I N A U G U R A L – D I S S E R T A T I O N

zur

Erlangung der Doktorwürde

der

Naturwissenschaftlich–Mathematischen Gesamtfakultät

der

R U P R E C H T – K A R L S – U N I V E R S I T Ä T

H E I D E L B E R G

vorgelegt von

M.Sc. Matthias Schlöder

aus Bonn

Tag der mündlichen Prüfung

.....

Numerical Methods for Optimal Control of Constrained Biomechanical Multi-Body Systems Appearing in Therapy Design of Cerebral Palsy

Advisor

Prof. Dr. Ekaterina Kostina

Zusammenfassung

In dieser Arbeit entwickeln wir neue mathematische Modelle und Methoden für die optimale Steuerung beschränkter biomechanischer Mehrkörpersysteme (MKS) bei Problemen, die in der Planung von Behandlungen bei Zerebralparese (CP) auftreten. Wir modellieren den menschlichen Körper beim Gehen als beschränktes starres MKS, und den Gang als Lösung eines Optimalsteuerungsproblems (OSP), dessen Dynamik durch die des MKS gegeben ist. Hierbei führen wechselnde Fuß-Boden-Kontakte zu Sprüngen in den differentiellen Zuständen. Nimmt man an, dass es möglich ist, ein individuell kalibriertes OSP bereitzustellen, dessen (ausgewählte) Lösung das individuelle Gangbild eines Patienten modelliert, so kann ein solches Optimalsteuerungsmodell (OSM) für die Vorhersage der Auswirkungen von medizinischen Behandlungen auf das Gangbild genutzt werden. In diesem Zusammenhang betrachten wir drei Aspekte: sich potentiell ändernde Abfolgen von Fuß-Boden-Kontakt-Arten als Folge von medizinischen Behandlungen, Worst-Case Szenarien im Fall von auftretenden Unsicherheiten, z. B. bei der Durchführung eines Eingriffs, und eine geeignete Übersetzung von Behandlungen in Änderungen des genutzten OSM.

Für den Fall, dass die Abfolge der auftretenden Fuß-Boden-Kontakt-Arten nach einer medizinischen Behandlung unbekannt ist, entwickeln wir einen Ansatz für die numerische Lösung von geschalteten OSP mit Schaltkosten sowie Sprüngen in den differentiellen Zuständen, die beim Schalten auftreten können. Hierzu betrachten wir ein gemischt-ganzzahliges OSP und erweitern den Partial Outer Convexification Ansatz. Wir entwickeln zwei Typen von Schalt-Indikatoren. Diese können als Auslöser für Ereignisse, die mit bestimmten Schaltereignissen assoziiert sind, sowie für die Berechnung von Schaltkosten genutzt werden.

In den betrachteten OSM können medizinische Behandlungen als Änderungen von Parametern modelliert werden, welche in dem für die Modellierung des Ganges eingesetzten OSP auftreten. In der medizinischen Praxis treten in der Durchführung von Eingriffen jedoch unvermeidbare Ungenauigkeiten auf. Daher untersuchen wir Worst-Case Szenarien für parametrische OSP mit Unsicherheiten in den Parametern. Wir entwickeln und untersuchen einen Ansatz für die Bestimmung von ungünstigsten Parameterrealisierungen und den dazugehörigen Lösungen des parametrischen OSP, der für die modellbasierte Planung von Behandlungen bei CP Patienten geeignet ist. Hierbei betrachten wir ein zweistufiges Optimierungsproblem mit einem OSP auf der unteren Ebene.

Um unseren Ansatz für die Behandlungsplanung unter der Berücksichtigung von Worst-Case Szenarien einzusetzen, entwickeln wir ein dafür geeignetes Modell für medizinische Eingriffe. Da viele Behandlungen im Zusammenhang mit CP letztendlich darauf abzielen die Bewegungsfreiheit in Gelenken zu erhöhen, präsentieren wir einen Modellierungsansatz, der Behandlungen dieser Art in Änderungen von Parametern übersetzt, welche in der Dynamik des für die Modellierung des Ganges eingesetzten OSP auftreten.

Wir demonstrieren den Nutzen der entwickelten Ansätze in zwei Fallstudien.

Abstract

In this thesis, we develop new mathematical models and methods for the Optimal Control of constrained biomechanical Multi-Body Systems (MBSs) for problems appearing in therapy design of Cerebral Palsy (CP). We model the human body while walking as a constrained rigid MBS, and the gait as a solution of an Optimal Control Problem (OCP) which is constrained by the dynamics of this MBS. Here, changing foot-ground contact configurations lead to jumps in the differential states. Assuming that it is possible to provide a patient-specifically calibrated OCP whose (selected) solution models the gait of a patient, such kind of Optimal Control model can be employed to predict the effect of medical treatments on the gait pattern. In this setting, we focus on three aspects: possibly changing sequences of foot-ground contact configurations due to medical interventions, worst-case scenarios in presence of uncertainties, e. g., in the applied medical treatments, and a suitable translation of interventions into changes of the employed Optimal Control model.

For the case that the sequence of foot-ground contact configurations after a medical treatment is unknown, we develop an approach for the numerical solution of OCPs with switches, switching costs, and jumps in the differential states, which can occur at switching. For this, we consider a Mixed-Integer Optimal Control Problem and extend the Partial Outer Convexification approach. We develop two types of so-called switching indicators which are utilized on the one hand as a trigger for events that are associated with certain types of switches, and on the other hand for the computation of switching costs.

In the considered setting, medical interventions can be seen as changes of parameters that enter the gait modeling OCP. However, in medical practice unavoidable inaccuracies can occur in the implementation of an intervention. Therefore, we study worst-case scenarios for parametric OCPs with parameter uncertainties. We develop and examine an approach for the determination of worst-possible parameter realizations and the according OCP solutions which is suited for model-based treatment planning of CP. Here, we deal with a bilevel optimization problem with an OCP on the lower level.

In order to apply our approach for worst-case treatment planning, we provide a suitable model for medical treatments. Since many interventions in CP management eventually aim at extending the ranges of motion of joints, we present a modeling approach that translates treatments of this kind into changes of parameters which enter the dynamics of the gait modeling OCP.

The usefulness of the developed approaches is demonstrated in two case studies.

Contents

Chapter 1

Introduction

1.1 Motivation and Goals

Cerebral Palsy (CP) is an umbrella term for multiple disabilities affecting a patient's nervous system, musculature, and skeletal system [43, p. 40]. It is the most frequent cause of motor disorders in childhood, see., e. g., [43, p. 44]. Patients exhibit motor disorders that impair the ability to walk and, in case of ambulatory patients, cause pathological gait patterns. It is not possible to remediate the causal brain damage. However, there are multiple treatments to ameliorate the patients' situations. A particular emphasis is put on improving the patients' gait patterns and, more generally, their ability to walk. Orthopedic interventions play an important role in the therapy of pathologies concerning the musculoskeletal system in CP, and various kinds of surgeries are routinely applied in medical practice, see, e. g., [7, p. 458] and [43]. For the diagnosis and quantification of gait patterns, elaborated procedures – so-called Gait Analyses (GAs) – are applied. They play an important role with regard to the question of the best treatment option. However, to our best knowledge, predictive modeling tools are not considered in clinical decision making. Although physicians accumulated a lot of experience over the past decades and the applied treatments are beneficial for many patients, it was observed that (despite making use of GAs) a significant amount of interventions still yields a negative outcome, see [131, p. 3] and [30, 36]. In view of this, the development of suitable mathematical models and the generation of a digital testing environment, which enables the evaluation and assessment of possible interventions in silico and particularly in advance, is highly desirable.

The aim of this thesis is to provide mathematical models and methods which contribute to achieving this ultimate goal. The foundations for the present work were laid in [71]. As done, e. g., in the latter reference, we rely on the assumption that every individual human gait is optimal with respect to certain individual criteria. This yields a mathematical model that describes the human gait as the solution of an Optimal Control Problem (OCP) which is governed by the dynamics of a constrained biomechanical Multi-Body System (MBS) that models the human body while walk-

ing, see, e. g., [71, 105]. In [71], Kathrin Hatz deals with the determination of individual optimality criteria and provides methods for patient-specific calibration of Optimal Control models using motion capture data from GA – a line of research that is further being worked on in parallel to the work on this thesis, see Section 1.3.

We go one step beyond. Presuming that it is possible to provide an individually calibrated model of a CP patient as a digital twin, we are interested in modeling the effect of medical treatments on the patient's gait, also in view of possible uncertainties, e. g., in the accuracy of the performed intervention. We focus on three aspects:

1. The integration of possible inaccuracies during a performed intervention into treatment planning. Taking into account worst-case scenarios would make treatment planning more robust and would reduce the amount of negative intervention outcomes. Assuming that a performed intervention suffers from a certain degree of uncertainty, e. g., in the performed accuracy, we intend to develop a mathematical framework for the computation of a worst-case treatment and the corresponding outcome. Mathematically, this leads to a bilevel optimization problem with an OCP on the lower level.

2. The development of an intervention model which is suitable for worst-case treatment planning. Medical treatments shall be translated to changes of parameters which occur in the OCP employed for gait modeling. The intervention model has to be suitable for the usage in the above mentioned bilevel optimization framework for worst-case treatment planning.

3. Changing foot-ground contacts due to intervention. In medical practice, one observes that the gait phases – which are associated with certain parts of the feet being in contact with the ground – can change due to treatment. Common approaches, in which the human gait is modeled as a solution of a multi-stage OCP with a predefined order of phases, are not suited for reflecting this phenomenon in a predictive modeling environment. Therefore, we aim for a mathematical framework which enables us to treat the gait as a solution of a free-phase OCP in which the order of gait phases is subject to optimization.

1.2 Contributions

In the course of this thesis, we develop mathematical models and numerical methods for the Optimal Control of constrained biomechanical MBS appearing in therapy design of CP. We summarize our main contributions.

Switching Indicators and Costs

We consider the Optimal Control of switched dynamical systems with switching costs that are associated with a change of the so-called operation modes of the system. Starting with a Mixed-Integer Optimal Control Problem (MIOCP), we use Partial Outer Convexification (POC) [127] to reformulate and subsequently relax the considered problem. In this setting, we develop two new types of switching indicators, i. e., variables which recognize switching events: the so-called omniscient switching indicators and the so-called subsequent switching indicators. Both encode the information whether a switch occurred at a time t_s or not and can be employed for the computation of associated switching costs. The omniscient indicators comprise the information of the order of modes involved in switching, while the subsequent switching indicators only hold the operation mode after a switch. In contrast to the indicators described in [80, sec. 2.5], the new ones can be utilized as a trigger for events which are associated with switches between a certain ordering of modes. We investigate and compare switching costs which are associated with different switching indicators. Details are given in Chapter 5, with a focus on Section 5.4.

Optimal Control Problems with Switches, Switching Costs, and Jumps

In a commonly applied approach, the human gait is modeled as a solution of a multi-stage OCP with a predefined order of model phases and jumps in the differential states at phase transition, cf., e. g., [71, 105]. The model phases correspond to the occurring foot-ground contact configurations during a gait cycle. However, as explained previously, it is observed that the order and number of model phases can change due to medical treatment. Thus, for a predictive modeling of intervention outcomes, such a model is only useful to a limited extend. One way to overcome this shortcoming is to model the human gait as the solution of an OCP that is governed by a switched dynamical system in which the number or order of model phases is free and subject to optimization. In [26] the authors investigate switched OCPs, however *without* considering jumps in the differential states at switching. We extend the framework presented in the latter reference to make it suitable for our purposes. In doing so, we extend the Partial Outer Convexification approach [127]. Switching indicators and switching costs play a crucial role in our approach. Altogether, we present a novel approach for the numerical solution of switched systems with switching costs and possible jumps in the differential states at phase transitions. Details are given in Chapter 5.

Worst-Case Treatment Planning by Bilevel Optimal Control

Assuming we are provided an individually calibrated OCP whose solution models the human gait, we can make use of this model to evaluate and assess the effect of possible treatment options. However, in medical practice inaccuracies can occur during the implementation of an intervention when treating CP patients. Uncertainties like this have to be taken into account to robustly judge whether a planned treatment is reasonable or not. We model interventions by changes of model parameters $\mathbf{p}$ which enter a gait modeling OCP, and uncertainties by means of an uncertainty set $\Omega_{\mathbf{p}} \ni \mathbf{p}$. In this setting, we present our new so-called Training Approach for modeling worst-possible interventions and the corresponding outcomes. Here, we assume that the patient's body adapts functionally to the changes resulting from a treatment after a training period. Hence, uncertainty is not present anymore after training. Mathematically, this yields a bilevel optimization problem with a parametric OCP on the lower level. We investigate the differences between the Training Approach and a common approach from the field of Robust Optimization (see, e. g., [40]) in terms of the feasible sets and the objective function values and explain why the Training Approach is preferable for the application of treatment planning. Furthermore, we apply both approaches to a test case to illustrate their fundamental difference. We remark, that the Training Approach is developed in a general setting and its applications are not restricted to the field of CP treatment planning. Details are given in Chapter 6.

A Model for Orthopedic Surgeries Affecting a Joint's Range of Motion which is Suitable for Worst-Case Treatment Planning

In order to apply the Training Approach for worst-case treatment planning, we have to translate medical treatments into our mathematical gait model in a suitable manner. Many treatments in CP management eventually aim at extending the ranges of motion of joints. We focus on such treatments. Again, we model the human gait as the solution of an OCP that is governed by the dynamics of a constrained biomechanical MBS which models the human body while walking. Inspired by [5, 103], we implement so-called *passive reset forces* which – simply put – push back the generalized coordinates that represent the rotational states of the considered joint into a desired domain when they are about to leave it. Here, we focus on domains whose (virtual) bounds are encoded in the parameters $\underline{p}$ and $\overline{p}$. Through the passive reset forces, both parameters enter the dynamics of the OCP that is employed for gait modeling. We propose to model an intervention as a change of $\underline{p}$ and $\overline{p}$. This way, a

change of parameters yields a change of the resulting gait pattern. Details are given in Chapter 4.

Numerical Investigations

We conduct two case studies to show the usefulness of the developed approaches. First, we model the gait of an elementary walker model as the solution of an MIOCP. We employ the developed approach for the numerical solution of OCPs with switches, switching costs, and jumps and solve the resulting optimization problem. This way, we generate a walking motion. Second, we apply the Training Approach for worst-case treatment planning to the case of a fictive CP patient who is forced into a crouch gait by the disease. The situation shall be ameliorated by the application of an orthopedic surgery which, however, suffers from a certain degree of accuracy. We model the surgery as a change of parameters that enter the dynamics of the gait modeling OCP (see the previous paragraph), and compute the worst possible intervention and the according outcome using our Training Approach. Details are given in Chapter 7.

1.3 Overall Project and Cooperation Partners

The results presented in this thesis were developed as part of the overall project "Numerical Methods for Diagnosis and Therapy Design of Cerebral Palsy by Bilevel Optimal Control of Constrained Biomechanical Multi-Body Systems". In our work, we focus on aspects which are related to the therapy of CP patients. Another subproject, which is running in parallel, deals with the identification of modeling parameters, this way contributing to the development of classification schemes for CP gaits. Here, similar to the present work the human gait is modeled as a solution of a parametric OCP which is governed by the dynamics of a constrained MBS that models the human body while walking. The unknown model parameters shall be identified by detecting the OCP solution which approximates given measurements best. Mathematically, this yields an inverse OCP, i. e., a bilevel optimization problem with a parameter estimation problem on the upper level and an OCP on the lower level. Being able to solve such problems reliably would enable us to provide individually calibrated gait models for arbitrary CP patients.

During the project, we collaborated with Apl. Prof. Dr. Sebastian Wolf, head of the Heidelberg MotionLab [151] which is part of the Department of Orthopaedics and Trauma Surgery of Heidelberg University Hospital. The Heidelberg MotionLab is the GA laboratory of the Heidelberg University Hospital. GAs are executed in daily rou-

tine in order to provide the clinical decision makers – amongst others – with spatiotemporal data of investigated gait patterns using 3D motion capture systems. Despite the time and effort a GA takes, it is well-established in the clinical routine of CP management. The GAs are one of the major components in clinical decision making and contribute significantly to the question of the best treatment. If a surgical treatment is applied, the result is again observed by means of a GA that is executed after the musculoskeletal system adapted to the physiological changes which result from the applied treatment.
In addition, we collaborated with Prof. Dr. Katja Mombaur, an expert in optimization and simulation of human motions and in particular human gaits, and the working group "Optimization in Robotics and Biomechanics" (ORB) at the Institute of Computer Engineering at Heidelberg University.

The lively exchange and discussions with our cooperation partners facilitated the generation of a profound understanding and gave a deep insight into medical aspects of the project on the one hand, and modeling aspects on the other hand.

1.4 Thesis Overview

This thesis comprises eight chapters and two appendices, and is organized as follows.

The introduction is followed by Chapter 2, in which we provide the mathematical foundations for this thesis. First, we introduce optimization problems in Banach spaces. Subsequently, we focus on Nonlinear Programming Problems (NLPs) and OCPs. For the latter, we discuss so-called direct solution approaches in which (infinite dimensional) OCPs are transcribed to (finite dimensional) NLPs. Finally, we give a concise introduction to the generation of numerical derivatives as these are crucial for most of the optimization methods we employ.

In Chapter 3, we give an overview on CP. We take a look at causes, classification schemes, symptoms, gait analysis, and medical treatments. In view of the goals of this thesis, we focus on impairments of the gait of ambulatory CP patients and on orthopedic treatments which aim at improving CP gait patterns.

In Chapter 4, we introduce our approach to model-based treatment planning. We model the human body as a rigid MBS and the human gait as solution of an OCP that is constrained by the dynamics of the MBS. Furthermore, we give an introduction to

model-based treatment planning and present the general approach we pursue in this thesis. Subsequently, we propose to model a class of medical treatments in CP as changes of certain parameters which enter the dynamics of the OCP employed for gait modeling.

In Chapter 5, we consider the Optimal Control of switched dynamical systems with an a priori unknown switching structure, switching costs, and possible jumps of the differential states at switching. We formulate the considered problem as an MIOCP. We relax the problem by means of so-called switching indicator functions and convexification techniques. Different switching indicators are introduced and compared with each other, in particular with regard to the respectively associated switching costs. We discretize the resulting problem and state an approach for the numerical solution of the discretized problem.

In Chapter 6, we consider methods to predict the worst possible outcome of orthopedic interventions which suffer from uncertainty. We model a patient's gait as a solution of a parametric OCP in which an orthopedic intervention is reflected by a change of parameters $\Delta\mathbf{p}$. Assuming that $\Delta\mathbf{p}$ lies in an uncertainty set, we aim to identify a worst-possible treatment option and the related gait pattern. We propose our so-called Training Approach for worst-case treatment planning, which yields a bilevel optimization problem with an OCP on the lower level. We compare the Training Approach to a common approach from the field of Robust Optimization, comment on a numerical solution approach for the considered bilevel problem, and give an outlook on how to employ the proposed approach for a real-world application.

In Chapter 7, we demonstrate the usefulness of the approaches from Chapters 5 and 6 by conducting two case studies. In a first example, we use the free-phase approach from Chapter 5 to model the gait of an elementary walker MBS and to compute a gait pattern. Second, we apply the Training Approach from Chapter 6 to a fictive scenario – a CP patient who is forced into a crouch gait by the disease and undergoes an orthopedic surgery to ameliorate the situation. However, the intervention suffers from a certain degree of uncertainty. We use the Training Approach to investigate whether the intervention is recommendable or not in view of the present uncertainties.

Chapter 8 recaps the findings of this thesis and presents conclusions.

Appendix A contains supplementary material regarding rigid MBS dynamics and gait generating Optimal Control models which are governed by MBS dynamics. We de-

rive explicit expressions for the dynamics of an elementary walker MBS and exemplarily set up an OCP whose solutions model gait patterns of this walker.

In Appendix B, we collect all proofs from Chapters 5 and 6.

1.5 Acknowledgments

During the work on this thesis I experienced a lot of support from many sides. I gratefully acknowledge the funding and support of the Deutsche Forschungsgemeinschaft (DFG) through Priority Programme 1962 “Non-Smooth and Complementarity-Based Distributed Parameter Systems: Simulation and Hierarchical Optimization”. Furthermore, some supporters from the university environment are mentioned explicitly in the following. I’d like to thank my advisors and mentors Ekaterina Kostina and Hans Georg Bock for their support in all matters and the always pleasant collaboration. Special thanks also goes to my cooperation partners Katja Mombaur and Sebastian Wolf and their working groups. Thank you for sharing your knowledge and experience with me and always giving me the feeling of being welcome. Furthermore, I thank Andreas Potschka for the helpful discussions, advices, and the wide-ranging support, Christian Kirches and Andreas Meyer for the inspiring and motivating collaboration when working on switched systems, Andreas Meyer, Johannes Herold, Ihno Schrot, and Robert Scholz for reading and commenting on parts of this thesis, and Herta Fitzer – personally and as a representative for the Heidelberg University administration – for the support in administrative matters. To all of those who were not mentioned explicitly but supported me in any way while working on my thesis, I wish to express my sincere gratitude.

Chapter 2

Mathematical Background

In this chapter, we concisely introduce the mathematical background necessary for the subsequent chapters. In Section 2.1, we introduce general optimization problems in Banach spaces and certain Banach spaces which will be of interest in the context of Optimal Control Problems (OCPs). Section 2.2 is dedicated to Nonlinear Programming. We state first-order necessary conditions for Nonlinear Programming Problems (NLPs) and present selected solution methods. In Section 2.3, we consider the optimal control of dynamic systems. We introduce different types of dynamic systems, present a general problem formulation for OCPs, and discuss techniques for problem transformations. For the numerical solution of OCPs, in Section 2.4 we present two so-called direct solution approaches, namely Direct Multiple Shooting and Direct Collocation. Section 2.5 deals with different approaches for the numerical computation of derivatives including sensitivities.

2.1 Optimization in Banach Spaces

In this section, we state a general problem formulation for optimization problems in Banach spaces and introduce certain Banach spaces of interest. We follow the presentation in [60, ch. 2] and [102, ch. 2-3].

In this thesis, we consider optimization problems which fit into the following setting: let $(X, \|\cdot\|)$ be a Banach space over $\mathbb{R}$, $\Sigma \subseteq X$ a non-empty subset and $f : X \to \mathbb{R}$ a functional. A general optimization problem is given by

$$\begin{aligned} &\min_{\mathbf{x} \in X} \; f(\mathbf{x}) \\ &\text{s.t. } \mathbf{x} \in \Sigma. \end{aligned} \tag{2.1}$$

The functional f is called objective function or cost function. The set Σ is called the feasible set, and a vector $\mathbf{x} \in X$ is feasible for Problem (2.1) if and only if $\mathbf{x} \in \Sigma$. We consider X together with the norm topology (i. e., the topology induced by the metric $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$) as a topological space.

Definition 2.1 (Solutions of Optimization Problems)
We call a vector $\mathbf{x}^* \in \Sigma \subseteq X$

- a local solution (or local minimum) of Problem (2.1) if there is an open neighborhood $U \subseteq X$ of $\mathbf{x}^*$ such that

  $$f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ for all } \mathbf{x} \in U \cap \Sigma,$$

  and a strict local solution (or strict local minimum) if the above inequality is strictly satisfied for all $\mathbf{x} \neq \mathbf{x}^*$ in $U \cap \Sigma$,

- a global solution (or global minimum) of Problem (2.1) if

  $$f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ for all } \mathbf{x} \in \Sigma,$$

  and the strict global solution (or strict global minimum) if the above inequality is strictly satisfied for all $\mathbf{x} \neq \mathbf{x}^*$ in Σ,

- the unique solution of Problem (2.1) if $\mathbf{x}^*$ is the only local solution. In this case we call Problem (2.1) uniquely solvable. △

When using the term "a solution" in this thesis, we consider a local solution unless stated otherwise. For results on the existence of solutions of Problem (2.1) and necessary conditions we refer to [60, sec. 2.3].

For the remainder of this section, let $\mathcal{T} = [t_0, t_f] \subset \mathbb{R}$ with $t_0 < t_f$. We now introduce certain Banach spaces which will be of interest later.

Definition 2.2

- Let $1 \leq p < \infty$. The Lebesgue space of equivalence classes of measurable functions $f : \mathcal{T} \to \mathbb{R}$ for which $|f(\cdot)|^p$ is Lebesgue-integrable is denoted by $L^p(\mathcal{T}, \mathbb{R})$. Here, we identify functions which equal each other almost everywhere in $\mathcal{T}$ with respect to the Lebesgue measure. We equip the latter space with the norm

  $$\|f(\cdot)\|_p = \int_{t_0}^{t_f} |f(t)|^p \, \mathrm{d}t.$$

- The Lebesgue space of equivalence classes of essentially bounded measurable functions $f : \mathcal{T} \to \mathbb{R}$ is denoted by $L^\infty(\mathcal{T}, \mathbb{R})$. Again, we identify functions which

equal each other almost everywhere in $\mathcal{T}$ with respect to the Lebesgue measure. We equip $L^\infty(\mathcal{T},\mathbb{R})$ with the norm

$$\|f(\cdot)\|_\infty = \operatorname*{ess\,sup}_{t\in\mathcal{T}} |f(t)|\,.$$

- For $p \in [1,\infty]$ we define

$$L^p\left(\mathcal{T},\mathbb{R}^n\right) \stackrel{\text{def}}{=} \underbrace{L^p(\mathcal{T},\mathbb{R}) \times \cdots \times L^p(\mathcal{T},\mathbb{R})}_{n\text{-times}}$$

and equip $L^p(\mathcal{T},\mathbb{R}^n) \ni \mathbf{f}(\cdot)$ with the norm $\|\mathbf{f}(\cdot)\|_p \stackrel{\text{def}}{=} \max_{j=1,\dots,n} \|\mathbf{f}_j(\cdot)\|_p$. △

For $p \in [1,\infty]$ the space $L^p(\mathcal{T},\mathbb{R})$ together with the stated norm is a Banach space, cf. [91, Theorem 2.8.2] and [91, Theorem 2.11.7], and hence the same holds for the product space $L^p(\mathcal{T},\mathbb{R}^n)$. In this thesis, particularly the case $p = \infty$ is of interest since in OCPs (see Section 2.3) the so-called control functions are typically chosen to be elements of $L^\infty(\mathcal{T},\mathbb{R}^n)$.

Definition 2.3
A function $f : \mathcal{T} = [t_0, t_f] \to \mathbb{R}$ is called absolutely continuous if for all $\varepsilon > 0$, there is a $\delta > 0$ such that for all $m \in \mathbb{N}$ and $t_0 \le a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_n < b_m \le t_f$ the implication

$$\sum_{i=1}^{m} |b_i - a_i| < \delta \implies \sum_{i=1}^{m} |f(b_i) - f(a_i)| < \varepsilon$$

holds. △

We summarize several results on absolutely continuous functions which can be found, e. g., in [109, ch. 9]. An absolutely continuous function $f : \mathcal{T} \to \mathbb{R}$ is continuous, in particular essentially bounded, and differentiable almost everywhere with Lebesgue-integrable derivative. We consider the derivative as an element of $L^1(\mathcal{T},\mathbb{R})$ and denote it by $\frac{\mathrm{d}}{\mathrm{d}t} f(\cdot)$. We have

$$f(t) = f(t_0) + \int_{t_0}^{t} \frac{\mathrm{d}}{\mathrm{d}t} f(\tau)\,\mathrm{d}\tau\,.$$

Furthermore, for any $g \in L^1(\mathcal{T},\mathbb{R})$ the function $G(t) = g(t_0) + \int_{t_0}^{t} g(\tau)\,\mathrm{d}\tau$ is absolutely continuous with $\frac{\mathrm{d}}{\mathrm{d}t} G(t) = g(t)$ as element of $L^1(\mathcal{T},\mathbb{R})$.

Definition 2.4
The space of absolutely continuous functions $f : \mathcal{T} \to \mathbb{R}$ with essentially bounded derivatives is denoted by $W^{1,\infty}(\mathcal{T},\mathbb{R})$. We equip $W^{1,\infty}(\mathcal{T},\mathbb{R})$ with the norm

$$\|f(\cdot)\|_{1,\infty} \overset{\text{def}}{=} \max\left(\|f(\cdot)\|_\infty, \left\|\frac{\mathrm{d}}{\mathrm{d}t} f(\cdot)\right\|_\infty\right).$$

Furthermore, we define

$$W^{1,\infty}\left(\mathcal{T},\mathbb{R}^n\right) \overset{\text{def}}{=} \overbrace{W^{1,\infty}(\mathcal{T},\mathbb{R}) \times \cdots \times W^{1,\infty}(\mathcal{T},\mathbb{R})}^{n\text{-times}}$$

and equip $W^{1,\infty}(\mathcal{T},\mathbb{R}^n) \ni \mathbf{f}(\cdot)$ with the norm $\|\mathbf{f}(\cdot)\|_{1,\infty} \overset{\text{def}}{=} \max_{j=1,\dots,n} \|\mathbf{f}_j(\cdot)\|_{1,\infty}$. △

The space $W^{1,\infty}(\mathcal{T},\mathbb{R})$ together with the stated norm is a Banach space, cf. [102, sec. 2.4] and the references therein, and hence the same holds for the product space $W^{1,\infty}(\mathcal{T},\mathbb{R}^n)$. In this thesis, these spaces are of interest since in OCPs (see Section 2.3) they are a natural choice for the so-called differential states which obey a differential equation.

2.2 Nonlinear Programming

In this section, we give a concise introduction to Nonlinear Programming. We introduce the problem formulation and state first-order necessary conditions for optimality. Subsequently, we present selected solution methods.

2.2.1 Problem Formulation and First-Order Necessary Conditions

This subsection is based on the textbooks [59], [110], and [144]. We consider a general optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) \tag{2.2a}$$
$$\text{s.t. } \mathbf{h}_j(\mathbf{x}) = 0, \quad j = 1,\dots,p, \tag{2.2b}$$
$$\mathbf{g}_i(\mathbf{x}) \le 0, \quad i = 1,\dots,m, \tag{2.2c}$$

with continuous (and potentially nonlinear) functions $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^p$, and $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$. Such a problem is called an NLP. Note that Problem (2.2) is of Form (2.1). We denote the feasible set by $\mathcal{F}$. For a given $\mathbf{x} \in \mathcal{F}$,

$$\mathcal{A}(\mathbf{x}) = \left\{ i \in \{1,\dots,m\} \,\middle|\, \mathbf{g}_i(\mathbf{x}) = 0 \right\}$$

is the index set of active inequality constraints at $\mathbf{x}$. In the below presentation, we assume $f(\cdot)$, $\mathbf{h}(\cdot)$, and $\mathbf{g}(\cdot)$ to be continuously differentiable. The Lagrangian (function) $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ of Problem (2.2) is given by

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \stackrel{\text{def}}{=} f(\mathbf{x}) + \sum_{i=1}^{m} \boldsymbol{\lambda}_i \mathbf{g}_i(\mathbf{x}) + \sum_{j=1}^{p} \boldsymbol{\mu}_j \mathbf{h}_j(\mathbf{x}) .$$

Definition 2.5
We consider Problem (2.2). Let $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\boldsymbol{\mu} \in \mathbb{R}^p$. The conditions

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{0}, \tag{2.3a}$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{2.3b}$$
$$\boldsymbol{\lambda}_i \geq 0, \quad i = 1, \ldots, m, \tag{2.3c}$$
$$\mathbf{g}_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, m, \tag{2.3d}$$
$$\boldsymbol{\lambda}_i \mathbf{g}_i(\mathbf{x}) = 0, \quad i = 1, \ldots, m \tag{2.3e}$$

are called Karush-Kuhn-Tucker (KKT) conditions of Problem (2.2). A vector $(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ which satisfies the KKT conditions is called a KKT point of Problem (2.2). △

If additional conditions – so-called Constraint Qualifications (CQs) – are satisfied, the KKT conditions are first-order necessary conditions for optimality. We take a look at CQs.

Definition 2.6
A subset $C \subseteq \mathbb{R}^n$ is called a cone, if $r\mathbf{x} \in C$ for all $r > 0$ and $\mathbf{x} \in C$. If $C \subseteq \mathbb{R}^n$ is a non-empty cone, the set

$$C^\circ = \left\{ \mathbf{y} \in \mathbb{R}^n \,\middle|\, \mathbf{y}^T \mathbf{x} \leq 0 \text{ for all } \mathbf{x} \in C \right\}$$

is called the polar cone of C. △

Definition 2.7
Let the feasible set $\mathcal{F}$ of Problem (2.2) be non-empty and $\mathbf{x} \in \mathcal{F}$. The set

$$\mathcal{T}(\mathbf{x}) \stackrel{\text{def}}{=} \left\{ \mathbf{d} \in \mathbb{R}^n \,\middle|\, \exists t_k > 0, \mathbf{x}^k \in \mathcal{F} \text{ with } \lim_{k\to\infty} \mathbf{x}^k = \mathbf{x}, \lim_{k\to\infty} t_k = 0, \text{ and } \lim_{k\to\infty} \frac{\mathbf{x}^k - \mathbf{x}}{t_k} = \mathbf{d} \right\}$$

is called the tangent cone at $\mathbf{x}$ and

$$\mathcal{T}_{\text{lin}}(\mathbf{x}) \stackrel{\text{def}}{=} \left\{ \mathbf{d} \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \nabla \mathbf{g}_i(\mathbf{x})^T \mathbf{d} = 0 & \text{for } i \in \mathcal{A}(\mathbf{x}), \\ \nabla \mathbf{h}_j(\mathbf{x})^T \mathbf{d} = 0 & \text{for } j = 1, \ldots, p \end{array} \right\}$$

the linearized tangent cone at $\mathbf{x}$. △

For a given $\mathbf{x} \in \mathcal{F}$, both sets $\mathcal{T}(\mathbf{x})$ and $\mathcal{T}_{\text{lin}}(\mathbf{x})$ are indeed non-empty cones. We have $\mathcal{T}(\mathbf{x}) \subseteq \mathcal{T}_{\text{lin}}(\mathbf{x})$, cf. [144, Lemma 16.5], and consequently $\mathcal{T}_{\text{lin}}(\mathbf{x})^\circ \subseteq \mathcal{T}(\mathbf{x})^\circ$ for the polar cones.

Definition 2.8
Let $\mathbf{x} \in \mathcal{F}$. The condition $\mathcal{T}_{\text{lin}}(\mathbf{x})^\circ = \mathcal{T}(\mathbf{x})^\circ$ is called Guignard Constraint Qualification (GCQ) at $\mathbf{x}$. Any condition which implies GCQ at $\mathbf{x}$ is called a CQ at $\mathbf{x}$. △

Definition 2.9
Let $\mathbf{x} \in \mathcal{F}$ and $\mathcal{A}(\mathbf{x})$ be the corresponding index set of active inequality constraints.

- If the set of gradients $\{\nabla \mathbf{g}_i(\mathbf{x}) \mid i \in \mathcal{A}(\mathbf{x})\} \cup \{\nabla \mathbf{h}_j(\mathbf{x}) \mid j = 1, \ldots, p\}$ is linearly independent, we say that the Linear Independence Constraint Qualification (LICQ) holds at $\mathbf{x}$.
- If the set of gradients $\{\nabla \mathbf{h}_j(\mathbf{x}) \mid j = 1, \ldots, p\}$ is linearly independent and there is a $\mathbf{d} \in \mathbb{R}^n$ with

  $$\nabla \mathbf{g}_i(\mathbf{x})^T \mathbf{d} < 0,\ i \in \mathcal{A}(\mathbf{x}) \quad \text{and} \quad \nabla \mathbf{h}_j(\mathbf{x})^T \mathbf{d} = 0,\ j = 1, \ldots, p,$$

  we say that the Mangasarian-Fromovitz Constraint Qualification (MFCQ) is satisfied at $\mathbf{x}$. △

Theorem 2.10
Let $\mathbf{x} \in \mathcal{F}$. We have

$$\text{LICQ holds at } \mathbf{x} \implies \text{MFCQ holds at } \mathbf{x} \implies \text{GCQ holds at } \mathbf{x}.$$

In particular, LICQ and MFCQ are CQs at $\mathbf{x}$.

Proof For the first implication see [110, sec. 12.6] and a proof of the second implication can be found in [144, sec. 16.1-16.2]. □

We can now state first-order necessary conditions for local solutions of Problem (2.2).

Theorem 2.11
Let $\mathbf{x}^* \in \mathcal{F}$ be a local solution of Problem (2.2) such that a CQ is satisfied at $\mathbf{x}$. Then there are $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^p$ such that $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a KKT point of Problem (2.2).

Proof See [144, sec. 16.1]. □

If $f(\cdot), \mathbf{h}(\cdot)$, and $\mathbf{g}(\cdot)$ are twice continuously differentiable, one can further derive second-order (necessary and sufficient) conditions. For according results we refer to [59, sec. 2.2.6] and [110, sec. 12.5], respectively.

2.2.2 Selected Solution Methods

We now present selected solution approaches for NLPs which will be applied in this thesis.

Interior-Point Methods

We sketch the idea of Interior-Point methods following [110, ch. 19] where the interested reader can find details.

The NLP (2.2) can be stated equivalently as

$$\min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{s}\in\mathbb{R}^m} \quad f(\mathbf{x}) \tag{2.4a}$$
$$\text{s.t.} \quad \mathbf{0} = \mathbf{h}(\mathbf{x}), \tag{2.4b}$$
$$\mathbf{0} = \mathbf{g}(\mathbf{x}) - \mathbf{s}, \tag{2.4c}$$
$$\mathbf{s}_i \leq 0, \quad i = 1, \dots, m, \tag{2.4d}$$

by means of a slack vector $\mathbf{s} \in \mathbb{R}^m$ with non-positive components. Let $L(\mathbf{x}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ denote the Lagrangian of the latter problem. The KKT conditions then read as

$$\mathbf{0} = \nabla_{\mathbf{x},\mathbf{s}} L(\mathbf{x}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}), \tag{2.5a}$$
$$\mathbf{0} = \mathbf{h}(\mathbf{x}), \tag{2.5b}$$
$$\mathbf{0} = \mathbf{g}(\mathbf{x}) - \mathbf{s}, \tag{2.5c}$$
$$\boldsymbol{\lambda}_i \geq 0, \quad i = 1, \dots, m, \tag{2.5d}$$
$$\mathbf{s}_i \leq 0, \quad i = 1, \dots, m, \tag{2.5e}$$
$$\mathbf{s}_i \boldsymbol{\lambda}_i = \varepsilon, \quad i = 1, \dots, m, \tag{2.5f}$$

if $\varepsilon = 0$. For $\varepsilon = 0$ it is challenging to find a solution $(\mathbf{x}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ of (2.5) due to the Complementarity Conditions (2.5d-2.5f) since they include the determination of the optimal index set of active inequality constraints – a task of combinatorial nature [110, p. 565]. To resolve this issue, the idea of Interior-Point methods is to (approximately) solve System (2.5) for a sequence of strictly negative ε_k in order to generate a sequence $(\mathbf{x}^k, \mathbf{s}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$ which ideally converges to a point that satisfies (2.5) with $\varepsilon = 0$ (i. e., is a KKT point of Problem (2.4)) and furthermore is a minimizer of Problem (2.4). For details – in particular on different algorithmic realizations and convergence properties – we refer to [110, ch. 19].

Sequential Quadratic Programming

We sketch the idea of Sequential Quadratic Programming (SQP). In this paragraph, we follow [59, sec. 5.5], [110, ch. 18], and [144, ch. 19] where details can be found, respectively.

We first consider an equality constrained NLP

$$\min_{\mathbf{x} \in \mathbb{R}^n} \; f(\mathbf{x}) \tag{2.6a}$$

$$\text{s.t. } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{2.6b}$$

with twice continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^p$. Let $L(\mathbf{x}, \boldsymbol{\mu})$ denote the Lagrangian of Problem (2.6). The KKT conditions of Problem (2.6) are satisfied if and only if $(\mathbf{x}, \boldsymbol{\mu})$ is a root of the nonlinear function

$$\mathbf{F} : \mathbb{R}^{n+p} \to \mathbb{R}^{n+p}, \;\; (\mathbf{x}, \boldsymbol{\mu}) \longmapsto \begin{pmatrix} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\mu}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}.$$

Determining a KKT point of Problem (2.6) by finding a root of $\mathbf{F}(\cdot)$ with Newton's method is known as the Lagrange-Newton method, see [59, sec. 5.5.2].

Let $(\mathbf{x}^k, \boldsymbol{\mu}^k)$ be an iterate of the Lagrange-Newton method. We determine the next iterate $(\mathbf{x}^{k+1}, \boldsymbol{\mu}^{k+1}) = (\mathbf{x}^k, \boldsymbol{\mu}^k) + (\Delta\mathbf{x}, \Delta\boldsymbol{\mu})$ by solving the linear system of equations

$$\begin{pmatrix} \nabla^2_{\mathbf{xx}} L(\mathbf{x}^k, \boldsymbol{\mu}^k) & \nabla \mathbf{h}(\mathbf{x}^k) \\ \nabla \mathbf{h}(\mathbf{x}^k)^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Delta\mathbf{x} \\ \Delta\boldsymbol{\mu} \end{pmatrix} = - \begin{pmatrix} \nabla_{\mathbf{x}} L(\mathbf{x}^k, \boldsymbol{\mu}^k) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}. \tag{2.7}$$

On the other hand, we consider the quadratic problem

$$\min_{\Delta\mathbf{x}\in\mathbb{R}^n} \nabla f(\mathbf{x}^k)^T\Delta\mathbf{x} + \tfrac{1}{2}\Delta\mathbf{x}^T\nabla^2_{\mathbf{xx}}L(\mathbf{x}^k,\boldsymbol{\mu}^k)\Delta\mathbf{x}^T \tag{2.8a}$$

$$\text{s.t. } 0 = \mathbf{h}_j(\mathbf{x}^k) + \nabla\mathbf{h}_j(\mathbf{x}^k)^T\Delta\mathbf{x}, \qquad j = 1,\dots,p. \tag{2.8b}$$

One can show that $(\Delta\mathbf{x}, \Delta\boldsymbol{\mu})$ solves (2.7) if and only if $(\Delta\mathbf{x}, \boldsymbol{\mu}^k + \Delta\boldsymbol{\mu})$ is a KKT point of Problem (2.8), cf. [144, Lemma 19.4]. Thus, simply put we can determine the iterates of the Lagrange-Newton method by solving a sequence of quadratic problems and in this way find a KKT point of Problem (2.6). An introduction to Quadratic Programming can be found in [110, ch. 16].

This idea can be transferred to equality *and* inequality constrained NLPs of Form (2.2). The result is a so-called local SQP method which – under appropriate assumptions – can be shown to be locally convergent with quadratic convergence rate, cf. [59, sec. 5.5.3]. Furthermore, the (local) method can be modified to achieve global convergence. For details – in particular on algorithmic realizations and convergence properties – we refer to [59, sec. 5.5] and [110, ch. 18].

Derivative-Free Optimization – Model-Based Approach

Interior-Point methods and SQP methods both rely on the availability of derivative information. In this paragraph however, we consider optimization problems for which reliable derivatives are *not* available in practice (at least at acceptable computational costs) or do not exist everywhere. This raises the need for Derivative-Free Optimization (DFO) methods. An introduction to DFO is given in [34] and, more concisely, in [110, ch. 9].

Various DFO methods exist. In this thesis, we make use of a so-called model-based DFO approach for box-constrained optimization problems of the form

$$\begin{aligned} &\min_{\mathbf{x}\in\mathbb{R}^n} \; f(\mathbf{x}) \\ &\text{s.t. } \mathbf{a}_i \le \mathbf{x}_i \le \mathbf{b}_i \quad i = 1,\dots,n, \end{aligned}$$

see [124]. For an introduction to model-based DFO methods for unconstrained optimization problems, we refer to [110, sec. 9.2]. We give a short outline of a class of solution methods which includes the ones described in [124] and in [110, sec. 9.2]. Until satisfaction of termination conditions, in each iteration one proceeds as follows. The objective function $f(\cdot)$ is locally approximated by a linear or quadratic model

function $m^k(\cdot)$ which interpolates $f(\cdot)$ at a set of interpolation points comprising the current iterate $\mathbf{x}^k$. The model $m^k(\cdot)$ is used to determine a (feasible) iterate $\mathbf{x}^{k+1}$, e. g., by solving a trust region subproblem. In addition, the set of interpolation points is updated.

2.3 Optimal Control of Dynamic Systems

The present thesis deals with Optimal Control of dynamic systems, i. e., the optimization of dynamic processes which are described by means of dynamic systems that can be influenced or steered by so-called control functions. Comprehensive introductions to the topic of Optimal Control can be found, e. g., in [60] and [83], respectively. In this section, we introduce different types of dynamic systems, present a general problem formulation for OCPs, and discuss techniques for problem transformations.

2.3.1 Dynamic Systems

Let $\mathcal{T} = [t_0, t_f] \subset \mathbb{R}$ with $t_0 < t_f$. In this thesis, we consider dynamic systems which can be steered by a so-called control function and that are governed by Ordinary Differential Equations (ODEs) of first order. More precisely, the systems can be described by means of differential states $\mathbf{x} : \mathcal{T} \to \mathbb{R}^{n_x}$, and for a given control function $\mathbf{u} : \mathcal{T} \to \mathbb{R}^{n_u}$ the states $\mathbf{x}(\cdot)$ satisfy a differential equation of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t)) \,, \quad t \in \mathcal{T},$$

with $\mathbf{f} : \mathcal{T} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$. If the initial state of a system is given we consider a so-called Initial Value Problem (IVP): for a given control function $\mathbf{u}(\cdot)$ and an initial value $\mathbf{x}_0$, find differential states $\mathbf{x}(\cdot)$ such that

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t)) \,, \quad t \in \mathcal{T}, \tag{2.9a}$$

$$\mathbf{x}(t_0) = \mathbf{x}_0 \,. \tag{2.9b}$$

If $\mathbf{f}(\cdot)$ is continuous and globally Lipschitz continuous with respect to the second argument for the given $\mathbf{u}(\cdot)$, the solution $\mathbf{x}(\cdot)$ of the IVP is unique and depends continuously on the initial value $\mathbf{x}_0$. Furthermore, under additional requirements on the functional matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$, for every $t \in \mathcal{T}$ the solution $\mathbf{x}(\cdot)$ of the IVP is continuously differentiable with respect to $\mathbf{x}_0$. For the mentioned results, see [141, sec. 7.1]. If we

replace the Initial Condition (2.9b) by a boundary condition of form

$$\mathbf{0} = \mathbf{r}\big(\mathbf{x}(t_0), \mathbf{x}(t_f)\big), \quad \text{or} \quad \mathbf{0} = \mathbf{r}\big(\mathbf{x}(t_0), \mathbf{x}(t_1), \ldots, \mathbf{x}(t_f)\big) \quad \text{with } t_0 < t_1 < \cdots < t_f,$$

we obtain a so-called boundary value problem or multi-point boundary value problem, respectively. Details on ODEs and their numerical treatment can be found, e. g., in [141, ch. 7].

More generally, we consider dynamic systems which are governed by Differential Algebraic Equations (DAEs). Here, the systems can be described by means of states $\mathbf{z} : \mathcal{T} \to \mathbb{R}^{n_z}$, and for a given control function $\mathbf{u} : \mathcal{T} \to \mathbb{R}^{n_u}$ the states $\mathbf{z}(\cdot)$ satisfy a differential equation of the form

$$\mathbf{F}(t, \mathbf{z}(t), \dot{\mathbf{z}}(t), \mathbf{u}(t)) = \mathbf{0}, \quad t \in \mathcal{T},$$

with $\mathbf{F} : \mathcal{T} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_u}$. Assuming that $\mathbf{u}(\cdot)$ is sufficiently smooth, we say that the above DAE is of differential index $k \in \mathbb{N}$ if k is the smallest number of differentiations with respect to time that is needed to transfer the equation system

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)^i \mathbf{F}(t, \mathbf{z}(t), \dot{\mathbf{z}}(t), \mathbf{u}(t)) = \mathbf{0}, \quad i = 0, \ldots, k$$

into an ODE system of form

$$\dot{\mathbf{z}}(t) = \tilde{\mathbf{f}}\left(t, \mathbf{z}(t), \mathbf{u}(t), \ldots, \left(\tfrac{\mathrm{d}}{\mathrm{d}t}\right)^k \mathbf{u}(t)\right)$$

by means of algebraic manipulations (cf. [60, p. 28] and [68, p. 455]). For instance, the DAEs we consider in this thesis arise from the dynamics of constrained Multi-Body Systems (see Section 4.1.1) and are of differential index 3. A comprehensive introduction to the numerical treatment of DAEs can be found, e. g., in [68, ch. VI & VII].

Switched Dynamic Systems

In this thesis, we consider switched (dynamic) systems. A concise introduction to switched systems can be found in [102, ch. 1] and for more details we refer to [63, 97, 145]. A switched system consists of a set of dynamic subsystems which represent the *operation modes* or simply *modes* the system can run in. A switch denotes a change of modes, and the timed sequence of modes is called switching sequence. A switch from one mode into another is triggered by a signal which can be caused

internally, i. e., by the value of the differential states or a time event, or externally, i. e., by a control function. Accordingly, an internally switched system is a switched system in which all switches are caused internally, and the term externally switched system is defined analogously.

In this thesis, we take a look at the ODE case in which the dimension of the differential states is constant throughout all modes. Hence, the modes can be identified with the set of possible right-hand side functions of the differential equation, and a switch can be seen as a change of the right-hand side. We are interested in systems which can run in a finite number of modes. For internally switched systems – also called implicitly switched systems – the differential equation can be described by means of a switching function $\boldsymbol{\sigma} : \mathcal{T} \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_\sigma}$ in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}\big(t, \mathbf{x}(t), \mathbf{u}(t), \operatorname{sgn}(\boldsymbol{\sigma}(t, \mathbf{x}(t)))\big) .$$

Here, the current mode of the system depends on the sign of the components of $\boldsymbol{\sigma}(t, \mathbf{x}(t))$, and the switching sequence depends on the initial value $\mathbf{x}(t_0)$ as well as on the control function $\mathbf{u}(\cdot)$. On the other hand, for externally switched systems the current mode of the system can be encoded in the value a discrete valued control function $\mathbf{v} : \mathcal{T} \to \mathfrak{D}$ with $|\mathfrak{D}| < \infty$ (e. g., $\mathfrak{D} \subset \mathbb{N}$ finite subset), and the differential equation can be expressed in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) .$$

Here, the overall control function comprises a continuously valued part $\mathbf{u}(\cdot)$ and a discrete valued part $\mathbf{v}(\cdot)$ where the latter determines the switching sequence.

In certain applications, a switch can induce a discontinuity of the differential states $\mathbf{x}(\cdot)$ which we call a jump in the following. At switching time t_s, the transition from the states before the jump, $\mathbf{x}(t_s^-)$, to the states after jump, $\mathbf{x}(t_s^+)$, can be expressed by means of a jump function $\boldsymbol{\Delta}(\cdot)$. Depending on the particular application, the jump function depends on the mode before the switch, after the switch, or both. For externally switched systems, the jump condition can be stated in the form

$$\mathbf{x}(t_s^+) = \boldsymbol{\Delta}\left(t_s, \mathbf{x}(t_s^-), \mathbf{v}(t_s^-), \mathbf{v}(t_s^+)\right) ,$$

where $\mathbf{v}(t_s^-)$ and $\mathbf{v}(t_s^+)$ encode the mode of the system before and after the switch, respectively.

2.3.2 Problem Formulation

We present a general problem formulation for OCPs. In words, we seek for so-called controls that steer a given dynamic process from an initial state to a terminal state in an optimal manner with respect to a performance criterion while satisfying constraints which are imposed on the process. More precisely, we consider optimization problems, e. g. (and w. l. o. g.) of a standardized form

$$\min_{\mathfrak{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot)} \Phi^M(\mathbf{x}(1)) + \int_0^1 \Phi^L(\mathbf{x}(t),\mathbf{u}(t))\,\mathrm{d}t \tag{2.10a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t)), \qquad t \in [0,1], \tag{2.10b}$$

$$\mathbf{0} \le \mathbf{c}(\mathbf{x}(t),\mathbf{u}(t)), \qquad t \in [0,1], \tag{2.10c}$$

$$\mathbf{0} \le \mathbf{r}(\mathbf{x}(0),\mathbf{x}(1),\mathfrak{u},\mathbf{p}), \tag{2.10d}$$

$$\mathfrak{u} \in \mathcal{P} \subseteq \mathbb{R}^{\mathfrak{n}_\mathfrak{u}}, \tag{2.10e}$$

$$\mathbf{u}(t) \in \mathcal{U} \subseteq \mathbb{R}^{n_u}, \qquad t \in [0,1], \tag{2.10f}$$

with

- so-called controllable parameters $\mathfrak{u} \in \mathbb{R}^{\mathfrak{n}_\mathfrak{u}}$,
- control function $\mathbf{u} : [0,1] \to \mathbb{R}^{n_u}$,
- differential states $\mathbf{x} : [0,1] \to \mathbb{R}^{n_x}$,
- (non-controllable) parameters $\mathbf{p} \in \mathbb{R}^{n_p}$,
- model functions $\Phi^M : \mathbb{R}^{n_x} \to \mathbb{R}$, $\Phi^L : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$, $\mathbf{f} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$, $\mathbf{c} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_c}$, and $\mathbf{r} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{\mathfrak{n}_\mathfrak{u}} \times \mathbb{R}^{n_p} \to \mathbb{R}^{n_r}$,

where all inequalities are assumed to hold component-wise. Remark, that in contrast to the non-controllable parameters $\mathbf{p}$, the controllable parameters $\mathfrak{u}$ are subject to optimization. Optimization problems of this kind are called Optimal Control Problems (OCPs). The above problem formulation covers many situations, such as externally switched systems, free time horizons, so-called multi-stage problems with possible jumps in the differential states (cf. Problem (2.14)), etc., as we will we see in the subsequent paragraphs. The terms $\Phi^M(\mathbf{x}(1))$ and $\int_0^1 \Phi^L(\mathbf{x}(t),\mathbf{u}(t))\,\mathrm{d}t$ in the Objective Function (2.10a) are called Mayer term and Lagrange term, respectively. Furthermore, we refer to the Constraints (2.10c) as path constraints and to (2.10d) as boundary constraints. Note that the above problem formulation formally includes

equality constraints as well, as they can be expressed by means of two opposing inequality constraints. We consider Problem (2.10) as an optimization problem in the Banach space

$$\mathbb{R}^{\mathfrak{n}_\mathfrak{u}} \times L^\infty\left([0,1],\mathbb{R}^{n_u}\right) \times W^{1,\infty}\left([0,1],\mathbb{R}^{n_x}\right) \ni (\mathfrak{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot)) \,,$$

see Section 2.1. Thus, Problem (2.10) is an infinite dimensional optimization problem of Form (2.1).

Mixed-Integer Optimal Control Problems

In Problem Formulation (2.10) we do not make assumptions on the sets $\mathcal{P} \ni \mathfrak{u}$ and $\mathcal{U} \ni \mathbf{u}(t)$. If one of the sets is discrete, we speak of a Mixed-Integer Optimal Control Problem (MIOCP) – an OCP which involves continuously valued and discrete valued control variables. MIOCPs arise, e. g., in connection with switched systems in which the switching sequence is free and subject to optimization by encoding the switching sequence in the time course of the control function (see, e. g., Chapter 5). Due to the discrete valued optimization variables, MIOCPs demand for a different treatment than continuous OCPs. As entry points to the topic we refer to [60, ch. 7], [81, ch. 2], and [127].

2.3.3 Transformation Techniques

OCPs can be formulated in many different ways. In the following, we present techniques to transfer frequently arising problem formulations to Form (2.10). Though this is important for theoretical considerations, in practice it is often beneficial to use the original problem formulation together with tailored solution methods.

Maximization. As the maximization of a real valued function $\Phi(\cdot)$ is equivalent to the minimization of $-\Phi(\cdot)$, OCPs in which the objective function is maximized can be equivalently transferred to a minimization problem by changing the sign of the cost function.

Time Horizon (see [60, sec. 1.2.1]). In Problem (2.10), we consider the fixed normalized time horizon $[0,1]$. However, we are also interested in similar problems with a time horizon $[t_0,t_f]$ where possibly both t_0 and t_f can either be fixed or free, respectively. In this situation, we employ a linear time transformation

$$t : [0,1] \to [t_0,t_f], \quad \tau \mapsto t_0 + \tau(t_f - t_0)$$

and define time transformed differential states $\tilde{\mathbf{x}} : [0,1] \to \mathbb{R}^{n_x}$ and control functions $\tilde{\mathbf{u}} : [0,1] \to \mathbb{R}^{n_u}$ by $\tilde{\mathbf{x}}(\tau) = \mathbf{x}(t(\tau))$ and $\tilde{\mathbf{u}}(\tau) = \mathbf{u}(t(\tau))$, respectively. According to the chain rule, we obtain for the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\tilde{\mathbf{x}}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\mathbf{x}(t(\tau)) = \dot{\mathbf{x}}(t(\tau))\frac{\mathrm{d}}{\mathrm{d}\tau}t(\tau) = (t_f - t_0)\mathbf{f}(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau)). \tag{2.11}$$

Path constraints as well as boundary constraints can be directly expressed in the Form (2.10c) and (2.10d), respectively, by means of $\tilde{\mathbf{x}}(\cdot)$ and $\tilde{\mathbf{u}}(\cdot)$, and the objective function is transformed to

$$\Phi^M(\tilde{\mathbf{x}}(1)) + \int_0^1 (t_f - t_0)\Phi^L(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau))\,\mathrm{d}\tau\,.$$

If t_0 or t_f are free variables, we add them to the vector of controllable parameters. Altogether, we obtain a problem of Form (2.10).

Parameter-Dependence of Model Functions. In Problem (2.10), the parameters $\mathbf{u}$ and $\mathbf{p}$ do not enter $\Phi^M(\cdot)$, $\Phi^L(\cdot)$, $\mathbf{f}(\cdot)$, $\mathbf{c}(\cdot)$, and solely appear in the Boundary Constraints (2.10d). In the following, we focus on $\mathbf{p}$ since $\mathbf{u}$ can be treated similarly. If we consider a more general OCP with the parameters $\mathbf{p}$ entering the objective function, $\mathbf{f}(\cdot)$, or $\mathbf{c}(\cdot)$, we introduce additional constant differential states $\tilde{\mathbf{p}}(\cdot)$ with the initial value $\mathbf{p}$. The augmented differential equation is given by

$$\begin{pmatrix}\dot{\mathbf{x}}(t)\\ \dot{\tilde{\mathbf{p}}}(t)\end{pmatrix} = \begin{pmatrix}\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p})\\ \mathbf{0}\end{pmatrix} = \begin{pmatrix}\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \tilde{\mathbf{p}}(t))\\ \mathbf{0}\end{pmatrix} \tag{2.12}$$

and can be stated in the form $\dot{\mathbf{y}} = \tilde{\mathbf{f}}(\mathbf{y}(t), \mathbf{u}(t))$ with the augmented states $\mathbf{y}(\cdot) = \begin{pmatrix}\mathbf{x}(\cdot)\\ \tilde{\mathbf{p}}(\cdot)\end{pmatrix}$. Path constraints and the objective function can be expressed in terms of $\mathbf{y}(\cdot)$ and $\mathbf{u}(\cdot)$. Furthermore, the boundary constraints are replaced by suitable constraints of the form

$$\mathbf{0} \leq \tilde{\mathbf{r}}(\mathbf{y}(0), \mathbf{y}(1), \mathbf{u}, \mathbf{p}) \tag{2.13}$$

which include the condition $\tilde{\mathbf{p}}(0) = \mathbf{p}$. Altogether, we get a problem of Form (2.10).

Non-Autonomous Problems (see [60, sec. 1.2.2]). In Problem (2.10), the model functions are not explicitly time-dependent which renders the problem a so-called autonomous problem. If one of the functions $\Phi^L(\cdot)$, $\mathbf{f}(\cdot)$, $\mathbf{c}(\cdot)$, $\mathbf{r}(\cdot)$ does explicitly depend on t, we introduce an additional differential state $\tilde{t} : [0,1] \to [0,1]$ with $\frac{\mathrm{d}}{\mathrm{d}t}\tilde{t}(t) = 1$ and

$\tilde{t}(0) = 0$. We augment the differential equation and the boundary constraints as in (2.12) and (2.13) and replace all explicit time dependencies by $\tilde{t}(t)$ to obtain a problem of Form (2.10).

Transformations of Objective Functions. In Problem (2.10), the Mayer term solely depends on $\mathbf{x}(1)$. If $\Phi^M(\cdot)$ depends on $\mathbf{x}(0)$ as well, we introduce an additional constant differential state with initial value $\mathbf{x}(0)$ which enables us to formulate the Problem in Form (2.10) again. Furthermore, we can eliminate the Lagrange term and focus on Mayer type objective functions by introducing an additional differential state $l(\cdot)$ with $\dot{l}(t) = \Phi^L(\mathbf{x}(t), \mathbf{u}(t))$ and $l(0) = 0$. The objective function is then given by the term $\Phi^M(\mathbf{x}(1)) + l(1)$ which can be brought in form $\tilde{\Phi}^M(\tilde{\mathbf{x}}(1))$ by introducing an additional differential state encoding the time, see the previous paragraph.

Multi-Stage OCPs (compare [60, sec. 1.2.5]). In this thesis, we deal with problems of the form

$$\min_{\substack{\mathbf{x}^1(\cdot),\dots,\mathbf{x}^n(\cdot),\\ \mathbf{u}^1(\cdot),\dots,\mathbf{u}^n(\cdot),\\ T_1,\dots,T_n}} \quad \sum_{j=1}^{n} \left[\Phi_j^M\left(T_j, \mathbf{x}^j(T_j), \mathbf{p}\right) + \int_{T_{j-1}}^{T_j} \Phi_j^L\left(\mathbf{x}^j(t), \mathbf{u}^j(t), \mathbf{p}\right) \mathrm{d}t \right] \tag{2.14a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}^j(t) = \mathbf{f}^j\left(\mathbf{x}^j(t), \mathbf{u}^j(t), \mathbf{p}\right), \quad t \in \mathcal{T}_j, \qquad j = 1,\dots,n, \tag{2.14b}$$

$$T_{j-1} \le T_j, \qquad j = 1,\dots,n, \tag{2.14c}$$

$$\mathbf{x}^{j+1}(T_j) = \mathbf{\Delta}^j\left(\mathbf{x}^j(T_j), \mathbf{p}\right), \qquad j = 1,\dots,n, \tag{2.14d}$$

$$\mathbf{0} \le \mathbf{c}^j\left(\mathbf{x}^j(t), \mathbf{u}^j(t), \mathbf{p}\right), \quad t \in \mathcal{T}_j, \tag{2.14e}$$

$$\mathbf{0} \le \mathbf{r}\left(\mathbf{x}^1(T_0), \mathbf{x}^1(T_1), \dots, \mathbf{x}^n(T_n), \mathbf{p}\right), \tag{2.14f}$$

with fixed $T_0 \in \mathbb{R}$ and $\mathcal{T}_j = [T_{j-1}, T_j]$. Problems of this kind are called multi-stage OCPs. They include $j = 1,\dots,n$ consecutive so-called model stages, also denoted by the term phases. Each model stage is assigned a time horizon $\mathcal{T}_j$, optimization variables $\mathbf{x}^j(\cdot) : \mathcal{T}_j \to \mathbb{R}^{n_{x,j}}$, $\mathbf{u}^j(\cdot) : \mathcal{T}_j \to \mathbb{R}^{n_{u,j}}$, $T_j \in \mathbb{R}$, a set of constraints and an objective function contribution. Furthermore, the Constraints (2.14d) describe the transition of the values of the differential states $\mathbf{x}^j(\cdot)$ at phase transition. Multi-stage OCPs arise in connection with switched dynamic systems with predefined sequences of modes and will be used for gait modeling in this thesis, see Section 4.1.2. If $n = 1$ and $\mathbf{\Delta}^1(\cdot) = \mathbf{Id}(\cdot)$, we obtain a so-called single-stage Problem.

We explain how Problem (2.14) can be reformulated into an equivalent problem of Form (2.10) by means of the techniques described in the previous paragraphs. To

this end, we introduce time transformations

$$t^j : [0,1] \to \mathcal{T}_j, \quad \tau \mapsto T_{j-1} + (T_j - T_{j-1})\tau, \quad j = 1, \dots, n,$$

as well as differential states $\tilde{\mathbf{x}}^j : [0,1] \to \mathbb{R}^{n_{x,j}}$ and control functions $\tilde{\mathbf{u}}^j : [0,1] \to \mathbb{R}^{n_{u,j}}$ which are given by $\tilde{\mathbf{x}}^j(\tau) = \mathbf{x}\big(t^j(\tau)\big)$ and $\tilde{\mathbf{u}}^j(\tau) = \mathbf{u}\big(t^j(\tau)\big)$, respectively. Similar to (2.11), we obtain the differential equations

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\tilde{\mathbf{x}}^j(\tau) = (T_j - T_{j-1})\mathbf{f}\big(\tilde{\mathbf{x}}^j(\tau), \tilde{\mathbf{u}}^j(\tau), \mathbf{p}\big).$$

We set

$$\mathbf{u} = \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}, \ \tilde{\mathbf{x}}(\tau) = \begin{pmatrix} \tilde{\mathbf{x}}^1(\tau) \\ \vdots \\ \tilde{\mathbf{x}}^n(\tau) \end{pmatrix}, \ \tilde{\mathbf{u}}(\tau) = \begin{pmatrix} \tilde{\mathbf{u}}^1(\tau) \\ \vdots \\ \tilde{\mathbf{u}}^n(\tau) \end{pmatrix}, \text{ and } \tilde{\mathbf{c}}(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau), \mathbf{p}) = \begin{pmatrix} \mathbf{c}^1(\tilde{\mathbf{x}}^1(\tau), \tilde{\mathbf{u}}^1(\tau), \mathbf{p}) \\ \vdots \\ \mathbf{c}^n(\tilde{\mathbf{x}}^n(\tau), \tilde{\mathbf{u}}^n(\tau), \mathbf{p}) \end{pmatrix}.$$

The Conditions (2.14c), (2.14d), and (2.14f) can be expressed together equivalently by a boundary constraint of the form $\mathbf{0} \le \tilde{\mathbf{r}}\big(\tilde{\mathbf{x}}(0), \tilde{\mathbf{x}}(1), \mathbf{u}, \mathbf{p}\big)$. Thus, Problem (2.14) can be transformed into a problem of the form

$$\min_{\mathbf{u}, \tilde{\mathbf{u}}(\cdot), \tilde{\mathbf{x}}(\cdot)} \ \tilde{\Phi}^M\big(\tilde{\mathbf{x}}(1), \mathbf{u}, \mathbf{p}\big) + \int_0^1 \tilde{\Phi}^L\big(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau), \mathbf{u}, \mathbf{p}\big)\,\mathrm{d}\tau \tag{2.15a}$$

$$\text{s.t. } \dot{\tilde{\mathbf{x}}}(\tau) = \tilde{\mathbf{f}}\big(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(t), \mathbf{u}, \mathbf{p}\big), \qquad \tau \in [0,1], \tag{2.15b}$$

$$\mathbf{0} \le \tilde{\mathbf{c}}\big(\tilde{\mathbf{x}}(\tau), \tilde{\mathbf{u}}(\tau), \mathbf{p}\big), \qquad \tau \in [0,1], \tag{2.15c}$$

$$\mathbf{0} \le \tilde{\mathbf{r}}\big(\tilde{\mathbf{x}}(0), \tilde{\mathbf{x}}(1), \mathbf{u}, \mathbf{p}\big). \tag{2.15d}$$

Finally, we use the technique described in paragraph "Parameter-Dependence of Model Functions" (see Page 23) to transfer all parameter dependencies to the boundary constraints which yields a problem of Form (2.10), as desired.

2.4 Direct Solution Approaches to Optimal Control Problems

We present solution approaches to continuous OCPs, i. e., OCPs with continuously valued optimization variables. For solution approaches to MIOCPs we refer to [60, ch. 7], [81, ch. 2], and [127]. As stated above, OCPs are infinite dimensional optimization problems. In this section, we concentrate on so-called *direct approaches*. Here, an OCP is transcribed into an NLP in a first step, and the resulting finite dimensional optimization problem is solved subsequently. For this reason, direct approaches are also referred to as "first discretize, then optimize" approaches. In contrast, in *indirect*

approaches one establishes necessary conditions, also called maximum or minimum principles, for continuous OCPs. This results in boundary value problems which have to be solved numerically in a second step. Therefore, indirect approaches are also known as "first optimize, then discretize" approaches. A further prominent solution approach for OCPs is *Dynamic Programming*. For indirect approaches we refer to [60] and [83, part III], and for Dynamic Programming to [83, ch. 3].

In this thesis we make use of two direct approaches, namely Direct Multiple Shooting and Direct Collocation, both of which we introduce in the following. To this end, we consider a continuous single-stage OCP of the form

$$\min_{\mathbf{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot)} \Phi^M(\mathbf{x}(1)) \tag{2.16a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t)), \qquad t \in [0,1], \tag{2.16b}$$

$$\mathbf{0} \le \mathbf{c}(\mathbf{x}(t),\mathbf{u}(t)), \qquad t \in [0,1], \tag{2.16c}$$

$$\mathbf{0} \le \mathbf{r}(\mathbf{x}(0),\mathbf{x}(1),\mathbf{u},\mathbf{p}), \tag{2.16d}$$

with controllable parameters $\mathbf{u} \in \mathbb{R}^{n_\mathbf{u}}$, control function $\mathbf{u}(\cdot) \in L^\infty([0,1],\mathbb{R}^{n_u})$, and differential states $\mathbf{x}(\cdot) \in W^{1,\infty}([0,1],\mathbb{R}^{n_x})$. We assume all model functions to be continuously differentiable.

2.4.1 Direct Multiple Shooting

The Direct Multiple Shooting Approach for solving OCPs was introduced in [25, 120] for OCPs constrained by ODEs. The software package `MUSCOD-II` [95] provides an implementation. In the following, we describe a so-called Multiple Shooting discretization for an OCP of Form (2.16) which transcribes the problem into an NLP. The resulting NLP is then solved, e. g., with a structure exploiting SQP method. For the discretization of multi-stage problems we refer to [93].

As a first step, we introduce a shooting grid of size $N+1$:

$$0 = \tau_0 < \tau_1 < \cdots < \tau_N = 1$$

Control Function Discretization

For each shooting interval $[\tau_i,\tau_{i+1}]$, $i = 0,\ldots,N-1$, the control function $\mathbf{u}(\cdot)$ is approximated by elements $\mathbf{U}^i(\cdot)$ of finite dimensional subspaces of $L^\infty([\tau_i,\tau_{i+1}],\mathbb{R}^{n_u})$. Popular choices for the interval-wise discretization of the components $\mathbf{u}_j(\cdot)$,

$j = 1, \ldots, n_u$, of $\mathbf{u}(\cdot)$ are piecewise constant functions,

$$\mathbf{U}_j^i(t) = q^{i,j} \quad \text{for } t \in [\tau_i, \tau_{i+1}]$$

with $q^{i,j} \in \mathbb{R}$, and piecewise linear functions,

$$\mathbf{U}_j^i(t) = \frac{t-\tau_{i+1}}{\tau_i - \tau_{i+1}} \mathbf{q}_1^{i,j} + \frac{t-\tau_i}{\tau_{i+1}-\tau_i} \mathbf{q}_2^{i,j} \quad \text{for } t \in [\tau_i, \tau_{i+1}]$$

with $\mathbf{q}^{i,j} \in \mathbb{R}^2$. In any case, for each shooting interval $[\tau_i, \tau_{i+1}]$ and each component $\mathbf{U}_j^i(\cdot)$, $j = 1, \ldots, n_u$, of $\mathbf{U}^i(\cdot)$ there are $\mathbf{q}^{i,j} \in \mathbb{R}^{n_{i,j}}$ and functions $\boldsymbol{\xi}_j^i : [\tau_i, \tau_{i+1}] \times \mathbb{R}^{n_{i,j}} \to \mathbb{R}$ such that

$$\mathbf{U}_j^i(t) = \boldsymbol{\xi}_j^i\left(t, \mathbf{q}^{i,j}\right) \quad \text{for } t \in [\tau_i, \tau_{i+1}].$$

Setting

$$\mathbf{q}^i = \begin{pmatrix} \mathbf{q}^{i,1} \\ \vdots \\ \mathbf{q}^{i,n_u} \end{pmatrix} \in \mathbb{R}^{\sum_j n_{i,j}} \quad \text{and} \quad \mathbf{q} = \begin{pmatrix} \mathbf{q}^0 \\ \vdots \\ \mathbf{q}^{N-1} \end{pmatrix} \in \mathbb{R}^{\sum_{i,j} n_{i,j}},$$

we obtain an approximation of $\mathbf{u}(\cdot)$ by the parameterized function

$$\mathbf{U}(t, \mathbf{q}) = \begin{cases} \mathbf{U}^i\left(t, \mathbf{q}^i\right) & \text{if } t \in [\tau_i, \tau_{i+1}) \text{ for } i = 0, \ldots, N-2, \\ \mathbf{U}^{N-1}\left(t, \mathbf{q}^{N-1}\right) & \text{if } t \in [\tau_{N-1}, \tau_N]. \end{cases}$$

If required, one can enforce continuity of a component $\mathbf{U}_j(\cdot)$ by imposing additional constraints of the form

$$\mathbf{U}_j^i\left(\tau_{i+1}, \mathbf{q}^{i,j}\right) = \mathbf{U}_j^{i+1}\left(\tau_{i+1}, \mathbf{q}^{i+1,j}\right), \quad i = 0, \ldots, N-2.$$

State Parametrization

For the differential states $\mathbf{x}(\cdot)$, we introduce variables $\mathbf{s}^i \in \mathbb{R}^{n_x}$, $i = 0, \ldots, N$, which represent the values of $\mathbf{x}(\cdot)$ at the shooting grid points and set

$$\mathbf{s} = \begin{pmatrix} \mathbf{s}^0 \\ \vdots \\ \mathbf{s}^N \end{pmatrix}.$$

For each shooting interval we consider an IVP

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}\Big(\mathbf{x}(t), \mathbf{U}^i\big(t, \mathbf{q}^i\big)\Big), \quad t \in [\tau_i, \tau_{i+1}], \\ \mathbf{x}(\tau_i) &= \mathbf{s}^i, \end{aligned} \tag{2.17}$$

and assume that for given initial values $\mathbf{s}^i$ and parameters $\mathbf{q}^i$ the solution of the IVP – which we denote by $\mathbf{x}\big(t; \mathbf{s}^i, \mathbf{q}^i\big)$ – exists, is unique, and continuously differentiable with respect to $\mathbf{s}^i$ and $\mathbf{q}^i$ for all $t \in [\tau_i, \tau_{i+1}]$, respectively. In practice, we solve the IVPs by numerical integration. In order to get a continuous trajectory we impose additional constraints, the so-called matching conditions

$$\mathbf{0} = \mathbf{x}\big(\tau_{i+1}; \mathbf{s}^i, \mathbf{q}^i\big) - \mathbf{s}^{i+1} \quad \text{for } i = 0, \dots, N-1\,.$$

In particular, we obtain $\mathbf{x}\big(1; \mathbf{s}^{N-1}, \mathbf{q}^{N-1}\big) = \mathbf{s}^N$.

Discretization of Objective Function, Path Constraints and Boundary Constraints

The objective function for the discretized problem is given by

$$\Phi^M\big(\mathbf{s}^N\big)$$

and the Boundary Constraints (2.16d) are transformed into

$$\mathbf{0} \leq \mathbf{r}\big(\mathbf{s}^0, \mathbf{s}^N, \mathbf{u}, \mathbf{p}\big)\,.$$

The Path Constraints (2.16c) are enforced to hold at the shooting grid points which yields

$$\begin{aligned} \mathbf{0} &\leq \mathbf{c}\Big(\mathbf{s}^i, \mathbf{U}^i\big(\tau_i, \mathbf{q}^i\big)\Big), \quad i = 0, \dots, N-1\,, \\ \mathbf{0} &\leq \mathbf{c}\big(\mathbf{s}^N, \mathbf{U}^{N-1}\big(\tau_N, \mathbf{q}^{N-1}\big)\big)\,. \end{aligned} \tag{2.18}$$

Though this is a relaxation of (2.16c), in many real world application it is sufficient to demand (2.18). However, if critical violations of the Constraints (2.16c) are observed, a straightforward approach is to adapt or refine the shooting grid. A more sophisticated method to overcome the issue can be found in [122, 123].

The Resulting Nonlinear Programming Problem

The NLP resulting from the control discretization and the state parametrization is given by

$$\min_{\mathbf{u},\mathbf{s},\mathbf{q}} \Phi^M\left(\mathbf{s}^N\right) \tag{2.19a}$$

$$\text{s.t.} \quad \mathbf{0} = \mathbf{x}\left(\tau_{i+1}; \mathbf{s}^i, \mathbf{q}^i\right) - \mathbf{s}^{i+1}, \qquad i = 0, \ldots, N-1, \tag{2.19b}$$

$$\mathbf{0} \leq \mathbf{c}\left(\mathbf{s}^i, \mathbf{U}^i\left(\tau_i, \mathbf{q}^i\right)\right), \qquad i = 0, \ldots, N-1, \tag{2.19c}$$

$$\mathbf{0} \leq \mathbf{c}\left(\mathbf{s}^N, \mathbf{U}^{N-1}\left(\tau_N, \mathbf{q}^{N-1}\right)\right), \tag{2.19d}$$

$$\mathbf{0} \leq \mathbf{r}\left(\mathbf{s}^0, \mathbf{s}^N, \mathbf{u}, \mathbf{p}\right). \tag{2.19e}$$

This is an NLP of Form (2.2) which can be solved with tailored solution methods that exploit the specific problem structure which results from the Multiple Shooting discretization, cf., e. g., [25, 94, 93]. Since the control function is discretized while the differential states are determined by solving the resulting dynamics equation, we speak of a reduced discretization approach.

2.4.2 Direct Collocation

In this section, we present the Direct Collocation Approach for solving OCPs [13, 20, 70]. In contrast to the Multiple Shooting approach, Direct Collocation is a full discretization approach in which both the control function $\mathbf{u}(\cdot)$ and the differential states $\mathbf{x}(\cdot)$ are discretized using polynomial approximations. The discretization yields an NLP which is solved in a second step. The software package `grc` [102] provides an implementation of such an approach. In the following, we describe a general collocation discretization scheme for an OCP of Form (2.16).

