Mathematical Programs with Complementarity Constraints in the Context of Inverse Optimal Control for Locomotion

Sebastian Albrecht[a] and Michael Ulbrich[a*]

[a] *Chair of Mathematical Optimization, Department of Mathematics, Technische Universität München, Boltzmannstr. 3, 85748 Garching, Germany*

In this paper an inverse optimal control problem in the form of a mathematical program with complementarity constraints (MPCC) is considered and numerical experiences are discussed. The inverse optimal control problem arises in the context of human navigation where the body is modeled as a dynamical system and it is assumed that the motions are optimally controlled with respect to an unknown cost function. The goal of the inversion is now to find a cost function within a given parametrized family of candidate cost functions such that the corresponding optimal motion minimizes the deviation from given data. MPCCs are known to be a challenging class of optimization problems typically violating all standard constraint qualifications. We show that under certain assumptions the resulting MPCC fulfills constraint qualifications for MPCCs being the basis for theory on MPCC optimality conditions and consequently for numerical solution techniques. Finally, numerical results are presented for the discretized inverse optimal control problem of locomotion using different solution techniques based on relaxation and lifting.

Keywords: inverse optimal control, mathematical program with complementarity constraints (MPCC), constraint qualification (CQ), constant positive-linear dependence (CPLD), locomotion

AMS Subject Classification: 49K15, 49M05, 90C33, 90C46

1. Introduction

Properties of human motions are analyzed in various disciplines ranging from biology to computer sciences. Naturally, the perspectives on this motions differs considerably; where one field might be interested in properties that can be obtained by analyzing recorded data, e.g., [22, 43], another might focus on postulating behavioral laws and comparing the output of the constructed system with the data, e.g., [5, 23, 34, 37]. In this paper we consider an inverse optimal control problem arising in the context of human locomotion, consequently, the goal is rather to find a suitable model than determine a biologically or psychologically plausible principle [20].

The locomotion problem is to move from a given start to a given end position without considering individual steps as a human would do them. This macroscopic perspective considers a plant with continuous dynamics that can be described by ordinary differential equations. If combined with a suitable cost function one obtains a standard optimal control problem discussed in detail in literature, e.g., [13, 44, 54, 57]. The direct approach to optimal control is chosen here and thus a combination of a discretization technique and a nonlinear optimization method is used. Two main approaches for discretizing the

---

*Corresponding author. Email: mulbrich@ma.tum.de

optimal control problem are multiple shooting, e.g., [7, 14, 48], and collocation, e.g., [7, 61]. Because the resulting problem is a standard nonlinear optimization problem a large number of optimization methods is available including sequential quadratic programming (SQP) methods, e.g., [28, 31, 49], and interior point methods (IPM), e.g., [30, 49, 58, 63].

The goal of the considered inverse optimal control task is to determine a cost function within a given parametrized family of cost functions such that the corresponding optimal control result has minimal distance to given data. In consequence, this problem is a special bilevel optimal control problem where the lower level is the optimal control problem and the upper level is the inversion problem. Some problems considered in game theory, e.g., [9, 16], are closely related to bilevel optimal control problems and the derivation of necessary optimality conditions, e.g., [65, 66], follows the line of standard bilevel problems. Bilevel optimal control problems where the lower level is the optimal control problem of a plane or a traveling-crane in a high-rack are considered in [21, 42]. The problem formulation of inverse optimal control for human arm motions is stated in [12]. In [47] a numerical method combining individual solvers for the two levels is presented for the locomotion problem. This approach is also used for the inversion of human arm motions in [10]. The approach of [1, 2, 35] is to use the first-order optimality conditions (KKT conditions) of the optimal control problem to transform the inverse optimal control problem into a standard nonlinear optimization problem. The inversion results of human data presented therein showed that this is a viable approach to tackle this challenging problem class. If the KKT conditions are used for the reformulation of the problem, the inequality constraints of the optimal control problem result in complementarity constraints. The implications following from these complementarity constraints on the numerical solution of the reformulated problem is the central aspect of this paper.

Mathematical programs with complementarity constraints (MPCCs) are a challenging problem class from both the theoretical and the numerical perspective because the complementarity constraints typically result in the violation of all standard constraint qualifications (CQs), e.g., [26, 50, 52], and consequently, the standard KKT theory cannot be applied to obtain optimality conditions. Therefore, the special MPCC structure has been used in a large number of publications to formulate MPCC-CQs, i.e., constraint qualifications that allow to derive suitable optimality conditions for MPCCs, e.g., [24–26, 50, 52, 67]. Naturally, these theoretical works are closely linked to the development of numerical methods for solving MPCCs, e.g., [6, 17, 27, 53].

The focus of this paper is on solving the MPCC resulting from inverse optimal control problems by applying different relaxation and lifting approaches. A relaxation scheme generates a sequence of problems where the complementarity condition is replaced by an approximation and the goal is that in the limit the solutions of the relaxed problems converge to a solution of the original MPCC. Two such relaxation approaches are considered here: the global relaxation approach of Scholtes [53] and the local relaxation approach of Steffensen and Ulbrich [55]. Instead of relaxing the kink of the complementarity condition lifting approaches introduce further variables, i.e., lifting the problem to higher dimensions. We will have a closer look on the lifting of Stein [56], where the kink is a projection of a smooth curve, and the lifting of Hatz et al. [35, 36], where a positivity condition on a complementarity pair is relaxed by an additional variable that is then penalized in the cost function.

The organization of the paper is as follows: First, the general structure of the discretized inverse optimal control problem in MPCC-form is introduced in section 2. This is followed by section 3 on the human locomotion problem: The dynamics are introduced, possible cost functions are discussed and distance measures are addressed. The next section 4 states some concepts of MPCC-CQs, which are the prerequisites of the numerical

optimization methods for MPCCs discussed in section 5, and shows that under certain assumptions such MPCC-CQs hold true for the considered discretized inverse optimal control problems. Numerical results of these methods in the context of inverse optimal control of locomotion are presented in the final section 6.

Most of the notation used in this paper is standard. By default a vector v is a column vector, i.e., $v \in \mathbb{R}^{n \times 1}$, with elements v_i, $i = 1, \ldots, n$. The Euclidean norm of the vector is denoted by $||v||_2$ and an inequality relating a vector to a scalar has to be understood elementwise, e.g., $v > 0 \Leftrightarrow v_i > 0 \; \forall i = 1, \ldots, n$. The notation $(v^{(k)})$ is used for a sequence of vectors $v^{(k)}$, $k \in \mathbb{N}$. The gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by $\nabla f(x) \in \mathbb{R}^n$ for $x \in \mathbb{R}^{n \times 1}$; for the Jacobian of a vector-valued function $g : \mathbb{R}^n \to \mathbb{R}^m$ the symbol $Dg(x) \in \mathbb{R}^{n \times m}$ with $x \in \mathbb{R}^n$ is used and the notation $\nabla g(x) := Dg(x)^T$ is in accordance with the definition of the gradient above if $m = 1$. For the Hessian of f the symbol $\nabla^2 f(x)$ is used and in case of a function $\widetilde{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ the notation $\nabla_{xx} \widetilde{f}(x, y)$ stands for the Hessian of $\widetilde{f}$ with respect to the x-variable only. Let $I \subseteq \{1, \ldots, n\} \subset \mathbb{N}$ be an index set with cardinality $|I| = k$ (the complement is $I^c := \{1, \ldots, n\} \setminus I$ with $|I^c| = n - k$), then the subvector corresponding to I of a vector $v \in \mathbb{R}^n$ is denoted by $v_I \in \mathbb{R}^k$. Given a matrix $M \in \mathbb{R}^{n \times m}$ and a second index set $J \subseteq \{1, \ldots, m\}$, the submatrix corresponding to row indices in I and column indices in J is stated in the form $M_{I,J}$, e.g., $Dg_{I,J}(x)$ denotes such a submatrix of the Jacobian, and if all column indices are considered, i.e., $J = \{1, \ldots, m\}$, the shortened notation M_I is used. However, for more complicated index sets we use the notation $M[I, J]$ or $M[I]$ instead. The symbol $\mathcal{I}$ is used for the squared identity matrix.

2. Inverse Optimal Control Problem Formulation

An optimal control problem describes the task of finding an optimal control $\overline{u}$ and a corresponding state $\overline{x}$ which minimize a given cost function $\overline{\phi}$ subject to boundary conditions b and the dynamics of the system φ. If the cost function is not assumed to be fixed but dependent on a parameter y, the corresponding (local) solutions $\overline{x}^*$ and $\overline{u}^*$ depend on this parameter, too. For the formulation of the problem suitable function spaces $\mathbb{X}$ and $\mathbb{U}$ have to be chosen for $\overline{x}$ and $\overline{u}$, respectively. For example, consider $\mathbb{X} = AC([0, 1], \mathbb{R}^{\overline{n}})$, the set of absolutely continuous functions, and for $\mathbb{U}$ the set of measurable functions $\overline{u} : [0, 1] \to \mathbb{R}^{\overline{m}}$ (cf. [15]).

Assuming that both $\overline{x}$ and $\overline{u}$ are functions of the independent time variable $t \in [0, 1]$, the corresponding optimal control problem with parameter reads:

$$
\begin{aligned}
\min \quad & \overline{\phi}(\overline{x}, \overline{u} \mid y) \\
\text{subject to} \quad & \overline{x}'(t) = \varphi(\overline{x}(t), \overline{u}(t)), \quad t \in [0, 1] \; a.e., \\
& g(\overline{x}(t), \overline{u}(t)) \leq 0, \quad t \in [0, 1] \; a.e., \\
& b(\overline{x}(0), \overline{x}(1)) = 0,
\end{aligned} \tag{1}
$$

where the parameter $y \in \mathbb{R}^m$ is given. Note that in this formulation the parameter y enters only the cost function, but generalizations to parameters in the constraints introduce no significant changes in the problem structure. If the cost function $\overline{\phi}$ has an integral cost term, one additionally has to demand that a feasible $\overline{x}$ and $\overline{u}$ guarantee the integrability of the integral cost term.

The existence of a solution for this type of optimal control problem can be proven by using the existence theorem of Filippov (cf. Theorem 9.3.i in [15]) if certain assumptions

on the problem describing functions and related sets hold including, e.g., continuity of φ and closedness of $\mathbb{B} := \left\{ (x_1, x_2) \in \mathbb{R}^{\overline{n} \times \overline{n}} \mid b(x_1, x_2) = 0 \right\}$.

The direct approach of optimal control is used for the optimal control problems in this paper, i.e., the continuous optimal control problem is first discretized and then optimized. The two main discretization approaches are multiple shooting and collocation. In the first case, a coarse partition $\Delta := \{t_i \mid i = 1, \ldots, \nu\}$ of the total time interval $[0, 1]$ is considered with $0 = t_1 < t_2 < \ldots < t_\nu = 1$ and the goal is to match the end state of one interval with the start state of the next where each end state is obtained by using a numerical integrator for the given ODE and start value of the interval. In the second case a slightly finer partition of $[0, 1]$ is used and the state is approximated on each interval by polynomials whose time derivatives equals φ for finitely many, specified collocation points and fulfill continuity conditions at the interval boundaries. Following the line of [61] suitable collocation strategies are, for example, obtained if a piecewise linear state approximation is combined with a piecewise constant control approximation or if a piecewise cubic state approximation is teamed with a piecewise linear control approximation. Convergence results of the optimality conditions for the discretized optimal control problem towards the optimality conditions for the continuous optimal control problem are discussed in [32, 33, 45, 61]. Assuming smooth state and control functions and control constraints only, pointwise convergence is proven for the above mentioned collocation strategies in [61].

Setting $n := (\overline{n} + \overline{m})(\nu - 1)$, the state and control variables at the boundaries of the intervals are stored in the combined vector $x \in \mathbb{R}^n$. Consequently, if ϕ is a suitable approximation of $\overline{\phi}$ and if h combines the equality constraints resulting from the discretization of the ODE with the boundary conditions b, the discretized optimal control problem with parameter reads:

$$
\begin{aligned}
\min \quad & \phi(x \mid y) \\
\text{subject to} \quad & h(x) = 0, \quad g(x) \le 0,
\end{aligned} \tag{2}
$$

with the given parameter $y \in \mathbb{R}^m$.

Let x^* be a local solution of the discretized optimal control problem for a given parameter y where a constraint qualification (CQ) is fulfilled (cf. Chapter 12 of [49]), then there exist Lagrange multipliers $\lambda^* \in \mathbb{R}^\nu$ and $\mu^* \in \mathbb{R}^\omega$ such that the Karush-Kuhn-Tucker (KKT) conditions hold:

$$
\begin{aligned}
\nabla \phi(x^* \mid y) + \nabla g(x^*)\lambda^* + \nabla h(x^*)\mu^* \quad &= \quad 0, \\
h(x^*) \quad &= \quad 0, \\
0 \le -g(x^*) \quad \perp \quad \lambda^* &\ge 0.
\end{aligned}
$$

The symbol $\perp$ is used to denote the orthogonality condition which here can, e.g., also be written as $g(x^*)^T \lambda^* = 0$.

