_n, \bar\mu) - \mathcal{L}(\bar\chi, \bar\mu)\right] \\ &\le \limsup_{n\to\infty} \frac{c}{\rho_n} |\log k_n|(k_n + h_n^2) + \partial_\chi \mathcal{L}(\bar\chi, \bar\mu) v = 0.\end{aligned}$$

Similarly, we calculate

$$\begin{aligned}g'(\bar\chi)v &= \lim_{n\to\infty} \frac{1}{\rho_n} \left[g(\bar\chi_n) - g(\bar\chi)\right] = \lim_{n\to\infty} \frac{1}{\rho_n} \left[(g_n(\bar\chi_n) - g(\bar\chi)) + (g(\bar\chi_n) - g_n(\bar\chi_n))\right] \\ &\le \limsup_{n\to\infty} \frac{c}{\rho_n} |\log k_n|(k_n + h_n^2) = 0.\end{aligned}$$

Hence, from

$$\partial_\chi \mathcal{L}(\bar\chi, \bar\mu) v = j'(\bar\chi)v + \bar\mu\, g'(\bar\chi)v = 0$$

and $\bar\mu > 0$ (see (3.4)), we conclude $g'(\bar\chi)v = 0$.

Step 3: $v \in C_{(\bar\nu,\bar q)}$. Because the set

$$\left\{ \delta q \in L^2(I\times\omega) \;\middle|\; \begin{matrix} \delta q \le 0 \text{ if } \bar q(t,x) = q_b \\ \delta q \ge 0 \text{ if } \bar q(t,x) = q_a \end{matrix} \right\},$$

is closed and convex, it is in particular weakly closed. Moreover, due to feasibility of q_n every $(q_n - \bar q)/\rho_n$ belongs to the set above, so does the weak limit. Thus, v satisfies $v^q \le 0$, if $\bar q(t,x) = q_b$, and $v^q \ge 0$, if $\bar q(t,x) = q_a$. For this reason, (5.1) implies

$$\int_0^1 \int_\omega \bar\nu(\alpha\bar q + B^*\bar z)v^q \,\mathrm{d}x\,\mathrm{d}t = \int_0^1 \int_\omega \bar\nu|(\alpha\bar q + B^*\bar z)v^q| \,\mathrm{d}x\,\mathrm{d}t.$$

Moreover, due to $\partial_\chi \mathcal{L}(\bar\chi,\bar\mu)v = 0$ and the first order necessary condition $\partial_\nu \mathcal{L}(\bar\chi,\bar\mu) = 0$ we have the equality

$$0 = \partial_q \mathcal{L}(\bar\chi,\bar\mu)v^q = \int_0^1 \bar\nu(\alpha\bar q + B^*\bar z, v^q)_{L^2(\omega)} \,\mathrm{d}t = \int_0^1 \int_\omega \bar\nu|(\alpha\bar q + B^*\bar z)v^q| \,\mathrm{d}x\,\mathrm{d}t.$$

Hence, $v^q = 0$, if $\alpha\bar q(t,x) + B^*\bar z(t,x) \neq 0$, and v^q satisfies the sign condition (3.11) as well. With step 1 we have proved that $v \in C_{(\bar\nu,\bar q)}$.

Step 4: $v = 0$. Since $\bar\chi_n \to \bar\chi$ in $\mathbb{R}\times L^2(I\times\omega)$, it holds $\check\chi_n \to \bar\chi$, where $\check\chi_n$ was defined in (5.35). Thus, continuity of $\partial_\chi\mathcal{L}$ in $\mathbb{R}\times L^2(I\times\omega)$ yields

$$\begin{aligned} \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2 &\ge \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi,\bar\mu)v_n^2 + \liminf_{n\to\infty} \partial_\chi^2[\mathcal{L}(\check\chi_n,\bar\mu) - \mathcal{L}(\bar\chi,\bar\mu)]v_n^2 \\ &= \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi,\bar\mu)v_n^2. \end{aligned} \tag{5.40}$$

Due to (5.29) and (5.37) we have

$$\begin{aligned} \frac{1}{\rho_n^2}\partial_\chi\left[\mathcal{L}(\bar\chi,\bar\mu_n) - \mathcal{L}(\bar\chi,\bar\mu)\right](\chi_n - \bar\chi) &= \frac{1}{\rho_n^2}(\bar\mu_n - \bar\mu)g'(\bar\chi)(\chi_n - \bar\chi) \\ &\le c\frac{|\bar\mu - \bar\mu_n|}{\|\bar\chi_n - \bar\chi\|}\frac{\|\chi_n - \bar\chi\|}{\|\bar\chi_n - \bar\chi\|} \le \frac{c}{\sqrt{n}}\left(\frac{|\log k_n|(k_n + h_n^2)}{\|\bar\chi_n - \bar\chi\|} + 1\right) \le \frac{c}{\sqrt{n}}. \end{aligned} \tag{5.41}$$

Similarly, using (5.29) and since $|g'(\bar\chi)v_n| \to 0$ by step 2, it holds

$$\begin{aligned} \frac{|\bar\mu - \bar\mu_n||g'(\bar\chi_n)(\bar\chi_n - \bar\chi)|}{\|\bar\chi_n - \bar\chi\|^2} &\le \frac{|\bar\mu - \bar\mu_n|}{\|\bar\chi_n - \bar\chi\|}\left(|g'(\bar\chi)v_n| + |[g'(\bar\chi_n) - g'(\bar\chi)]v_n|\right) \\ &\le c\left(\frac{|\log k_n|(k_n + h_n^2)}{\|\bar\chi_n - \bar\chi\|} + 1\right)\left(|g'(\bar\chi)v_n| + \|\bar\chi_n - \bar\chi\|\right) \to 0. \end{aligned} \tag{5.42}$$

Employing (5.40) and (5.36) we infer

$$\begin{aligned} \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi,\bar\mu)v_n^2 \le \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2 &\le \limsup_{n\to\infty} \partial_\chi^2 \mathcal{L}(\check\chi_n,\bar\mu)v_n^2 \\ \le \limsup_{n\to\infty}\Bigg(&\frac{c|\log k_n|(k_n + h_n^2)}{\|\bar\chi_n - \bar\chi\|}\left(1 + \frac{\|\chi_n - \bar\chi\|}{\|\bar\chi_n - \bar\chi\|}\right) + c\frac{\|\chi_n - \bar\chi\|}{\|\bar\chi_n - \bar\chi\|} \\ &+ \frac{\partial_\chi\mathcal{L}(\bar\chi,\bar\mu)(\chi_n - \bar\chi)}{\|\bar\chi_n - \bar\chi\|^2} + \frac{\partial_\chi\left[\mathcal{L}(\bar\chi,\bar\mu_n) - \mathcal{L}(\bar\chi,\bar\mu)\right](\chi_n - \bar\chi)}{\|\bar\chi_n - \bar\chi\|^2} \\ &+ \frac{|\bar\mu - \bar\mu_n||g'(\bar\chi_n)(\bar\chi_n - \bar\chi)|}{\|\bar\chi_n - \bar\chi\|^2}\Bigg) = 0. \end{aligned} \tag{5.43}$$

Here, we have used (5.37) to estimate the first three summands, (5.41) for the second last term, and (5.42) for the last term. Last, weak lower semicontinuity of j'' and g'', and Corollary 3.9 lead to

$$\partial_\chi^2 \mathcal{L}(\bar\chi, \bar\mu) v^2 \le \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi, \bar\mu) v_n^2 \le 0.$$

From the second order sufficient condition (3.15) we conclude $v = (v^\nu, v^q) = 0$. Note that this in particular implies $v^\nu \to 0$ in $\mathbb{R}$.

Step 5: Final contradiction. Using $\|(v_n^\nu, v_n^q)\| = 1$ and $v^\nu \to 0$ we obtain

$$0 < \alpha\bar\nu = \alpha\bar\nu \liminf_{n\to\infty} \|(v_n^\nu, v_n^q)\|^2 = \liminf_{n\to\infty} \alpha \int_0^1 \bar\nu \|v_n^q(t)\|_{L^2(\omega)}^2 \,\mathrm{d}t.$$

Using the specific structure of j'' and again strong convergence $v_n^\nu \to 0$ in $\mathbb{R}$, it holds

$$\liminf_{n\to\infty} \alpha \int_0^1 \bar\nu \|v_n^q(t)\|_{L^2(\omega)}^2 \,\mathrm{d}t = \liminf_{n\to\infty} j''(\bar\chi)[v_n^\nu, v_n^q]^2.$$

Due to $g''(\bar\chi)[0,0]^2 = 0$ and weak lower semicontinuity, see Corollary 3.9, we conclude

$$\begin{aligned} 0 &< \liminf_{n\to\infty} j''(\bar\chi) v_n^2 \le \liminf_{n\to\infty} j''(\bar\chi) v_n^2 + \bar\mu \liminf_{n\to\infty} g''(\bar\chi) v_n^2 \\ &\le \liminf_{n\to\infty} \partial_\chi^2 \mathcal{L}(\bar\chi, \bar\mu) v_n^2 \le 0, \end{aligned}$$

where we have used again (5.43) in the last inequality. □

Finally we prove the main result of this section, i.e. a priori discretization error estimates that are optimal with respect to the control variable. We consider different control discretization strategies.

Variational discretization of controls

As proposed in [78] for elliptic equations, cf. also [118] for parabolic equations, the state and adjoint equations are discretized, only. The control is then implicitly discretized employing the optimality conditions, precisely the discrete analogue to (5.2). In this case, the operator I_σ is the identity and $\sigma(k,h) = 0$.

Theorem 5.18 (Variational discretization)**.** *Let the assumptions of Lemma 5.15 hold and suppose the variational control discretization, i.e. $Q_\sigma(0,1) = Q(0,1)$. Then there is a constant $c > 0$ not depending on k, h, $\bar\nu_{kh}$, and $\bar q_{kh}$ such that*

$$|\bar\nu - \bar\nu_{kh}| + \|\bar q - \bar q_{kh}\|_{L^2(I\times\omega)} \le c|\log k|(k + h^2).$$

Proof. Lemma 5.15 guarantees the existence of a sequence of local solutions converging strongly in $\mathbb{R} \times L^2(I \times \omega)$. Hence, we can apply Lemma 5.17 with $q_{kh} = \bar q$. □

In case of purely time-dependent control, the set ω is already discrete and the space $L^2(\omega) \cong \mathbb{R}^{N_c}$ does not need to be discretized; cf. Assumption 5.1. Moreover, in view of the projection formula

$$\bar q_{kh} = P_{Q_{ad}} \left(-\frac{1}{\alpha} B^* \bar z_{kh} \right), \tag{5.44}$$

which can be deduced from (5.27) with $Q_{ad,\sigma}(0,1) = Q_{ad}(0,1)$, the optimal control $\bar q_{kh}$ obtained by the variational approach is piecewise constant in time with values in $\mathbb{R}^{N_c}$. Based on this observation, the controls constructed in Theorem 5.18 are already contained in a discrete space, and we obtain the following corollary.

Corollary 5.19 (Parameter control)**.** *Let the assumptions of Lemma 5.15 hold, suppose that ω is discrete, and choose the piecewise constant discrete control space*

$$Q_\sigma(0,1) = \left\{ v \in Q(0,1) \colon v|_{I_m} \in \mathcal{P}_0(I_m; \mathbb{R}^{N_c}),\ m = 1, 2, \dots, M \right\}.$$

Then there is a constant $c > 0$ not depending on k, h, $\bar\nu_{kh}$, and $\bar q_{kh}$ such that

$$|\bar\nu - \bar\nu_{kh}| + \|\bar q - \bar q_{kh}\|_{L^2(I;\mathbb{R}^{N_c})} \le c|\log k|(k + h^2).$$

In the case of a distributed control, the variational control discretization has an additional implementation effort in practice. Fully discrete strategies are therefore of independent interest and we will investigate different variants in the following.

Cellwise constant control approximation

As the case of purely time-dependent controls is already covered by Corollary 5.19, in the following we restrict to the situation of a distributed control on a subset $\omega \subset \Omega$ of the spatial domain. The discrete space of controls is defined as follows

$$Q_\sigma(0,1) = \{ v \in Q(0,1) \colon v|_{I_m \times K} \in \mathcal{P}_0(I_m \times K) \text{ for all } K \in \mathcal{T}_h^\omega,\ m = 1, 2, \dots, M \}.$$

Abbreviating $\mathcal{I}_k = \{ 1, 2, \dots, M \}$, on any tuple $(I_m, K) \in \mathcal{I}_k \times \mathcal{T}_h$ we define the piecewise constant projection $\Pi_{kh} \colon L^2(I \times \omega) \to Q_\sigma(0,1)$ via

$$(\Pi_{kh} v)(t,x) = \frac{1}{k_m |K|} \int_{I_m} \int_K v(s,\xi) \,\mathrm{d}\xi \,\mathrm{d}s, \quad (t,x) \in I_m \times K, \tag{5.45}$$

for $v \in L^2(I \times \omega)$. Moreover, we introduce the L^2-projection onto the piecewise constant functions in time as

$$(\Pi_k v)(t) = \frac{1}{k_m} \int_{I_m} v(\xi) \,\mathrm{d}\xi, \quad t \in I_m, \tag{5.46}$$

for every $v \in L^2(I; L^2)$ and $m \in \{ 1, 2, \dots, M \}$. Then, for any $v \in H^1(I; L^2) \cap L^2(I; H^1)$ we obtain the projection error estimate

$$\begin{aligned} \|\Pi_{kh} v - v\|_{L^2(I;L^2)} &\le \|\Pi_{kh} v - \Pi_k v\|_{L^2(I;L^2)} + \|\Pi_k v - v\|_{L^2(I;L^2)} \\ &\le ch \|\nabla v\|_{L^2(I;L^2)} + ck \|\partial_t v\|_{L^2(I;L^2)}. \end{aligned} \tag{5.47}$$

We obtain the following error estimate for the discretization by cellwise constant controls. Note that also in this case Lemma 5.15 only provides a suboptimal estimate of order $(k + h)^{1/2}$.

Theorem 5.20 (Cellwise constant controls)**.** *Let the assumptions of Lemma 5.15 hold and suppose the piecewise constant control discretization. Then there is a constant $c > 0$ not depending on k, h, $\bar\nu_{kh}$, and $\bar q_{kh}$ such that*

$$|\bar\nu - \bar\nu_{kh}| + \|\bar q - \bar q_{kh}\|_{L^2(I\times\omega)} \le c|\log k|(k + h).$$

Proof. We would like to apply Lemma 5.17 with $\mathrm{I}_\sigma = \Pi_{kh}$ and $q_{kh} = \mathrm{I}_\sigma \bar q$. Using the adjoint state, we write the derivative of the Lagrangian as

$$\partial_q \mathcal{L}(\bar\nu, \bar q, \bar\mu) v = \int_0^1 \bar\nu (\alpha \bar q + B^* \bar z, v)_{L^2(\omega)}.$$

Orthogonality of Π_{kh} and $\bar\nu \in \mathbb{R}$ yield

$$\begin{aligned}\partial_q \mathcal{L}(\bar\nu, \bar q, \bar\mu)(\mathrm{I}_\sigma \bar q - \bar q) &= \int_0^1 \bar\nu(\alpha\bar q + B^*\bar z, \mathrm{I}_\sigma \bar q - \bar q)_{L^2(\omega)} \\ &= \bar\nu \int_0^1 (\alpha\bar q + B^*\bar z - \mathrm{I}_\sigma(\alpha\bar q + B^*\bar z), \mathrm{I}_\sigma \bar q - \bar q)_{L^2(\omega)} \\ &= \bar\nu \int_0^1 (B^*\bar z - \mathrm{I}_\sigma B^*\bar z, \mathrm{I}_\sigma \bar q - \bar q)_{L^2(\omega)} - \alpha\bar\nu \|\mathrm{I}_\sigma \bar q - \bar q\|^2_{L^2(I\times\omega)}.\end{aligned}$$

Hence, the improved regularity $\bar q, B^*\bar z \in H^1(I; L^2(\omega)) \cap L^2(I; H^1(\omega))$, see Proposition 5.3, the fact that $\alpha, \bar\nu > 0$, and $\|\mathrm{I}_\sigma \bar q - \bar q\|_{L^2(I;L^2)} \le c(k+h)$ due to (5.47), imply the estimates

$$\begin{aligned}\partial_q \mathcal{L}(\bar\nu, \bar q, \bar\mu)(q_{kh} - \bar q) &\le \bar\nu \|B^*\bar z - \mathrm{I}_\sigma B^*\bar z\|_{L^2(I\times\omega)} \|\mathrm{I}_\sigma \bar q - \bar q\|_{L^2(I\times\omega)} \\ &\le c\,(k+h)^2\,.\end{aligned}$$

Lemma 5.17 yields the assertion. □

Cellwise linear control approximation

The discrete space of controls is defined as follows

$$\begin{aligned}Q_h &= \{v \in C(\overline\omega)\colon v|_K \in \mathcal{P}_1(K) \text{ for all } K \in \mathcal{T}_h^\omega\}\,, \\ Q_\sigma(0,1) &= \{v \in Q(0,1)\colon v|_{I_m\times K} \in \mathcal{P}_0(I_m; Q_h) \text{ for all } m = 1, 2, \dots, M\}\,.\end{aligned}$$

Let $\mathrm{I}_h\colon C(\overline\omega) \to Q_h$ denote the Lagrange interpolant on ω. As before, let $\mathcal{I}_k = \{1, 2, \dots, M\}$ and decompose the set $\mathcal{I}_k \times \mathcal{T}_h^\omega$ as

$$\begin{aligned}\mathcal{S}_1 &= \{(m,K) \in \mathcal{I}_k \times \mathcal{T}_h^\omega\colon |\alpha\bar q + B^*\bar z| > 0 \text{ a.e. in } I_m \times K\}\,, \\ \mathcal{S}_2 &= \{(m,K) \in \mathcal{I}_k \times \mathcal{T}_h^\omega\colon \alpha\bar q + B^*\bar z = 0 \text{ a.e. in } I_m \times K\}\,, \\ \mathcal{S}_3 &= (\mathcal{I}_k \times \mathcal{T}_h^\omega) \setminus (\mathcal{S}_1 \cup \mathcal{S}_2)\,.\end{aligned}$$

Under an additional assumption we obtain the following convergence result.

Theorem 5.21 (Cellwise linear controls)**.** *Adapt the assumption of Lemma 5.15 and suppose the temporal piecewise constant and spatial piecewise linear control discretization. Assume that there is $p > d$ such that $G'(\bar u(1))^* \in W_0^{1,p}(\Omega)$ and that there is $c > 0$ such that*

$$\sum_{(m,K)\in\mathcal{S}_3} k_m |K| \le ch. \tag{5.48}$$

Then there is a constant $c > 0$ not depending on k, h, $\bar\nu_{kh}$, and $\bar q_{kh}$ such that

$$|\bar\nu - \bar\nu_{kh}| + \|\bar q - \bar q_{kh}\|_{L^2(I\times\omega)} \le c|\log k|(k + h^{3/2-1/p}).$$

Remark 5.22. Since $\bar u(1) \in W_0^{1,p}$ for every $p \in (1,\infty)$, see Proposition A.22, for the prototypical problem the assumption on $G'(\bar u(1))^*$ reduces to the requirement $u_d \in W_0^{1,p}$. Thus, assuming $u_d \in W_0^{1,\infty}$, yields $\mathcal{O}(|\log k|(k + h^{3/2-\varepsilon}))$ for any $\varepsilon > 0$. However, the constant in Theorem 5.21 will depend on p, hence also on ε.

Similar assumptions to (5.48) have been used in related publications for cellwise linear control discretization; see, e.g., [118, Section 5.2] for a linear parabolic equation and [31, Theorem 4.5] for a quasilinear elliptic equation. The assumption can be justified by the observation that in practice the boundary of the active set of $\bar q$ often has zero measure.

Proof of Theorem 5.21. We set $\mathrm{I}_\sigma = \mathrm{I}_h\Pi_k$ with Π_k being the piecewise constant in time projection defined in (5.46) and apply Lemma 5.17 with $q_{kh} = \mathrm{I}_\sigma\bar q$.

Clearly, if $(m,K) \in \mathcal{S}_1$, then either $\bar q(t,x) = q_a$ or $\bar q(t,x) = q_b$ for $(t,x) \in I_m \times K$, whence $\bar q - \mathrm{I}_h\Pi_k\bar q \equiv 0$ in $I_m \times K$. If $(m,K) \in \mathcal{S}_2$, then it holds $\bar q(t,x) = -\alpha^{-1}B^*\bar z(t,x)$ for $(t,x) \in I_m \times K$. According to [24, Theorem 4.4.4] it holds

$$\|\Pi_k\bar q - \mathrm{I}_h\Pi_k\bar q\|_{L^2(I_m\times K)} \le ch^2\|\Pi_k\bar q\|_{L^2(I_m;H^2(K))} \le ch^2\|\bar q\|_{L^2(I_m;H^2(K))}.$$

Hence,

$$\sum_{(m,K)\in\mathcal{S}_2}\|\Pi_k\bar q - \mathrm{I}_\sigma\bar q\|^2_{L^2(I_m\times K)} \le ch^4\sum_{(m,K)\in\mathcal{S}_2}\|\bar z\|^2_{L^2(I_m;H^2(K))} \le ch^4\|\bar z\|^2_{L^2(I;H^2(\Omega))}.$$

Moreover, for $(m,K) \in \mathcal{S}_3$ we have the error estimate

$$\|\Pi_k\bar q - \mathrm{I}_h\Pi_k\bar q\|_{L^p(I_m\times K)} \le ch\|\Pi_k\bar q\|_{L^p(I_m;W^{1,p}(K))} \le ch\|\bar q\|_{L^p(I_m;W^{1,p}(K))}; \tag{5.49}$$

see [24, Theorem 4.4.4]. Using Hölder's inequality, we find

$$\begin{aligned}
\sum_{(m,K)\in\mathcal{S}_3}\|\Pi_k\bar q - \mathrm{I}_\sigma\bar q\|^2_{L^2(I_m\times K)} &\le \sum_{(m,K)\in\mathcal{S}_3}(k_m|K|)^{1-2/p}\|\Pi_k\bar q - \mathrm{I}_h\Pi_k\bar q\|^2_{L^p(I_m\times K)}\\
&\le ch^2\sum_{(m,K)\in\mathcal{S}_3}(k_m|K|)^{1-2/p}\|\bar q\|^2_{L^p(I_m;W^{1,p}(K))}\\
&\le ch^2\|\bar q\|^2_{L^p(I;W^{1,p}(\omega))}\left(\sum_{(m,K)\in\mathcal{S}_3}k_m|K|\right)^{1-2/p}\\
&\le ch^{3-2/p}\|\bar q\|^2_{L^p(I;W^{1,p}(\omega))},
\end{aligned}$$

due to $\bar q \in L^p(I;W^{1,p}(\omega))$; see Proposition 5.3. In summary, we obtain the estimate

$$\|\bar q - q_{kh}\|_{L^2(I\times\omega)} \le \|\bar q - \Pi_k\bar q\|_{L^2(I\times\omega)} + \|\Pi_k\bar q - \mathrm{I}_\sigma\bar q\|_{L^2(I\times\omega)} \le c(k + h^{3/2-1/p}).$$

Next, we consider the term $\partial_q\mathcal{L}(\bar\nu,\bar q,\bar\mu)(q_{kh} - \bar q)$. Using orthogonality of Π_k we find

$$\begin{aligned}
\partial_q\mathcal{L}(\bar\nu,\bar q,\bar\mu)(\Pi_k\bar q - \bar q) &= \bar\nu\int_0^1(\alpha\bar q + B^*\bar z, \Pi_k\bar q - \bar q)_{L^2(\omega)}\\
&= \bar\nu\int_0^1(\alpha\bar q + B^*\bar z - \Pi_k(\alpha\bar q + B^*\bar z), \Pi_k\bar q - \bar q)_{L^2(\omega)}\\
&= \alpha\bar\nu\int_0^1(\bar q - \Pi_k\bar q, \Pi_k\bar q - \bar q)_{L^2(\omega)} + \bar\nu\int_0^1(B^*(\bar z - \Pi_k\bar z), \Pi_k\bar q - \bar q)_{L^2(\omega)}\\
&\le \bar\nu\|\bar z - \Pi_k\bar z\|_{L^2(I;L^2)}\|\bar q - \Pi_k\bar q\|_{L^2(I;L^2)}\\
&\le ck^2\|\partial_t\bar z\|_{L^2(I;L^2)}\|\partial_t\bar q\|_{L^2(I;L^2)}.
\end{aligned}$$

Moreover, due to the definitions of $\mathcal{S}_1$ and $\mathcal{S}_2$ it holds

$$\sum_{(m,K)\in\mathcal{S}_1}\int_{I_m}(\alpha\bar q + B^*\bar z, \underbrace{\mathrm{I}_\sigma\bar q - \Pi_k\bar q}_{=0})_{L^2(K)} + \sum_{(m,K)\in\mathcal{S}_2}\int_{I_m}(\underbrace{\alpha\bar q + B^*\bar z}_{=0}, \mathrm{I}_\sigma\bar q - \Pi_k\bar q)_{L^2(K)} = 0.$$

Last, for each $(m,K) \in \mathcal{S}_3$ there is $(\hat t_m, \hat x_K)$ such that $\alpha\bar q(\hat t_m,\hat x_K) + B^*\bar z(\hat t_m,\hat x_K) = 0$. Therefore, abbreviating $w := \alpha\bar q + B^*\bar z$ we find

$$\begin{aligned}
\int_{I_m}(w, \mathrm{I}_\sigma\bar q - \Pi_k\bar q)_{L^2(K)} &= \int_{I_m}(w - w(\hat t_m,\hat x_K), \mathrm{I}_\sigma\bar q - \Pi_k\bar q)_{L^2(K)}\\
&\le (k_m|K|)^{1-2/p}\|w - w(\hat t_m,\hat x_K)\|_{L^p(I_m\times K)}\|\mathrm{I}_\sigma\bar q - \Pi_k\bar q\|_{L^p(I_m\times K)}.
\end{aligned}$$

Moreover, the improved regularity $\bar{q} \in C([0,1]; W^{1,p}(\omega))$ and $\bar{z} \in C([0,1]; W_0^{1,p})$, see Proposition 5.3, as well as $p > d+1$ imply

$$\|w - w(\hat{t}_m, \hat{x}_K)\|_{L^p(I_m \times K)} \leq c(k+h) \left(\|\partial_t w\|^p_{L^p(I_m \times K)} + \|\nabla w\|^p_{L^p(I_m \times K)} \right)^{1/p}.$$

Summation over all $(m,K) \in \mathcal{S}_3$, Hölder's inequality, and using assumption (5.48) as well as the error estimate (5.49) yield

$$\begin{aligned}
\int_I (w, \mathrm{I}_\sigma \bar{q} - \Pi_k \bar{q})_{L^2(\omega)} &= \sum_{(m,K) \in \mathcal{S}_3} \int_{I_m} (w, \mathrm{I}_\sigma \bar{q} - \Pi_k \bar{q})_{L^2(K)} \\
&\leq \sum_{(m,K) \in \mathcal{S}_3} (k_m |K|)^{1-2/p} \|w - w(\hat{t}_m, \hat{x}_K)\|_{L^p(I_m \times K)} \|\mathrm{I}_\sigma \bar{q} - \Pi_k \bar{q}\|_{L^p(I_m \times K)} \\
&\leq c(k+h) \left(\sum_{(m,K) \in \mathcal{S}_3} k_m |K| \right)^{1-2/p} \|\mathrm{I}_\sigma \bar{q} - \Pi_k \bar{q}\|_{L^p(I \times \omega)} \\
&\leq c h^{1-2/p} (k+h) h \leq c (k + h^{3/2-1/p})^2.
\end{aligned}$$

Lemma 5.17 yields the result. □

5.4. Numerical examples

To validate the theoretical findings of the preceding section in practice, we consider different numerical examples. All examples are implemented in MATLAB®. We use the augmented Lagrangian method as presented in Section 4.1 in order to deal with the state constraint, where we employ the parameter updates suggested in [15, Proposition 2] and [14, p. 414ff.]. The resulting optimal control problem is then solved using the trust-region semismooth Newton algorithm from [92] in a monolithic way, i.e. we optimize for the pair (ν, q) instead of a bilevel optimization. If the absolute value of the terminal constraint is smaller than 10^{-9}, the augmented Lagrangian method is stopped.

5.4.1. Example with analytic reference solution

First, we consider an academic test problem, where a solution can be given explicitly. Let

$$\begin{aligned}
\Omega = \omega = (0,1)^2, \quad \alpha = 1, \quad \delta_0 = \tfrac{1}{2}, \\
G(u) = \tfrac{1}{2}\|u - u_d\|^2_{L^2} - \tfrac{1}{2}\delta_0^2, \quad u_d(x) = -2\sin(\pi x_1)\sin(\pi x_2), \\
u_0(x) = \sin(\pi x_1)\sin(\pi x_2),
\end{aligned}$$

without control constraints. Moreover, we use the operator $-c\Delta$ with $c = 1/(2\pi^2)$ for convenience. The optimal state and adjoint state are given by

$$\bar{u}(t,x) = 2\left(\mathrm{e}^{-\bar{\nu}t} - \mathrm{e}^{\bar{\nu}(t-1)}\right) u_0(x), \quad \bar{z}(t,x) = 4\mathrm{e}^{\bar{\nu}(t-1)} u_0(x),$$

with optimal time $T = \bar{\nu} = \log(2)$. To verify the second order sufficient optimality condition, consider the second derivative of the Lagrange function that is given by

$$\partial^2_{(\nu,q)} \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu, \delta q]^2 = \bar{\nu}\|\delta q\|^2_{L^2(I \times \Omega)} + 2\delta\nu \int_0^1 (\delta q, \bar{q}) + \bar{\mu} g''(\bar{\nu}, \bar{q})[\delta\nu, \delta q]^2.$$

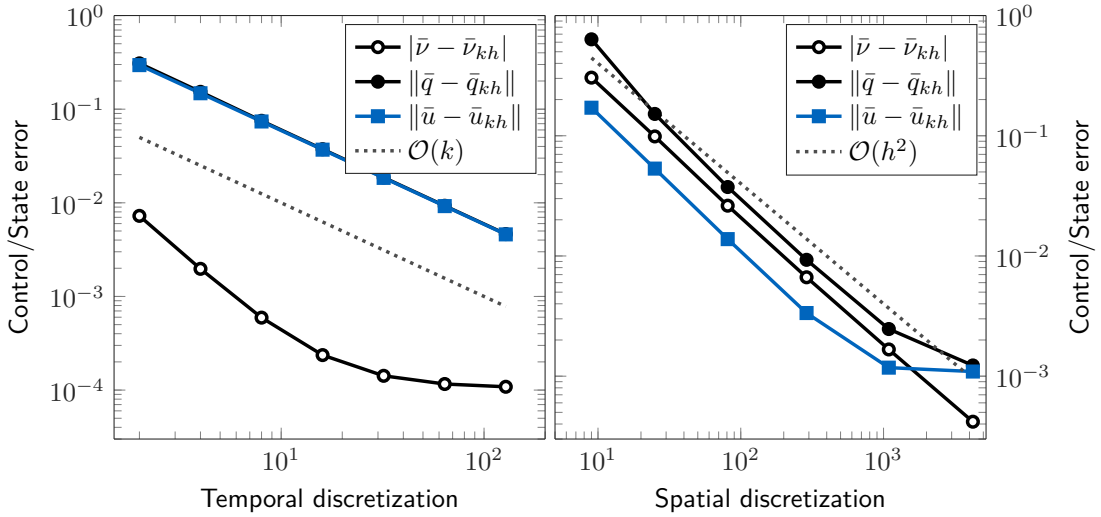

Figure 5.1.: Discretization error for Example 5.4.1 with variational control discretization and refinement of the time interval for $N = 16641$ nodes (left) and refinement of the spatial discretization for $M = 512$ time steps (right).

Introducing an additional adjoint state $\hat{z}$ defined as the solution to

$$-\partial_t \hat{z} - \bar{\nu} c \Delta \hat{z} = \Delta \bar{z}, \quad \hat{z}(1) = 0,$$

we obtain

$$\bar{\mu} g''(\bar{\nu}, \bar{q})[\delta\nu, \delta q]^2 = \bar{\mu} \|\delta u(1)\|_{L^2}^2 + 2\delta\nu \int_0^1 (\delta q, \bar{z} + \bar{\nu}\hat{z}) + 2\delta\nu^2 \int_0^1 (\bar{q} + c\Delta\bar{u}, \hat{z});$$

see Section 3.2.3 for details. Using that $\bar{\mu} > 0$, Cauchy's and Young's inequalities, and $\bar{q} = -\bar{z}$, we find

$$\begin{aligned}\partial^2_{(\nu,q)} \mathcal{L}(\bar{\nu}, \bar{q}, \bar{\mu})[\delta\nu, \delta q]^2 &\geq \bar{\nu} \|\delta q\|^2_{L^2(I\times\Omega)} + 2\bar{\nu}\delta\nu \int_0^1 (\delta q, \hat{z}) + 2\delta\nu^2 \int_0^1 (\bar{q} + c\Delta\bar{u}, \hat{z}) \\ &\geq -\bar{\nu}\delta\nu^2 \|\hat{z}\|^2_{L^2(I\times\Omega)} + 2\delta\nu^2 \int_0^1 (\bar{q} + c\Delta\bar{u}, \hat{z}).\end{aligned}$$

The solution of the additional adjoint equation is given by

$$\hat{z}(t, x) = 4(t-1)\mathrm{e}^{\bar{\nu}(t-1)} u_0(x).$$

We calculate

$$\begin{aligned}-\bar{\nu}\|\hat{z}\|^2_{L^2(I\times\Omega)} &= \frac{\log(2)\left(-6 + \log^2(4) + \log(16)\right)}{\log^3(4)} \approx -0.33968, \\ 2\int_0^1 (\bar{q} + c\Delta\bar{u}, \hat{z}) &= \frac{3 + \log^2(4) - \log(4)}{\log^2(4)} \approx 1.8397.\end{aligned}$$

Therefore, the second order sufficient optimality condition of Theorem 3.13 is satisfied on the whole space $\mathbb{R} \times L^2(I \times \Omega)$. Note that as in Lemma 3.18 it suffices to verify the second order sufficient optimality condition for $\delta\nu \neq 0$. In case $\delta\nu = 0$ the SSC is trivially fulfilled.

Since no control constraints are active this situation corresponds to the variational control discretization. We observe linear order of convergence with respect to the temporal and quadratic order of convergence with respect to the spatial discretization; see Figure 5.1.

M	N	$\alpha = 1$		$\alpha = 0.1$		$\alpha = 0.01$		$\alpha = 0.001$	
40	1089	7.439	4.548_{-1}	17.28	4.958_{-2}	1.523_{+3}	6.133_{-3}	2.422_{+6}	6.151_{-4}
80	1089	7.555	4.560_{-1}	17.66	4.895_{-2}	2.513_{+3}	6.052_{-3}	2.958_{+6}	6.062_{-4}
160	1089	7.546	4.531_{-1}	18.09	4.875_{-2}	2.505_{+3}	6.008_{-3}	1.369_{+6}	6.019_{-4}
320	1089	7.509	4.505_{-1}	18.29	4.864_{-2}	2.466_{+3}	5.986_{-3}	5.337_{+5}	5.997_{-4}
640	25	8.949	5.534_{-1}	19.22	5.812_{-2}	1.685_{+3}	6.829_{-3}	1.954_{+5}	6.844_{-4}
640	81	7.754	4.737_{-1}	18.44	5.080_{-2}	2.214_{+3}	6.182_{-3}	2.179_{+5}	6.194_{-4}
640	289	7.549	4.542_{-1}	18.30	4.896_{-2}	2.399_{+3}	6.016_{-3}	2.100_{+5}	6.027_{-4}
640	1089	7.507	4.495_{-1}	18.24	4.849_{-2}	2.473_{+3}	5.975_{-3}	2.953_{+5}	5.986_{-4}
Inactive constraints		96%		62%		5%		$< 1\%$	

Table 5.1.: Numerical verification of second order sufficient optimality condition for Example 5.4.2. Table shows the quantity (3.27) of Lemma 3.18 and the coercivity constant of Proposition 3.20 for different temporal and spatial degrees of freedoms and cost parameter α.

5.4.2. Example with purely time-dependent control

Next, we consider a time-optimal control problem with purely time-dependent controls for fixed spatially dependent functions. Let

$$\begin{aligned} \Omega = (0,1)^2, \quad \omega_1 = (0,0.5)\times(0,1), \quad \omega_2 = (0.5,1)\times(0,0.5), \quad \alpha = 10^{-2}, \\ B\colon \mathbb{R}^2 \to L^2(\Omega), \quad Bq = q_1 \mathbb{1}_{\omega_1} + q_2 \mathbb{1}_{\omega_2}, \\ G(u) = \tfrac{1}{2}\|u - u_d\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad u_d(x) = 0, \quad \delta_0 = \tfrac{1}{10}, \\ Q_{ad}(0,1) = \{q \in L^2(I;\mathbb{R}^2)\colon -1.5 \le q \le 0\}, \quad u_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2^3), \end{aligned}$$

where $\mathbb{1}_{\omega_1}$ and $\mathbb{1}_{\omega_2}$ denote the characteristic functions on ω_1 and ω_2. The spatial mesh is chosen such that the boundaries of ω_1 and ω_2 coincide with edges of the mesh, so that the control operator B can be easily implemented.

Since the control constraints are constants numbers, this corresponds to a variational control discretization and Theorem 5.18 applies. The solutions of the discrete problem are compared to a discrete solution calculated on a sufficiently fine mesh as we do not know the solution of the continuous problem. The optimal time is $T \approx 1.79931$. We observe linear convergence with respect to the temporal mesh size and quadratic order of convergence with respect to the spatial mesh size; see Figure 5.2.

To assess the validity of the second order sufficient optimality hypothesis, we verify the scalar condition of Lemma 3.18 for the discrete problem. Since the linear system (3.28) defines a symmetric but not a positive definite matrix, we calculate a solution using MINRES without assembling the matrix. We observe that for all choices of the cost parameter α the condition is satisfied on the discrete level; see Table 5.1. Note that the SSC for the discrete problem does not guarantee that the SSC for the continuous problem holds. However, the fact that the numbers are robust with respect to mesh refinement can serve as an indication for the continuous problem. In accordance with Proposition 3.20, we observe that the number of (3.27) from Lemma 3.18 increases as α decreases. In contrast, the constant $\bar{\gamma}$ increases. This can be explained as follows: Since we fix $\delta\nu = 1$, the variable $\delta\bar{q}$ has to counteract the decrease of $C_{\bar{q}}$ in order to satisfy the linear constraint $g'_{kh}(\bar{\nu}_{kh}, \bar{q}_{kh})(1, \delta\bar{q}) = 0$ resulting in an increase of the norm of $\delta\bar{q}$.

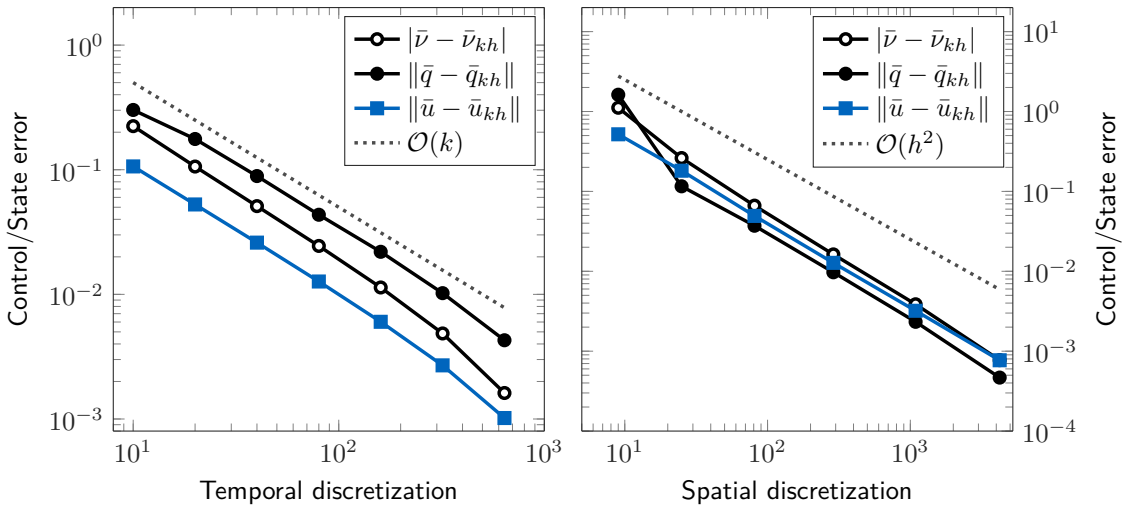

Figure 5.2.: Discretization error for Example 5.4.2 with variational control discretization and refinement of the time interval for $N = 1089$ nodes (left) and refinement of the spatial discretization for $M = 320$ time steps (right).

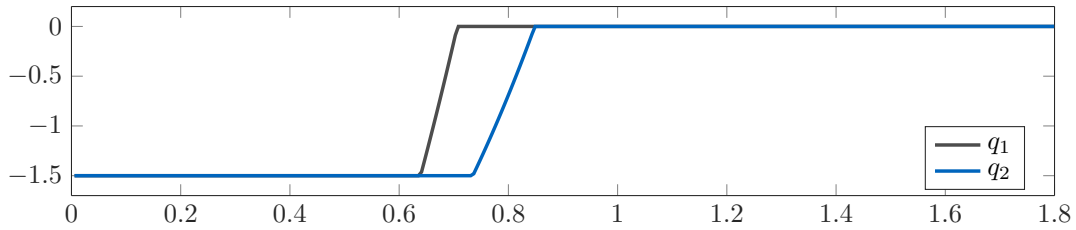

Figure 5.3.: Solution for example with purely time-dependent control.

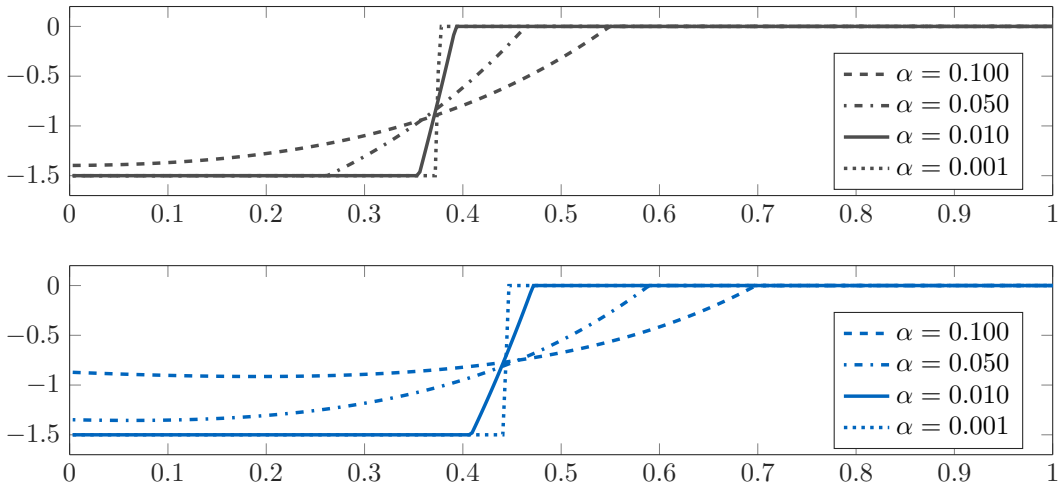

Figure 5.4.: Optimal control $\bar{q}_1$ (top) and $\bar{q}_2$ (bottom) for example with purely time-dependent control for different regularization parameters. In order to compare the solutions, the variables have been transformed to the reference time interval. The optimal times are approximately 1.8357, 1.8089, 1.7993, and 1.7991 ($M = 320$, $N = 1089$). Hence, the optimal T is not verify sensitive with respect to α for small α. We will investigate this behavior in detail in Section 5.5.