State and Control Discretization

To increase readability, we focus on global discretization schemes in which each component of the differential states and the control function is approximated by one polynomial on the entire time horizon. Furthermore, we assume the polynomial degree for the differential states and the controls to be constant throughout all components, respectively. More precisely, we approximate $\mathbf{x}(\cdot)$ by a function

$$\mathbf{X}(t) = \sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(t)$$

with parameters $\mathbf{a}^k \in \mathbb{R}^{n_x}$ and basis polynomials $\phi_k : [0,1] \to \mathbb{R}$, $k = 0,\dots,N_X$, in such a way that the components of $\mathbf{X}(\cdot)$ are polynomials of degree N_X. Similarly, $\mathbf{u}(\cdot)$ is approximated by

$$\mathbf{U}(t) = \sum_{k=0}^{N_U} \mathbf{b}^k \psi_k(t)$$

with parameters $\mathbf{b}^k \in \mathbb{R}^{n_u}$ and basis polynomials $\psi_k : [0,1] \to \mathbb{R}$, $k = 0,\dots,N_U$, such that the components $\mathbf{U}_j(t)$ are polynomials of degree N_U. In contrast to global discretization schemes, in local schemes we set up a grid $0 = \tau_0 < \tau_1 < \cdots < \tau_N = 1$ and approximate the components of the states and controls in the intervals $[\tau_i, \tau_{i+1}]$ by polynomials with possibly varying degree per interval.

Discretization of Objective Function and Constraints

We choose a set $\{t_l^c\}_{l=1}^{N_X} \subset [0,1]$ of cardinality N_X of so-called *collocation points* and demand the Differential Equation (2.16b) to hold at these points for the discretized states and controls,

$$\dot{\mathbf{X}}(t_l^c) = \mathbf{f}\big(\mathbf{X}(t_l^c), \mathbf{U}(t_l^c)\big) \quad \Longleftrightarrow \quad \sum_{k=0}^{N_X} \mathbf{a}^k \dot{\phi}_k(t_l^c) = \mathbf{f}\Big(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(t_l^c), \sum_{k=0}^{N_U} \mathbf{b}^k \psi_k(t_l^c)\Big)$$

for $l = 1,\dots,N_X$, such that the coefficients $\mathbf{a}^k$ are determined by given $\mathbf{b}^k$ and the value of $\mathbf{X}(\cdot)$ at the initial time $t = 0$.

For the Path Constraints (2.16c) we consider a set $\{t_l^e\}_{l=1}^{N_e} \subset [0,1]$ of evaluation points and demand

$$\mathbf{0} \le \mathbf{c}\big(\mathbf{X}(t_l^e), \mathbf{U}(t_l^e)\big) = \mathbf{c}\Big(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(t_l^e), \sum_{k=0}^{N_U} \mathbf{b}^k \psi_k(t_l^e)\Big), \quad l = 1,\dots,N_e\,.$$

The Boundary Constraints (2.16d) are transformed into

$$\mathbf{0} \le \mathbf{r}\big(\mathbf{X}(0), \mathbf{X}(1), \mathbf{u}, \mathbf{p}\big) = \mathbf{r}\Big(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(0), \sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(1), \mathbf{u}, \mathbf{p}\Big),$$

and the objective function of the discretized problem is given by

$$\Phi^M\big(\mathbf{X}(1)\big) = \Phi^M\Big(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(1)\Big)\,.$$

The Resulting Nonlinear Programming Problem

We set

$$\mathbf{a} = \begin{pmatrix} \mathbf{a}^0 \\ \vdots \\ \mathbf{a}^{N_X} \end{pmatrix} \in \mathbb{R}^{(N_X+1)n_x} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}^0 \\ \vdots \\ \mathbf{b}^{N_U} \end{pmatrix} \in \mathbb{R}^{(N_U+1)n_u}.$$

With the previously described discretization scheme, the original OCP transforms into

$$\min_{\mathbf{u},\mathbf{a},\mathbf{b}} \ \Phi^M\left(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(1)\right) \tag{2.20a}$$

$$\text{s.t. } \mathbf{0} = \textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \dot{\phi}_k(t_l^c) - \mathbf{f}\left(\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(t_l^c), \sum_{k=0}^{N_U} \mathbf{b}^k \psi_k(t_l^c)\right), \quad l = 1,\dots,N_x, \tag{2.20b}$$

$$\mathbf{0} \le \mathbf{c}\left(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(t_l^e), \sum_{k=0}^{N_U} \mathbf{b}^k \psi_k(t_l^e)\right), \qquad l = 1,\dots,N_e, \tag{2.20c}$$

$$\mathbf{0} \le \mathbf{r}\left(\textstyle\sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(0), \sum_{k=0}^{N_X} \mathbf{a}^k \phi_k(1), \mathbf{u}, \mathbf{p}\right). \tag{2.20d}$$

This is an NLP of Form (2.2) which can be solved with tailored solution methods, depending on the specific structure of the problem.

2.5 Derivative Generation

In the course of this thesis, we apply direct methods to OCPs and solve the arising discretized problems with SQP or Interior-Point methods, see Section 2.2.2. To this end, we need to efficiently provide (directional) derivatives of all functions which occur in the discretized problems. In the following, we state different strategies for the computation of partial derivatives of first order. This includes so-called sensitivities which arise in Direct Multiple Shooting. These considerations can be transferred to directional derivatives.

2.5.1 Three Approaches for Calculating Derivatives

For the numerical computation of derivatives of functions $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ we consider three approaches which are of interest in this thesis. We follow the introductions in [3, ch. 2] and [110, ch. 8] where the interested reader can find details and further references, respectively.

Symbolic Differentiation

Symbolic Differentiation, also called analytical differentiation, can be used if an explicit formula of the considered function is available. An explicit expression for the

derivatives is derived. This process can be done by hand which is, however, cumbersome and prone to error for complex functions. An alternative is the use of software tools such as MATLAB® [100]. As it provides explicit formulas for derivatives, symbolic differentiation enables us to compute derivative values exact up to machine precision.

Finite Differences

The finite differences approach allows to treat the function $\mathbf{f}(\cdot)$ as a black box. The partial derivatives of the components of $\mathbf{f}(\cdot)$ are approximated, e. g., by one-sided difference quotients

$$\frac{\mathbf{f}_j(\mathbf{x}+\varepsilon\mathbf{e}_i)-\mathbf{f}_j(\mathbf{x})}{\varepsilon} \approx \frac{\partial \mathbf{f}_j}{\partial \mathbf{x}_i}(\mathbf{x}) \tag{2.21}$$

with $\varepsilon > 0$ and $\mathbf{e}_i$ being the i-th unit vector. Another variant is to use central difference approximations of the form

$$\frac{\mathbf{f}_j(\mathbf{x}+\varepsilon\mathbf{e}_i)-\mathbf{f}_j(\mathbf{x}-\varepsilon\mathbf{e}_i)}{2\varepsilon} \approx \frac{\partial \mathbf{f}_j}{\partial \mathbf{x}_i}(\mathbf{x}) \tag{2.22}$$

which leads to a higher accuracy of the gradient approximation at the cost of more function evaluations, cf. [110, sec. 8.1]. Independently from the choice of the difference quotient, in theory the approximation becomes arbitrary good if ε is chosen small enough. However, in practice this is not true due to cancellation errors which become relevant when ε gets too small. Therefore, when computing derivatives with finite differences one has to accept a significant loss of accuracy. Even for an optimal choice of ε (which depends on $\mathbf{f}_j(\cdot)$) one has to expect a loss of about one half of the significant digits of the function evaluation for the one-sided Approximation Scheme (2.21) and of about one third in case of the Central Scheme (2.22). For more details we refer to [3, sec. 2.2] and [110, sec. 8.1]. The above considerations transfer to directional derivatives, cf. [3, sec. 2.3].

Automatic Differentiation

The basic idea of Automatic Differentiation (AD) is to view a function as a concatenation of elementary functions with known derivative and to make excessive use of the chain rule in order to compute (directional) derivatives. In contrast to symbolic differentiation, "[...] the chain rule is not applied to manipulate symbolic expressions but works on numerical values" [3, p. 26]. Given a piece of computer code

for the function evaluation, AD software tools (such as `Adol-C` [148]) break down the function evaluation into elementary operations and generate a computational graph which is used for the derivative generation. This way one can efficiently compute derivative values exact up to machine precision. As entry points to the topic we refer to [3, sec. 2.3] and [110, sec. 8.2], respectively, where the latter reference also comments on limitations of the usage of AD tools.

2.5.2 Sensitivity Generation

In this section we follow [3], [69, sec. II.6], and [81, sec. 3.4], where further information can be found, respectively. We again consider the Multiple Shooting Discretization. To solve the resulting NLP of Form (2.19) with gradient-based methods (e. g., Interior-Point or SQP methods), we have to deal with so-called *sensitivities*

$$\frac{\partial}{\partial \mathbf{s}^i}\mathbf{x}\left(\tau_{i+1};\mathbf{s}^i,\mathbf{q}^i\right) \quad \text{and} \quad \frac{\partial}{\partial \mathbf{q}^i}\mathbf{x}\left(\tau_{i+1};\mathbf{s}^i,\mathbf{q}^i\right)$$

where $\mathbf{x}\left(\tau_{i+1};\mathbf{s}^i,\mathbf{q}^i\right)$ denotes the solution of the IVP (2.17) at $t=\tau_{i+1}$. In practice, the expression $\mathbf{x}\left(\tau_{i+1};\mathbf{s}^i,\mathbf{q}^i\right)$ is evaluated by numerical integration. A straightforward approach for the computation of the sensitivities is the so-called external numerical differentiation. Here, the numerical integrator is treated a black box function which maps $(\mathbf{s}^i,\mathbf{q}^i)$ to an approximation of $\mathbf{x}\left(\tau_{i+1};\mathbf{s}^i,\mathbf{q}^i\right)$, and the sensitivities are approximated by finite differences. This approach is easy to implement. However, due to the adaptivity of the integrator, the integrator output cannot be assumed to depend smoothly on $\mathbf{s}^i$ and $\mathbf{q}^i$. This impairs the accuracy of the sensitivity approximation and raises the need for very tight integration tolerances, cf. [69, sec. II.6], which renders the approach unfavorable in practice.

An approach to overcome this issue is Internal Numerical Differentiation (IND) which is described in [22, 23]. The idea is to fix the adaptive elements of the integrator after computation of the nominal (i. e., unperturbed) trajectory and to differentiate the generated discretization scheme which can be seen as a concatenation of differentiable functions that map the values of the solution trajectory from one grid point to another. Thus, for instance one can apply AD to achieve the exact derivatives (up to machine precision) of the approximation of the nominal trajectory obtained by the integrator. Compared to external numerical differentiation, IND significantly reduces the accuracy requirements on the numerical integration which has a substantial impact on the computational effort.

Chapter 3

Cerebral Palsy

In this chapter, we give an overview of Cerebral Palsy (CP), in particular of causes, symptoms, diagnosis and treatments. The chapter is based on [2, 7, 43, 67, 71, 90] and the references therein, where the interested reader can find further information and details.

CP is an umbrella term for multiple disabilities affecting a patient's nervous system, musculature, and skeletal system [43, p. 40]. Patients exhibit "[...] complex and heterogeneous motor disorders [...]" [7, p. 448] that impair the ability to walk and, in case of ambulatory patients, cause deviations of the gait patterns, see, e. g., [7, 130]. To be more precise,

> "Cerebral palsy (CP) describes a group of permanent disorders of the development of movement and posture, causing activity limitation, that are attributed to non-progressive disturbances that occurred in the developing fetal or infant brain. The motor disorders of cerebral palsy are often accompanied by disturbances of sensation, perception, cognition, communication, and behaviour, by epilepsy, and by secondary musculoskeletal problems." [126, p. 9]

In CP, the brain damage does not worsen. In contrast, the impairments resulting from the brain lesion are progressive. For instance, the brain damage impairs the muscle functionality and can lead to spasticity, abnormal muscle tones, muscle weakness, and muscle imbalance. This results in pathological forces, permanently acting on a patients bones and joints, which over the years cause a deformity of the skeleton [43, sec. 4.4.1]. As a consequence of these deteriorations patients may lose their ability to walk.

CP is the most frequent cause of motor disorders in childhood, see, e. g., [43, p. 44]. In the industrialized countries the prevalence is about 2 – 3 per 1000 live births, see [43, p. 44] and the references therein.

3.1 Causes

CP is caused by a damage to the premature brain which subsequently impairs the development of the brain. As the development of the brain is not finished at the time of birth, the damage can be caused prenatally (before birth), perinatally (during birth), or postnatally (after birth) [90, p. 91]. The majority of CP cases – 70 – 80% according to [90, p. 91] – are caused prenatally. There are numerous causes, e. g., brain bleeding, infections (e. g., rubella, toxoplasmosis) of the mother, alcohol or nicotine consumption of the mother, or congenital malformation in the brain, just to mention a few. Perinatal causes are, e. g., delayed or complicated delivery, oxygen deficiency, or brain injury due to a mechanical trauma. Postnatal causes include infections (e. g., meningitis), mechanical trauma (e. g., falls or child abuse), near drowning, stroke, or metabolic-toxic impairments. For the mentioned causes see [90, p. 91], [43, sec. 4.2.1], and [67, p. 1007]. We also refer to the latter references for more extensive lists of causes.

One crucial risk factor is a child's birth weight. While the prevalence for CP is about 2 – 3 per 1000 live births in the industrialized countries as stated before, the prevalence rises significantly with decreasing birth weight and prematurity, see, e. g., [88, p. 39] and [114].

3.2 Classification

We follow [43, sec. 4.3] and [2, p. 78]. The disorder CP can be classified using different systems, depending on

- The location of the damage in the brain (e. g., cerebrum or brainstem).
- The kind of motor disorder:
  - Spastic CP (80% according to [2, p. 78]). In this most prevalent form patients exhibit spastic syndroms, e. g., increased muscle tone, cf. [43, sec. 4.4.2]. Movements seem effortful and slow, but are voluntary [67, p. 1007].
  - Dyskinetic or athetoid CP. Patients suffering from this form show unintended movements. They can exhibit involuntary contractions of certain muscles, e. g., in the face, simultaneous contractions of muscles and their antagonists [67, p. 1009], and sometimes "[...] slow, writhing movements [...]" [90, p. 92] of the limbs.

  - Ataxic CP (5 – 10%, cf. [90, p. 92]). This form affects coordinated movement [2, p. 78].
  - Not classifiable.

  However, there are cases in which the kind of motor disorder changed during the development of a child [43, p. 47].

- The affected parts of the body, e. g.,
  - monoplegia: one involved extremity,
  - hemiplegia: unilateral involvement,
  - paraplegia: lower body involved,
  - quadriplegia: all parts of the body involved,

  see [2, p. 78].

- The functional severity of the motor disorder, using classification systems such as the Gross Motor Function Classification System (GMFCS) [117], which categorizes patients based on the ability to perform self-initiated movements like walking or sitting. For a description of the GMFCS, we refer to [117, App. B] resp. [90, p. 96–97] and [43, p. 48-49].

3.3 Symptoms, Comorbidities, and Gait Patterns

CP shows a very heterogeneous clinical picture as the various ways of classification emphasize. The causal brain damage can lead to numerous consequential damages. It affects the musculature, leading to spasticity, abnormal muscle tone, muscle weakness, and muscle imbalance [43, sec. 4.4.2], resulting in motion disorders which in particular affect the ability to walk. Besides the impact on the musculoskeletal system, CP comes along with a bunch of comorbidities including impaired vision, impaired hearing, impairment of intelligence, epilepsy, oral-motor dysfunction (consequences: dystrophy and drooling), gastrointestinal problems, impairment of body awareness and pain sensation, and impairment of communication, cf., e. g., [43, ch. 4.4.4], [90, p. 94], [67, p. 1010] and [111], where the latter two references state occurring frequencies. In addition, patients often suffer from psychosocial problems due to a lack of social acceptance. A more extensive list of conditions associated with CP and according frequencies of occurrence can be found in the referenced literature.

Among all facets of the disorder, in this thesis we are interested in pathological gaits of ambulatory patients which are able to walk freely without the help of any assistance. In particular, we are interested in gait deviations which are (at least partially) resulting from musculoskeletal impairments, and thus can be improved by therapies concerning the musculoskeletal system like physiotherapy and orthopedic surgery.

Döderlein [43, sec. 5.2] states common gait deviations in spastic CP (which is by far the most frequent form of CP, see above). Following this reference, among these are

- equinus gait (unilateral or bilateral), a gait pattern in which patients walk on their toes on one or both feet,
- crouch gait, a gait disorder with bilateral knee flexion which can be caused by weakness, spastic deformity and/or contracture, but also by previously performed interventions,
- genu recurvatum gait, a gait pattern with a hyper-extended knee in stance phase,
- gait with pathologically (internally or externally) rotated legs which can emerge unilateral or bilateral, symmetric or asymmetric, and on one or multiple levels (pelvis, hip- and knee joints, lower leg, hindfoot, forefoot).

In practice, it can be difficult to determine the malpositions and in particular their magnitude with the naked eye. Toe walking, for instance, can be the consequence of a so-called equinus, but it can also result from other malpositions. In this case, this leads to the distinction between true and apparent equinus which is crucial when it comes to a medical treatment.

3.4 Gait Analysis

One important tool to analyze and quantify a patient's gait in more detail is Gait Analysis (GA). It is used in CP diagnosis and intervention planning, in particular for treatments which aim at improving a patient's gait. We follow [43, sec. 5.3] and [7] to give a short introduction.

During a GA, different examinations and measurements, static as well as dynamic, are performed. On the one hand, there are clinical examinations evaluating a patient's ranges of motion, muscular strength, degree of spasticity, coordination abilities as well as occurring foot deformities. In addition, further information including

Figure 3.1: A Cerebral Palsy patient who is equipped with motion capture markers for 3D Gait Analysis. Picture provided by the Heidelberg MotionLab [151].

body measurements like leg lengths, body height, and weight are recorded, and X-ray images of relevant body parts might be taken. The gait itself is assessed visually.

As walking is a dynamic process, the recording of dynamic data during the gait is of vital importance. In 3D GA, motion capture systems and force-plates are used to collect spatiotemporal data as well as ground-reaction forces. A picture of a patient who is equipped with markers for motion capture can be seen in Fig. 3.1. Using biomechanical models it is possible to extract kinematic and kinetic data, such as joint angles or torques acting at the joints, in all dimensions as an evolution of time. Furthermore, electromyographic data of relevant muscles can be recorded. In addition to these measurements, the patient's gait is recorded on video from different perspectives.

3D GA gives a deep insight into the gait of a patient. It helps physicians to correctly identify pathologies which are difficult to determine with the naked eye. In [130], the authors connect occurring kinematic deviations (i. e., the deviation of kinematic curves from the norm-curve) to probable causing impairments. For instance, an increased knee flexion during a certain phase of the the gait can be an indication for an overactivity of the hamstring. Armand et al. [7] caution however that the presented findings are "[...] based more on experience than evidence" [7, p. 453].

Despite the time and effort a complete GA takes, it is well-established in the clinical routine of CP management. For example, at the gait laboratory of the Heidelberg University Hospital (see Section 1.3), GA is performed in daily routine. The GAs are one of the major components in clinical decision making, contributing largely to the question of the best treatment. If a surgical treatment is performed, the result is again observed and assessed by a subsequent GA performed after the musculoskeletal system adapted to the medical changes.

3.5 Medical Treatments

Although it is not possible to remediate the brain damage there are multiple treatments aiming at improving a patient's situation. Gulati and Sondhi state very generally that for children "The management should be directed at stimulating the child's development with the aim to obtain maximal independence in activities of daily living" [67, p. 1010]. More precisely, following [43, sec. 7.5] the main goals for treatment of spastic CP are the lessening or suppression of spasm and the invigoration of paretic muscles. Further goals (for all kind of CP) are the improvement or recovery of pathological skeletal axis and limited joint flexibility, improvement of motor control, and pain relief as precondition for further assistance [43, p. 159]. As the impairments associated with CP often worsen progressively, patients frequently undergo therapy from an early age on in order to hinder the deterioration.

Available therapies can be divided into conservative (non-invasive) therapies and surgical therapies. Examples for conservative therapies are physiotherapy, ergotherapy, use of orthoses, and the administering of drugs (e. g., muscle relaxants or painkillers) [43, sec. 8.2]. Orthopedic surgical treatments among others aim at the correction of deformities, stabilization of joints, amendment of muscle imbalances, and the preservation or recovery of functionally important ranges of motion [43, p. 227]. Apart from these orthopedic surgeries, in selective dorsal rhizotomy – simply put – the spasticity is reduced or eliminated by intersecting appropriate nerve fibers at the lower part of the spinal cord [90, p. 98], see [43, sec. 15.5.3] for further information.

In this thesis, we are interested in therapies for ambulatory patients which aim at improving their gaits. To give an impression of the severity of the surgical interventions, we follow [7, p. 458] and the references therein and state possible treatment options for gait disorders frequently occurring in CP:

- *True equinus* (leads to toe walking): Treatment options include a lengthening of the gastrocnemius muscle, which is the larger part of the calf muscle, see, e. g., [9, 118], and a lengthening of the Achilles tendon.

- *Crouch gait*: Treatment options include a surgical correction of foot deformities, torsional tibial deformities, and the application of patella (kneecap) advancements. A further option is a hamstring lengthening.

- *Stiff knee gait*: For patients with GMFCS level I or II (see, e. g., [43, p. 48]), a so-called rectus femoris transfer is an option. In this treatment, the distal tendon of the rectus femoris muscle (which crosses the knee cap) is relieved and attached to another (resp. one part of a previously intersected) tendon which contributes to knee flexion (cf. [46, 149]).

- *Torsional deformities of tibia and femur* (shinbone and thighbone): Can be corrected by so-called derotational osteotomies. The affected bone is cut through and rotated. Then, both parts are again fixed to each other in the desired position.

To avoid the so-called birthday syndrome, meaning the high frequent hospitalization of patients due to many sequentially performed operations one year after another, often several surgeries at different parts of the body are combined. This reduces the number of overall surgeries but makes assessment more difficult since many factors – potentially interacting with each other – play a role.

Treatment Evaluation and Outcome

For the assessment of treatment outcomes multiple evaluation mechanism exist. In particular, for the assessment and quantification of gait patterns there are so-called gait scores as, e. g., the Gait Profile Score [8] which measures the quality of a gait by comparing it to an average healthy gait using data from GA. This rewards "healthy-looking" gaits. However, it is not clear whether such a style of walking is indeed advantageous for each individual patient. An introduction to the evaluation of CP therapies can be found in [43, ch. 9].

Despite the accumulated experience and the use of modern methods, CP management and in particular treatment planning, i. e., the choice of the appropriate treatments for a specific patient, is still difficult and prone to error. Although employing

elaborate 3D GAs, a significant amount of interventions still yields a negative outcome, see [131, p. 3] and [30, 36]. Furthermore, according to [43, p. 114], the long-term consequences of invasive treatments are rarely known, cf., e. g., [44, 45]. Novak et al. systematically review treatments for CP and their usefulness, and come to the conclusion that out of the considered CP therapies – including both, surgical but also non-surgical ones – only "[...] 24% are proven to be effective" [112, p. 886], while "70% have uncertain effects and routine outcome measurement is necessary" [112, p. 886]. Though, as Döderlein [43, p. 247] adds, this work can be seen critical, in any case it emphasizes that there is still a great potential for improvement in CP treatment planning.

Chapter 4

Model-Based Treatment Planning

In this chapter, we introduce the approach to model-based treatment planning we pursue in this thesis. We present a mathematical model for the human gait in Section 4.1 where we model the human body as a rigid Multi-Body System (MBS) and the human gait as solution of an Optimal Control Problem (OCP) which is constrained by the dynamics of the MBS. In Section 4.1.2, we give an introduction to model-based treatment planning for Cerebral Palsy (CP) and present the general approach we follow in this thesis. Subsequently, in Section 4.3 we propose a way to model a class of medical treatments in CP which is suitable in the context of our Optimal Control model.

4.1 An Optimal Control Model for the Human Gait

In this section, we present a mathematical model for the human gait. We explain how to assemble a biomechanical model of the human body and subsequently set up an OCP constrained by the dynamics of the biomechanical model. The human gait is then modeled by a solution of the OCP.

4.1.1 Biomechanical Model

We are interested in human walking resp. the human gait. Here, walking refers to a bipedal locomotion where at any point in time at least one foot is in contact with the ground. The human gait is commonly assumed to be cyclic and thus can be seen as succession of identical so-called gait cycles. A gait cycle in turn describes the sequence of movements during two subsequent steps where initial posture and terminal posture (approximately) coincide. The gait cycle of a healthy person is illustrated in Fig. 4.1. Sometimes, only half of the full cycle is taken into account (see, e. g., [77]) since in a healthy gait pattern the gait is commonly assumed to be symmetric. Beware however, that for disturbed gait patterns occurring, e. g., in CP this is not true in general. The gait cycle can be divided into phases in different ways, see, e. g., [43, sec. 3.2]. In this thesis, the phases are characterized by the parts of the human body which are in contact with the ground.

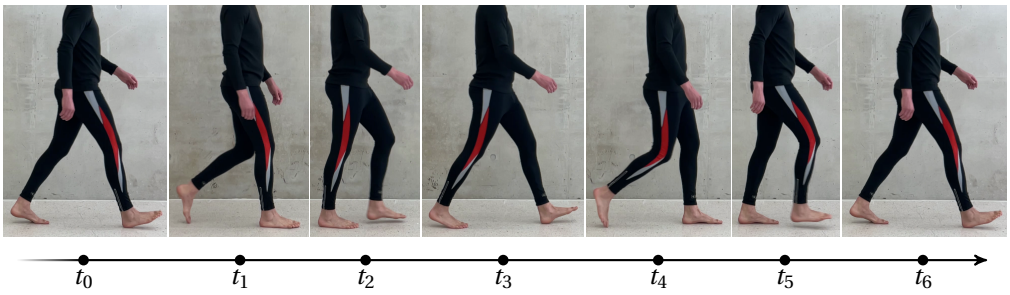
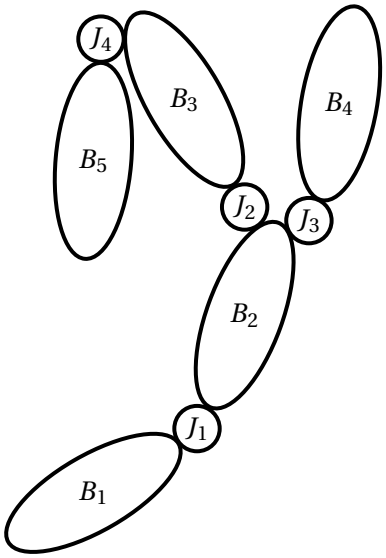


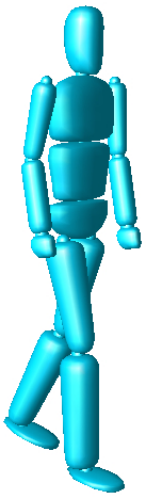
Figure 4.1: Illustration of the human gait cycle. We enter the gait cycle at time t_0 when the right foot strikes the ground. Until t_1, both feet are in contact with the ground. At t_1, the left foot lifts off the ground in order to swing forward, and strikes the ground again at t_3. During this so-called swing phase of the left foot, the right foot remains in contact with the ground. Subsequently, both feet reverse the roles. Finally, at t_6, the right foot strikes the ground again, and the gait cycle starts from the beginning.

We follow a common approach in biomechanics and model the human body as a rigid MBS actuated by torques (as in [48, 71]). In general, a rigid MBS consists of a set of rigid bodies which are interconnected by joints. Two exemplary MBSs are depicted in Fig. 4.2. A single rigid body can be described by several physical quantities, among these are its mass, position of center of mass in space, spatial orientation, inertia, and physical dimensions. Connecting several of these bodies with joints then establishes an MBS. Here, the term joint describes "[...] any possible kinematic relationship between a pair of rigid bodies" [47, p. 65]. In this thesis, we focus on rotational joints, i. e., revolute joints which allow for rotational movements about one axis and spherical joints for rotations in three dimensions, see Fig. 4.3 for an illustration. The mobility of an MBS is expressed in terms of so-called Degrees of Freedom (DoF) – the minimum number of independent parameters n_{dof} which is necessary to fully describe the *configuration* of a system, i. e., its position and orientation in space. Such a set of parameters is called *generalized coordinates* and denoted by $\mathbf{q} \in \mathbb{R}^{n_{\text{dof}}}$. The DoF of an MBS depend on the types of joints which connect the bodies to each other, and the physical properties of an MBS are determined by the properties of the sub-bodies as well as the configuration of the system. An illustrative example for an MBS with its corresponding generalized coordinates is given in Appendix A.1. Detailed introductions to rigid MBSs can be found in [47, 138, 150], respectively.

To model the human body as a rigid MBS, the body needs to be split up in segments in a suitable manner. We use the pelvis as a base segment which can move freely in



a) An exemplary rigid MBS comprising 5 bodies which are interconnected by 4 joints, similar to [76, Fig. 2.1].

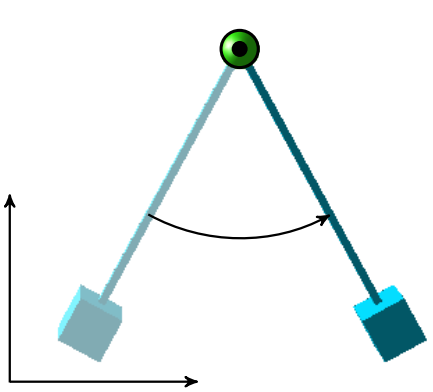


b) A rigid MBS roughly describing the topology of the human body. The displayed MBS is based on the `HeiMan` model [48, sec. 4.3]. Illustration created using `MeshUp` [48].

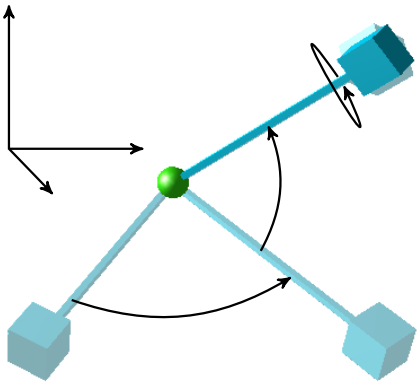
Figure 4.2: Two exemplary rigid MBSs.

space and connect adjacent segments successively. As connecting joints we employ rotational joints which is sufficient for our purposes. This way, we obtain a generic model for the human body which then needs to be calibrated person-specifically. For each limb of the considered person which is represented by a segment in our model we need to identify or guess its physical properties. While the physical dimensions could be measured directly, this is not possible for mass, center of mass, and inertia. Here, data from literature, e. g. [35], can be taken. However, in view of the topic of this thesis one should keep in mind that the values provided in the latter reference are not specific for CP. Furthermore, we refer to [48, ch. 4] resp. [51], where the authors use motion capture data – whose recording is part of a Gait Analysis (GA), see Section 3.4 – to generate subject specific models.

In order to reduce the computational cost and enable the numerical treatment of the mathematical problems we consider in this thesis, we refrain from using detailed



a) A revolute joint which allows for rotations about one axis. The axis of rotation runs through the green ball.



b) A spherical/ball joint – illustrated by the green ball – for rotations in three dimensions.

Figure 4.3: Two types of rotational joints. Illustration created using `MeshUp` [48].

models of the musculotendinous complexes in the following but summarize their effect in the resulting torque generated for each joint.

Dynamics of Rigid Multi-Body Systems

We are interested in the dynamics of the human gait. Since we model the human body as a rigid MBS with actuator torques, we give a concise overview on the dynamics of rigid MBSs under constraints in this section. More details can be found, e. g., in [47, 138, 150]. An illustrative example for the dynamics of a rigid MBS is given in Appendix A.1.

In the following we omit the argument for time-dependent variables. We consider a generic rigid MBS with n_{dof} DoF and generalized coordinates $\mathbf{q} \in \mathbb{R}^{n_{\text{dof}}}$. The time derivatives $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ are called generalized velocities and generalized accelerations, respectively. Physical parameters of the system (e. g., masses or dimensions of subbodies) are summarized in the vector $\bar{\mathbf{p}} \in \mathbb{R}^{n_{\bar{p}}}$. The MBS is governed by the rules of classical mechanics. The equations of motion can be expressed in the form

$$\mathbf{H}(\mathbf{q}, \bar{\mathbf{p}})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}, \bar{\mathbf{p}}) = \boldsymbol{\tau}\,, \tag{4.1}$$

with symmetric generalized inertia matrix $\mathbf{H}(\mathbf{q}, \bar{\mathbf{p}}) \in \mathbb{R}^{n_{\text{dof}} \times n_{\text{dof}}}$, generalized forces $\boldsymbol{\tau} \in \mathbb{R}^{n_{\text{dof}}}$, and generalized bias force $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}, \bar{\mathbf{p}}) \in \mathbb{R}^{n_{\text{dof}}}$, the latter comprising all other forces (e. g. Coriolis, centrifugal, gravitational) acting on the system besides $\boldsymbol{\tau}$. In the following we assume that all sub-bodies have non-zero mass. Then $\mathbf{H}(\mathbf{q}, \bar{\mathbf{p}})$ is positive definite and in particular regular. Hence, $\ddot{\mathbf{q}}$ can be computed directly from $\mathbf{q}$, $\dot{\mathbf{q}}$ and $\boldsymbol{\tau}$.

In the rigid MBSs we consider in later applications, position and orientation of the base segment (modeling the pelvis) are not actuated directly but experience actuation through interaction of the base segment with the adjacent bodies (similar to a trailer towed by a car). Thus, we have

$$\boldsymbol{\tau} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\tau}^a \end{pmatrix} \tag{4.2}$$

with actuated generalized forces $\boldsymbol{\tau}^a \in \mathbb{R}^{n_{\text{act}}}$ and $n_{\text{act}} < n_{\text{dof}}$. MBSs of this kind are called *underactuated*.

Moving forward while walking is possible due to the interaction of the feet with the ground. We model the ground contact by means of constraints which demand the positions of the contact points of a foot with the ground to be fixed for a certain period of time. Such a constraint can be expressed in terms of the generalized coordinates in the form

$$\mathbf{g}(\mathbf{q}, \bar{\mathbf{p}}) = \mathbf{0} \tag{4.3}$$

with $\mathbf{g}(\cdot) \in \mathbb{R}^{n_c}$, and is referred to as *external contact* in the following. The constraints induce constraint forces $\boldsymbol{\lambda} \in \mathbb{R}^{n_c}$ and the resulting equations of motion read as

$$\mathbf{H}(\mathbf{q}, \bar{\mathbf{p}})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}, \bar{\mathbf{p}}) = \boldsymbol{\tau} + \mathbf{G}(\mathbf{q}, \bar{\mathbf{p}})^T \boldsymbol{\lambda}, \tag{4.4a}$$

$$\mathbf{g}(\mathbf{q}, \bar{\mathbf{p}}) = \mathbf{0}, \tag{4.4b}$$

where $\mathbf{G}(\mathbf{q}, \bar{\mathbf{p}}) \overset{\text{def}}{=} \frac{\partial}{\partial \mathbf{q}} \mathbf{g}(\mathbf{q}, \bar{\mathbf{p}}) \in \mathbb{R}^{n_c \times n_{\text{dof}}}$ is the so-called *contact Jacobian*. The Equation System (4.4) is a Differential Algebraic Equation (DAE) of differential index 3. By differentiating (4.4b) twice with respect to time we get

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t} \mathbf{g}(\mathbf{q}, \bar{\mathbf{p}}) = \mathbf{G}(\mathbf{q}, \bar{\mathbf{p}})\dot{\mathbf{q}}, \tag{4.5a}$$

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}t} \left[\mathbf{G}(\mathbf{q}, \bar{\mathbf{p}})\dot{\mathbf{q}} \right] = \left[\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{G}(\mathbf{q}, \bar{\mathbf{p}}) \right] \dot{\mathbf{q}} + \mathbf{G}(\mathbf{q}, \bar{\mathbf{p}})\ddot{\mathbf{q}}, \tag{4.5b}$$

where $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{G}(\mathbf{q},\bar{\mathbf{p}})$ denotes the component-wise total differentiation of the matrix $\mathbf{G}(\mathbf{q},\bar{\mathbf{p}})$ with respect to time. We set

$$\boldsymbol{\gamma}(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}}) \stackrel{\text{def}}{=} -\left[\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{G}(\mathbf{q},\bar{\mathbf{p}})\right]\dot{\mathbf{q}} \in \mathbb{R}^{n_c}.$$

Then, aggregating Equations (4.4a) and (4.5b) yields the linear system

$$\begin{pmatrix}\mathbf{H}(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{G}(\mathbf{q},\bar{\mathbf{p}})^T\\ \mathbf{G}(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{0}\end{pmatrix}\begin{pmatrix}\ddot{\mathbf{q}}\\ -\boldsymbol{\lambda}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\tau}-\mathbf{C}(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}})\\ \boldsymbol{\gamma}(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}})\end{pmatrix}. \tag{4.6}$$

If the Constraints (4.3) are defined properly (in particular non-redundantly and such that $n_c \leq n_{\text{dof}}$), $\mathbf{G}(\mathbf{q},\bar{\mathbf{p}})$ has full rank. Hence, the matrix

$$\begin{pmatrix}\mathbf{H}(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{G}(\mathbf{q},\bar{\mathbf{p}})^T\\ \mathbf{G}(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{0}\end{pmatrix}$$

is non-singular and Equation (4.6) is uniquely solvable. The Differential Equation (4.6) ensures the vanishing of the term $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbf{g}(\mathbf{q},\bar{\mathbf{p}})$. Therefore, if (4.5a) and (4.3) hold in the beginning t' of our process, the solutions of the DAEs (4.4) and (4.6) coincide for all $t \geq t'$.

In general, instead of treating an MBS with external contacts it is also possible to incorporate the additional Contact Constraints (4.3) directly when choosing the generalized coordinates of the system. This leads to a system without external contacts but with reduced DoF. The equations of motion are then given in Form (4.1). Since the gait cycle comprises different phases, characterized by different contact constraints in this thesis, this results in phase-specific generalized coordinates with possibly altering dimensions. As a numerical method we use later (see Chapter 5) relies on constant dimensions of the generalized coordinates throughout all phases, this approach is not directly suitable for our purposes.

We model the gait phases by changing external contacts. Whenever a part of a body enters the ground contact, in reality high forces arise and deform the soft tissue of the according limb. As we choose to model the limbs by rigid and thus non-deformable bodies, we model this event by an instantaneously occurring perfect inelastic collision, resulting in an instantaneous jump of the generalized velocities. Let $\dot{\mathbf{q}}^-$ and $\dot{\mathbf{q}}^+$ be the generalized velocities instantly before or after the collision, respectively, and $\mathbf{G}^+(\mathbf{q},\bar{\mathbf{p}})$ the contact Jacobian belonging to the external contact $\mathbf{0} = \mathbf{g}^+(\mathbf{q},\bar{\mathbf{p}}) \in \mathbb{R}^{n_c^+}$ which holds *after* the collision. The transfer of velocities is then given by the equa-

tion

$$\begin{pmatrix} \mathbf{H}(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{G}^+(\mathbf{q},\bar{\mathbf{p}})^T \\ \mathbf{G}^+(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \dot{\mathbf{q}}^+ \\ -\boldsymbol{\Lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{H}(\mathbf{q},\bar{\mathbf{p}})\dot{\mathbf{q}}^- \\ \mathbf{0} \end{pmatrix}, \tag{4.7}$$

with *contact impulse* $\boldsymbol{\Lambda} \in \mathbb{R}^{n_c^+}$. Here, the matrix on the left-hand side is non-singular if $\mathbf{G}^+(\mathbf{q},\bar{\mathbf{p}})$ has full rank (which is the case for properly defined constraints). Furthermore, from (4.7) we have

$$\mathbf{0} = \mathbf{G}^+(\mathbf{q},\bar{\mathbf{p}})\dot{\mathbf{q}}^+ = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{g}^+(\mathbf{q},\bar{\mathbf{p}}).$$

Hence, (4.5a) is satisfied immediately after collision.

The equations of motion of an MBS can be established by different formalisms and algorithms, e. g., using Lagrangian or Hamiltonian mechanics. Setting up the equations of motion by hand is tedious and error-prone already for comparatively small systems which raises the need for computational support. In this thesis, we use the software library `RBDL` [49] – a "[...] framework with broad dissemination inside the robotics community" [29, p. 614] – which is based on the notation and algorithms presented in [47]. `RBDL` has proven its efficacy in many applications, in particular in the context of Optimal Control (see e. g., [6, 77, 104]), and is well-suited for our purposes.

4.1.2 The Human Gait as Solution of an Optimal Control Problem

In this section, we present a general Optimal Control model for the human gait. An illustrative example for the gait generation approach presented in this section is given in Appendix A.2.

Phase-Wise Dynamics with Jumps

As explained in the previous section, we model the human body while walking as rigid MBS with changing external contacts. Here, each contact configuration reflects a phase of the gait cycle. In each phase the dynamics of the MBS can be expressed as a DAE system. Let $\mathbf{q} = \mathbf{q}(t)$ denote the (phase-independent) generalized coordinates of the system, $\mathbf{g}^j(\mathbf{q},\bar{\mathbf{p}}) = \mathbf{0}$ the (properly defined) external contact constraint characterizing phase j with contact Jacobian $\mathbf{G}^j(\mathbf{q},\bar{\mathbf{p}})$, and $\boldsymbol{\lambda}^j = \boldsymbol{\lambda}^j(t)$ the corresponding

constraint forces. We define the differential states $\mathbf{x}(\cdot)$ by

$$\mathbf{x}_c(t) \overset{\text{def}}{=} \mathbf{q}(t), \quad \mathbf{x}_v(t) \overset{\text{def}}{=} \dot{\mathbf{q}}(t), \quad \mathbf{x}(t) \overset{\text{def}}{=} \begin{pmatrix} \mathbf{x}_c(t) \\ \mathbf{x}_v(t) \end{pmatrix}, \tag{4.8}$$

and phase-dependent algebraic states $\mathbf{z}^j(t) \overset{\text{def}}{=} \boldsymbol{\lambda}^j(t)$. Let phase j take place in the interval $\mathcal{T}_j = [T_{j-1}, T_j]$. According to the previous section, the dynamics during phase j can be described as

$$\begin{bmatrix} \dot{\mathbf{x}}_c(t) \\ \mathbf{M}_j(\mathbf{x}_c(t), \bar{\mathbf{p}}) \begin{pmatrix} \dot{\mathbf{x}}_v(t) \\ -\mathbf{z}^j(t) \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_v(t) \\ \boldsymbol{\tau}(t) - \mathbf{C}(\mathbf{x}(t), \bar{\mathbf{p}}) \\ \boldsymbol{\gamma}^j(\mathbf{x}(t), \bar{\mathbf{p}}) \end{bmatrix}, \qquad t \in \mathcal{T}_j, \tag{4.9a}$$

$$\mathbf{0} = \mathbf{g}^j\left(\mathbf{x}_c(T_{j-1}), \bar{\mathbf{p}}\right), \tag{4.9b}$$

$$\mathbf{0} = \mathbf{G}^j\left(\mathbf{x}_c(T_{j-1}), \bar{\mathbf{p}}\right) \mathbf{x}_v(T_{j-1}), \tag{4.9c}$$

with non-singular matrices $\mathbf{M}_j(\mathbf{x}_c(t), \bar{\mathbf{p}})$. This is a DAE of differential index 1. However, solving the (uniquely solvable) linear system

$$\mathbf{M}_j(\mathbf{x}_c(t), \bar{\mathbf{p}}) \begin{pmatrix} \dot{\mathbf{x}}_v(t) \\ -\mathbf{z}^j(t) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\tau}(t) - \mathbf{C}(\mathbf{x}(t), \bar{\mathbf{p}}) \\ \boldsymbol{\gamma}^j(\mathbf{x}(t), \bar{\mathbf{p}}) \end{pmatrix}$$

allows us to compute $\dot{\mathbf{x}}(t)$ and $\mathbf{z}^j(t)$ directly from $\mathbf{x}(t)$, $\boldsymbol{\tau}(t)$ and $\bar{\mathbf{p}}$. By incorporating the solution operator of the linear system, in the following we describe the dynamics using an Ordinary Differential Equation (ODE) formulation

$$\dot{\mathbf{x}}(t) = \bar{\mathbf{f}}^j\left(\mathbf{x}(t), \boldsymbol{\tau}(t), \bar{\mathbf{p}}\right), \qquad t \in \mathcal{T}_j, \tag{4.10a}$$

$$\mathbf{0} = \bar{\boldsymbol{\Gamma}}^j\left(\mathbf{x}(T_{j-1}), \bar{\mathbf{p}}\right) \tag{4.10b}$$

(see, e. g., [48, sec. 4.8]), where $\boldsymbol{\Gamma}^j(\cdot)$ summarizes the right-hand sides of the Constraints (4.9b) and (4.9c). Whenever the external contacts change, a jump in the generalized velocities is possible. The corresponding equation is given by (4.7). It can be expressed in terms of the differential states in the form

$$\mathbf{x}(T_j^+) = \bar{\boldsymbol{\Delta}}^j\left(\mathbf{x}(T_j^-), \bar{\mathbf{p}}\right), \tag{4.11}$$

for each change of contacts, where $\mathbf{x}(T_j^-) = \lim_{t \nearrow T_j} \mathbf{x}(t)$ and $\mathbf{x}(T_j^+) = \lim_{t \searrow T_j} \mathbf{x}(t)$ are the differential states instantly before and after the contact change, respectively. From (4.7), we see that – besides the arguments $\mathbf{x}(T_j^-)$ and $\bar{\mathbf{p}}$ – the so-called *jump*

function $\bar{\mathbf{\Delta}}^j(\cdot)$ depends on the external contact holding *after* change, whereas the external contact before change does not influence the jump function.

Modeling the Human Gait

After expressing the phase-wise defined dynamics of the MBS modeling the human body while walking, we model the human gait itself. For this, in accordance with optimality assumptions that are frequently made in modeling of processes in nature and particularly of human motions (see, e. g. [113, 143]), we claim (cf. [134, p. 1540]):

Assumption 4.1 (Optimality of Human Gaits)
Natural gaits are optimal with respect to a certain performance criterion depending on individual trait parameters. △

In other words, every person walks the way they does, because it is optimal for them. The involved optimality criteria are person-specific and a priori unknown. Criteria like energy efficiency and stability, but also pain and comfort are assumed to be of relevance, amongst others. Now having a dynamic model of the walking process at hand, in accordance with Assumption 4.1 we follow a common approach and model the human gait as a (local) solution of an OCP subject to the MBS dynamics and further constraints (cf., e. g., [71, 105]). In our Optimal Control model, the controls $\mathbf{u}(\cdot)$ generate the generalized forces of the MBS and this way actuate the system. More precisely, we have

$$\boldsymbol{\tau}(t) = \boldsymbol{\tau}\left(\mathbf{u}(t), \mathbf{p}_\tau\right), \tag{4.12}$$

where $\mathbf{p}_\tau \in \mathbb{R}^{n_{p_\tau}}$ denotes the model parameters involved in the generation of the generalized forces.

There exist different ways to set up an appropriate Optimal Control model for the human gait. A common assumption is that the number and order of model phases is known according to the gait cycle. In this case, let $j \in \{1, \dots, n\}$ enumerate the model phases, and let phase j take place in the interval $\mathcal{T}_j = [T_{j-1}, T_j]$. We summarize all occurring parameters in the parameter vector $\mathbf{p} \in \mathbb{R}^{n_p}$. The human gait can then be modeled as a solution of a multi-stage OCP of the general form

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{u}(\cdot),\\ T_1,\dots,T_n}} \quad \sum_{j=1}^{n}\left(\Phi_j^M\left(T_j,\mathbf{x}(T_j),\mathbf{p}\right)+\int_{T_{j-1}}^{T_j}\Phi_j^L\big(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\big)\,\mathrm{d}t\right) \tag{4.13a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t)=\mathbf{f}^j\left(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\right), \qquad t\in\mathcal{T}_j, \quad j=1,\dots,n, \tag{4.13b}$$

$$T_{j-1}\le T_j, \qquad j=1,\dots,n, \tag{4.13c}$$

$$\mathbf{x}(T_j^+)=\mathbf{\Delta}^j\left(\mathbf{x}(T_j^-),\mathbf{p}\right), \qquad j=1,\dots,n, \tag{4.13d}$$

$$\mathbf{0}\le\mathbf{c}^j\left(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\right), \qquad t\in\mathcal{T}_j, \tag{4.13e}$$

$$\mathbf{0}=\mathbf{r}^{\mathrm{eq}}\left(\mathbf{x}(T_0),\dots,\mathbf{x}(T_n),\mathbf{p}\right), \tag{4.13f}$$

$$\mathbf{0}\le\mathbf{r}^{\mathrm{ieq}}\left(\mathbf{x}(T_0),\dots,\mathbf{x}(T_n),\mathbf{p}\right), \tag{4.13g}$$

see, e. g., [105]. We discuss all occurring functions and quantities in the following.

The phase-wise defined Differential Equation (4.13b) reflects the dynamics of the system in ODE form and summarizes (4.10a) and (4.12). The Jump Condition (4.13d) describes the behaviour of the differential states at phase transition, see (4.11). For each phase, the Initial Constraints (4.9b) are encoded in (4.13f) which also includes the Constraint (4.9c) for Phase 1. For all subsequent phases, the satisfaction of (4.9c) is guaranteed by the jump condition. While the number and order of phases are fixed in this way of modeling, the phase durations are free and subject to optimization.

The phase-wise defined Path Constraints (4.13e) comprise inequality constraints which are required to hold during the process. They may include requirements as, e. g., collision avoidance or non self-penetration. The Point Constraints (4.13f) and (4.13g) can be used to define conditions whose satisfaction marks the end or the beginning of a phase. Furthermore, a common assumption in walking is periodicity. Constraints which couple $\mathbf{x}(T_0)$ and $\mathbf{x}(T_n)$ are also included in (4.13f) and (4.13g).

The Objective Function (4.13a) encodes a (individual) performance criterion, cf. Assumption 4.1. In this thesis, the Lagrange terms of the phases coincide and we consider a Mayer-term for the last phase only. Therefore, the objective function is given by

$$\Phi^M\left(T_n,\mathbf{x}(T_n),\mathbf{p}\right)+\int_{T_0}^{T_n}\Phi^L\left(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\right)\mathrm{d}t\,.$$

In [50], the authors compare gaits resulting from different objective functions. In the real world, we observe different styles of human walking as well. To reproduce these gaits using our modeling approach, the according person-specific objective func-

tions need to be determined – a challenging task. One way to do so is the Inverse Optimal Control approach, see, e. g., [71, 106, 107]. Here, the objective function is assumed to be a weighted sum of known criteria, and the weights for which an associated gait reproduces given measurement data best are determined by solving a bilevel optimization problem. This line of research is followed in another project which is running in parallel to the work performed in the course of this thesis, see Section 1.3. Consequently, in this thesis we assume that for each individual person's gait it is possible to provide a suitable objective function and therefore an individually calibrated Optimal Control model of Form (4.13). Furthermore, if the OCP modeling the gait has more than one local solution, we assume to know the solution which corresponds to the considered person's gait.

As mentioned above, modeling the human gait as a multi-stage OCP is based on the assumption that number and order of model phases are known. There are however situations (e. g., model-based treatment planning of CP, see Chapter 3), in which this assumption is not valid. In this case, free-phase formulations need to be considered. We derive such a free-phase formulation in Chapter 5 which leads to a Mixed-Integer Optimal Control Problem. Further free-phase approaches can be found in [99, 108, 121], respectively.

4.2 Model-Based Treatment Planning

In Chapter 3 we gave an overview of CP and noticed that there is a great potential for improvement in CP treatment planning. The goal of this thesis is to support this ongoing area of research by the development and application of mathematical models and methods. This section deals with model-based treatment planning in the context of Optimal Control models for the human gait. Model-based treatment planning strives to design a non-invasive computational testing environment for ex ante evaluation and assessment of potential surgery plans [134, p. 1538]. Ideally, such an environment would be able to predict the effect of a treatment on the human gait and thus could help to prevent surgeries with a non-desirable outcome.

General Approach and Literature

For model-based treatment planning three steps are of importance:

1. generation of a patient-specific model,
2. modeling of the treatment, and

3. prediction of the outcome.

In the following, we comment on these steps and give a short insight into corresponding literature.

Regarding Step 1, a suitable trade-off between model-accuracy and computational effort needs to be found. The used model should be as detailed as necessary to model the aspects of interest but as simple as possible in order to reduce the computational costs or enable the numerical treatment in practice.

In this thesis, we are interested in treatments concerning the musculoskeletal system which aim at improving a patient's gait. A frequently used approach is to set up a biomechanical model of the human body or the part of interest and study the effect of model alterations. Numerous contributions on patient-specific modeling as well as the outcome of surgeries – also for CP – can be found in the work of Delp [37].

One great unknown in Step 3 is the movement generation in the treated and thus anatomically altered body. This is directly connected with the prediction of the *dynamic* outcome of a treatment – in our case the human gait. One approach is to assume that, simply put, the pre-operative muscle excitations – generating the forces of the muscles which result in the considered movement – are assumed to equal the post-operative muscle excitations. Thus, they can be reused to study the outcome of a surgery which – due to the altered anatomy – changes the effect of the forces of the muscles and hence the resulting movement. For instance, in [54] the authors study the effect of rectus femoris transfer surgery (a treatment in CP management, see Section 3.5). In the reference, the surgery is modeled by an alteration of the spatial distribution of the rectus femoris muscle and its insertion point at the knee joint which results in an altered moment arm. To check the effect of the treatment, pre-operative muscle excitations are used again for the altered model in a forward simulation. While this approach can work out for the prediction of isolated movements, for the gait it will most likely fail due to constraint violations (e. g., ground penetration).

In [119], the authors present a framework which is close to the goal we pursue in our work. They study the effect of not only a single, but multiple surgeries performed during the same orthopedic intervention on the CP gait. Subsequently, their framework evaluates the post-operative gait performance using a measure which "*does not describe how a patient will move after treatment but how difficult it would be for*

him/her to achieve a normal gait pattern" [119, p. 17]. In particular, it cannot predict the patient's gait after surgery.

Other recent approaches make use of machine learning techniques to predict the outcome of surgeries, cf. [56] and [92]. In the latter reference, the authors work with a complex musculoskeletal model of the whole body, simulate different pathologic gaits occurring in CP, and are able to predict the resulting gait after different standard surgeries in CP management. One drawback of machine learning approaches is, however, that post-operative gait predictions – and consequently also possible treatment recommendations – come out of a black box which might cause ethical concerns in the context of the medical application.

Model-Based Treatment Planning in this Thesis

In this thesis, we model the human gait as solution of an OCP which is constrained by the dynamics of the MBS modeling the human body, see Section 4.1. The gait model depends on parameters $\mathbf{p} \in \mathbb{R}^{n_p}$ (including, e. g., physical parameters $\tilde{\mathbf{p}}$ of the MBS and parameters $\mathbf{p}_\tau$ which are involved in the generation of the generalized forces) which need to be calibrated patient-specifically. Another research project, running in parallel to the work on this thesis, focuses on the generation of patient-specific Optimal Control models, see Section 1.3. Therefore, throughout this thesis we assume that for a given patient we have a parametric and calibrated OCP at hand, and the gait of the patient is modeled by a – known – solution of this OCP. We then regard a treatment as a non-zero change of model parameters $\mathbf{\Delta p}$ (see Section 4.3 for a realization).

In this setting, we take an interest in three aspects of the prediction of treatment outcomes and present a mathematical model for each of them:

1. How does the resulting gait look like if an intervention is performed exactly as planned? – Prediction of treatment outcome for a change $\Delta\mathbf{p}$.

2. Assuming we are able to assess and quantify the result of a treatment: what is the worst resulting gait if an intervention is performed under uncertainty? – Worst possible treatment outcome if $\Delta\mathbf{p}$ lies in an uncertainty set $\Omega_{\mathbf{p}}$.

3. What is the ideal intervention for a specific patient in order to reach a certain goal? – Best possible treatment outcome for $\Delta\mathbf{p} \in \mathcal{I}_{\mathbf{p}}$ if $\mathcal{I}_{\mathbf{p}}$ models the set of possible interventions.

Prediction of Treatment Outcome

For this purpose, we have to solve an OCP – e. g., of Form (4.13) – with $\mathbf{p} = \mathbf{p}_{\text{pre}} + \Delta\mathbf{p}$, where $\mathbf{p}_{\text{pre}}$ is the parameter value which belongs to the pre-operative situation. For the numerical solution, a good initial guess for states and controls is required. A natural choice is the (known) solution of the problem before treatment which models the pre-operative gait. If this is not sufficient, a homotopy – approaching $\mathbf{p}_{\text{pre}} + \Delta\mathbf{p}$ more cautious – can be rewarding.

Worst Possible Treatment Outcome Under Uncertainty

This aspect is discussed in detail in Chapter 6. We model the worst possible outcome as solution of a bilevel optimization problem, either of max min Form (6.4) or of Form (6.5), if there is another measure for the success of a treatment than the optimal value of the objective function of the OCP modeling the gait. For instance, the evaluation and quantification of gait patterns is touched in Section 3.5. The resulting worst possible treatment outcome can then be incorporated in clinical decision making.

Best Possible Treatment

In this scenario, we seek an optimal intervention for a patient in order to achieve a certain goal which can be expressed by means of an objective function. For now, we assume that no uncertainty is involved. Then mathematically, this problem is similar to the worst-case optimization from the previous paragraph though the interpretation of the solution is different. Let $\Phi(\mathbf{x}(T_n), \mathbf{p})$ measure the quality of a gait – meaning the smaller the function value is, the better is the according gait – and let the set of possible interventions be encoded in $\mathcal{I}_{\mathbf{p}}$. For each $\mathbf{p} \in \mathcal{I}_{\mathbf{p}}$, the post-operative gait is modeled by the solution of an OCP, e. g. of Form (4.13), which we assume to be unique for the moment. Then solving a bilevel problem of the form

$$\begin{aligned} \min_{\substack{\mathbf{p}\in\mathcal{I}_{\mathbf{p}},\mathbf{x}(\cdot),\mathbf{u}(\cdot),\\ T_1,\dots,T_n}} \quad & \Phi(\mathbf{x}(T_n), \mathbf{p}) \\ \text{s.t.} \quad & \mathbf{x}(\cdot), \mathbf{u}(\cdot), T_1, \dots, T_n \text{ solve a problem of Form (4.13)} \end{aligned}$$

provides us with an optimal treatment plan for a given patient. Here, the lower level problem can be replaced by another suitable Optimal Control model for the human gait. For solution approaches to bilevel optimization problems, see Section 6.1.2. If uncertainties shall be considered additionally (e. g., in the accuracy of the treat-

ments), the approach can be combined with a worst-case optimization as described before which results in a three-level optimization problem. We do not treat such problems in this thesis.

4.3 Mathematical Modeling of Treatments

In this section, we propose a way for treatment modeling which is suitable for model-based treatment planning, in particular worst-case treatment planning, in the context of Optimal Control models for the human gait as described in the previous sections and Chapter 6. To this end, we model the gait of a patient as a solution of a parametric OCP of Form (4.13) which depends on patient- and in particular body-specific parameters $\mathbf{p} \in \Pi \subseteq \mathbb{R}^{n_p}$ that are altered through a medical intervention. Thus, the gait is represented by a solution $\left(\mathbf{T}^*, \mathbf{u}^*(\cdot), \mathbf{x}^*(\cdot)\right)$ – comprising phase durations $\mathbf{T}^* = (T_1^*, \ldots, T_n^*)$, controls $\mathbf{u}^*(\cdot)$, and differential states $\mathbf{x}^*(\cdot)$ – and we indicate the dependence on $\mathbf{p} \in \Pi$ by

$$\mathbf{T}^* = \mathbf{T}^*(\mathbf{p}), \quad \mathbf{u}^* = \mathbf{u}^*(\cdot;\mathbf{p}), \quad \text{and} \quad \mathbf{x}^* = \mathbf{x}^*(\cdot;\mathbf{p}).$$

For simplification of presentation, we assume the OCP to be uniquely solvable for each feasible $\mathbf{p}$ which relieves us from choosing the particular solution which corresponds to the true gait.

In real life, medical treatments result in altered gaits. To model these causalities mathematically, the implication

$$\mathbf{p} \neq \mathbf{p}' \implies \begin{pmatrix} \mathbf{T}^*(\mathbf{p}) \\ \mathbf{u}^*(\cdot;\mathbf{p}) \\ \mathbf{x}^*(\cdot;\mathbf{p}) \end{pmatrix} \neq \begin{pmatrix} \mathbf{T}^*(\mathbf{p}') \\ \mathbf{u}^*(\cdot;\mathbf{p}') \\ \mathbf{x}^*(\cdot;\mathbf{p}') \end{pmatrix} \tag{4.14}$$

needs to be transferred to the Optimal Control model, where $\mathbf{p}$ and $\mathbf{p}'$ correspond to the situation before and after the modeled treatment, respectively. In worst-case treatment planning as described in Chapter 6, we search for global solutions of optimization problems in which $\mathbf{p}$ is the optimization variable and $\mathbf{T}^*(\mathbf{p}), \mathbf{u}^*(\cdot;\mathbf{p}), \mathbf{x}^*(\cdot;\mathbf{p})$ are dependent variables. In view of this, it is particularly desirable that (4.14) holds for all feasible $\mathbf{p}$, $\mathbf{p}'$ with $\mathbf{p} \neq \mathbf{p}'$. In the following, we seek for ways of modeling treatments which satisfy this requirement.

A straightforward approach comprises a detailed modeling of affected parts of the body like joints or musculotendinous complexes. The anatomical changes per-

formed during an intervention then need to be translated into alterations of the corresponding model and its parameters. Drawbacks of this approach are the challenging and time-consuming calibration process on the one hand in which the model parameters – in particular those of the detailed submodels – are adapted to a specific patient, and the potentially high computational costs on the other hand which rise with the level of detail. Anyway, by the expertise of our cooperation partner (see Section 1.3) for many interventions their effect on the patient-specific parameters is still not clear. For instance, we consider the lengthening of a muscle in order to increase the flexibility of an adjacent joint: here, one method is – simply put – to incise certain parts of the muscle with multiple cuts, cf. [43, sec. 15.2]. Following our cooperation partner, during the procedure it is not perfectly known how a single cut increases the length of the muscle. Instead, the surgeons incise successively until the desired flexibility of the joint is reached. In this view the surgery is performed goal-oriented, meaning that the goal – a certain degree of flexibility – is more important than the exact implementation and in particular the final length of the muscle.

This is also the paradigm we follow in our "real-life oriented" modeling: we are interested in modeling the eventual effect of a surgery and not in the exact transfer of the surgery method to our model, as this might not be beneficial in view of the medical practice (see above). In the previous example, this means we are interested in the treatments ultimate effect, i. e., in the resulting (extended) ranges of motion, and not in the locations and the shape of the performed cuts or in the resulting length of the respective muscles.

By the expertise of our cooperation partner, many treatments in CP management eventually aim at extending the ranges of motion of joints. For the rest of this section, we are interested in modeling such interventions. In the Optimal Control models we use, the ranges of motion of the joints of the MBS can be expressed by means of those differential states $\mathbf{x}_i(\cdot)$ which represent generalized coordinates (cf. (4.8)) and according boundaries. We model them as box-constraints of the form

$$\underline{\boldsymbol{b}}_i' \leq \mathbf{x}_i(t) \leq \overline{\boldsymbol{b}}_i' \quad \text{for } t \in \mathcal{T}, \tag{4.15}$$

where $\mathcal{T} = [T_0, T_n]$ denotes the overall time horizon. One approach to model an intervention is to view $\underline{\boldsymbol{b}}_i'$ and $\overline{\boldsymbol{b}}_i'$ as mutable parameters. However, if none of the Constraints (4.15) are active in the solution $\left(\mathbf{T}^*, \mathbf{u}^*(\cdot), \mathbf{x}^*(\cdot)\right)$ of the OCP (i. e., hold with equality), the solution will not be affected by changes $\underline{\boldsymbol{b}}_i' + \Delta\underline{\boldsymbol{b}}_i'$ or $\overline{\boldsymbol{b}}_i' + \Delta\overline{\boldsymbol{b}}_i'$ as

long as

$$\underline{\boldsymbol{b}}_i' + \Delta\underline{\boldsymbol{b}}_i' \leq \inf_{t\in\mathcal{T}} \mathbf{x}_i^*(t) \quad \text{and} \quad \sup_{t\in\mathcal{T}} \mathbf{x}_i^*(t) \leq \overline{\boldsymbol{b}}_i' + \Delta\overline{\boldsymbol{b}}_i'. \tag{4.16}$$

Furthermore, if we solve the OCP numerically using a direct approach (see Section 2.4) without taking special care of the Constraints (4.15), they are assured to hold at certain evaluation points but can be violated in between. Thus, a change of parameters might not influence the numerical solution of the OCP although (4.16) is not satisfied. In sum, the approach does not satisfy our modeling requirement.

Alternatively, we propose a way of modeling in which the parameters, being subject of possible alterations by an intervention, enter the dynamics of the OCP. Here, we do not restrict the motion of a joint of interest to a fixed domain by Constraints (4.15) as in the previous paragraph. Instead, inspired by [5, 103] we implement so-called *passive reset forces* and represent the resulting generalized force as the sum of active and passive elements. Simply put, the passive reset forces push back the respective states $\mathbf{x}_i(\cdot)$ into a desired domain $[\underline{\boldsymbol{b}}_i, \overline{\boldsymbol{b}}_i]$ when they are about to leave it. Similar to, e. g., [5, 72, 155], we represent these forces by means of exponential functions. For the sake of brevity, in the following we omit the argument t for time-dependent variables. Let

$$\underline{\boldsymbol{\tau}}_i^{\text{pass}}(\mathbf{x}_i) \overset{\text{def}}{=} e^{-\underline{c}_i(\mathbf{x}_i - \underline{\boldsymbol{b}}_i)}, \tag{4.17a}$$

$$\overline{\boldsymbol{\tau}}_i^{\text{pass}}(\mathbf{x}_i) \overset{\text{def}}{=} -e^{\overline{c}_i(\mathbf{x}_i - \overline{\boldsymbol{b}}_i)}, \tag{4.17b}$$

(with $\underline{c}_i, \overline{c}_i > 0$) be the normalized passive reset forces. A schematic illustration of $|\underline{\boldsymbol{\tau}}_i^{\text{pass}}(\cdot)|$ and $|\overline{\boldsymbol{\tau}}_i^{\text{pass}}(\cdot)|$ is depicted in Fig. 4.4. If $\mathbf{x}_i$ is inside $(\underline{\boldsymbol{b}}_i, \overline{\boldsymbol{b}}_i)$ and sufficiently far away from $\underline{\boldsymbol{b}}_i$ and $\overline{\boldsymbol{b}}_i$, for suitable $\underline{c}_i$ and $\overline{c}_i$ both Forces (4.17) are negligibly small. If $\mathbf{x}_i$ approaches $\underline{\boldsymbol{b}}_i$ or $\overline{\boldsymbol{b}}_i$, the corresponding force strongly increases, respectively, and this way hinders $\mathbf{x}_i$ from leaving $[\underline{\boldsymbol{b}}_i, \overline{\boldsymbol{b}}_i]$. Thus, $\underline{\boldsymbol{b}}_i$ and $\overline{\boldsymbol{b}}_i$ can be seen as virtual bounds for $\mathbf{x}_i$. To prevent an effortless oscillation of $\mathbf{x}_i$ due to the passive reset forces, we include a (normalized) damping term

$$\boldsymbol{\tau}_i^{\text{damp}}(\dot{\mathbf{x}}_i) = -\beta_i \dot{\mathbf{x}}_i$$

(cf. [77, 103]) with a damping parameter $\beta_i > 0$. Here, the derivative $\dot{\mathbf{x}}_i$ entering the damping term is included in the differential states since $\mathbf{x}_i$ represents a generalized coordinate. More precisely, we have $\dot{\mathbf{x}}_i = \mathbf{x}_{i+n_{\text{dof}}}$, see (4.8).

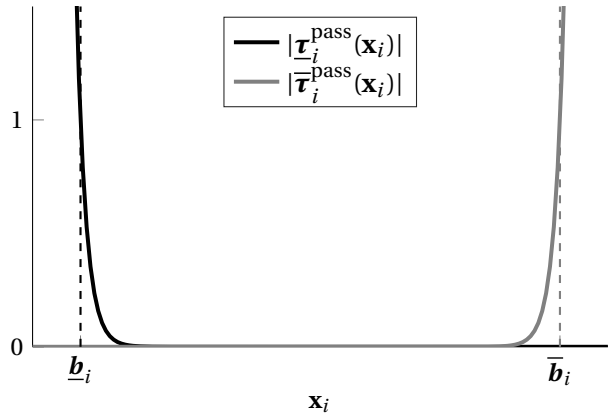


Figure 4.4: Schematic illustration of $|\underline{\tau}_i^{\text{pass}}|$ and $|\overline{\tau}_i^{\text{pass}}|$ – the absolute values of the Passive Reset Forces (4.17). The state $\mathbf{x}_i$ is not constrained to the interval $[\underline{b}_i, \overline{b}_i]$. However, when $\mathbf{x}_i$ approaches the virtual bounds $\underline{b}_i$ or $\overline{b}_i$, a passive reset force emerges and pushes $\mathbf{x}_i$ back to the interior of the U-shape.

Altogether, we put the resulting actuated generalized force as

$$\tau_i^a(\mathbf{u},\mathbf{x},\dot{\mathbf{x}}) = \tau_i^{a,\max}\left[\mathbf{u}_i + \underline{\tau}_{j(i)}^{\text{pass}}\left(\mathbf{x}_{j(i)}\right) + \overline{\tau}_{j(i)}^{\text{pass}}\left(\mathbf{x}_{j(i)}\right) + \tau_{j(i)}^{\text{damp}}\left(\dot{\mathbf{x}}_{j(i)}\right)\right] \tag{4.18}$$

where $\tau_i^{a,\max} > 0$ is the maximum active actuated generalized force, $\mathbf{u}_i \in [-1,1]$ the controlled normalized generalized force, and $j(i)$ the index of the differential states which belongs to i (according to (4.2) we have $j(i) = i + (n_{\text{dof}} - n_{\text{act}})$).

Now, the parameters $\tau_i^{a,\max}, \underline{b}_i, \overline{b}_i, \underline{c}_i, \overline{c}_i$, and β_i need to be chosen patient-specifically. After setting up reasonable values for $\tau_i^{a,\max}, \underline{b}_i$ and $\overline{b}_i$ (both latter can be guessed easily by physical examinations), we propose to estimate the intrinsic parameters $\underline{c}_i, \overline{c}_i$, and β_i from GA data using an adapted version of the so-called dynamics reconstruction (cf. [48, sec. 5.3] and [51]). This way, one can compute suitable control functions $\mathbf{u}$ and parameters $\underline{c}_i, \overline{c}_i$, and β_i, such that the motion capture data is reconstructed approximately.

In comparison to restricting the flexibility of $\mathbf{x}_i$ by state constraints as in (4.15), the described modeling is more realistic: No active force is needed to keep the states (mostly) inside the desired domain. This also holds in reality where passive torques

prevent, e. g., overstretching.

We model a treatment as a change of a subset of the virtual bound parameters $\underline{\boldsymbol{b}}_i$ and $\overline{\boldsymbol{b}}_i$ (assuming that all other involved parameters do not change significantly by the treatment). We define $\mathbf{p}$ – the parameter entering the OCP – to consist of all $\underline{\boldsymbol{b}}_i$ and $\overline{\boldsymbol{b}}_i$ which are changeable through the treatment. As these enter the dynamics through (4.18) we expect the Implication (4.14) to hold for all pairs of feasible parameters and our modeling requirement to be satisfied. The efficacy of the approach is demonstrated in Section 7.2.

Chapter 5

Numerical Solution of Optimal Control Problems with Switches, Switching Costs, and Jumps

In this thesis, we model the human gait as a solution of an Optimal Control Problem (OCP) with phase-wise defined dynamics and possible jumps of the differential states at phase transitions, cf. Section 4.1. The model phases are characterized by the foot-ground contact of the Multi-Body System (MBS) modeling the human body, and they switch whenever the contact points between the MBS and the ground change. Typically, number and order of phases during a gait cycle are assumed to be known, and the human gait can be modeled as a solution of a multi-stage OCP with predefined phase order, cf. Section 4.1.2. However, the situation is different in model-based treatment planning of Cerebral Palsy (CP). As mentioned in Chapter 3, a common gait pattern in CP is toe walking. By applying orthopedic changes to the musculoskeletal systems of patients, physicians aim at ameliorating this issue. After successful interventions, the patients' heels touch the ground while walking as well. Hence, the points of the feet which are in contact with the ground, and thus also number and order of model phases, change. In order to represent the gait before and after intervention, a model with a predefined sequence of phases is therefore not suitable.

These considerations motivate to study the Optimal Control of so-called switched systems, in which the number and order of model phases is subject to optimization. In this chapter, we present a novel approach for the numerical solution of switched systems with possible jumps in the differential states at phase transitions. An important component of this approach are so-called switching costs. We present two new approaches for the computation of switching costs after the discretization of a corresponding OCP.

This chapter is organized as follows: we give a literature overview on switched systems and the Optimal Control of those – also incorporating state jumps and switching costs – in Section 5.1. In Section 5.2, we introduce the problem formulation we are interested in. We reformulate the problem by means of so-called switching indi-

cator functions and the use of convexification techniques and discretize the control functions in Section 5.3. In Section 5.4, we present different approaches for handling switching costs in the context of OCPs and a comparison of those. Then, in Section 5.5, we complete the problem discretization, resulting in a Nonlinear Programming Problem (NLP) belonging to the class of Mathematical Programs with Vanishing Constraints (MPVCs). Section 5.6 is dedicated to MPVCs and we present an approach for their numerical treatment. Subsequently, in Section 5.7, we propose a strategy for the numerical solution of the resulting NLP from Section 5.5 and comment on a suitable software implementation in Section 5.8. Finally, we summarize the content of this Chapter in Section 5.9.

The content presented in this chapter can for the most part be found in the preprint [133] and parts of it are published in [134].

5.1 Literature Overview

In this chapter, we consider *hybrid systems* resp. *switched (dynamic) systems*. To be more precise, we are interested in dynamic systems which can switch between different operation modes and potentially allow for instantaneous changes of involved quantities on mode change, see Section 2.3.1. Examples are heating systems which regulate the temperature of a room automatically (cf. [145, sec. 2.2.4]) or mechanical systems under the influence of friction (cf. [26]) or with collision impacts (e. g., walking robots as in Section 4.1). Further examples are given in [145, ch. 2]. A concise introduction to switched systems can be found in [102, ch. 1] and for more details we refer to [63, 97, 145]. In addition, Zhu and Antsaklis [156] provide further references regarding the topic.

In particular, we take an interest in the Optimal Control of switched systems and in methods for computing such controls in practice. A survey on this topic is given in [156]. Necessary conditions for the solutions of switched OCPs can be found, e. g., in [42, 57, 142]. The literature on numerical approaches to the solution of switched OCPs distinguishes between *implicit* switching (also known as *state dependent* or *internally forced* switching) [27, 80] and *explicit* or *externally forced* switching [81, 127]. The former describes systems, where switches are triggered due to system-internal reasons, and the treatment of these system requires the solution of multi-point boundary value problems with switches and possible jumps in the differential states, cf. [21, 89]. In contrast, in explicitly switched systems the switching is steered by external input. However, in [26] the authors show how to transform implicitly

switched systems to explicit ones, and therefore we concentrate on explicit switching in the following. Two main categories exist for algorithms for the solution of OCPs with explicit switches: first, there are bilevel approaches (see, e. g., [65, 153, 154]) in which the order of operation modes is optimized on the upper level and the switching times as well as the control input function (for a given sequence of modes) on the lower level. Second, so-called *embedding transformation* methods (cf. [12, 82, 129]) relax a problem to obtain a continuously valued problem. The resulting problem can then be solved using standard approaches to Optimal Control (cf. Section 2.3), and the solution trajectories of the relaxed problem can be approximated arbitrary well using non-relaxed controls belonging to the original problem again, see [82, 127].

Only some of the available methods for the numerical solution of switched OCPs allow for state jumps on switching. In the literature, one can find methods which presume a predefined sequence of modes (resulting in a multi-stage optimization problem with state jumps, like Problem (4.13)) as described in [71], but also ones which are not bound to this restriction, e. g. [24, 98, 152]. In the latter two references, the authors focus on optimizing the switching sequence and do not consider a control input function in addition. Furthermore, the optimal number of switches needs to be known in advance. In contrast, the method proposed in [24] does not suffer from these disadvantages, but requires an integrator which is capable of handling the switching behaviour.

There are switched systems in which a switching between the operation modes of the system generates additional costs. For instance, for incandescent light bulbs every switch-on causes comparatively high abrasion and abbreviates the life span of a bulb. We refer to this phenomenon as switching costs, which in the context of Optimal Control can be modeled by a penalization of any change of modes in the objective function of the considered problem. Switching costs are addressed, e. g., in [16, 17, 41, 62, 64, 128] and [81, sec. 2.5].

5.2 Problem Formulation

In this section, we present the problem class of interest and formulate a representative as a Mixed-Integer Optimal Control Problem (MIOCP) with binary valued integer controls.

5.2.1 Optimal Control Problems with Switches, Switching Costs, and Jumps

We take an interest in OCPs in which the underlying dynamics can run in a finite number of different modes. Whenever the dynamics changes its mode, jumps in the differential states are possible.

We consider the time horizon $\mathcal{T} = [t_0, t_f]$, where both t_0 and t_f are fixed. The dynamic system we deal with can run in n different modes. We enumerate the modes and identify each mode by its corresponding number in $\{1, \dots, n\}$. For every $t \in \mathcal{T}$ the mode our system runs in is reflected by the value of a control function $w : \mathcal{T} \to \{1, \dots, n\}$ such that

$$\text{system is in mode } j \text{ at time } t \iff w(t) = j\,.$$

For the rest of this chapter, we assume the following:

Assumption 5.1 (Strictly Positive Dwell Time)
The considered system has a dwell time $\bar{\delta} > 0$, i. e., the system does not change its mode in $[t_0, t_0 + \bar{\delta})$ and whenever the system changes its mode at a time point t_s, it remains in the respective mode for at least all $t \in (t_s, t_s + \bar{\delta}) \subseteq \mathcal{T}$. △

For a subset $\mathcal{M} \subseteq \mathbb{R}^k$ we define

$$PC_{\bar{\delta}}(\mathcal{T}, \mathcal{M}) \overset{\text{def}}{=} \left\{ \boldsymbol{\rho} : \mathcal{T} \to \mathcal{M} \;\middle|\; \begin{array}{l} \bullet\, \forall\, t \in \mathcal{T} \setminus \{t_f\}\, \exists\, t_1, t_2 \in \mathcal{T} : t_2 - t_1 \geq \bar{\delta},\, t \in [t_1, t_2) \\ \quad \text{and } \boldsymbol{\rho}(t') = \boldsymbol{\rho}(t_1)\ \forall\, t' \in [t_1, t_2) \\ \bullet\, \boldsymbol{\rho}(t_f) = \boldsymbol{\rho}(t_f - \bar{\delta}) \end{array} \right\},$$

the *right-continuous piecewise constant functions* on $\mathcal{T}$ with values in $\mathcal{M}$ and dwell time $\bar{\delta}$. In accordance with Assumption 5.1, in the following we always demand $w(\cdot) \in PC_{\bar{\delta}}(\mathcal{T}, \{1, \dots, n\})$. Here, the right-continuity is a choice we make w. l. o. g.