If one is now interested in finding a suitable parameter $y \in \mathbb{R}^m$ such that a vector $x \in \mathbb{R}^n$ fulfilling the corresponding KKT conditions minimizes a distance measure Φ using given data Λ, the following problem of discretized inverse optimal control is obtained:

$$
\begin{aligned}
\min \quad & \Phi(x, y \mid \Lambda) \\
\text{subject to} \quad & \nabla_x \phi(x, y) + \nabla g(x)\lambda + \nabla h(x)\mu = 0, \\
& h(x) = 0, \quad H(y) = 0, \quad G(y) \leq 0, \\
& 0 \leq -g(x) \ \perp \ \lambda \geq 0,
\end{aligned}
\tag{3}
$$

where the functions H and G define the constraints on y. Note that the Lagrange multipliers λ and μ are optimization variables in addition to x and y.

If x is not assumed to be only a KKT point but a solution of the discretized optimal control problem (2), a bilevel optimization problem is obtained. Note that the above discretized inverse optimal control problem is not equivalent to such a bilevel optimization problem, because the optimal control problem is not convex and consequently, the KKT conditions are only necessary given a CQ but not sufficient. Bilevel problem are discussed in literature [4, 8, 18, 46] and under certain assumptions the existence of a global optimistic solution can be proven for the given setting [1, 2, 18].

The rest of this paper addresses the numerical solution of problem ((3)) and thus we introduce the defining function for our application example, e.g., the ODE φ and the cost function ϕ, in the following sections and then discuss solution techniques.

3. Human Locomotion

Human locomotion considers motions of the human body with the objective of changing the position in a terrestrial environment. We consider here the macroscopic problem of moving from a start to a goal position without paying attention to the complex dynamical problem of taking individual steps. Consequently, a simple model of the locomotion dynamics is discussed for a planar environment: the unicycle model, where the person is abstracted to a mass point with an orientation.

The idea of determining the optimal cost function used in human locomotion via inverse optimal control is introduced in [47]. There, obstacle-free paths are considered and the family of cost functions is given as linear combinations of five basic cost functions. The bilevel problem is solved by nesting the individual solvers for the data fitting problem and the optimal control problem. It is reported that the characteristics of the human motion data are met and the results are used to control a humanoid robot.

The inverse optimal control problem developed in the following differs, for example, in the choice of cost functions Φ and ϕ; especially, in [47] the parameter y appear only linearly in cost function ϕ, whereas here elements of y enter ϕ nonlinearly. However, the main difference is the inclusion of inequality constraints in the formulation of the optimal control problem (1) which results in the complementarity conditions of problem (3).

3.1 *Locomotion Dynamics*

Simplifying the human navigation problem to a two-dimensional problem, the configuration of the human can be described by his/her Cartesian coordinates $\mathcal{P}_P(t) = (\mathcal{P}_{Px}(t), \mathcal{P}_{Py}(t)) \in \mathbb{R}^2$ and its orientation $\beta_P(t) \in [0, 2\pi]$ at a time instance t; in the following, the direction given by the angle $\beta_P(t)$ is referred to as the forward direction.

The considered model assumes that the rigid body can only be (linearly) accelerated in the forward direction and consequently, the following ordinary differential equations
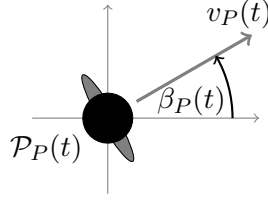
Figure 1. Schematic illustration of the unicycle model.

state the dynamics related to a translation of the rigid body:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_P(t) = (v_P(t)\cos(\beta_P(t)), v_P(t)\sin(\beta_P(t)))^T \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2}v_P(t) = \overline{u}_v(t),$$

in forward direction. The rotational dynamics are modeled by the simple differential equation $\frac{\mathrm{d}^3}{\mathrm{d}t^3}\beta_P(t) = \overline{u}_\beta(t)$,

where $\overline{u}_\beta$ being the jerk of the orientation angle β_P is the second control variable. Note that these simple integrator chains can be extended to a model using the mass and the inertia of the moving person. In a similar manner motions in sidewards direction can be included in this model. However, the focus on more general properties of human locomotion and therefore more complex models are set aside.

In consequence, the following system of first-order ordinary differential equations describes the model dynamics: $\frac{\mathrm{d}}{\mathrm{d}t}\overline{x}(t) = \overline{A}(\beta_P(t))\overline{x}(t) + \overline{B}\overline{u}(t)$, where the matrices are given by

$$\overline{A}(\beta_P(t)) := \begin{pmatrix} 0 & 0 & 0 & \cos(\beta_P(t)) & 0 & 0 & 0 \\ 0 & 0 & 0 & \sin(\beta_P(t)) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \overline{B} := \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the state $\overline{x}(t)$ and the control $\overline{u}(t)$ are defined by

$$\overline{x}(t) := (\mathcal{P}_{Px}(t), \mathcal{P}_{Py}(t), \beta_P(t), v_P(t), \beta_P{}'(t), v_P{}'(t), \beta_P{}''(t))^T$$

and $\overline{u}(t) := (\overline{u}_v(t), \overline{u}_\beta(t))^T$, accordingly.

3.2 Locomotion Cost Functions

The general idea of the inverse optimal control problems considered here is to find a cost function ϕ such that a corresponding KKT point minimizes the distance measure Φ. For the mathematical formulation of this problem a vector y is introduced that enters the cost function ϕ of the optimal control problem and certain constraints $G(y) \leq 0$ and $H(y) = 0$ have to be specified to guarantee a well-posed problem. Some candidates ϕ_i for cost functions describing human behavior can be found in the literature, e.g., [23, 59], others can easily be formulated that address certain aspects of locomotion [3, 47]. However, since no single cost function is known to reproduce human locomotion

behavior, a combination of such basic cost functions ϕ_i is considered here. One possibility is to use convex combinations:

$$\phi(x,y) = \sum_{i=1}^{k} w_i \phi_i(x, \pi_i), \quad H(y) = 1 - \sum_{i=1}^{k} w_i, \quad G(y) = \begin{pmatrix} -w \\ \pi - \pi_u \\ \pi_l - \pi \end{pmatrix},$$

where $y = (w^T, \pi^T)^T$ with the weights $w^T := (w_1, \ldots, w_k) \in \mathbb{R}^k$ and the parameters $\pi^T := (\pi_1, \ldots, \pi_k)$ that enter the cost functions nonlinearly. Note that the vectors π_u and π_l specify upper and lower bounds for each parameter π_i. However, for the later analysis the formulation relative to a given cost function ϕ_0 (i.e., for the function ϕ_0 no weight w_0 is introduced) is preferred with $\pi^T := (\pi_0, \pi_1, \ldots, \pi_k)$, cf. [47]:

$$\phi(x,y) = \phi_0(x, \pi_0) + \sum_{i=1}^{k} w_i \phi_i(x, \pi_i), \quad G(y) = \begin{pmatrix} -w \\ \pi - \pi_u \\ \pi_l - \pi \end{pmatrix}. \tag{4}$$

The starting point for defining a suitable set of basic cost functions ϕ_i is the minimization of the time integral of selected state or control variables.

$$f_{x,j}(\overline{x}) := \int_{t_0}^{t_f} \overline{x}_j(t)^2 \ \mathrm{d}t \ \text{ and } \ f_{u,j}(\overline{u}) := \int_{t_0}^{t_f} \overline{u}_j(t)^2 \ \mathrm{d}t.$$

These integral cost functions correspond to the continuous formulation of problem (1) and lead to candidates for $\overline{\phi}$; thus a discretization technique is needed to obtain the relevant cost functions ϕ_i. Two standard approaches are the transformation into a Meyer problem by introducing an additional ODE corresponding to the integrand or the application of an individual numerical quadrature method. Note that these general cost functions are denoted, e.g., by $f_{u,j}$ and later in the numerics section the specific choices for the ϕ_i are given, for example, ϕ_1 then corresponds to the sum of both $f_{u,j}(\overline{u})$. Considering not the value of a state variable itself, but the deviation from a reference value $r_j \in \mathbb{R}$ leads to the definition of further cost functions. A realization of such a cost function could be motivated by the tendency to walk at a comfortable walking speed:

$$f_{ref,j}(\overline{x}, r_j) := \int_{t_0}^{t_f} (\overline{x}_j(t) - r_j)^2 \ \mathrm{d}t.$$

Note that the reference value r_j corresponds to a parameter π_i being optimized. Another considered cost function is the deviation from a straight line connecting start and goal positions.

$$f_{line}(\overline{x}) := \int_{t_0}^{t_f} ||\mathcal{P}_P(t) - \mathbb{P}_{line}(\mathcal{P}_P(t))||^2 \ \mathrm{d}t,$$

where $\mathbb{P}_{line}$ is the projection on the straight line connecting the start and goal position. Furthermore, the cost function f_{goal} introduced by [47] integrates the squared difference between the current orientation $\beta_P(t)$ and the direction towards the goal position $\mathcal{P}_G =$

$(\mathcal{P}_{Gx}, \mathcal{P}_{Gy}) \in \mathbb{R}^2$:

$$f_{goal}(\overline{x}) := \int_{t_0}^{t_f} \left(\beta_P(t) - \arctan \left(\frac{\mathcal{P}_{Gy} - \mathcal{P}_{Py}(t)}{\mathcal{P}_{Gx} - \mathcal{P}_{Px}(t)} \right) \right)^2 \, dt.$$

A cost function popular for describing human arm motions is minimizing the jerk of the hand position [23]; if adapted to the locomotion problem the following cost function is obtained:

$$f_{jerk}(\overline{x}) := \int_{t_0}^{t_f} \left(\frac{d^3 \mathcal{P}_P(t)}{dt^3} \right)^2 \, dt.$$

Since the navigation tasks are considered to have a free final time t_f, the minimization of this final time gives rise to another cost function: $f_{time}(\overline{x}) := t_f$.

Note that the definition of problem (1) assumed a fixed time interval $[0, 1]$, therefore the problem is transformed by introducing the variable t_f as an additional state variable with a trivial ODE and scaling of the time derivatives.

A selection of the introduced cost function is used in the numerics section to define the basic cost functions leading via convex combinations to the parametrized family of considered cost functions in the inverse optimal control setup.

3.3   *Boundary Conditions and Inequality Constraints*

Two elements, the boundary conditions and possible inequality constraints, have still to be discussed for the complete statement of the optimal control problem. We consider here given index sets I_s and I_e and specified vectors x_s and x_e containing the state information for both start and end configuration to define the boundary conditions:

$$b(\overline{x}(0), \overline{x}(1)) := \left((\overline{x}_{I_s}(0) - x_s)^T, \quad (\overline{x}_{I_e}(1) - x_e)^T \right)^T.$$

Finally, inequality constraints are on a state variable, the forward velocity, are added: $0.2 \leq v_P(t) \leq 0.55$ [m/s].

These constraints result via the KKT conditions in the MPCC structure being the focus of this paper. Note that in the discretized optimal control problem (3) these inequalities have to be fulfilled only a the time instances t_i.

3.4   *Distance Measure*

For the specification of the discretized inverse optimal control problem (3) a suitable distance measure Φ has to be chosen, i.e., it has to be determined how two locomotion paths are compared in the data matching problem. Ideally one would like to match both the Cartesian path and the velocity profile, thus the choice for our numerical examples is to compare the positions of the computation $\mathcal{P}_P(t_i)$ with the position $\mathcal{P}_D(t_i)$ of the given data. The time instances t_i, $i = 1, \ldots, \overline{\nu}$, have to be specified in the interval $[t_0, t_f]$, however the motion time t_f is not fixed, but an optimization variable, too. In consequence, an interpolation approach is needed if not both values $\mathcal{P}_P(t_i)$ and $\mathcal{P}_D(t_i)$ are given at t_i:

$$\Phi(x, y \mid \Lambda) := \sum_{i=1}^{\bar{\nu}} (\mathcal{P}_P(t_i) - \mathcal{P}_D(t_i))^2.$$

Note that an adaption of the distance measure might be needed if locomotion data of real experiments is given for the problem (3), because the velocity profiles recorded in the experiments exhibit considerable oscillations corresponding to the individual steps of the participants. Since individual steps are not modeled in the dynamics, a combination of two distance measures decoupling positional and temporal information might be more appropriate (cf. [1, 3] for more information).

4. Mathematical Program with Complementarity Constraints

The discretized inverse optimal control problem (3) has to be handled with care due to its complementarity constraint

$$0 \leq -g(x) \perp \lambda \geq 0,$$

i.e., problem (3) is a mathematical program with complementarity constraints (MPCC). MPCCs are a challenging problem class because typically all standard constraint qualifications (CQs) are violated and therefore, optimality conditions cannot be obtained from the standard KKT theory, cf., e.g., [26, 50, 52]. In consequence, several works on special CQs using the MPCC problem structure explicitly have been published and suitable numerical optimization techniques have been proposed, e.g., [24–26, 50, 52, 67] (see also section 1).