M	N	$\alpha = 1$		$\alpha = 0.1$		$\alpha = 0.01$		$\alpha = 0.001$	
20	4225	19.68	2.456_{-1}	27.49	1.550_{-2}	4.641_{+2}	3.852_{-3}	1.622_{+4}	4.227_{-4}
40	4225	17.32	2.289_{-1}	27.11	1.511_{-2}	4.573_{+2}	3.765_{-3}	1.515_{+4}	4.135_{-4}
80	4225	16.14	2.203_{-1}	26.94	1.493_{-2}	4.567_{+2}	3.725_{-3}	1.470_{+4}	4.089_{-4}
160	4225	15.55	2.159_{-1}	26.90	1.487_{-2}	4.578_{+2}	3.706_{-3}	1.477_{+4}	4.067_{-4}
320	25	2.740	5.633_{-2}	23.62	1.201_{-2}	3.307_{+2}	2.681_{-3}	4.002_{+4}	2.957_{-4}
320	81	9.968	1.519_{-1}	26.54	1.344_{-2}	3.809_{+2}	3.275_{-3}	1.749_{+4}	3.655_{-4}
320	289	13.69	1.968_{-1}	26.52	1.437_{-2}	4.459_{+2}	3.588_{-3}	1.576_{+4}	3.946_{-4}
320	1089	14.92	2.102_{-1}	26.81	1.474_{-2}	4.602_{+2}	3.677_{-3}	1.390_{+4}	4.032_{-4}
Inactive constraints		98%		67%		19%		6%	

Table 5.2.: Numerical verification of second order sufficient optimality condition for Example 5.4.3. Table shows the quantity (3.27) of Lemma 3.18 and the coercivity constant of Proposition 3.20 for different temporal and spatial degrees of freedoms and cost parameter α.

5.4.3. Example with distributed control on subdomain

Last, we consider an example with distributed control on a subset of the domain. As before we compare to a reference solution obtained numerically on a fine grid. The problem data is given by

$$\begin{aligned} \Omega = (0,1)^2, \quad \omega = (0,0.75)^2, \quad \alpha = 10^{-2}, \quad \delta_0 = \tfrac{1}{10}, \\ G(u) = \tfrac{1}{2}\|u - u_d\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad u_d(x) = -2\min\{x_1, 1-x_1, x_2, 1-x_2\}, \\ Q_{ad}(0,1) = \{\, q \in L^2(I\times\omega) \colon -5 \le q \le 0 \,\}, \\ u_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2)^3. \end{aligned}$$

We consider the operator $-c\Delta$ with $c = 0.03$. Note that the control acts only on a subset $\omega \subsetneq \Omega$. Moreover, the control constraints as well as the regularization parameter are chosen in a way such that the constraints on the control are active in a large region.

The optimal time we obtain numerically is approximately $T \approx 1.22198$. Figures 5.5 and 5.7 show the optimal state and control. The control is discretized by cellwise constant functions in space. In accordance with Theorem 5.20 we observe linear convergence in time and space for the control variable; see Figure 5.8. In contrast, for the optimal time and the state we obtain quadratic order of convergence in h. The improved convergence rate cannot be explained by the theory so far. However, we expect that one can also prove full order of convergence for all variables if the control is post-processed in an appropriate way by using the projection formula for the optimal control; see, e.g., [118, 121].

As before, we assess the validity of the second order sufficient optimality hypothesis, by verifying the scalar condition of Lemma 3.18 for the discrete problem. For all choices of the regularization parameter α, we observe that the condition is satisfied; see Table 5.2. However, if the functional in (3.30) is minimized over the whole space $L^2(I\times\omega)$ instead of the subspace $C_{\bar q}$ we observe that this (strong) second order sufficient optimality condition is not satisfied for small α; see Table 5.3. Therefore, it is essential to work with the critical cone $C_{(\bar\nu,\bar q)}$ in the formulation of the second order conditions.

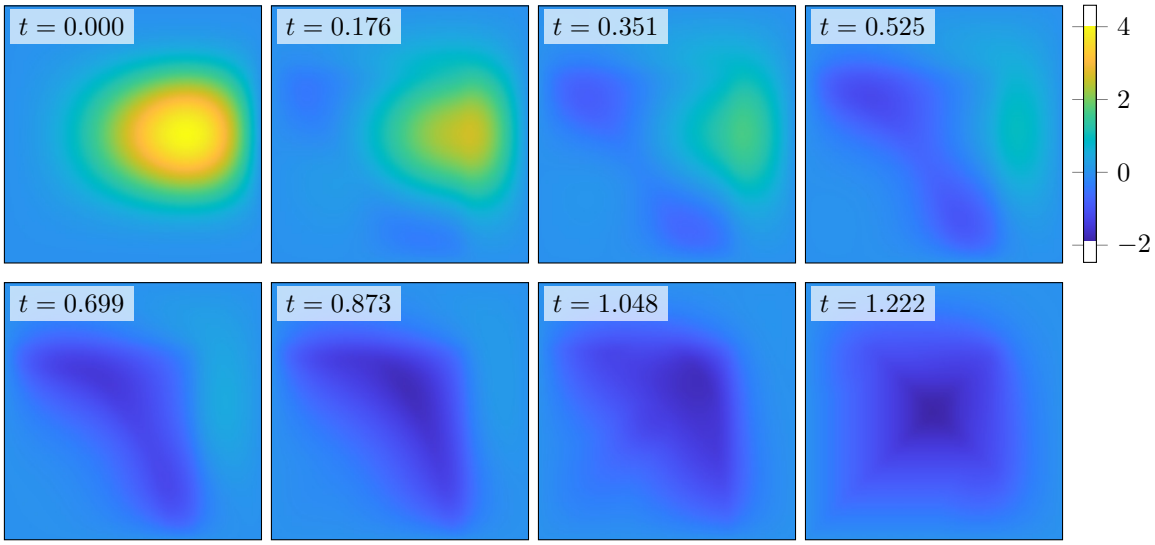

Figure 5.5.: Snapshots of optimal state for Example 5.4.3.

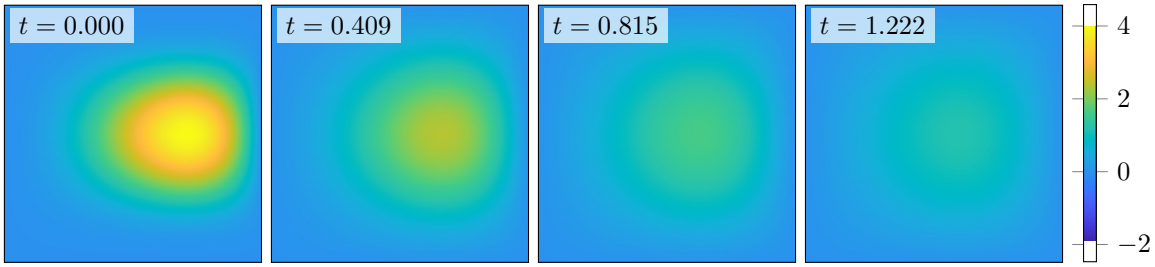

Figure 5.6.: Snapshots of state for uncontrolled equation for Example 5.4.3.

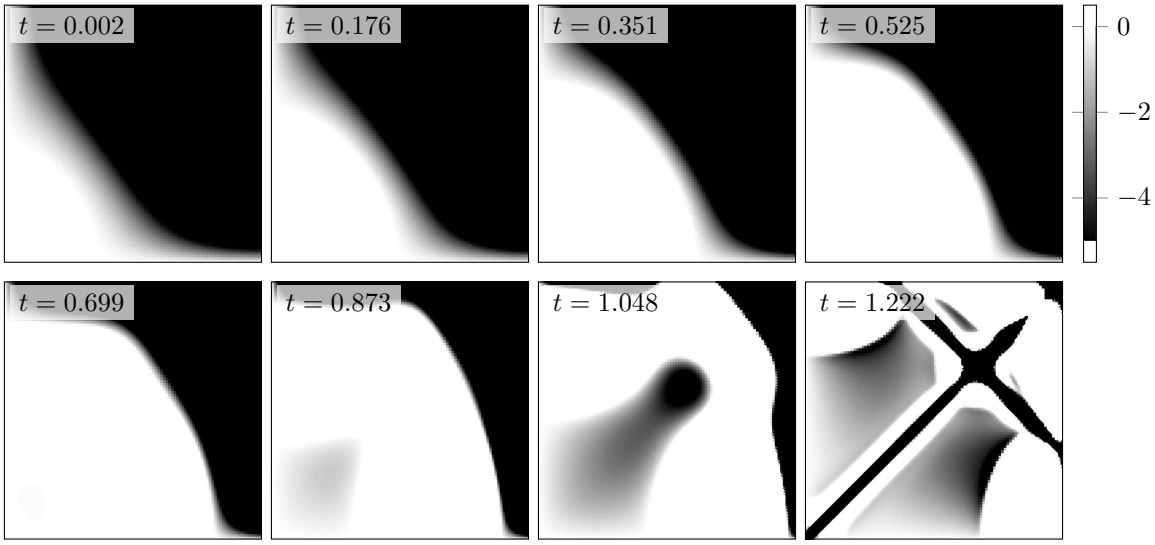

Figure 5.7.: Snapshots of optimal control for Example 5.4.3. Black and white denote the lower and the upper control bound, respectively.

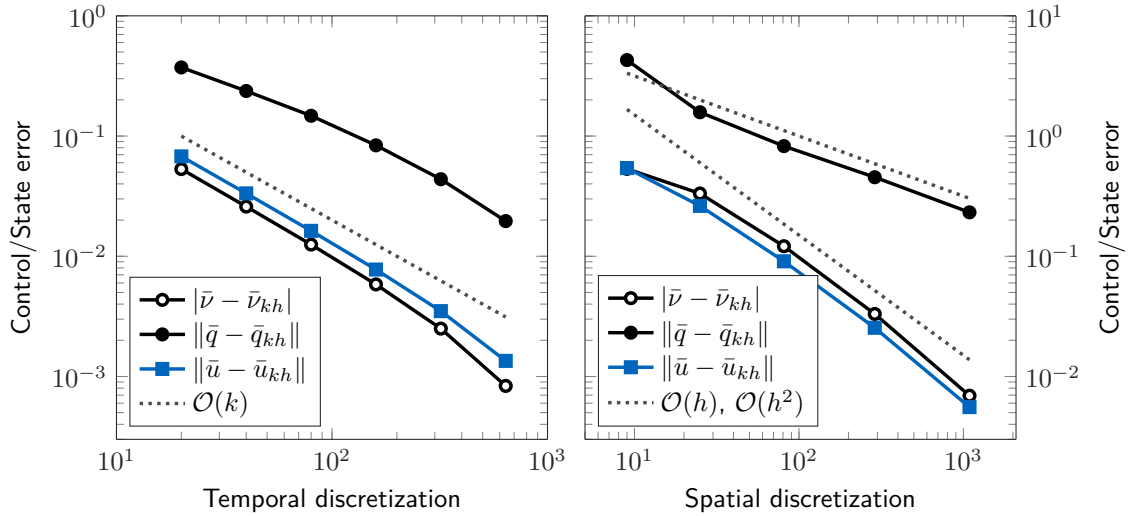

Figure 5.8.: Discretization error for Example 5.4.3 with cellwise constant control discretization and refinement of the time interval for $N = 1089$ nodes (left) and refinement of the spatial discretization for $M = 320$ time steps (right).

M	N	$\alpha = 1$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.001$
20	4225	20.50	5.795	2.444	−11.85
40	4225	18.15	5.028	1.638	−14.35
80	4225	16.99	4.674	1.251	−15.71
160	4225	16.40	4.503	1.055	−16.46
320	25	2.857	2.432	7.985_{-1}	−17.15
320	81	10.40	2.588	-4.119_{-1}	−19.84
320	289	14.40	3.764	4.669_{-1}	−17.53
320	1089	15.73	4.273	8.503_{-1}	−16.95

Table 5.3.: The quantity (3.27) of Lemma 3.18 as in Table 5.2 is shown, but here we minimize the functional in (3.30) over the whole space $L^2(I \times \omega)$ instead of the subspace $C_{\bar{q}}$. We observe that this second order sufficient criterion is not satisfied for small values of α. Therefore, a strong second order condition is not fulfilled in this example and for this reason it is essential to work with the critical cone $C_{(\bar{\nu},\bar{q})}$ in the formulation of the second order conditions.

5.5. Robust error estimates for bang-bang controls ($\alpha = 0$)

This section is devoted to discretization error estimates in case of bang-bang controls based on the structural assumption of the adjoint state (3.37). We first would like to motivate the "correct" choice of the norm. If the control is discretized by piecewise constant functions in time and space, then in case of bang-bang controls we cannot expect linear order of convergence in L^2. Precisely, we expect

$$\|\bar{q} - \bar{q}_{kh}\|_{L^2(I\times\omega)} \leq c(k+h)^{1/2}.$$

In contrast, changing the norm to L^1, we can get

$$\|\bar{q} - \bar{q}_{kh}\|_{L^1(I\times\omega)} \leq c(k+h).$$

Therefore, it seems to be reasonable to work with the L^1-norm instead of the L^2-norm as in the case with strictly positive regularization parameter α.

Combining the stability result Theorem 3.34 of the preceding chapter (with $\kappa = 1$) and the discretization error estimates of Section 5.3.3 directly implies the estimate

$$\begin{aligned}\|\bar{q} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)} &\leq \|\bar{q} - \bar{q}_{\alpha}\|_{L^1(I\times\omega)} + \|\bar{q}_{\alpha} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)} \\ &\leq c\alpha + c(\alpha)|\log k|(k+h^2),\end{aligned}$$

here for the variational control discretization, i.e. $\sigma(k,h) = 0$, for simplicity. However, the constant $c(\alpha)$ depends on α with $c(\alpha) \to \infty$ as $\alpha \to 0$. Therefore, the error due to regularization and the error due to discretization have to be balanced. Unfortunately, since the proof of the optimal order convergence result for the case $\alpha > 0$ relies on a contradiction argument, see Lemma 5.17, we cannot give the explicit dependence on α. Besides this, it would be desirable to have robust error estimates with respect to the regularization and the discretization, i.e. without any coupling between α, k, and h; cf. [152, 154].

In this section we suppose that the general assumptions from Sections 5.1 and 5.2 hold. Moreover, we assume that the projection Π_h is stable in H^1. To consider different control discretization schemes at the same time, we introduce the operator I_σ onto the (possibly discrete) control space $Q_\sigma(0,1) \subset L^2(I\times\omega)$ with an abstract parameter σ for the control discretization. In case of distributed control, we additionally assume that a subset denoted $\mathcal{T}_h^\omega$ of the mesh $\mathcal{T}_h$ is a non-overlapping cover of ω. As already mentioned, due to the bang-bang structure, we have to consider a different norm than L^2 in order to obtain optimal error estimates. We use the symbols $\sigma_1(k,h)$ and $\sigma_2(k,h)$ to denote the errors in $L^1(I\times\omega)$ and $L^2(I;H^{-1})$ due to control discretization. Concretely, we suppose

$$\|\bar{q} - \mathrm{I}_\sigma\bar{q}\|_{L^1(I\times\omega)} \leq \sigma_1(k,h)\|\bar{q}\|_{\sigma_1}, \tag{5.50}$$

$$\|B\,(\bar{q} - \mathrm{I}_\sigma\bar{q})\|_{L^2(I;H^{-1})} \leq \sigma_2(k,h)\|\bar{q}\|_{\sigma_2}, \tag{5.51}$$

where $\|\cdot\|_{\sigma_1}$ and $\|\cdot\|_{\sigma_2}$ stand for potentially different norms on $Q(0,1)$. We suppose that $\sigma_1(k,h) \to 0$ and $\sigma_2(k,h) \to 0$ as $k,h \to 0$ and $\mathrm{I}_\sigma Q_{ad}(0,1) \subset Q_{ad}(0,1)$. Moreover, we assume $\|\bar{q}\|_{\sigma_1} < \infty$ and $\|\bar{q}\|_{\sigma_2} < \infty$. For notational simplicity we write $\mathrm{I}_\sigma(\nu,q) = (\nu, \mathrm{I}_\sigma q)$ using the same symbol. Last, we define $Q_{ad,\sigma}(0,1) = Q_\sigma(0,1) \cap Q_{ad}(0,1)$. Concrete discretization strategies for the control will be discussed at the end of this section.

For any $\alpha \geq 0$ we define the discrete optimal control problem by

$$\inf_{\substack{\nu_{kh,\alpha}\in\mathbb{R}_+ \\ q_{kh,\alpha}\in Q_{ad,\sigma}(0,1)}} j_\alpha(\nu_{kh,\alpha}, q_{kh,\alpha}) \quad \text{subject to} \quad g_{kh}(\nu_{kh,\alpha}, q_{kh,\alpha}) \leq 0. \tag{$\hat{P}_{kh,\alpha}$}$$

As before, we apply a localization argument. Let $(\bar{\nu}, \bar{q}) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ be a local solution to $(\hat{P}_0)$. For $\rho > 0$ we introduce the problem

$$\inf_{\substack{\nu_{kh,\alpha}\in\mathbb{R}_+ \\ q_{kh,\alpha}\in Q_{ad,\sigma}(0,1)}} j_\alpha(\nu_{kh,\alpha}, q_{kh,\alpha}) \quad \text{subject to} \quad \begin{cases} g_{kh}(\nu_{kh,\alpha}, q_{kh,\alpha}) \leq 0, \\ \|(\nu_{kh,\alpha} - \bar{\nu}, q_{kh,\alpha} - \bar{q})\| \leq \rho, \end{cases} \tag{$\hat{P}^\rho_{kh,\alpha}$}$$

where we recall that the norm on the product space $\mathbb{R} \times Q(0,1)$ is given by

$$\|(\delta\nu, \delta q)\| = \left(|\delta\nu|^2 + \|\delta q\|^2_{L^2(I\times\omega)}\right)^{1/2}.$$

Similar as in Section 5.3.1 we construct two auxiliary sequences. First, we construct the sequence $\{(\nu_\gamma, q_\gamma)\}_{\gamma>0}$ converging to $(\bar{\nu}, \bar{q})$ as $\gamma \to 0$ that is feasible for the localized problem. In particular, this implies existence of solutions to $(\hat{P}^\rho_{kh,\alpha})$. Thereafter, we build a sequence $\{(\nu_\tau, q_\tau)\}_{\tau>0}$ converging to $(\bar{\nu}^\rho_{kh,\alpha}, \bar{q}^\rho_{kh,\alpha})$ as $\tau \to 0$ that is feasible for $(\hat{P}_\alpha)$. Since the solution operator to the state equation is continuous for right-hand sides from $L^2(I; H^{-1})$ into $W(0,1) \hookrightarrow C([0,1]; L^2)$, we may use (5.51) for all estimates concerning the state or the linearized state, whereas (5.50) is needed for the estimate for the controls in L^1. Note that all sequences constructed in Section 5.3.1 are independent of the regularization parameter α.

Proposition 5.23. *Let $(\bar{\nu}, \bar{q})$ be a locally optimal control of problem $(\hat{P}_0)$. There exists a sequence $\{(\nu_\gamma, q_\gamma)\}_{\gamma>0}$ of controls with $\gamma = \gamma(k,h)$ that are feasible for $(\hat{P}^\rho_{kh,\alpha})$ for k, h, ρ sufficiently small. Moreover,*

$$|\nu_\gamma - \bar{\nu}| + \|q_\gamma - \bar{q}\|_{L^1(I\times\omega)} \leq c\left(\sigma_1(k,h) + \sigma_2(k,h) + |\log k|(k + h^2)\right).$$

Proof. The sequence can be constructed as in Proposition 5.10 with slight modifications. In (5.21) we use σ_1 instead of σ. In (5.22) and (5.23) we replace σ by σ_2 that is allowed since the stability estimates for the discrete state equation also hold for $Bq \in L^2(I; H^{-1})$. □

In particular, Proposition 5.23 guarantees that for h, k, and ρ sufficiently small, the set of admissible controls of the discrete problem $(\hat{P}^\rho_{kh,\alpha})$ is nonempty. Hence, by standard arguments we obtain well-posedeness of the localized discrete problem; cf. Corollary 5.11.

Corollary 5.24. *Let h, k, and ρ be sufficiently small. Then there exists a solution $\bar{\chi}^\rho_{kh,\alpha} = (\bar{\nu}^\rho_{kh,\alpha}, \bar{q}^\rho_{kh,\alpha}) \in \mathbb{R}_+ \times Q_{ad,\sigma}(0,1)$ to $(\hat{P}^\rho_{kh,\alpha})$.*

Proposition 5.25. *Let $k, h, \rho > 0$ be sufficiently small. Moreover, let $(\bar{\nu}, \bar{q})$ be a locally optimal solution of $(\hat{P}_0)$ and let $(\bar{\nu}^\rho_{kh,\alpha}, \bar{q}^\rho_{kh,\alpha})$ be any globally optimal control of $(\hat{P}^\rho_{kh,\alpha})$. Then there exists a sequence $\{\nu_\tau\}_{\tau>0}$ with $\tau = \tau(k,h)$ such that $(\nu_\tau, \bar{q}^\rho_{kh,\alpha})$ is feasible for $(\hat{P}_0)$ and*

$$|\nu_\tau - \bar{\nu}^\rho_{kh,\alpha}| \leq c|\log k|(k + h^2).$$

Proof. This result can be proved as in Proposition 5.13. □

5.5.1. General regularization and discretization error estimates

First, we establish a robust error estimate with respect to regularization and discretization with a suboptimal rate concerning the control variable. Please note that Lemma 5.26 also holds in the case $\alpha = 0$ yielding an error estimate for the problems without regularization; cf. [47] for a linear-quadratic elliptic problem. Moreover, we emphasize that the convergence rate for the terminal time is independent of the value of κ from (3.37).

Lemma 5.26. *Let $(\bar\nu, \bar q)$ be a local solution to $(\hat P_0)$ satisfying the growth condition (3.41). Moreover, let $\{(k,h,\alpha)\}$ be a sequence of positive mesh sizes and regularization parameters converging to zero. Then there exists a sequence $\{(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})\}_{k,h,\alpha}$ of local solutions to $(\hat P_{kh,\alpha})$ converging to $(\bar\nu, \bar q)$ such that*

$$|\bar\nu - \bar\nu_{kh,\alpha}| + \|\bar q - \bar q_{kh,\alpha}\|^{1+1/\kappa}_{L^1(I\times\omega)} \le c\left(\alpha + \sigma_1(k,h) + \sigma_2(k,h) + |\log k|(k+h^2)\right), \tag{5.52}$$

where $c > 0$ is independent of k, h, α, $\bar\nu_{kh,\alpha}$, and $\bar q_{kh,\alpha}$. Moreover, there exists a Lagrange multiplier $\bar\mu_{kh,\alpha}$ such that the following optimality system is satisfied:

$$\bar\mu_{kh,\alpha} > 0, \tag{5.53}$$

$$\int_0^1 1 + \frac{\alpha}{2}\|\bar q_{kh,\alpha}\|^2_{L^2(\omega)} + \langle B\bar q_{kh,\alpha} + \Delta_h \bar u_{kh,\alpha}, \bar z_{kh,\alpha}\rangle \,\mathrm{d}t = 0, \tag{5.54}$$

$$\int_0^1 \langle \alpha \bar q_{kh,\alpha} + B^*\bar z_{kh,\alpha}, q - \bar q_{kh,\alpha}\rangle \,\mathrm{d}t \ge 0, \quad q \in Q_{ad,\sigma}(0,1), \tag{5.55}$$

$$G(\bar u_{kh,\alpha}(1)) = 0, \tag{5.56}$$

where $\bar u_{kh,\alpha} = S_{kh}(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})$ and $\bar z_{kh,\alpha} \in X_{k,h}$ is the solution to the discrete adjoint equation

$$\mathrm{B}(\bar\nu_{kh,\alpha}, \varphi_{kh}, \bar z_{kh,\alpha}) = \bar\mu_{kh,\alpha}(\bar u_{kh,\alpha}(1) - u_d, \varphi_{kh}(1)), \quad \varphi_{kh} \in X_{k,h}.$$

Proof. Let $\rho > 0$ be sufficiently small such that the quadratic growth condition (3.41) as well as Propositions 5.23 and 5.25 hold. Moreover, let $\{(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha})\}$ be a sequence of globally optimal solutions to $(\hat P^\rho_{kh,\alpha})$ that is guaranteed due to Corollary 5.24. Because the pair $(\nu_\tau, \bar q^\rho_{kh,\alpha})$ is feasible for $(\hat P_0)$, we may use the growth condition (3.41) to estimate

$$\begin{aligned} c\|\bar q - \bar q^\rho_{kh,\alpha}\|^{1+1/\kappa}_{L^1(I\times\omega)} &\le \nu_\tau - \bar\nu \le j_\alpha(\nu_\tau, \bar q^\rho_{kh,\alpha}) - j_\alpha(\bar\nu, \bar q) + \bar\nu\frac{\alpha}{2}\|\bar q\|^2_{L^2(I\times\omega)} \\ &\le j_\alpha(\nu_\tau, \bar q^\rho_{kh,\alpha}) - j_\alpha(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha}) + j_\alpha(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha}) - j_\alpha(\nu_\gamma, q_\gamma) \\ &\quad + j_\alpha(\nu_\gamma, q_\gamma) - j_\alpha(\bar\nu, \bar q) + c\alpha \\ &\le j_\alpha(\nu_\tau, \bar q^\rho_{kh,\alpha}) - j_\alpha(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha}) + j_\alpha(\nu_\gamma, q_\gamma) - j_\alpha(\bar\nu, \bar q) + c\alpha, \end{aligned} \tag{5.57}$$

where the last inequality follows from optimality of the pair $(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha})$ for $(\hat P^\rho_{kh,\alpha})$ and feasibility of (ν_γ, q_γ) for $(\hat P^\rho_{kh,\alpha})$. Then, we observe that

$$\begin{aligned} j_\alpha(\nu_\tau, \bar q^\rho_{kh,\alpha}) - j_\alpha(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha}) &= (\nu_\tau - \bar\nu^\rho_{kh,\alpha})\left(1 + \frac{\alpha}{2}\|\bar q^\rho_{kh,\alpha}\|^2_{L^2(I\times\omega)}\right) \\ &\le c\left(1 + \frac{\alpha}{2}\right)|\log k|(k+h^2) \end{aligned}$$

due to Proposition 5.25. Similarly,

$$\begin{aligned} j_\alpha(\nu_\gamma, q_\gamma) - j_\alpha(\bar\nu, \bar q) &\le (\nu_\gamma - \bar\nu)\left(1 + \frac{\alpha}{2}\|q_\gamma\|^2_{L^2(I\times\omega)}\right) \\ &\quad + \bar\nu\frac{\alpha}{2}\|q_\gamma + \bar q\|_{L^2(I\times\omega)}\|q_\gamma - \bar q\|_{L^2(I\times\omega)} \\ &\le c\left(\sigma_1(k,h) + \sigma_2(k,h) + \alpha + |\log k|(k+h^2)\right) \end{aligned}$$

employing Proposition 5.23 and boundedness of $Q_{ad}(0,1)$. Collecting all estimates above we arrive at

$$\|\bar q - \bar q^\rho_{kh,\alpha}\|^{1+1/\kappa}_{L^1(I\times\omega)} \le c\left(\alpha + \sigma_1(k,h) + \sigma_2(k,h) + |\log k|(k+h^2)\right).$$

Moreover, from Proposition 5.25 and (5.57) we further deduce

$$|\bar\nu^\rho_{kh,\alpha} - \bar\nu| \le |\bar\nu^\rho_{kh,\alpha} - \nu_\tau| + \nu_\tau - \bar\nu \le c\left(\alpha + \sigma_1(k,h) + \sigma_2(k,h) + |\log k|(k+h^2)\right).$$

In summary, the two preceding estimates establish the stated error estimate for the localized solutions. Furthermore, Hölder's inequality and uniform boundedness of $\bar q^\rho_{kh} \in Q_{ad}(0,1)$ in $L^\infty(I\times\omega)$ imply

$$\|\bar q - \bar q^\rho_{kh,\alpha}\|_{L^2(I\times\omega)} \le \|\bar q - \bar q^\rho_{kh,\alpha}\|^{1/2}_{L^1(I\times\omega)}\|\bar q - \bar q^\rho_{kh,\alpha}\|^{1/2}_{L^\infty(I\times\omega)} \le c\|\bar q - \bar q^\rho_{kh,\alpha}\|^{1/2}_{L^1(I\times\omega)}.$$

In particular, for $k, h, \alpha > 0$ sufficiently small the solution $(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha})$ does not lie on the boundary of the localization. Therefore, $(\bar\nu^\rho_{kh,\alpha}, \bar q^\rho_{kh,\alpha})$ is a local solution to $(\hat P_{kh,\alpha})$ and we can drop the super index ρ. Finally, the convergence result and the fact that $g'(\bar\nu, \bar q) \ne 0$ yields the optimality conditions in qualified form as stated above. □

Proposition 5.27. *Adopt the assumptions of Lemma 5.26. The Lagrange multipliers $\bar\mu_{kh,\alpha}$ satisfy $\bar\mu_{kh,\alpha} \to \bar\mu$ as $k, h, \alpha \to 0$.*

Proof. This follows as in Proposition 5.16 using the convergence result of Lemma 5.26. □

While the estimate of Lemma 5.26 is optimal for the terminal time in the case of a variational control discretization, it is suboptimal with respect to the control variable. Under certain conditions we will eventually provide an improved estimate that is based on the following result.

Proposition 5.28. *Adopt the assumptions of Lemma 5.26 and let* (3.37) *hold. Moreover, we assume that I_σ is an orthogonal projection onto $Q_\sigma(0,1)$ in $L^2(I\times\omega)$. In case of a distributed control, suppose in addition that $u_0 \in (L^p, \mathcal{D}_{L^p}(-\Delta))_{1-1/s,s}$ for $s, p \in (1,\infty)$ such that $d/(2p) + 1/s < 1$. There is a constant $c > 0$ independent of k, h, α, $\bar\nu_{kh,\alpha}$, and $\bar q_{kh,\alpha}$ such that*

$$\begin{aligned} \|\bar q - \bar q_{kh,\alpha}\|^{1/\kappa}_{L^1(I\times\omega)} \le c\Big(&\alpha + |\bar\nu - \bar\nu_{kh,\alpha}| + \|B^*\bar z - \mathrm{I}_\sigma B^*\bar z\|_{L^\infty(I\times\omega)} \\ &+ \|B^*\left(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)}\Big), \end{aligned}$$

where $z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}) \in W(0,1)$ denotes the solution to the adjoint equation with time transformation $\bar\nu_{kh,\alpha}$ and terminal value $\bar\mu_{kh,\alpha}(i_1 S(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}) - u_d)$.

For the proof of Proposition 5.28 we require the following Lipschitz estimate of the solution to the state equation with respect to the time transformation.

Proposition 5.29. *Let $\nu_{\max} > \nu_{\min} > 0$. There is $c > 0$ such that for any $u_0 \in L^2$, $f \in L^2(I; H^{-1})$, and $\nu_1, \nu_2 \in [\nu_{\min}, \nu_{\max}]$ the solutions to the state equation $u(\nu_1) = u(\nu_1, u_0, f)$ and $u(\nu_2) = u(\nu_2, u_0, f)$ satisfy the estimate*

$$\|u(\nu_1) - u(\nu_2)\|_{C([0,1];L^2)} \leq c|\nu_1 - \nu_2| \left(\|f\|_{L^2(I;H^{-1})} + \|u_0\|_{L^2} \right),$$

where $c > 0$ is independent of ν, f, and u_0.

Proof. Set $u_1 = u(\nu_1)$ and $u_2 = u(\nu_2)$. Then the difference $w = u_1 - u_2$ satisfies

$$\partial_t w - \nu_1 \Delta w = (\nu_1 - \nu_2)(\Delta u_2 + f), \quad w(0) = 0.$$

Hence, standard energy estimates lead to

$$\begin{aligned} \|w\|_{H^1(I;H^{-1}) \cap L^2(I;H^1)} &\leq c|\nu_1 - \nu_2| \| -\Delta u_2 + f\|_{L^2(I;H^{-1})} \\ &\leq c|\nu_1 - \nu_2| \left(\|f\|_{L^2(I;H^{-1})} + \|u_0\|_{L^2} \right). \end{aligned}$$

Last, the assertion follows from the embedding $H^1(I; H^{-1}) \cap L^2(I; H^1) \hookrightarrow C([0,1]; L^2)$. □

Proposition 5.30. *Let $\nu_{\max} > \nu_{\min} > 0$ and $s, p \in (1, \infty)$ such that $d/(2p) + 1/s < 1$. There is $c > 0$ such that for any $u_0 \in (L^p, \mathcal{D}_{L^p}(-\Delta))_{1-1/s,s}$, $f \in L^s(I; L^p)$, and $\nu_1, \nu_2 \in [\nu_{\min}, \nu_{\max}]$ the solutions to the state equation $u(\nu_1) = u(\nu_1, u_0, f)$ and $u(\nu_2) = u(\nu_2, u_0, f)$ satisfy the estimate*

$$\|u(\nu_1) - u(\nu_2)\|_{L^\infty(I \times \Omega)} \leq c|\nu_1 - \nu_2| \left(\|f\|_{L^s(I;L^p)} + \|u_0\|_{(L^p, \mathcal{D}_{L^p}(-\Delta))_{1-1/s,s}} \right),$$

where $c > 0$ is independent of ν, f, and u_0.

Proof. Maximal parabolic regularity of $-\Delta$ on L^p, see, e.g., [49, Theorem 2.9 b)], yields that the solution $u = u(\nu, f, u_0)$ satisfies the estimate

$$\|u\|_{W^{1,s}(I;L^p) \cap L^s(I;\mathcal{D}_{L^p}(-\Delta))} \leq c \left(\|f\|_{L^s(I;L^p)} + \|u_0\|_{(L^p, \mathcal{D}_{L^p}(-\Delta))_{1-1/s,s}} \right).$$

Moreover, continuity of $\nu \mapsto (\partial_t - \nu\Delta)^{-1}$, $\nu > 0$, as well as compactness of $[\nu_{\min}, \nu_{\max}]$ imply that the constant in the estimate above can be chosen uniformly with respect to ν. Set $u_1 = u(\nu_1)$ and $u_2 = u(\nu_2)$. Then the difference $w = u_1 - u_2$ satisfies

$$\partial_t w - \nu_1 \Delta w = (\nu_1 - \nu_2)(\Delta u_2 + f), \quad w(0) = 0.$$

Hence,

$$\begin{aligned} \|w\|_{W^{1,s}(I;L^p) \cap L^s(I;\mathcal{D}_{L^p}(-\Delta))} &\leq c|\nu_1 - \nu_2| \| -\Delta u_2 + f\|_{L^s(I;L^p)} \\ &\leq c|\nu_1 - \nu_2| \left(\|f\|_{L^s(I;L^p)} + \|u_0\|_{(L^p, \mathcal{D}_{L^p}(-\Delta))_{1-1/s,s}} \right). \end{aligned}$$

Finally, the assertion follows from the embedding

$$W^{1,s}(I; L^p) \cap L^s(I; \mathcal{D}_{L^p}(-\Delta)) \hookrightarrow C(\overline{I \times \Omega});$$

see the proof of [49, Theorem 3.1]. □

Proof of Proposition 5.28. We use ideas from the proof of [152, Theorem 31]. Setting $q = \bar{q}_{kh,\alpha}$ in (3.38) and multiplication by $\bar{\mu}_{kh,\alpha}/\bar{\mu} > 0$ yield

$$c\|\bar{q} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}^{1+1/\kappa} \leq -\int_0^1 (B^* z(\bar{\nu},\bar{q}), \bar{q} - \bar{q}_{kh,\alpha})_{L^2(\omega)}, \tag{5.58}$$

where $z(\bar{\nu},\bar{q})$ is the solution to the adjoint equation with time transformation $\bar{\nu}$ and terminal value $\bar{\mu}_{kh,\alpha}(i_1 S(\bar{\nu},\bar{q}) - u_d)$. Note that we have used Proposition 5.27 in order to guarantee that the constant c is independent of k, h, and α. The optimality condition (5.55) of Lemma 5.26 for $\bar{q}_{kh,\alpha}$ with $q = \mathrm{I}_\sigma \bar{q}$ can be written as

$$\alpha\|\mathrm{I}_\sigma\bar{q} - \bar{q}_{kh,\alpha}\|_{L^2(I\times\omega)}^2 \leq \int_0^1 (\alpha \mathrm{I}_\sigma\bar{q} + B^*\bar{z}_{kh,\alpha}, \mathrm{I}_\sigma\bar{q} - \bar{q}_{kh,\alpha})_{L^2(\omega)}. \tag{5.59}$$

Summation of (5.58) and (5.59) implies

$$\begin{aligned} c\|\bar{q} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}^{1+1/\kappa} + \alpha\|\mathrm{I}_\sigma\bar{q} - \bar{q}_{kh,\alpha}\|_{L^2(I\times\omega)}^2 \\ \leq \alpha\int_0^1 (\mathrm{I}_\sigma\bar{q}, \mathrm{I}_\sigma\bar{q} - \bar{q}_{kh,\alpha})_{L^2(\omega)} + \int_0^1 \left(B^*\left(\bar{z}_{kh,\alpha} - z(\bar{\nu},\bar{q})\right), \mathrm{I}_\sigma\bar{q} - \bar{q}_{kh,\alpha}\right)_{L^2(\omega)} \\ + \int_0^1 (B^* z(\bar{\nu},\bar{q}), \mathrm{I}_\sigma\bar{q} - \bar{q})_{L^2(\omega)}. \end{aligned} \tag{5.60}$$

We first consider the last term of the right-hand side of (5.60). Since I_σ is the $L^2(I\times\omega)$-projection onto $Q_\sigma(0,1)$, we have

$$\begin{aligned} \int_0^1 (B^* z(\bar{\nu},\bar{q}), \mathrm{I}_\sigma\bar{q} - \bar{q})_{L^2(\omega)} &= \int_0^1 (B^* z(\bar{\nu},\bar{q}) - \mathrm{I}_\sigma B^* z(\bar{\nu},\bar{q}), \bar{q}_{kh,\alpha} - \bar{q})_{L^2(\omega)} \\ &= \frac{\bar{\mu}_{kh,\alpha}}{\bar{\mu}} \int_0^1 (B^*\bar{z} - \mathrm{I}_\sigma B^*\bar{z}, \bar{q}_{kh,\alpha} - \bar{q})_{L^2(\omega)} \\ &\leq c\|B^*\bar{z} - \mathrm{I}_\sigma B^*\bar{z}\|_{L^\infty(I\times\omega)}\|\bar{q} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}. \end{aligned} \tag{5.61}$$

In the last step we have used that the multipliers $\bar{\mu}_{kh,\alpha}$ are uniformly bounded; see Proposition 5.27. The first term of (5.60) can be easily estimated by

$$\alpha\int_0^1 (\mathrm{I}_\sigma\bar{q}, \mathrm{I}_\sigma\bar{q} - \bar{q}_{kh,\alpha})_{L^2(\omega)} = \alpha\int_0^1 (\mathrm{I}_\sigma^*\mathrm{I}_\sigma\bar{q}, \bar{q} - \bar{q}_{kh,\alpha})_{L^2(\omega)} \leq c\alpha\|\bar{q} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}$$

using that I_σ is an orthogonal projection as well as $\mathrm{I}_\sigma\bar{q} \in Q_{ad}(0,1) \subset L^\infty(I\times\omega)$. Concerning the second term of the right-hand side of (5.60), we have

$$\begin{aligned} \int_0^1 \left(B^*\left(\bar{z}_{kh,\alpha} - z(\bar{\nu},\bar{q})\right), \bar{q} - \bar{q}_{kh,\alpha}\right)_{L^2(\omega)} &= \int_0^1 \left(B^*\left(\bar{z}_{kh,\alpha} - z(\bar{\nu}_{kh,\alpha},\bar{q}_{kh,\alpha})\right), \bar{q} - \bar{q}_{kh,\alpha}\right)_{L^2(\omega)} \\ &+ \int_0^1 \left(B^*\left(z(\bar{\nu}_{kh,\alpha},\bar{q}_{kh,\alpha}) - z(\bar{\nu},\bar{q}_{kh,\alpha})\right), \bar{q} - \bar{q}_{kh,\alpha}\right)_{L^2(\omega)} \\ &+ \int_0^1 \left(B^*\left(z(\bar{\nu},\bar{q}_{kh,\alpha}) - z(\bar{\nu},\bar{q})\right), \bar{q} - \bar{q}_{kh,\alpha}\right)_{L^2(\omega)}. \end{aligned}$$

Note that all adjoint states appearing above correspond to the same multiplier $\bar{\mu}_{kh,\alpha}$, which is uniformly bounded with respect to α, k, and h due to Proposition 5.27. For the first term on the right-hand side, we apply Hölder's inequality and obtain

$$\begin{aligned} \int_0^1 \left(B^*\left(\bar{z}_{kh,\alpha} - z(\bar{\nu}_{kh,\alpha},\bar{q}_{kh,\alpha})\right), \bar{q} - \bar{q}_{kh,\alpha}\right)_{L^2(\omega)} \\ \leq \|B^*\left(\bar{z}_{kh,\alpha} - z(\bar{\nu}_{kh,\alpha},\bar{q}_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)}\|\bar{q} - \bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}. \end{aligned}$$

The second term can be estimated using Proposition 5.29 for purely time-dependent control and Proposition 5.30 for distributed control as

$$\int_0^1 \left(B^*\left(z(\bar\nu_{kh,\alpha},\bar q_{kh,\alpha}) - z(\bar\nu,\bar q_{kh,\alpha})\right), \bar q - \bar q_{kh,\alpha}\right)_{L^2(\omega)} \le c|\bar\nu_{kh,\alpha} - \bar\nu| \|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}.$$

The third term is less than or equal to zero. In summary, we arrive at

$$\begin{aligned}\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}^{1+1/\kappa} \le c\Big(\alpha + |\bar\nu_{kh,\alpha} - \bar\nu| + \|B^*\left(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha},\bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)} \\ + \|B^*\bar z - \mathrm{I}_\sigma B^*\bar z\|_{L^\infty(I\times\omega)}\Big)\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}.\end{aligned}$$

Last, dividing by $\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}$ yields the desired estimate. □

If the controls are explicitly discretized by cellwise constant functions and if $\kappa < 1$, the term $\|B^*\bar z - \mathrm{I}_\sigma B^*\bar z\|_{L^\infty(I\times\omega)}$ limits the overall convergence rate in Proposition 5.28. Alternatively, in (5.61), we can estimate

$$\int_0^1 (B^* z(\bar\nu,\bar q), \mathrm{I}_\sigma\bar q - \bar q)_{L^2(\omega)} = \frac{\bar\mu_{kh,\alpha}}{\bar\mu}\int_0^1 (B^*\bar z, \mathrm{I}_\sigma\bar q - \bar q)_{L^2(\omega)} \le c|(B^*\bar z, \mathrm{I}_\sigma\bar q - \bar q)_{L^2(I\times\omega)}|.$$

Proceeding with the remaining terms as in the proof above, we in summary obtain

$$\begin{aligned}\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}^{1+1/\kappa} &\le c|(B^*\bar z, \mathrm{I}_\sigma\bar q - \bar q)_{L^2(I\times\omega)}| \\ &\quad + c\left(\alpha + |\bar\nu_{kh,\alpha} - \bar\nu| + \|B^*\left(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha},\bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)}\right)\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}.\end{aligned}$$

Furthermore, Young's inequality yields

$$\begin{aligned}\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}^{1+1/\kappa} &\le c|(B^*\bar z, \mathrm{I}_\sigma\bar q - \bar q)_{L^2(I\times\omega)}| \\ &\quad + c\left(\alpha + |\bar\nu_{kh,\alpha} - \bar\nu| + \|B^*\left(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha},\bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)}\right)^{1+\kappa}.\end{aligned}$$

Last, the fact that $(1+\kappa)/(1+1/\kappa) = \kappa$ implies the alternative estimate

$$\begin{aligned}\|\bar q - \bar q_{kh,\alpha}\|_{L^1(I\times\omega)}^{1/\kappa} \le c\left(\alpha + |\bar\nu - \bar\nu_{kh,\alpha}| + \|B^*\left(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha},\bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)}\right) \\ + c|(B^*\bar z, \mathrm{I}_\sigma\bar q - \bar q)_{L^2(I\times\omega)}|^{1/(1+\kappa)} \quad (5.62)\end{aligned}$$

under the same conditions as Proposition 5.28. If $\kappa < 1$, then (5.62) might lead to better estimates. However, also the convergence rate for ν in the theory so far is limited by σ_1 and σ_2. Therefore, we stay with the estimate in Proposition 5.28 and keep prospective improvements in mind. Note that in Theorem 5.57 we obtain the error estimate $k + h^{3/2}$ for ν under a different condition than the structural assumption. Indeed, in the numerical examples we always observe the full convergence rate $k + h^2$ for ν independent of the control discretization; see Section 5.7.