For any right-continuous function $\mathbf{h} : \mathcal{T} \to \mathbb{R}^k$, for which also the left-hand side limits $\mathbf{h}(t^-) \overset{\text{def}}{=} \lim_{t' \nearrow t} \mathbf{h}(t')$ exist for all $t \in \mathcal{T} \setminus \{t_0\}$, we define

$$\mathcal{S}(\mathbf{h}) \overset{\text{def}}{=} \left\{ t_s \in \mathcal{T} \setminus \{t_0\} \,\middle|\, \mathbf{h}(t_s^-) \neq \mathbf{h}(t_s) \right\}.$$

Due to Assumption 5.1, $t_f \notin \mathcal{S}(w)$ and the set $\mathcal{S}(w)$ is finite. We denote its cardinality by $|\mathcal{S}(w)|$. The elements of $\mathcal{S}(w)$ are called *switching points* since the mode of the system changes at these time points. A change of modes is called a *switch*. Instead

of saying *"the system switches from mode j_1 to mode j_2"*, we simply write "$j_1 \rightarrow_w j_2$", where the subscript indicates the dependency of the mode on the control function $w(\cdot)$.

When the system is in mode $j \in \{1,\dots,n\}$ it is governed by the Ordinary Differential Equation (ODE) system

$$\dot{\mathbf{x}}(t) = \mathbf{f}^j(\mathbf{x}(t), \mathbf{u}(t)),$$

with differential states $\mathbf{x} : \mathcal{T} \rightarrow \mathbb{R}^{n_x}$ and control function $\mathbf{u} : \mathcal{T} \rightarrow \mathbb{R}^{n_u}$. Whenever the system changes its mode – i. e., $w(\cdot)$ changes its value – at a switching point t_s jumps in the differential states may occur. We denote the differential states instantly before the switch resp. after the switch by

$$\mathbf{x}(t_s^-) = \lim_{t' \nearrow t_s} \mathbf{x}(t') \quad \text{resp.} \quad \mathbf{x}(t_s^+) = \lim_{t' \searrow t_s} \mathbf{x}(t').$$

Furthermore, we assume that for every ordered pair $(j_1, j_2) \in \{1,\dots,n\}^2$ with $j_1 \neq j_2$, there exists a so-called jump function $\mathbf{\Delta}_{j_1,j_2} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x}$, representing the potential jump in the differential states at switching:

$$\mathbf{x}(t_s^+) = \mathbf{\Delta}_{j_1,j_2}\left(\mathbf{x}(t_s^-)\right) \quad \text{if } j_1 \rightarrow_w j_2 \text{ at } t_s \in \mathcal{S}(w).$$

During the process, path constraints $\mathbf{0} \geq \mathbf{d}(\mathbf{x}(t), \mathbf{u}(t))$ must be satisfied, where $\mathbf{0}$ is the zero vector of appropriate size, $\mathbf{d} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_d}$, and all inequalities must hold component-wise. These constraints may include simple bounds of the form $\underline{\boldsymbol{b}} \leq \mathbf{x}(t) \leq \overline{\boldsymbol{b}}$, and accordingly for the values of the control function. Additionally, for each mode j there are path constraints $\mathbf{0} \geq \mathbf{c}^j(\mathbf{x}(t), \mathbf{u}(t))$ with $\mathbf{c}^j : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \rightarrow \mathbb{R}^{n_{c_j}}$, which are only required to hold if the system runs in the respective mode. In addition, boundary constraints $\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}(t_0), \mathbf{x}(t_f)\right)$ with $\mathbf{r} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_r}$ must hold.

We set up an MIOCP to find a continuously valued control function $\mathbf{u}(\cdot)$, a discrete valued control function $w(\cdot)$, and differential states $\mathbf{x}(\cdot)$, such that all constraints are satisfied and the value of an objective function is minimized. The objective function is built up by two contributions: The first contribution is given by a Mayer-term $\Phi(\mathbf{x}(t_f))$ (w. l. o. g., cf. Section 2.3.3) and the second contribution is given by the (finite) number of switching points $|\mathcal{S}(w)|$, multiplied by a penalization parameter $\pi \geq 0$. We denote the second contribution as *switching costs*. The resulting MIOCP takes the form

$$\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot),w(\cdot)} \quad \Phi(\mathbf{x}(t_f)) + \pi\,|\mathcal{S}(w)| \tag{5.1a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t) = \mathbf{f}^j\left(\mathbf{x}(t),\mathbf{u}(t)\right), \qquad \text{if } w(t) = j\,,\ t \in \mathcal{T}\,, \tag{5.1b}$$

$$\mathbf{x}(t_s^+) = \boldsymbol{\Delta}_{j_1,j_2}\left(\mathbf{x}(t_s^-)\right), \quad \text{if } j_1 \to_w j_2 \text{ at } t_s \in \mathcal{S}(w)\,, \tag{5.1c}$$

$$\mathbf{0} \geq \mathbf{c}^j\left(\mathbf{x}(t),\mathbf{u}(t)\right), \qquad \text{if } w(t) = j\,,\ t \in \mathcal{T}\,, \tag{5.1d}$$

$$\mathbf{0} \geq \mathbf{d}\left(\mathbf{x}(t),\mathbf{u}(t)\right), \qquad t \in \mathcal{T}\,, \tag{5.1e}$$

$$\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}(t_0),\mathbf{x}(t_f)\right). \tag{5.1f}$$

In the following, we suppose that all occurring functions are sufficiently smooth for our purposes. Some remarks regarding Problem (5.1):

- For numerical computations, the demand for a dwell time $\bar{\delta} > 0$ is not restrictive, as we can imagine $\bar{\delta}$ to be the maximum possible granularity of the time grid.

- Though in the presented problem formulation switches arise explicitly from a change of values of the control function $w(\cdot)$, systems with implicitly *and* explicitly forced switches can be treated as well using the above problem formulation, cf. [26].

- By setting $\pi = 0$ and $\boldsymbol{\Delta}_{j_1,j_2}(\cdot) = \mathbf{Id}(\cdot)$ for all pairs (j_1, j_2), the presented problem formulation also covers switched systems without switching costs and jumps, as treated in [26]. Hence, the framework presented in this chapter extends the one from the latter reference.

- The constraints $\mathbf{d}(\cdot)$ could also be embedded in the $\mathbf{c}^j(\cdot)$. However, as the numerical treatment differs in our approach, we keep them distinct.

5.2.2 Binary Valued Integer Control Function

We reformulate Problem (5.1) using convexification techniques (see, e. g., [96, ch. 6]) in order to achieve a binary valued mode-indicator function. To this aim, we define

$$\mathbb{S}^n \overset{\text{def}}{=} \left\{\mathbf{v} \in \{0,1\}^n \,\middle|\, \textstyle\sum_{j=1}^n \mathbf{v}_j = 1\right\} \quad \text{and} \quad \Omega^n \overset{\text{def}}{=} \left\{\boldsymbol{\omega} : \mathcal{T} \to \{0,1\}^n \,\middle|\, \boldsymbol{\omega}(t) \in \mathbb{S}^n \ \forall\, t \in \mathcal{T}\right\}.$$

We have

Lemma 5.2
The mapping

$$\varphi : \Omega^n \cap PC_{\tilde{\delta}}(\mathcal{T}, \{0,1\}^n) \longrightarrow PC_{\tilde{\delta}}(\mathcal{T}, \{1,\dots,n\}),$$
$$\boldsymbol{\omega}(\cdot) \longmapsto w(t) \stackrel{\text{def}}{=} \sum_{j=1}^{n} \boldsymbol{\omega}_j(t) \cdot j .$$

is a bijection.

Proof See Appendix B.1.1 . □

In accordance with Assumption 5.1 and the above lemma, in the following we always demand $\boldsymbol{\omega}(\cdot) \in PC_{\tilde{\delta}}(\mathcal{T}, \{0,1\}^n)$. For $\boldsymbol{\omega}(\cdot) \in \Omega^n$, we set

$$j_1 \rightarrow_{\boldsymbol{\omega}} j_2 \stackrel{def}{\Longleftrightarrow} j_1 \rightarrow_{\varphi(\boldsymbol{\omega})} j_2 .$$

We consider the following MIOCP:

$$\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot)} \quad \Phi(\mathbf{x}(t_f)) + \pi\,|\mathcal{S}(\boldsymbol{\omega})| \tag{5.2a}$$
$$\text{s.t.} \quad \boldsymbol{\omega}(t) \in \mathbb{S}^n, \qquad t \in \mathcal{T}, \tag{5.2b}$$
$$\dot{\mathbf{x}}(t) = \textstyle\sum_{j=1}^{n} \boldsymbol{\omega}_j(t) \cdot \mathbf{f}^j(\mathbf{x}(t), \mathbf{u}(t)), \qquad t \in \mathcal{T}, \tag{5.2c}$$
$$\mathbf{x}(t_s^+) = \boldsymbol{\Delta}_{j_1,j_2}\left(\mathbf{x}(t_s^-)\right), \qquad \text{if } j_1 \rightarrow_{\boldsymbol{\omega}} j_2 \text{ at } t_s \in \mathcal{S}(\boldsymbol{\omega}), \tag{5.2d}$$
$$\mathbf{0} \geq \boldsymbol{\omega}_j(t) \cdot \mathbf{c}^j(\mathbf{x}(t), \mathbf{u}(t)), \qquad t \in \mathcal{T}, \forall\, j, \tag{5.2e}$$
$$\mathbf{0} \geq \mathbf{d}(\mathbf{x}(t), \mathbf{u}(t)), \qquad t \in \mathcal{T}, \tag{5.2f}$$
$$\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}(t_0), \mathbf{x}(t_f)\right). \tag{5.2g}$$

Then we get

Proposition 5.3
Problem (5.2) and Problem (5.1) are equivalent in the following sense: $(\mathbf{x}, \mathbf{u}, w)$ is feasible for Problem (5.1) if and only if $(\mathbf{x}, \mathbf{u}, \varphi^{-1}(w))$ is feasible for Problem (5.2), and the values of the corresponding objective functions coincide.

Proof See Appendix B.1.2 . □

5.3 Reformulation, Relaxation, and Control Discretization

Formulation (5.2) is not immediately accessible to numerical solvers due to the binary valued $\boldsymbol{\omega}(\cdot)$, the switching costs entering the objective function and the Constraints (5.2d). To resolve this issue, in this section we reformulate and relax the problem in a suitable manner and subsequently discretize the controls in the resulting relaxed problem.

5.3.1 Reformulation and Relaxation

In the following, let $\mathcal{G} \subset \mathcal{T} \setminus \{t_0, t_f\}$ be an arbitrary finite subset. We take a look at the number of switching points for a given control function $\boldsymbol{\omega}(\cdot) \in \Omega^n \cap PC_{\tilde{\delta}}(\mathcal{T}, \{0,1\}^n)$. For every pair $(j_1, j_2) \in \{1, \dots, n\}^2$ with $j_1 \neq j_2$ we consider a *switching indicator function* $\boldsymbol{\theta}_{j_1,j_2} : \mathcal{T} \to [0,1]$ in the following. We set

$$\boldsymbol{\theta}_{j_1,j_2}(t) = \begin{cases} 1 & \text{if } j_1 \to_{\boldsymbol{\omega}} j_2 \text{ at } t, \\ 0 & \text{else}, \end{cases} \tag{5.3}$$

which is equivalent to

$$\boldsymbol{\theta}_{j_1,j_2}(t) = \begin{cases} \min\left(\boldsymbol{\omega}_{j_1}(t^-), \boldsymbol{\omega}_{j_2}(t^+)\right) & \text{if } t \in \mathcal{S}(\boldsymbol{\omega}) \cup \mathcal{G}, \\ 0 & \text{else}, \end{cases}$$

(independent from the choice of $\mathcal{G}$), as one can easily verify. A schematic illustration of $\boldsymbol{\theta}_{j_1,j_2}(t)$ is given in Fig. 5.1. The number of switches can then be expressed using the $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$ by

$$|\mathcal{S}(\boldsymbol{\omega})| = \sum_{t \in \mathcal{S}(\boldsymbol{\omega}) \cup \mathcal{G}} \sum_{\substack{j_1,j_2=1 \\ j_1 \neq j_2}}^{n} \boldsymbol{\theta}_{j_1,j_2}(t). \tag{5.4}$$

We define the aggregated jump function $\boldsymbol{\Delta} : \mathbb{R}^{n_x} \times [0,1]^{n\cdot(n-1)} \to \mathbb{R}^{n_x}$ by

$$\boldsymbol{\Delta}\left(\mathbf{x}, \left(\mathbf{z}_{j_1,j_2}\right)_{j_1 \neq j_2}\right) = \sum_{\substack{j_1,j_2=1 \\ j_1 \neq j_2}}^{n} \mathbf{z}_{j_1,j_2} \boldsymbol{\Delta}_{j_1,j_2}(\mathbf{x}) + \left(1 - \sum_{\substack{j_1,j_2=1 \\ j_1 \neq j_2}}^{n} \mathbf{z}_{j_1,j_2}\right) \mathbf{x}, \tag{5.5}$$

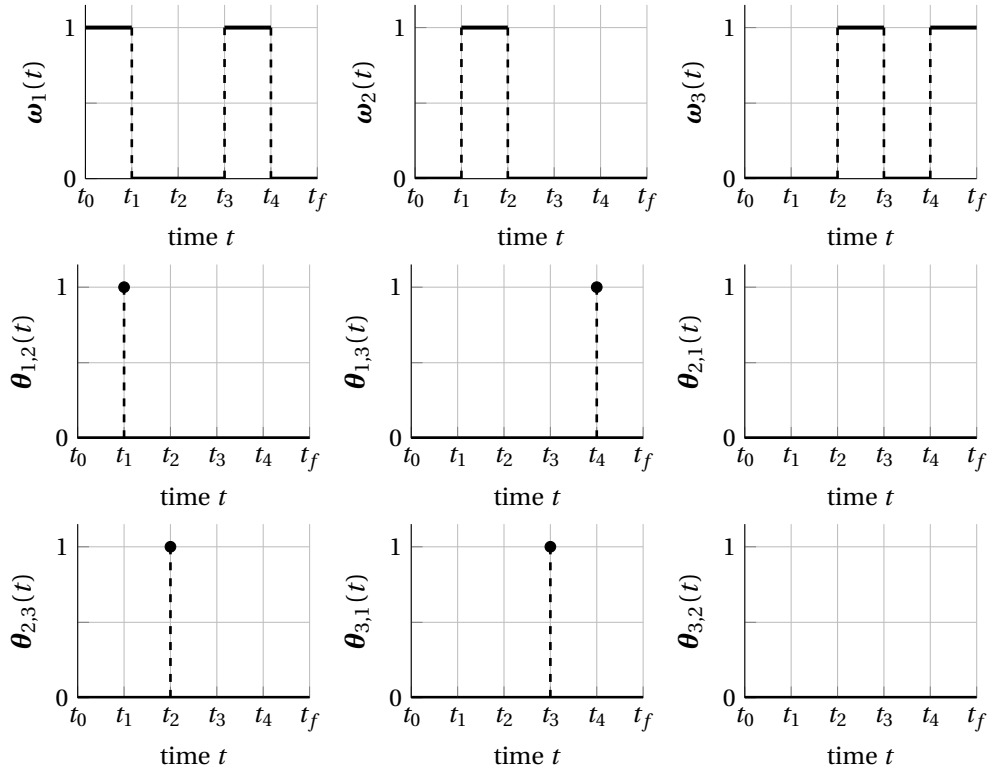


Figure 5.1: Schematic illustration of the family $\left(\boldsymbol{\theta}_{j_1,j_2}(\cdot)\right)_{j_1 \neq j_2}$ (as defined in (5.3)) in case $n = 3$ for a given sequence of modes, encoded in $\boldsymbol{\omega}(t)$. We highlight the non-zero values of $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$.

where the $\mathbf{\Delta}_{j_1,j_2}(\cdot)$ are the jump functions acting on the differential states $\mathbf{x}(t_s^-)$ at the switching points, and set up the following problem:

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot),\\ \boldsymbol{\theta}_{j_1,j_2}(\cdot)}} \Phi(\mathbf{x}(t_f)) + \pi \sum_{t\in\mathcal{S}(\boldsymbol{\omega})\cup\mathcal{G}} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \boldsymbol{\theta}_{j_1,j_2}(t) \tag{5.6a}$$

$$\text{s.t.} \quad \boldsymbol{\omega}(t) \in \mathbb{S}^n, \qquad t\in\mathcal{T}, \tag{5.6b}$$

$$\dot{\mathbf{x}}(t) = \textstyle\sum_{j=1}^{n} \boldsymbol{\omega}_j(t)\cdot\mathbf{f}^j(\mathbf{x}(t),\mathbf{u}(t)), \qquad t\in\mathcal{T}, \tag{5.6c}$$

$$\boldsymbol{\theta}_{j_1,j_2}(t) = 0, \qquad \text{if } t\notin\mathcal{S}(\boldsymbol{\omega})\cup\mathcal{G}, \tag{5.6d}$$

$$\boldsymbol{\theta}_{j_1,j_2}(t) = \min\left(\boldsymbol{\omega}_{j_1}(t^-), \boldsymbol{\omega}_{j_2}(t^+)\right), \qquad \text{if } t\in\mathcal{S}(\boldsymbol{\omega})\cup\mathcal{G}, \tag{5.6e}$$

$$\mathbf{x}(t^+) = \mathbf{\Delta}\left(\mathbf{x}(t^-), \left(\boldsymbol{\theta}_{j_1,j_2}(t)\right)_{j_1\neq j_2}\right), \qquad \text{if } t\in\mathcal{S}(\boldsymbol{\omega})\cup\mathcal{G}, \tag{5.6f}$$

$$\mathbf{0} \geq \boldsymbol{\omega}_j(t)\cdot\mathbf{c}^j(\mathbf{x}(t),\mathbf{u}(t)), \qquad t\in\mathcal{T}, \forall\, j, \tag{5.6g}$$

$$\mathbf{0} \geq \mathbf{d}(\mathbf{x}(t),\mathbf{u}(t)), \qquad t\in\mathcal{T}, \tag{5.6h}$$

$$\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}(t_0),\mathbf{x}(t_f)\right). \tag{5.6i}$$

Then we have

Proposition 5.4
For each finite subset $\mathcal{G}\subset\mathcal{T}\setminus\{t_0,t_f\}$ the Problems (5.2) and (5.6) are equivalent in the following sense: If $(\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot))$ is feasible for Problem (5.2), then there exist unique $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$ such that $\left(\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot),(\boldsymbol{\theta}_{j_1,j_2}(\cdot))_{j_1\neq j_2}\right)$ is feasible for Problem (5.6) and the values of the objective functions of the corresponding problems coincide. Vice versa, if $\left(\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot),(\boldsymbol{\theta}_{j_1,j_2}(\cdot))_{j_1\neq j_2}\right)$ is feasible for Problem (5.6), then $(\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot))$ is feasible for Problem (5.2) and the values of the objective functions coincide.

Proof See Appendix B.1.3 . □

Next, we relax Problem (5.6). More precisely, we replace the discrete valued control function $\boldsymbol{\omega}(\cdot)$ by a function $\boldsymbol{\alpha}(\cdot)$ with values in $[0,1]^n$, and furthermore relax Constraint (5.6e) to get rid of the min function. In the following, let $\boldsymbol{\alpha}\in PC_{\tilde{\delta}}(\mathcal{T},[0,1]^n)$ and $\boldsymbol{\beta}_{j_1,j_2}:\mathcal{T}\to[0,1]$ for $(j_1,j_2)\in\{1,\dots,n\}^2$ and $j_1\neq j_2$. We consider the problem

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{u}(\cdot),\boldsymbol{\alpha}(\cdot),\\ \boldsymbol{\beta}_{j_1,j_2}(\cdot),\boldsymbol{\theta}_{j_1,j_2}(\cdot)}} \Phi(\mathbf{x}(t_f)) + \pi \sum_{t\in\mathcal{S}(\boldsymbol{\alpha})\cup\mathcal{G}} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \boldsymbol{\theta}_{j_1,j_2}(t) \tag{5.7a}$$

$$\text{s.t.} \quad \boldsymbol{\alpha}(t) \in \operatorname{conv}\left(\mathbb{S}^n\right), \qquad t\in\mathcal{T}, \tag{5.7b}$$

$$\dot{\mathbf{x}}(t) = \textstyle\sum_{j=1}^{n} \boldsymbol{\alpha}_j(t)\cdot\mathbf{f}^j(\mathbf{x}(t),\mathbf{u}(t)), \qquad t\in\mathcal{T}, \tag{5.7c}$$

$$\boldsymbol{\beta}_{j_1,j_2}(t), \boldsymbol{\theta}_{j_1,j_2}(t) \in [0,1], \qquad t \in \mathcal{T}, \tag{5.7d}$$

$$0 = \boldsymbol{\beta}_{j_1,j_2}(t) = \boldsymbol{\theta}_{j_1,j_2}(t), \qquad \text{if } t \notin \mathcal{S}(\boldsymbol{\alpha}) \cup \mathcal{G}, \tag{5.7e}$$

$$\boldsymbol{\theta}_{j_1,j_2}(t) \geq \boldsymbol{\beta}_{j_1,j_2}(t)\boldsymbol{\alpha}_{j_1}(t^-) + \left(1 - \boldsymbol{\beta}_{j_1,j_2}(t)\right)\boldsymbol{\alpha}_{j_2}(t^+), \text{ if } t \in \mathcal{S}(\boldsymbol{\alpha}) \cup \mathcal{G}, \tag{5.7f}$$

$$\mathbf{x}(t^+) = \boldsymbol{\Delta}\left(\mathbf{x}(t^-), \left(\boldsymbol{\theta}_{j_1,j_2}(t)\right)_{j_1 \neq j_2}\right), \qquad \text{if } t \in \mathcal{S}(\boldsymbol{\alpha}) \cup \mathcal{G}, \tag{5.7g}$$

$$\mathbf{0} \geq \boldsymbol{\alpha}_j(t) \cdot \mathbf{c}^j(\mathbf{x}(t), \mathbf{u}(t)), \qquad t \in \mathcal{T}, \forall\, j, \tag{5.7h}$$

$$\mathbf{0} \geq \mathbf{d}(\mathbf{x}(t), \mathbf{u}(t)), \qquad t \in \mathcal{T}, \tag{5.7i}$$

$$\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}(t_0), \mathbf{x}(t_f)\right), \tag{5.7j}$$

where

$$\operatorname{conv}\left(\mathbb{S}^n\right) = \left\{ \mathbf{v} \in [0,1]^n \;\middle|\; \textstyle\sum_{j=1}^n \mathbf{v}_j = 1 \right\}$$

denotes the convex hull of $\mathbb{S}^n$. Then we have

Proposition 5.5
For every finite subset $\mathcal{G} \subset \mathcal{T} \setminus \{t_0, t_f\}$ Problem (5.7) is a relaxation of Problem (5.2) in the following sense: Let $\left(\mathbf{x}(\cdot), \mathbf{u}(\cdot), \boldsymbol{\omega}(\cdot)\right)$ be feasible for Problem (5.2) and set $\boldsymbol{\alpha}(\cdot) = \boldsymbol{\omega}(\cdot)$. Then there exist functions $\boldsymbol{\beta}_{j_1,j_2}(\cdot)$ and $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$, such that $\big(\mathbf{x}(\cdot), \mathbf{u}(\cdot), \boldsymbol{\alpha}(\cdot), (\boldsymbol{\beta}_{j_1,j_2}(\cdot))_{j_1 \neq j_2}, (\boldsymbol{\theta}_{j_1,j_2}(\cdot))_{j_1 \neq j_2}\big)$ is feasible for Problem (5.7), and the values of the objective functions of both problems coincide.

Proof See Appendix B.1.4 . □

In the context of Optimal Control, the technique of representing the dynamics of the system as the weighted sum of the dynamics belonging to each mode is known as Partial Outer Convexification (POC), cf. [127]. Subsequently, dropping the integrality constraint by replacing the binary valued control function $\boldsymbol{\omega}(\cdot) \in \Omega^n$ with a continuously valued control function $\boldsymbol{\alpha}(\cdot)$ yields a relaxed problem. Consider Problem (5.2) without switching costs, jumps, and the dwell time assumption. Dropping the integrality constraint transforms the MIOCP into a continuous OCP which is easier to solve. If we replace the zeros in the inequality constraints by $\varepsilon > 0$, one can show, that one can approximate all feasible tuples of the resulting relaxed problem arbitrarily well using binary feasible controls $\boldsymbol{\omega}(\cdot) \in \Omega^n$ again. In particular, this holds for an optimal solution and the according objective function value. Hence, POC and relaxation is a reasonable approach to solve MIOCPs. For details, see [82, 96, 127].

Regarding switching costs and excluding jumps in the differential states, recent approaches use relaxed MIOCP solutions to find binary feasible controls with minimal switching costs such that the resulting states approximate those of the relaxed

solution within a given accuracy [16, 17, 128], or to find binary feasible controls within given bounds on the switching costs that optimize the approximation accuracy of the states [128]. We are not aware of a modification of the theoretical result mentioned before or suitable rounding algorithms for problems in which the objective function depends on the integer controls (as for the switching costs) *and* state jumps, that are triggered by changes of the binary valued variable, occur. However, we will see later in Section 5.4 that after discretization, our way of penalizing switches in the objective function of Problem (5.7) hinders the occurrence of non-binary valued $\boldsymbol{\alpha}(\cdot)$ in a solution.

5.3.2 Control Discretization in Time

We intend to develop strategies for the numerical solution of Problem (5.7) using a direct approach ("first discretize, then optimize"). We discretize the control functions first. To do so, we introduce a time grid

$$\mathbb{G} = \{t_0 < t_1 < \cdots < t_N = t_f\}$$

with $\min_{i=1,\ldots,N} |t_i - t_{i-1}| \geq \bar{\delta}$. In accordance with Assumption 5.1, we restrict the control function $\boldsymbol{\alpha}(\cdot)$ to be locally constant on the grid intervals. Hence, we can parameterize $\boldsymbol{\alpha}(\cdot)$ using vectors $\mathbf{a}^0, \ldots, \mathbf{a}^{N-1} \in [0,1]^n$:

$$\begin{aligned} \boldsymbol{\alpha}(t) &= \mathbf{a}^i \qquad \text{for } t \in [t_i, t_{i+1}), i = 0, \ldots, N-2, \\ \boldsymbol{\alpha}(t) &= \mathbf{a}^{N-1} \quad \text{for } t \in [t_{N-1}, t_N]. \end{aligned}$$

We set $\mathcal{G} = \mathbb{G} \setminus \{t_0, t_f\}$. Observe, that due to the discretization, switches can only occur at the inner grid points, and therefore

$$\mathcal{S}(\boldsymbol{\alpha}) \subseteq \mathcal{G}.$$

The control functions $\boldsymbol{\beta}_{j_1,j_2}(\cdot)$ and $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$, which vanish outside $\mathcal{G}$ according to (5.7e), can be parameterized by $\boldsymbol{\beta}^i_{j_1,j_2}, \boldsymbol{\theta}^i_{j_1,j_2} \in [0,1]$, $i = 0, \ldots, N-2$, such that

$$\boldsymbol{\beta}^i_{j_1,j_2} = \boldsymbol{\beta}_{j_1,j_2}(t_{i+1}) \text{ and } \boldsymbol{\theta}^i_{j_1,j_2} = \boldsymbol{\theta}_{j_1,j_2}(t_{i+1}) \quad \text{for all } i = 0, \ldots, N-2.$$

The $\boldsymbol{\theta}^i_{j_1,j_2}$ are called *switching indicators*. Since $\boldsymbol{\alpha}(t^-_{i+1}) = \mathbf{a}^i$ and $\boldsymbol{\alpha}(t^+_{i+1}) = \mathbf{a}^{i+1}$, for every inner grid point the Constraints (5.7f) take the form

$$\boldsymbol{\theta}^i_{j_1,j_2} \geq \boldsymbol{\beta}^i_{j_1,j_2} \mathbf{a}^i_{j_1} + \left(1 - \boldsymbol{\beta}^i_{j_1,j_2}\right) \mathbf{a}^{i+1}_{j_2} \quad \text{for } j_1 \neq j_2 \text{ and } i = 0, \ldots, N-2.$$

Additionally, we represent the control function $\mathbf{u}(\cdot)$ by a function $\mathbf{U}(\cdot)$, which is determined by a finite number of parameters. The exact representation is discussed later in Section 5.5. After control discretization, the resulting problem takes the form

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{U}(\cdot),\boldsymbol{\alpha}(\cdot),\\ \mathbf{a}^i,\boldsymbol{\beta}^i_{j_1,j_2},\boldsymbol{\theta}^i_{j_1,j_2}}} \quad \Phi(\mathbf{x}(t_f)) + \pi \sum_{i=0}^{N-2} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \boldsymbol{\theta}^i_{j_1,j_2} \tag{5.8a}$$

$$\text{s.t.} \quad \mathbf{a}^i \in \operatorname{conv}\left(\mathbb{S}^n\right), \qquad i = 0,\dots,N-1, \tag{5.8b}$$

$$\boldsymbol{\alpha}(t) = \mathbf{a}^i \quad \text{for } t \in [t_i, t_{i+1}), \qquad i = 0,\dots,N-1, \tag{5.8c}$$

$$\dot{\mathbf{x}}(t) = \textstyle\sum_{j=1}^{n} \boldsymbol{\alpha}_j(t) \cdot \mathbf{f}^j(\mathbf{x}(t), \mathbf{U}(t)), \qquad t \in \mathcal{T}, \tag{5.8d}$$

$$\boldsymbol{\beta}^i_{j_1,j_2}, \boldsymbol{\theta}^i_{j_1,j_2} \in [0,1], \qquad i = 0,\dots,N-2, \tag{5.8e}$$

$$\boldsymbol{\theta}^i_{j_1,j_2} \geq \boldsymbol{\beta}^i_{j_1,j_2} \mathbf{a}^i_{j_1} + \left(1 - \boldsymbol{\beta}^i_{j_1,j_2}\right) \mathbf{a}^{i+1}_{j_2}, \quad i = 0,\dots,N-2, \tag{5.8f}$$

$$\mathbf{x}(t^+_{i+1}) = \boldsymbol{\Delta}\left(\mathbf{x}(t^-_{i+1}), \left(\boldsymbol{\theta}^i_{j_1,j_2}\right)_{j_1\neq j_2}\right), \qquad i = 0,\dots,N-2, \tag{5.8g}$$

$$\mathbf{0} \geq \boldsymbol{\alpha}_j(t) \cdot \mathbf{c}^j(\mathbf{x}(t), \mathbf{U}(t)), \qquad t \in \mathcal{T}, \forall j, \tag{5.8h}$$

$$\mathbf{0} \geq \mathbf{d}(\mathbf{x}(t), \mathbf{U}(t)), \qquad t \in \mathcal{T}, \tag{5.8i}$$

$$\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}(t_0), \mathbf{x}(t_f)\right). \tag{5.8j}$$

Observe, that for binary valued $\boldsymbol{\alpha}(\cdot)$, i. e. $\mathbf{a}^i \in \mathbb{S}^n$, we have

$$|\mathcal{S}(\boldsymbol{\alpha})| \leq \sum_{i=0}^{N-2} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \boldsymbol{\theta}^i_{j_1,j_2}, \tag{5.9}$$

and if the switching indicators $\boldsymbol{\theta}^i_{j_1,j_2}$ take their smallest possible value according to (5.8f), Equation (5.9) even holds with equality.

By reformulating, relaxing and partially discretizing Problem (5.1), we achieved the following: in Problem (5.1), switches – potentially enforced by the mode-dependent Constraints (5.1d) – can happen at any time $t \in [t_0 + \bar{\delta}, t_f - \bar{\delta}]$. We intend to use gradient-based optimization methods for the solution. In this case, it is in particular not clear how to handle the Jump Condition (5.1c) together with the switching costs contribution to the objective function numerically. In the relaxed and control-discretized Problem (5.8) however, switches can only occur at the (predefined) inner grid points, which simplifies the numerical treatment significantly. Moreover, our

approach leads to a differentiable formulation of all constraints – in particular the jump condition – and the objective function after full discretization (cf. Section 5.5). It thus allows for gradient-based optimization.

5.4 Switching Costs and Indicators

In the previous section, we expressed the switching costs in Problem (5.2) via Term (5.4), leading to a differentiable formulation in the relaxed and fully discretized problem, see Section 5.5. In this section, we additionally present two alternatives and compare all three expressions with each other. The approach presented in Section 5.4.2 was originally introduced in [81], and the other two were developed in the course of this thesis, generalizing the original idea.

For the rest of this section, let $\boldsymbol{\alpha} \in PC_{\bar{\delta}}(\mathcal{T}, [0,1]^n)$, such that $\boldsymbol{\alpha}(t') \in \operatorname{conv}(\mathbb{S}^n)$ for $t' \in \mathcal{T}$. Let $t \in \mathcal{T}$. In accordance with our previous notation, we say

$$\text{the system is in mode } j \text{ at } t \overset{def}{\iff} \boldsymbol{\alpha}_j(t) = 1 \text{ at } t\,.$$

If there is an index j with $\boldsymbol{\alpha}_j(t) \notin \{0,1\}$, we call this a *fractional mode*. Furthermore we extend our notation by

$$j_1 \rightarrow_{\boldsymbol{\alpha}} j_2 \text{ at } t_s \overset{def}{\iff} \boldsymbol{\alpha}_{j_1}(t_s^-) = \boldsymbol{\alpha}_{j_2}(t_s^+) = 1\,,$$

which again denotes a switch of modes at t_s.

In this section, we focus on switching costs. Therefore, we assume $\pi > 0$ and $\boldsymbol{\Delta}(\cdot) = \mathbf{Id}(\cdot)$ for the remainder of the section. In particular, we consider systems *without* jumps in the differential states. Nevertheless, for each of the approaches presented in the following, we comment on the suitability of the introduced switching indicators for the numerical treatment of switched OCPs with state-jumps.

5.4.1 Reformulation "Omniscient"

This reformulation was already used in Section 5.3. The members of the family $\left(\boldsymbol{\theta}^i_{j_1,j_2}\right)^i_{j_1 \neq j_2}$ occurring in the control-discretized Problem (5.8) are also called *"om-*

niscient" switching indicators, and the problem without jumps reads as

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{U}(\cdot),\boldsymbol{\alpha}(\cdot),\\ \mathbf{a}^i,\boldsymbol{\beta}^i_{j_1,j_2},\boldsymbol{\theta}^i_{j_1,j_2}}} \Phi(\mathbf{x}(t_f)) + \pi \sum_{i=0}^{N-2} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \boldsymbol{\theta}^i_{j_1,j_2} \qquad \text{(OCP-Omniscient)}$$

$$\begin{aligned}
\text{s.t.} \quad & \mathbf{a}^i \in \operatorname{conv}\left(\mathbb{S}^n\right), && i=0,\dots,N-1,\\
& \boldsymbol{\alpha}(t) = \mathbf{a}^i \ \text{ for } t\in[t_i,t_{i+1}), && i=0,\dots,N-1,\\
& \boldsymbol{\beta}^i_{j_1,j_2}, \boldsymbol{\theta}^i_{j_1,j_2} \in [0,1], && i=0,\dots,N-2,\\
& \boldsymbol{\theta}^i_{j_1,j_2} \geq \boldsymbol{\beta}^i_{j_1,j_2}\mathbf{a}^i_{j_1} + \left(1-\boldsymbol{\beta}^i_{j_1,j_2}\right)\mathbf{a}^{i+1}_{j_2}, && i=0,\dots,N-2, \qquad (5.10)\\
& (5.8\text{d}), (5.8\text{h})-(5.8\text{j}).
\end{aligned}$$

We refer to the term $\pi\sum_i\sum_{j_1\neq j_2}\boldsymbol{\theta}^i_{j_1,j_2}$ as *relaxed switching costs*. Let $\left(\mathbf{x}(\cdot),\mathbf{U}(\cdot),\boldsymbol{\alpha}(\cdot),\right.$ $\left.\mathbf{a}^i,\boldsymbol{\beta}^i_{j_1,j_2},\boldsymbol{\theta}^i_{j_1,j_2}\right)$ be feasible for Problem (OCP-Omniscient). Then

$$\boldsymbol{\theta}^i_{j_1,j_2} \geq \min\left(\mathbf{a}^i_{j_1}, \mathbf{a}^{i+1}_{j_2}\right) \quad \text{for } i=0,\dots,N-2 \qquad (5.11)$$

due to (5.10), and it is easy to see, that we can find feasible $\tilde{\boldsymbol{\beta}}^i_{j_1,j_2}$ and $\tilde{\boldsymbol{\theta}}^i_{j_1,j_2}$, such that (5.11) holds with equality. As $\pi>0$, this is enforced in an optimal solution of (OCP-Omniscient).

We will see later (in Section 5.4.4), that the relaxed switching costs contribute to hinder the occurrence of fractional modes in a solution of Problem (OCP-Omniscient). Now, we consider a solution of this problem. Let us assume that the corresponding $\boldsymbol{\alpha}^*(\cdot)$ is binary valued. Then for the respective optimal $\left(\boldsymbol{\theta}^i_{j_1,j_2}\right)^*$ (which are uniquely determined by $\boldsymbol{\alpha}^*(\cdot)$) we have

$$\left(\boldsymbol{\theta}^i_{j_1,j_2}\right)^* = \min\left[\left(\mathbf{a}^i_{j_1}\right)^*, \left(\mathbf{a}^{i+1}_{j_2}\right)^*\right] = \begin{cases} 1 & \text{if } j_1 \rightarrow_{\boldsymbol{\alpha}^*} j_2 \text{ at } t_{i+1},\\ 0 & \text{else}, \end{cases} \qquad \text{for } i=0,\dots,N-2.$$

In particular, in case of a switch the family of optimal switching indicators comprises the information which modes are involved in the switch in which order, i. e., if $j_1 \rightarrow_{\boldsymbol{\alpha}^*} j_2$ or $j_2 \rightarrow_{\boldsymbol{\alpha}^*} j_1$ at the according switching point. This justifies the naming "omniscient". Furthermore, we have

$$|\mathcal{S}(\boldsymbol{\alpha}^*)| = \sum_{i=0}^{N-2} \sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} \left(\boldsymbol{\theta}^i_{j_1,j_2}\right)^*.$$

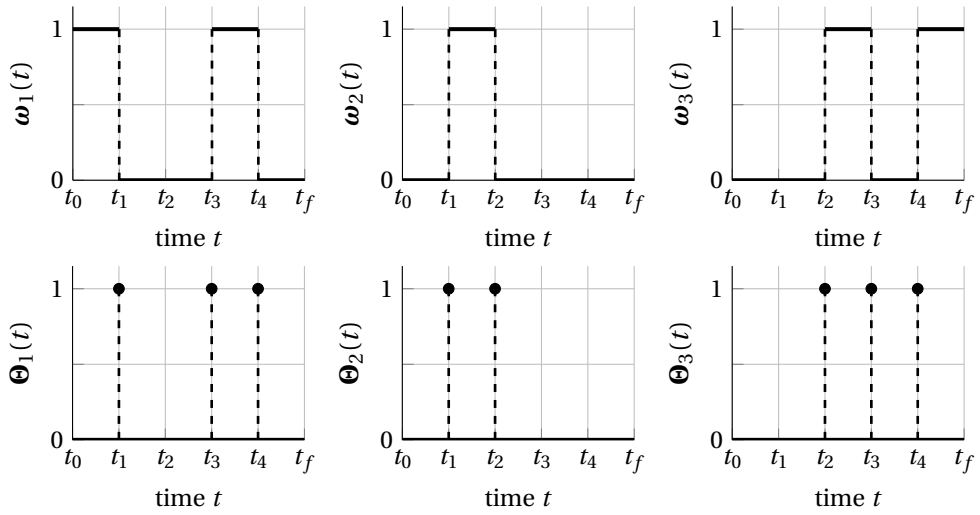


Figure 5.2: Schematic illustration of the family $\left(\boldsymbol{\Theta}_j(\cdot)\right)_j$ (see Section 5.4.2) in case $n = 3$ for a given sequence of modes, encoded in $\boldsymbol{\omega}(t)$. We highlight the non-zero values of $\boldsymbol{\Theta}_j(\cdot)$.

The omniscient switching indicators are suitable for the numerical treatment of possible jumps in the differential states as utilized in Problem (5.8).

5.4.2 Reformulation "Involved"

The switching indicators we present in this section were introduced in [81]. Let $\boldsymbol{\omega} \in \Omega^n \cap PC_{\tilde{\delta}}(\mathcal{T}, \{0,1\}^n)$ and $t \in \mathcal{T} \setminus \{t_0, t_f\}$. Then

$$\boldsymbol{\Theta}_j(t) \overset{\text{def}}{=} \min\left[\boldsymbol{\omega}_j(t^-) + \boldsymbol{\omega}_j(t^+), 2 - \boldsymbol{\omega}_j(t^-) - \boldsymbol{\omega}_j(t^+)\right] = \begin{cases} 1 & \text{if } j \rightarrow_{\boldsymbol{\omega}} j' \text{ at } t \text{ for } j' \neq j\,, \\ 1 & \text{if } j' \rightarrow_{\boldsymbol{\omega}} j \text{ at } t \text{ for } j' \neq j\,, \\ 0 & \text{else}\,, \end{cases}$$

for all j. A schematic illustration is given in Fig. 5.2. We get

$$|\mathcal{S}(\boldsymbol{\omega})| = \sum_{t_s \in \mathcal{S}(\boldsymbol{\omega})} \frac{1}{2} \sum_{j=1}^{n} \boldsymbol{\Theta}_j(t)\,.$$

If we use this idea and process Problem (5.2) without jumps in the differential states similarly as in Section 5.3, we obtain the control-discretized problem

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{U}(\cdot),\boldsymbol{\alpha}(\cdot),\\ \mathbf{a}^i,\boldsymbol{\beta}^i_j,\boldsymbol{\Theta}^i_j}} \quad \Phi(\mathbf{x}(t_f)) + \pi \sum_{i=0}^{N-2} \frac{1}{2} \sum_{j=1}^{n} \boldsymbol{\Theta}^i_j \tag{OCP-Involved}$$

$$\begin{aligned}
\text{s.t.} \quad & \mathbf{a}^i \in \operatorname{conv}\left(\mathbb{S}^n\right), && i = 0,\dots,N-1,\\
& \boldsymbol{\alpha}(t) = \mathbf{a}^i \ \text{ for } t \in [t_i, t_{i+1}), && i = 0,\dots,N-1,\\
& \boldsymbol{\beta}^i_j, \boldsymbol{\Theta}^i_j \in [0,1], && i = 0,\dots,N-2,\\
& \boldsymbol{\Theta}^i_j \geq \boldsymbol{\beta}^i_j \left(\mathbf{a}^i_j + \mathbf{a}^{i+1}_j\right) + \left(1-\boldsymbol{\beta}^i_j\right)\left(2-\mathbf{a}^i_j-\mathbf{a}^{i+1}_j\right), && i = 0,\dots,N-2, \qquad (5.12)\\
& (5.8\text{d}), (5.8\text{h}) - (5.8\text{j}),
\end{aligned}$$

The variables $\boldsymbol{\Theta}^i_j$ are called *"involved"* switching indicators. Again, we refer to the term $\pi \sum_{i=0}^{N-2} \frac{1}{2} \sum_{j=1}^{n} \boldsymbol{\Theta}^i_j$ as relaxed switching costs. For a feasible $\left(\mathbf{x}(\cdot), \mathbf{U}(\cdot), \boldsymbol{\alpha}(\cdot), \mathbf{a}^i, \boldsymbol{\beta}^i_j, \boldsymbol{\theta}^i_j\right)$, because of (5.12) we have

$$\boldsymbol{\Theta}^i_j \geq \min\left(\mathbf{a}^i_j + \mathbf{a}^{i+1}_j, 2 - \mathbf{a}^i_j - \mathbf{a}^{i+1}_j\right) \quad \text{for } i = 0,\dots,N-2, \tag{5.13}$$

and we can find feasible $\tilde{\boldsymbol{\beta}}^i_j$ and $\tilde{\boldsymbol{\Theta}}^i_j$ satisfying (5.13) with equality. Since $\pi > 0$, equality holds in an optimal solution of Problem (OCP-Involved).

As for the omniscient switching indicators, in Section 5.4.4 we will see that the relaxed switching costs hinder the occurrence of fractional modes in a solution of Problem (OCP-Involved). We consider a solution of this problem. Let us again assume that the corresponding $\boldsymbol{\alpha}^*(\cdot)$ is binary valued. Then, for the respective optimal $\left(\boldsymbol{\Theta}^i_j\right)^*$ (which are uniquely determined by $\boldsymbol{\alpha}^*(\cdot)$), we get

$$\left(\boldsymbol{\Theta}^i_j\right)^* = \begin{cases} 1 & \text{if } j \rightarrow_{\boldsymbol{\alpha}^*} j' \text{ at } t_{i+1} \text{ for } j' \neq j,\\ 1 & \text{if } j' \rightarrow_{\boldsymbol{\alpha}^*} j \text{ at } t_{i+1} \text{ for } j' \neq j,\\ 0 & \text{else}, \end{cases}$$

for $i = 0,\dots,N-2$, and consequently

$$|\mathcal{S}(\boldsymbol{\alpha}^*)| = \sum_{i=0}^{N-2} \frac{1}{2} \sum_{j=1}^{n} \left(\boldsymbol{\Theta}^i_j\right)^*.$$

In summary, for every inner grid point the optimal involved switching indicators hold the information whether a switch occurred or not, and if so, which modes are

involved in the switch (justifying the naming). In contrast to the omniscient switching indicators, the order of modes remains hidden.

Accordingly, using our approach in presence of jumps in the differential states, the involved switching indicators are useful, e. g. if

$$\boldsymbol{\Delta}_{j_1,j_2}(\cdot) = \boldsymbol{\Delta}_{j_2,j_1}(\cdot) \text{ for all } j_1 \neq j_2 .$$

In this case, instead of (5.5) we use the aggregated jump function

$$\boldsymbol{\Delta} : \mathbb{R}^{n_x} \times [0,1]^n \to \mathbb{R}^{n_x}, \quad \left(\mathbf{x}, \left(\mathbf{z}_j\right)_j\right) \mapsto \sum_{\substack{j_1,j_2=1 \\ j_1<j_2}}^{n} \mathbf{z}_{j_1}\mathbf{z}_{j_2}\boldsymbol{\Delta}_{j_1,j_2}(\mathbf{x}) + \left(1 - \sum_{\substack{j_1,j_2=1 \\ j_1<j_2}}^{n} \mathbf{z}_{j_1}\mathbf{z}_{j_2}\right)\mathbf{x}.$$

5.4.3 Reformulation "Subsequent"

Let again $\boldsymbol{\omega} \in \Omega^n \cap PC_{\bar{\delta}}(\mathcal{T}, \{0,1\}^n)$ and $t \in \mathcal{T} \setminus \{t_0, t_f\}$. Then

$$\boldsymbol{\theta}_j(t) \overset{\text{def}}{=} \min\left(\boldsymbol{\omega}_j(t^+), 1 - \boldsymbol{\omega}_j(t^-)\right) = \begin{cases} 1 & \text{if } j' \rightarrow_{\boldsymbol{\omega}} j \text{ at } t \text{ for some } j' \neq j , \\ 0 & \text{else} , \end{cases}$$

for all j. A schematic illustration is given in Fig. 5.3. We have

$$|\mathcal{S}(\boldsymbol{\omega})| = \sum_{t_s \in \mathcal{S}(\boldsymbol{\omega})} \sum_{j=1}^{n} \boldsymbol{\theta}_j(t) .$$

If we use this idea and process Problem (5.2) – without jumps in the differential states – as in Section 5.3, we get

$$\min_{\substack{\mathbf{x}(\cdot),\mathbf{U}(\cdot),\boldsymbol{\alpha}(\cdot), \\ \mathbf{a}^i,\boldsymbol{\beta}^i_j,\boldsymbol{\Theta}^i_j}} \quad \Phi(\mathbf{x}(t_f)) + \pi \sum_{i=0}^{N-2} \sum_{j=1}^{n} \boldsymbol{\theta}^i_j \tag{OCP-Subsequent}$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{a}^i \in \operatorname{conv}\left(\mathbb{S}^n\right), && i = 0,\dots,N-1, \\ & \boldsymbol{\alpha}(t) = \mathbf{a}^i \quad \text{for } t \in [t_i, t_{i+1}), && i = 0,\dots,N-1, \\ & \boldsymbol{\beta}^i_j, \boldsymbol{\theta}^i_j \in [0,1], && i = 0,\dots,N-2, \\ & \boldsymbol{\theta}^i_j \geq \boldsymbol{\beta}^i_j \mathbf{a}^{i+1}_j + \left(1 - \boldsymbol{\beta}^i_j\right)\left(1 - \mathbf{a}^i_j\right), && i = 0,\dots,N-2, \qquad (5.14) \\ & \text{(5.8d), (5.8h) – (5.8j)} . \end{aligned}$$

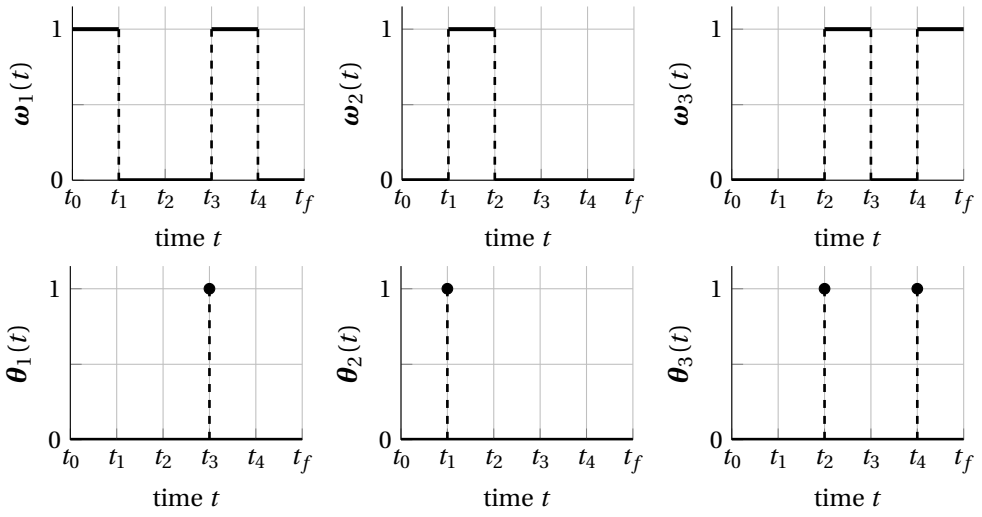


Figure 5.3: Schematic illustration of the family $\left(\boldsymbol{\theta}_j(\cdot)\right)_j$ (see Section 5.4.3) in case $n = 3$ for a given sequence of modes, encoded in $\boldsymbol{\omega}(t)$. We highlight the non-zero values of $\boldsymbol{\theta}_j(\cdot)$.

The variables $\boldsymbol{\theta}_j^i$ are called *"subsequent"* switching indicators. Again, we refer to the term $\pi \sum_{i=0}^{N-2} \frac{1}{2} \sum_{j=1}^{n} \boldsymbol{\theta}_j^i$ as relaxed switching costs. Let $\left(\mathbf{x}(\cdot), \mathbf{U}(\cdot), \boldsymbol{\alpha}(\cdot), \mathbf{a}^i, \boldsymbol{\beta}_j^i, \boldsymbol{\theta}_j^i\right)$ be feasible for Problem (OCP-Subsequent). From (5.14), we have

$$\boldsymbol{\theta}_j^i \geq \min\left(\mathbf{a}_j^{i+1}, 1 - \mathbf{a}_j^i\right) \quad \text{for } i = 0, \ldots, N-2, \tag{5.15}$$

and again there are feasible $\tilde{\boldsymbol{\beta}}_j^i$ and $\tilde{\boldsymbol{\theta}}_j^i$ satisfying (5.15) with equality. As $\pi > 0$, in an optimal solution of Problem (OCP-Subsequent) the Inequality (5.15) holds with equality.

Also for the subsequent switching indicators, in Section 5.4.4 we will see that the according relaxed switching costs hinder the occurrence of fractional modes in a solution of Problem (OCP-Subsequent). Now, we consider a solution of this problem and again assume, that the corresponding $\boldsymbol{\alpha}^*(\cdot)$ is binary valued. Then, for the respective optimal subsequent switching indicators $\left(\boldsymbol{\theta}_j^i\right)^*$ (which are uniquely determined by $\boldsymbol{\alpha}^*(\cdot)$), we have

$$\left(\boldsymbol{\theta}_j^i\right)^* = \begin{cases} 1 & \text{if } j' \rightarrow_{\boldsymbol{\alpha}^*} j \text{ at } t_{i+1} \text{ for } j' \neq j, \\ 0 & \text{else}, \end{cases}$$

for $i = 0, \dots, N-2$, and

$$|\mathcal{S}(\boldsymbol{\alpha}^*)| = \sum_{i=0}^{N-2} \sum_{j=1}^{n} \left(\boldsymbol{\theta}_j^i\right)^* .$$

Altogether, for every inner grid point the optimal subsequent switching indicators contain the information, whether a switch occurred or not, and if so, what is the mode in the *subsequent* grid interval after the switch (justifying the naming). The mode before the switch stays hidden.

In the presence of jumps in the differential states, the subsequent switching indicators are therefore useful, for instance, if the jump functions $\boldsymbol{\Delta}_{j_1,j_2}(\cdot)$ only depend on the mode after a switch, i. e., $\boldsymbol{\Delta}_{j_1,j_2}(\cdot) = \boldsymbol{\Delta}_{j_1',j_2}(\cdot)$ for all $j_1, j_1' \neq j_2$. In this particular situation, we set

$$\boldsymbol{\Delta}^j(\cdot) \overset{\text{def}}{=} \boldsymbol{\Delta}_{j',j}(\cdot) \quad \text{for some } j' \neq j\,,$$

and use the aggregated jump function

$$\boldsymbol{\Delta} : \mathbb{R}^{n_x} \times [0,1]^n \to \mathbb{R}^{n_x}, \quad \left(\mathbf{x}, \left(\mathbf{z}_j\right)_j\right) \mapsto \sum_{j=1}^{n} \mathbf{z}_j \boldsymbol{\Delta}^j(\mathbf{x}) + \left(1 - \sum_{j=1}^{n} \mathbf{z}_j\right)\mathbf{x}.$$

5.4.4 Properties of the Switching Indicators

We take a closer look at the three types of switching indicators introduced in the last sections. To this end, we define

$$\phi_{\text{omni}},\, \phi_{\text{inv}},\, \phi_{\text{subs}} : \operatorname{conv}\left(\mathbb{S}^n\right) \times \operatorname{conv}\left(\mathbb{S}^n\right) \longrightarrow \mathbb{R}$$

by

$$\begin{aligned}
\phi_{\text{omni}}(\mathbf{b},\mathbf{c}) &\overset{\text{def}}{=} \sum_{\substack{j_1,j_2=1 \\ j_1 \neq j_2}}^{n} \min(\mathbf{b}_{j_1}, \mathbf{c}_{j_2})\,,\\
\phi_{\text{inv}}(\mathbf{b},\mathbf{c}) &\overset{\text{def}}{=} \frac{1}{2} \sum_{j=1}^{n} \min(\mathbf{b}_j + \mathbf{c}_j, 2 - \mathbf{b}_j - \mathbf{c}_j)\,,\\
\phi_{\text{subs}}(\mathbf{b},\mathbf{c}) &\overset{\text{def}}{=} \sum_{j=1}^{n} \min(\mathbf{c}_j, 1 - \mathbf{b}_j)\,.
\end{aligned}$$

The minimal values of the respective relaxed switching costs for the Problems (OCP-Omniscient), (OCP-Involved), and (OCP-Subsequent) (which are taken in a so-

lution of each respective problem) are then given by

$$\pi \sum_{i=0}^{N-2} \phi_{\text{omni}}\left(\mathbf{a}^i, \mathbf{a}^{i+1}\right), \quad \pi \sum_{i=0}^{N-2} \phi_{\text{inv}}\left(\mathbf{a}^i, \mathbf{a}^{i+1}\right), \quad \pi \sum_{i=0}^{N-2} \phi_{\text{subs}}\left(\mathbf{a}^i, \mathbf{a}^{i+1}\right),$$

respectively, with $\mathbf{a}^i$ being the value of the mode-indicator function $\boldsymbol{\alpha}(\cdot)$ on the grid interval $[t_i, t_{i+1})$, cf. Sections 5.3-5.4.3.

We first investigate upper bounds of the three functions.

Proposition 5.6
Let $\mathbf{b}, \mathbf{c} \in \text{conv}(\mathbb{S}^n)$. We have $\phi_{\text{inv}}(\mathbf{b}, \mathbf{c})$, $\phi_{\text{subs}}(\mathbf{b}, \mathbf{c}) \leq 1$. If $\mathbf{b}_j + \mathbf{c}_j \leq 1$ for every component j, we get $\phi_{\text{inv}}(\mathbf{b}, \mathbf{c}) = \phi_{\text{subs}}(\mathbf{b}, \mathbf{c}) = 1$. For ϕ_{omni} we have

$$\max_{\mathbf{b},\mathbf{c} \in \text{conv}(\mathbb{S}^n)} \phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) = n - 1.$$

Proof See Appendix B.1.5. □

Second, we investigate lower bounds.

Proposition 5.7
We have $\phi_{\text{omni}}(\mathbf{b}, \mathbf{c})$, $\phi_{\text{inv}}(\mathbf{b}, \mathbf{c})$, $\phi_{\text{subs}}(\mathbf{b}, \mathbf{c}) \geq 0$ for all $\mathbf{b}, \mathbf{c} \in \text{conv}(\mathbb{S}^n)$ and

$$\phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) = \phi_{\text{inv}}(\mathbf{b}, \mathbf{c}) = \phi_{\text{inv}}(\mathbf{b}, \mathbf{c}) = 0 \quad \Longleftrightarrow \quad \mathbf{b}, \mathbf{c} \in \mathbb{S}^n \text{ and } \mathbf{b} = \mathbf{c}.$$

Proof See Appendix B.1.6. □

Consider the Problems (OCP-Omniscient), (OCP-Involved), and (OCP-Subsequent). The previous proposition states that in view of the (minimum possible) relaxed switching costs it is optimal to avoid fractional modes and to stay in the same mode for the whole time horizon. Nevertheless, due to constraints or the Mayer-term contribution in the respective objective functions, switches are unavoidable or desired, respectively.

Next, we investigate the incurring switching costs at two adjacent grid points.

Proposition 5.8
Let $\mathbf{b}, \mathbf{c}, \mathbf{d} \in \operatorname{conv}(\mathbb{S}^n)$. For $i \in \{\text{inv,subs}\}$, the "triangle inequality"

$$\phi_i(\mathbf{b}, \mathbf{d}) \leq \phi_i(\mathbf{b}, \mathbf{c}) + \phi_i(\mathbf{c}, \mathbf{d}) \tag{5.16}$$

holds. For ϕ_{omni} this is not true in general. However, if $\mathbf{b}, \mathbf{d} \in \mathbb{S}^n$, then (5.16) also holds for $i = \text{omni}$.

Proof See Appendix B.1.7. □

Consider the Problems (OCP-Omniscient), (OCP-Involved), and (OCP-Subsequent) again. Assume the system is in mode j_1 at time t_i and in mode j_2 at time t_{i+2}. The previous proposition states that with regard to the (minimum possible) relaxed switching costs it is not advantageous for the system to switch into some third (fractional or non-fractional) transition mode at time t_{i+1}. However, there are fractional modes for which this behavior is also not disadvantageous, as the next proposition shows.

Proposition 5.9
Let $\mathbf{b}, \mathbf{d} \in \mathbb{S}^n$, such that $\mathbf{b}_l = \mathbf{d}_k = 1$ for some $l \neq k$, and $\mathbf{c} \in \operatorname{conv}(\mathbb{S}^n)$. Then for $i \in \{\text{omni,inv,subs}\}$ we have

$$\phi_i(\mathbf{b}, \mathbf{d}) = \phi_i(\mathbf{b}, \mathbf{c}) + \phi_i(\mathbf{c}, \mathbf{d}) \quad \Longleftrightarrow \quad \mathbf{c}_l + \mathbf{c}_k = 1\,. \tag{5.17}$$

Proof See Appendix B.1.8. □

As a last step, we consider the special case $n = 2$. Here, the choice of switching indicators makes no difference in view of the minimal values of the relaxed switching costs:

Proposition 5.10
Let $n = 2$. Then for all $\mathbf{b}, \mathbf{c} \in \operatorname{conv}(\mathbb{S}^n)$ we have

$$\phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) = \phi_{\text{inv}}(\mathbf{b}, \mathbf{c}) = \phi_{\text{subs}}(\mathbf{b}, \mathbf{c})\,.$$

Proof See Appendix B.1.9.

5.5 State and Control Parametrization

We solve Problem (5.8) computationally using a Direct Collocation approach [20] (see also Section 2.4.2) to transcribe the control-discretized OCP to an NLP. In the

following, we give a detailed description of the parametrization. We extend the framework presented in [26] by additionally allowing for jumps in the differential states and stick close to the latter reference in what follows.

We have already introduced the time grid

$$\mathbb{G} = \{t_0 < t_1 < \cdots < t_N = t_f\}$$

in Section 5.3.2. For the representation of the differential states $\mathbf{x}(\cdot)$ and controls $\mathbf{U}(\cdot)$ we choose piecewise defined polynomials over the grid intervals $[t_i, t_{i+1}]$. Our framework admits to represent the different components $\mathbf{x}_j(\cdot)$ and $\mathbf{U}_j(\cdot)$ by polynomials of different degrees, respectively, cf. [102, ch. 7]. However, for the sake of a better readability, we assume in the following that all components are represented using polynomials of the same degree. Similar to [26], we use flipped Legendre-Gauss-Radau (LGR) points transformed to the grid intervals as collocation points. To be more precise, let $\mathcal{P}_l(\cdot)$ be the l-th Legendre Polynomial. Then the flipped LGR points are given by the l roots of $\mathcal{P}_{l-1}(\cdot)+\mathcal{P}_l(\cdot)$ mirrored at the origin and lie in $(-1,1]$. For $i = 0,\dots,N-1$, the map

$$[-1,1] \to [t_i, t_{i+1}], \quad t \mapsto t_i + (t_{i+1} - t_i)\frac{t+1}{2}$$

transforms the flipped LGR points to $[t_i, t_{i+1}]$ affinely. This way, we obtain collocation points $t_k^{(i)} \in (t_i, t_{i+1}]$, $k = 1,\dots,K_i$, for the discretization of the differential states, and $\bar{t}_m^{(i)} \in (t_i, t_{i+1}]$, $m = 1,\dots,\bar{K}_i$, for discretization of the control function $\mathbf{U}(\cdot)$, for each $i = 0,\dots,N-1$. To express the jump condition, in addition we set $t_0^{(i)} = t_i$ for each i.

Using these points, for each grid interval $[t_i, t_{i+1}]$ we consider Lagrange basis functions

$$\mathcal{L}_k^{(i)}(t) \overset{\text{def}}{=} \prod_{\substack{l=0 \\ l\neq k}}^{K_i} \frac{t - t_l^{(i)}}{t_k^{(i)} - t_l^{(i)}}, \quad k = 0,\dots,K_i,$$

and

$$\bar{\mathcal{L}}_m^{(i)}(t) \overset{\text{def}}{=} \prod_{\substack{l=1 \\ l\neq m}}^{\bar{K}_i} \frac{t - \bar{t}_l^{(i)}}{\bar{t}_m^{(i)} - \bar{t}_l^{(i)}}, \quad m = 1,\dots,\bar{K}_i.$$

In each grid interval, we represent the differential states by a polynomial

$$\mathbf{X}^{(i)}(t) \overset{\text{def}}{=} \sum_{k=0}^{K_i} \mathbf{x}_k^{(i)} \mathcal{L}_k^{(i)}(t), \quad t \in [t_i, t_{i+1}], \quad i = 0, \ldots, N-1,$$

using $K_i + 1$ points and nodal values $\mathbf{x}_k^{(i)} \in \mathbb{R}^{n_x}$. For the time derivatives, we get the polynomials

$$\dot{\mathbf{X}}^{(i)}(t) = \sum_{k=0}^{K_i} \mathbf{x}_k^{(i)} \dot{\mathcal{L}}_k^{(i)}(t), \quad t \in (t_i, t_{i+1}), \quad i = 0, \ldots, N-1,$$

which we extend continuously to the grid points. Similar to the representation of the differential states, in each grid interval the control function $\mathbf{U}(\cdot)$ is represented by

$$\mathbf{U}^{(i)}(t) \overset{\text{def}}{=} \sum_{m=1}^{\bar{K}_i} \mathbf{u}_m^{(i)} \bar{\mathcal{L}}_m^{(i)}(t), \quad t \in [t_i, t_{i+1}], \quad i = 0, \ldots, N-1, \tag{5.18}$$

using $\bar{K}_i$ points and nodal values $\mathbf{u}_m^{(i)} \in \mathbb{R}^{n_u}$.

We discretize the Mayer-term in the objective function as $\Phi\left(\mathbf{x}_{K_{N-1}}^{(N-1)}\right)$ and collocate the Differential Equation (5.8d) on each grid interval by

$$\begin{aligned} \mathbf{0} &= \dot{\mathbf{X}}^{(i)}\left(t_k^{(i)}\right) - \sum_{j=1}^{n} \mathbf{a}_j^i \mathbf{f}^j\left(\mathbf{X}^{(i)}\left(t_k^{(i)}\right), \mathbf{U}^{(i)}\left(t_k^{(i)}\right)\right) \\ \iff \quad \mathbf{0} &= \sum_{l=0}^{K_i} \mathbf{x}_l^{(i)} \dot{\mathcal{L}}_l^{(i)}\left(t_k^{(i)}\right) - \sum_{j=1}^{n} \mathbf{a}_j^i \mathbf{f}^j\left(\mathbf{x}_k^{(i)}, \mathbf{U}^{(i)}\left(t_k^{(i)}\right)\right), \end{aligned}$$

for $k = 1, \ldots, K_i$ and $i = 0, \ldots, N-1$. The Jump Conditions (5.8g) in discretized form are given by

$$\mathbf{x}_0^{(i+1)} = \boldsymbol{\Delta}\left(\mathbf{x}_{K_i}^{(i)}, (\boldsymbol{\theta}^i_{j_1, j_2})_{j_1 \neq j_2}\right), \quad i = 0, \ldots, N-2,$$

and discretizing the Boundary Constraints (5.8j) yields

$$\mathbf{0} \geq \mathbf{r}\left(\mathbf{x}_0^{(0)}, \mathbf{x}_{K_{N-1}}^{(N-1)}\right).$$

Simple bounds of the form $\underline{\boldsymbol{b}} \leq \begin{pmatrix}\mathbf{x}(t) & \mathbf{u}(t)\end{pmatrix}^T \leq \overline{\boldsymbol{b}}$ – which may be included in the Path Constraints (5.8i) – are directly transferred to the nodal values. The remaining path constraints in (5.8i) are demanded to be satisfied at a non-empty subset of the

collocation points and the points $t_0^{(i)}$,

$$\mathbf{0} \geq \mathbf{d}\left(\mathbf{x}_k^{(i)}, \mathbf{U}^{(i)}\left(t_k^{(i)}\right)\right), \quad k \in \mathcal{K}^{(i)} \subseteq \{0,\dots,K_i\},\ i = 0,\dots,N-1,$$

and the mode-dependent Path Constraints (5.8h) must hold at all grid points:

$$\mathbf{0} \geq \mathbf{a}_j^i \cdot \mathbf{c}^j\left(\mathbf{x}_0^{(i)}, \mathbf{U}^{(i)}\left(t_0^{(i)}\right)\right), \qquad j = 1,\dots,n,\ i = 0,\dots,N-1, \tag{5.19a}$$

$$\mathbf{0} \geq \mathbf{a}_j^{N-1} \cdot \mathbf{c}^j\left(\mathbf{x}_{K_{N-1}}^{(N-1)}, \mathbf{U}^{(N-1)}\left(t_{K_{N-1}}^{(N-1)}\right)\right), \qquad j = 1,\dots,n. \tag{5.19b}$$

5.6 Numerical Treatment of Vanishing Constraints

Constraints of the Form (5.19a) and (5.19b) are called Vanishing Constraints (VCs): if $\mathbf{a}_j^i = 0$, the corresponding constraint vanishes in the sense that it holds independently from the value of the second factor, which justifies the naming. The NLP resulting from the discretization described in the previous section is a Mathematical Program with Vanishing Constraints (MPVC). In the following, we give a brief introduction to MPVCs and present a relaxation strategy for their numerical treatment. This section is based on [26, 73]. The interested reader can find extensive information in the latter reference.

5.6.1 Mathematical Programs with Vanishing Constraints

An NLP of the form

$$\min_{\mathbf{x}\in\mathbb{R}^n} \quad f(\mathbf{x}) \tag{5.20a}$$

$$\text{s.t.} \quad \mathbf{g}_i(\mathbf{x}) \leq 0, \quad i = 1,\dots,m, \tag{5.20b}$$

$$\mathbf{h}_j(\mathbf{x}) = 0, \quad j = 1,\dots,p, \tag{5.20c}$$

$$\mathbf{H}_i(\mathbf{x}) \geq 0, \quad i = 1,\dots,l, \tag{5.20d}$$

$$\mathbf{H}_i(\mathbf{x})\mathbf{G}_i(\mathbf{x}) \leq 0, \quad i = 1,\dots,l, \tag{5.20e}$$

with continuously differentiable $f, \mathbf{g}_i, \mathbf{h}_j, \mathbf{H}_i, \mathbf{G}_i : \mathbb{R}^n \to \mathbb{R}$ is called an MPVC [1, 73]. MPVCs arise in many applications, e. g., in truss topology design [1], or more generally in discretized MIOCPs. In particular, the NLP resulting from the discretization of Problem (5.8) is an MPVC.

Due to the structure of the Constraints (5.20e), an MPVC is in general a non-convex problem [73, p. 2]. Furthermore, Constraint Qualifications (CQs) may be violated:

let $\mathbf{x}$ be a feasible point for Problem (5.20). If $\{i \mid \mathbf{H}_i(\mathbf{x}) = 0\} \neq \emptyset$, the Linear Independence Constraint Qualification (LICQ) is violated at $\mathbf{x}$, and $\{i \mid \mathbf{H}_i(\mathbf{x}) = 0 \text{ and } \mathbf{G}_i(\mathbf{x}) \geq 0\} \neq \emptyset$ even implies a violation of the weaker Mangasarian-Fromovitz Constraint Qualification (MFCQ), see [73, ch. 4]. However, it is reasonable to assume that the Guignard Constraint Qualification (GCQ) is satisfied, see [73, ch. 4]. In this case, local optima still satisfy the Karush-Kuhn-Tucker (KKT) conditions. The (potential) lack of strong CQs causes numerical problems and we expect standard NLP solvers to perform poorly or to fail.

5.6.2 Relaxation Approach

In view of the mentioned difficulties, a common solution approach – which was originally introduced for the field of Mathematical Programs with Complementarity Constraints in [135] – is to consider a family of relaxed problems

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{g}_i(\mathbf{x}) \leq 0, \qquad i = 1,\dots,m, \\ & \mathbf{h}_j(\mathbf{x}) = 0, \qquad j = 1,\dots,p, \\ & \mathbf{H}_i(\mathbf{x}) \geq 0, \qquad i = 1,\dots,l, \\ & \mathbf{H}_i(\mathbf{x})\mathbf{G}_i(\mathbf{x}) \leq \gamma, \qquad i = 1,\dots,l, \end{aligned} \tag{5.21}$$

with a VC parameter $\gamma > 0$, instead of tackling Problem (5.20) directly. The feasible set of the Family (5.21) approaches the feasible set of Problem (5.20) from the outside for $\gamma \searrow 0$. Under mild assumptions, the relaxed problems have advantageous properties in view of holding CQs. Details and convergence results can be found in [73, ch. 10].

5.7 Numerical Solution Approach

We propose to solve the collocation NLP resulting from Problem (5.8) using the relaxation approach resp. homotopy $\gamma \searrow 0$ described in Section 5.6.2 and give more details on the realization in the following. The presented approach is similar to the one described in [26] and we stick closely to this reference. Since we solve a discretized problem, switches can only happen at the inner points of the collocation grid. Thus, the need for mesh refinement arises. As in [26], we couple the discretization accuracy with the value of the VC parameter γ and propose to solve a sequence of NLPs, where γ is diminished while the grid is refined successively. For the solution

of the NLPs, we use state-of-the-art NLP solvers.

In the beginning of the homotopy, we choose an initial VC parameter γ_0 and solve the according NLP. If it turns out to be infeasible, the grid is refined and we attempt to solve the resulting NLP with the same VC parameter γ. If it is feasible, the grid is refined in order to to increase the solution accuracy, and we advance on the homotopy by reducing γ for the next NLP using the rule

$$\gamma_{\text{new}} = \rho\,\gamma_{\text{old}} \tag{5.22}$$

with a fixed $\rho \in (0,1)$. This procedure is repeated until an optimal solution of a feasible NLP with prescribed (problem-specific) termination tolerance $\gamma \leq \gamma_{\text{acc}}$ is found.

For the grid refinement as well as for the warm-start of the solver in subsequent iterations, in [26] the authors propose a strategy which, however, does not work properly anymore for our augmented framework due to the treatment of jumps. Therefore, we propose to assign a strategy for refining the grid as well as for warm-starting the solver to each problem individually, which enables us to include problem-specific knowledge.

Attracting Binary Solutions of the Relaxation

Although the relaxed switching costs hinder the occurrence of fractional modes in the solutions of the NLPs (as seen in Section 5.4.4), in general the Relaxation (5.8) permits $\boldsymbol{\alpha}(\cdot)$ to take fractional values. This can potentially lead to fractional values of $\boldsymbol{\theta}^i_{j_1,j_2}$ in turn, resulting in non-physical values of the jump function $\boldsymbol{\Delta}(\cdot)$, see (5.5) and (5.8g). We therefore strive to attract binary values. However, non-physical jumps can also occur in case of a binary valued $\boldsymbol{\alpha}(\cdot)$ if one of the $\boldsymbol{\theta}^i_{j_1,j_2}$ does not take its smallest possible value.

In case one of these issues occurs after the homotopy terminates, adapting the problem appropriately and restarting the homotopy might be expedient. We propose to augment the objective function with an additional term

$$\pi_2 \sum_{j=1}^{n} \int_{t_0}^{t_f} \boldsymbol{\alpha}_j(t)(1-\boldsymbol{\alpha}_j(t))\,dt \tag{5.23}$$

where $\pi_2 > 0$, cf. [127]. If the considerer problem behaves well, choosing π_2 large enough yields binary valued $\boldsymbol{\alpha}(\cdot)$. If π is chosen large enough as well, this yields bi-

nary valued $\boldsymbol{\theta}^i_{j_1,j_2}$ which take their smallest possible value in turn, due to the relaxed switching costs term $\pi \sum_i \sum_{j_1 \neq j_2} \boldsymbol{\theta}^i_{j_1,j_2}$ in the objective function.

In summary, if $\boldsymbol{\alpha}(\cdot)$ and $\boldsymbol{\theta}^i_{j_1,j_2}$ resulting from the homotopy do not meet our expectations, adding Term (5.23) with sufficiently large π_2 to the objective function while possibly adapting π can resolve the issue.

5.8 Implementation

In the previous sections, we described our approach for the numerical solution of OCPs with switching costs and jumps, extending the framework presented in [26]. The approach presented in the latter reference is implemented in the software package `grc`, see [102]. We extended `grc` by the `jump` add-on, which adds the required functionality to implement the approach described in this chapter. The enhanced software package allows to tackle generic switched OCPs with possible switching costs, and potentially occurring jumps in the differential states triggered by a switch of modes, i. e., problems of Form (5.1). The extension is implemented in the `MATLAB`® computing environment [100]. As in the original `grc` software, the arising NLPs can be solved with the software packages `SNOPT` [61] and `IPOPT` [146], implementing Sequential Quadratic Programming and Interior-Point methods, respectively.

When setting up a problem, the model functions – describing the objective function of the OCP (besides the switching costs), the dynamics including the aggregated jump function, path constraints, and boundary constraints – need to be specified in `C++` within the framework `SolvIND` [3], which is internally used for the generation of derivatives by means of the tool `Adol-C` [148] for automatic differentiation. Hence, if external libraries are used for setting up the model functions, the user needs to ensure the compatibility of these libraries with `Adol-C`.

5.9 Summary

In this chapter, we proposed a new approach for the numerical solution of switched OCPs with switching costs and jumps as well as two new approaches for the treatment of switching costs for discretized switched OCPs. We started our investigation with a general OCP of Form (5.1), where switches can happen at any time and finitely many times during the process while the order of modes and number of switches is left free for optimization. However, the latter is subject to penalization. In addition, jumps in the differential states are possible whenever the system changes its

mode. To make the problem accessible to gradient-based solvers, we reformulated and relaxed it by means of switching indicator functions and the use of convexification techniques. Discretizing the control functions of the resulting problem then yielded a problem, where switches can only happen at the inner grid points and are registered by switching indicators. We presented three different types of those indicators, whereof two were developed in the course of this thesis, and investigated their properties and suitability for the treatment of OCPs with jumps in the differential states. Finally, we fully discretized the control-discretized problem and ended up with an MPVC, for which we proposed a numerical solution approach. The presented method was implemented in the `jump` add-on of the software package `grc`. We demonstrate the efficacy of our approach in Section 7.1.

Chapter 6

Worst-Case Treatment Planning by Bilevel Optimal Control

We take an interest in the effect of orthopedic treatments on the gait of a Cerebral Palsy (CP) patient. In medical practice, inaccuracies can occur during the implementation of an intervention. We assume that the degree of possible inaccuracy is known. Then naturally the question arises if the planned surgery improves the patient's gait in any case – and in particular for the worst possible intervention outcome – considering the known uncertainty. Being able to answer this question makes treatment planning more robust and reduces the amount of negative outcomes after surgery.

Motivated by the above thoughts, in this chapter we develop an approach for the prediction of worst possible outcomes of orthopedic interventions which aim at improving the gait of a patient. As explained in Chapter 4, the gait of a patient is modeled as a solution of a parametric Optimal Control Problem (OCP) with parameters $\mathbf{p} \in \mathbb{R}^{n_p}$, and an orthopedic intervention is reflected by a non-zero change of parameters $\Delta\mathbf{p} \in \mathbb{R}^{n_p}$. Now, assuming that $\Delta\mathbf{p}$ lies in an uncertainty set $\Omega_{\mathbf{p}}$, our objective is to identify a worst possible treatment option $\Delta\mathbf{p} \in \Omega_{\mathbf{p}}$ and the corresponding gait pattern. Here, the term "worst" refers to a criterion which assesses the post-operative gait. Mathematically, this yields a bilevel optimization problem with an OCP on the lower level. We remark, that the approach presented in this chapter is rather general and its applications are not restricted to the field of CP treatment planning.