In order to discuss the numerical optimization approaches used to solve the MPCC (3) and to reduce notational complexity, we state a general formulation of a mathematical program with complementarity constraints (MPCC):

$$\begin{aligned} &\min_{z} \widetilde{\phi}(z) \\ \text{subject to} \quad &\widetilde{h}(z) = 0, \\ &\widetilde{g}(z) \leq 0, \\ &0 \leq \widetilde{G}(p) \perp \widetilde{H}(q) \geq 0, \end{aligned} \tag{5}$$

All functions are assumed to be at least once continuously differentiable and the optimization variable $z \in \mathbb{R}^{\widetilde{n}}$ is the concatenation of a vector $\gamma \in \mathbb{R}^{\widetilde{n} - 2\widetilde{s}}$ and the vectors $p, q \in \mathbb{R}^{\widetilde{s}}$ relevant for the complementarity conditions.

Let z^* be a feasible point for MPCC (5), then the following index sets are used to specify the different types of active inequality constraints (the notation follows [38]):

$$
\begin{aligned}
I_{\widetilde{g}}(z^*) &:= \{i \in \{1,\ldots,\widetilde{m}\} \mid \widetilde{g}_i(z^*) = 0\}, \\
I_{\widetilde{h}}(z^*) &:= \{1,\ldots,\widetilde{r}\}, \\
I_{00}(z^*) &:= \{i \in \{1,\ldots,\widetilde{s}\} \mid \widetilde{G}_i(p) = 0,\ \widetilde{H}_i(q) = 0\}, \\
I_{+0}(z^*) &:= \{i \in \{1,\ldots,\widetilde{s}\} \mid \widetilde{G}_i(p) > 0,\ \widetilde{H}_i(q) = 0\}, \\
I_{0+}(z^*) &:= \{i \in \{1,\ldots,\widetilde{s}\} \mid \widetilde{G}_i(p) = 0,\ \widetilde{H}_i(q) > 0\}, \\
I_{\widetilde{G}}(z^*) &:= I_{00}(z^*) \cup I_{0+}(z^*), \\
I_{\widetilde{H}}(z^*) &:= I_{00}(z^*) \cup I_{+0}(z^*).
\end{aligned}
$$

Without loss of generality we assume here that the index sets I_{0+}, I_{00} and I_{+0} have the following property:

$$
i_{0+} < i_{00} < i_{+0} \quad \forall i_{0+} \in I_{0+}(z^*),\ i_{00} \in I_{00}(z^*),\ i_{+0} \in I_{+0}(z^*).
$$

Three well-known stationarity concepts for MPCCs are weak stationarity, C-stationarity and M-stationarity: Let z^* be feasible for the MPCC (5), then z^* is

- weakly stationary if Lagrange-multipliers $\mu \in \mathbb{R}^{\widetilde{r}}$, $\lambda \in \mathbb{R}^{\widetilde{m}}$, $\psi_{\widetilde{G}}, \psi_{\widetilde{H}} \in \mathbb{R}^{\widetilde{s}}$ exist such that

$$
\begin{aligned}
\nabla\widetilde{\phi}(z^*) + \nabla\widetilde{g}(z^*)\lambda + \nabla\widetilde{h}(z^*)\mu - \nabla\widetilde{G}(p)\psi_{\widetilde{G}} - \nabla\widetilde{H}(q)\psi_{\widetilde{H}} &= 0, \\
\lambda &\geq 0, \\
\lambda^T \widetilde{g}(z^*) &= 0, \\
\psi_{\widetilde{G},i} &= 0, \quad i \in I_{0+}(z^*), \\
\psi_{\widetilde{H},i} &= 0, \quad i \in I_{+0}(z^*).
\end{aligned}
$$

- C-stationary if it is weakly stationary and if the products $\psi_{\widetilde{G},i}\psi_{\widetilde{H},i} \geq 0$ for all indices $i \in I_{00}$.
- M-stationary if it is weakly stationary and if either the product $\psi_{\widetilde{G},i}\psi_{\widetilde{H},i} = 0$ or both $\psi_{\widetilde{G},i} > 0$ and $\psi_{\widetilde{H},i} > 0$ for all indices $i \in I_{00}$.

Analogue to the case of standard nonlinear problems, MPCC-CQs are needed to guarantee that a local minimizer z^* of the MPCC (5) is weakly, C- or M-stationary. Several CQs known from a standard nonlinear optimization setup have been transferred to MPCCs by considering the following tightened nonlinear program (TNLP(z^*)). Given z^* is feasible for the MPCC (5), the TNLP(z^*) is obtained by transforming the inequalties corresponding to the index sets of active inequalities in z^* into equality constraints:

$$
\begin{aligned}
\min_{z} &\quad \widetilde{\phi}(z) \\
\text{subject to} &\quad \widetilde{h}(z) = 0, \\
&\quad \widetilde{g}(z) \leq 0, \\
&\quad \widetilde{G}_i(p) \geq 0,\quad \widetilde{H}_i(q) = 0,\quad \forall i \in I_{+0}(z^*), \\
&\quad \widetilde{G}_i(p) = 0,\quad \widetilde{H}_i(q) \geq 0,\quad \forall i \in I_{0+}(z^*), \\
&\quad \widetilde{G}_i(p) = 0,\quad \widetilde{H}_i(q) = 0,\quad \forall i \in I_{00}(z^*).
\end{aligned}
\tag{6}
$$

This TNLP(z^*) can now be used to transfer several well-known CQs for standard non-linear optimization problems, e.g., LICQ, MFCQ and CRCQ, to the MPCC setting: The corresponding MPCC-CQ is fulfilled in a feasible point z^* of the MPCC if the respective CQ is fulfilled for the tightened problem TNLP(z^*). Since the MPCC-CPLD is not that common, but will be used later on, the definition of positive-linear dependence (cf. Definition 2.1 in [51]) and the respective MPCC-CQ are stated explicitly (cf. Definition 3.2 in [38]).

DEFINITION 4.1 *(Positive-Linearly Dependent)*
Let two index sets I_α and I_β be given. Two sets of vectors $S_\alpha = \{v_i\}_{i \in I_\alpha} \subset \mathbb{R}^n$ and $S_\beta = \{w_i\}_{i \in I_\beta} \subset \mathbb{R}^n$ are positive-linearly dependent if scalars $\{\alpha_i\}_{i \in I_\alpha}$ and $\{\beta_i\}_{i \in I_\beta}$ exist with $\alpha_i \geq 0$ for all $i \in I_\alpha$, not all of them being zero, such that

$$\sum_{i \in I_\alpha} \alpha_i v_i + \sum_{i \in I_\beta} \beta_i w_i = 0.$$

Otherwise, the sets are positive-linearly independent.

Note that only the α_i are restricted to nonnegative scalars, not the β_i. For further usage we restate the definition of positive-linearly dependence for rows of two matrices $M_\alpha \in \mathbb{R}^{n_\alpha \times n}$ and $M_\beta \in \mathbb{R}^{n_\beta \times n}$: Let $I_\alpha \subseteq \{1, \ldots, n_\alpha\}$ and $I_\beta \subseteq \{1, \ldots, n_\beta\}$ be two sets of row indices of the corresponding matrices. The two matrices M_α and M_β are positive-linearly dependent with respect to I_α and I_β if vectors $\alpha \in \mathbb{R}^{n_\alpha}$ and $\beta \in \mathbb{R}^{n_\beta}$ fulfilling $||\alpha_{I_\alpha}||_2 + ||\beta_{I_\beta}||_2 \neq 0$, $\alpha \geq 0$, exist such that

$$\alpha_{I_\alpha}^T \, M_\alpha[I_\alpha] + \beta_{I_\beta}^T \, M_\beta[I_\beta] = 0.$$

The constant positive-linear dependence (CPLD) condition is originally defined for general nonlinear optimization problems (cf. Definition 2.6 in [51]), thus the structure of problem (2) is considered here:

A feasible point x^* is said to satisfy the CPLD if for all index sets $I_\alpha \subseteq \mathbb{A}(x^*) := \{i \mid g_i(x^*) = 0\} \subseteq \{1, \ldots, \nu\}$ and $I_\beta \subseteq \{1, \ldots, \omega\}$ such that $M_\alpha = Dg(x^*)$ and $M_\beta = Dh(x^*)$ are positive-linearly dependent with respect to I_α and I_β, there is a neighborhood $N(x^*)$ of x^* such that for any $x \in N(x^*)$ the rows of the matrix

$$\begin{pmatrix} Dg_{I_\alpha}(x) \\ Dh_{I_\beta}(x) \end{pmatrix}$$

are linearly dependent. If this definition is applied to TNLP(z^*), the following definition of MPCC-CPLD is obtained (cf. Definition 2.3 in [38]):

DEFINITION 4.2 *(MPCC-CPLD)*
Let z^* be feasible for the MPCC (5) and define

$$M_\alpha(z) := D\widetilde{g}(z) \quad and \quad M_\beta(z) := \begin{pmatrix} D\widetilde{h}(z) \\ D\widetilde{G}(z) \\ D\widetilde{H}(z) \end{pmatrix}.$$

Additionally, define the index set

$$I_T(z^*) = I_{\widetilde{h}}(z^*) \cup \{i + \widetilde{r} \mid i \in I_{\widetilde{G}}(z^*)\} \cup \{i + \widetilde{r} + \widetilde{s} \mid i \in I_{\widetilde{H}}(z^*)\}$$

corresponding to the equality constraints of TNLP(z^*) (6). The condition of constant positive-linear dependence for MPCCs (MPCC-CPLD) is fulfilled if for any subsets $I_\alpha \subseteq I_{\widetilde{g}}(z^*)$ and $I_\beta \subseteq I_T(z^*)$ such that the matrices $M_\alpha(z^*)$ and $M_\beta(z^*)$ are positive-linearly dependent with respect to I_α and I_β, there exists a neighborhood $N(z^*)$ of z^* such that the rows of the matrix

$$\begin{pmatrix} M_\alpha[I_\alpha](z) \\ M_\beta[I_\beta](z) \end{pmatrix}$$

are linearly dependent for all $z \in N(z^*)$.

The condition MPCC-CPLD is an MPCC-CQ for M-stationarity because MPCC-CPLD implies MPCC-ACQ (cf. Lemma 3.3 in [38] and Corollary 3.4 in [24]) and the MPCC-ACQ is a sufficient condition for M-stationarity of local optima according to Theorem 3.9 in [25].

LEMMA 4.3 *Every local minimizer z^* of the MPCC (5) satisfying MPCC-CPLD is M-stationary.*

In order to use standard numerical optimization methods to obtain a solution of an MPCC, one can either solve a sequence of relaxed problems, see sections 5.2 and 5.1, or transform the problem by lifting approaches, cf., sections 5.4 and 5.3. Certain emphasis lies here on the local relaxation approach of Steffensen and Ulbrich [60] being locally convergent to C-stationary points given the MPCC-CPLD condition (cf. Theorem 3.4 in [38]).

4.1 *Locomotion MPCC Structure*

The next goal is to show that under certain conditions on the discretized optimal control problem and on the inversion problem the locomotion problem (cf. section 3) fulfills the MPCC-CPLD condition and in some cases even the MPCC-LICQ.

The needed assumptions include a second order condition for the discretized optimal control problem (2).