5.5.2. Purely time-dependent controls

In case of purely time-dependent controls we immediately derive an error estimate (that is optimal if $\kappa = 1$) using the $L^\infty(I; L^2)$ discretization error estimate for the variational control discretization. Note that besides theoretical advantages purely time-dependent controls are also interesting in practice as distributed controls are typically difficult to implement.

Theorem 5.31 (Parameter control, variational)**.** *Adopt the assumptions of Lemma 5.26 and let* (3.37) *hold. Additionally, suppose purely time-dependent controls with variational control discretization, i.e.* $Q_\sigma(0,1) = Q(0,1)$*. There is a constant* $c > 0$ *not depending on* k*,* h*,* α*,* $\bar\nu_{kh,\alpha}$*, and* $\bar q_{kh,\alpha}$ *such that*

$$|\bar\nu - \bar\nu_{kh,\alpha}| + \|\bar q - \bar q_{kh,\alpha}\|^{1/\kappa}_{L^1(I\times\omega)} \le c\left(\alpha + |\log k|(k+h^2)\right).$$

Proof. This follows from Lemma 5.26 and Proposition 5.28, since in case of purely time-dependent control we may use the $L^\infty(I;L^2)$ discretization error estimate, see Lemma A.39, for the state and adjoint state equation to obtain

$$\|B^*\left(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)} \le c|\log k|(k+h^2).$$

In addition, $\mathrm{I}_\sigma = \mathrm{Id}$, $\sigma_1(k,h) = 0$, and $\sigma_2(k,h) = 0$, as we do not explicitly discretize the control variable. The remaining estimate for $\bar\nu$ is proved in Lemma 5.26. □

If $\alpha > 0$, by virtue of the projection formula

$$\bar q_{kh,\alpha} = P_{Q_{ad}}\left(-\frac{1}{\alpha}B^*\bar z_{kh,\alpha}\right), \tag{5.63}$$

which can be deduced from (5.55) with $Q_{ad,\sigma}(0,1) = Q_{ad}(0,1)$, the optimal control $\bar q_{kh,\alpha}$ obtained by the variational approach is piecewise constant in time with values in $\mathbb{R}^{N_c}$. Hence, in the case $\alpha > 0$, the variational control discretization is equivalent to the piecewise constant control discretization. However, in the case $\alpha = 0$, the estimate of Theorem 5.31 is still valid, but the discrete optimal control $\bar q_{kh,0}$ is not necessarily piecewise constant with the same time mesh as the state and adjoint state. Nevertheless, the optimality conditions for $\bar q_{kh,0}$ imply

$$\begin{aligned} B^*\bar z_{kh,0}|_{I_m} > 0 &\Rightarrow \bar q_{kh,0}|_{I_m} = q_a,\\ B^*\bar z_{kh,0}|_{I_m} < 0 &\Rightarrow \bar q_{kh,0}|_{I_m} = q_b, \end{aligned}$$

for all $m = 1, 2, \dots, M$, where the conditions are to be understood componentwise. Let Π_k denote the projection onto the piecewise constant functions in time, i.e.

$$(\Pi_k v)(t) = \frac{1}{k_m}\int_{I_m} v(\xi)\,\mathrm{d}\xi, \quad t \in I_m,$$

for every $v \in L^2(I;L^2)$ and $m \in \{1, 2, \dots, M\}$. Clearly, if $B^*\bar z_{kh,0}|_{I_m} > 0$, then $\Pi_k\bar q_{kh,0}|_{I_m} = q_a$, and if $B^*\bar z_{kh,0}|_{I_m} < 0$, then $\Pi_k\bar q_{kh,0}|_{I_m} = q_b$. For this reason, we are only interested in those time intervals I_m, where at least one component of $B^*\bar z_{kh,0}$ is identical zero. We define

$$\mathcal{S}_k = \{\, m = 1, 2, \dots, M \colon (B^*\bar z_{kh,0})(t_m, x) = 0, x \in \omega \,\},$$

and suppose that there exists $c > 0$ independent of k and h such that

$$\sum_{m\in\mathcal{S}_k} k_m \le ck^\kappa, \quad k > 0. \tag{5.64}$$

Note that a similar assumption has been used to prove optimal error estimates in Theorem 5.21 for cellwise linear control discretization. Employing the estimate of Theorem 5.31 we obtain

$$\begin{aligned} \|\bar q - \Pi_k\bar q_{kh,0}\|_{L^1(I\times\omega)} &\le \|\bar q - \bar q_{kh,0}\|_{L^1(I\times\omega)} + \|\bar q_{kh,0} - \Pi_k\bar q_{kh,0}\|_{L^1(I\times\omega)}\\ &\le c\left(|\log k|(k+h^2)\right)^\kappa + c\sum_{m\in\mathcal{S}_k} k_m \le c\left(|\log k|(k+h^2)\right)^\kappa. \end{aligned}$$

Furthermore, since Π_k is a projection, we have

$$(B\bar{q}_{kh,0}, \varphi_{kh})_{L^2(I;L^2)} = (B\Pi_k\bar{q}_{kh,0}, \varphi_{kh})_{L^2(I;L^2)} \quad \text{for all } \varphi_{kh} \in X_{k,h},$$

and the controls $\bar{q}_{kh,0}$ and $\Pi_k\bar{q}_{kh,0}$ have the same associated discrete state. In addition, the objective functional does not change, because of $\alpha = 0$. Therefore, the pair $(\bar{\nu}_{kh,0}, \Pi_k\bar{q}_{kh,0})$ is also optimal for $(\hat{P}_{kh,\alpha})$ with $\alpha = 0$. Based on this observation, we have the following corollary; cf. also Corollary 5.19.

Corollary 5.32 (Parameter control, discrete)**.** *Adopt the assumptions of Lemma 5.26 and let the assumption* (3.37) *hold. Moreover, suppose that ω is discrete, and choose the piecewise constant discrete control space*

$$Q_\sigma(0,1) = \left\{ v \in Q(0,1) \colon v|_{I_m} \in \mathcal{P}_0(I_m; \mathbb{R}^{N_c}),\ m = 1, 2, \dots, M \right\}.$$

If $\alpha = 0$ assume in addition that (5.64) *holds. Then there is a constant $c > 0$ not depending on k, h, $\bar{\nu}_{kh}$, and $\bar{q}_{kh}$ such that*

$$|\bar{\nu} - \bar{\nu}_{kh,\alpha}| + \|\bar{q} - \bar{q}_{kh,\alpha}\|^{1/\kappa}_{L^1(I;\mathbb{R}^{N_c})} \le c|\log k|(k + h^2).$$

5.5.3. Interlude: Interior pointwise error estimates

In order to apply Proposition 5.28 in case of a distributed control, we require pointwise error estimates for the solutions to the state and adjoint state equation. For simplicity, we consider the case of smooth initial data only. In the sequel we will prove the following interior pointwise error estimate that was obtained jointly with Dominik Hafemeyer. We generally assume that h is sufficiently small, precisely we suppose that $h < \mathrm{e}^{-4}$. Moreover, we assume that the family of triangulations is quasi-uniform; see Definition A.31.

Lemma 5.33. *Let $\nu \in [\nu_{\min}, \nu_{\max}]$ for fixed $0 < \nu_{\min} < \nu_{\max}$. Moreover, consider $\omega \subset \Omega$ open such that $\overline{\omega} \subset \Omega$. Given $f \in L^\infty(I \times \Omega)$ and $u_0 \in \mathcal{D}_{L^\infty}(-\Delta)$, let u be the solution to the state equation with right-hand side f, time transformation ν, initial value u_0, and u_{kh} its discrete counterpart. Then the estimate*

$$\|u - u_{kh}\|_{L^\infty(I\times\omega)} \le c|\log k|^2|\log h|^5(k + h^2)\left(\|f\|_{L^\infty(I\times\Omega)} + \|u_0\|_{\mathcal{D}_{L^\infty}(-\Delta)}\right)$$

holds, where the constant $c > 0$ is independent of k, h, ν, f, u_0, u, and u_{kh}.

For the proof of Lemma 5.33 we require several auxiliary results. We will frequently use the following embeddings for spaces of maximal parabolic regularity; see Proposition A.8. Let X and Y be Banach spaces such that $Y \hookrightarrow_d X$ and $s \in (1, \infty)$. Then

$$W^{1,s}(I;X) \cap L^s(I;Y) \hookrightarrow C([0,T]; (X,Y)_{1-1/s,s}). \tag{5.65}$$

If $\tau \in (0, 1 - \frac{1}{s})$, then

$$W^{1,s}(I;X) \cap L^s(I;Y) \hookrightarrow C^\alpha(I; (X,Y)_{\tau,1}), \quad 0 \le \alpha < 1 - \frac{1}{s} - \tau. \tag{5.66}$$

Furthermore, the constants for both embeddings can be chosen uniformly for all $s \in [2, \infty)$ and $\tau \in (0, 1)$.

Using the error estimates for the Lagrange interpolant I_h from Proposition A.32, we establish the following $L^\infty(I; L^2(\Omega))$ error estimates. Recall that $i_k \colon C([0,1]; V_h) \to X_{k,h}$ denotes the nodal interpolation defined by

$$i_k u(t_m) = u(t_m), \quad m = 1, 2, \dots, M.$$

Proposition 5.34. *Let $\nu \in [\nu_{\min}, \nu_{\max}]$ with $0 < \nu_{\min} < \nu_{\max}$. Given $f \in L^\infty(I; L^2(\Omega))$ and $u_0 \in \mathcal{D}_{L^2}(-\Delta)$, let u be the solution to the state equation with right-hand side f, time transformation ν, and initial value u_0. Then the estimate*

$$\begin{aligned}\|u - i_k \mathrm{I}_h u\|_{L^\infty(I;L^2(\Omega))} + h\|\nabla(u - i_k \mathrm{I}_h u)\|_{L^\infty(I;L^2(\Omega))} \\ \leq c|\log k||\log h|(k+h^2)\left(\|f\|_{L^\infty(I;L^2(\Omega))} + \|u_0\|_{\mathcal{D}_{L^2}(-\Delta))}\right)\end{aligned}$$

holds, where the constant $c > 0$ is independent of k, h, ν, f, u_0, and u.

Proof. First, we have the standard embedding

$$W^{1,r}(I; L^2(\Omega)) \hookrightarrow C^{1-1/r}(I; L^2(\Omega)), \quad r \in (1, \infty),$$

(with embedding constant one) that easily follows from

$$u(t_2) - u(t_1) = \int_{t_1}^{t_2} \partial_t u(\tau)\,\mathrm{d}\tau,$$

see, e.g., [4, Section III.1.2], and Hölder's inequality. Hence

$$\|u - i_k u\|_{L^\infty(I;L^2(\Omega))} \leq c k^{1-1/r} \|u\|_{W^{1,r}(I;L^2(\Omega))}$$

due to the definition of i_k. The norm of the right-hand side depends on r. Precisely, we have

$$\|u\|_{W^{1,r}(I;L^2(\Omega))\cap L^r(I;\mathcal{D}_{L^2}(-\Delta))} \leq c\frac{r^2}{r-1}\left(\|f\|_{L^r(I;L^2(\Omega))} + \|u_0\|_{(L^2(\Omega),\mathcal{D}_{L^2}(-\Delta))_{1-1/r,r}}\right);$$

see, e.g., [7, Theorem 1.3.2]. Thus, using the fact $\mathcal{D}_{L^2}(-\Delta) \hookrightarrow (L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{1-1/r,r}$ with uniform embedding constant for $r \geq r_0 > 1$ for some $r_0 > 1$ (see Proposition A.1 and Remark A.2) and taking $r = |\log k|$, we obtain

$$\|u - i_k u\|_{L^\infty(I;L^2(\Omega))} \leq c|\log k| k \left(\|f\|_{L^\infty(I;L^2(\Omega))} + \|u_0\|_{\mathcal{D}_{L^2}(-\Delta)}\right).$$

Note that we have used $k < \mathrm{e}^{-1}$ and $|\log k| > 1$ which holds since $k \leq 1/4 < \mathrm{e}^{-1}$. Similarly, according to the embedding (5.66) we have

$$W^{1,r}(I; L^2(\Omega)) \cap L^r(I; \mathcal{D}_{L^2}(-\Delta)) \hookrightarrow C^\alpha(I; (L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{\tau,1}), \quad \tau \in (0, 1 - 1/r),$$

for $0 \leq \alpha < 1 - 1/r - \tau$. Moreover,

$$(L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{\tau,1} \hookrightarrow \mathcal{D}_{L^2}((-\Delta)^\tau) \hookrightarrow \mathcal{D}_{L^2}((-\Delta)^{1/2}) = H_0^1(\Omega),$$

for $\tau > 1/2$; see Propositions A.12 and A.13. The embedding constant of the first injection is well-behaved by Remark A.14. Moreover, the embedding constant for the second injection is bounded by $\max\{1, \|(-\Delta)^{1/2-\tau}\|_{\mathcal{L}(L^2)}\}$ according to Proposition A.12. Since $-\Delta$ has bounded imaginary powers (see [110, Theorem 4.3.5]), the mapping $z \mapsto (-\Delta)^z$ is continuous on the half plane $\operatorname{Re} z \leq 0$; see [110, Lemma 4.2.5]. Hence, the second embedding constant is uniformly bounded if $\tau \to 1/2$. Thus, taking $\alpha = 1 - 2/r - \tau$ yields

$$\begin{aligned}\|\nabla u - i_k \nabla u\|_{L^\infty(I;L^2(\Omega))} &\leq c k^{1-2/r-\tau} \|\nabla u\|_{C^{1-2/r-\tau}(I;L^2(\Omega))} \\ &\leq c k^{1-2/r-\tau} \|u\|_{W^{1,r}(I;L^2(\Omega))\cap L^r(I;\mathcal{D}_{L^2}(-\Delta))}.\end{aligned}$$

Hence, with $r = |\log k|$ as before and letting $\tau \to 1/2$ we arrive at

$$\|\nabla u - i_k \nabla u\|_{L^\infty(I;L^2(\Omega))} \le c|\log k| k^{1/2} \left(\|f\|_{L^\infty(I;L^2(\Omega))} + \|u_0\|_{\mathcal{D}_{L^2}(-\Delta)} \right).$$

Next, we consider the error due to spatial discretization. First, according to Proposition A.5 we have

$$(L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{1-1/r,r} \hookrightarrow (L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{1-2/r,2}. \tag{5.67}$$

By Remark A.6 the embedding constant is uniformly bounded. Moreover, since Ω is convex, the characterization $\mathcal{D}_{L^2}(-\Delta) = H_0^1(\Omega) \cap H^2(\Omega)$ with equivalence of norms holds; see, e.g., [68, Theorem 3.2.1.2]. Thus, the definition of the interpolation space implies

$$(L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{1-2/r,2} \hookrightarrow (L^2(\Omega), H^2(\Omega))_{1-2/r,2}, \tag{5.68}$$

where the embedding constant is given by the embedding constant of $\mathcal{D}_{L^2}(-\Delta)$ into $H^2(\Omega)$. Last, we employ Proposition A.29 for $r > 4$

$$(L^2(\Omega), H^2(\Omega))_{1-2/r,2} \hookrightarrow W^{2-4/r,2}(\Omega), \tag{5.69}$$

where we have used the fact that Ω has a Lipschitz boundary, since it is convex; see [68, Corollary 1.2.2.3]. The embedding constant has the asymptotic behavior $\sim r$ for $r \to \infty$. Combining (5.67) – (5.69) we arrive at

$$(L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{1-1/r,r} \hookrightarrow W^{2-4/r,2}(\Omega),$$

with embedding constant $\sim r$ as $r \to \infty$. Using the embedding (5.65) that becomes in the particular case

$$W^{1,r}(I; L^2(\Omega)) \cap L^r(I; \mathcal{D}_{L^2}(-\Delta)) \hookrightarrow C([0,1]; (L^2(\Omega), \mathcal{D}_{L^2}(-\Delta))_{1-1/r,r}),$$

we find for $r > 4$ that

$$\begin{aligned} \|i_k (u - \mathrm{I}_h u)\|_{L^\infty(I;L^2(\Omega))} &\le \|u - \mathrm{I}_h u\|_{L^\infty(I;L^2(\Omega))} \\ &\le c r h^{2(1-2/r)} \|u\|_{C([0,1];W^{2(1-2/r),2}(\Omega))}, \end{aligned}$$

and

$$\begin{aligned} \|\nabla i_k (u - \mathrm{I}_h u)\|_{L^\infty(I;L^2(\Omega))} &\le \|\nabla (u - \mathrm{I}_h u)\|_{L^\infty(I;L^2(\Omega))} \\ &\le c r h^{2(1-2/r)-1} \|u\|_{C([0,1];W^{2(1-2/r),2}(\Omega))}, \end{aligned}$$

where we have used the error estimates of Proposition A.33. Now, we can argue as before (taking $r = |\log h|$) completing the proof. □

Proposition 5.35. *For all $p \in (1,\infty)$ and $\tau \in (0,1)$ such that $d/(2p) < \tau$ we have*

$$(L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{\tau,1} \hookrightarrow C(\overline{\Omega}).$$

Moreover, the embedding constant is bounded by

$$\frac{c\,\Gamma(\tau - d/(2p))}{\Gamma(\tau)}$$

with $c > 0$ independent of τ and p.

Proof. According to Proposition A.13 we have

$$(L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{\tau,1} \hookrightarrow \mathcal{D}_{L^p}((-\Delta)^\tau).$$

Note that the embedding constant can be bounded independently of τ; see Remark A.14. As in the proof of [49, Theorem 2.10 c)], for $\omega > 0$ to be specified later, we use the integral representation of the fractional operator

$$(-\Delta + \omega + 1)^{-\tau} = \frac{1}{\Gamma(\tau)} \int_0^\infty t^{\tau-1} \mathrm{e}^{-t(-\Delta+\omega+1)} \,\mathrm{d}t;$$

see, e.g., [128, Equation (6.9), Chapter 2]. Employing [49, Theorem 2.10 b)], there are $c > 0$ and $\omega > 0$ such that for $\kappa > 0$ sufficiently small we find

$$\|u\|_{C^\kappa(\Omega)} \leq \frac{c}{\Gamma(\tau)} \int_0^\infty t^{\tau-1} t^{-d/(2p)-\kappa/2} \mathrm{e}^{-t} \|(-\Delta + \omega + 1)^\tau u\|_{L^p(\Omega)} \,\mathrm{d}t,$$

where the constants $c > 0$ and ω are independent of κ, p, and τ. For the integral we have the expression

$$\int_0^\infty t^{\tau-1-d/(2p)-\kappa/2} \mathrm{e}^{-t} \,\mathrm{d}t = \Gamma(\tau - d/(2p) - \kappa/2).$$

Employing that $\mathcal{D}_{L^p}((-\Delta)^\tau) = \mathcal{D}_{L^p}((-\Delta + \omega + 1)^\tau)$ with equivalence of norms independent of p, see (A.11), we infer that

$$\mathcal{D}_{L^p}((-\Delta)^\tau) \hookrightarrow C(\overline{\Omega}), \quad d/(2p) < \tau.$$

Finally, going to the limit $\kappa \to 0$ yields the bound on the embedding constant as specified in the proposition. □

Remark 5.36. It is worth mentioning that Proposition 5.35 holds for fairly general domains and divergence form operators even with mixed boundary conditions. We will elaborate on the assumptions of [49] in our setting. In case of homogeneous Dirichlet conditions [49, Assumptions 2.3 and 2.5] are vacuously true. Moreover, [49, Assumptions 2.4] requires the Dirichlet boundary part to be a $(d-1)$-set; see [84, Chapter II]. Since Ω is a Lipschitz domain and there is no Neumann boundary part, from [120, Theorem 4.3] we conclude that $\partial\Omega$ is a $(d-1)$-set. Furthermore, [49] considers 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. For further details we also refer to [16, Appendix A] and the references given therein.

Proposition 5.37. *Let $\omega' \subset \Omega$ such that ω' has a C^∞-boundary. Then*

$$\|u\|_{W^{2,p}(\omega')} \leq c_p \left(\|u\|_{L^p(\Omega)} + \|-\Delta u\|_{L^p(\Omega)} \right), \quad u \in \mathcal{D}_{L^p}(-\Delta),$$

with $c_p \sim p$ as $p \to \infty$.

Proof. Let $u \in \mathcal{D}_{L^p(\Omega)}(-\Delta)$ and set $f := -\Delta u \in L^p(\Omega)$. Then the stated estimate follows from [59, Theorem 9.11]. The exact form of the constant c_p can be traced from the proof of [59, Theorem 9.9] and is given by the Hölder conjugate of the constant from the Marcinkiewicz interpolation theorem. □

Proof of Lemma 5.33. Since $\overline{\omega} \subset \Omega$, there is an open set $\overline{\omega} \subset \omega'$ such that $\overline{\omega'} \subset \Omega$ and ω' has a C^∞-boundary. We use the interior pointwise best approximation result [103, Theorem 2]

$$\|u - u_{kh}\|_{L^\infty(I\times\omega)} \le c|\log k||\log h| \inf_{\varphi_{kh}\in X_{k,h}} \left(\|u - \varphi_{kh}\|_{L^\infty(I\times\omega')} + \|u - \varphi_{kh}\|_{L^\infty(I;L^2(\Omega))} + h\|\nabla(u - \varphi_{kh})\|_{L^\infty(I;L^2(\Omega))}\right)$$

and would like to choose $\varphi_{kh} = i_k \mathrm{I}_h u$. Note that even though [103, Theorem 2] is formulated for ω' being a ball, its proof requires that $\overline{\omega} \subset \omega'$ and $\overline{\omega'} \subset \Omega$ only. The global errors on the right-hand side can be estimated using Proposition 5.34. Hence, we only have to estimate the first term on the right-hand side and consider the splitting

$$\|u - i_k \mathrm{I}_h u\|_{L^\infty(I\times\omega')} \le \|u - i_k u\|_{L^\infty(I\times\omega')} + \|i_k\,(u - \mathrm{I}_h u)\|_{L^\infty(I\times\omega')}.$$

Recall that due to (5.66), the continuous injection

$$W^{1,r}(I; L^p(\Omega)) \cap L^r(I; \mathcal{D}_{L^p}(-\Delta)) \hookrightarrow C^\alpha([0,1]; (L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{\tau,1}), \quad \tau \in (0, 1 - 1/r)$$

holds, where $0 \le \alpha < 1 - 1/r - \tau$. Furthermore, for $\tau > d/(2p)$, we have

$$(L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{\tau,1} \hookrightarrow C(\overline{\Omega});$$

see Proposition 5.35. Taking $\tau = d/p$, its embedding constant is bounded by

$$\frac{c\,\Gamma(\tau - d/(2p))}{\Gamma(\tau)} = \frac{c\,\Gamma(d/(2p))}{\Gamma(d/p)} \to 2 \quad \text{as } p \to \infty.$$

Hence, choosing $\alpha = 1 - 2/r - d/p$ with sufficiently large r we arrive at

$$\|u - i_k u\|_{L^\infty(I\times\omega')} \le c k^{1-2/r-d/p} \|u\|_{W^{1,r}(I;L^p(\Omega))\cap L^r(I;\mathcal{D}_{L^p}(-\Delta))}.$$

The r-dependence of the latter norm can be explicitly given as

$$\|u\|_{W^{1,r}(I;L^p(\Omega))\cap L^r(I;\mathcal{D}_{L^p}(-\Delta))} \le \frac{cr^2}{r-1} \left(\|f\|_{L^r(I;L^p(\Omega))} + \|u_0\|_{(L^p(\Omega),\mathcal{D}_{L^p}(-\Delta))_{1-1/r,r}}\right);$$

see, e.g., [7, Theorem 1.3.2]. Using the fact that

$$\mathcal{D}_{L^\infty}(-\Delta) \hookrightarrow \mathcal{D}_{L^p}(-\Delta) \hookrightarrow (L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{1-1/r,r}$$

with uniform embedding constants for $r \ge r_0 > 1$ for some $r_0 > 1$ (see Proposition A.1 and Remark A.2) and taking $r = |\log k| > 1$ yields

$$\|u - i_k u\|_{L^\infty(I\times\omega')} \le c|\log k| k \left(\|f\|_{L^\infty(I\times\Omega)} + \|u_0\|_{\mathcal{D}_{L^\infty}(-\Delta)}\right).$$

Next, we turn to the error due to spatial discretization. Using (5.66), we find the continuous injection

$$W^{1,r}(I; L^p(\omega')) \cap L^r(I; W^{2,p}(\omega')) \hookrightarrow C([0,1]; (L^p(\omega'), W^{2,p}(\omega'))_{1-1/r,r}).$$

Now let $p, r > 2 + d/2$. If $1 - 1/r > \tau > \max(1/2, d/(2p))$, we have according to Propositions A.5 and A.29

$$(L^p(\omega'), W^{2,p}(\omega'))_{1-1/r,r} \xrightarrow{c(\tau,1-1/r,r,p)} (L^p(\omega'), W^{2,p}(\omega'))_{\tau,p} \xrightarrow{c(\tau)} W^{2\tau,p}(\omega').$$

We take $\tau = 1 - 2/r$ and abbreviate $c_{r,p} = c(1 - 2/r, 1 - 1/r, r, p)$. The constant $c(\tau)$ from Proposition A.29 has the asymptotic behavior $\sim (1-\tau)^{-1}$ for $\tau \to 1$. Since $(1-\tau)^{-1} = r/2$, we obtain

$$\begin{aligned}\|i_k\,(u - \mathrm{I}_h u)\|_{L^\infty(I\times\omega')} &\le \|u - \mathrm{I}_h u\|_{L^\infty(I\times\omega')} \\ &\le c h^{2(1-2/r)-d/p} \|u\|_{C([0,1];W^{2(1-2/r),p}(\omega'))} \\ &\le c c_{r,p} r h^{2(1-2/r)-d/p} \|u\|_{W^{1,r}(I;L^p(\omega'))\cap L^r(I;W^{2,p}(\omega'))},\end{aligned}$$

where we have used the interpolation error estimate from Proposition A.33. Furthermore, Proposition 5.37 implies the estimate

$$\|u\|_{W^{1,r}(I;L^p(\omega'))\cap L^r(I;W^{2,p}(\omega'))} \le c_p \|u\|_{W^{1,r}(I;L^p(\Omega))\cap L^r(I;\mathcal{D}_{L^p}(-\Delta))},$$

where we have assumed without loss that $c_p \ge 1$. As above this yields

$$\|i_k\,(u - \mathrm{I}_h u)\|_{L^\infty(I\times\omega')} \le \frac{c c_{r,p} r c_p r^2}{r-1} \left(\|f\|_{L^r(I;L^p(\Omega))} + \|u_0\|_{(L^p(\Omega),\mathcal{D}_{L^p}(-\Delta))_{1-1/r,r}} \right). \tag{5.70}$$

Taking $r = p = |\log h| > 2 + d/2$, from Remark A.6 we infer the asymptotic behavior $c_{r,p} = c(1 - 1/r, 1 - 2/r, r, r) \sim r$ as $r \to \infty$. Hence, we have the estimate

$$\frac{c c_{r,p} c_p r^3}{r-1} \le c|\log h|^4,$$

where we remind the reader that $c_p \sim p$ as $p \to \infty$. Finally, (5.70) implies

$$\|i_k\,(u - \mathrm{I}_h u)\|_{L^\infty(I\times\omega')} \le c|\log h|^4 h^2 \left(\|f\|_{L^\infty(I\times\Omega)} + \|u_0\|_{\mathcal{D}_{L^\infty}(-\Delta)} \right).$$

This completes the proof. □

5.5.4. Variational control discretization

In the following we consider the case of a distributed control on a subset $\omega \subset \Omega$ starting with the variational control discretization. As before, suppose that the family of triangulations is quasi-uniform; see Definition A.31. For regularity reasons, we suppose that $\overline{\omega} \subset \Omega$. The following error estimates might also hold, if, e.g., ω touches Ω such that $\partial\Omega \cap \partial\omega$ is smooth. However, to avoid further technicalities, we stick to the case, when ω has a strict distance to the boundary of ω.

Introducing an additional term $z_{kh} = z_{kh}(\bar\nu_{kh,\alpha}, i_1 S(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}))$ being the solution to the discrete adjoint equation with terminal value $\bar\mu_{kh,\alpha}(i_1 S(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}) - u_d)$ and time transformation $\bar\nu_{kh,\alpha}$, we split the error

$$\begin{aligned}&\|B^*\,(\bar z_{kh,\alpha} - z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}))\|_{L^\infty(I\times\omega)} \\ &\qquad \le \|B^*\,(\bar z_{kh,\alpha} - z_{kh})\|_{L^\infty(I\times\omega)} + \|B^*\,(z_{kh} - z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}))\|_{L^\infty(I\times\omega)},\end{aligned} \tag{5.71}$$

i.e. z_{kh} is the discrete counterpart to the continuous adjoint state $z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})$ with discrete data. We will treat both terms of the right-hand side of (5.71) separately.

In order to apply the point-wise error estimate Lemma 5.33 for the second term of (5.71) we require $i_1 S(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}) - u_d \in \mathcal{D}_{L^\infty}(-\Delta)$. This will follow from the following proposition at the price of an additional logarithmic factor.

Proposition 5.38. *Let $p \in (1,\infty)$, $\nu \in \mathbb{R}_+$, $u_0 \in L^p(\Omega)$ and $f \in L^\infty(I;L^p(\Omega))$. For all $\tau \in (0,1)$ the solution u to the state equation with right-hand side f and initial value u_0 satisfies*

$$\|(-\Delta)^\tau u(1)\|_{L^p(\Omega)} \le c\nu^{-\tau}\left(\|u_0\|_{L^p(\Omega)} + \nu(1-\tau)^{-1}\|f\|_{L^\infty(I;L^p(\Omega))}\right)$$

with a constant $c > 0$ independent of ν, u_0, f, and τ.

Proof. Using [128, Theorem 2.6.13], we find for all $\tau \in (0,1)$ that

$$\begin{aligned}\|(-\Delta)^\tau u(1)\|_{L^p(\Omega)} &\le \|(-\Delta)^\tau \mathrm{e}^{\nu\Delta}u_0\|_{L^p(\Omega)} + \nu\int_0^1 \|(-\Delta)^\tau \mathrm{e}^{\nu(1-s)\Delta}\|_{\mathcal{L}(L^p)}\|f(s)\|_{L^p(\Omega)}\,\mathrm{d}s\\ &\le c_\tau\nu^{-\tau}\|u_0\|_{L^p(\Omega)} + \nu c_\tau\|f\|_{L^\infty(I;L^p(\Omega))}\nu^{-\tau}\int_0^1 (1-s)^{-\tau}\,\mathrm{d}s\\ &= c_\tau\nu^{-\tau}\left(\|u_0\|_{L^p(\Omega)} + \nu(1-\tau)^{-1}\|f\|_{L^\infty(I;L^p(\Omega))}\right).\end{aligned}$$

The constant c_τ depends on the resolvent estimate (A.13) for $-\Delta$ which does not depend on p. Last, the constant c_τ can be chosen to be independent of τ. □

Note that in the proof of Lemma 5.33 we have used the embedding

$$\mathcal{D}_{L^\infty}(-\Delta) \hookrightarrow (L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{1-1/r,r}.$$

Indeed, we have

$$\mathcal{D}_{L^p}((-\Delta)^\tau) \hookrightarrow (L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{\tau,\infty} \hookrightarrow (L^p(\Omega), \mathcal{D}_{L^p}(-\Delta))_{1-1/r,r}$$

for $\tau > 1 - 1/r$; see Propositions A.4 and A.15. We emphasize that the embedding constants do not depend on p; see Remark A.17. Choosing $\tau = 1 - 1/(2r)$, the embedding constant of the first injection is uniformly bounded as $r \to \infty$. Furthermore, the embedding constant of the second injection satisfies

$$c_r := c(1-1/r, 1-1/(2r), \infty, r) = \left(2r + \frac{r}{r-1}\right)\left[r\min\left\{1-\frac{1}{r}, \frac{1}{r}\right\}\right]^{1-1/r} \sim r$$

as $r \to \infty$. Using Proposition 5.38, the state $u = S(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})$ satisfies

$$\|u(1)\|_{(L^p(\Omega),\mathcal{D}_{L^p}(-\Delta))_{1-1/r,r}} \le cc_r(1-\tau)^{-1}, \quad \tau > 1 - 1/r.$$

Hence, choosing $\tau = 1 - 1/(2r)$, or, equivalently, $2r = (1-\tau)^{-1}$, and using the embedding above in the proof of Lemma 5.33, the second term of (5.71) can be estimated as

$$\begin{aligned}&\|B^*\left(z_{kh} - z(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})\right)\|_{L^\infty(I\times\omega)}\\ &\quad\le c|\log k|^4|\log h|^7\left(k + h^2\right)\left(\|u_0\|_{L^\infty(\Omega)} + \|B\bar q_{kh,\alpha}\|_{L^\infty(I;L^\infty(\Omega))} + \|u_d\|_{\mathcal{D}_{L^\infty}(-\Delta)}\right). \quad (5.72)\end{aligned}$$

Next, we consider the first term of the right-hand side of (5.71). For this, we require a pointwise stability result for the adjoint state equation.

Proposition 5.39. *Let $\nu \in [\nu_{\min}, \nu_{\max}]$ for fixed $0 < \nu_{\min} < \nu_{\max}$. Moreover, consider $\omega \subset \omega_1 \subset \Omega$ open sets, $\overline{\omega} \subset \omega_1$, $\overline{\omega_1} \subset \Omega$, and suppose that ω_1 has a smooth boundary. The solution to the discrete adjoint equation $z_{kh} \in X_{k,h}$ with terminal value z_1 and time transformation ν satisfies the estimate*

$$\|z_{kh}\|_{L^\infty(I\times\omega)} \le c\left(\|z\|_{L^\infty(I\times\omega_1)} + \|z\|_{L^\infty(I;L^2(\Omega))} + h\|\nabla z\|_{L^\infty(I;L^2(\Omega))}\right),$$

where z is the continuous counterpart to z_{kh}. The constant $c > 0$ is independent of k, h, ν, z_1, z, and z_{kh}.

Proof. The result is shown in the proof of [103, Theorem 2], where the stated estimate can be found at the bottom of page 1382. Again, even though [103, Theorem 2] is formulated for ω_1 being a ball, its proof requires that $\overline{\omega} \subset \omega_1$ and $\overline{\omega_1} \subset \Omega$, only. □

Proposition 5.40. *Let $\nu \in [\nu_{\min}, \nu_{\max}]$ for fixed $0 < \nu_{\min} < \nu_{\max}$ and $\omega_1 \subset \omega_2 \subset \Omega$ be open such that $\overline{\omega_1} \subset \omega_2$. Moreover, let $z_1 \in L^2(\Omega)$ such that $z_1|_{\omega_2} \in L^\infty(\omega_2)$. The solution z to*

$$-\partial_t z - \nu \Delta z = 0, \quad z(1) = z_1,$$

satisfies the estimate

$$\|z\|_{L^\infty(I\times\omega_1)} \leq c \left(\|z_1\|_{L^\infty(\omega_2)} + \|z_1\|_{L^2(\Omega)} \right)$$

with $c > 0$ independent of ν, z_1, and z.

For the proof of Proposition 5.40, we require the following standard stability estimate.

Proposition 5.41. *Let $\nu \in [\nu_{\min}, \nu_{\max}]$ for fixed $0 < \nu_{\min} < \nu_{\max}$, $f \in L^s(I; L^p)$, and $v_0 \in L^\infty$ with $d/(2p) < 1 - 1/s$ and $s, p \in (1, \infty)$. The solution v to*

$$\partial_t v - \nu \Delta v = f, \quad v(0) = v_0,$$

satisfies the estimate

$$\|v\|_{L^\infty(I\times\Omega)} \leq c \left(\|f\|_{L^s(I;L^p(\Omega))} + \|v_0\|_{L^\infty(\Omega)} \right)$$

with $c > 0$ independent of ν, f, v_0, and v.

Proof. If $f = 0$, then this follows from the fact that the semigroup generated by Δ is contractive on $L^\infty(\Omega)$; see, e.g., [66, Theorem 4.12]. If $v_0 = 0$, we apply [49, Theorem 3.1]. Superposition of both estimates yields the assertion for any fixed ν. Furthermore, continuity of the mapping $\nu \mapsto (\partial_t - \nu\Delta)^{-1}$ from $\mathbb{R}_+$ into $\mathcal{L}(L^s(I; L^p(\Omega)), L^\infty(I \times \Omega))$ and compactness of $[\nu_{\min}, \nu_{\max}]$ implies that the constant can be chosen to be independent of ν. □

Proof of Proposition 5.40. Let ω' be a further subdomain with smooth boundary such that $\omega_1 \subset \omega' \subset \omega_2$. Moreover, let $\xi \colon \Omega \to \mathbb{R}$ be a smooth cut-off function such that $\xi(x) = 1$ if $x \in \omega_1$ and $\xi(x) = 0$ if $x \in \Omega \setminus \omega'$. Then for all $\varphi \in H_0^1(\Omega)$ the expression

$$\begin{aligned} -\langle \Delta(\xi z), \varphi \rangle &= \langle \xi \nabla z, \nabla \varphi \rangle + \langle z \nabla \xi, \nabla \varphi \rangle \\ &= -\langle \Delta z, \xi \varphi \rangle - 2\langle \nabla z \cdot \nabla \xi, \varphi \rangle - \langle z \Delta \xi, \varphi \rangle \end{aligned}$$

holds. Hence,

$$-\partial_t(\xi z) - \nu \Delta(\xi z) = \xi(-\partial_t z - \nu \Delta z) - 2\nu \nabla z \cdot \nabla \xi - \nu z \Delta \xi = -2\nu \nabla z \cdot \nabla \xi - \nu z \Delta \xi,$$

i.e. $v = \xi z$ solves

$$-\partial_t v - \nu \Delta v = -2\nu \nabla z \cdot \nabla \xi - \nu z \Delta \xi, \quad v(1) = (\xi z)(1).$$

Using Proposition 5.41 with $p = 4$ and some $s \in (8/5, 2)$ we infer that

$$\|z\|_{L^\infty(I\times\omega_1)} \leq \|v\|_{L^\infty(I\times\omega')} \leq c \left(\|z\|_{L^s(I;W^{1,4}(\omega'))} + \|z_1\|_{L^\infty(\omega')} \right).$$

To arrive at the stated estimate, we have to bound the term $\|z\|_{L^s(I;W^{1,4}(\omega_2))}$. Consider a new cut-off function $\xi\colon \Omega \to \mathbb{R}$ such that $\xi(x) = 1$ if $x \in \omega'$ and $\xi(x) = 0$ if $x \in \Omega \setminus \omega_2$. Since $\xi z = z$ and $\nabla(\xi z) = \nabla z$ in ω', we have

$$\|z\|_{L^s(I;W^{1,4}(\omega'))} = \|\xi z\|_{L^s(I;W^{1,4}(\omega'))} \le \|\xi z\|_{L^s(I;W^{1,4}(\Omega))}.$$

Using that $\mathcal{D}_{W^{-1,4}}(-\Delta) = W_0^{1,4}(\Omega)$, see [45, Corollary 3.12], we deduce that

$$\|z\|_{L^s(I;W^{1,4}(\omega'))} \le c\|\xi z\|_{L^s(I;\mathcal{D}_{W^{-1,4}}(-\Delta))}.$$

Then maximal parabolic regularity of $-\Delta$ on $W^{-1,4}(\Omega)$, see, e.g., [8, Theorem 11.5], and the fact that $-2\nabla z \cdot \nabla\xi - z\Delta\xi = -2\nabla\cdot(z\nabla\xi) + z\Delta\xi$ imply

$$\|v\|_{L^s(I;\mathcal{D}_{W^{-1,4}}(-\Delta))} \le c\left(\|-2\nabla\cdot(z\nabla\xi) + z\Delta\xi\|_{L^s(I;W^{-1,4}(\Omega))} + \|\xi z_1\|_{(W^{-1,4}(\Omega),W_0^{1,4}(\Omega))_{1-1/s,s}}\right).$$

Since the mapping $\nu \mapsto (\partial_t - \nu\Delta)^{-1}$ is continuous, the constant above can be chosen uniformly with respect to $\nu \in [\nu_{\min}, \nu_{\max}]$. Moreover, according to [65, Lemma 3.4] and [146, Theorems 1.15.2 d), 1.3.3 e)] the embedding

$$\begin{aligned} L^4(\Omega) = [W^{-1,4}(\Omega), W_0^{1,4}(\Omega)]_{1/2} &\hookrightarrow (W^{-1,4}(\Omega), W_0^{1,4}(\Omega))_{1/2,\infty} \\ &\hookrightarrow (W^{-1,4}(\Omega), W_0^{1,4}(\Omega))_{1-1/s,s} \end{aligned}$$

holds, if $1/2 > 1 - 1/s$, or, equivalently, if $s < 2$. Thus,

$$\|\xi z_1\|_{(W^{-1,4}(\Omega),W_0^{1,4}(\Omega))_{1-1/s,s}} \le c\|z_1\|_{L^4(\omega_2)} \le c\|z_1\|_{L^\infty(\omega_2)}.$$

For the remaining term, we estimate

$$\|-2\nabla\cdot(z\nabla\xi) + z\Delta\xi\|_{L^s(I;W^{-1,4}(\Omega))} \le c\|z\|_{L^s(I;L^4(\Omega))} \le c\|z\|_{L^2(I;H_0^1(\Omega))} \le c\|z_1\|_{L^2(\Omega)},$$

where we have used the Sobolev embedding $H_0^1(\Omega) \hookrightarrow L^4(\Omega)$ in the second last step. □

Proposition 5.40 allows to estimate the $L^\infty(I \times \omega_1)$ term of the right-hand side of Proposition 5.39. To estimate the remaining terms of the right-hand side of Proposition 5.39, we observe that the solution z from Proposition 5.39 in addition satisfies the estimates

$$\begin{aligned} \|z\|_{L^\infty(I;L^2(\Omega))} &\le c\|z_1\|_{L^2(\Omega))}, \\ \|\nabla z\|_{L^\infty(I;L^2(\Omega))} &\le c\|\nabla z_1\|_{L^2(\Omega))}. \end{aligned}$$

Hence, combination of Propositions 5.39 and 5.40 immediately implies the estimate

$$\begin{aligned} \|B^*(\bar z_{kh,\alpha} - z)\|_{L^\infty(I\times\omega)} \le c\Big(&\|u_{kh}(1) - u(1)\|_{L^\infty(\omega_2)} \\ &+ \|u_{kh}(1) - u(1)\|_{L^2(\Omega)} + h\|\nabla(u_{kh}(1) - u(1))\|_{L^2(\Omega)}\Big), \end{aligned} \tag{5.73}$$

where we have set $u_{kh} = i_1 S_{kh}(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})$ and $u = S(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})$ for convenience. The first term of the right-hand side is estimated using Lemma 5.33 (with $\omega = \omega_2$). To estimate the remaining terms, we use Lemma A.39 and the following estimates: Let R_h denote the

Ritz projection. Then an inverse estimate, see, e.g., [24, Theorem 4.5.11], and the best approximation property of the Ritz projection in H_0^1 lead to

$$\begin{aligned}\|\nabla(u_{kh}(1)-u(1))\|_{L^2(\Omega)} &\le \|\nabla(u_{kh}(1)-R_h u(1))\|_{L^2(\Omega)} + \|\nabla(R_h u(1)-u(1))\|_{L^2(\Omega)}\\ &\le ch^{-1}\|u_{kh}(1)-R_h u(1)\|_{L^2(\Omega)} + \|\nabla(\mathrm{I}_h u(1)-u(1))\|_{L^2(\Omega)}.\end{aligned}$$

Using again Lemma A.39 and the error estimate (A.28) for R_h we find

$$\begin{aligned}\|u_{kh}(1)-R_h u(1)\|_{L^2(\Omega)} &\le \|u_{kh}(1)-u(1)\|_{L^2(\Omega)} + \|u(1)-R_h u(1)\|_{L^2(\Omega)}\\ &\le c|\log k|(k+h^2) + ch^{1+\tau}\|u(1)\|_{W^{1+\tau,2}(\Omega)}.\end{aligned}$$

Moreover, according to Proposition A.33, it holds

$$\|\nabla(\mathrm{I}_h u(1)-u(1))\|_{L^2(\Omega)} \le ch^{\tau}\|u(1)\|_{W^{1+\tau,2}(\Omega)}.$$

To estimate the $W^{1+\tau,2}$ norm, we employ Proposition A.16 and obtain

$$\mathcal{D}_{L^2}((-\Delta)^{\tau}) \hookrightarrow (H^1(\Omega), H^2(\Omega)\cap H_0^1(\Omega))_{\tau,2} \hookrightarrow (H^1(\Omega), H^2(\Omega))_{\tau,2} \hookrightarrow W^{1+\tau,2}(\Omega),$$

where we have used Proposition A.27 in the last step. Note that the constant of the first embedding constant is bounded by

$$1+(-2\cos(\pi\tau)\,\Gamma(-2\tau))^{1/2} \sim (1-\tau)^{-1/2} \quad \text{as } \tau \to 1.$$

The remaining embedding constants can be bounded uniformly. Finally, Proposition 5.38 with $(1-\tau)^{-1} = |\log h|$ implies

$$\|u(1)\|_{W^{1+\tau,2}(\Omega)} \le c|\log h|^{1/2}\left(\|B\bar{q}_{kh,\alpha}\|_{L^\infty(I;L^2(\Omega))} + \|u_0\|_{L^2(\Omega)}\right).$$

Collecting all estimates we have

$$\begin{aligned}&h\|\nabla(u_{kh}(1)-u(1))\|_{L^2(\Omega)}\\ &\qquad\le c|\log k||\log h|^{3/2}(k+h^2)\left(\|B\bar{q}_{kh,\alpha}\|_{L^\infty(I;L^2(\Omega))} + \|u_0\|_{L^2(\Omega)}\right). \quad (5.74)\end{aligned}$$

Hence, from (5.73), Lemmas 5.33 and A.39 as well as (5.74), we obtain

$$\|B^*(\bar{z}_{kh,\alpha}-z_{kh})\|_{L^\infty(I\times\omega)} \le c|\log k|^2|\log h|^5(k+h^2). \quad (5.75)$$

Finally, (5.72) and (5.75) yield the following estimate that we summarize for later reference.

