Pure and Applied Functional Analysis
Volume 6, Number 6, 2021, 1383–1414

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

ERROR ANALYSIS FOR THE IMPLICIT EULER DISCRETIZATION OF AFFINE OPTIMAL CONTROL PROBLEMS WITH INDEX TWO DAES

BJÖRN MARTENS AND CHRISTOPHER SCHNEIDER

Abstract. In this paper, we investigate the implicit Euler discretization of optimal control problems subject to index two differential algebraic equations and derive error estimates between the continuous and discrete solution. Herein, the cost functional and the dynamic are nonlinear with respect to the differential state and linear with respect to the algebraic state and control. Thus, classic second order sufficient conditions are not satisfied and the optimal control is discontinuous, i.e., convergence in the L_∞-norm cannot be expected. Furthermore, there is a discrepancy between the continuous and discrete necessary conditions. In order to derive first order error estimates with respect to the L_1-norm, we take the following steps: First, we derive discrete optimality conditions, which are consistent with the continuous necessary conditions. Then, using a growth property for the switching function at its zeros and a lower bound condition for the second derivatives of the Hamilton function, we show that the continuous KKT-conditions are strongly metrically sub-regular. Next, we prove that the discretized problem has a solution, which can get arbitrarily close to the continuous solution for decreasing mesh size. Finally, we estimate the discrete residuals in order to obtain convergence order of one in the L_1-norm. We conclude the paper with an illustrative example, which numerically confirms the theoretical findings of the paper.

1. Introduction

Research in direct discretization methods and their convergence properties is an important task in the area of optimal control, because discretization techniques like the Euler method offer a user-friendly way to approximately solve the infinite dimensional control problems by assigning them to discretized, finite dimensional optimization problems (see, e.g., Betts [9], Kraft [21], and von Stryk [40]). By doing so, naturally the following questions emerge:

(i) Do the discretized problems have solutions?
(ii) How accurate are those solutions, i.e., how far are they from the minimizers of the continuous-time control problem?

2020 *Mathematics Subject Classification.* 49J15, 49K15, 49M25, 34A09.
Key words and phrases. Optimal control, bang-bang control control, differential-algebraic equation, discrete approximations, convergence analysis.
This work was supported by the German Research Foundation DFG under the contract GE 1163/8-2.

The aim of this paper is to address both questions for a general class of nonlinear optimal control problems governed by index two differential algebraic equations (DAEs) where controls appear linearly and are box-constrained. Optimal control problems subject to DAEs can be seen as an extension to the classical case, where the dynamics of the control problem are given by ordinary differential equations (ODEs). Such problems arise, e.g., in mechanical engineering, path planning or process engineering. As reference for this topic, we point to the textbooks of Kunkel and Mehrmann [22] and Gerdts [18]. Due to the fact that the studied control problems are control-affine, they give rise to discontinuous optimal controls of bang-bang or bang-singular type. In this paper, we focus on the bang-bang case, which can be assured by certain typical assumptions on the problem data.

The discretization of nonlinear optimal control problems subject to ODEs together with its convergence analysis is well-studied for the case that the optimal control is sufficiently smooth, see, e.g., Dontchev et al [12, 14, 15] and Malanowski et al [23]. We also refer the reader to Tröltzsch [37] and the papers cited therein for similar results on control problems governed by PDEs. Very recently, Martens and Gerdts [24–26, 28] were able to transfer those results to control problems subject to index one and index two DAEs, respectively.

Usually, second-order optimality conditions are used for the error analysis of discretizations. Due to the lack of such conditions for bang-bang and bang-singular controls, initially only a few papers studied the discretization of control-affine problems, see, e.g., Alt and Mackenroth [4] and Veliov [38]. The development of new second-order optimality conditions for control-affine problems by Agrachev et al [1], Felgenhauer [17], and Osmolovskii and Maurer [29–31] supported the achievement of new discretization results. Alt et al [2, 6, 7] and Seydenschwanz [36] used these new second-order conditions to obtain error estimates for the discretization of linear-quadratic control problems with bang-bang solutions. While in [2], the explicit Euler method has been applied, [7, 36] used the implicit midpoint rule and [6] used the implicit Euler method. Again, by also using the implicit Euler scheme, Martens and Gerdts [27] were able to transfer those results to the case of linear-quadratic DAE control problems. Alt and Schneider [5] and Schneider and Wachsmuth [35] obtained error estimates for the discretization of linear-quadratic ODE control problems with additional L_1-sparsity terms in the cost functional. Veliov [39] obtained error estimates for convex control problems of Mayer type with controls appearing linearly by using Runge-Kutta methods of at least third order local consistency. In Haunschmied et al [20], these results have been extended by using the stability concept of strong bi-metric regularity. Pietrus et al [33] study higher order discretization schemes for Mayer type problems based on second order Volterra-Fliess approximations, see also Scarinci and Veliov [34].