It has to be assumed that all functions describing the nonlinear optimization problem are twice continuously differentiable. For the formulation of these assumptions we state the following cones of relevant directions

$$
\begin{aligned}
\mathbb{T}^+(g,h,x,\lambda) := \Big\{ d \in \mathbb{R}^{\widetilde{n}} \mid \nabla g_i(x)^T d \ &= \ 0, \ \text{if } i \in \mathbb{A}(x) \text{ and } \lambda_i > 0, \\
\nabla g_i(x)^T d \ &\leq \ 0, \ \text{if } i \in \mathbb{A}(x) \text{ and } \lambda_i = 0, \\
\nabla h(x)^T d \ &= \ 0 \Big\}, \\
\mathbb{T}(g,h,x,\lambda) := \Big\{ d \in \mathbb{R}^{\widetilde{n}} \mid \nabla g_i(x)^T d \ &= \ 0, \ \text{if } i \in \mathbb{A}(x) \text{ and } \lambda_i > 0, \\
\nabla h(x)^T d \ &= \ 0 \Big\}.
\end{aligned}
$$

Denoting the Lagrangian of problem (2) by $L(x, \lambda^*, \mu^*)$, the following second-order

condition is considered (cf., e.g., Section 2 in [40]): The second-order sufficient conditions (SOSC) hold at the point (x^*, λ^*, μ^*) if the following condition is fulfilled:

$$d^T \nabla_{xx} L\left(x^*, \lambda^*, \mu^*\right) d > 0 \quad \forall d \in \mathbb{T}^+(g, h, x^*, \lambda^*) \setminus \{0\} \,.$$

We recall that the SOSC and the KKT conditions together imply that x^* is a strict local minimum. If instead of $\mathbb{T}^+(g, h, x^*, \lambda^*)$ the cone $\mathbb{T}(g, h, x, \lambda)$ is considered, the following stronger second-order sufficient condition (SSOSC) is obtained:

$$d^T \nabla_{xx} L\left(x^*, \lambda^*, \mu^*\right) d > 0 \quad \forall d \in \mathbb{T}(g, h, x^*, \lambda^*) \setminus \{0\} \,.$$

The next step is to show MPCC-CPLD for the discretized inverse optimal control problem (3). We will use that the constraints of the inversion problem G and H fulfill LICQ by construction and that the following lemma holds for problem (2), cf., e.g., Proposition 4 in [40]:

LEMMA 4.4 *Let x^* be a local minimum of the discretized optimal control problem* (2) *for the given parameters y fulfilling LICQ and SSOSC with the corresponding Lagrange multipliers λ^* and μ^*. Using the abbreviations $z^T := (x^T, \mu^T, \lambda^T, y^T)$ and accordingly z^*, the following symmetric matrix is introduced:*

$$\overline{M}(z \mid \overline{I}) := \begin{pmatrix} \nabla_{xx} L(z) & \nabla h(x) & \nabla g_{I_{0+}(z^*)}(x) & \nabla g_{\overline{I}}(x) \\ Dh(x) & 0 & 0 & 0 \\ Dg_{I_{0+}(z^*)}(x) & 0 & 0 & 0 \\ Dg_{\overline{I}}(x) & 0 & 0 & 0 \end{pmatrix} \tag{7}$$

with $\overline{I} \subseteq I_{00}(z^*)$. Then the matrix $\overline{M}(z^* \mid \overline{I})$ has full rank for all $\overline{I} \subseteq I_{00}(z^*)$.

In the definition of $\overline{M}(z \mid \overline{I})$ the index sets of Dg (and ∇g) do not depend on z but are fixed to the index set related to z^*, which gives rise to the following corollary:

COROLLARY 4.5 *Let the problem-describing functions be twice continuously differentiable and let the prerequisites of Lemma 4.4 be fulfilled, then there exists a neighborhood $N(z^*)$ of z^* such that $\overline{M}(z \mid \overline{I})$ has full rank for all $z \in N(z^*)$.*

If one is interested in the rank of a submatrix of $\overline{M}(z^* \mid \overline{I})$, the following corollary is useful:

COROLLARY 4.6 *For each index set I the rows of $\overline{M}_I(z^* \mid \overline{I})$ are linearly independent and there exists an index set J such that $\overline{M}_{I,J}(z^* \mid \overline{I})$ is quadratic and has full rank.*

For the proofs of the following theorems we state two lemmas from linear algebra regarding linear (in)dependence:

LEMMA 4.7 *Let a matrix $M \in \mathbb{R}^{n \times m}$ be given and two index sets $I \subset \{1, \ldots, n\} =: \widehat{I}$ and $J \subset \{1, \ldots, m\}$. Additionally, assume that the submatrix $M_{I^c, J} = 0$ and that the rows of the submatrix M_{I^c, J^c} are linearly independent. Then the following holds:*

(a) *The rows of the matrix M are linearly independent if the rows of the submatrix $M_{I,J}$ are linearly independent.*

(b) *The rows of the matrix M are linearly dependent if the following properties hold true:*

    (i) *the rows of the submatrix $M_{I,J}$ are linearly dependent,*

(ii) the rank of the matrix M_{I^c,J^c} is equal to the column rank of $M_{\widehat{I},J^c}$.

With these lemmas the following theorem holds true for the case of strict complementarity. The prerequisite that the constraints H and G fulfill the LICQ at y is fulfilled if the problem structure (4) is given:

THEOREM 4.8
Let x^* be a local minimum of the discretized optimal control problem (2) for the given parameters y^* fulfilling LICQ and SOSC with the corresponding Lagrange multipliers μ^* and λ^* and let strict complementarity hold. If the constraints H and G assure the LICQ at y^*, then MPCC-LICQ is satisfied for the discretized inverse optimal control problem (3).

Proof:
The tightened nonlinear program (6) corresponding to the MPCC (3) of the discretized inverse optimal control problem can be formulated using the following functions with $z^T = (x^T, \mu^T, \lambda^T, y^T)$ and $z^{*T} = (x^{*T}, \mu^{*T}, \lambda^{*T}, y^{*T})$:
$\widetilde{\phi}(z) := \Phi(x, y \mid \Lambda)$, $\widetilde{g}(z) := G(y)$, $\widetilde{H}(z) := \lambda$, $\widetilde{G}(z) := -g(x)$ and

$$
\widetilde{h}(z) := \begin{pmatrix} \nabla_x \phi(x, y) + \nabla g(x)\lambda + \nabla h(x)\mu \\ h(x) \\ H(y) \end{pmatrix}.
$$

The Jacobians of the functions describing the tightened problem have the following structure:

$$
\begin{aligned}
D\widetilde{g}(z) &= (0, 0, 0, DG(y)), \\
D\widetilde{h}(z) &= \begin{pmatrix} \nabla_{xx}L(x, \mu, \lambda) & \nabla h(x) & \nabla g(x) & \nabla_{xy}\phi(x, y) \\ Dh(x) & 0 & 0 & 0 \\ 0 & 0 & 0 & DH(y) \end{pmatrix}, \\
D\widetilde{G}_{I_{\widetilde{G}}(z^*)}(z) &= \begin{pmatrix} -Dg_{I_{0+}(z^*)}(x) & 0 & 0 & 0 \\ -Dg_{I_{00}(z^*)}(x) & 0 & 0 & 0 \end{pmatrix}, \\
D\widetilde{H}_{I_{\widetilde{H}}(z^*)}(z) &= \begin{pmatrix} 0 & 0 & \mathcal{I}_{I_{00}(z^*)} & 0 \\ 0 & 0 & \mathcal{I}_{I_{+0}(z^*)} & 0 \end{pmatrix}.
\end{aligned}
$$

Here $\mathcal{I}$ is the $\nu \times \nu$ identity matrix.

The index set I_{00} is empty due to strict complementarity. Thus, the condition MPCC-LICQ is fulfilled if the rows of the following matrix are linearly independent in z^*:

$$
M = \begin{pmatrix} \nabla_{xx}L(z^*) & \nabla h(x^*) & \nabla g_{I_{0+}(z^*)}(x^*) & \nabla g_{I_{+0}(z^*)}(x^*) & \nabla_{xy}\phi(z^*) \\ Dh(x^*) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & DH(y^*) \\ Dg_{I_{0+}(z^*)}(x^*) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathcal{I} & 0 \\ 0 & 0 & 0 & 0 & DG_{\mathbb{A}}(y^*) \end{pmatrix},
$$

where $\mathbb{A}$ denotes the index set of active inequalities G_i in y^*. Next step is to cancel the

block rows 3, 5 and 6 together with the last two block columns, thus the index sets

$$
\begin{aligned}
I &= \{1, \ldots, n + \omega, n + \omega + \underline{q} + 1, \ldots, n + \omega + \underline{q} + |I_{0+}|\}, \\
J &= \{1, \ldots, n + \omega + |I_{0+}|\},
\end{aligned}
$$

are introduced. The resulting submatrix $M_{I,J}$ is identical to the matrix $\overline{M}(z^* \mid \overline{I})$ with $\overline{I} = \emptyset$, and therefore Lemma 4.4 guarantees that the rows of $M_{I,J}$ are linearly independent due to LICQ and SOSC for problem (2). The rows of the submatrix

$$
M_{I^c, J^c} = \begin{pmatrix} 0 & DH(y^*) \\ \mathcal{I} & 0 \\ 0 & DG_{\mathbb{A}}(y^*) \end{pmatrix}
$$

are linearly independent due to the LICQ fulfilled by H and G at y (cf. Lemma 4.7(a)). Since $M_{I^c, J} = 0$, Lemma 4.7(a) shows that the MPCC-LICQ holds true.   □

In case of degenerate components, i.e., $I_{00} \neq \emptyset$, MPCC-LICQ can be violated even if SSOSC is assumed. This problem could be avoided by introducing lifting variables (see Theorem 5.5). Considering the weaker MPCC-CPLD condition instead of the MPCC-LICQ, a corresponding theorem can be shown for the degenerate case if certain conditions are fulfilled:

DEFINITION 4.9   ***(Condition ISM)***
Define the matrices $\widetilde{M}(z)$ *and* $\widetilde{V}(z) \in \mathbb{R}^{n+\omega+|I_{0+}(z^*)|+|I_{00}(z^*)| \times |I_{00}(z^*)|}$ *by*

$$
\widetilde{M}(z) := \begin{pmatrix} \nabla_{xx} L(x, \mu, \lambda) & \nabla h(x) & \nabla g(x) & \nabla_{xy}\phi(x,y) \\ Dh(x) & 0 & 0 & 0 \\ Dg_{I_{0+}(z^*)}(x) & 0 & 0 & 0 \\ Dg_{I_{00}(z^*)}(x) & 0 & 0 & 0 \end{pmatrix}, \ \widetilde{V}(z) := \begin{pmatrix} \nabla g_{I_{00}(z^*)}(x) \\ 0 \\ 0 \\ 0 \end{pmatrix}.
$$

Condition ISM is fulfilled in z^* *if for arbitrary index sets* I *and* J *of* $\widetilde{M}$ *and* $j \in I_{00}(z^*)$ *the following holds true: If* $\widetilde{V}_{I,j}(z^*) \notin sp(\widetilde{M}_{I,J}(z^*))$, *then there exists a neighborhood* $N(z^*)$ *of* z^* *such that*

$$
\widetilde{V}_{I,j}(z) \notin sp(\widetilde{M}_{I,J}(z)) \quad \forall z \in N(z^*).
$$

THEOREM 4.10
Let x^* *be a local minimum of the discretized optimal control problem* (2) *for the given parameters* y^* *fulfilling LICQ and SSOSC with the corresponding Lagrange multipliers* λ^* *and* μ^*. *Let the structure of* (4) *be given and the condition ISM hold true in* z^*. *Then MPCC-CPLD is satisfied in* z^* *for the discretized inverse optimal control problem* (3).

Proof:
In the first part of the proof we show that for this problem the positive linear dependence of rows of the matrices $M_\alpha(z)$ and $M_\beta(z)$ (cf. Definition 4.2) implies the linear dependence of corresponding rows of a smaller matrix $\widehat{M}(z)$. Furthermore, we prove that the linear dependence of these rows of $\widehat{M}(z)$ implies linear dependence of the rows of the matrices $M_\alpha(z)$ and $M_\beta(z)$. In the second part of the proof the conditions LICQ, SSOSC and ISM are used to show that linear dependence of rows of $\widehat{M}(z^*)$ implies the existence of

a neighborhood $N(z^*)$ of z^* such that the rows of $\widehat{M}(z)$ are linearly dependend for all $z \in N(z^*)$. The combination of the two parts then shows that MPCC-CPLD is satisfied in z^* for the discretized inverse optimal control problem (3).

The functions used to formulate the MPCC (3) of the discretized inverse optimal control problem according to the structure of the tightened nonlinear program (6) have been introduced at the beginning of the proof of Theorem 4.8.

In consequence, the matrices $M_\alpha(z)$ and $M_\beta(z)$ of Definition 4.2 introducing the MPCC-CPLD have the following structure:

$M_\alpha(z) := D\widetilde{g}(z)$ and $M_\beta(z) :=$

$$
\begin{pmatrix}
\nabla_{xx}L(x,\mu,\lambda) & \nabla h(x) & \nabla g_{I_{0+}(z^*)}(x) & \nabla g_{I_{00}(z^*)}(x) & \nabla g_{I_{+0}(z^*)}(x) & \nabla_{xy}\phi(x,y) \\
Dh(x) & 0 & 0 & 0 & 0 & 0 \\
Dg_{I_{0+}(z^*)}(x) & 0 & 0 & 0 & 0 & 0 \\
Dg_{I_{00}(z^*)}(x) & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathcal{I} & 0 & 0 \\
0 & 0 & 0 & 0 & \mathcal{I} & 0
\end{pmatrix}.
$$

Note that the sign of the lines corresponding to $\widehat{I}_{g+}$ and $\widehat{I}_{g0}$ has been changed to obtain a more standard structure without changing the properties related to the MPCC-CPLD.

Assuming that index sets $I_\alpha \subseteq I_{\widetilde{g}}(z^*)$ and $I_\beta \subseteq I_T(z^*)$ exist such that the rows of the matrices $M_\alpha(z^*)$ and $M_\beta(z^*)$ are positive-linearly dependent with respect to I_α and I_β with vectors $\alpha \geq 0$ and β, it has to be shown that there exists a neighborhood $N(z^*)$ of z^* such that these rows are linearly dependent for all $z \in N(z^*)$ (cf. Definition 4.2).