Proposition 5.42. *Let $\overline{\omega} \subset \Omega$. Suppose that $u_0, u_d \in \mathcal{D}_{L^\infty}(-\Delta)$. Then there exists a constant $c>0$ such that*

$$\|B^*(\bar{z}_{kh,\alpha}-z(\bar{\nu}_{kh,\alpha},\bar{q}_{kh,\alpha}))\|_{L^\infty(I\times\omega)} \le c|\log k|^4|\log h|^7(k+h^2),$$

where $c>0$ is independent of k, h, α, $\bar{z}_{kh,\alpha}$, and $z(\bar{\nu}_{kh,\alpha},\bar{q}_{kh,\alpha})$.

By means of Propositions 5.28 and 5.42 we finally infer the following error estimate for the variational control discretization.

Theorem 5.43 (Variational discretization)**.** *Adopt the assumptions of Lemma 5.26 and let* (3.37) *hold. Moreover, suppose the variational control discretization, i.e. $Q_\sigma(0,1) = Q(0,1)$. In addition, assume $\overline{\omega} \subset \Omega$ as well as $u_0, u_d \in \mathcal{D}_{L^\infty}(-\Delta)$. Then there is a constant $c>0$ not depending on k, h, α, $\bar{\nu}_{kh,\alpha}$, and $\bar{q}_{kh,\alpha}$ such that*

$$|\bar{\nu}-\bar{\nu}_{kh,\alpha}| + \|\bar{q}-\bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}^{1/\kappa} \le c\left(\alpha + |\log k|^4|\log h|^7(k+h^2)\right).$$

Proof. This result immediately follows from Lemma 5.26 and Propositions 5.28 and 5.42, since for the variational control discretization we have $\mathrm{I}_\sigma = \mathrm{Id}$ and $\sigma_1(k,h) = \sigma_2(k,h) = 0$. □

5.5.5. Cellwise constant control discretization

Next, we consider the case of a distributed control on a subset $\omega \subset \Omega$ with controls discretized by cellwise constant functions in space. Recall that σ_1 denotes the projection error onto Q_σ measured in L^1 and σ_2 denotes the same error measured in $L^2(I; H^{-1})$; see (5.50) and (5.51). Since the control variable possesses less regularity (compared to the case $\alpha > 0$), for cellwise constant control discretization, we cannot expect order k of convergence in L^2 in time. We therefore propose a semivariational control discretization. To this end, let the discrete space of controls be defined as follows

$$\begin{aligned} Q_h &= \left\{ v \in L^2(\omega) \colon v|_K \in \mathcal{P}_0(K) \text{ for all } K \in \mathcal{T}_h^\omega \right\}, \\ Q_\sigma(0,1) &= L^2(I; Q_h). \end{aligned}$$

Hence, the controls are explicitly discretized in space but not explicitly discretized in time. Note that in case of $\alpha > 0$, then the optimal controls $\bar{q}_{kh,\alpha}$ are implicitly discretized by means of the projection formula (5.44); cf. also the discussion after Theorem 5.31. From the perspective of the numerical realization, one often uses $\alpha > 0$, because the problems for $\alpha = 0$ are typically difficult to solve numerically.

On any $K \in \mathcal{T}_h$ we define the cellwise constant projection $\Pi_{h,0}$ via

$$(\Pi_{h,0} v)(t,x) = \frac{1}{|K|} \int_K v(t,\xi) \, \mathrm{d}\xi, \quad t \in [0,1], x \in K.$$

Moreover, for almost every $t \in [0,1]$ we set

$$\mathcal{S}_{h,t} \coloneqq \mathcal{T}_h^\omega \setminus \{K \in \mathcal{T}_h^\omega \colon \bar{q}(t)|_K \equiv q_a \text{ or } \bar{q}(t)|_K \equiv q_b\},$$

for all $v \in L^2(I; L^2)$. We first establish the required estimates for σ_1 and σ_2 with $\mathrm{I}_\sigma = \Pi_{h,0}$.

Proposition 5.44. *Suppose there are functions $\delta_h \in L^1(I)$, $h > 0$, and a constant $c > 0$ such that*

$$\sum_{K \in \mathcal{S}_{h,t}} |K| \le \delta_h(t), \quad \text{a.e. } t \in [0,1], \quad h > 0, \tag{5.76}$$

and $\|\delta_h\|_{L^1(I)} \le ch$ for all $h > 0$. Then the estimates

$$\|\Pi_{h,0}\bar{q} - \bar{q}\|_{L^1(I\times\omega)} \le ch, \tag{5.77}$$

$$\|B\,(\Pi_{h,0}\bar{q} - \bar{q})\|_{L^2(I;H^{-1})} \le ch^{3/2}, \tag{5.78}$$

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

Proof. Because $\Pi_{h,0}\bar{q}(t)|_K \equiv q_a$, if $\bar{q}(t)|_K \equiv q_a$, and the same for the upper bound q_b, we obtain

$$\|\Pi_{h,0}\bar{q}(t) - \bar{q}(t)\|_{L^1(\omega)} \le c \sum_{K \in S_{h,t}} |K| \le \delta_h(t),$$

for almost every $t \in (0,1)$, where we have used that $\bar{q}(t) \in L^\infty(\omega)$ and (5.76). Integration with respect to t implies the first estimate (5.77). Moreover, for any $v \in H^1$ and $K \in \mathcal{T}_h^\omega$ we have

$$\begin{aligned} (\Pi_{h,0}\bar{q}(t) - \bar{q}(t), v)_{L^2(K)} &= (\Pi_{h,0}\bar{q}(t) - \bar{q}(t), v - \Pi_{h,0}v)_{L^2(K)} \\ &\le \|\Pi_{h,0}\bar{q}(t) - \bar{q}(t)\|_{L^2(K)} \|\Pi_{h,0}v - v\|_{L^2(K)} \\ &\le ch \|\Pi_{h,0}\bar{q}(t) - \bar{q}(t)\|_{L^2(K)} \|\nabla v\|_{L^2(K)}, \end{aligned}$$

since $\Pi_{h,0}$ is a projection. Hence, using Hölder's inequality, we infer that

$$\begin{aligned}\|B\left(\Pi_{h,0}\bar{q}(t)-\bar{q}(t)\right)\|_{H^{-1}} &= \sup_{v\in H_0^1}\frac{(\Pi_{h,0}\bar{q}(t)-\bar{q}(t),v)_{L^2}}{\|v\|_{H^1}}\\ &\le ch\sup_{v\in H_0^1}\|v\|_{H^1}^{-1}\sum_{K\in S_{h,t}}\|\Pi_{h,0}\bar{q}(t)-\bar{q}(t)\|_{L^2(K)}\|\nabla v\|_{L^2(K)}\\ &\le ch\sup_{v\in H_0^1}\|v\|_{H^1}^{-1}\left(\sum_{K\in S_{h,t}}\|\Pi_{h,0}\bar{q}(t)-\bar{q}(t)\|_{L^2(K)}^2\right)^{1/2}\|\nabla v\|_{L^2}\\ &\le ch\left(\sum_{K\in S_{h,t}}|K|\right)^{1/2}\le ch\left(\delta_h(t)\right)^{1/2},\end{aligned}$$

for almost every $t\in(0,1)$, where we have used (5.76) in the last estimate. Integration with respect to t leads to (5.78). □

Note that the condition (5.76) allows for certain accumulation of switching hyperplanes which might occur for bang-bang controls. Employing the structural assumption of the adjoint state (3.37), we can derive the following sufficient condition for (5.76); cf. also the proof of [36, Theorem 4.4]. We emphasize, that the condition (5.76) is less restrictive than to suppose that (3.37) holds with $\kappa=1$. This is also observed in the numerical examples; see Section 5.7.3.

Proposition 5.45. *If $B^*\bar{z}\in L^1(I;C^1(\overline{\omega}))$ and (3.37) holds with $\kappa=1$, then (5.76) is valid.*

Proof. Let $t\in[0,1]$ and $K\in\mathcal{S}_{h,t}$, i.e. $\bar{z}(t)$ changes sign in K. Hence, there exists $x_K\in K\subset\omega$ such that $B^*\bar{z}(t,x_K)=0$. Using the assumed regularity for $B^*\bar{z}$, we find for all $x\in K$ that

$$|B^*\bar{z}(t,x)| = |B^*\left(\bar{z}(t,x)-\bar{z}(t,x_K)\right)| \le ch\|B^*\bar{z}(t)\|_{C^1(\overline{\omega})}.$$

Thus,

$$\bigcup_{K\in\mathcal{S}_{h,t}}(t,K)\subset\left\{\,(t,x)\colon x\in\omega, |B^*\bar{z}(t,x)|\le ch\|B^*\bar{z}(t)\|_{C^1(\overline{\omega})}\,\right\}.$$

Now, the inclusion above and (3.37) with $\kappa=1$ imply

$$\sum_{K\in\mathcal{S}_{h,t}}|K|\le ch\|\bar{z}(t)\|_{C^1(\overline{\omega})}=:\delta_h(t).$$

Integration yields $\|\delta_h\|_{L^1}=ch$. □

Finally, we provide an error estimate for cellwise constant control discretization.

Theorem 5.46 (Cellwise constant controls)**.** *Adopt the assumptions of Lemma 5.26 and let (3.37) hold. Moreover, suppose the variational in time and cellwise constant control discretization in space, i.e. $Q_\sigma(0,1)=L^2(I;Q_h)$. In addition, assume $\overline{\omega}\subset\Omega$, $u_0,u_d\in\mathcal{D}_{L^\infty}(-\Delta)$, and that (5.76) is satisfied. There is a constant $c>0$ not depending on k, h, α, $\bar{\nu}_{kh,\alpha}$, and $\bar{q}_{kh,\alpha}$ such that*

$$|\bar{\nu}-\bar{\nu}_{kh,\alpha}|+\|\bar{q}-\bar{q}_{kh,\alpha}\|_{L^1(I\times\omega)}^{1/\kappa}\le c\left(\alpha+|\log k|^4|\log h|^7(k+h)\right).$$

Proof. We first note that for $p > d$ and using Proposition 5.37 we have the estimate

$$\|B^*\bar{z} - \Pi_{h,0}B^*\bar{z}\|_{L^\infty(I\times\omega)} \le ch\|\bar{z}\|_{L^\infty(I;W^{2,p}(\omega))} \le ch\|\bar{z}\|_{L^\infty(I;\mathcal{D}_{L^p}(-\Delta))} \le ch.$$

Hence, employing Lemma 5.26, Propositions 5.28 and 5.42 as well as the estimates for σ_1 and σ_2 from Proposition 5.44 yield the desired estimate. □

5.6. Robust error estimates for bang-bang controls ($\alpha = 0$) without sufficient optimality condition

While the error estimates of the preceding section essentially used a structural assumption on the adjoint state that is in general difficult to verify, in this section we will provide error estimates for the terminal time that rely on a condition that can be verified a priori. The estimates are based on the construction of feasible controls and crosswise testing. The techniques can be applied to relatively general problems and – because this will not lead to unnecessarily overloaded notation – we will discuss the main tool for a general autonomous evolution equation formulated in a Gelfand triple $V \hookrightarrow_c H \hookrightarrow V^*$. Moreover, the terminal set $U \subset H$ is assumed to be closed and convex.

Recall from Chapter 2, the *lower Hamiltonian* is defined by

$$h(u,\zeta) = \min_{q\in Q_{ad}} \langle Bq - Au, \zeta\rangle, \quad \text{for } u,\zeta \in V.$$

Suppose that P_U is stable in V and that there is $h_0 \ge 0$ such that for all $v \in V$ it holds

$$h(u,\zeta) \le -h_0\|\zeta\|, \quad \text{where } u = P_U(v),\ \zeta = v - u.$$

Then, according to Lemma 2.10, for each $u_0 \in H$ with $d_U(u_0)\,\omega_0 \le h_0$ there exists a control $q\colon [0,\infty) \to Q_{ad}$ such that the solution u to

$$\partial_t u + Au = Bq, \quad u(0) = u_0,$$

satisfies

$$d_U(u(t)) \le \max\{\,0,\ d_U(u_0) + (d_U(u_0)\,\omega_0 - h_0)\,t\,\}, \quad t \ge 0.$$

We will prove a discrete analog to Lemma 2.10 that will be used to construct feasible controls for the discrete problem. First, recall the Gårding inequality

$$\langle Au, u\rangle + \omega_0\|u\|^2 \ge \alpha_0\|u\|_V^2, \quad u \in V,$$

concerning the operator A. For $h > 0$, let $V_h \subset V$ be finite dimensional subspaces (equipped with the inner product and norm of V) and consider operators $A_h\colon V_h \to V_h^* \cong V_h$ satisfying Gårding's inequality on V_h, precisely

$$\langle A_h u, u\rangle + \omega_0\|u\|^2 \ge \alpha_0\|u\|_V^2, \quad u \in V_h,$$

for all $h > 0$ sufficiently small, where ω_0 and α_0 are fixed. For any $T > 0$, consider a partitioning of the time interval $[0,T]$ given as

$$[0,T] = \{0\} \cup I_1 \cup I_2 \cup \ldots \cup I_M$$

with disjoint subintervals $I_m = (t_{m-1}, t_m]$ of size k_m defined by the time points

$$0 = t_0 < t_1 < \ldots < t_{M-1} < t_M = T.$$

We abbreviate the time discretization by the parameter k defined as the piecewise constant function by setting $k|_{I_m} = k_m$ for all $m = 1, 2, \ldots, M$. Simultaneously, we denote by k the maximal size of the time steps, i.e. $k = \max k_m$. Given the temporal mesh, for any Banach space Y, we introduce the space of piecewise constant functions

$$X_k(Y) := \left\{ v \in L^2((0,T);Y) \colon v|_{(t_{m-1},t_m]} \in \mathcal{P}_0((t_{m-1},t_m];Y),\ m = 1,2,\ldots,M \right\}.$$

In addition, let $Q_h \subset Q$ be a subspace (not necessarily finite dimensional) and we define the set of admissible controls $Q_{ad,h} = Q_h \cap Q_{ad}$. For simplicity, we write $q \in X_k(Q_{ad,h})$ if $q \in X_k(Q_h)$ and $q|_{(t_{m-1},t_m]} \in \mathcal{P}_0((t_{m-1},t_m];Q_{ad,h})$ for all $m = 1, 2, \ldots, M$.

5.6.1. The discrete Hamiltonian and the contruction of feasible controls

For the construction of feasible points, we prove a discrete analog to Lemma 2.10. The proof presented here is based on a preliminary version of [18]. We introduce the *discrete lower Hamiltonian* for A_h on V_h as

$$h_h(u,\zeta) = \min_{q \in Q_{ad,h}} \langle Bq - A_h u, \zeta \rangle, \quad \text{for } u, \zeta \in V_h;$$

cf. the lower Hamiltonian on the continuous level.

Lemma 5.47. *Let P_U be stable in V_h, i.e. $P_U(V_h) \subset V_h$, and $k < 1/\omega_0$. Suppose there is $h_0 \geq 0$ such that for all $v \in V_h$ it holds*

$$h_h(u,\zeta) \leq -h_0\|\zeta\|, \quad \text{where } u = P_U(v),\ \zeta = v - u. \tag{5.79}$$

Then, for each $u_0 \in V_h$ with $d_U(u_0) \leq h_0/(4\omega_0)$ there exists a control $q_{kh} \in X_k(Q_{ad,h})$ such that the solution $u_{kh} \in X_k(V_h)$ to the discrete state equation, i.e.

$$\int_0^T \langle A_h u_{kh}, \varphi_{kh} \rangle + \sum_{m=2}^{M} ([u_{kh}]_{m-1}, \varphi_{kh,m}) + (u_{kh,1}, \varphi_{kh,1}) = (u_0, \varphi_{kh,1}) + \int_0^T \langle Bq_{kh}, \varphi_{kh} \rangle$$

for all $\varphi_{kh} \in X_k(V_h)$, satisfies

$$d_U(u_{kh}(t_m)) \leq \max\{\, 0,\ d_U(u_0) - (h_0/2)t_m \,\}, \quad m = 1, 2, \ldots, M.$$

To prove this result, we first regularize the distance function on U. For $\gamma \geq 0$ consider the mapping $\phi_\gamma \colon \mathbb{R}_+ \to \mathbb{R}_+$ defined as

$$\phi_\gamma(t) = \begin{cases} t^2/(2\gamma) & \text{if } 0 \leq t < \gamma, \\ t - \gamma/2 & \text{if } t \geq \gamma. \end{cases}$$

Then, the regularized distance function is given by $d_\gamma(u) = \phi_\gamma(d_U(u))$. Clearly, we have

$$d_U(u) - \gamma/2 \leq d_\gamma(u) \leq d_U(u) \quad \text{for all } u \in H.$$

Lemma 5.48. *For all $\gamma > 0$ and $u \in H$ the regularized distance function is differentiable with*

$$\nabla d_\gamma(u) = \frac{\phi_\gamma'(d_U(u))}{d_U(u)}(u - P_U(u)), \quad \text{where} \quad \phi_\gamma'(t) = \begin{cases} t/\gamma & \text{if } 0 \le t < \gamma, \\ 1 & \text{if } t \ge \gamma. \end{cases}$$

Furthermore, the gradient $\nabla d_\gamma \colon H \to H$ is Lipschitz continuous with

$$\|\nabla d_\gamma(u) - \nabla d_\gamma(v)\| \le \gamma^{-1}\|u - v\|.$$

Proof. First we note that the choice of d_γ is not arbitrary. In fact, this is precisely the Moreau-envelope for the parameter γ of the distance function

$$d_\gamma(u) = d_U^\gamma(u) = \min_{v \in H} \left[\frac{1}{2\gamma}\|v - u\|^2 + d_U(v) \right];$$

cf., e.g., [12, Section 12.4]. The differentiability and the Lipschitz continuity of the gradient follow directly from that; see, e.g., [12, Proposition 12.29]. Concerning the concrete form of the derivative, we note that for $d_U(u) > 0$, $d_U(u)$ is differentiable with $\nabla d_U(u) = (u - P_U(u))/d_U(u)$ (see, e.g., [12, Proposition 18.22 (i)]). Therefore, we can apply the chain rule. For $d_U(u) = 0$, we can provide a direct proof. □

Now, for $\zeta \in V_h$, we define for $\gamma > 0$ the controls of the form

$$q_\gamma = \Pi_{ad,h} \left\{ -\frac{1}{\gamma} B^* \zeta \right\}, \tag{5.80}$$

where $\Pi_{ad,h}$ denotes the Hilbert space projection onto $Q_{ad,h}$, i.e.

$$q = \Pi_{ad,h}\{z\} \quad \Leftrightarrow \quad (q - z, q' - q)_Q \ge 0 \quad \text{for all } q' \in Q_{ad,h}.$$

The controls q_γ approximate the minimizers from the definition of the discrete lower Hamiltonian.

Proposition 5.49. *For any $\zeta, u \in V_h$ and q_γ as in (5.80) we have*

$$\langle Bq_\gamma - A_h u, \zeta \rangle \le h_h(u, \zeta) + c\gamma, \tag{5.81}$$

where c solely depends on $Q_{ad,h}$.

Proof. This follows as in Proposition 2.13 noting that $Q_{ad,h} \subset Q_{ad}$ due to $Q_h \subset Q$. □

For the following considerations, we define $B_h \colon Q_h \to V_h^*$ as

$$\langle B_h q, \varphi_h \rangle = \langle Bq, \varphi_h \rangle \quad \text{for all } \varphi_h \in V_h.$$

Proposition 5.50. *Let $\gamma > 0$. Suppose that $P_U(V_h) \subset V_h$. For any $u_{m-1} \in V_h$ and $k_m \in (0, 1/\omega_0]$, the system of equations*

$$\begin{aligned} (I + k_m A_h) u_m &= u_{m-1} + k_m B_h q_m, \\ q_m &= \Pi_{ad,h} \left\{ -(1/\gamma) B_h^* \nabla d_\gamma(u_m) \right\}, \end{aligned} \tag{5.82}$$

possesses a solution $(u_m, q_m) \in V_h \times Q_{ad,h}$.

Proof. Consider the mapping

$$\mathcal{F}(q) \coloneqq \Pi_{ad,h}\left\{-\frac{1}{\gamma}B_h^*\left[\nabla d_\gamma\left((I+k_mA_h)^{-1}(u_{m-1}+k_mB_hq)\right)\right]\right\}.$$

First, because B is continuous from Q to V^*, we infer that B_h is continuous from Q_h to V_h^*. Moreover, since $(I+k_mA_h)$ is continuously invertible, we have $(I+k_mA_h)^{-1}\colon V_h^* \cong V_h \to V_h$. Additionally, ∇d_γ is Lipschitz continuous on H due to Lemma 5.48. Equivalence of norms in finite dimensions implies that ∇d_γ is continuous on V_h. Last, continuity of $B_h^*\colon V_h \to Q_h$ and $\Pi_{ad,h}$ on Q_h lead to continuity of $\mathcal{F}$ from Q_h into itself. Furthermore, since $V_h \subset V$ and V_h is finite dimensional, $\mathcal{F}(Q_{ad,h})$ is contained in a compact subset of Q_h.

In summary, $\mathcal{F}\colon Q_{ad,h} \to Q_{ad,h}$ is a continuous mapping with $\mathcal{F}(Q_{ad,h})$ compact. Therefore, Schauder's fixed point theorem (see, e.g., [163, Theorem 2.A]) yields the existence of a fixed point $\mathcal{F}(q_m) = q_m$. Setting $u_m = (I+k_mA_h)^{-1}(u_{m-1}+k_mB_hq_m)$ proves existence of a solution to (5.82). □

With this preparation, we are ready to prove the lemma.

Proof of Lemma 5.47. Let $u_0 \in V_h$ be given as specified. By iteration of (5.82), we construct a function $u = u_{kh} \in X_k(V_h)$ with

$$\begin{aligned}(I+k_mA_h)u_m &= u_{m-1}+k_mB_hq_m,\\ q_m &= \Pi_{ad,h}\left\{-\frac{1}{\gamma}B_h^*\nabla d_\gamma(u_m)\right\}.\end{aligned}$$

Then, convexity of d_γ and the definition of u_{m+1} yield

$$\begin{aligned}d_\gamma(u_{m+1}) &\le d_\gamma(u_m) + (\nabla d_\gamma(u_{m+1}), u_{m+1}-u_m)\\ &= d_\gamma(u_m) + k_{m+1}\langle B_hq_{m+1} - A_hu_{m+1}, \nabla d_\gamma(u_{m+1})\rangle\\ &= d_\gamma(u_m) + k_{m+1}\langle B_hq_{m+1} - A_hP_U(u_{m+1}), \nabla d_\gamma(u_{m+1})\rangle\\ &\quad + k_{m+1}\langle A_h(P_U(u_{m+1})-u_{m+1}), \nabla d_\gamma(u_{m+1})\rangle.\end{aligned}$$

Lemma 5.48 and the Gårding inequality further imply

$$\begin{aligned}\langle A_h(P_U(u_{m+1})-u_{m+1}), \nabla d_\gamma(u_{m+1})\rangle &\le -\alpha_0\frac{\phi_\gamma'(d_U(u_{m+1}))}{d_U(u_{m+1})}\|u_{m+1}-P_U(u_{m+1})\|_V^2\\ &\quad + \omega_0\frac{\phi_\gamma'(d_U(u_{m+1}))}{d_U(u_{m+1})}\|u_{m+1}-P_U(u_{m+1})\|^2\\ &\le \omega_0\phi_\gamma'(d_U(u_{m+1}))d_U(u_{m+1})\\ &\le \omega_0 d_U(u_{m+1})\end{aligned}$$

since $\phi_\gamma'(d_U(u_{m+1})) \le 1$. Setting $\zeta_{m+1} = \nabla d_\gamma(u_{m+1})$ and employing (5.81) and (5.79) we infer that

$$\begin{aligned}d_\gamma(u_{m+1}) &\le d_\gamma(u_m) + k_{m+1}\left[h_h(u_{m+1},\zeta_{m+1}) + c\gamma + \omega_0 d_U(u_{m+1})\right]\\ &\le d_\gamma(u_m) + k_{m+1}\left[-h_0\|\zeta_{m+1}\| + c\gamma + \omega_0 d_U(u_{m+1})\right]\\ &\le d_\gamma(u_m) + k_{m+1}\left[-h_0\phi_\gamma'(d_U(u_{m+1})) + \omega_0 d_\gamma(u_{m+1}) + (c+\omega_0/2)\gamma\right],\end{aligned}$$

since $\|\zeta_m\| = \phi'_\gamma(d_U(u_m))$ and $d_U(\cdot) \leq d_\gamma(\cdot) + \gamma/2$. In the following, we replace $(c + \omega_0/2)$ by the generic constant c, just depending on Q_{ad} and ω_0. Thus, we have shown that

$$d_\gamma(u_{m+1}) \leq d_\gamma(u_m) + k_{m+1}\left[-h_0\phi'_\gamma(d_U(u_{m+1})) + \omega_0 d_\gamma(u_{m+1}) + c\gamma\right]. \tag{5.83}$$

Since $k_m < 1/\omega_0$, we have the fundamental inequality

$$(1 - \omega_0 k_m)^{-1} \leq (1 + 2\omega_0 k_m) \leq \exp(2\omega_0 k_m),$$

which will be useful below.

By induction, we now show the following estimate for all m:

$$d_\gamma(u_m) \leq f_\gamma(t_m) := \begin{cases} c\gamma \exp(2\omega_0 t_m) t_m & \text{for } h_0 = 0, \\ \max\{\,\gamma,\ d_\gamma(u_0) - (h_0/2 - c\gamma)t_m\,\} & \text{for } h_0 > 0. \end{cases} \tag{5.84}$$

Clearly, the inequality holds for $m = 0$ due to the assumption $d_U(u_0) \leq h_0/(2\omega_0)$. In the following, we assume that the estimate $d_U(u_k) \leq f_\gamma(t_k)$ is established for all $k \leq m$, and proceed separately for $h_0 = 0$ and $h_0 > 0$.

Case $h_0 = 0$: From (5.83), we obtain

$$d_\gamma(u_{m+1}) \leq d_\gamma(u_m) + k_{m+1}\left[\omega_0 d_\gamma(u_{m+1}) + c\gamma\right],$$

which is equivalent to

$$d_\gamma(u_{m+1}) \leq (1 - \omega_0 k_{m+1})^{-1}(d_\gamma(u_m) + k_{m+1} c\gamma).$$

Now, by using $(1 - \omega_0 k_{m+1})^{-1} \leq \exp(2\omega_0 k_{m+1})$, the assumption $d_\gamma(u_m) \leq c\gamma \exp(2\omega_0 t_m) t_m$, and $k_{m+1} c\gamma \leq c\gamma \exp(2\omega_0 t_m) k_{m+1}$, and $t_{m+1} = t_m + k_{m+1}$, we obtain the desired inequality for $m + 1$.

Case $h_0 > 0$: In the following, we will choose γ sufficiently small such that $c\gamma < h_0/2$. For each $m+1$, we have two situations: Either, it holds $d_U(u_{m+1}) \leq \gamma$, which means that $d_\gamma(u_{m+1}) \leq \gamma$, and we are done. Or, we have $d_U(u_{m+1}) > \gamma$, and we can use $\phi'_\gamma(d_U(u_{m+1})) = 1$. In this situation, we rewrite (5.83) as

$$d_\gamma(u_{m+1}) \leq (1 - \omega_0 k_{m+1})^{-1}(d_\gamma(u_m) + k_{m+1}(c\gamma - h_0)).$$

Now, by using $(1 - \omega_0 k_{m+1})^{-1} \leq (1 + 2k_{m+1}\omega_0)$, we obtain

$$\begin{aligned} d_\gamma(u_{m+1}) &\leq d_\gamma(u_m) + 2\omega_0 k_{m+1} d_\gamma(u_m) + k_{m+1}(1 + 2\omega_0 k_{m+1})(c\gamma - h_0) \\ &= d_\gamma(u_m) + 2\omega_0 k_{m+1}(d_\gamma(u_m) + c\gamma - h_0) + k_{m+1}(c\gamma - h_0) \\ &\leq d_\gamma(u_m) + k_{m+1} h_0/2 + k_{m+1}(c\gamma - h_0) \\ &= d_\gamma(u_m) + k_{m+1}(c\gamma - h_0/2), \end{aligned}$$

using the hypotheses $d_\gamma(u_m) \leq f_\gamma(t_m) \leq d_\gamma(u_0) \leq h_0/(4\omega_0)$ and that $(c\gamma - h_0) \leq 0$. Employing the induction hypothesis, we obtain the desired estimate (5.84) for $m + 1$.

Finally, since

$$d_U(u_m) \leq d_\gamma(u_m) + \frac{\gamma}{2} \leq \max\{\,\gamma,\ d_\gamma(u_0) - (h_0/2 - c\gamma)t_m\,\} + \frac{\gamma}{2},$$

and going to the limit $\gamma \to 0$ proves Lemma 5.47. □

5.6.2. Robust regularization and discretization error estimates

Based on the discrete strengthened Hamiltonian condition (5.79) we now prove discretization error estimates for the optimal times. For simplicity, let again $A = -\Delta$ equipped with homogeneous Dirichlet boundary conditions as in Section 5.1. Hence, we choose $V = H_0^1(\Omega)$, $H = L^2(\Omega)$, and $V^* = H^{-1}(\Omega)$. It is worth pointing out that the techniques can be used for fairly general elliptic operators; see also Remark 5.53.

Suppose that the regularity conditions concerning the temporal mesh $0 = t_0 < t_1 < \ldots < t_{M-1} < t_M = 1$ and the spatial mesh $\mathcal{T}_h$ from Section 5.2 are satisfied. Let $V_h \subset V$ denote the subspace of continuous and cellwise linear functions associated with the mesh $\mathcal{T}_h$. Concerning the discretization of the controls, we propose a semivariational control discretization. Precisely, the controls are not explicitly discretized in time but can be explicitly discretized in space. The reasons are as follows: First, for a discretization in time, we would require an estimate for the controls in L^2 in time. However, for bang-bang controls we cannot expect an optimal order estimate in L^2. Second, as we consider the piecewise constant discretization of the state and adjoint state equation, in view of the projection formula (5.44), if $\alpha > 0$ the optimal controls to the discrete problem are piecewise constant as well. Hence, the semivariational and the discretization by piecewise constant functions in time are equivalent; cf. also Corollary 5.19. Last, there is a technical reason. Since the proof of the following error estimate is based on cross-wise testing, we have to extend an optimal control from the continuous problem in a way such that the auxiliary control is feasible for the discrete problem. In a semivariational control discretization we avoid the necessity of projecting the auxiliary control onto the given temporal mesh.

Recall that $Q_h \subset Q$ is a subspace (not necessarily finite dimensional) and define the set of admissible controls $Q_{ad,h} = Q_h \cap Q$. Moreover, we set

$$Q_{ad,h}(0,1) \coloneqq \{\, q \in Q(0,1) \colon q(t) \in Q_{ad,h} \ \text{ a.a. } t \in (0,1) \,\}\,.$$

For any $\alpha \geq 0$, the regularized and discretized problem reads as

$$\inf_{\substack{\nu_{kh,\alpha} \in \mathbb{R}_+ \\ q_{kh,\alpha} \in Q_{ad,h}(0,1)}} j_\alpha(\nu_{kh,\alpha}, q_{kh,\alpha}) \quad \text{subject to} \quad g_{kh}(\nu_{kh,\alpha}, q_{kh,\alpha}) \leq 0. \tag{5.85}$$

As for the continuous result, the discrete strengthened Hamiltonian condition in particular implies that the discrete problems are well-posed; cf. Remark 2.15. Even better, we obtain the following robust estimate that eventually leads to discretization error estimates.

Lemma 5.51. *Let $(\bar\nu, \bar q)$ be a global solution to $(\hat P_0)$ and $\{(k,h,\alpha)\}$ be a sequence of positive mesh sizes and regularization parameters converging to zero. Moreover, suppose that the conditions of Lemmas 2.10 and 5.47 hold with $h_0 > 0$. Then there exist $\delta > 0$ and a sequence $\{(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha})\}$ of global solutions to problem (5.85) such that*

$$|\bar\nu - \bar\nu_{kh,\alpha}| \leq c\left(\alpha + |\log k|(k + h^2) + \|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})}\right)$$

for any $q_h \in Q_{ad,h}(0,1)$ with $\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} \leq \delta$.

Proof. The proof is based on the construction of feasible controls and cross-wise testing. We start by constructing a feasible point for the discrete problem. Here we have to take care of the fact that we cannot simply add time steps to the temporal mesh. Instead we divide the temporal mesh in two parts; see also Figure 5.9.

Step 1: Feasible control for the discrete problem. Let $(\bar\nu, \bar q)$ be a (global) solution to $(\hat P_0)$. Moreover, let $m' \in \{1, 2, \dots, M\}$ be arbitrary that will be determined in the course of the proof. Given $q_h, \breve q \in Q_{ad,h}(0,1)$, we construct a new control by

$$q_h'(t) = \begin{cases} q_h(t_{m'}^{-1}\, t) & \text{if } t \le t_{m'}, \\ \breve q((t - t_{m'})(1 - t_{m'})^{-1}) & \text{else.} \end{cases}$$

Let u_{kh}' be the piecewise constant function with values in V_h satisfying

$$\bar\nu t_{m'}^{-1} \int_{t_{m-1}}^{t_m} (\nabla u_{kh}', \nabla \varphi_h) + ([u_{kh}']_{m-1}, \varphi_h) = \bar\nu t_{m'}^{-1} \int_{t_{m-1}}^{t_m} \langle B q_h', \varphi_h\rangle \quad \text{for all } \varphi_h \in V_h,$$

for all $m = 1, 2, \dots, m'$, i.e. u_{kh}' is the discrete state on the temporal mesh $t_0 < t_1 < t_2 < \dots < t_{m'}$ with time transformation $\bar\nu$ and the control q_h transformed to $(0, t_{m'})$. Define $\nu = (t_{m'}^{-1} - 1)\bar\nu$ or, equivalently, $\bar\nu t_{m'}^{-1} = \bar\nu + \nu$. Then

$$i_{t_{m'}} S_{kh}(\bar\nu + \nu, q_h') = u_{kh}'(t_{m'}).$$

Moreover,

$$i_{t_{m'}} S(\bar\nu + \nu, q_h') = i_1 S(\bar\nu, q_h).$$

Thus, we have the estimate

$$\begin{aligned} d_U(u_{kh}'(t_{m'})) &\le d_U(i_1 S(\bar\nu, \bar q)) + \|i_1 S(\bar\nu, q_h) - i_1 S(\bar\nu, \bar q)\|_{L^2} \\ &\quad + \|i_{t_{m'}} S_{kh}(\bar\nu + \nu, q_h') - i_{t_{m'}} S(\bar\nu + \nu, q_h')\|_{L^2} \\ &\le c\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} + \delta(k, h, m'), \end{aligned}$$

where we have used linearity of the solution operator (for fixed $\bar\nu$) and $\delta(k, h, m')$ denotes the discretization error to be discussed later. Now Lemma 5.47 guarantees the existence of a control $\breve q_{kh}$ such that the corresponding state $\breve u_{kh}$ as defined in Lemma 5.47 with initial state $u_{kh}'(t_{m'})$ satisfies

$$d_U(\breve u_{kh}(\breve t_m)) \le \max\{0, c\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} + \delta(k, h, m') - (h_0/2)\breve t_m\} \tag{5.86}$$

for $m = 1, 2, \dots$ on an arbitrary temporal mesh $0 = \breve t_0 < \breve t_1 < \dots$, because $\omega_0 = 0$ due to homogeneous Dirichlet boundary conditions. Let $\breve\nu > 0$ such that

$$c\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} + \delta(k, h, m') - (h_0/2)\breve\nu = 0.$$

Since $\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} \le \delta$, we obtain the upper bound

$$\breve\nu \le \frac{2}{h_0}\left(c\delta + \delta(k, h, m')\right). \tag{5.87}$$

For the following considerations we assume that $\breve\nu \le \bar\nu$ for δ, k, and h sufficiently small and postpone the rigorous proof of this estimate to step 2. Take $m' \in \{1, 2, \dots, M-1\}$ such that

$$\frac{\bar\nu}{\bar\nu + \breve\nu} = t_{m'} + \tau k_{m'+1}$$

with $t_{m'}$ from the reference time mesh and some $\tau \in [0, 1)$. Note that the case $m' = M$ is impossible due to $\breve\nu > 0$. We will argue that $\breve\nu$ can be slightly increased to some ν such that $\bar\nu/(\bar\nu + \nu) = t_{m'}$ as well as

$$\nu \le c\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} + ck + \delta(k, h, m')$$

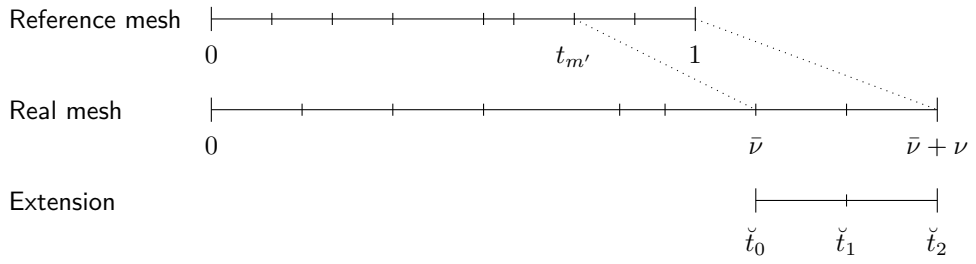

Figure 5.9.: Temporal meshes in the proof of Lemma 5.51.

(with a possibly different constant c) hold. A simple calculation reveals

$$\nu = \frac{1-t_{m'}}{t_{m'}}\bar\nu = \breve\nu + \tau k_{m'+1}\frac{\bar\nu+\breve\nu}{t_{m'}}.$$

Moreover, we have the lower bound

$$1 > t_{m'} = \frac{\bar\nu}{\bar\nu+\nu} - \tau k_{m'+1} \geq \frac{\bar\nu}{\bar\nu+\breve\nu} - k \geq \frac{1}{2} - k.$$

Since $k \leq 1/4$ we have $t_{m'} \geq 1/4$. Thus $\nu \leq \breve\nu + ck(\bar\nu+\breve\nu)$ and ν satisfies the required properties. Recall that for the constructed control $\breve q_{kh}$ from Lemma 5.47, we have the freedom to choose the temporal mesh. We set

$$\breve t_j = (\bar\nu+\nu)\sum_{i=1}^{j} k_{m'+i}, \quad j = 1,2,\dots,M-m';$$

see also Figure 5.9. This gives $\breve t_{M-m'} = \nu$ and $d_U(\breve u_{kh}(\breve t_{M-m'})) = 0$ due to estimate (5.86). We define a new control as above by

$$q_h'(t) = \begin{cases} q_h(t\,(\bar\nu+\nu)/\bar\nu) & \text{if } t \leq \bar\nu/(\bar\nu+\nu), \\ \breve q_{kh}(t\,(\bar\nu+\nu)/\nu - \bar\nu/\nu) & \text{else.} \end{cases}$$

Now since the new pair $(\bar\nu+\nu, q_h')$ is feasible for (5.85), we obtain

$$\begin{aligned} \bar\nu_{kh,\alpha} &\leq j_\alpha(\bar\nu_{kh,\alpha}, \bar q_{kh,\alpha}) \leq j_\alpha(\bar\nu+\nu, q_h') \\ &\leq \bar\nu + c\|B\,(q_h - \bar q)\|_{L^2(I;H^{-1})} + ck + \delta(k,h,m') + c\alpha. \end{aligned} \tag{5.88}$$

In particular, since there exist feasible controls for (5.85), we deduce that there is (at least) one optimal control for the discrete problem.