Most recently, Alt et al [3] prove convergence of order one w.r.t. the mesh size of the discretization for a general class of control-affine problems in Mayer form. Under slightly weaker assumptions, Osmolovskii and Veliov [32] prove a similar result for control-affine problems in Lagrange form, where they make use of a stability concept called strong metric sub-regularity (SMsR).

In the present paper, we combine ideas of [26] (discrete index reduction for DAEs), [3] (existence of discrete local solutions), and [32] (SMsR), to also prove convergence of order one for a general class of nonlinear control-affine problems governed by DAEs. We chose to analyze problems with index two DAEs for the following reasons:

(i) Problems with index one DAEs can be treated similarly to problems with explicit ODEs. In the cases, where the index of the DAE is two or higher, a discrepancy between the continuous and discrete necessary conditions occurs (compare Section 2.2).

(ii) For the implicit Euler method the algebraic state does not converge on the first step, if the index is three or higher. In this case, discretization schemes of higher order are required in order to obtain error estimates of order one (cf. Brenan et al [10]).

Therefore, we consider the index two case, where the difficulty for optimal control problems with DAEs arises and we can still apply the implicit Euler scheme.

This paper is structured as follows: In Section 2, we present the considered DAE control problem, its implicit discretization, and a comparison of the continuous and discrete optimality conditions. In Section 3, we introduce the assumptions we make throughout this paper and state the main result (Theorem 3.4). We prove strong metric sub-regularity (compare Definition 4.1) for the continuous KKT-conditions in Section 4 (Theorem 4.2). Then, in Section 5, we show that for a sufficiently small mesh-size, the distance between a local solution of the discretized problem with multipliers solving the alternative discrete necessary conditions (2.22)–(2.25) and a continuous KKT-point can be made arbitrarily small (Theorem 5.3). Using the results of Section 4 and Section 5, we are able to prove Theorem 3.4 in Section 6. An example, which numerically confirms the theoretical results of the previous sections, is presented in Section 7.

Notations. Throughout this paper, the n-dimensional Euclidean space with the norm $|\cdot|$ is denoted by $\mathbb{R}^n$. The space of $n \times m$-matrices A is equipped with the spectral norm $\|A\|$. We use $\Gamma, \Gamma_1, \Gamma_2, \ldots \in \mathbb{R}$ for generic, non-negative constants. For a metric or normed space X, we define by $\mathbf{B}_X(w;\nu)$ the closed ball with radius $\nu > 0$ and center $w \in X$ w.r.t. the metric or norm of X. Moreover, for a closed ball in, e.g., $\mathbb{R}^n$, we omit the dimension and just write $\mathbf{B}_{\mathbb{R}}(w;\nu)$. A set-valued mapping for spaces X, Y is denoted by $F\colon X \rightrightarrows Y$. For $\alpha \in [1,\infty]$ and vector functions $w\colon [0,1] \to \mathbb{R}^n$, we introduce the Banach spaces

L^n_α	...	space of equivalence classes, which consist of measurable functions that are bounded in the norm $\|\cdot\|_\alpha$,
$W^n_{1,\alpha}$	...	Sobolev space of absolutely continuous functions that are bounded in the norm $\|\cdot\|_{1,\alpha}$,
BV^n	...	space of functions that are of bounded variation,
BV^n_1	...	space of absolutely continuous functions with first derivative in BV^n,

with the norms

$$\|w\|_\alpha := \left(\int_0^1 |w(t)|^\alpha \,\mathrm{d}t\right)^{\frac{1}{\alpha}}, \qquad \alpha \in [1,\infty), \qquad \|w\|_\infty := \operatorname*{ess\,sup}_{t\in[0,1]} |w(t)|,$$

$$\|w\|_{1,\alpha} := (\|w\|_\alpha + \|\dot{w}\|_\alpha)^{\frac{1}{\alpha}}, \qquad \alpha \in [1,\infty), \qquad \|w\|_{1,\infty} := \|w\|_\infty + \|\dot{w}\|_\infty,$$

and the total variation of w on $[\tau_1, \tau_2] \subseteq [0,1]$, $\tau_1 < \tau_2$ is denoted by $\bigvee_{\tau_1}^{\tau_2} w$.

When it comes to the discretization of control problems, we associate discrete sequences $(w_i)_{i=0,\dots,N} \subset \mathbb{R}^n$ for $N \in \mathbb{N}$, $h = \frac{1}{N}$, and $\alpha \in [1,\infty]$ with the spaces

(1.1) $L^n_{\alpha,h} \subset L^n_\alpha$ ... space of functions that are piecewise constant on $(t_{i-1}, t_i]$ for $i = 1, \dots, N$,

(1.2) $W^n_{1,\alpha,h} \subset W^n_{1,\alpha}$ ... space of functions that are continuous and piecewise linear on $(t_{i-1}, t_i]$ for $i = 1, \dots, N$.