Two cases can be distinguished: In the first case, the rows of $M_\alpha[I_\alpha](z^*)$ are linearly dependent. Then the structure of the Jacobian (cf. (4))

$$
DG(y) = \begin{pmatrix} -\mathcal{I} & 0 \\ 0 & \mathcal{I} \\ 0 & -\mathcal{I} \end{pmatrix} \tag{8}
$$

implies that, for all z, the rows of $M_\alpha[I_\alpha](z)$ are linearly dependent.

In the second case, which is considered in the remaining proof, the rows of $M_\alpha[I_\alpha](z^*)$ are linearly independent. Introduce the following subsets I_L, I_h, I_{g+}, I_{g0}, $I_{\lambda 0}$ and $I_{\lambda+}$ for referencing the matrix blocks of $M_\beta[I_\beta](z)$ (the arguments are dropped for brevity):

$$
M_\beta[I_\beta](z) = \begin{pmatrix}
\nabla_{xx}L_{I_L} & \nabla h_{I_L} & \nabla g_{I_L, I_{0+}(z^*)} & \nabla g_{I_L, I_{00}(z^*)} & \nabla g_{I_L, I_{+0}(z^*)} & \nabla_{xy}\phi_{I_L} \\
Dh_{I_h} & 0 & 0 & 0 & 0 & 0 \\
Dg_{I_{g+}} & 0 & 0 & 0 & 0 & 0 \\
Dg_{I_{g0}} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \mathcal{I}_{I_{\lambda 0}} & 0 & 0 \\
0 & 0 & 0 & 0 & \mathcal{I}_{I_{\lambda+}} & 0
\end{pmatrix}.
$$

In the following we will show that positive-linearly dependence of the matrices $M_\alpha[I_\alpha](z^*)$ and $M_\beta[I_\beta](z^*)$ implies linear dependence of the rows of a matrix $\widehat{M}(z^*)$ defined in (9) and that linear dependence of the rows of $\widehat{M}(z)$ results in linear dependence of the corresponding rows of $M_\alpha[I_\alpha](z)$ and $M_\beta[I_\beta](z)$. The last part of the proof will then show: if the rows of $\widehat{M}(z^*)$ are linearly dependent, then a neighborhood $N(z^*)$ exists such that $\widehat{M}(z)$ has linear dependent rows for all $z \in N(z^*)$.

Define the following submatrix of $M_\beta[I_\beta](z)$ by:

$$\widehat{M}(z) := \begin{pmatrix} \nabla_{xx}L_{I_L} & \nabla h_{I_L} & \nabla g_{I_L, I_{0+}(z^*)} & \nabla g_{I_L, I_{\lambda 0}^c} & \nabla g_{I_L, I_{\lambda+}^c} & \nabla_{xy}\phi_{I_L, \widehat{I}_y} \\ Dh_{I_h} & 0 & 0 & 0 & 0 & 0 \\ Dg_{I_{g+}} & 0 & 0 & 0 & 0 & 0 \\ Dg_{I_{g0}} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (9)$$

where $\widehat{I}_y$ denotes the maximal set of column indices with $DG[I_\alpha, \widehat{I}_y](z^*) = 0$ (if $I_\alpha = \emptyset$, the index set $\widehat{I}_y$ contains all column indices of $\nabla_{xy}\phi(x, y)$). The matrix $\widehat{M}(z)$ is obtained from $M_\beta[I_\beta](z)$ by removing all columns where the last two block rows have a nonzero entry, all columns where $M_\alpha[I_\alpha](z)$ has a nonzero entry, and the last two block rows.

Then all prerequistists of Lemma 4.7 are fulfilled, and consequently the following two implications hold true: on the one hand, the rows of $\widehat{M}(z^*)$ are linearly dependent if the matrices $M_\alpha[I_\alpha](z^*)$ and $M_\beta[I_\beta](z^*)$ are positive-linearly dependent (cf. Lemma 4.7(a)). And on the other hand, see Lemma 4.7(b), if the rows of $\widehat{M}(z)$ are linearly dependent, the rows of $M_\alpha[I_\alpha](z)$ and $M_\beta[I_\beta](z)$ are linearly dependent, too.

To focus on the relevant structure of $\widehat{M}(z)$, we introduce the following additional index sets:

$$\begin{aligned} \widehat{I}_1 &= (I_{g0}(z^*) \cap I_{\lambda 0}^c(z^*)) \cup I_{g+}, \\ \widehat{I}_2 &= I_{g0}(z^*) \setminus I_{\lambda 0}^c(z^*), \\ \widehat{I}_3 &= (I_{g0}(z^*) \cap I_{\lambda 0}^c(z^*)) \cup I_{0+}(z^*), \\ \widehat{I}_4 &= (I_{\lambda 0}^c(z^*) \setminus I_{g0}(z^*)) \cup I_{\lambda+}^c(z^*), \end{aligned}$$

where the following inclusion holds true: $\widehat{I}_1 \subseteq \widehat{I}_3$. Consequently, the matrix $\widehat{M}(z)$ can be written using the following block structure:

$$\widehat{M}(z) = \left(\begin{array}{ccc|cc} \nabla_{xx}L_{I_L}(x, \mu, \lambda) & \nabla h_{I_L}(x) & \nabla g_{I_L, \widehat{I}_3}(x) & \nabla g_{I_L, \widehat{I}_4}(x) & \nabla_{xy}\phi_{I_L, \widehat{I}_y}(x, y) \\ Dh_{I_h}(x) & 0 & 0 & 0 & 0 \\ Dg_{\widehat{I}_1}(x) & 0 & 0 & 0 & 0 \\ \hline Dg_{\widehat{I}_2}(x) & 0 & 0 & 0 & 0 \end{array} \right).$$

Due to Corollary 4.5 and Corollary 4.6 not only the rows of the top left block are linearly independent in a neighborhood of z^*, but also the rows of the matrix

$$\begin{pmatrix} \nabla_{xx}L_{I_L}(x, \mu, \lambda) & \nabla h_{I_L}(x) & \nabla g_{I_L, \widehat{I}_3}(x) & \nabla g_{I_L, \widehat{I}_2}(x) \\ Dh_{I_h}(x) & 0 & 0 & 0 \\ Dg_{\widehat{I}_1}(x) & 0 & 0 & 0 \\ Dg_{\widehat{I}_2}(x) & 0 & 0 & 0 \end{pmatrix}$$

and there exists subsets $J_1 \subseteq \{1, \ldots, n\}$, $J_2 \subseteq \{1, \ldots, \omega\}$, $J_3 \subseteq \widehat{I}_3$ and $J_4 \subseteq \widehat{I}_2$ of the column indices such that a quadratic matrix with full rank is obtained in a neighborhood

of z^*:

$$
\left(
\begin{array}{ccc|c}
\nabla_{xx} L_{I_L, J_1}(x, \mu, \lambda) & \nabla h_{I_L, J_2}(x) & \nabla g_{I_L, J_3}(x) & \nabla g_{I_L, J_4}(x) \\
Dh_{I_h, J_1}(x) & 0 & 0 & 0 \\
Dg_{\widehat{I}_1, J_1}(x) & 0 & 0 & 0 \\
\hline
Dg_{\widehat{I}_2, J_1}(x) & 0 & 0 & 0
\end{array}
\right)
=:
\left(
\begin{array}{cc}
A(z) & B(z) \\
C(z) & 0
\end{array}
\right).
$$

$$(10)$$

Since there exists an inverse of matrix (10), it follows that the rows of $\widehat{M}(z)$ are linearly independent if and only if for all $i \in \{1, \ldots, |J_4|\}$: $\left((B(z)^T)_i, 0^T \right)^T \in \mathrm{sp}\left(\widehat{M}(z) \right)$.

By the first part of the proof positive linear dependence of arbitrarily chosen rows of the matrices $M_\alpha(z^*)$ and $M_\beta(z^*)$ implies linear dependence of the rows of the corresponding $\widehat{M}(z^*)$. Since $\widehat{M}(z^*)$ is a submatrix of $\widetilde{M}(z^*)$ (cf. Definition 4.9), define the index sets I and J of $\widetilde{M}(z^*)$ such that $\widetilde{M}_{I,J}(z^*) = \widehat{M}(z^*)$. Consequently, the linear dependence of the rows of $\widehat{M}(z^*)$ yields that there exists an index $j \in I_{00}(z^*)$ such that $\widetilde{V}_{I,j}(z^*) \notin \mathrm{sp}(\widetilde{M}_{I,J}(z^*))$. The condition ISM shows then that there exists a neighborhood $N(z^*)$ of z^* such that $\widetilde{V}_{I,j}(z) \notin \mathrm{sp}(\widetilde{M}_{I,J}(z)) \ \forall z \in N(z^*)$. Therefore, the rows of $\widehat{M}(z)$ are linearly dependent for all z in this neighborhood. In consequence, the chosen rows of $M_\alpha(z)$ and $M_\beta(z)$ are linearly dependent for $z \in N(z^*)$ which shows that the MPCC-CPLD hold true. $\qquad\square$

REMARK 4.11 *The proof shows that the MPCC-MFCQ may not hold for problem (3) if degenerate components exist, because one condition of MPCC-MFCQ is that the rows of $M_\beta(z^*)$ are linearly independent: If the number of degenerate components is greater than the number of considered cost functions, i.e., $|I_{00}(z^*)| > k+1$, the matrix $\widehat{M}(z^*)$ has more rows than columns, thus the rows are linearly dependent.*

REMARK 4.12 *Condition ISM holds true in z^* if there exists a neighborhood $N(z^*)$ of z^* such that every submatrix $\widetilde{M}_{I,J}(z)$ has constant rank for all $z \in N(z^*)$. However, this constant rank assumption is so strong that MPCC-CPLD can be proven without demanding LICQ and SSOSC: By the first part of the proof the positive linear dependence results in linear dependence of the rows of $\widehat{M}(z^*)$. The assumption of constant rank then shows that there exists a neighborhood $N(z^*)$ where the rows of $\widehat{M}(z)$ are linearly dependent for all $z \in N(z^*)$. In consequence the MPCC-CPLD holds true.*

5. Numerical Optimization Methods for MPCCs

A large number of numerical optimization methods for MPCC has been studied, see the introduction for some references. A popular approach is the relaxation technique of Scholtes [53] where the kink resulting from the complementarity conditions is globally regularized. In addition, the local relaxation approach of Steffensen and Ulbrich [55] considers only a region around the origin for locally smoothing the kink. Both relaxation methods consequently solve a series of relaxed problems where the update strategies for the relaxation parameters has to be attuned to the theoretical convergence considerations. On the other hand, the lifting approaches of Stein [56] and Hatz et al. [36] add additional variables to the constraints and possibly the cost function to avoid the standard formulation of the complementarity conditions.

5.1 *Relaxation Approach of Scholtes*

The regularization scheme of Scholtes [53] relaxes the complementarity condition to $\widetilde{G}_i(p) \geq 0, \quad \widetilde{H}_i(q) \geq 0, \quad \widetilde{G}_i(p)\widetilde{H}_i(q) \leq \delta_i$ by introducing a parameter $\delta_i > 0$. The feasible region of this regularization approach is displayed in figure 2.
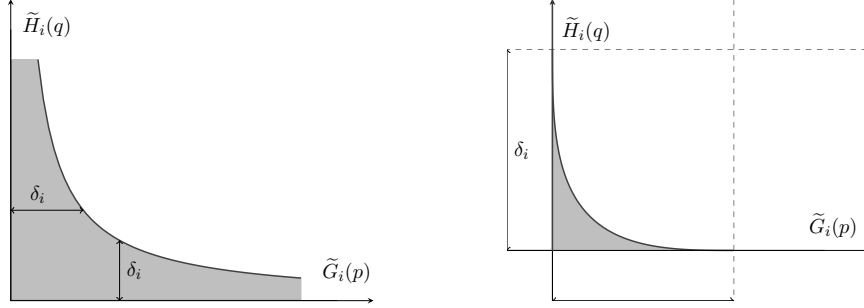

Figure 2. Illustration of the relaxation approach of [53] for one scalar complementarity condition, and of the relaxation approach of [60] for one scalar complementarity condition using the smoothing function β_p.

It is proven in [53] (Theorem 3.1) that under assumptions including the MPCC-LICQ the sequence of stationary points for the relaxed problems converge to a C-stationary point with unique multipliers; B-stationarity or M-stationarity follows if additional assumptions are fulfilled. Furthermore, it is shown that a piecewise smooth mapping relating the relaxation parameter δ_i to the corresponding stationary point exists if suitable assumptions are met. Note that in [39] (Theorem 3.1) it is shown that MPCC-MFCQ suffices to proof that each limit point is C-stationary.