This chapter is organized as follows: we give overviews on robust optimization and bilevel optimization in Section 6.1. In Section 6.2, we present our so-called Training Approach in which a worst possible gait under parameter uncertainty is modeled as a solution of a bilevel optimization problem with a parametric OCP on the lower level. Here, again the term "worst" refers to a criterion which assesses the post-operative gait, e. g., the optimal objective function value of the parametric OCP. Our approach for computing the worst possible treatment outcome is fundamentally different from the approaches commonly used in the field of robust optimization. To illustrate this difference, in Section 6.3 we consider a test case and compare the Train-

ing Approach with a frequently used approach in robust optimization. In Section 6.4, we present an approach for the numerical solution of the bilevel optimization problem resulting from the Training Approach. Finally, in Section 6.5 we give an outlook and explain how the Training Approach can be applied in a real-world scenario.

6.1 Overviews on Robust Optimization and Bilevel Optimization

In this section, we give brief overviews on robust optimization and bilevel optimization, both of which are related to worst-case treatment planning in the sense of this chapter.

6.1.1 Robust Optimization

Robust optimization is concerned with optimization problems which involve uncertain parameters whose value is a priori unknown. These uncertainties may influence the satisfaction of constraints but also the objective function value. The aim of robust optimization is to robustify or immunize a solution against uncertainty in terms of feasibility and optimality. We remark, that the dependence of the objective function on the uncertain parameter may be neglected as one can transfer the uncertainty to the constraints by using an equivalent problem formulation, see [66, p. 3]. However, for illustrative reasons we consider uncertainties in the objective function as well since this occurs in many applications. Comprehensive material on robust optimization can be found in the textbook [11], and surveys on the topic are given by [15, 19]. Furthermore, [66] provides hints regarding practical issues. Robust optimization has numerous applications in diverse areas, e. g., portfolio management in finance (uncertain mean returns and return covariance matrices of risky assets) [15, sec. 5.1.1], inventory control in supply chain management (uncertain demand) [14, 15], and truss topology design in engineering (uncertain load) [10, 15]. In view of the topic of this chapter, we also list the application of robust optimization methods to treatment planning under uncertainty in proton therapy (uncertain density), see, e. g. [55]. Further applications can be found in [15, sec. 5] and [19].

In general, one distinguishes between statistical or stochastical uncertainty models and deterministic uncertainty models. In this thesis, we focus on the latter, i. e., the uncertain parameters lie in a so-called uncertainty set. For deterministic uncertainty models, robustifying the solution of the considered problem means that the robustified solution yields feasible parameter-dependent variables for *all* possible realizations of the uncertain parameters, and that the robustified solution is opti-

mal with regard to the worst possible value the objective function can take due to uncertainty. As this approach takes into account all possible realizations of the uncertain parameters, it is rather conservative. Since our considerations are motivated by a medical application, we view conservatism as an advantage for ethical reasons. Hence, a deterministic uncertainty model is suitable for our purposes.

In this thesis, we are in particular interested in the robustification of Optimal Control models against uncertainties. Here, the uncertainty can enter the model dynamics in form of time-dependent disturbances, or by an uncertain parameter which enters the differential equation or – equivalently – the initial value of the differential states. Introductions to the robustification of OCPs and extensive lists of related literature can be found, e. g., in [75, 140]. For an example of an application, we refer to [40]. In the robustification of OCPs, one distinguishes between problems in which it is not possible to react to disturbances during the process, as in [40], and problems where feedback is available and the pursued strategy can be corrected during the process, see, e. g., [87, 137]. If no feedback is available, in case of deterministic uncertainty models the control functions and controllable parameters have to be chosen in a way that for *all* possible disturbances the resulting states are feasible – a strong limitation. In this thesis, our considerations are motivated by treatment planning of CP patients. As discussed in Section 3.2, the majority of these patients suffers from a form of CP in which movements are performed voluntarily and do not suffer from uncertainty. Therefore, we assume that the control function in our gait model (cf. Section 4.1) is not perturbed and we focus on parameter uncertainty.

6.1.2 Bilevel Optimization

We will see later that a robust optimization approach is not perfectly suitable for the application we have in mind (i. e., worst-case treatment planning for CP patients whose movements are performed voluntarily). Instead, the approach we propose in this chapter yields a bilevel optimization problem with a parametric OCP on the lower level whose influencing parameter is optimized on the upper level.

A bilevel optimization problem is an optimization problem in which an optimization problem enters the constraints. We refer to the former problem as upper level problem and to the latter one as lower level problem. A popular example is a so-called leader-follower game, in which two players – the leader and the follower – compete against each other. Both players try to optimize certain cost functions, each depending on the actions of both players. To any action of the leader the follower reacts in

an optimal manner. Knowing this, the leader will ideally choose an action which is optimal under the assumption that the follower chooses an optimal action subsequently – a bilevel optimization problem. Examples from the real world can be found, e. g., in environmental economics, where the government tries to achieve a certain environmental goal in an optimal way by taxing or subsidizing companies which in turn react to the governments decision by choosing an action which maximizes their profit, cf. [38, sec. 2.3].

Introductions to bilevel optimization, including the historic roots of the field, can be found in the reviews [33, 79, 139], and more extensively in [38]. Additionally, a large number of references related to the topic can be found in [39]. Besides the previously mentioned examples, there is a large variety of applications from diverse areas, e. g. defense, energy networks, toll setting, and optimal design. Related references can be found, e. g., in [39] and [139].

Since we model the human gait as solution of an OCP, cf. Section 4.1, in view of our application we are interested in Bilevel OCPs, i. e., bilevel optimization problems where at least one of the involved optimization problems is an OCP. Problems of this kind are treated, e. g., in [71, 84, 101], and more references regarding the topic can be found therein. As application, in [85] the author considers container cranes in industrial warehouses which transport goods from an initial to a desired position in a certain optimal way, while at any point during the transportation process an emergency stop has to be possible due to safety requirements. This emergency stop again needs to be performed in an optimal manner. Further examples can be found in the field of Inverse Optimal Control (IOC) as, e. g., in [4, 71, 106]. In IOC, the lower level problem models a process from the real world by an OCP with unknown optimization criterion and the upper level problem is a parameter identification problem. The solution of the bilevel problem then determines an objective function of the lower level OCP such that a solution of the OCP reproduces given measurements best. Furthermore, Bilevel OCPs also arise when robustifying OCPs, see, e. g. [40].

Solution Approaches to Bilevel Optimization Problems

A common approach for treating bilevel optimization problems it to transform the bilevel problem into a single level problem. Subsequently, the resulting problem is tackled by deriving optimality conditions or employing suitable (possibly established) optimization methods. A frequently applied single level reduction technique is to replace the lower level problem by its first-order necessary conditions,

cf. [4, 71, 85]. However, in general the resulting single level problem is not equivalent to the original bilevel problem. If inequality constraints are present on the lower level, this approach transforms the bilevel problem into a finite or infinite dimensional Mathematical Program with Complementarity Constraints (MPCC). MPCCs lack strong constraint qualifications. For an introduction to this challenging problem class we refer to [74, 132, 136] for the finite dimensional case and more generally to [101, ch. 3] and the references therein. Other single level reduction techniques are to view the lower level problem as a parametric optimization problem and to make use of its solution operator (if the lower level problem has a unique solution for each choice of upper level variables), see, e. g. [115] and also Section 6.3.2, or to express the demand for optimality on the lower level by introducing an additional constraint which incorporates the so-called optimal value function of the lower level problem, see, e. g. [115, 116]. We remark, however, that dealing with a reduced single level problem instead of the original bilevel problem does not automatically relieve us from repeatedly solving the lower level problem in the course of the solution process. Therefore, we distinguish between methods which actually retain the bilevel structure in the sense that for a change of upper level decision variables, the lower level problem has to be solved, and such ones in which this is not necessary. Following [71], we refer to the latter ones as simultaneous solution approaches.

In [71, sec. 4.4] the author gives an overview on solution approaches to bilevel OCPs. Besides the treatment of the bilevel structure, the methods stated therein differ in the treatment of the lower level OCP, the respective discretization method, the upper level treatment, and the methods for solving the resulting discretized problem. Examples for employing simultaneous solution approaches can be found in [4, 71, 85]. Conversely, the bilevel structure is retained in [52, 53] and [106], meaning that for every evaluation of the upper level objective function the parametric lower level problem is solved. In [106], a derivative-free method is used, while in [52] resp. [53] the author(s) make(s) use of a gradient-based method in which the gradient of the upper level objective function is computed by means of a so-called sensitivity analysis for the lower level problems.

6.2 Training Approach

In this chapter, we present a mathematical approach for predicting a worst possible treatment outcome for a CP patient who undergoes an orthopedic intervention. The quality of a treatment is measured on the basis of the affected patient's gait which changes due to the intervention. We model the CP gait as a solution of a parametric

OCP and an intervention as change of parameters $\mathbf{p} \in \mathbb{R}^{n_p}$, see Chapter 4. However, in our scenario the parameters which represent the treatment realization suffer from uncertainty that is modeled by means of an uncertainty set $\Omega_{\mathbf{p}} \ni \mathbf{p}$. Solving an OCP which includes uncertain parameters and the treatment of worst-case scenarios seems to be connected to robust optimization at first glance. In this section, we present two scenarios and corresponding modeling approaches for predicting a worst possible treatment outcome after intervention (in terms of the resulting gait): a classical approach from the field of robust optimization and our new so-called Training Approach. Furthermore, we justify the suitability of the latter one for the considered application.

From the Optimal Control perspective, two "choices" influence the gait pattern of a patient. On the one hand, the patient chooses a control function inducing the movement of the body (e. g., muscle excitations or torques) and controllable parameters (e. g., phase durations or initial values of the differential states). On the other hand, the "choice" resp. realization of the uncertain parameter $\mathbf{p}$ affects the resulting gait. The difference between a classical approach to robust optimization and our Training Approach is whether the patient has prior knowledge about the parameter realization or not when choosing the control function and the controllable parameters. We examine the difference in the following.

A Classical Approach to Robust Optimization

First, we consider a classical robust optimization approach with deterministically modeled uncertainty and without feedback, as in [40]. From a robust optimization perspective, the patient has no prior knowledge about the realization of $\mathbf{p} \in \Omega_{\mathbf{p}}$ when choosing the control function and the controllable parameters. Therefore, the patient needs to take into account all possible values of $\mathbf{p}$. Throughout this thesis we assume that human gaits are optimal, cf. Assumption 4.1. Hence, the patient will choose a control function and controllable parameters which optimize the worst case, i. e., the worst possible objective function value that can occur due to the parameter uncertainty. Furthermore, for all $\mathbf{p} \in \Omega_{\mathbf{p}}$ the resulting gait patterns have to be feasible.

We consider a parametric OCP with controllable parameters $\mathbf{u}$, control function $\mathbf{u}(\cdot)$, uncertain parameters $\mathbf{p} \in \Omega_{\mathbf{p}}$, differential states $\mathbf{x}(\cdot;\mathbf{p})$, a parameter-dependent set $\mathcal{F}(\mathbf{p})$ of feasible controllable parameters, controls, and differential states, and an objective function $\Phi(\cdot)$ of Mayer-type. Here, we assume that the differential states are

uniquely determined by given $\mathfrak{u}$, $\mathbf{u}(\cdot)$, and $\mathbf{p}$. In the present approach, $\mathfrak{u}$ and $\mathbf{u}(\cdot)$ are chosen independently from $\mathbf{p}$. If $\big(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p})\big) \in \mathcal{F}(\mathbf{p})$, then $\mathbf{x}(\cdot\,;\mathbf{p})$ denotes the differential states which are determined by $\mathfrak{u}$, $\mathbf{u}(\cdot)$, and $\mathbf{p}$. Due to the Optimality Assumption 4.1, the choice of the patient, i. e. controllable parameters and control function, is then given by a solution of the problem

$$\min_{\substack{\mathfrak{u},\mathbf{u}(\cdot),\\ \mathbf{p},\mathbf{x}(\cdot\,;\mathbf{p})}} \quad \Phi\big(\mathbf{x}(1;\mathbf{p})\big) \tag{6.1a}$$

$$\text{s.t. } \big(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p}')\big) \in \mathcal{F}(\mathbf{p}') \text{ for all } \mathbf{p}' \in \Omega_{\mathbf{p}}\text{, and} \tag{6.1b}$$

$$\big(\mathbf{p}, \mathbf{x}(\cdot\,;\mathbf{p})\big) \text{ globally solve} \tag{6.1c}$$

$$\max_{\mathbf{p}\in\Omega_{\mathbf{p}},\mathbf{x}(\cdot\,;\mathbf{p})} \Phi\big(\mathbf{x}(1;\mathbf{p})\big) \tag{6.1d}$$

$$\text{s.t.} \quad \big(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p})\big) \in \mathcal{F}(\mathbf{p})\,, \tag{6.1e}$$

which we abbreviate by

$$\min_{\mathfrak{u},\mathbf{u}(\cdot)} \max_{\substack{\mathbf{p}\in\Omega_{\mathbf{p}},\\ \mathbf{x}(\cdot\,;\mathbf{p})}} \quad \Phi\big(\mathbf{x}(1;\mathbf{p})\big) \tag{6.2a}$$

$$\text{s.t. } \big(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p}')\big) \in \mathcal{F}(\mathbf{p}') \text{ for all } \mathbf{p}' \in \Omega_{\mathbf{p}}\,. \tag{6.2b}$$

Here, we normalize the duration of the process to 1 w. l. o. g. For given $\mathfrak{u}$ and $\mathbf{u}(\cdot)$ we view the value of the objective function as a measure for the benignancy of a parameter realization. Thus, for any choice of $\mathfrak{u}$ and $\mathbf{u}(\cdot)$ a worst possible parameter realization is given by a global solution of the Lower Level Problem (6.1d-6.1e). A feasible choice of $\mathfrak{u}$ and $\mathbf{u}(\cdot)$ takes into account all possible parameter realizations in the sense that it yields feasible trajectories for all $\mathbf{p} \in \Omega_{\mathbf{p}}$, see (6.1b).

For modeling the worst possible post-operative gait, this approach has shortcomings since it does not take into account two important components, namely feedback and training. First, human walking is a process with feedback. In the real world, a patient does not set up a control strategy (e. g., a time history of muscle excitations or torques) in advance and implements this strategy independently from the course of the resulting walking process. Instead, the patient gets feedback from the sensory system and reacts to unfavorable movements by an adaption of the control strategy. Therefore, a robustified Optimal Control model (in the above sense) with feedback would be more a reasonable way for modeling a patient's gait shortly after treatment when the patient's locomotor system has not yet adapted to the anatomical changes of the body and uncertainty is still present, accordingly. Second, we are interested

in the post-operative gait *after* functional adaption of the patient to the changes, i. e., at a point in time when the patient underwent a training period and therefore is able to make optimal use of the altered anatomy. In particular and mathematically speaking, after the training period uncertainty is *not* present anymore in the choice of the controllable parameters $\mathbf{u}$ and the control function $\mathbf{u}(\cdot)$.

Training Approach

Because of the previously mentioned shortcomings, we develop a different approach for the worst-case prediction of the post-operative gait. In the real world, during an intervention a certain, but a priori unknown, parameter $\mathbf{p} \in \Omega_{\mathbf{p}}$ is realized. What follows is a training period in which the affected patient adapts functionally to the performed anatomical adjustment. After the training period, uncertainty is not present anymore in gait of the patient in the sense that the patient "knows" the value of the parameter realization due to the training and is able to react to it in an optimal manner. Hence, the post-operative gait can be modeled as a solution of the parametric OCP with an appropriate parameter value $\mathbf{p}$. For simplification of presentation, we assume the OCP to be uniquely solvable (i. e., to have exactly one local solution) for all parameter realizations. However, we cannot expect this assumption to be valid in practice, and we discuss the practical handling of this issue in Section 6.4. Using the notation from the previous section, the worst possible interventions can now be modeled as the global solutions of the problem

$$\max_{\substack{\mathbf{p}\in\Omega_{\mathbf{p}},\mathbf{u},\\ \mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p})}} \quad \Phi(\mathbf{x}(1;\mathbf{p})) \tag{6.3a}$$

$$\text{s.t.} \quad \big(\mathbf{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p})\big) \text{ solve} \tag{6.3b}$$

$$\min_{\substack{\mathbf{u},\mathbf{u}(\cdot),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \tag{6.3c}$$

$$\text{s.t.} \quad \big(\mathbf{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p})\big) \in \mathcal{F}(\mathbf{p})\,, \tag{6.3d}$$

if the value of the objective function is a measure for the quality of a gait. We abbreviate Problem (6.3) by

$$\max_{\mathbf{p}\in\Omega_{\mathbf{p}}} \min_{\substack{\mathbf{u},\mathbf{u}(\cdot),\\ \mathbf{x}(\cdot;\mathbf{p})}} \quad \Phi\big(\mathbf{x}(1;\mathbf{p})\big) \tag{6.4a}$$

$$\text{s.t.} \quad \big(\mathbf{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p})\big) \in \mathcal{F}(\mathbf{p})\,. \tag{6.4b}$$

For any parameter realization $\mathbf{p}$, the associated post-operative gait after training is modeled as the solution of the lower level resp. inner OCP. Due to the assumed training period, we call this kind of worst-case modeling *Training Approach*.

If the quality of a gait, or more generally the success of a treatment, can be measured by a more general assessment function $\varphi\left(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p}), \mathbf{p}\right)$ than the objective function $\Phi(\cdot)$ of the lower level problem, the worst possible treatments can be modeled as the global solutions of the problem

$$\max_{\substack{\mathbf{p}\in\Omega_{\mathbf{p}},\mathfrak{u},\\ \mathbf{u}(\cdot),\mathbf{x}(\cdot\,;\mathbf{p})}} \quad \varphi\left(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p}), \mathbf{p}\right) \tag{6.5a}$$

$$\text{s.t.} \quad \left(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p})\right) \text{ solve} \tag{6.5b}$$

$$\min_{\substack{\mathfrak{u},\mathbf{u}(\cdot),\\ \mathbf{x}(\cdot\,;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \tag{6.5c}$$

$$\text{s.t.} \quad \left(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p})\right) \in \mathcal{F}(\mathbf{p})\,. \tag{6.5d}$$

However, it is still an open question how to assess the success of treatments in CP. Criteria like, e. g., improved stability are only a conjecture. For this reason, we stick to the Modeling Approach (6.3) in the following, unless stated otherwise.

In case the lower level OCP which models the gait has more than one solution, in the Problems (6.3), (6.4), and (6.5) we have to consider the particular solution $\left(\mathfrak{u}, \mathbf{u}(\cdot), \mathbf{x}(\cdot\,;\mathbf{p})\right)$ of the lower level problem which models the actual establishing post-operative gait. As mentioned before, we discuss the handling of this issue in practice in Section 6.4.

Furthermore, we remark that we focus on patients which are able to perform voluntary, i. e. intended, movements in our considerations. This represents the majority of CP patients, see Section 3.2. If additionally we want to consider patients with involuntary movements, taking the Training Approach is not reasonable anymore as the basic assumption that a patient will be able to choose optimal $\mathfrak{u}$ and $\mathbf{u}(\cdot)$ for any parameter realization after a sufficiently long training period is not valid. Instead, in particular we have to deal with perturbations of the control function $\mathbf{u}(\cdot)$. In this situation, we propose to pursue a robust optimization approach with feedback.

6.3 Training Approach vs. Classical Approach

In this section, we compare the robustification approach from Section 6.2, to which we refer as *classical approach* in the following, to our Training Approach to demonstrate the difference in terms of the objective function values and the feasible sets. We will see that a solution obtained from the classical approach is worse than a one we receive from the Training Approach in terms of the respective objective function values. Furthermore, we conduct a case study to illustrate the fundamental difference between both approaches.

6.3.1 Comparison of Objective Function Values

For the comparability of the objective function values of the min max and max min problems, in this section we solely consider global optima. However, for the OCPs we investigate in Section 6.3.2 this is not a restriction as we will see later. We start our comparison with the following

Remark 6.1
Let $\Omega_{\mathbf{x}} \subset \mathbb{R}^{n_x}$ and $\Omega_{\mathbf{p}} \subset \mathbb{R}^{n_p}$ be compact subsets and $f : \Omega_{\mathbf{x}} \times \Omega_{\mathbf{p}} \to \mathbb{R}$ a continuous function. Then we have

$$\max_{\mathbf{p}\in\Omega_{\mathbf{p}}} \min_{\mathbf{x}\in\Omega_{\mathbf{x}}} f(\mathbf{x},\mathbf{p}) \leq \min_{\mathbf{x}\in\Omega_{\mathbf{x}}} \max_{\mathbf{p}\in\Omega_{\mathbf{p}}} f(\mathbf{x},\mathbf{p}) .$$

Proof See Appendix B.2.1. □

In the above remark, the optimal objective function value of the max min problem overestimates the one of the min max problem. It is easy to find examples in which the gap is indeed greater than zero. For instance, let $\Omega_x = [-5,5]$, $\Omega_p = [-1,1]$ and consider the function

$$f : \Omega_x \times \Omega_p \to \mathbb{R}, \ (x,p) \mapsto (x-p)^2 + p .$$

Then

$$\max_{p\in\Omega_p} \min_{x\in\Omega_x} f(x,p) = 1 < \tfrac{5}{4} = \min_{x\in\Omega_x} \max_{p\in\Omega_p} f(x,p) ,$$

see Appendix B.2.2 for a computation.

Remark 6.1 can be extended to OCPs. To this end, we consider a parametric OCP of the form

$$\min_{\mathfrak{u},\mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p})} \quad \Phi(\mathbf{x}(1;\mathbf{p})) \tag{6.6a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t;\mathbf{p}) = \mathbf{f}(\mathbf{x}(t;\mathbf{p}),\mathbf{u}(t)), \qquad t \in [0,1], \tag{6.6b}$$

$$\mathbf{x}(0;\mathbf{p}) = \mathbf{x}_0(\mathfrak{u},\mathbf{p}), \tag{6.6c}$$

$$\mathbf{0} \leq \mathbf{c}(\mathbf{x}(t;\mathbf{p}),\mathbf{u}(t)), \qquad t \in [0,1], \tag{6.6d}$$

$$\mathbf{0} \leq \mathbf{r}(\mathbf{x}(0;\mathbf{p}),\mathbf{x}(1;\mathbf{p})), \tag{6.6e}$$

$$\mathfrak{u} \in \mathcal{P} \subseteq \mathbb{R}^{\mathfrak{n}_\mathfrak{u}}, \tag{6.6f}$$

$$\mathbf{u}(t) \in \mathcal{U} \subseteq \mathbb{R}^{n_u}, \qquad t \in [0,1], \tag{6.6g}$$

with controllable parameters $\mathfrak{u} \in \mathcal{P} \subseteq \mathbb{R}^{\mathfrak{n}_\mathfrak{u}}$, control functions $\mathbf{u} : [0,1] \to \mathcal{U} \subseteq \mathbb{R}^{n_u}$, non-controllable parameters $\mathbf{p} \in \Omega_\mathbf{p} \subset \mathbb{R}^{n_\mathbf{p}}$, and differential states $\mathbf{x} : [0,1] \to \mathbb{R}^{n_x}$ which we assume to be uniquely determined by the choices of $\mathfrak{u}$, $\mathbf{u}(\cdot)$, and $\mathbf{p}$. As all occurring parameters can be regarded as constant differential states with appropriate initial value in general, limiting the influence of the parameters to the initial values of the differential states in not a restriction, see Section 2.3.3. For a given pair $(\mathfrak{u},\mathbf{u}(\cdot))$ the solution of the parametric Initial Value Problem (IVP) (6.6b-6.6c) is denoted by $\mathbf{x}(\cdot;\mathbf{p})$.

Let

$$\mathcal{C}(\mathbf{p}) \overset{\text{def}}{=} \left\{ (\mathfrak{u},\mathbf{u}(\cdot)) \;\middle|\; \begin{array}{l} \mathbf{0} \leq \mathbf{c}(\mathbf{x}(t;\mathbf{p}),\mathbf{u}(t)) \text{ for } t \in [0,1], \text{ and} \\ \mathbf{0} \leq \mathbf{r}(\mathbf{x}(0;\mathbf{p}),\mathbf{x}(1;\mathbf{p})), \text{ and} \\ \mathfrak{u} \in \mathcal{P}, \text{ and} \\ \mathbf{u}(t) \in \mathcal{U} \text{ for } t \in [0,1] \end{array} \right\}$$

be the set of feasible controllable parameters and control functions for Problem (6.6) for a given $\mathbf{p} \in \Omega_\mathbf{p}$, and

$$\tilde{\mathcal{C}}(\Omega_\mathbf{p}) \overset{\text{def}}{=} \bigcap_{\mathbf{p}\in\Omega_\mathbf{p}} \mathcal{C}(\mathbf{p}).$$

Robustifying Problem (6.6) according to the classical approach leads to the problem

$$\min_{(\mathfrak{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})} \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p}, \\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi\left(\mathbf{x}(1;\mathbf{p})\right). \tag{6.7}$$

For $\Omega'_\mathbf{p} \subseteq \Omega_\mathbf{p}$ we get $\tilde{\mathcal{C}}(\Omega_\mathbf{p}) \subseteq \tilde{\mathcal{C}}(\Omega'_\mathbf{p})$. Hence, the set of feasible control functions and controllable parameters cannot increase but potentially shrinks as the uncertainty

set grows. In Section 6.3.2 we will see an example where the feasible set is empty if the uncertainty set gets too large.

On the other hand, robustifying Problem (6.6) using the Training Approach yields the problem

$$\max_{\mathbf{p}\in\Omega_\mathbf{p}} \min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\mathcal{C}(\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi\left(\mathbf{x}(1;\mathbf{p})\right). \tag{6.8}$$

For the optimal solution values of the objective functions we get the following result which can seen as an analogon of Remark 6.1:

Proposition 6.2
Assume, that the extremal values

$$\min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\mathcal{C}(\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \quad \text{for all } \mathbf{p}\in\Omega_\mathbf{p}, \qquad \max_{\mathbf{p}\in\Omega_\mathbf{p}} \min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\mathcal{C}(\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})),$$

and

$$\max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \quad \text{for all } (\mathbf{u},\mathbf{u})\in\tilde{\mathcal{C}}(\Omega_\mathbf{p}), \qquad \min_{(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})} \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p}))$$

exist. Then we have

$$\max_{\mathbf{p}\in\Omega_\mathbf{p}} \min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\mathcal{C}(\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \le \min_{(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})} \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})).$$

Proof See Appendix B.2.3. □

Proposition 6.2 states that a solution we obtain from the classical approach is worse than a one we receive from the Training Approach in terms of the objective function values. As an application, we consider worst-case treatment planning for CP. Taking the objective function value obtained from the classical robustification approach as a measure for the worst expected treatment outcome therefore overestimates the measure for the actual worst-case outcome which is modeled by the Training Approach. This way, clinical decision makers could be prevented from recommending interventions which are in fact beneficial.

6.3.2 Case Study: State Constrained Rocket Car

To illustrate the fundamental difference between the classical robustification approach and the Training Approach, we consider a test case for which we investigate the difference between the robustification approaches in detail. In particular, we focus on the objective function values and the feasible sets. We consider a so-called "rocket-car" example with state constraints – the one-dimensional movement of a mass point under the influence of some constant deceleration, e. g. modeling headwind or sliding friction, which can accelerate and decelerate in order to reach a desired position. The mass of the car is normalized to 1 and the constant deceleration enters the model in form of an unknown parameter $p \in \mathbb{R}$ suffering from uncertainty, $p \in \Omega_p \subset \mathbb{R}$ with convex and compact uncertainty set Ω_p. Furthermore, we demand the velocity of the car to be bounded from above. We consider a problem in which the rocket car shall reach a final feasible position and velocity in minimum time:

$$\min_{T,u(\cdot),\mathbf{x}(\cdot;p)} \quad T \tag{6.9a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t;p) = T\begin{pmatrix}\mathbf{x}_2(t;p)\\ u(t)-p\end{pmatrix}, \quad t\in[0,1], \tag{6.9b}$$

$$\mathbf{x}(0;p) = \mathbf{0}, \tag{6.9c}$$

$$\mathbf{x}_2(t;p) \le 4, \qquad t\in[0,1], \tag{6.9d}$$

$$\mathbf{x}_1(1;p) \ge 10, \tag{6.9e}$$

$$\mathbf{x}_2(1;p) \le 0, \tag{6.9f}$$

$$T \ge 0, \tag{6.9g}$$

$$u(t) \in [-10,10], \qquad t\in[0,1]. \tag{6.9h}$$

In this time-transformed problem, the (time-transformed) position of the rocket car is encoded in $\mathbf{x}_1(\cdot;p)$, the (time-transformed) velocity in $\mathbf{x}_2(\cdot;p)$, and the (time-transformed) controlled acceleration and deceleration in $u(\cdot)$. The decision variables in the problem are the controllable parameter T, which encodes the process duration of the corresponding problem with free end time, and the control function $u : [0,1] \to \mathbb{R}$, while $\mathbf{x}(\cdot;p)$ is a dependent variable, uniquely determined by $T, u(\cdot)$, and p.

Robustifying Problem (6.9) using the classical approach yields

$$\min_{T,u(\cdot)} \max_{\substack{p\in\Omega_p,\\ \mathbf{x}(\cdot,p)}} \quad T \tag{6.10a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t;p) = T\begin{pmatrix}\mathbf{x}_2(t;p)\\ u(t)-p\end{pmatrix}, \quad t\in[0,1], \tag{6.10b}$$

$$\mathbf{x}(0;p) = \mathbf{0}, \tag{6.10c}$$

$$\mathbf{x}_2(t;p) \le 4, \qquad t\in[0,1], \text{ for all } p\in\Omega_p, \tag{6.10d}$$

$$\mathbf{x}_1(1;p) \ge 10, \qquad \text{for all } p\in\Omega_p, \tag{6.10e}$$

$$\mathbf{x}_2(1;p) \le 0, \qquad \text{for all } p\in\Omega_p, \tag{6.10f}$$

$$T \ge 0, \tag{6.10g}$$

$$u(t)\in[-10,10], \qquad t\in[0,1]. \tag{6.10h}$$

In this scenario, the set of feasible controllable parameters and control functions is given by those T and $u(\cdot)$ which yield feasible trajectories $\mathbf{x}(\cdot,p)$ for *all* $p\in\Omega_p$. The value of the objective function of the inner optimization problem does not depend on p and $\mathbf{x}(\cdot;p)$. A resulting optimal strategy is interesting, if the driver of the rocket car has no prior knowledge about the value of the parameter p and gets no feedback during the process, i. e., has to set up the driving strategy in advance.

In contrast to the classical approach, in the Training Approach we assume that the driver of the rocket car is able to perform optimally for every given p because of a preceding training period. Thus, the worst possible optimal performance is given by a solution of the problem

$$\max_{\substack{p\in\Omega_p,T,\\ u(\cdot),\mathbf{x}(\cdot;p)}} \quad T \tag{6.11a}$$

$$\text{s.t. } \big(T,u(\cdot),\mathbf{x}(\cdot;p)\big) \text{ solve Problem (6.9) for } p. \tag{6.11b}$$

In the following, we present the solutions of the Problems (6.9) (nominal problem), (6.10) (classically robustified problem), and (6.11) (Training Approach robustified problem), compare the resulting objective function values, and comment on the non-emptiness of the feasible sets for both robustified problems. We choose the uncertainty set

$$p\in\Omega_p = [p_l, p_u] \subseteq [0,9]$$

with $p_l < p_u$ to refrain from cumbersome case distinctions, cf. Remark 6.3.

Nominal Problem

First, we consider Problem (6.9) with $p \in [0,9]$. The optimization variables are

$$\left(T, u(\cdot), \mathbf{x}(\cdot\,; p)\right) \in \mathbb{R} \times L^\infty([0,1],\mathbb{R}) \times W^{1,\infty}\left([0,1],\mathbb{R}^2\right),$$

see Section 2.1 for the normed spaces. For proofs of the following statements, see Appendix B.2.4. Problem (6.9) has a unique global solution, and no further local solutions exist. The optimal controllable parameter is given by

$$T^* = T^*(p) = 2.5 + \frac{40}{100 - p^2}, \tag{6.12}$$

and the optimal control function $u^*(\cdot)$ $(= u^*(\cdot\,; p))$ by

$$u^*(t) = \begin{cases} 10 & \text{for } 0 \le t < \frac{4}{(10-p)\,T^*}, \\ p & \text{for } \frac{4}{(10-p)\,T^*} \le t < 1 - \frac{4}{(10+p)\,T^*}, \\ -10 & \text{for } 1 - \frac{4}{(10+p)\,T^*} \le t \le 1. \end{cases}$$

In words, we accelerate as strongly as possible until $\mathbf{x}_2^*(t;p) = 4$, then keep $\mathbf{x}_2^*(t;p)$ constant for a certain period of time, and eventually decelerate as strongly as possible, where $\mathbf{x}^*(\cdot\,;p)$ denotes the differential states which are determined by T^*, $u^*(\cdot)$, and p. The optimal differential states $\mathbf{x}^*(\cdot\,;p)$ and the optimal control function $u^*(\cdot)$ are illustrated in Fig. 6.1, and the dependence of the optimal objective function value $T^*(p)$ on p in Fig. 6.2.

Remark 6.3
For $10 > p > 2\sqrt{21}$ (> 9) the optimal strategy has to be adapted. In this case, the optimal objective function value is given by $T^*(p) = \frac{20}{\sqrt{100-p^2}}$ and the optimal control function is a Bang-Bang control, i. e. $u(t) \in \{-10, 10\}$ for $t \in [0,1]$, which takes its maximum value in the beginning and its minimum value in the remainder of the process. △

A sketch of a proof of Remark 6.3 can be found in Appendix B.2.5.

Training Approach

Let $\Omega_p = [p_l, p_u] \subseteq [0,9]$ with $p_l < p_u$. We consider Problem (6.11). As the nominal problem has a (unique) solution for each $p \in \Omega_p$, the feasible set of Problem (6.11) is

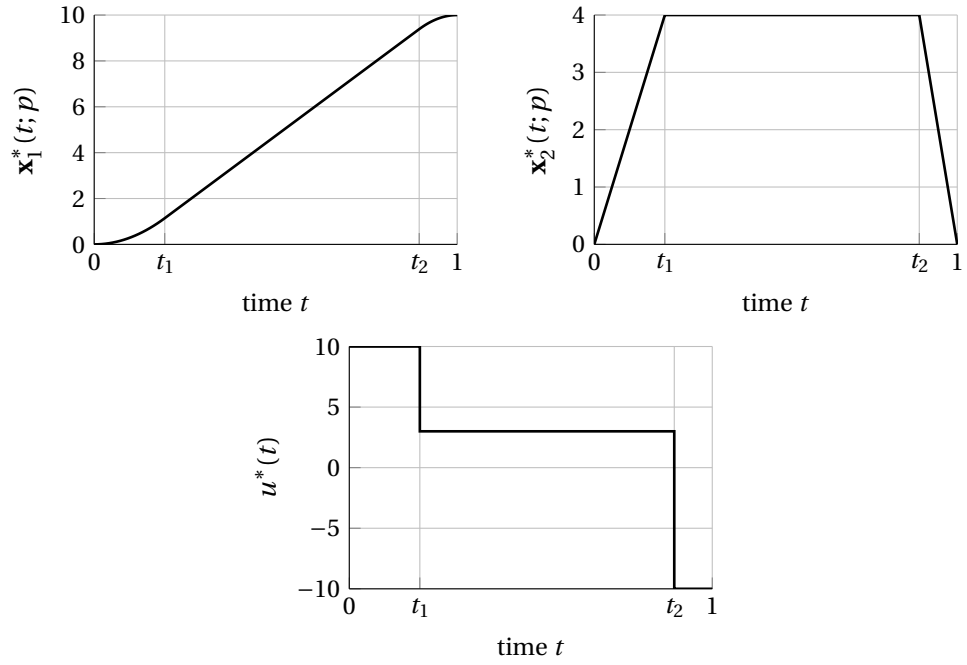


Figure 6.1: Optimal state trajectories $\mathbf{x}_1^*(\cdot;p), \mathbf{x}_2^*(\cdot;p)$, and control function $u^*(\cdot)$ of Problem (6.9) for $p = 3$. We have $t_1 = \frac{4}{(10-p)T^*}$ and $t_2 = 1 - \frac{4}{(10+p)T^*}$.

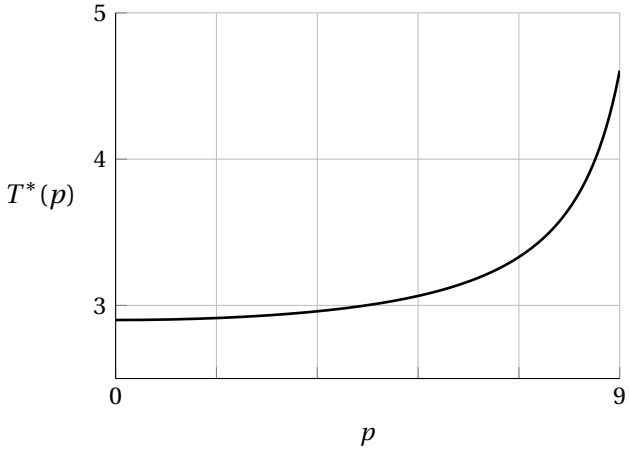


Figure 6.2: Dependence of the optimal objective function value $T^*(p)$ of Problem (6.9) on p.

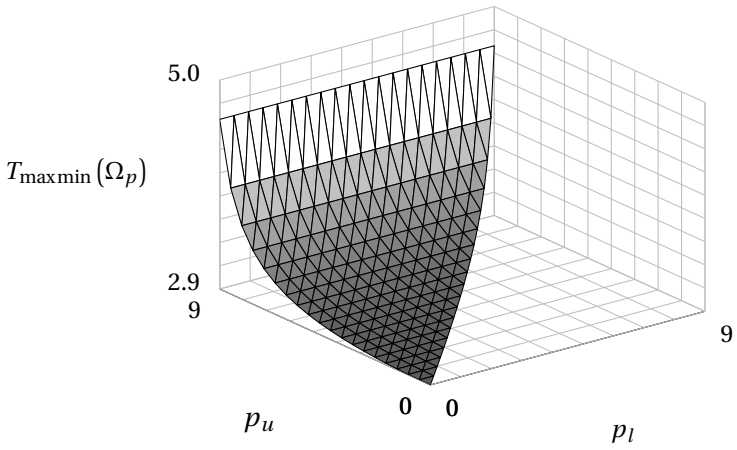


Figure 6.3: Maxmin worst-case value of the objective function of Problem (6.11) for all pairs $(p_l, p_u) \in [0,9]^2$ with $p_u \geq p_l$.

non-empty. Due to (6.12), the optimal objective function value of Problem (6.11) is given by

$$T_{\text{maxmin}}(\Omega_p) = T^*(p_u) = 2.5 + \frac{40}{100 - p_u^2},$$

and its dependence on Ω_p is depicted in Fig. 6.3. Hence, p_u solves the upper level problem and the corresponding solution of the lower level problem is given by the solution of Problem (6.9) for $p = p_u$.

Classical Approach

Let again $\Omega_p = [p_l, p_u] \subseteq [0,9]$ with $p_l < p_u$. We consider Problem (6.10). The optimization variables are given by

$$\left(T, u(\cdot), p, \mathbf{x}(\cdot\,; p)\right) \in \mathbb{R} \times L^\infty([0,1], \mathbb{R}) \times \mathbb{R} \times W^{1,\infty}\left([0,1], \mathbb{R}^2\right),$$

see Section 2.1 for the normed spaces. For proofs of the following statements, see Appendix B.2.6. In contrast to the Training Approach, for the classical approach the feasible set is empty if $\Omega_p = [p_l, p_u]$ becomes too large. More specifically, the feasible set is non-empty if and only if

$$p_u \leq p_l + \frac{8}{10 + \frac{8}{10 - p_l} + \frac{8}{10 + p_l}}.$$

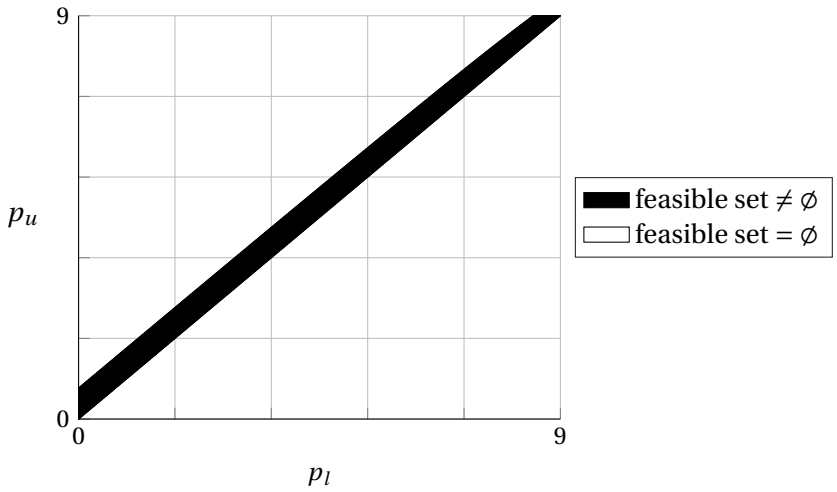


Figure 6.4: Non-emptiness of the feasible set of Problem (6.10) depending on $\Omega_p = [p_l, p_u] \subseteq [0,9]$. For a pair (p_l, p_u), the corresponding point in the graph is colored black if the feasible set is non-empty and white otherwise.

The non-emptiness of the feasible set depending on Ω_p is depicted in Fig. 6.4.

Now let the feasible set be non-empty. We can show that the optimal T^* $(= T^*(\Omega_p))$ and $u^*(\cdot)$ $(= u^*(\cdot;\Omega_p))$ are uniquely determined. Let $\mathbf{x}^*(\cdot;p)$ denote the differential states which are determined by T^*, $u^*(\cdot)$, and p. Then, the global solutions of Problem (6.10) are given by

$$\left\{\left(T^*, u^*(\cdot), p, \mathbf{x}^*(\cdot;p)\right) \,\middle|\, p \in \Omega_p\right\}.$$

Furthermore, every local solution of Problem (6.10) is a global solution.

It remains to state the optimal T^* and $u^*(\cdot)$. Let $g : \mathbb{R} \to \mathbb{R}$ be given by

$$g\left(y;\Omega_p\right) = -\frac{1}{2}(p_u - p_l)\,y^2 + 4y - \frac{8}{10-p_l} - \frac{8}{10+p_l} - 10.$$

If the feasible set is non-empty, the (globally) optimal objective function value $T^* = T^*\left(\Omega_p\right)$ of Problem (6.10) is given by

$$T^*\left(\Omega_p\right) = T_{\min\max}\left(\Omega_p\right) = \min\left\{y \in \mathbb{R} \,|\, g\left(y;\Omega_p\right) = 0\right\},$$

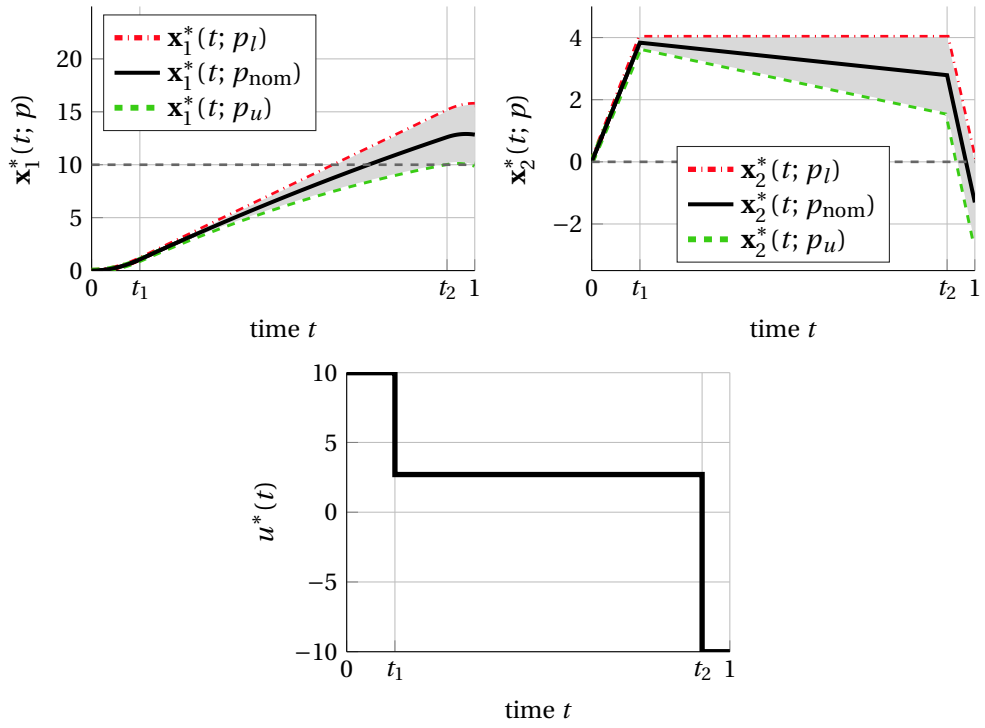


Figure 6.5: Optimal state trajectories $\mathbf{x}_1^*(\cdot;p), \mathbf{x}_2^*(\cdot;p)$ and optimal control function $u^*(\cdot)$ of Problem (6.10) for $p \in \Omega_p = [2.7, 3.3] = [p_l, p_u]$. We have $t_1 = \frac{4}{(10-p_l)T^*}$ and $t_2 = 1 - \frac{4}{(10+p_l)T^*}$. For the state trajectories, the black (solid) lines refer to the trajectories belonging to $p_{\text{nom}} = 3$, the red (dash-dotted) lines to $p = p_l$, and the green (dashed) lines to $p = p_u$. The shaded area includes the possible state trajectories for all $p \in \Omega_p$ for the given control law.

and the optimal control function by

$$u^*(t) = \begin{cases} 10 & \text{for } 0 \le t < \frac{4}{(10-p_l)T^*}, \\ p_l & \text{for } \frac{4}{(10-p_l)T^*} \le t < 1 - \frac{4}{(10+p_l)T^*}, \\ -10 & \text{for } 1 - \frac{4}{(10+p_l)T^*} \le t \le 1. \end{cases}$$

In words, we accelerate as strongly as possible until $\mathbf{x}_2^*(t;p_l) = 4$, then keep $\mathbf{x}_2^*(t;p_l)$ constant for a certain period of time, and eventually decelerate as strongly as possible. The solution trajectories for an illustrative solution of Problem (6.10) are depicted in Fig. 6.5, and a graph displaying the dependence of $T_{\min\max}(\Omega_p)$ on Ω_p is given in Fig. 6.6.

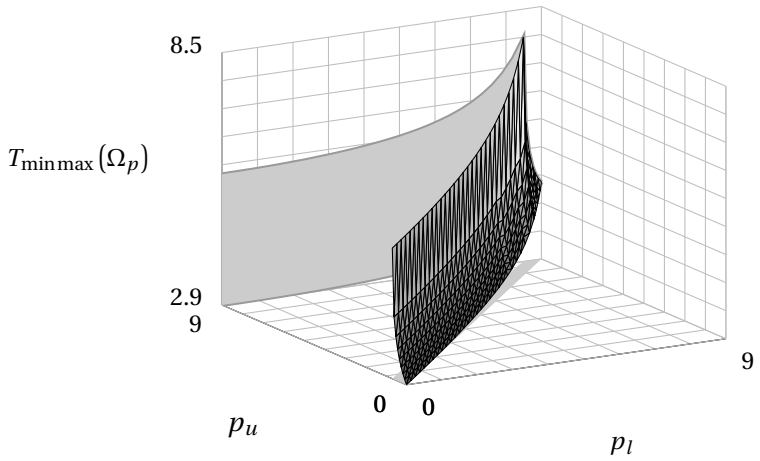


Figure 6.6: Min max optimal value $T_{\min\max}(\Omega_p)$ of the objective function of Problem (6.10) for uncertainty sets $\Omega_p = [p_l, p_u] \subseteq [0,9]$ which yield a non-empty feasible set. The shaded area on the bottom of the graph depicts those pairs (p_l, p_u) for which the feasible set in non-empty. For the sake of a better readability, the shaded area in the plane defined by $p_u = 9$ comprises the projected objective function values.

Comparison of Classically and Training Approach Robustified Solutions

We compare the robust solutions obtained by the classical approach and our Training Approach in case of the rocket-car example and comment on the non-emptiness of the feasible sets of the respective problems. For the Training Approach, the feasible set is non-empty for any non-empty uncertainty set $\Omega_p = [p_l, p_u] \subseteq [0,9]$. This is because the nominal problem has a solution for each $p \in [0,9]$. In contrast, for the classically robustified problem the feasible set is empty if the uncertainty gets too large. This is due to the fact that the controllable parameter T and the control function $u(\cdot)$ have to be chosen independently from the realization of the uncertain parameter $p \in \Omega_p$. A feasible choice of T and $u(\cdot)$ has to ensure that the parametric trajectory $\mathbf{x}(\cdot\,; p)$ satisfies the Constraints (6.9d-6.9f) for *all* $p \in \Omega_p$. This limitation renders Problem (6.10) infeasible if the uncertainty set becomes too large, see Fig. 6.4. In this case, there are no feasible T and $u(\cdot)$ which can guarantee feasible trajectories $\mathbf{x}(\cdot\,; p)$ for all $p \in \Omega_p$. For a non-empty feasible set, the solution of the classically robustified problem therefore represents the best feasible choices of T and $u(\cdot)$ that ensure feasible trajectories $\mathbf{x}(\cdot\,; p)$ for any parameter realization. In contrast, the solution of the Training Approach robustified Problem (6.11) comprises the best fea-

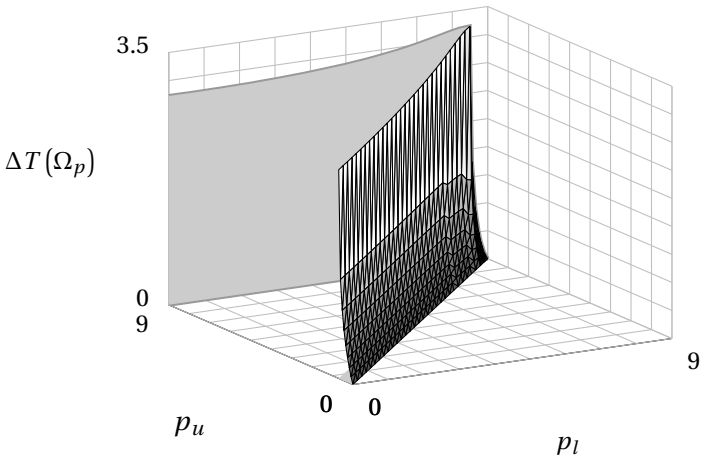


Figure 6.7: Gap $\Delta T(\Omega_p) = T_{\min\max}(\Omega_p) - T_{\max\min}(\Omega_p)$ between the optimal values $T_{\min\max}(\Omega_p)$ and $T_{\max\min}(\Omega_p)$ of the objective functions of Problems (6.10) and (6.11), respectively, for uncertainty sets $\Omega_p = [p_l, p_u] \subseteq [0,9]$ which admit a solution of both problems. For a better readability, the shaded area in the plane defined by $p_u = 9$ comprises the projected values of $\Delta T(\Omega_p)$.

sible choices T and $u(\cdot)$ in response to the given worst possible parameter realization p^*, and feasibility of trajectories is only required for $\mathbf{x}(\cdot; p^*)$.

Next, we consider the objective function values. We have

$$T_{\max\min}(\Omega_p) \leq T_{\min\max}(\Omega_p),$$

in accordance with Proposition 6.2, i. e., the worst-case objective function value obtained by the classical approach, $T_{\min\max}(\Omega_p)$, overestimates the one we receive from the Training Approach, $T_{\max\min}(\Omega_p)$. Fig. 6.7 illustrates the gap between both worst-case objective function values depending on the uncertainty set $\Omega_p = [p_l, p_u]$. The gap increases strongly for a growing difference $p_u - p_l$ and reaches a similar magnitude as the optimal values of the respective objective functions.

6.4 Numerical Solution Approach

In this section, we state a general bilevel OCP for worst-case treatment planning which results from the Training Approach (see Section 6.2), and describe an approach for its numerical solution.

6.4.1 A Bilevel Optimal Control Problem for Worst-Case Treatment Planning

As explained in Chapter 4 (see also [105]), we model the human gait as a solution of a parametric multi-stage OCP of the form

$$\min_{\substack{T_1,\dots,T_n,\\ \mathbf{u}(\cdot),\mathbf{x}(\cdot)}} \quad \Phi^M\left(T_n,\mathbf{x}(T_n),\mathbf{p}\right)+\int_{T_0}^{T_n}\Phi^L\big(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\big)\,\mathrm{d}t \tag{6.13a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}(t)=\mathbf{f}^j\left(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\right), \qquad t\in\mathcal{T}_j, \quad j=1,\dots,n, \tag{6.13b}$$

$$T_{j-1}\le T_j, \qquad j=1,\dots,n, \tag{6.13c}$$

$$\mathbf{x}(T_j^+)=\boldsymbol{\Delta}^j\left(\mathbf{x}(T_j^-),\mathbf{p}\right), \qquad j=1,\dots,n, \tag{6.13d}$$

$$\mathbf{0}\le\mathbf{c}^j\left(\mathbf{x}(t),\mathbf{u}(t),\mathbf{p}\right), \qquad t\in\mathcal{T}_j, \tag{6.13e}$$

$$\mathbf{0}=\mathbf{r}^{\mathrm{eq}}\left(\mathbf{x}(T_0),\dots,\mathbf{x}(T_n),\mathbf{p}\right), \tag{6.13f}$$

$$\mathbf{0}\le\mathbf{r}^{\mathrm{ieq}}\left(\mathbf{x}(T_0),\dots,\mathbf{x}(T_n),\mathbf{p}\right), \tag{6.13g}$$

with $\mathcal{T}_j=[T_{j-1},T_j]$, where the differential equation describes the dynamics of the rigid Multi-Body System (MBS) modeling the human body. The parameters enter the dynamics of the OCP and are interpreted as patient-specific properties which are altered through a medical intervention. In general, the above problem could be replaced by any other suitable OCP modeling the gait. However, multi-stage formulations have proven their usefulness in the context of bilevel optimization and gait modeling, cf. [31, 32, 71], and we stick to this kind of modeling in the following. In particular, for each realization of $\mathbf{p}$ we assume the model phases to coincide.

As described in Section 6.2, an application of the Training Approach yields a bilevel OCP for worst-case treatment planning of the general form

$$\max_{\substack{\mathbf{p}\in\Omega_{\mathbf{p}},T_1,\dots,T_n,\\ \mathbf{u}(\cdot),\mathbf{x}(\cdot)}} \quad \varphi\big(T_1,\dots,T_n,\mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p}),\mathbf{p}\big) \tag{6.14a}$$

$$\text{s.t.} \quad \big(T_1,\dots,T_n,\mathbf{u}(\cdot),\mathbf{x}(\cdot;\mathbf{p})\big) \text{ solves Problem (6.13) for } \mathbf{p}, \tag{6.14b}$$

$$\text{and models the (post-operative) human gait.} \tag{6.14c}$$

Here, for a given $\mathbf{p} \in \Omega_{\mathbf{p}}$ the function $\varphi(\cdot)$ assesses the quality of a treatment outcome by means of $\mathbf{p}$ and the solution of the lower level OCP which models the post-operative gait for the intervention encoded in $\mathbf{p}$. Problem (6.14) allows for the incorporation of arbitrary assessment functions $\varphi(\cdot)$. In practice, if no other meaningful measure for the quality of a treatment outcome is available, we propose to choose $\varphi(\cdot)$ to coincide with the Optimization Criterion (6.13a). Furthermore, we focus on box-shaped uncertainty sets $\Omega_{\mathbf{p}}$, i. e.

$$\Omega_{\mathbf{p}} = \left\{ \mathbf{p} \in \mathbb{R}^{n_p} \;\middle|\; \mathbf{p}_i^l \leq \mathbf{p}_i \leq \mathbf{p}_i^u \text{ for } i = 1, \ldots, n_p \right\} \tag{6.15}$$

for some $\mathbf{p}^l, \mathbf{p}^u \in \mathbb{R}^{n_p}$.

6.4.2 Numerical Solution Approach to Problem (6.14)

We describe an approach for the numerical solution of Problem (6.14). The ultimate goal is to provide a tool for worst-case treatment planning which is routinely applicable in medical practice. In view of this, we seek for a solution approach which is regularly applicable to varying problems of Form (6.14). The efficacy of the presented approach is demonstrated in Section 7.2.

If the Lower Level Problem (6.13) has exactly one solution for each $\mathbf{p} \in \Omega_{\mathbf{p}}$, the constraint (6.14c) is redundant and can be dropped. For the moment, unless stated otherwise we assume that this is the case which relieves us from choosing the particular solution of the lower level problem which corresponds to the actual post-operative gait establishing after intervention.

Treatment of Lower Level Problem and Bilevel Structure

An overview on solution approaches to bilevel optimization problems and in particular to bilevel OCPs can be found in Section 6.1.2. In view of the complexity of the parametric lower level OCP – which includes state and control constraints – we refrain from applying an indirect approach. This relieves us from dealing with a potentially very ill-conditioned multi-point boundary value problem with possibly complex and unknown switching structure (regarding the active inequality constraints) and jumps in the adjoint variables, for which it can be very challenging to determine a sufficiently good initial guess. For difficulties concerning the application of an indirect approach, we refer to [18, sec. 4.3]. Furthermore, it is hardly possible to derive an analytical expression for neither the parametric solution of the lower level problem nor its value function which would relieve us from solving the lower level OCP

numerically. Hence, we tackle the lower level problem with a direct approach.

Replacing the resulting discretized lower level problem by its Karush-Kuhn-Tucker (KKT) conditions would yield a (numerically challenging) MPCC for which we, however, cannot expect equivalence to the original problem. Hence, we retain the bilevel structure of the problem, meaning that for every evaluation of the upper level objective function we solve the lower level problem. In case the lower level problem has more than one solution (see later, Page 118), solving it for each upper level objective evaluation furthermore facilitates to ensure that one does not lose track of the actual gait modeling OCP solution during optimization. For comparability of the obtained solutions of the parametric lower level problem, we employ the same solution approach and the same discretization scheme for each numerical solution of the OCPs. In particular, the resulting vector of optimization variables is of the same dimension for each $\mathbf{p}$, and its components can be similarly interpreted.

Issues Concerning Derivative Generation

Let $\mathbf{y}^*(\mathbf{p}) \in \mathbb{R}^{n_y}$ denote the parameter-dependent solution vector of the discretized lower level OCP. In order to employ a gradient-based solution approach, we have to compute the derivatives $\nabla_{\mathbf{p}}\varphi\left(\mathbf{y}^*(\mathbf{p}),\mathbf{p}\right)$ and $\nabla_{\mathbf{p}}\mathbf{y}^*(\mathbf{p})$. Here, the vector $\mathbf{y}^*(\mathbf{p}) \in \mathbb{R}^{n_y}$ results from a complex iterative solution process and Automatic Differentiation (AD) (see Section 2.5) is not applicable.

We consider the finite differences approach, see Section 2.5. A finite differences approximation of the gradient comes along with a significant loss of precision, cf., e. g., [110, sec. 8.1]. Remember $\Omega_{\mathbf{p}} \subset \mathbb{R}^{n_p}$. For instance, to compute the gradient $\nabla_{\mathbf{p}}\mathbf{y}^*(\mathbf{p})$ with total error $\leq \delta$, we expect that we have to determine either n_p+1 solutions of the discretized lower level OCP with approximate relative error $\leq \delta^2$, or $2n_p$ solutions with approximate relative error $\leq \delta^{\frac{3}{2}}$, depending on the choice of the difference scheme, cf. arguments in [110, sec. 8.1]. In practice, $\delta \leq 10^{-5}$ is a reasonable choice, and the computation of the required OCP solutions with according precisions – if possible – would yield high numerical costs which we consider to be prohibitive, in particular for complex gait models in which $\mathbf{p}$ enters the dynamics. Hence, we refrain from approximating the required gradients by finite differences.

Furthermore, we remark that in general the function which maps $\mathbf{p}$ to $\varphi\left(\mathbf{y}^*(\mathbf{p}),\mathbf{p}\right)$ cannot be assumed to be smooth in $\mathbf{p}$, also if we choose $\varphi\left(\mathbf{y}^*(\mathbf{p}),\mathbf{p}\right)$ to coincide with the optimal value of the lower level objective function (cf., e. g. [116]).

A Derivative-Free Approach

For a general applicability, we propose the usage of a Derivative-Free Optimization (DFO) method. DFO methods are often applied to optimization problems in which the objective function is non-smooth (or more generally, cannot be assumed to be smooth), or the function evaluations suffer from inaccuracies, as in our case. We identify the solution approach proposed in [106] (see also [31]) to be suitable for our problem. We describe the method in the following.

For the optimization of the upper level problem, we use the `BOBYQA` algorithm [124]. We stick to the latter reference and give a concise overview on the algorithm. `BOBYQA` is designed for problems of the form

$$\begin{aligned} &\min_{\mathbf{z}\in\mathbb{R}^n} F(\mathbf{z}) \\ &\text{s.t. } \mathbf{a}_j \le \mathbf{z}_j \le \mathbf{b}_j, \quad i = 1,\dots,n, \end{aligned} \tag{6.16}$$

in which the function $F(\cdot)$ is considered as a black box. As stated before, for a given $\mathbf{p}$ we solve the parametric lower level OCP (6.13) with a direct approach and view the resulting (finite dimensional) solution as dependent variable. Furthermore, the uncertainty set $\Omega_{\mathbf{p}}$ is box-shaped, see (6.15). Hence, `BOBYQA` is applicable to our problem class.

In the method of `BOBYQA`, in each iteration k the objective function is approximated by a sequence of quadratic functions $Q_k(\cdot)$, such that

$$Q_k(\mathbf{z}^{k,i}) = F(\mathbf{z}^{k,i}), \quad i = 1,\dots,m,$$

for interpolation points $\mathbf{z}^{k,i} \in \mathbb{R}^n$. The number of interpolation points m is constant and can be choosen between $n+2$ and $\frac{1}{2}(n+1)(n+2)$. In [124] the author proposes, e. g., the choice $m = 2n+1$. Let $\mathbf{x}_k \in \operatorname{argmin}\left\{Q_k(\mathbf{z}^{k,i}) \,\middle|\, i = 1,\dots,m\right\}$. In every iteration k, by means of the quadratic model one computes a feasible step $\mathbf{d}_k$ which is inside a "trust-region radius" Δ_k, i. e., $\|\mathbf{d}_k\| \le \Delta_k$. Subsequently, the function $F(\cdot)$ is evaluated at $\mathbf{x}_k + \mathbf{d}_k$, one interpolation point $\mathbf{z}^{k,i}$ is replaced by $\mathbf{x}_k + \mathbf{d}_k$, and the quadratic model is updated. The sequence $\mathbf{x}_k$ is expected to approach a solution of Problem (6.16). For details regarding `BOBYQA`, we refer to [124].

After setting up the initial quadratic model, in every iteration one interpolation point is replaced. Consequently, if we apply `BOBYQA` to the Bilevel Problem (6.14), in total $m + k_{\text{end}}$ lower level OCPs have to be solved if k_{end} denotes the number of iterations.

In contrast to the finite differences approximation of the gradients with respect to $\mathbf{p}$, no increased precision of the OCP solutions is required. For the solutions of the OCPs, we use the Multiple Shooting approach [25] (see also Section 2.4.1) together with a Sequential Quadratic Programming (SQP) method, see [94, 95]. The description of the method proposed in [106] is complete.

We remark that `BOBYQA` computes local extrema while in our application, the upper level problem needs to be solved globally. However, we assume that our model behaves benign in the sense that if the uncertainty set is of moderate size, then solely one local maximum exists. Alternatives are the provision of a good initial guess for the global optimum or the use of global optimization routines on the upper level. For the latter, however, increasing computational costs have to be expected.

Possible Non-Unique Solvability of Lower Level Problem

Till now, we assumed that the Lower Level Problem (6.13) has exactly one solution for each $\mathbf{p} \in \Omega_{\mathbf{p}}$. However, in general we cannot expect this assumption to be valid. Thus, we need to ensure that the numerical solver for the lower level OCP finds that local minimum which corresponds to the human gait. For any possible parameter $\mathbf{p}$, let

$$\mathfrak{G}(\mathbf{p}) = \left\{\text{local solutions of Problem (6.13) for } \mathbf{p}\right\}.$$

By $\mathfrak{g}(\mathbf{p}) \in \mathfrak{G}(\mathbf{p})$ we denote the element which describes the actual establishing gait for a given $\mathbf{p}$. We assume the considered parametric OCP to behave benign in the following sense:

Assumption 6.4
Let $\mathbf{p}$ be any parameter with known corresponding gait $\mathfrak{g}(\mathbf{p})$ and $\Delta\mathbf{p}$ a change of moderate size. We consider Problem (6.13) for $\mathbf{p} + \Delta\mathbf{p}$. Then the employed method for solving the problem converges to $\mathfrak{g}(\mathbf{p} + \Delta\mathbf{p})$ if we use $\mathfrak{g}(\mathbf{p})$ as initial guess. △

Under this assumption, for the solution of the Bilevel Problem (6.14) we propose to proceed as follows. Let $\mathbf{p}_{\text{pre}}$ be the parameter before intervention and $\mathbf{p}_{\text{nom}}$ encode the planned (i. e., nominal) parameter value after intervention. In our setting, the gait before intervention $\mathfrak{g}(\mathbf{p}_{\text{pre}})$ is known. We approach $\mathbf{p}_{\text{nom}}$, which can be far away from $\mathbf{p}_{\text{pre}}$, in sufficiently small steps,

$$\mathbf{p}^i = \mathbf{p}^{i-1} + \tfrac{1}{n_s}\left(\mathbf{p}_{\text{nom}} - \mathbf{p}_{\text{pre}}\right), \quad i = 1, \dots, n_s,$$

with $\mathbf{p}^0 = \mathbf{p}_{\text{pre}}$. For each $\mathbf{p}^i$, we solve the corresponding OCP of Form (6.13), where we use the determined solution for $\mathbf{p}^{i-1}$ as initial guess. Then by Assumption 6.4, we iteratively obtain the solutions $\mathfrak{g}(\mathbf{p}^i)$, and finally $\mathfrak{g}(\mathbf{p}_{\text{nom}})$. After determining $\mathfrak{g}(\mathbf{p}_{\text{nom}})$, we start our optimization routine for the solution of Problem (6.14) in $\mathbf{p}_{\text{nom}}$. For uncertainty sets of moderate size, large distances between the evaluation points do not occur. Thus, by Assumption 6.4, for each parameter value $\mathbf{p}^k$ that is investigated throughout the optimization process, the OCP solver finds the lower level solution $\mathfrak{g}(\mathbf{p}^k)$ if we use the previously computed OCP solution $\mathfrak{g}(\mathbf{p}^{k-1})$ as initial guess. Consequently, the finally obtained solution of the bilevel problem indeed encodes a worst possible intervention. For the case that large distances between the evaluation points of the upper level problem occur, we refer to the next paragraph.

Summary and Outlook on Algorithmic Variants

Summing up, for the numerical solution of Problem (6.14) we propose the following procedure. We retain the bilevel structure of the problem and use the DFO method of `BOBYQA` for the solution of the upper level problem. For each evaluation of the upper level problem, the lower level OCP has to be solved. For this, we use the Direct Multiple Shooting approach. Since for each $\mathbf{p}$ the lower level OCP can have more than one solution, we have to take care of selecting the one which corresponds to the actual establishing gait. We rely on Assumption 6.4 to handle this issue during the optimization process. However, large distances are possible between the nominal parameter value $\mathbf{p}_{\text{nom}}$ in which the optimization routine is started and the value $\mathbf{p}_{\text{pre}}$ for which the solution $\mathfrak{g}(\mathbf{p}_{\text{pre}})$ – which corresponds to the pre-operative gait – is assumed to be known. To provide a suitable initial guess at the beginning of the solution procedure, we approach the initial parameter $\mathbf{p}_{\text{nom}}$ using a homotopy and solve a sequence of OCPs.

For the case study which is presented in Section 7.2, the described strategy is suitable and works well. In the following, we propose adaptions of the algorithm which can be beneficial if certain issues arise in future applications.

Choice of DFO method. If the proposed method performs poorly, one reason could be that the method of `BOBYQA` is not a good choice for the treatment of the upper level problem. This can eventuate, e. g., if discontinuities occur in the vicinity of a upper level problem solution, or the local approximation by quadratic models is not useful for other reasons. In this situation, employing a different (appropriate) DFO

method can be beneficial. For an overview on DFO methods, we refer to [125].

Incorporating derivative information. If the proposed method performs poorly and discontinuities are not expected to play a crucial role, a further option is to make use of gradient-based methods. As explained before, standard approaches for derivative generation (see Section 2.5) are not applicable or expected to cause prohibitive numerical costs. However, besides the previously discussed approaches, a more sophisticated approach is to compute the required derivatives by means of a so-called sensitivity analysis for the lower level problem, see, e. g. [53, ch. 5] and [28]. In the described situation, we propose to incorporate this technique.

Large steps in $\mathbf{p}$. As explained before, we cannot assume the lower level OCP to have exactly one solution for each $\mathbf{p} \in \Omega_{\mathbf{p}}$. In particular, if large steps in $\mathbf{p}$ – or more generally, large distances between the current evaluation point and the previous evaluation points of the upper level problem – occur during the solution process (for any choice of method), we have to pay attention that the OCP solver does not lose track, and does not converge or selects a local solution different from the desired $\mathfrak{g}(\mathbf{p})$. This can be important for large uncertainty sets. If the occurring distances between evaluation points are suspected to cause problems, we propose a similar proceeding as for the computation of the nominal gait pattern, whereby again, we rely on the validity of Assumption 6.4. Let $\mathbf{p}_{\text{prev}}$ be the previous evaluation point, for which we assume to know $\mathfrak{g}(\mathbf{p}_{\text{prev}})$, and $\mathbf{p}_{\text{cur}}$ the current evaluation point for which $\mathfrak{g}(\mathbf{p}_{\text{cur}})$ has to be computed. If $\mathbf{p}_{\text{cur}} - \mathbf{p}_{\text{prev}}$ is large, we make use of a homotopy and approach $\mathbf{p}_{\text{cur}}$ in sufficiently small steps $\mathbf{p}^i$, starting at $\mathbf{p}_{\text{prev}}$. In each step, we solve the OCP using the solution from the previous step as initial guess. By Assumption 6.4, this way we finally determine $\mathfrak{g}(\mathbf{p}_{\text{cur}})$, as desired, at the cost of computing additional OCP solutions.

For given $\mathbf{p}_{\text{cur}}$, the parameter $\mathbf{p}_{\text{prev}}$ could either be chosen as the closest parameter value or as the last parameter value for which $\mathfrak{g}(\mathbf{p})$ has been computed in an earlier calculation. If storage is a limitation, the latter can be advantageous.

6.5 Outlook: Application of Training Approach to CP Treatment Planning

In this section, we explain how the Training Approach can be applied for worst-case treatment planning of CP in a real-world scenario. An application of the approach in a case study can be found in Section 7.2 where we consider a fictive CP patient who

is forced into a crouch gait by the disease.

We consider a CP patient who underwent the procedure of a Gait Analysis (GA), see Section 3.4. The treating physicians propose an orthopedic intervention in order to improve the patient's gait. However, the intervention cannot be performed with perfect accuracy. Our goal is to compute a worst possible treatment and the corresponding outcome in view of the occurring uncertainty and the possible resulting post-operative gait patterns. If the worst possible post-operative gaits are still better than the pre-operative one in terms of a given measure, the intervention seems reasonable. In order to reach the stated goal by an application of the Training Approach, several steps have to be carried out.

1. An MBS Model for the Patient's Body

First, we set up a suitable rigid MBS to model the patient's body, cf. Section 4.1.1. In particular, the physical properties of the rigid segments which model the single parts of the body need to be calibrated patient-specifically. For the calibration, some properties (such as physical dimensions) can be measured directly while for others data from literature, e. g. [35], can be taken into account. Furthermore, one can make use of the motion capture data from GA, see [48, ch. 4] and [51].

2. A Parametric OCP for the Pre- and Post-operative Gait

We model the patient's gait as a solution of a parametric OCP subject to the MBS dynamics and further constraints, cf. Section 4.1.2. To model the gait, we set up a multi-stage OCP. Based on the data from GA, appropriate model stages have to be identified, and we need to impose a set of suitable constraints to model the process of walking. Specific attention needs to be paid to the a priori unknown optimization criterion. We propose to use a weighted sum of suitable criteria which needs to be calibrated later. Altogether, so far three kinds of parameters occur in the OCP: treatment parameters $\mathbf{p}$ which represent those properties of the patient-specific model that are altered by the considered orthopedic intervention, parameters $\mathbf{p}_\Phi$ which determine the objective function, and further modeling parameters $\mathbf{p}_\mathrm{M}$. Both $\mathbf{p}_\Phi$ and $\mathbf{p}_\mathrm{M}$ need to calibrated patient-specifically, see Step 3. After calibration, we obtain a parametric OCP of Form (6.13) in the parameter $\mathbf{p}$.

The intervention itself is accordingly modeled as a change of $\mathbf{p}$. The modeling of the intervention is incorporated into the gait model this way, and the calibrated OCP is able to describe the patient's pre-operative *and* post-operative gait, respectively,

depending on the value of $\mathbf{p}$. An example for incorporating the modeling of interventions into the MBS dynamics can be found in Section 4.3.

3. Calibration of the OCP

After setting up the OCP for modeling the gait, it needs to be calibrated patient-specifically – a challenging task. On the one hand, we need to determine the modeling parameters $\mathbf{p}_{\mathrm{M}}$ as well as the value of the treatment parameters $\mathbf{p}$ for the pre-operative situation, which we denote by $\mathbf{p}_{\mathrm{pre}}$. On the other hand, we have to determine $\mathbf{p}_{\Phi}$ in order to identify an objective function which yields the pre-operative gait. First, for the identification of $\mathbf{p}_{\mathrm{M}}$ and $\mathbf{p}_{\mathrm{pre}}$, depending on the way of modeling different techniques can be used. For instance, we propose to use medical examinations in combination with an adapted version of the so-called dynamics reconstruction (cf. [48, sec. 5.3] and [51]), where the latter computes parameters (and controls) which yield an approximation of the data from GA. Second, for the determination of an objective function which generates the patient's gait pattern an IOC approach (cf., e. g., [71, 106]) can be used. An identification of $\mathbf{p}_{\mathrm{M}}$, $\mathbf{p}_{\mathrm{pre}}$, and $\mathbf{p}_{\Phi}$ at the same time by means of an IOC approach could also be beneficial. Furthermore, the IOC approach provides us with an OCP solution modeling the pre-operative gait.

4. Determination of Nominal Treatment Parameter, Uncertainty Set, and Treatment Assessment Criteria

After a successful model calibration we have a suitable parametric OCP at hand and know the OCP solution which corresponds to the pre-operative gait. With the aid of the physicians, we need to identify a nominal post-operative treatment parameter value $\mathbf{p}_{\mathrm{nom}}$ – encoding the intervention according to plan – and a set $\Omega_{\mathbf{p}}$ of possible treatment parameter realizations which represents the uncertainty. Furthermore, an assessment function which quantifies the quality of a gait – in terms of an OCP solution – needs to be provided. If no other meaningful measure for the quality of a treatment outcome is available, we propose to choose the objective function of the OCP modeling the gait. Since we know the OCP solution which models to the pre-operative gait, we can quantify and assess the patient's gait pattern before intervention.

5. Application of Training Approach and Interpretation of Results

Finally, we can apply the Training Approach. To compute a worst possible intervention, the corresponding post-operative gait, as well as the corresponding worst pos-

sible value of the assessment function, we have to find a global solution of a bilevel problem of Form (6.5). If the worst possible post-operative gaits are assessed better than the pre-operative one, the planned intervention seems reasonable despite the considered uncertainty. In the other case, the intervention is not recommendable.

Chapter 7

Case Studies

In this chapter, we demonstrate the usefulness of the approaches from Chapters 5 and 6 by conducting two case studies.

7.1 Optimal Control Problems with Switches, Switching Costs, and Jumps – A Walking Motion

In this section, we set up an actuated rigid Multi-Body System (MBS) and an Optimal Control Problem (OCP) whose solutions describe locally optimal walking-like motions of the MBS. In Appendix A.2, we present a multi-stage OCP whose solutions model gait patterns for the case that number and order of model phases are specified in advance. In contrast, in this section we refrain from using a multi-stage formulation with a predetermined order of phases but use the free-phase approach from Chapter 5.

Parts of the content presented in this section can be found in [134].

7.1.1 The Simplest Walker Model

We consider the 2D "Simplest Walker" model as in [58, 134] and Appendix A.1 – a rigid MBS comprising three point masses connected by massless rods. Individual trait parameters are normalized to 1 kg for the weight of all mass points and 1 m for the length of the rods. We view the MBS as a stick man with two stiff legs and refer to the bottom point masses as feet and to the top point mass as head. The respective time-dependent positions in space are given by $\big(x_h(t), y_h(t)\big)$ (head), $\big(x_l(t), y_l(t)\big)$ (left foot), and $\big(x_r(t), y_r(t)\big)$ (right foot). The rotation of the left and right foot around the head is described by the angles $\varphi_l(t)$ and $\varphi_r(t)$, respectively. An illustration of the simplest walker MBS can be found in Fig. 7.1

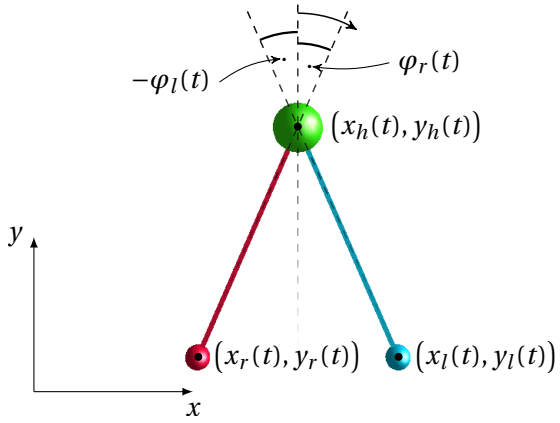


Figure 7.1: The simplest walker modeled by a rigid multi-body system. The right leg is represented by red segments and the left leg by blue segments. Illustration created using `MeshUp` [48].

Table 7.1: Generalized coordinates of the simplest walker MBS.

$\mathbf{q}_1(\cdot) = x_h(\cdot)$	horizontal position of the head
$\mathbf{q}_2(\cdot) = y_h(\cdot)$	vertical position of the head
$\mathbf{q}_3(\cdot) = \varphi_l(\cdot)$	angle describing rotation of left leg around head pivot
$\mathbf{q}_4(\cdot) = \varphi_r(\cdot)$	angle describing rotation of right leg around head pivot

7.1.2 Mode-Dependent Dynamics of the Simplest Walker Model

The dynamics of a general rigid MBS are described in Section 4.1, and the mechanical equations of motion for the simplest walker model during "walking" are derived in Appendix A.1. We give a concise summary and refer to the mentioned sections for more details.