Step 2: Discretization error $\delta(k,h,m')$. We are left with the task of determining the discretization error $\delta(k,h,m')$. Similar as in Lemma A.39, we can prove the error estimate

$$\begin{aligned} \delta(k,h,m') &= \|i_{t_{m'}} S_{kh}(\bar\nu+\nu, q_h') - i_{t_{m'}} S(\bar\nu+\nu, q_h')\|_{L^2} \\ &\leq c|\log k|(k+h^2)(1+\bar\nu+\nu) \\ &\leq c|\log k|(k+h^2)(1+\delta+\delta(k,h,m')) \end{aligned}$$

for all $m' \in \{1,2,\dots M\}$, where we have used the upper bound (5.87) for $\breve\nu$. Note that the constant is independent of q_h and $\breve q$ due to boundedness of $Q_{ad,h}(0,1)$. In particular,

the above estimate implies that $\delta(k, h, m')$ is arbitrary small for k and h sufficiently small. Hence, $\breve{\nu} \le \bar{\nu}$ is guaranteed for δ, k, and h small. Finally, from (5.88) we infer that

$$\bar{\nu}_{kh,\alpha} - \bar{\nu} \le c \left(\alpha + \|B\,(q_h - \bar{q})\|_{L^2(I;H^{-1})} + |\log k|(k + h^2) \right).$$

Step 3: Feasible control for the continuous problem. For the reverse inequality we proceed in a similar way. Let $(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})$ be a solution to (5.85) and set $\bar{u}_{kh,\alpha} = S(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})$. According to the error estimate Lemma A.39 we get

$$\begin{aligned} d_U(\bar{u}_{kh,\alpha}(1)) &\le d_U(i_1 S_{kh}(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})) + \|i_1 S(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha}) - i_1 S_{kh}(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})\|_{L^2} \\ &\le c|\log k|(k + h^2). \end{aligned}$$

By means of Lemma 2.10 there exists an admissible control $\breve{q}$ such that the corresponding solution to the state equation $\breve{u}$ satisfies

$$d_U(\breve{u}(t)) \le \max\{\, 0, c|\log k|(k + h^2) - h_0 t \,\}$$

for all $t \ge 0$. Hence, setting $\nu = c|\log k|(k + h^2)/h_0$ and

$$q'(t) = \begin{cases} \bar{q}_{kh,\alpha}(t\,(\bar{\nu}_{kh,\alpha} + \nu)/\bar{\nu}_{kh,\alpha}) & \text{if } t \le \bar{\nu}_{kh,\alpha}/(\bar{\nu}_{kh,\alpha} + \nu), \\ \breve{q}(t\,(\bar{\nu}_{kh,\alpha} + \nu)/\nu - \bar{\nu}_{kh,\alpha}/\nu) & \text{else,} \end{cases}$$

the pair $(\bar{\nu}_{kh,\alpha} + \nu, q')$ is feasible for $(\hat{P}_\alpha)$, and we obtain

$$\bar{\nu} \le \bar{\nu}_{kh,\alpha} + \nu = \bar{\nu}_{kh,\alpha} + c|\log k|(k + h^2). \tag{5.89}$$

Combination of the estimates (5.88) and (5.89) proves the assertion. □

Similar as in Section 2.4, the assumptions of Lemma 5.51 can be explicitly verified in concrete situations. For the particular case that $A = -\Delta$, we can state the following sufficient condition.

Proposition 5.52. *Suppose that $U = \mathcal{B}_{\delta_0}(u_d)$ with $u_d \in V_h \subset V$ and $\delta_0 > 0$. If there exists a control $\breve{q}_h \in Q_{ad,h} \subset Q_{ad}$ such that*

$$\|B\breve{q}_h + \Delta_h u_d\|_{H^{-1}} < \frac{c_P^2}{1 + c_P^2}\,\delta_0, \tag{5.90}$$

with c_P denoting the Poincaré constant, then the assumptions of Lemmas 2.10 and 5.47 are satisfied.

Proof. We first note that in our case $\alpha_0 = c_P^2/(1 + c_P^2)$. The assumptions of Lemma 2.10 are verified in Proposition 2.37 provided that (5.90) holds. Since $u_d \in V_h$, for the verification of the discrete strengthened Hamiltonian condition (5.79) we can use similar arguments as in Proposition 2.37. □

Remark 5.53. (i) The techniques used in the proof of Lemma 5.51 can also be applied to obtain error estimates for Neumann boundary control and for general autonomous parabolic equations, if discretization error estimates for the state equation are available. The only requirements are the strengthened Hamiltonian conditions in Lemmas 2.10 and 5.47 that can be verified to hold for fairly general operators and other control situations. In particular, the classical problem with $U = \mathcal{B}_{\delta_0}(0)$ and distributed control on a subset of the spatial domain is included in our setting; see Proposition 5.52.

(ii) Note that all conditions of Lemma 5.51 except for 'k, h, and δ sufficiently small' can be verified a priori, in contrast to Proposition 5.28 that relied on the structural assumption that can be hardly checked a priori.

(iii) Lemma 5.51 generalizes the convergence result of [87, Theorem 4] to more general terminal sets than the L^2-ball centered at zero. The proof of [87, Theorem 4] essentially relies on the fact that Δ generates an exponentially stable semigroup on L^2 to construct feasible controls; cf. also Proposition A.21. In our framework this is hidden in Lemmas 2.10 and 5.47.

(iv) The case with $\alpha = 0$ is included in Lemma 5.51.

(v) Lemma 5.47 requires the stability of P_U in V_h. For the prototypical example $U = \mathcal{B}_{\delta_0}(u_d)$ this is equivalent to $u_d \in V_h$. The proof of Lemma 5.51 can be easily modified if $u_d \in V \setminus V_h$, by using the projected desired state $\Pi_h u_d$ with a corresponding error estimate for u_d.

While Lemma 5.51 potentially provides optimal error estimates for the optimal times, we cannot show strong convergence of the controls without any additional assumption as in Section 3.3.2 or Section 5.5; cf. also [37, 47, 148, 162]. By standard arguments, merely weak convergence to a control $q^* \in Q_{ad}(0,1)$ that is also optimal for $(\hat{P}_0)$ is guaranteed.

Proposition 5.54. *Adapt the assumptions of Lemma 5.51. If there is a sequence $(q_h)_{h>0}$, $q_h \in Q_{ad,h}(0,1)$, such that $\|B(q_h - \bar{q})\|_{L^2(I;H^{-1})} \to 0$ as $h \to 0$. Then $\bar{\nu}_{kh,\alpha} \to \bar{\nu}$ and $\bar{q}_{kh,\alpha} \rightharpoonup q^*$ in $L^r(I;Q)$ for any $r \in (1,\infty)$. Moreover, the pair $(\bar{\nu}, q^*)$ is optimal for $(\hat{P}_0)$.*

Proof. First, Lemma 5.51 and the supposition imply $\bar{\nu}_{kh,\alpha} \to \bar{\nu}$. From uniform boundedness of $\bar{q}_{kh,\alpha} \in Q_{ad,h}(0,1) \subset Q_{ad}(0,1)$ we conclude the existence of a subsequence converging weakly to some $q^* \in Q_{ad}(0,1)$ in $L^r(I;Q)$ for $r > 2$. Feasibility of q^* for $(\hat{P}_0)$ follows from

$$\begin{aligned} d_U(i_1 S(\bar{\nu}, q^*)) \le{}& d_U(i_1 S_{kh}(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})) + c\|i_1 S(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha}) - i_1 S_{kh}(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})\|_H \\ &+ c\|i_1 S(\bar{\nu}, q^*) - i_1 S(\bar{\nu}_{kh,\alpha}, \bar{q}_{kh,\alpha})\|_H, \end{aligned}$$

the error estimate Lemma A.39, and complete continuity of the control-to-state mapping Proposition A.20. Due to $\alpha = 0$, the pair $(\bar{\nu}, q^*)$ is also optimal for $(\hat{P}_0)$. □

However, under additional assumptions, we can verify strong convergence of the controls for the unregularized problems; cf. the proof of [100, Theorem 3.1]. Without restriction suppose that the control bounds are symmetric (i.e. $-q_a = q_b$) by adding a fixed right-hand side to the state equation which does not affect the preceding results. If $\bar{q}$ is bang-bang, which in particular implies uniqueness of $\bar{q}$ (cf. Proposition 3.26), then we automatically have $q^* = \bar{q}$. Hence,

$$\begin{aligned} \|\bar{q}_{kh,0} - \bar{q}\|^2_{L^2(I\times\omega)} &= \|\bar{q}_{kh,0}\|^2_{L^2(I\times\omega)} - 2(\bar{q}_{kh,0}, \bar{q})_{L^2(I\times\omega)} + \|\bar{q}\|^2_{L^2(I\times\omega)} \\ &\le 2|I\times\omega|^2 q_b^2 - 2(\bar{q}_{kh,0}, \bar{q})_{L^2(I\times\omega)}. \end{aligned}$$

Thus, weak convergence of $\bar{q}_{kh,0}$ to $\bar{q}$ implies strong convergence in $L^2(I \times \omega)$.

Variational discretization of controls

Since Q_h was not assumed to be finite dimensional, we can still take $Q_h = Q$ and directly obtain an error estimate for the variational control discretization as proposed in [78] for elliptic equations, cf. also [118] for parabolic equations.

Theorem 5.55 (Variational discretization)**.** *Let the assumptions of Lemma 5.51 hold and suppose the variational control discretization, i.e. $Q_h = Q$. Then there is a constant $c > 0$ not depending on k, h, α, and $\bar{\nu}_{kh,\alpha}$ such that*

$$|\bar{\nu} - \bar{\nu}_{kh,\alpha}| \le c \left(\alpha + |\log k|(k + h^2) \right).$$

Proof. We apply Lemma 5.51 with $q_h = \bar{q}$ that is allowed due to $Q_h = Q$. □

In case of purely time-dependent controls and if $\alpha > 0$, the variational control discretization and the discretization by piecewise constant functions in time are equivalent due to the projection formula (5.63). Whereas if $\alpha = 0$, the optimal control $\bar{q}_{kh,0}$ to $(\hat{P}_{kh,0})$ is not necessarily piecewise constant with the same time mesh. Defining a new control $\Pi_k \bar{q}_{kh,0}$ that is the projection of $\bar{q}_{kh,0}$ onto the space of piecewise constant functions in time, we observe

$$(B\bar{q}_{kh,0}, \varphi_{kh})_{L^2(I;L^2)} = (B\Pi_k \bar{q}_{kh,0}, \varphi_{kh})_{L^2(I;L^2)} \quad \text{for all } \varphi_{kh} \in X_{k,h},$$

i.e. the controls $\bar{q}_{kh,0}$ and $\Pi_k \bar{q}_{kh,0}$ have the same associated discrete state. Hence, in case $\alpha = 0$, we can always find a feasible control that belongs to the discrete space of controls with the same objective function value. In contrast to the comment after Theorem 5.31, we do not require any assumption on the set of switching points, since convergence of the controls cannot be guaranteed anyway by the techniques in this section. Based on this observation we obtain the following corollary.

Corollary 5.56 (Parameter control)**.** *Let the assumptions of Lemma 5.51 hold, suppose that ω is discrete, and choose the piecewise constant discrete control space*

$$Q_\sigma(0,1) = \left\{ v \in Q(0,1) \colon v|_{I_m} \in \mathcal{P}_0(I_m; \mathbb{R}^{N_c}),\ m = 1, 2, \dots, M \right\}.$$

Then there is a constant $c > 0$ not depending on k, h, $\bar{\nu}_{kh}$, and $\bar{q}_{kh}$ such that

$$|\bar{\nu} - \bar{\nu}_{kh,\alpha}| \le c \left(\alpha + |\log k|(k + h^2) \right).$$

Cellwise constant control approximation

Last, we consider the explicit discretization of controls by cellwise constant functions. Note that we still do not discretize the controls explicitly in time. The discrete space of controls is defined as follows

$$Q_h = \{ v \in Q \colon v|_K \in \mathcal{P}_0(K) \text{ for all } K \in \mathcal{T}_h^\omega \}.$$

On any $K \in \mathcal{T}_h$ we define the piecewise constant projection $\Pi_{h,0}$ via

$$(\Pi_{h,0} v)(t,x) = \frac{1}{|K|} \int_K v(t,\xi) \, \mathrm{d}\xi, \quad x \in K.$$

Moreover, for each $t \in [0,1]$ we set

$$\mathcal{S}_{h,t} \coloneqq \mathcal{T}_h^\omega \setminus \{K \in \mathcal{T}_h^\omega \colon \bar{q}(t)|_K \equiv q_a \text{ or } \bar{q}(t)|_K \equiv q_b\}.$$

Under a structural assumption on the set with switching we can derive the following discretization error estimate; cf. Theorem 5.46.

Theorem 5.57 (Cellwise constant controls)**.** *Let the assumptions of Lemma 5.51 hold and suppose the cellwise constant control discretization. Moreover, suppose that* (5.76) *holds, i.e. there are functions $\delta_h \in L^1(I)$, $h > 0$, and a constant $c > 0$ such that*

$$\sum_{K \in \mathcal{S}_{h,t}} |K| \leq \delta_h(t), \quad \text{a.e. } t \in [0,1], \quad h > 0,$$

and $\|\delta_h\|_{L^1(I)} \leq ch$ for all $h > 0$. Then there is a constant $c > 0$ not depending on k, h, α, and $\bar{\nu}_{kh,\alpha}$ such that

$$|\bar{\nu} - \bar{\nu}_{kh,\alpha}| \leq c\left(\alpha + |\log k|(k + h^{3/2})\right).$$

Proof. According to the supposition on $\mathcal{S}_{h,t}$ and Proposition 5.44, we have

$$\|\Pi_{h,0}\bar{q} - \bar{q}\|_{L^2(I;H^{-1})} \leq ch^{3/2}.$$

Since $\|\Pi_{h,0}\bar{q} - \bar{q}\|_{L^2(I;H^{-1})} \to 0$ as $h \to 0$, we can apply Lemma 5.51 with $q_h = \Pi_{h,0}\bar{q}$, which yields the desired estimate. □

5.7. Numerical examples for bang-bang controls

We continue the numerical examples of Section 5.4 for the case of bang-bang controls. The aim is again the numerical verification of the theoretically obtained error estimates for regularization and discretization. In all examples, we consider the operator $-c\Delta$ with $c = 0.03$ instead of $-\Delta$, which clearly does not effect the results of this chapter.

5.7.1. Example with purely time-dependent control

We take again the example from Section 5.4.2 with purely time-dependent controls for fixed spatially dependent functions. Let

$$\begin{aligned} \Omega &= (0,1)^2, \quad \omega_1 = (0,0.5) \times (0,1), \quad \omega_2 = (0.5,1) \times (0,0.5), \\ B &\colon \mathbb{R}^2 \to L^2(\Omega), \quad Bq = q_1 \mathbb{1}_{\omega_1} + q_2 \mathbb{1}_{\omega_2}, \\ G(u) &= \tfrac{1}{2}\|u - u_d\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad u_d(x) = 0, \quad \delta_0 = \tfrac{1}{10}, \\ Q_{ad}(0,1) &= \{q \in L^2((0,1);\mathbb{R}^2) \colon -1.5 \leq q \leq 0\}, \quad u_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2^3), \end{aligned}$$

where $\mathbb{1}_{\omega_1}$ and $\mathbb{1}_{\omega_2}$ denote the characteristic functions on ω_1 and ω_2. The spatial mesh is chosen such that the boundaries of ω_1 and ω_2 coincide with edges of the mesh, which ensures that B can be easily implemented.

Since the exact solution is unknown, we calculate a numerical solution on a sufficiently fine grid for a small regularization parameter. In accordance with Theorem 5.31 (provided that (3.37) holds with $\kappa = 1$), we observe linear convergence in all variables with respect to α up to a threshold, where the error due to discretization dominates to total error; cf. Figure 5.10. Concerning the discretization error, in Figure 5.11 we observe linear order in k for the temporal discretization and quadratic order in h for the spatial discretization.

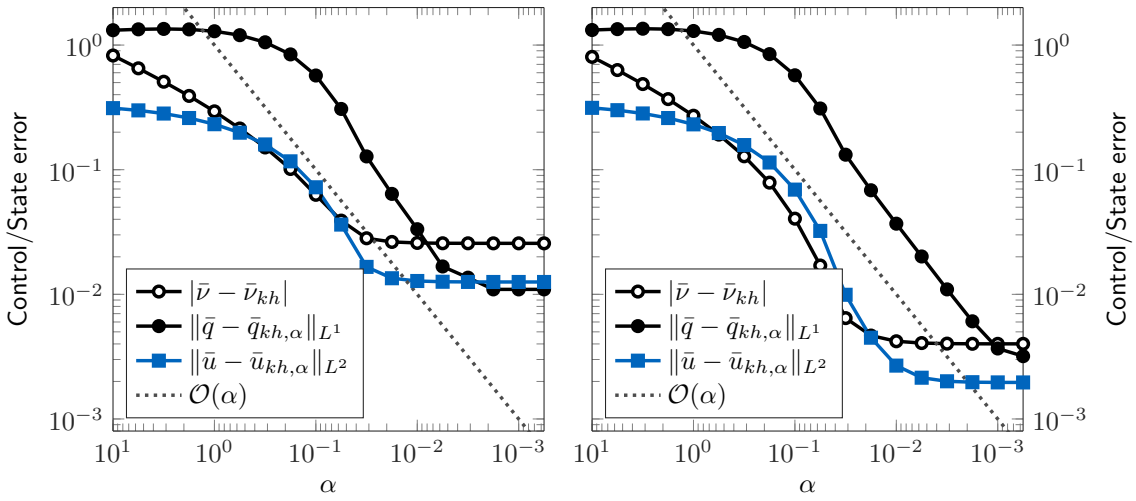

Figure 5.10.: Discretization error for Example 5.7.1 with variational control discretization and refinement of the regularization parameter for $N = 289$ nodes and $M = 80$ time steps (left) and $N = 4225$ nodes and $M = 320$ time steps (right).

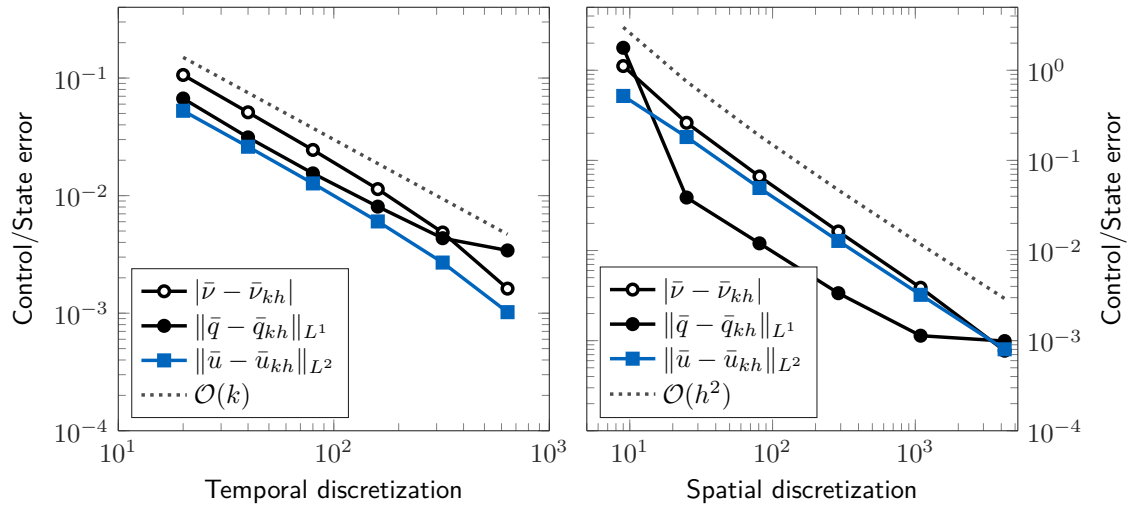

Figure 5.11.: Discretization error for Example 5.7.1 with variational control discretization and refinement of the time interval for $N = 4225$ nodes (left) and refinement of the spatial discretization for $M = 320$ time steps (right) for $\alpha = 10^{-3}$. The reference solution is calculated for $\alpha = 0$ using the algorithmic approach from Section 4.2.

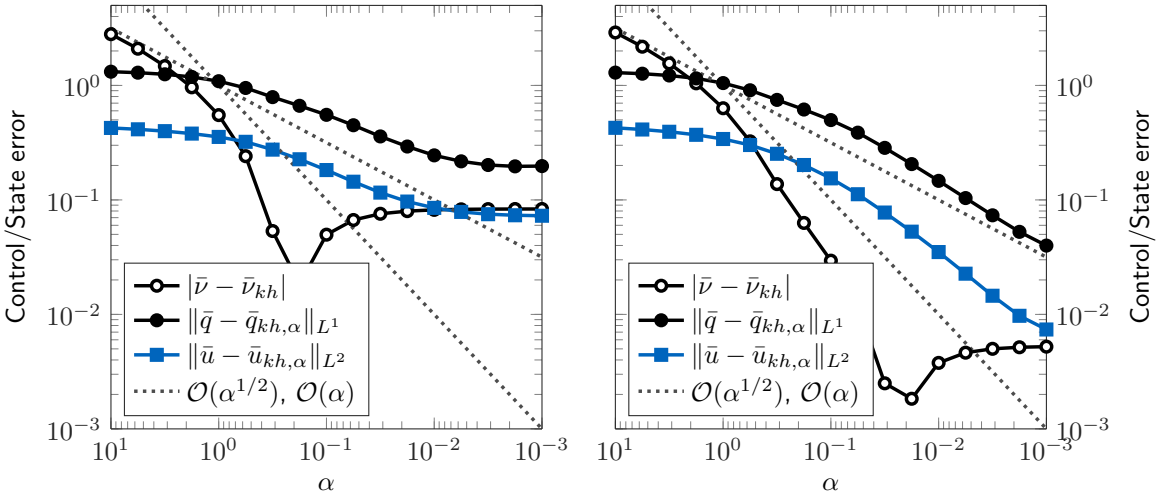

Figure 5.12.: Discretization error for Example 5.7.2 with cellwise constant control discretization and refinement of the regularization parameter for $N = 81$ nodes and $M = 80$ time steps (left) and $N = 1089$ nodes and $M = 320$ time steps (right).

5.7.2. Example with distributed control on subdomain

Next, we consider the example from Section 5.4.3 with distributed control on a subset of the domain. As before we compare to a reference solution obtained numerically on a fine mesh for a small regularization parameter. The problem data is

$$\begin{aligned} \Omega = (0,1)^2, \quad \omega = (0,0.75)^2, \quad \delta_0 = \tfrac{1}{10}, \\ G(u) = \tfrac{1}{2}\|u - u_d\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad u_d(x) = -2\min\{x_1, 1-x_1, x_2, 1-x_2\}, \\ Q_{ad}(0,1) = \{q \in L^2(I\times\omega) \colon -5 \le q \le 0\}, \\ u_0(x) = 4\sin(\pi x_1^2)\sin(\pi x_2)^3. \end{aligned}$$

The mesh is chosen such that the boundary of the control domain coincides with edges of the spatial mesh. We use cellwise constant functions for the discretization of the control variable. Since ω does not have a strict distance to the boundary of the spatial domain, this example does not fit into the setting considered in Section 5.5.5. However, we expect that one can show similar results for the peculiar problem on the unit square.

From Figure 5.12 we approximately deduce the convergence rate 1/2 with respect to α for the control variable measured in $L^1(I\times\omega)$. Moreover, we observe approximately order 1/2 for the error due to temporal discretization and linear order for the error due to spatial discretization of the control variable; see Figure 5.13. As already observed in Section 5.4.3, the error for the terminal time decreases at the full rate $k + h^2$. Taking into account the structure of the adjoint state depicted in Figure 5.15, it seems that the structural assumption (3.37) is not satisfied with $\kappa = 1$ in this case. For this reason, Theorem 5.46 does not guarantee the rate $\alpha + k + h$ for the control variable. Numerically evaluating the condition (3.37) seems to confirm the hypothesis. In Example 5.7.1 we observe linear decrease while in Example 5.7.2 it is hard to determine the rate of decrease; see Figure 5.14. Nevertheless, the decrease for Example 5.7.2 is for sure less than linear. Note that for the terminal time we only require (3.37) to hold for some $\kappa > 0$ to obtain the rate α due to Lemma 5.26.

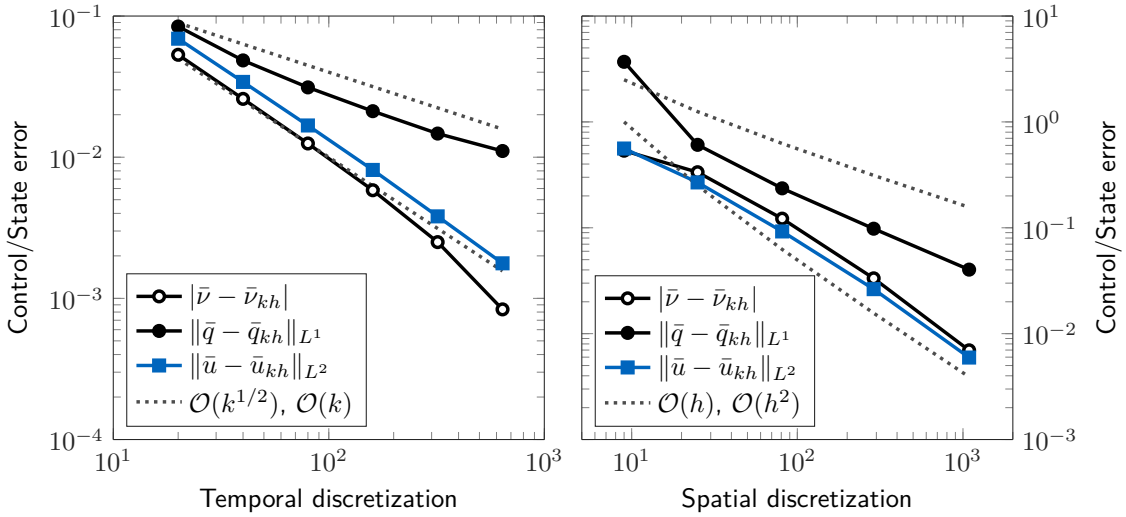

Figure 5.13.: Discretization error for Example 5.7.2 with cellwise constant control discretization and refinement of the time interval for $N = 1089$ nodes (left) and refinement of the spatial discretization for $M = 160$ time steps (right) for $\alpha = 10^{-4}$. The reference solution is calculated for $\alpha = 10^{-5}$.

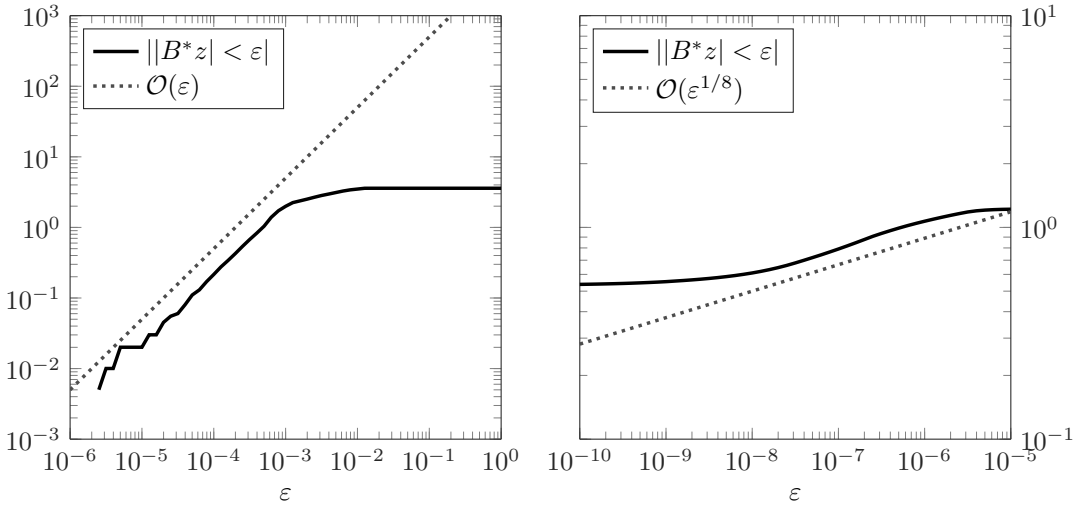

Figure 5.14.: Numerical verification of structural assumption on the adjoint state (3.37) for Example 5.7.1 (left) and Example 5.7.2 (right).

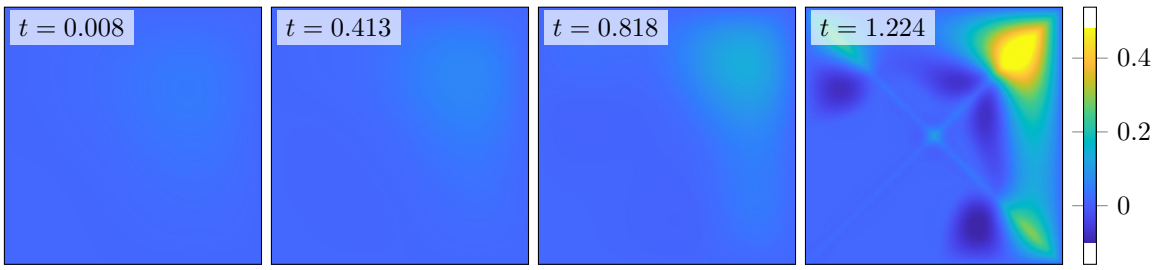

Figure 5.15.: Snapshots of adjoint state of Example 5.7.2 for $\alpha = 10^{-5}$.

5.7.3. Example with distributed control on domain

Last, let us consider again the example from Section 4.2.7 with distributed control on the whole domain. The main difference to the preceding example is that we take $u_d \equiv 0$ to ease the computation. As before we compare to a reference solution obtained numerically on a fine mesh for a small regularization parameter. The problem data is

$$\begin{aligned} \Omega = (0,1)^2 = \omega, \quad \delta_0 = \tfrac{1}{10}, \\ G(u) = \tfrac{1}{2}\|u - u_d\|_{L^2}^2 - \tfrac{1}{2}\delta_0^2, \quad u_d(x) = 0, \\ Q_{ad}(0,1) = \{q \in L^2(I \times \omega) \colon -2 \le q \le 1\}, \\ u_0(x) = 10 \sin(\pi x_1^2) \sin(\pi x_2)^3. \end{aligned}$$

The control variable is discretized by cellwise constant functions in space. We observe approximately order $1/2$ of convergence with respect to α for the control variable in $L^1(I \times \omega)$; see Figure 5.16. Moreover, in Figure 5.17 we observe order $1/2$ of convergence with respect to k and linear order with respect to h. Concerning the structural assumption (3.37), from Figure 5.18 we numerically find $\kappa \approx 1/2$. Hence, the convergence rate for the regularization error seems to be in accordance with the theory.

Additionally, from the numerical verification Figure 5.18 (right) it seems that the assumption (5.76) used in the proof for cellwise constant control discretization is fulfilled. Hence, by virtue of Theorem 5.46, we can expect the overall convergence rate $k^{1/2} + h^{1/2}$ for the control variable. While the convergence rate for the temporal discretization is in accordance with the theory, for the spatial discretization we observe better order of convergence. In the theory there are two reasons for the limited convergence rate in h: First, we expect that the estimate in (5.61) can be improved in the case $\kappa < 1$; see also the comment after the proof of Proposition 5.28. Second, in the numerical examples we always observe the full rate $k + h^2$ for ν. However, Lemma 5.26 guarantees the rate $k + h$, only. As this quantity directly enters into the estimate of Proposition 5.28, the suboptimal rate for ν limits the convergence rate for q. It is worth mentioning that from Theorem 5.57 we could expect the rate $k + h^{3/2}$ for the optimal times, which is better but still not optimal.

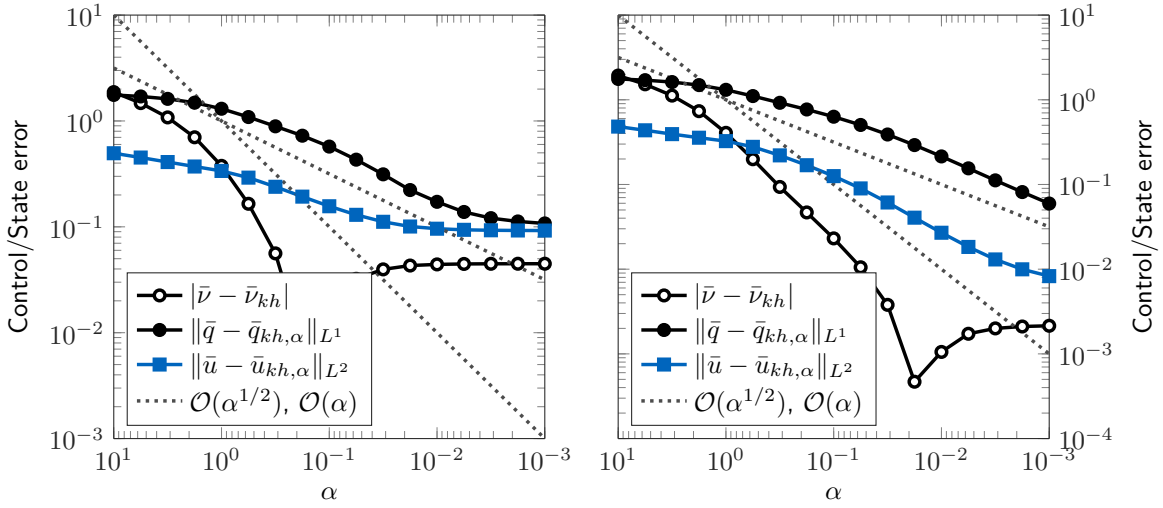

Figure 5.16.: Discretization error for Example 5.7.3 with cellwise constant control discretization and refinement of the regularization parameter for $N = 81$ nodes and $M = 80$ time steps (left) and $N = 1089$ nodes and $M = 320$ time steps (right).

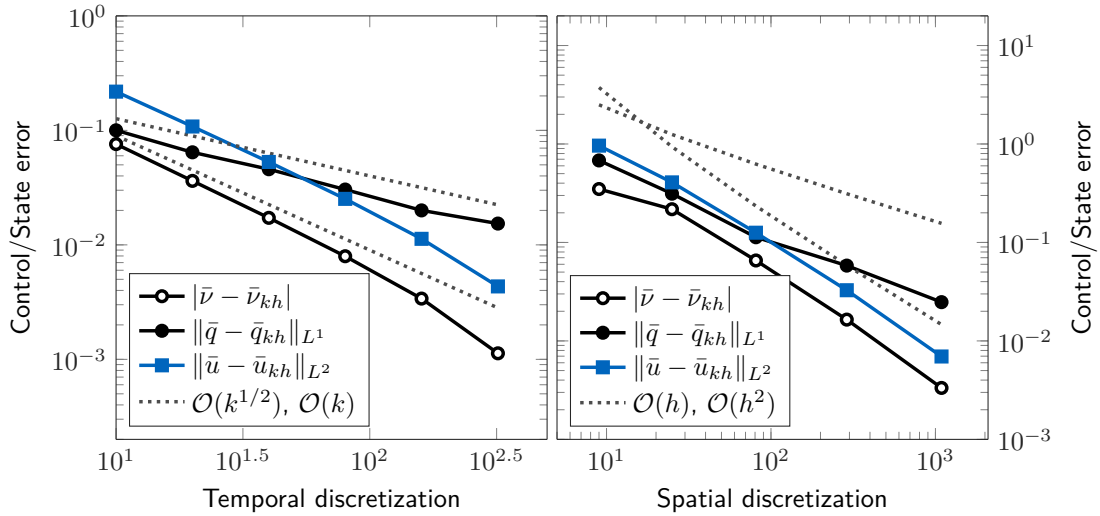

Figure 5.17.: Discretization error for Example 5.7.3 with cellwise constant control discretization and refinement of the time interval for $N = 289$ nodes (left) and refinement of the spatial discretization for $M = 160$ time steps (right) for $\alpha = 10^{-4}$. The reference solution is calculated for $\alpha = 0$ using the algorithmic approach from Section 4.2.

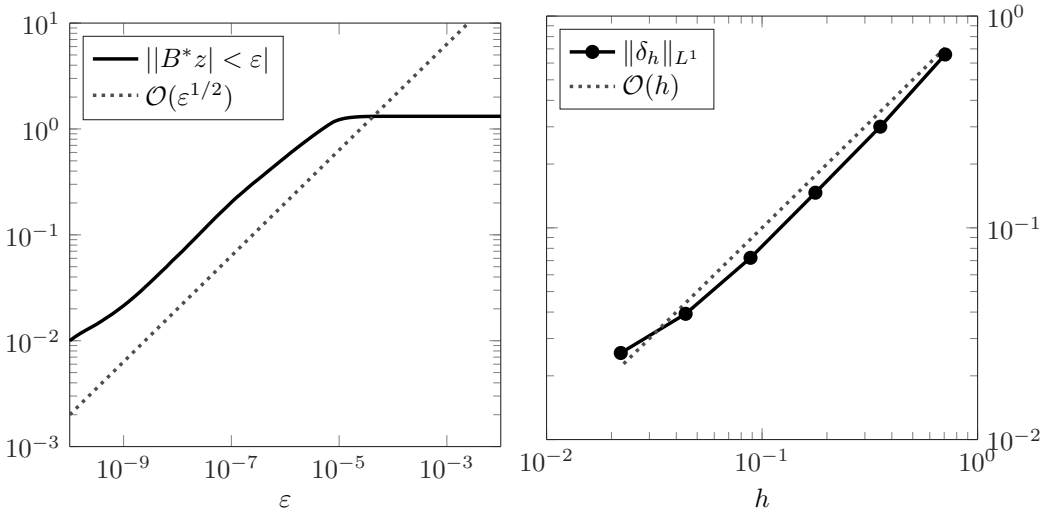

Figure 5.18.: Numerical verification of structural assumption on adjoint state (3.37) (left) and assumption (5.76) (right) for Example 5.7.3. Quantities estimated from numerical solution for $\alpha = 0$, $N = 4225$ nodes, and $M = 160$ time steps that was calculated by the algorithmic approach from Section 4.2 with $\varepsilon_{\mathsf{tol}} = 10^{-8}$.

6. Outlook and perspectives

In this thesis we considered the numerical analysis and algorithmic solution of time-optimal control problems. Especially we focused on discretizing both the temporal and the spatial component of the involved partial differential equation by means of the finite element method. There are of course many open questions that could not be tackled in this thesis leading to several possible directions for future investigations on this research topic. As already mentioned in the introduction, different objective functionals than the here regarded L^2-norm of the controls could be of interest. Other norms may be more appropriate to represent inherent control costs or may lead to a more natural regularization strategy. For example one could choose the L^1-norm of the control in the objective; see [29] for corresponding second order optimality conditions.

Since many processes in natural sciences or engineering require nonlinear models and also time-optimal control formulations are of interest, this gives rise to the numerical analysis of time-optimal control problems subject to nonlinear state equations. For example quasilinear parabolic partial differential equations that arise in, e.g., heat conduction problems in electrical engineering [90] and semiconductors [143] are important in applications. First and second order optimality conditions for an optimal control problem on a fixed time horizon without state constraints have been analyzed in [16] for quasilinear elliptic operators of divergence type; see also [27]. It would be interesting to combine these results with those of this thesis for the numerical analysis of time-optimal control problems subject to nonlinear state equations.

Moreover, pointwise constraints for the state are important, both theoretically and practically; see, e.g., [119]. To extend the results presented in this thesis to pointwise state constrained optimal control problems, one could rely on recent advances in the regularity theory for parabolic partial differential equations. In particular, Hölder continuity in time and space can be guaranteed for very general spatial domains and nonsmooth right-hand sides; see [49]. Corresponding results for discrete solutions for finite element discretizations have been proved recently; see, e.g., [102, 103]. However, as we exploited the regularity of the Lagrange multiplier in several arguments in this thesis, a direct extension of the presented results to pointwise state constraints is not straightforward.

Even though in many applications regular controls are needed, there is an independent interest in bang-bang control problems. For its efficient algorithmic solution in the context of time-optimal control, adaptive mesh refinement strategies should be considered. This seems to be particularly important, as in the numerical examples we observed that the control variable is relatively constant for large parts of the time horizon but tends to vary towards its end. Hence, adaptive algorithms may pay of and help to reduce the computational cost.

Furthermore, in the context of bang-bang controls, we proposed an algorithm based on an equivalent reformulation of the optimization problem; see Section 4.2. Concretely, we search for a root of a certain value function. To evaluate the value function, we have to solve convex and control constrained optimization problems. Different methods for the solution

of the inner problem have been considered. In particular, for the solution by means of the conditional gradient method equipped with an acceleration strategy we obtained promising results numerically. Hence, it would be desirable to investigate conditions under which fast convergence of the accelerated conditional gradient method in infinite dimensional spaces can be guaranteed. This would be of independent interest, as the pure conditional gradient method exhibits in general slow convergence, but higher order methods are not be applicable in every situation.

A. Appendix

A.1. Interpolation spaces

We collect several well-known properties of interpolation spaces. For further information we refer to the monographs [13, 110, 146]. 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. This section is part of a joint work with Dominik Hafemeyer.

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

$$K(t, u, X, Y) = \inf_{x\in X, u-x\in Y} \left[\|x\|_X + t\|u - x\|_Y \right].$$

For $\tau \in (0,1)$ and $1 \le s \le \infty$ we define the real interpolation space

$$(X,Y)_{\tau,s} \coloneqq \{\, u \in X + Y : t \mapsto t^{-\tau} K(t,u,X,Y) \in L^s_*(\mathbb{R}_+) \,\}$$

equipped with the norm

$$\|u\|_{\tau,s} = \|t^{-\tau} K(t,u,X,Y)\|_{L^s_*(\mathbb{R}_+)};$$

see, e.g., [110, 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 notes the norm of the real interpolation space is always defined by the K-functional as above, if not indicated otherwise.

Proposition A.1. *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 A.2. If $s \in (s_0, \infty)$ for some $s_0 > 1$ and $\tau = 1 - 1/s$, then the constant from Proposition A.1 remains bounded for large s. This follows easily from the estimate

$$\left(\frac{s}{(s-\tau)\tau}\right)^{1/s} = \left(\frac{s}{(s-1+1/s)(1-1/s)}\right)^{1/s} \le \left(\frac{s^2}{(s-1)^2}\right)^{1/s} \le \frac{1}{(s_0-1)^2} \left(s^{1/s}\right)^2.$$

Proof of Proposition A.1. 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

$$\|u\|^s_{\tau,s} \le \int_0^C t^{-\tau} t^s \|u\|^s_Y \frac{\mathrm{d}t}{t} + \int_C^\infty t^{-\tau} C^s \|u\|^s_Y \frac{\mathrm{d}t}{t} = \left(\frac{1}{s-\tau} + \frac{1}{\tau}\right) C^{s-\tau} \|u\|^s_Y.$$

Taking the s-th root yields the claim. □

Proposition A.3. *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 $c(\tau,s_1,s_2) = [s_1 \min\{\, \tau, 1-\tau \,\}]^{1/s_1 - 1/s_2}$.

Proof. See proof of [110, Proposition 1.1.3]. □

Proposition A.4. *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 [110, Proposition 1.1.4]. □

Combination of Propositions A.3 and A.4 immediately implies the following embedding; see also [146, Theorem 1.3.3 e)].