REMARK 5.1 *Keeping in mind that in numerical computations for each relaxation step only an approximate stationary point is computed (up to the given tolerance ϵ), the C-stationarity of the limit point can still be proven if the MPCC-MFCQ holds and $\epsilon_i = o(\delta_i)$, confer Theorem 3.2 in [41].*

5.2 *Relaxation Approach of Steffensen and Ulbrich*

The basic idea of the relaxation scheme of Steffensen and Ulbrich [55, 60], which can be seen as a combination of the relaxation approach of [53] and the regularization approach of [29], is to relax the complementarity condition for each pair $\widetilde{G}_i(p)$ and $\widetilde{H}_i(q)$, $i = 1, \ldots, \widetilde{s}$, only on a subset of the triangle with the vertices $(0,0)$, $(\delta_i, 0)$ and $(0, \delta_i)$. Note that for a sufficiently small relaxation parameter $\delta_i > 0$ the complementarity conditions are only modified for degenerate components, i.e., $\widetilde{G}_i(p) = \widetilde{H}_i(q) = 0$, and for the choice of $\delta_i = 0$ the original complementarity condition is obtained. Since the approach is independent for each of the scalar complementarity conditions, it is sufficient to discuss how to handle one of them and therefore, the range of the index i is not mentioned at each instance in the following. The presentation follows closely the one of [55, 60].

A reparametrization of the complementarity problem $\widetilde{G}_i(p) \geq 0$, $\widetilde{H}_i(q) \geq 0$, $\widetilde{G}_i(p)\widetilde{H}_i(q) = 0$ into the problem $\underline{z} = |\overline{z}|$ by introducing $\underline{z} := \widetilde{G}_i(p) + \widetilde{H}_i(q)$ and $\overline{z} := \widetilde{G}_i(p) - \widetilde{H}_i(q)$ allows to post the relaxation problem as a problem of smoothing the absolute value function within the interval $[-\delta_i, \delta_i]$. Consequently, one is interested in smoothing the kink of the absolute value function. A function $\beta : I \to \mathbb{R}$ defined on an open interval I with $[-1, 1] \subset I \subset \mathbb{R}$ is called an kink smoothing function if the following

conditions hold: $\beta|_{[-1,1]} \in \mathcal{C}^2([-1,1], \mathbb{R})$, $\beta(-1) = \beta(1) = 1$, $\beta'(-1) = -1$, $\beta'(1) = 1$, $\beta''(-1) = \beta''(1) = 0$, and $\beta''(\overline{z}) > 0 \ \ \forall \overline{z} \in (-1,1)$.

Two of such kink smoothing functions are introduced in [60]:

$$\beta_s(\overline{z}) := \frac{2}{\pi} \sin\left(\frac{(\overline{z} + 3)\pi}{2}\right) + 1 \ \ \text{and} \ \ \beta_p(\overline{z}) := \frac{1}{8}\left(-\overline{z}^4 + 6\overline{z}^2 + 3\right).$$

Using an kink smoothing function for the scaled interval $[-\delta_i, \delta_i]$ and the absolute value function for the complement, the following function $\xi \in \mathcal{C}^2(\mathbb{R} \times \mathbb{R}_{\geq 0}, \mathbb{R})$ with

$$\xi(\overline{z}, \delta_i) := \begin{cases} |\overline{z}| & \text{for} & |\overline{z}| \geq \delta_i, \\ \\ \delta_i \beta\left(\delta_i^{-1}\overline{z}\right) & \text{for} & |\overline{z}| < \delta_i, \end{cases}$$

is obtained and can be used to write a relaxed version of the complementarity condition $\underline{z} = |\overline{z}|$: $\underline{z} \geq -\overline{z}$, $\underline{z} \geq \overline{z}$, $\underline{z} \leq \xi(\overline{z}, \delta_i)$. Switching back to the original variables $\widetilde{G}_i(p)$ and $\widetilde{H}_i(q)$, the function $\Xi_i \in \mathcal{C}^2(\mathbb{R} \times \mathbb{R} \times \mathbb{R}_{\geq 0}, \mathbb{R})$ defined by

$$\Xi_i(\widetilde{G}_i(p), \widetilde{H}_i(q), \delta_i) := \widetilde{G}_i(p) + \widetilde{H}_i(q) - \xi(\widetilde{G}_i(p) - \widetilde{H}_i(q), \delta_i)$$

allows to state the complementarity condition for $\widetilde{G}_i(p)$ and $\widetilde{H}_i(q)$ in the form $\widetilde{G}_i(p) \geq 0$, $\widetilde{H}_i(q) \geq 0$, $\Xi_i(\widetilde{G}_i(p), \widetilde{H}_i(q), \delta_i) \leq 0$. Combining the individual functions Ξ_i for all $i = 1, \ldots, \widetilde{s}$, the definition of the vector-valued function $\Xi \in \mathcal{C}^2(\mathbb{R}^{\widetilde{s}} \times \mathbb{R}^{\widetilde{s}} \times \mathbb{R}^{\widetilde{s}}_{\geq 0}, \mathbb{R}^{\widetilde{s}})$ is straightforward and results in $\widetilde{G}(p) \geq 0$, $\widetilde{H}(q) \geq 0$, $\Xi(\widetilde{G}(p), \widetilde{H}(q), \delta) \leq 0$. Note that the vector $\delta \in \mathbb{R}^{\widetilde{s}}_{\geq 0}$ allows to individually determine a suitable relaxation for each scalar complementarity condition.

In consequence, the following parametric nonlinear optimization problem $\mathcal{R}(\delta)$ is obtained:

DEFINITION 5.2   **(Relaxed Problem $\mathcal{R}(\delta)$)**

$$\min_z \widetilde{\phi}(z)$$

$$\text{subject to} \quad \widetilde{h}(z) = 0, \ \ \widetilde{g}(z) \leq 0,$$
$$\widetilde{G}(p) \geq 0, \ \ \widetilde{H}(q) \geq 0, \ \ \Xi(\widetilde{G}(p), \widetilde{H}(q), \delta) \leq 0.$$

By construction the relaxation properties of the scheme are evident: If a variable z^* is feasible for the original MPCC problem (5), then it also feasible for the relaxed problem 5.2. Denote the feasible set of $\mathcal{R}(\delta)$ by if the inequality $\delta^{II} \leq \delta^{I}$ holds componentwise, then the following inclusion for the feasible set results: $\mathbb{X}_{\mathcal{R}}(\delta^{II}) \subseteq \mathbb{X}_{\mathcal{R}}(\delta^{I})$. Furthermore, if a strict local solution z^* of $\mathcal{R}(\delta^{I})$ is feasible for the MPCC problem (5), then z^* is a strict local solution for all δ^{II} with $0 \leq \delta^{II} \leq \delta^{I}$. For more details and the proofs of these properties see [60].

THEOREM 5.3   **(Convergence to C-stationary Point)**
Let $(\delta^{(i)}) \subset \mathbb{R}^+$ be a sequence satisfying $\delta^{(i)} \to 0$ and let $(z^{(i)}) \subset \mathbb{R}^{\widetilde{n}}$ be a sequence of stationary points of $\mathcal{R}(\delta^{(i)})$ that satisfies $z^{(i)} \to z^*$. If MPCC-CPLD holds in z^*, then z^* is a C-stationary point of the MPCC (5).

Proof:

If the MPCC-CPLD is replaced by the slightly more restrictive MPCC-CRCQ (constant rank constraint qualification for MPCCs), a proof can be found in [55] (cf. Theorem 5.2 in [55]). As noted in section 3.2 in [38], the proof has not to be modified for the MPCC-CPLD assumption, because the essential step of the proof uses a set being by construction not only linearly dependent but positive-linearly dependent and the only drawn deduction is the linear dependency for all elements in the neighborhood, which exactly matches the MPCC-CPLD. Consequently, confer the proof in [55]. □

Since MPCC-MFCQ implies MPCC-CPLD, this theorem shows the convergence of stationary points for the relaxed problems to a C-stationary point for the locomotion problem.

REMARK 5.4 *In general, only approximate stationary points can be obtained (up to the given tolerance ϵ) with numerical computations and in consequence, the convergence to a C-stationary point might be lost and then only weak stationarity can be proven if the MPCC-MFCQ holds (cf. Theorem 5.3 in [41]). In order to also guarantee C-stationarity for the approximate case additional conditions considerably restricting the feasible region of the complementarity condition have to be introduced (for details see [41]).*

The relaxation scheme of Steffensen and Ulbrich can result in problems where no sequence of proper interior points of the feasible region converging to a stationary point exists. Therefore, a two-sided relaxation variant is introduce in [60] for interior point methods. However, whether the inequalities resulting from the complementarity conditions are realized via barrier terms or penalty terms highly depends on the realization of the interior-point methods. Where in the first case the two-sided approach is indispensable, in the second case the numerical tolerance ϵ might already make this advanced approach dispensable.

5.3   *Lifting Approach of Stein*

The lifting approach of Stein [56] is based on the construction of a smooth set in $\mathbb{R}^3$ for each complementarity condition normally stated in $\mathbb{R}^2$ such that the projection of this set on to $\mathbb{R}^2$ coincides with the typical complementarity set with the kink $\mathbb{L} := \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 \perp x_2\}$.

The basis element for the construction is a at least continuously differentiable function $s : \mathbb{R} \to \mathbb{R}$ being surjective from $\mathbb{R}_+$ to $\mathbb{R}_+$ fulfilling $s(x) = 0 \quad \forall x \leq 0, s(x) > 0 \quad \forall x > 0$. Possible candidates are $s(x) = (\max\{0, x\})^p$ with $p \in \mathbb{N}$ and $p > 1$. This can be used to introduce the function $\psi : \mathbb{R}^3 \to \mathbb{R}^2$ by

$$\psi(x_1, x_2, x_3) := \begin{pmatrix} x_1 - s(x_3) \\ x_2 - s(-x_3) \end{pmatrix}$$

and the corresponding zero set $\mathbb{S} := \{x \in \mathbb{R}^3 \mid \psi(x) = 0\}$. Note that $\mathbb{S}$ is a smooth curve in $\mathbb{R}^3$ and the project of $\mathbb{S}$ on the first two components equals $\mathbb{L}$ (cf. Lemma 2.1 in [56]). If the function ψ is used to lift the complementarity conditions of an MPCC by introducing the corresponding number of additional variables, the minimizers of the MPCC are the first components of minimizers of the lifted problem. It can be shown that a point is stationary for the lifted problem if and only if the first components are weakly stationary for MPCC (cf. Proposition 4.1 in [56]).

In order to proof a stronger concept than weak stationarity, the lifted problem is regu-

larized by coupling the lifting variables to the cost function and a corresponding stability concept is introduced, the tilting stability (cf. Definition 5.1 in [56]). This concept is used to relate stationary points of the lifted problem to non-degenerate C-stationary points of MPCC, see [56] for more details.

5.4　*Lifting Approach of Hatz et al.*

Considering the complementarity constraint of the discretized inverse optimal control problem (3) the lifting approach of Hatz et al. [35, 36] introduces the additional vector of variables $v \in \mathbb{R}^\nu$ to modify the left-hand term: $v \leq -g(x) \perp \lambda \geq 0$. Using a penalty parameter $\chi > 0$ and a penalty function P, e.g., $P(v) = ||v||_1$ or $P(v) = ||v||_2^2$, the following modified cost function is considered: $\Phi(x, y \mid \Lambda) + \chi P(v)$. The main advantage of the lifted MPCC is the following theorem proven in [35, 36] (cf. Theorem 2.1 in [36]):

THEOREM 5.5
If at a given point both LICQ and SOSC are fulfilled for the discretized optimal control problem with the modified inequality constraints $v \leq -g(x)$, then the lifted MPCC satisfies MPCC-LICQ at this point.

This theorem guarantees the setting used by [29] to proof convergence results for SQP methods applied to MPCCs under relatively mild conditions.

However, the choice of the penalty function poses some problems: if the exact penalty function $P(v) = ||v||_1$ is used, a nonsmooth problem is obtained. This could be avoided by constraining the sign of the lifting variables v by $v \leq 0$, but then the theorem 5.5 does not hold in full generality anymore, because constellations with $v_i = g_i(x) = \lambda_i = 0$ violated the MPCC-LICQ. On the other hand, using the inexact penalty function $P(v) = ||v||_2^2$ results in a smooth optimization problem, but the penalty parameter χ has to tend to infinity to obtain convergence of the lifted problem toward the original MPCC. However, for increasing penalty parameters the condition of the Hessian worsens and consequently, further problem reformulations might be needed as known from standard nonlinear penalty problems (cf. [49]).

6.　Numerical Results

In this section numerical results for locomotion problems are stated considering the relaxation scheme of Steffensen and Ulbrich and the lifting approaches of Stein and Hatz et al. and comparing them to obtained results for the non-relaxed and non-lifted problem. In addition to the problem properties specified in the sections 2 and 3, further details are now provided on the chosen time discretization, the boundary conditions of the locomotion tasks and the set of basic cost functions. Additionally, the numerical optimization method used to solve the resulting nonlinear optimization problems has to be specified and the data generation has to be discussed.

6.1　*Time Discretization*

The discretization of the time interval has to be sufficiently fine to assure that the solution of the discretized optimal control problem comes close to the solution of the original optimal control problem. On the other hand using a fine uniform discretization increases the problem size significantly and consequently, a strategy is needed that assures

a reasonable trade-off between the two.