We allow for movements in two dimensions. The MBS has four degrees of freedom, comprising the position of the head in 2D and the respective rotation of the legs around the head pivot. Thus, the system can be described by means of four generalized coordinates, summarized in $\mathbf{q}(t)$, cf. Tab. 7.1, and the corresponding generalized velocities. We include both in the differential states $\mathbf{x}(\cdot)$, which we define by

$$\mathbf{x}_c(\cdot) \overset{\text{def}}{=} \mathbf{q}(\cdot), \quad \mathbf{x}_v(\cdot) \overset{\text{def}}{=} \dot{\mathbf{q}}(\cdot), \quad \mathbf{x}(\cdot) \overset{\text{def}}{=} \begin{pmatrix} \mathbf{x}_c(\cdot) \\ \mathbf{x}_v(\cdot) \end{pmatrix},$$

cf. (4.8). The walker is able to accelerate its feet by controlling rotational torques applied to the legs which we summarize in the control function $\mathbf{u}(\cdot)$. The generalized forces acting on the MBS at time t are then given by

$$\boldsymbol{\tau}(t) = \begin{pmatrix} 0 & 0 & \mathbf{u}_1(t) & \mathbf{u}_2(t) \end{pmatrix}^T \in \mathbb{R}^4.$$

In a walking motion (in contrast to running), either one of the two feet must be fixed to the ground. Here, fixed to the ground means that the corresponding point foot is in touch with the ground and does not change its position. The walking motion can hence be realized by alternating between two possible modes of the system:

- Mode 1: the left foot is fixed to the ground.
- Mode 2: the right foot is fixed to the ground.

Due to the stiff legs of the walker, a situation in which both feet are fixed to the ground arises only momentarily as an isolated point of transition between Modes 1 and 2. Therefore, it is not handled as a separate mode in the sense of Chapter 5.

The varying modes are modeled as varying external contacts which act on the MBS as additional constraints. The external contacts can be expressed by means of the differential states in form

$$\mathbf{0} = \mathbf{g}^j(\mathbf{x}_c(t)),$$

where j indicates the contact configuration at time t. For each mode, the governing dynamics of the system are given by

$$\begin{bmatrix} \dot{\mathbf{x}}_c(t) \\ \mathbf{M}_j(\mathbf{x}_c(t)) \begin{pmatrix} \dot{\mathbf{x}}_v(t) \\ -\mathbf{z}^j(t) \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_v(t) \\ \boldsymbol{\tau}(t) - \mathbf{C}(\mathbf{x}(t)) \\ -\left[\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{G}^j(\mathbf{x}_c(t)) \right] \mathbf{x}_v(t) \end{bmatrix}, \tag{7.1a}$$

$$\mathbf{0} = \mathbf{g}^j(\mathbf{x}_c(t)), \tag{7.1b}$$

$$\mathbf{0} = \mathbf{G}^j(\mathbf{x}_c(t))\mathbf{x}_v(t), \tag{7.1c}$$

with a regular matrix $\mathbf{M}_j(\cdot)$, contact forces $\mathbf{z}^j(\cdot)$, applied generalized forces $\boldsymbol{\tau}(\cdot)$, generalized bias force $\mathbf{C}(\cdot)$, and contact Jacobians $\mathbf{G}^j(\mathbf{x}_c(t)) = \frac{\partial}{\partial \mathbf{x}_c} \mathbf{g}^j(\mathbf{x}_c(t))$. Explicit expressions for $\mathbf{M}_j(\cdot)$, $\mathbf{C}(\cdot)$, $\mathbf{g}^j(\cdot)$, $\mathbf{G}^j(\cdot)$, as well as the inverse of $\mathbf{M}_j(\cdot)$ can be found in Appendix A.1. If the Constraints (7.1b-7.1c) hold at any time point, the Differen-

tial Algebraic Equation (7.1) ensures that the constraints also hold for all subsequent time points as long as the mode of the system does not change. As the matrices $\mathbf{M}_j(\cdot)$ are regular, we can describe the dynamics in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}^j(\mathbf{x}(t), \mathbf{u}(t)), \quad t \in [t_i, t_{i+1}), \tag{7.2a}$$

$$\mathbf{0} = \mathbf{\Gamma}^j(\mathbf{x}(t_i)), \tag{7.2b}$$

if the system is in Mode j on the interval $[t_i, t_{i+1})$, where $\mathbf{\Gamma}^j(\cdot)$ summarizes the right hand sides of the Constraints (7.1b) and (7.1c). Whenever the mode – i. e., the external contact – changes (e. g., if a foot hits the ground after swinging freely before) at a time point t_s, a collision impact takes place and transfers the generalized velocities before the collision, $\mathbf{x}_v(t_s^-) = \lim_{t \nearrow t_s} \mathbf{x}_v(t)$, to those after the collision, $\mathbf{x}_v(t_s^+) = \lim_{t \searrow t_s} \mathbf{x}_v(t)$. We model the impact as a perfect inelastic collision. The transfer of velocities can be expressed in form

$$\mathbf{M}_j(\mathbf{x}_c(t_s)) \begin{pmatrix} \mathbf{x}_v(t_s^+) \\ -\mathbf{\Lambda}^j \end{pmatrix} = \begin{pmatrix} \mathbf{H}(\mathbf{x}_c(t_s))\mathbf{x}_v(t_s^-) \\ \mathbf{0} \end{pmatrix},$$

with contact impulse $\mathbf{\Lambda}^j$ and generalized inertia matrix $\mathbf{H}(\cdot)$. The index j corresponds to the mode *after* the impact. Since the matrices $\mathbf{M}_j(\cdot)$ are regular, the transfer of velocities can be incorporated into a jump function of the form

$$\mathbf{x}(t_s^+) = \mathbf{\Delta}^j\left(\mathbf{x}(t_s^-)\right),$$

mapping the differential states before a change of modes $\mathbf{x}(t_s^-)$ to the differential states after the change $\mathbf{x}(t_s^+)$.

7.1.3 Walking Motion Constraint Model

We propose a set of constraints imposed to model the process of walking. The walking motion spans a time interval $\mathcal{T} = [0, t_f]$ with $t_f \geq 0$ being a free end time which is determined by optimization later. For illustrative reasons, we use the notation from Section 7.1.1 to formulate the constraints. However, all constraints can be expressed in terms of the differential states $\mathbf{x}(\cdot)$.

Boundary Constraints

At time $t = 0$, the constraints

$$x_h(0) = y_l(0) = y_r(0) = 0, \tag{7.3a}$$

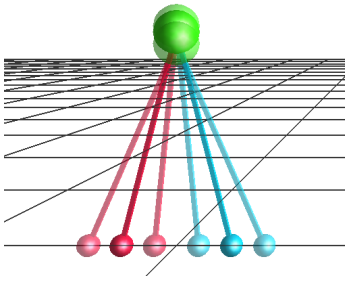


Figure 7.2: Possible initial postures of the simplest walker MBS according to the Constraints (7.3). Illustration created using `MeshUp` [48].

$$0 \leq y_h(0), \tag{7.3b}$$
$$0.2 \leq x_l(0) - x_r(0) \leq 0.8, \tag{7.3c}$$
$$-\pi \leq \varphi_l(0), \varphi_r(0) \leq \pi, \tag{7.3d}$$
$$-5 \leq \dot{x}_h(0), \dot{y}_h(0), \dot{\varphi}_l(0), \dot{\varphi}_r(0) \leq 5 \tag{7.3e}$$

force the walker to start in a reasonable initial configuration, cf. Fig. 7.2. The resulting walking motion will be determined by optimization later. In order to generate a homogeneous walking pattern in which initial (as well as terminal, cf. Constraints (7.5)) posture and velocities do not stand out, respectively, we leave some freedom for optimizing the initial posture and velocities as well.

At the end of the time horizon, the constraint

$$1.8 \leq x_h(t_f) \tag{7.4}$$

prescribes a final position which forces the walker to move. Furthermore, we want the modeled movement to be cyclic up to a certain accuracy, i. e., the posture of the walker in the beginning and the end of the observed interval should approximately coincide. The same must hold for the velocities of all sub-bodies. To achieve this, we demand

$$-\varepsilon_{\text{tol}}^{\text{c}} \leq \varphi_l(0) - \varphi_l(t_f) \leq \varepsilon_{\text{tol}}^{\text{c}}, \qquad -\varepsilon_{\text{tol}}^{\text{c}} \leq \varphi_r(0) - \varphi_r(t_f) \leq \varepsilon_{\text{tol}}^{\text{c}}, \tag{7.5a}$$
$$-\varepsilon_{\text{tol}}^{\text{c}} \leq \dot{x}_h(0) - \dot{x}_h(t_f) \leq \varepsilon_{\text{tol}}^{\text{c}}, \qquad -\varepsilon_{\text{tol}}^{\text{c}} \leq \dot{y}_h(0) - \dot{y}_h(t_f) \leq \varepsilon_{\text{tol}}^{\text{c}}, \tag{7.5b}$$
$$-\varepsilon_{\text{tol}}^{\text{c}} \leq \dot{\varphi}_l(0) - \dot{\varphi}_l(t_f) \leq \varepsilon_{\text{tol}}^{\text{c}}, \qquad -\varepsilon_{\text{tol}}^{\text{c}} \leq \dot{\varphi}_r(0) - \dot{\varphi}_r(t_f) \leq \varepsilon_{\text{tol}}^{\text{c}}, \tag{7.5c}$$

for some $\varepsilon^{\mathrm{c}}_{\mathrm{tol}} > 0$. We remark that the posture of the walker only depends on φ_l and φ_r. Thus, if $\big(\varphi_l(0), \varphi_r(0)\big)$ is close to $\big(\varphi_l(t_f), \varphi_r(t_f)\big)$, then also $y_h(0)$ is close to $y_h(t_f)$, which is why we do not impose additional cyclicity constraints on $y_h(\cdot)$.

Path Constraints

In order to generate a natural looking walking-like motion, we demand the head of the walker to stay above a certain level, and the feet should not penetrate the ground. However, the considered stick man is not able to walk in a reasonable way without penetrating the ground due to its stiff legs. Thus, we set up a tolerance $\varepsilon^{\mathrm{p}}_{\mathrm{tol}} > 0$ and demand

$$-\varepsilon^{\mathrm{p}}_{\mathrm{tol}} \leq y_l(t),\, y_r(t), \quad 0.8 \leq y_h(t)\,, \quad t \in \mathcal{T}\,. \tag{7.6}$$

For the initial time this is ensured by (7.3).

Mode-Dependent Path Constraints

As stated in Section 7.1.2, each mode is characterized by a fixation of a respective foot to the ground. For instance, during Mode 1 the left foot is fixed to the ground. Hence, at each time point during Mode 1 we demand

$$\mathbf{c}^1(\mathbf{x}(t)) \overset{\mathrm{def}}{=} \begin{pmatrix} y_l(t) & \dot{x}_l(t) & \dot{y}_l(t) \end{pmatrix}^T = \mathbf{0}\,. \tag{7.7}$$

This way, Equation (7.7) corresponds to Constraints (7.1b-7.1c) (and (7.2b), respectively), which in turn belong to the mode-dependent dynamics.

Accordingly, at each time point during Mode 2 we demand

$$\mathbf{c}^2(\mathbf{x}(t)) \overset{\mathrm{def}}{=} \begin{pmatrix} y_r(t) & \dot{x}_r(t) & \dot{y}_r(t) \end{pmatrix}^T = \mathbf{0}\,. \tag{7.8}$$

7.1.4 An Optimal Control Model for a Walking-Like Motion

We set up an appropriate OCP of Form (5.2) to generate a gait of the simplest walker MBS. In contrast to the approach pursued in Appendix A.2, the order and number of modes are not determined in advance.

Optimization Criterion

We consider an optimization criterion which represents a compromise between three criteria: mechanical effort (encoded in $\int_0^{t_f} \mathbf{u}_1(t)^2 + \mathbf{u}_2(t)^2 \mathrm{d}t$), duration of the walking motion (encoded in the process duration t_f), and the number of switches, i. e. mode changes (encoded in $|\mathcal{S}(\boldsymbol{\omega})|$, where $\boldsymbol{\omega}(\cdot)$ denotes the mode-indicator functions from the next paragraph), the latter being closely related to the number of steps of a gait pattern. From our numerical experiments, for natural looking gait patterns we expect all three criteria to take values of the same magnitude. We weight them equally which yields the optimization criterion

$$\int_0^{t_f} \mathbf{u}_1(t)^2 + \mathbf{u}_2(t)^2 \mathrm{d}t + t_f + |\mathcal{S}(\boldsymbol{\omega})| \,. \tag{7.9}$$

Optimal Control Model

It is reasonable to assume that the switched MBS has a strictly positive dwell time $\bar{\delta}$, such that only finitely many switches occur, but not in 0 or t_f. The differential equation together with the mode-dependent Path Constraints (7.7-7.8) can then be written in form

$$\begin{aligned}
\boldsymbol{\omega}(t) &\in \mathbb{S}^2\,, && t \in \mathcal{T}\,,\\
\dot{\mathbf{x}}(t) &= \textstyle\sum_{j=1}^{2} \boldsymbol{\omega}_j(t) \cdot \mathbf{f}^j(\mathbf{x}(t), \mathbf{u}(t))\,, && t \in \mathcal{T}\,,\\
\mathbf{0} &\geq \pm \boldsymbol{\omega}_j(t) \cdot \mathbf{c}^j(\mathbf{x}(t)), && t \in \mathcal{T}\,, \text{for } j = 1,2\,,
\end{aligned}$$

with mode-indicator functions $\boldsymbol{\omega}(\cdot) \in PC_{\bar{\delta}}(\mathcal{T}, \{0,1\}^2)$, and the summands in the right-hand side of the differential equation corresponding to (7.2a).

Our approach, as described in Chapter 5, neither allows for a free end time nor for a Lagrange term in the objective function. This issue can be resolved by a time transformation and the introduction of additional differential states, cf. Section 2.3.3. To this end, we consider the differential states

$$\begin{pmatrix} \mathbf{y}(t) \\ \mathbf{z}(t) \end{pmatrix} \in \mathbb{R}^8 \times \mathbb{R}^2\,,$$

together with the differential equation

$$\begin{bmatrix} \dot{\mathbf{y}}(t) \\ \dot{\mathbf{z}}_1(t) \\ \dot{\mathbf{z}}_2(t) \end{bmatrix} = \mathbf{z}_2(t) \begin{bmatrix} \sum_{j=1}^{2} \boldsymbol{\omega}_j(t) \cdot \mathbf{f}^j(\mathbf{y}(t), \mathbf{u}(t)) \\ \mathbf{u}_1(t)^2 + \mathbf{u}_2(t)^2 \\ 0 \end{bmatrix}, \quad t \in \mathcal{T} = [0,1]\,. \tag{7.10}$$

The value of the constant state $\mathbf{z}_2$ encodes the process duration and for the initial values of $\mathbf{z}(\cdot)$ we demand

$$\mathbf{z}_1(0) = 0 \;\text{ and }\; \mathbf{z}_2(0) \geq 0\,. \tag{7.11}$$

The Optimization Criterion (7.9) transforms into

$$\mathbf{z}_1(1) + \mathbf{z}_2(1) + |\mathcal{S}(\boldsymbol{\omega})|\,. \tag{7.12}$$

We remark that in applications where the value of the dwell time is of vital importance, the dwell time needs to be adapted after time transformation. However, in the considered application only the existence of a dwell time matters and not its particular value. Hence, no adaption is required. We assume the dwell time to be sufficiently small such that there is no conflict with any grid resolutions arising during the solution process.

We set the parameter values to

$$\varepsilon_{\text{tol}}^{\text{c}} = \frac{1}{20}, \quad \varepsilon_{\text{tol}}^{\text{p}} = \frac{1}{10},$$

which have proven to be favorable in practice. In summary, we obtain the following Mixed-Integer Optimal Control Problem for modeling a gait of the walker MBS:

$$\min_{\substack{\mathbf{y}(\cdot),\mathbf{z}(\cdot),\\ \mathbf{u}(\cdot),\boldsymbol{\omega}(\cdot)}} \mathbf{z}_1(1) + \mathbf{z}_2(1) + |\mathcal{S}(\boldsymbol{\omega})| \tag{7.13a}$$

$$\begin{aligned}
\text{s.t. } & \boldsymbol{\omega}(t) \in \mathbb{S}^2, && t \in \mathcal{T}, \\
& \dot{\mathbf{y}}(t) = \mathbf{z}_2(t) \left[\textstyle\sum_{j=1}^{2} \boldsymbol{\omega}_j(t) \cdot \mathbf{f}^j(\mathbf{y}(t), \mathbf{u}(t)) \right], && t \in \mathcal{T}, \\
& \dot{\mathbf{z}}(t) = \mathbf{z}_2(t) \begin{bmatrix} \mathbf{u}_1(t)^2 + \mathbf{u}_2(t)^2 \\ 0 \end{bmatrix}, && t \in \mathcal{T}, \\
& \mathbf{0} \geq \pm \boldsymbol{\omega}_j(t) \mathbf{c}^j(\mathbf{y}(t)) && t \in \mathcal{T}, \forall\, j, \\
& \mathbf{y}(t_s^+) = \boldsymbol{\Delta}_{j_2}\left(\mathbf{y}(t_s^-)\right), && \text{if } j_1 \rightarrow_{\boldsymbol{\omega}} j_2 \text{ at } t_s \in \mathcal{S}(\boldsymbol{\omega}),
\end{aligned}$$

$$\mathbf{0} \geq \mathbf{d}(\mathbf{y}(t), \mathbf{u}(t)), \qquad t \in \mathcal{T}, \tag{7.13b}$$

$$\mathbf{0} \geq \mathbf{r}(\mathbf{y}(0), \mathbf{z}(0), \mathbf{y}(1)), \tag{7.13c}$$

where (7.13b) comprises the Path Constraints (7.6), while (7.13c) summarizes the Boundary Constraints (7.3-7.5) and (7.11).

7.1.5 Implementation Details

We use the approach described in Chapter 5 to tackle Problem (7.13). We successively reformulate, relax, and discretize Problem (7.13) as explained in Sections 5.3 and 5.5 and solve the resulting problem with the approach from Section 5.7 using our software implementation (see Section 5.8). In the following, we comment on further details regarding the solution process.

MBS Dynamics Computations

The computation of, e. g., the dynamics equations by hand is cumbersome and prone to error, leading to a significant effort if the considered MBS is altered. Therefore, we modified the MBS software library `RBDL` [49] to render it compatible to the automatic differentiation tool `Adol-C` [148], which `grc` – the software package we use to tackle Problem (7.13), cf. Section 5.8 – uses internally for derivative generation. In our implementation for the present application, we use this modified version of `RBDL` for all computations in the context of the MBS, as it enables the treatment of altered MBSs without the need to reimplement MBS-specific expressions.

Switching Indicators

In our application, the jump function only depends on the mode after a switch. Per se, it would be possible to make use of the subsequent switching indicators and an according jump function, as introduced in Section 5.4.3. If one considers MBSs with more than two contact modes, using the subsequent switching indicators is beneficial as it reduces the number of variables in comparison to a usage of the omniscient switching indicators. However, as we consider two modes, both sets of switching indicators coincide, and we use the omniscient indicators in our application.

Discretization

We follow Sections 5.3.2 and 5.5 for the problem discretization. After applying the approach stated therein, besides the choice of the time grid, it remains to determine

the exact representation of each component of the differential states $\big(\mathbf{y}(\cdot), \mathbf{z}(\cdot)\big)$ and the control function $\mathbf{u}(\cdot)$, as well as the exact time points at which we evaluate the Path Constraints (7.13b).

For any grid occurring during the solution process, we represent the components of $\mathbf{y}(\cdot)$ and $\mathbf{z}_1(\cdot)$ per grid interval by cubic polynomials. The state $\mathbf{z}_2(\cdot)$ (corresponding to the constant process duration) and both components of the control function $\mathbf{u}(\cdot)$ are represented by piecewise linear polynomials. We recall that the Path Constraints (7.13b) are already satisfied at the first grid point $t = 0$ due to the boundary constraints. Hence, they are evaluated at all grid points except for the first one. As initial time grid, we choose the equidistant grid $\mathbb{G} = \left\{ \frac{i}{N} \,\middle|\, i = 0, \ldots, N \right\}$ with $N = 20$.

Homotopy, Refinement, and Warmstart

As stated in Section 5.7, the strategies for homotopy, refinement, and warm-start must be set up for each problem individually. We state a strategy for the present application example.

For the initial vanishing constraint parameter we choose the value $\gamma_0 = 5 \cdot 10^{-3}$. The reduction factor for the homotopy parameter (cf. (5.22)) is set to $\rho = 0.5$, and we find $\gamma_{\text{acc}} = 10^{-5}$ to be a suitable termination threshold.

For the refinement, we proceed as follows: we consider an inner grid point t_i, $i \in \{1, \ldots, N-1\}$. Two observations can be made. First, let $\left(\boldsymbol{\theta}^i_{j_1,j_2}\right)_{j_1 \neq j_2}$ be the switching indicators resulting from the output of the Nonlinear Programming Problem (NLP) solver. If the corresponding vector of NLP variables is feasible, from Proposition 5.7 we conclude that fractional modes inside the grid intervals $[t_{i-1}, t_i]$ or $[t_i, t_{i+1}]$ as well as a switch at t_i yield a non-zero $\boldsymbol{\theta}^{i-1}_{j_1,j_2}$. A non-zero $\boldsymbol{\theta}^{i-1}_{j_1,j_2}$ in turn results in a non-trivial jump function acting at the grid point t_i. Second, from our experience with walking motions we expect the control functions $\mathbf{u}_1(t)$ and $\mathbf{u}_2(t)$ to be continuous as long as the mode of the system does not change. Let $\mathbf{U}^{(i-1)}(t)$ and $\mathbf{U}^{(i)}(t)$ be the linear polynomials representing $\mathbf{u}(t)$ on the grid intervals $[t_{i-1}, t_i]$ and $[t_i, t_{i+1}]$, respectively, cf. (5.18). Continuity of the control functions yields the condition

$$\mathbf{U}^{(i-1)}(t_i) = \mathbf{U}^{(i)}(t_i)\,,$$

which can be expressed in terms of the NLP variables and the collocation points.

We set up tolerances $\varepsilon_{1,\text{ref}}, \varepsilon_{2,\text{ref}} > 0$. Based on the observations, for all $i = 1, \dots, N-1$ we check

$$\left\| \left(\boldsymbol{\theta}^{i-1}_{j_1,j_2} \right)_{j_1 \neq j_2} \right\|_\infty > \varepsilon_{1,\text{ref}}, \tag{7.14a}$$

$$\left\| \mathbf{U}^{(i-1)}(t_i) - \mathbf{U}^{(i)}(t_i) \right\|_\infty > \varepsilon_{2,\text{ref}}. \tag{7.14b}$$

If one of the conditions is satisfied, the intervals $[t_{i-1}, t_i]$ and $[t_i, t_i]$ are marked for refinement. After checking the Conditions (7.14) for each i the marked intervals are bisected. Thereby, it is not important if an interval is marked for refinement once or twice. It remains to choose $\varepsilon_{1,\text{ref}}$ and $\varepsilon_{2,\text{ref}}$. For this, we consider the vanishing constraint parameter γ. For decreasing γ we expect the solutions of the NLPs to become more accurate. Thus, it is reasonable to couple the values of $\varepsilon_{1,\text{ref}}$ and $\varepsilon_{2,\text{ref}}$ with the current value of γ. We find $\varepsilon_{1,\text{ref}} = \varepsilon_{1,\text{ref}}(\gamma) = \frac{\gamma}{10}$ and $\varepsilon_{2,\text{ref}} = \varepsilon_{2,\text{ref}}(\gamma) = 10\gamma$ to be suitable choices.

After refining the grid we set up a new guess for the subsequent NLP resulting from the refinement. This so-called warm-starting procedure works as follows. We interpolate all components of the differential states $\left(\mathbf{y}(\cdot), \mathbf{z}(\cdot)\right)$ as well as the control functions $\mathbf{u}_1(\cdot)$ and $\mathbf{u}_2(\cdot)$ and $\boldsymbol{\alpha}_1(\cdot)$ and $\boldsymbol{\alpha}_2(\cdot)$. The control parameters $\boldsymbol{\beta}^i_{j_1,j_2}$ and $\boldsymbol{\theta}^i_{j_1,j_2}$ are then initialized in such a way that

$$\boldsymbol{\theta}^i_{j_1,j_2} = \min\left[\boldsymbol{\alpha}_{j_1}(t_i), \boldsymbol{\alpha}_{j_2}(t_{i+1}) \right]$$

holds for all $i = 0, \dots, N-2$.

7.1.6 Results

We solve the arising NLPs with an Interior-Point method. We use the software package `IPOPT` [146] with standard settings except for the desired convergence tolerance (parameter `tol` in [147], resp. error tolerance ϵ_{tol} in [146]) which is set to 10^{-6}. In the 10th iteration, we have $\gamma \leq \gamma_{\text{acc}}$ and we receive the solution depicted in Fig. 7.3 and Fig. 7.4. Our solver determines the process duration – encoded in the value of the constant state $\mathbf{z}_2$ – to be ≈ 5.436 seconds. We see that the homotopy together with the grid refinement successively reduces the deviation of the resulting $\boldsymbol{\theta}^i_{j_1,j_2} \in [0,1]$ and $\mathbf{a}^i_j \in [0,1]$ from $\{0,1\}$. To illustrate this behavior, the values of

$$\max_{i, j_1 \neq j_2} \min\left(\boldsymbol{\theta}^i_{j_1,j_2}, 1 - \boldsymbol{\theta}^i_{j_1,j_2} \right), \quad \max_{i,j} \min\left(\mathbf{a}^i_j, 1 - \mathbf{a}^i_j \right),$$

per iteration are depicted in Fig. 7.5. Furthermore, a visualization of the postures of the simplest walker during the walking process is shown in Fig. 7.6.

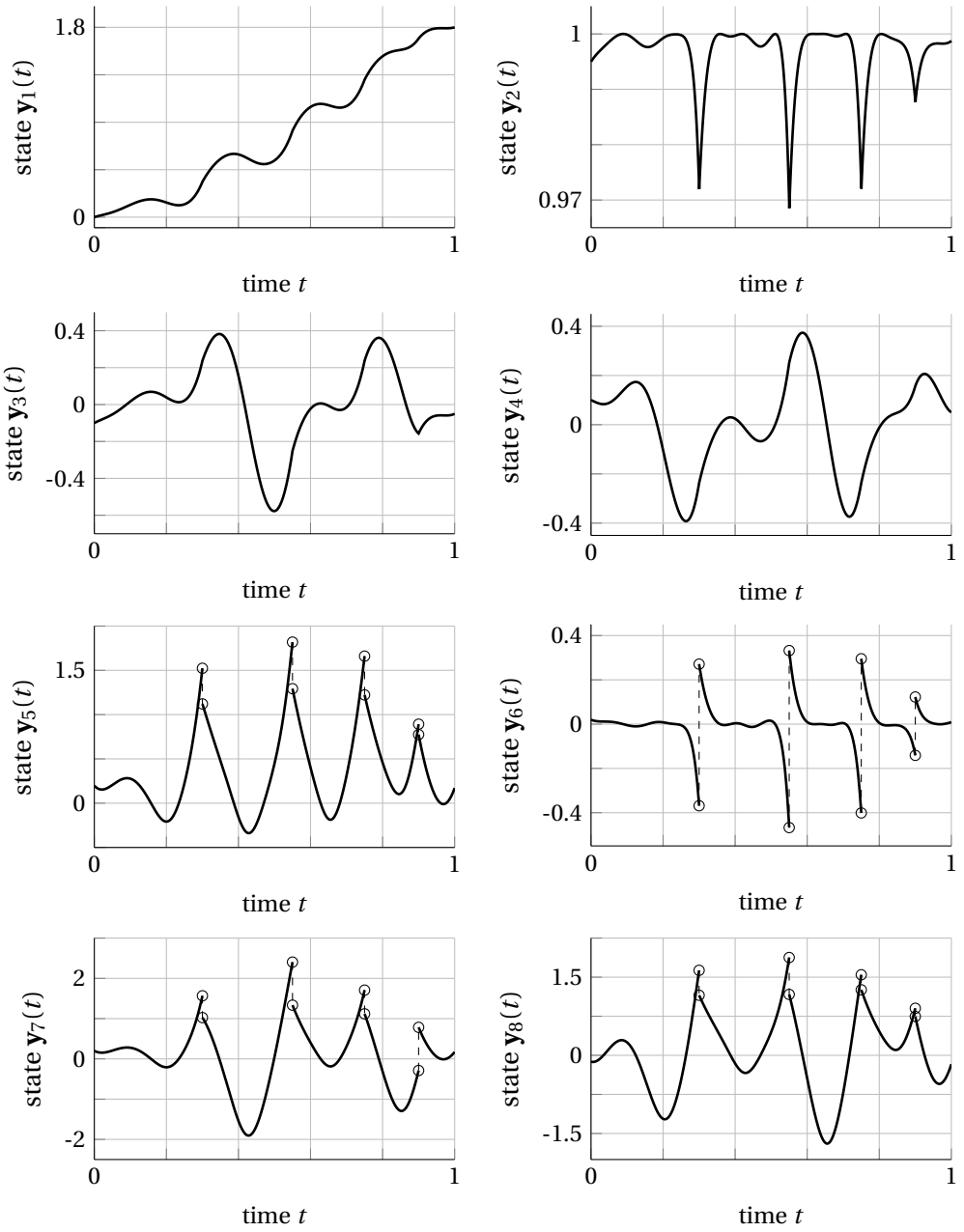
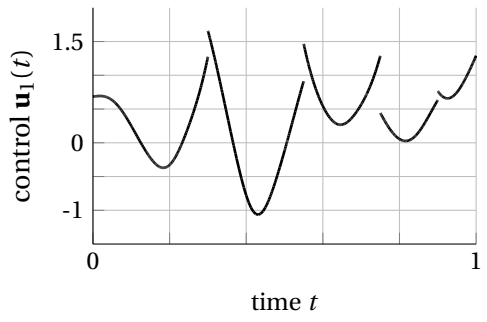
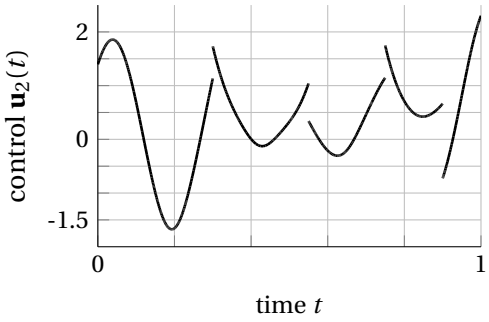


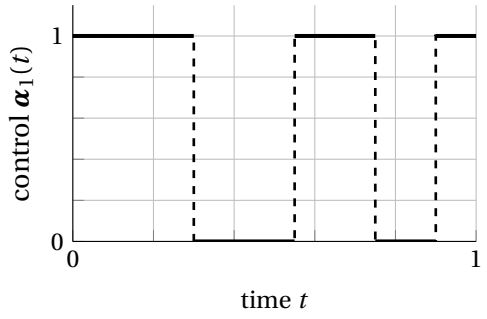
Figure 7.3: Differential states corresponding to the time-transformed generalized coordinates $\left(\mathbf{y}_1(\cdot),\ldots,\mathbf{y}_4(\cdot)\right)$ and velocities $\left(\mathbf{y}_5(\cdot),\ldots,\mathbf{y}_8(\cdot)\right)$ of the simplest walker multi-body system (cf. Tab. 7.1 for their meaning) in the solution after the 10th iteration. Jumps occur in the generalized velocities of the multi-body system whenever a foot hits the ground.



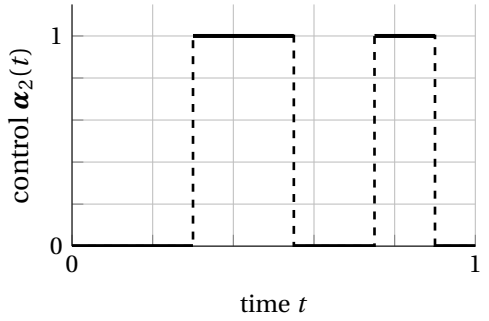
a) Left foot torque $\boldsymbol{\tau}_3(t)$.



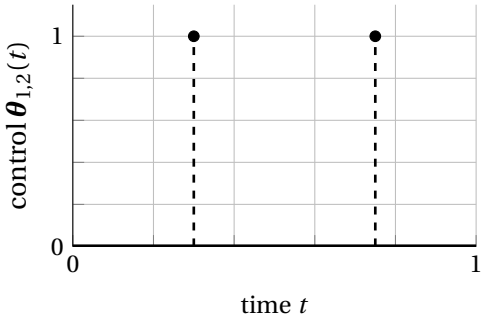
b) Right foot torque $\boldsymbol{\tau}_4(t)$.



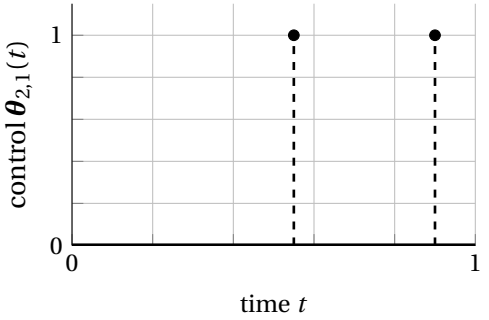
c) Indicator $\boldsymbol{\alpha}_1(\cdot)$ for Mode 1 (left foot fixed to the ground).



d) Indicator $\boldsymbol{\alpha}_2(\cdot)$ for Mode 2 (right foot fixed to the ground).

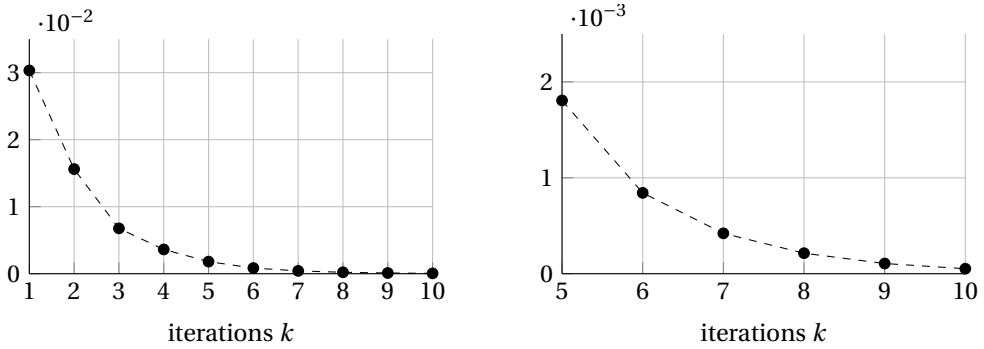


e) $\boldsymbol{\theta}_{1,2}(\cdot)$ indicating switches from Mode 1 to Mode 2.

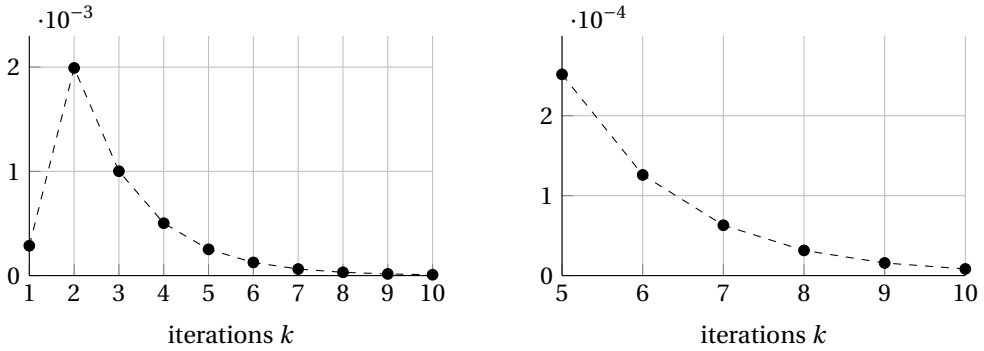


f) $\boldsymbol{\theta}_{2,1}(\cdot)$ indicating switches from Mode 2 to Mode 1.

Figure 7.4: Trajectories of controls $\mathbf{u}(\cdot)$ (the time-transformed actuator torques of the MBS), mode-indicators $\boldsymbol{\alpha}(\cdot)$, and switching indicators $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$ in the solution after the 10th iteration. In the trajectories of the switching indicators $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$, we highlight the values at mode change.

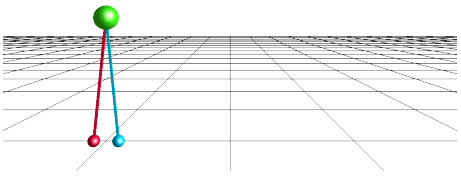


a) Value of $\max_{i,j_1\neq j_2} \min\left(\boldsymbol{\theta}^i_{j_1,j_2}, 1-\boldsymbol{\theta}^i_{j_1,j_2}\right)$ per iteration, measuring the deviation of the $\boldsymbol{\theta}^i_{j_1,j_2}$ from $\{0,1\}$. The left plot depicts the respective value for all iterations while the right plot focuses on the iterations $k \geq 5$.

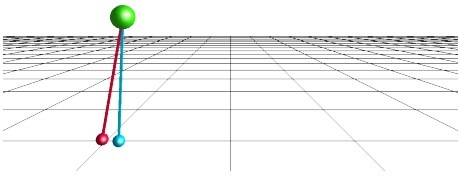


b) Value of $\max_{i,j} \min\left(\mathbf{a}^i_j, 1-\mathbf{a}^i_j\right)$ per iteration, measuring the deviation of $\mathbf{a}^i_j$ from $\{0,1\}$. The left plot depicts the respective value for all iterations while the right plot focuses on the iterations $k \geq 5$.

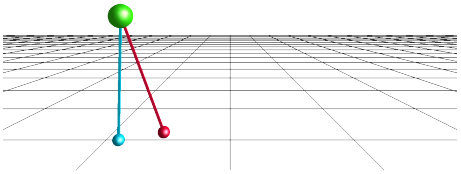
Figure 7.5: Effect of homotopy and grid refinement on the deviation of the resulting $\boldsymbol{\theta}^i_{j_1,j_2} \in [0,1]$ and $\mathbf{a}^i_j \in [0,1]$ from $\{0,1\}$.



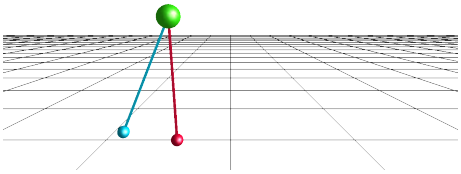
a) $t = 0$: initial posture.



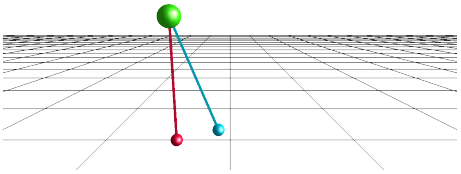
b) $t = 0.564$: walker in Mode 1. The left foot is fixed to the ground.



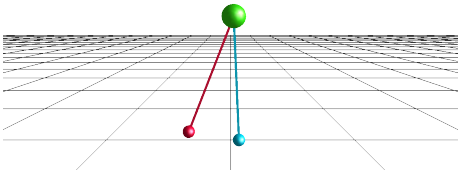
c) $t = 1.288$.



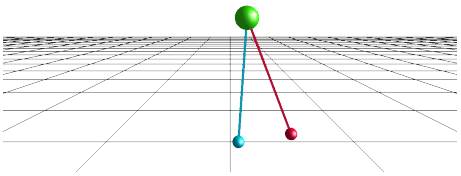
d) $t = 1.772$: walker in Mode 2 after a collision impact took place. Now, the right foot is fixed to the ground.



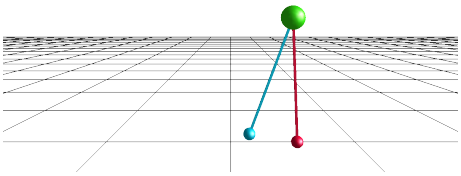
e) $t = 2.456$



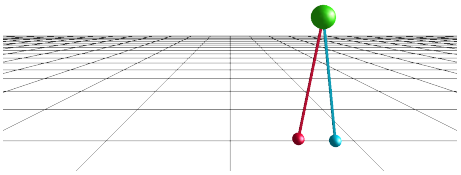
f) $t = 3.181$: again, a switch occurred.



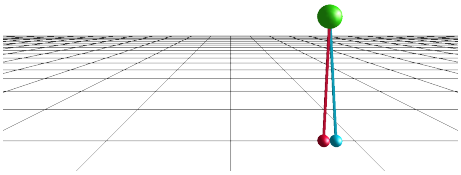
g) $t = 3.946$



h) $t = 4.268$



i) $t = 4.953$



j) $t = 5.436$: terminal posture.

Figure 7.6: Postures of the simplest walker at various time points during the walking process. Here, the time is scaled to the process duration encoded in the value of the state $\mathbf{z}_2$. The right leg is represented by the red segments and the left leg by the blue segments. Initial and terminal posture coincide approximately due to the Constraints (7.5). Visualization created with `MeshUp` [48].

7.2 Worst-Case Treatment Planning by Bilevel Optimal Control

In this section, we consider a fictive patient suffering from Cerebral Palsy (CP) who is forced into a crouch gait by the disease. In the real world, physicians would aim at ameliorating the gait by applying orthopedic surgeries. For instance, a hamstring lengthening alters the flexibility of a patient's knee joint and this way gives them the possibility to stand and walk more upright. However, we assume that the intervention cannot be performed with absolute accuracy and thus suffers from a certain degree of uncertainty. We apply the Training Approach from Chapter 6 to this scenario to demonstrate its efficacy.

7.2.1 A Multi-Body System Modeling the Human Body

We consider a walking motion in 2D. The Multi-Body System (MBS) we use to model the human body is based on the `HeiMan` model [48, sec. 4.3]. The following description of the MBS is inspired by [86], where a similar MBS is employed. We consider a rigid MBS comprising 7 bodies, representing the upper body, thighs, shanks, and feet. All bodies except for the feet are spatially extended bodies with non-zero inertia. In contrast, the feet are modeled as mass points in order to reduce the number of contact configurations and accordingly the number of model stages in the Optimal Control Problem (OCP) we will set up later (see Section 7.2.2). Upper body, thighs, and shanks are interconnected via rotational joints, modeling the hip and knee joints. The feet – being represented by mass points – are fixed to the shanks. The sub-body which represents the upper part of the human body is composed of 10 rigid bodies. However, all of them are rigidly fixed to each other and we focus on the lower body in our example. An illustration of the MBS topology of the walker model can be seen in Fig. 7.7.

We consider a 2D walking motion in the sagittal plane, i. e., the (x, y)-plane in Fig. 7.7. The segment which represents the upper body acts as a base segment for the MBS. It is connected to the origin of the global coordinate system with a so-called Floating Base Joint which allows for free translations and rotations about the origin of the upper body $(x_p, y_p, 0)$ in the (x, y)-plane. The joints which model the left and right hip joint are located in the upper body and connect the upper body with both thighs. The knee joints in turn are located in the respective thigh and connect the thighs with the rigid bodies which summarize the respective shank and foot. Both hip and knee joints allow for rotations in the (x, y)-plane. The precise joint locations are stated in Tab. 7.2. Here, the positions of the joints are given relative to stated ref-

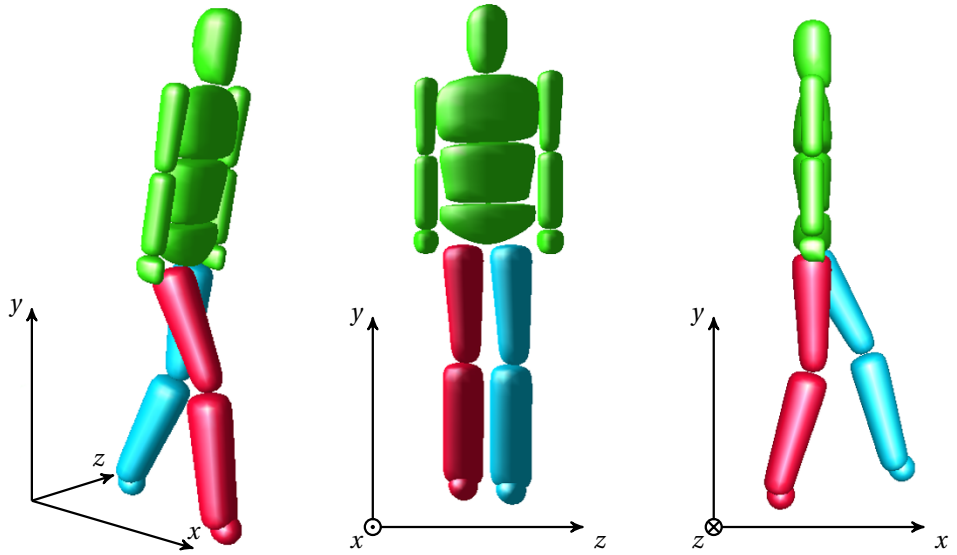


Figure 7.7: The walker model from Section 7.2 in different perspectives. The right leg is represented by red segments and the left leg by blue segments. Illustrations created using `MeshUp` [48].

Table 7.2: Joint locations in the MBS given relative to a reference point in the non-rotated segment in which the respective joint is located. For a visualization, see Fig. 7.8.

Joint Name	**Located in**	**Reference Point**	**Relative Position** in m^3
Floating Base Joint	Global Coord. System	$(0 \quad 0 \quad 0)^T$	$(x_p \quad y_p \quad 0)^T$
Left Hip Joint	Upper Body	Floating Base Joint	$(0 \quad 0 \quad 0.08091)^T$
Left Knee Joint	Left Thigh	Left Hip Joint	$(0 \quad -0.4220 \quad 0)^T$
Right Hip Joint	Upper Body	Floating Base Joint	$(0 \quad 0 \quad -0.08091)^T$
Right Knee Joint	Right Thigh	Right Hip Joint	$(0 \quad -0.4220 \quad 0)^T$

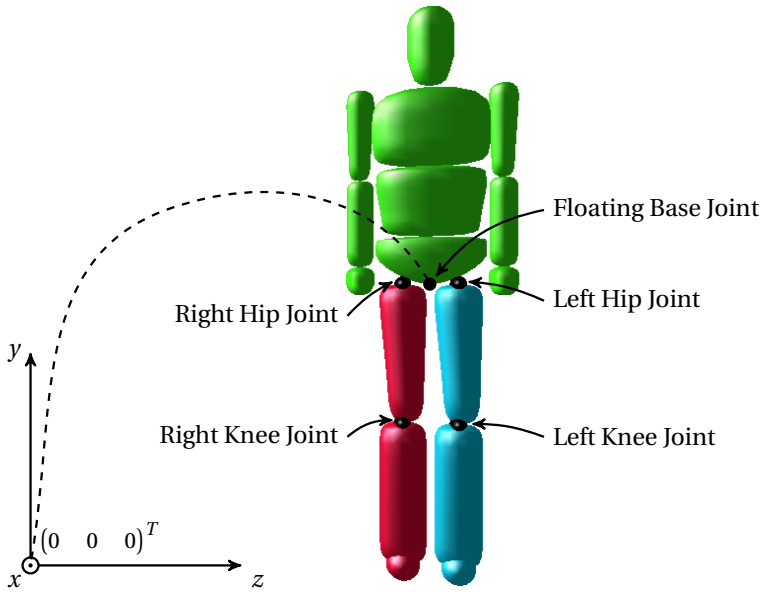


Figure 7.8: Joint Locations in the walker MBS. Illustration created using `MeshUp` [48].

erence points in the non-rotated segments. An illustration can be found in Fig. 7.8.

As we consider motions in the (x, y)-plane, the inertia properties of the single segments can be described by means of their masses, mass centers and the respective rotational inertias for rotations in the (x, y)-plane about the respective centers of mass. All quantities regarding the inertia properties of the segments are given in Tab. 7.3.

The segment which represents the upper body can freely translate and rotate in the (x, y)-plane, and the segments representing the thighs and shanks (with feet) can rotate freely around the hip and knee joints, respectively, in the (x, y)-plane. Hence, the considered motion of the MBS can be described by means of seven generalized coordinates

$$\mathbf{q}(\cdot) = \begin{pmatrix} x_p(\cdot) & y_p(\cdot) & \varphi_p(\cdot) & \varphi_{h,l}(\cdot) & \varphi_{k,l}(\cdot) & \varphi_{h,r}(\cdot) & \varphi_{k,r}(\cdot) \end{pmatrix}^T, \tag{7.15}$$

Table 7.3: MBS segment properties. The mass centers are given relative to the stated origins of the non-rotated segments. The rotational inertias describe the moments of inertia for rotations in the (x, y)-plane about the respective centers of mass. All properties are given in SI or SI derived units.

Segment	**Mass** in kg	**Origin**	**Center of Mass** in m^3	**Rot. Inertia** in $\mathrm{kg}\cdot\mathrm{m}^3$
Upper Body	44.61	Floating Base Joint	$\begin{pmatrix}0 & 0.3436 & 0\end{pmatrix}^T$	0.2176
Left Thigh	10.48	Left Hip Joint	$\begin{pmatrix}0 & -0.1728 & 0\end{pmatrix}^T$	0.1388
Left Shank + Foot	4.218	Left Knee Joint	$\begin{pmatrix}0 & -0.2548 & 0\end{pmatrix}^T$	0.1413
Right Thigh	10.48	Right Hip Joint	$\begin{pmatrix}0 & -0.1728 & 0\end{pmatrix}^T$	0.1388
Right Shank + Foot	4.218	Right Knee Joint	$\begin{pmatrix}0 & -0.2548 & 0\end{pmatrix}^T$	0.1413

Table 7.4: Generalized coordinates of the walker MBS.

$\mathbf{q}_1(\cdot) = x_p(\cdot)$	Horizontal (i. e., x-) position of the origin of the upper body
$\mathbf{q}_2(\cdot) = y_p(\cdot)$	Vertical (i. e., y-) position of the origin of the upper body
$\mathbf{q}_3(\cdot) = \varphi_p(\cdot)$	Rotation of upper body about the origin pivot
$\mathbf{q}_4(\cdot) = \varphi_{h,l}(\cdot)$	Rotation of left thigh about left hip joint
$\mathbf{q}_5(\cdot) = \varphi_{k,l}(\cdot)$	Rotation of left shank about left knee joint
$\mathbf{q}_6(\cdot) = \varphi_{h,r}(\cdot)$	Rotation of right thigh about right hip joint
$\mathbf{q}_7(\cdot) = \varphi_{k,r}(\cdot)$	Rotation of right shank about right knee joint

encoding these translations and rotations. For the precise meaning of the single coordinates see Tab. 7.4 and Fig. 7.9.

The movements of the walker are caused by torques acting through the (rotational) hip and knee joints, respectively. These torques are denoted by $\boldsymbol{\tau}_{h,l}(\cdot)$, $\boldsymbol{\tau}_{k,l}(\cdot)$, $\boldsymbol{\tau}_{h,r}(\cdot)$, and $\boldsymbol{\tau}_{k,r}(\cdot)$, see Tab. 7.5 for their meaning. The upper body itself is actuated indirectly as a result of the interaction of the feet with the ground. Thus, we consider an underactuated MBS and the generalized forces at time t are given by

$$\boldsymbol{\tau}(t) = \begin{pmatrix}\mathbf{0} & \overbrace{\boldsymbol{\tau}_{h,l}(t) & \boldsymbol{\tau}_{k,l}(t) & \boldsymbol{\tau}_{h,r}(t) & \boldsymbol{\tau}_{k,r}(t)}^{\boldsymbol{\tau}^a(t)^T}\end{pmatrix}^T = \begin{pmatrix}\mathbf{0} \\ \boldsymbol{\tau}^a(t)\end{pmatrix} \in \mathbb{R}^3 \times \mathbb{R}^4.$$

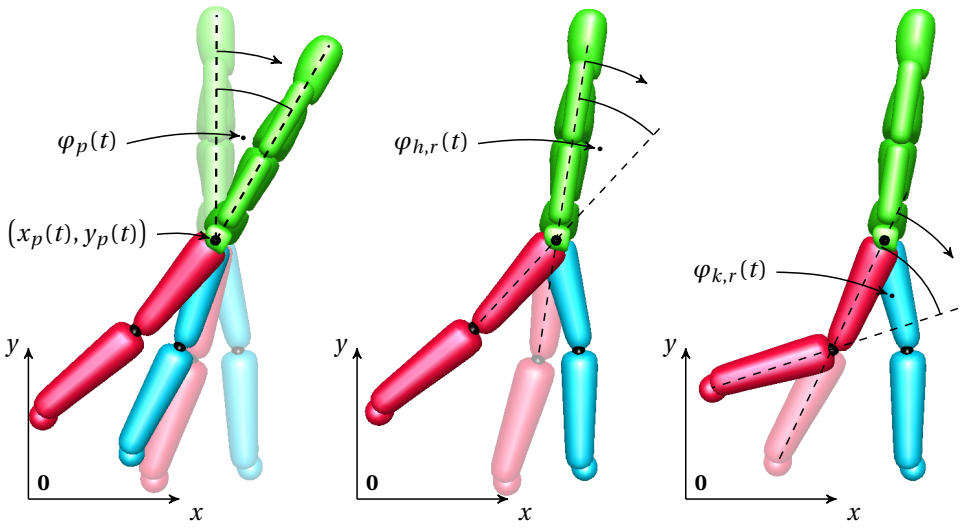


Figure 7.9: Illustration of generalized coordinates $x_p(t)$, $y_p(t)$, $\varphi_p(t)$, $\varphi_{h,r}(t)$, and $\varphi_{k,r}(t)$ of the walker MBS. The right leg is represented by red segments and the left leg by blue segments. The arrows next to the angles indicate the directions of rotation. The coordinates $\varphi_{h,l}(t)$ and $\varphi_{k,l}(t)$ can be visualized analogously to $\varphi_{h,r}(t)$ and $\varphi_{k,r}(t)$, respectively. Illustrations created using `MeshUp` [48].

Table 7.5: Torques actuating the walker MBS.

$\boldsymbol{\tau}_{h,l}(\cdot)$	Torque acting through left hip joint
$\boldsymbol{\tau}_{k,l}(\cdot)$	Torque acting through left knee joint
$\boldsymbol{\tau}_{h,r}(\cdot)$	Torque acting through right hip joint
$\boldsymbol{\tau}_{k,l}(\cdot)$	Torque acting through right knee joint

For the mechanical equations of motion for rigid MBSs, see Section 4.1.

7.2.2 A Parametric Optimal Control Problem for the Patient's Gait

We set up a parametric OCP whose solutions model optimal gait patterns of the considered walker MBS during a single gait cycle. As described in Chapter 6, we use a multi-stage formulation for the OCP as this has produced favorable results in the context of bilevel optimization in practice [31, 32, 71]. We take a look at the torque generation, consider constraints which force the MBS into a walking motion, and finally set up an Optimal Control model. For the sake of brevity, in the following we sometimes omit the argument t for time-dependent variables.

Torque Generation

The walker is actuated by torques $\boldsymbol{\tau}_{h,l}(\cdot)$, $\boldsymbol{\tau}_{k,l}(\cdot)$, $\boldsymbol{\tau}_{h,r}(\cdot)$, and $\boldsymbol{\tau}_{k,r}(\cdot)$ which act through the hip and knee joints. We follow the modeling approach from Section 4.3. Each torque is made up out of an active part – modeling the effect of the involved muscles – and a passive part. The passive part contains a damping term and passive reset forces, where the latter – simply put – ensure that the respective joint does not leave a certain movement range during walking. In the Optimal Control model we set up later, the active parts are represented by the normalized control function $\mathbf{u}(\cdot)$ whose values lie in $[-1,1]^4$.

In the present example, the movement ranges of the hip joints are not the subject of interest and will later be chosen such that they do not influence the resulting optimal gait (see (7.29)). Therefore, we do not consider passive reset forces in the hip joints. However, for the knee joints the movement ranges and thus passive reset forces need to be incorporated. Let $[\underline{\boldsymbol{p}}_{k,l}, \overline{\boldsymbol{p}}_{k,l}]$ and $[\underline{\boldsymbol{p}}_{k,r}, \overline{\boldsymbol{p}}_{k,r}]$ be the modeled ranges of motion of the left and right knee, respectively. According to Section 4.3, the resulting torques are then given by

$$\boldsymbol{\tau}_{h,l} = \boldsymbol{\tau}_1^a = \boldsymbol{\tau}_1^{a,\max}\left[\mathbf{u}_1 - \beta_1 \dot{\varphi}_{h,l}\right], \tag{7.16a}$$

$$\boldsymbol{\tau}_{k,l} = \boldsymbol{\tau}_2^a = \boldsymbol{\tau}_2^{a,\max}\left[\mathbf{u}_2 + e^{-\underline{c}_2(\varphi_{k,l}-\underline{\boldsymbol{p}}_{k,l})} - e^{\overline{c}_2(\varphi_{k,l}-\overline{\boldsymbol{p}}_{k,l})} - \beta_2 \dot{\varphi}_{k,l}\right], \tag{7.16b}$$

$$\boldsymbol{\tau}_{h,r} = \boldsymbol{\tau}_3^a = \boldsymbol{\tau}_3^{a,\max}\left[\mathbf{u}_3 - \beta_3 \dot{\varphi}_{h,r}\right], \tag{7.16c}$$

$$\boldsymbol{\tau}_{k,r} = \boldsymbol{\tau}_4^a = \boldsymbol{\tau}_4^{a,\max}\left[\mathbf{u}_4 + e^{-\underline{c}_4(\varphi_{k,r}-\underline{\boldsymbol{p}}_{k,r})} - e^{\overline{c}_4(\varphi_{k,r}-\overline{\boldsymbol{p}}_{k,r})} - \beta_4 \dot{\varphi}_{k,r}\right], \tag{7.16d}$$

with $\beta_i > 0$ (damping parameter), $\underline{c}_i, \overline{c}_i > 0$ (passive reset forces curvature) and $\boldsymbol{\tau}_i^{a,\max} > 0$ (maximum active actuated torque). Summarizing all occurring param-

eters in the vector $\mathbf{p}$ yields

$$\boldsymbol{\tau}^a = \boldsymbol{\tau}^a\left(\mathbf{u}, \mathbf{q}, \dot{\mathbf{q}}, \mathbf{p}\right).$$

Walking Motion Constraints

In general, the gait cycle contains phases in which exactly one foot is in contact with the ground – so called single support phases – as well as phases in which both feet are. However, according to [71, p. 174] the latter only represent a small portion of the whole gait cycle. Therefore, we follow [71] and consider a sequence of single support phases of free duration together with the according phase transitions that take place at isolated time points at which both feet are in contact with the ground.

We define the ground as the (x,z)-plane of the global coordinate system. Foot-ground contacts are modeled by external contacts of the MBS (cf. Section 4.1.1). For later usage, we denote the time-dependent position of the left and right foot in the (x,y)-plane by $\left(x_l(t), y_l(t)\right)$ and $\left(x_r(t), y_r(t)\right)$, respectively. Furthermore, we define differential states $\mathbf{x}(\cdot)$ using the generalized coordinates (7.15) via

$$\mathbf{x}(t) = \begin{pmatrix} \mathbf{q}(t) \\ \dot{\mathbf{q}}(t) \end{pmatrix} \in \mathbb{R}^{14}.$$

Initial Position, Posture, and Velocities

We set the initial time T_0 to 0. In our model, the walker enters the gait cycle at the beginning of the single support phase assigned to the right foot and thus immediately after the corresponding phase transition. Hence, at $t = T_0$ both feet are in touch with the ground and the velocities of the right foot equal zero:

$$y_l(0) = y_r(0) = 0\,, \tag{7.17a}$$
$$\dot{x}_r(0) = \dot{y}_r(0) = 0\,. \tag{7.17b}$$

We demand the origin of the upper body to start in zero x-position and the left foot to be placed behind the right foot in x-direction:

$$x_p(0) = 0\,, \tag{7.18a}$$
$$x_r(0) \geq x_l(0)\,. \tag{7.18b}$$

If the angles $\varphi_{k,l}(t)$ and $\varphi_{k,r}(t)$ – describing the rotation of the knee joints – approach or even exceed the borders of the intervals $\left[\underline{\boldsymbol{p}}_{k,l}, \overline{\boldsymbol{p}}_{k,l}\right]$ and $\left[\underline{\boldsymbol{p}}_{k,r}, \overline{\boldsymbol{p}}_{k,r}\right]$, respectively,

this results in high passive reset forces. In the Optimal Control Model (7.32) we set up later, the Objective Function (7.31) encodes a compromise between the duration of the modeled gait cycle and the mechanical effort during walking. From an optimization perspective, high initial passive reset forces therefore – simply put – energize the system "for free". Hence, we want to avoid such a state in the beginning of the walking process. We set up a threshold $\varepsilon_{\text{pass}} > 0$ and bound the normalized passive reset forces (cf. (4.17)) by

$$\begin{aligned} e^{-\underline{c}_2\left(\varphi_{k,l}(0)-\underline{\boldsymbol{p}}_{k,l}\right)},\ e^{\overline{c}_2\left(\varphi_{k,l}(0)-\overline{\boldsymbol{p}}_{k,l}\right)} &\leq \varepsilon_{\text{pass}}, \\ e^{-\underline{c}_4\left(\varphi_{k,r}(0)-\underline{\boldsymbol{p}}_{k,r}\right)},\ e^{\overline{c}_4\left(\varphi_{k,r}(0)-\overline{\boldsymbol{p}}_{k,r}\right)} &\leq \varepsilon_{\text{pass}}, \end{aligned}$$

which is equivalent to

$$-\frac{1}{\underline{c}_2}\log\varepsilon_{\text{pass}} + \underline{\boldsymbol{p}}_{k,l} \leq \varphi_{k,l}(0) \leq \frac{1}{\overline{c}_2}\log\varepsilon_{\text{pass}} + \overline{\boldsymbol{p}}_{k,l}, \tag{7.19a}$$

$$-\frac{1}{\underline{c}_4}\log\varepsilon_{\text{pass}} + \underline{\boldsymbol{p}}_{k,r} \leq \varphi_{k,r}(0) \leq \frac{1}{\overline{c}_4}\log\varepsilon_{\text{pass}} + \overline{\boldsymbol{p}}_{k,r}. \tag{7.19b}$$

Phase 1: Right Foot Fixed to the Ground

Phase 1 starts at $T_0 = 0$ and ends at $T_1 \geq T_0$. Let $\mathcal{T}_1 = [0, T_1]$. We model the foot-ground contact using external contacts. Due to arising constraint forces, the mass point representing the right foot is not accelerated. The mechanical equations of motion for the MBS can be stated by means of the differential states $\mathbf{x}(\cdot)$ in form

$$\dot{\mathbf{x}}(t) = \mathbf{f}^{\text{I}}\left(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}\right), \quad t \in \mathcal{T}_1,$$

cf. Section 4.1. Because of (7.17) they ensure a fixation of the position of the right foot to the ground during Phase 1 (cf. Section 4.1.1).

We impose constraints on the constraint force $\boldsymbol{\lambda}^r(\cdot) = \boldsymbol{\lambda}^r\left(\mathbf{x}(\cdot), \mathbf{u}(\cdot), \mathbf{p}\right)$ (cf. Section 4.1) that keeps the right foot fixed to the ground. Here, $\boldsymbol{\lambda}^r(\cdot)$ models the ground reaction force acting on the right foot. On the one hand, the vertical component $\boldsymbol{\lambda}^r_y(\cdot)$ must always be non-negative:

$$\boldsymbol{\lambda}^r_y(t) \geq 0, \quad t \in \mathcal{T}_1. \tag{7.20}$$

On the other hand, the right foot must not slip away due to dry friction. Let $\mu_{\text{fric}} > 0$ be the friction coefficient for the foot-ground contact. We demand

$$\boldsymbol{\lambda}_x^r(t) \leq \mu_{\text{fric}} \boldsymbol{\lambda}_y^r(t), \quad t \in \mathcal{T}_1, \tag{7.21}$$

where $\boldsymbol{\lambda}_x^r(\cdot)$ denotes the horizontal component of the constraint force.

Furthermore, the vertical position of the left foot – which can move freely – must not enter the ground in Phase 1:

$$y_l(t) \geq 0, \quad t \in \mathcal{T}_1. \tag{7.22}$$

Phase Transition: Left Foot Hits the Ground

At $t = T_1$ the left foot hits the ground, i. e.,

$$y_l(T_1) = 0, \tag{7.23a}$$
$$\dot{y}_l(T_1) \leq 0, \tag{7.23b}$$

and a collision impact occurs. As explained before, we do not consider (non-instantaneous) gait phases in which both feet are fixed to the ground. Thus, in Phase 2 the left foot is fixed to a position on the ground which we realize again using external contacts. The transition of velocities at collision impact can be expressed in form

$$\mathbf{x}(T_1^+) = \boldsymbol{\Delta}^1\left(\mathbf{x}(T_1^-)\right),$$

where $\boldsymbol{\Delta}^1(\cdot)$ transfers the differential states instantly before the impact, $\mathbf{x}(T_1^-)$, to those instantly after the impact, $\mathbf{x}(T_1^+)$, according to the rules of mechanics for perfect inelastic collisions, see Section 4.1, Equation (4.7). We automatically get $\dot{x}_l(T_1^+) = \dot{y}_l(T_1^+) = 0$.

Phase 2: Left Foot Fixed to the Ground

Phase 2 takes place in the (non-empty) time interval $\mathcal{T}_2 = [T_1, T_2]$. As in Phase 1, we model the foot-ground contact by external contacts. The left foot stays at its fixed position due to a constraint force $\boldsymbol{\lambda}^l(\cdot) = \boldsymbol{\lambda}^l\left(\mathbf{x}(\cdot), \mathbf{u}(\cdot), \mathbf{p}\right)$. The equations of motion can be stated in form

$$\dot{\mathbf{x}}(t) = \mathbf{f}^2\left(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}\right), \quad t \in \mathcal{T}_2.$$

Similar to Phase 1, we impose constraints

$$\boldsymbol{\lambda}_y^l(t) \geq 0\,, \tag{7.24a}$$
$$\boldsymbol{\lambda}_x^l(t) \leq \mu_{\text{fric}} \boldsymbol{\lambda}_y^l(t)\,, \tag{7.24b}$$
$$y_r(t) \geq 0\,, \tag{7.24c}$$

for $t \in \mathcal{T}_2$, respectively.

Phase Transition: Right Foot Hits the Ground

At $t = T_2$ the right foot hits the ground after swinging freely before:

$$y_r(T_2) = 0\,, \tag{7.25a}$$
$$\dot{y}_r(T_2) \leq 0\,. \tag{7.25b}$$

Similar to the previous phase transition, the transition of velocities at collision impact can be expressed in form

$$\mathbf{x}(T_2^+) = \boldsymbol{\Delta}^2\left(\mathbf{x}(T_2^-)\right).$$

Terminal Position and Posture

By demanding

$$x_p(T_2) \geq x_{\text{end}} \tag{7.26}$$

for some $x_{\text{end}} > 0$ we force the walker to leave its initial position and move in positive x-direction.

We want the resulting movement to be cyclic, meaning that the initial and terminal postures of the walker coincide, and the same shall hold for the velocities. The considered walker MBS has seven degrees of freedom. At $t = T_2$ both feet are in touch with the ground. Hence, in the terminal configuration there are five degrees of freedom left. A translation in x-direction does not influence the posture of the walker. Thus, it is sufficient to demand cyclicity for four generalized coordinates:

$$y_p(0) = y_p(T_2)\,, \tag{7.27a}$$
$$\varphi_p(0) = \varphi_p(T_2)\,, \tag{7.27b}$$
$$\varphi_{h,l}(0) = \varphi_{h,l}(T_2)\,, \tag{7.27c}$$

$$\varphi_{h,r}(0) = \varphi_{h,r}(T_2)\,. \tag{7.27d}$$

We remark that in theory it is possible to have $\varphi_{k,i}(0) \neq \varphi_{k,i}(T_2)$ for $i \in \{l,r\}$ even if (7.27) is satisfied. However, in the present example it turns out that the Constraints (7.27) yield cyclic optimal gaits in practice.

For the velocities of the system, by the properties of the function $\mathbf{\Delta}^2(\cdot)$ – modeling the change of velocities at phase transition – we know that $\dot{x}_r(t) = \dot{y}_r(t) = 0$ after the right foot hit the ground. Arguing similarly as before, we see that it suffices to demand

$$\dot{x}_p(0) = \dot{x}_p(T_2)\,, \tag{7.28a}$$
$$\dot{y}_p(0) = \dot{y}_p(T_2)\,, \tag{7.28b}$$
$$\dot{\varphi}_p(0) = \dot{\varphi}_p(T_2)\,, \tag{7.28c}$$
$$\dot{\varphi}_{h,l}(0) = \dot{\varphi}_{h,l}(T_2)\,, \tag{7.28d}$$
$$\dot{\varphi}_{k,l}(0) = \dot{\varphi}_{k,l}(T_2) \tag{7.28e}$$

to achieve cyclicity of the generalized velocities.

Bounds

In addition to the constraints we imposed above, we set up simple bounds for the variables which will be subject of optimization later. For the phase durations, we demand $0 = T_0 \leq T_1 \leq T_2$. The control function $\mathbf{u}(\cdot)$ is normalized such that $\mathbf{u}(t) \in [-1,1]^4$ for $t \in \mathcal{T} = [T_0, T_2]$. For the Generalized Coordinates (7.15) and velocities – both summarized in the differential states $\mathbf{x}(\cdot)$ – we demand

$$-3 \leq \varphi_i(t) \leq 3\,, \tag{7.29}$$

for $i \in \{p, \{h,l\}, \{h,r\}, \{k,l\}, \{k,r\}\}$ and $t \in \mathcal{T}$, and

$$-10 \leq \dot{x}_p(t), \dot{y}_p(t), \dot{\varphi}_i(t) \leq 10 \tag{7.30}$$

for $i \in \{p, \{h,l\}, \{h,r\}, \{k,l\}, \{k,r\}\}$ and $t \in \mathcal{T}$.

Optimal Control Model

We set up an OCP to model walking-like motions of the considered MBS. As optimization criterion we consider a compromise between the mechanical effort and the duration of the modeled gait cycle (the latter being closely related to walking

speed), encoded in

$$\underbrace{\int_0^{T_2} \sum_{i=1}^{4} \mathbf{u}(t)^2 \,\mathrm{d}t}_{\widehat{=}\text{ mechanical effort}} + \frac{1}{10} \underbrace{T_2}_{\widehat{=}\text{ duration of gait cycle}}. \tag{7.31}$$

The employed weighting leads to contributions of similar magnitude in the considered solutions (for the most values $\underline{\boldsymbol{p}}$ we are interested in, see later). We set the parameter values to

$$\begin{aligned}
\boldsymbol{\tau}_i^{a,\max} &= 100 \quad \text{for } i = 1,\dots,4\,,\\
\beta_i &= 0.025 \quad \text{for } i = 1,\dots,4\,,\\
\underline{c}_i = \overline{c}_i &= 30 \quad \text{for } i = 2,4\,,\\
\overline{\boldsymbol{p}}_{k,l} = \overline{\boldsymbol{p}}_{k,r} &= \tfrac{2}{3}\pi \ \left(\widehat{=}120 \text{ degrees}\right),\\
\varepsilon_{\text{pass}} &= 0.1\,,\\
\mu_{\text{fric}} &= 0.65\,,\\
x_{\text{end}} &= 0.9\,,
\end{aligned}$$

which have proven to be favorable in practice. The parameters $\underline{\boldsymbol{p}}_{k,l}$ and $\underline{\boldsymbol{p}}_{k,r}$ – modeling the maximum possible knee extension – will be used for intervention modeling later, see the Section 7.2.3. They will become optimization parameters in the upper level of Problem (7.33). We summarize both in the parameter $\underline{\boldsymbol{p}}$.

Altogether, the resulting parametric multi-stage OCP modeling the gait of the considered (fictive) patient is given by

$$\min_{\substack{\mathbf{x}(\cdot;\underline{\boldsymbol{p}}),\\ \mathbf{u}(\cdot),T_1,T_2}} \quad \frac{1}{10} T_2 + \int_0^{T_2} \sum_{i=1}^{4} \mathbf{u}(t)^2 \,\mathrm{d}t \tag{7.32a}$$

$$\text{s.t.} \quad 0 \le T_1 \le T_2\,, \tag{7.32b}$$

$$\mathbf{u}(t) \in [-1,1]^4\,, \qquad t \in \mathcal{T}\,, \tag{7.32c}$$

$$\underline{\boldsymbol{b}}_i \le \mathbf{x}_i(t) \le \overline{\boldsymbol{b}}_i\,, \qquad t \in \mathcal{T}\,, \quad i = 3,\dots,14\,, \tag{7.32d}$$

$$\dot{\mathbf{x}}(t) = \mathbf{f}^j\left(\mathbf{x}(t),\mathbf{u}(t),\underline{\boldsymbol{p}}\right), \qquad t \in \mathcal{T}_j\,, \quad j = 1,2\,, \tag{7.32e}$$

$$\mathbf{x}(T_j^+) = \boldsymbol{\Delta}^j\left(\mathbf{x}(T_j^-)\right), \qquad j = 1,2\,, \tag{7.32f}$$

$$\mathbf{0} \le \mathbf{c}^j\left(\mathbf{x}(t),\mathbf{u}(t),\underline{\boldsymbol{p}}\right), \qquad t \in \mathcal{T}_j\,, \tag{7.32g}$$

$$\mathbf{0} = \mathbf{r}^{\text{eq}}\left(\mathbf{x}(0),\mathbf{x}(T_1),\mathbf{x}(T_2)\right), \tag{7.32h}$$

$$\mathbf{0} \le \mathbf{r}^{\text{ieq}}\left(\mathbf{x}(0),\mathbf{x}(T_1),\mathbf{x}(T_2),\underline{\boldsymbol{p}}\right). \tag{7.32i}$$

where

- $\mathcal{T} = [0, T_2]$, $\mathcal{T}_j = [T_{j-1}, T_j]$,
- (7.32d) summarizes the Box-Constraints (7.29-7.30),
- $\mathbf{c}^1(\cdot)$ summarizes the Constraints (7.20-7.22) and $\mathbf{c}^2(\cdot)$ the Constraints (7.24),
- $\mathbf{r}^{\text{eq}}(\cdot)$ sums up (7.17-7.18a), (7.23a), (7.25a), (7.27-7.28), and
- $\mathbf{r}^{\text{ieq}}(\cdot)$ summarizes (7.18b-7.19), (7.23b), (7.25b) and (7.26).

The parameter $\underline{\boldsymbol{p}}$ enters the Differential Equation (7.32e), the Path Constraints (7.32g), and the Boundary Constraints (7.32i).

7.2.3 Intervention, Uncertainty, and Worst Possible Intervention

We follow Section 4.3. In the model we described in the previous sections the ranges of motion of both knee joints are limited. The degrees of limitation are encoded in the intervals

$$\mathcal{I}_l = [\underline{\boldsymbol{p}}_{k,l}, \overline{\boldsymbol{p}}_{k,l}] = \left[\underline{\boldsymbol{p}}_{k,l}, \tfrac{2}{3}\pi\right] \quad \text{and} \quad \mathcal{I}_r = [\underline{\boldsymbol{p}}_{k,r}, \overline{\boldsymbol{p}}_{k,r}] = \left[\underline{\boldsymbol{p}}_{k,r}, \tfrac{2}{3}\pi\right],$$

respectively. Whenever $\varphi_{k,l}(t)$ approaches or exceeds the boundaries of $\mathcal{I}_l$, high passive reset forces arise and push $\varphi_{k,l}(t)$ back into the interior of $\mathcal{I}_l$. The same holds for $\varphi_{k,r}(t)$ and $\mathcal{I}_r$. Recall that for $\varphi_l(t) = 0$ ($\varphi_r(t) = 0$), the left (right) knee is fully extended, and for $0 < \varphi_l(t)$ ($\varphi_r(t)$) $< \pi$, the left (right) knee is flexed (see Fig. 7.9 for a visualization of $\varphi_{k,r}(t)$). Hence, if $0 < \underline{\boldsymbol{p}}_{k,l}, \underline{\boldsymbol{p}}_{k,r}$, the walker is forced into a gait pattern with flexed knees and the closer $\underline{\boldsymbol{p}}_{k,l}, \underline{\boldsymbol{p}}_{k,r}$ approach zero, the more upright the walker is able to stand and walk.

Before the intervention, the walker shows a crouch gait pattern and the pre-operative parameter values are given by $\underline{\boldsymbol{p}}_{k,l} = \underline{\boldsymbol{p}}_{k,r} = \frac{\pi}{9}$ ($\widehat{=}$ 20 degrees), which we summarize in $\underline{\boldsymbol{p}}_{\text{pre}}$. The goal of the considered intervention is to ameliorate the pre-operative gait. We model the intervention by an alteration of

$$\underline{\boldsymbol{p}} = \begin{pmatrix} \underline{\boldsymbol{p}}_{k,l} \\ \underline{\boldsymbol{p}}_{k,r} \end{pmatrix}.$$

To achieve a symmetric situation in both legs we need to set $\underline{\boldsymbol{p}}_{k,l} = \underline{\boldsymbol{p}}_{k,r}$.

However, in the present example we assume that the intervention cannot be performed with absolute accuracy. We further assume that deviations of five degrees $\left(\widehat{=}\frac{\pi}{36}\right)$ from the targeted nominal result are possible for each of the two parameters. To avoid overcorrections for both knees, the nominal value – encoding the intervention as planned – is set to

$$\underline{\boldsymbol{p}}_{\text{nom}} = \begin{pmatrix} \frac{\pi}{36} \\ \frac{\pi}{36} \end{pmatrix}.$$

Hence, due to uncertainty the possible realizations reside in the uncertainty set

$$\underline{\boldsymbol{p}} \in \Omega_{\mathbf{p}} = \left[0, \tfrac{\pi}{18}\right]^2.$$

In particular, we cannot expect $\underline{\boldsymbol{p}}_{k,l} = \underline{\boldsymbol{p}}_{k,r}$.