Proposition A.5. *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 A.6. 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 $r \ge p > 1$, the embedding constant of Proposition A.5 is bounded by

$$\begin{aligned} c(1-2/r,1-1/r,r,p) &= \left[r\min\left\{1-\frac{1}{r},\frac{1}{r}\right\}\right]^{1/r}\left(r+\left(1-\frac{1}{r}\right)^{-1}\right)\left[p\min\left\{1-\frac{2}{r},\frac{2}{r}\right\}\right]^{1-1/p} \\ &\le r^{1/r}(r+1)\min\left\{p-\frac{2p}{r},\frac{2p}{r}\right\} \le cp. \end{aligned}$$

The trace method

Let i_0 denote the trace mapping, i.e. $i_0 u = u(0)$. 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

$$V(s,1-\tau,Y,X) := \{\, i_0 v \colon v_{0,1-\tau} \in L^s_*(\mathbb{R}_+;Y), v_{1,1-\tau} \in L^s_*(\mathbb{R}_+;X) \,\},$$

equipped with the norm

$$\|u\|^{\mathrm{Tr}}_{\tau,s} = \inf\{\, \|v_{0,1-\tau}\|_{L^s_*(\mathbb{R}_+;Y)} + \|v_{1,1-\tau}\|_{L^s_*(\mathbb{R}_+;X)} \colon i_0 v = u \,\}.$$

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 A.7. *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 [110, 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 A.8. *Let $T > 0$ and X, Y be Banach spaces such that $Y \hookrightarrow_d X$. If $s \in (1, \infty)$, then*

$$W^{1,s}((0,T);X) \cap L^s((0,T);Y) \hookrightarrow C([0,T];(X,Y)_{1-1/s,s}). \tag{A.1}$$

If $\tau \in (0, 1 - \frac{1}{s})$, then

$$W^{1,s}((0,T);X) \cap L^s((0,T);Y) \hookrightarrow C^\alpha((0,T);(X,Y)_{\tau,1}), \qquad 0 \le \alpha < 1 - \frac{1}{s} - \tau. \tag{A.2}$$

Moreover, the embedding constants are bounded by

$$c_{(\mathrm{A}.1)}(s) = \frac{cs}{s-1} \quad \text{and} \quad c_{(\mathrm{A}.2)}(\tau, s) = 2\left(c_{(\mathrm{A}.1)}(s)\right)^{\tau/(1-1/s)}.$$

Proof. The embedding constant for (A.2) is explicitly verified in [49, Lemma 3.4 b)]. Precisely, the constant for (A.2) is bounded by $2c^\lambda$ with $\lambda = \tau/(1-1/s)$ and c from (A.1). For these reasons, it remains to verify the dependencies of (A.1), where we follow the ideas of [4, Theorem III.4.10.2].

For the particular choice $\tau = 1 - 1/s$, the trace space becomes

$$V(s, 1/s, Y, X) \coloneqq \{\, i_0 v \colon v \in W^{1,s}(\mathbb{R}_+; X) \cap L^s(\mathbb{R}_+; Y) \,\},$$

equipped with the norm

$$\|u\|^{\mathrm{Tr}}_{1-1/s,s} = \inf \{\, \|v\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \colon i_0 v = u \,\}.$$

Clearly, the trace mapping $i_0 \colon W^{1,s}(\mathbb{R}_+; X) \cap L^s(\mathbb{R}_+; Y) \to V(s, 1/s, Y, X)$ 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 $t, t' \ge 0$. 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. [4, Lemma III.4.10.1 (i)]. Noting that $i_0 \lambda_t u = u(t)$, we infer

$$\|u(t)\|^{\mathrm{Tr}}_{1-1/s,s} \le \|\lambda_t u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \le \|u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)}, \quad t \ge 0.$$

Furthermore, if $0 \le t < t' < \infty$, we have

$$\begin{aligned} \|u(t') - u(t)\|^{\mathrm{Tr}}_{1-1/s,s} &\le \|\lambda_t(\lambda_{t'-t} - 1)u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \\ &\le \|(\lambda_{t'-t} - 1)u\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \end{aligned}$$

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 (A.1), 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

$$\hat{u}(t) = \begin{cases} u(t), & \text{if } 0 \le t \le T, \\ u(2T - t), & \text{if } T < t \le 2T. \end{cases}$$

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 $W^{1,s}((0,T);X)\cap L^s((0,T);Y)$, we obtain

$$\|Eu\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)} \le 2\|\eta\|_{C^1(\mathbb{R}_+)}\|u\|_{W^{1,s}((0,T);X)\cap L^s((0,T);Y)}$$

for all $u \in W^{1,s}((0,T);X)\cap L^s((0,T);Y)$. Thus, for any $t \in [0,T)$,

$$\begin{aligned}\|u(t)\|^{\mathrm{Tr}}_{1-1/s,s} = \|(Eu)(t)\|^{\mathrm{Tr}}_{1-1/s,s} &\le \|Eu\|_{W^{1,s}(\mathbb{R}_+;X)\cap L^s(\mathbb{R}_+;Y)}\\ &\le c\|u\|_{W^{1,s}((0,T);X)\cap L^s((0,T);Y)},\end{aligned}$$

with $c = 2\|\eta\|_{C^1(\mathbb{R}_+)}$, which is independent of s. Finally, according to Proposition A.7 it holds $V(s,1/s,Y,X) = (X,Y)_{1-1/s,s}$ and

$$\|u\|_{1-1/s,s} \le \frac{s}{s-1}\|u\|^{\mathrm{Tr}}_{1-1/s,s},$$

which yields (A.1). □

Intermediate spaces and the reiteration theorem

Let $\{X,Y\}$ be an interpolation couple, $0 \le \theta \le 1$, 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

$$\|x\|_E \le c\|x\|_X^{1-\theta}\|x\|_Y^{\theta}, \quad x \in X\cap Y.$$

We write $E \in J_\theta(X,Y)$ for short. The following result is one half of the reiteration theorem for real interpolation spaces.

Proposition A.9. *Let $0 \le \theta_0 < \theta_1 \le 1$ 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

$$2(\theta_1-\theta_0)^{-1-1/s}(1+3\theta^{-1})(c_0+c_1(1-\tau)^{-1})\tau^{-2}\left(2+\tau^{-1}\right)$$

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 A.9, we have to trace the constants mentioned in [110, Remark 1.2.4].

Proposition A.10. *For each $v \in V(p,1-\theta,Y,X)$, with $\theta \in (0,1)$, the mean of v defined by*

$$w(t) := \frac{1}{t}\int_0^t v(s)\,\mathrm{d}s, \quad t>0,$$

satisfies the estimate

$$\begin{aligned}\|t^{1-\theta}w\|_{L^p_*(\mathbb{R}_+;Y)} + \|t^{2-\theta}w'\|_{L^p_*(\mathbb{R}_+;Y)} + \|t^{1-\theta}w'\|_{L^p_*(\mathbb{R}_+;X)}&\\ \le (1+3/\theta)\left(\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)} + \|v_{1,1-\theta}\|_{L^p_*(\mathbb{R}_+;X)}\right).&\quad (A.3)\end{aligned}$$

We also have $i_0 w = i_0 v$.

Proof. Let v and w be as above. First note that the derivative of w is given by

$$\begin{aligned} w'(t) &= -\frac{1}{t^2}\int_0^t v(s)\,\mathrm{d}s + \frac{1}{t}v(t) \\ &= \frac{1}{t}\left(-w(t)+v(t)\right) = \frac{1}{t^2}\int_0^t -v(s)+v(t)\,\mathrm{d}s. \end{aligned} \tag{A.4}$$

We estimate the first summand in (A.3), using [110, Corollary A.3.1]

$$\|t^{1-\theta}w\|_{L^p_*(\mathbb{R}_+;Y)} \le \frac{1}{\theta}\|t^{1-\theta}v\|_{L^p_*(\mathbb{R}_+;Y)} = \frac{1}{\theta}\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)}. \tag{A.5}$$

As a consequence, the second summand in (A.3) can now be estimated as

$$\begin{aligned} \|t^{2-\theta}w'\|_{L^p_*(\mathbb{R}_+;Y)} &= \|t^{1-\theta}(-w(t)+v(t))\|_{L^p_*(\mathbb{R}_+;Y)} \\ &\le \|t^{1-\theta}w(t)\|_{L^p_*(\mathbb{R}_+;Y)} + \|t^{1-\theta}v(t)\|_{L^p_*(\mathbb{R}_+;Y)} \\ &\le \left(1+\theta^{-1}\right)\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)}. \end{aligned} \tag{A.6}$$

The third summand in (A.3) can be estimated employing the last expression of (A.4). Thus,

$$\begin{aligned} \|w'(t)\|_X &\le \frac{1}{t^2}\|\int_0^t\int_s^t v'(\sigma)\,\mathrm{d}\sigma\,\mathrm{d}s\|_X \\ &\le \frac{1}{t^2}\int_0^t\int_0^t \|v'(\sigma)\|_X\,\mathrm{d}\sigma\,\mathrm{d}s \le \frac{1}{t}\int_0^t \|v'(\sigma)\|_X\,\mathrm{d}\sigma. \end{aligned}$$

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., [110, Equation (A.3.1)], leads to

$$\begin{aligned} \|t^{1-\theta}w'\|^p_{L^p_*(\mathbb{R}_+;X)} &\le \theta^{-p}\int_0^\infty s^{-\theta p}\left(s\|v'(s)\|_X\right)^p\frac{\mathrm{d}s}{s} \\ &= \theta^{-p}\int_0^\infty s^{(1-\theta)p}\|v'(s)\|^p_X\frac{\mathrm{d}s}{s} = \theta^{-p}\|v_{1,1-\theta}\|^p_{L^p_*(\mathbb{R}_+;X)}. \end{aligned} \tag{A.7}$$

Thus, the desired inequality follows by adding (A.5), (A.6), and the p-th root of (A.7). The last statement directly follows from continuity of v: For $t>0$, we have

$$|\frac{1}{t}\int_0^t v(s)\,\mathrm{d}s - i_0 v| = |\frac{1}{t}\int_0^t v(s)-v(0)\,\mathrm{d}s| \le \sup_{s\in[0,t]}|v(s)-v(0)|.$$

Continuity of v on $[0,\infty)$ and going to the limit $t\to 0$ in the inequality above yields

$$i_0 w = \lim_{t\to 0} w(t) = \lim_{t\to 0}\frac{1}{t}\int_0^t v(s)\,\mathrm{d}s = i_0 v,$$

concluding the proof. □

Proof of Proposition A.9. This is a standard result in interpolation theory. To trace the constants, we follow the proof of [110, 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. $i_0 v = u$. Defining w by the mean of v as in Proposition A.10 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 $c(\theta, v) := (1 + 3\theta^{-1}) \left(\|v_{0,1-\theta}\|_{L^s_*(\mathbb{R}_+;Y)} + \|v_{1,1-\theta}\|_{L^s_*(\mathbb{R}_+;X)} \right)$ 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 $u = i_0 v = i_0 w = i_0 g$ 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

$$\|t^{1-\tau(\theta_1-\theta_0)} w'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} \le c_0 c(\theta, v), \tag{A.8}$$

$$\|t^{1+(1-\tau)(\theta_1-\theta_0)} w'(t)\|_{L^s_*(\mathbb{R}_+;E_1)} \le c_1 c(\theta, v). \tag{A.9}$$

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 $w(t) = -\int_t^\infty w'(\sigma)\, \mathrm{d}\sigma$, inequality (A.9), 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 (A.8),

$$\begin{aligned} \|t^{1-\tau} g'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} &= (\theta_1 - \theta_0)^{-1-1/s} \|t^{1-\tau(\theta_1-\theta_0)} w'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} \\ &\le (\theta_1 - \theta_0)^{-1-1/s} c_0 c(\theta, v). \end{aligned}$$

This and (A.1) 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 $i_0 g = u$) and its equivalence to the K-method, see Proposition A.7, that

$$\begin{aligned}\|u\|_{(E_0,E_1)_{\tau,s}} &\le \tau^{-1}\|u\|^{\mathrm{Tr}}_{(E_0,E_1)_{\tau,s}}\\ &\le \tau^{-1}\left(\|t^{1-\tau}g'(t)\|_{L^s_*(\mathbb{R}_+;E_0)} + \|t^{1-\tau}g(t)\|_{L^s_*(\mathbb{R}_+;E_1)}\right),\\ &\le \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0+c_1(1-\tau)^{-1})c(\theta,v),\\ &= \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0+c_1(1-\tau)^{-1})\\ &\quad \left(1+3\theta^{-1}\right)\left(\|v_{0,1-\theta}\|_{L^p_*(\mathbb{R}_+;Y)} + \|v_{1,1-\theta}\|_{L^p_*(\mathbb{R}_+;X)}\right).\end{aligned}$$

Finally, taking the infimum over all v with $i_0 v = u$, we find

$$\begin{aligned}\|u\|_{(E_0,E_1)_{\tau,s}} &\le \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0+c_1(1-\tau)^{-1})\left(1+3\theta^{-1}\right)\|u\|^{\mathrm{Tr}}_{(X,Y)_{\tau,s}}\\ &\le \tau^{-1}(\theta_1-\theta_0)^{-1-1/s}(c_0+c_1(1-\tau)^{-1})\left(1+3\theta^{-1}\right)\frac{2}{\tau}\left(2+\tau^{-1}\right)\|u\|_{(X,Y)_{\tau,s}}\end{aligned}$$

concluding the proof. □

The real interpolation method and domains of fractional operators

In this paragraph we consider a linear operator 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 A.11. *Let $\tau \in (0,1)$ and $1 \le s \le \infty$. Then*

$$(X, \mathcal{D}_X(A))_{\tau,s} = \left\{x \in X \colon t \mapsto x_\tau(t) \coloneqq t^\tau \|AR(-t,A)x\|_X \in L^s_*(\mathbb{R}_+)\right\},$$

and the norms $\|\cdot\|_{\tau,s}$ and

$$\|x\|^*_{\tau,s} \coloneqq \|x\|_X + \|x_\tau\|_{L^s_*(\mathbb{R}_+)}.$$

are equivalent. Precisely,

$$\|x\|_{\tau,s} \le \left(2 + M((1-\tau)s)^{-1/s}\right)\|x\|^*_{\tau,s}, \quad \|x\|^*_{\tau,s} \le (M+1)\|x\|_{\tau,s}.$$

Proof. This follows from the proof of [110, Proposition 3.1.1]. □

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., [146, Section 1.15] and [110, 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 A.12. *Let $z_1, z_2 \in \mathbb{C}$ such that $\operatorname{Re} z_1 < \operatorname{Re} z_2$. Then*

$$\mathcal{D}_X(A^{z_2}) \hookrightarrow \mathcal{D}_X(A^{z_1})$$

and the embedding constant is bounded by $\max\{1, \|A^{z_1-z_2}\|_{\mathcal{L}(X)}\}$.

Proof. From the proof of [110, Theorem 4.1.6] we have

$$\|A^{z_1}x\|_X \le \|A^{z_1-z_2}\|_{\mathcal{L}(X)}\|A^{z_2}x\|_X$$

for all $x \in \mathcal{D}_X(A^{z_2})$. □

Proposition A.13. *Let A be a sectorial operator on a Banach space X. Then*

$$(X, \mathcal{D}_X(A))_{\tau,1} \hookrightarrow \mathcal{D}_X(A^\tau), \quad \tau \in (0,1),$$

where the embedding constant is bounded by $(M+1)\max\{1, (\Gamma(\tau)\,\Gamma(1-\tau))^{-1}\}$.

Remark A.14. For $\tau \in (0,1)$ we have $\Gamma(\tau) \ge 1 - \mathrm{e}^{-1}$ and thus $\max(1, (\Gamma(\tau)\,\Gamma(1-\tau))^{-1})$ in Proposition A.13 is bounded by a constant independently of τ. This can be seen from the definition of the gamma function

$$\Gamma(\tau) = \int_0^\infty t^{\tau-1}\mathrm{e}^{-t}\,\mathrm{d}t \ge \int_0^1 t^{\tau-1}\mathrm{e}^{-t}\,\mathrm{d}t \ge \int_0^1 \mathrm{e}^{-t}\,\mathrm{d}t = 1 - \mathrm{e}^{-1} > 0.$$

Proof of Proposition A.13. We closely follow the proof of [110, Proposition 4.1.7]. Consider $x \in (X, \mathcal{D}_X(A))_{\tau,1}$. Due to Proposition A.11, 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., [110, 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 A.11 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 A.15. *Let A be a sectorial operator on a Banach space X. Then*

$$\mathcal{D}_X(A^\tau) \hookrightarrow (X, \mathcal{D}_X(A))_{\tau,\infty}, \quad \tau \in (0,1),$$

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 [110, Proposition 4.1.7]. Let $x \in \mathcal{D}_X(A^\tau)$. According to Proposition A.11 we have

$$\|x\|_{\tau,s} \le (2+M) \sup_{t>0} t^\tau \|AR(-t,A)x\|_X.$$

Using the representation formula

$$A^{-\tau-1}x = \frac{1}{\Gamma(1-\tau)\,\Gamma(1+\tau)} \int_0^\infty z^{-\tau} R(-z,A)^2 x \,\mathrm{d}z,$$

see [110, Equation (4.1.8)], we obtain

$$AR(-t,A)x = \frac{A^2 R(-t,A)}{\Gamma(1-\tau)\,\Gamma(1+\tau)} \int_0^\infty z^{-\tau} R(-z,A)^2 A^\tau x \,\mathrm{d}z.$$

Moreover, for any $t > 0$ we estimate

$$\|AR(-t,A)x\|_X \leq \frac{M}{1+t} \int_0^t z^{-\alpha}(M+1)^2 \|A^\tau x\|_X + (M+1) \int_t^\infty z^{-\alpha} \frac{M(M+1)}{1+z} \|A^\tau x\|_X.$$

Hence,

$$t^\tau \|AR(-t,A)x\|_X \leq \frac{M(M+1)^2}{\Gamma(1-\tau)\,\Gamma(1+\tau)} \left(\frac{t}{1+t} \frac{1}{1-\tau} + \frac{1}{\tau} \right) \|A^\tau x\|_X$$

concluding the proof. □

In the Hilbert space case, we can give the following embedding.

Proposition A.16. *Let A be a sectorial, self-adjoint operator on a Hilbert space H. Then*

$$\mathcal{D}_H(A^\tau) \hookrightarrow (H, \mathcal{D}_H(A))_{\tau,2},$$

where the embedding constant is bounded by

$$1 + (-2\cos(\pi\tau)\,\Gamma(-2\tau))^{1/2}.$$

Proof. Following the proof of [146, Theorem 1.18.10], the constant c in step 2 is given by

$$\int_0^\infty \frac{|\mathrm{e}^{\mathrm{i}t\mu} - 1|^2}{(t\mu)^{2\tau}} \frac{\mathrm{d}t}{t} = -2\cos(\pi\tau)\,\Gamma(-2\tau).$$

Taking square roots yields the bound. □

Last, we verify the resolvent estimates of this subsection for the concrete example $A = -\Delta$ that we will consider in the main text on error estimates. Note that the arguments do not employ the homogeneous Dirichlet boundary conditions of $-\Delta$, which allows applying them to fairly general operators and different boundary conditions.

Remark A.17. First, according to [66, Theorem 5.1], we have the estimate

$$\|R(z,-\Delta)\|_{\mathcal{L}(L^p)} \leq \frac{M}{|z|}, \quad \operatorname{Re} z \leq 0, \quad z \neq 0, \tag{A.10}$$

where the constant M can be chosen to be independent of $p \in (1,\infty)$. The resolvent estimate (A.10) ensures that for all fixed $\omega_0 > 0$ we have

$$\mathcal{D}_{L^p}((-\Delta+\omega)^\tau) = \mathcal{D}_{L^p}((-\Delta)^\tau) \tag{A.11}$$

with equivalence of norms independent of $p \in (1,\infty)$, $\tau \in [0,1]$, and $\omega \in [0,\omega_0]$; see, e.g., [110, Lemma 4.1.11]. Moreover, [66, Theorem 5.1] in addition yields resolvent estimates for $-\Delta + 1$, precisely

$$\|R(z,-\Delta+1)\|_{\mathcal{L}(L^p)} \leq \frac{M}{1+|z|}, \quad \operatorname{Re} z \leq 0. \tag{A.12}$$

Using the equivalence of norms (A.11), we can equivalently consider $-\Delta$ or $-\Delta+1$ in the results of this section.

Employing the power series expansion, see, e.g., [71, Proposition A.2.3], the estimates (A.10) and (A.12) can be extended to hold on a sector. To this end, consider $z = \lambda \mathrm{e}^{\mathrm{i}\varphi+\mathrm{i}\pi/2}$ for $\lambda > 0$ and $\varphi \in (-\pi/2,\pi/2)$. Then

$$R(\lambda \mathrm{e}^{\mathrm{i}\varphi+\mathrm{i}\pi/2},-\Delta) = R(\mathrm{i}\lambda,-\Delta)\sum_{m=0}^{\infty}(1-\mathrm{e}^{\mathrm{i}\varphi})^m[\mathrm{i}\lambda R(\mathrm{i}\lambda,-\Delta)]^m.$$

Choose $\theta \in (0,\pi/2)$ sufficiently small such that $|1-\mathrm{e}^{\mathrm{i}\varphi}| < (2M)^{-1}$ for all $|\varphi| < \theta$. Using the submultiplicity of the operator norm and the resolvent estimate (A.10) we obtain

$$\|R(\lambda \mathrm{e}^{\mathrm{i}\varphi+\mathrm{i}\pi/2},-\Delta)\|_{\mathcal{L}(L^p)} \leq \|R(\mathrm{i}\lambda,-\Delta)\|_{\mathcal{L}(L^p)}\sum_{m=0}^{\infty}2^{-m} \leq \frac{2M}{|z|}.$$

Hence, there are constants $M' > 0$ and $\theta \in (0,\pi/2)$ such that

$$\|R(z,-\Delta)\|_{\mathcal{L}(L^p)} \leq \frac{M'}{|z|}, \quad z \in \mathbb{C}\setminus\Sigma_\theta, \quad z \neq 0, \tag{A.13}$$

where

$$\Sigma_\theta \coloneqq \{z \in \mathbb{C}\setminus\{0\} \colon |\arg z| < \theta\}.$$

Analogously, we obtain sectoriality of $-\Delta+1$ by extending (A.12) to a sector.

A.2. Regularity of the state equation

We summarize several regularity results for the state equation and give short proofs. Throughout this section we assume that the operator A is given by a bilinear form satisfying Gårdings inequality; see Assumption 2.1. The symbol $\mathrm{e}^{-\cdot A}$ denotes the semigroup generated by $-A$.

Proposition A.18. *Let $T > 0$, $\theta \in (0,1/2]$, $f \in L^\infty((0,T);X_\theta)$, $u_0 \in V^*$. Consider the solution u to*

$$\partial_t u + Au = f, \quad u(0) = u_0.$$

Then:

(i) If $u_0 \in V$, then u is continuous from $[0,T]$ into V,

(ii) If $u_0 \in H$, then u is continuous from $(0,T]$ into V,

(iii) If $u_0 = 0$ and $\gamma \in [\theta,1]$, then

$$\|u(t)\|_{X_\gamma} \leq c\|f\|_{L^\infty((0,T);X_\theta)}\, t^{1+\theta-\gamma}, \quad 0 \leq t \leq T,$$

with $c > 0$ depending on θ,γ, but independent of f.

Proof. The unique solution is given by the variation of constants formula

$$u(t) = \mathrm{e}^{-tA}u_0 + \int_0^t \mathrm{e}^{-(t-s)A} f(s)\,\mathrm{d}s, \quad t \in [0,T]. \tag{A.14}$$

According to Theorem 2.6.13 c) in [128], for $\theta > 0$ there is a constant $M_\theta > 0$ such that it holds

$$\mathrm{e}^{-\omega_0 t}\|\mathrm{e}^{-tA}v\|_{X_\theta} = \|(A+\omega_0)^\theta \mathrm{e}^{-t(A+\omega_0)}v\|_{V^*} \le M_\theta t^{-\theta}\|v\|_{V^*} \tag{A.15}$$

for all $v \in V^*$ and $t > 0$.

(iii): Employing (A.15) we obtain

$$\begin{aligned}\|u(t)\|_{X_\gamma} &= \|\int_0^t \mathrm{e}^{-(t-s)A} f(s)\,\mathrm{d}s\|_{X_\gamma} \le \int_0^t \|(A+\omega_0)^{\gamma-\theta}\mathrm{e}^{-(t-s)A}(A+\omega_0)^\theta f(s)\|_{V^*}\,\mathrm{d}s \\ &\le M_{\gamma-\theta}\mathrm{e}^{t\omega_0}\|(A+\omega_0)^\theta f\|_{L^\infty((0,t);V^*)}\int_0^t s^{\theta-\gamma}\,\mathrm{d}s \le c\,t^{1+\theta-\gamma}\|f\|_{L^\infty((0,t);X_\theta)}.\end{aligned}$$

(i), (ii): If $u_0 \in V$, it holds $(A+\omega_0)\mathrm{e}^{-tA}u_0 = \mathrm{e}^{-tA}(A+\omega_0)u_0$; see, e.g., [128, Theorem 2.6.13 b)]. Whence, continuity of $t \mapsto \mathrm{e}^{-tA}u_0$ from $[0,T]$ into V follows from [128, Corollary 1.2.3]. If $u_0 \in H$, we find for any $t, \tau > 0$ that

$$\|\left(\mathrm{e}^{-(t+\tau)A} - \mathrm{e}^{-tA}\right)u_0\|_V = \|\mathrm{e}^{-tA}\left(\mathrm{e}^{-\tau A} - 1\right)u_0\|_V \le M_{1/2}\mathrm{e}^{\omega_0 t}t^{-1/2}\|\left(\mathrm{e}^{-\tau A} - 1\right)u_0\|_{X_{1/2}}.$$

This proves continuity of $t \mapsto \mathrm{e}^{-tA}u_0$ in V for $t > 0$, using that $-A$ induces a continuous semigroup also on $H = X_{1/2}$.

Now we turn to the second term of (A.14). Since A exhibits maximal parabolic regularity, both on V^* and H, it also possesses maximal regularity on the interpolation space X_θ; see [73, Lemma 5.3]. Hence, for $f \in L^r((0,T), X_\theta)$, the function $\breve{u}(t) = \int_0^t \mathrm{e}^{-(t-s)A}f(s)\,\mathrm{d}s$ has the regularity $\breve{u} \in W^{1,r}((0,T);X_\theta) \cap L^r((0,T);X_{1+\theta})$ for any $r \in (1,\infty)$. Furthermore, by the trace theorem, there holds the embedding

$$W^{1,r}((0,T);X_\theta) \cap L^r((0,T);X_{1+\theta}) \hookrightarrow C([0,T];(X_\theta, X_{1+\theta})_{1-1/r,r});$$

see, e.g., [4, Theorem III.4.10.2]. Choose $r > 1/\theta$, which is equivalent to $1-\theta < 1-1/r$. Thus,

$$(X_\theta, X_{1+\theta})_{1-1/r,r} \hookrightarrow (X_\theta, X_{1+\theta})_{1-\theta,1} \hookrightarrow [X_\theta, X_{1+\theta}]_{1-\theta} = \mathcal{D}_{V^*}(A+\omega_0) = V$$

due to [146, Theorems 1.3.3 e), 1.15.2 d) and 1.15.3]. In summary, we conclude the proof of (i) and (ii). □

Proposition A.19. *Let $T > 0$ and $u_0 \in H$. The solution operator $f \mapsto u$ with*

$$\partial_t u + Au = f, \quad u(0) = u_0,$$

is continuous and compact from $L^2((0,T);X_{\theta_0})$ into $L^2((0,T);V)$.

Proof. Let $\mathcal{S}$ denote the solution operator of the parabolic state equation, i.e. $u = \mathcal{S}(u_0, f)$ satisfies $\partial_t u + Au = f$, $u(0) = u_0$. Since A exhibits maximal parabolic regularity, both on V^* and H, it also possesses maximal regularity on the interpolation space X_{θ_0}; see, e.g., [73, Lemma 5.3]. Hence, $f \mapsto \mathcal{S}(0,f)$ is continuous from $L^2((0,T);X_{\theta_0})$ into $H^1((0,T);X_{\theta_0}) \cap$

$L^2((0,T);X_{1+\theta_0})$, where we have used the identification $\mathcal{D}_{X_{\theta_0}}(A) = X_{1+\theta_0}$. Clearly, $X_{1+\theta_0} \hookrightarrow \mathcal{D}_{V^*}(A) = V \hookrightarrow_c H \hookrightarrow X_{\theta_0}$. Employing [4, Theorem I.2.11.1] we deduce $X_{1+\theta_0} \hookrightarrow_c [X_{\theta_0}, X_{1+\theta_0}]_{1-\theta_0} = V$, where we have used [146, Theorem 1.15.3] in the last step. Therefore, the Aubin-Lions Lemma (see, e.g., [107, Théorème I.5.1]) yields the compact injection

$$H^1((0,T);X_{\theta_0}) \cap L^2((0,T);X_{1+\theta_0}) \hookrightarrow_c L^2((0,T);V).$$

Furthermore, $\mathcal{S}(u_0,0) \in W(0,T) \hookrightarrow L^2((0,T);V)$. Whence, the assertion follows from the splitting $\mathcal{S}(u_0,f) = \mathcal{S}(u_0,0) + \mathcal{S}(0,f)$. □

Proposition A.20. *Let $s > 2$ and $u_0 \in H$. The mapping $(\nu,q) \mapsto S(\nu,q)$ is completely continuous from $\mathbb{R} \times L^s(I;Q)$ into $C([0,1];H)$.*

Proof. Consider first the case $u_0 = 0$. According to [3, Theorem 3] we have the compact embedding

$$W^{1,s}(I;V^*) \cap L^s(I;V) \hookrightarrow_c C^\alpha(I;(V^*,V)_{\tau,1}), \quad 0 \le \alpha < 1 - \frac{1}{s} - \tau,$$

for any $\tau \in (0, 1-1/s)$ due to $V \hookrightarrow_c V^*$. Since $s > 2$ we may choose $\tau = 1/2$ and obtain

$$(V^*,V)_{\tau,1} \hookrightarrow \mathcal{D}_{V^*}(A^{1/2}) = H;$$

see [146, Theorem 1.15.2d)]. Therefore, for each fixed $\nu \in \mathbb{R}_+$ we find that $q \mapsto S(\nu,q)$ is completely continuous from $L^s(I;Q)$ into $C([0,1];H)$. Note that continuity of the control operator from Q into V^* is sufficient for the argument above. Moreover, using that $\mathbb{R}$ is finite dimensional, we conclude that the mapping $(\nu,q) \mapsto S(\nu,q)$ is completely continuous from $\mathbb{R}_+ \times Q_{ad}(0,1)$ into $C([0,1];H)$. If $u_0 \neq 0$, then the variation of constants formula yields the additional term $\mathrm{e}^{-\nu A}u_0$, which is continuous in ν. We conclude the proof by superposition of both cases. □

Proposition A.21. *Let Gårdings inequality hold with $\omega_0 = 0$. Then*

$$\|\mathrm{e}^{-tA}\|_{\mathcal{L}(H)} \le \mathrm{e}^{-\alpha_0 t} \quad \textit{for all} \quad t > 0.$$

Proof. Let $\rho \in [0,\alpha_0)$. Then the form $b(u,v) \coloneqq a(u,v) - \rho(u,v)_H$ is coercive. Let B denote the operator associated with the form b. Employing [6, Theorem 4.2] we infer that $-B$ generates a contractive semigroup $\mathrm{e}^{-\cdot B}$. Hence

$$\|\mathrm{e}^{-tA}\|_{\mathcal{L}(H)} = \mathrm{e}^{-\rho t}\|\mathrm{e}^{-tB}\|_{\mathcal{L}(H)} \le \mathrm{e}^{-\rho t},$$

where we have used the representation $-A = -B - \rho I$. Choosing a sequence $\rho_n \in [0,\alpha_0)$ such that $\rho_n \to \alpha_0$ and using the estimate above yields the result. □

Moreover, for discretization error estimates with cellwise linear controls, see Theorem 5.21, we have to assume improved regularity of the adjoint state. It can be reduced to a regularity assumption on the desired state by means of the following proposition; see Remark 5.22.

Proposition A.22. *Let $\Omega \subset \mathbb{R}^d$ be a bounded and convex domain with polygonal boundary. Suppose that $A = -\Delta$ is equipped with homogeneous Dirichlet boundary conditions. In addition, assume distributed control, i.e. $B\colon L^s(\omega) \to L^s(\Omega)$ for all $s \in (1,\infty]$. For any tuple $(\nu,q) \in \mathbb{R}_+ \times Q_{ad}(0,1)$ and $p \in (1,\infty)$ the solution u to the state equation has the improved regularity $u(1) \in W_0^{1,p}(\Omega)$.*

Proof. First, it holds $\mathrm{e}^{\nu\Delta}u_0 \in \mathcal{D}_{W^{-1,p}}(-\Delta) = W_0^{1,p}$, see [128, Theorem 2.6.13 (a)], since $u_0 \in H_0^1 \hookrightarrow W^{-1,p}$. Furthermore, due to $\mathcal{D}_{W^{-1,p}}(-\Delta) = W_0^{1,p}$, see [45, Corollary 3.12], the operator $-\Delta\colon W_0^{1,p} \to W^{-1,p}$ is an isomorphism. Hence, putting $X = [L^p, W^{-1,p}]_\zeta$ with $\zeta = 1 - 1/2p$ we obtain for $\tau = 1 - 1/4p$ that

$$(X, \mathcal{D}_X(-\Delta))_{\tau,1} \hookrightarrow W_0^{1,p};$$

see [73, Lemma 6.6 (i)]. Since $-\Delta$ satisfies maximal parabolic regularity both on L^p and $W^{-1,p}$, also on the complex interpolation space X; see [73, Lemma 5.3]. Thus, the solution to the state equation with homogeneous initial condition belongs to

$$W^{1,s}(I;X) \cap L^s(I;\mathcal{D}_X(-\Delta)) \hookrightarrow C^\kappa(I;(X,\mathcal{D}_X(-\Delta))_{\tau,1})$$

with $\kappa > 0$ small, provided that $0 < \tau < 1 - 1/s$; see [3, Theorem 3]. Due to the control constraints, we can take $s \in (1,\infty)$ arbitrary large so that $\tau = 1 - 1/4p < 1 - 1/s$. Combination of both embeddings yields $u(1) \in W_0^{1,p}$. □

A.3. Clarke's generalized subdifferential

The *generalized directional derivative* at x from a Banach space X for any function $f\colon X \to \mathbb{R}$ that is Lipschitz near x is given by [40, Section 10.1]

$$f^\circ(x;v) \coloneqq \limsup_{y\to x,\ \tau\downarrow 0} \tau^{-1}\left[f(y+\tau v) - f(y)\right]. \tag{A.16}$$

Then $\zeta \in X^*$ belongs to the *generalized gradient* $\partial_{\mathrm{C}} f(x)$ if and only if $f^\circ(x;v) \geq \langle \zeta, v\rangle$ for all $v \in X$.

Let X_1, X_2 be Banach spaces and $f\colon X_1 \times X_2 \to \mathbb{R}$ Lipschitz near $x_1 \in X_1$ and $x_2 \in X_2$. We define the partial generalized directional derivatives and partial generalized gradients $f_{x_1}^\circ$, $f_{x_2}^\circ$, $\partial_{\mathrm{C},x_1} f$, and $\partial_{\mathrm{C},x_2} f$ analogously to (A.16).

Proposition A.23. *If $f_{x_1}^\circ(x_1,x_2;v_1) = f^\circ(x_1,x_2;v_1,0)$ and $f_{x_2}^\circ(x_1,x_2;v_2) = f^\circ(x_1,x_2;0,v_2)$ for all $v_1 \in X_1$ and $v_2 \in X_2$, then*

$$\partial_{\mathrm{C}} f(x_1,x_2) \subset \partial_{\mathrm{C},x_1} f(x_1,x_2) \times \partial_{\mathrm{C},x_2} f(x_1,x_2).$$

Proof. $\zeta \in \partial_{\mathrm{C}} f(x_1,x_2)$ if and only if $f^\circ(x_1,x_2;v_1,v_2) \geq \langle \zeta_1, v_1\rangle + \langle \zeta_2, v_2\rangle$ for all $v_1 \in X_1$ and $v_2 \in X_2$. Taking $v_1 = 0$ and $v_2 = 0$ implies $f^\circ(x_1,x_2;v_1,0) \geq \langle \zeta_1, v_1\rangle$ for all $v_1 \in X_1$ and $f^\circ(x_1,x_2;0,v_2) \geq \langle \zeta_2, v_2\rangle$ for all $v_2 \in X_2$. Using the suppositions on $f_{x_1}^\circ$ and $f_{x_2}^\circ$ we finish the proof. □

Proposition A.24. *For j from problem $(\hat{P})$, it holds*

$$\partial_{\mathrm{C}} j(\bar\nu,\bar q) \subset \partial_{\mathrm{C},\nu} j(\bar\nu,\bar q) \times \partial_{\mathrm{C},q} j(\bar\nu,\bar q).$$

Proof. In our case the assumptions of the preceding proposition are satisfied for j. Regarding the differentials with respect to ν, we obtain for all $\delta\nu \in L^\infty(0,1)$ that

$$\begin{aligned} j^\circ(\bar\nu,\bar q;\delta\nu,0) &= \limsup_{\nu\to\bar\nu,\ q\to\bar q,\ \tau\downarrow 0} \tau^{-1}\left[j(\nu+\tau\delta\nu,q) - j(\nu,q)\right] = \limsup_{q\to\bar q} \int_0^1 \delta\nu(1+L(q))\,\mathrm{d}t \\ &= \int_0^1 \delta\nu(1+L(\bar q))\,\mathrm{d}t = j_\nu^\circ(\bar\nu,\bar q;\delta\nu), \end{aligned}$$

using the fact that j is linear in ν in the first and last step. In the other case, we estimate

$$\begin{aligned} j_q^\circ(\bar\nu,\bar q;\delta q) &= \limsup_{q\to\bar q,\ \tau\downarrow 0} \tau^{-1}\left[j(\bar\nu, q+\tau\delta q) - j(\bar\nu,q)\right] \le j^\circ(\bar\nu,\bar q;0,\delta q) \\ &= \limsup_{\nu\to\bar\nu,\ q\to\bar q,\ \tau\downarrow 0} \tau^{-1}\int_0^1 \nu\left[L(q+\tau\delta q) - L(q)\right]\mathrm{d}t \\ &\le j_q^\circ(\bar\nu,\bar q;\delta q) + \limsup_{\nu\to\bar\nu,\ q\to\bar q,\ \tau\downarrow 0} \tau^{-1}\int_0^1 [\nu-\bar\nu]\left[L(q+\tau\delta q) - L(q)\right]\mathrm{d}t \\ &\le j_q^\circ(\bar\nu,\bar q;\delta q) + \limsup_{\nu\to\bar\nu} c_L\int_0^1 |\nu-\bar\nu|\|\delta q\|_Q\,\mathrm{d}t = j_q^\circ(\bar\nu,\bar q;\delta q), \end{aligned}$$

for all $\delta q \in Q(0,1)$, where c_L is the Lipschitz constant of L. □

A.4. Comparison principle

For any $\omega_0 \ge 0$, define $\phi\colon \mathbb{R}_+ \to \mathbb{R}_+$ by

$$\phi(t) = \omega_0^{-1}(\mathrm{e}^{\omega_0 t} - 1), \text{ if } \omega_0 > 0, \text{ and } \phi(t) = t, \text{ if } \omega_0 = 0.$$

We easily verify that $\phi(t) \ge t$ for all $t \ge 0$.

Proposition A.25. *Let $c, \gamma > 0$ and $\omega_0, h_0 \ge 0$. Moreover, let d_γ be continuously differentiable on $(0,\infty)$ and continuous on $[0,\infty)$ with $d_\gamma \ge 0$ such that*

$$d_\gamma'(t) \le \omega_0 d_\gamma(t) + c\gamma/d_\gamma(t) - h_0 \quad \text{on } \{\, t \mid d_\gamma(t) > 0 \,\}. \tag{A.17}$$

Then it holds

$$d_\gamma(t) \le \max\{\,\sqrt{\gamma}, (d_\gamma(0)+\sqrt{\gamma})\mathrm{e}^{\omega_0 t} + (c\sqrt{\gamma} - h_0)\phi(t)\,\} =: D_\gamma(t). \tag{A.18}$$

Proof. We argue by contradiction: Suppose that (A.18) is not satisfied and let t_0 be the first time such that $d_\gamma(t_0) = D_\gamma(t_0)$ and $d_\gamma(t) > D_\gamma(t)$ for $t \in (t_0, t_1)$. This implies $d_\gamma(t) > \sqrt{\gamma}$ and therefore from (A.17) we infer $d_\gamma'(t) \le \omega_0 d_\gamma(t) + c\sqrt{\gamma} - h_0$ for $t \in (t_0, t_1)$.

The unique solution of $z'(t) = \omega_0 z(t) + c\sqrt{\gamma} - h_0$ with $z(t_0) = d_\gamma(t_0)$ is given by

$$z(t) = d_\gamma(t_0)\mathrm{e}^{\omega_0(t-t_0)} + (c\sqrt{\gamma} - h_0)\phi(t - t_0).$$

The comparison principle yields $d_\gamma(t) \le z(t)$ for $t \in [t_0, t_1)$. Now we distinguish two cases: If $d_\gamma(t_0) = D_\gamma(t_0) = (d_\gamma(0)+\sqrt{\gamma})\mathrm{e}^{\omega_0 t_0} + (c\sqrt{\gamma} - h_0)\phi(t_0)$, we obtain

$$\begin{aligned} d_\gamma(t) \le z(t) &= d_\gamma(t_0)\mathrm{e}^{\omega_0(t-t_0)} + (c\sqrt{\gamma} - h_0)\phi(t-t_0) \\ &= (d_\gamma(0)+\sqrt{\gamma})\mathrm{e}^{\omega_0 t} + (c\sqrt{\gamma} - h_0)\phi(t_0)\mathrm{e}^{\omega_0(t-t_0)} + (c\sqrt{\gamma} - h_0)\phi(t-t_0) \\ &= (d_\gamma(0)+\sqrt{\gamma})\mathrm{e}^{\omega_0 t} + (c\sqrt{\gamma} - h_0)\phi(t) \\ &\le D_\gamma(t) < d_\gamma(t), \end{aligned}$$

for $t \in (t_0, t_1)$, yielding a contradiction. Otherwise, it holds

$$\begin{aligned} \sqrt{\gamma} = d_\gamma(t_0) = D_\gamma(t_0) &> (d_\gamma(0)+\sqrt{\gamma})\mathrm{e}^{\omega_0 t_0} + (c\sqrt{\gamma} - h_0)\phi(t_0) \\ &= d_\gamma(0) + \sqrt{\gamma} + \left((c+\omega_0)\sqrt{\gamma} + \omega_0 d_\gamma(0) - h_0\right)\phi(t_0) \end{aligned}$$

and we necessarily must have $((c+\omega_0)\sqrt{\gamma} + \omega_0 d_\gamma(0) - h_0) < 0$. Thus, we have

$$\sqrt{\gamma} < d_\gamma(t) \le z(t) = \sqrt{\gamma} + ((c+\omega_0)\sqrt{\gamma} - h_0)\phi(t-t_0) < \sqrt{\gamma},$$

for $t \in (t_0, t_1)$, also yielding a contradiction. □

A.5. Stability estimates

We collect stability estimates for the state and the linearized state. Suppose that the assumptions from Section 2.1 hold and recall that $I = (0,1)$ is the reference time interval.