Furthermore, to simplify notation, we often use the abbreviation $F[t]$ for functions of type $F(w(t))$ (usually evaluated at an optimal solution or KKT-point).

2. The Continuous Problem and Its Discretization

In this section, we formulate the DAE optimal control problem that is analyzed throughout the paper. We then use the implicit Euler scheme to obtain an discretized version of the time-continuous problem. Finally, we discuss the resulting optimality conditions.

2.1. **The Continuous DAE Optimal Control Problem.** Let $X = W^{n_x}_{1,1} \times L^{n_y}_1 \times L^{n_u}_1$. For a control $u(t) \in \mathbb{R}^{n_u}$, a differential state $x(t) \in \mathbb{R}^{n_x}$, and an algebraic state $y(t) \in \mathbb{R}^{n_y}$ at time $t \in [0,1]$, we consider the following optimal control problem

$$\text{(OCP)} \qquad \underset{(x,y,u)\,\in\,X}{\text{Minimize}} \quad \int_0^1 f_0\left(x(t), y(t), u(t)\right) \mathrm{d}t$$

$$(2.1) \qquad \text{subject to} \quad \dot{x}(t) = f\left(x(t), y(t), u(t)\right) \qquad \text{a.e. in } [0,1]\,,$$

$$(2.2) \qquad 0 = g\left(x(t)\right) \qquad \text{in } [0,1]\,,$$

$$(2.3) \qquad 0 = D\left(x(0) - x^0\right),$$

$$(2.4) \qquad u(t) \in U \qquad \text{a.e. in } [0,1]\,.$$

The running costs $f_0 \colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_u} \to \mathbb{R}$ of the objective functional,

$$(2.5) \qquad f_0(x,y,u) := p(x) + q(x)^\top y + r(x)^\top u,$$

and the right hand side $f \colon \mathbb{R}^{n_x} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ of the system equation (2.1),

$$(2.6) \qquad f(x,y,u) := a(x) + B(x)\,y + C(x)\,u\,,$$

are affine linear w.r.t. y and u and may be nonlinear w.r.t. x. For the initial condition (2.3) we have $D \in \mathbb{R}^{(n_x - n_y)\times n_x}$. Furthermore, the algebraic equation (2.2) is defined by $g \colon \mathbb{R}^{n_x} \to \mathbb{R}^{n_y}$. The control set $U \subset \mathbb{R}^{n_u}$ in (2.4) is defined by box constraints, i.e.,

$$U := \{u \in \mathbb{R}^{n_u} \mid b_\ell \le u \le b_u\} \quad \text{with } b_\ell, b_u \in \mathbb{R}^{n_u} \text{ and } b_\ell < b_u\,.$$

Here, all inequalities are to be understood componentwise.
Note that the dynamics in Problem (OCP) are given by a differential algebraic equation (DAE), which consists of a differential equation (2.1) and an algebraic equation (2.2) in semi-explicit form. A DAE is usually characterized by its index,

which has multiple concepts (cf. [22]). By differentiating the algebraic equation (2.2) twice with respect to time t, we get

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}t} g\left(x(t)\right) = g'\left(x(t)\right) f\left(x(t), y(t), u(t)\right), \\
0 &= \frac{\mathrm{d}^2}{\mathrm{d}t^2} g\left(x(t)\right) \\
&= g''\left(x(t)\right) f\left(x(t), y(t), u(t)\right) f\left(x(t), y(t), u(t)\right) \\
&\quad + g'\left(x(t)\right) f'_x\left(x(t), y(t), u(t)\right) f\left(x(t), y(t), u(t)\right) \\
&\quad + g'\left(x(t)\right) f'_y\left(x(t), y(t), u(t)\right) \dot{y}(t) + g'\left(x(t)\right) f'_u\left(x(t), y(t), u(t)\right) \dot{u}(t).
\end{aligned}$$

We can solve the latter equation for $\dot{y}$, if the matrix $g'\left(x\right) f'_y\left(x, y, u\right) = g'\left(x\right) B\left(x\right)$ is non-singular with a bounded inverse along a trajectory (compare Assumption (A2) below). Thus, after two differentiations of the algebraic equation we obtain an explicit differential equation for the algebraic variable y and therefore the DAE has the (differentiation) index two.