We use a static adaptation strategy for the optimal control problem which updates the time discretization after having solved the discretized problem for the current discretization. Other discretization strategies are known where the time discretization is updated during the optimization of an optimal control problem (e.g., [61]), but this introduces further nonlinearities and additional conditions have to be added to the discretized optimal control problem.

Following the line of [11], a (relative) local discretization error is used to formulate the adaptation strategy for the time discretization. The goal is to refine a given discretization by subdividing the intervals with a large relative local discretization error. The most simple approach is to bisect these intervals, but, if one wants to add more than one intermediate discretization point, the error reduction has to be predicted. Consequently, the type of discretization strategy influences the choice of the time discretization. Using the order of consistency for a given Runge-Kutta method, the iterative adaptation strategy of [11] allows to divide intervals into multiple subintervals such that the maximal discretization error is minimized.

Here we use this adaptation strategy for solving the optimal control problems (1), however we assume for the discretized inverse optimal control problem (3) that the specified time grid is fixed and suitable for the corresponding optimal control problem. The adaptation strategy could also be applied to the inverse optimal control problem, however this will result in different time grids for each solution approach of the corresponding MPCC and thus make comparisons of these approaches significantly harder.

6.2   *Basic Cost Functions*

The following basic cost functions ϕ_i are chosen in the numerical examples, where the integrals of the definitions in section 3.2 are replaced by suitable approximations using the discretized state and control values (this is indicated by the superscript a):
$\phi_1(x) := f_{u,1}^a(x) + f_{u,2}^a(x)$, $\phi_2(x, \underline{v}) := f_{ref,4}^a(x, \underline{v})$, $\phi_3(x) := f_{time}^a(x)$, $\phi_4(x) := f_{line}^a(x)$, $\phi_5(x) := f_{jerk}^a(x)$. Consequently, $w^T := (w_1, \ldots, w_5)$ and $\pi := \underline{v}$.

6.3   *Interior-Point Method IPOPT*

The interior-point method `IPOPT` of Wächter and Biegler [64] is used here to solve the nonlinear optimization problems arising in the examples. The basic idea of interior-point methods is to solve a sequence of barrier problems whose solutions define the so-called central path. The goal is to stay in a suitable neighborhood of the central path and successively reduce to barrier parameter. `IPOPT` is based an a combination of outer iterations modifying the barrier parameter and inner iterations for solving the corresponding barrier problem up to a suitable tolerance. In addition, a filter approach is used to determine whether new iterates guarantee sufficient decrease in the cost function or the violation of the constraints in order to proof local and global convergence properties of the method [62, 63]. An implementation of the algorithm is available under an open source license.

6.4   *Data Generation*

The focus of the numerical examples presented here is on the comparison of different approaches to handle the MPCC-structure of the discretized inverse optimal control

problem (3), therefore we consider here only synthetic data, i.e., the data is computed by solving the optimal control problem (1) for a given vector y. To avoid possible numerical artifacts resulting from the perfect fit of the model and the computed data, a certain amount of Gaussian white noise is added to the computed data. This approach allows to avoid model errors and the obtained results are solely based on the problem structure, the solution technique and the added noise level.

If data of recorded human motions is to be considered, differences between the model of the dynamics and the true dynamics of the human body are obvious. To handle the resulting modeling errors, the data needs to be processed, e.g., the velocity profiles of recorded human motions exhibit characteristic peaks where the individual steps are made, but the unicycle model generates smooth velocity profiles, cf., [3]. The most important consequence is that specially tailored distance measures are needed. Thus all these technicalities needed for recorded data are avoided here to concentrate on the basic mathematical problem. Please, confer the publications [1, 3] for inversion results of human data.

6.5   *Three Motions Example*

The considered locomotion example combines three motions from a common starting point to three different goal positions. Note that the structure of the discretized inverse optimal control problem (3) when instead of a single motion multiple motions that share a common y are considered. The corresponding boundary conditions are the following ones:

$\mathcal{P}_{Px}(t_0)$	$\mathcal{P}_{Py}(t_0)$	$\beta_P(t_0)$	$v_P(t_0)$	$\mathcal{P}_{Px}(t_f)$	$\mathcal{P}_{Py}(t_f)$	$\beta_P(t_f)$	$v_P(t_f)$
0	0	0	0.5	0	4	0.5π	0.5
0	0	0	0.5	1	2	π	0.5
0	0	0	0.5	1	3	0	0.5

Confer Figure 3 for motions fulfilling these boundary conditions for a given choice of cost function. For the three motions common upper and lower bounds are introduced for all state variables. However, only the bounds on the forward velocity will become active and the others simply increase the performance of the numerical methods by specifing a suitable range for these values:

$\mathcal{P}_{Px}(t)$	$\mathcal{P}_{Py}(t)$	$\beta_P(t)$	$v_P(t)$	$\beta_P{}'(t)$	$v_P{}'(t)$	$\beta_P{}''(t)$	t
5	5	4	0.55	3	3	3	20
-5	-5	-4	0.20	-3	-3	-3	0.1

In this example we use the values $w^T := (0.1, 0.05, 0.3, 0.2, 0.35)$ and $\pi = \underline{v} = 0.8$ for the data generation and add to the solution of the corresponding optimal control problem Gaussian white noise with a standard deviation of 10^{-4}, which corresponds to an accuracy in the measurement in the range of a millimeter.

The velocity profiles clearly show that the upper constraint on the forward velocity becomes active for each of the three motions. Note that the time grids are not chosen or adapted in such a way that the time instances of the continuous solution of the optimal control problem where the constraint becomes active or inactive are part of the grid, i.e., the inequality constraint might be violated in the interior of the intervals and in most cases problems with strict complementarity will result.
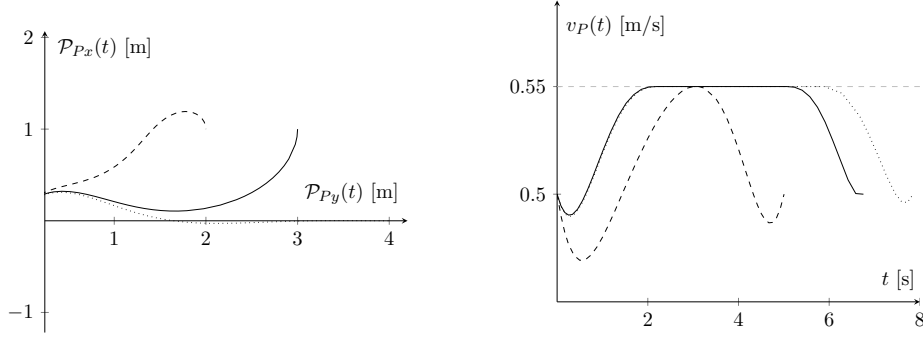
Figure 3. Trajectories and velocity profiles for the three data motions.

In order to achieve a good trade-off between problem sizes and integration accuracy, the time grid of the optimal control problem for the data generation is non-uniformly refined, cf. section 6.1. However, the adapted time grids are then fixed for the inversion part to avoid introducing a further source of measurement errors due to differing grids. The grids displayed in figure 4 are used for this example (each has 40 subintervals).
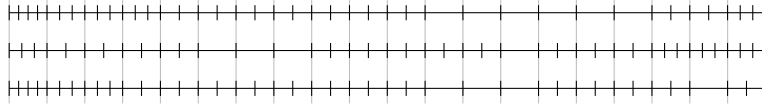

Figure 4. Refined time-meshes for the three individual motions.

The starting values for the inversion are generated by solving the optimal control problem for random w and π values, i.e., uniformly distributed $w_i \in [0, 1]$ with $\sum_i w_i = 1$ and $\underline{v} \in [0.7, 0.9]$. The considered upper and lower bounds for $\underline{v}$ in the optimization are 0.5 and 1.0, i.e., the generated starting values are feasible.
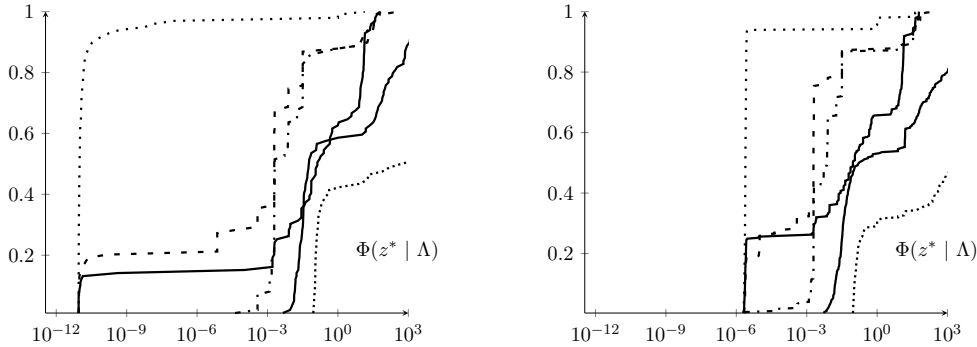

Figure 5. Performance profile with respect to value of cost function for various starting values for data with (right) and without (left) noise

Numerical results for these multiple starting values are displayed in the following using performance profiles of the Dolan-Moré-type [19]: The x-axes gives the values of some measure to compare the obtained solution to the generated data and the y-axis shows the relative number of starting values which resulted in value smaller or equal than the corresponding x-value.

The three figures (Figures 5 to 7) display results obtained by the interior-point method

IPOPT [64] considering data without noise (left plots) and with a noise level of 10^{-3} which results in an optimal value for the distance measure about 10^{-6} (right ones). The following color code was used for the six tested approaches: the non-modified problem —— , the lifting approach of Hatz with l_2-penalty term —— and with l_1-penalty term ·····, the lifting approach of Stein with exponent 2 - - and with exponent 3 - · -, and the relaxation approach of Steffensen and Ulbrich · · · ·.

The result show that about twenty percent of the starting values yield an optimal reconstruction result with respect to the distance measure for the non-modified problem, whereas it seems to be a rather hard problem for the other starting values. The lifting approach of Stein with exponent 2 seem to considerably increase the performance compared to non-modified version and outperforms the lifting with exponent 3 considerably. On the other hand, the lifting approach of Hatz et al. results in scaling problems for the interior-point method IPOPT and yields stationary points for the discretized inverse optimal control problem that are rather different from the values used to generate the data. The relaxation approach of Steffensen and Ulbrich yields the best reconstruction results for about 90 percent of the random starting values.
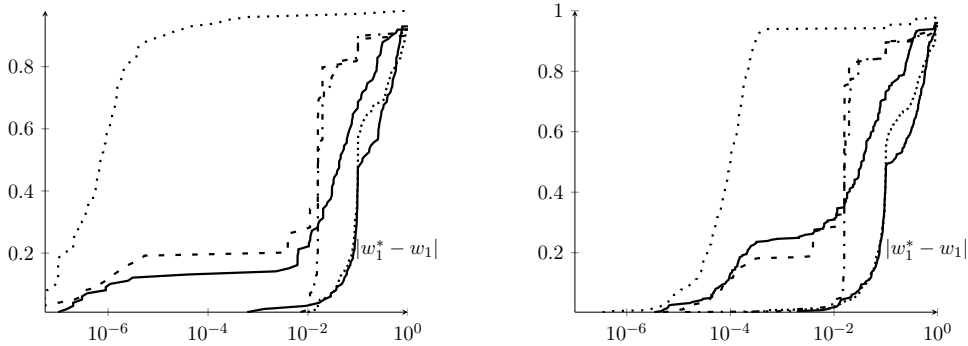

Figure 6. Performance profile with respect to value of the weighting factor w_1 for various starting values for data with (right) and without (left) noise
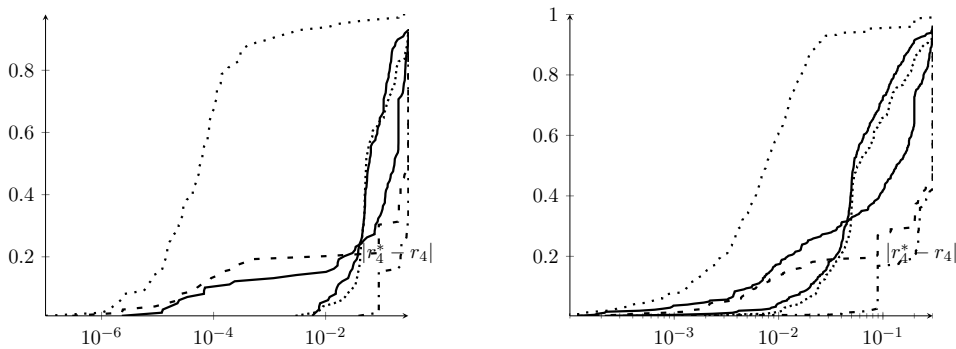

Figure 7. Performance profile with respect to value of the nonlinear parameter r_4 for various starting values for data with (right) and without (left) noise

Figure 6 shows values of the distance measure $\Phi(z^* \mid \Lambda)$ directly relate to the obtained weight distributions y^*; here the weight w_1 of the first cost function is chosen, but the other weight show similar behavior. The differences displayed in Figure 7 indicate that the nonlinear parameter r_4 is matched best by the relaxation approach of Steffensen and Ulbrich, whereas the lifting approaches yields larger deviations. Naturally, a misfit in this

nonlinear parameter results in considerable errors with respect to $\Phi(z^* \mid \Lambda)$. However, if this parameter is fixed, the performance profiles are not that different from the case with an unknown parameter (cf. left plot in Figure 6.5). If the upper bound on the velocity is considerably increased such that the bound does not become active for the three motions, i.e., no active inequality constraints, then the structure of the reconstruction problem simplifies and the advantage of the relaxation formulation compared to the non-modified problem reduces considerably.
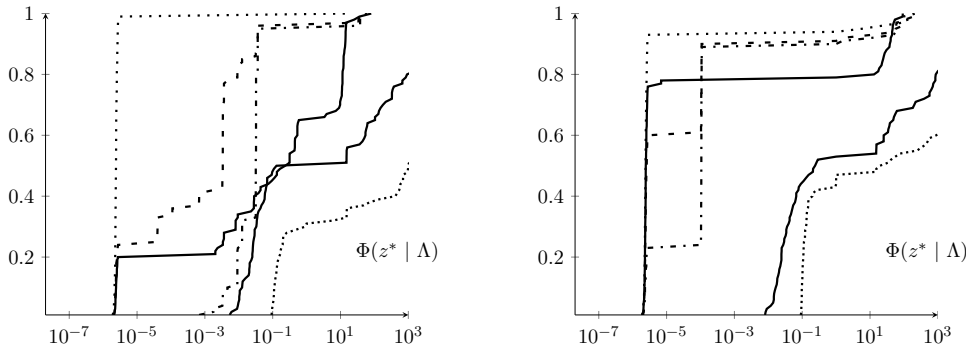

Figure 8. Performance profile with respect to value of cost function for various starting values for data with noise. The left plot shows results for a fixed nonlinear parameter r_4 and for the right plot he upper bound v_{max} is additionally increased.

