Universität der Bundeswehr München
Fakultät für Luft- und Raumfahrttechnik
Institut für Mathematik und Rechneranwendung

Regularized Newton Methods for Linear Quadratic Optimal Control Problems With Applications in Model Predictive Controllers

Dipl.-Math. Björn Hüpping

Vollständiger Abdruck der bei der Fakultät
für Luft- und Raumfahrttechnik der
Universität der Bundeswehr München zur
Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

eingereichten Dissertation

Vorsitzender:	Univ.-Prof. Dr. Berthold Färber
1. Berichterstatter:	Univ.-Prof. Dr. Matthias Gerdts
2. Berichterstatter:	Univ.-Prof. Dr. Hans-Josef Pesch, Universität Bayreuth

Diese Dissertation wurde am 10.11.2011 bei der Universität der Bundeswehr München, 85577 Neubiberg eingereicht und durch die Fakultät für Luft- und Raumfahrttechnik am 16.11.2011 angenommen. Die mündliche Prüfung fand am 25.04.2012 statt.

Acknowledgements

First of all, I would like to thank Prof. Dr. Gerdts for his guidance, advice and support during the time I worked on my thesis. Besides, I am thankful that he gave me the opportunity to improve my teaching skills not least because his own lectures set a high didactic standard.

I am very grateful to Prof. Dr. Pesch for his willingness to become the second assessor for this thesis.

The main part of this thesis was written at the University of Birmingham, where I found the friendliest and most open-hearted postgraduate community imaginable. I remember the warm welcome by Ben Fairbairn, as well as his introduction into the community and various aspects of British culture (e.g. whisky, the British sense of humor, and many more). I also appreciate some mathematical discussions I had with Oliver Cooley. Even more importantly, I thank him and Rebecca Waldecker, for all the fun, and I say thank you for the music.

I would also like to show my gratitude to Jan Fiala, with whom I developed Lemma 2.2 during a break. Since without a first lemma my thesis would have looked pretty poor, I believe this to be of particular importance.

It is a pleasure for me to thank my German colleagues who accompanied me through part of this chapter of my life: Martin Kunkel, with whom it was fun to learn a lot of what I know now; Eggert Rose, an unexpected brother in arms and a good friend; Martin Schlüter, with whom I still enjoy an occasional virtual coffee; and the colleagues from the ZAIK in Cologne, who granted me asylum for quite a long time.

I would like to thank my family for their encouragement and patience through the years.

Most of all, I want to say thank you to Kirsten Albrecht, who certainly supported me the most, and with whom I laughed the most. That aside, I also won't forget that with you, I had the best proof reader in the world.

Finally, I would like to dedicate this work to a former mathematics teacher of mine, Josef Schraub. This way, I want to express my gratitude for his excellent teaching.

Contents

Notation

$(x)_{i=0}^N$	if $x_i \in \mathbb{R}\ \forall i$, $(x)_{i=0}^N$ denotes the vector $(x_0, x_1, \ldots, x_N)^\top \in \mathbb{R}^{N+1}$
	$x_i \in \mathbb{R}^n\ \forall i$, $(x)_{i=0}^N$ denotes the vector $(x_0, x_1, \ldots, x_N)^\top \in \mathbb{R}^{(N+1)\cdot n}$
$\lim_{x \searrow a}$, $\lim_{x \nearrow a}$	the limit from above/below
$\|\cdot\|_X$	norm on the Banach space X
$\|\cdot\|$	norm in $\mathbb{R}^{n_x}$, usually the Euklidian norm
$\langle \cdot, \cdot \rangle$	the scalar product in the respective space
Id	identiy map, $Id(x) = x$
$o, \mathcal{O}$	Landau symbols
I	identity matrix
$\mathcal{H}$	Hamiltonian of an optimal control problem
$\operatorname{im} F$	image of F
$\ker F$	kernel of F
$\operatorname{ess\,sup}$	essential supremum
$B_r(x)$	open Ball with radius r and center x
$\mathcal{F}(X, Y)$	set of mappings from X to Y
$\mathcal{L}(X, Y)$	set of linear continuous mappings from X to Y
$\mathcal{C}(X, Y)$	continuous functions from X to Y
$\mathcal{C}^n(X, Y)$	n-times continuously differentiable functions from X to Y
$L^p(X, Y)$	Lebesgue measurable functions from X to Y
	with finite $\|\cdot\|_p$-norm
$W^{p,q}(X, Y)$	q-times weakly differentiable function with derivative in L^p
$f^{(i)}$	the i-th derivative of the function f
$\|f\|_p$	$(\int_a^b f^p(t)dt)^{(1/p)}$
$\|f\|_{q,p}$	$\left(\sum_{i=0}^q \|f^{(i)}\|_p^p\right)^{1/p}$
$BV([a, b], \mathbb{R}^n)$	functions of bounded variation
$NBV([a, b], \mathbb{R}^n)$	normed functions of bounded variation
$TV(f, a, b)$	total variation of f on $[a, b]$
$\int_a^b f(t)d\alpha(t)$	Stieltjes Integral
$\mathcal{H}[t]$	abbreviated notation indicating that $\mathcal{H}$ is evaluated along
	the optimal solution
$\mathcal{L}(X, Y)$	the space of linear continuous mappings from X to Y
$X^* = \mathcal{L}(X, \mathbb{R})$	the dual space of X, i.e. the space of continuous linear
	functionals from X to $\mathbb{R}$
$F\|_\Omega$	the restriction of $F : X \to Y$ to $\Omega \subset X$
$F'(x)(h)$	Fréchet derivative of F at x in direction h
$F'_x(x, y)(h_x)$	partial Fréchet of F at (x, y) in direction h_x

1. Introduction

Since the necessary optimality conditions for optimal control problems have been introduced by Pontrjagin et. al., two different approaches have been developed for the numerical solution of this class of problems. The direct approach, in which the optimal control problem is discretized and then regarded as an optimal control problem in finite dimensional spaces, has become popular, as it can be used for a broad class of problems without making assumptions about the structure of the optimal control.

Indirect approaches that are based upon the optimality conditions for the original problem often tried to solve the complementarity problem by finding solutions for different control structures. This work focuses on another approach: The necessary optimality conditions are rearranged as an operator equation in function spaces. In the next step, a Newton method for function spaces is applied to the equation. An advantage of this procedure is that the arising method can be used without knowledge about the structure of the optimal solution. At the same time, theoretical problems emerge when applying the Newton method in infinite dimensional spaces.

In Chapter 2, basics from functional analysis and control theory are introduced. This helps to clarify the notation used in this work and provides the reader with examples and theorems that are needed for the complete understanding the theory.

The necessary optimality conditions, also know as the minimum principle of optimal control, is derived in Chapter 3. While most authors concentrate on problems with either set constrained controls or problems with mixed control state constraints, we state the principle for the generic case in which both types of constraints may appear and extend the theory by introducing knowledge from control theory to this field. This helps to simplify some assumptions needed to ensure the validity of the optimality conditions.

Chapter 4 presents an application of the minimum principle: It is used to prove convergence results for the virtual control concept, a regularization technique that turns problems with pure state constraints into problems with mixed control-state constraints. The main advantage of this regularization in the context of this work is that the latter problem type can be solved by an adequate indirect solution approach.

Solution techniques that make use of this approach are presented in Chapter 5. In order to apply the Newton method to function spaces, the complementarity problem derived in Chapter 3 are turned into an operator equation. As the operators are necessarily nonsmooth, we regularize the equation, before algorithms based on the Newton method are introduced.

The numerical realization of the Newton methods is shown and compared to the direct discretization approach in Chapter 6.

Finally, the algorithms and the regularizations are tested in Chapter 7 in the context of Linear Quadratic Model Predictive Controllers. The virtual control concept in this case has the advantage that all problems that may arise during the regulation are solvable, and that the regulation is independent from the system of ordinary differential equations that describes the physical system. This chapter is divided into examples with mixed control-state constraints and examples with pure state constraints. This allows to independently observe the influence of the regularization techniques in use.

2. Basics

The purpose of this chapter is to gather most fundamental definitions and statements in one place, so as to improve the reading flow in later chapters. Also, some examples (e.g. Example 2.20.1) are easy to prove, but hard to find in literature.

2.1. Analysis

The first lemma belongs to the cathegory "easy to prove but hard to find". It allows to deduce the existence of a limes of function from the function's Hölder continuity. This proves useful when a convergence result for a regularized complementarity problem is derived in Theorem 5.14.

The definition of Hölder continuity can be found in [Dob06, p. 36]:

Definition 2.1 (Hölder continuity)
Let X, Y be normed spaces and $\Omega \subset X$. A function $F : \Omega \to Y$ is called Hölder continuous (with exponent $\alpha \in (0, 1)$)*, if for all $x_1, x_2 \in \Omega$ it holds that*

$$\|F(x_1) - F(x_2)\|_Y \leq L \cdot \|x_1 - x_2\|_X^\alpha$$

for some constant L that is independent of x_1 and x_2.

Lemma 2.2
Let X be a Banach space, and let $a, b \in \mathbb{R}$, such that $a < b$. Let $F : [a, b] \to X$ be Hölder continuous on $(a, b]$. Then $\lim_{t \searrow a} F(t) \in X$ exists.

Proof.
Let $(t_i)_{i\in\mathbb{N}}$ be a sequence with $\lim\limits_{i\to\infty} t_i = a$, $t_i > a$. Then

$$||F(t_n) - F(t_m)|| \leq L \cdot |t_n - t_m|^\alpha.$$

So $\left(F(t_i)\right)_{i\in\mathbb{N}}$ is Cauchy and therefore converges. Now suppose the limit value was not unique. Then let $(t_i)_{i\in\mathbb{N}}$ and $(\tilde{t}_i)_{i\in\mathbb{N}}$ be two sequences, such that $\lim\limits_{i\to\infty} t_i = \lim\limits_{i\to\infty} \tilde{t}_i = a$, but $\lim\limits_{i\to\infty} F(t_i) = f \neq \tilde{f} = \lim\limits_{i\to\infty} F(\tilde{t}_i)$. Now

$$\begin{aligned} ||f - \tilde{f}|| &\leq ||f - F(t_n) + F(\tilde{t}_n) - \tilde{f} + F(t_n) - F(\tilde{t}_n)|| \\ &\leq ||f - F(t_n)|| + ||F(\tilde{t}_n) - \tilde{f}|| + ||F(t_n) - F(\tilde{t}_n)||, \end{aligned}$$

where the right hand side vanishes for large n. □

The concept of a normal cone (cf. [GV03, Definition 1.1]) is needed in order to derive optimality conditions for optimality problems in infinite dimensions.

Definition 2.3 (Normal Cone)
Let $U \subset \mathbb{R}^n$ be a closed set, and $\hat{u} \in U$. Then v in $\mathbb{R}^n$ is normal to U at $\hat{u}$ *if there exist series $(v_i)_{i\in\mathbb{N}}$, $v_i \to v$ and $(u_i)_{i\in\mathbb{N}}$, $u_i \to \hat{u}$ (in U), such that for i*

$$\langle v_i, u - u_i\rangle \leq o(|u - u_i|).$$

The normal cone $N_U(\hat{u})$ *is the set of all normals to U in $\hat{u}$.*

The Banach spaces in which the optimization problems are stated are Lebesgue spaces and Sobolev spaces, that are presented in the following definitions. The variables in optimal control problems can be chosen from these spaces (cf. [Ger06, Section 2.3]).

Definition 2.4 (Lebesgue and Sobolev Spaces and Their Respective Norms)
Let $1 \leq p \leq \infty$. The set of functions needed for the definition of $L^p([a,b],\mathbb{R})$ is

$$\tilde{L}^p([a,b],\mathbb{R}) := \{f : [a,b] \to \mathbb{R} \mid f \text{ is measurable with } \|f\|_p < \infty\},$$

where

$$\|f\|_p := \begin{cases} \left(\int_a^b |f(t)|^p dt\right)^{1/p} & \text{if } p < \infty \\ \operatorname{ess\,sup}_{a\leq t\leq b} |f(t)| & \text{if } p = \infty \end{cases}.$$

Then the Lebesgue space L^p is the space of equivalence classes in $\tilde{L}^p$ with respect to the $\|\cdot\|_p$-norm.

Let $1 \leq p, q \leq \infty$. The space $W^{q,p}([a,b],\mathbb{R})$ consists of all absolutely continuous functions $f : [a,b] \to \mathbb{R}$ with absolutely continuous derivatives up to order $q-1$ and

$$\|f\|_{q,p} < \infty,$$

where the norm $\|\cdot\|_{q,p}$ is defined by

$$\|f\|_{q,p} := \begin{cases} \left(\sum_{i=0}^q \|f^{(i)}\|_p^p\right)^{1/p} & \text{if } p < \infty \\ \max\limits_{0\leq i\leq q} \|f^{(i)}\|_\infty & \text{if } p = \infty \end{cases}.$$

The spaces L^p as well as $W^{q,p}$ with their respective norms are Banach spaces. The spaces $W^{q,2}([a,b],\mathbb{R})$ are Hilbert spaces with the scalar product

$$\langle f, g\rangle_{W^{q,2}} := \sum_{i=0}^q \int_a^b f^{(i)}(t) g^{(i)}(t) dt.$$

Hölder's inequality (see [Alt06, Lemma 1.16, p. 51]) as well as the subsequent embedding theorem are useful in some examples of linear and differential operators.

Lemma 2.5 (Hölder's Inequality)
Let $m \in \mathbb{N}$ and $f_i \in L^{p_i}(\mathbb{R}, \mathbb{R})$ for $i = 1, \ldots, m$ with $1 \leq p_i \leq \infty$, and $1 \leq q \leq \infty$ with

$$\sum_{i=1}^{m} \frac{1}{p_i} = \frac{1}{q}.$$

Then the product $\Pi_{i=1}^{m} f_i$ is in $L^q(\mathbb{R}, \mathbb{R})$, and it holds

$$\|\Pi_{i=1}^{m} f_i\| \leq \Pi_{i=1}^{m} \|f_i\|_{L^{p_1}}$$

The following theorem [Alt06, Theorem 8.9, p. 328] covers embeddings of Sobolev spaces and how their respective norms are associated.

Theorem 2.6 (Embeddings of Sobolev Spaces)
Let $\Omega \subset \mathbb{R}^n$ be open and bounded with Lipschitz boundary. Let m_1 and m_2 be integers, and $1 \leq p_1 \leq \infty$ and $1 \leq p_2 \leq \infty$. Then it holds:

2.6.1 *If*

$$m_1 - \frac{n}{p_1} \geq m_2 - \frac{n}{p_2}, \qquad \text{as well as } m_1 \geq m_2,$$

then the embedding

$$Id : W^{m_1,p_1}(\Omega) \to W^{m_2,p_2}(\Omega)$$

exists and is continuous. For $u \in W^{m_1,p_1}(\Omega)$, there exists a constant depending on n, Ω, m_1, p_1, m_2, p_2, such that

$$\|u\|_{W^{m_2,p_2}(\Omega)} \leq C \|u\|_{W^{m_1,p_1}(\Omega)}.$$

2.6.2 *If*

$$m_1 - \frac{n}{p_1} \geq m_2 - \frac{n}{p_2}, \qquad \text{as well as } m_1 > m_2,$$

then the embedding

$$Id : W^{m_1,p_1}(\Omega) \to W^{m_2,p_2}(\Omega)$$

exists and is continuous and compact.

The following lemma is needed in the proof of the minimum principle. According to this lemma, dual elements of functions with disjoint support can be investigated independent from each others. The dual space of X is denoted by X^*. The space of mappings $F : X \to Y$ is denoted by $\mathcal{F}(X, Y)$.

Lemma 2.7 (Dual Space and Support of Functions)
Let $S \subset \mathcal{F}(A \cup B, K)$, $A \cap B = \emptyset$. Let $y^* \in S^*$, such that $y^* \bar{x} = 0$ for all $\bar{x}$ with $\bar{x}(t) = 0 \quad \forall t \in A$. Then there exists an element $y_A^* \in (S|_A)^*$, so that $y^* x = y_A^* x|_A$ for all $x \in S$.

Proof.
For any subset X of $A \cup B$, let

$$x_X(t) := \begin{cases} x(t) & \text{if } t \in X \\ 0 & \text{otherwise.} \end{cases}$$

Then $y^* x = y^*(x_A + x_B) = y^* x_A + y^* x_B = y^* x_A$, hence $y^* \in S_A^*$. The assertion follows since S_A and $S|_A$ are isomorphic. □

Another expression for the essence of this lemma is $[\mathcal{F}(A+B,K)]^* = [\mathcal{F}(A,K)]^* + [\mathcal{F}(B,K)]^*$.

This lemma will be used later when an element of the dual space of a subset of L^∞ is analyzed. The dual element will be written as a sum of two elements, each of which maps functions with support A and B, respectively, to 0. The argument derived from this lemma will be that the dual elements only need to be investigated on these respective time sets.

2.1.1. The Stieltjes integral and functions of Bounded Variation

The Stieltjes integral occurs naturally when optimal control problems with state constraints are considered. One way to go without this tool would be to make regularity assumptions on the control problems. As shown in [DH98, Lemma 3.11], Lipschitz continuity for the state multiplier can be ensured under such assumptions. However, the said assumptions involve uniform independence conditions ([DH98, p. 699]) that imply that the active state constraints are of first order. Problems with higher order state constraints may have a more complicated structure. As the formulae needed for the proofs in this work can be applied for the Stieltjes integral, the notions necessary for its definition as well as basic properties shall be introduced briefly. The definitions and properties gathered in this section have been collected from [Wid46], [Nat75] and [Ger06].

Definition 2.8 (Subdivision)
A subdivision $\mathbb{G}$ of an interval $[a,b]$ is an $(n+2)$-tuple $\mathbb{G} = (t_i)_{i=0,\dots,n+1}$ of points with $a = t_0 < \cdots < t_{n+1} = b$. The coarseness δ of a subdivision is defined by $\delta(\mathbb{G}) := \max_{i=0,\dots,n}(t_{i+1} - t_i)$.

The notion of a subdivision is needed for the definition of functions of bounded variation, a function space of great significance in the theory of Stieltjes integration. The definition is cited from [Ger06, p. 21] and [Ger06, p. 24]:

Definition 2.9 (Functions of Bounded Variation)
A function $f : [a,b] \to \mathbb{R}$ is of bounded variation*, if there exists a constant K, such that for any partition*

$$\mathbb{G}_n := \{a = t_0 < \cdots < t_{n+1} = b\}$$

of $[a,b]$ it holds that

$$\sum_{i=1}^{n+1} |f(t_i) - f(t_{i-1}| \leq K.$$

The total variation of f *is*

$$TV(f,a,b) := \sup_{\mathbb{G}_\kappa} \sum_{i=1}^{n+1} |f(t_i) - f(t_{i-1})|.$$

The space $BV([a,b],\mathbb{R})$ consists of all functions of bounded variation on $[a,b]$. The space of normalized functions of bounded variation *$NBV([a,b],\mathbb{R})$ consists of all functions f of bounded variation that are continuous from the right on (a,b) and satisfy $f(a) = 0$.*

The following definition is cited from [Wid46, p. 4]:

Definition 2.10 (Stieltjes Integral)
If the limit

$$\lim_{\delta(\mathbb{G})\to 0} \sum_{i=0}^{n} f(\tau_i)[\alpha(t_{i+1}) - \alpha(t_i)],$$

where

$$t_i \leq \tau_i \leq t_{i+1} \qquad (i = 0, \ldots, n),$$

exists independently of the choice of subdivision $\mathbb{G}$ and of the choice of the numbers τ_i, then the limit is called the Stieltjes integral of f with respect to α from a to b and is denoted by

$$\int_a^b f(t)d\alpha(t).$$

The existence of the Stieltjes integral can be shown if a continuous function is integrated with respect to a function of bounded variation [Wid46, Th. 4a]:

Theorem 2.11 (Existence of the Stieltjes Integral)
If f is continuous and α is of bounded variation in (a, b), then the Stieltjes integral of f with respect to α from a to b exists.

Theorem 2.11 provides sufficient conditions for the existence of the Stieltjes integral that are not necessarily fulfilled by f and α if the integral exists. Thus, the partial integration rule is cited from [Nat75, p. 257, point 5], since it makes weaker assumptions than the analog rule presented in [Wid46]:

Lemma 2.12 (Integration by Parts)
If one of the integrals $\int_a^b f(t)dg(t)$ or $\int_a^b g(t)df(t)$ exists, then so does the other, and it holds:

$$\int_a^b f(t)dg(t) + \int_a^b g(t)df(t) = [f(t)g(t)]_a^b,$$

where

$$[f(t)g(t)]_a^b = f(b)g(b) - f(a)g(a).$$

The following lemmata are cited from [Ger06, p. 22-23]. The first lemma deals with Stieltjes integrals in which the function μ is itself defined by a Stieltjes integral.

Lemma 2.13
Let g be continuous and h of bounded variation in $[a, b]$. Let

$$\mu(t) = \int_c^t g(\tau)dh(\tau), \qquad a \leq c \leq b, a \leq t \leq b,$$

then

$$\int_a^b f(t)d\mu(t) = \int_a^b f(t)g(t)dh(t).$$

Consequently, one can ask how functions that are expressed in terms of a Lebesgue integral behave if used in the Stieljes integral. In this case, the integral in question can also be expressed as a Lebesgue integral.

Lemma 2.14 (Stieltjes integral and Lebesgue integral)
If f is of bounded variation and μ is absolutely continuous on $[a,b]$, then

$$\int_a^b f(t)d\mu(t) = \int_a^b f(t)\mu'(t)dt,$$

where the integral on the right is a Lebesgue integral.

Riesz' Theorem states that analogous to absolutely continuous functions that can be expressed by means of the Lebesgue integral, continuous functions can be expressed by the Stieltjes integral:

Theorem 2.15 (Riesz)
Let the functional $\varphi : \mathcal{C}([a,b],\mathbb{R}) \to \mathbb{R}$ be continuous. Then there exists a unique function $\mu \in NBV([a,b])$, such that

$$\varphi(f) = \int_a^b f(t)d\mu(t) \qquad \forall f \in \mathcal{C}([a,b],\mathbb{R}).$$

As a consequence of this theorem, dual elements of continuous functions can be expressed using functions of bounded variation and the Stieltjes integral. This theorem provides a natural representation of multipliers for state constraints in the minimum principle.

2.1.2. Linear Operators and Fréchet differentiability

The following definition provides a concept of differentiability on normed spaces as it is needed in optimal control theory in infinite dimensional spaces. In this definition, the space $\mathcal{L}(X,Y)$ consists of all linear continuous functions from X to Y. An equivalent definition can be found in [Wer07, p. 113]. The equivalence is shown in [Wer07, Lem. III.5.2].

Definition 2.16 ((Continuous) Fréchet Differentiability)
Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces. The function $F : X \to Y$ is Fréchet differentiable in x_0, if there exists an operator $F'(x_0) \in \mathcal{L}(X,Y)$, such that

$$F(x_0 + h) = F(x_0) + F'(x_0)(h) + o(\|h\|_X) \tag{2.1}$$

for $h \to 0$. If it exists for all $x_0 \in U \subset X$, with U an open subset of X, then

$$F' : U \to \mathcal{L}(X,Y), \quad x \mapsto F'(x)$$

is called the Fréchet derivative of F in U.

F is continuously Fréchet differentiable in $U \subset X$ if F is differentiable and the derivative is continuous.

One of the most important tools for calculations with Fréchet differentiability is the chain rule (cf. [IT79, p. 27]), the assertion about continuous differentiability is an immediate implication:

Lemma 2.17 (Chain Rule)
Let $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$ and $(Z, \|\cdot\|_Z)$ be Banach spaces. Let $U \subset X$ and $V \subset Y$ be open subsets of X and Y, respectively. Let $F : U \to Y$ and $G : V \to Z$. Assume that there exists a point $x \in U$, such that $F(x) \in V$. If F is Fréchet differentiable in x and G is Fréchet differentiable in $F(x)$, then the mapping $H := G \circ F$ is Fréchet differentiable in x, and

$$H'(x) = G'(F(x)) \circ F'(x).$$

If F is continuously differentiable on U and G' is continuous on $F(U)$, then H is continuously differentiable on U.

Differentiability is inherited by subspaces, as it is shown in the following lemma:

Lemma 2.18
Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, and let $T : X \to Y$. Let $(U, \|\cdot\|_U)$ be a third normed space with $U \subset X$, and suppose that there exists some $C > 0$, such that $\|u\|_U \geq C \cdot \|u\|_X$ for $u \in U$. If T is Fréchet differentiable in u_0 with respect to $(X, \|\cdot\|_X)$, then $T|_U : U \to Y$ is Fréchet differentiable in u_0 with respect to the $\|\cdot\|_U$-norm, and the derivative is inherited,

$$T|_U{}'(u_0) = (T'(u_0))|_U.$$

Proof.

1. The linearity of $(T'(u_0))|_U$ is clear. The continuity is also easily shown, since for $h \in U$, it holds:
$$\|T'(u_0)(h)\|_Y \leq C_T \|h\|_X \leq C_T/C \cdot \|h\|_U,$$
where C is the constant mentioned in the assumptions of the lemma.

2. In order to show that (2.1) holds, the inequality
$$\lim_{\|h\|_U \to 0} \frac{\|T(u_0+h) - T(u_0) - T'(u_0)(h)\|_Y}{\|h\|_U} \leq \lim_{\substack{h \in U \\ \|h\|_X \to 0}} \frac{\|T(u_0+h) - T(u_0) - T'(u_0)(h)\|_Y}{C \cdot \|h\|_X}$$
can be analyzed. Note that $\|h\|_U \to 0 \Rightarrow \|h\|_X \to 0$. Since T is Fréchet differentiable in $(X, \|\cdot\|_X)$, the right hand side equals 0, and hence so does the left hand side. □

Lemma 2.18 leads to an important finding: If an operator $T : L^\infty([t_0, t_f], \mathbb{R}^n) \to Y$ is differentiable in some point x_0, then $T|_{W^{1,\infty}([t_0,t_f],\mathbb{R}^n)}$ is also differentiable, and

$$(T|_{W^{1,\infty}([t_0,t_f],\mathbb{R}^n)})' = T'|_{W^{1,\infty}([t_0,t_f],\mathbb{R}^n)}.$$

In other words, in order to calculate derivatives for operators that map $W^{1,\infty}$ into some space, it is sufficient to show that the operator that maps L^∞ into the same space is differentiable.

Example 2.19 (Examples: Linear Continuous Operators)
The following are examples of linear continuous operators. According to the definition of differentiability, they remain invariant under differentiation.

2.19.1 *Let $p \in \mathbb{N} \cup \{\infty\}$, and let T be the derivation operator for $W^{1,p}$, i.e. $T : W^{1,p} \to L^p$, $x \mapsto \dot{x}$. Then T is linear and continuous since $\|T(x) - T(y)\|_p \leq \|x - y\|_{1,p}$. Hence the Fréchet derivative $T'(x_0)$ is T itself (compare [Wer07, p. 149]).*

2.19.2 *Let $p \in \mathbb{N} \cup \{\infty\}$, and $T : L^p([t_0, t_f], \mathbb{R}^n) \to W^{1,p}([t_0, t_f], \mathbb{R}^n)$, with $T(f)(t) := \int_{t_0}^{t} f(\tau)d\tau$. Then T is linear and continuous since $\|T(x - y)\|_\infty \leq \int_{t_0}^{t_f} |x(t) - y(t)|dt = \|x - y\|_1 \leq C \cdot \|x - y\|_p$. Again, the Fréchet derivative $T'(x_0)$ is T itself.*

2.19.3 *Let F^τ be the operator that evaluates a function in $W^{1,p}$ at a given time τ, i.e. $F^\tau : W^{1,p} \to \mathbb{R}^n$, $x \mapsto x(\tau)$. Then F^τ is linear and differentiable with $(F^\tau)'(x_0) = F^\tau$, since F^τ is continuous with respect to this norm: Application of theorem 2.6 yields $\|F^\tau(x) - F^\tau(y)\| = \|x(\tau) - y(\tau)\| \leq \|x - y\|_\infty \leq \|x - y\|_{1,p}$.*

2.19.4 *Let $f \in L^\infty([t_0, t_f], \mathbb{R})$ and $F : L^\infty([t_0, t_f], \mathbb{R}) \to L^\infty([t_0, t_f], \mathbb{R})$, $x(\cdot) \mapsto f(\cdot)x(\cdot)$. Then F is linear and continuous since*

$$\|F(y) - F(x)\|_\infty = \|f(\cdot)y(\cdot) - f(\cdot)x(\cdot)\|_\infty = \|f(\cdot)\left[y(\cdot) - x(\cdot)\right]\|_\infty \leq \|f\|_\infty \cdot \|y - x\|_\infty.$$

Example 2.20 (Examples: Differentiable operators)
2.20.1 *Let $n, m \in \mathbb{N}$ and $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^m$ such that f is continuous and continuously differentiable with respect to x. Then $T : L^\infty([t_0, t_f], \mathbb{R}^n) \to L^\infty([t_0, t_f], \mathbb{R}^m)$, defined by $T(x)(t) := f(t, x(t))$, is continuously differentiable.*

The Fréchet derivative is $T'(x_0)(h)(t) = f'_x(t, x_0(t))h(t)$.

Proof.
2.20.1 Compare [KWW78, lemma 1.4a]. Obviously, $T'(x_0)$ defined as

$$T'(x_0)(h)(t) := f'_x(t, x_0(t))h(t)$$

is linear. It first remains to show that $T'(x_0)$ is continuous with respect to h, and that (2.1) is satisfied. Finally, it is shown that the derivative is continuous (with respect to x).

1. Continuity with respect to h can be shown as follows:

$$\begin{aligned}\|T'(x_0)(h)\|_\infty &= \|f'_x(\cdot, x_0(\cdot))h(\cdot)\|_\infty \\ &\leq \|f'_x(\cdot, x_0(\cdot))\|_{\mathcal{L}(L^\infty, L^\infty)} \cdot \|h\|_\infty\end{aligned}$$

   Now it holds that $\|x_0\|_\infty \leq C_{x_0}$. Since f'_x is continuous, there exists a constant C_f, such that $|f'_x(t, x)| \leq C_f$ for all $t \in [t_0, t_f]$, $x \leq C_{x_0}$, which yields:

$$\|T'(x_0)(h)\|_\infty \leq C_f \|h\|_\infty$$

   This shows that $T'(x_0)$ is continuous.

2. For (2.1), consider

$$\begin{aligned}&\|T(x_0 + h)(t) - T(x_0)(t) - T'(x_0)(h)(t)\|_{\mathbb{R}^m} \\ &\quad = \|f(t, x_0(t) + h(t)) - f(t, x_0(t)) - f'_x(t, x_0(t))h(t)\|_{\mathbb{R}^m}\end{aligned}$$

We use the mean value theorem to estimate the norm:

$$\begin{aligned}
&\|T(x_0+h)(t) - T(x_0)(t) - T'(x_0)(h)(t)\|_{\mathbb{R}^m} \\
&\quad = \left\| \int_0^1 f'_x(t, x_0(t) + \tau h(t))h(t)d\tau - f'_x(t, x_0(t))h(t) \right\|_{\mathbb{R}^m} \\
&\quad = \left\| \int_0^1 f'_x(t, x_0(t) + \tau h(t)) - f'_x(t, x_0(t))d\tau \cdot h(t) \right\|_{\mathbb{R}^m} \\
&\quad \leq \left\| \int_0^1 f'_x(t, x_0(t) + \tau h(t)) - f'_x(t, x_0(t))d\tau \right\|_{\mathbb{R}^{n\times m}} \cdot \|h(t)\|_{\mathbb{R}^m} \qquad (2.2)
\end{aligned}$$

Let $\tilde{\tau}(t)$ be defined as

$$\tilde{\tau}(t) := \arg\max_{\tau\in[0,1]} \|f'_x(t, x_0(t) + \tau h(t)) - f'_x(t, x_0(t))\|_{\mathbb{R}^{n\times m}},$$

and $\tilde{x}(t) := x_0(t) + \tilde{\tau}h(t)$, so that $\tilde{x}(t) \in [x_0(t), x_0(t) + h(t)]$. Therefore, $\|\tilde{x}(t) - x_0(t)\| \leq \|h\|_\infty$ for all $t \in [t_0, t_f]$.

Applying the essential supremum on both sides of inequality (2.2) and dividing by $\|h\|_\infty$ yields

$$\|T(x_0+h) - T(x_0) - T'(x_0)(h)\|_\infty \cdot (\|h\|_\infty)^{-1} \leq \|f'_x(\cdot, \tilde{x}(\cdot)) - f'_x(\cdot, x_0(\cdot))\|_\infty,$$

For sufficiently small h, $(t, x_0(t))$ and $(t, \tilde{x}(t))$ remain on the compact set $\{(t, x)|t \in [t_0, t_f], \|x(t) - x_0(t)\| \leq 1\}$. According to Cantor's theorem, f'_x is uniformly continuous on this set, and it holds

$$\lim_{h\to 0} \|f'_x(\cdot, \tilde{x}(\cdot)) - f'_x(\cdot, x_0(\cdot))\|_\infty = 0,$$

which shows that

$$\lim_{h\to 0} \|T(x_0+h) - T(x_0) - T'(x_0)(h)\|_\infty \cdot (\|h\|_\infty)^{-1} = 0.$$

3. For the continuity of $T' : L^\infty \to \mathcal{L}(L^\infty, L^\infty)$, note that

$$\begin{aligned}
\|T'(x) - T'(y)\|_{\mathcal{L}(L^\infty,L^\infty)} &= \sup_{\|h\|_\infty=1} \|[f'_x(\cdot, x(\cdot)) - f'_x(\cdot, y(\cdot))]h(\cdot)\|_\infty \\
&\leq \sup_{\|h\|_\infty=1} \|f'_x(\cdot, x(\cdot)) - f'_x(\cdot, y(\cdot))\|_\infty \cdot \|h(\cdot)\|_\infty \\
&= \|f'_x(\cdot, x(\cdot)) - f'_x(\cdot, y(\cdot))\|_\infty.
\end{aligned}$$

Hence, application of Cantor's theorem yields

$$\lim_{x\to y} \|T'(x) - T'(y)\|_{\mathcal{L}(L^\infty,L^\infty)} \leq \lim_{x\to y} \|f'_x(\cdot, x(\cdot)) - f'_x(\cdot, y(\cdot))\|_\infty = 0. \qquad \square$$

The theorem below is cited from [Ger06, Theorem 2.2.8] and later used in the same context: The theorem gives conditions under which the image of an operator is closed. These assumptions are therefore useful for the proof of normality of optimality problems.

Theorem 2.21
Let X and Y be Banach spaces. Let $F : X \to Y \times \mathbb{R}^n$ be defined by $F(x) = (G(x), H(x))$, where $G : X \to Y$ is a linear, continuous and surjective operator and $H : X \to \mathbb{R}^n$ is linear and continuous. Then $\mathrm{im}(F)$ *is closed in $Y \times \mathbb{R}^n$.*

The following is the implicit function theorem, which plays an important role for regularization of the complementarity systems as shown in chapter 5. The first part (the existence) is cited from [Wer07, Theorem III.5.4]. The derivative of the implicitly defined function g can be obtained from deriving the equality $F(x, g(x)) = 0$.

Theorem 2.22 (Implicit Function Theorem)
Let X, Y and Z be complete and $F : X \times Y \supset U \times V \to Z$ continuously differentiable with $F(x_0, y_0) = 0$. Let the derivative of $y \mapsto F(x_0, y)$ be an isomorphism of Y on Z. Then there exist neighborhoods U_0 of x_0 and V_0 of y_0, such that for all $x \in U_0$, the equation $F(x, y) = 0$ has a unique solution $y =: g(x)$ in V_0, and the so defined function $g : U_0 \to V_0$ is continuously differentiable with $g'_x = -F'^{-1}_y F'_x$.

Carathéodory's existence Theorem (cf. [Wal00, Theorem 18, p. 128]) ensures the existence of solution to an initial value problem

$$\dot{x} = f(t, x(t), u(t)), \qquad x(t_0) = x_0 : \tag{IVP}$$

Theorem 2.23 (Carathéodory's Existence Theorem)
Let $f : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ be continuous and locally Lipschitz continuous in x, i.e. for every $R > 0$ there exists a constant $L_R > 0$ such that

$$\|f(t, x_1, u) - f(t, x_2, u)\| \leq L_R \cdot \|x_1 - x_2\|$$

for all $x_1, x_2 \in \mathbb{R}^{n_x}$ and all $u \in \mathbb{R}^{n_u}$ with $\|x_1\|, \|x_2\|, \|u\| < R$, everywhere in $[t_0, t_f]$. Then for any $x_0 \in \mathbb{R}^{n_x}$ and any control $u \in L^\infty([t_0, t_f], \mathbb{R}^{n_u})$ there exists a function $x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$ that satisfies IVP for almost all $t \in [t_0, t_f]$.

In the following lemma, we examine linear boundary value problems. A more general version of this lemma (with application to Differential Algebraic Equations) can be found in [Ger06, Lemma 4.1.6].

Lemma 2.24
Let $A \in L^\infty([t_0, t_f], \mathbb{R}^{n_x \times n_x})$, $h_1 \in L^\infty([t_0, t_f], \mathbb{R}^{n_x})$, $C_0, C_f \in \mathbb{R}^{r \times n_x}$ and $h_2 \in \mathbb{R}^r$. Consider the linear boundary value problem

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + h_1(t) \\ h_2 &= C_0 x(t_0) + C_f x(t_f). \end{aligned}$$

2.24.1 *If the given problem is an initial value problem, i.e. $r = n_x$, $C_0 = I$ and $C_f = 0$, then there is a unique solution for x, given by*

$$x(t) = \Phi(t) \left(h_2 + \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau \right).$$

Here, Φ denotes the fundamental solution of the differential equation, i.e. $\Phi(t_0) = I$, $\Phi'(t) = A(t)\Phi(t)$.

2.24.2 *If*

$$\operatorname{rank}\left(C_0\Phi(t_0) + C_f\Phi(t_f)\right) = r,$$

where Φ *again is the fundamental solution as above, then the boundary value problem has a solution.*

A parameter free version of the following lemma can be found in [Ger08]. This version follows straightforwardly from the lemma since the estimate holds for each parameter.

Theorem 2.25 (boundary value problems)
Let P *be some arbitrary parameter set. Consider the boundary value problem* $G(p)(\xi) = 0$*, defined by the parameterized operator*

$$G : P \times W^{1,\infty}([t_0, t_f], \mathbb{R}^n) \to L^\infty([t_0, t_f], \mathbb{R}^n) \times \mathbb{R}^m,$$

where

$$G(p)(\xi) = \begin{pmatrix} \xi'(t) - B(p)(t)\xi(t) \\ E_0\xi(t_0) + E_1\xi(t_f) \end{pmatrix}.$$

Let the following assumptions be satisfied.

1. *There exists* C *such that for all* $p \in P$ *and a.e. in* $[t_0, t_f]$ *it holds* $||B(p)(t)|| \leq C$*.*
2. *There exists* $\kappa > 0$ *such that for all* $p \in P$ *and all* $\zeta \in \mathbb{R}^n$ *it holds*

$$||(E_0\theta_p(t_0) + E_1\theta_p(t_f))\zeta|| \geq \kappa||\zeta||,$$

where θ_p *is a fundamental solution with* $\theta_p'(t) = B(p)(t)\theta_p(t), \quad \theta_p(t_0) = I.$

Then the inverse operator $G(p)^{-1}$ *exists and it holds* $||G(p)^{-1}|| \leq K$ *for some constant* K*, independent from the parameter* p*.*

For the following lemma is taken from [Kön00, p. 103]. The proof remains the same:

Lemma 2.26
Let X*,* Y *be Banach. A Fréchet differentiable function* $F : X \to Y$ *with* $\|F'\|_{\mathcal{L}(X,Y)} \leq L$ *for some* $L > 0$ *is Lipschitz continuous, with*

$$\|F(x) - F(y)\|_Y \leq L \cdot \|x - y\|_X$$

Proof.
Let $\gamma(t) := y + t(x - y)$. For $\varepsilon \in \mathbb{R}$ let $F_\varepsilon : [0, 1] \to \mathbb{R}$,

$$F_\varepsilon(t) := \|F(\gamma(t)) - F(y)\|_Y - t \cdot (L + \varepsilon)\|x - y\|_X.$$

Assume that $F_\varepsilon(1) > 0$ for some $\varepsilon > 0$. Since F is continuous, so is F_ε, and hence there exists a time $t_0 \in (0, 1]$, such that $F(t_0) < F(t)$ for all $t \in (t_0, 1]$.

Hence $\frac{F_\varepsilon(t) - F_\varepsilon(t_0)}{t - t_0} > 0$ for all $t \in (t_0, 1]$, and

$$0 < \frac{F_\varepsilon(t) - F_\varepsilon(t_0)}{t - t_0}$$

$$= \frac{\|F(\gamma(t)) - F(y)\|_Y - \|F(\gamma(t_0)) - F(y)\|_Y}{t - t_0} - (L + \varepsilon)\|x - y\|_X$$
$$\leq \frac{\|F(\gamma(t)) - F(\gamma(t_0))\|_Y}{t - t_0} - (L + \varepsilon)\|x - y\|_X,$$

which implies that

$$(L + \varepsilon)\|x - y\|_X \leq \lim_{t \searrow t_0} \frac{\|F(\gamma(t)) - F(\gamma(t_0))\|_Y}{t - t_0}$$

At the same time, due to the definition of the derivative, it holds

$$\lim_{t \searrow t_0} \frac{\|F(\gamma(t)) - F(\gamma(t_0))\|_Y}{t - t_0} = \|F'(\gamma(t_0))(x - y)\|_Y \leq L\|x - y\|_X,$$

and the last to inequalities are inconsistent.

Summarizing, it holds $F_\varepsilon(1) \leq 0$ for all $\varepsilon > 0$, which shows the assertion. □

2.2. Control Theory

The concept of controllability is used in various situations as a controllability assumption is usually needed in order to assert regularity properties of the problem and its associated multipliers. We will first briefly introduce the necessary concepts from control theory as far as they are needed on the way to the definition of controllability.

The main merit of this introduction will be a weak regularity assumption for optimal control problems that guarantees the validity of necessary optimal control conditions and can even be checked in practice. The definitions and theorems can be found in [Son98].

Definition 2.27 (Continuous-time Control System)
Let $x_0 \in \mathbb{R}^{n_x}$ and $f : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ be continuous and continuously differentiable with respect to x and u, i.e., its second and third argument. Let $[t_0, t_f]$ be an interval.

Let $\xi(\cdot, t_*, x_*, u)$ denote the solution of the initial value problem

$$\dot{x}(t) = f(t, x(t), u(t)),$$
$$x(t_*) = x_*.$$

Then the $\Sigma_f := ([t_0, t_f], \mathbb{R}^{n_x}, \mathbb{R}^{n_u}, \xi)$, which consists of the time set, the state and control space and the solution mapping of the ordinary differential equation, is called a continuous-time control system with the right hand side f.

Definition 2.28 (Linear Continuous-time Control System)
Let Σ_f be a continuous-time control system, where the right hand side f is in the form

$$f(t, x, u) = A(t)x + B(t)u$$

with $A \in \mathcal{C}^1([t_0, t_f], \mathbb{R}^{n_x \times n_x})$, $B \in \mathcal{C}^1([t_0, t_f], \mathbb{R}^{n_x \times n_u})$.

The terms "event" and "reachability" make the definition of controllability easier. The definitions chosen for this work applies the more general definitions in [Son98, Definition 3.1.1] to the case of continuous-time control systems:

Definition 2.29 (Event, Reachability)
Let $\Sigma_f = ([t_0, t_f], \mathbb{R}^{n_x}, \mathbb{R}^{n_u}, \xi)$ be a continuous-time control system.

An event is a pair $(x, t) \in \mathbb{R}^{n_x} \times [t_0, t_f]$.

The event (z, τ) can be reached from the event (x, σ) if there is a path of Σ_f on $[\sigma, \tau]$ whose initial state is x and final state is z, that is, if there exists a $u : [\sigma, \tau) \to \mathbb{R}^{n_u}$, such that

$$z = \xi(\tau, \sigma, x, u).$$

Definition 2.30 (Controllability)
The control system $\Sigma_f = ([t_0, t_f], \mathbb{R}^{n_x}, \mathbb{R}^{n_u}, \xi)$ is controllable on the interval $[\sigma, \tau]$ if for each $x, z \in \mathbb{R}^{n_x}$ it holds that (z, τ) can be reached from (x, σ).

The last definition (cited from [Son98, Definition 3.1.6]) together with the definition of linear continuous-time control systems, leads to the question under which conditions these are controllable on the interval $[t_0, t_f]$, since this question is particularly interesting for developing necessary optimality conditions under mild assumptions. The following theorem sums up the results from Proposition 3.5.16 and Corollary 3.5.18 and Remark 3.5.19 in [Son98, p. 113, p. 115].

Theorem 2.31
Let Σ_f be a continuous-time linear system with right hand side

$$f(t, x, u) = A(t)x + B(t)u.$$

Let $k > 0$ be an integer, such that $A \in \mathcal{C}^k([t_0, t_f], \mathbb{R}^{n_x \times n_x})$ and $B \in \mathcal{C}^k([t_0, t_f], \mathbb{R}^{n_x \times n_u})$.

For $i = 0, \ldots, k-1$ let

$$B_0(t) := B(t)$$
$$B_{i+1}(t) := A(t)B_i(t) - \frac{d}{dt}B_i(t)$$

If there exists $\tau \in [t_0, t_f]$, for which

$$\operatorname{rank}[B_0(\tau), B_1(\tau), \ldots, B_k(\tau)] = n_x,$$

then Σ_f is controllable on $[t_0, t_f]$.

3. The Minimum Principle

3.1. The Optimal Control Problem and Infinite Dimensional Optimalization Problems

The most important tool for analyzing Optimal Control Problems are necessary optimality conditions. The general idea for proving necessary conditions in terms of a maximum principle was first invented by Pontrjagin et al. [PBGM64]. Later, the principle was adapted to many classes of problems. An overview of adaptions can be found in [HSV95]. In [Ger06], optimality principles were proved for problems with either control-state constraints or control set constraints.

The necessary conditions, that are derived in this section, will be applied to Linear Quadratic Optimal Control Problems with pure state constraints and to problems with control-state and control set constraints. The assumption that makes the problem accessible for an analysis analogous to the one found in [Ger06, Chap. 4] is that the control set constraint refers to components of the control that do not occur in the mixed control state constraints. We consider the following OCP:

Problem 3.1 (Optimal Control Problem (OCP))

$$\text{min!} \quad J(x,u,v) := \varphi(x(t_0),x(t_f)) + \int_{t_0}^{t_f} f_0(t,x(t),u(t),v(t))dt$$

$$\begin{aligned} \textit{with respect to the state function} \quad & x \in W^{1,\infty}([t_0,t_f],\mathbb{R}^{n_x}) \\ \textit{and the control functions} \quad & u \in L^{\infty}([t_0,t_f],\mathbb{R}^{n_u}) \\ \textit{and} \quad & v \in L^{\infty}([t_0,t_f],\mathbb{R}^{n_v}) \end{aligned}$$

subject to the differential equation

$$\dot{x}(t) = f(t,x(t),u(t),v(t)) \quad \textit{a.e. in } [t_0,t_f],$$

boundary conditions

$$\Psi(x(t_0),x(t_f)) = 0,$$

mixed control state constraints for v

$$c(t,x(t),v(t)) \leq 0,$$

pure state constraints

$$s(t,x(t)) \leq 0$$

and set constraints for u

$$u(t) \in U(t) \subset \mathbb{R}^{n_u} \quad \textit{a.e. in } [t_0,t_f]$$

Necessary optimality conditions are derived from conditions for infinite optimization problems. Therefore, the objective function as well as the constraints have to be embedded into suitable spaces. For simplicity, the following set is defined:

$$U_{ad} := \{\nu \in L^\infty([t_0, t_f], \mathbb{R}^{n_u}) | \nu(t) \in U(t) \text{ a.e. on } [t_0, t_f]\}.$$

In order to exploit Fréchet differentiability of the corresponding functions when applying the necessary optimality conditions for optimization problems, the following smoothness assumptions will be made:

Assumption 3.2 (Smoothness)
3.2.1 *The function $\varphi : (\mathbb{R}^{n_x})^2 \to \mathbb{R}$ is differentiable.*

3.2.2 *The mapping $f_0 : [t_0, t_f] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \to \mathbb{R}$ is continuous and continuously differentiable with respect to (x, u, v).*

3.2.3 *The ODE defining function $f : [t_0, t_f] \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_x}$ is continuous and continuously differentiable with respect to (x, u, v).*

3.2.4 *The function $\Psi : (\mathbb{R}^{n_x})^2 \to \mathbb{R}^{n_\Psi}$ that defines the boundary conditions is continuously differentiable.*

3.2.5 *The control state constraint $c : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_c}$ is continuous and continuously differentiable with respect to (x, v), and the state constraint $s : \mathbb{R} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_s}$ is continuous and continuously differentiable with respect to x.*

3.2.6 *There exists a function $\hat{u} \in U_{ad}$, such that $u \in U_{ad}$ for all u with $||u - \hat{u}||_\infty \leq \varepsilon$ for some $\varepsilon > 0$. The set U_{ad} is closed and convex.*

The following Banach spaces will be used throughout the remainder of this chapter:

Definition 3.3 (The spaces X, Y, Z, the cone K and the admissible set S)
3.3.1 *The space X denotes the space of optimization variables, i.e. $(x, u, v) \in X$, with*

$$X := W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x}) \times L^\infty([t_0, t_f], \mathbb{R}^{n_u}) \times L^\infty([t_0, t_f], \mathbb{R}^{n_v}).$$

Together with the norm

$$\|(x, u, v)\|_X := \max\{\|x\|_{1,\infty}, \|u\|_\infty, \|v\|_\infty\},$$

the tuple $(X, \|\cdot\|_X)$ becomes a Banach space.

3.3.2 *The space Z will be used as an image space for the equality constraints,*

$$Z := L^\infty([t_0, t_f], \mathbb{R}^{n_x}) \times \mathbb{R}^{n_\Psi}.$$

The natural norm that makes $(Z, \|\cdot\|_Z)$ a Banach space is

$$\|(z_1, z_2)\|_Z := \max\{\|z_1\|_\infty, \|z_2\|\}.$$

The equality constraints can be expressed as $H(x, u, v) = 0$ with

$$\begin{aligned} H &:= (H_1, H_2) : X \to Z, \\ H_1(x, u, v) &:= f(\cdot, x(\cdot), u(\cdot), v(\cdot)) - \dot{x}(\cdot), \\ H_2(x, u, v) &:= -\Psi(x(t_0), x(t_f)). \end{aligned}$$

3.3.3 *The inequality constraints can be handled in a similar manner as above, introducing the space Y and the cone K. Let*

$$Y := L^\infty([t_0, t_f], \mathbb{R}^{n_c}) \times \mathcal{C}([t_0, t_f], \mathbb{R}^{n_s}),$$

which is a Banach space with

$$\|(y_1, y_2)\|_Y := \max\{\|y_1\|_\infty, \|y_2\|_\infty\}.$$

Let

$$\begin{aligned} G &:= (G_1, G_2) : X \to Y \\ G_1(x, u, v) &:= -c(\cdot, x(\cdot), v(\cdot)) \\ G_2(x, u, v) &:= -s(\cdot, x(\cdot)). \end{aligned}$$

Using the cone K, defined as

$$\begin{aligned} K &:= K_1 \times K_2 \subset Y \\ K_1 &:= \{z \in L^\infty([t_0, t_f], \mathbb{R}^{n_c}) | z(t) \geq 0_{n_c} \text{ a.e. in } [t_o, t_f]\} \\ K_2 &:= \{z \in \mathcal{C}([t_0, t_f], \mathbb{R}^{n_s}) | z(t) \geq 0_{n_s} \text{ a.e. in } [t_0, t_f]\}, \end{aligned}$$

the inequality constraints are equivalent to

$$G(x, u, v) \in K.$$

3.3.4 *The set S of admissible optimization variables is*

$$S := W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x}) \times U_{ad} \times L^\infty([t_0, t_f], \mathbb{R}^{n_v}).$$

Lemma 3.4 (Properties)
3.4.1 *The objective function $J : X \to \mathbb{R}$ is Fréchet differentiable if the smoothness Assumptions 3.2.1 and 3.2.2 hold, with derivative*

$$\begin{aligned} J'(\hat{x}, \hat{u}, \hat{v})(x, u, v) =& \varphi'_{x_0}(\hat{x}(t_0), \hat{x}(t_f))x(t_0) + \varphi'_{x_f}(\hat{x}(t_0), \hat{x}(t_f))x(t_f) \\ &+ \int_{t_0}^{t_f} f_{0x}'(t, \hat{x}, \hat{u}, \hat{v})x(t) + f_{0u}'(t, \hat{x}, \hat{u}, \hat{v})u(t) + f_{0v}'(t, \hat{x}, \hat{u}, \hat{v})v(t)dt. \end{aligned}$$

3.4.2 *If Assumptions 3.2.3 and 3.2.4 hold, then H is continuously Fréchet differentiable, and $H' = (H_1', H_2')$ with*

$$\begin{aligned} H_1'(\hat{x}, \hat{u}, \hat{v})(x, u, v) =& f_x'(\cdot, \hat{x}(\cdot), \hat{u}(\cdot), \hat{v}(\cdot))x(\cdot) + f_u'(\cdot, \hat{x}(\cdot), \hat{u}(\cdot), \hat{v}(\cdot))u(\cdot) \\ &+ f_v'(\cdot, \hat{x}(\cdot), \hat{u}(\cdot), \hat{v}(\cdot))v(\cdot) - \dot{x}(\cdot) \\ H_2'(\hat{x}, \hat{u}, \hat{v})(x, u, v) =& -\Psi'_{x_0}(\hat{x}(t_0), \hat{x}(t_f))x(t_0) - \Psi'_{x_f}(\hat{x}(t_0), \hat{x}(t_f))x(t_f). \end{aligned}$$

3.4.3 *G is continuously Fréchet differentiable if Assumption 3.2.5 holds, with*

$$\begin{aligned} G' &= (G_1', G_2'), \\ G_1'(\hat{x}, \hat{u}, \hat{v})(x, u, v) &= -c_x'(\cdot, \hat{x}(\cdot), \hat{v}(\cdot))x(\cdot) - c_v'(\cdot, \hat{x}(\cdot), \hat{v}(\cdot))v(\cdot), \\ G_2'(\hat{x}, \hat{u}, \hat{v})(x, u, v) &= -s_x'(\cdot, \hat{x}(\cdot))x(\cdot). \end{aligned}$$

3.4.4 *If Assumption 3.2.6 holds, then* $\operatorname{int}(S) \neq \emptyset$.

3.4.5 *Under Assumption 3.2.3,* $\operatorname{im}(H')$ *is closed.*

Proof.
3.4.1 Let
$J_1 : W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x}) \to \mathbb{R}^{2n_x} \qquad x \mapsto (x(t_0), x(t_f))$,
$J_2 : \mathbb{R}^{2n_x} \to \mathbb{R} \qquad (x_0, x_f) \mapsto \varphi(x_0, x_f)$.
Furthermore, let
$J_3 : L^\infty([t_0, t_f], \mathbb{R}^{n_x+n_u+n_v}) \to L^\infty([t_0, t_f], \mathbb{R})$
$(x(\cdot), u(\cdot), v(\cdot)) \mapsto f_0(\cdot, x(\cdot), u(\cdot), v(\cdot))$,
$J_4 : L^\infty([t_0, t_f], \mathbb{R}) \to W^{1,\infty}([t_0, t_f], \mathbb{R}) \qquad f \mapsto \int_{t_0}^{\cdot} f(\tau) d\tau$,
$J_5 : W^{1,\infty}([t_0, t_f], \mathbb{R}) \to \mathbb{R} \qquad f \mapsto f(t_f)$.