In principle, without further theoretical investigations we cannot expect the OCP (7.32) to have exactly one local solution for all relevant values of $\underline{\boldsymbol{p}}$. Therefore we have to consider the case that various solutions exist (though we did not come across differing ones in the course of our investigations). In this fictive example, the OCP solution which we assume to model the pre-operative gait is known by computation, see Section 7.2.4. For any $\underline{\boldsymbol{p}}$, we denote the OCP solution which models the actual establishing (possible post-operative) gait by $\mathfrak{g}(\underline{\boldsymbol{p}})$. We choose the corresponding Objective Function Value (7.32a) to assess the gait pattern encoded in $\mathfrak{g}(\underline{\boldsymbol{p}})$. As explained in Sections 6.2 and 6.4, a worst possible intervention due to uncertainty is then given by a global solution of the bilevel problem

$$\max_{\substack{\underline{\boldsymbol{p}} \in \Omega_{\mathbf{p}}, T_1, T_2, \\ \mathbf{u}(\cdot), \mathbf{x}(\cdot; \underline{\boldsymbol{p}})}} \quad \frac{1}{10} T_2 + \int_0^{T_2} \sum_{i=1}^{4} \mathbf{u}(t)^2 \, \mathrm{d}t \tag{7.33a}$$

$$\text{s.t.} \quad \left(T_1, T_2, \mathbf{u}(\cdot), \mathbf{x}\left(\cdot; \underline{\boldsymbol{p}}\right)\right) \text{ solves Problem (7.32) for } \underline{\boldsymbol{p}}, \text{ and} \tag{7.33b}$$

$$\left(T_1, T_2, \mathbf{u}(\cdot), \mathbf{x}\left(\cdot; \underline{\boldsymbol{p}}\right)\right) = \mathfrak{g}(\underline{\boldsymbol{p}}). \tag{7.33c}$$

Such a solution $\left(\underline{\boldsymbol{p}}^*, T_1^*, T_2^*, \mathbf{u}^*(\cdot), \mathbf{x}^*\left(\cdot; \underline{\boldsymbol{p}}^*\right)\right) = \left(\underline{\boldsymbol{p}}^*, \mathfrak{g}(\underline{\boldsymbol{p}}^*)\right)$ of the problem models a worst possible treatment option and the associated gait pattern. For a given $\underline{\boldsymbol{p}} \in \Omega_{\mathbf{p}}$ we denote the corresponding value of the Objective Function (7.33a) by $\varphi\left(\mathfrak{g}(\underline{\boldsymbol{p}})\right)$.

7.2.4 Solution Approach and Implementation

We follow the approach we described in detail in Section 6.4. In the present example, we first determine the OCP solution which we assume to model the pre-operative gait by solving the OCP (7.32) for the pre-operative parameter $\underline{\boldsymbol{p}}_{\text{pre}}$. Subsequently, we approach the nominal solution which corresponds to the nominal post-operative gait (i. e., the gait resulting from the intervention as planned), by means of a homotopy, starting in $\underline{\boldsymbol{p}}_{\text{pre}}$ and ending in $\underline{\boldsymbol{p}}_{\text{nom}}$.

We tackle Problem (7.33) using the approach proposed in [106] and described in Section 6.4. We use the Derivative-Free Optimization method of `BOBYQA` [124] with the implementation provided by the `NLOPT` library [78]. We start the optimization routine in $\underline{\boldsymbol{p}}_{\text{nom}}$. As termination criterion, we set up a relative tolerance of 10^{-5}.
For any evaluation of the objective function we solve the OCP (7.32) using the Direct Multiple Shooting approach [25] together with a Sequential Quadratic Programming method, see [94]. In our implementation, and more generally for every realization of the parameter $\underline{\boldsymbol{p}}$ occurring while generating the results presented in Section 7.2.5, we employ the OCP solver `MUSCOD-II` [95]. Here, we set up a multi-stage problem with four model stages. Two of them model the Phases 1 and 2 (see Section 7.2.2 for the phase descriptions) and have a free phase duration while the remaining stages correspond to the phase transitions and have zero duration. The stages modeling Phase 1 and 2 are discretized using a multiple shooting discretization with 10 shooting intervals and piecewise linear and stage-wise continuous control functions. As termination criterion, we use an acceptable KKT tolerance of 10^{-6}. For all computations in the context of rigid MBSs we employ the MBS dynamics library `RBDL` [49].

Our implementation employs software modules developed by our cooperation partner Prof. Dr. Katja Mombaur and the working group “Optimization in Robotics and Biomechanics” at Heidelberg University.

7.2.5 Results

Using the approach described in the previous section, we determine the optimal solution of Problem (7.33). Furthermore, we present the gait modeling solutions of the lower level OCP (7.32) for certain values of the parameter $\underline{\boldsymbol{p}}$.

We first focus on the OCP solutions which model the pre-operative gait ($\underline{\boldsymbol{p}} = \underline{\boldsymbol{p}}_{\text{pre}}$) and the nominal post-operative gait ($\underline{\boldsymbol{p}} = \underline{\boldsymbol{p}}_{\text{nom}}$). The corresponding phase durations and optimal objective function values can be found in Tab. 7.6, and the values of

generalized coordinates and control functions are depicted in Fig. 7.10 and Fig. 7.11, respectively. Furthermore, Fig. 7.12 illustrates the postures of the walker during the nominal post-operative gait.

For the Bilevel Problem (7.33) we (approximately) determine the solution parameter

$$\underline{\boldsymbol{p}}^* = \begin{pmatrix} \frac{\pi}{18} \\ \frac{\pi}{18} \end{pmatrix},$$

i. e., both optimization parameters take their maximum possible value with respect to the box-shaped uncertainty set $\Omega_{\mathbf{p}}$. During the optimization process, `BOBYQA` demands for 23 objective function evaluations, and for each evaluation the lower level OCP has to be solved. A graph of $\varphi\big(\mathfrak{g}(\cdot)\big)$ on $\Omega_{\mathbf{p}}$ (objective function value of Problem (7.33) depending on $\underline{\boldsymbol{p}}$) can be seen in Fig. 7.13. The function $\varphi\big(\mathfrak{g}(\cdot)\big)$ takes its minimal value on $\Omega_{\mathbf{p}}$ for $\underline{\boldsymbol{p}} = \mathbf{0}$.

To demonstrate the variety of possible intervention results due to uncertainty, we consider the gait modeling solutions of Problem (7.32) for $\underline{\boldsymbol{p}} \in \{\mathbf{0}, \underline{\boldsymbol{p}}_{\text{nom}}, \underline{\boldsymbol{p}}^*\}$ which represent the outcome of the best possible, nominal (= planned), and worst possible intervention, respectively. The corresponding phase durations as well as the objective function values can be found in Tab. 7.6, and the differential states, controls, and respective sums of the normalized passive reset forces

$$\boldsymbol{\tau}_{k,l}^{\text{pass}}(t) = e^{-30\left(\varphi_{k,l}(t) - \underline{\boldsymbol{p}}_{k,l}\right)} - e^{30\left(\varphi_{k,l}(t) - \frac{2}{3}\pi\right)}, \tag{7.34a}$$

$$\boldsymbol{\tau}_{k,r}^{\text{pass}}(t) = e^{-30\left(\varphi_{k,l}(t) - \underline{\boldsymbol{p}}_{k,r}\right)} - e^{30\left(\varphi_{k,r}(t) - \frac{2}{3}\pi\right)} \tag{7.34b}$$

(cf. (4.17) and (7.16)) are depicted in the Figures 7.14-7.17.

7.2.6 Discussion

First, we take a look at the effect of the parameter alteration – modeling the intervention – on the solution of the OCP (7.32) which models the patient's gait. We consider the graphs of $y_p(\cdot)$ (vertical position of origin of the upper body), $\varphi_{k,l}(\cdot)$ (rotation of left shank about the corresponding knee joint), and $\varphi_{k,r}(\cdot)$ (rotation of right shank about the corresponding knee joint) in Figures 7.10 and 7.14. A simultaneous raise of the values of the parameters $\underline{\boldsymbol{p}}_{k,l}$ and $\underline{\boldsymbol{p}}_{k,r}$ leads to an increased knee flexion on the one hand and, consequently, a decreased vertical position of the upper body while walking on the other hand. In other words, the crouch gait pattern intensifies for

increasing parameter values, as expected.

Furthermore, the values of the optimal control functions depend on $\underline{\boldsymbol{p}}$ as well, see Figures 7.11 and 7.16. During Phase 1 – in which the right foot is fixed to the ground – the absolute values of the control function $\mathbf{u}_4(\cdot)$ – modeling the controlled normalized generalized force acting through the right knee joint – decrease with decreasing parameter values, and the same holds for Phase 2 (left foot fixed to the ground) and $\mathbf{u}_2(\cdot)$ (controlled normalized generalized force acting through the left knee joint). This observation can be explained as follows. We consider a situation in which the walker stands on one foot in a steady state. The greater the knee flexion, the higher is the torque acting through the corresponding knee joint which is necessary in order to keep the system in the steady state. This consideration can be transferred to a walking motion. The observation corresponds to the well-known fact that walking in a crouch gait is more exhausting than implementing a more upright gait pattern. Corresponding to this, the objective function values of Problem (7.33) – encoding a compromise between mechanical effort and duration of the considered gait cycle – increase for increasing parameter values despite decreasing duration, see Tab. 7.6 and Fig. 7.13.

Besides the direct effect of the intervention on the gait pattern of the walker, we want to assess the considered treatment and in particular evaluate whether it is recommendable in view of the occurring uncertainties. As as measure for the quality of the gait, we use the objective function value of the OCP (7.32). Our solver (approximately) determines $\underline{\boldsymbol{p}}^* = \begin{pmatrix} \frac{\pi}{18} & \frac{\pi}{18} \end{pmatrix}^T$ as the solution parameter of the Bilevel Problem (7.33) and Fig. 7.13 justifies this result. In particular, we have

$$\varphi\big(\mathfrak{g}(\underline{\boldsymbol{p}})\big) \leq \varphi\big(\mathfrak{g}(\underline{\boldsymbol{p}}^*)\big) \approx 0.3375 < 0.5314 \approx \varphi\big(\mathfrak{g}(\underline{\boldsymbol{p}}_{\text{pre}})\big) \quad \text{for all } \underline{\boldsymbol{p}} \in \Omega_{\mathbf{p}},$$

i. e., the post-operative gait improves significantly in comparison to the pre-operative gait in any case in view of the employed assessment criterion. Hence, the intervention seems reasonable despite the considered uncertainty.

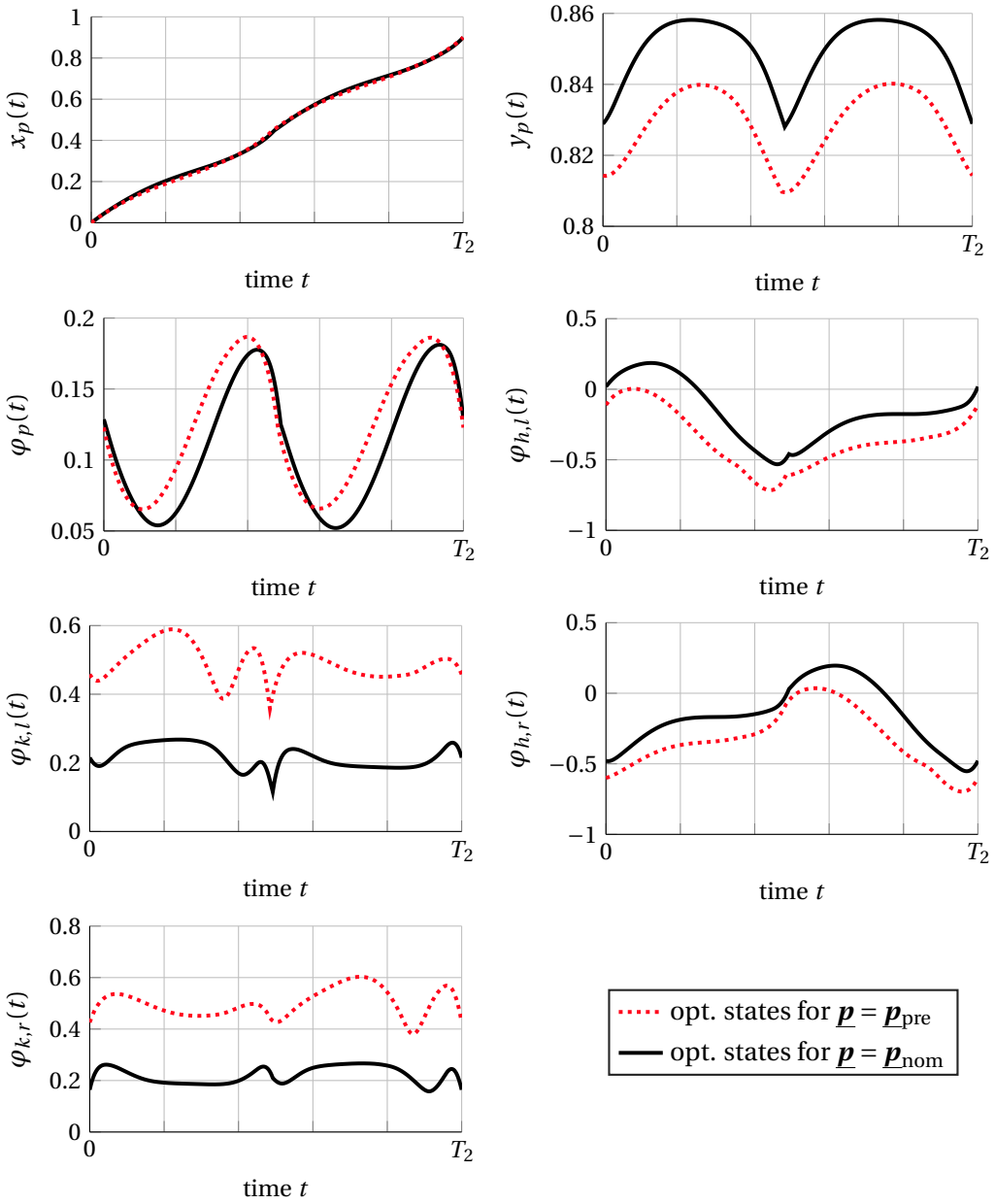


Figure 7.10: Comparison of pre-operative gait and nominal post-operative gait: values of generalized coordinates. The red (dotted) lines refer to the solution belonging to the pre-operative parameter $\underline{\boldsymbol{p}}_{\text{pre}}$ and the black (solid) lines to the nominal parameter $\underline{\boldsymbol{p}}_{\text{nom}}$. For both parameters the graphs are scaled to the same length.

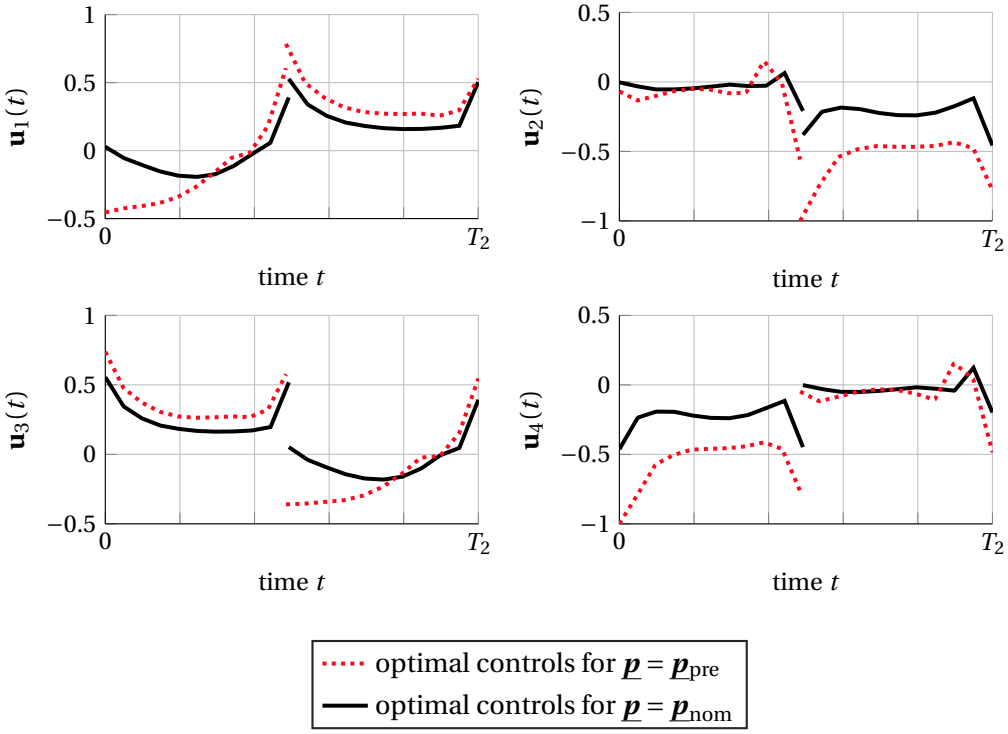
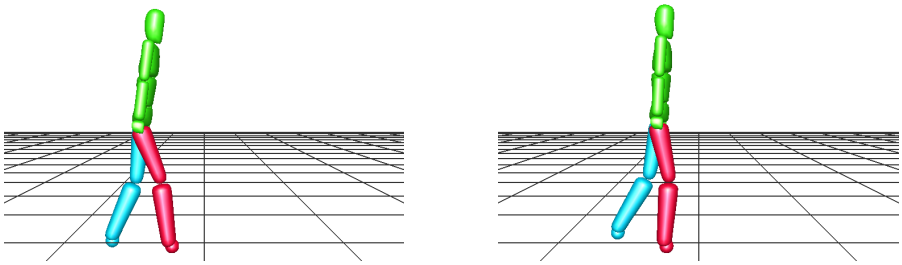


Figure 7.11: Comparison of pre-operative gait and nominal post-operative gait: values of $\mathbf{u}(t)$ representing the controlled normalized generalized forces. The red (dotted) lines refer to the solution belonging to the pre-operative parameter $\underline{\boldsymbol{p}}_{\text{pre}}$ and the black (solid) lines to the nominal parameter $\underline{\boldsymbol{p}}_{\text{nom}}$. For both parameters the graphs are scaled to the same length.

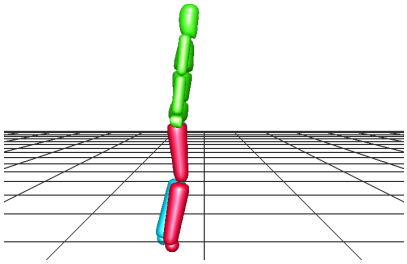
Table 7.6: Phase durations $T_1 - 0 = T_1$ and $T_2 - T_1$ (in seconds) and objective function values in the considered solutions of Problem (7.32) for $\underline{\boldsymbol{p}} \in \{\mathbf{0}, \underline{\boldsymbol{p}}_{\text{nom}}, \underline{\boldsymbol{p}}^*, \underline{\boldsymbol{p}}_{\text{pre}}\}$.

$\underline{\boldsymbol{p}}$	T_1	$T_2 - T_1$	$\varphi\big(\mathfrak{g}(\underline{\boldsymbol{p}})\big)$
$\mathbf{0}$	0.604	0.614	0.2216
$\underline{\boldsymbol{p}}_{\text{nom}} = \left(\frac{\pi}{36} \quad \frac{\pi}{36}\right)^T$	0.545	0.563	0.2681
$\underline{\boldsymbol{p}}^* = \left(\frac{\pi}{18} \quad \frac{\pi}{18}\right)^T$	0.487	0.508	0.3375
$\underline{\boldsymbol{p}}_{\text{pre}} = \left(\frac{\pi}{9} \quad \frac{\pi}{9}\right)^T$	0.396	0.421	0.5314

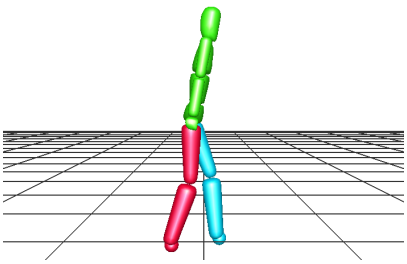


a) $t = 0$: initial posture.

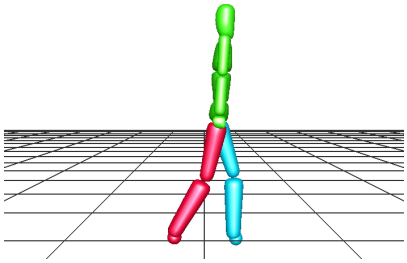
b) $t = 0.160$.



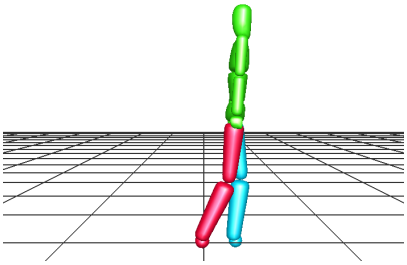
c) $t = 0.313$.



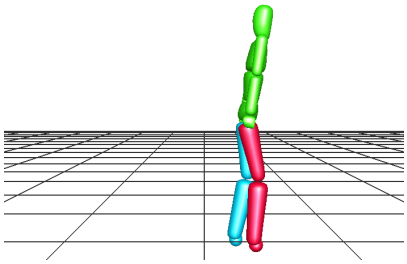
d) $t = 0.473$.



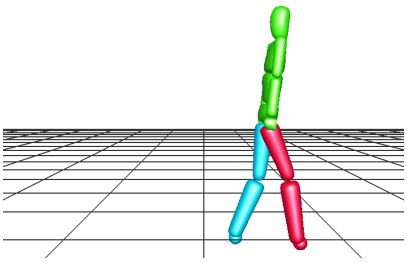
e) $t = 0.633$.



f) $t = 0.793$.



g) $t = 0.946$.



h) $t = T_2 = 1.108$: terminal posture.

Figure 7.12: Postures of the walker at various time points during the computed walking process belonging to $\underline{\boldsymbol{p}} = \underline{\boldsymbol{p}}_{\text{nom}}$. The right leg is represented by the red segments and the left leg by the blue segments. Visualization created using `MeshUp` [48].

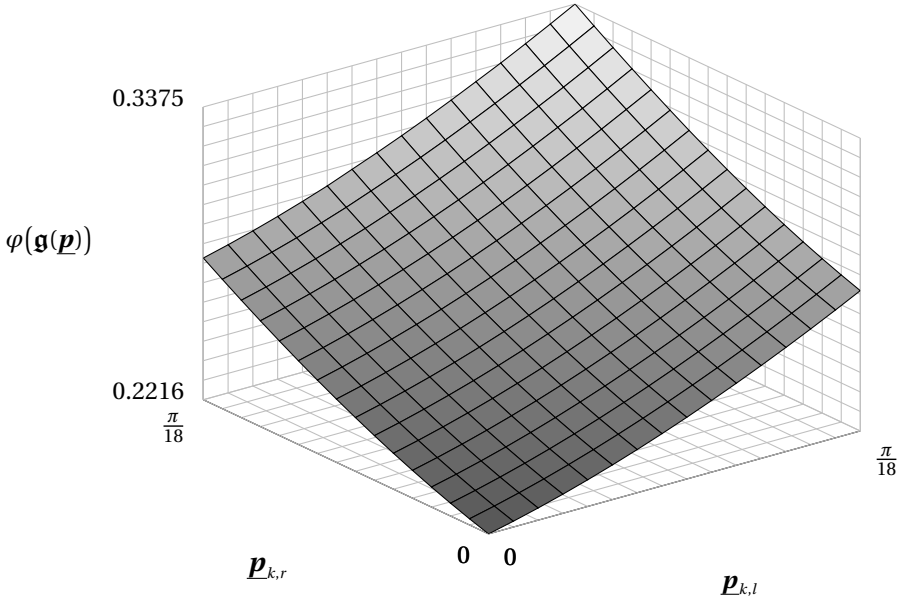


Figure 7.13: Objective function value $\varphi(\mathfrak{g}(\underline{\boldsymbol{p}}))$ of Problem (7.33) for parameters $\underline{\boldsymbol{p}}$ in the uncertainty set $\Omega_{\mathbf{p}}$.

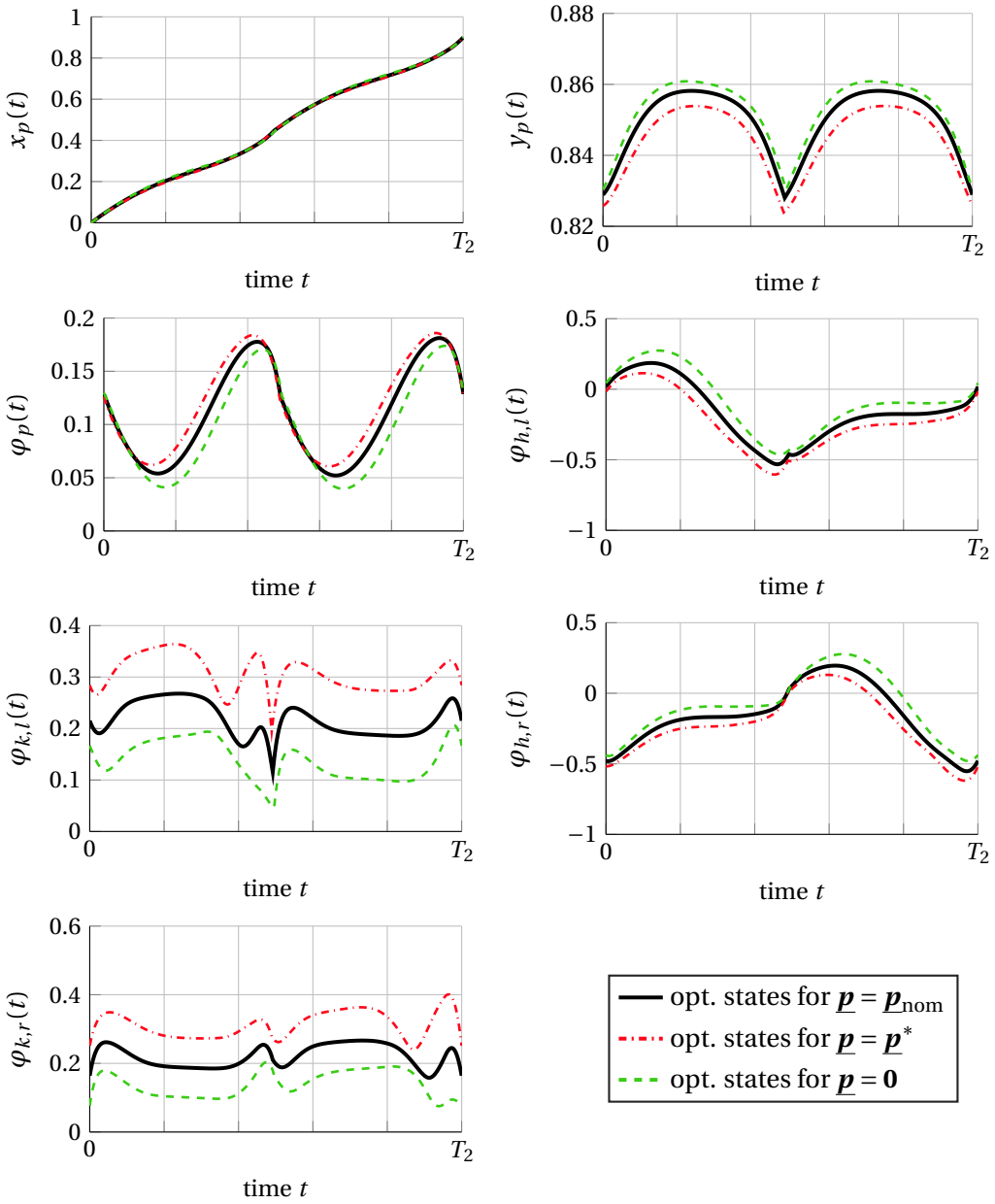


Figure 7.14: Comparison of possible treatment outcomes: values of generalized coordinates during the computed gaits. The black (solid) lines refer to the solution belonging to the nominal parameter $\underline{\boldsymbol{p}}_{\text{nom}}$, the red (dash-dotted) lines to the worst possible parameter realization $\underline{\boldsymbol{p}}^*$, and the green (dashed) lines to the best possible parameter realization $\underline{\boldsymbol{p}} = \mathbf{0}$. For each parameter the graphs are scaled to the same length.

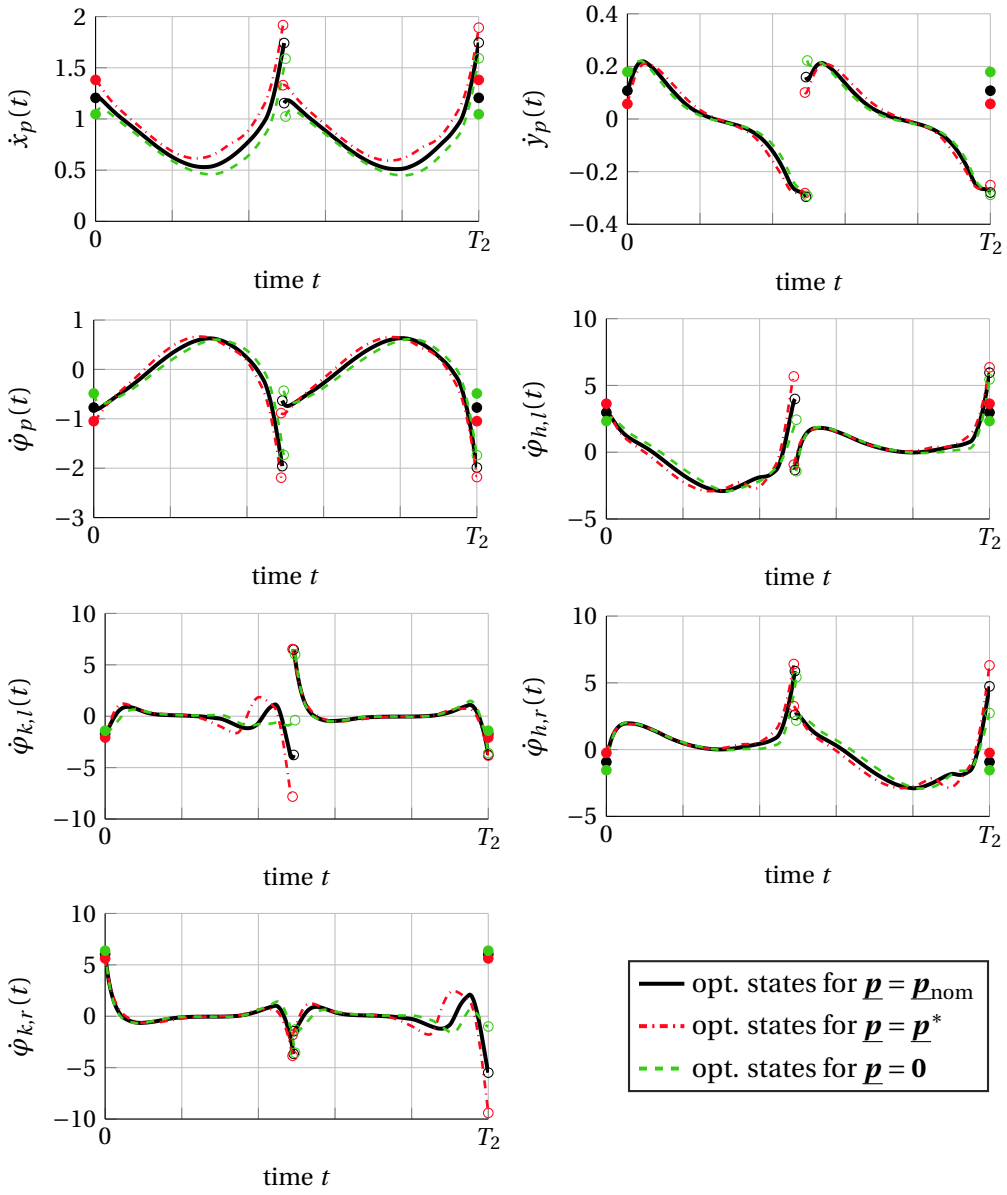


Figure 7.15: Comparison of possible treatment outcomes: values of generalized velocities during the computed gaits. Jumps occur at the phase transitions. The filled dots mark the initial and terminal values of the respective curves. The black (solid) lines refer to the solution belonging to the nominal parameter $\underline{\boldsymbol{p}}_{\text{nom}}$, the red (dash-dotted) lines to the worst possible parameter realization $\underline{\boldsymbol{p}}^*$, and the green (dashed) lines to the best possible parameter realization $\underline{\boldsymbol{p}} = \mathbf{0}$. For each parameter the graphs are scaled to the same length.

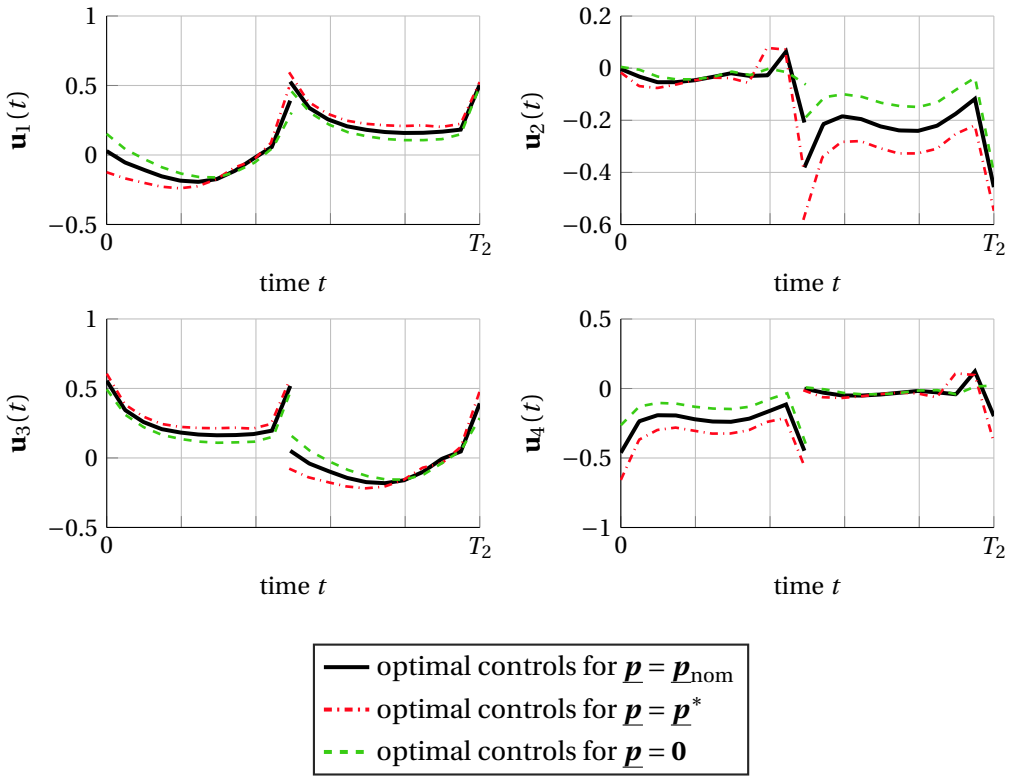


Figure 7.16: Comparison of possible treatment outcomes: values of $\mathbf{u}(t)$ – representing the controlled normalized generalized forces – during the computed gaits. The black (solid) lines refer to the solution belonging to the nominal parameter $\underline{\boldsymbol{p}}_{\text{nom}}$, the red (dash-dotted) lines to the worst possible parameter realization $\underline{\boldsymbol{p}}^*$, and the green (dashed) lines to the best possible parameter realization $\underline{\boldsymbol{p}} = \mathbf{0}$. For each parameter the graphs are scaled to the same length.

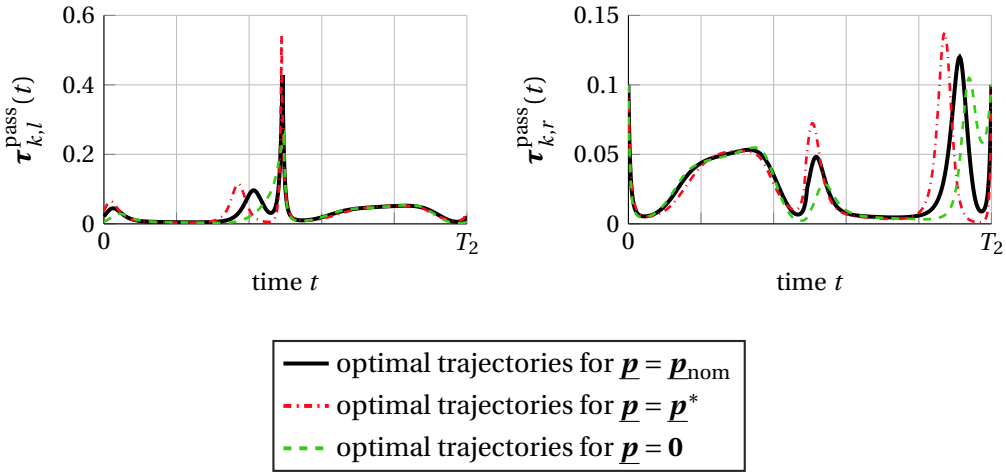


Figure 7.17: Comparison of possible treatment outcomes: values of sums of normalized passive reset forces $\boldsymbol{\tau}_{k,l}^{\text{pass}}(t)$ and $\boldsymbol{\tau}_{k,r}^{\text{pass}}(t)$ (cf. (7.34)) during the computed gaits. $\boldsymbol{\tau}_{k,l}^{\text{pass}}(t)$ and $\boldsymbol{\tau}_{k,r}^{\text{pass}}(t)$ act through the left and right knee joint, respectively. The black (solid) lines refer to the solution belonging to the nominal parameter $\underline{\boldsymbol{p}}_{\text{nom}}$, the red (dash-dotted) lines to the worst possible parameter realization $\underline{\boldsymbol{p}}^*$, and the green (dashed) lines to the best possible parameter realization $\underline{\boldsymbol{p}} = \mathbf{0}$. For each parameter the graphs are scaled to the same length.

Chapter 8

Conclusion

Summary

In this thesis, we developed mathematical models and numerical methods for the Optimal Control of constrained biomechanical Multi-Body Systems (MBSs), which can be employed in model-based treatment planning of Cerebral Palsy (CP). Our approach to model-based treatment planning is based on the following idea. We assume that the human gait can be modeled as a solution of an individually calibrated Optimal Control Problem (OCP) with phase-wise defined dynamics and possible jumps in the differential states at phase transition. Medical treatments can be encoded as changes of parameters, which enter the OCP. The resulting parametric OCP can then be employed to predict the outcome of medical interventions with regard to the resulting gait patterns. In the present work, we dealt with three aspects of model-based treatment planning regarding the described modeling environment.

1. In medical practice, one observes that foot-ground-contact patterns can change due to medical treatment – e. g., toe walking shall be corrected by interventions, such that the patient's heels touch the ground afterwards. Common approaches, in which the human gait is modeled as a solution of a multi-stage OCP with prescribed model phases, are not directly suitable to reflect this behavior. To incorporate the phenomenon into a predictive modeling environment, we developed a new numerical solution approach for OCPs with switches, switching costs, and jumps in the differential states, where the order of model phases is subject to optimization. To this end, we extended the Partial Outer Convexification (POC) approach, cf. [127], and the framework presented in [26] with regard to switchings costs and jumps in the differential states. We developed two kinds of so-called switching indicators, which can be employed as a trigger for events that are caused by certain changes of model-phases, as well as for the computation of switching costs. The developed free-phase approach is not tailored to gait modeling and can also be employed in other applications.

2. In principle, the result of a treatment can be assessed by examining the appropriate solution of the gait modeling parametric OCP with a suitable parameter value. However, in practice it is hardly possible to implement medical interventions exactly as planned, and uncertainties are expected in the accuracy of an implemented surgical treatment. The robustness of treatment plans is of fundamental importance to avoid negative surgery outcomes. Thus, all possible outcomes due to uncertainty have to be considered. In particular, the question arises whether a planned treatment is reasonable or not in view of the expected uncertainty. Here, one is interested in the post-operative treatment outcome *after* the patient's musculoskeletal system adapted functionally to the physiological changes due to a training period. To handle this issue, we proposed a bilevel OCP for robust treatment planning, where the upper level optimizes on the parameters encoding the medical treatment, and the lower level problem is given by a parametric OCP which models the gait. A global solution of the bilevel problem encodes a worst possible treatment and the corresponding gait pattern. The proposed modeling approach is suited for taking into account uncertainties in model-based treatment planning, but also for any other application in which one seeks for the optimal design of a process that can be described by a solution of an OCP.

3. Many interventions in CP management eventually aim at extending the ranges of motion of certain joints that are limited by the disorder. We modeled the range of motion of a joint by virtual bounds, encoded in model parameters, and so-called passive reset forces which appear in the neighborhood of these bounds. Subsequently, we proposed to model a treatment as a change of the parameters which represent the virtual bounds. This way, interventions can be translated into changing OCP dynamics. As a result, altered parameters – used for intervention modeling – yield altered gait patterns.

Limitations and Future Work

Medical Application

Two prerequisites are essential for the application of the developed and proposed methods for model-based treatment planning in practice. First, we assume a successful model calibration, i. e., the existence of a calibrated OCP and the knowledge of a corresponding solution that accurately models the human gait for the preoperative situation. We emphasize that the generation of such a calibrated model – in particular with an appropriate optimization criterion – is highly non-trivial, cf.

[71]. Second, we assume that the proposed way of modeling medical interventions or rather the resulting parametric OCP is suitable to reliably predict the outcome of an intervention (that is performed with perfect accuracy). In future work, the predictive character of such parametric OCPs has to be validated with real-world data. This is a challenging task. On the one hand, nowadays physicians often combine several medical treatments into one surgical event in order to avoid a high frequent hospitalization of the patients. Ideally, suitable data for the model evaluation of single treatments have to be found and made accessible. On the other hand, we are interested in the gait after functional adaption to the applied medical changes. The adaption can take several months, and often the subject of interventions are children, for whom the time period for rehabilitation plays a non negligible role regarding their physical development. Thus, for model-based treatment planning of children, physical development needs to be incorporated into the predictive model in a suitable way. Another open question concerns a proper assessment of surgery outcomes. In future work, suitable criteria have to be developed to quantify and evaluate the quality of post-operative gait patterns.

Mathematical models and methods

Besides the open questions regarding the medical application, we comment on possible next steps concerning the presented mathematical models and methods. Regarding the bilevel Optimal Control approach for worst-case treatment planning (see Chapter 6), we propose to consider further examples. In doing so, different candidates for assessment functions measuring the success of a treatment need to be identified and tested. Special attention has to be paid to applications of practical relevance in which multiple local maxima arise. Furthermore, a speed-up by the employment of gradient-based sequential methods [52, 53] or so-called all-at-once approaches [71] has to be tested.

In the presented solution approach to switched OCPs with switching costs and jumps (see Chapter 5), the employed switching costs hinder the occurrence of non-binary mode- and switching indicators in the solution of the arising relaxed and discretized OCPs. However, non-binary values cannot be excluded in general. Strategies for the handling of this issue in presence of state jumps need to be developed. Here, rounding algorithms in connection with the Partial Outer Convexification approach, e. g., [16, 17, 82, 127] (the former two taking into account switching costs), can serve as a starting point for further investigations. Furthermore, it would be desirable to enhance the sequential numerical solution approach in order to provide generalized

strategies for the successive grid adaption and the warm-starting procedure and to improve the detection of optimal switching points. For the grid adaption strategy, we expect that the removal of dispensable grid points can significantly reduce the computational effort. In addition, instead of the applied collocation discretization scheme, a multiple shooting discretization can be employed. This could improve feasibility properties, e. g., for problems with MBS dynamics and external contacts.

Concluding Remarks

In this thesis, we developed general mathematical approaches, whose possible fields of application include CP treatment planning. We believe that the presented ideas and approaches have a great potential to improve the robustness of treatment plans if the developed Optimal Control models or modifications of them are verified in practice for the predictive modeling of post-operative gait patterns. This way, the amount of negative treatment outcomes could be reduced noticeably. The present work made promising steps towards this goal.

Appendices

Appendix A

Simplest Walker Dynamics and Gait Model

A.1 Simplest Walker Dynamics

In this section, we set up the equations of motion for a basic MBS, the 2D "Simplest Walker" model as in [58, 134]. The model consists of three point masses which are connected by massless rods of length l, as displayed in Fig. A.1. We refer to the middle point mass as *head*, and to the other ones as *left foot* and *right foot*, according to the positions in Fig. A.1. The time-dependent positions of the respective bodies are $\big(x_h(t), y_h(t)\big)$ (head), $\big(x_l(t), y_l(t)\big)$ (left foot) and $\big(x_r(t), y_r(t)\big)$ (right foot), and their masses are given by M (head), and $m_l = m_r = m$ (left and right foot). The system parameters are

$$\bar{\mathbf{p}} = \begin{pmatrix} M & m & l \end{pmatrix}^T . \tag{A.1}$$

For the remainder of this section, we omit the arguments for time-dependent variables. During walking, different *modes* occur: Either the left foot is fixed to the ground (Mode 1), or the right foot is (Mode 2), where the term *fixed to the ground* means

$$x_j = \text{const.}, \tag{A.2a}$$
$$y_j = 0, \tag{A.2b}$$

$j \in \{l, r\}$, during the respective mode. A third mode, in which both feet are fixed to the ground, arises only momentarily as an isolated point of transition between the other modes, and thus is neglected. For each mode, Garcia et al. [58] treats the MBS as a double pendulum which can be described by means of two generalized coordinates. As explained in Section 4.1.1, we take another approach in which we consider the MBS without ground contact constraints first, and add the contact constraints as external contacts later.

We introduce generalized coordinates for the MBS without contact constraints. The head is the base segment of our walker. In general, a rigid body in 2D has three

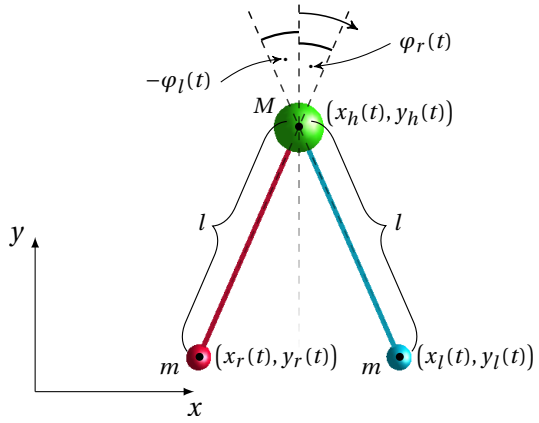


Figure A.1: The simplest walker modeled by a rigid multi-body system. Illustration created using `MeshUp` [48].

Degrees of Freedom (DoF), two of them describing the position of the center of mass of the body and one for its orientation resp. rotation in space. Since we only consider point masses in this example, the orientation of the head is not relevant, and two generalized coordinates (x_h, y_h) for the head, describing its position in space, are sufficient. The legs are connected to the head by revolute joints. Hence, per foot we need one additional generalized coordinate φ_l resp. φ_r, each describing the respective foot's rotation around the head. Altogether, the generalized coordinates are given by

$$\mathbf{q} = \begin{pmatrix} x_h & y_h & \varphi_l & \varphi_r \end{pmatrix}^T,$$

and the MBS has four DoF. In the following, we derive the equations of motion for the considered MBS. All physical quantities of interest are summarized in Tab. A.1.

In this example, the walker can accelerate its feet by applying torques acting through the rotational joints. The head itself is only accelerated indirectly as a result of the feet's interaction with the ground. In particular, the MBS is underactuated and the generalized forces $\boldsymbol{\tau}$ are given by

$$\boldsymbol{\tau} = \begin{pmatrix} 0 & 0 & \boldsymbol{\tau}_1^a & \boldsymbol{\tau}_2^a \end{pmatrix}^T \in \mathbb{R}^4. \tag{A.3}$$

Table A.1: Physical quantities occurring in the dynamics of the simplest walker.

M	mass of head
m	mass of each foot
l	length of legs
(x_h, y_h)	position of head in space
(x_l, y_l)	position of left foot in space
(x_r, y_r)	position of right foot in space
φ_l	angle describing the rotation of the left foot around the head
φ_r	angle describing the rotation of the right foot around the head

Equations of Motion

We derive the equations of motion. During Mode 1, the left foot is fixed to the ground at a position $(x_0, 0)^T$ (cf. (A.2)). The Cartesian coordinates of the left foot are given by

$$x_l = x_h - l\sin\varphi_l\,,$$
$$y_l = y_h - l\cos\varphi_l\,.$$

Therefore, in the notation of Section 4.1.1 we have

$$\mathbf{g}^1(\mathbf{q},\bar{\mathbf{p}}) = \begin{pmatrix} x_h - l\sin\varphi_l \\ y_h - l\cos\varphi_l \end{pmatrix} - \begin{pmatrix} x_0 \\ 0 \end{pmatrix},$$

the contact Jacobian is given by

$$\mathbf{G}^1(\mathbf{q},\bar{\mathbf{p}}) = \frac{\partial}{\partial \mathbf{q}}\mathbf{g}^1(\mathbf{q},\bar{\mathbf{p}}) = \begin{pmatrix} 1 & 0 & -l\cos\varphi_l & 0 \\ 0 & 1 & l\sin\varphi_l & 0 \end{pmatrix},$$

and we get

$$\boldsymbol{\gamma}^1(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}}) = -\left(\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{G}^1(\mathbf{q},\bar{\mathbf{p}})\right)\dot{\mathbf{q}} = \begin{pmatrix} -l\dot{\varphi}_l^2\sin\varphi_l \\ -l\dot{\varphi}_l^2\cos\varphi_l \end{pmatrix}.$$

We set up the equations of motion using Lagrangian mechanics. The Lagrangian L is given by the difference of the total kinetic energy T and the total potential energy U,

$$L = T - U\,.$$

Since our system contains redundant coordinates due to the external contacts (if one foot is fixed to the ground, the resulting system has two DoF left in contrast to the four coordinates we use), we apply Lagrange's equation of the first kind, given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right)^T - \left(\frac{\partial L}{\partial \mathbf{q}}\right)^T = \boldsymbol{\tau} + \mathbf{G}^1(\mathbf{q}, \bar{\mathbf{p}})^T \boldsymbol{\lambda}\,, \tag{A.4}$$

with constraint forces $\boldsymbol{\lambda} \in \mathbb{R}^2$. The total potential energy is given by

$$U = Mgy_h + mgy_l + mgy_r = (M+2m)gy_h - mgl(\cos\varphi_l + \cos\varphi_r)\,,$$

where $g > 0$ is the absolute value of the gravitational acceleration. Since we only consider point masses, for each body the rotational energy is zero. Hence, the total kinetic energy is given by

$$\begin{aligned} T &= \frac{1}{2}M(\dot{x}_h^2 + \dot{y}_h^2) + \frac{1}{2}m({\dot{x}_l}^2 + {\dot{y}_l}^2) + \frac{1}{2}m({\dot{x}_r}^2 + {\dot{y}_r}^2) \\ &= \frac{M+2m}{2}(\dot{x}_h^2 + \dot{y}_h^2) + \frac{1}{2}ml^2(\dot{\varphi}_l^2 + \dot{\varphi}_r^2) \\ &\quad + ml\dot{\varphi}_l\left(-\dot{x}_h\cos\varphi_l + \dot{y}_h\sin\varphi_l\right) + ml\dot{\varphi}_r\left(-\dot{x}_h\cos\varphi_r + \dot{y}_h\sin\varphi_r\right). \end{aligned}$$

Applying Lagrange's Equation (A.4), we eventually obtain

$$\begin{aligned} \boldsymbol{\tau} + \mathbf{G}^1(\mathbf{q}, \bar{\mathbf{p}})^T \boldsymbol{\lambda} = &\begin{pmatrix} M+2m & 0 & -ml\cos\varphi_l & -ml\cos\varphi_r \\ 0 & M+2m & ml\sin\varphi_l & ml\sin\varphi_r \\ -ml\cos\varphi_l & ml\sin\varphi_l & ml^2 & 0 \\ -ml\cos\varphi_r & ml\sin\varphi_r & 0 & ml^2 \end{pmatrix} \ddot{\mathbf{q}} \\ &+ \begin{pmatrix} ml\dot{\varphi}_l^2\sin\varphi_l + ml\dot{\varphi}_r^2\sin\varphi_r \\ ml\dot{\varphi}_l^2\cos\varphi_l + ml\dot{\varphi}_r^2\cos\varphi_r + (M+2m)g \\ mgl\sin\varphi_l \\ mgl\sin\varphi_r \end{pmatrix}. \end{aligned}$$

By setting

$$\mathbf{H}(\mathbf{q}, \bar{\mathbf{p}}) = \begin{pmatrix} M+2m & 0 & -ml\cos\varphi_l & -ml\cos\varphi_r \\ 0 & M+2m & ml\sin\varphi_l & ml\sin\varphi_r \\ -ml\cos\varphi_l & ml\sin\varphi_l & ml^2 & 0 \\ -ml\cos\varphi_r & ml\sin\varphi_r & 0 & ml^2 \end{pmatrix}$$

and

$$\mathbf{C}(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}}) = \begin{pmatrix} ml\dot{\varphi}_l^2 \sin\varphi_l + ml\dot{\varphi}_r^2 \sin\varphi_r \\ ml\dot{\varphi}_l^2 \cos\varphi_l + ml\dot{\varphi}_r^2 \cos\varphi_r + (M+2m)g \\ mgl\sin\varphi_l \\ mgl\sin\varphi_r \end{pmatrix},$$

we can put the equations of motion in Form (4.6).

During Mode 2, the right foot is fixed to the ground. In the belonging equations of motion, $\mathbf{H}(\mathbf{q},\bar{\mathbf{p}})$ and $\mathbf{C}(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}})$ stay the same, but the contact Jacobian and accordingly $\boldsymbol{\gamma}(\cdot)$ need to be adapted. We get

$$\mathbf{G}^2(\mathbf{q},\bar{\mathbf{p}}) = \begin{pmatrix} 1 & 0 & 0 & -l\cos\varphi_r \\ 0 & 1 & 0 & l\sin\varphi_r \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma}^2(\mathbf{q},\dot{\mathbf{q}},\bar{\mathbf{p}}) = \begin{pmatrix} -l\dot{\varphi}_r^2\sin\varphi_r \\ -l\dot{\varphi}_r^2\cos\varphi_r \end{pmatrix},$$

which again enables us to state the equations of motion for mode 2 in Form (4.6).

From what we have seen before, we have the required quantities at hand to put the equations for the transfer of generalized velocities at mode transitions, cf. Equation (4.7), which occur whenever the foot being in contact with ground (i. e., the external contact) changes. Hence, we provided a complete description of the simplest walker dynamics while walking. An Optimal Control model for generating a gait of the simplest walker is set up in Appendix A.2.

Towards the Equations of Motion in Explicit Form

Till now, we stated the equations of motion (including the jump conditions) in implicit form. To compute the quantities $\ddot{\mathbf{q}}(\mathbf{q},\dot{\mathbf{q}},\boldsymbol{\tau},\bar{\mathbf{p}})$ or $\dot{\mathbf{q}}^+(\mathbf{q},\dot{\mathbf{q}}^-,\bar{\mathbf{p}})$ (using the notation from Section 4.1.1) in case of a jump, respectively, a system of linear equations needs to be solved. However, there are situations in which it is advantageous to have an explicit formula for $\ddot{\mathbf{q}}(\cdot)$ and $\dot{\mathbf{q}}^+(\cdot)$ at hand, for instance if we want to compute derivatives of these quantities using automatic differentiation tools, such as `Adol-C` [148]. The main work in deriving an explicit formula from the implicit formulation is to compute the inverse of the matrices

$$\mathbf{M}_j \overset{\text{def}}{=} \begin{pmatrix} \mathbf{H}(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{G}^j(\mathbf{q},\bar{\mathbf{p}})^T \\ \mathbf{G}^j(\mathbf{q},\bar{\mathbf{p}}) & \mathbf{0} \end{pmatrix}$$

for $j = \{1,2\}$. As the $\mathbf{M}_j$ are symmetric, the same holds for their inverses. We state their upper triangular parts in the following. For this, let

$$d = M + m\sin^2(\varphi_r - \varphi_l)\,.$$

We denote the entry of $\left(\mathbf{M}_j\right)^{-1}$ which is located in the l-th column of the k-th row by $\left(\mathbf{M}_j\right)^{-1}(k,l)$.

Inverse of $\mathbf{M}_1$

For $(\mathbf{M}_1)^{-1}$, for the first row we get the entries

$$\begin{aligned}
(\mathbf{M}_1)^{-1}(1,1) &= \frac{1}{d}\cos^2\varphi_l\,,\\
(\mathbf{M}_1)^{-1}(1,2) &= -\frac{1}{2d}\sin(2\varphi_l)\,,\\
(\mathbf{M}_1)^{-1}(1,3) &= \frac{1}{ld}\cos\varphi_l\,,\\
(\mathbf{M}_1)^{-1}(1,4) &= \frac{1}{ld}\cos\varphi_l\cos(\varphi_r-\varphi_l)\,,\\
(\mathbf{M}_1)^{-1}(1,5) &= \frac{1}{d}\sin\varphi_l\left[\sin\varphi_l\left(M+m\cos^2\varphi_r\right)-\frac{1}{2}m\cos\varphi_l\sin(2\varphi_r)\right]\,,\\
(\mathbf{M}_1)^{-1}(1,6) &= \frac{1}{d}\cos\varphi_l\left[\sin\varphi_l\,(m+M)-m\sin\varphi_r\cos(\varphi_r-\varphi_l)\right]\,,
\end{aligned}$$

for the second row we get

$$\begin{aligned}
(\mathbf{M}_1)^{-1}(2,2) &= \frac{1}{d}\sin^2\varphi_l\,,\\
(\mathbf{M}_1)^{-1}(2,3) &= -\frac{1}{ld}\sin\varphi_l\,,\\
(\mathbf{M}_1)^{-1}(2,4) &= -\frac{1}{ld}\sin\varphi_l\cos(\varphi_r-\varphi_l)\,,\\
(\mathbf{M}_1)^{-1}(2,5) &= \frac{1}{d}\sin\varphi_l\left[\cos\varphi_l(m+M)-m\cos\varphi_r\cos(\varphi_r-\varphi_l)\right]\,,\\
(\mathbf{M}_1)^{-1}(2,6) &= \frac{1}{d}\cos\varphi_l\left[M\cos\varphi_l+m\sin\varphi_r\sin(\varphi_r-\varphi_l)\right]\,,
\end{aligned}$$

for the third row we get

$$(\mathbf{M}_1)^{-1}(3,3) = \frac{1}{l^2 d}\,,$$

$$(\mathbf{M}_1)^{-1}(3,4) = \frac{1}{l^2 d}\cos(\varphi_r - \varphi_l),$$
$$(\mathbf{M}_1)^{-1}(3,5) = -\frac{1}{ld}\left[M\cos\varphi_l + m\sin\varphi_r \sin(\varphi_r - \varphi_l)\right],$$
$$(\mathbf{M}_1)^{-1}(3,6) = \frac{1}{ld}\left[\sin\varphi_l(m+M) - m\sin\varphi_r\cos(\varphi_r - \varphi_l)\right],$$

for the fourth row we get

$$(\mathbf{M}_1)^{-1}(4,4) = \frac{1}{ml^2 d}(M+m),$$
$$(\mathbf{M}_1)^{-1}(4,5) = -\frac{1}{ld}\sin\varphi_l \sin(\varphi_r - \varphi_l)(m+M),$$
$$(\mathbf{M}_1)^{-1}(4,6) = -\frac{1}{ld}\cos\varphi_l \sin(\varphi_r - \varphi_l)(m+M),$$

for the fifth row we get

$$(\mathbf{M}_1)^{-1}(5,5) = -\frac{1}{d}\left[m^2\sin^2(\varphi_r - \varphi_l) + Mm(1+\sin^2\varphi_l) + M^2\sin^2\varphi_l\right],$$
$$(\mathbf{M}_1)^{-1}(5,6) = -\frac{1}{2d}\sin(2\varphi_l)M(m+M),$$

and for the last row

$$(\mathbf{M}_1)^{-1}(6,6) = -\frac{1}{d}\left[m^2\sin^2(\varphi_r - \varphi_l) + Mm(1+\cos^2\varphi_l) + M^2\cos^2\varphi_l\right].$$

Inverse of $\mathbf{M}_2$

For $(\mathbf{M}_2)^{-1}$, for the first row we get

$$(\mathbf{M}_2)^{-1}(1,1) = \frac{1}{d}\cos^2\varphi_r,$$
$$(\mathbf{M}_2)^{-1}(1,2) = -\frac{1}{2d}\left[\sin(2\varphi_r)\right],$$
$$(\mathbf{M}_2)^{-1}(1,3) = \frac{1}{ld}\cos\varphi_r\cos(\varphi_r - \varphi_l),$$
$$(\mathbf{M}_2)^{-1}(1,4) = \frac{1}{ld}\cos\varphi_r,$$
$$(\mathbf{M}_2)^{-1}(1,5) = \frac{1}{d}\sin\varphi_r\left[\sin\varphi_r(M + m\cos^2\varphi_l) - \frac{1}{2}m\cos\varphi_r\sin(2\varphi_l)\right],$$
$$(\mathbf{M}_2)^{-1}(1,6) = \frac{1}{d}\cos\varphi_r\left[M\sin\varphi_r + m\cos\varphi_l\sin(\varphi_r - \varphi_l)\right],$$

for the second row we get

$$
\begin{aligned}
(\mathbf{M}_2)^{-1}(2,2) &= \frac{1}{d}\sin^2\varphi_r\,,\\
(\mathbf{M}_2)^{-1}(2,3) &= -\frac{1}{ld}\sin\varphi_r\cos(\varphi_r-\varphi_l)\,,\\
(\mathbf{M}_2)^{-1}(2,4) &= -\frac{1}{ld}\sin\varphi_r\,,\\
(\mathbf{M}_2)^{-1}(2,5) &= \frac{1}{d}\sin\varphi_r\left[M\cos\varphi_r - m\sin\varphi_l\sin(\varphi_r-\varphi_l)\right],\\
(\mathbf{M}_2)^{-1}(2,6) &= \frac{1}{d}\cos\varphi_r\left[\cos\varphi_r(m\sin^2\varphi_l+M)-\frac{1}{2}m\sin(2\varphi_l)\sin\varphi_r\right],
\end{aligned}
$$

for the third row we get

$$
\begin{aligned}
(\mathbf{M}_2)^{-1}(3,3) &= \frac{1}{ml^2d}(M+m)\,,\\
(\mathbf{M}_2)^{-1}(3,4) &= \frac{1}{l^2d}\cos(\varphi_r-\varphi_l)\,,\\
(\mathbf{M}_2)^{-1}(3,5) &= \frac{1}{ld}\sin\varphi_r\sin(\varphi_r-\varphi_l)(m+M)\,,\\
(\mathbf{M}_2)^{-1}(3,6) &= \frac{1}{ld}\cos\varphi_r\sin(\varphi_r-\varphi_l)(m+M)\,,
\end{aligned}
$$

for the fourth row we get

$$
\begin{aligned}
(\mathbf{M}_2)^{-1}(4,4) &= \frac{1}{l^2d}\,,\\
(\mathbf{M}_2)^{-1}(4,5) &= -\frac{1}{ld}\left[M\cos\varphi_r - m\sin\varphi_l\sin(\varphi_r-\varphi_l)\right],\\
(\mathbf{M}_2)^{-1}(4,6) &= \frac{1}{ld}\left[M\sin\varphi_r + m\cos\varphi_l\sin(\varphi_r-\varphi_l)\right],
\end{aligned}
$$

for the fifth row we get

$$
\begin{aligned}
(\mathbf{M}_2)^{-1}(5,5) &= -\frac{1}{d}\left[m^2\sin^2(\varphi_r-\varphi_l)+Mm(1+\sin^2\varphi_r)+M^2\sin^2\varphi_r\right],\\
(\mathbf{M}_2)^{-1}(5,6) &= -\frac{1}{2d}\sin(2\varphi_r)M(M+m)\,,
\end{aligned}
$$

and for the last row

$$
(\mathbf{M}_2)^{-1}(6,6) = -\frac{1}{d}\left[m^2\sin^2(\varphi_r-\varphi_l)+Mm(1+\cos^2\varphi_r)+M^2\cos^2\varphi_r\right].
$$

A.2 A Multi-Stage Optimal Control Model for a Simplest Walker's Gait

To give an illustrative example for the approach described in Section 4.1.2, inspired by [48, sec. 5.4] and [134] we provide a gait model for the simplest walker MBS from Appendix A.1, where we already derived the equations of motion of the MBS. In the following, we use the notation from Section 4.1.2 and the variables introduced in Appendix A.1.

The MBS is controlled directly by its actuated generalized forces, see (A.3). Thus, we have $\mathbf{u}(\cdot) \in \mathbb{R}^2$ and

$$\boldsymbol{\tau}(\mathbf{u}(t), \mathbf{p}) = \begin{pmatrix} 0 \\ 0 \\ \mathbf{u}_1(t) \\ \mathbf{u}_2(t) \end{pmatrix}.$$

The simplest walker's gait cycle comprises two phases and corresponding phase transitions. In the following, we describe the phases and holding constraints in more detail, and assemble the resulting OCP.

Initial Position, Posture, and Velocities

In the beginning of the gait cycle, we impose constraints on the position, posture, and velocities of the walker, respectively, namely

$$x_h(T_0) = y_l(T_0) = y_r(T_0) = 0\,, \tag{A.5a}$$
$$0.2l \leq x_l(T_0) - x_r(T_0) \leq 0.8l\,, \tag{A.5b}$$
$$-\pi \leq \varphi_l(T_0), \varphi_r(T_0) \leq \pi\,, \tag{A.5c}$$
$$-5l \leq \dot{x}_h(T_0), \dot{y}_h(T_0), \dot{\varphi}_l(T_0), \dot{\varphi}_r(T_0) \leq 5l\,, \tag{A.5d}$$

which force the walker to start in a reasonable initial configuration. The Constraints (A.5b) and (A.5d) are scaled with the length of the legs to account for parameter dependencies. The resulting walking motion will be determined by optimization later. In order to generate a homogeneous walking pattern in which the initial (as well as the terminal, cf. Constraints (A.11) and (A.12)) posture and velocities do not stand out, respectively, we leave some freedom for optimizing the initial posture and velocities.

Phase 1: Left Foot Fixed to the Ground

During the first phase, the left foot is fixed to the ground, i. e.

$$x_l(t) = \text{const.}, \quad t \in \mathcal{T}_1, \tag{A.6a}$$
$$y_l(t) = 0, \quad t \in \mathcal{T}_1, \tag{A.6b}$$

and hence

$$\dot{x}_l(t) = 0, \quad t \in \mathcal{T}_1, \tag{A.7a}$$
$$\dot{y}_l(t) = 0, \quad t \in \mathcal{T}_1. \tag{A.7b}$$

The corresponding differential equation has been stated in Appendix A.1. As explained in Section 4.1.2, we treat it as an Ordinary Differential Equation (ODE)

$$\dot{\mathbf{x}}(t) = \mathbf{f}^{\mathrm{I}}\left(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}\right), \quad t \in \mathcal{T}_1,$$

with 8 differential states

$$\mathbf{x}(\cdot) = \begin{pmatrix} x_h(\cdot) & y_h(\cdot) & \varphi_l(\cdot) & \varphi_r(\cdot) & \dot{x}_h(\cdot) & \dot{y}_h(\cdot) & \dot{\varphi}_l(\cdot) & \dot{\varphi}_r(\cdot) \end{pmatrix}^T. \tag{A.8}$$

The differential equation ensures, that the second time derivative of (A.6) vanishes for $t \in \mathcal{T}_1$, cf. Section 4.1.1. Hence, it suffices to demand (A.7) and (A.6) to hold only for $t = T_0$ instead of the whole interval $\mathcal{T}_1$. Equations (A.6a) and (A.7a) are equivalent, and at $t = T_0$ (A.6b) is already satisfied by the initial constraints. Thus, it remains to add (A.7), evaluated at $t = T_0$, to the point constraints.

In order to generate a natural looking walking-like motion, we demand the head of the walker to stay above a certain level, and we want the feet not to penetrate the ground. However, since the considered stick man is not able to walk in a reasonable way without penetrating the ground due to its stiff legs, we set up a tolerance $\varepsilon_{\text{tol}} = 0.1l$, and demand the path constraints

$$-\varepsilon_{\text{tol}} \leq y_r(t), \quad 0.8l \leq y_h(t), \quad t \in \mathcal{T}_1$$

to hold.

Phase Transition: Right Foot Hits the Ground

Phase 1 ends, when the right foot hits the ground from above (i. e. with a negative vertical component of the velocity). Thus, we demand

$$y_r(T_1) = 0\,, \quad \dot{y}_r(T_1) \leq 0\,. \tag{A.9}$$

The transition of velocities can be expressed using a jump function,

$$\mathbf{x}(T_1^+) = \boldsymbol{\Delta}^1(\mathbf{x}(T_1^-), \mathbf{p})\,,$$

transferring the differential states instantly before the jump to the differential states instantly after the jump, cf. Equation (4.7).

Phase 2: Right Foot Fixed to the Ground

In Phase 2, the differential equation in ODE form is given by

$$\dot{\mathbf{x}}(t) = \mathbf{f}^2\left(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}\right)\,, \quad t \in \mathcal{T}_2\,.$$

Due to the Conditions (A.9) demanded at the end of Phase 1 and the nature of the jump function, the position of the right foot is fixed to the ground at the beginning of Phase 2, and the corresponding derivatives (meaning the counterpart of (A.7)) equal zero, too. Thus, no additional point constraints are needed. Similar to Phase 1, we furthermore demand the path constraints

$$-\varepsilon_{\text{tol}} \leq y_l(t)\,, \quad 0.8l \leq y_h(t)\,, \quad t \in \mathcal{T}_2$$

to hold.

Phase Transition: Left Foot Hits the Ground

Similar to the first phase transition described above, the end of Phase 2 is marked by the conditions

$$y_l(T_2) = 0\,, \quad \dot{y}_l(T_2) \leq 0\,, \tag{A.10}$$

and the transition of velocities can be expressed in form

$$\mathbf{x}(T_2^+) = \boldsymbol{\Delta}^2(\mathbf{x}(T_2^-), \mathbf{p})\,.$$

We remark, that this phase transition at the end of the walking process is not necessary to describe the dynamics, but simplifies the formulation of periodic constraints, see Constraints (A.12).

Terminal Position and Posture

By demanding

$$x_{\text{end}} \leq x_h(T_2)$$

for a $x_{\text{end}} > 0$, we force the walker to move a sufficient distance in positive x-direction. Here, x_{end} needs to be chosen suitable according to the leg length l and the desired step length. We find $x_{\text{end}} = l$ to be a suitable choice.

Due to (A.10) and the ground contact in Phase 2, at $t = T_2$ both feet are in touch with the ground. For the cyclicity of the posture of the walker it is sufficient to demand

$$\begin{aligned} \varphi_l(T_2) &= \varphi_l(T_0), \\ \varphi_r(T_2) &= \varphi_r(T_0), \end{aligned} \tag{A.11}$$

as this already implies $y(T_2) = y(T_0)$. Similar, for the velocities of the system we demand

$$\begin{aligned} \dot{x}_h(T_2) &= \dot{x}_h(T_0), \\ \dot{y}_h(T_2) &= \dot{y}_h(T_0), \\ \dot{\varphi}_l(T_2) &= \dot{\varphi}_l(T_0), \\ \dot{\varphi}_r(T_2) &= \dot{\varphi}_r(T_0). \end{aligned} \tag{A.12}$$

The Resulting Optimal Control Problem

All above constraints can be expressed in terms of the differential states $\mathbf{x}(\cdot)$, see (A.8), and the system parameters $\mathbf{p}$, see (A.1). We summarize the constraints in path and point constraints using functions $\mathbf{c}^1(\cdot)$, $\mathbf{c}^2(\cdot)$, $\mathbf{r}^{\text{eq}}(\cdot)$, and $\mathbf{r}^{\text{ieq}}(\cdot)$ as occurring in Problem (4.13).

It remains to set up an objective function. For instance, a compromise between minimum walking process duration (which is related to maximum walking speed) and minimum torques (which encodes the mechanical effort), combined in a weighted

objective function of the form

$$\gamma T_2 + (1-\gamma)\int_{T_0}^{T_2} \mathbf{u}_1^2(t) + \mathbf{u}_2^2(t)\mathrm{d}t\,, \tag{A.13}$$

with $\gamma \in [0,1]$, leads to natural- and smooth-looking gaits. A proper choice of γ depends on the specific application.

Altogether, we obtain an OCP of Form (4.13) whose solutions model gait patterns of the simplest walker.

Appendix B

Proofs

B.1 Proofs for Chapter 5

B.1.1 Proof of Lemma 5.2

Let $w \in PC_{\tilde{\delta}}(\mathcal{T}, \{1, \dots, n\})$. We set

$$\boldsymbol{\omega}(t) = \begin{pmatrix} \delta_{1\, w(t)} \\ \vdots \\ \delta_{n\, w(t)} \end{pmatrix}$$

using the *Kronecker delta*. Then obviously we have $\boldsymbol{\omega}(\cdot) \in PC_{\tilde{\delta}}(\mathcal{T}, \{0,1\}^n)$. Let $t \in \mathcal{T}$ and $w(t) = j'$ for some $j' \in \{1, \dots, n\}$. Then $\sum_{j=1}^{n} \boldsymbol{\omega}_j(t) = \boldsymbol{\omega}_{j'}(t) = 1$, ergo $\boldsymbol{\omega}(t) \in \mathbb{S}^n$, and $\sum_{j=1}^{n} \boldsymbol{\omega}_j(t) \cdot j = \boldsymbol{\omega}_{j'}(t) \cdot j' = j' = w(t)$. Hence, $w = \varphi(\boldsymbol{\omega})$ and φ is surjective.

To show the injectivity, let $\boldsymbol{\omega}^1(\cdot), \boldsymbol{\omega}^2(\cdot) \in PC_{\tilde{\delta}}(\mathcal{T}, \{0,1\}^n) \cap \Omega^n$ with $\boldsymbol{\omega}^1(\cdot) \neq \boldsymbol{\omega}^2(\cdot)$. Then there is a $t \in \mathcal{T}$ and distinct indices $j_1, j_2 \in \{1, ..., n\}$ such that $\boldsymbol{\omega}^1_{j_1}(t) = 1 = \boldsymbol{\omega}^2_{j_2}(t)$, and all other entries are zero, respectively. Hence

$$\varphi\left(\boldsymbol{\omega}^1\right)(t) = \sum_{j=1}^{n} \boldsymbol{\omega}^1_j(t) \cdot j = j_1 \neq j_2 = \varphi\left(\boldsymbol{\omega}^2\right)(t),$$

which finishes the proof.

B.1.2 Proof of Proposition 5.3

For the first direction, let $(\mathbf{x}, \mathbf{u}, w)$ be feasible for Problem (5.1). We define

$$\boldsymbol{\omega}(t) = \begin{pmatrix} \delta_{1\, w(t)} \\ \vdots \\ \delta_{n\, w(t)} \end{pmatrix}.$$

By the proof of Lemma 5.2 we know $\boldsymbol{\omega} = \varphi^{-1}(w)$, and by construction we have $\mathcal{S}(w) = \mathcal{S}(\boldsymbol{\omega})$. Thus the values of both cost functions coincide. Since $\boldsymbol{\omega}_j(t) = \delta_{j\,w(t)}$, we have

$$\mathbf{f}^j(\mathbf{x}(t), \mathbf{u}(t)) = \sum_{j=1}^{n} \boldsymbol{\omega}_j(t) \cdot \mathbf{f}^j(\mathbf{x}(t), \mathbf{u}(t)) \quad \text{if } w(t) = j\,,$$

and the right-hand sides of the differential equations of both problems coincide (almost everywhere). It remains to show that (5.2e) holds. Let $t \in \mathcal{T}$ such that $w(t) = j'$ and $\mathbf{0} \geq \mathbf{c}^{j'}(\mathbf{x}(t), \mathbf{u}(t))$. Then $\boldsymbol{\omega}_{j'}(t) = 1$ and $\boldsymbol{\omega}_j(t) = 0$ for all $j \neq j'$. Therefore

$$\mathbf{0} \geq \boldsymbol{\omega}_j(t) \cdot \mathbf{c}^j(\mathbf{x}(t), \mathbf{u}(t))$$

indeed holds for all $j \in \{1, \ldots, n\}$. The proof for the reverse direction works similarly.

B.1.3 Proof of Proposition 5.4

We take a look at the first statement. Let $(\mathbf{x}(\cdot), \mathbf{u}(\cdot), \boldsymbol{\omega}(\cdot))$ be feasible for Problem (5.2). For Problem (5.6), the only feasible $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$ are uniquely determined by (5.6d) and (5.6e). For each $t_s \in \mathcal{S}(\boldsymbol{\omega})$, we claim

$$\boldsymbol{\Delta}\left(\mathbf{x}(t_s^-), \left(\boldsymbol{\theta}_{j_1,j_2}(t_s)\right)_{j_1,j_2}\right) = \boldsymbol{\Delta}_{j_1,j_2}\left(\mathbf{x}(t_s^-)\right) \quad \text{if } j_1 \rightarrow_{\boldsymbol{\omega}} j_2 \text{ at } t_s\,. \tag{B.1}$$

Indeed, if $j_1 \rightarrow_{\boldsymbol{\omega}} j_2$ at t_s, we have $\boldsymbol{\omega}_{j_1}(t_s^-) = \boldsymbol{\omega}_{j_2}(t_s^+) = 1$, $\boldsymbol{\omega}_{j'}(t_s^-) = 0$ for all $j' \neq j_1$ and $\boldsymbol{\omega}_{j'}(t_s^+) = 0$ for all $j' \neq j_2$. Due to (5.6e), therefore

$$\boldsymbol{\theta}_{j_1',j_2'}(t_s) = \begin{cases} 1 & \text{if } j_1' = j_1 \text{ and } j_2' = j_2\,, \\ 0 & \text{else}\,, \end{cases}$$

and (B.1) holds as one easily verifies. For $t \in \mathcal{G} \setminus \mathcal{S}(\boldsymbol{\omega})$ on the other hand, we have $\boldsymbol{\theta}_{j_1,j_2}(t) = 0$ for all $j_1 \neq j_2$ according to (5.6e). This yields

$$\boldsymbol{\Delta}\left(\mathbf{x}(t^-), \left(\boldsymbol{\theta}_{j_1,j_2}(t)\right)_{j_1,j_2}\right) = \mathbf{x}(t^-)\,,$$

and no jump in the differential states occurs, as desired. Therefore $\left(\mathbf{x}(\cdot), \mathbf{u}(\cdot), \boldsymbol{\omega}(\cdot), (\boldsymbol{\theta}_{j_1,j_2}(\cdot))_{j_1 \neq j_2}\right)$ is feasible for Problem (5.6), and the objective function values coincide because of (5.4).

The second statement can be proven in a similar fashion.

B.1.4 Proof of Proposition 5.5

For every $\alpha_1, \alpha_2 \in \mathbb{R}$, there exists a $\beta \in [0,1]$ such that

$$\min(\alpha_1, \alpha_2) = \beta\alpha_1 + (1-\beta)\alpha_2 .$$

Using this and Proposition 5.4, the statement follows.

B.1.5 Proof of Proposition 5.6

We take a look at ϕ_{subs} first:

$$\phi_{\text{subs}}(\mathbf{b}, \mathbf{c}) = \sum_{j=1}^{n} \min(\mathbf{c}_j, 1 - \mathbf{b}_j) \leq \sum_{j=1}^{n} \mathbf{c}_j = 1 ,$$

since $\mathbf{c} \in \operatorname{conv}(\mathbb{S}^n)$. We define

$$J_1 \stackrel{\text{def}}{=} \left\{ j \in \{1, \ldots, n\} \,\middle|\, \mathbf{b}_j + \mathbf{c}_j \leq 1 \right\} \quad \text{and} \quad J_2 \stackrel{\text{def}}{=} \{1, \ldots, n\} \setminus J_1 .$$

Then

$$\phi_{\text{subs}}(\mathbf{b}, \mathbf{c}) = \sum_{j=1}^{n} \min(\mathbf{c}_j, 1 - \mathbf{b}_j) = \sum_{j \in J_1} \mathbf{c}_j + \sum_{j \in J_2} (1 - \mathbf{b}_j) .$$

We conclude: If $\mathbf{b}_j + \mathbf{c}_j \leq 1$ for all j, i. e. $J_1 = \{1, \ldots, n\}$ and $J_2 = \emptyset$, then $\phi_{\text{subs}}(\mathbf{b}, \mathbf{c}) = 1$. For ϕ_{inv} the proof works similarly.