References

[1] S. Albrecht, *Modeling and numerical solution of inverse optimal control problems for the analysis of human motions*, Dissertation, Fakultät für Mathematik, Technische Universität München (2013).

[2] S. Albrecht, M. Leibold, and M. Ulbrich, *A bilevel optimization approach to obtain optimal cost functions for human arm movements*, Numerical Algebra, Control and Optimization 2 (2012), pp. 105–127.

[3] S. Albrecht, P. Basili, S. Glasauer, M. Leibold, and M. Ulbrich, *Modeling and analysis of human navigation with crossing interferer using inverse optimal control*, in *Proceedings of the Vienna International Conference on Mathematical Modeling*, 2012.

[4] G. Anandalingam and T. Friesz (eds.), *Hierarchical Optimization*, Annals of Operations Research, Vol. 34, Springer, 1992.

[5] F. Anderson and M. Pandy, *Dynamic optimization of human walking*, Biomechanical Engineering 123 (2001), pp. 381–390.

[6] M. Anitescu, *On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints*, SIAM Journal on Optimization 15 (2005), pp. 1203–1236.

[7] U. Ascher, R. Mattheij, and R. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*, Society for Industrial and Applied Mathematics (SIAM), 1995.

[8] J. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Springer, 1998.

[9] T. Başar and G. Olsder, *Dynamic Noncooperative Game Theory*, Society for Industrial and Applied Mathematics (SIAM), 1999.

[10] B. Berret, E. Chiovetto, F. Nori, and T. Pozzo, *Evidence for composite cost functions in arm movement planning: an inverse optimal control approach*, PLoS Computational Biology 7 (2011), pp. e1002183(1–18).

[11] J. Betts, *Practical Methods for Optimal Control and Estimation using Nonlinear Programming*, Society for Industrial and Applied Mathematics (SIAM), 2010.

[12] C. Bottasso, B. Prilutsky, A. Croce, E. Imberti, and S. Sartirana, *A numerical procedure for inferring from experimental data the optimization cost functions using a multibody model of the neuro-musculoskeletal system*, Multibody System Dynamics 16 (2006), pp. 123–154.

[13] A. Bryson and Y.C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*, Hemisphere Publishing Corporation, 1975.

[14] R. Bulirsch, *Die Mehrzielmethode zur numerischen Lösung von nichtlinearen Randwertproblemen und Aufgaben der optimalen Steuerung*, Report der Carl-Cranz-Gesellschaft (1971).

[15] L. Cesari, *Optimization - Theory and Applications*, Springer, 1983.

[16] C. Chen and J. Cruz, *Stackelberg solution for two-person games with biased information patterns*, IEEE Transactions on Automatic Control 17 (1972), pp. 791–798.

[17] V. DeMiguel, M. Friedlander, F. Nogales, and S. Scholtes, *A two-sided relaxation scheme for mathematical programs with equilibrium constraints*, SIAM Journal on Optimization 16 (2005), pp. 587–609.

[18] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, 2002.

[19] E. Dolan and J. Moré;, *Benchmarking optimization software with performance profiles*, Mathematical Programming 91 (2002), p. 201213.

[20] S. Engelbrecht, *Minimum principles in motor control*, Mathematical Psychology 45 (2001), pp. 497–542.

[21] F. Fisch, *Development of a framework for the solution of high-fidelity trajectory optimization problems and bilevel optimal control problems*, Dissertation, Lehrstuhl für Flugsystemdynamik, Technische Universität München (2011).

[22] P. Fitts, *The information capacity of the human motor system in controlling the amplitude of movement*, Experimental Psychology 47 (1954), pp. 381–391.

[23] T. Flash and N. Hogan, *The coordination of arm movements: an experimentally confirmed mathematical model*, Neuroscience 5 (1985), pp. 1688–1703.

[24] M. Flegel and C. Kanzow, *Abadie-type constraint qualification for mathematical programs with equilibrium constraints*, Optimization Theory and Applications 124 (2005), pp. 595–614.

[25] M. Flegel and C. Kanzow, *On M-stationary points for mathematical programs with equilibrium constraints*, Mathematical Analysis and Applications 310 (2005), pp. 286–302.

[26] M. Flegel and C. Kanzow, *On the Guignard constraint qualification for mathematical programs with equilibrium constraints*, Optimization 54 (2005), pp. 517–534.

[27] R. Fletcher and S. Leyffer, *Solving mathematical programs with complementarity constraints as nonlinear programs*, Optimization Methods and Software 19 (2004), pp. 15–40.

[28] R. Fletcher, N. Gould, S. Leyffer, P. Toint, and A. Wächter, *Global convergence of trust-region SQP-filter algorithms for general nonlinear programming*, SIAM Journal on Optimization 13 (2002), pp. 635–659.

[29] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local convergence of SQP methods for mathematical programs with equilibrium constraints*, SIAM Journal on Optimization 17 (2007), pp. 259–286.

[30] A. Forsgren, P. Gill, and M. Wright, *Interior methods for nonlinear optimization*, SIAM Review 44 (2002), pp. 525–597.

[31] P. Gill, W. Murray, and M. Saunders, *Snopt: An SQP algorithm for large-scale constrained optimization*, SIAM Review 47 (2005), pp. 99–131.

[32] W. Grimm and A. Markl, *Adjoint estimation from a direct multiple shooting method*, Optimization Theory and Applications 92 (1997), pp. 263–283.

[33] W. Hager, *Runge-Kutta methods in optimal control and the transformed adjoint system*, Numerische Mathematik 87 (2000), pp. 247–282.

[34] C. Harris and D. Wolpert, *Signal-dependent noise determines motor planning*, Nature 394 (1998), pp. 780–784.

[35] K. Hatz, *Efficient numerical methods for hierarchical dynamic optimization with application to cerebral palsy gait modeling*, Dissertation, Naturwissenschaftlich-Mathematische Gesamtfakultät, Ruprecht-Karls-Universität Heidelberg (2014).

[36] K. Hatz, S. Leyffer, J. Schlöder, and H. Bock, *Regularizing bilevel nonlinear programs by lifting*, Tech. Rep., Technical Report ANL/MCS-P4076-0613, Argonne National Laboratory, Mathematics and Computer Science Division, 2013.

[37] H. Hatze and J. Buys, *Energy-optimal controls in the mammalian neuromuscular system*, Biological Cybernetics 27 (1977), pp. 9–20.

[38] T. Hoheisel, C. Kanzow, and A. Schwartz, *Convergence of a local regularization approach for mathematical programmes with complementarity or vanishing constraints*, Optimization Methods and Software 27 (2012), pp. 483–512.

[39] T. Hoheisel, C. Kanzow, and A. Schwartz, *Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints*, Mathematical Programming 137 (2013), pp. 257–288.

[40] A. Izmailov and M. Solodov, *Karush-kuhn-tucker systems: regularity conditions, error bounds and a class of newton-type methods*, Mathematical Programming 95 (2003), pp. 631–650.

[41] C. Kanzow and A. Schwartz, *The price of inexactness: Convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited*, Mathematics of Operations Research (2014).

[42] M. Knauer and C. Büskens, *Hybrid solution methods for bilevel optimal control problems with time dependent coupling*, in *Recent Advances in Optimization and its Applications in Engineering*, M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, eds., Springer, 2010, pp. 237–246.

[43] R. Leigh and D. Zee, *The Neurology of Eye Movements*, Oxford University Press, 1999.

[44] F. Lewis and V. Syrmos, *Optimal Control*, Wiley-Interscience, 1995.

[45] K. Malanowski, C. Büskens, and H. Maurer, *Convergence of approximations to nonlinear optimal control problems*, in *Mathematical programming with data perturbations*, A. Fiacco, ed., Lecture Notes in Pure and Applied Mathematics, Vol. 195, Dekker, 1997, pp. 253–284.

[46] A. Migdalas, P. Pardalos, and P. Värbrand (eds.), *Multilevel Optimization: Algorithms and Applications*, Kluwer Academic Publishers, 1998.

[47] K. Mombaur, A. Truong, and J. Laumond, *From human to humanoid locomotion: an inverse optimal control approach*, Autonomous Robots 28 (2010), pp. 369–383.

[48] D. Morrison, J. Riley, and J. Zancanaro, *Multiple shooting method for two-point boundary value problems*, Communications of the ACM 5 (1962), pp. 613–614.

[49] J. Nocedal and S. Wright, *Numerical Optimization*, Springer, 1999.

[50] J. Pang and M. Fukushima, *Complementarity constraint qualifications and simplified b-stationarity conditions for mathematical programs with equilibrium constraints*, Computational Optimization and Applications 13 (1999), pp. 111–136.

[51] L. Qi and Z. Wei, *On the constant positive linear dependence condition and its application to SQP methods*, SIAM Journal on Optimization 10 (2000), pp. 963–981.

[52] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, Mathematics of Operations Research 25 (2000), pp. 1–22.

[53] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM Journal on Optimization 11 (2001), pp. 918–936.

[54] J. Speyer and D. Jacobson, *Primer on Optimal Control Theory*, Society for Industrial and Applied Mathematics (SIAM), 2010.

[55] S. Steffensen and M. Ulbrich, *A new relaxation scheme for mathematical programs with equilibrium constraints*, SIAM Journal on Optimization 20 (2010), pp. 2504–2539.

[56] O. Stein, *Lifting mathematical programs with complementarity constraints*, Mathematical Programming 131 (2012), pp. 71–94.

[57] J. Troutman, *Variational Calculus and Optimal Control*, Springer, 1996.

[58] M. Ulbrich, S. Ulbrich, and L. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Mathematical Programming 100 (2004), pp. 379–410.

[59] Y. Uno, M. Kawato, and R. Suzuki, *Formation and control of optimal trajectory in human multijoint arm movement*, Biological Cybernetics 61 (1989), pp. 89–101.

[60] S. Veelken, *A new relaxation scheme for mathematical programs with equilibrium constraints: Theory and numerical experience*, Dissertation, Fakultät für Mathematik, Technische Universität München (2009).

[61] O. von Stryk, *Numerische Lösung optimaler Steuerungsprobleme: Diskretisierung, Parameteroptimierung und Berechnung der adjungierten Variablen*, in *Fortschritt-Berichte VDI*, no. 441 in Reihe 8, VDI-Verlag, 1995.

[62] A. Wächter and L. Biegler, *Line search filter methods for nonlinear programming: Local convergence*, SIAM Journal on Optimization 16 (2006), pp. 32–48.

[63] A. Wächter and L. Biegler, *Line search filter methods for nonlinear programming: Motivation and global convergence*, SIAM Journal on Optimization 16 (2006), pp. 1–31.

[64] A. Wächter and L. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Mathematical Programming 106 (2006), pp. 25–57.

[65] J. Ye, *Necessary conditions for bilevel dynamic optimization problems*, SIAM Journal on Control and Optimization 33 (1995), pp. 1208–1223.

[66] J. Ye, *Optimal strategies for bilevel dynamic problems*, SIAM Journal on Control and Optimization 35 (1997), pp. 512–531.

[67] J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, Mathematical Analysis and Applications 307 (2005), pp. 350–369.