Then J can be seen as the composition and sum $J = J_2 \circ J_1 + J_5 \circ J_4 \circ J_3$. Each component is Fréchet differentiable according to Examples 2.20.1 (J_3), 2.19.3 (J_1, J_5), 2.19.2 (J_4) and Lemma 2.18. The composition then is differentiable by Lemma 2.17.

The smoothness assertions in 3.4.2 and 3.4.3 follow analogously. The assertion 3.4.4 is equivalent to Assumption 3.2.6.

Finally, Lemma 2.24.1 shows that H_1 is surjective. Since H_1 and H_2 are linear and continuous, Theorem 2.21 yields that $\operatorname{im}(H')$ is closed, as claimed in 3.4.5. □

The resulting optimization problem in infinite dimensional spaces (OP) reads

Problem 3.5 (OP)

$$\text{min!} \quad F(x, u, v)$$

with respect to the variables $(x, u, v) \in X$

subject to the conical constraints

$$G(x, u, v) \in K$$

equality constraints

$$H(x, u, v) = 0$$

and set constraints

$$(x, u, v) \in S$$

The following necessary optimality conditions [Ger06, Th. 3.4.2] hold for infinite dimensional optimization problems in the form of OP:

Theorem 3.6
Let $F : X \to \mathbb{R}$ *and* $G : X \to Y$ *be Fréchet differentiable and* $H : X \to Z$ *continuously Fréchet differentiable. Let* $\hat{x} \in X$ *be a local minimum of OP,* $\operatorname{int}(S) \neq \emptyset$ *and* $\operatorname{int}(K) \neq \emptyset$.

Let S be a closed and convex set and K a closed convex cone. Assume that $\operatorname{im}(H'(\hat{x}))$ *is not a proper dense subset of Z. Then there exist nontrivial multipliers $(l_0, \lambda^*, \mu^*) \in \mathbb{R} \times Y^* \times Z^*$, $(l_0, \lambda^*, \mu^*) \neq 0$, such that*

$$\begin{aligned} l_0 &\geq 0 \\ \lambda^* &\in K^+ \\ \lambda^*(G(\hat{x})) &= 0 \\ l_0 F'(\hat{x})(d) - \lambda^*(G(\hat{x}))(d) - \mu^*(H(\hat{x}))(d) &\geq 0 \quad \forall d \in S - \{\hat{x}\} \end{aligned}$$

Using Definition 3.3, our aim is to derive necessary optimality conditions for Problem 3.1 from Theorem 3.6. According to Lemma 3.4, the differentiability assumptions as well as the condition $\operatorname{int}(S) \neq \emptyset$ are already satisfied if Assumption 3.2 holds.

Applying Theorem 3.6 to the spaces and operators defined in Definition 3.3 leads to the corollary

Corollary 3.7
Let the smoothness Assumptions 3.2 hold for the OCP and $\operatorname{int}(K) \neq \emptyset$. *Let $(\hat{x}, \hat{u}, \hat{v}) \in X$ be a local weak minimum.*

Then there exist nontrivial multipliers $l_0 \in \mathbb{R}$, $\eta^* \in Y^*$ and $\lambda^* \in Z^*$, such that

$$\begin{aligned} l_0 &\geq 0 \\ \eta^* &\in K^+ \\ \eta^*(G(\hat{x}, \hat{u}, \hat{v})) &= 0 \\ l_0 F'(\hat{x}, \hat{u}, \hat{v})(x - \hat{x}, u - \hat{u}, v - \hat{v}) & \\ -\eta^*(G'(\hat{x}, \hat{u}, \hat{v})(x - \hat{x}, u - \hat{u}, v - \hat{v})) & \\ -\lambda^*(H'(\hat{x}, \hat{u}, \hat{v})(x - \hat{x}, u - \hat{u}, v - \hat{v})) &\geq 0 \quad \forall (x, u, v) \in S. \end{aligned}$$

3.2. Smoothness and representation of the multipliers

In this section, representations for the multipliers from Corollary 3.7 as measurable functions and functions of bounded variation are derived in contrast to their representation as elements of the dual spaces Y^* and Z^*.

Let an OCP in the form of Problem 3.1 be given, where all functions fulfill the smoothness Assumptions 3.2. Let $(\hat{x}, \hat{u}, \hat{v}) \in X$ be a weak local minimum, and let $l_0 \in \mathbb{R}$, $\eta^* \in Y^*$ and $\lambda^* \in Z^*$ be multipliers with

$$l_0 \geq 0 \tag{3.1}$$
$$\eta^* \in K^+ \tag{3.2}$$
$$\eta^*(G(\hat{x}, \hat{u}, \hat{v})) = 0 \tag{3.3}$$

$$\begin{aligned} & l_0 F'(\hat{x},\hat{u},\hat{v})(x-\hat{x},u-\hat{u},v-\hat{v}) \\ -&\eta^*(G'(\hat{x},\hat{u},\hat{v})(x-\hat{x},u-\hat{u},v-\hat{v})) \\ -&\lambda^*(H'(\hat{x},\hat{u},\hat{v})(x-\hat{x},u-\hat{u},v-\hat{v})) \geq 0 \quad \forall (x,u,v) \in S. \end{aligned} \tag{3.4}$$

Riesz' Theorem 2.15 yields another representation for the second component of the multiplier $\eta^* \in Y^*$ with $\eta^* = (\eta_1^*, \eta_2^*) \in (L^\infty([t_0,t_f],\mathbb{R}^{n_c}))^* \times (\mathcal{C}([t_0,t_f],\mathbb{R}^{n_s}))^*$: According to Theorem 2.15, there exist unique functions $\mu_i \in NBV([t_0,t_f],\mathbb{R})$, $i = 1,...,n_s$ with

$$\eta_2^*(f) = \sum_{i=1}^{n_s} \int_{t_0}^{t_f} f_i(t) d\mu_i(t).$$

The second component of the multiplier $\lambda^* \in Z^*$ also has a simple representation, since

$$\lambda^* =: (\lambda_f^*, \sigma) \in (L^\infty([t_o,t_f],\mathbb{R}^{n_x}))^* \times \mathbb{R}^{n_\Psi}.$$

If the inequality (3.4) is satisfied for all $(x,u,v) \in S$, then by setting $u = \hat{u}$, $v = \hat{v}$ it follows:

$$l_0 F_x'(\hat{x},\hat{u},\hat{v})(x-\hat{x}) - \eta^*\left(G_x'(\hat{x},\hat{u},\hat{v})(x-\hat{x})\right) - \lambda^*\left(H_x'(\hat{x},\hat{u},\hat{v})(x-\hat{x})\right) \geq 0 \qquad \forall x \in W^{1,\infty}([t_0,t_f],\mathbb{R}^{n_x}),$$

which is equivalent to

$$\begin{aligned} & (l_0 \varphi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)))x(t_0) \\ +&(l_0 \varphi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)))x(t_f) \\ +&\int_{t_0}^{t_f} l_0 f_{0x}'(t,\hat{x}(t),\hat{u}(t),\hat{v}(t))x(t)dt + \sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_{ix}'(t,\hat{x}(t))x(t)d\mu_i(t) \\ +&\eta_1^*\left(c_x'(\cdot,\hat{x}(\cdot),\hat{v}(\cdot))x\right) + \lambda_f^*(\dot{x} - f_x'(\cdot,\hat{x}(\cdot),\hat{u}(\cdot),\hat{v}(\cdot))x) = 0 \qquad \forall x \in W^{1,\infty}([t_0,t_f],\mathbb{R}^{n_x}). \end{aligned} \tag{3.5}$$

Analogously, setting $x = \hat{x}$, $u = \hat{u}$ yields

$$l_0 \int_{t_0}^{t_f} f_{0v}'(t,\hat{x}(t),\hat{u}(t),\hat{v}(t))v(t)dt + \eta_1^*\left(c_v'(\cdot,\hat{x}(\cdot),\hat{v}(\cdot))v\right) - \lambda_f^*\left(f_v'(\cdot,\hat{x}(\cdot),\hat{u}(\cdot),\hat{v}(\cdot))v\right) = 0 \qquad \forall v \in L^\infty([t_0,t_f],\mathbb{R}^{n_v}), \tag{3.6}$$

and finally, with $x = \hat{x}$, $v = \hat{v}$ it follows from (3.4) that

$$l_0 \int_{t_0}^{t_f} f_{0u}'(t,\hat{x}(t),\hat{u}(t),\hat{v}(t))u(t)dt - \lambda_f^*\left(f_u'(\cdot,\hat{x}(\cdot),\hat{u}(\cdot),\hat{v}(\cdot))u\right) \geq 0 \quad \forall u \in U_{ad} - \{\hat{u}\}. \tag{3.7}$$

For further investigation of the smoothness of the multipliers, equation (3.5) is analyzed, introducing $\bar{x}$ as the solution of the initial value problem

$$\dot{x} = f_x'[t]x + h_1(t), \qquad x(t_0) = 0 \tag{3.8}$$

with an arbitrary function $h_1 \in L^\infty([t_0, t_f], \mathbb{R}^{n_x})$. Here, the expression $f'_x[t]$ denotes the function f'_x along the weak local minimum $(\hat{x}, \hat{u}, \hat{v})$, i.e. $f'_x[t] := f'_x(t, \hat{x}(t), \hat{u}(t), \hat{v}(t))$. Analogous notations will from now on be used for f, s and c. Also, the dependence of φ and Ψ on variables will be omitted when they are evaluated in $(\hat{x}(t_0), \hat{x}(t_f))$.

According to Lemma 2.24, the solution of (3.8) is

$$\bar{x}(t) = \Phi(t) \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau,$$

where Φ solves $\Phi(0) = I$, $\Phi'(t) = f'_x[t]\Phi(t)$.

Hence, substituting x in (3.5) with $\bar{x}$ yields

$$\begin{aligned} 0 = {} & (l_0 \varphi'_{x_f} + \sigma^\top \Psi'_{x_f}) \Phi(t_f) \int_{t_0}^{t_f} \Phi^{-1}(\tau) h_1(\tau) d\tau \\ & + \int_{t_0}^{t_f} l_0 f_{0x}'[t] \Phi(t) \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau dt \\ & + \sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_{ix}'[t] \Phi(t) \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau d\mu_i(t) \\ & + \eta_1^* \left(c'_x \Phi(\cdot) \int_{t_0}^{\cdot} \Phi^{-1}(\tau) h_1(\tau) d\tau \right) + \lambda_f^* (h_1) \,. \end{aligned} \tag{3.9}$$

Integration by parts of the second term of the right hand side leads to

$$\begin{aligned} & \int_{t_0}^{t_f} l_0 f_{0x}'[t] \Phi(t) \left(\int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau \right) dt \\ & \quad = \left[\left(\int_{t_0}^{t} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \right) \cdot \left(\int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau \right) \right]_{t_0}^{t_f} \\ & \qquad - \int_{t_0}^{t_f} \left(\int_{t_0}^{t} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \right) \Phi^{-1}(t) h_1(t) dt \\ & \quad = \int_{t_0}^{t_f} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \cdot \int_{t_0}^{t_f} \Phi^{-1}(t) h_1(t) dt \\ & \qquad - \int_{t_0}^{t_f} \int_{t_0}^{t} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \cdot \Phi^{-1}(t) h_1(t) dt \\ & \quad = \int_{t_0}^{t_f} \int_{t_0}^{t_f} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \cdot \Phi^{-1}(t) h_1(t) dt \\ & \qquad - \int_{t_0}^{t_f} \int_{t_0}^{t} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \cdot \Phi^{-1}(t) h_1(t) dt \\ & \quad = \int_{t_0}^{t_f} \int_{t}^{t_f} l_0 f_{0x}'[\tau] \Phi(\tau) d\tau \cdot \Phi^{-1}(t) h_1(t) dt \end{aligned}$$

For processing the third term

$$\sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_{ix}'[t] \Phi(t) \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau d\mu_i(t),$$

we first make the following general observation: For three arbitrary functions a, b and μ, $a \in \mathcal{C}([t_0, t_f], \mathbb{R}^n)$, $b \in W^{1,\infty}([t_0, t_f], \mathbb{R}^n)$, $\mu \in BV([t_0, t_f])$, it holds

$$\begin{aligned}\int_{t_0}^{t_f} a(t)^\top b(t) d\mu(t) =& \int_{t_0}^{t_f} \sum_{i=1}^{n} a_i(t) b_i(t) d\mu(t) \\ =& \sum_{i=1}^{n} \int_{t_0}^{t_f} a_i(t) b_i(t) d\mu(t),\end{aligned}$$

and by Lemma 2.13, we deduce:

$$\int_{t_0}^{t_f} a(t)^\top b(t) d\mu(t) = \sum_{i=1}^{n} \int_{t_0}^{t_f} b_i(t) d\left(\int_{t_0}^{t} a_i(\tau) d\mu(\tau) \right).$$

Using integration by parts for the Stieltjes integral shows that:

$$\begin{aligned}\int_{t_0}^{t_f} a(t)^\top b(t) d\mu(t) =& \sum_{i=1}^{n} \int_{t_0}^{t_f} a_i(\tau) d\mu(\tau) \cdot b_i(t_f) - \sum_{i=1}^{n} \int_{t_0}^{t_f} \left(\int_{t_0}^{t} a_i(\tau) d\mu(\tau) \right) db_i(t) \\ =& \left(\int_{t_0}^{t_f} a(\tau)^\top d\mu(\tau) \right) \cdot b(t_f) - \int_{t_0}^{t_f} \left(\int_{t_0}^{t} a(\tau)^\top d\mu(\tau) \right) b'(t) dt.\end{aligned}$$

Now, inserting

$$\begin{aligned} a(t) &:= (s_{i_x}'[t]\Phi(t))^\top \\ \text{and} \quad b(t) &:= \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau \end{aligned}$$

into this formula yields:

$$\begin{aligned}\int_{t_0}^{t_f} s_{i_x}'[t]\Phi(t) \int_{t_0}^{t} \Phi^{-1}(\tau) h_1(\tau) d\tau d\mu_i(t) =& \left(\int_{t_0}^{t_f} s_{i_x}'[\tau]\Phi(\tau) d\mu_i(\tau) \right) \cdot \int_{t_0}^{t_f} \Phi^{-1}(\tau) h_1(\tau) d\tau \\ & - \int_{t_0}^{t_f} \left(\int_{t_0}^{t} s_{i_x}'[\tau]\Phi(\tau) d\mu_i(\tau) \right) \Phi^{-1}(t) h_1(t) dt \\ =& \int_{t_0}^{t_f} \int_{t}^{t_f} s_{i_x}'[\tau]\Phi(\tau) d\mu_i(\tau) \Phi^{-1}(t) h_1(t) dt.\end{aligned}$$

Next, both substitutions are inserted in (3.9):

$$\begin{aligned} 0 =& (l_0 \varphi_{x_f}' + \sigma^\top \Psi_{x_f}') \Phi(t_f) \int_{t_0}^{t_f} \Phi^{-1}(\tau) h_1(\tau) d\tau \\ &+ \int_{t_0}^{t_f} \int_{t}^{t_f} l_0 f_{0x}'[\tau]\Phi(\tau) d\tau \cdot \Phi^{-1}(t) h_1(t) dt \\ &+ \sum_{i=1}^{n_s} \int_{t_0}^{t_f} \int_{t}^{t_f} s_{i_x}'[\tau]\Phi(\tau) d\mu_i(\tau) \Phi^{-1}(t) h_1(t) dt \\ &+ \eta_1^* \left(c_x'[\cdot]\bar{x}(\cdot) \right) + \lambda_f^* \left(h_1 \right) \\ =& \int_{t_0}^{t_f} \Bigg[(l_0 \varphi_{x_f}' + \sigma^\top \Psi_{x_f}') \Phi(t_f) + \int_{t}^{t_f} l_0 f_{0x}'[\tau]\Phi(\tau) d\tau \end{aligned}$$

$$+\sum_{i=1}^{n_s}\int_t^{t_f} s_{i_x}'[\tau]\Phi(\tau)d\mu_i(\tau)\Big]\cdot\Phi^{-1}(t)h_1(t)dt$$
$$+\eta_1^*\left(c_x'[\cdot]\bar{x}(\cdot)\right)+\lambda_f^*\left(h_1\right)$$

This can be written equivalently as

$$\int_{t_0}^{t_f} p_f(t)^\top h_1(t)dt+\eta_1^*\left(c_x'[\cdot]\bar{x}(\cdot)\right)+\lambda_f^*\left(h_1\right)=0, \tag{3.10}$$

where

$$p_f(t)^\top:=\left((l_0\varphi_{x_f}'+\sigma^\top\Psi_{x_f}')\Phi(t_f)+\int_t^{t_f} l_0 f_{0x}'[\tau]\Phi(\tau)d\tau+\sum_{i=1}^{n_s}\int_t^{t_f} s_{i_x}'[\tau]\Phi(\tau)d\mu_i(\tau)\right]\cdot\Phi^{-1}(t).$$

Definition 3.8 (Pseudo Inverse)
The notation $(\cdot)^+$ is used for the pseudo inverse of a matrix, i.e. $(A)^+:=A^\top(AA^\top)^{-1}$ for $A\in\mathbb{R}^{m\times n}$, if $AA^\top$ is invertible.

Lemma 3.9 (Existence of the Pseudo Inverse)
Let $A:\mathbb{R}\to\mathbb{R}^{m\times n}$, $n\geq m$, and assume that there exists a constant $C\geq 0$, such that

$$\left\|A(t)^\top x\right\|\geq C\|x\|\qquad\forall t\in\mathbb{R}, x\in\mathbb{R}^m.$$

Then the pseudo inverse of $A(t)$ exists for all times t, and $\|(A(t))^+\|\leq C_2$ for some $C_2\in\mathbb{R}$.

Proof.
The fact that for any t it holds $\left\|A(t)^\top x\right\|^2=x^\top A(t)A(t)^\top x\geq C^2\|x\|^2$ shows that the minimal eigenvalue of $A(t)A(t)^\top$ (which is a symmetric matrix) is $\lambda_{\min}\geq C^2$, hence its inverse exists and is bounded. □

Corollary 3.10
Let an OCP in the form of Problem 3.1 be given, where all functions fulfill the smoothness Assumptions 3.2. Let $(\hat{x},\hat{u},\hat{v})\in X$ be a weak local minimum, let $l_0\in\mathbb{R}$, $\eta^*\in Y^*$ and $\lambda^*\in Z^*$ be multipliers that solve (3.1)–(3.4). Further assume that the pseudo inverse $(c_v'[t])^+$ exists and that $\|(c_v'[t])^+\|\leq C$ holds for some $C\in\mathbb{R}$ for almost all times $t\in[t_0,t_f]$.

Then there exist functions $\hat{p}_f\in BV([t_0,t_f],\mathbb{R}^{n_x})$ and $\eta\in L^\infty([t_0,t_f],\mathbb{R}^{n_c})$ with

$$\eta_1^*(k)=\int_{t_0}^{t_f}\eta(t)^\top k(t)dt\qquad\forall k\in L^\infty([t_0,t_f],\mathbb{R}^{n_c}) \tag{3.11}$$

$$\lambda_f^*(h_1)=-\int_{t_0}^{t_f}\hat{p}_f(t)^\top h_1(t)dt\qquad\forall h_1\in L^\infty([t_0,t_f],\mathbb{R}^{n_x}), \tag{3.12}$$

where η_1^* and λ_f^* satisfy (3.5) and (3.6).

Proof.
1. Equation (3.10) holds for every $h_1\in L^\infty([t_0,t_f],\mathbb{R}^{n_x})$ and $\bar{x}$ being the solution of the initial value problem (3.8). Inserting $h_1(t):=f_v'[t]v(t)$ into this equation yields

$$0=\lambda_f^*\left(f_v'[\cdot]v(\cdot)\right)+\int_{t_0}^{t_f} p_f(t)^\top f_v'[t]v(t)dt+\eta_1^*\left(c_x'[\cdot]\bar{x}(\cdot)\right).$$

Now insert the resulting expression for $\lambda_f^* \left(f_v'[\cdot]v(\cdot) \right)$ into (3.6):

$$0 = \int_{t_0}^{t_f} l_0 f_{0v}'[t]v(t)dt + \eta_1^*(c_v'[\cdot]v(\cdot) + c_x'[\cdot]\bar{x}(\cdot)) + \int_{t_0}^{t_f} p_f(t)^\top f_v'[t]v(t)dt. \quad (3.13)$$

Let $k \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})$ be arbitrary. The equation

$$k(t) = c_x'[t]\bar{x}(t) + c_v'[t]\bar{v}(t)$$

can be solved for $\bar{v}$ due to the assumption that $\|(c_v'[t])^+\| \leq C$: Let

$$\bar{v}(t) := (c_v'[t])^+(k(t) - c_x'[t]\bar{x}(t)).$$

Inserting $\bar{v}$ into (3.13) yields

$$0 = \int_{t_0}^{t_f} (l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+(k(t) - c_x'[t]\bar{x}(t))dt + \eta_1^*(k). \quad (3.14)$$

Finally, $\bar{x}$ can also be expressed as a function dependent on k,

$$\begin{aligned}
\dot{\bar{x}}(t) &= f_x'[t]\bar{x}(t) + f_v'[t]\bar{v}(t) \\
&= f_x'[t]\bar{x}(t) + f_v'[t](c_v'[t])^+(k(t) - c_x'[t]\bar{x}(t)) \\
&= (f_x'[t] - f_v'[t](c_v'[t])^+c_x'[t])\bar{x}(t) + f_v'[t](c_v'[t])^+k(t) \\
&= \hat{f}_x(t)\bar{x}(t) + \hat{f}_k(t)k(t),
\end{aligned}$$

where $\hat{f}_x(t) := f_x'[t] - f_v'[t](c_v'[t])^+c_x'[t]$ and $\hat{f}_k(t) := f_v'[t](c_v'[t])^+$.

The solution $\bar{x}$ of the initial value problem can be expressed depending on k as

$$\begin{aligned}
\bar{x}(t) &= \hat{\Phi}(t) \int_{t_0}^{t} \hat{\Phi}^{-1}(\tau)\hat{f}_k(\tau)k(\tau)d\tau \\
\text{with} \qquad \hat{\Phi}(0) &= I, \qquad \hat{\Phi}'(t) = \hat{f}_x(t)\hat{\Phi}(t).
\end{aligned}$$

Inserting $\bar{x}$ into (3.14) yields:

$$\begin{aligned}
0 = &\int_{t_0}^{t_f} (l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+ \left(k(t) - c_x'[t]\hat{\Phi}(t) \int_{t_0}^{t} \hat{\Phi}^{-1}(\tau)\hat{f}_k(\tau)k(\tau)d\tau \right) dt \\
&+ \eta_1^*(k) \\
= &\int_{t_0}^{t_f} (l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+k(t)dt \\
&- \int_{t_0}^{t_f} (l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+c_x'[t]\hat{\Phi}(t) \left(\int_{t_0}^{t} \hat{\Phi}^{-1}(\tau)\hat{f}_k(\tau)k(\tau)d\tau \right) dt \\
&+ \eta_1^*(k),
\end{aligned}$$

integration by parts shows that:

$$0 = \int_{t_0}^{t_f} (l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+k(t)dt$$

$$
\begin{aligned}
&- \int_{t_0}^{t_f} \left(\int_t^{t_f} (l_0 f_{0v}'[\tau] + p_f(\tau)^\top f_v'[\tau])(c_v'[\tau])^+ c_x'[\tau]\hat{\Phi}(\tau)d\tau \right) \hat{\Phi}^{-1}(t)\hat{f}_k(t)k(t)dt \\
&+ \eta_1^*(k) \\
=& \int_{t_0}^{t_f} \Bigg[(l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+ \\
&- \left(\int_t^{t_f} (l_0 f_{0v}'[\tau] + p_f(\tau)^\top f_v'[\tau])(c_v'[\tau])^+ c_x'[\tau]\hat{\Phi}(\tau)d\tau \right) \hat{\Phi}^{-1}(t)\hat{f}_k(t) \Bigg] k(t)dt \\
&+ \eta_1^*(k).
\end{aligned}
$$

Setting

$$
\begin{aligned}
\eta(t)^\top :=& - (l_0 f_{0v}'[t] + p_f(t)^\top f_v'[t])(c_v'[t])^+ \\
&+ \left(\int_t^{t_f} (l_0 f_{0v}'[\tau] + p_f(\tau)^\top f_v'[\tau])(c_v'[\tau])^+ c_x'[\tau]\hat{\Phi}(\tau)d\tau \right) \hat{\Phi}^{-1}(t)\hat{f}_k(t),
\end{aligned}
$$

the multiplier η_1^* possesses the representation

$$
\eta_1^*(k) = \int_{t_0}^{t_f} \eta(t)^\top k(t)dt
$$

for all $k \in L^\infty([t_0, t_f], \mathbb{R}^{n_c}$, this is equation (3.11).

2. With the first assertion, 3.10 becomes:

$$
\int_{t_0}^{t_f} p_f(t)^\top h_1(t)dt + \int_{t_0}^{t_f} \eta(t)^\top c_x'[t]\bar{x}(t)dt + \lambda_f^*(h_1) = 0.
$$

Again, this equation holds for general h_1 and $\bar{x}$ satisfying

$$
\dot{x}(t) = f_x'[t]x(t) + h_1(t),\ x(t_0) = 0.
$$

The solution $\bar{x}$ of this initial value problem is inserted:

$$
\begin{aligned}
0 =& \int_{t_0}^{t_f} p_f(t)^\top h_1(t)dt + \int_{t_0}^{t_f} \eta(t)^\top c_x'[t]\Phi(t) \int_{t_0}^{t} \Phi^{-1}(\tau)h_1(\tau)d\tau dt + \lambda_f^*(h_1), \\
=& \int_{t_0}^{t_f} p_f(t)^\top h_1(t)dt + \int_{t_0}^{t_f} \int_t^{t_f} \eta(\tau)^\top c_x'[\tau]\Phi(\tau)d\tau \cdot \Phi^{-1}(t)h_1(t)dt + \lambda_f^*(h_1) \\
=& \int_{t_0}^{t_f} \left(p_f(t)^\top + \int_t^{t_f} \eta(\tau)^\top c_x'[\tau]\Phi(\tau)d\tau \cdot \Phi^{-1}(t) \right) h_1(t)dt + \lambda_f^*(h_1)
\end{aligned}
$$

This shows that

$$
\hat{p}_f(t)^\top := p_f(t)^\top + \int_t^{t_f} \eta(\tau)^\top c_x'[\tau]\Phi(\tau)d\tau \cdot \Phi^{-1}(t)
$$

satisfies equation (3.12) for arbitrary $h_1 \in L^\infty([t_0, t_f], \mathbb{R}^{n_x})$, and since p_f is of bounded variation, so is $\hat{p}_f$. □

3.3. Minimum principle for OCP

The smoothness results of Corollary 3.10, together with some variational deliberations, lead to Theorem 3.15. The following lemma cited from [Ger06, section 2.8] are needed as background for the proof:

Lemma 3.11
Let $f, g \in L^\infty([t_0, t_f], \mathbb{R})$, $s \in \mathcal{C}([t_0, t_f], \mathbb{R})$ and $\mu \in BV([t_0, t_f], \mathbb{R})$. If

$$\int_{t_0}^{t_f} f(t)h(t) + g(t)\dot{h}(t)dt + \int_{t_0}^{t_f} s(t)h(t)d\mu(t) = 0$$

for every $h \in W^{1,\infty}([t_0, t_f], \mathbb{R})$ with $h(t_0) = h(t_f) = 0$, then there exists a function $\hat{g} \in BV([t_0, t_f], \mathbb{R})$, such that $\hat{g}(t) = g(t)$ a.e. in $[t_0, t_f]$ and $\hat{g}(t) = \int_{t_0}^{t} f(\tau)d\tau + \int_{t_0}^{t} s(\tau)d\mu(\tau)$.

The proof of the next lemma makes use of Lusin's Theorem. This can be found in [Alt, p. 210]:

Theorem 3.12 (Lusin's theorem)
Let μ be a regular σ-additive measure $S \to \mathbb{R}^+$ over a linear space S, and Y a Banach space.

Every μ- measurable function $f : S \to Y$ is μ- almost continuous, i.e. for every μ-measurable set E and every $\varepsilon > 0$ there is a compact set $K \subset E$ with $\mu(E \setminus K) \leq \varepsilon$ such that $f \restriction_K$ is continuous (on K).

Lemma 3.13
Let $[a, b]$ be an interval and $f \in L^\infty([a, b], \mathbb{R}^n)$. Let $\hat{g}$ be a function $\hat{g} \in U \subset L^\infty([a, b], \mathbb{R}^n)$.

If $\int_a^b f(t)^\top (g(t) - \hat{g}(t))dt \geq 0$ for all $g \in U$, then $f(t)^\top (g(t) - \hat{g}(t)) \geq 0$ a.e. on $[a, b]$.

The proof essentially follows [Lem72, p. 66].

Proof.
Let μ be the Lebesgue measure, and $A_1 \subset [a, b]$, such that $\mu(A_1) > 0$ but $f(t)^\top (g(t) - \hat{g}(t)) < 0$ a.e. on A_1 with $g \in U$. According to Lusin's Theorem, there are continuous functions $h_f, h_g : [a, b] \to \mathbb{R}^n$ that are equal to f and $g - \hat{g}$, respectively, on A_1 except for small subsets of $[a, b]$, i.e.

$$\begin{aligned} h_f(t) &= f(t) && \text{a.e. on } [a, b] \setminus B_f && \text{for some } B_f \text{ with } \mu(B_f) < \mu(A_1)/2 \\ h_g(t) &= g(t) - \hat{g}(t) && \text{a.e. on } [a, b] \setminus B_g && \text{for some } B_g \text{ with } \mu(B_g) < \mu(A_1)/2 \end{aligned}$$

Now since $\mu(A_1 \setminus (B_f \cup B_g)) \geq \mu(A_1) - \mu(B_f) - \mu(B_g) > 0$, there must be some t_0 with $f(t_0)^\top (g(t_0) - \hat{g}(t_0)) =: -\varepsilon < 0$, so that the intersection of a neighborhood of t_0 with the set $A_1 \setminus (B_f \cup B_g)$ is not a set of measure zero[1]:

$$\mu((A_1 \setminus (B_f \cup B_g)) \cap [t_0 - \delta, t_0 + \delta]) > 0 \qquad \forall \delta > 0$$

[1] Assume that this is not true. Then for each $t_0 \in [a, b]$, there is an $\varepsilon > 0$, such that $\mu((A \setminus (B_f \cup B_g)) \cap [t_0 - \varepsilon, t_0 + \varepsilon]) = 0$. The family $\{(A \setminus (B_f \cup B_g)) \cap [t_0 - \varepsilon(t_0), t_0 + \varepsilon(t_0)] : t_0 \in \mathbb{Q}\}$ is a countable cover of $A \setminus (B_f \cup B_g)$. Since μ is σ-additive, one can estimate $\mu(A \setminus (B_f \cup B_g)) \leq \sum_{t_0 \in \mathbb{Q}} \mu((A \setminus (B_f \cup B_g)) \cap [t_0 - \varepsilon(t_0), t_0 + \varepsilon(t_0)]) = 0$.

Since $h_f^\top h_g$ is continuous, there is a $\hat{\delta} > 0$, such that $h_f(t)^\top h_g(t) < -\varepsilon/2$ for all $t \in [t_0 - \hat{\delta}, t_0 + \hat{\delta}]$. Now we define

$$A_2 := (A_1 \setminus (B_f \cup B_g) \cap [t_0 - \hat{\delta}, t_0 + \hat{\delta}])$$

Consider the function $\bar{g} \in U$, defined by

$$\bar{g}(t) := \begin{cases} g(t) & t \in A_2 \\ \hat{g}(t) & \text{otherwise} \end{cases}$$

The integral over $f^\top(\bar{g} - \hat{g})$ is negative:

$$\begin{aligned} \int_a^b f(t)^\top(\bar{g}(t) - \hat{g}(t))dt &= \int_{A_2} f(t)^\top(g(t) - \hat{g}(t))dt \\ &\leq -\varepsilon/2 \cdot \mu(A_2) \\ &< 0 \end{aligned}$$

□

The (augmented) Hamilton function is introduced which allows characterization of the minimum principle in a plain form:

Definition 3.14 (Hamilton function)

1. *The* Hamilton function $\mathcal{H} : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_x} \times \mathbb{R} \to \mathbb{R}$ *for a given OCP is defined as*

$$\mathcal{H}(t, x, u, v, \lambda, l_0) := l_0 f_0(t, x, u, v) + \lambda^\top f(t, x, u, v).$$

2. *The* augmented Hamilton function $\hat{\mathcal{H}} : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_c} \times \mathbb{R} \to \mathbb{R}$ *is defined as*

$$\hat{\mathcal{H}}(t, x, u, v, \lambda, \eta, l_0) := \mathcal{H}(t, x, u, v, \lambda, l_0) + \eta^\top c(t, x, v).$$

Theorem 3.15 (Minimum Principle for OCP)
Consider the OCP 3.1 where the problem defining functions φ, f_0, f, Ψ, c, s and U_{ad} fulfill the smoothness Assumption 3.2. Let $(\hat{x}, \hat{u}, \hat{v})$ be a weak local minimum of OCP. Assume that the pseudo inverse $(c_v'[t])^+$ exists and that there exists a constant $C \in \mathbb{R}$, such that

$$\|(c_v'[t])^+\| \leq C \quad \textit{for all } t \in [t_0, t_f].$$

Then there exist multipliers

$$l_0 \in \mathbb{R}, \lambda \in BV([t_0, t_f], \mathbb{R}^{n_x}), \eta \in L^\infty([t_0, t_f], \mathbb{R}^{n_c}), \mu \in NBV([t_0, t_f], \mathbb{R}^{n_s}) \textit{ and } \sigma \in \mathbb{R}^{n_\Psi}$$

that satisfy the following conditions:

1. *Nontriviality:*

$$l_0 \geq 0, \quad (l_0, \lambda, \eta, \mu, \sigma) \neq 0 \tag{3.15}$$

2. Adjoint equation:

$$\begin{aligned}\lambda(t) &= \lambda(t_f) + \int_t^{t_f} \hat{\mathcal{H}}'_x(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{v}(\tau), \lambda(\tau), \eta(\tau), l_0)^\top d\tau \\ &\quad + \sum_{i=1}^{n_s} \int_t^{t_f} s_{i_x}'(\tau, \hat{x}(\tau))^\top d\mu_i(\tau) \qquad t \in [t_0, t_f] \\ &= \lambda(t_0) - \int_{t_0}^{t} \hat{\mathcal{H}}'_x(\tau, \hat{x}(\tau), \hat{u}(\tau), \hat{v}(\tau), \lambda(\tau), \eta(\tau), l_0)^\top d\tau \\ &\quad - \sum_{i=1}^{n_s} \int_{t_0}^{t} s_{i_x}'(\tau, \hat{x}(\tau))^\top d\mu_i(\tau) \qquad t \in [t_0, t_f]\end{aligned} \tag{3.16}$$

3. Transversality conditions:

$$\lambda(t_0)^\top = -\left(l_0 \varphi'_{x_0}(\hat{x}(t_0), \hat{x}(t_f)) + \sigma^\top \Psi'_{x_0}(\hat{x}(t_0), \hat{x}(t_f))\right) \tag{3.17}$$

$$\lambda(t_f)^\top = \quad l_0 \varphi'_{x_f}(\hat{x}(t_0), \hat{x}(t_f)) + \sigma^\top \Psi'_{x_f}(\hat{x}(t_0), \hat{x}(t_f)) \tag{3.18}$$

4. Optimality conditions:

$$\hat{\mathcal{H}}'_v(t, \hat{x}(t), \hat{u}(t), \hat{v}(t), \lambda(t), l_0) = 0 \qquad \text{a.e. in } [t_0, t_f] \tag{3.19}$$

$$\mathcal{H}'_u(t, \hat{x}(t), \hat{u}(t), \hat{v}(t), \lambda(t), l_0)(u - \hat{u}(t)) \geq 0 \qquad \forall u \in U(t), \text{ a.e. in } [t_0, t_f] \tag{3.20}$$

5. Complementarity conditions:

$$\eta(t)^\top c(t, \hat{x}(t), \hat{v}(t)) = 0, \qquad \eta(t) \geq 0 \qquad \text{a.e. in } [t_0, t_f]. \tag{3.21}$$

$$\sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_i(t, \hat{x}(t)) d\mu_i(t) = 0, \tag{3.22}$$

and

$$\sum_{i=1}^{n_s} \int_{t_0}^{t_f} z_i(t) d\mu_i(t) \geq 0 \qquad \text{for all } z \in \mathcal{C}([t_0, t_f], \mathbb{R}^{n_s}) \text{ with } z(t) \geq 0 \text{ for } t \in [t_0, t_f].$$

Proof.

- By Corollary 3.10, there exist multipliers $\lambda \in BV([t_0, t_f]$ and $\eta \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})$, such that for all $h_1 \in L^\infty([t_0, t_f], \mathbb{R}^{n_x})$ and all $k \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})$ it holds

$$\lambda_f^*(h_1) = -\int_{t_0}^{t_f} \hat{p}_f(t)^\top h_1(t) dt$$
$$\eta_1^*(k) = \int_{t_0}^{t_f} \eta(t)^\top k(t) dt.$$

  In view of this representation of the multipliers λ_f^* and η_1^*, equation (3.5) implies that for all $x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$:

$$\begin{aligned}0 = &\left(l_0 \varphi'_{x_0}(\hat{x}(t_0), \hat{x}(t_f)) + \sigma^\top \Psi'_{x_0}(\hat{x}(t_0), \hat{x}(t_f))\right) x(t_0) \\ &+ \left(l_0 \varphi'_{x_f}(\hat{x}(t_0), \hat{x}(t_f)) + \sigma^\top \Psi'_{x_f}(\hat{x}(t_0), \hat{x}(t_f))\right) x(t_f)\end{aligned}$$

$$
\begin{aligned}
&+ \int_{t_0}^{t_f} l_0 f_{0x}'[t]x(t)dt + \sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_{ix}'[t]x(t)d\mu_i(t) \\
&+ \int_{t_0}^{t_f} \eta(t)^\top c_x'[t]x(t)dt + \int_{t_0}^{t_f} \lambda(t)^\top (f_x'[t]x(t) - \dot{x}(t))dt \\
&= (l_0\varphi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)))x(t_0) \\
&\quad + (l_0\varphi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)))x(t_f) \\
&\quad + \int_{t_0}^{t_f} \hat{\mathcal{H}}_x'[t]x(t)dt + \sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_{ix}'[t]x(t)d\mu_i(t) - \int_{t_0}^{t_f} \lambda(t)^\top \dot{x}(t)dt
\end{aligned}
$$

The last term is

$$
\begin{aligned}
-\int_{t_0}^{t_f} \lambda(t)^\top \dot{x}(t)dt &= -\int_{t_0}^{t_f} \lambda(t)^\top dx(t) \\
&= \int_{t_0}^{t_f} x(t)^\top d\lambda(t) - \left[\lambda(t)^\top x(t)\right]_{t_0}^{t_f} \\
&= \int_{t_0}^{t_f} x(t)^\top d\lambda(t) - \lambda(t_f)^\top x(t_f) + \lambda(t_0)^\top x(t_0).
\end{aligned}
$$

Thus

$$
\begin{aligned}
0 &= (l_0\varphi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)) + \lambda(t_0)^\top)x(t_0) \\
&\quad + (l_0\varphi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)) - \lambda(t_f)^\top)x(t_f) \\
&\quad + \int_{t_0}^{t_f} \hat{\mathcal{H}}_x'[t]x(t)dt + \sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_{ix}'[t]x(t)d\mu_i(t) + \int_{t_0}^{t_f} x(t)^\top d\lambda(t) \\
&= (l_0\varphi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_0}'(\hat{x}(t_0),\hat{x}(t_f)) + \lambda(t_0)^\top)x(t_0) \\
&\quad + (l_0\varphi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)) + \sigma^\top \Psi_{x_f}'(\hat{x}(t_0),\hat{x}(t_f)) - \lambda(t_f)^\top)x(t_f) \\
&\quad + \int_{t_0}^{t_f} x(t)^\top d\left(\lambda(t) - \int_t^{t_f} \hat{\mathcal{H}}_x'[\tau]^\top d\tau - \sum_{i=1}^{n_s} \int_t^{t_f} s_{ix}'[\tau]^\top d\mu_i(\tau)\right)
\end{aligned}
$$

for all $x \in W^{1,\infty}([t_0,t_f],\mathbb{R}^{n_x})$. From the fact that $x(t_0)$ and $x(t_f)$ can be chosen arbitrarily, the transversality conditions (3.17) and (3.18) follow. Lemma 3.11 implies that there is a vector C, such that

$$
C = \lambda(t_f) - \int_t^{t_f} \hat{\mathcal{H}}_x'[\tau]^\top d\tau - \sum_{i=1}^{n_s} \int_t^{t_f} s_{ix}'[\tau]^\top d\mu_i(\tau),
$$

which proves (3.16), since $t = t_f$ shows that $C = \lambda(t_f)$.

- In the next step, (3.6) and (3.7) are reformulated using the smooth multipliers λ and η_1. This yields

$$
\int_{t_0}^{t_f} \hat{\mathcal{H}}_v'[t]v(t)dt = 0 \qquad \forall v \in L^\infty([t_0,t_f],\mathbb{R}^{n_v})
$$

and

$$
\int_{t_0}^{t_f} \hat{\mathcal{H}}_u'[t](u(t) - \hat{u}(t))dt \geq 0 \qquad \forall u \in U_{ad}.
$$

The first equation can be written equivalently as two inequalities, so that both inequalities have to hold pointwise by Lemma 3.13.[2] Summarizing both inequalities then yields

$$\hat{\mathcal{H}}_v'[t]v(t) = 0 \qquad \text{a.e. in } [t_0, t_f],\ \forall v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$$

and hence

$$\hat{\mathcal{H}}_v'[t] = 0 \qquad \text{a.e. in } [t_0, t_f],$$

which proves the optimality condition for v, (3.19). The second inequality together with Lemma 3.13 immediately shows (3.20).

- Since $\eta^* \in K$, it follows

$$\int_{t_0}^{t_f} \eta(t)^\top z(t) dt \geq 0$$

for all $z \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})$ with $z(t) \geq 0$ a.e. in $[t_0, t_f]$. Hence $\eta(t) \geq 0$ a.e. in $[t_0, t_f]$.

Also, $\eta^*(G(\hat{x}, \hat{u}, \hat{v})) = 0$, so $\int_{t_0}^{t_f} \eta(t)^\top c(t, \hat{x}(t), \hat{v}(t)) dt = 0$. By Lemma 3.13, this yields (3.21).

Similarly, it holds that $\eta_2 \in K_2$, so

$$\sum_{i=1}^{n_s} \int_{t_0}^{t_f} z_i(t) d\mu_i(t) \geq 0$$

for all $z \in \mathcal{C}([t_0, t_f], \mathbb{R}^{n_s})$ with $z(t) \geq 0$ on $[t_0, t_f]$. This implies that

$$\eta_2^*\left(s(\cdot, \hat{x}(\cdot))\right) = \sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_i(t, \hat{x}(t)) d\mu_i(t) = 0,$$

which completes the proof. □

The following lemmata (cf. [Ger06, Lemma 2.8.5 and Lemma 2.8.6]) are stated for the sake of completeness. In [Ger06], the complementarity conditions for problems with state constraints are stated in a different form.

Lemma 3.16
Let $\mu \in BV([t_0, t_f], \mathbb{R})$. If

$$\int_{t_0}^{t_f} f(t) d\mu(t) \geq 0$$

holds for every non-negative function $f \in \mathcal{C}([t_0, t_f], \mathbb{R})$, then μ is non-decreasing in $[t_0, t_f]$.

Hence, the complementarity condition for the multiplier μ stated here implies the respective complementarity condition in [Ger06].

[2] The equation $\int_{t_0}^{t_f} \hat{\mathcal{H}}_v'[t]v(t)dt = 0 \quad \forall v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$ implies that $\int_{t_0}^{t_f} \hat{\mathcal{H}}_v'[t]v(t)dt \geq 0$ holds for all $v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$. Lemma 3.13 yields $\hat{\mathcal{H}}_v'[t]v(t) \geq 0$ a.e. on $[t_0, t_f]$ $\forall v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$. The same reasoning yields that $\hat{\mathcal{H}}_v'[t]v(t) \leq 0$ a.e. on $[t_0, t_f]$ $\forall v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$.

Lemma 3.17
Let $\mu \in BV([t_0, t_f], \mathbb{R})$ be non-decreasing and $f \in \mathcal{C}([t_0, t_f], \mathbb{R})$ non-positive. If

$$\int_{t_0}^{t_f} f(t) d\mu(t) = 0$$

holds, then μ is constant on every interval $[a, b] \subset [t_0, t_f]$ with $a < b$ and $f(t) < 0$ in $[a, b]$.

Hence, the complementarity conditions (3.22) imply that for every $i = 1, \dots, n_s$, the multiplier is constant on every interval $[a, b] \subseteq [t_0, t_f]$ on which $s_i[t] < 0$ holds.

It can be shown that this statement is equivalent to the formulation used here:

Lemma 3.18
Let $\mu : [t_0, t_f] \to \mathbb{R}$ and $f : [t_0, t_f] \to \mathbb{R}$ be any two functions for which the Stieltjes integral $\int_{t_0}^{t_f} f(t) d\mu(t)$ exists. If μ is constant on every interval $[a, b] \subset [t_0, t_f]$ with $a < b$ and $f(t) \neq 0$ in $[a, b]$, then it holds

$$\int_{t_0}^{t_f} f(t) d\mu(t) = 0.$$

Proof.
Let $\mathbb{G} = (t_i)_{i=0,\dots,n+1}$ be any subdivision of the interval $[t_0, t_f]$. Let $j_0, \dots, j_m$ denote the indices for which a zero of f exists in $[t_{j_i}, t_{j_i+1}]$ for $i = 0, \dots, m-1$, and define ξ_{j_i} as such a zero. Let $j_{m+1}, \dots, j_n$ denote the indices for which no zero of f exists in $[t_{j_i}, t_{j_i+1}]$, $i = m+1, \dots, n$. On these intervals, μ is constant according to the assumptions. Let ξ_{j_i} be arbitrary for $i = m+1, \dots, n$. Then

$$\begin{aligned}
\sum_{i=0}^{n} f(\xi_i)[\mu(t_{i+1}) - \mu(t_i)] &= \sum_{i=0}^{n} f(\xi_{j_i})[\mu(t_{j_i+1}) - \mu(t_{j_i})] \\
&= \sum_{i=0}^{m} \underbrace{f(\xi_{j_i})}_{=0}[\mu(t_{j_i+1}) - \mu(t_{j_i})] + \sum_{i=m+1}^{n} f(\xi_{j_i}) \underbrace{[\mu(t_{j_i+1}) - \mu(t_{j_i})]}_{=0} \\
&= 0
\end{aligned}$$

This argument holds for any subdivision $\mathbb{G}$ and a particular choice of ξ. Hence, for this choice of ξ, the limit $\lim_{\delta(\mathbb{G})\to 0}$ equals zero. Since the limit exists and is equal for any choice of $\mathbb{G}$ and ξ, the Stieltjes integral itself equals zero. □

3.3.1. Weaker assumptions for the control state constraints

One issue of Theorem 3.15 is the strong assumption that the pseudo inverse $(c_v'[t])^+$ exists with

$$\|(c_v'[t])^+\| \leq C \quad \text{for all } t \in [t_0, t_f]$$

for some constant C. This assumption is violated e.g. if box constraints of the form $v_{\min} \leq v \leq v_{\max}$ are included as mixed control state constraints.

However, a thought experiment shows that this assumption can be weakened: Note that the local minimum under consideration does not change if the mixed control state

constraints that are "plainly inactive" are altered. The same holds for the multipliers. Consequently, it should be sufficient to postulate assumptions only for the α-active parts of the constraints.

In this section, this idea shall be made concrete following ideas introduced in [Mal03].

Definition 3.19 (α-active constraints)
Let $(\hat{x}, \hat{u}, \hat{v})$ be a local minimum of Problem 3.1. For $\alpha \geq 0$ and $t \in [t_0, t_f]$, define

$$
\begin{aligned}
I_\alpha(t) &:= \{i \in 1, \dots, n_c | c_i(t, \hat{x}(t), \hat{v}(t)) \geq -\alpha\} \\
c_\alpha^i(t) &:= \begin{cases} 0 & \text{if } c_i(t, \hat{x}(t), \hat{v}(t)) \geq -\alpha \\ -1 & \text{if } c_i(t, \hat{x}(t), \hat{v}(t)) < -\alpha \end{cases}, \qquad i \in \{1, \dots, n_c\} \\
S_\alpha(t) &:= \operatorname{diag}\left(c_\alpha^i(t)\right)_{i=1}^{n_c}
\end{aligned}
$$

Now, for a local minimizer $(\hat{x}, \hat{u}, \hat{v})$ and $\alpha \geq 0$, define the auxiliary problem $(\tilde{P}_\alpha)$:

Problem 3.20 ($\tilde{P}_\alpha(\hat{x}, \hat{u}, \hat{v})$)

$$
\text{min!} \quad J_\alpha(x, u, v, \pi) := \varphi(x(t_0), x(t_f)) + \int_{t_0}^{t_f} f_0(t, x(t), u(t), v(t))dt + \frac{1}{2}\|\pi\|_2^2
$$

$$
\begin{aligned}
\textit{with respect to the state function} \quad & x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x}) \\
\textit{and the control functions} \quad & u \in L^\infty([t_0, t_f], \mathbb{R}^{n_u}) \\
\textit{and} \quad & v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v}) \\
\textit{and slack variables} \quad & \pi \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})
\end{aligned}
$$

subject to the differential equation

$$
\dot{x}(t) = f(t, x(t), u(t), v(t)) \quad \textit{a.e. in } [t_0, t_f],
$$

boundary conditions

$$
\Psi(x(t_0), x(t_f)) = 0,
$$

relaxed mixed control state constraints

$$
c(t, x(t), v(t)) + S_\alpha(t)\pi(t) \leq 0,
$$

pure state constraints

$$
s(t, x(t)) \leq 0
$$

and set constraints for u

$$
u(t) \in U(t) \subset \mathbb{R}^{n_u} \quad \textit{a.e. in } [t_0, t_f]
$$

Lemma 3.21
Let $c_1, c_2 > 0$. Then there exists some constant $c_3 > 0$, such that for all $a, b \geq 0$ it holds:

$$
\max\{c_1 a - c_2 b, b\} \geq c_3 \cdot \max\{a, b\}.
$$

Proof.

- $b \geq a$: In this case, $\max\{c_1 a - c_2 b, b\} \geq b = 1 \cdot \max\{a, b\}$.
- $a \geq b \wedge \max\{c_1 a - c_2 b, b\} = c_1 a - c_2 b$: It holds

$$c_1 a - c_2 b \geq b \;\Rightarrow\; b \leq a \frac{c_1}{1 + c_2}.$$

  Hence

$$\begin{aligned} c_1 a - c_2 b &\geq c_1 a - c_2 \frac{c_1}{1 + c_2} a \\ &= a \left(c_1 \left(1 - \frac{c_2}{1 + c_2} \right) \right) \\ &= c_1 \left(1 - \frac{c_2}{1 + c_2} \right) \cdot \max\{a, b\}. \end{aligned}$$

- $a \geq b \wedge \max\{c_1 a - c_2 b, b\} = b$: Then, $b \geq \frac{c_1}{1+c_2} a$, hence

$$\max\{c_1 a - c_2 b, b\} \geq \frac{c_1}{1 + c_2} \cdot \max\{a, b\}.$$

Thus, $c_3 := \min\left\{1, c_1 \left(1 - \frac{c_2}{1+c_2}\right), \frac{c_1}{1+c_2}\right\}$ satisfies the inequality. □

Remark 3.22 (Smoothness)
Note that Problem $\tilde{P}_\alpha(\hat{x}, \hat{u}, \hat{v})$ can be written in the form of Problem 3.1, but the smoothness Assumption 3.2.5 will be violated. However, the operator G that models the inequality constraints is still continuously differentiable according to Example 2.19.4.

If $\|c'_{v I_\alpha(t)}[t]^\top \xi\| \geq C \cdot \|\xi\|$ for ξ of appropriate dimension (i.e. $\#I_\alpha(t)$) and some $C > 0$, independent of t, then the linearization of the mixed control state constraints fulfills

$$\begin{aligned} \left\| \begin{pmatrix} c'_v[t]^\top \\ S_\alpha(t)^\top \end{pmatrix} \xi \right\| &= \left\| \begin{pmatrix} c'_{v I_\alpha(t)}[t]^\top \xi_{I_\alpha(t)} + c'_{v I_\alpha^c(t)}[t]^\top \xi_{I_\alpha^c(t)} \\ -\xi_{I_\alpha^c(t)} \end{pmatrix} \right\| \\ &= \max\left\{ \left\| c'_{v I_\alpha(t)}[t]^\top \xi_{I_\alpha(t)} + c'_{v I_\alpha^c(t)}[t]^\top \xi_{I_\alpha^c(t)} \right\|, \left\| \xi_{I_\alpha^c(t)} \right\| \right\} \\ &\geq \max\left\{ C \left\| \xi_{I_\alpha(t)} \right\| - C_2 \left\| \xi_{I_\alpha^c(t)} \right\|, \left\| \xi_{I_\alpha^c(t)} \right\| \right\} \\ &\geq C_3 \|\xi\| \end{aligned}$$

for some constant $C_3 > 0$. The last step follows from Lemma 3.21. Hence, the precondition of Lemma 3.9 is satisfied; the pseudo inverse is bounded.

Therefore, in a local minimum of Problem 3.20, all assertions of Theorem 3.15 hold.

The following lemma (cf. [Mal03, Lemma 3.3]) states the relation between the original Problem 3.1 and the auxiliarity Problem 3.20:

Lemma 3.23
Let all functions of Problem 3.1 satisfy the smoothness Assumptions 3.2.

3.23.1 *Let $(\hat{x}, \hat{u}, \hat{v})$ be a local minimizer of Problem 3.1. Then $(\hat{x}, \hat{u}, \hat{v}, 0)$ is a local minimizer of Problem 3.20.*

3.23.2 *The elements $(l_0, \lambda, \mu, \sigma)$ are multipliers for the local minimizer $(\hat{x}, \hat{u}, \hat{v})$ of Problem 3.1 satisfying equations* (3.15)-(3.22) *if and only if they are multipliers for the local minimizer $(\hat{x}, \hat{u}, \hat{v}, 0)$ of Problem 3.20.*

Proof.

3.23.1 We show that if a point (x, u, v, π) in a sufficiently small neighborhood of $(\hat{x}, \hat{u}, \hat{v}, 0)$ is feasible for Problem 3.20, then (x, u, v) is feasible for Problem 3.1:

The point $(\hat{x}, \hat{u}, \hat{v}, 0)$ is feasible for Problem 3.20.