For ϕ_{omni} we have

$$\phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) = \sum_{\substack{j_1, j_2 = 1 \\ j_1 \neq j_2}}^{n} \min(\mathbf{b}_{j_1}, \mathbf{c}_{j_2}) \leq \sum_{\substack{j_1, j_2 = 1 \\ j_1 \neq j_2}}^{n} \mathbf{b}_{j_1} = (n-1) \sum_{j_1=1}^{n} \mathbf{b}_{j_1} = n - 1 , \tag{B.2}$$

and if we set $\mathbf{b}_j = \mathbf{c}_j = \frac{1}{n}$ for all j, the inequality in (B.2) becomes an equality, which closes the proof.

B.1.6 Proof of Proposition 5.7

The first statement is obviously true. For the proof of the second statement, we first take a look at "$\Leftarrow$":

If $\mathbf{b},\mathbf{c} \in \mathbb{S}^n$ and $\mathbf{b} = \mathbf{c}$, for every j we have either $\mathbf{b}_j = \mathbf{c}_j = 1$ or $\mathbf{b}_j = \mathbf{c}_j = 0$. Therefore $\phi_{\text{inv}}(\mathbf{b},\mathbf{c}) = \phi_{\text{subs}}(\mathbf{b},\mathbf{c}) = 0$ from the function's definitions. Also for every pair (j_1, j_2) with $j_1 \neq j_2$, either $\mathbf{b}_{j_1} = 0$ or $\mathbf{c}_{j_2} = 0$, and hence $\phi_{\text{omni}}(\mathbf{b},\mathbf{c}) = 0$.

"⇒": We take a look at ϕ_{inv} first. If $\phi_{\text{inv}}(\mathbf{b},\mathbf{c}) = 0$, then $\min(\mathbf{b}_j + \mathbf{c}_j, 2 - \mathbf{b}_j - \mathbf{c}_j) = 0$ for all j. Since $\mathbf{b}_j, \mathbf{c}_j \in [0,1]$, this is only possible if $\mathbf{b}_j = \mathbf{c}_j \in \{0,1\}$ for all j. Thus $\mathbf{b} = \mathbf{c} \in \{0,1\}^n \cap \text{conv}(\mathbb{S}^n) = \mathbb{S}^n$, which proves the statement for ϕ_{inv}.

Next we consider ϕ_{subs}. Similar as before, if $\phi_{\text{subs}}(\mathbf{b},\mathbf{c}) = 0$ then $\min(\mathbf{c}_j, 1 - \mathbf{b}_j) = 0$ for all j. Thus for every j either $\mathbf{c}_j = 0$ or $\mathbf{b}_j = 1$. Since $\mathbf{c} \in \text{conv}(\mathbb{S}^n)$, there must be a j with $\mathbf{c}_j > 0$, and hence $\mathbf{b}_j = 1$. Since $\mathbf{b} \in \text{conv}(\mathbb{S}^n)$, it follows $\mathbf{b}_{j'} = 0$ for all $j' \neq j$. Ergo $\mathbf{c}_{j'} = 0$ for all $j' \neq j$, and consequently $\mathbf{c}_j = 1$, which shows the result for ϕ_{subs}.

If $\phi_{\text{omni}}(\mathbf{b},\mathbf{c}) = 0$, we have $\min(\mathbf{b}_{j_1}, \mathbf{c}_{j_2}) = 0$ for all (j_1, j_2) with $j_1 \neq j_2$. Since $\mathbf{b} \in \text{conv}(\mathbb{S}^n)$, there is a j with $\mathbf{b}_j > 0$. This yields $\mathbf{c}_{j'} = 0$ for all $j' \neq j$ and therefore $\mathbf{c}_j = 1$. Now using the same arguments again, we conclude $\mathbf{b}_j = 1$ and $\mathbf{b}_{j'} = 0$ for all $j' \neq j$, in particular $\mathbf{b} = \mathbf{c} \in \mathbb{S}^n$.

B.1.7 Proof of Proposition 5.8

To proof the statement, we take a look at several distinct cases. We first consider $i = \text{inv}$. Let $\mathbf{b},\mathbf{c},\mathbf{d} \in \text{conv}(\mathbb{S}^n)$. It is sufficient to show

$$\min(\mathbf{b}_j + \mathbf{d}_j, 2 - \mathbf{b}_j - \mathbf{d}_j) \leq \min(\mathbf{b}_j + \mathbf{c}_j, 2 - \mathbf{b}_j - \mathbf{c}_j) + \min(\mathbf{c}_j + \mathbf{d}_j, 2 - \mathbf{c}_j - \mathbf{d}_j)$$

for all j.

Case i): Let $\min(\mathbf{b}_j + \mathbf{d}_j, 2 - \mathbf{b}_j - \mathbf{d}_j) = \mathbf{b}_j + \mathbf{d}_j$, i. e. $\mathbf{b}_j + \mathbf{d}_j \leq 2 - \mathbf{b}_j - \mathbf{d}_j$. We have

$$\begin{aligned}
&\mathbf{b}_j + \mathbf{d}_j \leq (\mathbf{b}_j + \mathbf{c}_j) + (\mathbf{c}_j + \mathbf{d}_j),\\
&\mathbf{b}_j + \mathbf{d}_j \leq \mathbf{b}_j + 2 - \mathbf{d}_j = (\mathbf{b}_j + \mathbf{c}_j) + (2 - \mathbf{c}_j - \mathbf{d}_j),\\
&\mathbf{b}_j + \mathbf{d}_j \leq 2 - \mathbf{b}_j + \mathbf{d}_j = (2 - \mathbf{b}_j - \mathbf{c}_j) + (\mathbf{c}_j + \mathbf{d}_j),\\
&\mathbf{b}_j + \mathbf{d}_j \overset{i)}{\leq} 2 - \mathbf{b}_j - \mathbf{d}_j \leq 2 - \mathbf{b}_j - \mathbf{d}_j + 2 - 2\mathbf{c}_j = (2 - \mathbf{b}_j - \mathbf{c}_j) + (2 - \mathbf{c}_j - \mathbf{d}_j).
\end{aligned}$$

Case ii): Let $\min(\mathbf{b}_j + \mathbf{d}_j, 2 - \mathbf{b}_j - \mathbf{d}_j) = 2 - \mathbf{b}_j - \mathbf{d}_j$. We get

$$\begin{aligned}
&2 - \mathbf{b}_j - \mathbf{d}_j \overset{ii)}{\leq} \mathbf{b}_j + \mathbf{d}_j \leq (\mathbf{b}_j + \mathbf{c}_j) + (\mathbf{c}_j + \mathbf{d}_j),\\
&2 - \mathbf{b}_j - \mathbf{d}_j \leq \mathbf{b}_j + 2 - \mathbf{d}_j = (\mathbf{b}_j + \mathbf{c}_j) + (2 - \mathbf{c}_j - \mathbf{d}_j),
\end{aligned}$$

$$2-\mathbf{b}_j-\mathbf{d}_j \le 2-\mathbf{b}_j+\mathbf{d}_j = (2-\mathbf{b}_j-\mathbf{c}_j)+(\mathbf{c}_j+\mathbf{d}_j),$$
$$2-\mathbf{b}_j-\mathbf{d}_j \le 2-\mathbf{b}_j-\mathbf{d}_j+2-2\mathbf{c}_j = (2-\mathbf{b}_j-\mathbf{c}_j)+(2-\mathbf{c}_j-\mathbf{d}_j).$$

Altogether, we see

$$\min(\mathbf{b}_j+\mathbf{d}_j, 2-\mathbf{b}_j-\mathbf{d}_j) \le \min(\mathbf{b}_j+\mathbf{c}_j, 2-\mathbf{b}_j-\mathbf{c}_j)+\min(\mathbf{c}_j+\mathbf{d}_j, 2-\mathbf{c}_j-\mathbf{d}_j),$$

which proves the first statement for $i = \mathrm{inv}$.

Next we consider $i = \mathrm{subs}$. Let again $\mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathrm{conv}(\mathbb{S}^n)$. It is sufficient to show

$$\min(\mathbf{d}_j, 1-\mathbf{b}_j) \le \min(\mathbf{c}_j, 1-\mathbf{b}_j)+\min(\mathbf{d}_j, 1-\mathbf{c}_j)$$

for all j.

Case i): Let $\min(\mathbf{d}_j, 1-\mathbf{b}_j) = \mathbf{d}_j$. Then

$$\mathbf{d}_j \le \mathbf{c}_j+\mathbf{d}_j,$$
$$\mathbf{d}_j \le 1 = \mathbf{c}_j+(1-\mathbf{c}_j),$$
$$\mathbf{d}_j \le (1-\mathbf{b}_j)+\mathbf{d}_j,$$
$$\mathbf{d}_j \overset{i)}{\le} 1-\mathbf{b}_j \le (1-\mathbf{b}_j)+(1-\mathbf{c}_j).$$

Case ii): Now let $\min(\mathbf{d}_j, 1-\mathbf{b}_j) = 1-\mathbf{b}_j$. We find

$$1-\mathbf{b}_j \overset{ii)}{\le} \mathbf{d}_j \le \mathbf{c}_j+\mathbf{d}_j,$$
$$1-\mathbf{b}_j \le 1 = \mathbf{c}_j+(1-\mathbf{c}_j),$$
$$1-\mathbf{b}_j \le (1-\mathbf{b}_j)+\mathbf{d}_j,$$
$$1-\mathbf{b}_j \le (1-\mathbf{b}_j)+(1-\mathbf{c}_j).$$

Altogether

$$\min(\mathbf{d}_j, 1-\mathbf{b}_j) \le \min(\mathbf{c}_j, 1-\mathbf{b}_j)+\min(\mathbf{d}_j, 1-\mathbf{c}_j),$$

which shows (5.16) for $i = \mathrm{subs}$.

For $i = \text{omni}$, the triangle inequality does not hold in general. As a counterexample, we consider

$$\mathbf{b} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \\ 0 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{d} = \begin{pmatrix} \frac{2}{5} \\ \frac{2}{5} \\ \frac{1}{5} \end{pmatrix}.$$

Then $\mathbf{b}, \mathbf{c}, \mathbf{d} \in \operatorname{conv}(\mathbb{S}^n)$, and we get

$$\begin{aligned}
\phi_{\text{omni}}(\mathbf{b}, \mathbf{d}) &= \sum_{j_1 \neq j_2} \min(\mathbf{b}_{j_1}, \mathbf{d}_{j_2}) = \left(\frac{1}{3} + \frac{1}{5}\right) + \left(\frac{2}{5} + \frac{1}{5}\right) + 0 > \frac{1}{3} + \frac{3}{5}, \\
\phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) &= \sum_{j_1 \neq j_2} \min(\mathbf{b}_{j_1}, \mathbf{c}_{j_2}) = \frac{1}{3} + 0 + 0 = \frac{1}{3}, \\
\phi_{\text{omni}}(\mathbf{c}, \mathbf{d}) &= \sum_{j_1 \neq j_2} \min(\mathbf{c}_{j_1}, \mathbf{d}_{j_2}) = 0 + \left(\frac{2}{5} + \frac{1}{5}\right) + 0 = \frac{3}{5},
\end{aligned}$$

and the triangle inequality does not hold.

To prove the last statement, let $\mathbf{b}, \mathbf{d} \in \mathbb{S}^n$ with $\mathbf{b}_l = \mathbf{d}_k = 1$ and $\mathbf{c} \in \operatorname{conv}(\mathbb{S}^n)$. Then

$$\phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) = \sum_{\substack{j_1, j_2 = 1 \\ j_1 \neq j_2}}^{n} \min(\mathbf{b}_{j_1}, \mathbf{c}_{j_2}) = \sum_{\substack{j_2 = 1 \\ j_2 \neq l}}^{n} \min(1, \mathbf{c}_{j_2}) = \sum_{\substack{j_2 = 1 \\ j_2 \neq l}}^{n} \mathbf{c}_{j_2} = 1 - \mathbf{c}_l, \tag{B.3}$$

and similar $\phi_{\text{omni}}(\mathbf{c}, \mathbf{d}) = 1 - \mathbf{c}_k$. For $\mathbf{b} = \mathbf{d}$, we have $\phi_{\text{omni}}(\mathbf{b}, \mathbf{d}) = 0$ according to Proposition 5.7, and for $\mathbf{b} \neq \mathbf{d}$ obviously $\phi_{\text{omni}}(\mathbf{b}, \mathbf{d}) = 1$. This yields

$$\phi_{\text{omni}}(\mathbf{b}, \mathbf{d}) \leq 1 \leq 2 - (\mathbf{c}_l + \mathbf{c}_k) = (1 - \mathbf{c}_l) + (1 - \mathbf{c}_k) \overset{(B.3)}{=} \phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) + \phi_{\text{omni}}(\mathbf{c}, \mathbf{d}),$$

which closes the proof.

B.1.8 Proof of Proposition 5.9

For $i \in \{\text{inv,subs}\}$ we have $\phi_i(\mathbf{b}, \mathbf{d}) = 1$ by Proposition 5.6, which is also true for $i = \text{omni}$, as one easily verifies.

Let us consider $i = \text{omni}$ first. As seen in the proof of Proposition 5.8, we have $\phi_{\text{omni}}(\mathbf{b}, \mathbf{c}) = 1 - \mathbf{c}_l$ and $\phi_{\text{omni}}(\mathbf{c}, \mathbf{d}) = 1 - \mathbf{c}_k$. Hence, Statement (5.17) is equivalent to

$$1 = 2 - (\mathbf{c}_l + \mathbf{c}_k) \quad \Longleftrightarrow \quad \mathbf{c}_l + \mathbf{c}_k = 1, \tag{B.4}$$

which is obviously true.

Next we take a look at $i = \text{inv}$. We find

$$\begin{aligned} 2\phi_{\text{inv}}(\mathbf{b},\mathbf{c}) &= \sum_{j=1}^{n} \min(\mathbf{b}_j + \mathbf{c}_j, 2 - \mathbf{b}_j - \mathbf{c}_j) \\ &= \left[\sum_{j\neq l} \min(\mathbf{c}_j, 2 - \mathbf{c}_j)\right] + \min(1 + \mathbf{c}_l, 1 - \mathbf{c}_l) = 1 - \mathbf{c}_l + \sum_{j\neq l} \mathbf{c}_j \\ &= 1 - 2\mathbf{c}_l + \sum_{j=1}^{n} \mathbf{c}_j = 2 - 2\mathbf{c}_l\,, \end{aligned}$$

and thus $\phi_{\text{inv}}(\mathbf{b},\mathbf{c}) = 1 - \mathbf{c}_l$. Similarly, we get $\phi_{\text{inv}}(\mathbf{c},\mathbf{d}) = 1 - \mathbf{c}_k$. Again, Statement (5.17) reduces to the valid Statement (B.4).

For $i = \text{subs}$, the proof works similarly.

B.1.9 Proof of Proposition 5.10

Since $n = 2$, we have $\mathbf{b}_1 + \mathbf{b}_2 = 1 = \mathbf{c}_1 + \mathbf{c}_2$. Therefore

$$\begin{aligned} \phi_{\text{subs}}(\mathbf{b},\mathbf{c}) &= \min(\mathbf{c}_1, 1 - \mathbf{b}_1) + \min(\mathbf{c}_2, 1 - \mathbf{b}_2) \\ &= \min(\mathbf{c}_1, \mathbf{b}_2) + \min(\mathbf{c}_2, \mathbf{b}_1) = \phi_{\text{omni}}(\mathbf{b},\mathbf{c})\,. \end{aligned}$$

As $\mathbf{b}_2 + \mathbf{c}_2 = 2 - \mathbf{b}_1 - \mathbf{c}_1$ and $2 - \mathbf{b}_2 - \mathbf{c}_2 = \mathbf{b}_1 + \mathbf{c}_1$, we furthermore have

$$\begin{aligned} \phi_{\text{inv}}(\mathbf{b},\mathbf{c}) &= \frac{1}{2}\left[\min(\mathbf{b}_1 + \mathbf{c}_1, 2 - \mathbf{b}_1 - \mathbf{c}_1) + \min(\mathbf{b}_2 + \mathbf{c}_2, 2 - \mathbf{b}_2 - \mathbf{c}_2)\right] \\ &= \min(\mathbf{b}_1 + \mathbf{c}_1, 2 - \mathbf{b}_1 - \mathbf{c}_1) = \begin{cases} \mathbf{b}_1 + \mathbf{c}_1 & \text{if } \mathbf{b}_1 + \mathbf{c}_1 \leq 1\,, \\ 2 - \mathbf{b}_1 - \mathbf{c}_1 & \text{if } \mathbf{b}_1 + \mathbf{c}_1 > 1\,. \end{cases} \end{aligned}$$

On the other hand, we also find

$$\begin{aligned} \phi_{\text{omni}}(\mathbf{b},\mathbf{c}) &= \min(\mathbf{b}_1, \mathbf{c}_2) + \min(\mathbf{b}_2, \mathbf{c}_1) = \min(\mathbf{b}_1, 1 - \mathbf{c}_1) + \min(1 - \mathbf{b}_1, \mathbf{c}_1) \\ &= \min(\mathbf{b}_1 + \mathbf{c}_1, 1) - \mathbf{c}_1 + \min(1, \mathbf{c}_1 + \mathbf{b}_1) - \mathbf{b}_1 \\ &= 2\min(\mathbf{b}_1 + \mathbf{c}_1, 1) - (\mathbf{b}_1 + \mathbf{c}_1) = \begin{cases} \mathbf{b}_1 + \mathbf{c}_1 & \text{if } \mathbf{b}_1 + \mathbf{c}_1 \leq 1\,, \\ 2 - \mathbf{b}_1 - \mathbf{c}_1 & \text{if } \mathbf{b}_1 + \mathbf{c}_1 > 1\,, \end{cases} \end{aligned}$$

and hence $\phi_{\text{inv}}(\mathbf{b},\mathbf{c}) = \phi_{\text{omni}}(\mathbf{b},\mathbf{c})$, which completes the proof.

B.2 Proofs for Section 6.3

B.2.1 Proof of Remark 6.1

For the proof, we need the following

Lemma B.1
Let $\Omega_{\mathbf{x}} \subset \mathbb{R}^{n_x}$ and $\Omega_{\mathbf{p}} \subset \mathbb{R}^{n_p}$ be compact subsets, and $h : \Omega_{\mathbf{x}} \times \Omega_{\mathbf{p}} \to \mathbb{R}$ a continuous function. Then

$$g : \Omega_{\mathbf{x}} \to \mathbb{R}, \ \mathbf{x} \mapsto \max_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}, \mathbf{p})$$

is continuous, and the same holds for $g' : \Omega_{\mathbf{x}} \to \mathbb{R}$, $\mathbf{x} \mapsto \min_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}, \mathbf{p})$.

Proof As $h(\cdot)$ is continuous on a compact set, it is uniformly continuous (Heine-Cantor theorem). Let $\mathbf{x} \in \Omega_{\mathbf{x}}$ and $(\mathbf{x}^n)_{n \in \mathbb{N}}$ be a sequence in $\Omega_{\mathbf{x}}$ converging to $\mathbf{x}$. As $h(\cdot)$ is uniformly continuous we have

$$\forall \varepsilon > 0 \, \exists N(\varepsilon) \in \mathbb{N} \text{ s. t. } \left| h(\mathbf{x}^n, \mathbf{p}) - h(\mathbf{x}, \mathbf{p}) \right| < \varepsilon \ \forall n \geq N(\varepsilon) \, \forall \mathbf{p} \in \Omega_{\mathbf{p}} .$$

In particular, $N(\varepsilon)$ does not depend on $\mathbf{p}$. Let $\varepsilon > 0$. We have

$$h(\mathbf{x}, \mathbf{p}) - \varepsilon < h(\mathbf{x}^n, \mathbf{p}) < h(\mathbf{x}, \mathbf{p}) + \varepsilon \ \ \forall n \geq N(\varepsilon)$$

for all $\mathbf{p} \in \Omega_{\mathbf{p}}$, and therefore also

$$\max_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}, \mathbf{p}) - \varepsilon < \max_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}^n, \mathbf{p}) < \max_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}, \mathbf{p}) + \varepsilon \ \ \forall n \geq N(\varepsilon)$$

(note that a continuous function takes its maximum on a compact set according to a generalization of the extreme value theorem). Altogether, we get

$$\left| g(\mathbf{x}^n) - g(\mathbf{x}) \right| = \left| \max_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}^n, \mathbf{p}) - \max_{\mathbf{p} \in \Omega_{\mathbf{p}}} h(\mathbf{x}, \mathbf{p}) \right| < \varepsilon \ \ \forall n \geq N(\varepsilon) ,$$

which shows the continuity of $g(\cdot)$. The continuity of $g'(\cdot)$ can be proven in a similar manner. □

Now, we are able to prove Remark 6.1. We have

$$f(\mathbf{x}, \mathbf{p}) \leq \max_{\mathbf{p} \in \Omega_{\mathbf{p}}} f(\mathbf{x}, \mathbf{p})$$

for all $\mathbf{x} \in \Omega_\mathbf{x}$ and $\mathbf{p} \in \Omega_\mathbf{p}$, where $\max_{\mathbf{p}\in\Omega_\mathbf{p}} f(\mathbf{x},\mathbf{p})$ depends on $\mathbf{x}$ only. Consequently,

$$\min_{\mathbf{x}\in\Omega_\mathbf{x}} f(\mathbf{x},\mathbf{p}) \leq \min_{\mathbf{x}\in\Omega_\mathbf{x}} \max_{\mathbf{p}\in\Omega_\mathbf{p}} f(\mathbf{x},\mathbf{p})$$

for all $\mathbf{p} \in \Omega_\mathbf{p}$, and thus

$$\max_{\mathbf{p}\in\Omega_\mathbf{p}} \min_{\mathbf{x}\in\Omega_\mathbf{x}} f(\mathbf{x},\mathbf{p}) \leq \min_{\mathbf{x}\in\Omega_\mathbf{x}} \max_{\mathbf{p}\in\Omega_\mathbf{p}} f(\mathbf{x},\mathbf{p}) \,.$$

The existence of all above maxima and minima follows from Lemma B.1 and a generalization of the extreme value theorem.

B.2.2 max min vs min max NLP – Objective Function Value

For $\Omega_x = [-5,5]$, $\Omega_p = [-1,1]$ we consider the function

$$f : \Omega_x \times \Omega_p \to \mathbb{R}, \;\; (x,p) \mapsto (x-p)^2 + p \,.$$

Since $\Omega_p \subset \Omega_x$ we have $\min_{x\in\Omega_x} f(x,p) = f(p,p) = p$ and hence

$$\max_{p\in\Omega_p} \min_{x\in\Omega_x} f(x,p) = 1 \,.$$

For the min max problem we consider

$$f(x,p) = (x-p)^2 + p = p^2 - p(2x-1) + x^2 \,.$$

For a fixed $\tilde{x}$, the function $f(\tilde{x},\cdot)$ is a convex parabola with vertex $\bar{p}(\tilde{x}) = \frac{2\tilde{x}-1}{2}$. Since $\Omega_p = [-1,1]$, by the symmetry of the parabola we get

$$g(x) \stackrel{\text{def}}{=} \max_{p\in\Omega_p} f(x,p) = \begin{cases} f(x,1) & \text{if } \bar{p}(x) \leq 0 \\ f(x,-1) & \text{if } \bar{p}(x) > 0 \end{cases} = \begin{cases} (x-1)^2+1 & \text{if } x \leq 0.5 \\ (x+1)^2-1 & \text{if } x > 0.5 \end{cases} .$$

For the derivative of $g(\cdot)$, we get

$$\frac{\mathrm{d}}{\mathrm{d}x} g(x) = \begin{cases} 2(x-1) & \text{if } x < 0.5 \\ \text{undefined} & \text{if } x = 0.5 \\ 2(x+1) & \text{if } x > 0.5 \end{cases} .$$

Hence, the continuous function $g(\cdot)$ is strictly monotonically decreasing for $x < 0.5$ and strictly monotonically increasing for $x > 0.5$, thus having a minimum at $x = 0.5$.

Altogether, we get

$$\max_{p\in\Omega_p}\min_{x\in\Omega_x} f(x,p) = 1 < 1.25 = f(0.5,1) = \min_{x\in\Omega_x}\max_{p\in\Omega_p} f(x,p)\,.$$

B.2.3 Proof of Proposition 6.2

The proof works similar to the one of Remark 6.1. We have

$$\Phi(\mathbf{x}(1;\mathbf{p})) \le \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p}))$$

for all $\mathbf{p}\in\Omega_\mathbf{p}$ and $(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})$. Thus,

$$\min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \le \min_{(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})} \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p}))$$

for all $\mathbf{p}\in\Omega_\mathbf{p}$. As $\tilde{\mathcal{C}}(\Omega_\mathbf{p})\subseteq\mathcal{C}(\mathbf{p})$, we get

$$\min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\mathcal{C}(\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \le \min_{(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})} \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p}))$$

for all $\mathbf{p}\in\Omega_\mathbf{p}$, and consequently

$$\max_{\mathbf{p}\in\Omega_\mathbf{p}}\min_{\substack{(\mathbf{u},\mathbf{u}(\cdot))\in\mathcal{C}(\mathbf{p}),\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p})) \le \min_{(\mathbf{u},\mathbf{u}(\cdot))\in\tilde{\mathcal{C}}(\Omega_\mathbf{p})} \max_{\substack{\mathbf{p}\in\Omega_\mathbf{p},\\ \mathbf{x}(\cdot;\mathbf{p})}} \Phi(\mathbf{x}(1;\mathbf{p}))\,.$$

All maxima and minima above exist per assumption.

B.2.4 Solution of Problem (6.9)

Let $p\in[0,9]$. We consider Problem (6.9) with variables

$$\big(T, u(\cdot), \mathbf{x}(\cdot\,;p)\big) \in \mathbb{R}\times L^\infty([0,1],\mathbb{R})\times W^{1,\infty}\big([0,1],\mathbb{R}^2\big)\,,$$

see Section 2.1 for the spaces and corresponding norms, where $\mathbf{x}(\cdot\,;p)$ denotes the (unique) solution of Initial Value Problem (IVP) (6.9b-6.9c) for given T, $u(\cdot)$, and p. The product space $\mathbb{R}\times L^\infty([0,1],\mathbb{R})\times W^{1,\infty}\big([0,1],\mathbb{R}^2\big)$, equipped with the norm

$$\big\|\big(T, u(\cdot), \mathbf{x}(\cdot\,;p)\big)\big\| = \max\big[|T|, \|u(\cdot)\|_\infty, \|\mathbf{x}(\cdot\,;p)\|_{1,\infty}\big]\,,$$

is a Banach space. In this section, we state the unique globally optimal solution of Problem (6.9), prove its optimality, and show that no different local optima exist. The main results of this section can be found in Corollary B.4 and Proposition B.9.

The Globally Optimal Solution of Problem (6.9)

We first consider the global optimum and start with

Lemma B.2
Let $\big(T, u(\cdot), \mathbf{x}(\cdot\,; p)\big)$ be feasible for Problem (6.9). Then $T > \frac{4}{10+p} + \frac{4}{10-p}$.

Proof Due to the Boundary Conditions (6.9c) and (6.9e) we have $T > 0$. Furthermore,

$$\mathbf{x}_2(t; p) = \mathbf{x}_2(t_0; p) + \int_{t_0}^{t} T(u(\tau) - p)\,\mathrm{d}\tau$$

for all $0 \le t_0 \le t \le 1$. Since $\mathbf{x}_2(0; p) = 0$ and $u(t) \le 10$ we get

$$\mathbf{x}_2(t; p) \le (10 - p)\,T\,t\,.$$

On the other hand, because of $\mathbf{x}_2(1; p) \le 0$ and $u(t) \ge -10$ we have

$$\mathbf{x}_2(t; p) \le (10 + p)\,T - (10 + p)\,T\,t$$

for all $t \in [0, 1]$. Otherwise, there exists a t_0 with $\mathbf{x}_2(t_0; p) > (10+p)\,T - (10+p)\,T\,t_0$ and we get

$$\begin{aligned}\mathbf{x}_2(1; p) &> (10 + p)\,T - (10 + p)\,T\,t_0 + \int_{t_0}^{1} T(u(\tau) - p)\,\mathrm{d}\tau \\ &\ge (10 + p)\,T - (10 + p)\,T\,t_0 - (10 + p)\,T(1 - t_0) = 0\end{aligned}$$

which contradicts the Terminal Condition (6.9f). We conclude

$$\begin{aligned}\mathbf{x}_2(t; p) &\le \min\big[(10 - p)\,T\,t,\, (10 + p)\,T - (10 + p)\,T\,t\big] \\ &= \begin{cases} (10 - p)\,T\,t & \text{if } 0 \le t \le \frac{10+p}{20}, \\ (10 + p)\,T - (10 + p)\,T\,t & \text{else,} \end{cases}\end{aligned}$$

and

$$\begin{aligned}\mathbf{x}_1(1;p) = T\int_0^1 \mathbf{x}_2(t;p)\,\mathrm{d}t &\le T\int_0^{\frac{10+p}{20}} (10-p)\,T\,t\,\mathrm{d}t + T\int_{\frac{10+p}{20}}^1 (10+p)\,T - (10+p)\,T\,t\,\mathrm{d}t \\ &= \frac{1}{2}(10-p)\,T^2\left(\frac{10+p}{20}\right)^2 + \frac{1}{2}(10-p)\,T^2\left(\frac{10+p}{20}\right)\left(\frac{10-p}{20}\right).\end{aligned}$$

Now, if we had $T \le \frac{4}{10+p} + \frac{4}{10-p} = \frac{80}{(10-p)(10+p)}$, because of $0 \le p \le 9$ we would obtain

$$\mathbf{x}_1(1;p) \le \frac{8}{10-p} + \frac{8}{10+p} \le 8 + \frac{8}{10} < 10$$

which contradicts the Terminal Condition (6.9e). Hence, $T > \frac{4}{10+p} + \frac{4}{10-p}$. □

To state the global optimum and to prove its optimality, we consider a certain class of control functions and the corresponding differential states. Let $T > \frac{4}{10+p} + \frac{4}{10-p}$. We set

$$t_1 = t_1(T,p) = \frac{4}{T(10-p)} \quad \text{and} \quad t_2 = t_2(T,p) = 1 - \frac{4}{T(10+p)}.$$

Then $0 < t_1 < t_2 < 1$ and we define

$$u_T(t;p) \stackrel{\text{def}}{=} \begin{cases} 10 & \text{for } 0 \le t < t_1(T,p), \\ p & \text{for } t_1(T,p) \le t < t_2(T,p), \\ -10 & \text{for } t_2(T,p) \le t \le 1. \end{cases} \tag{B.5}$$

Let $\mathbf{x}_T(\cdot;p)$ denote the differential states which are determined by p, T, and $u_T(\cdot;p)$. We have

$$\mathbf{x}_{T,2}(t;p) = \begin{cases} T(10-p)\,t & \text{for } 0 \le t < t_1(T,p), \\ 4 & \text{for } t_1(T,p) \le t < t_2(T,p), \\ 4 - T(10+p)\,\big(t - t_2(T,p)\big) & \text{for } t_2(T,p) \le t \le 1. \end{cases}$$

In particular, $\mathbf{x}_{T,2}(t;p) \le 4$ for all $t \in [0,1]$ and $\mathbf{x}_{T,2}(1;p) = 0$. Furthermore, we get

$$\begin{aligned}\mathbf{x}_{T,1}(t_1;p) - \mathbf{x}_{T,1}(0;p) &= \frac{1}{2}(10-p)\,T^2 t_1^2 = \frac{8}{10-p} \\ \mathbf{x}_{T,1}(t_2;p) - \mathbf{x}_{T,1}(t_1;p) &= 4T(t_2 - t_1) = 4T - \frac{16}{10-p} - \frac{16}{10+p}, \\ \mathbf{x}_{T,1}(1;p) - \mathbf{x}_{T,1}(t_2;p) &= \frac{1}{2}4T\,(1-t_2) = \frac{8}{10+p}\end{aligned}$$

(e. g., by geometrical considerations) and therefore

$$\mathbf{x}_{T,1}(1;p) = \mathbf{x}_{T,1}(1;p) - \mathbf{x}_{T,1}(0;p) = 4T - \frac{8}{10-p} - \frac{8}{10+p} \tag{B.6}$$

since $\mathbf{x}_{T,1}(0;p) = 0$. In particular,

$$\mathbf{x}_{T,1}(1;p) = 10 \iff T = T^*(p) \stackrel{\text{def}}{=} 2.5 + \frac{40}{100-p^2}, \tag{B.7}$$

$\mathbf{x}_{T,1}(1;p) < 10$ for $\frac{4}{10+p} + \frac{4}{10-p} < T < T^*(p)$ and $\mathbf{x}_{T,1}(1;p) > 10$ for $T > T^*(p)$. Hence, the tuple $\big(T, u_T(\cdot;p), \mathbf{x}_T(\cdot;p)\big)$ is feasible if and only if $T \geq T^*(p)$. We define

$$T^* \stackrel{\text{def}}{=} T^*(p), \quad u^*(\cdot) = u^*(\cdot;p) \stackrel{\text{def}}{=} u_{T^*}(\cdot;p), \quad \text{and} \quad \mathbf{x}^*(\cdot;p) \stackrel{\text{def}}{=} \mathbf{x}_{T^*}(\cdot;p).$$

Proposition B.3
Let $\big(T, u(\cdot), \mathbf{x}(\cdot;p)\big)$ be feasible for Problem (6.9) with $u(\cdot) \neq u_T(\cdot;p)$ (in $L^\infty([0,1],\mathbb{R})$), where $u_T(\cdot;p)$ is given by (B.5). Then $\big(T, u_T(\cdot;p), \mathbf{x}_T(\cdot;p)\big)$ is feasible as well and we have

$$10 \leq \mathbf{x}_1(1;p) < \mathbf{x}_{T,1}(1;p).$$

Proof From Lemma B.2 we get $T > \frac{4}{10+p} + \frac{4}{10-p}$, and $u_T(\cdot;p)$ is well-defined. First, we show that $\mathbf{x}_2(t;p) \leq \mathbf{x}_{T,2}(t;p)$ for all $t \in [0,1]$. Let $t_1 = \frac{4}{T(10-p)}$ and $t_2 = 1 - \frac{4}{T(10+p)}$. For $t \in [0, t_1]$, we have

$$\mathbf{x}_2(t;p) = T\int_0^t u(t) - p\,\mathrm{d}\tau \leq T\int_0^t 10 - p\,\mathrm{d}\tau = \mathbf{x}_{T,2}(t;p),$$

and due to feasibility $\mathbf{x}_2(t;p) \leq 4 = \mathbf{x}_{T,2}(t;p)$ for $t \in [t_1, t_2]$. If there was a $t' \in [t_2, 1]$ with $\mathbf{x}_2(t';p) > \mathbf{x}_{T,2}(t';p)$, we would get

$$\mathbf{x}_2(1;p) = \mathbf{x}_2(t';p) + T\int_{t'}^1 u(t) - p\,\mathrm{d}t > \mathbf{x}_{T,2}(t';p) + T\int_{t'}^1 -10 - p\,\mathrm{d}t = \mathbf{x}_{T,2}(1;p) = 0$$

which contradicts the feasibility of $\big(T, u(\cdot), \mathbf{x}(\cdot;p)\big)$. Thus, we have $\mathbf{x}_2(t;p) \leq \mathbf{x}_{T,2}(t;p)$ for all $t \in [0,1]$.

Next, we show that there is a $t' \in [0,1]$ with $\mathbf{x}_2(t';p) < \mathbf{x}_{T,2}(t';p)$. By assumption, we have $u(\cdot) \neq u_T(\cdot;p)$. We distinct three cases:

Case 1): $u(\cdot) \not\equiv u_T(\cdot;p)$ in $[0,t_1]$ (almost surely). Then there is an $\varepsilon > 0$ and a subset $\mathcal{A} \subseteq [0,t_1]$ with non-zero measure such that $u(t) < 10 - \varepsilon = u_T(t;p) - \varepsilon$ for $t \in \mathcal{A}$ (almost surely). Thus, we get $\mathbf{x}_2(t_1;p) < \mathbf{x}_{T,2}(t_1;p)$.

Case 2): $u(\cdot) \equiv u_T(\cdot;p)$ in $[0,t_1]$ and $u(\cdot) \not\equiv u_T(\cdot;p)$ in $[t_1,t_2]$ (almost surely, respectively). Then $\mathbf{x}_2(t_1;p) = 4$. We claim that there is an $\varepsilon > 0$ and a subset $\mathcal{A} \subseteq [t_1,t_2]$ with non-zero measure such that $u(t) < p - \varepsilon = u_T(t;p) - \varepsilon$ for $t \in \mathcal{A}$. Indeed, if such a subset does not exist, we have $u(t) \geq u_T(t;p)$ in $[t_1,t_2]$ (almost surely), and since the control functions differ on a set with non-zero measure we get $\mathbf{x}_2(t;p) > 4$ for some $t \in [t_1,t_2]$. This contradicts the feasibility of $\big(T,u(\cdot),\mathbf{x}(\cdot;p)\big)$. Hence such a subset exists and consequently $\mathbf{x}_2(t;p) < \mathbf{x}_{T,2}(t;p)$ for some t' in $[t_1,t_2]$.

Case 3): $u(\cdot) \equiv u_T(\cdot;p)$ in $[0,t_2]$ and $u(\cdot) \not\equiv u_T(\cdot;p)$ in $[t_2,1]$ (almost surely, respectively). Then we have $\mathbf{x}_2(t_2;p) = \mathbf{x}_{T,2}(t_2;p)$ and there is an $\varepsilon > 0$ and a subset $\mathcal{A} \subseteq [t_2,1]$ with non-zero measure such that $u(t) > u_T(t;p) + \varepsilon = -10 + \varepsilon$ for $t \in \mathcal{A}$ (almost surely). Consequently, $\mathbf{x}_2(1;p) > \mathbf{x}_{T,2}(1;p) = 0$ which contradicts the feasibility of $\big(T,u(\cdot),\mathbf{x}(\cdot;p)\big)$. Thus, Case 3) does not occur.

Altogether, we have seen that $\mathbf{x}_2(t;p) \leq \mathbf{x}_{T,2}(t;p)$ for all $t \in [0,1]$ and there is a t' with $\mathbf{x}_2(t';p) < \mathbf{x}_{T,2}(t';p)$. By the continuity of $\mathbf{x}_2(\cdot;p)$ and $\mathbf{x}_{T,2}(\cdot;p)$ we conclude

$$\mathbf{x}_{T,1}(1;p) > \mathbf{x}_1(1;p) \geq 10\,,$$

which shows the feasibility of $\big(T,u_T(\cdot;p),\mathbf{x}_T(\cdot;p)\big)$ and completes the proof. □

Corollary B.4
The tuple $\big(T^*,u^*(\cdot),\mathbf{x}^*(\cdot;p)\big)$ is the unique global optimum of Problem (6.9).

Proof By construction, $\big(T^*,u^*(\cdot),\mathbf{x}^*(\cdot;p)\big)$ is feasible with $\mathbf{x}_1^*(1;p) = 10$ and we have $\mathbf{x}_{T,1}(1;p) < 10$ for all $\frac{4}{10+p} + \frac{4}{10-p} < T < T^*$, see (B.6) and (B.7). Let $\big(T,u(\cdot),\mathbf{x}(\cdot;p)\big)$ be any feasible tuple for Problem (6.9). If $T < T^*$, from Proposition B.3 we get

$$10 \leq \mathbf{x}_1(1;p) \leq \mathbf{x}_{T,1}(1;p) < 10\,,$$

which is a contradiction. Hence, $T \geq T^*$ and $\big(T^*,u^*(\cdot),\mathbf{x}^*(\cdot;p)\big)$ is globally optimal. Furthermore, let $\big(T,u(\cdot),\mathbf{x}(\cdot;p)\big) \neq \big(T^*,u^*(\cdot),\mathbf{x}^*(\cdot;p)\big)$ be feasible for Problem (6.9) with $T = T^*$ and $u(\cdot) \neq u^*(\cdot;p) = u_{T^*}(\cdot;p)$. Again we apply Proposition B.3 and find

the contradiction

$$10 \leq \mathbf{x}_1(1;p) < \mathbf{x}_{T,1}(1;p) = \mathbf{x}_{T^*,1}(1;p) = 10\,,$$

which shows the uniqueness of the global optimum $\left(T^*, u^*(\cdot), \mathbf{x}^*(\cdot;p)\right)$. □

The Unique Solvability of Problem (6.9)

We show that $\left(T^*, u^*(\cdot), \mathbf{x}^*(\cdot;p)\right)$ is the only local optimum of Problem (6.9) in the considered normed space. For a proof we need several auxiliary results.

Lemma B.5
Let $T > 0$, $u(\cdot) \in L^\infty([0,1],\mathbb{R})$, and $\mathbf{x}(\cdot;p) \in W^{1,\infty}\left([0,1],\mathbb{R}^2\right)$ be the differential states which are determined by T, $u(\cdot)$, and p. Let $\varepsilon > 0$. Then there exist $\delta_T, \delta_u > 0$ such that

$$\|\mathbf{x}'(\cdot;p) - \mathbf{x}(\cdot;p)\|_{1,\infty} < \varepsilon$$

for all T', $u'(\cdot)$ with $|T' - T| < \delta_T$ and $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$, where $\mathbf{x}'(\cdot;p)$ denotes the differential states which are determined by T', $u'(\cdot)$, and p.

Proof Let $\bar{\delta}_T, \bar{\delta}_u > 0$, $|T' - T| < \bar{\delta}_T$, and $\|u'(\cdot) - u(\cdot)\|_\infty < \bar{\delta}_u$. Then

$$\begin{aligned}
\left\|\dot{\mathbf{x}}_2'(\cdot;p) - \dot{\mathbf{x}}_2(\cdot;p)\right\|_\infty &= \left\|T'\left(u'(\cdot) - p\right) - T\left(u(\cdot) - p\right)\right\|_\infty \\
&= \left\|(T' - T + T)\left(u'(\cdot) - p\right) - T\left(u(\cdot) - p\right)\right\|_\infty \\
&\leq |T' - T| \left\|u'(\cdot) - p\right\|_\infty + |T| \left\|u'(\cdot) - u(\cdot)\right\|_\infty \\
&= |T' - T| \left\|u'(\cdot) - u(\cdot) + u(\cdot) - p\right\|_\infty + |T| \left\|u'(\cdot) - u(\cdot)\right\|_\infty \\
&< \left(\bar{\delta}_T + |T|\right)\bar{\delta}_u + \bar{\delta}_T \left\|u(t) - p\right\|_\infty \\
&\leq \left(\bar{\delta}_T + |T|\right)\bar{\delta}_u + \bar{\delta}_T \left(\|u(\cdot)\|_\infty + 9\right), \\
\left\|\mathbf{x}_2'(\cdot;p) - \mathbf{x}_2(\cdot;p)\right\|_\infty &= \sup_{t\in[0,1]} \left| T' \int_0^t \dot{\mathbf{x}}_2'(\tau;p)\,\mathrm{d}\tau - T\int_0^t \dot{\mathbf{x}}_2(\tau;p)\,\mathrm{d}\tau \right| \\
&\leq \sup_{t\in[0,1]} \int_0^t \left\| T'\dot{\mathbf{x}}_2'(\cdot;p) - T\dot{\mathbf{x}}_2(\cdot;p)\right\|_\infty \mathrm{d}\tau \\
&= \left\| T'\dot{\mathbf{x}}_2'(\cdot;p) - T\dot{\mathbf{x}}_2(\cdot;p)\right\|_\infty \\
&< \left(\bar{\delta}_T + |T|\right)\left\|\dot{\mathbf{x}}_2'(\cdot;p) - \dot{\mathbf{x}}_2(\cdot;p)\right\|_\infty + \bar{\delta}_T \left\|\dot{\mathbf{x}}_2(\cdot;p)\right\|_\infty .
\end{aligned}$$

Furthermore, similar to what we have seen before we get

$$\begin{aligned}
\left\|\dot{\mathbf{x}}_1'(\cdot;p)-\dot{\mathbf{x}}_1(\cdot;p)\right\|_\infty &= \left\|T'\mathbf{x}_2'(\cdot;p)-T\mathbf{x}_2(\cdot;p)\right\|_\infty \\
&< \left(\bar{\delta}_T+|T|\right)\left\|\mathbf{x}_2'(\cdot;p)-\mathbf{x}_2(\cdot;p)\right\|_\infty+\bar{\delta}_T\left\|\mathbf{x}_2(\cdot;p)\right\|_\infty, \\
\left\|\mathbf{x}_1'(\cdot;p)-\mathbf{x}_1(\cdot;p)\right\|_\infty &= \sup_{t\in[0,1]}\left|\int_0^t T'\dot{\mathbf{x}}_1'(\tau;p)\,\mathrm{d}\tau-\int_0^t T\dot{\mathbf{x}}_1(\tau;p)\,\mathrm{d}\tau\right| \\
&< \left(\bar{\delta}_T+|T|\right)\left\|\dot{\mathbf{x}}_1'(\cdot;p)-\dot{\mathbf{x}}_1(\cdot;p)\right\|_\infty+\bar{\delta}_T\left\|\dot{\mathbf{x}}_1(\cdot;p)\right\|_\infty.
\end{aligned}$$

We see that $\left\|\dot{\mathbf{x}}_2'(\cdot;p)-\dot{\mathbf{x}}_2(\cdot;p)\right\|_\infty \to 0$ for $\bar{\delta}_T,\bar{\delta}_u \to 0$. Consequently, $\bar{\delta}_T,\bar{\delta}_u \to 0$ also successively implies

$$\left\|\mathbf{x}_2'(\cdot;p)-\mathbf{x}_2(\cdot;p)\right\|_\infty \to 0,\ \left\|\dot{\mathbf{x}}_1'(\cdot;p)-\dot{\mathbf{x}}_1(\cdot;p)\right\|_\infty \to 0,\ \text{ and } \left\|\mathbf{x}_1'(\cdot;p)-\mathbf{x}_1(\cdot;p)\right\|_\infty \to 0.$$

Altogether, we get

$$\left\|\mathbf{x}'(\cdot;p)-\mathbf{x}(\cdot;p)\right\|_{1,\infty} \to 0 \ \text{ for } \ \bar{\delta}_T,\bar{\delta}_u \to 0,$$

and the statement of the lemma follows. □

Due to Lemma B.5, it is sufficient to show that for each feasible $\left(T,u(\cdot),\mathbf{x}(\cdot;p)\right) \neq \left(T^*,u^*(\cdot),\mathbf{x}^*(\cdot;p)\right)$ and each $\delta_T,\delta_u>0$ there is a feasible tuple $\left(T',u'(\cdot),\mathbf{x}'(\cdot;p)\right)$ with $|T'-T|<\delta_T$, $\|u'(\cdot)-u(\cdot)\|_\infty<\delta_u$, and $T'<T$. In the following, we prove this claim. We start by investigating three different cases.

Lemma B.6
Let $\delta_u>0$ and $\left(T,u(\cdot),\mathbf{x}(\cdot;p)\right)$ be feasible for Problem (6.9) with

$$\mathbf{x}_2(1;p)<0.$$

Then there is a feasible $\left(T',u'(\cdot),\mathbf{x}'(\cdot;p)\right)$ with $T'=T$, $\|u'(\cdot)-u(\cdot)\|_\infty<\delta_u$, and $\mathbf{x}_1'(1;p)>10$.

Proof Due to $\mathbf{x}_2(1;p)<0$, by the continuity of $\mathbf{x}_2(\cdot;p)$ there is a $\varepsilon>0$ such that $\mathbf{x}_2(t;p)<0$ for all $t\in[1-\varepsilon,1]$ and $\mathbf{x}_2(1;p)<\mathbf{x}_2(1-\varepsilon;p)<0$. If we had $u(t)\geq p$ for $t\in[1-\varepsilon,1]$ (almost surely), then $\mathbf{x}_2(1;p)\geq\mathbf{x}_2(1-\varepsilon;p)$ which is a contradiction. Hence, there is a $\delta>0$ with $\delta<\min\left(\delta_u,4,|\mathbf{x}_2(1;p)|\right)$ and a subset $\mathcal{A}\subseteq[1-\varepsilon,1]$ with non-zero (Lebesgue-)measure $\lambda(\mathcal{A})$ such that $u(t)<p-\delta<p$ for $t\in\mathcal{A}$ (almost surely).

Let $\chi_{\mathcal{A}}(\cdot)$ be the characteristic function on the set $\mathcal{A}$. As $T^* = 2.5 + \frac{40}{100-p^2}$ is the global optimum of Problem (6.9) according to Corollary B.4, we have $T \geq T^* > 1$. We set

$$u'(t) = u(t) + \frac{\delta}{T}\chi_{\mathcal{A}}(t)\,.$$

Then by $T > 1$ and the choice of δ we get $u'(t) \in [-10, 10]$ and $\|u'(\cdot) - u(\cdot)\|_\infty = \delta < \delta_u$. Let $\mathbf{x}'(\cdot\,; p)$ denote the differential states which are determined by T, $u'(\cdot)$, and p. We have $\mathbf{x}_2'(t; p) = \mathbf{x}_2(t; p)$ for all $t \in [0, 1-\varepsilon]$, and for $t \in [1-\varepsilon, 1]$ we get

$$\begin{aligned}
\mathbf{x}_2'(t; p) &= \mathbf{x}_2'(1-\varepsilon; p) + T\int_{1-\varepsilon}^{t} u'(\tau) - p\,\mathrm{d}\tau \\
&= \mathbf{x}_2(1-\varepsilon; p) + T\int_{1-\varepsilon}^{t} u(\tau) - p + \frac{\delta}{T}\chi_{\mathcal{A}}(\tau)\,\mathrm{d}\tau \\
&= \mathbf{x}_2(t; p) + \delta\lambda\left(\mathcal{A} \cap [1-\varepsilon, t]\right)\,.
\end{aligned}$$

We conclude $\mathbf{x}_2'(t; p) \geq \mathbf{x}_2(t; p)$ for all $t \in [0, 1]$ and $\mathbf{x}_2'(t; p) \leq \mathbf{x}_2(t; p) + \delta < \delta$ for all $t \in [1-\varepsilon, 1]$. Thus, by the choice of δ we get $\mathbf{x}_2'(t; p) \leq 4$ for all $t \in [0, 1]$, and furthermore

$$\mathbf{x}_2(1; p) < \mathbf{x}_2'(1; p) = \mathbf{x}_2(1; p) + \delta\lambda(\mathcal{A}) \leq \mathbf{x}_2(1; p) + \delta < 0\,.$$

As $\mathbf{x}_2(\cdot\,; p)$ and $\mathbf{x}_2'(\cdot\,; p)$ are continuous functions, we have $10 \leq \mathbf{x}_1(1; p) < \mathbf{x}_1'(1; p)$. In particular, $\left(T, u'(\cdot), \mathbf{x}'(\cdot\,; p)\right)$ is feasible and has the desired properties. □

Lemma B.7
Let $\delta_T > 0$ and $\left(T, u(\cdot), \mathbf{x}(\cdot\,; p)\right)$ be feasible for Problem (6.9) with

$$\mathbf{x}_1(1\,; p) > 10\,.$$

Then there is a feasible $\left(T', u'(\cdot), \mathbf{x}'(\cdot\,; p)\right)$ with $u'(\cdot) = u(\cdot)$, $|T' - T| < \delta_T$, and $T' < T$.

Proof Since $\mathbf{x}_1(1; p) > 10$ there is a $0 < T' < T$ with $\left(\frac{T'}{T}\right)^2 \mathbf{x}_1(1\,; p) \geq 10$ and $|T' - T| < \delta_T$. Let $\mathbf{x}'(\cdot\,; p)$ denote the differential states which are determined by T', $u(\cdot)$, and p. We have

$$\mathbf{x}_2'(t; p) = T'\int_0^t u(\tau) - p\,\mathrm{d}\tau = \frac{T'}{T}\mathbf{x}_2(t; p) \leq \mathbf{x}_2(t; p)$$

for all $t \in [0, 1]$, and

$$\mathbf{x}_1'(1; p) = T'\int_0^1 \mathbf{x}_2'(t; p)\,\mathrm{d}t = \left(\frac{T'}{T}\right)^2 T\int_0^1 \mathbf{x}_2(t; p)\,\mathrm{d}t = \left(\frac{T'}{T}\right)^2 \mathbf{x}_1(1; p) \geq 10\,.$$

Thus, $\big(T', u(\cdot), \mathbf{x}'(\cdot\,; p)\big)$ is feasible. □

Proposition B.8
Let $\delta_u > 0$ and $\big(T, u(\cdot), \mathbf{x}(\cdot\,; p)\big) \neq \big(T^*, u^*(\cdot), \mathbf{x}^*(\cdot\,; p)\big)$ be feasible for Problem (6.9) with

$$\mathbf{x}_1(1\,; p) = 10 \quad \text{and} \quad \mathbf{x}_2(1\,; p) = 0\,.$$

Then there is a feasible tuple $\big(T', u'(\cdot), \mathbf{x}'(\cdot\,; p)\big)$ with $T' = T$, $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$, and $\mathbf{x}_1'(1; p) > 10$.

Proof We have $T > T^*$ as the global optimum of Problem (6.9) is unique, see Corollary B.4. Let $t_1 = \frac{4}{T(10-p)}$ and $t_2 = 1 - \frac{4}{T(10+p)}$. From Lemma B.2 we get $0 < t_1 < t_2 < 1$. We distinct two cases:

Case 1): $u(\cdot) \not\equiv 10$ on $[0, t_1]$ (almost surely). Similar to the proof of Proposition B.3 we show $\mathbf{x}_2(t; p) < 4$ for all $t \in [0, t_1]$. In particular, there is $\delta > 0$ with $\mathbf{x}_2(t; p) \leq 4 - \delta$ for all $t \in [0, t_1]$. Furthermore, there is a $0 < \bar{\delta}_1 < \frac{\delta}{2}$ and a subset $\mathcal{A}_1 \subseteq [0, t_1]$ with non-zero (Lebesgue-)measure $\lambda(\mathcal{A}_1)$ (< 1) such that $u(t) < 10 - \bar{\delta}_1$ for $t \in \mathcal{A}_1$ (almost surely). Let $\chi_{\mathcal{A}_1}(\cdot)$ denote the characteristic function on the set $\mathcal{A}_1$. Then, for all $t \in [0, t_1]$ and $0 < \delta' < \bar{\delta}_1$ we get

$$\mathbf{x}_2(t; p) \leq T \int_0^t \left(u(\tau) - p + \frac{1}{T}\delta' \chi_{\mathcal{A}_1}(\tau) \right) \mathrm{d}\tau \leq \mathbf{x}_2(t; p) + \delta' \lambda(\mathcal{A}_1) < \mathbf{x}_2(t; p) + \frac{\delta}{2} < 4\,, \quad \text{(B.8)}$$

which we will need later in the course of the proof. Again, we distinct two cases:

Case i): $\mathbf{x}_2(t; p) + \frac{\delta}{2} \leq 4$ for all $t \geq t_1$. If we had $u(\cdot) \equiv -10$ on $[t_1, 1]$ (almost surely), due to $t_2 > t_1$ we would get

$$\begin{aligned} \mathbf{x}_2(1; p) &< 4 + T \int_{t_1}^{t_2} -10 - p \,\mathrm{d}t + T \int_{t_2}^{1} -10 - p \,\mathrm{d}t \\ &= 4 - T(10 + p)(1 - t_2) + T \int_{t_1}^{t_2} -10 - p \,\mathrm{d}t = -T \int_{t_1}^{t_2} 10 + p \,\mathrm{d}t < 0 \end{aligned}$$

which contradicts the prerequisites of the proposition. Thus, there is a $\bar{\delta}_2 > 0$ and a subset $\mathcal{A}_2 \subseteq [t_1, 1]$ with non-zero measure $\lambda(\mathcal{A}_2)$ (< 1) such that $u(t) > -10 + \bar{\delta}_2$ on $\mathcal{A}_2$ (almost surely). We set $\delta' = \frac{1}{2} \min\big(\bar{\delta}_1, \bar{\delta}_2, \delta_u\big)$ $\big(< \frac{\delta}{2}\big)$. Then

$$u(t) < 10 - \delta' \text{ for } t \in \mathcal{A}_1 \subseteq [0, t_1] \quad \text{and} \quad u(t) > -10 + \delta' \text{ for } t \in \mathcal{A}_2 \subseteq [t_1, 1]$$

(almost surely, respectively). We can choose $\mathcal{A}_1$ and $\mathcal{A}_2$ such that $\lambda(\mathcal{A}_1) = \lambda(\mathcal{A}_2)$ and define

$$u'(t) = u(t) + \frac{1}{T}\delta'\chi_{\mathcal{A}_1}(t) - \frac{1}{T}\delta'\chi_{\mathcal{A}_2}(t)\,.$$

Since $T > T^* > 1$, we have $u'(t) \in [-10, 10]$ and $\|u(\cdot) - u'(\cdot)\|_\infty < \delta' < \delta_u$. Let $\mathbf{x}'(\cdot\,; p)$ denote the differential states which are determined by T, $u'(\cdot)$, and p. Due to (B.8), for all $t \in [0, t_1]$ we have

$$\mathbf{x}_2(t;p) \le T\int_0^t \left(u(\tau) - p + \frac{1}{T}\delta'\chi_{\mathcal{A}_1}(\tau)\right)\mathrm{d}\tau = \mathbf{x}_2'(t;p) < \mathbf{x}_2(t;p) + \delta' < \mathbf{x}_2(t;p) + \frac{\delta}{2} < 4$$

and $\mathbf{x}_2'(t_1;p) > \mathbf{x}_2(t_1;p)$. Furthermore, for $t \in [t_1, 1]$ we get

$$\begin{aligned}\mathbf{x}_2'(t;p) &= \mathbf{x}_2(t;p) + \delta'\lambda(\mathcal{A}_1) - T\int_{t_1}^t \frac{1}{T}\delta'\chi_{\mathcal{A}_2}(\tau)\,\mathrm{d}\tau \\ &= \mathbf{x}_2(t;p) + \delta'\lambda(\mathcal{A}_1) - \delta'\lambda(\mathcal{A}_2 \cap [t_1, t]) \ge \mathbf{x}_2(t;p)\,.\end{aligned}$$

Altogether, we see $\mathbf{x}_2'(1;p) = \mathbf{x}_2(1;p) = 0$ and

$$\mathbf{x}_2(t;p) \le \mathbf{x}_2'(t;p) < \mathbf{x}_2(t;p) + \delta' < \mathbf{x}_2(t;p) + \frac{\delta}{2} \le 4$$

for all $t \in [0, 1]$ due to the assumption in the beginning of Case i). Since $\mathbf{x}_2(t_1;p) < \mathbf{x}_2'(t_1;p)$, by the continuity of $\mathbf{x}_2(\cdot\,;p)$ and $\mathbf{x}_2'(\cdot\,;p)$ we get $\mathbf{x}_1'(1;p) > \mathbf{x}_1(1;p) = 10$, and $\big(T, u'(\cdot), \mathbf{x}'(\cdot\,;p)\big)$ is feasible.

Case ii): There is a $t \in [t_1, 1]$ with $\mathbf{x}_2(t;p) > 4 - \frac{\delta}{2}$. Since $\mathbf{x}_2(t_1;p) < 4 - \frac{\delta}{2}$ according to (B.8) and $\mathbf{x}_2(\cdot\,;p)$ is continuous, the minimum

$$\bar{t} = \min_{t\in[t_1,1]} \left\{ t \,\middle|\, \mathbf{x}_2(t;p) = 4 - \frac{\delta}{2} \right\}$$

exists and we have $t_1 < \bar{t}$. In particular, $\mathbf{x}_2(t;p) < 4 - \frac{\delta}{2}$ for all $t \in [t_1, \bar{t})$ and $\mathbf{x}_2(\bar{t};p) = 4 - \frac{\delta}{2}$. Hence, there exists a subset $\mathcal{A}_2 \subseteq [t_1, \bar{t}] \subseteq [t_1, 1]$ with non-zero measure $\lambda(\mathcal{A}_2)$ such that $u(t) > p$ on $\mathcal{A}_2$ (almost surely). We set $\delta' = \frac{1}{2}\min\big(\delta_1, \delta_u\big)\big(< \frac{\delta}{2}\big)$. Then – as in Case i) – we have

$$u(t) < 10 - \delta' \text{ for } t \in \mathcal{A}_1 \subseteq [0, t_1] \quad \text{and} \quad u(t) > -10 + \delta' \text{ for } t \in \mathcal{A}_2 \subseteq [t_1, 1]\,.$$

We can choose $\mathcal{A}_1$ and $\mathcal{A}_2$ such that $\lambda(\mathcal{A}_1) = \lambda(\mathcal{A}_2)$ and define

$$u'(t) = u(t) + \frac{1}{T}\delta'\chi_{\mathcal{A}_1}(t) - \frac{1}{T}\delta'\chi_{\mathcal{A}_2}(t).$$

As in Case i), we have $u'(t) \in [-10, 10]$ (almost surely) and $\|u(\cdot) - u'(\cdot)\|_\infty < \delta' < \delta_u$. Let $\mathbf{x}'(\cdot\,; p)$ denote the differential states which are determined by T, $u'(\cdot)$, and p. Then similar to Case i) we have $\mathbf{x}'_2(t; p) \leq \mathbf{x}_2(t; p) + \frac{\delta}{2} < 4$ for all $t \in [0, t_1]$ and $\mathbf{x}'_2(t_1; p) > \mathbf{x}_2(t_1; p)$. Furthermore, for $t \in [t_1, \bar{t}]$ we get

$$\mathbf{x}'_2(t; p) = \mathbf{x}_2(t; p) + \delta'\lambda(\mathcal{A}_1) - \delta'\lambda(\mathcal{A}_2 \cap [t_1, \bar{t}]) \geq \mathbf{x}_2(t; p).$$

In particular, $\mathbf{x}'_2(t; p) = \mathbf{x}_2(t; p)$ for all $t \geq \bar{t}$. As in Case i), we conclude that the tuple $\left(T, u'(\cdot), \mathbf{x}'(\cdot\,; p)\right)$ is feasible with $\mathbf{x}'_1(1; p) > \mathbf{x}_1(1; p) = 10$.

Case 2): $u(\cdot) \equiv 10$ in $[0, t_1]$ (almost surely). Then we have $\mathbf{x}_2(t_1; p) = 4$. If $\mathbf{x}_2(t; p) = 4$ for all $t \in [t_1, t_2]$, then $u(\cdot) \equiv -10$ in $[t_2, 1]$ (almost surely) due to the terminal constraint $\mathbf{x}_2(1; p) \leq 0$. Thus $u(\cdot) = u_T(\cdot\,; p)$ and $\mathbf{x}_1(1; p) > 10$ by construction since $T > T^*$. This contradicts the prerequisites of the proposition. Hence, there is a $t \in (t_1, t_2]$ with $\mathbf{x}_2(t; p) < 4$. Let

$$\bar{t} = \inf_{t \in [t_1, t_2]} \left\{ t \,\middle|\, \mathbf{x}_2(t; p) < 4 \right\}.$$

By the continuity of $\mathbf{x}_2(t; p)$ we get $t_1 \leq \bar{t} < t_2$ and $\mathbf{x}_2(t; p) = 4$ for all $t \in [t_1, \bar{t}]$. There is a $\varepsilon > 0$ with $\bar{t} + \varepsilon < t_2 - \varepsilon$. Again by the continuity of $\mathbf{x}_2(\cdot\,; p)$, there is a

$$t_l \in \arg\min_{t \in [\bar{t}, \bar{t}+\varepsilon]} \mathbf{x}_2(t; p),$$

and by the choice of $\bar{t}$ we have $t_l > \bar{t}$ and $\mathbf{x}_2(t_l; p) < 4$. Furthermore, there are $\varepsilon', \delta > 0$ such that

$$\mathbf{x}_2(t; p) < 4 - \delta \quad \text{for all } t \in [t_l - \varepsilon', t_l + \varepsilon'] \cap [\bar{t}, \bar{t} + \varepsilon] = [\tau_1, \tau_2].$$

We have $\tau_1 < t_l \leq \tau_2 \leq \bar{t} + \varepsilon < t_2 - \varepsilon$. Since $\mathbf{x}_2(t_l; p) \leq \mathbf{x}_2(t; p)$ for all $t \in [\tau_1, \tau_2] \subseteq [\bar{t}, \bar{t} + \varepsilon]$, there is a subset $\mathcal{A}_1 \subseteq [\tau_1, \tau_2]$ with non-zero measure such that $u(t) \leq p$ on $\mathcal{A}_1$. Otherwise, due to $\tau_1 < t_l$ we would get $\mathbf{x}_2(\tau_1; p) < \mathbf{x}_2(t_l; p)$. In particular, there is a $\bar{\delta}_1$ with $0 < \bar{\delta}_1 < \frac{\delta}{2}$ such that $u(t) < 10 - \bar{\delta}_1$ on $\mathcal{A}_1$ (almost surely). Similar to Case 1),

for all $t \in [\tau_1, \tau_2]$ and $0 < \delta' < \bar{\delta}_1$ we get

$$\mathbf{x}_2(\tau_1; p) + T \int_{\tau_1}^{t} \left(u(\tau) - p + \frac{1}{T} \delta' \chi_{\mathcal{A}_1}(\tau) \right) \mathrm{d}\tau \leq \mathbf{x}_2(t; p) + \delta' \lambda(\mathcal{A}_1) < \mathbf{x}_2(t; p) + \frac{\delta}{2} < 4 \,.$$

Again we distinct two cases: either $\mathbf{x}_2(t; p) + \frac{\delta}{2} \leq 4$ for all $t \geq \tau_2$, or there is a $t \in [\tau_2, 1]$ with $\mathbf{x}_2(t; p) + \frac{\delta}{2} > 4$. Since $\tau_2 \leq t_2 - \varepsilon$, we can argue as in Case 1), subcases i)-ii), and construct a feasible tuple a feasible $\left(T', u'(\cdot), \mathbf{x}'(\cdot; p)\right)$ with $T' = T$, $\| u'(\cdot) - u(\cdot) \|_\infty < \delta_u$, and $\mathbf{x}'_1(1; p) > 10$. □

We are now able to prove that Problem (6.9) is uniquely solvable:

Proposition B.9
The tuple $\left(T^*, u^*(\cdot), \mathbf{x}^*(\cdot; p)\right)$ is the only local optimum of Problem (6.9) in the considered normed space.

Proof We know that $\left(T^*, u^*(\cdot), \mathbf{x}^*(\cdot; p)\right)$ is the global minumum of Problem (6.9), see Proposition B.4. Now let $\left(T, u(\cdot), \mathbf{x}(\cdot; p)\right) \neq \left(T^*, u^*(\cdot), \mathbf{x}^*(\cdot; p)\right)$ be feasible for Problem (6.9). Since the global minimum is unique, we have $T > T^*$. Let $\delta_T, \delta_u > 0$. We distinct two cases:

Case 1): $\mathbf{x}_1(1; p) = 10$. We can apply Lemma B.6 or Proposition B.8 to construct a feasible $\left(T', u'(\cdot), \mathbf{x}'(\cdot; p)\right)$ with $T' = T$, $\| u'(\cdot) - u(\cdot) \|_\infty < \delta_u$, and $\mathbf{x}'_1(1; p) > 10$. Subsequently, we apply Lemma B.7 and find a feasible tuple $\left(T'', u''(\cdot), \mathbf{x}''(\cdot; p)\right)$ with $\| u''(\cdot) - u(\cdot) \|_\infty = \| u'(\cdot) - u(\cdot) \|_\infty < \delta_u$, $|T'' - T| = |T'' - T'| < \delta_T$, and $T'' < T$.

Case 2): $\mathbf{x}_1(1; p) > 10$. We can apply Lemma B.7 and find a feasible $\left(T', u'(\cdot), \mathbf{x}'(\cdot; p)\right)$ with $u'(\cdot) = u(\cdot)$, $|T' - T| < \delta_T$, and $T' < T$.

Using Lemma B.5 we conclude that $\left(T, u(\cdot), \mathbf{x}(\cdot; p)\right)$ is not a local minimum. □

B.2.5 Sketch of Proof of Remark 6.3

Let $p \geq 0$, $T = T^*(p) = 2.5 + \frac{40}{100-p^2}$ and $t_1 = \frac{4}{T(10-p)}$ and $t_2 = 1 - \frac{4}{T(10+p)}$ as in Appendix B.2.4. Then we get

$$\begin{aligned}
& & t_2 &> t_1 \\
\iff & & T &> \tfrac{4}{10-p} + \tfrac{4}{10+p} \\
\iff & 2.5 + \tfrac{40}{100-p^2} &&> \tfrac{4}{10-p} + \tfrac{4}{10+p} = \tfrac{80}{100-p^2} \\
\iff & 2.5 &&> \tfrac{40}{100-p^2} \\
\iff & 100 - p^2 &&> 16 \\
\iff & 84 &&> p^2 \\
\iff & 2\sqrt{21} &&> p,
\end{aligned}$$

as $p \geq 0$. Hence, for $p \geq 2\sqrt{21}$ the function $u_T(\cdot\,; p)$, see (B.5), is not well-defined anymore and the optimal strategy needs to be adapted. If $2\sqrt{21} \leq p < 10$, for the optimal controllable parameter we get $T^*(p) = \frac{20}{\sqrt{100-p^2}}$ and the optimal control function $u^*(\cdot)$ is given by

$$u^*(t) = \begin{cases} 10 & \text{for } 0 \leq t < \frac{10+p}{20}, \\ -10 & \text{for } \frac{10+p}{20} \leq t \leq 1. \end{cases}$$

The proof of this statement works similarly as the one given in Appendix B.2.4.

B.2.6 Solution of Problem (6.10)

Let $\Omega_p = [p_l, p_u] \subseteq [0,9]$ with $p_l < p_u$. In this section, we consider Problem (6.10) with variables

$$\left(T, u(\cdot), p, \mathbf{x}(\cdot\,; p)\right) \in \mathbb{R} \times L^\infty([0,1], \mathbb{R}) \times \mathbb{R} \times W^{1,\infty}\left([0,1], \mathbb{R}^2\right),$$

see Section 2.1 for the spaces and corresponding norms, where $\mathbf{x}(\cdot\,; p)$ denotes the (unique) solution of IVP (6.10b-6.10c) for given T, $u(\cdot)$, and $p \in \Omega_p$. The product space $\mathbb{R} \times L^\infty([0,1], \mathbb{R}) \times \mathbb{R} \times W^{1,\infty}\left([0,1], \mathbb{R}^2\right)$, equipped with the norm

$$\left\| \left(T, u(\cdot), p, \mathbf{x}(\cdot\,; p)\right) \right\| = \max\left[|T|, \|u(\cdot)\|_\infty, |p|, \|\mathbf{x}(\cdot\,; p)\|_{1,\infty} \right],$$

is a Banach space. In this section, we derive a condition for the non-emptiness of the feasible set of Problem (6.10) and compute the set of globally optimal solutions for the case that the feasible set is non-empty. Furthermore, if $\left(T, u(\cdot), p, \mathbf{x}(\cdot\,; p)\right)$ is a (locally) optimal solution, then $T = T^*$ and $u(\cdot) = u^*(\cdot)$ for a certain parameter T^*

and a certain control function $u^*(\cdot)$. The main results of this section can be found in Proposition B.14 and Corollary B.22.

A Reformulation of Problem (6.10)

We state a problem reformulation which will be useful for the subsequent investigations. Let $T \geq 0$ and $u : [0,1] \to [-10,10]$. For all $p \in \Omega_p = [p_l, p_u]$ we have

$$u(t) - p_u \leq u(t) - p \leq u(t) - p_l\,, \quad t \in [0,1]\,.$$

By the monotonicity of the integral, we get

$$\mathbf{x}_2(t;p_u) \leq \mathbf{x}_2(t;p) \leq \mathbf{x}_2(t;p_l)\,, \quad t \in [0,1]\,,$$

and thus also

$$\mathbf{x}_1(t;p_u) \leq \mathbf{x}_1(t;p) \leq \mathbf{x}_1(t;p_l)\,, \quad t \in [0,1]\,.$$

In particular

$$\begin{bmatrix} \mathbf{x}_2(t;p) \leq 4\,, & t \in [0,1]\,, \text{ for all } p \in \Omega_p\,, \\ \mathbf{x}_1(1;p) \geq 10\,, & \text{for all } p \in \Omega_p\,, \\ \mathbf{x}_2(1;p) \leq 0\,, & \text{for all } p \in \Omega_p \end{bmatrix} \iff \begin{bmatrix} \mathbf{x}_2(t;p_l) \leq 4\,, & t \in [0,1]\,, \\ \mathbf{x}_1(1;p_u) \geq 10\,, & \\ \mathbf{x}_2(1;p_l) \leq 0 & \end{bmatrix}. \tag{B.9}$$

We consider the problem

$$\min_{T,u(\cdot),\mathbf{x}(\cdot;\Omega_p)} \quad T \tag{B.10a}$$

$$\text{s.t.} \quad \dot{\mathbf{x}}\left(t;\Omega_p\right) = T \begin{pmatrix} \mathbf{x}_2\left(t;\Omega_p\right) \\ u(t) - p_l \\ \mathbf{x}_4\left(t;\Omega_p\right) \\ u(t) - p_u \end{pmatrix}, \quad t \in [0,1]\,, \tag{B.10b}$$

$$\mathbf{x}\left(0;\Omega_p\right) = \mathbf{0}\,, \tag{B.10c}$$

$$\mathbf{x}_2\left(t;\Omega_p\right) \leq 4\,, \qquad t \in [0,1]\,, \tag{B.10d}$$

$$\mathbf{x}_3\left(1;\Omega_p\right) \geq 10\,, \tag{B.10e}$$

$$\mathbf{x}_2\left(1;\Omega_p\right) \leq 0\,, \tag{B.10f}$$

$$T \geq 0\,, \tag{B.10g}$$

$$u(t) \in [-10,10]\,, \qquad t \in [0,1]\,, \tag{B.10h}$$

with variables

$$\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right) \in \mathbb{R} \times L^\infty\left([0,1],\mathbb{R}\right) \times W^{1,\infty}\left([0,1],\mathbb{R}^4\right)$$

where $\mathbf{x}\left(\cdot;\Omega_p\right)$ denotes the (unique) solution of the IVP (B.10b-B.10c) for given T, $u(\cdot)$, and Ω_p. If we equip the above space with the norm

$$\left\|\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)\right\| = \max\left[|T|, \|u(\cdot)\|_\infty, \left\|\mathbf{x}\left(\cdot;\Omega_p\right)\right\|_{1,\infty}\right],$$

it is a Banach space. We get

Lemma B.10
Let $T > 0$, $u(\cdot) \in L^\infty\left([0,1],\mathbb{R}\right)$, and $\mathbf{x}\left(\cdot;\Omega_p\right) \in W^{1,\infty}\left([0,1],\mathbb{R}^4\right)$ the differential states which are determined by T, $u(\cdot)$, and Ω_p. Let $\varepsilon > 0$. Then there exist $\delta_T, \delta_u > 0$ such that

$$\left\|\mathbf{x}'\left(\cdot;\Omega_p\right) - \mathbf{x}\left(\cdot;\Omega_p\right)\right\|_{1,\infty} < \varepsilon$$

for all T', $u'(\cdot)$ with $|T' - T| < \delta_T$ and $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$, where $\mathbf{x}'\left(\cdot;\Omega_p\right)$ denotes the differential states which are determined by T', $u'(\cdot)$, and Ω_p.

Proof Similar to proof of Lemma B.5. □

We find the following equivalence of the Problems (B.10) and (6.10):

Lemma B.11
Problem (B.10) is equivalent to Problem (6.10) in the following sense: let $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ be feasible for Problem (B.10). Then for every $p \in \Omega_p = [p_l, p_u]$, the tuple $\left(T, u(\cdot), p, \mathbf{x}(\cdot;p)\right)$ is feasible for Problem (6.10) and the values of the objective functions coincide. Vice versa, let $\left(T, u(\cdot), p, \mathbf{x}(\cdot;p)\right)$ be feasible for Problem (6.10). Then $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ is feasible for Problem (B.10) and the values of the objective functions coincide. In particular, $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ is a local minimum of Problem (B.10) if and only if $\left(T, u(\cdot), p, \mathbf{x}(\cdot;p)\right)$ is a local minimum of Problem (6.10) for each $p \in \Omega_p$. The same holds for global minima.

Proof The feasibility assertions follow from the Equivalence (B.9) and the fact that every pair $\left(p, \mathbf{x}(\cdot;p)\right)$ is a (global) maximizer of the lower level problem of Problem (6.10) for given T and $u(\cdot)$. The accordance of the respective values of the objective functions is clear.

Let $\big(T, u(\cdot), \mathbf{x}\big(\cdot;\Omega_p\big)\big)$ be a local minimum of Problem (B.10), and $p \in \Omega_p$. If the tuple $\big(T, u(\cdot), p, \mathbf{x}(\cdot;p)\big)$ would not be a local minimum of Problem (6.10), then for every $\varepsilon > 0$ we find a feasible $\big(T'_\varepsilon, u'_\varepsilon(\cdot), p'_\varepsilon, \mathbf{x}'_\varepsilon(\cdot;p)\big)$ with $|T - T'_\varepsilon| < \varepsilon$, $\|u(\cdot) - u'_\varepsilon(\cdot)\|_\infty < \varepsilon$ and $T'_\varepsilon < T$. Let $\mathbf{x}'_\varepsilon\big(\cdot;\Omega_p\big)$ denote the solution of IVP (B.10b-B.10c) which is determined by T'_ε, $u'_\varepsilon(\cdot)$, and Ω_p. By the already proven part of the lemma, $\big(T'_\varepsilon, u'_\varepsilon(\cdot), \mathbf{x}'_\varepsilon\big(\cdot;\Omega_p\big)\big)$ is feasible for Problem (B.10). By Lemma B.10, $\big(T'_\varepsilon, u'_\varepsilon(\cdot), \mathbf{x}'_\varepsilon\big(\cdot;\Omega_p\big)\big)$ lies in an arbitrary small neighborhood of $\big(T, u(\cdot), \mathbf{x}\big(\cdot, \Omega_p\big)\big)$ if ε is small enough. Thus, $\big(T, u(\cdot), \mathbf{x}\big(\cdot;\Omega_p\big)\big)$ is no local minimum of Problem (B.10) which is a contradiction.

The transfer of local minima from Problem (6.10) to Problem (B.10) can be shown similarly, and the transfer of global minima from Problem (6.10) to Problem (B.10) and vice versa follows from the first part of the lemma. □

Justified by the previous Lemma, we focus on Problem (B.10) in the following.