Remark 2.1. In theory, with the above index two condition, it would be possible to first solve the time derivative of (2.2) for the algebraic state y depending on x and u, to insert the expression into (2.1) in order to obtain an explicit ODE, and to then apply the implicit Euler scheme. However, as outlined in [26, Remark 1], considering discretizations of this reduced problem has numerous drawbacks, e.g., the drift-off effect (cf. [11, 19]).

Denoting the feasible set of Problem (OCP) by

$$\mathcal{F} := \big\{(x, y, u) \in X \mid \dot{x}(t) = f\left(x(t), y(t), u(t)\right) \text{ a.e. in } [0,1], \; g\left(x(t)\right) = 0 \text{ in } [0,1], \\ D\left(x(0) - x^0\right) = 0, \text{ and } u(t) \in U \text{ a.e. in } [0,1]\big\},$$

we introduce the concept of a local solution of Problem (OCP).

Definition 2.2. A tuple $(\hat{x}, \hat{y}, \hat{u}) \in \mathcal{F}$ is called a local solution of Problem (OCP), if there exists some $\varepsilon > 0$, such that

$$\int_0^1 f_0\left(\hat{x}(t), \hat{y}(t), \hat{u}(t)\right) \mathrm{d}t \leq \int_0^1 f_0\left(x(t), y(t), u(t)\right) \mathrm{d}t$$

for all $(x, y, u) \in \mathcal{F} \cap \mathbf{B}_X\left((\hat{x}, \hat{y}, \hat{u}); \varepsilon\right)$.

In order to formulate optimality conditions for Problem (OCP), we introduce the Hamilton function

$$\mathcal{H}\left(x, y, u, \lambda, \mu\right) := f_0\left(x, y, u\right) + \lambda^\top f\left(x, y, u\right) + \mu^\top g'\left(x\right) f\left(x, y, u\right). \tag{2.7}$$

Let $(\hat{x}, \hat{y}, \hat{u}) \in \mathcal{F}$ be a local solution of Problem (OCP), then there exist functions $\hat{\lambda} \in W_{1,1}^{n_x}$ and $\hat{\mu} \in L_1^{n_y}$, such that the following necessary conditions hold (cf. [18, Theorem 3.1.11]):

$$0 = \dot{\hat{\lambda}}(t) + \nabla_x \mathcal{H}\left(\hat{x}(t), \hat{y}(t), \hat{u}(t), \hat{\lambda}(t), \hat{\mu}(t)\right) \quad \text{a.e. in } [0,1]\,, \tag{2.8}$$

$$0 = \nabla_y \mathcal{H}\left(\hat{x}(t), \hat{y}(t), \hat{u}(t), \hat{\lambda}(t), \hat{\mu}(t)\right) \quad \text{a.e. in } [0,1]\,, \tag{2.9}$$

$$0 = \hat{\lambda}(1), \tag{2.10}$$

$$0 \in \nabla_u \mathcal{H}\left(\hat{x}(t), \hat{y}(t), \hat{u}(t), \hat{\lambda}(t), \hat{\mu}(t)\right) + \mathcal{N}_U\left(\hat{u}(t)\right) \quad \text{a.e. in } [0,1]\,, \tag{2.11}$$

where the normal cone operator to U at u is defined by

$$\mathcal{N}_U(u) := \begin{cases} \{w \in \mathbb{R}^{n_u} \mid \langle w, v-u\rangle \le 0 \text{ for all } v \in U\}\,, & \text{if } u \in U\,, \\ \emptyset\,, & \text{otherwise.} \end{cases} \tag{2.12}$$

2.2. **Implicit Euler Discretization of Problem** (OCP)**.** For $N \in \mathbb{N}$ with $N \ge 2$ and the mesh size $h = 1/N$, let $t_i = i\,h$ for $i = 0, \dots, N$ be the grid points of the discretization. We approximate the controls and the algebraic states by piecewise constant functions in $L^{n_u}_{1,h}$ and $L^{n_y}_{1,h}$ (compare (1.1)) represented by their values $u(t_i) = u_i$ and $y(t_i) = y_i$, resp., at the grid points. Further, we approximate the differential states by continuous, piecewise linear functions in $W^{n_x}_{1,1,h}$ (compare (1.2)) represented by their values $x(t_i) = x_i$ at the grid points. With the backwards difference approximation

$$x_i' := \frac{x_i - x_{i-1}}{h}\,, \quad i = 1, \dots, N\,,$$

we obtain the implicit Euler discretization of (OCP):

$$\text{(DOCP)} \qquad \text{Minimize} \quad h \sum_{i=1}^{N} f_0\left(x_i, y_i, u_i\right)$$

$$\text{subject to} \quad x_i' = f\left(x_i, y_i, u_i\right), \qquad i = 1, \dots, N\,, \tag{2.13}$$

$$0 = g\left(x_i\right), \qquad i = 0, \dots, N\,, \tag{2.14}$$

$$0 = D\left(x_0 - x^0\right), \tag{2.15}$$

$$u_i \in U\,, \qquad i = 1, \dots, N\,. \tag{2.16}$$

In order to derive convergence properties for solutions of (DOCP), we compare the respective KKT-conditions of (OCP) and (DOCP). For the discretized problem (DOCP) we have the Hamilton function

$$\mathbb{H}\left(x, y, u, \varphi, \psi\right) := f_0\left(x, y, u\right) + \varphi^\top f\left(x, y, u\right) + \psi^\top g\left(x\right). \tag{2.17}$$