Choose $\delta > 0$ so small that

$$\|c(\cdot, x(\cdot), v(\cdot)) - c(\cdot, \hat{x}(\cdot), \hat{v}(\cdot))\|_\infty \leq \alpha$$

for all (x, u, v) with $\|(x, u, v) - (\hat{x}, \hat{u}, \hat{v})\|_X \leq \delta$. This is possible since c is continuous with respect to x, v and t.

Then if (x, u, v, π) is feasible for Problem 3.20, it holds for $i \in I_\alpha(t)$ that

$$0 \geq c_i(t, x(t), v(t)) + (S_\alpha(t)\pi(t))_i = c_i(t, x(t), v(t)) + c^i_\alpha(t)\pi_i(t) = c_i(t, x(t), v(t)).$$

For $i \notin I_\alpha(t)$, $c_i(t, \hat{x}(t), \hat{v}(t)) < -\alpha$, hence $c_i(t, x(t), v(t)) \leq 0$ according to the choice of δ.

Hence, if (x, u, v, π) lies in this neighborhood of $(\hat{x}, \hat{u}, \hat{v}, 0)$, then (x, u, v) is feasible for Problem 3.1.

Suppose that $(\hat{x}, \hat{u}, \hat{v})$ is a local minimum of Problem 3.1, but $(\hat{x}, \hat{u}, \hat{v}, 0)$ is not a local minimum of Problem 3.20. Then for any neighborhood of $(\hat{x}, \hat{u}, \hat{v}, 0)$, there exists a point $(\tilde{x}, \tilde{u}, \tilde{v}, \tilde{\pi})$, such that $J_\alpha(\tilde{x}, \tilde{u}, \tilde{v}, \tilde{\pi}) < J_\alpha(\hat{x}, \hat{u}, \hat{v}, 0)$. If the neighborhood is small enough, then $(\tilde{x}, \tilde{u}, \tilde{v})$ is feasible for Problem 3.1, and it holds

$$\begin{aligned}
J(\tilde{x}, \tilde{u}, \tilde{v}) &= \varphi(\tilde{x}(t_0), \tilde{x}(t_f)) + \int_{t_0}^{t_f} f_0(t, \tilde{x}(t), \tilde{u}(t), \tilde{v}(t))dt \\
&\leq \varphi(\tilde{x}(t_0), \tilde{x}(t_f)) + \int_{t_0}^{t_f} f_0(t, \tilde{x}(t), \tilde{u}(t), \tilde{v}(t))dt + \frac{1}{2}\|\tilde{\pi}\|_2^2 \\
&= J_\alpha(\tilde{x}, \tilde{u}, \tilde{v}, \tilde{\pi}) \\
&< J_\alpha(\hat{x}, \hat{u}, \hat{v}, 0) \\
&= \varphi(\hat{x}(t_0), \hat{x}(t_f)) + \int_{t_0}^{t_f} f_0(t, \hat{x}(t), \hat{u}(t), \hat{v}(t))dt \\
&= J(\hat{x}, \hat{u}, \hat{v}).
\end{aligned}$$

Hence $J(\tilde{x}, \tilde{u}, \tilde{v}) < J(\hat{x}, \hat{u}, \hat{v})$ and $(\tilde{x}, \tilde{u}, \tilde{v})$ is feasible for 3.1, which contradicts the assumption that $(\hat{x}, \hat{u}, \hat{v})$ is a local minimum.

3.23.2 Let $(l_0, \lambda, \mu, \sigma)$ be multipliers for a local minimum $(\hat{x}, \hat{u}, \hat{v}, 0)$ of Problem 3.20. This means that $(l_0, \lambda, \mu, \sigma) \neq 0$ with $l_0 \geq 0$, $\eta(t) \geq 0$ and μ_i monotonically increasing on $[t_0, t_f]$ satisfy

$$\lambda(t) = \lambda(t_f) + \int_t^{t_f} l_0 f_{0x}'[\tau]^\top + f_x'[\tau]^\top \lambda(\tau) + c_x'[\tau]^\top \eta(\tau) d\tau$$

$$
\begin{aligned}
&+ \sum_{i=1}^{n_s} \int_t^{t_f} s_{ix}'[\tau]^\top d\mu_i(\tau) \qquad t \in [t_0, t_f],\\
\lambda(t_0)^\top =& -(l_0 \varphi_{x_0}' + \sigma^\top \Psi_{x_0}'),\\
\lambda(t_f)^\top =& l_0 \varphi_{x_f}' + \sigma^\top \Psi_{x_f}',
\end{aligned}
$$

$$
\begin{aligned}
l_0 f_{0v}'[t] + \lambda(t)^\top f_v'[t] + \eta(t)^\top c_v'[t] =& 0 \qquad \text{a.e. in} [t_0, t_f],\\
S_\alpha(t)^\top \eta(t) =& 0 \qquad \text{a.e. in} [t_0, t_f],\\
\hat{\mathcal{H}}_u'(t, \hat{x}(t), \hat{u}(t), \hat{v}(t), \lambda(t), l_0)(u - \hat{u}(t)) \geq& 0 \qquad \forall u \in U(t), \text{ a.e. in } [t_0, t_f],\\
\eta(t)^\top c[t] =& 0 \qquad \text{a.e. in } [t_0, t_f],\\
\sum_{i=1}^{n_s} \int_{t_0}^{t_f} s_i(t, \hat{x}(t)) d\mu_i(t) =& 0.
\end{aligned}
\tag{3.23}
$$

These conditions coincide with the necessary optimality conditions (3.15)-(3.22), apart from the supplementary condition (3.23).

It remains to show that multipliers $(l_0, \lambda, \mu, \sigma)$ for a local minimum $(\hat{x}, \hat{u}, \hat{v})$ of 3.1 satisfy equation (3.23).

This equation is satisfied, since:

$$
\begin{aligned}
(S_\alpha(t))_{ii} \neq 0 &\Leftrightarrow c_\alpha^i(t) \neq 0\\
&\Leftrightarrow c_i[t] < -\alpha,
\end{aligned}
$$

so $(S_\alpha(t))_{ii} \neq 0$ implies that $\eta_i(t) = 0$ due to the complementarity condition (3.21).□

Corollary 3.24
Let an OCP 3.1 be given, where φ, f_0, f, Ψ, c, s and U_{ad} satisfy the smoothness Assumptions 3.2. Let $(\hat{x}, \hat{u}, \hat{v})$ be a weak local minimum of the OCP. Assume that the pseudo inverse $(c_{vI_\alpha(t)}'[t])^+$ exists and that there exists a constant $C \in \mathbb{R}$, such that

$$\|(c_{vI_\alpha(t)}'[t])^+\| \leq C \quad \textit{for all } t \in [t_0, t_f]$$

for some $\alpha \geq 0$. Then the assertions of Theorem 3.15 hold.

3.3.2. Normality of the multipliers

In many applications, it is useful to assume that there exist "normal multipliers" for the OCP problem 3.1, i.e. multipliers with $l_0 = 1$. In this section, we derive conditions under which this can be asserted.

As in [Ger06, sections 3.5 and 4.1.3], the following corollary cited from [Ger06, Corollary 3.5.4] gives a condition under which normal multipliers exist. This condition is the Mangasarian-Fromowitz Condition:

Corollary 3.25
Let $G : X \to Y$ and $H : X \to Z$ be Fréchet differentiable at $\hat{x}$, $K \subseteq Y$ a closed convex cone with vertex at zero and $\operatorname{int}(K) \neq \emptyset$, $G(\hat{x}) \in K$, $H(\hat{x}) = 0$. Furthermore, let the following conditions be fulfilled:

1. *Let H' be surjective.*
2. *Let there exist some $\hat{d} \in \operatorname{int}(S - \{\hat{x}\})$ with*

$$\begin{aligned} H'(\hat{x})(\hat{d}) &= 0, \\ G'(\hat{x})(\hat{d}) &\in \operatorname{int}(K - \{G(\hat{x})\}). \end{aligned}$$

Then the assertions of Theorem 3.6 hold with $l_0 = 1$.

The following lemma states a condition under which H' is surjective. The idea for the proof was taken from [Mal03, proof of Lemma 4.1].

Lemma 3.26 (Surjectivity of H')
Let $(\hat{x}, \hat{u}, \hat{v})$ be a local minimum for Problem 3.1. Assume that for every vector $g \in \mathbb{R}^{n_\Psi}$, there exists a solution (x, u, v) to the problem

$$\begin{aligned} \dot{x}(t) &= f'_x[t]x(t) + f'_u[t]u(t) + f'_v[t]v(t), \\ \Psi'_{x_0}x(t_0) + \Psi'_{x_f}x(t_f) &= g. \end{aligned}$$

Then the linearized operator H' of the equality constraints is surjective.

Proof.
Let h_1 be a function $h_1 \in L^\infty([t_0, t_f], \mathbb{R}^{n_x})$. According to Lemma 2.24.1, there exists a solution $w \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$ to the initial value problem

$$\dot{w}(t) = f'_x[t]w(t) - h_1(t), \qquad w(t_0) = 0.$$

Let z, u, v be a solution to

$$\dot{z}(t) = f'_x[t]z(t) + f'_u[t]u(t) + f'_v[t]v(t), \qquad \Psi'_{x_0}z(t_0) + \Psi'_{x_f}z(t_f) = h_2 - \Psi'_{x_f}w(t_f),$$

then with $x := z + w$, it holds

$$\begin{aligned} \dot{x}(t) = (z+w)'(t) &= f'_x[t](z+w)(t) + f'_u[t]u(t) + f'_v[t]v(t) - h_1(t) \\ &= f'_x[t]x(t) + f'_u[t]u(t) + f'_v[t]v(t) - h_1(t), \end{aligned}$$

and

$$\Psi'_{x_0}x(t_0) + \Psi'_{x_f}x(t_f) = \Psi'_{x_0}z(t_0) + \Psi'_{x_f}z(t_f) + \Psi'_{x_f}w(t_f) = h_2.$$

This shows that the equation $H'(x, u, v) = (h_1, h_2)^\top$ is solvable for any h_1, h_2. □

Summarizing, these observations yield the following general conditions for normality (cf. [Ger06, Theorem 4.1.15]):

Theorem 3.27 (Normality I)
Let the functions φ, f_0, f, Ψ, c, s and U_{ad} fulfill the smoothness Assumptions 3.2. Let $(\hat{x}, \hat{u}, \hat{v})$ be a weak local minimum of the OCP and let the following conditions hold:

3.27.1 *There exists a constant $C \in \mathbb{R}$, such that for some $\alpha > 0$, it holds*

$$\|c'_{vI_\alpha(t)}[t]^\top \xi\| \geq C\|\xi\| \quad \text{for all } \xi \in \mathbb{R}^{|I_\alpha(t)|}, \text{ and almost all } t \in [t_0, t_f].$$

3.27.2 *For every vector $g \in \mathbb{R}^{n_\Psi}$, there exists a solution (x, u, v) to the problem*

$$\begin{aligned} \dot{x}(t) &= f'_x[t]x(t) + f'_u[t]u(t) + f'_v[t]v(t), \\ \Psi'_{x_0}x(t_0) + \Psi'_{x_f}x(t_f) &= g. \end{aligned}$$

3.27.3 *For some $\varepsilon > 0$, there exist functions $x_0 \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$, $u_0 \in \operatorname{int}(U_{ad} - \hat{u})$ and $v_0 \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$, such that a.e. in $[t_0, t_f]$, it holds:*

$$\begin{aligned} c[t] + c'_x[t]x_0 + c'_v[t]v_0 &\leq -\varepsilon e \\ s[t] + s'_x[t]x_0 &< 0, \\ \dot{x}_0(t) = f'_x[t]x_0 + f'_u[t]u_0 + f'_v[t]v_0, \\ \Psi'_{x_0}x_0(t_0) + \Psi'_{x_f}x_0(t_f) &= 0. \end{aligned}$$

Then there exist multipliers

$$l_0 = 1, \lambda \in BV([t_0, t_f], \mathbb{R}^{n_x}), \eta \in L^\infty([t_0, t_f], \mathbb{R}^{n_c}), \mu \in NBV([t_0, t_f], \mathbb{R}^{n_s}) \text{ and } \sigma \in \mathbb{R}^{n_\Psi},$$

such that (3.15)-(3.22) *are fulfilled.*

The following corollary cites a condition analogous to [Ger06, Theorem 4.1.14] that is sufficient for condition 3.27.2:

Corollary 3.28 (Normality II)
Let

$$\operatorname{rank}\left(\Psi'_{x_0}\Phi(t_0) + \Psi'_{x_f}\Phi(t_f)\right) = n_\Psi,$$

where Φ solves

$$\Phi'(t) = f'_x[t]\Phi(t), \qquad \Phi(t_0) = I_{n_x}.$$

Then condition 3.27.2 is fulfilled.

From Theorem 2.31, we can derive a different assumption under which this condition is met:

Corollary 3.29 (Normality III)
Let the partial derivatives of f be sufficiently smooth, i.e. let

$$f'_x[t] \in C^k([t_0, t_f], \mathbb{R}^{n_x \times n_x}),\ f'_u[t] \in C^k([t_0, t_f], \mathbb{R}^{n_x \times n_u}) \text{ and } f'_v[t] \in C^k([t_0, t_f], \mathbb{R}^{n_x \times n_v})$$

for some integer $k > 0$.

Let

$$B_0(t) := \left(f'_u[t], f'_v[t]\right), \qquad B_{i+1}(t) := f'_x[t]B_i(t) - \frac{d}{dt}B_i(t)$$

for $i = 0, \ldots, k-1$.

Assume that there exists a $\tau \in [t_0, t_f]$, *for which*

$$\operatorname{rank}\left(B_0(\tau), B_1(\tau), \ldots, B_k(\tau)\right) = n_x.$$

Let Ψ'_{x_0} *and* Ψ'_{x_f} *satisfy*

$$\operatorname{rank}\left(\Psi'_{x_0}, \Psi'_{x_f}\right) = n_\Psi.$$

Then condition 3.27.2 is fulfilled.

Proof.
If the assumptions of corrollary 3.29 are satisfied, then the continuous-time linear system with right hand side

$$f^{lin}(t, x, u, v) = f'_x[t]x + f'_{u,v}[t](u, v)^\top$$

is controllable. Hence, given any initial vector x_0 and final vector x_f, there exists a trajectory $(\tilde{x}, \tilde{u}, \tilde{v})$ that satisfies the differential equation

$$\dot{\tilde{x}}(t) = f^{lin}(t, \tilde{x}(t), \tilde{u}(t), \tilde{v}(t)) \quad \text{a.e. on } [t_0, t_f] \tag{3.24}$$

as well as the boundary conditions

$$\tilde{x}(t_0) = x_0, \qquad \tilde{x}(t_f) = x_f.$$

Since $\operatorname{rank}\left(\Psi'_{x_0}, \Psi'_{x_f}\right) = n_\Psi$, the linear equation $\Psi'_{x_0}x_0 + \Psi'_{x_f}x_f = g$ admits a solution for any right hand side $g \in \mathbb{R}^{n_\Psi}$. With this solution (x_0, x_f), $(\tilde{x}, \tilde{u}, \tilde{v})$ satisfies (3.24) as well as the boundary conditions. □

Remark 3.30 (Independence of Conditions)
Neither of the conditions stated in Corollaries 3.28 and 3.29 implies the other.

Corollary 3.28 implies that any problem, where the boundary conditions take the form of $x(t_0) = x_0$, *i.e. start conditions, satisfies 3.27.2. The rank assumption implies that* $n_\Psi \leq n_x$.

Corollary 3.29 states that all problems where the linearization is controllable fulfill 3.27.2. This implies $n_\Psi \leq 2 \cdot n_x$.

4. Linear Quadratic Optimal Control Problems

An important part of this work is the investigation of Linear Quadratic Optimal Control Problems ($LQOCP$), for which, in section 4.2, we introduce a regularization concept, cf. [GHed].

A Linear Quadratic OCP with control set constraints, mixed control state constraints and pure state constraints is a problem of the form:

Problem 4.1 (*LQOCP*)

$$\begin{aligned}\text{min!} \quad J^{LQP}(x,u,v) :=& \frac{1}{2}x(t_f)^\top Q_f x(t_f) \\ &+\frac{1}{2}\int_{t_0}^{t_f} \left(x(t)^\top, u(t)^\top, v(t)^\top\right) \begin{pmatrix} Q(t) & R_u(t) & R_v(t) \\ R_u(t)^\top & S_u(t) & 0 \\ R_v(t)^\top & 0 & S_v(t) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ v(t) \end{pmatrix} dt\end{aligned}$$

$$\begin{aligned} \textit{with respect to the state function} \quad & x \in W^{1,\infty}([t_0,t_f],\mathbb{R}^{n_x}) \\ \textit{and the control functions} \quad & u \in L^{\infty}([t_0,t_f],\mathbb{R}^{n_u}) \\ \textit{and} \quad & v \in L^{\infty}([t_0,t_f],\mathbb{R}^{n_v}) \end{aligned}$$

subject to the differential equation

$$\dot{x}(t) = A(t)x(t) + B(t)\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} \quad \textit{a.e. in } [t_0,t_f],$$

boundary conditions

$$E_0 x(t_0) + E_1 x(t_f) = f,$$

mixed control state constraints

$$G(t)x(t) + H(t)v(t) \leq l(t) \quad \textit{a.e. in } [t_0,t_f],$$

pure state constraints

$$C(t)x(t) \leq d(t) \quad \textit{in } [t_0,t_f],$$

and control set constraints

$$u(t) \in U(t) \subset \mathbb{R}^{n_u} \quad \textit{a.e. in } [t_0,t_f].$$

Let B be partitioned, $B(t) = (B_u(t), B_v(t))$. The weighting matrix will be named W:

$$W(t) := \begin{pmatrix} Q(t) & R_u(t) & R_v(t) \\ R_u(t)^\top & S_u(t) & 0 \\ R_v(t)^\top & 0 & S_v(t) \end{pmatrix}.$$

For any positive semidefinite square Matrix A, $\|\cdot\|_A$ is defined as

$$\|x\|_A := \sqrt{x^\top A x}.$$

If A is positive (semi-) definite, then $\|\cdot\|_A$ is a (half) norm.

Also, for any square matrix function $A \in L^\infty([t_0, t_f], \mathbb{R}^{n\times n})$, $\|\cdot\|_A$ will denote

$$\|x\|_A := \left(\int_{t_0}^{t_f} x(t)^\top A(t) x(t) dt\right)^{\frac{1}{2}}.$$

The Hamilton function for Problem 4.1 reads

$$\begin{aligned}\hat{\mathcal{H}}(t, x, u, v, \lambda, \eta, l_0) := &\frac{1}{2} l_0 \|(x, u, v)\|^2_{W(t)} + \lambda^\top (A(t)x + B_u(t)u + B_v(t)v) \\ &+ \eta^\top (G(t)x + H(t)v - l(t)).\end{aligned}$$

In the remainder of this work, the data of the problem will be assumed to satisfy the following smoothness assumption:

Assumption 4.2 (Smoothness)
The matrix $Q_f \in \mathbb{R}^{n_x \times n_x}$ as well as the matrix functions $Q : [t_0, t_f] \to \mathbb{R}^{n_x \times n_x}$, $S_u : [t_0, t_f] \to \mathbb{R}^{n_u \times n_u}$ and $S_v : [t_0, t_f] \to \mathbb{R}^{n_v \times n_v}$ are symmetric. The matrix Q_f is positive semidefinite, and the weighting matrix $W(t)$ is positive semidefinite for all $t \in [t_0, t_f]$.

All of the following functions are continuous:

4.2.1 $Q : [t_0, t_f] \to \mathbb{R}^{n_x \times n_x}$, $R_u : [t_0, t_f] \to \mathbb{R}^{n_x \times n_u}$, $R_v : [t_0, t_f] \to \mathbb{R}^{n_x \times n_v}$, $S_u : [t_0, t_f] \to \mathbb{R}^{n_u \times n_u}$, $S_v : [t_0, t_f] \to \mathbb{R}^{n_v \times n_v}$

4.2.2 $A : [t_0, t_f] \to \mathbb{R}^{n_x \times n_x}$, $B_u : [t_0, t_f] \to \mathbb{R}^{n_x \times n_u}$, $B_v : [t_0, t_f] \to \mathbb{R}^{n_x \times n_v}$,

4.2.3 $G : [t_0, t_f] \to \mathbb{R}^{n_c \times n_x}$, $H : [t_0, t_f] \to \mathbb{R}^{n_c \times n_v}$, $l : [t_0, t_f] \to \mathbb{R}^{n_c}$,

4.2.4 $C : [t_0, t_f] \to \mathbb{R}^{n_s \times n_x}$ *and* $d : [t_0, t_f] \to \mathbb{R}^{n_s}$.

4.1. Properties of the Problem

For the analysis of the problem, the existence of normal multipliers in the necessary optimality conditions is essential. Hence, the assumption that guarantees normality is stated in the form of Theorem 3.27:

Assumption 4.3 (LQOCP Normality)
Let the data of the Linear Quadratic Problem 4.1 satisfy the smoothness Assumption 4.2. Let $(\hat{x}, \hat{u}, \hat{v})$ be a local minimum and $\alpha > 0$, such that the following conditions are fulfilled:

4.3.1 *The pseudo inverse* $(H_{I_\alpha(t)}(t))^+$ *exists, and there exists a constant* $C \in \mathbb{R}$, *such that* $\|(H_{I_\alpha(t)}(t))^+\| \leq C \quad$ *for all* $t \in [t_0, t_f]$

4.3.2 *For any vector* $g \in \mathbb{R}^{n_\Psi}$, *there exists a solution to the boundary value problem*

$$\dot{x}(t) = A(t)x(t) + B(t)\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} \quad \text{a.e. in } [t_0, t_f]$$
$$E_0 x(t_0) + E_1 x(t_f) = g.$$

4.3.3 *There exists an* $\varepsilon > 0$ *and a solution* (x_0, u_0, v_0), *where* $x_0 \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$, $u_0 \in \operatorname{int}(U_{ad} - \hat{u})$, $v_0 \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$, *to the system*

$$G(t)(\hat{x}(t) + x_0(t)) + H(t)(\hat{v}(t) + v_0(t)) \leq l(t) - \varepsilon e_{n_c} \quad \text{a.e. in } [t_0, t_f],$$
$$C(t)(\hat{x}(t) + x_0(t)) < d(t) \quad \text{in } [t_0, t_f],$$
$$\dot{x}_0(t) = A(t)x_0(t) + B(t)\begin{pmatrix} u_0(t) \\ v_0(t) \end{pmatrix} \quad \text{a.e. in } [t_0, t_f],$$
$$E_0 x_0(t_0) + E_1 x_0(t_f) = 0.$$

Remark 4.4
Seemingly, part 4.3.3 of the normality conditions could be replaced by the following system that is independent of the local minimum in investigation:

$$G(t)x_0(t) + H(t)v_0(t) \leq -\varepsilon e \quad \text{a.e. in } [t_0, t_f],$$
$$C(t)x_0(t) < 0 \quad \text{in } [t_0, t_f],$$
$$\dot{x}_0(t) = A(t)x_0(t) + B(t)\begin{pmatrix} u_0(t) \\ v_0(t) \end{pmatrix} \quad \text{a.e. in } [t_0, t_f],$$
$$E_0 x_0(t_0) + E_1 x_0(t_f) = 0. \tag{4.1}$$

Indeed, if the above system is solvable, then Assumption 4.3.3 is satisfied. However, this assumption is a lot stronger. Consider the case when $E_0 = I$, $E_1 = 0$, *i.e. there exists a specific start value for the system. Then for any* x_0 *that satisfies the above system it holds* $x_0(t_0) = 0$ *due to equation* (4.1), *so that* $C(t_0)x_0(t_0) = 0$. *Hence the above conditions are violated if state constraints are present, while there still may be a solution to the system in 4.3.3 if the state constraints are inactive in* t_0.

Unlike in [GHed], we consider mixed control state constraints instead of control set constraints in this work, since the latter are usually modeled using mixed control state constraints. The most common example of control set constraints are box constraints of the form $v_{\min}(t) \leq v(t) \leq v_{\max}(t)$. Constraints of this type do satisfy the rank condition 4.3.1:

Lemma 4.5 (Box Constraints and the Rank Condition)
Assume that the mixed control state constraints are given in the form of

$$\begin{pmatrix} I \\ -I \end{pmatrix} v(t) \leq \begin{pmatrix} v_{\max}(t) \\ v_{\min}(t) \end{pmatrix},$$

where $v_{\max}(t) - v_{\min}(t) \geq \varepsilon > 0$ *a.e. in* $[t_0, t_f]$. *Then it holds that* $(H_{I_\alpha(t)}(t)H_{I_\alpha(t)}(t)^\top) = I_{|I_\alpha(t)|}$ *for some* $\alpha > 0$.

Proof.
Let $\alpha := \varepsilon/2$, then for any control v, it is impossible for any index $i \in \{1, \dots, n_v\}$ that $i \in I_\alpha(t)$ and $i + n_v \in I_\alpha(t)$ at the same time $t \in [t_0, t_f]$. Therefore, each column of $H_{I_\alpha(t)}(t)$ contains at most one element in $\{-1, 1\}$. At the same time, each row in $H_{I_\alpha(t)}(t)$ contains exactly one element in $\{-1, 1\}$ (since the constraint described by this row is α-active). Hence there exists a permutation matrix P_c, such that $H_{I_\alpha(t)}(t) = (J, 0)P_c$, where $J = \operatorname{diag}(j_l)_{l=1}^{|I_\alpha(t)|}$, with $j_l \in \{-1, 1\}$ for $l = 1, \dots, |I_\alpha(t)|$, and it holds

$$\begin{aligned} H_{I_\alpha(t)}(t)(H_{I_\alpha(t)}(t))^\top =& ((J, 0)P_c)((J, 0)P_c)^\top \\ =& (J, 0)P_c P_c^\top (J, 0)^\top = (J, 0)(J, 0)^\top = I_{|I_\alpha(t)|}. \end{aligned}$$ □

In the linear quadratic case, it is often convenient to make use of Corollary 3.29, especially if the matrices A and B are autonomous:

Lemma 4.6
4.6.1 *Let $A \in \mathcal{C}^\infty([t_0, t_f], \mathbb{R}^{n_x \times n_x})$, $B \in \mathcal{C}^\infty([t_0, t_f], \mathbb{R}^{n_x \times (n_u + n_v)})$. Let*

$$B_0(t) := B(t), \qquad B_{i+1}(t) := A(t)B_i(t) - \frac{d}{dt}B_i(t), \quad \text{for } i \in \mathbb{N}.$$

Assume that there exist $\tau \in [t_0, t_f]$ and $k \in \mathbb{N}$, with

$$\operatorname{rank}(B_0(\tau), \dots, B_k(\tau)) = n_x.$$

Let E_0 and E_1 satisfy

$$\operatorname{rank}(E_0, E_1) = n_\Psi.$$

Then Assumption 4.3.2 holds.

4.6.2 *Let A and B be constant. Assume that*

$$\operatorname{rank}(B, AB, \dots, A^{n_x - 1}B) = n_x$$

and

$$\operatorname{rank}(E_0, E_1) = n_\Psi.$$

Then Assumption 4.3.2 holds.

Proof.
The assertion of Lemma 4.6.1 is a direct implication of Corollary 3.29. Lemma 4.6.2 includes the assertion that only the powers of A up to $A^{n_x - 1}$ need to be considered.

Observe that if a vector $A^{k+1}B_j$ is linearly independent of $(A^0B, \dots, A^kB)$, then A^kB_j is linearly independent of $(A^0B, \dots, A^{k-1}B)$.[1] Hence, if

$$\operatorname{rank}(A^0B, \dots, A^kB) = \operatorname{rank}(A^0B, \dots, A^{k+1}B) = m$$

for some m, then $\operatorname{rank}(A^0B, \dots, A^iB) = m$ for all $i \geq k$. Since the dimension of this family of vectors cannot exceed n_x, this proves the assertion. □

[1] Otherwise let $A^kB_j = \sum_{i=0}^{k-1} A^iB\lambda_i$, where $\lambda_i \in \mathbb{R}^{n_u + n_v}$, then $A^{k+1}B_j = \sum_{i=0}^{k-1} AA^iB\lambda_i = \sum_{i=1}^{k} A^iB\lambda_{i-1}$ is a linear combination as well.

The implications of the minimum principle 3.24 and the regularity lead to several observations, e.g. uniqueness and continuity of solutions under appropriate assumptions.

Corollary 4.7 (Necessary Optimality Conditions for LQOCP)
Let $(\hat{x}, \hat{u}, \hat{v})$ be a weak local minimum of LQOCP, satisfying Assumptions 4.2 and 4.3.

Then there exist multipliers λ, η, μ and σ, $\lambda \in BV([t_0, t_f], \mathbb{R}^{n_x})$, $\eta \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})$, $\mu \in NBV([t_0, t_f], \mathbb{R}^{n_s})$ and $\sigma \in \mathbb{R}^{n_\Psi}$ that satisfy

1. *Adjoint equation:*

$$\begin{aligned}\lambda(t) = \lambda(t_0) - \int_{t_0}^{t} Q(\tau)\hat{x}(\tau) + R_u(\tau)\hat{u}(\tau) + R_v(\tau)\hat{v}(\tau) + A(\tau)^\top\lambda(\tau) + G(\tau)^\top\eta(\tau)d\tau \\ - \int_{t_0}^{t} C(\tau)^\top d\mu(\tau) \qquad t \in [t_0, t_f]\end{aligned} \tag{4.2}$$

2. *Transversality conditions:*

$$\lambda(t_0) = -E_0^\top\sigma \tag{4.3}$$
$$\lambda(t_f) = Q_f\hat{x}(t_f) + E_1^\top\sigma \tag{4.4}$$

3. *Optimality conditions:*

$$0 = S_v(t)\hat{v}(t) + R_v(t)^\top\hat{x}(t) + B_v(t)^\top\lambda(t) + H(t)^\top\eta(t) \tag{4.5}$$
$$0 \leq \left(\hat{u}(t)^\top S_u(t) + \hat{x}(t)^\top R_u(t) + \lambda(t)^\top B_u(t)\right)(u - \hat{u}(t)) \qquad \forall u \in U(t) \tag{4.6}$$

4. *Complementarity conditions:*

$$\eta(t)^\top(G(t)\hat{x}(t) + H(t)\hat{v}(t) - l(t)) = 0, \qquad \eta(t) \geq 0 \qquad \text{a.e. in } [t_0, t_f] \tag{4.7}$$

The multipliers μ satisfy

$$0 \leq \int_{t_0}^{t_f} z(t)^\top d\mu(t) \qquad \forall z \in \{z \in \mathcal{C}([t_0, t_f], \mathbb{R}^{n_s}) | z(\cdot) \geq 0\} \tag{4.8}$$

and

$$\int_{t_0}^{t_f} \left(C(t)\hat{x}(t) - d(t)\right)^\top d\mu(t) = 0. \tag{4.9}$$

Lemma 4.8 (Sufficiency of the Optimality Conditions)
Let $(\hat{x}, \hat{u}, \hat{v})$ be a weak local minimum of LQP, satisfying Assumptions 4.2 and 4.3, and let $\hat{\lambda} \in BV([t_0, t_f], \mathbb{R}^{n_x})$, $\hat{\eta} \in L^\infty([t_0, t_f], \mathbb{R}^{n_c})$, $\hat{\mu} \in NBV([t_0, t_f], \mathbb{R}^{n_s})$ and $\hat{\sigma} \in \mathbb{R}^{n_\Psi}$ be the associated multipliers as in Corollary 4.7. Let the matrix $W(t)$ be positive semidefinite for every $t \in [t_0, t_f]$.

If $(x, u, v) \neq (\hat{x}, \hat{u}, \hat{v})$ is a feasible point, then $J^{LQP}(x, u, v) \geq J^{LQP}(\hat{x}, \hat{u}, \hat{v})$. If $W(t)$ is positive definite, then $J^{LQP}(x, u, v) > J^{LQP}(\hat{x}, \hat{u}, \hat{v})$.

Proof.
Let $(x,u,v) \in W^{1,\infty} \times L^\infty \times L^\infty$ be feasible, i.e.

$$\begin{aligned}
\dot{x}(t) = A(t)x(t) + B_u(t)u(t) + B_v(t)v(t) \quad &\text{a.e. in } [t_0,t_f],\\
E_0x(t_0) + E_1x(t_f) &= f,\\
G(t)x(t) + H(t)v(t) \leq l(t) \quad &\text{a.e. in } [t_0,t_f],\\
C(t)x(t) \leq d(t) \quad &\text{in } [t_0,t_f],\\
u(t) \in U(t) \quad &\text{a.e. in } [t_0,t_f].
\end{aligned}$$

Then for the multipliers $\hat{\lambda}, \hat{\eta}$ as in Corollary 4.7 it holds:

$$\begin{aligned}
J^{LQP}(x,u,v) = \frac{1}{2}x(t_f)^\top Q_f x(t_f)\\
+ \int_{t_0}^{t_f} \hat{\mathcal{H}}(t,x,u,v,\hat{\lambda},\hat{\eta},1) - \hat{\lambda}^\top \dot{x} - \hat{\eta}^\top(G(t)x(t) + H(t)v(t) - l(t))dt,
\end{aligned}$$

so that

$$\begin{aligned}
&J^{LQP}(x,u,v) - J^{LQP}(\hat{x},\hat{u},\hat{v})\\
&\overset{Q_f=Q_f^\top}{=} \underbrace{\frac{1}{2}(x(t_f)-\hat{x}(t_f))^\top Q_f(x(t_f)-\hat{x}(t_f))}_{\geq 0, \text{ since } Q_f \geq 0}\\
&\quad + \hat{x}(t_f)^\top Q_f(x(t_f)-\hat{x}(t_f))\\
&\quad + \int_{t_0}^{t_f} \hat{\mathcal{H}}(t,x(t),u(t),v(t),\hat{\lambda}(t),\hat{\eta}(t),1)\\
&\qquad - \hat{\mathcal{H}}(t,\hat{x}(t),\hat{u}(t),\hat{v}(t),\hat{\lambda}(t),\hat{\eta}(t),1) - \hat{\lambda}(t)^\top(\dot{x}(t)-\dot{\hat{x}}(t))\\
&\qquad - \underbrace{\hat{\eta}(t)^\top}_{\geq 0}\underbrace{(G(t)x(t)+H(t)v(t)-l(t))}_{\leq 0}\\
&\qquad + \underbrace{\hat{\eta}(t)^\top(G(t)\hat{x}(t)+H(t)\hat{v}(t)-l(t))}_{=0}dt\\
&\geq \int_{t_0}^{t_f} \hat{\mathcal{H}}(t,x(t),u(t),v(t),\hat{\lambda}(t),\hat{\eta}(t),1)\\
&\qquad - \hat{\mathcal{H}}(t,\hat{x}(t),\hat{u}(t),\hat{v}(t),\hat{\lambda}(t),\hat{\eta}(t),1)dt\\
&\qquad - \int_{t_0}^{t_f} \hat{\lambda}(t)^\top(\dot{x}(t)-\dot{\hat{x}}(t))dt + \hat{x}(t_f)^\top Q_f(x(t_f)-\hat{x}(t_f)).
\end{aligned}$$

Partial integration of the term $\int_{t_0}^{t_f} \hat{\lambda}(t)^\top(\dot{x}(t)-\dot{\hat{x}}(t))dt$ yields

$$\int_{t_0}^{t_f} \hat{\lambda}(t)^\top(\dot{x}(t)-\dot{\hat{x}}(t))dt = [\hat{\lambda}^\top(x-\hat{x})]_{t_0}^{t_f} - \int_{t_0}^{t_f}(x(t)-\hat{x}(t))^\top d\hat{\lambda}(t),$$

where

$$\begin{aligned}[\hat{\lambda}^\top(x-\hat{x})]_{t_0}^{t_f} =&(\hat{x}(t_f)^\top Q_f + \hat{\sigma}^\top E_1)(x(t_f)-\hat{x}(t_f)) + \hat{\sigma}^\top E_0(x(t_0)-\hat{x}(t_0))\\ =&\hat{x}(t_f)^\top Q_f(x(t_f)-\hat{x}(t_f))\\ &+\hat{\sigma}^\top(E_0x(t_0)+E_1x(t_f)) - \hat{\sigma}^\top(E_0\hat{x}(t_0)+E_1\hat{x}(t_f))\\ =&\hat{x}(t_f)^\top Q_f(x(t_f)-\hat{x}(t_f)) + \hat{\sigma}(f-f)\\ =&\hat{x}(t_f)^\top Q_f(x(t_f)-\hat{x}(t_f))\end{aligned}$$

and

$$\begin{aligned}\int_{t_0}^{t_f}(x-\hat{x})^\top d\hat{\lambda}(t) =& -\int_{t_0}^{t_f}(x-\hat{x})^\top\left[Q\hat{x}+R_u\hat{u}+R_v\hat{v}+A^\top\hat{\lambda}+G^\top\hat{\eta}\right]dt\\ &-\int_{t_0}^{t_f}(x-\hat{x})^\top C^\top d\hat{\mu}(t)\\ =&-\int_{t_0}^{t_f}(x-\hat{x})^\top\left[Q\hat{x}+R_u\hat{u}+R_v\hat{v}+A^\top\hat{\lambda}+G^\top\hat{\eta}\right]dt\\ &+\int_{t_0}^{t_f}\underbrace{(d-Cx)}_{\geq 0}^\top d\hat{\mu}(t) + \underbrace{\int_{t_0}^{t_f}(C\hat{x}-d)^\top d\hat{\mu}(t)}_{=0}\\ \geq&-\int_{t_0}^{t_f}(x-\hat{x})^\top\left[Q\hat{x}+R_u\hat{u}+R_v\hat{v}+A^\top\hat{\lambda}+G^\top\hat{\eta}\right]dt.\end{aligned}$$

The difference of the Hamilton functions is

$$\begin{aligned}&\int_{t_0}^{t_f}\hat{\mathcal{H}}(t,x,u,v,\hat{\lambda},\hat{\eta},1)-\hat{\mathcal{H}}(t,\hat{x},\hat{u},\hat{v},\hat{\lambda},\hat{\eta},1)dt\\ &\quad=\frac{1}{2}\|(x,u,v)\|_W^2-\frac{1}{2}\|(\hat{x},\hat{u},\hat{v})\|_W^2\\ &\qquad+\int_{t_0}^{t_f}\hat{\lambda}^\top(A(x-\hat{x})+B_u(u-\hat{u})+B_v(v-\hat{v}))+\hat{\eta}^\top(G(x-\hat{x})+H(v-\hat{v}))dt,\end{aligned}$$

and for each $t\in[t_0,t_f]$ we get:

$$\begin{aligned}&\frac{1}{2}\|(x,u,v)\|_{W(t)}^2-\frac{1}{2}\|(\hat{x},\hat{u},\hat{v})\|_{W(t)}^2\\ &\quad=\frac{1}{2}\|(x-\hat{x},u-\hat{u},v-\hat{v})\|_{W(t)}^2+(\hat{x},\hat{u},\hat{v})^\top W(t)\begin{pmatrix}x-\hat{x}\\u-\hat{u}\\v-\hat{v}\end{pmatrix}\\ &\quad=\frac{1}{2}\|(x-\hat{x},u-\hat{u},v-\hat{v})\|_{W(t)}^2+(\hat{x}^\top Q+\hat{u}^\top R_u^\top+\hat{v}^\top R_v^\top)(x-\hat{x})\\ &\qquad+(\hat{x}^\top R_u+\hat{u}^\top S_u)(u-\hat{u})+(\hat{x}^\top R_v+\hat{v}^\top S_v)(v-\hat{v}).\end{aligned}$$

Summarizing, we have

$$\begin{aligned}J^{LQP}(x,u,v)-J^{LQP}(\hat{x},\hat{u},\hat{v})\geq&\frac{1}{2}\|(x-\hat{x},u-\hat{u},v-\hat{v})\|_W^2\\ &+\int_{t_0}^{t_f}\left(\hat{x}^\top Q+\hat{u}^\top R_u^\top+\hat{v}^\top R_v^\top\right)(x-\hat{x})\end{aligned}$$

$$
\begin{aligned}
&+ \left(\hat{x}^\top R_u + \hat{u}^\top S_u\right)(u-\hat{u}) + \left(\hat{x}^\top R_v + \hat{v}^\top S_v\right)(v-\hat{v})dt \\
&+ \int_{t_o}^{t_f} \hat{\lambda}^\top \left(A(x-\hat{x}) + B_u(u-\hat{u}) + B_v(v-\hat{v})\right) \\
&+ \hat{\eta}^\top \left(G(x-\hat{x}) + H(v-\hat{v})\right) dt \\
&+ \hat{x}(t_f)^\top Q(x(t_f) - \hat{x}(t_f)) - \hat{x}(t_f)^\top Q(x(t_f) - \hat{x}(t_f)) \\
&- \int_{t_0}^{t_f} (x-\hat{x})^\top \left(Q\hat{x} + R_u\hat{u} + R_v\hat{v} + A^\top\hat{\lambda} + G^\top\hat{\eta}\right) dt \\
=&\frac{1}{2}\|(x-\hat{x}, u-\hat{u}, v-\hat{v})\|_W^2 \\
&+ \int_{t_0}^{t_f} \underbrace{\left(\hat{x}^\top R_u + \hat{u}^\top S_u + \hat{\lambda}^\top B_u\right)(u-\hat{u})}_{\geq 0 \quad \text{a.e. in } [t_0,t_f]} dt \\
&+ \int_{t_0}^{t_f} \underbrace{\left(\hat{x}^\top R_v + \hat{v}^\top S_v + \hat{\lambda}^\top B_v + \hat{\eta}^\top H\right)}_{=0}(v-\hat{v})dt \\
\geq&\frac{1}{2}\|(x-\hat{x}, u-\hat{u}, v-\hat{v})\|_W^2.
\end{aligned}
$$

This shows the assertions, since $\|(x-\hat{x}, u-\hat{u}, v-\hat{v})\|_W^2 \geq 0$ and $\|(x-\hat{x}, u-\hat{u}, v-\hat{v})\|_W^2 = 0 \Leftrightarrow (x-\hat{x}, u-\hat{u}, v-\hat{v}) = 0 \quad$ a.e. on $[t_0, t_f]$, if $W(t)$ is positive definite. □

In [Hag79], a theoretic result was introduced that became the basis of several investigations about the continuity of optimal control functions, and a continuity result was presented for a special class of problems. In this class, the constraints take the form of pure control constraints and pure state constraints. Control set constraints are not present, and the boundary conditions are initial conditions. The objective function does not include a Mayer term, i.e. $Q_f = 0$. For unconstrained problems with boundary conditions $x(t_0) = x_0$, $x(t_f) = x_1$, continuity has been investigated in [CV90]. Lemma 4.10 is taken from [GV03, Theorem 3.1 and comment (c)].

Definition 4.9
The set $J(t,x) \subset \{1,\dots,n_s\}$ denotes the index set of active state constraints:

$$J(t,x) : [t_0,t_f] \times \mathbb{R}^{n_x} \to 2^{\{1,\dots,n_s\}}, \qquad (t,x) \mapsto \{i \in \{1,\dots,n_s\} | C^i(t)x = d_i(t)\}.$$

Lemma 4.10 (Continuity of Solutions)
Assume that in Problem 4.1 it holds $n_v = 0$. Let $U(t) = U \quad \forall t \in [t_0,t_f]$ for some closed convex constant set U, and assume that A and B are locally Lipschitz continuous, H is differentiable with Lipschitz continuous gradient and S_u is Lipschitz continuous and positive definite. Let $(\hat{x},\hat{u})$ be a local minimum of LQOCP, satisfying Assumptions 4.2 and 4.3. Moreover, assume that

$$C(t)_{J(t,\hat{x}(t))}G(t)\xi \notin \operatorname{span} N_U(\hat{u})$$

for all $\xi \in \mathbb{R}_+^{|J(t,\hat{x}(t))|}$. Then $\hat{u}$ is Lipschitz continuous.

4.2. Virtual Control as a Regularization Concept

In [GHed], a regularization concept that had been introduced for PDE constrained problems in [KR08] and [CKR08] was applied to linear quadratic optimal control problems. This concept allowed to treat pure state constraints as mixed control state constraints. So far, we considered mixed control state constraints as well as control set constraints under the assumption that the respective controls on which these constraints are imposed are independent. Consequently, the regularization will be applied to this class of problems.

In the virtual control concept, the control of the problem under investigation is augmented by as many control variables as there are state constraints. These new variables are subtracted from the left hand side of the state constraints, such that a violation of the state constraints can be compensated by these new controls. In order to encourage compliance with these constraints, an L_2 penalty term is added to the objective function. The influence of the new controls on the constraints as well as the "cost" imposed on the usage of them can be influenced by a regularization parameter α. It is also possible to model influence on the differential equation.

As there are three possibilities of inserting the regularization in the problem, we introduce three functions depending on the actual regularization parameter α:

- γ models the actual regularization, i.e. $\gamma(\alpha) \cdot w_i$, where w is the virtual control, is substracted from the left hand side of the state constraint. This way, it is made easier to satisfy the constraint if $\gamma(\alpha) > 0$ is satisfied.
- ϕ regulates the cost of the regularization. The factor $\frac{\phi(\alpha)}{2} \cdot \int_{t_0}^{t_f} \|w\|_2^2 dt$ is added to the objective function. This term makes it expensive to use the virtual control w if $\phi(\alpha) > 0$ holds.
- κ can be used to make it easier for the system to satisfy the state constraints. The term $\kappa(\alpha) \cdot \sum_{i=1}^{n_s} w_i$ is substracted from the right hand side of the differential equation. This way, the growth of the state trajectory can be lowered. It is within the decision of the user to make $\kappa(\alpha)$ greater or smaller than zero or even to set $\kappa(\alpha) = 0$.

For Problem 4.1, the augmented problem LQOCP$_\alpha$ reads:

Problem 4.11 (LQOCP$_\alpha$)

$$\begin{aligned}\text{min!} \quad J_\alpha^{LQP}(x,u,v,w) :=& \frac{1}{2} x(t_f)^\top Q_f x(t_f) \\ &+ \frac{1}{2} \int_{t_0}^{t_f} \left(x(t)^\top, u(t)^\top, v(t)^\top\right) \begin{pmatrix} Q(t) & R_u(t) & R_v(t) \\ R_u(t)^\top & S_u(t) & 0 \\ R_v(t)^\top & 0 & S_v(t) \end{pmatrix} \begin{pmatrix} x(t) \\ u(t) \\ v(t) \end{pmatrix} dt \\ &+ \frac{\phi(\alpha)}{2} \int_{t_0}^{t_f} \|w(t)\|_2^2 dt\end{aligned}$$

with respect to the state function $x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$,
the control functions $u \in L^\infty([t_0, t_f], \mathbb{R}^{n_u})$
and $v \in L^\infty([t_0, t_f], \mathbb{R}^{n_v})$
and the virtual control $w \in L^\infty([t_0, t_f], \mathbb{R}^{n_s})$

subject to the differential equation

$$\dot{x}(t) = A(t)x(t) + B(t)\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} - \kappa(\alpha) e_{n_x} e_{n_s}^\top w(t) \quad \text{a.e. in } [t_0, t_f],$$

boundary conditions

$$E_0 x(t_0) + E_1 x(t_f) = f,$$

mixed control state constraints

$$\begin{pmatrix} G(t) \\ C(t) \end{pmatrix} x(t) + \begin{pmatrix} H(t) & 0 \\ 0 & -\gamma(\alpha) \end{pmatrix} \begin{pmatrix} v(t) \\ w(t) \end{pmatrix} \leq \begin{pmatrix} l(t) \\ d(t) \end{pmatrix}$$

and control set constraints

$$u(t) \in U(t) \subset \mathbb{R}^{n_u} \quad \text{a.e. in } [t_0, t_f]$$

The second mixed control state constraint $C(t)x(t) - \gamma(\alpha)w(t) \leq d(t)$ fixes the value for $\hat{w}_\alpha$, as it is shown in the following lemma:

Lemma 4.12
Let $C \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_s \times n_x})$ and $d \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_s})$ and $\gamma(\alpha) > 0$.
Let $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be a local minimum of Problem 4.11. Then it holds that $\hat{w}_\alpha \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_s})$ (more precisely: there exists a representative of the class that is an element of $W^{1,\infty}$), and

$$\gamma(\alpha)\hat{w}_\alpha(t) = \max\{0, C(t)\hat{x}_\alpha(t) - d(t)\}.$$

Proof.
It suffices to show the above equation. If the equation holds, then $\hat{w}_\alpha$ belongs to $W^{1,\infty}$, since the maximum function is Lipschitz continuous and the composition operator of a Lipschitz continuous function maps into $W^{1,\infty}$, cf. [Mer91].

For any admissible $(x_\alpha, u_\alpha, v_\alpha, w_\alpha)$, it follows that $\gamma(\alpha)w_\alpha(t) \geq C(t)x_\alpha(t) - d(t)$ holds a.e. in $[t_0, t_f]$. Assume that $(\tilde{x}_\alpha, \tilde{u}_\alpha, \tilde{v}_\alpha, \tilde{w}_\alpha)$ is optimal for Problem 4.11, and that $\tilde{w}_\alpha$ does disobey the above equation on a set with nonzero measure. Then $(\tilde{x}_\alpha, \tilde{u}_\alpha, \tilde{v}_\alpha, \bar{w}_\alpha)$ with

$$\bar{w}_\alpha := \frac{1}{\gamma(\alpha)} \max\{0, C(t)\tilde{x}_\alpha(t) - d(t)\}$$

is still admissible but further reduces the objective function value, since $\|\bar{w}_\alpha\|_2^2 < \|\tilde{w}_\alpha\|_2^2$. This contradiction shows the assertion. □

Remark 4.13
In the case $\kappa \equiv 0$, the virtual control regularization is equivalent to using the L_2 penalty term

$$\frac{\phi(\alpha)}{2\gamma^2(\alpha)} \int_{t_0}^{t_f} \max\{0, C(t)x(t) - d(t)\}^2 dt.$$

This follows directly from Lemma 4.12 by inserting the shown representation of $\hat{w}_\alpha$ into the penalty term $\frac{\phi(\alpha)}{2}\|\hat{w}_\alpha\|_2^2$. An advantage of the form of Problem 4.11 is that all data of the problem remain twice differentiable.

A simple connection between the solutions of Problem 4.1 and 4.11 results directly from the shape of the admissible sets, as shown in the following lemma [GHed, cf. Lemma 3]:

Lemma 4.14
Let the matrix W be positive semidefinite almost everywhere in $[t_0, t_f]$. Let $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be an optimal solution for Problem 4.11, and let $(\hat{x}, \hat{u}, \hat{v})$ be optimal for Problem 4.1. Then it holds that

$$\frac{\phi(\alpha)}{2}\|\hat{w}_\alpha\|_2^2 \leq J^{LQP}(\hat{x}, \hat{u}, \hat{v}).$$

If $\phi(\alpha) \geq \delta_\phi$ for some $\delta_\phi > 0$ independent of α, then the optimal virtual control $\hat{w}_\alpha$ remains bounded with respect to the $\|\cdot\|_2$-norm, independent of α.

If $\phi(\alpha) \to \infty$ for $\alpha \to 0$, then $\lim_{\alpha\to 0}\|\hat{w}_\alpha\|_2 = 0$.

Proof.
Let $(\hat{x}, \hat{u}, \hat{v})$ be admissible for Problem 4.1, i.e., let $(\hat{x}, \hat{u}, \hat{v})$ satisfy all constraints. Then $(\hat{x}, \hat{u}, \hat{v}, 0)$ is admissible for Problem 4.11, for any α. Hence,

$$\begin{aligned} J_\alpha^{LQP}(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha) &= \frac{1}{2}\|\hat{x}_\alpha(t_f)\|_{Q_f}^2 + \frac{1}{2}\|(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha)\|_W^2 + \frac{\phi(\alpha)}{2}\|\hat{w}_\alpha\|_2^2 \\ &\leq J^{LQP}(\hat{x}, \hat{u}, \hat{v}). \end{aligned}$$

The rest follows directly from this inequality. □

The error analysis between the optimal solutions for Problem 4.1 and Problem 4.11 relies on the normality of the multipliers. The smoothness assumption, together with appropriate normality conditions ensure the existence of normal multipliers for both problems according to Theorem 3.15, Theorem 3.27 and Corollary 3.29 (compare [GHed, Lemmas 1 and 2]):

Lemma 4.15
Let LQOCP be a problem where the given data satisfy the smoothness assumptions 4.2, and let $(\hat{x}, \hat{u}, \hat{v})$ be a local minimum that satisfies the LQOCP normality conditions 4.3. Furthermore, let $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be a local minimum for $LQOCP_\alpha$, such that there exists a $\delta > 0$ and a constant $C_H > 0$, so that $(H(t)_{I_\delta(t)})^+$ exists with $\|(H(t)_{I_\delta(t)})^+\| \leq C_H$. Then:

4.15.1 *There exist multipliers*

$$\hat{\lambda} \in BV([t_0, t_f], \mathbb{R}^{n_x}), \hat{\eta} \in L^\infty([t_0, t_f], \mathbb{R}^{n_c}), \hat{\mu} \in NBV([t_0, t_f], \mathbb{R}^{n_s}) \text{ and } \hat{\sigma} \in \mathbb{R}^{n_\Psi}$$

that satisfy (4.2)-(4.9) *for the local minimum $(\hat{x}, \hat{u}, \hat{v})$.*

4.15.2 *There exist multipliers*

$$\hat{\lambda}_\alpha \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x}), \hat{\eta}_\alpha \in L^\infty([t_0, t_f], \mathbb{R}^{n_c}), \hat{\nu}_\alpha \in L^\infty([t_0, t_f], \mathbb{R}^{n_s}) \text{ and } \hat{\sigma}_\alpha \in \mathbb{R}^{n_\Psi},$$

such that

$$\dot{\hat{\lambda}}_\alpha(t) = -\left(Q(t)\hat{x}_\alpha(t) + R_u(t)\hat{u}_\alpha(t) + R_v(t)\hat{v}_\alpha(t)\right.$$

$$+A(t)^\top \hat{\lambda}_\alpha(t) + G(t)^\top \hat{\eta}_\alpha(t) + C(t)^\top \hat{\nu}_\alpha(t)\Big) \tag{4.10}$$

$$\hat{\lambda}_\alpha(t_0) = -E_0^\top \hat{\sigma}_\alpha \tag{4.11}$$

$$\hat{\lambda}_\alpha(t_f) = Q_f \hat{x}_\alpha(t_f) + E_1^\top \hat{\sigma}_\alpha$$

$$0 = S_v(t)\hat{v}_\alpha(t) + R_v(t)^\top \hat{x}_\alpha(t) + B_v(t)^\top \hat{\lambda}_\alpha(t) + H(t)^\top \hat{\eta}_\alpha(t) \tag{4.12}$$

$$0 \leq \left(\hat{u}_\alpha(t)^\top S_u(t) + \hat{x}_\alpha(t)^\top R_u(t) + \hat{\lambda}_\alpha(t)^\top B_u(t)\right)(u - \hat{u}_\alpha(t)) \tag{4.13}$$

$$0 = \phi(\alpha)\hat{w}_\alpha(t) - \kappa(\alpha) e_{n_s} e_{n_x}^\top \hat{\lambda}_\alpha(t) - \gamma(\alpha)\hat{\nu}_\alpha(t) \tag{4.14}$$

$$0 \leq \hat{\eta}_\alpha(t) \perp l(t) - G(t)\hat{x}_\alpha(t) - H(t)\hat{v}_\alpha(t) \geq 0 \tag{4.15}$$

$$0 \leq \hat{\nu}_\alpha(t) \perp d(t) - C(t)\hat{x}_\alpha(t) + \gamma(\alpha)\hat{w}_\alpha(t) \geq 0 \tag{4.16}$$

Proof.
Problem 4.11 can be rewritten in the form of Problem 4.1. It remains to show that the normality conditions 4.3 of the original problem are sufficient for the respective normality conditions of the new problem.