Proposition A.26. *There exists a constant $c > 0$ such that for all $\nu > 0$, $q \in Q(0,1)$, and initial conditions $u_0 \in H$ the estimates*

$$\|u\|_{C([0,1];H)} + \sqrt{\nu}\|u\|_{L^2(I;V)} \le c\left(\sqrt{\nu}\|Bq\|_{L^2(I;V^*)} + \|u_0\|_H\right), \tag{A.19}$$

$$\|\delta u\|_{C([0,1];H)} + \sqrt{\nu}\|\delta u\|_{L^2(I;V)} \le c\frac{|\delta\nu|}{\sqrt{\nu}}\left(\|Bq\|_{L^2(I;V^*)} + \|u\|_{L^2(I;V)}\right) + \sqrt{\nu}\|B\delta q\|_{L^2(I;V^*)}, \tag{A.20}$$

$$\|\delta\tilde{u}\|_{C([0,1];H)} + \sqrt{\nu}\|\delta\tilde{u}\|_{L^2(I;V)} \le c\frac{|\delta\nu|}{\sqrt{\nu}}\left(\|B\delta q\|_{L^2(I;V^*)} + \|\delta u\|_{L^2(I;V)}\right), \tag{A.21}$$

hold, where $u = S(\nu,q)$, $\delta u = S'(\nu,q)(\delta\nu,\delta q)$ and $\delta\tilde{u} = S''(\nu,q)[\delta\nu,\delta q]^2$ for $\delta\nu \in \mathbb{R}$ and $\delta q \in L^2(I;V^*)$. Furthermore, for $q_1, q_2 \in Q_{ad}(0,1)$ we have

$$\|u_1 - u_2\|_{C([0,1];H)} + \sqrt{\nu_1}\|u_1 - u_2\|_{L^2(I;V)} \le c_0\left(|\nu_1 - \nu_2| + \|B(q_1 - q_2)\|_{L^2(I;V^*)}\right), \tag{A.22}$$

$$\|\delta u_1 - \delta u_2\|_{C([0,1];H)} \le c_1\left(|\nu_1 - \nu_2| + \|B(q_1 - q_2)\|_{L^2(I;V^*)}\right)\left(|\delta\nu| + \|B\delta q\|_{L^2(I;V^*)}\right), \tag{A.23}$$

where $u_i = S(\nu_i, q_i)$ and $\delta u_i = S'(\nu_i, q_i)(\delta\nu, \delta q)$ for $i = 1, 2$ and

$$c_0 = c_0(\nu_1,\nu_2) = c/\sqrt{\nu_1}\max\{1, 1/\sqrt{\nu_2}, \nu_2\},$$
$$c_1 = c_1(\nu_1,\nu_2) = c/\sqrt{\nu_1}\max\{1, 1/\nu_1, 1/(\nu_1\sqrt{\nu_2}), 1/\nu_2, 1/\nu_2^{3/2}, \nu_2/\nu_1\}.$$

The constant $c > 0$ depends exclusively on Poincaré's constant, Q_{ad}, and u_0.

Proof. For $\nu, \delta\nu \in \mathbb{R}$ with $\nu > 0$ and $f, g \in L^2(I;V^*)$ and $v_0 \in H$ consider the solution v to the linear parabolic equation

$$\partial_t v + \nu Av = \nu f + \delta\nu g, \quad v(0) = v_0.$$

Standard energy estimates, see for instance [161, §26], yield

$$\|v\|_{C([0,1];H)} + \sqrt{\nu}\|v\|_{L^2(I;V)} \le c\left(\sqrt{\nu}\|f\|_{L^2(I;V^*)} + \frac{|\delta\nu|}{\sqrt{\nu}}\|g\|_{L^2(I;V^*)} + \|u_0\|_H\right), \tag{A.24}$$

with c depending exclusively on Poincaré's constant. This establishes (A.19) – (A.21). Concerning (A.22), set $u_1 = S(\nu_1, q_1)$ and $u_2 = S(\nu_2, q_2)$. The difference $w = u_1 - u_2$ satisfies

$$\partial_t w + \nu_1 Aw = (\nu_2 - \nu_1)Au_2 + \nu_1 Bq_1 - \nu_2 Bq_2, \quad w(0) = 0.$$

Estimating the right-hand side yields

$$\|(\nu_2 - \nu_1)Au_2 + \nu_1 Bq_1 - \nu_2 Bq_2\|_{L^2(I;V^*)} \le \left(|\nu_1 - \nu_2|\left(\|Au_2 + Bq_1\|_{L^2(I;V^*)}\right) + \nu_2\|B(q_1 - q_2)\|_{L^2(I;V^*)}\right).$$

From (A.24), the estimate

$$\|Au_2\|_{L^2(I;V^*)} \le c\|Bq_2\|_{L^2(I;V^*)} + \frac{c}{\sqrt{\nu_2}}\|u_0\|_H,$$

and boundedness of Q_{ad} we conclude (A.22). Concerning the last estimate, the difference $\delta w = \delta u_1 - \delta u_2$ satisfies $\delta w(0) = 0$ and

$$\partial_t \delta w + \nu_1 A\delta w = (\nu_2 - \nu_1)A\delta u_2 + \delta\nu B(q_1 - q_2) + \delta\nu A(u_1 - u_2) + (\nu_1 - \nu_2)B\delta q.$$

Similarly as above, the estimate (A.23) follows from (A.19), (A.22), and (A.20). □

A.6. 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 [1, Chapter 7]; see also [48] for an introduction to this topic. This section is part of a joint work with Dominik Hafemeyer.

Let $\Omega \subseteq \mathbb{R}^d$ be an open set with $d \in \mathbb{N}$. For $\theta \in (0,1)$ and $p \in [1,\infty)$ we define

$$[f]_{\theta,p,\Omega} \coloneqq \left(\int_\Omega \int_\Omega \frac{|f(x) - f(y)|^p}{|x-y|^{d+\theta p}} \,\mathrm{d}x \,\mathrm{d}y \right)^{1/p}$$

the Gagliardo (semi)norm of f and define the norm of the fractional Sobolev space on Ω denoted $W^{\theta,p}(\Omega)$ by

$$\|f\|_{W^{\theta,p}(\Omega)} \coloneqq \left(\|f\|_{L^p(\Omega)}^p + [f]_{\theta,p,\Omega}^p \right)^{1/p}.$$

If $\theta > 1$ and θ is not an integer, then we write $\theta = m + \sigma$ with $m \in \mathbb{N}$ and $\sigma \in (0,1)$, and define the norm of $W^{\theta,p}(\Omega)$ by

$$\|f\|_{W^{\theta,p}(\Omega)} \coloneqq \left(\|f\|_{W^{m,p}(\Omega)}^p + \sum_{|\alpha|=m} [D^\alpha f]_{\sigma,p,\Omega}^p \right)^{1/p}.$$

Here, α denotes the multindex and $|\alpha| = \sum_{j=1}^d \alpha_j$. It is worth mentioning 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. [22, Remark 5] and [115, Theorem 1].

For the point-wise error estimates in Section 5.5.3 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 A.27. *For any $p \in [1,\infty)$ and $\theta \in (0,1)$ one has*

$$(W^{m,p}(\mathbb{R}^d), W^{m+1,p}(\mathbb{R}^d))_{\theta,p} \hookrightarrow W^{m+\theta,p}(\mathbb{R}^d), \quad m \in \mathbb{N}, \tag{A.25}$$

where the embedding constant is bounded by

$$\left(\min\{\theta, 1-\theta\}\, p + 2^{2p} c_{d,m} \right)^{1/p},$$

and $c_{d,m}$ exclusively depends on the spatial dimension d and the parameter m.

Proof. We follow the proof of [110, Example 1.1.8]; cf. also [1, Theorem 7.47].

Step 1: $m = 0$. Let $u \in (L^p(\mathbb{R}^d), W^{1,p}(\mathbb{R}^d))_{\theta,p}$. Consider a splitting $u = v + w$ with $v \in L^p(\mathbb{R}^d)$ and $w \in W^{1,p}(\mathbb{R}^d)$. Recall that

$$\int_{\mathbb{R}^d} |w(x+h) - w(x)|^p \,\mathrm{d}x \leq |h|^p \|\nabla w\|_{L^p}^p.$$

Therefore, using Jensen's inequality twice, we see that

$$\begin{aligned}[u]_{\theta,p}^p &\le 2^{p-1}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{|v(x+h)-v(x)|^p}{|h|^{d+\theta p}}\,\mathrm{d}x\,\mathrm{d}h + 2^{p-1}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\frac{|w(x+h)-w(x)|^p}{|h|^{d+\theta p}}\,\mathrm{d}x\,\mathrm{d}h\\ &\le \int_{\mathbb{R}^d}|h|^{-d-\theta p}\left(2^{2p-2}\|v\|_{L^p}^p + 2^{p-1}|h|^p\|w\|_{W^{1,p}}^p\right)\,\mathrm{d}h\\ &\le 2^{2p-2}\int_{\mathbb{R}^d}|h|^{-d-\theta p}\left(\|v\|_{L^p}+|h|\|w\|_{W^{1,p}}\right)^p\,\mathrm{d}h.\end{aligned}$$

Hence, by means of the definition of the K-functional, we obtain

$$\begin{aligned}[u]_{\theta,p}^p &\le 2^{2p-2}\int_{\mathbb{R}^d}|h|^{-d-\theta p}K(|h|,u)^p\,\mathrm{d}h\\ &\le 2^{2p-2}\int_0^\infty t^{-1-\theta p}K(t,u)^p\,\mathrm{d}t\int_{\partial\mathcal{B}_1(0)}\mathrm{d}\sigma = 2^{2p-2}c_d\|u\|_{\theta,p}^p,\end{aligned}$$

where the constant c_d exclusively depends on the spatial dimension d. 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 A.3. 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 spaces 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 $[u]_{\theta,p}$. Thus,

$$[D^\alpha u]_{\theta,p}^p\le 2^{2p-2}\int_{\mathbb{R}^d}|h|^{-d-\theta p}\left(\|v\|_{W^{m,p}}+|h|\|w\|_{W^{m+1,p}}\right)^p\,\mathrm{d}h.$$

Using that the number of multiindices with $|\alpha|=m$ only depends on d and m, the $W^{\theta,p}$-norm can be estimated as in the first step with a constant $c_{d,m}$ (instead of c_d). □

Lemma A.28. *For any $p\in[1,\infty)$ and $\theta\in(0,1)\setminus\{1/2\}$ one has*

$$(L^p(\mathbb{R}^d),W^{2,p}(\mathbb{R}^d))_{\theta,p}\hookrightarrow W^{2\theta,p}(\mathbb{R}^d).$$

Furthermore, the embedding constant is bounded by $c(\theta)$ that 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 $\theta\to 1/2$.

Proof. According to [114, Corollary 1.4.7.1] we have

$$\|\nabla u\|_{L^p}\le c_p\|u\|_{W^{2,p}}^{1/2}\|u\|_{L^p}^{1/2}\quad\text{for all } u\in W^{2,p}(\mathbb{R}^d),$$

where $c_p\le cK_d^{1/p}$ and K_d denotes the volume of the d dimensional unit ball. Thus,

$$\|u\|_{W^{1,p}}\le(1+cK_d^{1/p})\|u\|_{W^{2,p}}^{1/2}\|u\|_{L^p}^{1/2}\quad\text{for all } u\in W^{2,p}(\mathbb{R}^d).$$

Whence, the space $W^{1,p}(\mathbb{R}^d)$ belongs to the class $J_{1/2}(L^p(\mathbb{R}^d),W^{2,p}(\mathbb{R}^d))$. For these reasons, if $\theta>1/2$, then the reiteration theorem Proposition A.9 (with $\theta_0=1/2$ and $\theta_1=1$) and the embedding (A.25) 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 A.27 we obtain the asymptotic behavior of the embedding constant as stated in the proposition. □

Proposition A.29. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary. For all $\theta \in (0,1) \setminus \{1/2\}$ and $p \in (1,\infty)$ the embedding*

$$(L^p(\Omega), W^{2,p}(\Omega))_{\theta,p} \hookrightarrow W^{2\theta,p}(\Omega)$$

holds. Furthermore, the embedding constant is bounded by $c(\theta)$ that 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 $\theta \to 1/2$.

For the proof, we require the extension theorem due to Stein:

Lemma A.30. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $m \in \mathbb{N}$. Then there exists an extension operator E mapping $W^{k,p}(\Omega)$ continuously into $W^{k,p}(\mathbb{R}^d)$ for all $k = 0,1,\dots,m$ and $p \in [1,\infty)$. Moreover, there is $c > 0$ such that*

$$\|Ef\|_{W^{k,p}(\mathbb{R}^d)} \le c\|f\|_{W^{k,p}(\Omega)}, \quad f \in W^{k,p}(\Omega),$$

and the constant is independent of p, k, and f.

Proof. This result is proved in [144, Theorem VI.3.5]. The bound on the norm of E as stated above can be found in [144, Chapter VI.3, Equation (32)]. □

Proof of Proposition A.29. The proof is based on the corresponding result on $\mathbb{R}^d$ by first extending the functions from Ω to $\mathbb{R}^d$ and retraction afterwards. According to Lemma A.30, there exists an extension operator $E\colon W^{k,p}(\Omega) \to W^{k,p}(\mathbb{R}^d)$ for all $k = 0,1,2$ and its norm is independent of p. Hence

$$\begin{aligned}\|f\|_{W^{2\tau,p}(\Omega)} &= \|Ef\|_{W^{2\tau,p}(\Omega)} \le \|Ef\|_{W^{2\tau,p}(\mathbb{R}^d)} \\ &\le c(\tau)\|Ef\|_{(L^p(\mathbb{R}^d),W^{2,p}(\mathbb{R}^d))_{\tau,p}} \le c(\tau)\|f\|_{(L^p(\mathbb{R}^d),W^{2,p}(\mathbb{R}^d))_{\tau,p}},\end{aligned}$$

where we have used the interpolation result Lemma A.28 on $\mathbb{R}^d$ in the second inequality and a general interpolation principle for linear operators, see, e.g., [146, 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. □

A.7. Discretization error estimates for the state equation

We collect general discretization error estimates for the state equation. To this end, we first summarize error estimates for finite element discretizations of elliptic equations. For further information we refer to, e.g., the monographs [24, 38].

Consider a discretization of the convex and polygonal domain $\Omega \subset \mathbb{R}^d$, $d \in \{2, 3\}$, consisting of triangular or tetrahedral cells K that constitute a non-overlapping cover of the domain. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$. Let h_K denote the diameter of the cell $K \in \mathcal{T}_h$ and let ρ_K denote the diameter of the largest ball that can be inscribed in K. We define the discretization parameter h as the cellwise constant function $h|_K = h_K$. Simultaneously, we denote by h the maximal diameter, i.e. $h = \max h_K$.

Definition A.31. Let $\{\mathcal{T}_h\}_{h>0}$ be a family of triangulations.

(i) The family is called *regular*, if there exists a constant $\sigma > 0$ such that $\rho_K \geq \sigma h_K$ for all cells $K \in \mathcal{T}_h$ and $h \in (0, 1]$.

(ii) The family is called *quasi-uniform*, if the exists a $\sigma > 0$ such that $\rho_K \geq \sigma h$ for all cells $K \in \mathcal{T}_h$ and $h \in (0, 1]$.

Associated with the mesh $\mathcal{T}_h$, we define $V_h \subset H_0^1$ as the subspace of continuous and cellwise linear functions. Let $\mathrm{I}_h \colon C(\overline{\Omega}) \to V_h$ denote the Lagrange interpolant on Ω; see, e.g., [24, Definition 3.3.9]. We have the following interpolation error estimate.

Proposition A.32 ([24, Theorem 4.4.20])**.** *Let $\{\mathcal{T}_h\}_{h>0}$ be a family of regular triangulations. Then there is $c > 0$ such that*

$$\|u - \mathrm{I}_h u\|_{L^2} \leq ch\|u - \mathrm{I}_h u\|_{H^1} \leq ch^2 \|\nabla^2 u\|_{L^2}, \quad u \in H^2 \cap H_0^1. \tag{A.26}$$

Moreover, we require interpolation error estimates with fractional Sobolev spaces; see Appendix A.6. Note that the following estimate also follows from the corresponding result for Sobolev spaces with integer differentiability index and real interpolation. However, since we are interested in estimates that are uniform in τ and p, we directly use the fractional norm to avoid the identification of the real interpolation space with the fractional Sobolev space.

Proposition A.33. *Let $\{\mathcal{T}_h\}_{h>0}$ be a family of regular triangulations, $p \in [1, \infty)$, and $2 > \tau > d/p$. Then for all $u \in W^{\tau,p} \cap H_0^1$, the interpolation error estimate*

$$\begin{aligned} \|u - \mathrm{I}_h u\|_{L^p} + h\|\nabla (u - \mathrm{I}_h u)\|_{L^p} &\leq ch^\tau \|u\|_{W^{\tau,p}}, \\ \|u - \mathrm{I}_h u\|_{L^\infty} &\leq ch^{\tau - d/p} \|u\|_{W^{\tau,p}}. \end{aligned}$$

is valid, where the constant $c > 0$ is independent of h, p, and u. Moreover, for fixed $\tau_\bullet > d/p$ the constant c can be chosen uniformly with respect to $\tau \in [\tau_\bullet, 1)$.

Proof. This well-known result is proved by transformation to a reference cell and using [51, Theorem 6.1]. Back transformation to the cell K follows as in [51, Example 3]. □

Next, we define the Ritz projection $R_h \colon H_0^1 \to V_h$ by

$$(\nabla(u - R_h u), \nabla \varphi_h) = 0 \quad \text{for all } \varphi_h \in V_h.$$

The following projection error estimate is valid.

Proposition A.34. *Let $\{\mathcal{T}_h\}_{h>0}$ be a family of regular triangulations. Then there is $c > 0$ such that*

$$\|u - R_h u\|_{L^2} \le ch\|u - R_h u\|_{H^1} \le ch^2 \|\nabla^2 u\|_{L^2}, \quad u \in H^2 \cap H_0^1. \tag{A.27}$$

Moreover, the estimate

$$\|u - R_h u\|_{L^2} \le ch^{1+\tau} \|u\|_{W^{1+\tau,2}}, \quad u \in W^{\tau,2} \cap H_0^1, \tag{A.28}$$

holds for all $2 > \tau > d/2$.

Proof. The first estimate is proved in [24, Theorem 5.4.8] based on (A.26). The second estimate follows from the first and Proposition A.33. □

Define the spatial L^2-projection $\Pi_h \colon L^2 \to V_h$ by

$$(u - \Pi_h u, \varphi)_{L^2} = 0 \quad \text{for all } \varphi \in V_h.$$

We have the following projection error estimate.

Proposition A.35. *Let $\{\mathcal{T}_h\}_{h>0}$ be a family of regular triangulations. Then there is $c > 0$ such that*

$$\|u - \Pi_h u\|_{L^2} \le ch^2 \|\nabla^2 u\|_{L^2}, \quad u \in H^2 \cap H_0^1. \tag{A.29}$$

If in addition, the projection Π_h is stable in H^1, then

$$\|\nabla(u - \Pi_h u)\|_{L^2} \le ch\|\nabla^2 u\|_{L^2}, \quad u \in H^2 \cap H_0^1. \tag{A.30}$$

For quasi-uniform meshes, the stability of Π_h in H^1 directly follows from an inverse estimate and an error estimate for Π_h in L^2. However, weaker conditions are known such as local quasi-uniformity; cf. [23].

Proof of Proposition A.35. For the estimate (A.29) we use the best approximation property of Π_h in L^2 and the error estimate (A.26) for I_h to deduce that

$$\|u - \Pi_h u\|_{L^2} \le \|u - \mathrm{I}_h u\|_{L^2} \le ch^2 \|\nabla^2 u\|_{L^2}.$$

To show (A.30), we calculate

$$\begin{aligned} \|\nabla(u - \Pi_h u)\|_{L^2} &\le \|\nabla(u - \Pi_h R_h u)\|_{L^2} + \|\nabla \Pi_h (u - R_h u)\|_{L^2} \\ &\le (1 + c)\|\nabla(u - R_h u)\|_{L^2} \le (1 + c)h\|\nabla^2 u\|_{L^2}, \end{aligned}$$

where we have used the projection property and the stability of Π_h as well as the error estimate (A.27) for R_h. □

Discretization error estimates for the state in $L^2(I;L^2)$ and $L^2(I;H^1)$

Our objective is to show the following error estimates for the state equation measured in $L^2(I;L^2)$ and $L^2(I;H^1)$. Thereafter, we provide discretization error estimates for the state evaluated at the terminal time. We suppose throughout that $\{\,\mathcal{T}_h\,\}_{h>0}$ is a family of regular triangulations.

Lemma A.36. *Let $\nu \in \mathbb{R}_+$ and $f \in L^2((0,1);L^2)$. For the solution $u = u(\nu, f)$ to the state equation with right-hand side f and the discrete solution $u_{kh} = u_{kh}(\nu, f)$ to equation (5.4) with right-hand side f the estimate*

$$\|u - u_{kh}\|_{L^2(I;L^2)} \le c\left(k\|\partial_t u\|_{L^2(I;L^2)} + h^2\|\Delta u\|_{L^2(I;L^2)}\right) \tag{A.31}$$

holds. If additionally Π_h is stable in H^1, then

$$\|\nabla u - \nabla u_{kh}\|_{L^2(I;L^2)} \le c(k^{1/2} + h)\left(\|\partial_t u\|_{L^2(I;L^2)} + \|\Delta u\|_{L^2(I;L^2)}\right). \tag{A.32}$$

The constant $c > 0$ is independent of k, h, ν, f, u_0, u, and u_{kh}.

To discuss the error due to temporal and spatial discretization separately, let us introduce the nodal interpolation $i_k \colon C([0,1];H_0^1) \to X_k$ as

$$i_k u(t_m) = u(t_m), \quad m = 1, 2, \dots, M, \tag{A.33}$$

where

$$X_k = \left\{v_k \in L^2(I;H_0^1) \colon v_k|_{I_m} \in \mathcal{P}_0(I_m;H_0^1),\, m = 1,2,\dots,M\right\}$$

is the semi-discrete state space. The following interpolation error estimates are valid.

Proposition A.37 ([130, Lemma A.7])**.** *If $u \in H^1(I;L^2) \cap L^2(I;H^2 \cap H_0^1)$, then*

$$\|u - i_k u\|_{L^2(I;L^2)} \le ck\|\partial_t u\|_{L^2(I;L^2)}, \tag{A.34}$$

$$\|u - i_k u\|_{L^2(I;H^1)} \le ck^{1/2}\left(\|\partial_t u\|_{L^2(I;L^2)} + \|\Delta u\|_{L^2(I;L^2)}\right). \tag{A.35}$$

The constant $c > 0$ is independent of k, ν, and u.

Let u be the solution to the continuous state equation for $(\nu, q) \in \mathbb{R}_+ \times Q(0,1)$ and $u_{kh} \in X_{k,h}$ the corresponding discrete solution to (5.4). Using a density argument one may show that u satisfies the discrete equation (5.4) as well. Therefore, the Galerkin orthogonality

$$\mathrm{B}(\nu, u - u_{kh}, \varphi_{kh}) = 0 \quad \text{for all } \varphi_{kh} \in X_{k,h} \tag{A.36}$$

holds. We consider the splitting

$$u - u_{kh} = u - \Pi_h i_k u + \Pi_h i_k u - u_{kh} =: \zeta_{kh} + \xi_{kh}.$$

Proposition A.38. *Let the terms ζ_{kh} and ξ_{kh} be defined as above. Then*

$$\mathrm{B}(\nu, \zeta_{kh}, \varphi_{kh}) = \nu(\nabla\zeta_{kh}, \nabla\varphi_{kh})_{L^2(I;L^2)} \quad \textit{for all } \varphi_{kh} \in X_{k,h}, \tag{A.37}$$

and

$$\|\nabla\xi_{kh}\|_{L^2(I;L^2)} \le \|\nabla\zeta_{kh}\|_{L^2(I;L^2)}. \tag{A.38}$$

Proof. The follows as in [117, Section 5.1]. First note that for all $\varphi_{kh}, \psi_{kh} \in X_{k,h}$ we have

$$\mathrm{B}(\nu, \varphi_{kh}, \psi_{kh}) = (\nu \nabla \varphi_{kh}, \nabla \psi_{kh})_{L^2(I;L^2)} - \sum_{m=1}^{M-1} (\varphi_{kh,m}, [\psi_{kh}]_m) + (\varphi_{kh}(1), \psi_{kh}(1)). \quad \text{(A.39)}$$

Consider the splitting $\zeta_{kh} = u - i_k u + i_k u - \Pi_h i_k u =: \zeta_k + \zeta_h$. Then $\zeta_{k,m} = 0$ due to the definition of i_k and $(\zeta_{h,m}, [\psi_{kh}]_m) = 0$ according to the definition of Π_h. Using (A.39) we conclude the first identity (A.37). Moreover, using (5.3), we obtain

$$\mathrm{B}(\nu, \varphi_{kh}, \varphi_{kh}) = \nu (\nabla \varphi_{kh}, \nabla \varphi_{kh})_{L^2(I;L^2)} + \sum_{m=1}^{M-1} ([\varphi_{kh}]_m, \varphi_{kh,m+1}) + (\varphi_{kh,1}, \varphi_{kh,1}). \quad \text{(A.40)}$$

Thus, testing (A.39) with $\psi_{kh} = \varphi_{kh}$ and summation of (A.39) and (A.40) implies

$$\mathrm{B}(\nu, \varphi_{kh}, \varphi_{kh}) \geq \nu \|\nabla \varphi_{kh}\|^2_{L^2(I;L^2)} \quad \text{(A.41)}$$

for all $\varphi_{kh} \in X_{k,h}$. Therefore, using Galerkin orthogonality we find

$$\nu \|\nabla \xi_{kh}\|^2_{L^2(I;L^2)} \leq \mathrm{B}(\nu, \xi_{kh}, \xi_{kh}) = -\mathrm{B}(\nu, \zeta_{kh}, \xi_{kh}) = -\nu (\nabla \zeta_{kh}, \nabla \xi_{kh})_{L^2(I;L^2)},$$

where we have used (A.37) in the last step and $\xi_{kh} \in X_{k,h}$. Finally, (A.38) follows from the Cauchy-Schwarz inequality. □

Proof of Lemma A.36, Estimate (A.31). The estimate (A.31) follows by standard arguments; see, e.g., [117, Section 5.1]. We give a detailed proof to clearly see the dependence on ν. Consider

$$\nu \|u - u_{kh}\|^2_{L^2(I;L^2)} = \nu (\zeta_{kh}, u - u_{kh})_{L^2(I;L^2)} + \nu (\xi_{kh}, u - u_{kh})_{L^2(I;L^2)} =: J_1 + J_2.$$

Using the Cauchy-Schwarz inequality and stability of the projection Π_h in L^2, we find

$$\begin{aligned} J_1 &\leq \nu \|\zeta_{kh}\|_{L^2(I;L^2)} \|u - u_{kh}\|_{L^2(I;L^2)} \\ &\leq \nu \left(\|u - \Pi_h u\|_{L^2(I;L^2)} + \|u - i_k u\|_{L^2(I;L^2)} \right) \|u - u_{kh}\|_{L^2(I;L^2)}. \end{aligned}$$

To estimate J_2, consider $\tilde{z}_{kh} \in X_{k,h}$ the solution to

$$\mathrm{B}(\nu, \varphi_{kh}, \tilde{z}_{kh}) = \nu (\varphi_{kh}, u - u_{kh})_{L^2(I;L^2)}, \quad \varphi_{kh} \in X_{k,h}.$$

Due to Galerkin orthogonality, (A.37), the properties of the Ritz projection R_h, and the definition of the discrete Laplacian $-\Delta_h$, we obtain that

$$\begin{aligned} \nu (\xi_{kh}, u - u_{kh})_{L^2(I;L^2)} &= \mathrm{B}(\nu, \xi_{kh}, \tilde{z}_{kh}) = -\mathrm{B}(\nu, \zeta_{kh}, \tilde{z}_{kh}) \\ &= -\nu (\nabla (u - \Pi_h i_k u), \nabla \tilde{z}_{kh})_{L^2(I;L^2)} \\ &= \nu (R_h u - \Pi_h i_k u, \Delta_h \tilde{z}_{kh})_{L^2(I;L^2)} \\ &\leq \nu \|\Pi_h (R_h u - i_k u)\|_{L^2(I;L^2)} \|\Delta_h \tilde{z}_{kh}\|_{L^2(I;L^2)} \\ &\leq \nu \|R_h u - i_k u\|_{L^2(I;L^2)} \|u - u_{kh}\|_{L^2(I;L^2)} \\ &\leq \nu \left(\|R_h u - u\|_{L^2(I;L^2)} + \|u - i_k u\|_{L^2(I;L^2)} \right) \|u - u_{kh}\|_{L^2(I;L^2)}, \end{aligned}$$

where we have used the stability estimate (5.10) for $\Delta_h \tilde{z}_{kh}$ and stability of the projection Π_h in L^2. Employing the interpolation and projection error estimates (A.34), (A.29), and (A.27) we obtain

$$\begin{aligned} J_1 + J_2 &\leq \nu c \left(\|u - \Pi_h u\|_{L^2(I;L^2)} + \|u - i_k u\|_{L^2(I;L^2)} + \|R_h u - u\|_{L^2(I;L^2)} \right) \|u - u_{kh}\|_{L^2(I;L^2)} \\ &\leq c\nu \left(k \|\partial_t u\|_{L^2(I;L^2)} + h^2 \|\nabla^2 u\|_{L^2(I;L^2)} \right) \|u - u_{kh}\|_{L^2(I;L^2)}. \end{aligned}$$

Finally, elliptic regularity theory yields the estimate $\|\nabla^2 u\|_{L^2} \leq c\|\Delta u\|_{L^2}$, see, e.g., [68, Theorem 3.1.1.2], completing the proof of (A.31).

Estimate (A.32). We observe

$$\begin{aligned} \|\nabla u - \nabla u_{kh}\|^2_{L^2(I;L^2)} &= (\nabla u - \nabla u_{kh}, \nabla \zeta_{kh})_{L^2(I;L^2)} + (\nabla u - \nabla u_{kh}, \nabla \xi_{kh})_{L^2(I;L^2)} \\ &\leq \|\nabla u - \nabla u_{kh}\|_{L^2(I;L^2)} \left(\|\nabla \zeta_{kh}\|_{L^2(I;L^2)} + \|\nabla \xi_{kh}\|_{L^2(I;L^2)} \right) \\ &\leq 2 \|\nabla u - \nabla u_{kh}\|_{L^2(I;L^2)} \|\nabla \zeta_{kh}\|_{L^2(I;L^2)}, \end{aligned}$$

where we have used (A.38). From the stability of the L^2-projection in H^1 and the interpolation and projection error estimates (A.35) and (A.30) we obtain

$$\begin{aligned} \|\nabla \zeta_{kh}\|_{L^2(I;L^2)} &\leq \|\nabla (u - \Pi_h u)\|_{L^2(I;L^2)} + \|\nabla \Pi_h (u - i_k u)\|_{L^2(I;L^2)} \\ &\leq c \left(h \|\nabla^2 u\|_{L^2(I;L^2)} + k^{1/2} \left(\|\partial_t u\|_{L^2(I;L^2)} + \|\Delta u\|_{L^2(I;L^2)} \right) \right). \end{aligned}$$

Again elliptic regularity theory yields (A.32). □

Discretization error estimates for the state at the terminal time

Furthermore, we require estimates for the discretization error at the terminal time that will be verified subsequently. We generally suppose that the regularity conditions concerning the temporal mesh from Section 5.2 are valid and $\{ \mathcal{T}_h \}_{h>0}$ is a family of regular triangulations.

Lemma A.39. *Let $\nu \in \mathbb{R}_+$ and $f \in L^\infty((0,1);L^2)$. For the solution $u = u(\nu, f)$ to the state equation with right-hand side f and the discrete solution $u_{kh} = u_{kh}(\nu, f)$ to equation (5.4) with right-hand side f the estimates*

$$\|u - u_{kh}\|_{L^\infty(I;L^2)} \leq c|\log k| \left(k + h^2 \right) \left((1+\nu) \|f\|_{L^\infty(I;L^2)} + \nu^{-1} \|u_0\|_{L^2} \right) \tag{A.42}$$

$$\|u - u_{kh}\|_{L^\infty(I;L^2)} \leq c|\log k| \left(k + h^2 \right) (1+\nu) \left(\|f\|_{L^\infty(I;L^2)} + \|\Delta u_0\|_{L^2} \right) \tag{A.43}$$

hold, where the constant $c > 0$ is independent of k, h, ν, f, u_0, u, and u_{kh}.

To prove the estimates, we need several auxiliary results for solutions to dual equations.

Proposition A.40. *For $z_1 \in L^2$ let $z \in H^1(I;L^2) \cap L^2(I;H^2 \cap H_0^1)$ the continuous and $z_k \in X_k$ denote the semidiscrete adjoint state with let $z(1) = z_1$ and $z_k(1) = z_1$, i.e.*

$$\mathrm{B}(\nu, \varphi_k, z_k) = (z_1, \varphi_{k,M}) \quad \text{for all } \varphi_k \in X_k.$$

Then

$$\|z - z_k\|_{L^1(I;L^2)} \leq c \left(1 + \nu^{-1/2} \right) k |\log k| \|z_1\|_{L^2}, \tag{A.44}$$

$$\|z(0) - z_{k,1}\|_{H^{-2}} \leq c\nu k \|z_1\|_{L^2}, \tag{A.45}$$

where the constant $c > 0$ is independent of k, ν, z_1, z, and z_k.

Proof. This is the assertion of [116, Lemma 5.2] and we give the proof to clearly see the dependence on ν. To this end, we introduce the nodal interpolation as

$$i_k^*\colon C([0,1];H_0^1) \to X_k, \quad i_k^* u|_{I_m} = u(t_{m-1}), \quad m = 1,2,\dots,M.$$

Define $\zeta_k \coloneqq i_k^* z - z_k$. By means of Galerkin orthogonality and the definition of i_k^*, for all $\varphi_k \in X_k \cap L^2(I;H^2)$ it holds

$$\begin{aligned}
\mathrm{B}(\nu,\varphi_k,\zeta_k) &= -\mathrm{B}(\nu,\varphi_k,z-i_k^*z) = \nu\int_0^1 (\Delta\varphi_k, z-i_k^*z)_{L^2}\,\mathrm{d}t \\
&= \nu\sum_{m=1}^M \left(\int_{I_m} (\Delta\varphi_{k,m}, z(t))_{L^2}\,\mathrm{d}t - k_m(\Delta\varphi_k, z(t_{m-1}))_{L^2}\right) \\
&= \nu\sum_{m=1}^M \int_{I_m} (t_m - t)(\Delta\varphi_{k,m}, \partial_t z(t))_{L^2}\,\mathrm{d}t.
\end{aligned}$$

This expression is equivalent to the following set of equations

$$\nu\int_{I_m} (\nabla\varphi, \nabla\zeta_k)_{L^2} - (\varphi_m, [\zeta_k]_m)_{L^2} = \nu\int_{I_m} (t_m - t)(\Delta\varphi, \partial_t z(t))_{L^2}\,\mathrm{d}t, \tag{A.46}$$

for all $\varphi \in \mathcal{P}_0(I_m; H^2\cap H_0^1)$ and for all $m = 1,2,\dots,M$.

Estimate (A.45). Testing in (A.46) with $\varphi = \Delta^{-2}\zeta_k$, integrating by parts, and using the identity $\partial_t z = -\nu\Delta z$ we find

$$-\nu\int_{I_m} (\Delta^{-1}\zeta_k, \zeta_k)_{L^2} - (\Delta^{-1}\zeta_k, [\Delta^{-1}\zeta_k]_m)_{L^2} = -\nu^2\int_{I_m} (t_m - t)(\zeta_k, z(t))_{L^2}\,\mathrm{d}t. \tag{A.47}$$

The right-hand side can be estimated as

$$\begin{aligned}
-\nu^2\int_{I_m} (t_m - t)(\zeta_k, z(t))_{L^2}\,\mathrm{d}t &= \nu^2\int_{I_m} (t_m - t)(\nabla\Delta^{-1}\zeta_k, \nabla z(t))_{L^2}\,\mathrm{d}t \\
&\le \frac{\nu}{2}\int_{I_m} \|\nabla\Delta^{-1}\zeta_k\|_{L^2}^2 + \frac{\nu^3 k_m^2}{2}\int_{I_m} \|\nabla z(t)\|_{L^2}^2\,\mathrm{d}t.
\end{aligned}$$

Applying the identity

$$\frac{1}{2}\left(\|\varphi_{m+1}\|_{L^2}^2 - \|[\varphi]_m\|_{L^2}^2 - \|\varphi_m\|_{L^2}^2\right) = ([\varphi]_m, \varphi_m)_{L^2} \tag{A.48}$$

to the left-hand side of (A.47) and using that

$$-\nu\int_{I_m} (\Delta^{-1}\zeta_k, \zeta_k)_{L^2} = \nu\int_{I_m} \|\nabla\Delta^{-1}\zeta_k\|_{L^2}^2,$$

yield

$$\|\Delta^{-1}\zeta_{k,m}\|_{L^2}^2 + \nu\int_{I_m} \|\nabla\Delta^{-1}\zeta_k\|_{L^2}^2 \le \|\Delta^{-1}\zeta_{k,m+1}\|_{L^2}^2 + \nu^3 k^2\int_{I_m} \|\nabla z(t)\|_{L^2}^2\,\mathrm{d}t.$$

Summation of the above inequality for all $m = 1,\dots,M$, the stability estimate (A.19), as well as equivalence of the norms $\|\Delta^{-1}\cdot\|_{L^2}$ and $\|\cdot\|_{H^{-2}}$ imply

$$\|\zeta_{k,1}\|_{H^{-2}} + \nu\|\nabla\Delta^{-1}\zeta_k\|_{L^2(I;L^2)} \le c\nu^{3/2}k\|\nabla z\|_{L^2(I;L^2)} \le c\nu k\|z_1\|_{L^2}. \tag{A.49}$$

Since $z(0) - z_{k,1} = \zeta_{k,1}$, this concludes the proof of (A.45).

Estimate (A.44). First, it holds

$$\begin{aligned}\|z - i_k^* z\|_{L^1(I;L^2)} &\leq \int_{I\setminus I_M} \|z - i_k^* z\|_{L^2} + \int_{I_M} \|z - i_k^* z\|_{L^2} \\ &\leq ck \left(\int_{I\setminus I_M} \|\partial_t z\|_{L^2} + \sup_{t\in I} \|z(t)\|_{L^2} \right).\end{aligned}$$

Hence, [116, Theorem 4.4] and the stability estimate (A.19) imply

$$\|z - i_k^* z\|_{L^1(I;L^2)} \leq ck|\log k|^{1/2} \|z_1\|_{L^2}.$$

Testing in (A.46) with $\varphi = -\Delta^{-1}\zeta_k$ after integrating by parts yields

$$\nu \int_{I_m} \|\zeta_k\|_{L^2}^2 - (\nabla\Delta^{-1}\zeta_{k,m}, [\nabla\Delta^{-1}\zeta_k]_m)_{L^2} = -\nu \int_{I_m} (t_m - t)(\zeta_k, \partial_t z(t))_{L^2}.$$

Estimating the right-hand side as

$$-\nu \int_{I_m} (t_m - t)(\zeta_k, \partial_t z(t))_{L^2} \leq \frac{\nu}{2} \int_{I_m} \|\zeta_k\|_{L^2}^2 + \frac{\nu}{2} \int_{I_m} (t_m - t)^2 \|\partial_t z\|_{L^2}^2,$$

and applying the identity (A.48) implies

$$\nu \int_{I_m} \|\zeta_k\|_{L^2}^2 + \|\nabla\Delta^{-1}\zeta_{k,m}\|_{L^2}^2 \leq \|\nabla\Delta^{-1}\zeta_{k,m+1}\|_{L^2}^2 + \nu \int_{I_m} (t_m - t)^2 \|\partial_t z\|_{L^2}^2.$$

Multiplication by $(1 - t_{m-1})$, using that $(1 - t_m) = (1 - t_{m-1}) + k_m$, and summation for all $m = 1, 2, \ldots, M$ yield

$$\begin{aligned}\nu \sum_{m=1}^{M} (1 - t_{m-1}) \int_{I_m} \|\zeta_k\|_{L^2}^2 + \|\nabla\Delta^{-1}\zeta_{k,1}\|_{L^2}^2 \\ \leq \sum_{m=1}^{M} k_m \|\nabla\Delta^{-1}\zeta_{k,m+1}\|_{L^2}^2 + \nu \sum_{m=1}^{M} (1 - t_{m-1}) \int_{I_m} (t_m - t)^2 \|\partial_t z\|_{L^2}^2.\end{aligned}$$

Moreover, since $k_m \leq k_{\text{ratio}} k_{m+1}$, and using (A.49) we estimate

$$\sum_{m=1}^{M} k_m \|\nabla\Delta^{-1}\zeta_{k,m+1}\|_{L^2}^2 \leq k_{\text{ratio}} \|\nabla\Delta^{-1}\zeta_k\|_{L^2(I;L^2)}^2 \leq ck^2 \|z_1\|_{L^2}^2.$$

For any $m \leq M - 1$ and $t \in I_m$, we have

$$1 - t_{m-1} \leq 1 - t_m + k_{\text{ratio}} k_{m+1} \leq (1 - t_m)(1 + k_{\text{ratio}}) \leq (1 - t)(1 + k_{\text{ratio}}).$$

Hence,

$$\begin{aligned}\sum_{m=1}^{M} (1 - t_{m-1}) \int_{I_m} (t_m - t)^2 \|\partial_t z\|_{L^2}^2 \leq \sum_{m=1}^{M-1} k_m^2 \int_{I_m} (1 - t_{m-1}) \|\partial_t z\|_{L^2}^2 \\ + k_M^2 \int_{I_M} (1 - t) \|\partial_t z\|_{L^2}^2 \leq (1 + k_{\text{ratio}}) k^2 \int_I (1 - t) \|\partial_t z\|_{L^2}^2.\end{aligned}$$

Therefore, from [116, Theorem 4.4] we infer

$$\sum_{m=1}^{M}(1-t_{m-1})\int_{I_m}\|\zeta_k\|_{L^2}^2 \le c\left(1+\nu^{-1}\right)k^2\|z_1\|_{L^2}^2.$$

In summary, we have

$$\begin{aligned}\|\zeta_k\|_{L^1(I;L^2)}^2 &\le \left(\sum_{m=1}^{M}\frac{k_m}{1-t_{m-1}}\right)\left(\sum_{m=1}^{M}(1-t_{m-1})k_m\|\zeta_{k,m}\|_{L^2}^2\right)\\ &\le c\left(1+\nu^{-1}\right)|\log k|k^2\|z_1\|_{L^2}^2,\end{aligned}$$

where we have used $k \le 1/4$ in the last step. Finally, we obtain

$$\begin{aligned}\|z-z_k\|_{L^1(I;L^2)} &\le \|z-i_k^*z\|_{L^1(I;L^2)} + \|i_k^*z-z_k\|_{L^1(I;L^2)}\\ &\le c\left(1+\nu^{-1/2}\right)|\log k|k\|z_1\|_{L^2},\end{aligned}$$

concluding the proof. □

Proposition A.41. *For $z_1 \in L^2$ let $z_k \in X_k$ and $z_{kh} \in X_{k,h}$ denote the semidiscrete, respectively, discrete adjoint state with $z_k(1) = z_1$ and $z_{kh}(1) = z_1$. Then*

$$\|z_k - z_{kh}\|_{L^1(I;L^2)} \le c|\log k|h^2\|z_1\|_{L^2}, \tag{A.50}$$

$$\|z_{k,1} - z_{kh,1}\|_{H^{-2}} \le ch^2\|z_1\|_{L^2}, \tag{A.51}$$

where the constant $c > 0$ is independent of k, h, ν, z_1, z_k, and z_{kh}.