One would expect the continuous necessary conditions to hold for (2.17), where the algebraic constraint (2.2) is directly adjoined to the Hamilton function. However, in [8, Example 3.16] it was shown that the continuous necessary conditions (2.8)–(2.11) are not satisfied for (2.17) in general. Thus, there is a discrepancy between the continuous necessary conditions formulated with the Hamilton function (2.7) and the discrete necessary conditions which use the Hamilton function (2.17) (cf. [18, Theorem 5.4.4]). This originates from an implicit index reduction that occurs for the continuous necessary conditions. If the DAE (2.1), (2.2) has index two, then the adjoint DAE (2.8), (2.9) has only index one. Furthermore, the necessary conditions

(2.8)–(2.11) correspond to the equivalent index one problem, where we replaced the algebraic equation (2.2) in (OCP) with the equivalent constraints

$$
\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}t} g\left(x(t)\right) = g'\left(x(t)\right) f\left(x(t), y(t), u(t)\right) \\
&= g'\left(x(t)\right)\left[a\left(x(t)\right) + B\left(x(t)\right) y(t) + C\left(x(t)\right) u(t)\right] \quad \text{a.e. in } [0,1]\,, \qquad (2.18)\\
0 &= g\left(x(0)\right). \qquad (2.19)
\end{aligned}
$$

Note that the index reduced algebraic constraint (2.18) is adjoined in (2.7) instead of (2.2). In order to obtain consistent necessary conditions for (OCP) and (DOCP), we emulate the index reduction in (2.18), (2.19) for the discrete algebraic equation (2.14) in (DOCP) (cf. [26]), and obtain

$$
\begin{aligned}
& 0 = g\left(x_i\right), && i = 0, \dots, N \\
\iff\quad & 0 = \frac{1}{h}\left[g\left(x_i\right) - g\left(x_{i-1}\right)\right], && i = 1, \dots, N\,, \quad 0 = g\left(x_0\right). \qquad (2.20)
\end{aligned}
$$

Now, with the relation

$$x_{i-1} = x_i - h\, f\left(x_i, y_i, u_i\right), \qquad i = 1, \dots, N$$

derived from the difference equation (2.13), and the discrete approximation of the time derivative of $g\left(x(\cdot)\right)$,

$$\tilde{g}_h\left(x, y, u\right) := \frac{1}{h}\left[g\left(x\right) - g\left(x - hf\left(x, y, u\right)\right)\right],$$

we introduce the alternative discrete Hamilton function

$$\mathcal{H}_h\left(x, y, u, \lambda, \mu\right) := f_0\left(x, y, u\right) + \lambda^\top f\left(x, y, u\right) + \mu^\top \tilde{g}_h\left(x, y, u\right). \qquad (2.21)$$

We obtain the alternative discrete necessary conditions (cf. [18, Theorem 5.4.4])

$$
\begin{aligned}
(2.22)\quad 0 &= \lambda_i' + \nabla_x \mathcal{H}_h\left(x_i, y_i, u_i, \lambda_{i-1}, \mu_{i-1}\right), && i = 1, \dots, N\,,\\
(2.23)\quad 0 &= \nabla_y \mathcal{H}_h\left(x_i, y_i, u_i, \lambda_{i-1}, \mu_{i-1}\right), && i = 1, \dots, N\,,\\
(2.24)\quad 0 &= \lambda_N\,,\\
(2.25)\quad 0 &\in \nabla_u \mathcal{H}_h\left(x_i, y_i, u_i, \lambda_{i-1}, \mu_{i-1}\right) + \mathcal{N}_U\left(u_i\right), && i = 1, \dots, N\,.
\end{aligned}
$$

In addition, we have the following relationships between the multipliers associated with (2.17) and (2.21):

$$
\begin{aligned}
&\varphi_i = \lambda_i + g'\left(x_i\right)^\top \mu_i\,, \quad \lambda_i = \varphi_i - g'\left(x_i\right)^\top \sum_{k=i}^{N-1} h\psi_k\,, \quad i = 0, \dots, N\,,\\
&\psi_{i-1} = -\frac{\mu_i - \mu_{i-1}}{h}\,, \qquad \mu_{i-1} = \sum_{k=i-1}^{N-1} h\psi_k\,, \qquad i = 1, \dots, N\,, \quad \mu_N = 0\,.
\end{aligned}
$$

It is easy to show that the usual discrete necessary conditions formulated with the Hamilton function (2.17) are equivalent to the necessary conditions (2.22)–(2.25) with the alternative discrete Hamilton function (2.21), i.e., using the relations between the multipliers, we can transform the usual discrete necessary conditions into the alternative ones and vice versa.