The normality conditions 4.3.1 and 4.3.2 are obviously satisfied. A solution for the system in 4.3.3 is given by $x_{\alpha 0} := \frac{1}{2}(\hat{x} + x_0) - \frac{1}{2}\hat{x}_\alpha$, $u_{\alpha 0} := \frac{1}{2}(\hat{u} + u_0) - \frac{1}{2}\hat{u}_\alpha$, $v_{\alpha 0} := \frac{1}{2}(\hat{v} + v_0) - \frac{1}{2}\hat{v}_\alpha$ and $w_{\alpha 0} = 0$, where (x_0, u_0, v_0) with $\hat{u} + u_0 \in \operatorname{int} U_{ad}$ solves the respective system for Problem 4.1, since

$\hat{u}_\alpha + u_0$ is an interior point:[2]

$$\hat{u}_\alpha + u_{\alpha 0} = \frac{1}{2}u_\alpha + \frac{1}{2}(\hat{u} + u_0) \in \operatorname{int}(U_{ad})$$

and

$$G(t)\left(\hat{x}_\alpha(t) + x_{\alpha 0}(t)\right) + H(t)\left(\hat{v}_\alpha(t) + v_{\alpha 0}(t)\right) \leq l(t) - \frac{1}{2}\varepsilon e_{n_c}$$

$$C(t)\left(\hat{x}_\alpha(t) + x_{\alpha 0}(t)\right) = \frac{1}{2}C(t)\hat{x}_\alpha(t) + \frac{1}{2}C(t)\left(\hat{x}(t) + x_0(t)\right) < d(t),$$

which implies

$$C(t)\left(\hat{x}_\alpha(t) + x_{\alpha 0}(t)\right) \leq d(t) - \varepsilon_2 e_{n_s}$$

for some $\varepsilon > 0$, since the left hand side is continuous.

Due to the linearity of the differential equations for $\hat{x}$, $\hat{x}_\alpha$ and x_0, it holds

$$\dot{x}_{\alpha 0}(t) = A(t)x_{\alpha 0}(t) + B(t)\begin{pmatrix} u_{\alpha 0} \\ v_{\alpha 0} \end{pmatrix}$$

and

$$\begin{aligned} E_0 x_{\alpha 0}(t_0) + E_1 x_{\alpha 0}(t_f) = \; & E_0\left(\frac{1}{2}(\hat{x}(t_0) + x_0(t_0)) - \frac{1}{2}\hat{x}_\alpha(t_0)\right) \\ & + E_1\left(\frac{1}{2}(\hat{x}(t_f) + x_0(t_f)) - \frac{1}{2}\hat{x}_\alpha(t_f)\right) \end{aligned}$$

[2]This argument is explained in more detail in Lemma A.1 in the appendix.

$$
\begin{aligned}
&= \frac{1}{2}\left(E_0 x_0(t_0) + E_1 x_0(t_f)\right) \\
&\quad + \frac{1}{2}\left(E_0 \hat{x}(t_0) + E_1 \hat{x}(t_f)\right) \\
&\quad - \frac{1}{2}\left(E_0 \hat{x}_\alpha(t_0) + E_1 \hat{x}_\alpha(t_f)\right) \\
&= 0. \qquad \square
\end{aligned}
$$

Lemma 4.15 can be used to derive a first estimation for the deviation $\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2$, which leads to the main result of this section:

Lemma 4.16
Let $(\hat{x}, \hat{u}, \hat{v})$ and $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be local minima of Problems 4.1 and 4.11, respectively, that satisfy the assumptions of Lemma 4.15. Then the following estimation holds:

$$
\begin{aligned}
\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 \leq & - \int_{t_0}^{t_f} (\hat{\eta}_\alpha - \hat{\eta})^\top \left(G(\hat{x}_\alpha - \hat{x}) + H(\hat{\eta}_\alpha - \hat{\eta})\right) dt \\
& - \kappa(\alpha) \int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top e_{n_x} e_{n_s}^\top \hat{w}_\alpha dt \\
& - \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top \hat{\nu}_\alpha dt + \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top d\hat{\mu}(t)
\end{aligned}
$$

Proof.
As in [GHed], the dependence of all functions on t is omitted in this proof.

Partial integration for the Stieltjes integral (cf. Lemma 2.12) yields

$$
\begin{aligned}
& \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top d(\hat{\lambda}_\alpha - \hat{\lambda}) + \int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top d(\hat{x}_\alpha - \hat{x}) \qquad (4.17) \\
= \quad & (\hat{x}_\alpha(t_f) - \hat{x}(t_f))^\top (\hat{\lambda}_\alpha(t_f) - \hat{\lambda}(t_f)) - (\hat{x}_\alpha(t_0) - \hat{x}(t_0))^\top (\hat{\lambda}_\alpha(t_0) - \hat{\lambda}(t_0)) \\
\overset{(4.3),(4.11)}{=} & (\hat{x}_\alpha(t_f) - \hat{x}(t_f))^\top (E_1^\top \hat{\sigma}_\alpha - E_1^\top \hat{\sigma} + Q_f \hat{x}_\alpha(t_f) - Q_f \hat{x}(t_f)) \\
& + (\hat{x}_\alpha(t_0) - \hat{x}(t_0))^\top (E_0^\top \hat{\sigma}_\alpha - E_0^\top \hat{\sigma}) \\
= \quad & (E_0 \hat{x}_\alpha(t_0) + E_1 \hat{x}_\alpha(t_f))^\top (\hat{\sigma}_\alpha - \hat{\sigma}) \\
& - (E_0 \hat{x}(t_0) + E_1 \hat{x}(t_f))^\top (\hat{\sigma}_\alpha - \hat{\sigma}) \\
& + (\hat{x}_\alpha(t_f) - \hat{x}(t_f))^\top Q_f (\hat{x}_\alpha(t_f) - \hat{x}(t_f)) \\
= \quad & f^\top (\hat{\sigma}_\alpha - \hat{\sigma}) - f^\top (\hat{\sigma}_\alpha - \hat{\sigma}) + (\hat{x}_\alpha(t_f) - \hat{x}(t_f))^\top Q_f (\hat{x}_\alpha(t_f) - \hat{x}(t_f)) \\
\geq \quad & 0.
\end{aligned}
$$

Applying equations (4.2) and (4.10) on the terms of the left hand side shows that

$$
\begin{aligned}
& \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top d(\hat{\lambda}_\alpha - \hat{\lambda}) \\
= & \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top (-Q(\hat{x}_\alpha - \hat{x}) - R_u(\hat{u}_\alpha - \hat{u}) - R_v(\hat{v}_\alpha - \hat{v}) - A^\top (\hat{\lambda}_\alpha - \hat{\lambda}) - G^\top (\hat{\eta}_\alpha - \hat{\eta})) dt
\end{aligned}
$$

$$- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top \hat{\nu}_\alpha dt + \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top d\hat{\mu}(t)$$

and

$$\begin{aligned}
&\int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top d(\hat{x}_\alpha - \hat{x}) \\
&= \int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top \left(A(\hat{x}_\alpha - \hat{x}) + B_u(\hat{u}_\alpha - \hat{u}) + B_v(\hat{v}_\alpha - \hat{v}) - \kappa(\alpha) e_{n_x} e_{n_s}^\top \hat{w}_\alpha \right) dt.
\end{aligned}$$

Summarizing these results, (4.17) becomes

$$\begin{aligned}
0 \leq &- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top \left(Q(\hat{x}_\alpha - \hat{x}) + R_u(\hat{u}_\alpha - \hat{u}) + R_v(\hat{v}_\alpha - \hat{v}) + G^\top(\hat{\eta}_\alpha - \hat{\eta}) \right) dt \\
&+ \int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top \left(B_u(\hat{u}_\alpha - \hat{u}) + B_v(\hat{v}_\alpha - \hat{v}) - \kappa(\alpha) e_{n_x} e_{n_s}^\top \hat{w}_\alpha \right) dt \\
&- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top \hat{\nu}_\alpha dt + \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top d\hat{\mu}(t),
\end{aligned}$$

which can be solved for $\int_{t_0}^{t_f} (\hat{\lambda} - \hat{\lambda}_\alpha)^\top B_u(\hat{u}_\alpha - \hat{u}) dt$:

$$\begin{aligned}
&\int_{t_0}^{t_f} (\hat{\lambda} - \hat{\lambda}_\alpha)^\top B_u(\hat{u}_\alpha - \hat{u}) dt \qquad (4.18) \\
\leq &- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top \left(Q(\hat{x}_\alpha - \hat{x}) + R_u(\hat{u}_\alpha - \hat{u}) + R_v(\hat{v}_\alpha - \hat{v}) + G^\top(\hat{\eta}_\alpha - \hat{\eta}) \right) dt \\
&+ \int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top \left(B_v(\hat{v}_\alpha - \hat{v}) - \kappa(\alpha) e_{n_x} e_{n_s}^\top \hat{w}_\alpha \right) dt \\
&- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top \hat{\nu}_\alpha dt + \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top d\hat{\mu}(t).
\end{aligned}$$

From the optimality conditions for $\hat{v}$, (4.5), and $\hat{v}_\alpha$, (4.12), we derive

$$0 = (\hat{v}_\alpha - \hat{v})^\top S_v^\top + (\hat{x}_\alpha - \hat{x})^\top R_v + (\hat{\lambda}_\alpha - \hat{\lambda})^\top B_v + (\hat{\eta}_\alpha - \hat{\eta})^\top H,$$

so (4.18) becomes

$$\begin{aligned}
&\int_{t_0}^{t_f} (\hat{\lambda} - \hat{\lambda}_\alpha)^\top B_u(\hat{u}_\alpha - \hat{u}) dt \qquad (4.19) \\
\leq &- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top \left(Q(\hat{x}_\alpha - \hat{x}) + R_u(\hat{u}_\alpha - \hat{u}) + 2R_v(\hat{v}_\alpha - \hat{v}) + G^\top(\hat{\eta}_\alpha - \hat{\eta}) \right) dt \\
&- \int_{t_0}^{t_f} (\hat{v}_\alpha - \hat{v})^\top \left(S_v(\hat{v}_\alpha - \hat{v}) + H^\top(\hat{\eta}_\alpha - \hat{\eta}) \right) dt \\
&- \kappa(\alpha) \int_{t_0}^{t_f} (\hat{\lambda}_\alpha - \hat{\lambda})^\top e_{n_x} e_{n_s}^\top \hat{w}_\alpha dt \\
&- \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top \hat{\nu}_\alpha dt + \int_{t_0}^{t_f} (\hat{x}_\alpha - \hat{x})^\top C^\top d\hat{\mu}(t).
\end{aligned}$$

Inserting the optimal control $\hat{u}_\alpha$ for Problem 4.11 into the optimality condition (4.6) for Problem 4.1 and vice versa for inequality (4.13) yields

$$0 \leq \int_{t_0}^{t_f} \left(S_u^\top(\hat{u} - \hat{u}_\alpha) + R_u^\top(\hat{x} - \hat{x}_\alpha) + B_u^\top(\hat{\lambda} - \hat{\lambda}_\alpha) \right)^\top (\hat{u}_\alpha - \hat{u}) dt.$$

Substituting $\int_{t_0}^{t_f}(\hat{\lambda}-\hat{\lambda}_\alpha)^\top B_u(\hat{u}_\alpha-\hat{u})dt$ in this expression according to (4.19) leads to

$$\begin{aligned}
&\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top Q(\hat{x}_\alpha-\hat{x})+2(\hat{x}_\alpha-\hat{x})^\top R_u(\hat{u}_\alpha-\hat{u})+2(\hat{x}_\alpha-\hat{x})^\top R_v(\hat{v}_\alpha-\hat{v})\\
&+(\hat{u}_\alpha-\hat{u})^\top S_u(\hat{u}_\alpha-\hat{u})+(\hat{v}_\alpha-\hat{v})^\top S_v(\hat{v}_\alpha-\hat{v})dt\\
\leq&-\int_{t_0}^{t_f}(\hat{\eta}_\alpha-\hat{\eta})^\top\left(G(\hat{x}_\alpha-\hat{x})+H(\hat{v}_\alpha-\hat{v})\right)dt\\
&-\kappa(\alpha)\int_{t_0}^{t_f}(\hat{\lambda}_\alpha-\hat{\lambda})^\top e_{n_x}e_{n_s}^\top\hat{w}_\alpha dt\\
&-\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top C^\top\hat{\nu}_\alpha dt+\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top C^\top d\hat{\mu}(t).
\end{aligned}$$

This proves the assertion of Lemma 4.16. □

Analogue to [GHed, Theorem 2], we can further simplify the right hand side of the estimation in Lemma 4.16, using the complementarity conditions (4.7), (4.8), (4.15) and (4.16).

Lemma 4.17
Let $(\hat{x},\hat{u},\hat{v})$ and $(\hat{x}_\alpha,\hat{u}_\alpha,\hat{v}_\alpha,\hat{w}_\alpha)$ be defined as in Lemma 4.16. Then then following inequalities hold:

$$\begin{aligned}
&-\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top C^\top\hat{\nu}_\alpha dt+\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top C^\top d\hat{\mu}(t)-\kappa(\alpha)\int_{t_0}^{t_f}\hat{\lambda}_\alpha^\top e_{n_x}e_{n_s}^\top\hat{w}_\alpha dt\\
&\leq-\phi(\alpha)\left\|\hat{w}_\alpha\right\|_2^2+\gamma(\alpha)\int_{t_0}^{t_f}\hat{w}_\alpha^\top d\hat{\mu}(t),
\end{aligned}\tag{4.20}$$

$$0\geq-\int_{t_0}^{t_f}(\hat{\eta}_\alpha-\hat{\eta})^\top\left(G(\hat{x}_\alpha-\hat{x})+H(\hat{v}_\alpha-\hat{v})\right)dt\tag{4.21}$$

Proof.
The complementarity conditions (4.7), (4.8), (4.15) and (4.16), together with the optimality condition (4.14) yield:

(4.20):

$$\begin{aligned}
&-\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top C^\top\hat{\nu}_\alpha dt+\int_{t_0}^{t_f}(\hat{x}_\alpha-\hat{x})^\top C^\top d\hat{\mu}(t)-\kappa(\alpha)\int_{t_0}^{t_f}\hat{\lambda}_\alpha^\top e_{n_x}e_{n_s}^\top\hat{w}_\alpha dt\\
=&-\int_{t_0}^{t_f}\underbrace{(C\hat{x}_\alpha-\gamma(\alpha)\hat{w}_\alpha-d)^\top\hat{\nu}_\alpha}_{\overset{(4.16)}{=}0}dt+\int_{t_0}^{t_f}\underbrace{(C\hat{x}-d)^\top}_{\leq0}\underbrace{\hat{\nu}_\alpha}_{\geq0}dt\\
&+\underbrace{\int_{t_0}^{t_f}(C\hat{x}_\alpha-\gamma(\alpha)\hat{w}_\alpha-d)^\top d\hat{\mu}(t)}_{\leq0,\quad\hat{w}_\alpha\text{ continuous}}-\underbrace{\int_{t_0}^{t_f}(C\hat{x}-d)^\top d\hat{\mu}(t)}_{\overset{(4.9)}{=}0}\\
&-\kappa(\alpha)\int_{t_0}^{t_f}\hat{\lambda}_\alpha^\top e_{n_x}e_{n_s}^\top\hat{w}_\alpha dt-\gamma(\alpha)\int_{t_0}^{t_f}\hat{w}_\alpha^\top\hat{\nu}_\alpha dt+\gamma(\alpha)\int_{t_0}^{t_f}\hat{w}_\alpha^\top d\hat{\mu}(t)\\
\leq&-\kappa(\alpha)\int_{t_0}^{t_f}\hat{\lambda}_\alpha^\top e_{n_x}e_{n_s}^\top\hat{w}_\alpha dt-\gamma(\alpha)\int_{t_0}^{t_f}\hat{w}_\alpha^\top\hat{\nu}_\alpha dt+\gamma(\alpha)\int_{t_0}^{t_f}\hat{w}_\alpha^\top d\hat{\mu}(t)\\
\overset{(4.14)}{=}&-\phi(\alpha)\left\|\hat{w}_\alpha\right\|_2^2+\gamma(\alpha)\int_{t_0}^{t_f}\hat{w}_\alpha^\top d\hat{\mu}(t),
\end{aligned}$$

(4.21):

$$\begin{aligned}
&-\int_{t_0}^{t_f} (\hat{\eta}_\alpha - \hat{\eta})^\top \left(G(\hat{x}_\alpha - \hat{x}) + H(\hat{v}_\alpha - \hat{v})\right) dt \\
=&-\int_{t_0}^{t_f} \underbrace{\hat{\eta}_\alpha^\top (G\hat{x}_\alpha + H\hat{v}_\alpha - l)}_{\overset{(4.15)}{=}0} dt - \int_{t_0}^{t_f} \underbrace{\hat{\eta}^\top (G\hat{x} + H\hat{v} - l)}_{\overset{(4.7)}{=}0} dt \\
&+\int_{t_0}^{t_f} \underbrace{\hat{\eta}_\alpha^\top}_{\geq 0} \underbrace{(G\hat{x} + H\hat{v} - l)}_{\leq 0} dt + \int_{t_0}^{t_f} \underbrace{\hat{\eta}^\top}_{\geq 0} \underbrace{(G\hat{x}_\alpha + H\hat{v}_\alpha - l)}_{\leq 0} dt \\
\leq& 0.
\end{aligned}$$

□

Theorem 4.18 sums up the results from Lemma 4.16 and Lemma 4.17:

Theorem 4.18
Let $(\hat{x}, \hat{u}, \hat{v})$ and $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be local minima of the Problems 4.1 and 4.11, respectively, that satisfy the assumptions of Lemma 4.15 for any $\alpha > 0$. Then it holds:

$$\begin{aligned}
&\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 + \phi(\alpha)\|\hat{w}_\alpha\|_2^2 \\
&\quad \leq \gamma(\alpha) \int_{t_0}^{t_f} \hat{w}_\alpha^\top d\hat{\mu}(t) + \kappa(\alpha) \int_{t_0}^{t_f} \hat{\lambda}^\top e_{n_x} e_{n_s}^\top \hat{w}_\alpha dt.
\end{aligned}$$

The conclusions that can be drawn from Theorem 4.18 depend on the smoothness of $\hat{\mu}$ and the problem data. The simplest case is $\hat{\mu} \in W^{1,2}$:

Theorem 4.19
Let $(\hat{x}, \hat{u}, \hat{v})$ and $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be defined as in Theorem 4.18, and assume that $\hat{\mu} \in W^{1,2}([t_0, t_f], \mathbb{R}^{n_s})$.

If $\lim_{\alpha \to 0} \frac{\gamma(\alpha)+\kappa(\alpha)}{\phi(\alpha)} = 0$, then $\lim_{\alpha \to 0} \|\hat{w}_\alpha\|_2 = 0$. If additionally $\kappa(\alpha) \leq C_R$ and $\gamma(\alpha) \leq C_R$ for some $C_R \in \mathbb{R}$, independent of α, then $\lim_{\alpha \to 0} \|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 = 0$.

Proof.
According to Theorem 4.18, it holds that

$$\begin{aligned}
&\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 + \phi(\alpha)\|\hat{w}_\alpha\|_2^2 \\
&\quad \leq \gamma(\alpha) \int_{t_0}^{t_f} \hat{w}_\alpha^\top d\hat{\mu}(t) + \kappa(\alpha) \int_{t_0}^{t_f} \hat{\lambda}^\top e_{n_x} e_{n_s}^\top \hat{w}_\alpha dt \\
&\quad \leq \gamma(\alpha) \int_{t_0}^{t_f} \hat{w}_\alpha^\top \dot{\hat{\mu}} dt + \kappa(\alpha) \int_{t_0}^{t_f} C_\lambda e_{n_s}^\top \hat{w}_\alpha dt \\
&\quad \leq \gamma(\alpha)\|\hat{w}_\alpha\|_2 \|\dot{\hat{\mu}}\|_2 + \kappa(\alpha) C_\lambda \|\hat{w}_\alpha\|_2 \\
&\quad \leq C_M \|\hat{w}_\alpha\|_2 (\gamma(\alpha) + \kappa(\alpha)),
\end{aligned}$$

which particularly implies

$$\phi(\alpha)\|\hat{w}_\alpha\|_2^2 \leq C_M \|\hat{w}_\alpha\|_2 (\gamma(\alpha) + \kappa(\alpha)),$$

and therefore

$$\|\hat{w}_\alpha\|_2 \leq C_M \frac{\gamma(\alpha)+\kappa(\alpha)}{\phi(\alpha)},$$

which proves the first assertion.

If $\kappa(\alpha) \leq C_R$ and $\gamma(\alpha) \leq C_R$, then

$$\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 \leq 2 \cdot C_M \cdot C_R \cdot \|\hat{w}_\alpha\|_2$$

according to the first inequality. Since $\|\hat{w}_\alpha\|_2 \to 0$ for $\alpha \to 0$, this completes the proof. □

If μ does not comply with the smoothness assumption, then it can be replaced by the assumption that $\hat{w}_\alpha$ remains bounded with respect to the $\|\cdot\|_\infty$ norm.

Theorem 4.20
Let $(\hat{x}, \hat{u}, \hat{v})$ and $(\hat{x}_\alpha, \hat{u}_\alpha, \hat{v}_\alpha, \hat{w}_\alpha)$ be defined as in Theorem 4.18. Assume that $\|\hat{w}_\alpha\|_\infty \leq C_{w\infty}$ for some constant $C_{w\infty}$.

If $\lim_{\alpha\to 0} \frac{\gamma(\alpha)+\kappa(\alpha)}{\phi(\alpha)} = 0$, then $\lim_{\alpha\to 0} \|\hat{w}_\alpha\|_2 = 0$. If additionally $\lim_{\alpha\to 0} \kappa(\alpha) = 0$ and $\lim_{\alpha\to 0} \gamma(\alpha) = 0$, then $\lim_{\alpha\to 0} \|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 = 0$.

Proof.
Theorem 4.18 again assures that

$$\begin{aligned}
&\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 + \phi(\alpha)\|\hat{w}_\alpha\|_2^2 \\
&\quad\leq \gamma(\alpha) \int_{t_0}^{t_f} \hat{w}_\alpha^\top d\hat{\mu}(t) + \kappa(\alpha) \int_{t_0}^{t_f} \hat{\lambda}^\top e_{n_x} e_{n_s}^\top \hat{w}_\alpha dt \\
&\quad\leq \gamma(\alpha) \cdot \|\hat{w}_\alpha\|_\infty \cdot TV(\hat{\mu}, [t_0, t_f]) + \kappa(\alpha) C_\lambda \|\hat{w}_\alpha\|_2 \\
&\quad\leq \gamma(\alpha) \cdot \|\hat{w}_\alpha\|_\infty \cdot TV(\hat{\mu}, [t_0, t_f]) + \kappa(\alpha) C_\lambda \|\hat{w}_\alpha\|_\infty \\
&\quad\leq C(\gamma(\alpha) + \kappa(\alpha)).
\end{aligned}$$

The first assertion follows analogly to the proof of Theorem 4.19, as the right hand side in $\|\hat{w}_\alpha\|_2^2 \leq C \frac{\gamma(\alpha)+\kappa(\alpha)}{\phi(\alpha)}$ vanishes. The second assertion follows directly from

$$\|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_W^2 \leq C(\gamma(\alpha) + \kappa(\alpha)).$$ □

Remark 4.21
Theorems 4.19 and 4.20 yield convergence properties in the $\|\cdot\|_W$ (half) norm. Consequently, if W is uniformly positive definite, i.e. $(x,u,v)^\top W(t)(x,u,v) \geq \delta\|(x,u,v)^\top\|_2^2$ for any vector $(x,u,v) \in \mathbb{R}^{n_x+n_u+n_v}$, independent of $t \in [t_0, t_f]$, then

$$\lim_{\alpha\to 0} \|(\hat{x}_\alpha - \hat{x}, \hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_2 = 0.$$

Otherwise, if W is only positive semidefinite with either

- *$(x,u,v)^\top W(t)(x,u,v) \geq \delta\|x\|_2^2$, then $\lim_{\alpha\to 0} \|\hat{x}_\alpha - \hat{x}\|_2 = 0$, that is, the state deviation vanishes (with respect to the $\|\cdot\|_2$ norm) for decreasing α, or*
- *$(x,u,v)^\top W(t)(x,u,v) \geq \delta\|(u,v)\|_2^2$, then $\lim_{\alpha\to 0} \|(\hat{u}_\alpha - \hat{u}, \hat{v}_\alpha - \hat{v})\|_2 = 0$, and the control deviation vanishes (again with respect to the $\|\cdot\|_2$ norm).*

4.3. Examples

The solutions to the following examples have been calculated using the combined Newton method described in chapter 5. Numerical results based on the globalized semismooth Newton method (cf. [Ger08]) have been presented in [GHed]. The parameter functions were set to

$$\kappa(\alpha) := 0, \qquad \varphi(\alpha) := 1, \qquad \gamma(\alpha) := \alpha.$$

For both problems, the solutions were calculated on an equidistant grid consisting of 501 grid points. This fineness has been chosen in order to eliminate effects that originate from the discretization.

4.3.1. Minimum Energy Problem

The Minimum Energy Problem (cf. [BH75, p. 120], [GHed]) is a linear quadratic optimal control problem with a second order state constraint.

The task is to find the form of a homogenous stick under tension. Both ends of the stick are attached to the ground in a given angle of $\pi/4$. The time coordinate t in this case represents the first space dimension in which the stick expands. The height of the stick at a given point in the first space dimension is represented by the first state x. Its derivative, the slope of the stick, is the second state y. The fact that the stick is attached to the ground at the two points $t_0 = 0$ and $t_f = 1$ translates to the boundary conditions $x(0) = x(1) = 0$, and the angle conditions can be expressed as $y(0) = -y(1) = \tan(\pi/4)$.

The control u describes the derivative of the slope y, i.e. the bend of the stick. The integral of its square is to be minimized. The state constraint models a height restriction. The linear quadratic optimal control problem reads:

Problem 4.22 (Minimum Energy Problem)

$$\text{min!} \quad \frac{1}{2}\int_0^1 u(t)^2 dt$$

subject to

$$\begin{aligned} \dot{x}(t) &= y(t), & x(0) &= \quad x(1) = 0, \\ \dot{y}(t) &= u(t), & y(0) &= -y(1) = 1 \end{aligned}$$

and

$$x(t) \leq \frac{1}{9}.$$

The regularized optimal control problem with regularization parameters as above reads

α	$\|w_\alpha\|_2$	$\|w_\alpha\|_\infty$	$\|F(z)\|_{Y^\infty}$
$1E-1$	$7.312E-01$	$1.158E+00$	$1.903E-06$
$1E-2$	$4.433E-01$	$7.216E-01$	$2.234E-06$
$1E-3$	$8.508E-02$	$2.684E-01$	$3.908E-07$
$1E-4$	$1.541E-02$	$8.687E-02$	$1.129E-06$
$1E-5$	$2.714E-03$	$2.682E-02$	$3.640E-07$
$1E-6$	$3.929E-04$	$5.254E-03$	$3.343E-06$
$1E-7$	$4.233E-05$	$6.073E-04$	$6.103E-06$

Table 4.1.: Norms of w_α for the Minimum Energy Problem

Problem 4.23 (Regularized Minimum Energy Problem)

$$\text{min!} \quad \frac{1}{2}\int_0^1 u(t)^2 dt + \frac{1}{2}\int_0^1 w_\alpha(t)^2 dt$$

subject to

$$\begin{aligned} \dot{x}(t) &= y(t) & x(0) &= \quad x(1) = 0 \\ \dot{y}(t) &= u(t) & y(0) &= -y(1) = 1 \end{aligned}$$

and

$$x(t) - \alpha \cdot w_\alpha(t) \leq \frac{1}{9}.$$

Figure 4.1 shows the plots of solutions to the regularized problem for different values of α. Additionally, table 4.1 lists the norms of the virtual control as well as the residua $\|F(z)\|_\infty$ of the calculations.

Both Table 4.1 and Figure 4.1 confirm the convergence results of Theorem 4.20 and show that the virtual control vanishes even in the $\|\cdot\|_\infty$-norm. At the same time, the multiplier η_α explodes. This effect is due to the fact that η_α is an approximation for $\dot{\mu}$ in the original problem, and μ is piecewise continuous.

On the other hand, this example shows that the uniform convergence (i.e. convergence in the L^∞ sense) result cannot be transferred to the multipliers: The first adjoints λ_1 of the regularized problems are continuous. Uniform convergence of continuous functions would imply that their limit function, which is the adjoint of the original problem, was continuous, which is not the case.

4.3.2. Simplified Trolley Problem

This example is a simplified model of a trolley crane (cf. [Kim02, p. 18], [GHed]). A weight is attached to the crane by means of a string. The task in this example is to carry the weight over a unified distance. The first state x_1 represents position of the trolley on the track, and x_2 is its velocity. The acceleration of the trolley is the control variable u. The

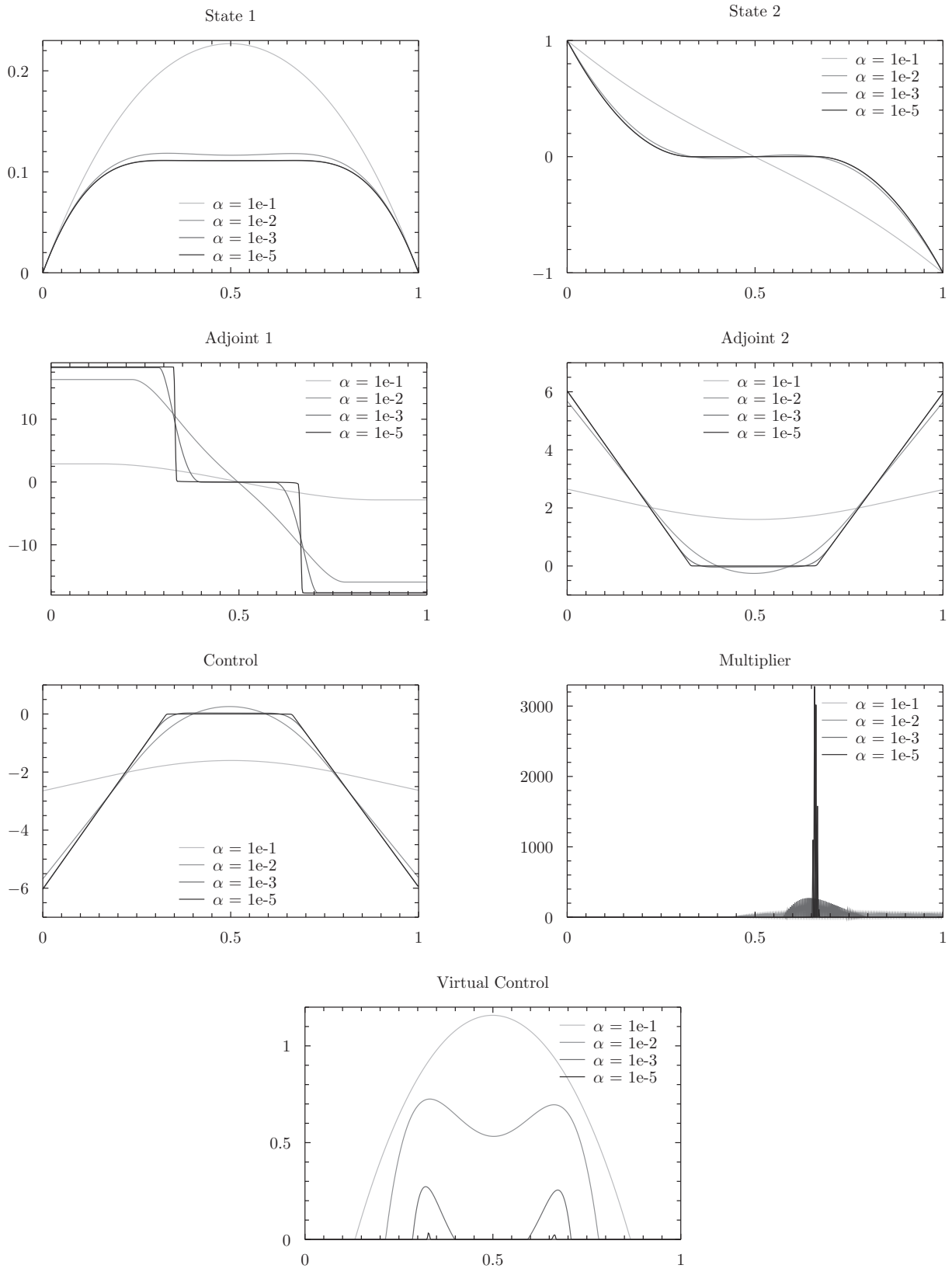


Figure 4.1.: Solutions of LQR_α for the Minimum Energy Problem for different values of α

displacement of the weight is described by x_3, and its derivative x_4 is also influenced by the acceleration u of the trolley.

Here, we assume that the weight does not itself accelerate the trolley. This assumption is justifiable for small attached weights (or heavy trolleys, respectively), and it has the advantage that the model together with an adequate objective function results in a linear quadratic optimal controls problem. In chapter 7, a more sophisticated model that does not fit in this form is investigated. The numbers used at this point are not fitted to the physical interpretation, but are chosen to construct a problem with a second order state constraint[3]. This example is stated in the following

Problem 4.24 (Simplified Trolley Problem)

$$\text{min!} \quad \frac{1}{2}\int_0^1 \|x(t)\|_2^2 + \|u(t)\|_2^2 dt$$

subject to

$$\begin{aligned}
&\dot{x}_1(t) = x_2(t), && x_1(0) = 0, && x_1(1) = 1,\\
&\dot{x}_2(t) = u(t), && x_2(0) = 0, && x_2(1) = 0,\\
&\dot{x}_3(t) = x_4(t), && x_3(0) = 0, && x_3(1) = 0,\\
&\dot{x}_4(t) = u(t) - x_3(t), && x_4(0) = 0, && x_4(1) = 0
\end{aligned}$$

and

$$x_1(t) \leq 5.$$

This leads to the regularized version of the Trolley Problem:

Problem 4.25 (Regularized Simplified Trolley Problem)

$$\text{min!} \quad \frac{1}{2}\int_0^1 \|x(t)\|_2^2 + \|u(t)\|_2^2 dt + \frac{1}{2}\int_0^1 w_\alpha(t)^2 dt$$

subject to

$$\begin{aligned}
&\dot{x}_1(t) = x_2(t), && x_1(0) = 0, && x_1(1) = 1,\\
&\dot{x}_2(t) = u(t), && x_2(0) = 0, && x_2(1) = 0,\\
&\dot{x}_3(t) = x_4(t), && x_3(0) = 0, && x_3(1) = 0,\\
&\dot{x}_4(t) = u(t) - x_3(t), && x_4(0) = 0, && x_4(0) = 0
\end{aligned}$$

and

$$x_1(t) - \alpha \cdot w_\alpha(t) \leq 5.$$

[3] In this case, a bound on the first state corresponds to the requirement that the trolley should not move further than a given mark. For physically justifiable numbers however, the first state can be expected to stay inside the interval $[0, 1]$. Therefore, the small end time in this example enforces an extreme movement of the trolley, so that the second order state constraint becomes active.

α	$\|w_\alpha\|_2$	$\|w_\alpha\|_\infty$	$\|F(z)\|_{Y^\infty}$
$1E-1$	$9.358E+00$	$2.602E+01$	$2.714E-08$
$1E-2$	$8.384E+01$	$2.321E+02$	$2.434E-10$
$1E-3$	$7.935E+01$	$2.244E+02$	$3.296E-08$
$1E-4$	$1.161E+01$	$6.218E+01$	$4.223E-08$
$1E-5$	$1.416E+00$	$1.068E+01$	$5.511E-06$
$1E-6$	$1.420E-01$	$1.074E+00$	$9.424E-06$
$1E-7$	$1.420E-02$	$1.074E-01$	$4.799e-06$

Table 4.2.: Norms of w_α for the Simplified Trolley Problem

Again, the plots in figure 4.2 as well as the values in table 4.2 show that the virtual control vanishes in both the $\|\cdot\|_2$-norm and the $\|\cdot\|_\infty$-norm. The numerical difficulties in this example do not allow for the solutions to the regularized problem with $\alpha = 1E-7$ to be calculated using the combined Newton method as the condition of the matrix for the linear equation becomes too large.

This example illustrates that the convergence is not necessarily uniform: Note that in all plots, the solutions for $\alpha = 10^{-2}$ and $\alpha = 10^{-3}$ are separated by a huge gap that does not occur between e.g. $\alpha = 10^{-1}$ and $\alpha = 10^{-2}$.

As in the Minimum Energy Problem, the multiplier η_α explodes for vanishing values of the parameter α.

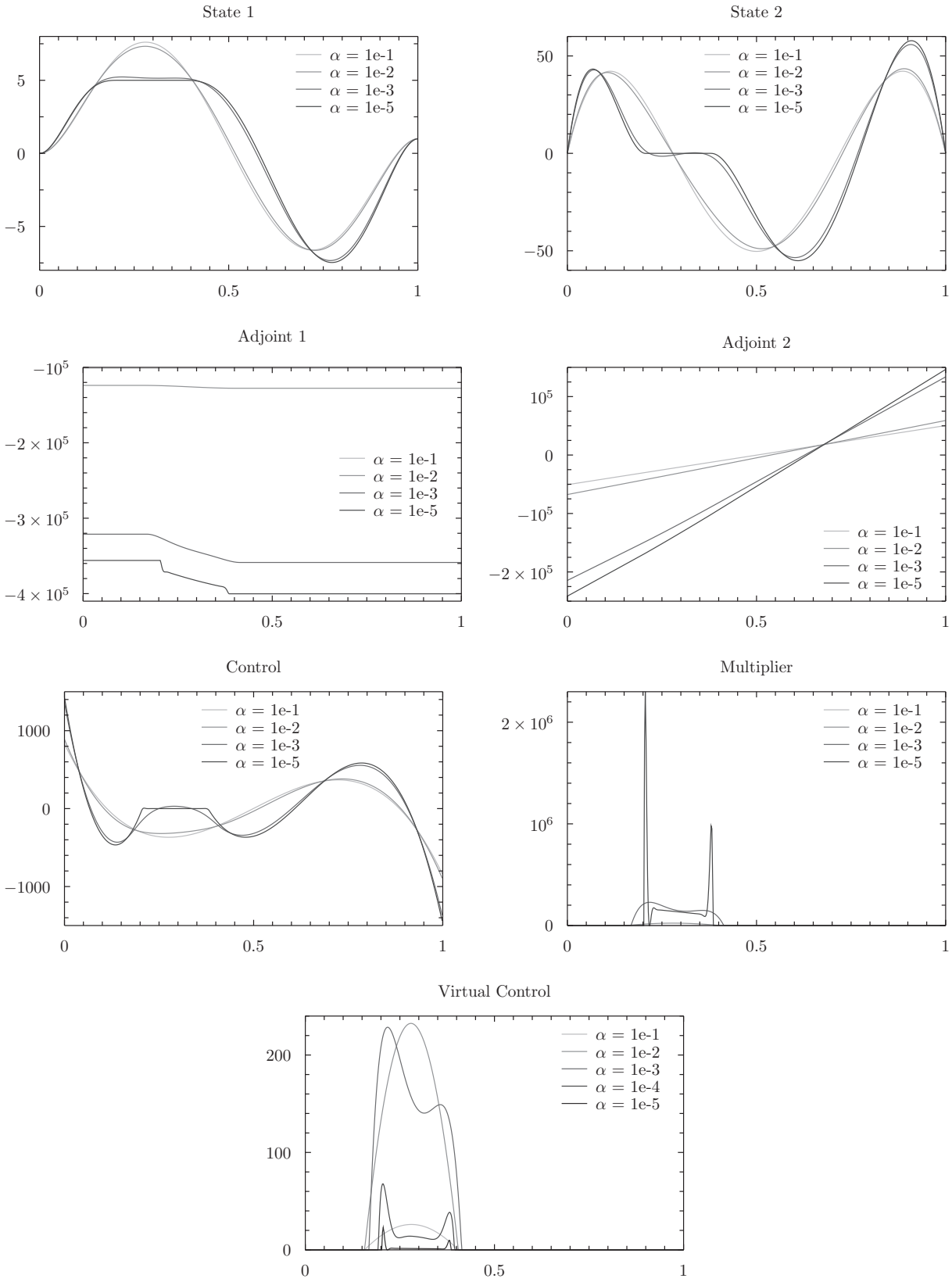


Figure 4.2.: Solutions of LQR_α for the Simplified Trolley Problem for different values of α

5. Solving Optimal Control Problems

In this chapter, we address the problem of finding numerical solutions to the Optimal Control Problem. The first goal is the introduction of a local Newton method in appropriate function spaces. It can be used for finding a solution to the necessary optimality conditions from chapter 3. As a direct application however leads to theoretical problems, we start by developing a regularization for the complementarity problem. We conclude the chapter proposing two globalization approaches for the method.

5.1. Regularizing the Complementarity Problem

In this chapter, we introduce an algorithm for solving linear quadratic optimal control problems of the form 4.1 numerically, with the restriction that we assume $n_u = n_s = 0$. Applying the necessary optimality conditions from corollary 4.7, the problem of finding an optimal control and state trajectory is transformed to solving a complementarity problem. In finite dimensions, several approaches for this class of problems have been investigated, see [FK98] for a survey.

Problem 5.1 (LQOCP$_s$)

$$\begin{aligned} \min! \quad J^{LQP_s}(x,v) :=& \frac{1}{2} x(t_f)^\top Q_f x(t_f) \\ &+ \frac{1}{2}\int_{t_0}^{t_f} \left(x(t)^\top, v(t)^\top\right) \begin{pmatrix} Q(t) & R_v(t) \\ R_v(t)^\top & S(t) \end{pmatrix} \begin{pmatrix} x(t) \\ v(t) \end{pmatrix} dt \end{aligned}$$

with respect to the state function $x \in W^{1,\infty}([t_0,t_f],\mathbb{R}^{n_x})$
and the control function $v \in L^\infty([t_0,t_f],\mathbb{R}^{n_v})$

subject to the differential equation

$$\dot{x}(t) = A(t)x(t) + B_v(t)v(t) \quad a.e.\ in\ [t_0,t_f],$$

boundary conditions

$$E_0 x(t_0) + E_1 x(t_f) = f$$

and mixed control state constraints

$$G(t)x(t) + H(t)v(t) \leq l(t).$$

A plausible approach for solving this problem is to transform the necessary optimality conditions from Corollary 4.7 into an operator equation which then can be solved using numerical methods (in this case, we choose the Newton method). Since state constraints (of higher order than one) lead to the multiplier λ merely being of bounded variation, which in turn is intricate to model numerically, we will assume that no pure state constraints are involved, i.e. $n_s = 0$. In the previous chapter, a method for regularizing state constrained problems was introduced, so that problems with state constraints can be approximated. We assume that no control set constraints occur, i.e. $n_u = 0$, also for computational reasons.

In order to find an equation equivalent to 4.7, the complementarity conditions (4.7) can be transferred using an NCP function:

Definition 5.2 (NCP function)
An NCP function *is a function $\varrho : \mathbb{R}^2 \to \mathbb{R}$, satisfying*

$$\varrho(a,b) = 0 \quad \Longleftrightarrow \quad a \geq 0 \ \wedge \ b \geq 0 \ \wedge \ ab = 0.$$

The following are examples of NCP functions.

1. The min function $\varrho_{\min}$: $\varrho_{\min}(a,b) := \min(a,b)$
2. The Fischer-Burmeister function ϱ_{FB}: $\varrho_{\mathrm{FB}}(a,b) := \sqrt{a^2+b^2} - a - b$

The function $\varrho_{\min}$ was investigated in e.g. [HIK03] (more precisely, the equivalent max-function $\varrho_{\max}(a,b) = a - \max(0, a-b) = \min(a,b) = \varrho_{\min}(a,b)$ was analyzed).

Given an NCP function ϱ, we define the NCP operator ω (induced by ϱ):

Definition 5.3 (NCP Operator)
For an NCP function ϱ, the NCP operator ω for an LQP on the interval $[0,t_f]$ with n_c mixed control-state constraints is defined as the mapping

$$\omega : (L^\infty([0,t_f],\mathbb{R}^{n_c}))^2 \to L^\infty([0,t_f],\mathbb{R}^{n_c}), \qquad (a,b)(\cdot) \mapsto (\varrho(a_i(\cdot),b_i(\cdot)))_{i=1,\dots,n_c}.$$

Both NCP functions used here suffer from their lack of smoothness. However, this is a general problem of NCP functions: If an NCP function ϱ is differentiable, then $\varrho'(0,0) = 0$ (cf. [Kun06, Proposition 2.18]), which is disadvantageous if combined with the Newton method, since the inverse of the Jacobian in this algorithm is supposed to be uniformly bounded. On the other hand, weaker properties like semismoothness or slant differentiability are generally not inherited by the superposition operator $\omega : (L^\infty)^2 \to L^\infty$. In [HIK03, Proposition 4.1], it was shown that $\omega_{\min}$ is only semismooth as an operator $L^q \to L^p$, where $q > p$. The necessity of this norm gap was shown for ω_{FB} in [Ulb03, Example 5.11].

This motivates using a regularized NCP function. The following regularization for the Fischer-Burmeister function (cf. [Ulb03, p. 808]) will be used in the remainder of this work:

Definition 5.4 (Regularized Fischer-Burmeister Function)
Let $\beta > 0$. The regularized Fischer-Burmeister function $\varrho^\beta : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is defined by

$$\varrho^\beta(a, b) := \sqrt{(a^2 + b^2 + \beta)} - a - b.$$

The regularized NCP operator ω_β is the mapping

$$\omega_\beta : (L^\infty([0, t_f], \mathbb{R}^{n_c}))^2 \to L^\infty([0, t_f], \mathbb{R}^{n_c}), \qquad (a, b)(\cdot) \mapsto \left(\varrho^\beta(a_i(\cdot), b_i(\cdot))\right)_{i=1,\dots,n_c}.$$

Using definition 5.4, the necessary optimality conditions can be stated as an equation. Firstly, the definition space X_∞ and image space Y_∞ are defined, then we define the operator F_β:

Definition 5.5 (The spaces X_∞ and Y_∞ and the operator F_β)
Let the spaces X_∞ and Y_∞ be defined as

$$\begin{aligned} X_\infty &:= \left(W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})\right)^2 \times L^\infty([t_0, t_f], \mathbb{R}^{n_v}) \times L^\infty([t_0, t_f], \mathbb{R}^{n_c}) \times \mathbb{R}^{n_c} \\ Y_\infty &:= \left(L^\infty([t_0, t_f], \mathbb{R}^{n_x})\right)^2 \times \mathbb{R}^{n_E} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times L^\infty([t_0, t_f], \mathbb{R}^{n_v}) \times L^\infty([t_0, t_f], \mathbb{R}^{n_c}). \end{aligned}$$

Together with the norms

$$\|(x, \lambda, v, \eta, \sigma)\|_X := \max\{\|x\|_{1,\infty}, \|\lambda\|_{1,\infty}, \|v\|_\infty, \|\eta\|_\infty, |\sigma|\}$$

and

$$\|(y_1, y_2, y_3, y_4, y_5, y_6, y_7)\|_Y := \max\{\|y_1\|_\infty, \|y_2\|_\infty, |y_3|, |y_4|, |y_5|, \|y_6\|_\infty, \|y_7\|_\infty\},$$

$(X_\infty, \|\cdot\|_X)$ and $(Y_\infty, \|\cdot\|_Y)$ become Banach spaces.

Let $F_\beta : X_\infty \to Y_\infty$ be the operator, defined by

$$F_\beta(x, \lambda, v, \eta, \sigma) := \begin{pmatrix} \dot{x}(\cdot) - A(\cdot)x(\cdot) - B_v(\cdot)v(\cdot) \\ \dot{\lambda}(\cdot) + Q(\cdot)x(\cdot) + R_v(\cdot)v(\cdot) + A(\cdot)^\top\lambda(\cdot) + G(\cdot)^\top\eta(\cdot) \\ E_0x(t_0) + E_1x(t_f) - f \\ \lambda(t_0) + E_0^\top\sigma \\ \lambda(t_f) - Q_fx(t_f) - E_1^\top\sigma \\ S_v(\cdot)v(\cdot) + R_v(\cdot)^\top x(\cdot) + B_v(\cdot)^\top\lambda(\cdot) + H(\cdot)^\top\eta(\cdot) \\ \omega_\beta(\eta, l - Gx - Hv) \end{pmatrix}.$$

As ω_β is continuously differentiable for any $\beta > 0$, the operator F_β is continuously Fréchet differentiable with respect to $z := (x, \lambda, v, \eta, \sigma)$ (cf. example 2.20.1).

For any $\beta \geq 0$, F_β is Lipschitz continuous. In fact, we will show in Lemma 5.18, that the Lipschitz continuity is uniform, i.e. the Lipschitz constant does not depend on β.

Let

$$\begin{aligned} r_i(\cdot) &:= \varrho^{\beta\prime}{}_a(\eta(\cdot), l(\cdot) - G(\cdot)x(\cdot) - H(\cdot)v(\cdot)), & \mathbf{r} &:= \operatorname{diag}(r_1, \dots, r_{n_c}), \\ s_i(\cdot) &:= \varrho^{\beta\prime}{}_b(\eta(\cdot), l(\cdot) - G(\cdot)x(\cdot) - H(\cdot)v(\cdot)), & \mathbf{s} &:= \operatorname{diag}(s_1, \dots, s_{n_c}), \end{aligned}$$

then the equation

$$F_{\beta}{}'_{z}(x,\lambda,v,\eta,\sigma)(h_x,h_\lambda,h_v,h_\eta,h_\sigma) = (d_{y_1},d_{y_2},d_{y_3},d_{y_4},d_{y_5},d_{y_6},d_{y_7})^\top$$

reads

$$\begin{pmatrix}\dot{h_x}\\ \dot{h_\lambda}\end{pmatrix} = \begin{pmatrix}A & 0\\ -Q & -A^\top\end{pmatrix}\begin{pmatrix}h_x\\ h_\lambda\end{pmatrix} + \begin{pmatrix}B_v & 0\\ -R_v & -G^\top\end{pmatrix}\begin{pmatrix}h_v\\ h_\eta\end{pmatrix} + \begin{pmatrix}d_{y_1}\\ d_{y_2}\end{pmatrix} \tag{5.1}$$

$$E_0 h_x(t_0) + E_1 h_x(t_f) = d_{y_3} \tag{5.2}$$

$$h_\lambda(t_0) + E_0^\top h_\sigma = d_{y_4} \tag{5.3}$$

$$h_\lambda(t_f) - Q_f h_x(t_f) - E_1^\top h_\sigma = d_{y_5} \tag{5.4}$$

$$\begin{pmatrix}S_v & H^\top\\ -\mathbf{s}H & \mathbf{r}\end{pmatrix}\begin{pmatrix}h_v\\ h_\eta\end{pmatrix} = \begin{pmatrix}-R_v^\top & -B_v^\top\\ \mathbf{s}G & 0\end{pmatrix}\begin{pmatrix}h_x\\ h_\lambda\end{pmatrix} + \begin{pmatrix}d_{y_6}\\ d_{y_7}\end{pmatrix}. \tag{5.5}$$

Given that $\mathcal{A}_\beta$ is invertible with $\|\mathcal{A}_\beta^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)} \le C_\beta$, where

$$\mathcal{A}_\beta := \begin{pmatrix}S_v & H^\top\\ -\mathbf{s}H & \mathbf{r}\end{pmatrix} \qquad \mathcal{A}_\beta^{-1} := \begin{pmatrix}\mathcal{V}_{11} & \mathcal{V}_{12}\\ \mathcal{V}_{21} & \mathcal{V}_{22}\end{pmatrix}, \tag{5.6}$$

equation (5.5) can be solved for (h_v, h_η):

$$\begin{pmatrix}h_v\\ h_\eta\end{pmatrix} = \begin{pmatrix}-\mathcal{V}_{11}R_v^\top + \mathcal{V}_{12}\mathbf{s}G & -\mathcal{V}_{11}B_v^\top\\ -\mathcal{V}_{21}R_v^\top + \mathcal{V}_{22}\mathbf{s}G & -\mathcal{V}_{21}B_v^\top\end{pmatrix}\begin{pmatrix}h_x\\ h_\lambda\end{pmatrix} + \begin{pmatrix}\mathcal{V}_{11}d_{y_6} + \mathcal{V}_{12}d_{y_7}\\ \mathcal{V}_{21}d_{y_6} + \mathcal{V}_{22}d_{y_7}\end{pmatrix}. \tag{5.7}$$

Conditions for the boundedness of $\mathcal{A}_\beta^{-1}$ can be derived in the same way as in [Ger08, Theorem 3.2]. For this result, we need another index set, since the assumptions on the linear independence of the control space constraints will have to be connected to the multiplier η:

Assumption 5.6
Let $z = (x,\lambda,v,\eta,\sigma) \in X_\infty$, and

$$J_\gamma(t) := \{i \in \{1,\dots,n_c\} \mid |G_i(t)x(t) + H_i(t)v(t) - l_i(t)| \le \gamma\eta_i(t),\ \eta_i(t) \ge 0\}.$$

There exist $\gamma > 0$ and $\delta > 0$, such that $\|H_{J_\gamma(t)}(t)^\top \xi\| \ge \delta\|\xi\|$ for all $\xi \in \mathbb{R}^{|J_{\gamma(t)}(t)|}$, for all times $t \in [t_0,t_f]$.

Lemma 5.7
Let the data of the $LQOCP_s$ satisfy the smoothness conditions 4.2, as well as the normality conditions 4.3. Let S_v be bounded and uniformly positive definite. Furthermore, assume that assumption 5.6 is satisfied.

Then for ϱ_{FB}^β, it holds that

$$\|\mathcal{A}_\beta^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)} \le C,$$

independent from the regularization parameter β.

Proof.
Analog to definition 3.19, let $I_\varepsilon(t) := \{i \in \{1, \dots, n_c\} | r_i \geq -\varepsilon\}$. The time t will be omitted for convenience in the remainder of this proof.

The function $\varrho^\beta_{\text{FB}}$ is symmetric, i.e. $\varrho^\beta_{\text{FB}}(a,b) = \varrho^\beta_{\text{FB}}(b,a)$, and if $\beta > 0$ then $\varrho^\beta_{\text{FB}}$ is continuously differentiable with

$$\varrho^{\beta\,\prime}_{\text{FB}a}(a,b) = \frac{a}{\sqrt{a^2+b^2+\beta}} - 1,$$

so that $(r_i+1)^2 + (s_i+1)^2 = 1-\beta$ for all $i = 1, \dots, n_c$ and it holds that $-2 \leq s_i \leq 0$ and $-2 \leq r_i \leq 0$, independent from β.