Non-Emptiness of Feasible Set and Global Optimum of Problem (B.10)

First, we investigate the feasible set of Problem (B.10). We start with

Lemma B.12
Let $\big(T, u(\cdot), \mathbf{x}\big(\cdot;\Omega_p\big)\big)$ be feasible for Problem (B.10). Then $T > \frac{4}{10+p_l} + \frac{4}{10-p_l}$.

Proof We proceed similar as in the proof of Lemma B.2 for $p = p_l$ and make use of $\mathbf{x}_3\big(t;\Omega_p\big) \leq \mathbf{x}_1\big(t;\Omega_p\big)$. □

Let $T > \frac{4}{10+p_l} + \frac{4}{10-p_l}$. We set

$$t_1(T,\Omega_p) = \frac{4}{T(10-p_l)} \quad \text{and} \quad t_2(T,\Omega_p) = 1 - \frac{4}{T(10+p_l)}\,.$$

Then $0 < t_1(T,\Omega_p) < t_2(T,\Omega_p) < 1$. We define

$$u_T\big(t;\Omega_p\big) = \begin{cases} 10 & \text{for } 0 \leq t < t_1(T,\Omega_p)\,, \\ p_l & \text{for } t_1(T,\Omega_p) \leq t < t_2(T,\Omega_p)\,, \\ -10 & \text{for } t_2(T,\Omega_p) \leq t \leq 1\,. \end{cases}$$

Let $\mathbf{x}_T(\cdot;\Omega_p) \in \mathbb{R}^4$ denote the differential states which are determined by Ω_p, T, and $u_T(\cdot;\Omega_p)$. Then we have

$$\mathbf{x}_{T,2}(t;\Omega_p) = \begin{cases} T(10-p_l)\,t & \text{for } 0 \le t < t_1(T,\Omega_p), \\ 4 & \text{for } t_1(T,\Omega_p) \le t < t_2(T,\Omega_p), \\ 4 - T(10+p_l)\,(t - t_2(T,\Omega_p)) & \text{for } t_2(T,\Omega_p) \le t \le 1. \end{cases}$$

In particular, $\mathbf{x}_{T,2}(t;\Omega_p) \le 4$ for all $t \in [0,1]$ and $\mathbf{x}_{T,2}(1;\Omega_p) = 0$. Similar as in Appendix B.2.4, Equation (B.6), we get

$$\mathbf{x}_{T,1}(1;\Omega_p) = 4T - \frac{8}{10-p_l} - \frac{8}{10+p_l}.$$

Furthermore, $\mathbf{x}_{T,4}(t;\Omega_p) = \mathbf{x}_{T,2}(t) - (p_u - p_l)\,T\,t$ and consequently

$$\begin{aligned} \mathbf{x}_{T,3}(1;\Omega_p) &= \mathbf{x}_{T,1}(1;\Omega_p) - \frac{1}{2}(p_u - p_l)\,T^2 \\ &= -\frac{1}{2}(p_u - p_l)\,T^2 + 4T - \frac{8}{10-p_l} - \frac{8}{10+p_l}. \end{aligned}$$

Thus, the tuple $\left(T, u_T(\cdot;\Omega_p), \mathbf{x}_T(\cdot;\Omega_p)\right)$ is feasible if and only if $\mathbf{x}_{T,3}(1;\Omega_p) \ge 10$. Similar to Appendix B.2.4, Proposition B.3, we have

Proposition B.13
Let $\left(T, u(\cdot), \mathbf{x}(\cdot;\Omega_p)\right)$ be feasible for Problem (B.10) with $u(\cdot) \neq u_T(\cdot;\Omega_p)$ (as elements of $L^\infty([0,1],\mathbb{R})$). Then, $\left(T, u_T(\cdot;\Omega_p), \mathbf{x}_T(\cdot;\Omega_p)\right)$ is feasible and we have

$$10 \le \mathbf{x}_3(1;\Omega_p) < \mathbf{x}_{T,3}(1;\Omega_p).$$

Proof We have $\mathbf{x}_2(t;\Omega_p) = \mathbf{x}_4(t;\Omega_p) + (p_u - p_l)\,T\,t$, and equally for $\mathbf{x}_T(\cdot;\Omega_p)$. Hence $\mathbf{x}_2(t;\Omega_p) \le \mathbf{x}_{T,2}(t;\Omega_p)$ implies $\mathbf{x}_4(t;\Omega_p) \le \mathbf{x}_{T,4}(t;\Omega_p)$ and the same holds for strict inequality. We use this and proceed similarly as in the proof of Proposition B.3. □

We define

$$g : \mathbb{R} \to \mathbb{R},\ T \longmapsto g(T;\Omega_p) = -\frac{1}{2}(p_u - p_l)\,T^2 + 4T - \frac{8}{10-p_l} - \frac{8}{10+p_l} - 10.$$

Let $\underline{T} = \frac{4}{10-p_l} + \frac{4}{10+p_l}$. Then $\underline{T} = \frac{80}{100-p_l^2} \le \frac{80}{19} < 5$ and, as $p_u > p_l$,

$$g(\underline{T};\Omega_p) = -\frac{1}{2}(p_u - p_l)\,\underline{T}^2 + 2\underline{T} - 10 \le 2\underline{T} - 10 < 0. \tag{B.11}$$

For $T > \underline{T}$ we have

$$g\left(T;\Omega_p\right) = \mathbf{x}_{T,3}\left(1;\Omega_p\right) - 10\,.$$

Since $p_u > p_l$, the map $g\left(\cdot;\Omega_p\right)$ is concave quadratic function with argument of the maximum $T_{\max} = \frac{4}{p_u - p_l}$ and corresponding value

$$g\left(T_{\max};\Omega_p\right) = \frac{8}{p_u - p_l} - \frac{8}{10 - p_l} - \frac{8}{10 + p_l} - 10\,. \tag{B.12}$$

A straightforward calculation shows

$$g\left(T_{\max};\Omega_p\right) \geq 0 \iff p_u \leq p_l + \frac{8}{10 + \frac{8}{10-p_l} + \frac{8}{10+p_l}}\,. \tag{B.13}$$

We can now characterize the non-emptiness of the feasible set of Problem (B.10).

Proposition B.14
The feasible set of Problem (B.10) is non-empty if and only if

$$p_u \leq p_l + \frac{8}{10 + \frac{8}{10-p_l} + \frac{8}{10+p_l}}\,. \tag{B.14}$$

The same holds for the feasible set of Problem (6.10). If the feasible sets are non-empty, we have

$$T_{\max} = \frac{4}{p_u - p_l} \geq 5 + \frac{4}{10 - p_l} + \frac{4}{10 + p_l}\,. \tag{B.15}$$

Proof By Lemma B.11, the feasible set of Problem (B.10) is non-empty if and only if the feasible set of Problem (6.10) is non-empty. If the feasible set of Problem (B.10) is non-empty, there is a feasible tuple $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$. By Proposition B.13, the tuple $\left(T, u_T\left(\cdot;\Omega_p\right), \mathbf{x}_T\left(\cdot;\Omega_p\right)\right)$ is feasible. In particular, we have

$$0 \leq \mathbf{x}_{T,3}\left(1;\Omega_p\right) - 10 = g\left(T;\Omega_p\right) \leq g\left(T_{\max};\Omega_p\right) = \frac{8}{p_u - p_l} - \frac{8}{10 - p_l} - \frac{8}{10 + p_l} - 10\,,$$

and (B.14) follows from the Equivalence (B.13). Conversely, we assume that (B.14) holds. From the Equivalence (B.13) and Equation (B.12) we get

$$g\left(T_{\max};\Omega_p\right) = 2\left(\frac{4}{p_u - p_l} - \frac{4}{10 - p_l} - \frac{4}{10 + p_l} - 5\right) \geq 0\,,$$

and in particular

$$T = T_{\max} = \frac{4}{p_u - p_l} \geq 5 + \frac{4}{10 - p_l} + \frac{4}{10 + p_l} > \frac{4}{10 - p_l} + \frac{4}{10 + p_l}.$$

Thus, $u_T(\cdot;\Omega_p)$ is well-defined and $\mathbf{x}_{T,3}(1;\Omega_p) = g(T_{\max};\Omega_p) + 10 \geq 10$. In particular, $(T, u_T(\cdot;\Omega_p), \mathbf{x}_T(\cdot;\Omega_p))$ is feasible for Problem (B.10) and the feasible set is non-empty. □

Next, we compute the unique global optimum of Problem (B.10). Let the feasible set of Problem (B.10) be non-empty. Then by the previous Proposition and Equation (B.13), we have $g\left(\frac{4}{p_u - p_l};\Omega_p\right) \geq 0$ and $\frac{4}{p_u - p_l} > \frac{4}{10 - p_l} + \frac{4}{10 + p_l}$. On the other hand, $g\left(\frac{4}{10 - p_l} + \frac{4}{10 + p_l};\Omega_p\right) < 0$, see (B.11). Let $\mathcal{Z} = \{T \in \mathbb{R} \mid g(T;\Omega_p) = 0\}$. As $g(\cdot;\Omega_p)$ is a (non-constant) concave quadratic function, for the cardinality of $\mathcal{Z}$ we have $|\mathcal{Z}| \in \{1,2\}$. We set

$$z(\Omega_p) \stackrel{\text{def}}{=} \min_{z \in \mathcal{Z}} z > \frac{4}{10 - p_l} + \frac{4}{10 + p_l}. \tag{B.16}$$

By definition of $z(\Omega_p)$ and the properties of $g(\cdot;\Omega_p)$, we get the implication

$$g(T;\Omega_p) \geq 0 \implies z(\Omega_p) \leq T. \tag{B.17}$$

We define

$$T^* = T^*(\Omega_p) \stackrel{\text{def}}{=} z(\Omega_p), \quad u^*(\cdot) = u^*(\cdot;\Omega_p) \stackrel{\text{def}}{=} u_{T^*}(\cdot;\Omega_p), \quad \text{and} \quad \mathbf{x}^*(\cdot;\Omega_p) \stackrel{\text{def}}{=} \mathbf{x}_{T^*}(\cdot;\Omega_p).$$

By construction, $(T^*, u^*(\cdot), \mathbf{x}^*(\cdot;\Omega_p))$ is feasible for Problem (B.10) and we get

Corollary B.15
Let the feasible set of Problem (B.10) be non-empty. Then $(T^*, u^*(\cdot), \mathbf{x}^*(\cdot;\Omega_p))$ is the unique global optimum of Problem (B.10).

Proof As the feasible set of Problem (B.10) is non-empty, $(T^*, u^*(\cdot), \mathbf{x}^*(\cdot;\Omega_p))$ is well-defined, see the preceding considerations. Let $(T, u(\cdot), \mathbf{x}(\cdot;\Omega_p))$ be any other feasible point. We note that $u(\cdot) = u_T(\cdot;\Omega_p)$ is not excluded. Thus, by Proposition B.13 we have

$$0 \leq \mathbf{x}_3(1;\Omega_p) - 10 \leq \mathbf{x}_{T,3}(1;\Omega_p) - 10 = g(T;\Omega_p)$$

and consequently $T^* = z(\Omega_p) \leq T$, see (B.17). Hence, $(T^*, u^*(\cdot), \mathbf{x}^*(\cdot;\Omega_p))$ is a global optimum. To show the uniqueness, we assume that there is another feasible point

$\left(T^*, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ with $u(\cdot) \neq u^*(\cdot) = u_{T^*}\left(\cdot;\Omega_p\right)$. We apply Proposition B.13 again and see

$$0 \leq \mathbf{x}_3\left(1;\Omega_p\right) - 10 < \mathbf{x}_{T^*,3}\left(1;\Omega_p\right) - 10 = g\left(T^*;\Omega_p\right) = 0$$

which is a contradiction. Thus, such a point does not exist and the global minimum is unique. □

The Unique Solvability of Problem (B.10)

In the following, we show that the global optimum of Problem (B.10) is the only local optimum in the considered normed space. We proceed similar as in Appendix B.2.4 and adapt the proofs where needed.

Lemma B.16
Let $\delta_u > 0$ and $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ be feasible for Problem (B.10) with

$$\mathbf{x}_2\left(1;\Omega_p\right) < 0\,.$$

Then there is a feasible tuple $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ with $T' = T$, $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$, and $\mathbf{x}'_3\left(1;\Omega_p\right) > 10$.

Proof We have $\mathbf{x}_2\left(t;\Omega_p\right) = \mathbf{x}_4\left(t;\Omega_p\right) + (p_u - p_l)\,T\,t$, and equally for $\mathbf{x}'\left(\cdot;\Omega_p\right)$. Hence, $\mathbf{x}_2(t;\Omega_p) \leq \mathbf{x}'_2(t;\Omega_p)$ implies $\mathbf{x}_4(t;\Omega_p) \leq \mathbf{x}'_4(t;\Omega_p)$ and the same holds for strict inequality. We exploit this fact and proceed similarly as in the proof of Lemma B.6. □

Lemma B.17
Let $\delta_T > 0$ and $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ be feasible for Problem (B.10) with

$$\mathbf{x}_3\left(1;\Omega_p\right) > 10\,.$$

Then there is a feasible $\left(T', u'(\cdot), \mathbf{x}'(\cdot)\right)$ with $u'(\cdot) = u(\cdot)$, $|T' - T| < \delta_T$, and $T' < T$.

Proof Similar to proof of Lemma B.7. □

It remains to investigate the case of a feasible, non-optimal tuple $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ for which the terminal constraints are satisfied with equality, see Proposition B.20. Proposition B.20 can be seen as the counterpart of Proposition B.8 in Appendix B.2.4. However in contrast to Appendix B.2.4, $T > T^*$ does not imply $\mathbf{x}_{T,3}\left(1;\Omega_p\right) > 10$. This is because $\mathbf{x}_{T,3}\left(1;\Omega_p\right) - 10 = g\left(T;\Omega_p\right)$ for $T \geq T^*$, $g\left(T^*;\Omega_p\right) = 0$, and $g\left(\cdot;\Omega_p\right)$ is a

concave quadratic function. Therefore, the proof of Proposition B.8 cannot be transferred directly to Proposition B.20 and needs to be adapted. For this, we need two auxiliary results:

Lemma B.18
Let $T > T^*$ such that $\left(T, u_T\left(\cdot;\Omega_p\right), \mathbf{x}_T\left(\cdot;\Omega_p\right)\right)$ is feasible for Problem (B.10) and $\mathbf{x}_{T,3}\left(\cdot;\Omega_p\right) = 10$. Then

$$T > \frac{4}{p_u - p_l}.$$

Furthermore, for $t_1 = \frac{4}{T(10-p_l)}$ and $t_2 = 1 - \frac{4}{T(10+p_l)}$ we have

$$T\left(t_2 - t_1\right) > 5.$$

Proof We have $g\left(T;\Omega_p\right) = \mathbf{x}_{T,3}\left(1;\Omega_p\right) - 10 = 0$, i. e., T is a zero of $g\left(\cdot;\Omega_p\right)$. Since $T > T^* = \min\left\{T \in \mathbb{R} \,\middle|\, g\left(T;\Omega_p\right) = 0\right\}$, the concave quadratic function $g(\cdot)$ has two zeros, namely T and T^*. The vertex of $g\left(\cdot;\Omega_p\right)$ is given by $T_{\max} = \frac{4}{p_u - p_l}$ and we have $T^* < T_{\max} < T$. Furthermore,

$$T(t_2 - t_1) = T - \frac{4}{10 + p_l} - \frac{4}{10 - p_l} > T_{\max} - \frac{4}{10 + p_l} - \frac{4}{10 - p_l} \geq 5,$$

where the latter inequality is due to Proposition B.14, Inequality (B.15). □

Proposition B.19
Let $\delta_T, \delta_u > 0$, and $T > T^*$ such that $\left(T, u_T\left(\cdot;\Omega_p\right), \mathbf{x}_T\left(\cdot;\Omega_p\right)\right)$ is feasible for Problem (B.10) with $\mathbf{x}_{T,3}\left(1;\Omega_p\right) = 10$. Then there is a feasible $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ with $\mathbf{x}'_3\left(1;\Omega_p\right) > 10$, $\left\| u'(\cdot) - u_T\left(\cdot;\Omega_p\right)\right\|_\infty < \delta_u$, $|T' - T| < \delta_T$, and $T' < T$.

Proof We start with technical preparations which become relevant later in the proof. Let $t_1 = \frac{4}{T(10-p_l)}$ and $t_2 = 1 - \frac{4}{T(10+p_l)}$. We have $T\left(t_2 - t_1\right) > 5$ according to Lemma B.18. Consequently, we can choose a $n \in \mathbb{N}$, $n > 2$, with

$$T\left(t_2 - t_1\right) > \frac{5n}{n-1}.$$

Next, we consider the quadratic function

$$h : \mathbb{R} \to \mathbb{R},\ y \mapsto (T - y)^2 \frac{10}{T^2} + (T - y) y \frac{4(n-1)}{Tn}\left(t_2 - t_1\right). \tag{B.18}$$

Then $h(0) = 10$ and

$$\frac{\mathrm{d}}{\mathrm{d}y} h(y)\Big|_{y=0} = -\frac{20}{T} + \frac{4(n-1)}{Tn} T(t_2 - t_1) > -\frac{20}{T} + \frac{4(n-1)}{Tn} \frac{5n}{n-1} = 0$$

by the choice of n. In particular, there is a $\varepsilon' > 0$ such that $h(y) > 10$ for all $0 < y < \varepsilon'$.

In the following, we construct a feasible $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ with the desired properties. Let

$$t_1' = t_1 + \frac{1}{n}(t_2 - t_1) \quad \text{and} \quad t_2' = t_1 + \frac{n-1}{n}(t_2 - t_1). \tag{B.19}$$

Then $t_1 < t_1' < t_2' < t_2$ (since $n > 2$). Furthermore, we set

$$\varepsilon = \frac{1}{2} \min\left(\delta_T, \frac{5}{4(n-1)}\delta_u, \frac{5}{4(n-1)}(10 - p_l), \varepsilon', 1\right)$$

and define

$$T' = T - \varepsilon \quad \text{and} \quad \delta = \frac{4n}{T'T(t_2 - t_1)}\varepsilon.$$

Then $|T' - T| < \delta_T$. From Lemma B.18 and Proposition B.14, Inequality (B.15), we get

$$T' = T - \varepsilon > T - 1 > \frac{4}{p_u - p_l} - 1 > 5 - 1 = 4,$$

and therefore

$$0 < \delta < \frac{n\varepsilon}{T(t_2 - t_1)} < \frac{4n\varepsilon}{T(t_2 - t_1)} < \frac{n-1}{5n} 4n\varepsilon = \frac{4(n-1)}{5}\varepsilon.$$

In particular, by the definition of ε we get

$$\delta < \delta_u \quad \text{and} \quad \delta < 10 - p_l. \tag{B.20}$$

We define

$$u'(t) = u_T\left(t;\Omega_p\right) + \delta\chi_{[t_1, t_1']}(t) - \delta\chi_{[t_2', t_2]}(t).$$

From (B.20) we get $\left\|u'(\cdot) - u_T\left(\cdot;\Omega_p\right)\right\|_\infty < \delta_u$ and, as $u_T\left(t;\Omega_p\right) = p_l$ for $t \in [t_1, t_1'] \cup [t_2', t_2]$, $u'(t) \in [-10, 10]$ for $t \in [0, 1]$.

Let $\mathbf{x}'(\cdot;\Omega_p)$ denote the differential states which are determined by T', $u'(\cdot)$, and Ω_p. Since $u'(t) = u_T(t;\Omega_p)$ in $[0, t_1]$, we have

$$\mathbf{x}_2'(t;\Omega_p) = T'\int_0^t u'(\tau) - p_l\,\mathrm{d}\tau = \frac{T'}{T}T\int_0^t u_T(\tau;\Omega_p) - p_l\,\mathrm{d}\tau = \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p)$$

for all $t \in [0, t_1]$.

For $t \in [t_1, t_1']$ we have $\mathbf{x}_{T,2}(t;\Omega_p) = \mathbf{x}_{T,2}(t_1;\Omega_p) = 4$ and therefore

$$\begin{aligned}\mathbf{x}_2'(t;\Omega_p) &= \mathbf{x}_2'(t_1;\Omega_p) + T'\delta(t-t_1) = \frac{T'}{T}\mathbf{x}_{T,2}(t_1;\Omega_p) + T'\delta(t-t_1)\\ &= \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p) + T'\delta(t-t_1) \le \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p) + T'\delta(t_1'-t_1)\\ &= 4\frac{T'}{T} + \frac{4n\varepsilon}{T(t_2-t_1)}\frac{t_2-t_1}{n} = 4\frac{T'}{T} + 4\frac{\varepsilon}{T} = 4.\end{aligned}$$

At $t = t_1'$ we have equality, $\mathbf{x}_2'(t_1';\Omega_p) = 4 = \frac{T'}{T}\mathbf{x}_{T,2}(t_1';\Omega_p) + 4\frac{\varepsilon}{T}$.

In $[t_1', t_2']$ we have $u'(t) = u_T(t;\Omega_p) = p_l$ and $\mathbf{x}_{T,2}(t;\Omega_p) = 4$. Thus we can rewrite $\mathbf{x}_2'(\cdot;\Omega_p)$ as

$$\mathbf{x}_2'(t;\Omega_p) = 4 = \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p) + 4\frac{\varepsilon}{T} \quad \text{for } t \in [t_1', t_2'].$$

For $t \in [t_2', t_2]$ we compute

$$\begin{aligned}\mathbf{x}_2'(t;\Omega_p) &= \mathbf{x}_2'(t_2';\Omega_p) - T'\delta(t-t_2')\\ &= \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p) + 4\frac{\varepsilon}{T} - T'\delta(t-t_2')\\ &\ge \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p) + 4\frac{\varepsilon}{T} - T'\delta(t_2-t_2')\\ &= \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p) + 4\frac{\varepsilon}{T} - \frac{4n\varepsilon}{T(t_2-t_1)}\frac{t_2-t_1}{n} = \frac{T'}{T}\mathbf{x}_{T,2}(t;\Omega_p),\end{aligned}$$

as $\mathbf{x}_{T,2}(t;\Omega_p) = \mathbf{x}_{T,2}(t_2';\Omega_p)$ in $[t_2', t_2]$. In particular, since $\mathbf{x}_2'(\cdot;\Omega_p)$ is monotonically decreasing in $[t_2', t_2]$ we get $\mathbf{x}_2'(t;\Omega_p) \le \mathbf{x}_2'(t_2';\Omega_p) = 4$ for $t \in [t_2', t_2]$. Furthermore, at $t = t_2'$ we have the equality

$$\mathbf{x}_2'(t_2;\Omega_p) = \frac{T'}{T}\mathbf{x}_{T,2}(t_2;\Omega_p).$$

For $t \in [t_2, 1]$ we have $u'(t) = u_T\left(t;\Omega_p\right)$ and therefore

$$\begin{aligned}\mathbf{x}_2'\left(t;\Omega_p\right) &= \mathbf{x}_2'\left(t_2;\Omega_p\right) + T'\int_{t_2}^{t} u'(\tau) - p_l \,\mathrm{d}\tau \\ &= \frac{T'}{T}\mathbf{x}_{T,2}\left(t_2;\Omega_p\right) + \frac{T'}{T}\left[\mathbf{x}_{T,2}\left(t;\Omega_p\right) - \mathbf{x}_{T,2}\left(t_2;\Omega_p\right)\right] = \frac{T'}{T}\mathbf{x}_{T,2}\left(t;\Omega_p\right).\end{aligned}$$

Altogether, we have

$$\mathbf{x}_2'\left(t;\Omega_p\right) = \frac{T'}{T}\mathbf{x}_{T,2}\left(t;\Omega_p\right) + D(t)$$

with the continuous, non-negative function

$$D(t) = \begin{cases} 0 & \text{for } t \in [0, t_1], \\ T'\delta(t - t_1) & \text{for } t \in (t_1, t_1'], \\ 4\frac{\varepsilon}{T} & \text{for } t \in (t_1', t_2'], \\ 4\frac{\varepsilon}{T} - T'\delta(t - t_2') & \text{for } t \in (t_2', t_2], \\ 0 & \text{for } t \in (t_2, 1]. \end{cases}$$

From what we have seen above we conclude

$$\mathbf{x}_2'\left(t;\Omega_p\right) \le 4 \quad \text{for all } t \in [0,1] \quad \text{and} \quad \mathbf{x}_2'\left(1;\Omega_p\right) = \frac{T'}{T}\mathbf{x}_{T,2}\left(1;\Omega_p\right) = 0.$$

In the remainder of the proof, we investigate $\mathbf{x}_3'\left(1;\Omega_p\right)$. By (B.19) we get

$$t_1' - t_1 = t_2 - t_2' = \frac{1}{n}(t_2 - t_1).$$

Hence,

$$T'\int_0^1 D(t)\,\mathrm{d}t = \frac{1}{2}T'^2\delta\left(t_1' - t_1\right)^2 + 4\frac{T'}{T}\varepsilon\left(t_2 - t_1'\right) - \frac{1}{2}T'^2\delta\left(t_2 - t_2'\right)^2 = 4\frac{T'}{T}\varepsilon\frac{n-1}{n}\left(t_2 - t_1\right).$$

Since $\mathbf{x}_4'\left(t;\Omega_p\right) = \mathbf{x}_2'\left(t;\Omega_p\right) - T'(p_u - p_l)\,t$, we have

$$\mathbf{x}_3'\left(1;\Omega_p\right) = \mathbf{x}_1'\left(1;\Omega_p\right) - \frac{1}{2}T'^2(p_u - p_l),$$

and similarly $\mathbf{x}_{T,3}\left(1;\Omega_p\right) = \mathbf{x}_{T,1}\left(1;\Omega_p\right) - \frac{1}{2}T^2(p_u - p_l)$. Thus,

$$\begin{aligned}
\mathbf{x}_3'\left(1;\Omega_p\right) &= \frac{T'^2}{T^2}\mathbf{x}_{T,1}\left(1;\Omega_p\right) + T'\int_0^1 D(t)\,\mathrm{d}t - \frac{1}{2}T'^2(p_u - p_l) \\
&= \frac{T'^2}{T^2}\mathbf{x}_{T,1}\left(1;\Omega_p\right) + 4\frac{T'}{T}\varepsilon\frac{n-1}{n}\left(t_2 - t_1\right) - \frac{1}{2}T'^2(p_u - p_l) \\
&= \frac{T'^2}{T^2}\left(\mathbf{x}_{T,1}\left(1;\Omega_p\right) - \frac{1}{2}T^2(p_u - p_l)\right) + 4\frac{T'}{T}\varepsilon\frac{n-1}{n}\left(t_2 - t_1\right) \\
&= \frac{T'^2}{T^2}\mathbf{x}_{T,3}\left(1;\Omega_p\right) + 4\frac{T'}{T}\varepsilon\frac{n-1}{n}\left(t_2 - t_1\right) \\
&= 10\frac{(T-\varepsilon)^2}{T^2} + 4\frac{T-\varepsilon}{T}\varepsilon\frac{n-1}{n}\left(t_2 - t_1\right). \\
&= h(\varepsilon)
\end{aligned}$$

by definition of $h(\cdot)$, see (B.18). From the properties of $h(\cdot)$, since $0 < \varepsilon < \varepsilon'$ we get

$$\mathbf{x}_3'\left(1;\Omega_p\right) = h(\varepsilon) > 10\,.$$

To sum up, the tuple $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ is feasible, and we have $\mathbf{x}_3'\left(1;\Omega_p\right) > 10$, $\left\|u'(\cdot) - u_T\left(\cdot;\Omega_p\right)\right\|_\infty < \delta_u$, $|T' - T| < \delta_T$, and $T' < T$, as desired. □

Now, we are able to consider the case of a non-optimal tuple for which the Terminal Constraints (B.10e) and (B.10f) are satisfied with equality:

Proposition B.20
Let $\delta_T, \delta_u > 0$, and $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right) \neq \left(T^*, u^*(\cdot), \mathbf{x}^*\left(\cdot;\Omega_p\right)\right)$ be feasible for Problem (B.10) with

$$\mathbf{x}_3\left(1;\Omega_p\right) = 10 \quad \text{and} \quad \mathbf{x}_2\left(1;\Omega_p\right) = 0\,.$$

Then there is a feasible $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ with $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$, $|T' - T| < \delta_T$, $T' \leq T$, and $\mathbf{x}_3'\left(1;\Omega_p\right) > 10$.

Proof Since the global optimum of Problem (B.10) is unique (see Corollary B.15) we have $T > T^*$. Let $t_1 = \frac{4}{T(10-p_l)}$ and $t_2 = 1 - \frac{4}{T(10+p_l)}$. We make a case distinction as in the proof of Proposition B.8.

Case 1): $u(\cdot) \not\equiv 10$ on $[0, t_1]$ (almost surely). Similar to Case 1) in the proof of Proposition B.8, there is a control function $u'(\cdot)$ with $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$ such that $\left(T, u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ is feasible, where $\mathbf{x}'\left(\cdot;\Omega_p\right)$ denotes the differential states which are determined by T, $u'(\cdot)$, and Ω_p. We have $\mathbf{x}_2\left(t;\Omega_p\right) \leq \mathbf{x}_2'\left(t;\Omega_p\right)$ for all $t \in [0,1]$ and $\mathbf{x}_2\left(t';\Omega_p\right) < \mathbf{x}_2'\left(t';\Omega_p\right)$ for some $t' \in [0,1]$. Due to $\mathbf{x}_2\left(t;\Omega_p\right) = \mathbf{x}_4\left(t;\Omega_p\right) + (p_u - p_l)\,T\,t$

(and similarly for $\mathbf{x}'\left(\cdot;\Omega_p\right)$) and $\mathbf{x}_3\left(1;\Omega_p\right) = 10$ we get $\mathbf{x}_3'\left(1;\Omega_p\right) > 10$.

Case 2): $u(\cdot) \equiv 10$ on $[0, t_1]$ (almost surely). If we have $\mathbf{x}_2\left(t;\Omega_p\right) = 4$ for $t \in [t_1, t_2]$, then $u(\cdot) \equiv p_l$ on $[t_1, t_2]$ (almost surely) and furthermore $u(\cdot) \equiv -10$ on $[t_2, 1]$ (almost surely) due to the terminal constraint $\mathbf{x}_2\left(1;\Omega_p\right) \leq 0$. In particular, $u(\cdot) = u_T\left(\cdot;\Omega_p\right)$. We can apply Proposition B.19 to achieve the targeted result. For the remainder, we assume that there is a $t \in [t_1, t_2]$ with $\mathbf{x}_2\left(t;\Omega_p\right) < 4$. We can proceed similar to Case 2) in the proof of Proposition B.8 to show that there is a control function $u'(\cdot)$ with $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$ such that $\left(T, u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ is feasible, if $\mathbf{x}'\left(\cdot;\Omega_p\right)$ denotes the differential states which are determined by T, $u'(\cdot)$, and Ω_p. We have $\mathbf{x}_2\left(t;\Omega_p\right) \leq \mathbf{x}_2'\left(t;\Omega_p\right)$ for all $t \in [0,1]$ and $\mathbf{x}_2\left(t';\Omega_p\right) < \mathbf{x}_2'\left(t';\Omega_p\right)$ for some t'. This implies $\mathbf{x}_3'\left(1;\Omega_p\right) > 10$ as in Case 1), see above. □

Next, we show that Problem (B.10) has exactly one local minimum.

Proposition B.21
Let the feasible set of Problem (B.10) be non-empty. The tuple $\left(T^*, u^*(\cdot), \mathbf{x}^*\left(\cdot;\Omega_p\right)\right)$ is the only local optimum of Problem (B.10) in the considered normed space.

Proof The proof works similar to the proof of Proposition B.9. We know that the tuple $\left(T^*, u^*(\cdot), \mathbf{x}^*\left(\cdot;\Omega_p\right)\right)$ is the unique global minimum of Problem (6.9), see Corollary B.15. Let $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right) \neq \left(T^*, u^*(\cdot), \mathbf{x}^*\left(\cdot;\Omega_p\right)\right)$ be feasible for Problem (6.9). Since the global minimum is unique, we have $T > T^*$. Let $\delta_T, \delta_u > 0$. We distinct two cases:

Case 1): $\mathbf{x}_3\left(1;\Omega_p\right) = 10$. Depending on whether Constraint (B.10f) is satisfied with equality or not, we can apply Lemma B.16 or Proposition B.20 to get a feasible tuple $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ with $\|u'(\cdot) - u(\cdot)\|_\infty < \delta_u$, $|T' - T| < \frac{\delta_T}{2}$, $T' \leq T$, and $\mathbf{x}_3'\left(1;\Omega_p\right) > 10$. Subsequently, we apply Lemma B.17 and find a feasible $\left(T'', u''(\cdot), \mathbf{x}''\left(\cdot;\Omega_p\right)\right)$ with $u''(\cdot) = u'(\cdot)$, $|T'' - T| \leq |T'' - T'| + |T' - T| < \delta_T$, and $T'' < T' \leq T$.

Case 2): $\mathbf{x}_3\left(1;\Omega_p\right) > 10$. In this case, we can directly apply Lemma B.17 and find a feasible $\left(T', u'(\cdot), \mathbf{x}'\left(\cdot;\Omega_p\right)\right)$ with $u'(\cdot) = u(\cdot)$, $|T' - T| < \delta_T$, and $T' < T$.

Using Lemma B.10 we conclude that $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ is not a local minimum. □

It remains to transfer the established results for Problem (B.10) to the original Problem (6.10). We get

Corollary B.22
Let the feasible set of Problem (B.10) be non-empty. Let T^* and $u^*(\cdot)$ be the globally optimal controllable parameter and control function of Problem (B.10). For given T^*, $u^*(\cdot)$, and any $p \in \Omega_p$, we denote the solution of the resulting IVP (6.10b-6.10c) by $\mathbf{x}^*(\cdot;p)$. Then the global solutions of Problem (6.10) are given by

$$\left\{\left(T^*, u^*(\cdot), p, \mathbf{x}^*(\cdot;p)\right) \,\middle|\, p \in \Omega_p\right\}.$$

Furthermore, if $\left(T, u(\cdot), p, \mathbf{x}(\cdot;p)\right)$ is a local solution of Problem (6.10) in the considered normed space, then $T = T^*$ and $u(\cdot) = u^*(\cdot)$. In particular, every local solution is a global solution.

Proof Due to Corollary B.15 and Lemma B.11 the global solutions of Problem (6.10) are given by

$$\left\{\left(T^*, u^*(\cdot), p, \mathbf{x}^*(\cdot;p)\right) \,\middle|\, p \in \Omega_p\right\}.$$

If $\left(T, u(\cdot), p, \mathbf{x}(\cdot;p)\right)$ is a local solution of Problem (6.10), again by Lemma B.11 we know that $\left(T, u(\cdot), \mathbf{x}\left(\cdot;\Omega_p\right)\right)$ is a local solution of Problem (B.10). From Proposition B.21 we get $T = T^*$ and $u(\cdot) = u^*(\cdot)$, as claimed. □

Bibliography

[1] W. Achtziger and C. Kanzow. Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications. *Mathematical Programming*, 114(1):69–99, 2008.

[2] A. Agarwal and I. Verma. Cerebral palsy in children: An overview. *Journal of Clinical Orthopaedics and Trauma*, 3(2):77–81, 2012.

[3] J. Albersmeyer. *Adjoint-based algorithms and numerical methods for sensitivity generation and optimization of large scale dynamic systems*. PhD thesis, Heidelberg University, 2010.

[4] S. Albrecht, K. Ramírez-Amaro, F. Ruiz-Ugalde, D. Weikersdorfer, M. Leibold, M. Ulbrich, and M. Beetz. Imitating human reaching motions using physically inspired optimization principles. In *2011 11th IEEE-RAS International Conference on Humanoid Robots*, pages 602–607, 2011.

[5] D. E. Anderson, M. L. Madigan, and M. A. Nussbaum. Maximum voluntary joint torque as a function of joint angle and angular velocity: Model development and application to the lower limb. *Journal of Biomechanics*, 40(14): 3105–3113, 2007.

[6] T. Apgar, P. Clary, K. Green, A. Fern, and J. Hurst. Fast Online Trajectory Optimization for the Bipedal Robot Cassie. In *Proceedings of Robotics: Science and Systems XIV*, 2018.

[7] S. Armand, G. Decoulon, and A. Bonnefoy-Mazure. Gait analysis in children with cerebral palsy. *EFORT Open Reviews*, 1(12):448–460, 2016.

[8] R. Baker, J. L. McGinley, M. H. Schwartz, S. Beynon, A. Rozumalski, H. K. Graham, and O. Tirosh. The Gait Profile Score and Movement Analysis Profile. *Gait & Posture*, 30(3):265–269, 2009.

[9] J. U. Baumann and H. G. Koch. Ventrale aponeurotische Verlängerung des Musculus gastrocnemius. *Operative Orthopädie und Traumatologie*, 1(4):254–258, 1989.

[10] A. Ben-Tal and A. Nemirovski. Robust Truss Topology Design via Semidefinite Programming. *SIAM Journal on Optimization*, 7(4):991–1016, 1997.

[11] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[12] S. C. Bengea and R. A. DeCarlo. Optimal control of switching systems. *Automatica*, 41(1):11–27, 2005.

[13] D. A. Benson, G. T. Huntington, T. P. Thorvaldsen, and A. V. Rao. Direct Trajectory Optimization and Costate Estimation via an Orthogonal Collocation Method. *Journal of Guidance, Control, and Dynamics*, 29(6):1435–1440, 2006.

[14] D. Bertsimas and A. Thiele. A Robust Optimization Approach to Supply Chain Management. In D. Bienstock and G. Nemhauser, editors, *Integer Programming and Combinatorial Optimization: 10th International IPCO Conference*, pages 86–100. Springer, 2004.

[15] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and Applications of Robust Optimization. *SIAM Review*, 53(3):464–501, 2011.

[16] F. Bestehorn, C. Hansknecht, C. Kirches, and P. Manns. A switching cost aware rounding method for relaxations of mixed-integer optimal control problems. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 7134–7139, 2019.

[17] F. Bestehorn, C. Hansknecht, C. Kirches, and P. Manns. Mixed-integer optimal control problems with switching costs: a shortest path approach. *Mathematical Programming*, 188(2):621–652, 2021.

[18] J. T. Betts. *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. SIAM, 2nd edition, 2010.

[19] H.-G. Beyer and B. Sendhoff. Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering*, 196(33-34):3190–3218, 2007.

[20] L. T. Biegler. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers & Chemical Engineering*, 8(3-4):243–247, 1984.

[21] H. G. Bock. Numerical Solution of Nonlinear Multipoint Boundary Value Problems with Applications to Optimal Control. *Zeitschrift für Angewandte Mathematik und Mechanik*, 58(7):T407–T409, 1978.

[22] H. G. Bock. Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics. In K. H. Ebert, P. Deuflhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, pages 102–125. Springer, 1981.

[23] H. G. Bock. Recent Advances in Parameteridentification Techniques for O.D.E. In P. Deuflhard and E. Hairer, editors, *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, pages 95–121. Birkhäuser, 1983.

[24] H. G. Bock. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen*, volume 183 of *Bonner Mathematische Schriften*. Rheinische Friedrich–Wilhelms–Universität Bonn, 1987.

[25] H. G. Bock and K. J. Plitt. A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems. In *9th IFAC World Congress: A Bridge Between Control Science and Technology*, pages 1603–1608, 1984.

[26] H. G. Bock, C. Kirches, A. Meyer, and A. Potschka. Numerical solution of optimal control problems with explicit and implicit switches. *Optimization Methods & Software*, 33(3):450–474, 2018.

[27] U. Brandt-Pollmann. *Numerical solution of optimal control problems with implicitly defined discontinuities with applications in engineering*. PhD thesis, Heidelberg University, 2004.

[28] C. Büskens and H. Maurer. SQP-methods for solving optimal control problems with control and state constraints: adjoint variables, sensitivity analysis and real-time control. *Journal of Computational and Applied Mathematics*, 120 (1-2):85–108, 2000.

[29] J. Carpentier, G. Saurel, G. Buondonno, J. Mirabel, F. Lamiraux, O. Stasse, and N. Mansard. The Pinocchio C++ library: A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives. In *2019 IEEE/SICE International Symposium on System Integration (SII)*, pages 614–619, 2019.

[30] F. M. Chang, A. J. Seidl, K. Muthusamy, A. K. Meininger, and J. J. Carollo. Effectiveness of Instrumented Gait Analysis in Children With Cerebral Palsy –

Comparison of Outcomes. *Journal of Pediatric Orthopaedics*, 26(5):612–616, 2006.

[31] D. Clever and K. D. Mombaur. An Inverse Optimal Control Approach for the Transfer of Human Walking Motions in Constrained Environment to Humanoid Robots. In *Proceedings of Robotics: Science and Systems XII*, 2016.

[32] D. Clever, R. M. Schemschat, M. L. Felis, and K. Mombaur. Inverse Optimal Control Based Identification of Optimality Criteria in Whole-Body Human Walking on Level Ground. In *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 1192–1199, 2016.

[33] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153:235–256, 2007.

[34] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization*. SIAM, 2009.

[35] P. de Leva. Adjustments to Zatsiorsky-Seluyanov's segment inertia parameters. *Journal of Biomechanics*, 29(9):1223–1230, 1996.

[36] M. C. de Morais Filho, R. Yoshida, W. da Silva Carvalho, H. E. Stein, and N. F. Novo. Are the recommendations from three-dimensional gait analysis associated with better postoperative outcomes in patients with cerebral palsy? *Gait & Posture*, 28(2):316–322, 2008.

[37] S. L. Delp. Publications of Scott L. Delp. Stanford University, Department of Bioengeneering. URL `https://profiles.stanford.edu/scott-delp?tab=publications`. Accessed: 2020-08-31.

[38] S. Dempe. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, 2002.

[39] S. Dempe. Bilevel Optimization: Theory, Algorithms, Applications and a Bibliography. In S. Dempe and A. Zemhoko, editors, *Bilevel Optimization: Advances and Next Challenges*, pages 581–672. Springer, 2020.

[40] M. Diehl, H. G. Bock, and E. Kostina. An approximation technique for robust nonlinear optimization. *Mathematical Programming*, 107(1-2):213–230, 2006.

[41] X. C. Ding, Y. Wardi, D. Taylor, and M. Egerstedt. Optimization of Switched-Mode Systems with Switching Costs. In *Proceedings of the 2008 American Control Conference*, pages 3965–3970, 2008.

[42] A. V. Dmitruk and A. M. Kaganovich. The Hybrid Maximum Principle is a consequence of Pontryagin Maximum Principle. *Systems & Control Letters*, 57(11): 964–970, 2008.

[43] L. Döderlein. *Infantile Zerebralparese: Diagnostik, konservative und operative Therapie*. Springer, 2nd edition, 2015.

[44] T. Dreher, T. Buccoliero, S. I. Wolf, D. Heitzmann, S. Gantz, F. Braatz, and W. Wenz. Long-Term Results After Gastrocnemius-Soleus Intramuscular Aponeurotic Recession as a Part of Multilevel Surgery in Spastic Diplegic Cerebral Palsy. *The Journal of Bone & Joint Surgery*, 94(7):627–637, 2012.

[45] T. Dreher, S. I. Wolf, M. Maier, S. Hagmann, D. Vegvari, S. Gantz, D. Heitzmann, W. Wenz, and F. Braatz. Long-Term Results After Distal Rectus Femoris Transfer as a Part of Multilevel Surgery for the Correction of Stiff-Knee Gait in Spastic Diplegic Cerebral Palsy. *The Journal of Bone & Joint Surgery*, 94(19):e142, 2012.

[46] T. Dreher, F. Braatz, S. I. Wolf, V. Ewerbeck, D. Heitzmann, W. Wenz, and L. Döderlein. Distal Rectus Femoris Tendon Transfer for the Correction of Stiff-Knee Gait in Cerebral Palsy. *JBJS Essential Surgical Techniques*, 3(1), 2013.

[47] R. Featherstone. *Rigid Body Dynamics Algorithms*. Springer, 2008.

[48] M. L. Felis. *Modeling Emotional Aspects in Human Locomotion*. PhD thesis, Heidelberg University, 2015.

[49] M. L. Felis. RBDL: an efficient rigid-body dynamics library using recursive algorithms. *Autonomous Robots*, 41(2):495–511, 2017.

[50] M. L. Felis and K. Mombaur. Using Optimal Control Methods to Generate Human Walking Motions. In *Proceedings of the 5th International Conference on Motion in Games*, pages 197–207, 2012.

[51] M. L. Felis, K. Mombaur, and A. Berthoz. An Optimal Control Approach to Reconstruct Human Gait Dynamics from Kinematic Data. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pages 1044–1051, 2015.

[52] F. Fisch. *Development of a Framework for the Solution of High-Fidelity Trajectory Optimization Problems and Bilevel Optimal Control Problems*. PhD thesis, Technical University of Munich, 2011.

[53] F. Fisch, J. Lenz, F. Holzapfel, and G. Sachs. On the solution of bilevel optimal control problems to increase the fairness in air races. *Journal of Guidance, Control, and Dynamics*, 35(4):1292–1298, 2012.

[54] M. D. Fox, J. A. Reinbolt, S. Õunpuu, and S. L. Delp. Mechanisms of improved knee flexion after rectus femoris transfer surgery. *Journal of Biomechanics*, 42 (5):614–619, 2009.

[55] A. Fredriksson, A. Forsgren, and B. Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical Physics*, 38(3): 1672–1684, 2011.

[56] O. A. Galarraga C., V. Vigneron, B. Dorizzi, N. Khouri, and E. Desailly. Predicting postoperative gait in cerebral palsy. *Gait & Posture*, 52:45–51, 2017.

[57] M. Garavello and B. Piccoli. Hybrid Necessary Principle. *SIAM Journal on Control and Optimization*, 43(5):1867–1887, 2005.

[58] M. Garcia, A. Chatterjee, A. Ruina, and M. Coleman. The Simplest Walking Model: Stability, Complexity, and Scaling. *Journal of Biomechanical Engineering*, 120(2):281–288, 1998.

[59] C. Geiger and C. Kanzow. *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer, 2002.

[60] M. Gerdts. *Optimal control of ODEs and DAEs*. De Gruyter, 2011.

[61] P. E. Gill, W. Murray, and M. A. Saunders. SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization. *SIAM Review*, 47(1):99–131, 2005.

[62] A. Giua, C. Seatzu, and C. Van Der Mee. Optimal control of autonomous linear systems switched with a pre-assigned finite sequence. In *Proceeding of the 2001 IEEE International Symposium on Intelligent Control (ISIC 01)*, pages 144–149, 2001.

[63] R. Goebel, R. G. Sanfelice, and A. R. Teel. Hybrid Dynamical Systems. *IEEE control systems magazine*, 29(2):28–93, 2009.

[64] H. Gonzalez, R. Vasudevan, M. Kamgarpour, S. S. Sastry, R. Bajcsy, and C. Tomlin. A Numerical Method for the Optimal Control of Switched Systems. In *49th IEEE Conference on Decision and Control (CDC)*, pages 7519–7526, 2010.

[65] H. Gonzalez, R. Vasudevan, M. Kamgarpour, S. S. Sastry, R. Bajcsy, and C. J. Tomlin. A Descent Algorithm for the Optimal Control of Constrained Nonlinear Switched Dynamical Systems. In *Proceedings of the 13th ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, pages 51–60, 2010.

[66] B. L. Gorissen, İ. Yanıkoğlu, and D. den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.

[67] S. Gulati and V. Sondhi. Cerebral Palsy: An Overview. *Indian Journal of Pediatrics*, 85(11):1006–1016, 2018.

[68] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer, 2nd edition, 1996.

[69] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer, 2nd edition, 1993.

[70] C. R. Hargraves and S. W. Paris. Direct Trajectory Optimization Using Nonlinear Programming and Collocation. *Journal of Guidance, Control, and Dynamics*, 10(4):338–342, 1987.

[71] K. Hatz. *Efficient numerical methods for hierarchical dynamic optimization with application to cerebral palsy gait modeling*. PhD thesis, Heidelberg University, 2014.

[72] P. D. Hoang, R. B. Gorman, G. Todd, S. C. Gandevia, and R. D. Herbert. A new method for measuring passive length–tension properties of human gastrocnemius muscle in vivo. *Journal of Biomechanics*, 38(6):1333–1341, 2005.

[73] T. Hoheisel. *Mathematical Programs with Vanishing Constraints*. PhD thesis, University of Würzburg, 2009.

[74] T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Mathematical Programming*, 137(1-2):257–288, 2013.

[75] B. Houska. *Robust Optimization of Dynamic Systems*. PhD thesis, Katholieke Universiteit Leuven, 2011.

[76] Y. Hu. *The role of compliance in humans and humanoid robots locomotion*. PhD thesis, Heidelberg University, 2017.

[77] Y. Hu and K. Mombaur. Analysis of human leg joints compliance in different walking scenarios with an optimal control approach. *IFAC-PapersOnLine*, 49 (14):99–106, 2016.

[78] S. G. Johnson. The NLopt nonlinear-optimization package. URL `https://github.com/stevengj/nlopt`.

[79] V. V. Kalashnikov, S. Dempe, G. A. Pérez-Valdés, N. I. Kalashnykova, and J.-F. Camacho-Vallejo. Bilevel Programming and Applications. *Mathematical Problems in Engineering*, 2015, 2015.

[80] C. Kirches. A Numerical Method for Nonlinear Robust Optimal Control with Implicit Discontinuities and an Application to Powertrain Oscillations. Diploma Thesis, Heidelberg University, 2006.

[81] C. Kirches. *Fast Numerical Methods for Mixed–Integer Nonlinear Model–Predictive Control*. PhD thesis, Heidelberg University, 2010.

[82] C. Kirches, F. Lenders, and P. Manns. Approximation Properties and Tight Bounds for Constrained Mixed-Integer Optimal Control. *SIAM Journal on Control and Optimization*, 58(3):1371–1402, 2020.

[83] D. E. Kirk. *Optimal Control Theory: An Introduction*. Prentice-Hall, 1970.

[84] M. Knauer. *Bilevel-Optimalsteuerung mittels hybrider Lösungsmethoden am Beispiel eines deckengeführten Regalbediengerätes in einem Hochregallager*. PhD thesis, University of Bremen, 2009.

[85] M. Knauer. Fast and save container cranes as bilevel optimal control problems. *Mathematical and Computer Modelling of Dynamical Systems*, 18(4):465–486, 2012.

[86] R. M. Kopitzsch. *Analysis of Human Push Recovery Motions Based on Optimization*. PhD thesis, Heidelberg University, 2020.

[87] O. Kostyukova and E. Kostina. Robust optimal feedback for terminal linear-quadratic control problems under disturbances. *Mathematical Programming*, 107(1-2):131–153, 2006.

[88] I. Krägeloh-Mann. Klassifikation, Epidemiologie, Pathogenese und Klinik. In F. Heinen and B. Werner, editors, *Das Kind und die Spastik: Erkenntnisse der Evidenced-based Medicine zur Cerebralparese*, pages 37–48. Verlag Hans Huber, 2001.

[89] P. Krämer-Eis. *Ein Mehrzielverfahren Zur Numerischen Berechnung Optimaler Feedback-Steuerungen Bei Beschränkten Nichtlinearen Problemen*, volume 166 of *Bonner Mathematische Schriften*. Rheinische Friedrich–Wilhelms–Universität Bonn, 1985.

[90] K. W. Krigger. Cerebral Palsy: An Overview. *American Family Physician*, 73(1): 91–100, 2006.

[91] A. Kufner, O. John, and S. Fučik. *Function Spaces*. Noordhoff International Publishing, 1977.

[92] S. Lee, M. Park, K. Lee, and J. Lee. Scalable Muscle-Actuated Human Simulation and Control. *ACM Transactions on Graphics (TOG)*, 38(4):Article No. 73, 1–13, 2019.

[93] D. B. Leineweber. *Efficient Reduced SQP Methods for the Optimization of Chemical Processes Described by Large Sparse DAE models*. PhD thesis, Heidelberg University, 1999.

[94] D. B. Leineweber, I. Bauer, H. G. Bock, and J. P. Schlöder. An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization. Part 1: theoretical aspects. *Computers & Chemical Engineering*, 27 (1):157–166, 2003.

[95] D. B. Leineweber, A. Schäfer, H. G. Bock, and J. P. Schlöder. An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization: Part II: Software aspects and applications. *Computers & Chemical Engineering*, 27(2):167–174, 2003.

[96] F. Lenders. *Numerical Methods for Mixed-Integer Optimal Control with Combinatorial Constraints*. PhD thesis, Heidelberg University, 2018.

[97] D. Liberzon. *Switching in Systems and Control*. Birkhäuser, 2003.

[98] R. C. Loxton, K. L. Teo, and V. Rehbock. Computational Method for a Class of Switched System Optimal Control Problems. *IEEE Transactions on Automatic Control*, 54(10):2455–2460, 2009.

[99] Z. Manchester and S. Kuindersma. Variational Contact-Implicit Trajectory Optimization. In N. Amato, G. Hager, S. Thomas, and M. Torres-Torriti, editors, *Robotics Research: The 18th International Symposium ISRR*, pages 985–1000. Springer, 2020.

[100] MATLAB®. *Version 9.3 (R2017b)*. The MathWorks Inc., Natick, Massachusetts, USA, 2017.

[101] P. Mehlitz. *Contributions to complementarity and bilevel programming in Banach spaces*. PhD thesis, TU Bergakademie Freiberg, 2017.

[102] A. Meyer. *Numerical Solution of Optimal Control Problems with Explicit and Implicit Switches*. PhD thesis, Heidelberg University, 2020.

[103] M. Millard, T. Uchida, A. Seth, and S. L. Delp. Flexing Computational Muscle: Modeling and Simulation of Musculotendon Dynamics. *Journal of Biomechanical Engineering*, 135(2):021005, 2013.

[104] M. Millard, M. Sreenivasa, and K. Mombaur. Predicting the Motions and Forces of Wearable Robotic systems Using optimal Control. *Frontiers in Robotics and AI*, 4(41), 2017.

[105] K. Mombaur. Optimal Control for Applications in Medical and Rehabilitation Technology: Challenges and Solutions. In J.-B. Hiriart-Urruty, A. Korytowski, H. Maurer, and M. Szymkat, editors, *Advances in Mathematical Modeling, Optimization and Optimal Control*, pages 103–145. Springer, 2016.

[106] K. Mombaur, A. Truong, and J.-P. Laumond. From human to humanoid locomotion – an inverse optimal control approach. *Autonomous Robots*, 28(3): 369–383, 2010.

[107] K. Mombaur, A.-H. Olivier, and A. Crétual. Forward and Inverse Optimal Control of Bipedal Running. In K. Mombaur and K. Berns, editors, *Modeling, Simulation and Optimization of Bipedal Walking*, pages 165–179. Springer, 2013.

[108] I. Mordatch, Z. Popović, and E. Todorov. Contact-Invariant Optimization for Hand Manipulation. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA)*, pages 137–144, 2012.

[109] I. P. Natanson. *Theorie der Funktionen einer reellen Veränderlichen*. Verlag Harri Deutsch, 4th edition, 1975.

[110] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2nd edition, 2006.

[111] I. Novak, M. Hines, S. Goldsmith, and R. Barclay. Clinical Prognostic Messages From a Systematic Review on Cerebral Palsy. *Pediatrics*, 130(5):e1285–e1312, 2012.

[112] I. Novak, S. McIntyre, C. Morgan, L. Campbell, L. Dark, N. Morton, E. Stumbles, S.-A. Wilson, and S. Goldsmith. A systematic review of interventions for children with cerebral palsy: state of the evidence. *Developmental Medicine & Child Neurology*, 55(10):885–910, 2013.

[113] Y. Nubar and R. Contini. A minimal principle in biomechanics. *Bulletin of Mathematical Biophysics*, 23(4):377–391, 1961.

[114] M. Oskoui, F. Coutinho, J. Dykeman, N. Jetté, and T. Pringsheim. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Developmental Medicine & Child Neurology*, 55(6):509–519, 2013.

[115] J. V. Outrata. On the Numerical Solution of a Class of Stackelberg Problems. *Mathematical Methods of Operations Research*, 34(4):255–277, 1990.

[116] K. Palagachev and M. Gerdts. Exploitation of the Value Function in a Bilevel Optimal Control Problem. In L. Bociu, J.-A. Désidéri, and A. Habbal, editors, *System Modeling and Optimization: 27th IFIP TC 7 Conference, CSMO 2015*, pages 410–419, 2016.

[117] R. Palisano, P. Rosenbaum, S. Walter, D. Russell, E. Wood, and B. Galuppi. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Developmental Medicine & Child Neurology*, 39(4): 214–223, 1997.

[118] S. J. Pinney, B. J. Sangeorzan, and S. T. Hansen Jr. Surgical Anatomy of the Gastrocnemius Recession (Strayer Procedure). *Foot & Ankle International*, 25 (4):247–250, 2004.

[119] L. Pitto, H. Kainz, A. Falisse, M. Wesseling, S. Van Rossom, H. Hoang, E. Papageorgiou, A. Hallemans, K. Desloovere, G. Molenaers, A. Van Campenhout, F. De Groote, and I. Jonkers. *SimCP*: A Simulation Platform to Predict Gait Performance Following Orthopedic Intervention in Children With Cerebral Palsy. *Frontiers in Neurorobotics*, 13(54), 2019.

[120] K. J. Plitt. Ein Superlinear Konvergentes Mehrzielverfahren Zur Direkten Berechnung Beschränkter Optimaler Steuerungen. Diploma thesis, University of Bonn, 1981.

[121] M. Posa, C. Cantu, and R. Tedrake. A direct method for trajectory optimization of rigid bodies through contact. *The International Journal of Robotics Research*, 33(1):69–81, 2014.

[122] A. Potschka. Handling Path Constraints in a Direct Multiple Shooting Method for Optimal Control Problems. Diploma thesis, Heidelberg University, 2006.

[123] A. Potschka, H. G. Bock, and J. P. Schlöder. A minima tracking variant of semi-infinite programming for the treatment of path constraints within direct solution of optimal control problems. *Optimization Methods & Software*, 24(2): 237–252, 2009.

[124] M. J. D. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. Technical Report DAMTP 2009/NA06, Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, 2009.

[125] L. M. Rios and N. V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293, 2013.

[126] P. Rosenbaum, N. Paneth, A. Leviton, M. Goldstein, M. Bax, D. Damiano, B. Dan, and B. Jacobsson. A report: the definition and classification of cerebral palsy April 2006. *Developmental Medicine & Child Neurology*, 49(s109):8–14, 2007.

[127] S. Sager. *Numerical methods for mixed–integer optimal control problems*. PhD thesis, Heidelberg University, 2006.

[128] S. Sager and C. Zeile. On mixed-integer optimal control with constrained total variation of the integer control. *Computational Optimization and Applications*, 78(2):575–623, 2021.

[129] S. Sager, H. G. Bock, and G. Reinelt. Direct methods with maximal lower bound for mixed-integer optimal control problems. *Mathematical Programming*, 118 (1):109–149, 2009.

[130] M. Sangeux and S. Armand. Kinematic Deviations in Children with Cerebral Palsy. In F. Canavese and J. Deslandes, editors, *Orthopedic Management of Children with Cerebral Palsy: A Comprehensive Approach*, pages 241–253. Nova Science Publishers, 2015.

[131] M. Sartori, J. W. Fernandez, L. Modenese, C. P. Carty, L. A. Barber, K. Oberhofer, J. Zhang, G. G. Handsfield, N. S. Stott, T. F. Besier, D. Farina, and D. G. Lloyd. Toward modeling locomotion using electromyography-informed 3D models:

application to cerebral palsy. *WIREs Systems Biology and Medicine*, 9(2):e1368, 2017.

[132] H. Scheel and S. Scholtes. Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity. *Mathematics of Operations Research*, 25(1):1–22, 2000.

[133] M. Schlöder, C. Kirches, E. Kostina, and A. Meyer. Numerical Solution of Optimal Control Problems with Switches, Switching Costs and Jumps. Optimization Online Preprint 6888, 2018. URL `http://www.optimization-online.org/DB_FILE/2018/10/6888.pdf`.

[134] M. Schlöder, C. Kirches, E. Kostina, and A. Meyer. Generation of Optimal Walking-Like Motions Using Dynamic Models with Switches, Switch Costs, and State Jumps. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1538–1543, 2019.

[135] S. Scholtes. Convergence Properties of a Regularization Scheme for Mathematical Programs with Complementarity Constraints. *SIAM Journal on Optimization*, 11(4):918–936, 2001.

[136] A. Schwartz. Mathematical Programs with Complementarity Constraints and Related Problems. Course Notes, Graduate School CE, Technical University of Darmstadt, 2018.

[137] P. O. M. Scokaert and D. Q. Mayne. Min-max feedback model predictive control for constrained linear systems. *IEEE Transactions on Automatic control*, 43(8):1136–1142, 1998.

[138] A. A. Shabana. *Dynamics of Multibody Systems*. Cambridge University Press, 3rd edition, 2005.

[139] A. Sinha, P. Malo, and K. Deb. A Review on Bilevel Optimization: From Classicalto Evolutionary Approaches and Applications. *IEEE Transactions on Evolutionary Computation*, 22(2):276–295, 2018.

[140] H. I. Stibbe. *Special Bilevel Quadratic Problems for Construction of Worst-Case Feedback Control in Linear-Quadratic Optimal Control Problems under Uncertainties*. PhD thesis, University of Marburg, 2019.

[141] J. Stoer and R. Bulirsch. *Numerische Mathematik 2*. Springer, 5th edition, 2005.

[142] H. J. Sussmann. A maximum principle for hybrid optimal control problems. In *Proceedings of the 38th IEEE Conference on Decision and Control (CDC)*, pages 425–430, 1999.

[143] E. Todorov. Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9):907–915, 2004.

[144] M. Ulbrich and S. Ulbrich. *Nichtlineare Optimierung*. Birkhäuser Science, 2012.

[145] A. J. van der Schaft and H. Schumacher. *An Introduction to Hybrid Dynamical Systems*. Springer, 2000.

[146] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[147] A. Wächter et al. Ipopt Documentation. URL `https://coin-or.github.io/Ipopt/index.html`. Accessed: 2021-09-03.

[148] A. Walther and A. Griewank. Getting Started with ADOL-C. In U. Naumann and O. Schenk, editors, *Combinatorial Scientific Computing*, pages 181–202. Chapman & Hall/CRC, 2012.

[149] W. Wenz and L. Döderlein. Die Verpflanzung der Sehne des Musculus rectus femoris bei Patienten mit spastischer Diparese. *Operative Orthopädie und Traumatologie*, 11(3):213–222, 1999.

[150] C. Woernle. *Mehrkörpersysteme: Eine Einführung in die Kinematik und Dynamik von Systemen starrer Körper*. Springer, 2nd edition, 2016.

[151] S. I. Wolf. Heidelberg MotionLab. Heidelberg University Hospital, Department of Orthopaedics and Trauma Surgery. URL `https://www.heidel-motionlab.de/`. Accessed: 2021-09-03.

[152] C. Wu and K. Teo. Global impulsive optimal control computation. *Journal of Industrial & Management Optimization*, 2(4):435–450, 2006.

[153] X. Xu and P. J. Antsaklis. Optimal Control of Switched Systems: New Results and Open Problems. In *Proceedings of the 2000 American Control Conference*, pages 2683–2687, 2000.

[154] X. Xu and P. J. Antsaklis. Optimal Control of Switched Systems Based on Parameterization of the Switching Instants. *IEEE Transactions on Automatic Control*, 49(1):2–16, 2004.

[155] Y.-S. Yoon and J. Mansour. The passive elastic moment at the hip. *Journal of Biomechanics*, 15(12):905–910, 1982.

[156] F. Zhu and P. J. Antsaklis. Optimal control of hybrid switched systems: A brief survey. *Discrete Event Dynamic Systems*, 25(3):345–364, 2015.

List of Acronyms

AD Automatic Differentiation

CQ Constraint Qualification

CP Cerebral Palsy

DAE Differential Algebraic Equation

DFO Derivative-Free Optimization

DoF Degree of Freedom

GA Gait Analysis

GCQ Guignard Constraint Qualification

GMFCS Gross Motor Function Classification System

IND Internal Numerical Differentiation

IVP Initial Value Problem

IOC Inverse Optimal Control

KKT Karush-Kuhn-Tucker

LGR Legendre-Gauss-Radau

LICQ Linear Independence Constraint Qualification

MBS Multi-Body System

MFCQ Mangasarian-Fromovitz Constraint Qualification

MIOCP Mixed-Integer Optimal Control Problem

MPCC Mathematical Program with Complementarity Constraints

MPVC Mathematical Program with Vanishing Constraints

NLP Nonlinear Programming Problem

OCP Optimal Control Problem

ODE Ordinary Differential Equation

SQP Sequential Quadratic Programming

VC Vanishing Constraint

POC Partial Outer Convexification

Nomenclature – Selected Symbols

Throughout this thesis, vectors and vector valued functions are denoted by bold symbols. Vectors are represented as column vectors. Functions and other dependencies are denoted by appending the symbol $(\cdot)$ to the function or variable designator.

General symbols

$\subseteq$	Subset
$\subset$	Strict subset
$\triangle$	End of definition or assumption
$\triangle$	End of definition or assumption
$\square$	End of proof
$\mathbf{0}$	Vector of zeros of appropriate (context-dependent) size
$\mathcal{T}$	Time horizon
$\lvert\mathcal{M}\rvert$	Cardinality of a set $\mathcal{M}$
$\nabla_{\mathbf{x}} g(\mathbf{x})$	Gradient of a scalar valued function $g(\cdot)$ w. r. t. $\mathbf{x}$, represented as column vector. If the context is clear, the subscript can be omitted.
$\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x})$	Transposed of Jacobian of a (vector valued) function $\mathbf{f}(\cdot)$ (w. r. t. $\mathbf{x}$)
$\nabla^2_{\mathbf{xx}} g(\mathbf{x})$	Hessian of a scalar valued function $g(\cdot)$ (w. r. t. $\mathbf{x}$)
$\operatorname{sgn}(\cdot)$	Sign function
$\dot{\mathbf{f}}(t)$	Abbreviation for $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{f}(t)$ for a time-dependent function $\mathbf{f}(\cdot)$
$\varepsilon \searrow 0$	Scalar variable ε approaching 0 from above resp. right
$\lim_{\tau \searrow t} \mathbf{f}(\tau)$	One-sided limit from the right for a function $\mathbf{f}(\cdot)$ at point t
$\mathbf{f}(t^+)$	Abbreviation for $\lim_{\tau \searrow t} \mathbf{f}(\tau)$
$\lim_{\tau \nearrow t} \mathbf{f}(\tau)$	One-sided limit from the left for a function $\mathbf{f}(\cdot)$ at point t
$\mathbf{f}(t^-)$	Abbreviation for $\lim_{\tau \nearrow t} \mathbf{f}(\tau)$

Function spaces and associated norms

L^∞	Lebesgue space of essentially bounded measurable functions, see Definition 2.2
$\lVert\cdot\rVert_\infty$	L^∞ norm, see Definition 2.2

$W^{1,\infty}$	Space of absolutely continuous functions with essentially bounded derivatives, see Definition 2.4
$\|\cdot\|_{1,\infty}$	$W^{1,\infty}$ norm, see Definition 2.4

Optimal Control

$\mathbf{x}(\cdot)$	Differential states
$\mathbf{u}(\cdot)$	Control functions
$\mathfrak{u}$	Controllable parameters
$\mathbf{p}$	Non-controllable parameters
$\Phi^M(\cdot)$	Mayer-type objective function contribution
$\Phi^L(\cdot)$	Lagrange-type objective function contribution
$\mathbf{f}(\cdot)$	Right-hand side of Ordinary Differential Equation
$\boldsymbol{\Delta}(\cdot)$	Jump-function, modeling a discontinuity in $\mathbf{x}(\cdot)$
$\mathbf{c}(\cdot)$	Path constraint function
$\mathbf{r}(\cdot)$	Boundary constraint function
$T_1, \ldots, T_n$	Stage beginnings/endings in multi-stage OCPs
T_0	Initial time
$\mathcal{T}_j = [T_{j-1}, T_j]$	Time horizon of stage j in multi-stage OCPs
$\mathbf{f}^j(\cdot), \boldsymbol{\Delta}^j(\cdot), \ldots$	Phase-dependent model functions in multi-stage OCPs
n_x	Number of differential states
n_u	Number of control functions
$\mathfrak{n}_{\mathbf{u}}$	Number of controllable parameters
n_p	Number of non-controllable parameters

Multi-Body Systems

$\mathbf{q}(\cdot), \dot{\mathbf{q}}(\cdot), \ddot{\mathbf{q}}(\cdot)$	Generalized coordinates/velocities/accelerations
n_{dof}	Number of generalized coordinates
$\mathbf{H}(\cdot)$	Generalized inertia matrix
$\mathbf{C}(\cdot)$	Generalized bias force
$\boldsymbol{\tau}(\cdot)$	Generalized forces
$\boldsymbol{\tau}^a(\cdot)$	Actuated generalized forces
n_{act}	Number of actuated generalized forces
$\mathbf{g}(\cdot)$	External contact constraint
$\mathbf{G}(\cdot)$	Contact Jacobian $\frac{\partial}{\partial \mathbf{q}}\mathbf{g}(\cdot)$
$\boldsymbol{\gamma}(\cdot)$	The expression $\left(-\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{G}(\cdot)\right)\dot{\mathbf{q}}(\cdot)$

$\boldsymbol{\lambda}(\cdot)$	Constraint forces
$\mathbf{g}^j(\cdot), \mathbf{G}^j(\cdot), \ldots$	According quantities for phase-varying external contacts
$\mathbf{g}^-(\cdot), \mathbf{g}^+(\cdot)$	External contact before/after a change of contacts
$\mathbf{G}^-(\cdot), \mathbf{G}^+(\cdot)$	Contact Jacobian belonging to $\mathbf{g}^-(\cdot)/\mathbf{g}^+(\cdot)$
$\dot{\mathbf{q}}^-(\cdot), \dot{\mathbf{q}}^+(\cdot)$	Gen. velocities immediately before/after a change of contacts
$\boldsymbol{\Lambda}$	Contact Impulse

Switches, Switching Costs, and Jumps – Chapter 5 and Section 7.1

$w(\cdot)$	Mode-indicator function
n	Number of modes
$\boldsymbol{\omega}(\cdot)$	Binary valued mode indicator function
$\mathcal{T} = [t_0, t_f]$	Time horizon $[t_0, t_f]$ with initial time t_0 and final time t_f
$\bar{\delta}$	Dwell time
$PC_{\bar{\delta}}$	Right-continuous piecewise constant functions with dwell time $\bar{\delta}$, see Section 5.2, p. 66
$\mathcal{S}(\cdot)$	Set of discontinuities, e. g., for mode-indicator functions, see Section 5.2, p. 66
$\mathcal{S}(w), \mathcal{S}(\boldsymbol{\omega})$	Sets of switching points
$\mathbf{f}^j(\cdot)$	Ordinary Differential Equation right-hand side during mode j
$j_1 \rightarrow_w j_2$	Change of modes from j_1 to j_2
$j_1 \rightarrow_{\boldsymbol{\omega}} j_2$	Similar meaning as $j_1 \rightarrow_w j_2$
t_s	Switching point
$\boldsymbol{\Delta}_{j_1,j_2}(\cdot)$	Jump function acting if $j_1 \rightarrow_w j_2$
$\mathbf{d}(\cdot)$	Mode-independent path constraint function
$\mathbf{c}^j(\cdot)$	mode-dependent path constraint functions
$\mathbf{r}(\cdot)$	Boundary constraint function
$\Phi(\cdot)$	Mayer-type objective function contribution
$\lvert\mathcal{S}(w)\rvert, \lvert\mathcal{S}(\boldsymbol{\omega})\rvert$	Number of switching points
π	penalization parameter
$\mathbb{S}^n$	The set $\left\{\mathbf{v} \in \{0,1\}^n \,\middle|\, \sum_{j=1}^n \mathbf{v}_j = 1\right\}$
$\operatorname{conv}(\mathbb{S}^n)$	Convex hull of $\mathbb{S}^n$
Ω^n	The set $\left\{\boldsymbol{\omega} : \mathcal{T} \rightarrow \{0,1\}^n \mid \boldsymbol{\omega}(t) \in \mathbb{S}^n \; \forall\, t \in \mathcal{T}\right\}$
$\mathcal{G}$	Finite subset of $\mathcal{T} \setminus \{t_0, t_f\}$
$\boldsymbol{\theta}_{j_1,j_2}(\cdot)$	Switching indicator function belonging to "omniscient" indicators
$\boldsymbol{\Delta}(\cdot)$	Aggregated jump function
$\boldsymbol{\alpha}(\cdot)$	Relaxed mode-indicator function
$\mathbb{G}$	Time grid

t_i	Grid points in $\mathbb{G}$
$N+1$	Number of grid points of $\mathbb{G}$
$\mathbf{a}^i$	Values of $\boldsymbol{\alpha}(\cdot)$ in half-open grid interval $[t_i, t_{i+1})$ after discretization
$\boldsymbol{\beta}_{j_1,j_2}(\cdot)$	Auxiliary function
$\boldsymbol{\theta}^i_{j_1,j_2}$	"Omniscient" switching indicators, describing the discretization of $\boldsymbol{\theta}_{j_1,j_2}(\cdot)$
$\boldsymbol{\beta}^i_{j_1,j_2}$	Parameters describing the discretization of $\boldsymbol{\beta}_{j_1,j_2}(\cdot)$
$\mathbf{U}(\cdot)$	Representation of control functions $\mathbf{u}(\cdot)$ after parametrization
$\boldsymbol{\Theta}(\cdot)$	Switching indicator function belonging to "involved" indicators
$\boldsymbol{\Theta}^i_j$	"Involved" switching indicators, describing the discretization of $\boldsymbol{\Theta}_j(\cdot)$
$\boldsymbol{\theta}(\cdot)$	Switching indicator function belonging to "subsequent" indicators
$\boldsymbol{\theta}^i_j$	"Subsequent" switching indicators, describing the discretization of $\boldsymbol{\theta}_j(\cdot)$
γ	Vanishing constraint relaxation parameter
γ_0	Initial value of γ for homotopy $\gamma \searrow 0$, see Section 5.7
γ_{acc}	Termination tolerance for homotopy $\gamma \searrow 0$, see Section 5.7

Section 7.1 and Appendix A

$x_h(\cdot), y_h(\cdot)$	Horizontal/vertical position of head of "Simplest Walker" MBS
$\varphi_l(\cdot), \varphi_r(\cdot)$	Angle describing rotation of left/right leg
$x_l(\cdot), y_l(\cdot)$	Horizontal/vertical position of left foot
$x_r(\cdot), y_r(\cdot)$	Horizontal/vertical position of right foot
$\varepsilon^c_{\text{tol}}$	Cyclicity constraint tolerance
$\varepsilon^p_{\text{tol}}$	Ground penetration tolerance
Mode 1	Left foot is fixed to ground
Mode 2	Right foot is fixed to ground
$\mathbf{y}(\cdot)$	Vector of time-transformed generalized coordinates and velocities
$\mathbf{z}_2(\cdot)$	Differential state encoding process duration
$\mathbf{z}_1(1)$	Overall mechanical effort

Treatment Planning – Chapters 4, 6, and Section 7.2

$\Delta\mathbf{p}$	Change of parameters
$\mathbf{p}_{\text{nom}}$	Nominal post-operative parameter value
$\mathbf{p}_{\text{pre}}$	Pre-operative parameter value

$\Omega_{\mathbf{p}}$	Uncertainty set
$\underline{\boldsymbol{\tau}}_i^{\text{pass}}(\cdot)$	Passive reset force acting near lower virtual bound
$\overline{\boldsymbol{\tau}}_i^{\text{pass}}(\cdot)$	Passive reset force acting near upper virtual bound
$\underline{c}_i, \overline{c}_i$	Respective curvatures of passive reset forces
β_i	Damping parameters
$\boldsymbol{\tau}_i^{a,\max}$	Maximum active actuated generalized forces
$\varphi(\cdot)$	Assessment function for quality of gait patterns
$\mathbf{g}(\mathbf{p})$	Actual establishing gait pattern for given $\mathbf{p}$

Section 7.2

$x_p(\cdot), y_p(\cdot)$	Horizontal/vertical position of origin of walker MBS upper body
$\varphi_p(\cdot)$	Rotation of upper body about origin
$\varphi_{h,l}(\cdot), \varphi_{h,r}(\cdot)$	Rotation of left/right thigh about left/right hip joint
$\varphi_{k,l}(\cdot), \varphi_{k,r}(\cdot)$	Rotation of left/right shank about left/right knee joint
$x_l(\cdot), y_l(\cdot)$	Horizontal/vertical position of left foot
$x_r(\cdot), y_r(\cdot)$	Horizontal/vertical position of right foot
$\boldsymbol{\tau}_{h,l}(\cdot), \boldsymbol{\tau}_{h,r}(\cdot)$	Torque acting through left/right hip joint
$\boldsymbol{\tau}_{k,l}(\cdot), \boldsymbol{\tau}_{k,r}(\cdot)$	Torque acting through left/right knee joint
$\underline{\boldsymbol{p}}_{k,l}, \overline{\boldsymbol{p}}_{k,l}$	Lower/upper virtual bound for range of motion of left knee
$\underline{\boldsymbol{p}}_{k,r}, \overline{\boldsymbol{p}}_{k,r}$	Lower/upper virtual bound for range of motion of right knee
$\varepsilon_{\text{pass}}$	Threshold for normalized passive reset forces at $t = 0$
Phase 1	Right foot fixed to ground
Phase 2	Left foot fixed to ground
$\mathcal{T} = [T_0, T_2]$	Overall time horizon
$\boldsymbol{\lambda}^r(\cdot)$	Constraint force fixing right foot to ground during Phase 1
$\boldsymbol{\lambda}_x^r(\cdot), \boldsymbol{\lambda}_y^r(\cdot)$	Horizontal/vertical component of $\boldsymbol{\lambda}^r(\cdot)$
μ_{fric}	Friction coefficient for foot-ground contact
$\boldsymbol{\Delta}^j(\cdot)$	Jump function encoding discontinuity in $\mathbf{x}(\cdot)$ at end of Phase j
$\boldsymbol{\lambda}^l(\cdot)$	Constraint force fixing left foot to ground during Phase 2
$\boldsymbol{\lambda}_x^l(\cdot), \boldsymbol{\lambda}_y^l(\cdot)$	Horizontal/vertical component of $\boldsymbol{\lambda}^l(\cdot)$
x_{end}	Minimal terminal horizontal position of $x_p(\cdot)$
$\underline{\boldsymbol{p}}$	Vector aggregating $\underline{\boldsymbol{p}}_{k,l}$ and $\underline{\boldsymbol{p}}_{k,r}$
$\underline{\boldsymbol{p}}_{\text{pre}}$	Pre-operative value of $\underline{\boldsymbol{p}}$
$\underline{\boldsymbol{p}}_{\text{nom}}$	Nominal post-operative value of $\underline{\boldsymbol{p}}$
$\underline{\boldsymbol{p}}^*$	Parameter encoding worst-case treatment

List of Figures

List of Tables