3. Assumptions and Main Result

In this section, we first present the assumptions we make throughout this paper, and then we introduce the main result, which will be proven in the following sections.

(A1) There exists a local solution $(\hat{x}, \hat{y}, \hat{u}) \in W_{1,1}^{n_x} \times L_1^{n_y} \times L_1^{n_u}$ of (OCP) and a convex compact set $\mathcal{M}$ with $\mathbf{B}_{\mathbb{R}}(\hat{x}(t); 1) \subset \mathcal{M}$ for all $t \in [0,1]$. The functions p, q, r, a, B, C are twice differentiable and g is three times differentiable with respect to x on $\mathcal{M}$. Furthermore, p, q, r, a, B, C and their first two derivatives, g and its first three derivatives are Lipschitz continuous on $\mathcal{M}$ with constant $\mathbb{L}$.

(A2) The matrix

$$g'(\hat{x}(t)) f_y'(\hat{x}(t), \hat{y}(t), \hat{u}(t)) = g'(\hat{x}(t)) B(\hat{x}(t)) \in \mathbb{R}^{n_y \times n_y}$$

is non-singular and the inverse is continuous and uniformly bounded for all $t \in [0,1]$. Moreover, the matrix $\begin{pmatrix} g'(\hat{x}(0)) \\ D \end{pmatrix} \in \mathbb{R}^{n_x \times n_x}$ is non-singular.

Remark 3.1.

(i) According to (A1) and (A2), there exist Lagrange multipliers $\left(\hat{\lambda}, \hat{\mu}\right) \in W_{1,1}^{n_x} \times L_1^{n_y}$ solving (2.8)–(2.11). Moreover, we can enlarge $\mathcal{M}$ in (A1) if necessary such that $\mathbf{B}_{\mathbb{R}}\left(\hat{\lambda}(t); 1\right) \subset \mathcal{M}$ for all $t \in [0,1]$. In addition, since $\hat{u}(t) \in U$ for almost every $t \in [0,1]$, we have

$$\hat{u} \in \mathcal{U} := \{u \in L_\infty^{n_u} \mid u(t) \in U \text{ a.e. in } [0,1]\}.$$

Exploiting (A2), we can solve (2.18) (the time derivative of the algebraic equation (2.2)) for $\hat{y}$ and the algebraic equation (2.9) for $\hat{\mu}$ to obtain

$$\hat{y}(\cdot) = -\left(g'(\hat{x}(\cdot)) B(\hat{x}(\cdot))\right)^{-1} g'(\hat{x}(\cdot)) \left[a(\hat{x}(\cdot)) + C(\hat{x}(\cdot)) \hat{u}(\cdot)\right], \tag{3.1}$$

$$\hat{\mu}(\cdot) = -\left[\left(g'(\hat{x}(\cdot)) B(\hat{x}(\cdot))\right)^{-1}\right]^\top \left[q(\hat{x}(\cdot)) + B(\hat{x}(\cdot))^\top \hat{\lambda}(\cdot)\right], \tag{3.2}$$

which implies $\hat{y} \in L_\infty^{n_y}$ and $\hat{\mu} \in W_{1,1}^{n_y}$. Then, the differential equations (2.1) and (2.8) yield $\hat{x} \in W_{1,\infty}^{n_x}$ and $\hat{\lambda} \in W_{1,\infty}^{n_x}$. Exploiting (3.2) again, gives us $\hat{\mu} \in W_{1,\infty}^{n_y}$.

(ii) For every $(y, u, \mu) \in \mathbf{B}_{L_1}(\hat{y}; \epsilon) \times \mathbf{B}_{L_1}(\hat{u}; \epsilon) \cap \mathcal{U} \times \mathbf{B}_{L_1}(\hat{\mu}; \epsilon)$, where $\epsilon > 0$ is sufficiently small, the differential equations (2.1) and (2.8) have solutions $x \in W_{1,1}^{n_x}$ and $\lambda \in W_{1,1}^{n_x}$ with $\mathbf{B}_{\mathbb{R}}\left(x(t); \frac{1}{2}\right) \cup \mathbf{B}_{\mathbb{R}}\left(\lambda(t); \frac{1}{2}\right) \subset \mathcal{M}$ for all $t \in [0,1]$. We denote by $\mathbb{M}$ the bound of p, q, r, a, B, C and their first two derivatives, g and its first three derivatives on $\mathcal{M}$.