For $i \in I_\varepsilon$, it holds that $-\varepsilon \leq r_i \leq 0$, and $(s_i+1)^2 \leq 1-(r_i+1)^2$ yields

$$(s_i+1)^2 \leq 1-(1-\varepsilon)^2 = \varepsilon(2-\varepsilon),$$

hence $|s_i+1| \leq \sqrt{\varepsilon(2-\varepsilon)}$.

Summarizing, it holds that

$$\|\mathbf{r}_{I_\varepsilon}\| \leq \varepsilon, \qquad \varepsilon \leq \|\mathbf{r}_{I_\varepsilon^C}\| \leq 2, \qquad \frac{1}{2} \leq \|\mathbf{r}^{-1}_{I_\varepsilon^C}\| \leq \frac{1}{\varepsilon},$$
$$0 \leq \|\mathbf{s}_{I_\varepsilon^C}\| \leq 2, \qquad 1-\sqrt{\varepsilon(2-\varepsilon)} \leq \|\mathbf{s}_{I_\varepsilon}\| \leq 2, \qquad \frac{1}{2}\|\mathbf{s}^{-1}_{I_\varepsilon}\| \leq \frac{1}{1-\sqrt{\varepsilon(2-\varepsilon)}},$$

and the same reasoning as in the proof of [Ger08, Theorem 3.2] can be applied. This shows the boundedness of $\mathcal{A}_\beta^{-1}$ for the regularized Fischer-Burmeister function. □

For the sake of completeness, the form of $\mathcal{A}^{-1}$ can be explicitely stated. The Schur complement turns out to be useful here. Its definition is taken from [BV04, Appendix C.4.1]. In [HJ85, Section 0.8.5], the notion *Schur complement* has been defined for more general index sets.

Definition 5.8
Let the matrix $M \in \mathbb{R}^{n+m \times n+m}$ be partitioned as

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

with $M_{11} \in \mathbb{R}^{n\times n}$, $M_{12} \in \mathbb{R}^{n\times m}$, $M_{21} \in \mathbb{R}^{m\times n}$ and $M_{22} \in \mathbb{R}^{m\times m}$.

If $\det M_{11} \neq 0$*, then*

$$K := M_{22} - M_{21}M_{11}^{-1}M_{12}$$

is called the Schur complement *of M_{11} in M.*

Using the Schur complement, the inverse of a block matrix can be expressed as (cf. [BV04, Appendix C.4.1]):

Lemma 5.9
Let M be as in definition 5.8, with $\det M_{11} \neq 0$. Then the Schur complement K of M_{11} is nonsingular if and only if M is nonsingular.

If M is nonsingular, then M^{-1} takes the form

$$M^{-1} = \begin{pmatrix} M_{11}^{-1} + M_{11}^{-1} M_{12} K^{-1} M_{21} M_{11}^{-1} & -M_{11}^{-1} M_{12} K^{-1} \\ -K^{-1} M_{21} M_{11}^{-1} & K^{-1} \end{pmatrix}.$$

Lemma 5.9 leads to the conclusion that if $\mathcal{A}$ as in (5.6) is invertible, then it holds that

$$\mathcal{A}^{-1} = \begin{pmatrix} \mathcal{V}_{11} & \mathcal{V}_{12} \\ \mathcal{V}_{21} & \mathcal{V}_{22} \end{pmatrix} = \begin{pmatrix} S_v^{-1} - S_v^{-1} H^\top K^{-1} \mathbf{s} H S_v^{-1} & -S_v^{-1} H^\top K^{-1} \\ K^{-1} \mathbf{s} H S_v^{-1} & K^{-1} \end{pmatrix},$$

where $K = \mathbf{r} + \mathbf{s} H S_v^{-1} H^\top$. Here, S_v^{-1} exists, since S_v was assumed to be symmetric and positive definite.

Lemma 5.7, together with theorem 2.25 paves the way for the proof of the first intermediate result: Under suitable conditions, $\|{F'_\beta}^{-1}\|$ is bounded, independent from β. In order to formulate the assumptions necessary for this estimate, we introduce the operators that are needed to transform the system (5.1)-(5.5):

Definition 5.10
The operator $\mathcal{B}$, the function $\mathbf{c}$, the matrices $\mathcal{E}_0$ and $\mathcal{E}_1$ and the vector $\mathbf{f}$ are defined as follows:

$$\mathcal{B} := \begin{pmatrix} A & 0 & 0 \\ -Q & -A^\top & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -R_v & -G^\top \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -\mathcal{V}_{11} R_v^\top + \mathcal{V}_{12} \mathbf{s} G & -\mathcal{V}_{11} B_v^\top \\ -\mathcal{V}_{21} R_v^\top + \mathcal{V}_{22} \mathbf{s} G & -\mathcal{V}_{21} B_v^\top \end{pmatrix},$$

$$\mathbf{c} := \begin{pmatrix} 0 & 0 \\ -R_v & -G^\top \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{V}_{11} d_{y_6} + \mathcal{V}_{12} d_{y_7} \\ \mathcal{V}_{21} d_{y_6} + \mathcal{V}_{22} d_{y_7} \end{pmatrix} + \begin{pmatrix} d_{y_1} \\ d_{y_2} \\ 0 \end{pmatrix},$$

$$\mathcal{E}_0 := \begin{pmatrix} E_0 & 0 & 0 \\ 0 & I & E_0^\top \\ 0 & 0 & -I \end{pmatrix}, \qquad \mathcal{E}_1 := \begin{pmatrix} E_1 & 0 & 0 \\ 0 & 0 & 0 \\ -Q_f & I & -E_1^\top \end{pmatrix}, \qquad \mathbf{f} := \begin{pmatrix} d_{y_3} \\ d_{y_4} \\ d_{y_5} \end{pmatrix}.$$

In terms of definition 5.10, we can formulate the assumption that ensures that solutions of system (5.1)-(5.5) remain bounded (theorem 2.25):

Assumption 5.11
Let $\mathcal{B}$, $\mathcal{E}_0$ and $\mathcal{E}_1$ be defined as in definition 5.10. Let there exist a constant $C > 0$ such that $\|\mathcal{B}(t)\| \leq C$, a.e. in $[t_0, t_f]$.

Let there exist $\kappa > 0$ independent from β, such that for all $\xi \in \mathbb{R}^{n_x}$ it holds that

$$\|(\mathcal{E}_0 \Theta(t_0) + \mathcal{E}_1 \Theta(t_f))\xi\| \geq \kappa \|\xi\|,$$

where Θ solves $\dot{\Theta}(t) = \mathcal{B}(t)\Theta(t)$, $\Theta(t_0) = I$.

Lemma 5.12
Let the data of the LQOCP$_s$ satisfy the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, and let assumption 5.11 be satisfied.

Then it holds that

$$\|{F_{\beta}'_z}^{-1}\|_{\mathcal{L}(Y_\infty, X_\infty)} \leq C_F$$

for some constant $C_F > 0$.

Proof.
According to lemma 5.7, the operator $\mathcal{A}_\beta$ can be inverted, and its inverse remains bounded with respect to the operator norm.

The system (5.1)-(5.5) can therefore be transformed into a linear boundary value problem using equation (5.7), if h_σ is interpreted as a constant function. The new system reads

$$\begin{pmatrix} \dot{h}_x \\ \dot{h}_\lambda \\ \dot{h}_\sigma \end{pmatrix} = \mathcal{B} \begin{pmatrix} h_x \\ h_\lambda \\ h_\sigma \end{pmatrix} + \mathbf{c} \qquad \mathcal{E}_0 \begin{pmatrix} h_x(t_0) \\ h_\lambda(t_0) \\ h_\sigma(t_0) \end{pmatrix} + \mathcal{E}_1 \begin{pmatrix} h_x(t_f) \\ h_\lambda(t_f) \\ h_\sigma(t_f) \end{pmatrix} = \mathbf{f},$$

using the notation from definition 5.10. Theorem 2.25 ensures that the solution $(h_x, h_\lambda, h_\sigma)$ exists, and its norm is bounded linearly by the norm of $\mathbf{c}$ and $\mathbf{f}$, i.e.

$$\|(h_x, h_\lambda, h_\sigma)\| \leq K \cdot \|(\mathbf{c}, \mathbf{f})\|$$

for some constant K. As $\|\mathcal{A}^{-1}\|_{\mathcal{L}(Y_\infty, x_\infty)} \leq C_A$ for some $C_A > 0$, this implies that there exists a constant C_F, such that

$$\|(h_x, h_\lambda, h_v, h_\eta, h_\sigma)\|_{X_\infty} \leq C_F \cdot \|(d_{y_1}, d_{y_2}, d_{y_3}, d_{y_4}, d_{y_5}, d_{y_6}, d_{y_7})\|_{X_\infty},$$

which was the claim of lemma 5.12. □

The next step is to apply theorem 2.22 to the equation

$$F_\beta(z) = 0$$

in order to derive a formula for the dependence of a solution $z \in X_\infty$ on the regularization parameter β.

Lemma 5.13
Let the data of the LQOCP$_s$ satisfy the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, and let assumption 5.11 be satisfied for all $\beta \in (0, \delta)$, with $\delta > 0$.

Then the solution $g(\beta)$ of $F_\beta(g(\beta)) = 0$ depends Hölder continuously on β for $\beta \in (0, \delta)$. There exists a constant C_g, such that for β_1, β_2 the estimation

$$\|g(\beta_1) - g(\beta_2)\|_{X_\infty} \leq C_g \cdot |\beta_1 - \beta_2|^{\frac{1}{2}}$$

holds.

Proof.
In the scope of this proof, let $F : X_\infty \times (0,\delta) \to Y_\infty$, $(z,\beta) \mapsto F_\beta(z)$, with the norm $\|(z,\beta)\|_{X_\infty \times \mathbb{R}} := \|z\| + |\beta|$. Then F is continuously differentiable, and the derivative of $z \mapsto F(z,\beta_0)$ in z_0, i.e. $F'_z(z_0,\beta_0) : X_\infty \to Y_\infty$, is an isomorphism, as shown in lemma 5.12.

The derivative of ϱ_β with respect to β is $\varrho'_{\beta\beta}(a,b) = \frac{1}{2\sqrt{a^2+b^2+\beta}}$. This means that F is continuously differentiable on $(0,\delta)$, with derivative

$$F'_\beta(z,\beta) = \left(0,0,0,0,0,0, \left(\tfrac{1}{2}((\eta_i)^2 + (l - Gx - Hv)_i^2 + \beta)^{-\frac{1}{2}}\right)_{i=1,\dots,n_c}\right)^\top.$$

As we see, the inequality

$$\|F'_\beta(z,\beta)\|_{\mathcal{L}((0,\delta),Y_\infty)} \leq C_\beta \cdot \beta^{-\frac{1}{2}}$$

holds on the interval $(0,\delta)$, independent of z. At the same time, Lemma 5.12 assures that F'^{-1}_z is also bounded, independent of β.

Consequently Theorem 2.22 assures that there exist neighborhoods U_0 of z_0 and V_0 of β_0, such that for any $z \in U_0$, the equation $F(z,\beta) = 0$ admits a unique solution $z =: g(\beta) \in V_0$, where $g'_\beta(\beta) = (F'_z(z,\beta))^{-1} F'_\beta(z,\beta)$.

Hence

$$\begin{aligned}
\|g'_\beta(\beta)\|_{\mathcal{L}((0,\delta),X_\infty)} &= \|(F'_z(z,\beta))^{-1} F'_\beta(z,\beta)\|_{\mathcal{L}((0,\delta),X_\infty)} \\
&\leq \|(F'_z(z,\beta))^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)} \cdot \|F'_\beta(z,\beta)\|_{\mathcal{L}((0,\delta),Y_\infty)} \\
&\leq C_F \cdot C_\beta \cdot \beta^{-\frac{1}{2}}.
\end{aligned}$$

Now it holds

$$\begin{aligned}
\|g(\beta_1) - g(\beta_2)\|_{X_\infty} &\leq \int_{\beta_1}^{\beta_2} \|g'_\beta(\beta)\|_{X_\infty} d\beta \\
&\leq \int_{\beta_1}^{\beta_2} \|(F'_z(z,\beta))^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)} \cdot \|F'_\beta(z,\beta)\|_{\mathcal{L}((0,\delta),Y_\infty)} d\beta \\
&\leq C_F \cdot C_\beta \cdot |\beta_1 - \beta_2|^{\frac{1}{2}},
\end{aligned}$$

so g is Hölder continuous. □

Together with Lemma 2.2, Lemma 5.13 proves the main result of this section:

Theorem 5.14
Let the data of the LQOCP$_s$ satisfy the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, and let assumption 5.11 be satisfied for all $\beta \in (0,\delta)$ with $\delta > 0$.

Let $g(\beta) : (0,\delta) \to X_\infty$ be implicitly defined by $F_\beta(g(\beta)) = 0$. Then $\hat{z} := \lim_{\beta \searrow 0} g(\beta)$ exists and $F_0(\hat{z}) = 0$. The function g is Hölder continuous on $[0,\delta)$.

Proof.
The existence is the direct consequence of lemma 2.2 and lemma 5.13. The limit $\hat{z}$ solves $F_0(\hat{z}) = 0$ since F_0 is continuous, so that $F_0(\hat{z}) = \lim_{\beta \searrow 0} F_0(g(\beta)) = 0$.

With $g(0) := \lim_{\beta \searrow 0} g(\beta)$, g is continuous on $[0, \delta)$, and for $\beta \in (0, \delta)$ and sufficiently small $\hat{\beta}$, it holds that

$$\begin{aligned} \|g(0) - g(\beta)\| &\leq \|g(0) - g(\hat{\beta})\| + \|g(\hat{\beta}) - g(\beta)\| \\ &\leq \varepsilon + \sqrt{\beta - \hat{\beta}} \\ &\leq \varepsilon + \sqrt{\beta}. \end{aligned}$$

Taking the limit for $\varepsilon \to 0$ shows the Hölder continuity of g on $[0, \delta)$. □

Remark 5.15 (Property: Feasibility of Solutions)
An important property of the regularized Fischer-Burmeister function is its effect on the feasibility of solutions. Note that for $\beta > 0$ and $a, b \in \mathbb{R}$, it holds

$$\sqrt{a^2 + b^2 + \beta} - a - b = 0 \quad \Leftrightarrow \quad a > 0,\ b > 0,\ 2ab = \beta.$$

Therefore, all solutions of the regularized problem with $\beta > 0$ are strictly feasible (a.e. on $[t_0, t_f]$).

5.2. Solving the Regularized Problem

In this step, the regularized problem $F_\beta(z) = 0$ with $F_\beta : X_\infty \to Y_\infty$ as in definition 5.5 is solved by the means of the Newton method for fixed $\beta > 0$. The convergence results of the local method are based on the results in [Wan99], [Wan00] and [WL03].

5.2.1. The Newton method and its convergence radius

We write $\omega_\beta(z)$ synonymous for $\omega_\beta(\eta, l - Gx - Hv)$, where $z = (x, \lambda, v, \eta, \sigma) \in X_\infty$.

The local Newton method reads

Algorithm 5.16 (Local Newton Method)

1. *Choose $z_0 \in X_\infty$, $\beta > 0$ and $\epsilon > 0$. Let $k = 0$.*
2. *If $\|F_\beta(z_k)\|_{Y_\infty} \leq \epsilon$, stop.*
3. *Compute the search direction d_k as*

$$d_k = -{F_\beta}'_z(z_k)^{-1} F_\beta(z_k). \tag{5.8}$$

4. *Set $z_{k+1} := z_k + d_k$, $k := k + 1$, and go to step 2*

The following Theorem is cited from [TW79, Theorem 2.1], since it answers the two most important questions:

- What assumptions are needed to guarantee that $F_\beta{}'(z_k)$ is invertible?
- What is the convergence radius for the algorithm?

Theorem 5.17
Let X, Y be Banach spaces and F_β be a mapping $F_\beta : X \to Y$. Let $\delta > 0$ and x_β be a simple zero of F_β. Assume that $F_\beta{}'$ exists in $B_\delta(x_\beta)$ with

$$A_2 = A_2(\delta) = \sup_{x,y \in B_\delta(x_\beta)} \frac{\|F_\beta{}'(x_\beta)^{-1} [F_\beta{}'(x) - F_\beta{}'(y)]\|_{\mathcal{L}(X,X)}}{2 \|x - y\|_X},$$

and let $A_2\delta \leq \frac{q}{1+2q}$ for some $0 < q < 1$. Then the Newton iterations are well defined for any start value $x^0 \in B_\delta(x_\beta)$, with

$$\lim_{i\to\infty} x_i = x_\beta, \qquad \|x_{i+1} - x_\beta\|_X \leq q \cdot \|x_i - x_\beta\|_X \quad \forall i$$
$$\text{and} \quad \|x_{i+1} - x_\beta\|_X \leq C_i \|x_i - x_\beta\|_X^2 \quad \forall i, \text{ with } C_i := A_2/(1 - 2A_2\|x_i - x_\beta\|_X).$$

It first remains to find appropriate estimates for the bound A_2. Since we restrict our investigation to linear quadratic problems, the derivative of one part of the operator does not depend on the point in which it is evaluated. In order to exploit this fact, we divide the operator in two parts:

$$F_\beta = (F_{\beta_1}, F_{\beta_2})^\top : X^\infty \to Y_1^\infty \times Y_2^\infty, \tag{5.9}$$

$$F_{\beta_1}(z) = \begin{pmatrix} \dot{x}(\cdot) - A(\cdot)x(\cdot) - B_v(\cdot)v(\cdot) \\ \dot{\lambda}(\cdot) + Q(\cdot)x(\cdot) + R_v(\cdot)v(\cdot) + A(\cdot)^\top\lambda(\cdot) + G(\cdot)^\top\eta(\cdot) \\ E_0x(t_0) + E_1x(t_f) - f \\ \lambda(t_0) + E_0^\top\sigma \\ \lambda(t_f) - Q_fx(t_f) - E_1^\top\sigma \\ S_v(\cdot)v(\cdot) + R_v(\cdot)^\top x(\cdot) + B_v(\cdot)^\top\lambda(\cdot) + H(\cdot)^\top\eta(\cdot) \end{pmatrix}, \tag{5.10}$$

$$F_{\beta_2}(z) = \omega_\beta(\eta, l - Gx - Hv). \tag{5.11}$$

This notation turns out to be very handy when examining the properties of the operator. Firstly, we note that F_β is uniformly Lipschitz continuous:

Lemma 5.18
Let the smoothness Assumptions 4.2 be satisfied. Then the regularized operator F_β is uniformly Lipschitz continuous, i.e. the Lipschitz constant does not depend on β.

Proof.
We note that F_{β_1} is since it is linear and bounded, so that the (bounded) derivative does not depend on z. Hence, according to Lemma 2.26, F_{β_1} is Lipschitz continuous. Naturally, the Lipschitz constant is independent from β, as the parameter does not appear in the function.

The following consideration enables us to exploit this property: Consider a function $f : X \to Y_1 \times Y_2$, $x \mapsto (f_1(x), f_2(x))$, where X, Y_1 and Y_2 are Banach spaces and $\|(y_1, y_2)\|_Y := \max(\|y_1\|_{Y_1}, \|y_2\|_{Y_2})$. If f_1 and f_2 are both Lipschitz continuous with Lipschitz constants L_1 and L_2, respectively, then f is also Lipschitz continuous with Lipschitz constant $L_f := \max(L_1, L_2)$:

Let $x_1, x_2 \in X$. Then it holds that

$$\begin{aligned}\|f(x_1) - f(x_2)\|_Y &= \max(\|f_1(x_1) - f_1(x_2)\|_{Y_1},\ \|f_2(x_1) - f_2(x_2)\|_{Y_2}) \\ &\leq \max(L_1 \cdot \|x_1 - x_2\|_X, L_2 \cdot \|x_1 - x_2\|_X) \\ &= \max(L_1, L_2) \cdot \|x_1 - x_2\|_X.\end{aligned}$$

Hence, it only remains to show that F_{β_2} is Lipschitz continuous as well. For $a_1, b_1, a_2, b_2 \in \mathbb{R}$ and $\beta \geq 0$ it holds that

$$\begin{aligned}\left|\sqrt{a_1^2 + b_1^2 + \beta} - \sqrt{a_2^2 + b_2^2 + \beta}\right| &\leq \left|\sqrt{a_1^2 + b_1^2} - \sqrt{a_2^2 + b_2^2}\right| \\ &\leq \sqrt{|a_1^2 - a_2^2| + |b_1^2 - b_2^2|} \\ &\leq \sqrt{2} \cdot \max(|a_1 - a_2|,\ |b_1 - b_2|)\end{aligned}$$

due to the triangle inequalities in $\mathbb{R}^2$ and $\mathbb{R}^3$ and the equivalence of norms in $\mathbb{R}^2$.

Let $\beta \geq 0$. For any $z_1, z_2 \in X_\infty$, $z_1 = (x_1, \lambda_1, v_1, \eta_1, \sigma_1)$, $z_2 = (x_2, \lambda_2, v_2, \eta_2, \sigma_2)$, it holds

$$\begin{aligned}&\|F_{\beta_2}(z_1) - F_{\beta_2}(z_2)\|_{Y_\infty} \\ &\quad= \|\omega_\beta(\eta_1, l - Gx_1 - Hv_1) - \omega_\beta(\eta_2, l - Gx_2 - Hv_2)\|_{Y_\infty} \\ &\quad= \max_i \operatorname*{ess\,sup}_t \Big|\sqrt{\eta_{1i}^2 + (l - G^i x_1 - H^i v_1)^2 + \beta} - \eta_{1i} - (l - G^i x_1 - H^i v_1) \\ &\qquad\qquad -\sqrt{\eta_{2i}^2 + (l - G^i x_2 - H^i v_2)^2 + \beta} - \eta_{2i} - (l - G^i x_2 - H^i v_2)\Big| \\ &\quad\leq \max_i \operatorname*{ess\,sup}_t \Big|\sqrt{2} \cdot \max\left(|\eta_{1i} - \eta_{2i}|, |G^i(x_1 - x_2) + H^i(v_1 - v_2)|\right) \\ &\qquad\qquad + |\eta_{1i} - \eta_{2i}| + |G^i(x_1 - x_2) + H^i(v_1 - v_2)|\Big|,\end{aligned}$$

and finally, there exists a constant $C > 0$, such that $C \cdot \|z_1 - z_2\|$ yields an upper estimate for the right hand side. □

Obviously, ${F_{\beta_1}}'(z_1) - {F_{\beta_1}}'(z_2) = 0$ for any two $z_1, z_2 \in X_\infty$, as the derivative does not depend on the point in which it is evaluated. The second part can be estimated:

Lemma 5.19
Let $F_\beta = (F_{\beta_1}, F_{\beta_2})^\top$. Assume that the LQOCP$_s$ satisfies the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, let assumption 5.11 be satisfied and let $\|z_\beta\|_{X_\infty}$ be bounded for all β. Then there exists a constant $C_L \in \mathbb{R}$ independent of z_1 and z_2, such that

$$\left\|{F_\beta}'(z_\beta)^{-1}\left({F_\beta}'(z_1) - {F_\beta}'(z_2)\right)\right\|_{\mathcal{L}(X_\infty, X_\infty)} \leq C_L \cdot \|z_1 - z_2\|_{X_\infty} \cdot \beta^{-\frac{1}{2}}$$

for all $z_1, z_2 \in X_\infty$.

Proof.
Note that

$$\begin{aligned}&\|F_\beta{}'(z_\beta)^{-1}\left(F_\beta{}'(z_1)-F_\beta{}'(z_2)\right)\|_{\mathcal{L}(X_\infty,X_\infty)}\\ &\quad\leq\|F_\beta{}'(z_\beta)^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)}\cdot\|F_\beta{}'(z_1)-F_\beta{}'(z_2)\|_{\mathcal{L}(X_\infty,Y_\infty)}\\ &\quad=\|F_\beta{}'(z_\beta)^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)}\cdot\|F_{\beta_2}{}'(z_1)-F_{\beta_2}{}'(z_2)\|_{\mathcal{L}(X_\infty,Y_\infty^2)},\end{aligned}$$

where Y_∞^2 denotes the image space of F_{β_2}.

Deriving F_{β_2} in the direction of $h_z=(h_x,h_\lambda,h_v,h_\eta,h_\sigma)\in X_\infty$ yields

$$\begin{aligned}&\left\|\left(F_{\beta_{2i}}{}'(z_1)-F_{\beta_{2i}}{}'(z_2)\right)(h_z)\right\|_{Y_{\infty i}^2}\\ &\quad=\Bigg\|\left(\frac{\eta_{1i}}{\sqrt{\eta_{1i}^2+(l_i-G^ix_1-H^iv_1)^2+\beta}}-\frac{\eta_{2i}}{\sqrt{\eta_{2i}^2+(l_i-G^ix_2-H^iv_2)^2+\beta}}\right)h_{\eta_i}\\ &\qquad+\left(\frac{l_i-G^ix_2-H^iv_2}{\sqrt{\eta_{2i}^2+(l_i-G^ix_2-H^iv_2)^2+\beta}}-\frac{l_i-G^ix_1-H^iv_1}{\sqrt{\eta_{1i}^2+(l_i-G^ix_1-H^iv_1)^2+\beta}}\right)\\ &\qquad\cdot(G^ih_x+H^ih_v)\Bigg\|_{Y_{2i}^\infty}.\end{aligned}\tag{5.12}$$

Let $a_1,a_2,b_1,b_2\in\mathbb{R}$, then it holds that

$$\left|\frac{a_1}{\sqrt{a_1^2+b_1^2+\beta}}-\frac{a_2}{\sqrt{a_2^2+b_2^2+\beta}}\right|\leq\frac{|a_1-a_2|+|b_1-b_2|}{\sqrt{\beta}}.$$

This inequality is derived in the Appendix in Lemma A.2. Applying this formula to equation (5.12) yields

$$\begin{aligned}&\left\|\left(F_{\beta_{2i}}{}'(z_1)-F_{\beta_{2i}}{}'(z_2)\right)(h_z)\right\|_{Y_{\infty i}^2}\\ &\quad\leq\left\|\left(|\eta_{1i}-\eta_{2i}|\cdot h_{\eta_i}+\left|G^i(x_1-x_2)+H^i(v_1-v_2)\right|\cdot(G^ih_x+H^ih_v)\right)\cdot\beta^{-\frac{1}{2}}\right\|_{Y_{\infty i}^2}.\end{aligned}$$

The assertion follows by taking the essential supremum of z_1-z_2 and h_z. □

Lemma 5.19, together with Theorem 5.17 allows a qualitative estimate of the convergence radius of the Newton method in dependence on β:

Corollary 5.20
Assume that the LQOCP$_s$ satisfies the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, let assumption 5.11 be satisfied and assume that z_β remains bounded for small β. Then there exists a constant $C_\delta>0$ independent of β, such that the sequence $(z_i)_{i\in\mathbb{N}}$ generated by the Newton method with the initial value z_0 converges quadratically towards the solution z_β of $F_\beta(z_\beta)=0$, if $z_0\in B_\delta(z_\beta)$, where δ satisfies

$$\delta<C_\delta\cdot\beta^{\frac{1}{2}}.$$

Proof.
According to Lemma 5.19, A_2 in Theorem 5.17 is bounded by

$$\begin{aligned} A_2(\delta) = &\sup_{z_1,z_2\in B_\delta(z_\beta)} \frac{\left\|{F_\beta}'(z_\beta)^{-1}\left[{F_\beta}'(z_1) - {F_\beta}'(z_2)\right]\right\|_{\mathcal{L}(X_\infty,X_\infty)}}{2\left\|z_1 - z_2\right\|_{X_\infty}} \\ \leq &\frac{C_L}{2}\beta^{-\frac{1}{2}}. \end{aligned}$$

The condition for convergence is then fulfilled if

$$\frac{C_L}{2}\cdot\beta^{-\frac{1}{2}}\cdot\delta < \frac{1}{3}$$

since then the inequality $\frac{C_L}{2}\cdot\beta^{-\frac{1}{2}}\cdot\delta < \frac{q}{1+2q}$ admits a solution. Solving this inequality for δ yields the assertion with $C_\delta := \frac{2}{3C_L}$. □

Corollary 5.20 leads to some important observations:

Remark 5.21 (Convergence of the Newton method for different values of β)
5.21.1 *The estimation for the convergence radius of the Newton method depends linearly on $\sqrt{\beta}$.*

5.21.2 *If $\beta > 0$, then there exists a radius $\delta > 0$, such that the method converges if started with some initial value $z^0 \in B_\delta(z_\beta)$. For this result, no assumption about the behaviour of the iterations has to be made.*

5.21.3 *For a given initial value z_0, it is always possible to find a β, such that the Newton method converges, as long as the solutions z_β remain uniformly bounded.*

Finally, it should be mentioned that all convergence results also hold for the residual values $\|F_\beta\|_\infty$. The proof is analogous to the proof of [Ger08, Theorem 2.3]:

Corollary 5.22
Let the assumptions of Corollary 5.20 be satisfied. If the sequence $(z_i)_{i\in\mathbb{N}}$ generated by the Newton method converges quadratically towards the solution z_β of $F_\beta(z_\beta) = 0$, then either the resiual values also converge quadratically:

$$\frac{\|F_\beta(z_{k+1})\|_{Y_\infty}}{\|F_\beta(z_k)\|_{Y_\infty}^2} \leq C \quad \forall i$$

or vanish in finite time, i.e. $F_\beta(z_i) = 0$ for some $i \in \mathbb{N}$.

For sufficiently large $i \in \mathbb{N}$, there exists a constant C_r, such that

$$\|z_i - z_\beta\|_{X_\infty} \leq C_r\|F_\beta(z_i)\|_{Y_\infty}$$

holds, so that $\|F_\beta(z_i)\|_{Y_\infty}$ provides an estimate for the distance of the current iterate from the solution.

Proof.
It is assumed that $(z_i)_{i\in\mathbb{N}}$ converges quadratically towards z_β, i.e.

$$\frac{\|z_{i+1} - z_\beta\|_{X_\infty}}{\|z_i - z_\beta\|_{X_\infty}^2} \leq q \quad \forall i$$

with $q < 1$. The convergence of the residuals therefore follows from the continuity of F_β. Due to the assumptions, it holds that $\|F_\beta{}_z'(z_i)^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)} \leq C_F$ for some constant C_F. Therefore,

$$\|z_{i+1} - z_i\|_{X_\infty} = \|F_\beta{}_z'(z_i)^{-1} F_\beta(z_i)\|_{X_\infty} \leq \|F_\beta{}_z'(z_i)^{-1}\|_{\mathcal{L}(Y_\infty,X_\infty)} \cdot \|F_\beta(z_i)\|_{Y_\infty} \leq C_F \|F_\beta(z_i)\|_{Y_\infty}$$

and as z_i converges to z_β at more than superlinear rate, it holds for sufficiently large $i \in \mathbb{N}$:

$$\begin{aligned}\|z_i - z_\beta\|_{X_\infty} \leq& \|z_{i+1} - z_i\|_{X_\infty} + \|z_{i+1} - z_\beta\|_{X_\infty} \\ \leq& C_F \|F_\beta(z_i)\|_{Y_\infty} + \epsilon \|z_i - z_\beta\|_{X_\infty},\end{aligned}$$

hence

$$\|z_i - z_\beta\|_{X_\infty} \leq \frac{C_F}{1-\epsilon} \|F_\beta(z_i)\|_{Y_\infty},$$

where the right hand side remains bounded for small values of ϵ. This proves the inequality.

The above consideration, together with the fact that F_β is uniformly Lipschitz continuous (according to Lemma 5.18) yields

$$\begin{aligned}\|F_\beta(z_{i+1})\|_{X_\infty} &= \|F_\beta(z_{i+1}) - F_\beta(z_\beta)\|_{X_\infty} \\ &\leq L_F \|z_{i+1} - z_\beta\|_{X_\infty} \\ &\leq q L_F \|z_i - z_\beta\|_{X_\infty}^2 \\ &\leq \frac{q L_F C_F^2}{(1-\epsilon)^2} \|F_\beta(z_i)\|_{Y_\infty}^2.\end{aligned}$$

As the right hand side remains bounded for small ϵ, this shows the assertion. □

5.2.2. Example: Regularized Minimum Energy Problem

The following numerical example shows the fast convergence of the local Newton method. At the same time, it becomes evident that the convergence radius is limited. This motivates the derivation of a global method in Section 5.3.

This example is the regularized Minimum Energy Problem 4.23, with parameters

$$\kappa(\alpha) = 0, \qquad \varphi(\alpha) = 1, \qquad \gamma(\alpha) = \alpha := 10^{-5}, \qquad \beta = 10^{-2}, \qquad \epsilon = 10^{-9}.$$

The plots and tables have been calculated for 501 time steps. Table 5.1a and Figure 5.1 show the first iterations of the local Newton method that has been started in $z_0 = 0$. The lighter grey lines are plots of earlier iterations, and the black line indicates the last iteration. Apparently, the start value lies outside of the convergence radius, as the iterations seem to diverge. However, this sequence does find the numerical solution after 25 iterations.

#It	$\|F_\beta(z_k)\|_\infty$	$\|d_k\|_\infty$
0	$1.00000E+00$	$2.01940E+00$
1	$2.35475E-01$	$4.81094E-01$
2	$1.61896E-01$	$3.81774E+00$
3	$1.33029E-01$	$5.09223E+01$
4	$3.51663E-02$	$1.86798E+02$
5	$1.43450E+02$	$4.70267E+02$
6	$7.11620E+02$	$1.14120E+03$
7	$1.42233E+03$	$1.35245E+03$
	...	

(a) Iterations and errors for $z_0 = 0$

#It	$\|F_\beta(z_k)\|_\infty$	$\|d_k\|_\infty$
0	$8.00000E-01$	$1.21596E+03$
1	$2.51245E-02$	$8.58279E+02$
2	$9.32281E-03$	$5.47601E+02$
3	$2.97069E-03$	$2.65740E+02$
4	$6.55716E-04$	$3.48458E+01$
5	$6.49627E-05$	$2.51709E+00$
6	$1.10786E-06$	$3.42440E-02$
7	$4.76022E-10$	

(b) Iterations and errors for $z_0 = 0.2 \cdot \hat{z}$

Table 5.1.: Iterations for the Minimum Energy Problem: divergence and convergence

The second Table 5.1b and Figure 5.2 show the iterations for the algorithm for $z_0 = 0.2 \cdot \hat{z}$, where $\hat{z}$ is a numerical solution (up to a residual error of 10^{-9}). The algorithm shows the expected fast convergence[1]. The figure suggests that the components of z_k that slow down the convergence are the multiplier η and the virtual control. The fact that these variables are in a way closely related to the regularization parameter supports the results in [GHed, Ex. 4.1]. There, the numerical experiments with a related globalized Newton method showed that the number of iterations grew with decreasing regularization parameter α.

5.3. Synthesis: Newton Method for the Unregularized Problem

So far, Corollary 5.20 gives a result about the convergence radius of the Newton method for the regularized problem. In certain applications, this may be sufficient, cf. Chapter 7. As the examples illustrate, a globally convergent algorithm is needed. This can easily be derived in combination with Theorem 5.14.

The obvious approach is to choose an initial regularization parameter and let the Newton method converge up to a given tolerance. Afterwards, both the parameter and the tolerance are decreased, and the procedure starts over again. The residua $\|F(z_k)\|_{Y_\infty}$ and $\|F_\beta(z_k)\|_{Y_\infty}$ in this context serve as measurements for $\|z_k - z_0\|_{X_\infty}$ and $\|z_k - z_\beta\|_{X_\infty}$, respectively.

Algorithm 5.23 (Combined Newton Method)

1. *Choose* $z_0 \in X_\infty$, $\beta_0 := \beta > 0$, $C_{tol} > 0$, $c_\beta \in (0,1)$ *and* $\epsilon > 0$.
2. *If* $\|F(z_k)\|_{Y_\infty} \leq \epsilon$, *stop.*

[1] Due to the small regularization parameter however, the radius of quadratic convergence is very small, so that the convergence rate increases quite late.

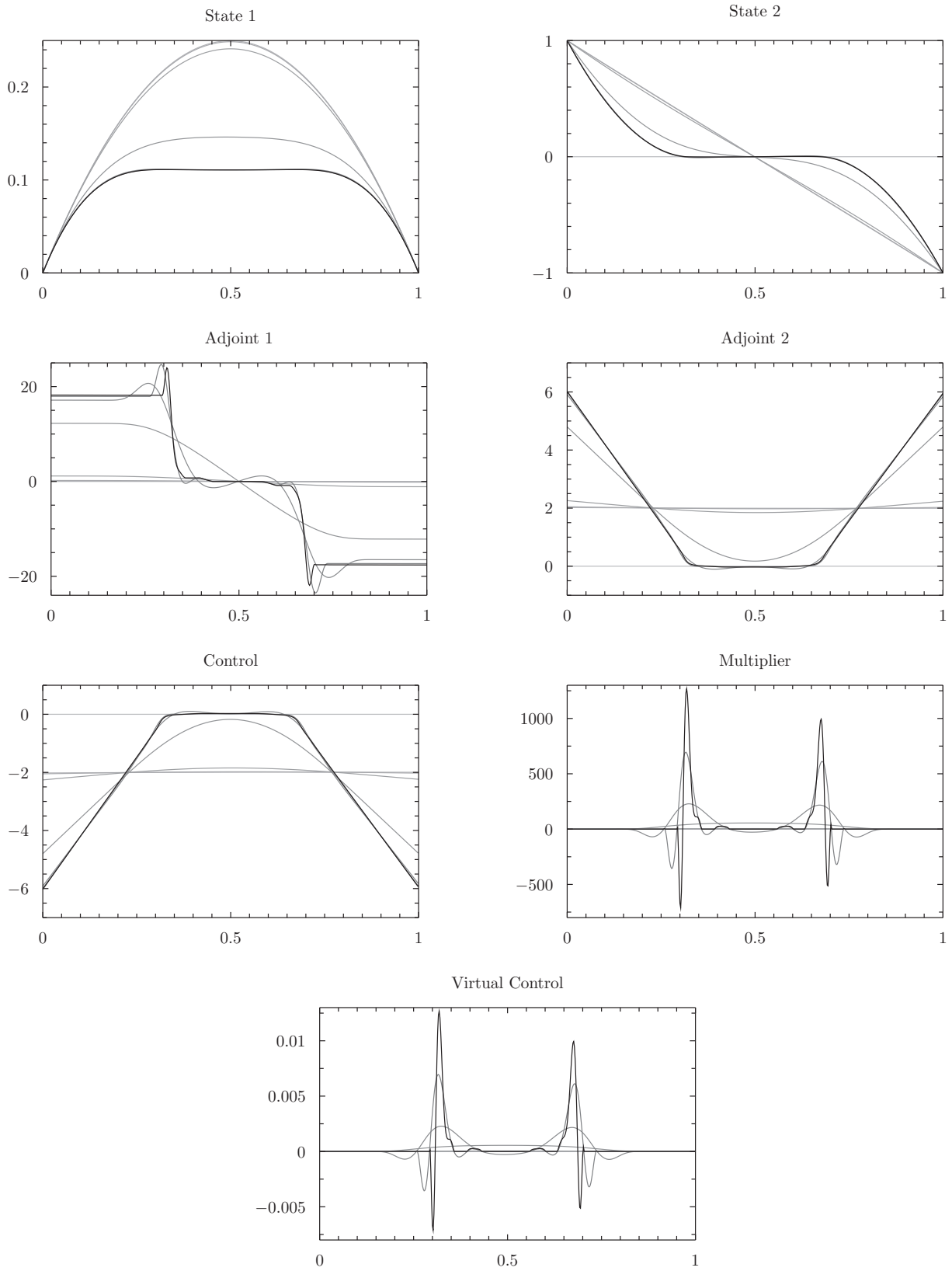


Figure 5.1.: Iterations for the regularized Minimum Energy Problem with $z_0 = 0$

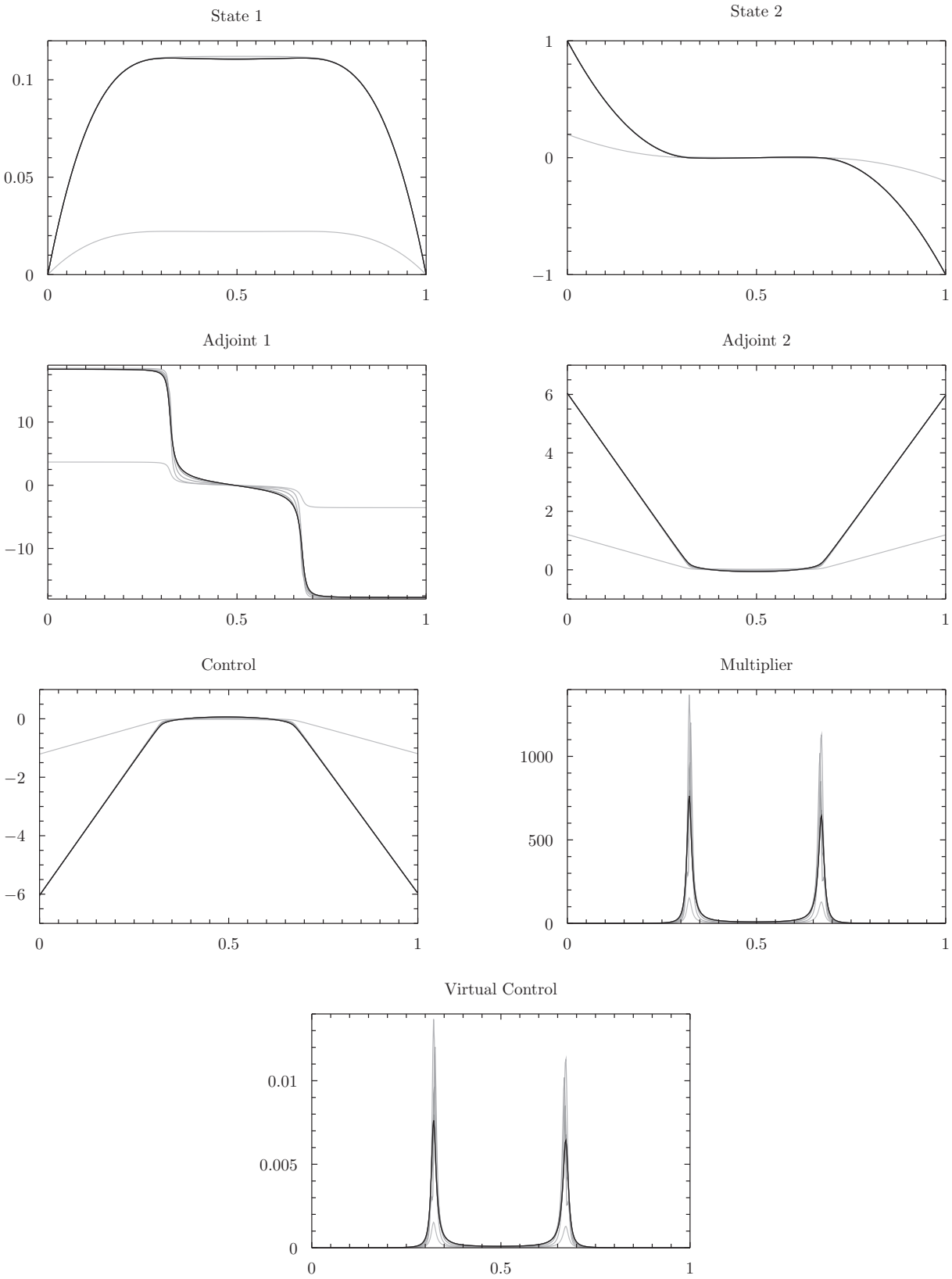


Figure 5.2.: Iterations for the regularized Minimum Energy Problem with $z_0 = 0.2 \cdot \hat{z}$

3. *If* $\|F_{\beta_{k-1}}(z_k)\|_{Y_\infty} \leq C_{tol} \cdot \sqrt{\beta_{k-1}}$ *set* $\beta_k := \beta_{k-1} \cdot c_\beta$, *otherwise set* $\beta_k := \beta_{k-1}$.
4. *Compute the search direction* $d_k = -{F_{\beta_k}}'(z_k)^{-1} F_{\beta_k}(z_k)$.
5. *Set* $z_{k+1} := z_k + d_k$, $k := k+1$ *and go to step 2.*

Theorem 5.24 (Convergence of the Combined Newton Method)
Assume that the LQOCP$_s$ satisfies the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, let assumption 5.11 be satisfied and assume that z_β remains bounded.

If z_0 lies in a sufficiently small neighborhood of the solution $\hat{z}_{\beta_0}$ of $F_{\beta_0}(z) = 0$, the Combined Newton Method stops in finite time. The final iterate z_k lies in a neighborhood of the solution $\hat{z}$ of $F_0(z) = 0$.

Proof.
First note that steps 4 and 5 form the local Newton method for the regularized problem. We assume that the first iterate under consideration, z_l fulfills $\|z_l - \hat{z}_{\beta_l}\| < C_\delta \cdot \sqrt{\beta_l}$, where C_δ is the constant from Corollary 5.20. The algorithm will converge quadratically towards $\hat{z}_{\beta_l}$, until $\|F_{\beta_l}(z_m)\|_{Y_\infty} < C_{tol} \cdot \sqrt{\beta_l}$ holds for some $m \in \mathbb{N}$. In this situation, β will be updated to $\beta_m := \beta_l \cdot c_\beta$. It remains to show that the first iterate that satisfies this condition, z_m, lies in the convergence radius of the new problem. The distance between the solutions $\hat{z}_{\beta_l}$ and $\hat{z}_{\beta_m}$ can be estimated using the inequality of Lemma 5.13:

$$\|\hat{z}_{\beta_l} - \hat{z}_{\beta_m}\|_{X_\infty} \leq \sqrt{\beta_l - \beta_m} \cdot C_g,$$

hence

$$\begin{aligned}
\|z_m - \hat{z}_{\beta_m}\|_{X_\infty} &\leq \|z_m - \hat{z}_{\beta_l}\|_{X_\infty} + \|\hat{z}_{\beta_l} - \hat{z}_{\beta_m}\|_{X_\infty} \\
&< C_{tol}\sqrt{\beta_l} + \sqrt{\beta_l - \beta_m} \cdot C_g \\
&= C_{tol}\sqrt{c_\beta^{-1}}\sqrt{\beta_m} + C_g\sqrt{c_\beta^{-1} - 1}\sqrt{\beta_m} \\
&= \left(C_{tol}\sqrt{c_\beta^{-1}} + C_g\sqrt{c_\beta^{-1} - 1}\right) \cdot \sqrt{\beta_m}
\end{aligned}$$

If C_{tol} and c_β are chosen appropriately (i.e. c_β sufficiently close to 1 and C_{tol} sufficiently small), the right hand side satisfies

$$(C_{tol}\sqrt{c_\beta^{-1}} + C_g\sqrt{c_\beta^{-1} - 1}) \cdot \sqrt{\beta} < C_\delta \cdot \sqrt{\beta_m}.$$

If that is the case, the first iterate that satisfies the condition in step 3 lies in the convergence radius of the Newton method for the decreased parameter β_m. □

Remark 5.25
For the convergence result of Theorem 5.24, the assumption is made that z_0 lies in a neighborhood of z_{β_0}. If on the other hand the solutions z_β of the equation $F_\beta(z_\beta)$ remain bounded independent from β, then z_0 will eventually lie in a sufficiently small neighborhood if β_0 is chosen large enough as the convergence radius of the local Newton method increases.

#IT	β_k	$\|F_\beta(z_k)\|_{Y^\infty}$	$\|F_0(z_k)\|_{Y^\infty}$	$\|F_\beta(z_k)\|_{Y^2}$	$\|F_0(z_k)\|_{Y^2}$
0	$5.0000E-01$	$1.0000E+00$	$1.0000E+00$	$1.4773E+00$	$1.4142E+00$
1	$2.5000E-01$	$4.0970E-01$	$1.4826E-01$	$1.5679E-01$	$6.8559E-02$
2	$1.2500E-01$	$1.8327E-01$	$1.3611E-01$	$7.6128E-02$	$6.4080E-02$
3	$6.2500E-02$	$1.1197E-01$	$1.0800E-01$	$5.0453E-02$	$5.2440E-02$
4	$6.2500E-02$	$1.7143E+00$	$1.6785E+00$	$3.3872E-01$	$3.2406E-01$
5	$6.2500E-02$	$1.2925E+01$	$1.2920E+01$	$1.6761E+00$	$1.6733E+00$
6	$3.1250E-02$	$2.4467E-01$	$8.9959E-02$	$3.8394E-02$	$2.1185E-02$
7	$1.5625E-02$	$1.3198E-01$	$5.5248E-02$	$1.4071E-02$	$1.6275E-02$
			...		
61	$8.8818E-16$	$7.2672E-09$	$1.0591E-08$	$2.8217E-10$	$7.4474E-10$
62	$4.4409E-16$	$1.1384E-10$	$1.0324E-08$	$2.5895E-10$	$5.4729E-10$
63	$2.2204E-16$	$5.2358E-09$	$5.8675E-09$	$7.6201E-11$	$2.9929E-10$
64	$1.1102E-16$	$1.2620E-11$	$6.0590E-09$	$1.2960E-10$	$2.8104E-10$
65	$5.5511E-17$	$6.8183E-10$	$3.0895E-09$	$5.2056E-11$	$1.4291E-10$
66	$2.7756E-17$	$6.6370E-11$	$1.9847E-09$	$4.2054E-11$	$9.0331E-11$
67	$1.3878E-17$	$4.3615E-11$	$1.0321E-09$	$2.2026E-11$	$4.6892E-11$
68	$1.3878E-17$	$1.0714E-11$	$5.3982E-10$	$4.7196E-13$	$2.4470E-11$

Table 5.2.: Errors of the combined method for the regularized Problem

5.3.1. Example: Regularized Minimum Energy Problem

The Regularized Minimum Energy Problem from Section 5.2.2 can now be solved using the Combined Newton method 5.23. The algorithm was started with the following parameters:

$$\beta_0 = 1, \qquad \epsilon = 10^{-9}, \qquad C_{tol} = 1, \qquad c_\beta = 0.5$$

and with $z_0 = 0$.

The plots in Figure 5.3 show the convergence of the iterates. Table 5.2 lists the errors of the iterates in the $\|\cdot\|_\infty$-norm as well as the $\|\cdot\|_2$-norm. On one hand, the convergence seems to be quite slow. Also, at some of the iterates, the error increases significantly (e.g. 4). On the other hand, the differential equations for the system becomes stiff as the regularization parameter for the state constraints is $\alpha = 10^{-4}$, which is rather small. For the unregularized problem, the iterations cannot converge at all in the $\|\cdot\|_\infty$-norm sense, as the first adjoint λ_1 is discontinuous and all iterates are continuous. Bearing this in mind, it is remarkable that the iterations remain smoother than the ones produced by the local Newton method.

5.4. Alternative: A Globalized Approach

While the analytical properties of the algorithm proved to be advantageous in Example 5.3.1, the convergence rate is slower than it could be expected in direct discretization algorithms

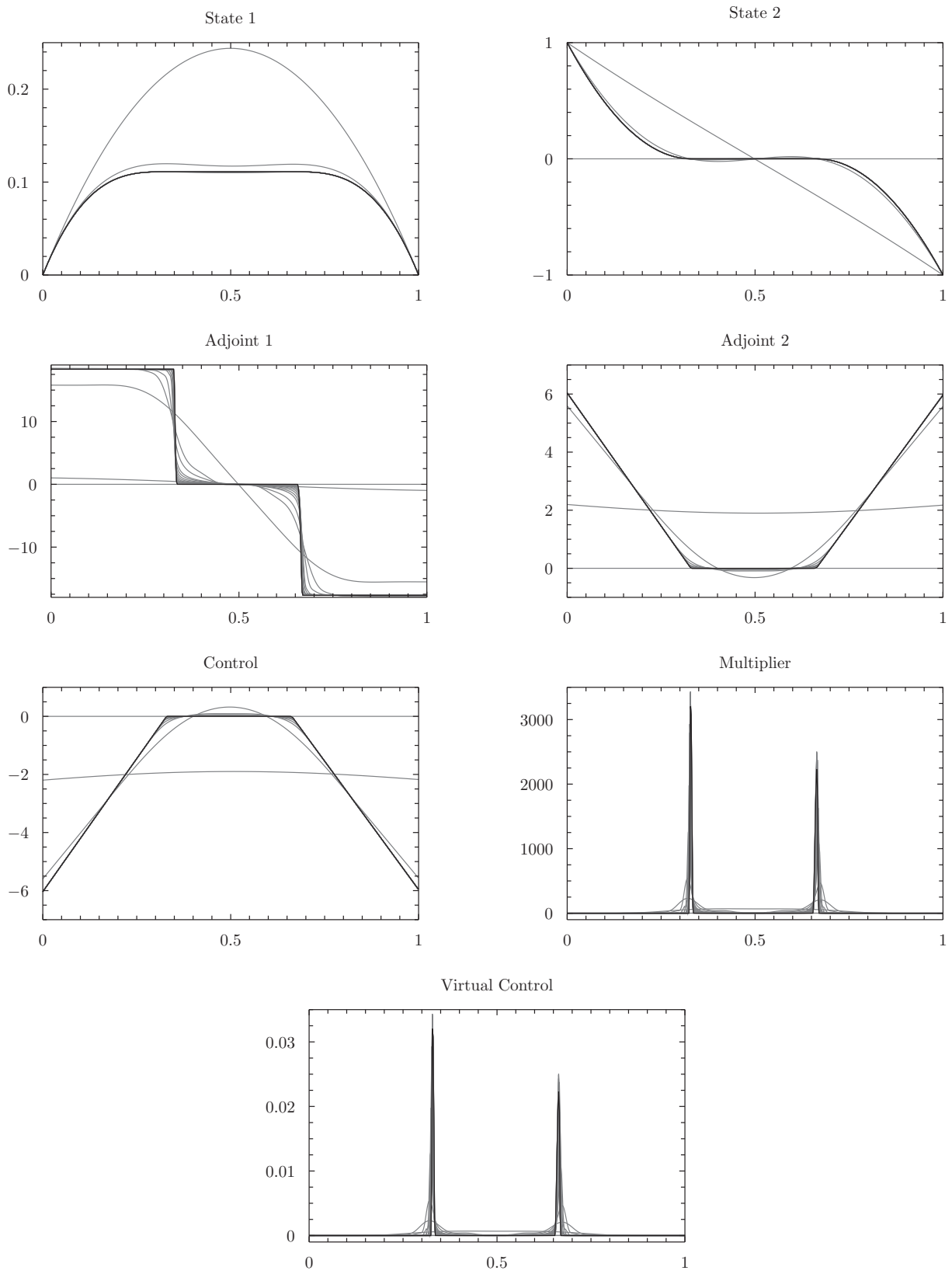


Figure 5.3.: Iterations of the combined method for the regularized Minimum Energy Problem

(cf. Table 5.2). Also, since the constants C_{tol} and c_β that have a significant influence on the convergence, they have to be guessed correctly.

A straightforward approach would be to use the results from [Ger08] to improve this, namely by using the globalized Newton method for the local problems. The method is inspired by a globalized nonsmooth Newton method in finite spaces. This way, the constants C_{tol} and c_β do not need to be guessed correctly in order to assure global convergence as any local problem can always be solved globally.