Proof. Estimate (A.51) is proved in [116, Lemma 5.8] with a constant $c > 0$ that can be checked to be independent of k, h, ν, z_1, z_k, and z_{kh}. Estimate (A.50) is proved as in [116, Theorem 5.10]: Using [116, Lemmas 5.9 and 5.8] we have

$$\begin{aligned}\|z_k - z_{kh}\|_{L^1(I;L^2)} &\le \sum_{m=1}^{M} k_m(1-t_{m-1})^{-1}\max_{m=1,2,\dots,M}\|z_{k,m}-z_{kh,m}\|_{L^2}\\ &\le c|\log k|h^2\|z_{k,M}-z_{kh,M}\|_{L^2} \le c|\log k|h^2\|z_1\|_{L^2},\end{aligned}$$

where we have used that $k \le 1/4$ in the second last step. □

Proof Lemma A.39, Estimate (A.43). For simplicity, we only consider the last time interval. Let $\tilde z \in H^1(I;L^2)\cap L^2(I;H^2\cap H_0^1)$ and $\tilde z_{kh} \in X_{k,h}$ be the solutions to the adjoint equation with $\tilde z(1) = \tilde z_{kh}(1) = u(1) - u_{kh}(1)$. Due to Galerkin orthogonality we obtain

$$\begin{aligned}\|u(1)-u_{kh}(1)\|_{L^2}^2 &= \mathrm{B}(\nu, u-u_{kh}, \tilde z) = \mathrm{B}(\nu, u-u_{kh}, \tilde z - \tilde z_{kh}) = \mathrm{B}(\nu, u, \tilde z - \tilde z_{kh})\\ &= \nu\int_0^1 (f, \tilde z - \tilde z_{kh})_{L^2} + (u_0, \tilde z(0) - \tilde z_{kh}(0))\\ &\le \nu\|f\|_{L^\infty(I;L^2)}\|\tilde z - \tilde z_{kh}\|_{L^1(I;L^2)} + \|\Delta u_0\|_{L^2}\|\tilde z(0) - \tilde z_{kh,1}\|_{H^{-2}}.\end{aligned}$$

Propositions A.40 and A.41 and dividing by $\|u(1)-u_{kh}(1)\|_{L^2}$ imply the result, where we have used the estimate $\nu^{1/2} + \nu \le 1 + 2\nu$.

Estimate (A.42). Consider first the case $u_0 = 0$. Then this is exactly (A.43). In the case $q = 0$, we combine Theorems 1 and 2 from [111] with clearly stated time dependency. Superposition of both estimates yields (A.42). □

Bibliography

[1] R. A. Adams and J. J. F. Fournier. *Sobolev spaces*. Second. Vol. 140. Pure and Applied Mathematics. Elsevier/Academic Press, Amsterdam, 2003, pp. xiv+305.

[2] K. Altmann, S. Stingelin, and F. Tröltzsch. "On some optimal control problems for electrical circuits". In: *International Journal of Circuit Theory and Applications* 42.8 (2014), pp. 808–830. DOI: 10.1002/cta.1889.

[3] H. Amann. "Linear parabolic problems involving measures". In: *RACSAM. Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat.* 95.1 (2001), pp. 85–119.

[4] H. Amann. *Linear and quasilinear parabolic problems. Vol. I*. Vol. 89. Monographs in Mathematics. Abstract linear theory. Birkhäuser Boston, Inc., Boston, MA, 1995, pp. xxxvi+335. DOI: 10.1007/978-3-0348-9221-6.

[5] H. Amann and J. Escher. *Analysis. I*. Birkhäuser Verlag, Basel, 2005, pp. xiv+426. DOI: 10.1007/b137107.

[6] W. Arendt and A. F. M. ter Elst. "From forms to semigroups". In: *Spectral theory, mathematical system theory, evolution equations, differential and difference equations*. Vol. 221. Oper. Theory Adv. Appl. Birkhäuser/Springer Basel AG, Basel, 2012, pp. 47–69. DOI: 10.1007/978-3-0348-0297-0_4.

[7] A. Ashyralyev and P. E. Sobolevskiı̆. *Well-posedness of parabolic difference equations*. Vol. 69. Operator Theory: Advances and Applications. Birkhäuser Verlag, Basel, 1994, pp. xiv+349. DOI: 10.1007/978-3-0348-8518-8.

[8] P. Auscher, N. Badr, R. Haller-Dintelmann, and J. Rehberg. "The square root problem for second-order, divergence form operators with mixed boundary conditions on L^p". In: *J. Evol. Equ.* 15.1 (2015), pp. 165–208. DOI: 10.1007/s00028-014-0255-1.

[9] M. Badra and T. Takahashi. "On the Fattorini criterion for approximate controllability and stabilizability of parabolic systems". In: *ESAIM Control Optim. Calc. Var.* 20.3 (2014), pp. 924–956. DOI: 10.1051/cocv/2014002.

[10] V. Barbu. *Analysis and control of nonlinear infinite-dimensional systems*. Vol. 190. Mathematics in Science and Engineering. Academic Press, Boston, MA, 1993, pp. x+476.

[11] V. Barbu. "The time optimal control of Navier-Stokes equations". In: *Systems Control Lett.* 30.2-3 (1997), pp. 93–100. DOI: 10.1016/S0167-6911(96)00083-7.

[12] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. With a foreword by Hédy Attouch. Springer, New York, 2011, pp. xvi+468. DOI: 10.1007/978-1-4419-9467-7.

[13] J. Bergh and J. Löfström. *Interpolation spaces. An introduction*. Grundlehren der Mathematischen Wissenschaften, No. 223. Springer-Verlag, Berlin-New York, 1976, pp. x+207.

[14] D. P. Bertsekas. *Nonlinear Programming*. 2. ed. Athena Scientific, 1999.

[15] D. P. Bertsekas. "On penalty and multiplier methods for constrained minimization". In: *SIAM J. Control Optimization* 14.2 (1976), pp. 216–235.

[16] L. Bonifacius and I. Neitzel. "Second Order Optimality Conditions for Optimal Control of Quasilinear Parabolic Equations". In: *Math. Control Relat. Fields* 8 (1 2018), pp. 1–34. DOI: 10.3934/mcrf.2018001.

[17] L. Bonifacius, K. Pieper, and B. Vexler. "A priori Error Estimates for Space-Time Finite Element Discretization of Parabolic Time-Optimal Control Problems". In: *ArXiv e-prints* (Feb. 2018). arXiv: 1802.00611 [math.OC].

[18] L. Bonifacius and K. Pieper. "Strong Stability of Linear Parabolic Time-Optimal Control Problems". In: *ESAIM: Control, Optimisation and Calculus of Variations* (). DOI: 10.1051/cocv/2017079.

[19] F. Bonnans and E. Casas. "An extension of Pontryagin's principle for state-constrained optimal control of semilinear elliptic equations and variational inequalities". In: *SIAM J. Control Optim.* 33.1 (1995), pp. 274–298. DOI: 10.1137/S0363012992237777.

[20] J. F. Bonnans and A. Shapiro. "Optimization problems with perturbations: a guided tour". In: *SIAM Rev.* 40.2 (1998), pp. 228–264.

[21] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000, pp. xviii+601. DOI: 10.1007/978-1-4612-1394-9.

[22] J. Bourgain, H. Brezis, and P. Mironescu. "Another look at Sobolev spaces". In: *Optimal control and partial differential equations*. IOS, Amsterdam, 2001, pp. 439–455.

[23] J. H. Bramble, J. E. Pasciak, and O. Steinbach. "On the stability of the L^2 projection in $H^1(\Omega)$". In: *Math. Comp.* 71.237 (2002), 147–156 (electronic). DOI: 10.1090/S0025-5718-01-01314-X.

[24] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*. Third. Vol. 15. Texts in Applied Mathematics. Springer, New York, 2008, pp. xviii+397. DOI: 10.1007/978-0-387-75934-0.

[25] J. V. Burke. "Calmness and exact penalization". In: *SIAM J. Control Optim.* 29.2 (1991), pp. 493–497. DOI: 10.1137/0329027.

[26] O. Cârjă. "The minimal time function in infinite dimensions". In: *SIAM J. Control Optim.* 31.5 (1993), pp. 1103–1114. DOI: 10.1137/0331051.

[27] E. Casas and K. Chrysafinos. "Analysis and optimal control of some quasilinear parabolic equations". Personal communication. 2017.

[28] E. Casas. "Second order analysis for bang-bang control problems of PDEs". In: *SIAM J. Control Optim.* 50.4 (2012), pp. 2355–2372. DOI: 10.1137/120862892.

[29] E. Casas, R. Herzog, and G. Wachsmuth. "Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional". In: *SIAM J. Optim.* 22.3 (2012), pp. 795–820. DOI: 10.1137/110834366.

[30] E. Casas, J. C. de los Reyes, and F. Tröltzsch. "Sufficient second-order optimality conditions for semilinear control problems with pointwise state constraints". In: *SIAM J. Optim.* 19.2 (2008), pp. 616–643. DOI: 10.1137/07068240X.

[31] E. Casas and F. Tröltzsch. "A general theorem on error estimates with application to a quasilinear elliptic optimal control problem". In: *Comput. Optim. Appl.* 53.1 (2012), pp. 173–206. DOI: 10.1007/s10589-011-9453-8.

[32] E. Casas and F. Tröltzsch. "Error estimates for the finite-element approximation of a semilinear elliptic control problem". In: *Control Cybernet.* 31.3 (2002). Well-posedness in optimization and related topics (Warsaw, 2001), pp. 695–712.

[33] E. Casas and F. Tröltzsch. "Second order analysis for optimal control problems: improving results expected from abstract theory". In: *SIAM J. Optim.* 22.1 (2012), pp. 261–279. DOI: 10.1137/110840406.

[34] E. Casas and F. Tröltzsch. "Second Order Optimality Conditions and Their Role in PDE Control". In: *Jahresber. Dtsch. Math.-Ver.* 117.1 (2015), pp. 3–44. DOI: 10.1365/s13291-014-0109-3.

[35] E. Casas and F. Tröltzsch. "Second-order necessary and sufficient optimality conditions for optimization problems and applications to control theory". In: *SIAM J. Optim.* 13.2 (2002), pp. 406–431. DOI: 10.1137/S1052623400367698.

[36] E. Casas, D. Wachsmuth, and G. Wachsmuth. "Second-Order Analysis and Numerical Approximation for Bang-Bang Bilinear Control Problems". In: (July 21, 2017). arXiv: 1707.06880v1 [math.OC].

[37] E. Casas, D. Wachsmuth, and G. Wachsmuth. "Sufficient Second-Order Conditions for Bang-Bang Control Problems". In: *SIAM J. Control Optim.* 55.5 (2017), pp. 3066–3090. DOI: 10.1137/16M1099674.

[38] P. G. Ciarlet. *The finite element method for elliptic problems.* Vol. 40. Classics in Applied Mathematics. Reprint of the 1978 original [North-Holland, Amsterdam; MR0520174 (58 #25001)]. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002, pp. xxviii+530.

[39] F. H. Clarke. "A new approach to Lagrange multipliers". In: *Math. Oper. Res.* 1.2 (1976), pp. 165–174.

[40] F. H. Clarke. *Functional analysis, calculus of variations and optimal control.* Vol. 264. Graduate Texts in Mathematics. Springer, London, 2013, pp. xiv+591. DOI: 10.1007/978-1-4471-4820-3.

[41] F. H. Clarke, L. Rifford, and R. J. Stern. "Feedback in state constrained optimal control". In: *ESAIM Control Optim. Calc. Var.* 7 (2002), pp. 97–133. DOI: 10.1051/cocv:2002005.

[42] G. M. Coclite and M. Garavello. "A time-dependent optimal harvesting problem with measure-valued solutions". In: *SIAM J. Control Optim.* 55.2 (2017), pp. 913–935.

[43] J. Daafouz, M. Tucsnak, and J. Valein. "Nonlinear control of a coupled PDE/ODE system modeling a switched power converter with a transmission line". In: *Systems Control Lett.* 70 (2014), pp. 92–99. DOI: 10.1016/j.sysconle.2014.05.009.

[44] N. von Daniels. "Tikhonov regularization of control-constrained optimal control problems". Apr. 2017.

[45] M. Dauge. "Neumann and mixed problems on curvilinear polyhedra". In: *Integral Equations Operator Theory* 15.2 (1992), pp. 227–261. DOI: 10.1007/BF01204238.

[46] R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology. Vol. 5.* Evolution problems. I, With the collaboration of Michel Artola, Michel Cessenat and Hélène Lanchon. Springer-Verlag, Berlin, 1992, pp. xiv+709. DOI: 10.1007/978-3-642-58090-1.

[47] K. Deckelnick and M. Hinze. "A note on the approximation of elliptic control problems with bang-bang controls". In: *Comput. Optim. Appl.* 51.2 (2012), pp. 931–939. DOI: 10.1007/s10589-010-9365-z.

[48] E. Di Nezza, G. Palatucci, and E. Valdinoci. "Hitchhiker's guide to the fractional Sobolev spaces". In: *Bull. Sci. Math.* 136.5 (2012), pp. 521–573.

[49] K. Disser, A. F. M. ter Elst, and J. Rehberg. "Hölder estimates for parabolic operators on domains with rough boundary". In: *Ann. Sc. Norm. Sup. Pisa* (2015). DOI: 10.2422/2036-2145/201503-013.

[50] J. C. Dunn. "Convergence rates for conditional gradient sequences generated by implicit step length rules". In: *SIAM J. Control Optim.* 18.5 (1980), pp. 473–487. DOI: 10.1137/0318035.

[51] T. Dupont and R. Scott. "Polynomial approximation of functions in Sobolev spaces". In: *Math. Comp.* 34.150 (1980), pp. 441–463. DOI: 10.2307/2006095.

[52] H. C. Elman, D. J. Silvester, and A. J. Wathen. *Finite elements and fast iterative solvers: with applications in incompressible fluid dynamics*. Second. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford, 2014, pp. xiv+479. DOI: 10.1093/acprof:oso/9780199678792.001.0001.

[53] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992, pp. viii+268.

[54] H. O. Fattorini. *Infinite dimensional linear control systems*. Vol. 201. North-Holland Mathematics Studies. The time optimal and norm optimal problems. Elsevier Science B.V., Amsterdam, 2005, pp. xii+320.

[55] U. Felgenhauer. "On stability of bang-bang type controls". In: *SIAM J. Control Optim.* 41.6 (2003), pp. 1843–1867. DOI: 10.1137/S0363012901399271.

[56] E. Fernández-Cara and E. Zuazua. "The cost of approximate controllability for heat equations: the linear case". In: *Adv. Differential Equations* 5.4-6 (2000), pp. 465–514.

[57] M. Fortin and R. Glowinski. *Augmented Lagrangian methods*. Vol. 15. Studies in Mathematics and its Applications. Applications to the numerical solution of boundary value problems. North-Holland Publishing Co., Amsterdam, 1983, pp. xix+340.

[58] A. V. Fursikov. "Stabilizability of a quasi-linear parabolic equation by means of a boundary control with feedback". In: *Sbornik: Mathematics* 192.4 (2001), pp. 593–639.

[59] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Reprint of the 1998 edition. Springer-Verlag, Berlin, 2001, pp. xiv+517.

[60] W. Gong and N. Yan. "Finite Element Method and its error estimates for the time optimal control of heat equation". In: *International Journal of Numerical Analysis & Modeling* 13.2 (2016).

[61] M. S. Gowda. "A characterization of positive semidefinite operators on a Hilbert space". In: *J. Optim. Theory Appl.* 48.3 (1986), pp. 419–425.

[62] F. Gozzi and P. Loreti. "Regularity of the minimum time function and minimum energy problems: the linear case". In: *SIAM J. Control Optim.* 37.4 (1999), pp. 1195–1221. DOI: 10.1137/S0363012996312763.

[63] M. Grant and S. Boyd. *CVX: Matlab Software for Disciplined Convex Programming, version 2.1*. http://cvxr.com/cvx. Mar. 2014.

[64] M. Grant and S. Boyd. “Graph implementations for nonsmooth convex programs”. In: *Recent Advances in Learning and Control*. Ed. by V. Blondel, S. Boyd, and H. Kimura. Lecture Notes in Control and Information Sciences. http://stanford.edu/~boyd/graph_dcp.html. Springer-Verlag Limited, 2008, pp. 95–110.

[65] J. A. Griepentrog, K. Gröger, H.-C. Kaiser, and J. Rehberg. “Interpolation for function spaces related to mixed boundary value problems”. In: *Math. Nachr.* 241 (2002), pp. 110–120. DOI: 10.1002/1522-2616(200207)241:1<110::AID-MANA110>3.0.CO;2-R.

[66] J. A. Griepentrog, H.-C. Kaiser, and J. Rehberg. “Heat kernel and resolvent properties for second order elliptic differential operators with general boundary conditions on L^p”. In: *Adv. Math. Sci. Appl.* 11.1 (2001), pp. 87–112.

[67] R. Griesse and B. Vexler. “Numerical sensitivity analysis for the quantity of interest in PDE-constrained optimization”. In: *SIAM J. Sci. Comput.* 29.1 (2007), pp. 22–48. DOI: 10.1137/050637273.

[68] P. Grisvard. *Elliptic problems in nonsmooth domains*. Vol. 24. Monographs and Studies in Mathematics. Pitman (Advanced Publishing Program), Boston, MA, 1985, pp. xiv+410. DOI: 10.1137/1.9781611972030.

[69] K. Gröger. “A $W^{1,p}$-estimate for solutions to mixed boundary value problems for second order elliptic differential equations”. In: *Math. Ann.* 283.4 (1989), pp. 679–687. DOI: 10.1007/BF01442860.

[70] M. Gugat. “A Newton method for the computation of time-optimal boundary controls of one-dimensional vibrating systems”. In: *J. Comput. Appl. Math.* 114.1 (2000). Control of partial differential equations (Jacksonville, FL, 1998), pp. 103–119. DOI: 10.1016/S0377-0427(99)00291-5.

[71] M. Haase. *The functional calculus for sectorial operators*. Vol. 169. Operator Theory: Advances and Applications. Birkhäuser Verlag, Basel, 2006, pp. xiv+392. DOI: 10.1007/3-7643-7698-8.

[72] R. Haller-Dintelmann, C. Meyer, J. Rehberg, and A. Schiela. “Hölder continuity and optimal control for nonsmooth elliptic problems”. In: *Appl. Math. Optim.* 60.3 (2009), pp. 397–428. DOI: 10.1007/s00245-009-9077-x.

[73] R. Haller-Dintelmann and J. Rehberg. “Maximal parabolic regularity for divergence operators including mixed boundary conditions”. In: *J. Differential Equations* 247.5 (2009), pp. 1354–1396. DOI: 10.1016/j.jde.2009.06.001.

[74] H. Hermes and J. P. LaSalle. *Functional analysis and time optimal control*. Mathematics in Science and Engineering, Vol. 56. Academic Press, New York-London, 1969, pp. viii+136.

[75] M. R. Hestenes. “Multiplier and gradient methods”. In: *J. Optimization Theory Appl.* 4 (1969), pp. 303–320.

[76] M. Hintermüller, K. Ito, and K. Kunisch. “The primal-dual active set strategy as a semismooth Newton method”. In: *SIAM J. Optim.* 13.3 (2002), 865–888 (2003). DOI: 10.1137/S1052623401383558.

[77] M. Hintermüller and K. Kunisch. “Path-following methods for a class of constrained minimization problems in function space”. In: *SIAM J. Optim.* 17.1 (2006), pp. 159–187. DOI: 10.1137/040611598.

[78] M. Hinze. "A variational discretization concept in control constrained optimization: the linear-quadratic case". In: *Comput. Optim. Appl.* 30.1 (2005), pp. 45–61. DOI: 10.1007/s10589-005-4559-5.

[79] L. Hörmander. *Linear partial differential operators*. Die Grundlehren der mathematischen Wissenschaften, Bd. 116. Academic Press, Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg, 1963, pp. vii+287.

[80] K. Ito and F. Kappel. *Evolution equations and approximations*. Vol. 61. Series on Advances in Mathematics for Applied Sciences. World Scientific Publishing Co., Inc., River Edge, NJ, 2002, pp. xiv+498. DOI: 10.1142/9789812777294.

[81] K. Ito and K. Kunisch. *Lagrange multiplier approach to variational problems and applications*. Vol. 15. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008, pp. xviii+341. DOI: 10.1137/1.9780898718614.

[82] K. Ito and K. Kunisch. "Semismooth Newton methods for time-optimal control for a class of ODEs". In: *SIAM J. Control Optim.* 48.6 (2010), pp. 3997–4013. DOI: 10.1137/090753905.

[83] K. Ito and K. Kunisch. "Semi-smooth Newton methods for variational inequalities of the first kind". In: *M2AN Math. Model. Numer. Anal.* 37.1 (2003), pp. 41–62.

[84] A. Jonsson and H. Wallin. *Function spaces on subsets of* $\mathbb{R}^n$. Vol. 2. 1. 1984, pp. xiv+221.

[85] C. Y. Kaya and J. L. Noakes. "Computational method for time-optimal switching control". In: *J. Optim. Theory Appl.* 117.1 (2003), pp. 69–92.

[86] C. Y. Kaya and J. L. Noakes. "Computations and time-optimal controls". In: *Optimal Control Appl. Methods* 17.3 (1996), pp. 171–185.

[87] G. Knowles. "Finite element approximation of parabolic time optimal control problems". In: *SIAM J. Control Optim.* 20.3 (1982), pp. 414–427. DOI: 10.1137/0320032.

[88] G. Knowles. "Some problems in the control of distributed systems, and their numerical solution". In: *SIAM J. Control Optim.* 17.1 (1979), pp. 5–22. DOI: 10.1137/0317002.

[89] W. Krabs. "Optimal control of processes governed by partial differential equations. I. Heating processes". In: *Z. Oper. Res. Ser. A-B* 26.1 (1982), A21–A48.

[90] M. Krízek and P. Neittaanmäki. *Mathematical and numerical modelling in electrical engineering*. Vol. 1. Mathematical Modelling: Theory and Applications. Theory and applications, With a foreword by Ivo Babuška. Kluwer Academic Publishers, Dordrecht, 1996, pp. xiv+300.

[91] K. Kunisch and A. Rösch. "Primal-dual active set strategy for a general class of constrained optimal control problems". In: *SIAM J. Optim.* 13.2 (2002), pp. 321–334.

[92] K. Kunisch, K. Pieper, and A. Rund. "Time optimal control for a reaction diffusion system arising in cardiac electrophysiology—a monolithic approach". In: *ESAIM Math. Model. Numer. Anal.* 50.2 (2016), pp. 381–414. DOI: 10.1051/m2an/2015048.

[93] K. Kunisch and A. Rund. "Time optimal control of the monodomain model in cardiac electrophysiology". In: *IMA J. Appl. Math.* 80.6 (2015), pp. 1664–1683. DOI: 10.1093/imamat/hxv010.

[94] K. Kunisch and D. Wachsmuth. "On time optimal control of the wave equation and its numerical realization as parametric optimization problem". In: *SIAM J. Control Optim.* 51.2 (2013), pp. 1232–1262. DOI: 10.1137/120877520.

[95] K. Kunisch and D. Wachsmuth. "On time optimal control of the wave equation, its regularization and optimality system". In: *ESAIM Control Optim. Calc. Var.* 19.2 (2013), pp. 317–336. DOI: 10.1051/cocv/2012010.

[96] K. Kunisch and L. Wang. "Bang-bang property of time optimal controls of semilinear parabolic equation". In: *Discrete Contin. Dyn. Syst.* 36.1 (2016), pp. 279–302.

[97] K. Kunisch and L. Wang. "Time optimal control of the heat equation with pointwise control constraints". In: *ESAIM Control Optim. Calc. Var.* 19.2 (2013), pp. 460–485. DOI: 10.1051/cocv/2012017.

[98] S. Lacoste-Julien and M. Jaggi. "On the Global Linear Convergence of Frank-Wolfe Optimization Variants". In: *ArXiv e-prints* (Nov. 2015). arXiv: 1511.05932 [math.OC].

[99] D. Lamberton. "Équations d'évolution linéaires associées à des semi-groupes de contractions dans les espaces L^p". In: *J. Funct. Anal.* 72.2 (1987), pp. 252–262. DOI: 10.1016/0022-1236(87)90088-7.

[100] I. Lasiecka. "Ritz-Galerkin approximation of the time optimal boundary control problem for parabolic systems with Dirichlet boundary conditions". In: *SIAM J. Control Optim.* 22.3 (1984), pp. 477–500. DOI: 10.1137/0322029.

[101] I. Lasiecka and R. Triggiani. *Control theory for partial differential equations: continuous and approximation theories. I.* Vol. 74. Encyclopedia of Mathematics and its Applications. Abstract parabolic systems. Cambridge University Press, Cambridge, 2000, pp. xxii+644+I4.

[102] D. Leykekhman and B. Vexler. "Discrete maximal parabolic regularity for Galerkin finite element methods". In: *Numerische Mathematik* (2016), pp. 1–30. DOI: 10.1007/s00211-016-0821-2.

[103] D. Leykekhman and B. Vexler. "Pointwise best approximation results for Galerkin finite element solutions of parabolic problems". In: *SIAM J. Numer. Anal.* 54.3 (2016), pp. 1365–1384. DOI: 10.1137/15M103412X.

[104] X. J. Li and J. M. Yong. *Optimal control theory for infinite-dimensional systems.* Systems & Control: Foundations & Applications. Birkhäuser Boston, Inc., Boston, MA, 1995, pp. xii+448. DOI: 10.1007/978-1-4612-4260-4.

[105] F.-H. Lin. "A uniqueness theorem for parabolic equations". In: *Comm. Pure Appl. Math.* 43.1 (1990), pp. 127–136.

[106] J.-L. Lions. *Optimal control of systems governed by partial differential equations.* Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin, 1971, pp. xi+396.

[107] J.-L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires.* Dunod; Gauthier-Villars, Paris, 1969, pp. xx+554.

[108] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I.* Die Grundlehren der mathematischen Wissenschaften, Band 181. Springer-Verlag, New York-Heidelberg, 1972, pp. xvi+357.

[109] X. Lu, L. Wang, and Q. Yan. "Computation of time optimal control problems governed by linear ordinary differential equations". In: *J. Sci. Comput.* 73.1 (2017), pp. 1–25.

[110] A. Lunardi. *Interpolation theory.* Second. Appunti. Scuola Normale Superiore di Pisa (Nuova Serie). [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)]. Edizioni della Normale, Pisa, 2009, pp. xiv+191.

[111] M. Luskin and R. Rannacher. “On the smoothing property of the Galerkin method for parabolic equations”. In: *SIAM J. Numer. Anal.* 19.1 (1982), pp. 93–113. DOI: 10.1137/0719003.

[112] J. W. Macki and A. Strauss. *Introduction to optimal control theory*. Undergraduate Texts in Mathematics. Springer-Verlag, New York-Berlin, 1982, pp. xiii+165.

[113] H. Maurer and N. P. Osmolovskii. “Second order sufficient conditions for time-optimal bang-bang control”. In: *SIAM J. Control Optim.* 42.6 (2004), pp. 2239–2263. DOI: 10.1137/S0363012902402578.

[114] V. G. Maz’ya. *Sobolev spaces*. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1985, pp. xix+486. DOI: 10.1007/978-3-662-09922-3.

[115] V. G. Maz’ya and T. Shaposhnikova. “On the Bourgain, Brezis, and Mironescu theorem concerning limiting embeddings of fractional Sobolev spaces”. In: *J. Funct. Anal.* 195.2 (2002), pp. 230–238.

[116] 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”. In: *SIAM J. Control Optim.* 49.5 (2011), pp. 1961–1997. DOI: 10.1137/100793888.

[117] D. Meidner and B. Vexler. “A priori error estimates for space-time finite element discretization of parabolic optimal control problems. I. Problems without control constraints”. In: *SIAM J. Control Optim.* 47.3 (2008), pp. 1150–1177. DOI: 10.1137/070694016.

[118] D. Meidner and B. Vexler. “A priori error estimates for space-time finite element discretization of parabolic optimal control problems. II. Problems with control constraints”. In: *SIAM J. Control Optim.* 47.3 (2008), pp. 1301–1329. DOI: 10.1137/070694028.

[119] H. Meinlschmidt, C. Meyer, and J. Rehberg. “Optimal control of the thermistor problem in three spatial dimensions, Part 1: Existence of optimal solutions”. In: *SIAM J. Control Optim.* 55.5 (2017), pp. 2876–2904.

[120] H. Meinlschmidt and J. Rehberg. “Hölder-estimates for non-autonomous parabolic problems with rough data”. In: *Evolution Equations and Control Theory* 5.1 (2016), pp. 147–184. DOI: 10.3934/eect.2016.5.147.

[121] C. Meyer and A. Rösch. “Superconvergence properties of optimal control problems”. In: *SIAM J. Control Optim.* 43.3 (2004), pp. 970–985. DOI: 10.1137/S0363012903431608.

[122] S. Micu, I. Roventa, and M. Tucsnak. “Time optimal boundary controls for the heat equation”. In: *J. Funct. Anal.* 263.1 (2012), pp. 25–49. DOI: 10.1016/j.jfa.2012.04.009.

[123] I. Neitzel, J. Pfefferer, and A. Rösch. “Finite Element Discretization of State-Constrained Elliptic Optimal Control Problems with Semilinear State Equation”. In: *SIAM J. Control Optim.* 53.2 (2015), pp. 874–904. DOI: 10.1137/140960645.

[124] I. Neitzel and B. Vexler. “A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems”. In: *Numer. Math.* 120.2 (2012), pp. 345–386. DOI: 10.1007/s00211-011-0409-9.

[125] S. Nicaise, S. Stingelin, and F. Tröltzsch. “On two optimal control problems for magnetic fields”. In: *Comput. Methods Appl. Math.* 14.4 (2014), pp. 555–573.

[126] J. Nocedal and S. J. Wright. *Numerical optimization*. Second. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2006, pp. xxii+664.

[127] E. M. Ouhabaz. *Analysis of heat equations on domains*. Vol. 31. London Mathematical Society Monographs Series. Princeton University Press, Princeton, NJ, 2005, pp. xiv+284.

[128] A. Pazy. *Semigroups of linear operators and applications to partial differential equations*. Vol. 44. Applied Mathematical Sciences. Springer-Verlag, New York, 1983, pp. viii+279. DOI: 10.1007/978-1-4612-5561-1.

[129] M. D. Perlman. "Jensen's inequality for a convex vector-valued function on an infinite-dimensional space". In: *J. Multivariate Anal.* 4 (1974), pp. 52–65.

[130] K. Pieper. "Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems". PhD thesis. Technische Universität München, 2015.

[131] B. T. Poljak and N. V. Tret'jakov. "A method of penalty estimates for conditional extremum problems". In: *Ž. Vyčisl. Mat. i Mat. Fiz.* 13 (1973), pp. 34–46, 267.

[132] H. A. Poonawala and M. W. Spong. "Time-optimal velocity tracking control for differential drive robots". In: *Automatica J. IFAC* 85 (2017), pp. 153–157.

[133] M. J. D. Powell. "A method for nonlinear constraints in minimization problems". In: *Optimization (Sympos., Univ. Keele, Keele, 1968)*. Academic Press, London, 1969, pp. 283–298.

[134] J. P. Raymond and H. Zidani. "Pontryagin's principle for time-optimal problems". In: *J. Optim. Theory Appl.* 101.2 (1999), pp. 375–402. DOI: 10.1023/A:1021793611520.

[135] S. M. Robinson. "Normal maps induced by linear transformations". In: *Math. Oper. Res.* 17.3 (1992), pp. 691–714.

[136] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Vol. 317. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1998, pp. xiv+733. DOI: 10.1007/978-3-642-02431-3.

[137] A. Rösch and D. Wachsmuth. "Numerical Verification of Optimality Conditions". In: *SIAM Journal on Control and Optimization* 47.5 (Jan. 2008), pp. 2557–2581. DOI: 10.1137/060663714.

[138] L. Rosier. "A survey of controllability and stabilization results for partial differential equations". In: *Journal européen des systèmes automatisés* 41.3/4 (2007), p. 365.

[139] Y. Saad. *Iterative methods for sparse linear systems*. Second. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2003, pp. xviii+528. DOI: 10.1137/1.9780898718003.

[140] K. Schittkowski. "Numerical solution of a time-optimal parabolic boundary value control problem". In: *J. Optim. Theory Appl.* 27.2 (1979), pp. 271–290. DOI: 10.1007/BF00933231.

[141] E. J. P. G. Schmidt and R. J. Stern. "Invariance theory for infinite-dimensional linear control systems". In: *Appl. Math. Optim.* 6.2 (1980), pp. 113–122. DOI: 10.1007/BF01442887.

[142] D. Sciutti. "On a characterization of convergence for the Hestenes method of multipliers". In: *J. Optimization Theory Appl.* 22.2 (1977), pp. 227–237.

[143] S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Springer, 1984. DOI: 10.1007/978-3-7091-8752-4.

[144] E. M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970, pp. xiv+290.

[145] N. Thanh Qui and D. Wachsmuth. "Stability for Bang-Bang Control Problems of Partial Differential Equations". In: *ArXiv e-prints* (July 2017). arXiv: 1707.03698 [math.OC].

[146] H. Triebel. *Interpolation theory, function spaces, differential operators*. Vol. 18. North-Holland Mathematical Library. North-Holland Publishing Co., Amsterdam-New York, 1978, pp. 1–528.

[147] F. Tröltzsch. *Optimal control of partial differential equations*. Vol. 112. Graduate Studies in Mathematics. Theory, methods and applications. American Mathematical Society, Providence, RI, 2010, pp. xvi+399. DOI: 10.1090/gsm/112.

[148] M. Tucsnak, J. Valein, and C.-T. Wu. "Finite dimensional approximations for a class of infinite dimensional time optimal control problems". In: *International Journal of Control* (2016). (published online), pp. 1–13. DOI: 10.1080/00207179.2016.1228122.

[149] M. Tucsnak, G. Wang, and C.-T. Wu. "Perturbations of Time Optimal Control Problems for a Class of Abstract Parabolic Systems". In: *SIAM J. Control Optim.* 54.6 (2016), pp. 2965–2991. DOI: 10.1137/15M101909X.

[150] M. Tucsnak and G. Weiss. *Observation and control for operator semigroups*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Verlag, Basel, 2009, pp. xii+483.

[151] M. Ulbrich. "On a nonsmooth Newton method for nonlinear complementarity problems in function space with applications to optimal control". In: *Complementarity: applications, algorithms and extensions (Madison, WI, 1999)*. Vol. 50. Appl. Optim. Kluwer Acad. Publ., Dordrecht, 2001, pp. 341–360. DOI: 10.1007/978-1-4757-3279-5_16.

[152] N. von Daniels and M. Hinze. "Variational discretization of a control-constrained parabolic bang-bang optimal control problem". July 2017.

[153] G. Vossen and H. Maurer. "On L^1-minimization in optimal control and applications to robotics". In: *Optimal Control Appl. Methods* 27.6 (2006), pp. 301–321.

[154] D. Wachsmuth. "Robust error estimates for regularization and discretization of bang-bang control problems". In: *Comput. Optim. Appl.* 62.1 (2015), pp. 271–289. DOI: 10.1007/s10589-014-9645-0.

[155] D. Wachsmuth and G. Wachsmuth. "Necessary conditions for convergence rates of regularizations of optimal control problems". In: *System modeling and optimization*. Vol. 391. IFIP Adv. Inf. Commun. Technol. Springer, Heidelberg, 2013, pp. 145–154. DOI: 10.1007/978-3-642-36062-6_15.

[156] G. Wachsmuth and D. Wachsmuth. "Convergence and regularization results for optimal control problems with sparsity functional". In: *ESAIM Control Optim. Calc. Var.* 17.3 (2011), pp. 858–886. DOI: 10.1051/cocv/2010027.

[157] G. Wang and L. Wang. "The bang-bang principle of time optimal controls for the heat equation with internal controls". In: *Systems Control Lett.* 56.11-12 (2007), pp. 709–713.

[158] G. Wang and Y. Xu. "Equivalence of three different kinds of optimal control problems for heat equations and its applications". In: *SIAM J. Control Optim.* 51.2 (2013), pp. 848–880. DOI: 10.1137/110852449.

[159] G. Wang and G. Zheng. "An approach to the optimal time for a time optimal control problem of an internally controlled heat equation". In: *SIAM J. Control Optim.* 50.2 (2012), pp. 601–628. DOI: 10.1137/100793645.

[160] G. Wang and E. Zuazua. "On the equivalence of minimal time and minimal norm controls for internally controlled heat equations". In: *SIAM J. Control Optim.* 50.5 (2012), pp. 2938–2958. DOI: 10.1137/110857398.

[161] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987, pp. xii+518. DOI: 10.1017/CBO9781139171755.

[162] H. Yu. "Approximation of time optimal controls for heat equations with perturbations in the system potential". In: *SIAM J. Control Optim.* 52.3 (2014), pp. 1663–1692. DOI: 10.1137/120904251.

[163] E. Zeidler. *Nonlinear functional analysis and its applications. I.* Fixed-point theorems. Springer-Verlag, New York, 1986, pp. xxi+897. DOI: 10.1007/978-1-4612-4838-5.

[164] C. Zhang. "The time optimal control with constraints of the rectangular type for linear time-varying ODEs". In: *SIAM J. Control Optim.* 51.2 (2013), pp. 1528–1542.

[165] G. Zheng and J. Yin. "Numerical approximation for a time optimal control problems governed by semi-linear heat equations". In: *Adv. Difference Equ.* (2014), 2014:94, 7. DOI: 10.1186/1687-1847-2014-94.

[166] J. Zowe and S. Kurcyusz. "Regularity and stability for the mathematical programming problem in Banach spaces". In: *Appl. Math. Optim.* 5.1 (1979), pp. 49–62. DOI: 10.1007/BF01442543.

[167] E. Zuazua. "Controllability and observability of partial differential equations: some results and open problems". In: *Handbook of differential equations: evolutionary equations. Vol. III.* Handb. Differ. Equ. Elsevier/North-Holland, Amsterdam, 2007, pp. 527–621. DOI: 10.1016/S1874-5717(07)80010-7.

Symbols

General

$\mathbb{N}, \mathbb{R}, \mathbb{R}_+, \mathbb{C}$	Natural numbers, real numbers, (strictly) positive real numbers, complex numbers	
x^+	Positive part of x, i.e. $x^+ = \max\{0, x\}$	
Re	Real part of complex number	
Γ	Gamma function	
∂	Convex subdifferential	
$\partial_{\mathrm{C}}, \partial_{\mathrm{C},x}$	Clarke's generalized subdifferential	180
$\mathrm{d}^+, \mathrm{d}^-$	Directional derivatives in positive and negative direction	
$d_U(\cdot), d_U^H(\cdot)$	Distance function to U in Hilbert space H	9
$N_U(u)$	Normal cone to U at the point u	9
P_U, P_U^H	Minimizing projection onto U in Hilbert space H	9
$T_U(u)$	Tangent cone to U at the point u	

Linear operators, Function spaces, and Interpolation

$\hookrightarrow$	Continuous embedding	
$\hookrightarrow_c$	Continuous and compact embedding	
$\hookrightarrow_d$	Continuous and dense embedding	
$[X, Y]_\theta$	Complex interpolation space	
$(X, Y)_{\theta,p}$	Real interpolation space	167
X^*	Dual space to X	
$\mathcal{L}(X, Y)$	Linear and bounded operators between X and Y	
$\mathcal{D}_X(A)$	Domain of operator A in Banach space X	
$\rho(A)$	Resolvent set of A	
$R(z, A)$	Resolvent of A	
$\ker A$	Null space of linear operator A	
$\operatorname{ran} A$	Range of linear operator A	
span	Linear span of vectors	
i_t	Trace mapping, i.e. if $u\colon [0,T] \to X$, then $i_t u = u(t)$ for $t \in [0,T]$	9
$C^\alpha(I; X)$	Hölder continuous functions on I with values in X	

$H^{\theta,p}$	Bessel potential space	39
$H^1(I;X)$	Short for $W^{1,2}(I;X)$	
$L^p(I;X)$	Lebesgue p-integrable functions on I with values in X	
$W^{k,p}$	Sobolev/Sobolev-Slobodeckij space	39, 183
$W^{k,p}(I;X)$	Sobolev space on I with values in X	
$W(0,T)$	Short for $H^1((0,T);V^*) \cap L^2((0,T);V)$	

Optimal control problem

$\langle \cdot, \cdot \rangle$	Duality pairing between V^* and V	8
$(\cdot, \cdot)$	Inner product in H	8
(ω, ϱ)	Measure space for control space	36
$\|\cdot\|$	Norm on H	8
$\|(\cdot, \cdot)\|$	Norm on product space $\mathbb{R} \times L^2(I \times \omega)$	41
$A, a(\cdot, \cdot)$	Weakly coercive operator $A\colon V \to V^*$ defined by bilinear form $a\colon V \times V \to \mathbb{R}$	8
B	Control operator from Q into X_{θ_0}	9
$C_{(\bar{\nu},\bar{q})}$	Critical cone	42
$h(\cdot, \cdot)$	Lower Hamiltonian	13
H_{μ_0}	Hamiltonian	21
H	Pivot space of Gelfand triple $V \hookrightarrow_c H \hookrightarrow V^*$	8
N_{ad}	Set of admissible scaling functions	19
Q, Q_{ad}	Space of controls and subset of admissible controls	9, 36
$P_{Q_{ad}}$	Pointwise projection onto set of admissible controls	38, 101
S	Control-to-state mapping	19
X_θ	Domain of fractional powers of A, i.e. $X_\theta = \mathcal{D}_{V^*}((A+\omega_0)^\theta)$	8
V	Domain of linear operator A constituting the Gelfand triple $V \hookrightarrow_c H \hookrightarrow V^*$	8

Discretization

$\mathrm{B}(\cdot, \cdot, \cdot)$	Trilinear form for Galerkin scheme	102
$[\,\cdot\,]_m$	Jump terms in discontinuous Galerkin scheme	102
Δ_h	Discrete Laplace operator	103
$h_h(\cdot, \cdot)$	Discrete lower Hamiltonian	148
i_k	Interpolation onto piecewise constant functions in time	188
I_h	Interpolation onto cellwise linear and continuous functions	186
I_σ	Projection/Interpolation operator onto set of controls Q_σ	107, 127
Π_k	L^2-projection onto piecewise constant functions in time	117

Π_h	L^2-projection onto V_h	187
$\Pi_{h,0}$	L^2-projection onto cellwise constant functions in space	145
Π_{kh}	L^2-projection onto piecewise constant functions in time and cellwise constant functions in space	117
$\Pi_{ad,h}$	Hilbert space projection onto $Q_{ad,h}$	149
$\mathcal{P}_0(I;X)$	Space of constant functions on I with values in X	102
$Q_\sigma, Q_{ad,\sigma}$	Space of temporally and spatially discrete controls and subset of admissible controls	107, 127
$Q_h, Q_{ad,h}$	Space of spatially discrete controls and subset of admissible controls	148
R_h	Ritz projection	186
$\sigma(k,h)$	Projection/Interpolation error of I_σ in $L^2(I\times\omega)$	107
$\sigma_1(k,h)$	Projection/Interpolation error of I_σ in $L^1(I\times\omega)$	127
$\sigma_2(k,h)$	Projection/Interpolation error of I_σ in $L^2(I;H^{-1})$	127
S_{kh}	Discrete control-to-state mapping	103
$\mathcal{T}_h, \mathcal{T}_h^\omega$	Spatial mesh for finite element discretization	102, 107
V_h	Space of continuous and cellwise linear functions	102
X_k	Semi-discrete state space	188
$X_k(Y)$	Piecewise constant functions with values in Y	148
$X_{k,h}$	Discrete state space	102