The general idea is to use a merit function for the local problems that allows the use of a sensible step size strategy. As this leads to a globally convergent algorithm, the only remaining question is how the regularization parameter should be adapted. The approach that is used in the context of this work is to update the parameter so that we can hope that the solution for the new parameter solves the original problem sufficiently well.

The merit function used for the globalization in this context is

$$
\begin{aligned}
\Theta_\beta(z) :=& \|F_\beta(z)\|_{Y^2}^2 \qquad (5.13)\\
=\ & \frac{1}{2}\sum_{i=1}^{n_x}\int_{t_0}^{t_f}\left(\dot{x}_i(t) - A^i(t)x(t) - B_v^i(t)v(t)\right)^2 dt\\
&+ \frac{1}{2}\sum_{i=1}^{n_x}\int_{t_0}^{t_f}\left(\dot{\lambda}_i(t) + Q^i(t)x(t) + R_v^i(t)v(t) + A_i^\top(t)\lambda(t) + G_i^\top(t)\eta(t)\right)^2 dt\\
&+ \sum_{i=1}^{n_E}\left(E_0^i x(t_0) + E_1^i x(t_f) - f_i\right)^2\\
&+ \sum_{i=1}^{n_x}\left(\lambda_i(t_0) + {E_0}_i^\top \sigma\right)^2\\
&+ \frac{1}{2}\sum_{i=1}^{n_v}\int_{t_0}^{t_f}\left(S_v^i(t)v(t) + {R_v}_i^\top(t)x(t) + {B_v}_i^\top(t)\lambda(t) + H_i^\top(t)\eta(t)\right)^2 dt\\
&+ \frac{1}{2}\sum_{i=1}^{n_c}\int_{t_0}^{t_f}{\omega_\beta}_i^2(\eta(t), l(t) - G(t)x(t) - H_v(t)v(t))dt.
\end{aligned}
$$

As in [Ger08], this merit function seems particularly natural for the problem as the partial derivative at a point z_k in the Newton direction d_k equals the L^2-norm of the residuum in z_k:

Lemma 5.26
Let $z_k \in X_\infty$, and let d_k be the Newton direction $d_k := -{F_\beta}'(z_k)^{-1}F_\beta(z_k)$. Then the partial derivative of Θ_β, defined as in Equation (5.13)*, satisfies*

$$\Theta_\beta{}'(z_k)(d_k) = -2\Theta_\beta(z_k) = -\|F_\beta(z_k)\|_{Y_2}^2.$$

Proof.
Let $\langle\cdot,\cdot\rangle_{Y_2}$ denote the scalar product in Y_2 with $\Theta_\beta(z) = \langle z, z\rangle_{Y_2}$. Then it holds that

$$\Theta_\beta'(z_k)(d_k) = \frac{d}{dz_k}\left\langle F_\beta(z_k), F_\beta(z_k)\right\rangle_{Y_2}(d_k)$$

$$
\begin{aligned}
&= 2\ \langle F_\beta(z_k), {F_\beta}'(z_k) d_k \rangle_{Y_2} \\
&= 2\ \left\langle F_\beta(z_k), -{F_\beta}'(z_k) {F_\beta}'(z_k)^{-1} F_\beta(z_k) \right\rangle_{Y_2} \\
&= -2\Theta_\beta(z_k)\ = -\|F_\beta(z_k)\|_{Y_2}^2.
\end{aligned}
$$

□

Lemma 5.26 motivates the use of Θ_β as a merit function in a globalization approach for the local problems that uses Armijo's step size, as the directional derivative does not need to be calculated seperately, analogous to [Ger08]:

Algorithm 5.27 (Globalized Newton Method for regularized problems)

1. *Choose $z_0 \in X_\infty$, $\beta_N \in (0,1)$, $\sigma_N \in (0, 1/2)$ and $\epsilon > 0$.*
2. *If $\|F_\beta(z_k)\|_{Y_\infty} \leq \epsilon$, stop.*
3. *Compute the search direction d_k as*

$$d_k = -{F_\beta}'(z_k)^{-1} F_\beta(z_k)$$

4. *Find the smallest $i_k \in \mathbb{N}_0$, such that*

$$\Theta_\beta(z_k + \beta_N^{i_k} d_k) \leq \Theta_\beta(z_k) + \sigma_N \beta_N^{i_k} {\Theta_\beta}'(z_k)(d_k) \tag{5.14}$$

   and set $\alpha_N^k := \beta_N^{i_k}$.
5. *Set $z_{k+1} := z_k + \alpha_N^k d_k$, $k := k+1$, and go to step 2.*

According to Lemma 5.26, the partial direction in step 4 can replaced by ${\Theta_\beta}'(z_k)(d_k) = -2\Theta_\beta(z_k)$, so that Inequality (5.14) can equivalently be replaced by

$$\Theta_\beta(z_k + \beta_N^{i_k} d_k) \leq (1 - 2\sigma_N \beta_N^{i_k}) \Theta_\beta(z_k).$$

As the merit function shows the same attributes as in the aforementioned work, Theorem 4.2 from [Ger08] can be transfered to our setting:

Theorem 5.28
Assume that the LQOCP$_s$ satisfies the smoothness conditions 4.2, the normality conditions 4.3 and the connected rank assumptions 5.6. Let S_v be bounded and uniformly positive definite, let assumption 5.11 be satisfied and assume that z_β remains bounded.

Let $\epsilon = 0$. Let z_β^* be an accumulation point of the sequence $(z_k)_{k \in \mathbb{N}}$ generated by Algorithm 5.27. Then z_β^* is a zero of F_β.

The proof is omitted here, as it can be literally copied from [Ger08, Theorem 4.2]. Consequently, the following convergence result can be transferred to this setting analogously (cf. [Ger08, Theorem 4.3]):

Theorem 5.29
Let the assumptions from Theorem 5.28 be valid. Assume that there exists a constant $K \in \mathbb{R}$, such that

$$\|F(z_k)\|_{Y_\infty} \leq K \cdot \|F(z_k)\|_{Y_2} \tag{5.15}$$

holds for the sequence $(z_k)_{k\in\mathbb{N}}$ with $z_k \to \hat{z}_\beta$. Then for sufficiently large k, the step size $\alpha_N^k = 1$ is accepted, so that the global method coincides with the local method. The local quadratic convergence is therefore inherited.

Remark 5.30 (The two-norm discrepancy in Theorem 5.29)
The problem with Theorem 5.29 is that Equation (5.15) *cannot yet be guaranteed to hold for the whole sequence by any other sensible assumption. The obvious assumption that the equation is satisfied by the whole space is clearly wrong, as the two norms are not equivalent. This is known as a "two-norm discrepancy".*

In the numerical calculations however, the two-norm discrepancy hardly matters as the iterates are calculated on a discrete grid. All functions are therefore approximated as a part of the $\mathbb{R}^n$, where all norms are equivalent. The merit function should then be chosen so that it is compatible with the discretization of the Newton direction. Naturally, the expected problems with the two norm discrepancy may return if the discretization grid is refined during the process.

Thus, we concentrate on a fixed grid for the remainder of this work, bearing in mind that this merely helps accelerating the numerical solution of the original problem.

As we are interested in the solution of the original problem with $\beta = 0$ rather than the regulated problem, we use an adaptive regularization. Similarly to the combined Newton method, we let the local method converge up to a tolerance that takes the current regularization parameter β_k into account: If $\|F_\beta(z_k)\|_{Y_2} \leq C_{loc} \cdot \sqrt{\beta_k}$ holds, we decrease the parameter. Here, we already make use of the assumed property (5.15), as it is essential for the fast convergence of the algorithm.

Then we set a decreased regularization parameter β^{new}, so that, according to our knowledge of the problem, the next iterate $\hat{z}$ that satisfies the stopping criterion satisfies $\|F_0(\hat{z})\|_{X_\infty} \leq \epsilon$. As our estimations of the global error depend linearly on the square root of the regularization parameter β, it seems straightforward to choose a stopping criterion for the local problem that also depends linearly on the square root of β.

Let $\tilde{z} \in X_\infty$ be an iterate that satisfies $\|F_{\beta^{new}}(\tilde{z})\|_{Y_2} \leq C_{loc} \cdot \sqrt{\beta^{new}}$. Then it holds for the solution $\hat{z}_0$ of $F_0(z) = 0$, that

$$\begin{aligned}
\|F_0(\tilde{z})\|_{Y^\infty} &\leq C_F \cdot \|\tilde{z} - \hat{z}_0\|_{X_\infty} \\
&\leq C_F \cdot \|\tilde{z} - \hat{z}_{\beta^{new}}\|_{X_\infty} + C_F \cdot \|\hat{z}_{\beta^{new}} - \hat{z}_0\|_{X_\infty} \\
&\leq C_F \cdot C_\tau \cdot \|F_{\beta^{new}}(\tilde{z})\|_{Y_\infty} + C_F \cdot C_g \sqrt{\beta^{new}} \\
&\leq C_F \cdot C_\tau \cdot K \cdot \|F_{\beta^{new}}(\tilde{z})\|_{Y_2} + C_F \cdot C_g \sqrt{\beta^{new}} \\
&\leq (C_F \cdot C_\tau \cdot K \cdot C_{loc} + C_F \cdot C_g) \cdot \sqrt{\beta^{new}}.
\end{aligned}$$

Hence there exists a constant C, such that

$$\frac{\|F_0(\tilde{z})\|_{Y_\infty}}{\sqrt{\beta^{new}}} \leq C.$$

As an analogous inequality holds for β_k whenever the iterate z_k satisfies $z_k \leq C_{loc} \cdot \sqrt{\beta_k}$, we simply use $\|F_0(z_k)\|_{Y_\infty}/\sqrt{\beta_k}$ as an estimation of this constant. This leads to the condition

$$\frac{\|F_0(z_k)\|_{Y_\infty}}{\sqrt{\beta_k}} \cdot \sqrt{\beta^{new}} \leq \epsilon$$

for the new parameter β^{new}, so that we choose

$$\beta^{new} := \min\left\{\frac{\epsilon^2 \cdot \beta_k}{\|F_0(z_k)\|^2_{Y_\infty}}, \frac{\beta_k}{2}\right\}.$$

The minimum is used to guarantee that $\beta^{new} < \beta_k$ holds. This is necessary, as the regularization parameter is only updated if the stopping criterion for the regularized problem is fulfilled and the residuum of the original problem is not sufficiently small.

Algorithm 5.31 (Globalized Newton Method)

1. *Choose* $z_0 \in X_\infty$, $\beta_N \in (0,1)$, $\sigma_N \in (0, 1/2)$, β_0 *and* $\epsilon > 0$. *Let* $k = 0$.
2. *If* $\|F_0(z_k)\|_{Y_\infty} \leq \epsilon$, *stop.*
3. *If* $\|F_{\beta_k}(z_k)\|_{Y_2} \leq C_{loc} \cdot \sqrt{\beta_k}$, *set* $\beta_{k+1} := \min\left\{\frac{\epsilon^2 \cdot \beta_k}{\|F_0(z_k)\|^2_{Y_\infty}}, \frac{\beta_k}{2}\right\}$, $z_{k+1} := z_k$, $k := k+1$.
4. *Compute the search direction* d_k *as*

$$d_k = -{F_\beta}'(z_k)^{-1} F_\beta(z_k)$$

5. *Find the smallest* $i_k \in \mathbb{N}_0$, *such that*

$$\Theta_{\beta_k}(z_k + \beta_N^{i_k} d_k) \leq (1 - 2\sigma_N \beta_N^{i_k}) \Theta_{\beta_k}(z_k)$$

   and set $\alpha_N^k := \beta_N^{i_k}$.
6. *Set* $z_{k+1} := z_k + \alpha_N^k d_k$, $k := k+1$, *and go to step 2.*

Again, Step 5 is equivalent to the Armijo step width, according to Lemma 5.26.

5.4.1. Example: Regularized Minimum Energy Problem

Finally, we compare the performance of the Combined Newton Method from the previous Section 5.2.2 with the Globalized Method 5.31.

The algorithm was started with the following parameters:

$$\beta_0 = 1, \qquad \epsilon = 10^{-9}, \qquad \beta_N = 0.8, \qquad \sigma_N = 10^{-2}.$$

#IT	β_k	$\|F_\beta(z_k)\|_\infty$	$\|F_0(z_k)\|_\infty$	$\|F_\beta(z_k)\|_2$	$\|F_0(z_k)\|_2$
0	$1.0000E+00$	$1.0000E+00$	$1.0000E+00$	$1.5491E+00$	$1.4142E+00$
1	$1.0000E+00$	$5.2499E-01$	$1.4434E-01$	$3.2741E-01$	$6.7182E-02$
2	$1.0000E+00$	$2.8997E-01$	$1.3398E-01$	$1.6986E-01$	$6.3318E-02$
3	$1.0000E+00$	$1.7567E-01$	$1.1935E-01$	$9.5312E-02$	$5.7463E-02$
3	$7.0197E-17$	$1.1935E-01$	$1.1935E-01$	$5.7463E-02$	$5.7463E-02$
4	$7.0197E-17$	$1.1935E-01$	$1.1935E-01$	$5.7460E-02$	$5.7460E-02$
5	$7.0197E-17$	$1.1934E-01$	$1.1934E-01$	$5.7455E-02$	$5.7455E-02$
6	$7.0197E-17$	$1.1933E-01$	$1.1933E-01$	$5.7449E-02$	$5.7449E-02$
7	$7.0197E-17$	$1.1930E-01$	$1.1930E-01$	$5.7437E-02$	$5.7437E-02$
			...		
37	$7.0197E-17$	$1.6849E-05$	$1.6849E-05$	$6.5117E-06$	$6.5117E-06$
37	$2.4729E-25$	$1.6849E-05$	$1.6849E-05$	$6.5117E-06$	$6.5117E-06$
38	$2.4729E-25$	$9.4896E-05$	$9.4896E-05$	$3.4182E-06$	$3.4182E-06$
39	$2.4729E-25$	$3.6907E-07$	$3.6907E-07$	$1.2626E-07$	$1.2626E-07$
40	$2.4729E-25$	$1.6948E-08$	$1.6948E-08$	$6.3788E-09$	$6.3788E-09$
40	$8.6096E-28$	$1.6948E-08$	$1.6948E-08$	$6.3788E-09$	$6.3788E-09$
40	$2.9976E-30$	$1.6948E-08$	$1.6948E-08$	$6.3788E-09$	$6.3788E-09$
41	$2.9976E-30$	$2.8507E-10$	$2.8507E-10$	$8.7271E-22$	$8.7271E-22$

Table 5.3.: Errors of the Globalized Method for the regularized problem

Naturally, the algorithm was again started with $z^0 = 0$.

Figure 5.4 depicts the iterates generated by the algorithm. The first column in Table 5.3 shows the regularization parameter β, and the subsequent columns show the local and global residua, in both the $\|\cdot\|_\infty$-norm and the $\|\cdot\|_2$-norm.

Whenever β was updated, a new row was used as the local errors changed (in this case, the iteration number as well as the global errors remain the same). The number of iterations needed for $\|F_0(z_k)\|_\infty$ to decrease below 10^{-9} is smaller than in the previous methods. A reason for this effect is certainly that the regularization parameter β is set to 10^{-17} after the third iteration, so that its influence on the numerics pratically vanishes. The plots as well as the $\|\cdot\|_{Y^2}$-norm residua show that the merit function actually slows down the convergence after this point, so that it takes more than 30 iterations until notable progress is made. Since in the early iterations only small step sizes are accepted, the program has to check many possible step sizes, which leads to a longer computation time. In fact, the computations with the Combined Newton method took 9.81 seconds on an i7 processor (at 2.93Ghz), while the Globalized method took 12.65 seconds on the same machine. The price to pay for the global convergence seems to be slower convergence under rough conditions. A positive aspect of this phenomenon this is that the local and global residua decrease consistently (although this is not guaranteed by the algorithm).

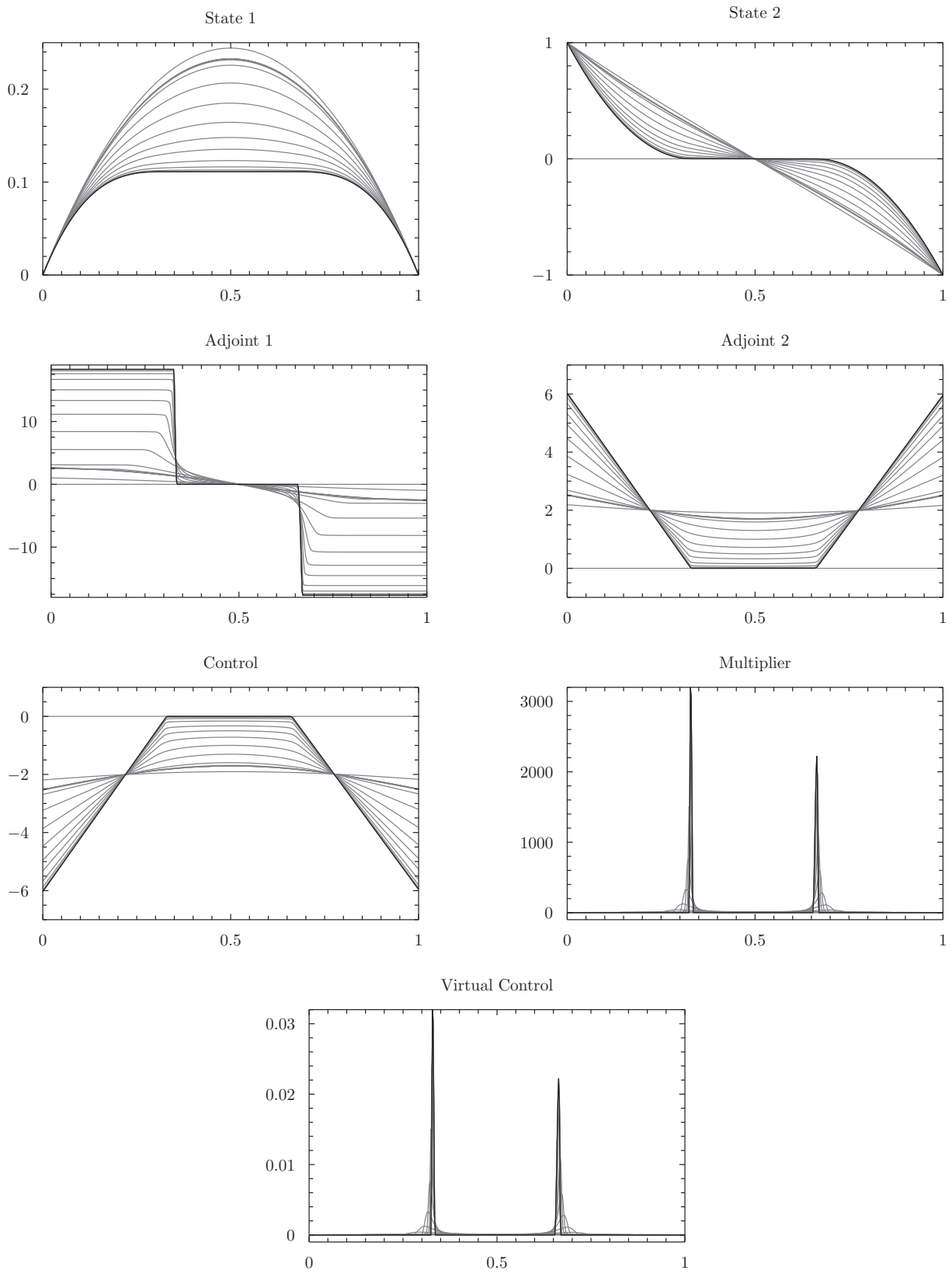


Figure 5.4.: Iterations of the Globalized Method for the regularized Minimum Energy Problem

6. Numerical Aspects

The focus of this work lies in the indirect solution approach to optimal control problems. Hence, the theory focuses on the analytic properties of the problems and necessary optimality conditions that are used to derive the various algorithms. Numerical solutions of the problems necessarily involve discretization, which for this type of approach means that the Newton direction is calculated by discretizing Equation (5.8).

In contrast to our approach, direct discretization method is introduces in the first part of this chapter. Here, the optimal control problem itself is discretized, so that necessary optimality conditions in finite dimensions can be used to create solution algorithms. The second part of this chapter shows without proof of convergence how the Newton direction is calculated in the indirect method. As the Globalized Newton Method 5.27 from the previous chapter makes use of a merit function, the discretization of the Newton direction in this context is fitted to the direct approach in order to insure that the step size $\alpha_N = 1$ is accepted.

6.1. The Direct Discretization Approach

The basic idea of this approach is to discretize the optimal control problem directly, yielding an optimization problem in finite dimensions. In the next step, Newton's method is used to find a solution to the necessary optimality conditions for the finite problem.

In [MBM97] and [Ger06], the minimum principle and convergence results for the Euler discretized problem are presented and proved. As we deal with problems of the class LQOCP$_s$ 5.1, the results are presented for this particular class of problems.

For the Euler discretization, the discretized LQOCP reads:

Problem 6.1 (*DLQOCP*)

$$\begin{aligned} \text{min!} \quad J^h(x,v) :=& \frac{1}{2} x(t_N)^\top Q_f x(t_N) \\ &+ \frac{1}{2} h \sum_{i=0}^{N-1} (x_i^\top, v_i^\top) \begin{pmatrix} Q(t_i) & R_v(t_i) \\ R_v(t_i)^\top & S_v(t_i) \end{pmatrix} \begin{pmatrix} x_i \\ v_i \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \textit{with respect to the grid functions} \quad & x_h : \mathbb{G}_h \to \mathbb{R}^{n_x}, \ x_i := x_h(t_i), \ i = 0, \dots, N \\ \textit{and} \quad & v_h : \mathbb{G}_h \to \mathbb{R}^{n_v}, \ v_i := v_h(t_i), \ i = 0, \dots, N-1 \end{aligned}$$

subject to the difference equations

$$\frac{x_{i+1} - x_i}{h} = A(t_i)x_i + B_v(t_i)v_i \quad i = 0, \ldots, N-1.$$

boundary conditions

$$E_0 x_0 + E_1 x_N = f,$$

and mixed control state constraints

$$G(t_i)x(t_i) + H(t_i)v_i \leq l(t_i) \quad i = 0, \ldots, N-1.$$

The following assumptions are crucial for the proof of convergence for the above problem and were derived from the article by Malanowski et al. [MBM97, Assumptions II.1–II.6]. Firstly, the functions occuring in the problem need to be sufficiently smooth.

Assumption 6.2

1. *The functions Q, R_v, S_v, A, B_v as well as G and H are constant in time. The matrices S_v and Q are symmetric. E_0 and E_1 have the structure*

$$E_0 = \begin{pmatrix} I_{n_x} \\ 0_{n_x} \end{pmatrix}, \qquad E_1 = \begin{pmatrix} 0 \\ E' \end{pmatrix},$$

   with $E' \in \mathbb{R}^{n_c \times n_x}$ arbitrary. We write $f = (f_1, f_2)^\top$ with $f_1, f_2 \in \mathbb{R}^{n_x}$.

2. *There exists a possibly local solution*

$$(x_0, v_0) \in \mathcal{C}^1([t_0, t_f], \mathbb{R}^{n_x}) \times \mathcal{C}([t_0, t_f], \mathbb{R}^{n_v}).$$

As the class of problems in this chapter is restricted to optimality problems with linear constraints, normality of the multipliers in the necessary optimality conditions holds, without further regularity assumptions [GK02, Th. 2.42].

The KKT conditions for the discretized problem read (cf. [MBM97, Equ. (4.8),(4.9)]):

Theorem 6.3
Let Assumption 6.2 be satisfied by the data of the optimal control problem.

Let $(\hat{x}_h, \hat{v}_h)$ be a local minimum of problem 4.1.

Then there exist multipliers $\kappa \in \mathbb{R}^{n_E}$, $\lambda = (\lambda_0, \ldots, \lambda_N) \in \mathbb{R}^{n_x \times (N+1)}$, $\zeta = (\zeta_0, \ldots, \zeta_{N-1}) \in \mathbb{R}^{n_c \times N}$, $\kappa_1, \kappa_2 \in \mathbb{R}^{n_x}$, such that the following conditions are satisfied:

1. *Discrete adjoint equation:*

$$\frac{\lambda_{i+1} - \lambda_i}{h} = -\left(Q_i \hat{x}_{hi} + R_i \hat{v}_{hi} + A_i^\top \lambda_{i+1} + G_i^\top \zeta_i\right) \qquad \text{for all } i = 0, \ldots, N-1$$

2. *Transversality conditions:*

$$\begin{aligned} \lambda_0 &= -\kappa_1 \\ \lambda_N &= \ Q_f \hat{x}_{hN} + {E'}^\top \kappa_2 \end{aligned}$$

3. *Optimality conditions:*

$$R_i^\top \hat{x}_{hi} + S_i \hat{v}_{hi} + B_i^\top \lambda_{i+1} + H_i^\top \zeta_i = 0 \qquad \text{for all } i = 0, \dots, N-1$$

4. *Complementarity conditions:*

$$0 \le \zeta_i \perp l(t_i) - G(t_i)\hat{x}_{hi} - H(t_i)\hat{v}_{hi} \ge 0 \qquad \text{for all } i = 0, \dots, N-1$$

The following assumptions make use of the definition of the set of (α-) active controls, cf. 3.19. They firstly insure that the control v has a sufficient influence on the active constraints (this influence is defined by the matrix H). The second part is a controllability assumption.

Assumption 6.4

1. *There exists a positive constant $\gamma > 0$, such that*

$$\left| H_{I_0(t)}(t)^\top z \right| \ge \gamma |z| \quad \text{for all } z \in \mathbb{R}^{\#I_\sigma(t)} \text{ and almost all } t \in [t_0, t_f].$$

2. *For any $e \in \mathbb{R}^{n_c}$, the following boundary value system has a solution:*

$$\begin{aligned} \dot{y}(t) &= \tilde{A}(t)y(t) + \tilde{B}_v(t) \\ y(t_0) &= 0, \quad E'y(t_f) = e, \end{aligned}$$

where

$$\begin{aligned} \tilde{A}(t) =& A(t) - B_v(t)H_{I_0(t)}^\top (H_{I_0(t)}H_{I_0(t)}^\top)^{-1} H_{I_0(t)}, \\ \tilde{B}_v(t) =& B_v(t)\left(I_{n_v} - H_{I_0(t)}^\top (H_{I_0(t)}H_{I_0(t)}^\top)^{-1} H_{I_0(t)}\right). \end{aligned}$$

Finally, we make an assumption that guarantees the solvability of a linearization. In order to formulate the according assumption, we introduce the following index set of positive multipliers:

Definition 6.5
Let

$$I_+ := \{i \in I_0(t) | \eta_i(t) > 0\}$$

denote the set of indices with a positive multiplier. The multiplier η in this context denotes the multiplier for the mixed control state constraint.

With this new set of indices, we present the coercivity condition and the Riccati equation that needs to be solvable:

Assumption 6.6

1. *The Matrix S_v is positive definite.*

2. The Riccati equation

$$\begin{aligned}\dot{P}(t) =& - P(t)A(t) - A(t)^\top P(t) - Q(t) \\ &+ \left[\begin{pmatrix} R_v(t)^\top \\ G_{I_+(t)}(t)\end{pmatrix}^\top + P(t)\begin{pmatrix} B_v(t)^\top \\ 0\end{pmatrix}^\top\right] \cdot \begin{pmatrix} S_v(t) & H_{I_+(t)}(t)^\top \\ H_{I_+(t)}(t) & 0\end{pmatrix} \\ &\cdot \left[\begin{pmatrix} B_v(t)^\top \\ 0\end{pmatrix} P(t) + \begin{pmatrix} R_v(t)^\top \\ G_{I_+(t)}(t)\end{pmatrix}\right]\end{aligned}$$

possesses a bounded solution P on $[t_0, t_f]$ that satisfies

$$d^\top(Q_f - P(t_f))d \geq 0 \quad \forall d \in \mathbb{R}^{n_x} : E_1 d = 0.$$

Remark 6.7
Assumption 6.2 already requires S_v to be constant in time. Therefore, the coercivity condition in [MBM97] and [Ger06] is always satisfied if S_v is positive definite, as $\mathcal{H}''_{vv} = S_v$.

Under the above assumptions, convergence of the Euler discretization has been shown in [MBM97, Theorem 5.7]:

Theorem 6.8 (Convergence of the Euler discretization)
If Assumptions 6.2, 6.4 and 6.6 hold, then there exists $\tilde{h} > 0$ such that for each $h < \tilde{h}$, there exists a locally unique KKT point $(x_h, v_h, \kappa_h, \lambda_h, \zeta_h)$ of $(DLQOCP)$, and

$$\begin{aligned}\|x_h - x_0\|_{1,\infty}, \|v_h - v_0\|_\infty &\leq l' |h|, \\ |\kappa_h - \kappa_0|, \|\lambda_h - \lambda_0\|_{1,\infty}, \|\zeta_h - \zeta_0\|_\infty &\leq l' |h|,\end{aligned}$$

where l' is independent of h.

Theorem 6.8 together with Theorem 6.3 lead to another way of solving the Optimal Control Problem: After transforming the KKT conditions into equations, we can solve the conditions using a globalized Newton's method. At the same time, a comparison between the direct and the indirect approach leads to a formula for the discretized problem and a merit function that fits the problem exactly. In order to compare the Newton direction for the direct approach with the Newton direction for the indirect approach, we again use the regularization parameter β and define the operator F_h^β:

Definition 6.9
Let

$$\begin{aligned}z_h &:= (x_h, v_h, \kappa, \lambda, \zeta)^\top \in X_h, \\ X_h &:= \mathbb{R}^{n_x \cdot (N+1)} \times \mathbb{R}^{n_v \cdot (N+1)} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x \cdot (N+1)} \times \mathbb{R}^{n_c \cdot (N+1)}, \\ Y_{h1} &:= \mathbb{R}^{n_x \cdot N} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x \cdot N} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_v \cdot (N+1)}, \\ Y_{h2} &:= \mathbb{R}^{n_c \cdot (N+1)}.\end{aligned}$$

Let the (regularized) discrete NCP function $\omega_h^\beta : \left(\mathbb{R}^{n_c \cdot N}\right)^2 \to \mathbb{R}^{n_c \cdot N}$ be defined as

$$\omega_h^\beta\left((a_i)_{i=0}^{n_c \cdot (N-1)}, (b_i)_{i=0}^{n_c \cdot (N-1)}\right) := (\phi_\beta(a_i, b_i))_{i=0}^{n_c \cdot (N-1)}$$

with $\phi_\beta : \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ (a,b) \mapsto \sqrt{a^2+b^2+\beta} - a - b.$

Let $F_h^\beta = (F_{h\,1}^\beta, F_{h\,2}^\beta)^\top : X_h \to Y_{h1} \times Y_{h2}$ *be defined by*

$$F_{h\,1}^\beta(z_h) := \begin{pmatrix} \left(\frac{x_{hi+1}-x_{hi}}{h} - A_i x_{hi} - B_{vi} v_{hi}\right)_{i=0}^{N-1} \\ x_{h0} - f_1 \\ E' x_{hN} - f_2 \\ \left(\frac{\lambda_{i+1}-\lambda_i}{h} + Q_i x_{hi} + R_{vi} v_{hi} + A_i^\top \lambda_{i+1} + G_i^\top \zeta_i\right)_{i=0}^{N-1} \\ \lambda_0 + \kappa_1 \\ \lambda_N - Q_f x_{hN} - {E'}^\top \kappa_2 \\ \left(R_{v_i}^\top x_{hi} + S_{vi} v_{hi} + B_{v_i}^\top \lambda_{i+1} + H_i^\top \zeta_i\right)_{i=0}^{N-1} \end{pmatrix},$$

$$F_{h\,2}^\beta(z_h) := \omega_h^\beta\left(\zeta, (l(t_i) - G(t_i)x_{hi} - H(t_i)v_{hi})_{i=0}^{N-1}\right).$$

Now we need to find an algorithm that solves the equation $F_h^\beta(z) = 0$. As so far, algorithms based on the Newton method have been used in the function space setting, it seems natural to also use a suitable Newton method in the context of finite dimensional spaces.

The unregularized function F_h^0 is not differentiable, but it is still differentiable in any direction. In this case, an algorithm based on a Newton method can make use of the Bouligand subdifferential (see [Taw09]):

Definition 6.10 (Bouligand Subdifferential)
Let $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ *be locally Lipschitzian at each point of an open set* Ω*. For* $x^* \in \Omega$*, define the Bouligand subdifferential of* F *at* x^* *by*

$$\partial_B F(x^*) = \left\{ \lim_{k\to\infty} \nabla F(x^k) : x_k \to x^*, x_k \in \Omega_F \right\},$$

where Ω_F *is the set of all points in* Ω *where* F *is Fréchet differentiable.*

In [Kan00], a globalized nonsmooth Newton method for mixed complementarity problems is presented. The necessary optimality conditions can be stated in this form, and the cited Algorithm [Kan00, Algorithm 2.1] works by applying and globalizing a Newton method to the unregularized equation $F_h^0(z) = 0$.

The merit function for the globalization is defined analogously to the merit function in the indirect approach:

For $z \in X_h$, let

$$\Theta(z) := \frac{1}{2}\left\|F_h^0(z)\right\|_2^2.$$

Algorithm 6.11 (Globalized Semismooth Newton Method in Finite Spaces)

1. *Choose* $z_0 \in \mathbb{R}^n$*,* $\rho > 0$*,* $\beta \in (0,1)$*,* $\sigma_N \in (0, 1/2)$*,* $p > 2$ *and* $\varepsilon > 0$*. Let* $k := 0$*.*
2. *If* $\|F_h^0(z_k)\| \le \varepsilon$*, stop.*

3. *Select an element $V_k \in \partial_B F_h^0(z_k)$. Compute the Newton direction d_k as a solution to*

$$V_k d = -F_h^0(z_k). \tag{6.1}$$

If the system is not solvable or the condition

$$\nabla\Theta(z_k)^\top d_k \leq -\rho\|d_k\|^p$$

is not satisfied, set $d_k := -\nabla\Theta(z_k)$.

4. *Find the smallest $i_k \in N_0$, such that*

$$\Theta(z_k + \beta_N^{i_k} d_k) \leq \Theta(z_k) + \beta_N^{i_k} \nabla\Theta(z_k)^\top d_k.$$

5. *Set $z_{k+1} := z_k + \beta_N^{i_k} d_k$, $k = k+1$, and go to 2.*

The convergence properties are claimed to hold without very rigid assumptions (see [Kan00, Theorem 2.2]):

Theorem 6.12
Every accumulation point of a sequence (z_k) generated by Algorithm 6.11 is a stationary point of Θ, and such a stationary point is a solution of the mixed complementarity problem under relative mild assumptions.

Finally, we state the linear system (6.1) that determines the Newton direction. Let z_k be the current iterate, and let

$$\begin{aligned} r_j(t_i) &\in \partial_B \varrho_{0a}'(\eta_k(t_i), l(t_i) - G(t_i)x_k(t_i) - H(t_i)v_k(t_i)), \qquad \mathbf{r} := \operatorname{diag}(r_1, \ldots, r_{n_c}), \\ s_j(t_i) &\in \partial_B \varrho_{0b}'(\eta_k(t_i), l(t_i) - G(t_i)x_k(t_i) - H(t_i)v_k(t_i)), \qquad \mathbf{s} := \operatorname{diag}(s_1, \ldots, s_{n_c}). \end{aligned}$$

Then the system reads

$$\frac{d_{xi+1} - d_{xi}}{h} - A_i d_{xi} - B_{vi} d_{vi} = y_{1i}, \quad i = 0, \ldots, N-1, \tag{6.2}$$

$$\frac{d_{\lambda i+1} - d_{\lambda i}}{h} + Q_i d_{xi} + R_{vi} d_{vi} + A_i^\top d_{\lambda i+1} + G_i^\top d_{\zeta_i} = y_{2i}, \quad i = 0, \ldots, N-1, \tag{6.3}$$

$$d_{x0} = y_3, \tag{6.4}$$

$$E' d_{xN} = y_4, \tag{6.5}$$

$$d_{\lambda 0} + d_{\kappa 1} = y_5, \tag{6.6}$$

$$d_{\lambda N} - Q_f d_{xN} - {E'}^\top d_{\kappa 2} = y_6, \tag{6.7}$$

$$R_{vi}^\top d_{xi} + S_{vi} d_{vi} + B_{vi}^\top d_{\lambda i+1} + H_i^\top d_{\zeta_i} = y_{7i}, \quad i = 0, \ldots, N-1, \tag{6.8}$$

$$-\mathbf{s}_i H_i d_{vi} + \mathbf{r}_i d_{\zeta_i} - \mathbf{s}_i G_i d_{xi} = y_{8i}, \quad i = 0, \ldots, N-1, \tag{6.9}$$

where the right hand sides y_{j_i} and y_j are defined as follows:

$$y_{1i} := -\frac{x_{ki+1} - x_{ki}}{h} + A_i x_{ki} + B_{vi} v_{ki}, \qquad i = 0, \ldots, N-1,$$

$$
\begin{aligned}
y_{2i} &:= -\frac{\lambda_{ki+1} - \lambda_{ki}}{h} - Q_i x_{ki} - R_{vi} v_{ki} - A_i^\top \lambda_{ki+1} - G_i^\top \zeta_{ki}, \qquad i = 0, \dots, N-1,\\
y_3 &:= -x_{k0} + f_1,\\
y_4 &:= -E' x_{kN} + f_2,\\
y_5 &:= -\lambda_{k0} - \kappa_{k1},\\
y_6 &:= -\lambda_{kN} + Q_f x_{kN} - {E'}^\top \kappa_{k2},\\
y_{7i} &:= -{R_{vi}}^\top x_{ki} - S_{vi} v_{ki} - {B_{vi}}^\top \lambda_{ki+1} - H_i^\top \zeta_{ki}, \qquad i = 0, \dots, N-1,\\
y_{8i} &:= -\omega_h^0(\zeta_k, (l_i - G_i x_{ki} - H_i v_{ki})_{i=0}^{N-1}).
\end{aligned}
$$

6.2. Computing the Search Direction

In this section, we discuss algorithms that can be used to find the search direction of the Newton method. Let an LQOCP$_s$ of the form 5.1 be given. Let $z_k = (x_k, \lambda_k, v_k, \sigma_k, \eta_k)$ be the current iterate. Then the search direction $d_k = (d_x, d_\lambda, d_v, d_\sigma, d_\eta)$ is

$$d_k = -F_\beta'(z_k)^{-1} F_\beta(z_k). \tag{5.8}$$

Again with

$$
\begin{aligned}
r_i(\cdot) &:= {\varrho_\beta}_a'(\eta_k(\cdot), l(\cdot) - G(\cdot)x_k(\cdot) - H(\cdot)v_k(\cdot)), \qquad & \mathbf{r} &:= \operatorname{diag}(r_1, \dots, r_{n_c}),\\
s_i(\cdot) &:= {\varrho_\beta}_b'(\eta_k(\cdot), l(\cdot) - G(\cdot)x_k(\cdot) - H(\cdot)v_k(\cdot)), \qquad & \mathbf{s} &:= \operatorname{diag}(s_1, \dots, s_{n_c}),
\end{aligned}
$$

this is a solution to the differential algebraic equation

$$
\begin{pmatrix} \dot{d_x} \\ \dot{d_\lambda} \end{pmatrix} = \begin{pmatrix} A & 0 \\ -Q & -A^\top \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} + \begin{pmatrix} B_v & 0 \\ -R_v^\top & -G^\top \end{pmatrix} \begin{pmatrix} d_v \\ d_\eta \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \tag{6.10}
$$

$$E_0 d_x(t_0) + E_1 d_x(t_f) = y_3, \tag{6.11}$$

$$d_\lambda(t_0) + E_0^\top d_\sigma = y_4, \tag{6.12}$$

$$d_\lambda(t_f) - Q_f d_x(t_f) - E_1^\top d_\sigma = y_5, \tag{6.13}$$

$$
\begin{pmatrix} S_v & H^\top \\ -\mathbf{s}H & \mathbf{r} \end{pmatrix} \begin{pmatrix} d_v \\ d_\eta \end{pmatrix} = \begin{pmatrix} -R_v^\top & -B_v^\top \\ \mathbf{s}G & 0 \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} + \begin{pmatrix} y_6 \\ y_7 \end{pmatrix}, \tag{6.14}
$$

where

$$
\begin{aligned}
y_1 &:= -\dot{x}_k + Ax_k + B_v v_k & y_5 &:= -\lambda_k(t_f) + Q_f x_k(t_f) + E_1^\top \sigma_k\\
y_2 &:= -\dot{\lambda}_k - Qx_k - R_v v_k - A^\top \lambda_k - G^\top \eta_k & y_6 &:= -S_v v_k - R_v^\top x_k - B_v^\top \lambda_k - H^\top \eta_k\\
y_3 &:= -E_0 x_k(t_0) - E_1 x_k(t_f) + f & y_7 &:= -\omega_\beta(\eta_k, l - Gx_k - Hv_k).\\
y_4 &:= -\lambda_k(t_0) - E_0^\top \sigma_k
\end{aligned}
$$

As the above system is linear in d_z, discretization leads to a system of linear equations. We apply the Euler method to the differential equation for x, and, inspired by the direct discretization approach, we alter the discretization of the adjoint equation to match (6.3). Then we transform the algebraic equations (6.14) into algebraic equations on the grid.

Let $n \in \mathbb{N}$, $h := \frac{t_f - t_0}{n-1}$ and let Γ be the grid that devides the interval $[t_0, t_f]$ into n equidistant time steps, i.e. $\Gamma := \{t_0 + h \cdot i \mid i = 0, \dots, n-1\}$.

For $k \in \mathbb{N}$, the last (descrete) iterate is denoted by $z_k^\Gamma = (x_k^\Gamma, \lambda_k^\Gamma, v_k^\Gamma, \sigma_k^\Gamma, \eta_k^\Gamma)$, with $x_k^\Gamma \in \mathbb{R}^{n \cdot n_x}$, $\lambda_k^\Gamma \in \mathbb{R}^{n \cdot n_x}$, $v_k^\Gamma \in \mathbb{R}^{(n-1) \cdot n_v}$, $\sigma_k^\Gamma \in \mathbb{R}^{n_E}$ and $\eta_k^\Gamma \in \mathbb{R}^{(n-1) \cdot n_c}$.

We denote the discrete search direction by $d_z^\Gamma = (d_x^\Gamma, d_\lambda^\Gamma, d_v^\Gamma, d_\sigma^\Gamma, d_\eta^\Gamma)$, again with $d_x^\Gamma \in \mathbb{R}^{n \cdot n_x}$, $d_\lambda^\Gamma \in \mathbb{R}^{n \cdot n_x}$, $d_v^\Gamma \in \mathbb{R}^{(n-1) \cdot n_v}$, $d_\sigma^\Gamma \in \mathbb{R}^{n_E}$ and $d_\eta^\Gamma \in \mathbb{R}^{(n-1) \cdot n_c}$.

In order to find a consistent discretization for the differential equation (6.10), we use the same method as in equation (6.2). The discretized differential equations for the discrete search direction of the state d_x^Γ and the adjoint d_λ^Γ read:

$$\frac{d_{x\,i+1}^\Gamma - d_{x\,i}^\Gamma}{h} - A_i d_{x\,i}^\Gamma - B_{v\,i} d_{v\,i}^\Gamma = y_{1i}, \quad i = 0, \dots, n-2, \tag{6.15a}$$

$$\frac{d_{\lambda\,i+1}^\Gamma - d_{\lambda\,i}^\Gamma}{h} + Q_i d_{x\,i}^\Gamma + R_{v\,i} d_{v\,i}^\Gamma + A_i^\top d_{\lambda\,i+1}^\Gamma + G_i^\top d_{\eta\,i}^\Gamma = y_{2i}, \quad i = 0, \dots, n-2, \tag{6.15b}$$

for $i = 0, \dots, n-2$, where the right hand sides are

$$\begin{aligned} y_{1i} &:= -\frac{x_{k\,i+1}^\Gamma - x_{k\,i}^\Gamma}{h} + A_i x_{k\,i}^\Gamma + B_{v\,i} v_{k\,i}^\Gamma, \qquad i = 0, \dots, n-2, \\ y_{2i} &:= -\frac{\lambda_{k\,i+1}^\Gamma - \lambda_{k\,i}^\Gamma}{h} - Q_i x_{k\,i}^\Gamma - R_v v_{k\,i}^\Gamma - A_i^\top d_{\lambda\,i+1}^\Gamma - G_i^\top d_{\eta\,i}^\Gamma, \qquad i = 0, \dots, n-2. \end{aligned}$$

Note that in the Euler discretization the control variables only influence the right hand side of the differential equation up to the time step t_{n-2}. Hence, the control variables that are calculated are $v_0, \dots, v_{n-2}$. Consequently, the search direction is $d_{v0}, \dots, d_{vn-2}$ as well. The boundary conditions (6.11)-(6.13) with respect to the discretized variables read

$$\begin{pmatrix} E_0 & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix} \begin{pmatrix} d_{x\,0}^\Gamma \\ d_{\lambda\,0}^\Gamma \end{pmatrix} + \begin{pmatrix} E_1 & 0 \\ 0 & 0 \\ -Q_f & I \end{pmatrix} \begin{pmatrix} d_{x\,n-1}^\Gamma \\ d_{\lambda\,n-1}^\Gamma \end{pmatrix} + \begin{pmatrix} 0 \\ E_0^\top \\ -E_1^\top \end{pmatrix} d_\sigma^\Gamma = \begin{pmatrix} y_3 \\ y_4 \\ y_5 \end{pmatrix} \tag{6.16}$$

with

$$\begin{aligned} y_3 &:= -E_0 x_{k\,0}^\Gamma - E_1 x_{k\,n-1}^\Gamma + f \\ y_4 &:= -\lambda_{k\,0}^\Gamma - E_0^\top \sigma_k^\Gamma \\ y_5 &:= -\lambda_{k\,n-1}^\Gamma + Q_f x_{k\,n-1}^\Gamma + E_1^\top . \sigma_k^\Gamma \end{aligned}$$

Finally, the algebraic equation (6.14) has to hold where the control variables are calculated, i.e. from the time step t_0 to t_{n-2}:

$$\begin{pmatrix} S_{v_i} & H_i^\top \\ -\mathbf{s}_i H_i & \mathbf{r}_i \end{pmatrix} \begin{pmatrix} d_{v\,i}^\Gamma \\ d_{\eta_i}^\Gamma \end{pmatrix} = \begin{pmatrix} -R_{vi}^\top & -B_{vi}^\top \\ \mathbf{s}_i G_i & 0 \end{pmatrix} \begin{pmatrix} d_{x\,i}^\Gamma \\ d_{\lambda\,i+1}^\Gamma \end{pmatrix} + \begin{pmatrix} y_{6i} \\ y_{7i} \end{pmatrix} \tag{6.17}$$

for $i = 0, \dots, n-2$ with the right hand sides

$$y_{6i} := -S_{vi} v_{k\,i}^\Gamma - R_{vi}{}^\top x_{k\,i}^\Gamma - B_{vi}{}^\top \lambda_{k\,i+1}^\Gamma - H_i^\top \eta_{k\,i}^\Gamma$$

$$y_{7i} := -\varrho_\beta(\eta^\Gamma_{k\,i}, l(t_i) - G_i x^\Gamma_{k\,i} - H_i v^\Gamma_{k\,i}).$$

Equations (6.15)-(6.17) can be combined in one linear equation. Let

$$M := \begin{pmatrix} M_{11} & M_{12} & 0 \\ M_{21} & M_{22} & 0 \\ M_{31} & 0 & M_{33} \end{pmatrix},$$

where

$$M_{11} := \begin{pmatrix} \begin{pmatrix} -I_{n_x} - hA_0 & 0 \\ Q_0 & -I_{n_x} \end{pmatrix} & \begin{pmatrix} I_{n_x} & 0 \\ 0 & I_{n_x} + hA_0^\top \end{pmatrix} & & \\ & \ddots & \ddots & \\ & & \begin{pmatrix} -I_{n_x} - hA_{n-2} & 0 \\ Q_{n-2} & -I_{n_x} \end{pmatrix} & \begin{pmatrix} I_{n_x} & 0 \\ 0 & I_{n_x} + hA_{n-2}^\top \end{pmatrix} \end{pmatrix},$$

$$M_{12} := \begin{pmatrix} h\begin{pmatrix} -B_{v0} & 0 \\ R_{v0} & G_0^\top \end{pmatrix} & & \\ & \ddots & \\ & & h\begin{pmatrix} -B_{vn-2} & 0 \\ R_{vn-2} & G_{n-2}^\top \end{pmatrix} \end{pmatrix},$$

$$M_{21} := \begin{pmatrix} \begin{pmatrix} R_{v0}^\top & 0 \\ -\mathbf{s}_0 G_0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & B_{v0}^\top \\ 0 & 0 \end{pmatrix} & & \\ & \ddots & \ddots & \\ & & \begin{pmatrix} R_{vn-2}^\top & 0 \\ -\mathbf{s}_{n-2} G_{n-2} & 0 \end{pmatrix} & \begin{pmatrix} 0 & B_{vn-2}^\top \\ 0 & 0 \end{pmatrix} \end{pmatrix},$$

$$M_{22} := \begin{pmatrix} \begin{pmatrix} S_{v0} & H_0^\top \\ -\mathbf{s}_0 H_0 & \mathbf{r}_0 \end{pmatrix} & & \\ & \ddots & \\ & & \begin{pmatrix} S_{vn-2} & H_{n-2}^\top \\ -\mathbf{s}_{n-2} H_{n-2} & \mathbf{r}_{n-2} \end{pmatrix} \end{pmatrix},$$

$$M_{31} := \begin{pmatrix} E_0 & 0 & 0 & \dots & 0 & E_1 & 0 \\ 0 & I_{n_x} & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & -Q_f & I_{n_x} \end{pmatrix}$$

$$M_{33} := \begin{pmatrix} 0 \\ E_0^\top \\ -E_1^\top \end{pmatrix}$$

so that the block matrices have the following dimensions:

$$\begin{aligned} &M_{11} \in \mathbb{R}^{(n_t-1)\cdot(2n_x) \times n_t\cdot(2n_x)}, && M_{12} \in \mathbb{R}^{(n_t-1)\cdot(2n_x) \times (n_t-1)\cdot(n_v+n_c)}, \\ &M_{21} \in \mathbb{R}^{(n_t-1)\cdot(n_v+n_c) \times n_t\cdot(2n_x)}, && M_{22} \in \mathbb{R}^{(n_t-1)\cdot(n_v+n_c)\times(n_t-1)\cdot(n_v+n_c)}, \\ &M_{31} \in \mathbb{R}^{(n_E+2n_x) \times n_t\cdot(2n_x)}, && M_{33} \in \mathbb{R}^{(n_E+2n_x) \times n_E}. \end{aligned}$$

Let $r \in \mathbb{R}^{(n_t-1)\cdot(2n_x)+(n_t-1)\cdot(n_v+n_c)+n_E+n_x+n_x}$,

$$r := (y_{10}, y_{20}, y_{11}, y_{21}, \ldots, y_{1n-1}, y_{2n-1},\ y_{60}, y_{70}, \ldots, y_{6n-2}, y_{7n-2},\ y_3, y_4, y_5)^\top.$$

Then the discretized search direction $d^\Gamma \in \mathbb{R}^{n_t\cdot(2n_x)+(n_t-1)\cdot(n_v+n_c)+n_E}$

$$d_z^\Gamma = (d_{x_0}^\Gamma, d_{\lambda_0}^\Gamma, \ldots, d_{x_{n-1}}^\Gamma, d_{\lambda_{n-1}}^\Gamma,\ d_{v_0}^\Gamma, d_{\eta_0}^\Gamma, \ldots, d_{v_{n-2}}^\Gamma, d_{\eta_{n-2}}^\Gamma,\ d_\sigma^\Gamma)^\top$$

solves

$$M d_z^\Gamma = r. \tag{6.18}$$

In [Ger08], another approach has been introduced that reduces the size of the linear system by reducing the DAE to a linear boundary value problem. If the inverse of

$$\mathcal{A} := \begin{pmatrix} S & H^\top \\ -\mathbf{s}H & \mathbf{r} \end{pmatrix}$$

exists, then the algebraic equation (6.14) can be solved for d_v and d_η.

Substituting the respective terms in the ODE (6.10) yields:

$$\begin{aligned}
\begin{pmatrix} \dot{d}_x \\ \dot{d}_\lambda \end{pmatrix} &= \begin{pmatrix} A & 0 \\ -Q & -A^\top \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} \\
&\quad + \begin{pmatrix} B_v & 0 \\ -R_v^\top & -G^\top \end{pmatrix} \mathcal{A}^{-1} \left[\begin{pmatrix} -R_v^\top & -B_v^\top \\ \mathbf{s}G & 0 \end{pmatrix} \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} + \begin{pmatrix} y_6 \\ y_7 \end{pmatrix} \right] + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
&= \left[\begin{pmatrix} A & 0 \\ -Q & -A^\top \end{pmatrix} + \begin{pmatrix} B_v & 0 \\ -R_v^\top & -G^\top \end{pmatrix} \mathcal{A}^{-1} \begin{pmatrix} -R_v^\top & -B_v^\top \\ \mathbf{s}G & 0 \end{pmatrix} \right] \begin{pmatrix} d_x \\ d_\lambda \end{pmatrix} \\
&\quad + \begin{pmatrix} B_v & 0 \\ -R_v^\top & -G^\top \end{pmatrix} \mathcal{A}^{-1} \begin{pmatrix} y_6 \\ y_7 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},
\end{aligned}$$

which can be rewritten using

$$\begin{aligned}
\mathbf{B} &:= \begin{pmatrix} A & 0 \\ -Q & -A^\top \end{pmatrix} + \begin{pmatrix} B_v & 0 \\ -R_v^\top & -G^\top \end{pmatrix} \mathcal{A}^{-1} \begin{pmatrix} -R_v^\top & -B_v^\top \\ \mathbf{s}G & 0 \end{pmatrix} \\
\mathbf{b} &:= \begin{pmatrix} B_v & 0 \\ -R_v^\top & -G^\top \end{pmatrix} \mathcal{A}^{-1} \begin{pmatrix} y_6 \\ y_7 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.
\end{aligned}$$

Application of the single shooting method with Euler discretization on the grid Γ for this problem leads to the linear equation for d_x, d_λ and d_σ:

$$\begin{pmatrix} I + h_0\mathbf{B}_0 & -I & & & & & \\ & I + h_1\mathbf{B}_1 & -I & & & & \\ & & \ddots & \ddots & & & \\ & & & I + h_{n-2}\mathbf{B}_{n-2} & -I & & \\ \begin{pmatrix} E_0 & 0 \\ 0 & I \\ 0 & 0 \end{pmatrix} & & & & \begin{pmatrix} E_1 & 0 \\ 0 & 0 \\ -Q_f & I \end{pmatrix} & \begin{pmatrix} 0 \\ E_0^\top \\ -E_1^\top \end{pmatrix} & \end{pmatrix} \begin{pmatrix} d_{x0}^\Gamma \\ d_{\lambda 0}^\Gamma \\ \vdots \\ d_{xn-1}^\Gamma \\ d_{\lambda n-1}^\Gamma \\ d_\sigma^\Gamma \end{pmatrix}$$

$$= \begin{pmatrix} h_0\mathbf{b}_0 \\ \vdots \\ \vdots \\ h_{n-2}\mathbf{b}_{n-2} \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$$

Finally, the discretized control d_v^Γ and the multiplier d_η^Γ can be calculated solving the algebraic equation 6.14:

$$\begin{pmatrix} d_{vi}^\Gamma \\ d_{\eta i}^\Gamma \end{pmatrix} = \mathcal{A}_i^{-1} \left[\begin{pmatrix} -R_{vi}^\top & -B_{vi}^\top \\ \mathbf{s}_i G_i & 0 \end{pmatrix} \begin{pmatrix} d_{xi}^\Gamma \\ d_{\lambda i}^\Gamma \end{pmatrix} + \begin{pmatrix} y_{6i} \\ y_{7i} \end{pmatrix} \right]$$

for all $i = 0, \ldots, n-2$.

Naturally, this approach could also be fitted to the discretization scheme that was used to derive linear system (6.18). Here, we made use of the simpler explicit Euler method for simplicity. The numerical examples stated in this work were calculated by solving the linear system (6.18).

7. Application: LQ Controller Design

In this chapter, the concept of a Linear Quadratic Controller for control-state constrained systems is introduced. This presents an application for function space Newton methods. In numerical tests, the globalized Newton method for a fixed regularization of the NCP function showed the best performance and stability, so that this solver was used for the subsequent examples. Problems with pure state constraints may become infeasible during the passage of time. The Virtual Control Concept from chapter 4.2 turns out to fix this problem so that a solution can be calculated. Since the regularized problem then is constrained by mixed control-state constraints, the said method can be used again.

The concept of simulation shall be explained as well, before numerical examples show the effects of the different parameters. In this chapter, we use the letter u for the control, as this is traditionally used in controller design.

7.1. Controller design and Simulation

For the controller, Newton's method is used in the context of a linear quadratic model predictive controller, cf. [GH10]. A system's dynamic is described by ordinary differential equations, where the right hand side can be influenced via an L^∞ control function. The task of the controller is to regulate the system, i.e. to either track a reference trajectory or to efficiently reach an equilibrium state.

Figure 7.1 (cf. [GH10, Figure 1]) depicts the concept of a predictive controller. The system's dynamic is described by an ordinary differential equation. Given a measured initial state, the behavior of the system is predicted over some time horizon (t_{calc} in the figure). The objective function for this prediction is then optimized with respect to the control function. The dashed line in the figure shows the predicted state trajectory of the system for the optimal control. Then the calculated control is applied to the system for some time (t_{appl}), leading to the solid line as the real system's behavior under this control. The time horizon for the application of the control does not coincide with the calculation time horizon. Afterwards, the state of the system is measured and used as the initial state, and the algorithm starts again. The prediction can be quite accurate if it is possible to get a good measurement of the initial state.

For this concept of controller design, it is crucial that the optimal control is calculated quickly, as the optimization process cannot start until a measurement of the initial state for the problem has been made. This leads to a time gap between the calculation and the

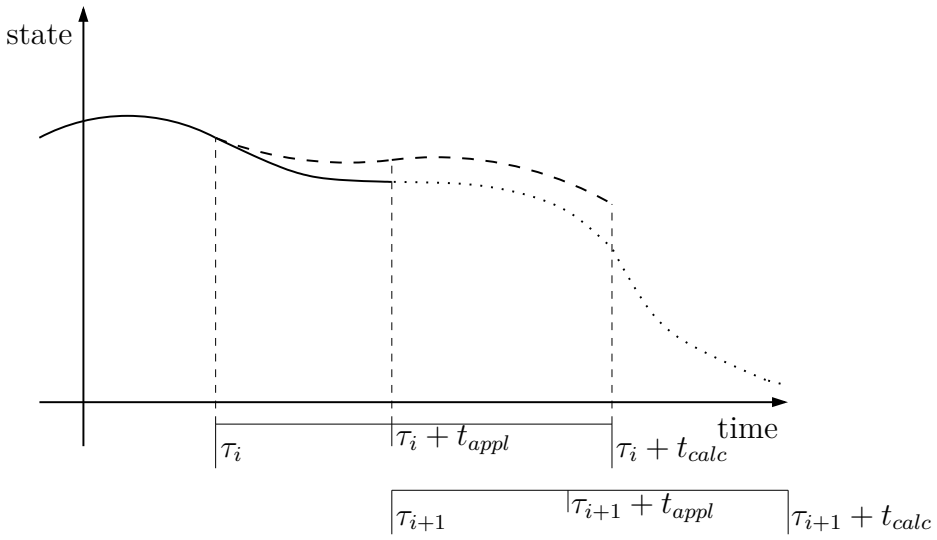


Figure 7.1.: Concept of control

actual application of the control. In this time gap, the system may already divert from its prediction.

Two models for controller design are commonly used in this context:

Model Predictive Control In this approach, the underlying dynamic is left unchanged, while the objective function is often modelled as a weighted L^2-norm. Control and control-state constraints can be taken into account when determining the optimal control. The calculation needed for determining the optimal control may be quite expensive, so that embedded systems may struggle to solve the problems in due time.

LQ Control In order to reduce the calculation time, one can also look at a simplified dynamic. While the objective function is a weighted L^2-norm, the system's dynamic is linearized along the reference trajectory or in the equilibrium state, depending on the present problem. One advantage of this method is that its solution can be calculated offline. The optimal control can then be determined by multiplying the state deviation from the reference trajectory by a time-dependent matrix. When tracking an equilibrium state, this matrix is even independent of time, so that the online calculation effort is reduced to a simple matrix multiplication. For this approach to be applicable, the state deviation from the original trajectory must not be too large. Otherwise, the linearization may become inaccurate.

The controller design proposed in this work is a mixture between the above approaches. The system's dynamic is linearized as in the LQ control, but linear state and control-state constraints are taken into account. In contrast to the LQ control case, the optimal control problem cannot be solved analytically any more, but has to be solved numerically.

Problem 7.1 (Original Problem)
Let the physical system be described by the ordinary differential equation

$$\dot{x}(t) = f(t, x(t), u(t)) \quad t \in [t_0, t_f],$$

with $x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$ *and* $u \in L^\infty([t_0, t_f], \mathbb{R}^{n_u})$*, together with a given initial condition*

$$x(t_0) = x_0,$$

for some $x_0 \in \mathbb{R}^{n_x}$*. Let state and control-state constraints be modelled as*

$$\begin{aligned} c(t, x(t), u(t)) &\leq 0 \quad \textit{a.e. in } [t_0, t_f], \\ s(t, x(t)) &\leq 0 \quad \textit{for } t \in [t_0, t_f]. \end{aligned}$$

The task is to track a given reference trajectory (x_{ref}, u_{ref})*,* $x_{ref} \in W^{1,\infty}([t_0, t_f], \mathbb{R}^{n_x})$*,* $u_{ref} \in L^\infty([t_0, t_f], R^{n_u})$*.*

Unfortunately, the controller algorithm cannot guarantee that the state constraints will be statisfied. One reason is that the state cannot be exactly determined due to measurement errors. The second reason is that a linearized version of the dynamics is being used. Finally, the calculated control is applied over a given time interval, during which the state constraint may be violated. It may happen that a feasible solution to the above problem does not even exist, because the constraints are already violated in the initial state.

In this case, it is desirable to get a solution that reduces the violation and also gives a good objective function value. A regularization that renders the problem feasible is therefore needed in this case. Hence, the virtual control technique as presented in chapter 4.2 is applied to Problem 7.2 with a positive regularization parameter α.

In all examples, the parameter functions were again set to

$$\kappa(\alpha) := 0, \qquad \phi(\alpha) := 1, \qquad \gamma(\alpha) := \alpha.$$

The Linear Quadratic Regulation Problem for the initial state ξ_i on the interval $[\tau_i, \tau_f]$ (with $\tau_f := \min\{\tau_i + t_{calc}, t_f\}$) reads:

Problem 7.2 (Linear Quadratic Regulation Problem *LQRP*)
Let $\Delta x_i := \xi_i - x_{ref}(\tau_i)$ *denote the deviation of the measured state from the reference state at time* τ_i*. Find a control correction* $\Delta\hat{u} \in L^\infty([\tau_i, \tau_f], \mathbb{R}^{n_u})$*, a virtual control* $\hat{w}_\alpha \in L^\infty([\tau_i, \tau_f], \mathbb{R}^{n_s})$ *and a state correction* $\Delta\hat{x} \in W^{1,\infty}([\tau_i, \tau_f], \mathbb{R}^{n_x})$*, that minimize*

$$\begin{aligned} J(\Delta x, \Delta u) := &\frac{1}{2}\Delta x(\tau_f)^\top Q_f \Delta x(\tau_f) \\ &+ \frac{1}{2}\int_{\tau_i}^{\tau_f} (\Delta x(t)^\top, \Delta u(t)^\top) \begin{pmatrix} Q(t) & R(t) \\ R(t)^\top & S(t) \end{pmatrix} (\Delta x(t), \Delta u(t)) dt \\ &+ \frac{1}{2}\int_{\tau_i}^{\tau_f} \|w_\alpha(t)\|_2^2 dt \end{aligned}$$

under the constraints

$$\begin{aligned} \Delta\dot{x}(t) &= A(t)\Delta x(t) + B(t)\Delta u(t), \\ \Delta x(\tau_i) &= \Delta x_i, \end{aligned}$$

$$C(t)\Delta x(t) - w_\alpha(t) \leq d(t),$$
$$G(t)\Delta x(t) + H(t)\Delta u(t) \leq l(t),$$

where A, B, C, d, G, H and l are defined as

$$A(t) := f'_x(t, x_{ref}(t), u_{ref}(t)), \qquad B(t) := f'_u(t, x_{ref}(t), u_{ref}(t)),$$
$$C(t) := s'_x(t, x_{ref}(t)), \qquad d(t) := -s(t, x_{ref}(t)),$$
$$G(t) := c'_x(t, x_{ref}(t), u_{ref}(t)), \qquad H(t) := c'_u(t, x_{ref}(t), u_{ref}(t)),$$
$$l(t) := -c(t, x_{ref}(t), u_{ref}(t)).$$

Then the proposed controller works as follows:

Algorithm 7.3 (Linear Quadratic Controller for Constrained Problems)

1. *Let $i := 0$, $\tau_0 := t_0$, and let ξ_0 be the measured system state in τ_0.*
2. *Calculate Δx, Δu as an approximate solution to the Linear Quadratic Regulation Problem 7.2 in $[\tau_i, \tau_f]$, with $\tau_f := \min\{\tau_i + t_{calc}, t_f\}$.*
3. *Apply the control $u_{ref} + \Delta u$ in the time interval $[\tau_i, \tau_i + t_{apply})$. Measure the final state $\xi_{i+1} := x(\tau_i + t_{apply})$.*
4. *Let $\tau_{i+1} := \tau_i + t_{apply}$, $i := i + 1$, and go to step 2.*

This controller algorithm will be referred to as LQC controller in the remainder in contrast to the LQ algorithm (or "LQR" for "Linear Quadratic Regulator") that does not take the constraints into account.

In step 2, a numerical solution to the Regulation Problem 7.2 is needed. We apply a function space method presented in chapter 5 and discuss the effects of this approach in the following examples.

For the simulation of the following examples, the discretization method described in chapter 6 is used on a time grid with the same equidistant step size as for the calculation. In the first examples, where control constraints in the form of box constraints $u_{i\min} \leq u_i \leq u_{i\max}$ are present, the controls u_i are cut off to fit in the given box, independent from the control method in use. The reason for this is that this sort of constraints usually reflects physical constraints of the controls (e.g., the maximum velocity/acceleration of a motor).

It should be mentioned that the control scheme of the proposed algorithm is applied to the traditional LQ controller as well. That means that the LQ controller is not applied as a feedback control (where the behavior of the system would have been predicted for an infinite time horizon or the remaining time horizon in the case of a given reference trajectory) but as a closed loop control, where again the optimal control is calculated over a bigger horizon, and the control is then applied over a smaller interval. As the main calculations for the traditional LQ controller can be done before the simulation starts, the actual on line calculation time for this approach negligible.

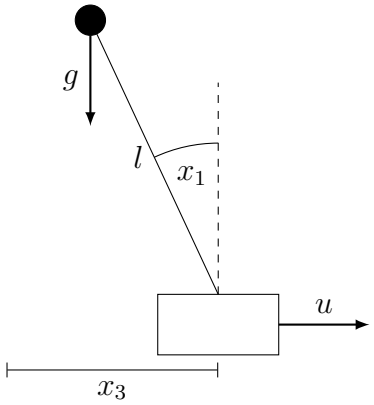


Figure 7.2.: Sketch: The Inverse Pendulum

7.2. Examples: LQC Controller With Control Constraints

First, the controller is tested for a simple equilibrium example. At the same time, this example is used to showcase the simulation technique used for all numerical controller examples. The effect of the virtual control in comparison to the LQ approach as well as the effect of the regularization of the Fischer-Burmeister function when solving the Regulation Problem is observed.

In the subsequent example, a reference trajectory will be tracked.

7.2.1. Inverse Pendulum With Control Constraints

This is a (2D-) model of an inverse pendulum, mounted on a wagon (cf. Figure 7.2). The weight at the top can be balanced by accelerating the wagon. In this model, we assume that the weight of the wagon is is much higher than the weight of the pendulum. This leads to the simplifying assumption that the pendulum has no influence on the wagon:

Problem 7.4 (Original Inverse Pendulum)

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -kx_2 + g\sin x_1 + u\cos x_1 \\ \dot{x}_3 &= x_4 \\ \dot{x}_4 &= u \\ & \quad a.e.\ in\ [0, t_f], \end{aligned}$$

$$x_1(0) = x_{10},\ x_2(0) = x_{20},\ x_3(0) = x_{30},\ x_4(0) = x_{40}$$

In this model, the states x_1 and x_2 do not depend on the position x_3 or the velocity x_4 of the wagon. The task is to keep the pendulum in the upper position $x_1 = x_2 = 0$ for as long as possible, despite of possible disturbances.

The Linear Quadratic Regulation Problem for a given initial state x_0 reads:

Problem 7.5 (Inverse Pendulum: *LQRP*)

$$\begin{aligned} \text{min!} \quad F(\Delta x, \Delta u) &:= \Delta x_1(\tau_f)^2 + \Delta x_2(\tau_f)^2 \\ &\quad + \frac{1}{2}\int_{\tau_i}^{\tau_f} \frac{1}{10}\Delta u(t)^2 + \Delta x_1(t)^2 + \Delta x_2(t)^2 dt \end{aligned} \tag{7.6}$$

$$\begin{aligned} \textit{with respect to} \quad & \Delta x \in W^{1,\infty}([\tau_i, \tau_f], \mathbb{R}^4) \\ \textit{and} \quad & \Delta u \in L^{\infty}([\tau_i, t_f], \mathbb{R}^1), \end{aligned}$$

s.t.

$$\Delta x(0) = \Delta x_0$$

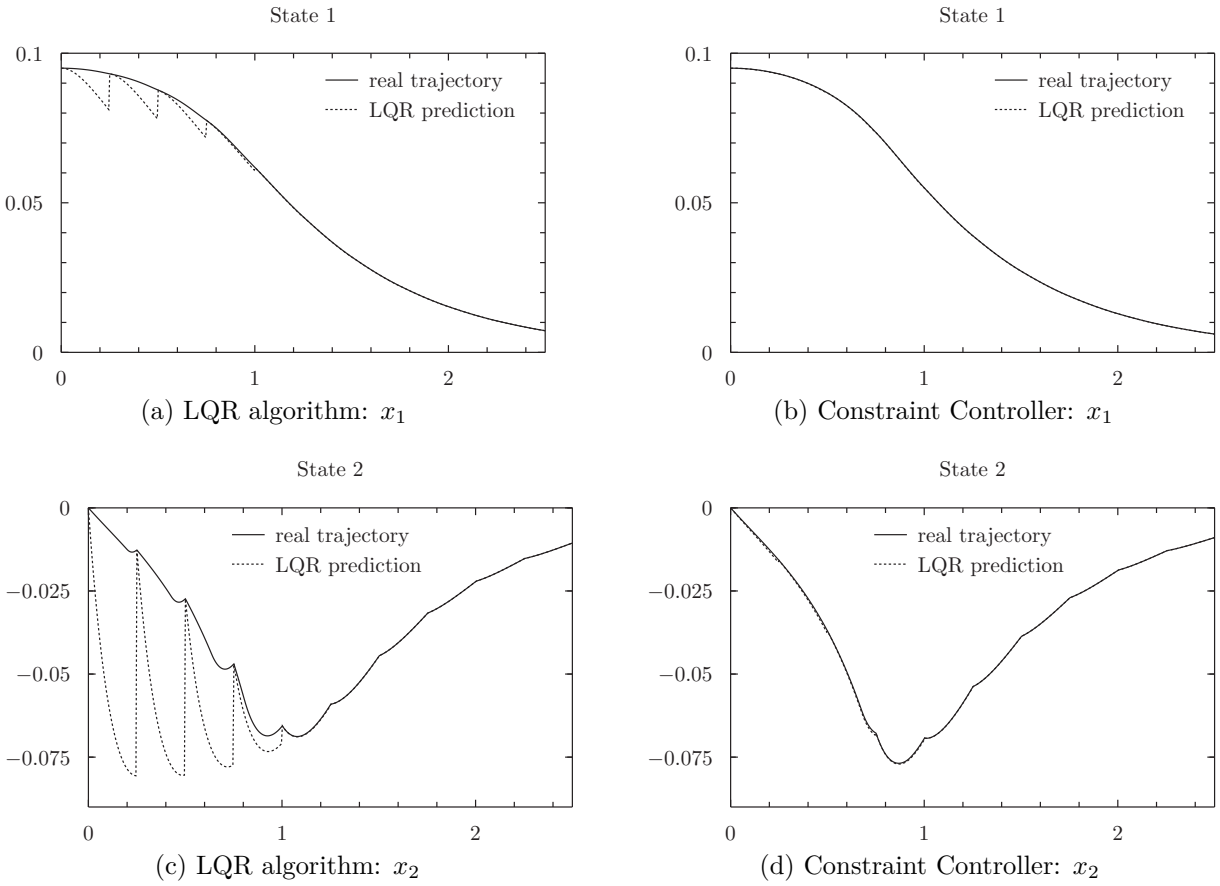
$$\begin{aligned} \Delta\dot{x}_1 &= \Delta x_2 \\ \Delta\dot{x}_2 &= g\Delta x_1 - k\Delta x_2 + u \\ \Delta\dot{x}_3 &= \Delta x_4 \\ \Delta\dot{x}_4 &= \Delta u \\ & \quad \textit{a.e. in } [\tau_i, \tau_f], \end{aligned}$$

$$-1 \le \Delta u \le 1 \qquad \textit{a.e. in } [\tau_i, \tau_f].$$

For both implementations, we used the parameters $k = 1$, $g = 9.81$. For the controller, we chose $t_{calc} = 0.75$, $t_{apply} = 0.25$ over a total time of 2.5, starting at $x_0 = (0.095, 0, 0, 0)^\top$. The constraint $-1 \le u(t) \le 1$ was enforced. Each LQRP was solved using the discretization method from chapter 6 with 150 time steps. The simulations plotted in Figures 7.3 and 7.4 have been calculated with $\beta = 10^{-5}$. As the system proved to be quite stable with respect to the calculation errors, the tolerance was set to $\varepsilon = 10^{-3}$. The overall calculation time for the LQC simulation was 1.134 seconds, while the maximum time needed for one time step was 0.287 seconds[1].

The most important aspect that can be expected to be advantageous about the LQC algorithm is that the future behavior of the system can be predicted more precisely near boundaries, since the predictions of the traditional algorithm lead to optimal controls that are not necessarily feasible and have to be altered before they are applied to the system. This becomes apparent in Figure 7.3: The trajectories generated by the classical

[1] All calculation times in this chapter were measured on a Dell XPS laptop running at 2.00 Ghz. These values includes the time needed for memory allocation and finding the norm of the direction, which in real life applications can be simplified. It should also be noted that the calculations were made using Scilab instead of a lower level language.



(a) LQR algorithm: x_1

(b) Constraint Controller: x_1

(c) LQR algorithm: x_2

(d) Constraint Controller: x_2

Figure 7.3.: Inverse Pendulum: Prediction and Trajectories for the unconstrained and the constrained controller

LQ controller differ significantly from the mathematical prediction (shown in 7.3a and 7.3c). The predictions of the proposed algorithm are significantly more precise (shown in 7.3b and 7.3d). This leads to a lower functional value for the controller.

Figure 7.3c together with Figure 7.4 reveal the reason for the difference between the two approaches: The optimal control in the unconstrained LQ algorithm is infeasible, so that it was projected into the feasible box before the simulation. This leads to unsmooth "hooks"[2] in the plot whenever the calculation is restarted. The control plots in Figure 7.4 show that at the end of the first $[t_i, t_i + t_{appl}]$ intervals, the control already starts getting smaller in the LQ calculation, as the predicted trajectory has already decreased significantly.

Effects of the Regularization on the Controller

Finally, the question remains how the solution depends on the regularization parameter β. It turns out that for sufficiently small β, the solutions seem to converge. If on the other hand β is too large, then the LQC controller does not yield any advantage over the LQ approach at all. In fact, the trajectories for $\beta = 10^{-2}$ are even less advantageous as the

[2]This effect will be explained in more detail at the end of the example on page 112.

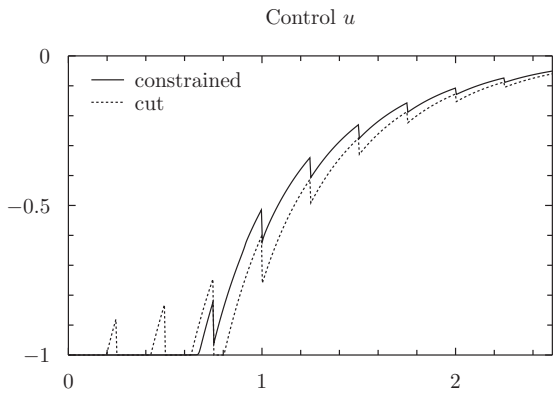


Figure 7.4.: Inverse Pendulum: Controls calculated by the unconstrained and the constrained controller

LQ control. For the start value $x_0 = (0.1, 0, 0, 0)^\top$, the controller with $\beta = 10^{-2}$ even fails to track the equilibrium at all; the applied control was not sufficient to keep the pendulum upright.

The explanation for this effect lies in the nature of the example. The inverse pendulum tends to "tip over" and move towards the stable equilibrium $(-\pi, 0, x_3, x_4)^\top$. If the applied control is not powerful enough, then the algorithm fails. At this point, the constraints can render the problem absurd when the control is constrained in a way that makes it impossible to reach the upper equilibrium. In this regard, the property mentioned in Remark 5.15 has a negative aspect in this scenario: The generated control is strictly feasible, in other words, the regularized solutions stay away from the boundaries of the set of feasible controls. This leads to the calculated control being weaker than the control given by the LQ algorithm, so that the set of start values that still lead to tracking trajectories decreases. For the given start value and $\beta = 10^{-2}$, the effect is that the control is too weak to regulate the system more efficiently than the LQ controller does.

Other parameters of the Controller Algorithm

Finally, it remains to point out how the general parameters of the controller effect the behavior of the simulated system. In particular, the claim that the "hooks" (cf. footnote 2 on page 111) arise from the difference between calculated and applied control needs support, as even the control calculated by the proposed algorithm shows points of discontinuity (cf. Figure 7.4 and the last plot in Figure 7.5).

Evidence for this claim is given in Figure 7.6. These plots have been made using a longer prediction horizon, $t_{calc} = 1.5$ as opposed to $t_{calc} = 0.75$ in the previous examples. It can clearly be seen that the points of discontinuity in the control of the LQ algorithm remain in a neighborhood of the constraints. This supports the claim that these points arise from the projection into the box of feasible constraints. All other "hooks" seem to vanish, in the LQ algorithm as well as in the proposed LQC controller. The reason for this is that the bounded horizon leads to the algorithm calculating a control that is only optimal inside the bounded time interval $[\tau_i, \tau_i + t_{calc}]$. After the application of the control, the

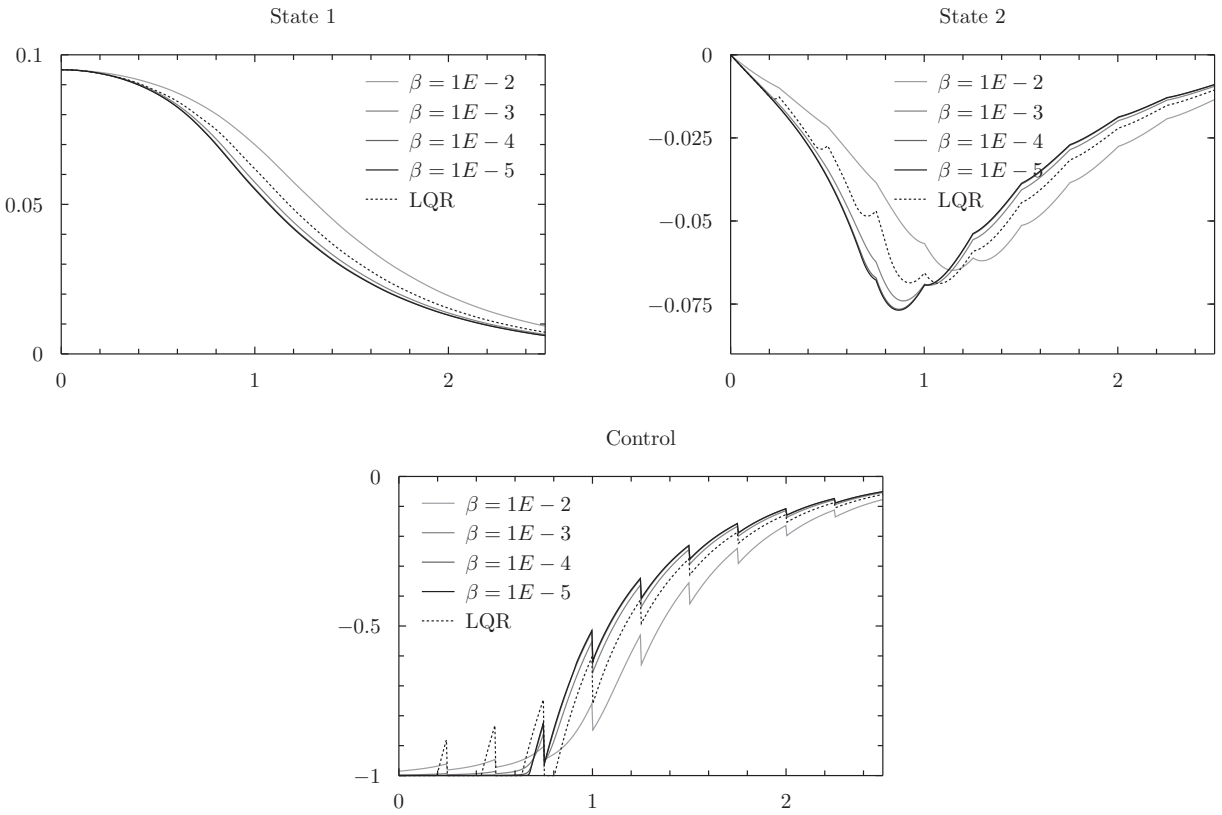


Figure 7.5.: Inverse Pendulum: Regularized calculations with $\beta = 10^{-2}, \ldots, 10^{-5}$

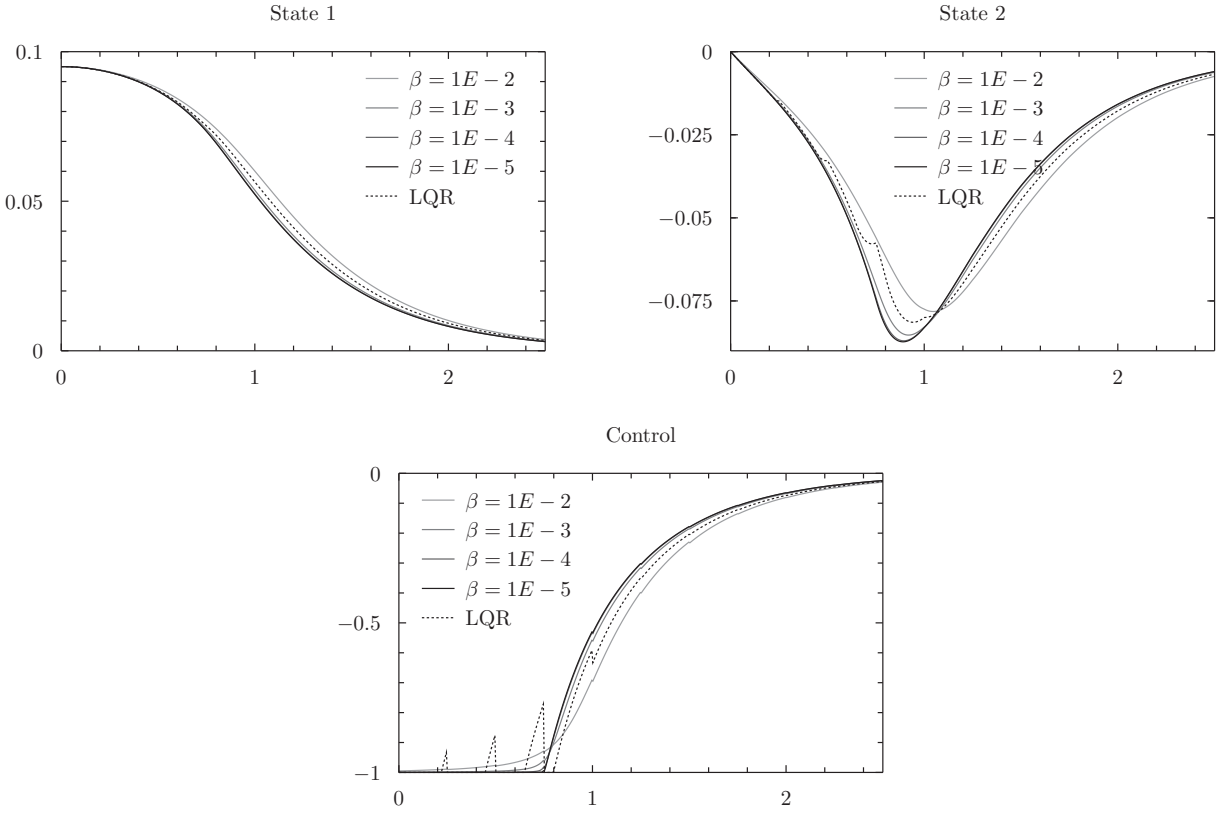


Figure 7.6.: Inverse Pendulum: Simulations for a longer time horizon t_{calc}

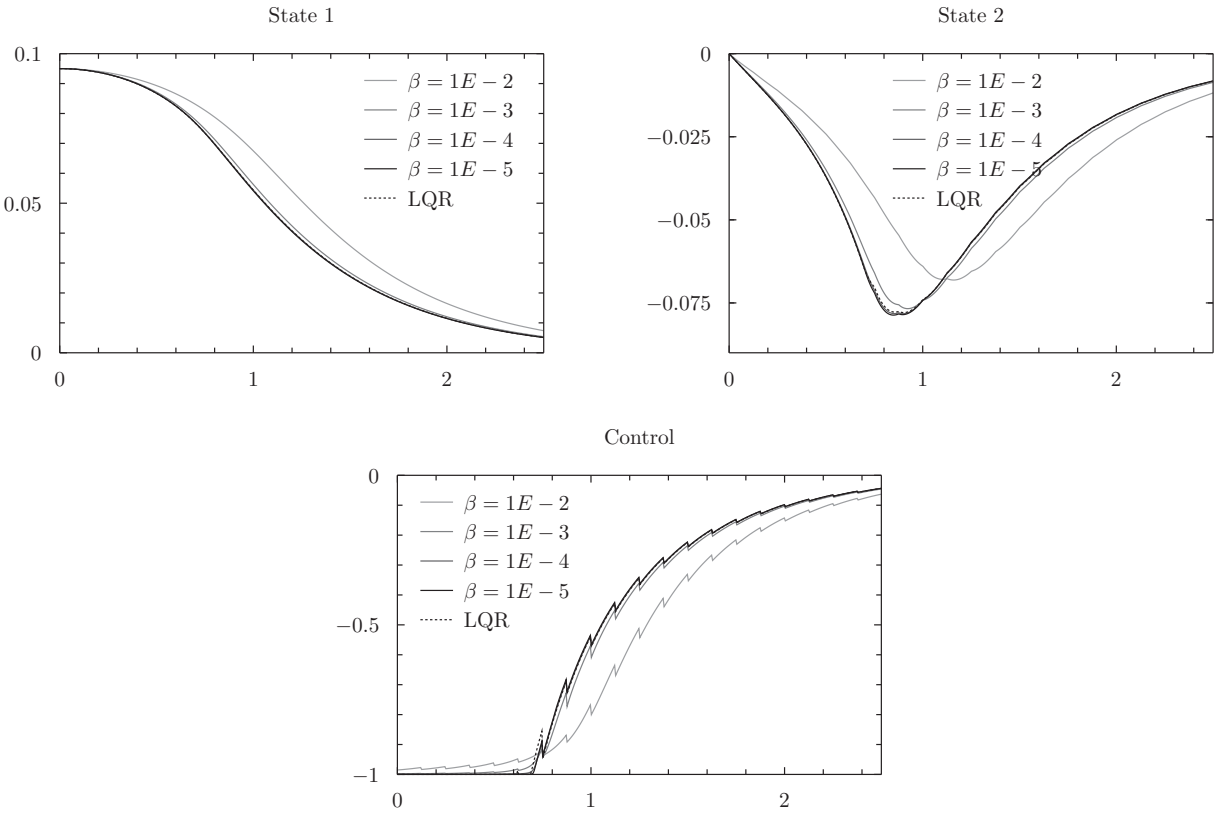


Figure 7.7.: Inverse Pendulum: Simulations for a smaller application time t_{appl}

time interval is shifted, which leads to a different optimal control problem. For larger time horizons, the difference between both problems vanishes up to the difference that arises from the fact that the problems are calculated using a linearization of the system's dynamic.

The price that needs to be paid for the higher level of smoothness are longer calculation times. As the time interval of the optimal control problem becomes twice as long, so do the calculation times in the best case.

Since the argument of the control problems becoming more similar holds analogously when the application time t_{apply} decreases, it can be expected that the "hooks" also decrease for smaller values of t_{apply}. Figure 7.7 shows that in fact, the simulations behave even better than expected: The aforementioned discontinuities in general decrease, and even the "hooks" near active constraints in the control of the LQ simulation vanish. The reason for this is simple: In this example, the projection of the calculated control into the feasible set yields the optimal control for the shown time intervals $[\tau_i, \tau_i + t_{calc}]$.

Several problems may render a reduction of t_{apply} inapplicable in real life applications:

- The overall computation times increase. While in each step the optimal control problem remains unchanged in size, the number of optimal control problems doubles when t_{apply} is halved.
- As t_{apply} is reduced, the importance of the calculation time increases. The calculated

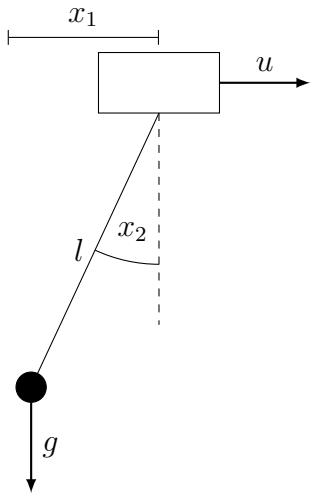


Figure 7.8.: Sketch: The Trolley Model

control can only be invoked after it has been calculated. This leads to an offset between the measurement of the initial state of the Linear Quadratic Regulation Problem and the application of the control. Sensibly, t_{apply} should stay significantly larger than the time needed for solving the optimal control problems.

- A reduction of t_{apply} also means that the control algorithm has to be called more often. As the matrices that store the linearization of the system's dynamic may have to be updated (this is the case when either the dynamic or the reference trajectory is dependent on time), an overhead occurs each time the calculation routine is called.

7.2.2. Trolley Problem With Control Constraints

This example is a linear quadratic controller for a trolley. A weight m_2 is attached to a rope of length l hanging from the trolley. The trolley itself has mass m_1. The task is to move the weight quickly over a distance of 1, without causing too much oscillation of the weight (cf. Problem 7.6). The acceleration of the trolley can be controlled. Figure 7.8 shows a sketch of the trolley.

Problem 7.6 (Original Trolley problem)

$$\text{min!} \quad J(x,u) := \frac{1}{2}\int_0^{t_f} (u(t)^2 + 1000x_4(t)^2)dt + t_f$$

$$\begin{aligned} \textit{with respect to} \quad & x \in W^{1,\infty}([0,t_f],\mathbb{R}^4) \\ \textit{and} \quad & u \in L^{\infty}([0,t_f],\mathbb{R}^1), \end{aligned}$$

s.t.

$$\begin{aligned} \dot{x}_1 &= x_3 \\ \dot{x}_2 &= x_4 \end{aligned}$$

$$
\begin{aligned}
\dot{x}_3 &= \frac{(m_2^2 l^3 x_4^2 + m_2 I_{y_2} l x_4^2 + m_2^2 l^2 g\cos(x_2))\sin(x_2) - (m_2 l^2 + I_{y_2})u}{-m_1 m_2 l^2 - m_1 I_{y_2} - m_2^2 l^2 - m_2 I_{y_2} + m_2^2 l^2 \cos(x_2)^2} \\
\dot{x}_4 &= \frac{m_2 l(m_2 l\cos(x_2)^2 x_4^2 \sin(x_2) + g\sin(x_2)(m_1+m_2) - \cos(x_2)u)}{-m_1 m_2 l^2 - m_1 I_{y_2} - m_2^2 l^2 - m_2 I_{y_2} + m_2^2 l^2 \cos(x_2)^2} \\
& \textit{a.e. in } [0, t_f],
\end{aligned}
$$

$$
\begin{aligned}
x_1(0) = x_2(0) = x_3(0) = x_4(0) = 0 \\
x_1(t_f) = 1, x_2(t_f) = x_3(t_f) = x_4(t_f) = 0
\end{aligned}
$$

and

$$
-0.2 \leq u \leq 0.2
$$

Values for and the meaning of the different parameters of the problem are shown in table 7.1.

Parameter	Meaning	Value
g	gravitational acceleration	9.81
m_1	mass of the trolley	0.6
m_2	mass of the weight	0.62
l	length of the rope	0.73
r	radius of the (spherical) weight	0.1
I_{y_2}	moment of inertia of a spherical weight	$0.4 * m_2 * r^2$

Table 7.1.: Parameters of the Trolley Problem

The reference trajectory (x_{ref}, u_{ref}) was calculated as the optimal solution for this problem using the OC-ODE library written by Matthias Gerdts. In this example, the job of the controller algorithm is to find optimal *deviations* of the controls and states for given initial *deviations* of the state.

The problem is linearized along the calculated reference trajectory. The objective for the controller is to minimize the deviation from the states and the control. After choosing weights for these objectives, the Regulation Problem reads:

Problem 7.7 (Trolley: *LQRP*)

$$
\text{min!} \qquad J(\Delta x, \Delta u) := \frac{1}{2}\|\Delta x(t_f)\|_2^2 + \frac{1}{2}\int_{t_0}^{t_f} \|\Delta x(t)\|_2^2 + \frac{1}{10}\|\Delta u(t)\|_2^2 dt
$$

$$
\begin{aligned}
\textit{with respect to} \quad & \Delta x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^4) \\
\textit{and} \quad & \Delta u \in L^{\infty}([t_0, t_f], \mathbb{R}^1),
\end{aligned}
$$

s.t.

$$\begin{aligned} \Delta\dot{x} &= f'_x(x_{ref}, u_{ref})\Delta x + f'_u(x_{ref}, u_{ref})\Delta u \\ &\quad a.e.\ in\ [t_0, t_f], \end{aligned}$$

with

$$-0.2 \leq u_{ref} + \Delta u \leq 0.2$$

The original trajectory was calculated on 500 time steps on the time interval $[0, 5.680]$. For the LQ algorithm, the same time steps were used, and the whole interval was divided into 10 subintervals for application, while the calculation was performed for a time horizon three times the length of the application time. Hence, each LQRP interval starts at some time t_{n_0}, then the optimal control over the next $3 \cdot 50 = 150$ time steps is calculated and the control is applied for 50 steps, before the calculation starts again at time t_{n_0+50}. Again, the constrained LQC controller was calculated with a regularization parameter $\beta = 10^{-5}$. With a tolerance of $\varepsilon = 10^{-5}$, the calculations took 1.647 seconds in total. The maximal time used for one prediction step was 0.185 seconds.

Figure 7.9 shows the trajectories and the control for the LQ algorithm as well as the proposed LQC controller. The latter controller performs better in the sense that the reference trajectory gets tracked faster (this effect is best visible in states x_2 and x_4, cf. Fig. 7.9b and 7.9d). Also, the maximum deviation in state 4 remains significantly smaller, cf. Fig. 7.9d.

Figure 7.10 gives the reason for the difference. As in the previous example, the traditional LQ algorithm cannot predict the system's behavior where the constraints become active. In such cases, the prediction and the real trajectory (i.e. the trajectory of the simulated system) differ.

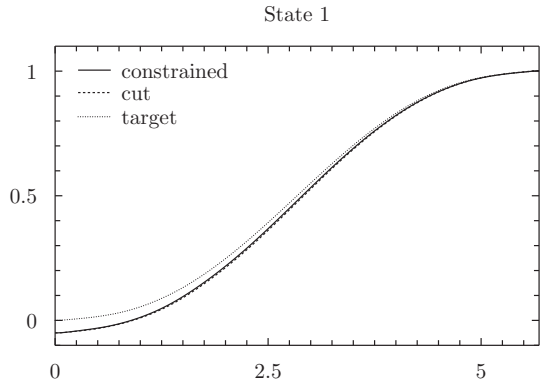
7.3. Examples: LQ Control With State Constraints

In this section, the examples from section 7.2 are revised with different constraints. In the previous section, the new LQC controller performed slightly better than the traditional unconstrained LQ controller, but it remained questionable if the better performance would compensate the higher computational effort.

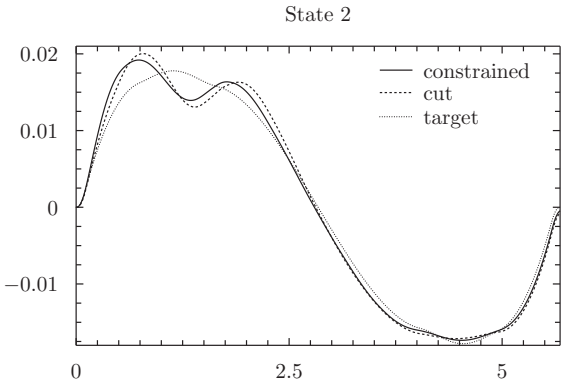
In the first example, the inverse pendulum is simulated in a constrained space. Still, the task is to track the equilibrium $(x_1, x_2) = (0, 0)$, while the constrained variable is x_2, the angular velocity[3]. The drawback of the traditional LQ approach is that these state constraints cannot be explicitly taken into account when calculating a controller.

A natural way of handling this situation is to add a penalty term for the constrained variable in the LQRP weighting function (or in this case increase the existing term), so that the LQRP objective function reads

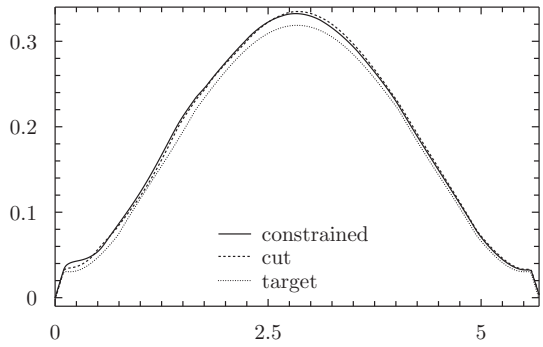
[3] At first glance, it appears more natural to restrict the velocity or the available space for the wagon. On second thought, it becomes clear that for such a task, other controller design approaches are more suitable. The problem is discussed in adequate depth in Appendix B.



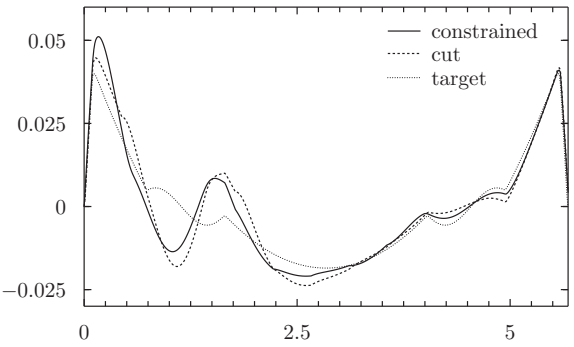
(a) LQ, LQC, target: x_1



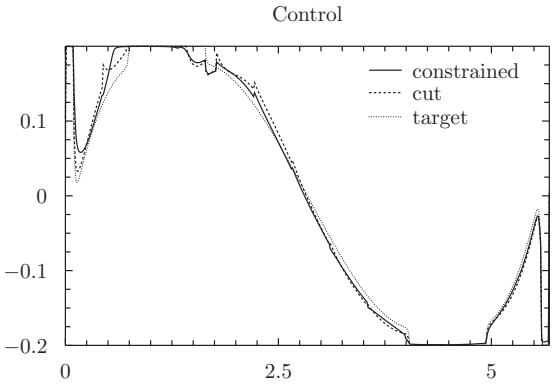
(b) LQ, LQC, target: x_2



(c) LQ, LQC, target: x_3

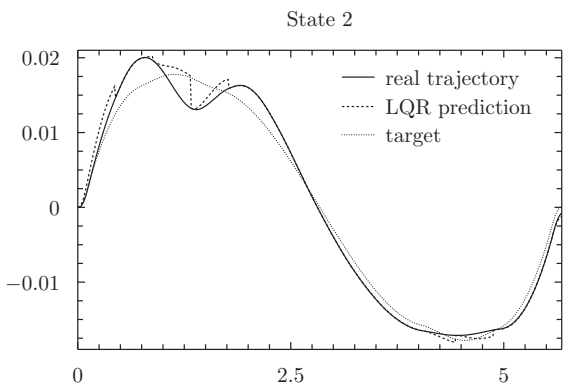


(d) LQ, LQC, target: x_4

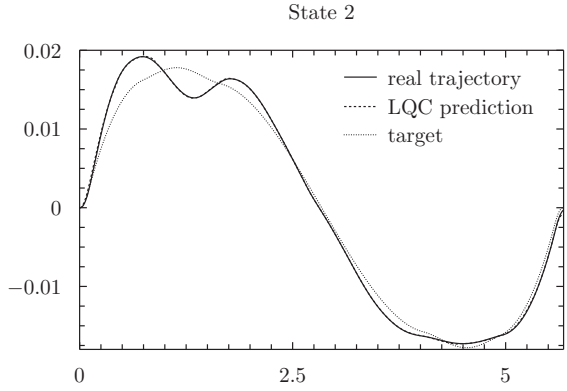


(e) LQ, LQC, target: u

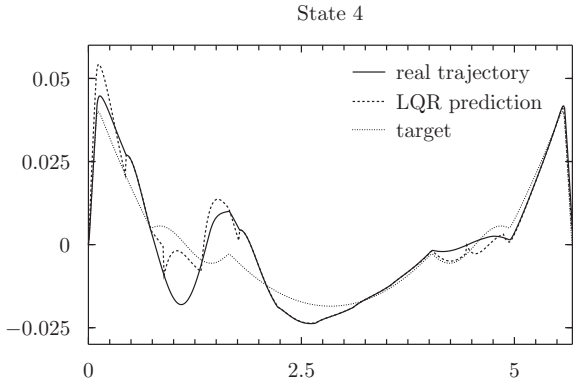
Figure 7.9.: Trolley Problem: Trajectories for the LQ and the LQC controller



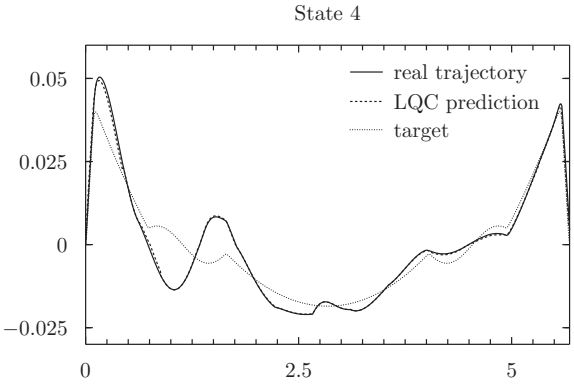
(a) LQ controller: x_2



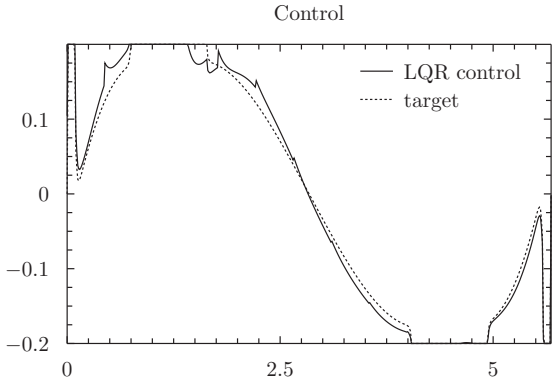
(b) LQC controller: x_2



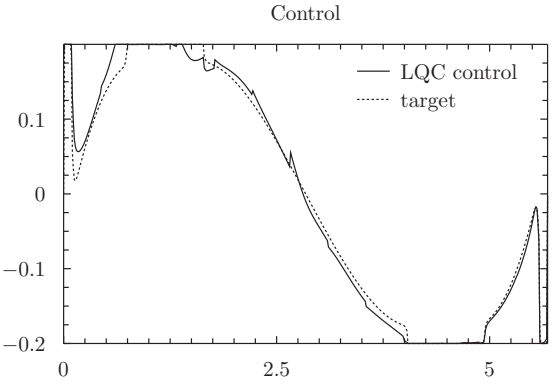
(c) LQ controller: x_4



(d) LQC controller: x_4



(e) LQ controller: u



(f) LQC controller: u

Figure 7.10.: Trolley Problem: Prediction and trajectories for the LQ and the LQC controller

$$\begin{aligned}\text{min!} \quad F(\Delta x, \Delta u) := \Delta x_1(\tau_f)^2 + \Delta x_2(\tau_f)^2 \\ + \frac{1}{2}\int_{\tau_i}^{\tau_f} \frac{1}{10}\Delta u(t)^2 + \Delta x_1(t)^2 + c \cdot \Delta x_2(t)^2 dt\end{aligned} \tag{7.19}$$

for some constant $c > 1$. The downside of this is that deviations from x_2 are penalized in general, which means that the angular speed of the pendulum tends to 0 more than it would be optimal for the original task. The negative aspect of this model becomes apparent when the constraints are not symmetric, or when a variable is constrained that does not need to be minimized.

The LQC controller admits a more natural approach. The problem is augmented by a virtual control variable that models an L^2-penalty term, thus adding a cost when the constraint is violated, so that the virtual control does allow violation of the constraints. The form of the cost for the violation is the same as the form of the cost for a deviation from the reference trajectory. At the same time, the fact that the regularized Fischer-Burmeister function is being used constitutes a counterweight to the relaxation. As mentioned in Remark 5.15, the regularized complementarity conditions are equivalent to $C^i x \cdot \mu_i = \beta$, so that the calculated trajectories remain strictly feasible. In this sense, the regularized NCP-function acts as an antagonist to the virtual control relaxation.

7.3.1. Inverse Pendulum With State Constraints

Again, we start with an example in which an equilibrium should be tracked, as this type of example has the simplest structure. In real life applications, problems with state constraints will often come with a given reference trajectory.

The state constraint for this problem is $x_2(t) \geq -0.08$.

The objective function for the for the controller is analog to the objective function (7.6). The LQRP with the virtual control for the state constraint reads:

Problem 7.8 (Inverse Pendulum with State Constraints: *LQRP*)

$$\begin{aligned}\text{min!} \quad F(\Delta x, \Delta u, w_\alpha) :=& \Delta x_1(\tau_f)^2 + \Delta x_2(\tau_f)^2 + \frac{1}{2}\int_{\tau_i}^{\tau_f} \frac{1}{10}\Delta u(t)^2 + \Delta x_1(t)^2 + \Delta x_2(t)^2 dt \\ &+ \frac{1}{2}\int_{\tau_i}^{\tau_f} \|w_\alpha(t)\|_2^2 dt\end{aligned}$$

$$\begin{aligned}\textit{with respect to} \quad & \Delta x \in W^{1,\infty}([\tau_i, \tau_f], \mathbb{R}^4), \\ & \Delta u \in L^\infty([\tau_i, t_f], \mathbb{R}^1), \\ \textit{and} \quad & w_\alpha \in L^\infty([\tau_i, \tau_f], \mathbb{R}^1),\end{aligned}$$

s.t.

$$\Delta x(0) = \Delta x_0$$

$$\begin{aligned} \Delta\dot{x}_1 &= \Delta x_2 & \Delta\dot{x}_2 &= g\Delta x_1 - k\Delta x_2 + u \\ \Delta\dot{x}_3 &= \Delta x_4 & \Delta\dot{x}_4 &= \Delta u \end{aligned}$$
$$\text{a.e. in } [\tau_i, \tau_f],$$

and

$$\Delta x_2 + \alpha w_\alpha \geq -0.08 \qquad \text{in } [\tau_i, \tau_f].$$

Analog to the examples from section 4.3, we first investigate the influence of the virtual control on the behavior of the system and compare it the behavior of the system, regulated by an LQ Controller.

The simulations in Figure 7.11 were calculated using a global algorithm, so that the regularization parameter β did not have any measurable influence on the outcome. The parameter for the virtual control was set to $\alpha = 10^{-3}$, and both simulations were calculated over $t_{calc} = 1.2$ seconds, before the control was applied for $t_{appl} = 0.4$ seconds. In this case, the virtual control did not measurably weaken the constraints; the virtual control variable remained smaller than 10^{-3}. As this scenario was only interesting for illustrating the influence of α on the control when the regularization of the NCP function vanishes, the code for this example was not optimized for speed. The start iteration for each calculation was set to 0, and the globalized Newton method was used, which lead to calculation times of about 20 seconds.

For the LQ controller, the objective function was altered in order to satisfy the constraint, cf. (7.19): The constant that the Δx_2 part of the objective function was multiplied with is set to $c = 3.5$, as for this value, the constraint was nearly satisfied. Figure 7.11 shows the effect that has already been mentioned: The altered objective function leads to the effect that the original task cannot be performed as well, as part of the efford is diverted on the minimization of Δx_2. Different simulation parameters like longer calculation horizons or shorter application times smoothed out the hooks in the control of the simulation but did not have any influence on this effect. The performance on the LQ algorithm was improved, but only in the sense of its specific task, which already had to be changed from the original problem.

Figure 7.13 shows the influence of the virtual control parameter α on the system. For fixed $\beta = 10^{-3}$, the state x_2 as well as the control u seem nearly unaffected by the virtual control, which is why only the values $\alpha = 10^{-1}, 10^{-5}$ are shown. The LQ solution is plotted for comparison. The virtual control w_α itself is already small for $\alpha = 10^{-1}$, as the third plot shows. For values below 10^{-3}, the virtual control was not visible at the given scale (cf. Figure 7.12).

Finally, the influence of the parameter β is visualized in Figure 7.14. The calculations are made with fixed $\alpha = 10^{-3}$, and different values of β are compared. As in the previous examples, large values of β lead to the trajectories being similar to the LQ outcome, as

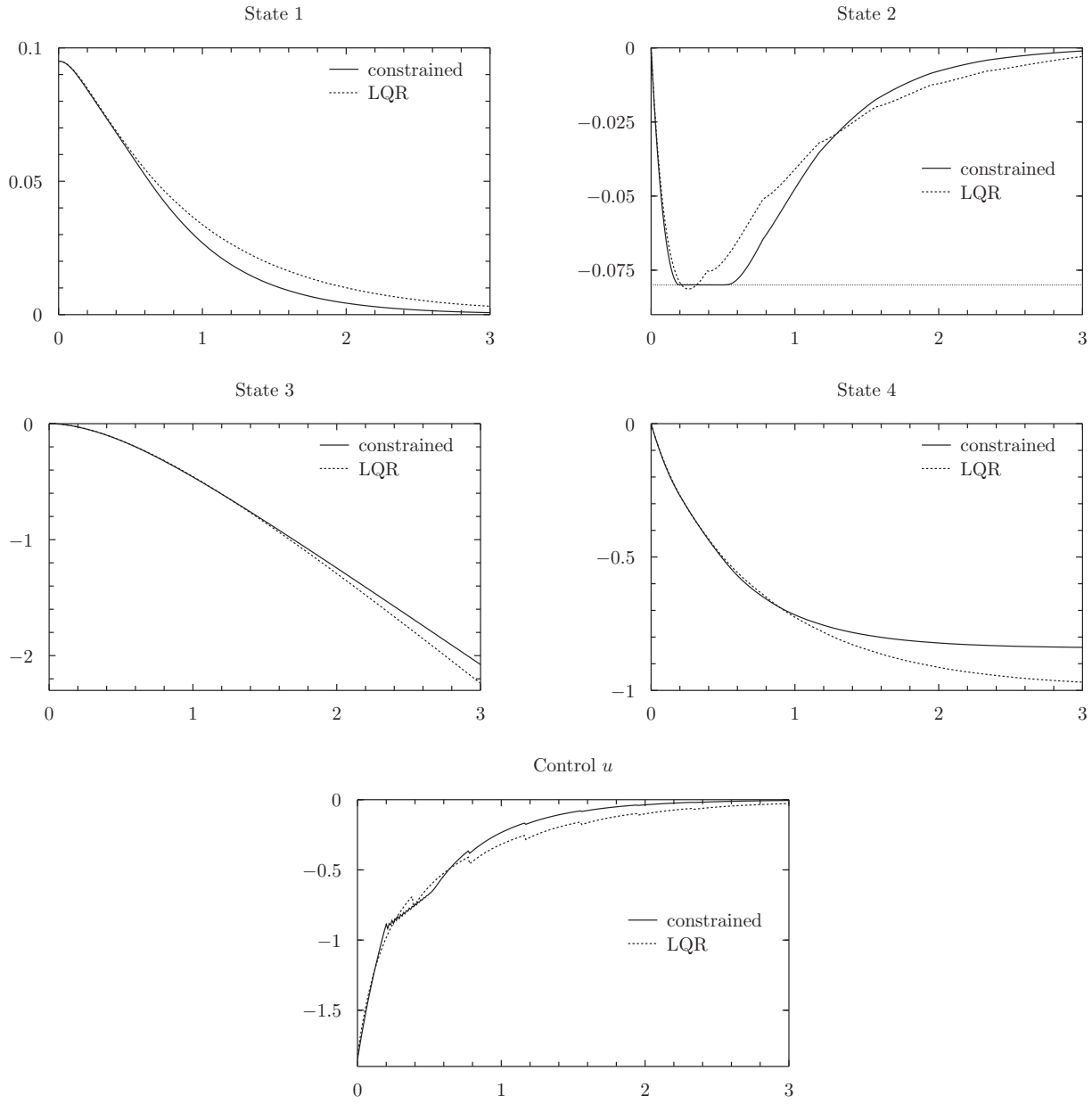


Figure 7.11.: Inverse Pendulum With State Constraints: Simulations for LQC and LQ

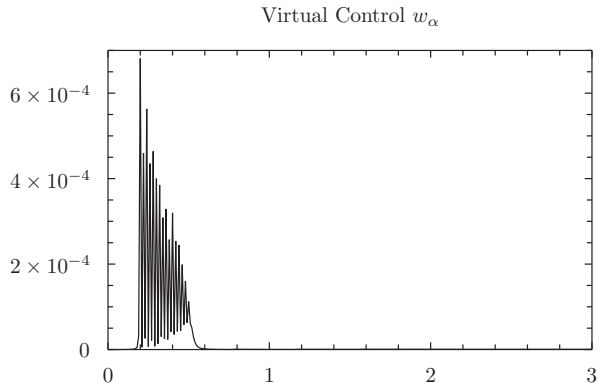


Figure 7.12.: Inverse Pendulum With State Constraints: Virtual Control

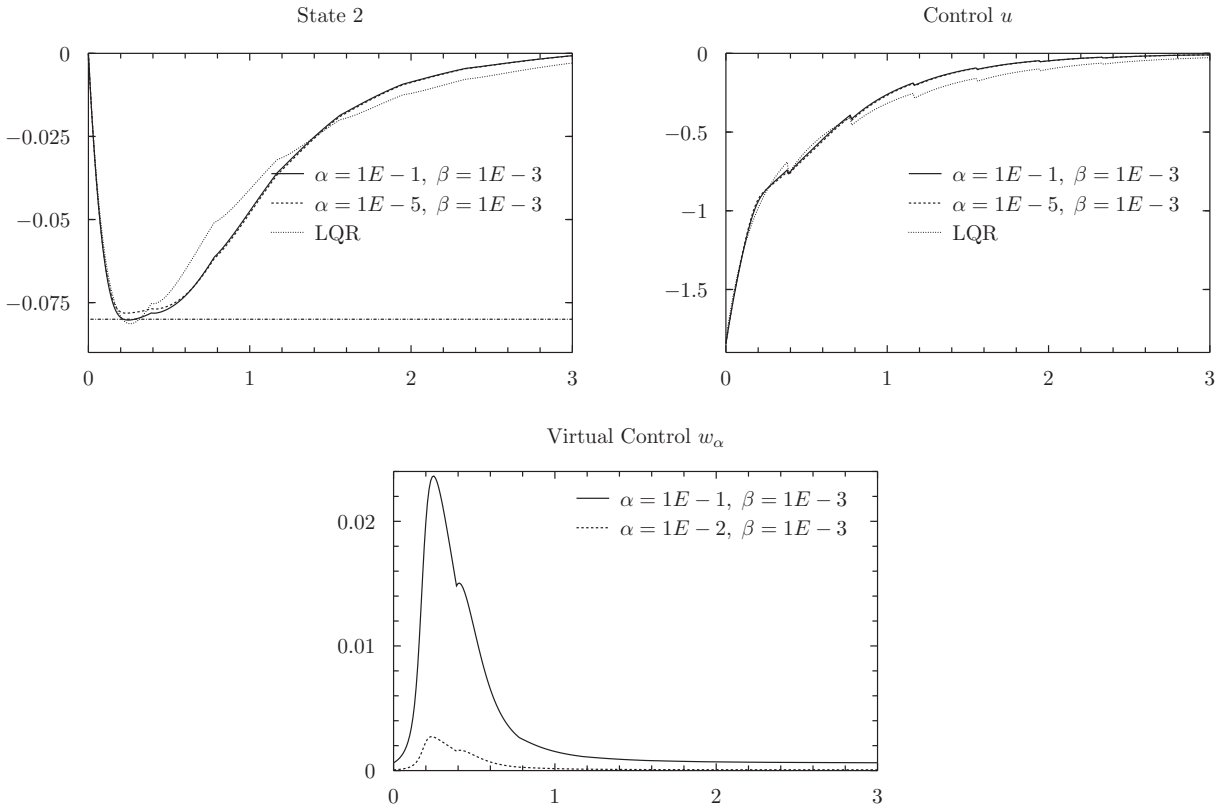


Figure 7.13.: Inverse Pendulum With State Constraints: Simulations for different values of α

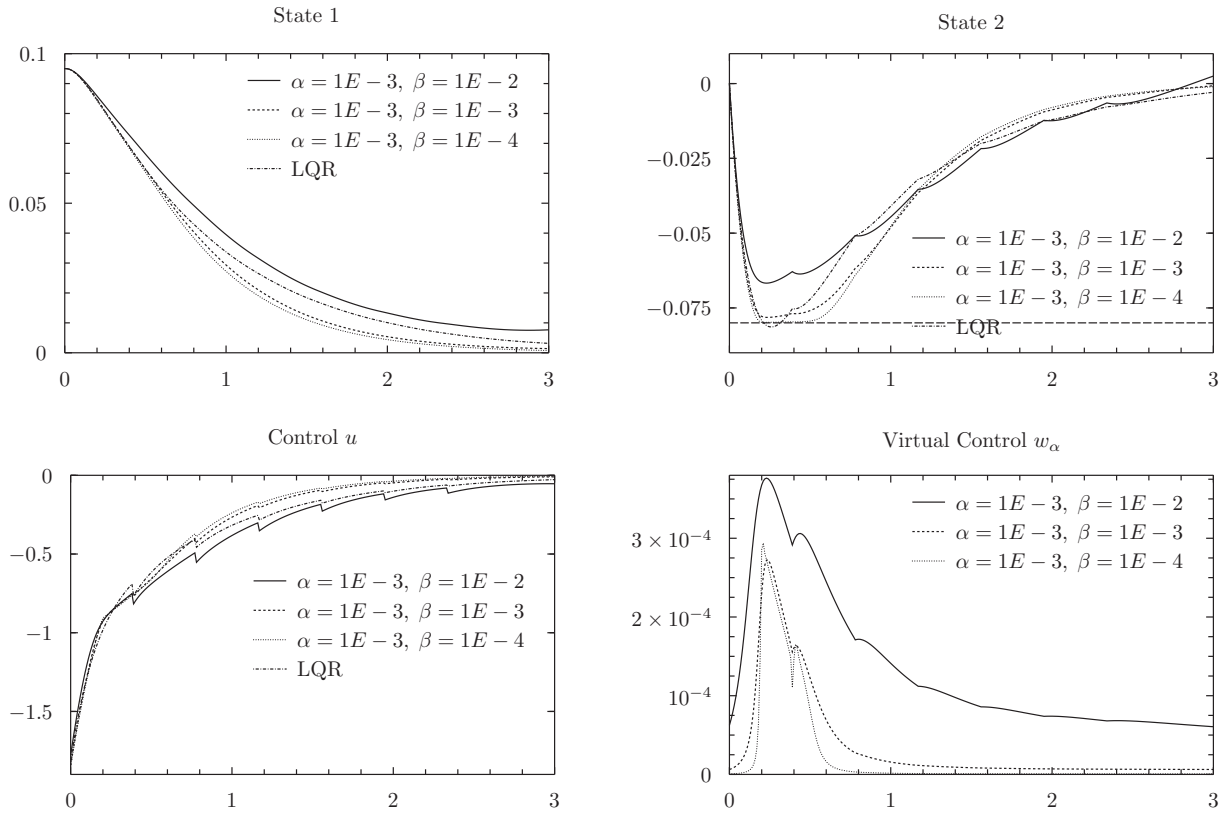


Figure 7.14.: Inverse Pendulum With State Constraints: Simulations for different values of β

active constraints in general are avoided. For $\beta \leq 10^{-3}$, the LQC algorithm performs significantly better than the traditional controller. The plot of the virtual control in the same figure suggests that for $\beta = 10^{-2}$, the virtual control variable is even needed when the trajectory is relatively far from being active. The conclusion to draw from this example is that the regularization of the Fischer-Burmeister function does lead to results that faintly reminiscent of inner point methods, as the constraints are avoided. In order to achieve better results than with the LQ approach, small values of β have to be used. This is in fact unsurprising, as the regularization parameter only appears in the square root, so that effectively the square root of the parameter is an indicator for its influence.

The question of applicability is addressed in Table 7.2. The table shows the iterations needed during the calculation process for $\alpha = 10^{-3}$ and $\beta = 10^{-3}$, as these values lead to very satisfying results in the simulation. The maximal time used to find an optimal solution was 0.333 seconds, and the total calculation time was 1.549 seconds. Given that the number of time steps for the calculation was quite high, i.e. $n_{t_{calc}} = 121$, in order to produce pleasant plots, the calculation time suggests that the algorithm is well applicable even for systems with limited computing power.

In the given example, it is not surprising that the number of iterations depends on the intervals in which the constraint becomes active. In the second step, in which the time interval $[0.4, 1.6]$ is predicted, six iterations are needed, as the trajectory hits and leaves the constraint. Later, the calculations are easier to perform, even though the dimensions

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$2.00000E-01$	$1.95735E+00$
1	$3.57617E-04$	$5.63265E-02$
2	$5.93604E-05$	$1.04266E-01$
3	$5.24942E-06$	$8.03924E-02$
4	$1.86715E-08$	$1.23143E-02$
5	$7.93431E-13$	

(a) First Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$4.26015E-01$	$9.89940E-01$
1	$2.65456E-03$	$9.15294E-02$
2	$5.74878E-05$	$4.93657E-02$
3	$2.97132E-06$	$4.61292E-02$
4	$5.27395E-08$	$3.11659E-02$
5	$4.22407E-10$	$9.50863E-03$
6	$8.50779E-13$	

(b) Second Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$2.26002E-01$	$7.89571E-01$
1	$2.95960E-02$	$6.19402E-01$
2	$7.24123E-05$	$2.24016E-02$
3	$3.03829E-06$	$1.05638E-02$
4	$4.20130E-08$	$2.54524E-03$
5	$3.85125E-11$	

(c) Third Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$2.78895E-01$	$3.35391E-01$
1	$1.66753E-05$	$1.85661E-02$
2	$2.16874E-07$	$3.66097E-03$
3	$1.83028E-10$	$1.50931E-04$
4	$3.14266E-16$	

(d) Fourth Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$3.10643E-01$	$3.39366E-01$
1	$7.30048E-08$	$1.32900E-03$
2	$4.09751E-12$	

(e) Fifth Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$3.26300E-01$	$3.41456E-01$
1	$1.51797E-09$	$1.72744E-04$
2	$1.13469E-15$	

(f) Sixth Step

Table 7.2.: Iterations for the Inverse Pendulum: Calculation Steps

of the problem remain the same.

7.3.2. Trolley With State Constraints

Again, the trolley example is revised under state constraints. As the reference trajectory is calculated before the control and simulation process, the state constraints were also taken into account for the calculation of said trajectory. The state constraint that was imposed on the problem is $x_3 \leq 0.25$. This is a natural constraint, as in the context of the model, this means that the speed of the wagon is limited. Thus, the LQRP for this problem reads

Problem 7.9 (Trolley: *LQRP*)

$$\begin{aligned} \text{min!} \qquad J(\Delta x, \Delta u) :=& \frac{1}{2}\|\Delta x(t_f)\|_2^2 + \frac{1}{2}\int_{t_0}^{t_f} \|\Delta x(t)\|_2^2 + \frac{1}{10}\|\Delta u(t)\|_2^2 dt \\ &+ \frac{1}{2}\int_{\tau_i}^{\tau_f} \|w_\alpha(t)\|_2^2 dt \end{aligned}$$

$$\begin{aligned} \textit{with respect to} \quad & \Delta x \in W^{1,\infty}([t_0, t_f], \mathbb{R}^4), \\ & \Delta u \in L^\infty([t_0, t_f], \mathbb{R}^1), \\ \textit{and} \quad & w_\alpha \in L^\infty([\tau_i, \tau_f], \mathbb{R}^1), \end{aligned}$$

s.t.

$$\begin{aligned} \Delta\dot{x} \;=\; & f'_x(x_{ref}, u_{ref})\Delta x + f'_u(x_{ref}, u_{ref})\Delta u \\ & \textit{a.e. in } [t_0, t_f], \end{aligned}$$

and

$$\Delta x_3 + x_{ref} - \alpha w_\alpha \leq 0.25 \qquad \textit{in } [\tau_i, \tau_f],$$

where again f is defined as in the original trolley problem 7.6.

The first Figure 7.15 compares the outcome of the simulations for different virtual control parameters α. For $\alpha < 10^{-3}$, the virtual control values remained the same, so that no notable difference could be seen in the plots. In this example, the constraint was not obeyed completely for any value of α. The violation is, as described earlier, owed to the linearization of the system's dynamic. For this first set of plots, β was chosen so small that its influence on the simulations was neglectable. At first sight, it is visible that α again does not dramatically change the plots for values of the chosen size. For $\alpha = 1$, the plots looked more like the LQ simulation, but the constraint was basically ignored. As the LQ simulation is plotted for comparison, where the weight for the third state Q_{33} was set to 5 in order to encourage trajectories that would obey the constraint, it becomes clear that for the chosen weight, the violation of the constraint is not very satisfactory. The reason why the said weight was chosen lies in the plot of the first state. As the control process was startet in the initial state $\Delta x(t_0) = (-0.05; 0; 0; 0)^\top$, the task was to possibly go back to 0. As the plot shows, the LQ trajectory remained noticably below the other simulations, which means that setting $Q_{33} := 5$ did improve the violation slightly, but it also distracted the controller from its actual task.

Figure 7.16 shows that again the use of sufficiently small values for β is essential for good solutions. While for $\beta = 10^{-2}$, the LQC simulation lead to much worse results than the LQ simulation, the results remained pratically unchanged if β was chosen below 10^{-4}. However, for all values of β, the constraint was in this plot strictly obeyed.

The conclusion that can be drawn from these figures is that for sufficiently small values of α and β, the LQC showed a notable improvement in the constraint violations as well as in the original task. As both controller designs can be considered independent from the algorithm that is used for solving the occurring problem, they might also be taken into consideration in other regulation settings like real model predictive controllers.

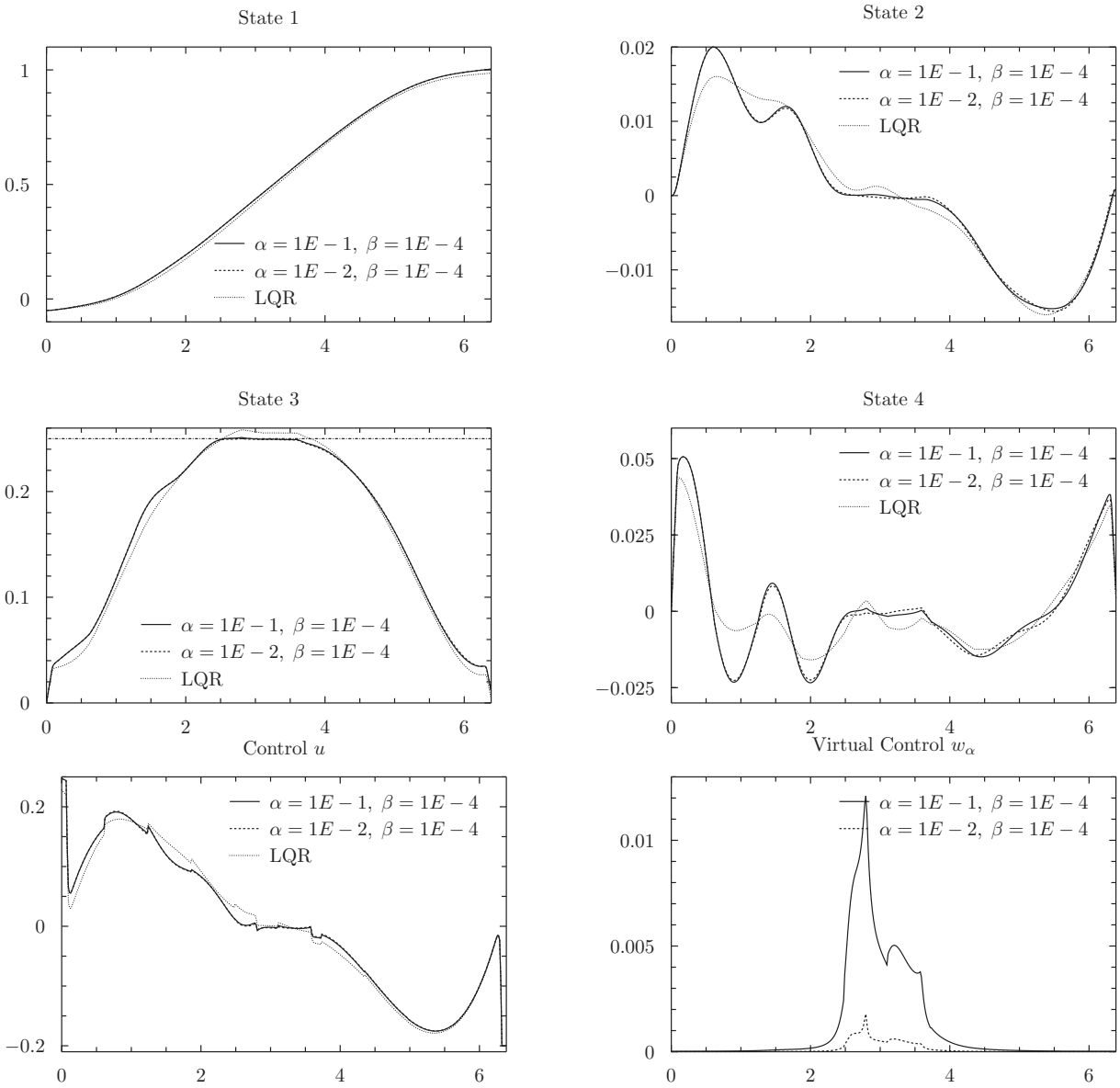


Figure 7.15.: Trolley Problem With State Constraints: Simulations for different values of the parameter α

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$2.00000E-01$	$3.85632E-01$
1	$6.28510E-02$	$2.54463E-01$
2	$1.50044E-03$	$8.95896E-02$
3	$2.65896E-05$	$1.81585E-02$
4	$2.98504E-07$	$4.72707E-03$
5	$5.05438E-10$	$2.92018E-04$
6	$3.01915E-15$	

(a) First Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$2.44547E-02$	$5.07445E-02$
1	$1.33512E-05$	$6.41324E-02$
2	$1.71759E-07$	$6.42286E-02$
3	$6.65935E-10$	$2.41058E-02$
4	$1.94245E-12$	

(b) Second Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$2.38481E-02$	$4.57406E-01$
1	$2.29196E-05$	$2.39113E-01$
2	$4.89248E-07$	$2.25262E-02$
3	$4.92583E-09$	$1.32603E-02$
4	$2.58371E-12$	

(c) Third Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$1.82405E-02$	$5.47085E-02$
1	$1.77612E-05$	$2.55310E-02$
2	$3.39932E-07$	$7.22392E-03$
3	$2.22879E-10$	$3.82003E-04$
4	$3.99205E-16$	

(d) Fourth Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$1.62306E-02$	$6.69842E-02$
1	$1.65866E-05$	$4.14300E-02$
2	$8.01618E-07$	$1.90887E-02$
3	$3.81022E-09$	$1.91570E-03$
4	$1.63971E-13$	

(e) Fifth Step

#It	$\|F_\beta(z^k)\|_2^2$	$\|d^k\|_\infty$
0	$1.37460E-02$	$7.30710E-02$
1	$1.90938E-06$	$1.41719E-02$
2	$5.30438E-09$	$3.38793E-03$
3	$3.73864E-12$	

(f) Sixth Step

Table 7.3.: Iterations for the Trolley Problem: Calculation Steps

The Table 7.3 shows the iterations needed during the calculation for the first six steps, for $\alpha = 10^{-3}$ and $\beta = 10^{-3}$. Again, the quadratic convergence is visible in all steps. The overall simulation intervall $[0, 6.370]$ was divided into 400 time steps.

The local problems were, as in the previous example, calculated with 120 steps (equivalent to 1.911 seconds), and the maximal time used for the calculation of a solution in each step was 0.235 seconds. As the resulting optimal control was applied for 40 time steps (equivalent to 0.637 seconds), the calculation time was significantly lower than the application time. In total, the calculations over the whole interval took 1.332 seconds. Using calculation environments with higher performance as well as a coarser grid for the interval would certainly render the controller design applicable in a real life application.

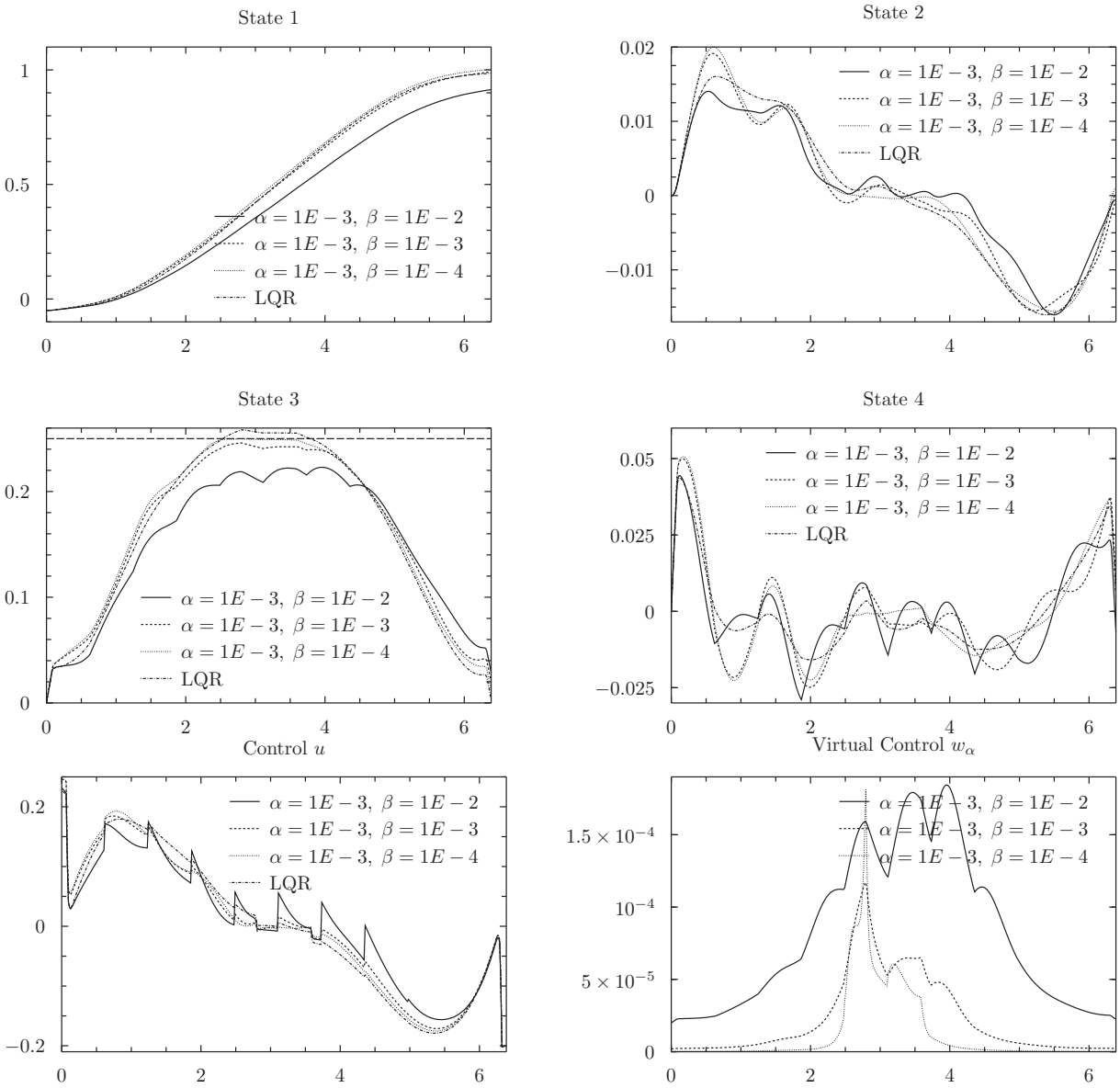


Figure 7.16.: Trolley Problem With State Constraints: Simulations for different values of the parameter β

8. Conclusions

In Chapter 3, the minimum principle was generalized for problems where mixed state control constraints as well as control set constraints are present. As the principle had already been developed for both types of constraints and due to the assumption was made that the different constraints only affected different controls, the main improvement in the theory lies in the usage of control theory for the proof of normality for the multipliers.

The virtual control concept in Chapter 4 was introduced for linear quadratic optimal control problems. By intuition, one would expect the convergence results to hold for general problems as well, but the restriction to linear quadratic problems turned out to be essential for this regularization. It is conceivable that the results may be generalized in the future. However, the theory has to be fitted in a nontrivial way. As the focus in this work lies on the application in linear quadratic regulation, the concept proved useful nevertheless.

The regularized Fischer-Burmeister function led to several globalizations of the Newton method. The results for the Combined Newton method were comparable to the globalized approach but suffered from the fact that suitable constant have to be chosen. The globalized method on the other hand seemed to slow down convergence at the beginning (one should mention that the example for which the iterations were evaluated is extremely ill-conditioned), but a good reason for its usage is that the iterations converge independently from the chosen constants. For this reason, in Chapter 7, the globalized method for fixed regularization parameters was used, although local Newton methods worked in most cases as well.

In the numerical experiments, the LQC algorithm worked well in the case of state constrained systems, while in the control constrained cases, similar results could be attained by exploiting the smaller computation effort of the LQ algorithm. A promising field for experiments might be the use of the described algorithms for fixed numbers of iterations. So far, the iterations were calculated until some tolerance was reached. As the linear system of differential equations is solved by any iteration in the Newton method, further iterations just improve the complementarity conditions. For problems where the constraints cannot be satisfied, as it may occur during the regulation process, the effect of regulating using just the first (or the first few) iterates may be worth observing. Also, it remains to research the question of how the computation time is affected if a nonlinear system of equations is used in the regualtion process, which leads to actual model predictive controllers instead of the linear quadratic approach. In that case, the iterations would probably have to be calculated with more accuracy, as the solution has to satisfy the differential equation. Appendix B finally shows that even in linear quadratic control, where the problems arising in applications seem to be of simple structure, unexpected effect may occur.

A. Auxiliary Proofs

This appendix contains proofs that would have disrupted the thesis if they had been proved in detail where they were used.

The first proof is a lemma about convex sets that is needed in lemma 4.15 that targeted on showing that the normality conditions for the original problem are also sufficient for normality of the regularized problem. The property that is needed in the proof is that a convex combination of an inner point and an arbitrary (in particular, a boundary point) of a convex set is an inner point:

Lemma A.1
Let S be a convex set in a normed vector space, and $x \in \partial S$, $y \in \operatorname{int} S$. If $\lambda \in (0,1)$, then $z := \lambda x + (1-\lambda)y \in \operatorname{int} S$.

For simplicity, the idea of the proof is depicted in Figure A.1. If $z \notin \operatorname{int} S$, then there exists a point z_ϵ nearby with $z_\epsilon \notin S$. A projection y_ϵ of such a point that is constructed using the principle of intersecting lines would then lie in a neighborhood of y, and therefore $y_\epsilon \in S$ if ϵ is chosen sufficiently small. Consequently, $z_\epsilon = \lambda x + (1-\lambda)y_\epsilon$ (by construction of y_ϵ) would belong to the convex set S.

Proof.
Assume that $z \notin \operatorname{int} S$, then it must hold $z \in \partial S$, since S is convex. For any $\varepsilon > 0$ there must be a point $z_\epsilon \in B_\epsilon(z)$ with $z_\epsilon \notin S$.

Let $y_\epsilon := \frac{1}{1-\lambda}(z_\epsilon - \lambda x)$. For proving that y_ϵ gets arbitrarily close to y, we first express y by means of x and z: As $\lambda \in (0,1)$, we find that

$$y = \frac{1}{1-\lambda}(z - \lambda x)$$

The distance of the new point from y is

$$\|y - y_\epsilon\| = \frac{1}{1-\lambda}\|z - \lambda x - z_\epsilon + \lambda x\|$$

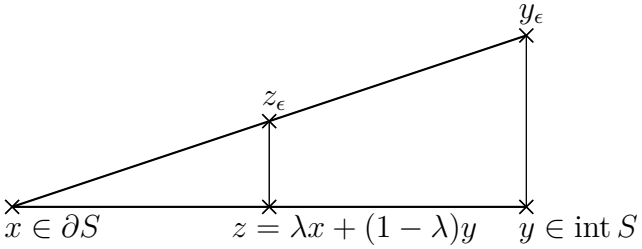


Figure A.1.: Construction of y_ϵ

$$\leq \frac{\epsilon}{1-\lambda}$$

This shows that for small ϵ, the point y_ϵ gets close to y. As y is an interior point, it must eventually hold that $y_\epsilon \in S$. Therefore, the point $z_\epsilon = \lambda x + (1-\lambda)y_\epsilon$ lies in the convex set S, which contradicts $z_\epsilon \notin S$. The assumption $z \notin \operatorname{int} S$ must be wrong. □

The second proof is a simple result that helps finding an estimate for the convergence radius of the Newton method.

Lemma A.2
Let $a_1, a_2, b_1, b_2 \in \mathbb{R}$ and $\beta > 0$. Then the estimate

$$\left| \frac{a_1}{\sqrt{a_1^2 + b_1^2 + \beta}} - \frac{a_2}{\sqrt{a_2^2 + b_2^2 + \beta}} \right| \leq \frac{|a_1 - a_2| + |b_1 - b_2|}{\sqrt{\beta}}$$

holds.

Proof.
Let $\beta > 0$, and $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be defined as

$$f(a,b) := \frac{a}{\sqrt{a^2 + b^2 + \beta}}.$$

For the partial derivatives f'_a and f'_b, it holds

$$\begin{aligned} |f'_a(a,b)| &= \left| \frac{b^2 + \beta}{\sqrt{a^2 + b^2 + \beta}(a^2 + b^2 + \beta)} \right| \leq \frac{1}{\sqrt{\beta}} \\ |f'_b(a,b)| &= \left| \frac{ab}{\sqrt{a^2 + b^2 + \beta}(a^2 + b^2 + \beta)} \right| \leq \left| \frac{\max\{a^2, b^2\}}{\sqrt{a^2 + b^2 + \beta}(a^2 + b^2 + \beta)} \right| \leq \frac{1}{\sqrt{\beta}}. \end{aligned}$$

For $(a_1, b_1), (a_2, b_2) \in \mathbb{R}^2$ we get

$$\begin{aligned} &\left| \frac{a_1}{\sqrt{a_1^2 + b_1^2 + \beta}} - \frac{a_2}{\sqrt{a_2^2 + b_2^2 + \beta}} \right| \\ &\quad = |f(a_1, b_1) - f(a_2, b_2)| \\ &\quad \leq |f(a_1, b_1) - f(a_2, b_1)| + |f(a_2, b_1) - f(a_2, b_2)| \end{aligned}$$

Now f is totally differentiable. Hence, according to the mean value theorem, there exist $\xi_a \in (a_1, a_2)$ and $\xi_b \in (b_1, b_2)$, such that

$$\begin{aligned} &|f(a_1, b_1) - f(a_2, b_1)| + |f(a_2, b_1) - f(a_2, b_2)| \\ &\quad = |f'_a(\xi_a, b_1)(a_2 - a_1)| + |f'_b(a_2, \xi_b)(b_2 - b_1)| \\ &\quad \leq \left| \frac{a_2 - a_1}{\sqrt{\beta}} \right| + \left| \frac{b_2 - b_1}{\sqrt{\beta}} \right| \end{aligned}$$

□

B. A Curious Regulation Example

In this chapter, another example for regulation is introduced, based on the state constraint variant of the trolley from section 7.3.2. The original moptivation for this example lies in the idea of dealing with higher order state constraints.

The physical system remains the same, and the constraint under consideration is $x_3(t) \geq -0.3$. Still, the only requirement that has to be met by the regularor is that the states x_1 and x_2 are tracked. At first glance, the constraint is of the same kind as any other state constraint that has been introduced so far, and the traditional LQR approach can be altered so that the state constraint is obeyed. Again, the price for altering the objective function of the regulator is that the tracking of the "important" first two states works less efficiently. However, there appears a quite unique problem in this example.

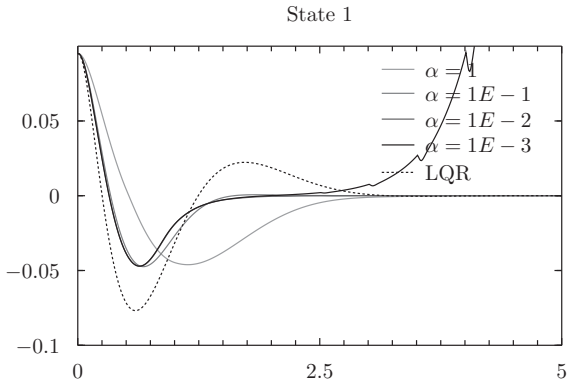
In order to understand the difficulties of the problem, we have to go back to the physical meaning of the task: An inverse pendulum is meant to be juggled so that it remains in (or near) the upright position. The constraint imposed on the third state means that the space available for the movement of the trolley is bounded (cf. Figure 7.2). Of course, the constraint $x_3 \geq -0.3$ is chosen in a fashion that ensures that the constraint becomes active. This means that the control of the pendulum has to be increased in comparision with the unconstrained control; the wagon has to be accelerated more, which increases the objective function, as the acceleration appears quadratically in the objective function. Consequently, a linear quadratic regulator that only aims at obeying the constraint and minimizing the objective function value will lead the wagon to the state $x_3 = -0.3$ and remain in this position as moving away from it only increases the costs. More precisely, the quadratic appearance of the control in the objective function leads to an oscillation around or near this point; an oscillation is tolerated (even if the constraint is violated), if it remains small and the control that is necessary to bring the system back to the upright position remains sufficiently small with respect to the L^2-norm.

The fact that the behavior of the system is restricted turns out fatal in the simulations: A violation of the constraint is tolerated (it is only taken into consideration in the L^2-norm sense), which leads to slight violations. Increasing controls have to be used to regulate this, which also affects the objective function, so they are kept low. As the pendulum states x_1 and x_2 are independent from the constrained state x_3, a deviation in the first state is accepted, and the pendulum tilts more. The tilt of the pendulum is considered less important than the violations, so that at some point the pendulum tips over. At the point when the angle of the pendulum becomes big enough to influence the objective function so much that the regulation reacts (approximately at $t = 3.5$), the tilt has already grown to much to be regulated again. During the subsequent intervals, the control explodes at the beginning (again due to the L^2-norm in the objective function), but the impulses from the

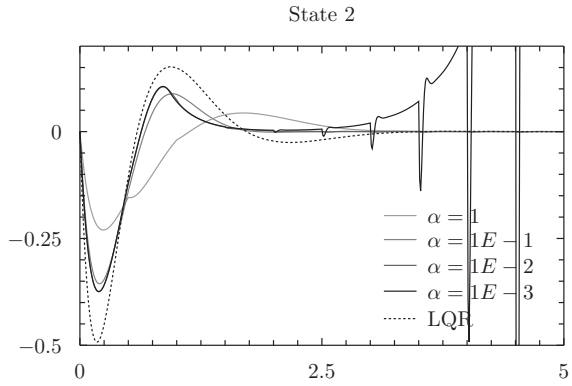
control are not sufficient to track the instable upper position of the pendulum.

A deduction that can be drawn from this example is that the system to be regulated needs to be deeply understood before an algorithm and an objective function for regulation is chosen. In this particular case, two options can be used to avoid the difficulties:

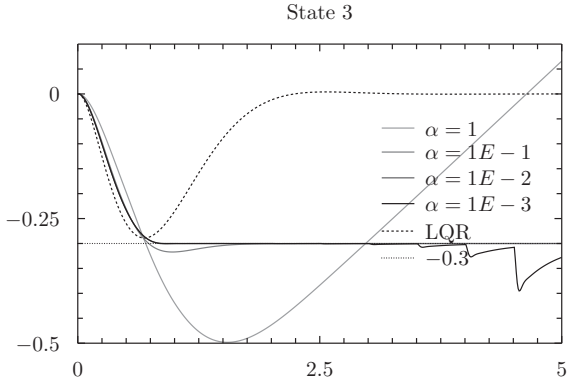
- The easiest way to regulate the given example is to use the traditional LQR algorithm with an altered objective function. The positive aspect is that pendulum stays in the middle of the available space (if the pendulum and the space are modeled accordingly), so that even consecutive occurrences of noise are regulated.
- Alternatively, the LQRC algorithm can be used if only small disturbances are to be expected. The parameter α for the virtual control may then be chosen sufficiently big, so that the system accepts bigger violations. This can be compensated by using more rigorous constraints in the calculation, like $x_3 \geq -0.25$, so that trajectories that come close to the constraints are already penalized.



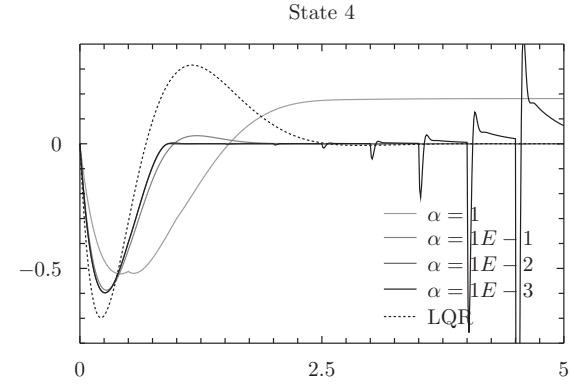
(a) LQR and global LQRC: x_1



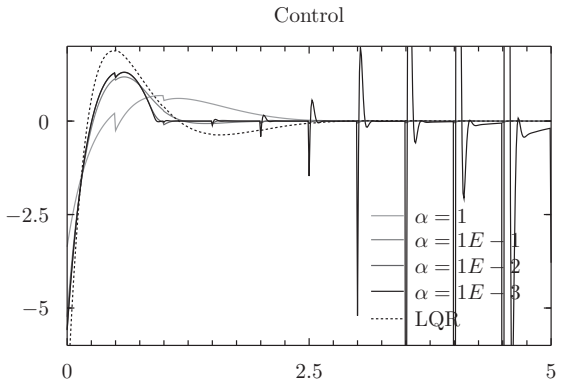
(b) LQR and global LQRC: x_2



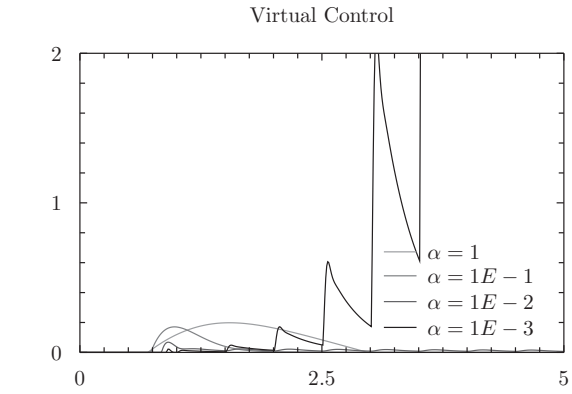
(c) LQR and global LQRC: x_3



(d) LQR and global LQRC: x_4



(e) LQR and global LQRC: u



(f) LQR and global LQRC: w_α

Figure B.1.: LQRC and LQR for the pendulum in a constrained space

C. The Controller Software

This chapter gives an overview of the Optimal Control software created in the frame of this work. The software is written in Scilab, an open source software package for numerical computations. The plots were made using PyXPlot. Alternatively, the data files created by the software kernel can be visualized using any other plotting software.

Each problem is placed in an own subdirectory of "`problems`". The `runme` script in the directory starts the calculations in Scilab and plots the data with PyXPlot[1]. For some examples, plots are made using different values for the parameters, e.g. for different values of α, the parameter for the virtual control. The name of the Scilab function that calls the example for different parameters start with "`article`". The solvers and wrapper scripts that are called can be found in the "`kernel`" directory.

First, the solvers shall be described in more detail. All solvers are written for linear quadratic optimal control problem with mixed control state constraints, but with no state constraints. As problems with state constraints are handled by regularization in the frame of this work, they are dealt with by the wrapper scripts. There also exists a script that facilitates the simulation process for the model predictive controller. The wrappers are described in the subsequent section.

Solvers

All solvers share a common structure: their return values are

1. the number of iterations needed,
2. the final residuum value with respect to the $\|\cdot\|_{Y^\infty}$-norm,
3. the final iterate
4. and the calculation time.

The parameters are `flags`, `t`, `Qf`, `Q`, `R`, `S`, `A`, `B`, `E0`, `E1`, `f`, `G`, `H`, `l`[2]. The array `t` contains all time steps, and all other functions that depend on the time are expected to be defined with respect to this grid, e.g. `Q(i,:,:)` $= Q(t_i)$, `l(i)` $= l(t_i)$. The names of the parameters in this list are chosen to coincide with their meaning in problem 4.1. As pure state constraints are not handled by the numerical solvers directly, only the functions that define mixed

[1] In order for this to work, the runme script as well as the pyxplot files have to be marked executable when used in Linux. More adjustments may be necessary for other operating systems.

[2] The names used in the function declaration differ from this list for historic reasons. For the function call however, the names used in the declaration are irrelevant.

function	Algorithm
`lq_alt`	Globalized Newton method 5.27.
`lq_comb`	Combined Newton method 5.23.
`lq_globloc`	Globalized Newton method for fixed values of β.
`lq_locloc`	Local Newton method for fixed values of β 5.16.

Table C.1.: Algorithm description for the solver functions

function	`flags`
`lq_alt`	verbosity, *tolerance, *start value
`lq_comb`	verbosity, *tolerance, *β_0, **C_{tol}, **c_β, *start value
`lq_globloc`	verbosity, *tolerance, *β, *start value
`lq_locloc`	verbosity, *tolerance, *β, *start value

Table C.2.: The flags lists for the solver functions

control state constraints are expected by the solvers. Pure state constraints are handled by a wrapper script, namely `lq_statesolv`.

The `flags` variable contains a list of parameters for the solver algorithms. The list in table C.1 explains the algorithm realized by the functions, and table C.2 contains descriptions of the flags list that these functions expect. Values marked with an asterisk (*) are optional arguments. For the function `lq_comb`, the double asterisk (**) indicates that the arguments C_{tol} and c_β are optional but can only be used together. All optional parameters have to be filled up from the left to the right. E.g. in order to use `lq_comb` with a specific start value, the user has to supply all other parameters as well.

Wrapper scripts

`lq_statesolv` Regularizes a linear quadratic problem. The actual solver can be changed easily[3]. The second parameter is used as the regularization parameter α, with $\kappa(\alpha) := 0$, $\phi(\alpha) := 1$, $\gamma(\alpha) := \alpha$. The `flags` list is forwarded (with the second parameter missing) to the solver. The other parameters include the matrix valued function C and the vector valued function d that define the state constraints. Altogether, the parameters are `flags`, `t`, `Qf`, `Q`, `R`, `S`, `A`, `B`, `E0`, `E1`, `f`, `C`, `d`, `G`, `H`, `l`, with the notation from problem 4.1.

`lq_regulate` Simulates a system controlled by the linear quadratic model predictive controller. The script expects the same flags as `lq_statesolv`, even though the α-flag is not used by this script. This happens in order to guarantee compatibility with `lq_stateregulate`. Apart from the `flags`, this script expects the number of application time steps `nt_apply`, the number of prediction time steps `ntlqr`, the

[3]This can be done in line 52. It is important to note that the solver used in this wrapper has to be initialized (using the exec-command) first.

time steps `t` that define the interval $[t_0, t_f]$, the right hand side of the ordinary differential equation `odef` and its derivatives `fx` and `fu`, the initial state `x0`, the functions `C` and `d` as in the problem definition (they are expected to have 0 columns, as `lq_stateregulate` should be used otherwise), `umin` and `umax` for the box constraints, the reference trajectory `prex` and `preu` as well as the matrices `Qf`, `Q` and `S`. The return value is a matrix that contains `t` as well as the simulated state and the control.

`lq_stateregulate` Simulates a state constrained system. It is used in the same fashion as `lq_regulate`. The expected parameters are the same. The return value is again a matrix describing the simulated system. An important difference between the two scripts is that `lq_stateregulate` can handle unconstraint controls: If it is called with `umin = umax`, then the simulation ignores any control constraints.

The reason for the existence of different wrappers for controller scenarios with and without state constraints is a glitch in Scilab: The sizes of tensors are not correctly calculated if one of the sizes are 0. Therefore, a state constraint defining matrix function $C \in \mathbb{R}^{n_t,0,n_x}$ could not be inserted into a block matrix. The same glitch makes it necessary to insert inactive constraints when the traditional LQR controller needs to be simulated. In this case, the lines 81 and 82 in `lq_regulate` have to be changed additionally, so that the control is cut off according to the box constraints.

Bibliography

[Alt06] Hans Wilhelm Alt. *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung.* Springer-Verlag Berlin Heidelberg, 5th edition, 2006.

[BH75] Arthur E. Bryson and Yu-Chi Ho. *Applied Optimal Control: Optimization, Estimation and Control.* Taylor and Francis Group, 1975.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[CKR08] S. Cherednichenko, K. Krumbiegel, and A. Rösch. Error estimates for the Lavrientiev regularization of elliptic optimal control problems. *Inverse Problems*, 24:1–21, 2008.

[CV90] F.H. Clarke and R.B. Vinter. Regularity properties of optimal controls. *SIAM J. Control Optimization*, 28(4):980–997, 1990.

[DH98] A. L. Dontchev and W. W. Hager. Lipschitzian stability for state constrained nonlinear optimal control. *SIAM J. Control Optimization*, 36(2):698–718, 1998.

[Dob06] Manfred Dobrowolski. *Angewandte Funktionalanalysis: Funktionalanalysis, Sobolev-Räume und ellipische Differentialgleichungen.* Springer, 2006.

[FK98] M. C. Ferris and C. Kanzow. Complementarity and related problems: A survey. *Mathematical Programming Technical Report*, 98-17, 1998.

[Ger06] Matthias Gerdts. *Optimal Control of Ordinary Differential Equations and Differential-Algebraic Equations.* Habilitation, Fakultät für Mathematik und Physik, Universität Bayreuth, 2006.

[Ger08] Matthias Gerdts. Global convergence of a nonsmooth Newton method for control-state constrained optimal control problems. *SIAM J. Optim.*, 19(1):326–350, 2008.

[GH10] Matthias Gerdts and Björn Hüpping. A linear-quadratic model-predictive controller for control and state constrained nonlinear control problems. pages 319–328, 2010.

[GHed] M. Gerdts and B. Hüpping. Virtual control regularization of state constrained linear quadratic optimal control problems. *Computational Optimization and Applications*, (accepted).

[GK02] Carl Geiger and Christian Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben.* Springer, 2002.

[GV03] Grant N. Galbraith and Richard B. Vinter. Lipschitz continuity of optimal controls for state constrained problems. *SIAM J. Control Optimization*, 42(5):1727–1744, 2003.

[Hag79] William W. Hager. Lipschitz continuity for constrained processes. *SIAM J. Control Optimization*, 17:321–338, 1979.

[HIK03] H. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semismooth newton method. *SIAM J. Optimization*, 13(3):865–888, 2003.

[HJ85] Roger A. Horn and Charles A. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[HSV95] Richard F. Hartl, Suresh P. Sethi, and Raymond G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Rev.*, 37(2):181–218, 1995.

[IT79] A.D. Ioffe and V.M. Tihomirov. *Theory of extremal problems*. Studies in Mathematics and its Applications, Vol. 6. Amsterdam, New York, Oxford: North-Holland Publishing Company. XII, 460 p., 1979.

[Kan00] Christian Kanzow. Global optimization techniques for mixed complementarity problems. *J. Glob. Optim.*, 16(1):1–21, 2000.

[Kim02] J.-H. R. Kim. *Optimierungsmethoden und Sensitivitätsanalyse für optimale bang-bang Steuerungen mit Anwendungen in der nichtlinearen Optik.* Dissertation, Universität Münster, 2002.

[Kön00] Konrad Königsberger. *Analysis 2*. Springer-Verlag, 3rd edition, 2000.

[KR08] K. Krumbiegel and A. Rösch. On the regularization error of state constrained Neumann control problems. *Control and Cybernetics*, 37(2):369–392, 2008.

[Kun06] Martin Kunkel. *Anwendung nichtglatter Newton-Verfahren auf Optimalsteuerungsprobleme mit reinen Zustandsbeschränkungen*. Diplomarbeit, Universität Hamburg, 2006.

[KWW78] Andreas Kirsch, Wolfgang Warth, and Jochen Werner. *Notwendige Optimalitätsbedingungen und ihre Anwendung.* Lecture Notes in Economics and Mathematical Systems. 152. Berlin-Heidelberg-New York: Springer-Verlag. 157 S., 1978.

[Lem72] Frank Lempio. *Tangentialmannigfaltigkeiten und Infinite Optimierung*. Habilitation, Universität Hamburg, 1972.

[Mal03] K. Malanowski. On normality of Lagrange multipliers for state constrained optimal control problems. *Optimization*, 52(1):75–91, 2003.

[MBM97] K. Malanowski, C. Büskens, and H. Maurer. Convergence of approximations to nonlinear optimal control problems. *Mathematical programming with data perturbations*, 195:253–284, 1997.

[Mer91] Nelson Merentes. On the composition operator in $\mathcal{AC}[a, b]$. *Collect. Math.*, 42(3):237–243, 1991.

[Nat75] I. P. Natanson. *Theorie der Funktionen einer reellen Veränderlichen*. Verlag Harri Deutsch, Zürich-Frankfurt-Thun, 1975.

[PBGM64] L.S. Pontrjagin, V.G. Boltjanskij, R.V. Gamkrelidze, and E.F. Mišcenko. *Mathematische Theorie optimaler Prozesse.* München-Wien: R. Oldenbourg., 1964.

[Son98] Eduardo D. Sontag. *Mathematical control theory. Deterministic finite dimensional systems. 2nd ed.* Texts in Applied Mathematics. Springer, New York, 1998.

[Taw09] M. A. Tawhid. An unconstrained optimization technique for nonsmooth nonlinear complementarity problems. *Journal of inequalities in pure and applied mathematics*, 10(3):14 pp., 2009.

[TW79] J. F. Traub and H. Wozniakowski. Convergence and complexity of newton iteration. *J. Assoc. Comput. Math.*, 29:250–258, 1979.

[Ulb03] Michael Ulbrich. Semismooth newton methods for operator equations in function spaces. *Siam J. Optimization*, 13(3):805–841, 2003.

[Wal00] Wolfgang Walter. *Gewöhnliche Differentialgleichungen*. Springer Berlin Heidelberg, 7th edition, 2000.

[Wan99] Xinghua Wang. Convergence of Newton's method and inverse function theorem in Banach space. *Math. Comput.*, 68(225):169–186, 1999.

[Wan00] Xinghua Wang. Convergence of Newton's method and uniqueness of the solution of equations in Banach space. *IMA J. Numer. Anal.*, 20(1):123–134, 2000.

[Wer07] Dirk Werner. *Functionalanalysis*. Springer-Lehrbuch. Berlin: Springer. xiii, 531 p., 6th edition, 2007.

[Wid46] D. V. Widder. *The Laplace Transform*. Princeton University Press, Princeton, 1946.

[WL03] Xinghua Wang and Chong Li. Convergence of Newton's method and uniqueness of the solution of equations in Banach spaces. II. *Acta Math. Sin., Engl. Ser.*, 19(2):405–412, 2003.