Let us denote the switching function by

$$\sigma(\cdot) := \nabla_u \mathcal{H}[\cdot] = r(\hat{x}(\cdot)) + C(\hat{x}(\cdot))^\top \hat{\lambda}(\cdot) + C(\hat{x}(\cdot))^\top g'(\hat{x}(\cdot))^\top \hat{\mu}(\cdot) \tag{3.3}$$

and assume the following growth condition:

(A3) Let there exist constants $\varsigma, \rho > 0$ such that for every $j \in \{1, \ldots, n_u\}$ and $s \in [0,1]$ with $\sigma_j(s) = 0$ we have

$$|\sigma_j(t)| \geq \varsigma |t - s|, \quad \text{for all } t \in [s - \rho, s + \rho] \cap [0,1].$$

Since $\left(\hat{x}, \hat{\lambda}, \hat{\mu}\right) \in W_{1,\infty}^{n_x} \times W_{1,\infty}^{n_x} \times W_{1,\infty}^{n_y}$ by Remark 3.1(i), we have $\sigma \in W_{1,\infty}^{n_u}$. In addition, according to (A3), each component of the switching function has at most $\frac{1}{\rho}$ isolated zeros, which implies the following:

Lemma 3.2. *If Assumptions (A1)–(A3) are satisfied, then the optimal control is of bang-bang-type and we have*

$$\left(\hat{x}, \hat{y}, \hat{u}, \hat{\lambda}, \hat{\mu}\right) \in BV_1^{n_x} \times BV^{n_y} \times BV^{n_u} \times BV_1^{n_x} \times BV_1^{n_y}\,. \tag{3.4}$$

Proof. The structure of the optimal control $\hat{u}$ is determined by the switching function (3.3) and condition (2.11). Thus, for $j = 1, \ldots, n_u$ we have

$$u_j(t) = \begin{cases} b_{\ell,j}\,, & \text{if } \sigma_j(t) > 0\,, \\ b_{u,j}\,, & \text{if } \sigma_j(t) < 0\,, \\ \text{undetermined}, & \text{if } \sigma_j(t) = 0\,. \end{cases} \tag{3.5}$$

Since the switching function has finitely many, isolated zeros, the optimal control is bang-bang, i.e., $u_j(t) \in \{b_{\ell,j}, b_{u,j}\}$ a.e. in $[0,1]$ for $j = 1, \ldots, n_u$. Hence, $\hat{u} \in BV^{n_u}$. Equation (3.1) implies $\hat{y} \in BV^{n_y}$. Exploiting the differential equations (2.1) and (2.8) we get $\hat{x}, \hat{\lambda} \in BV_1^{n_x}$, since $\hat{\mu} \in W_{1,\infty}^{n_y} \subset BV^{n_y}$. Finally, (3.2) yields $\hat{\mu} \in BV_1^{n_y}$, which proves the assertion. ☐

According to [17, Lemma 3.3] and [32, Proposition 4.1], Assumptions (A1)–(A3) are sufficient for the following lower bound condition for the switching function (3.3):

Lemma 3.3. *Let (A1)–(A3) be satisfied. Then, there exists $\eta_1 > 0$ such that for every $u \in \mathcal{U}$ we have*

$$\int_0^1 \langle \sigma(t), u(t) - \hat{u}(t) \rangle \,\mathrm{d}t \geq \eta_1 \left\| u - \hat{u} \right\|_1^2. \tag{3.6}$$

Next, we define the quadratic functional $\mathcal{P} : W_{1,1}^{n_x} \times L_1^{n_y} \times L_1^{n_u} \to \mathbb{R}$ by

$$\begin{aligned} \mathcal{P}\left(\delta x, \delta y, \delta u\right) := \frac{1}{2} \int_0^1 & \delta x(t)^\top \nabla_{xx}^2 \mathcal{H}\left[t\right] \delta x(t) \\ & + 2\delta x(t)^\top \left(\nabla_{xy}^2 \mathcal{H}\left[t\right] \delta y(t) + \nabla_{xu}^2 \mathcal{H}\left[t\right] \delta u(t)\right) \mathrm{d}t\,, \end{aligned} \tag{3.7}$$

where $\mathcal{H}[t] := \mathcal{H}\left(\hat{x}(t), \hat{y}(t), \hat{u}(t), \hat{\lambda}(t), \hat{\mu}(t)\right)$. Note that $\nabla_{(y,u)(y,u)}^2 \mathcal{H}\left[t\right] = 0$ by (2.5), (2.6). For this functional we assume the following:

(A4) There exists $\eta_2 \in (0, \eta_1)$, where $\eta_1 > 0$ is the constant in Lemma 3.3, such that

$$2\mathcal{P}\left(\delta x, \delta y, \delta u\right) \geq -\eta_2 \left\| \delta u \right\|_1^2 \tag{3.8}$$

for every $(\delta x, \delta y, \delta u) \in W_{1,1}^{n_x} \times L_1^{n_y} \times L_1^{n_u}$ satisfying

$$\delta \dot{x}(t) = f_x' [t] \, \delta x(t) + B' [t] \, \delta y(t) + C' [t] \, \delta u(t) \quad \text{a.e. in } [0,1] , \tag{3.9}$$

$$0 = g' [t] \, \delta x(t) \quad \text{in } [0,1] , \tag{3.10}$$

$$0 = D \delta x(0) , \tag{3.11}$$

$$\delta u(t) \in U - \hat{u}(t) \quad \text{a.e. in } [0,1] . \tag{3.12}$$

An immediate consequence of Lemma 3.3 and Assumption (A4) is

$$\int_0^1 \langle \sigma(t), \delta u(t) \rangle \, \mathrm{d}t + 2\mathcal{P} (\delta x, \delta y, \delta u) \geq \eta \, \|\delta u\|_1^2 \tag{3.13}$$

for $\eta := \eta_1 - \eta_2 > 0$ and every $(\delta x, \delta y, \delta u)$ satisfying (3.9)–(3.12). This in turn implies that $(\hat{x}, \hat{y}, \hat{u})$ is a strict local solution of Problem (OCP) and, according to [32, Corollary 2.1], we have the following quadratic growth condition for the objective functional:

$$\int_0^1 f_0 (x(t), y(t), u(t)) \, \mathrm{d}t - \int_0^1 f_0 (\hat{x}(t), \hat{y}(t), \hat{u}(t)) \, \mathrm{d}t \geq \frac{\eta}{4} \, \|\delta u\|_1^2 \tag{3.14}$$

for all feasible $(x, y, u) \in W_{1,1}^{n_x} \times L_1^{n_y} \times L_1^{n_u}$ with

$$\delta u = u - \hat{u} \in (\mathcal{U} - \hat{u}) \cap \mathbf{B}_{L_1} (0; \nu) \setminus \{0\} ,$$

where $\nu > 0$ is sufficiently small.

Assumptions (A1)–(A4) allow us to state the main result of this paper, whose proof will be given in Section 6:

Theorem 3.4. *Let (A1)–(A4) be satisfied and $h > 0$ be sufficiently small. Then,* (DOCP) *has a solution $(\hat{x}_h, \hat{y}_h, \hat{u}_h)$ with unique multipliers $\left(\hat{\lambda}_h, \hat{\mu}_h\right)$ satisfying the alternative necessary conditions* (2.22)–(2.25)*, and for $\Gamma \geq 0$ independent of h, we have the error estimates*

$$\|\hat{x}_h - \hat{x}\|_{1,1} + \|\hat{y}_h - \hat{y}\|_1 + \|\hat{u}_h - \hat{u}\|_1 + \left\|\hat{\lambda}_h - \hat{\lambda}\right\|_{1,1} + \|\hat{\mu}_h - \hat{\mu}\|_\infty \leq \Gamma \, h . \tag{3.15}$$

4. Strong Metric sub-Regularity

In this section, we aim to show that the continuous KKT-conditions satisfy a certain stability property, if Assumptions (A1)–(A4) hold. To that end, we introduce the notion of strong metric sub-regularity (cf. [16, p. 202], [32, Definition 1.1]).

Definition 4.1. Let (Ξ, d_Ξ) and (Ω, d_Ω) be two metric spaces. A set-valued mapping $\mathcal{F} \colon \Xi \rightrightarrows \Omega$ is strongly metrically sub-regular (SMsR) at $\hat{\xi} \in \Xi$ for $\hat{\omega} \in \Omega$, if $\hat{\omega} \in \mathcal{F}\left(\hat{\xi}\right)$ and there exist constants $\alpha, \beta, \gamma > 0$ such that for any $\omega \in \mathbf{B}_\Omega (\hat{\omega}; \alpha)$ and for any solution $\xi \in \mathbf{B}_\Xi \left(\hat{\xi}; \beta\right)$ of $\omega \in \mathcal{F}(\xi)$ the inequality

$$d_\Xi \left(\xi, \hat{\xi}\right) \leq \gamma \, d_\Omega (\omega, \hat{\omega}) \tag{4.1}$$

is satisfied.

In order to prove that the continuous KKT-conditions are strongly metrically sub-regular, we denote the following abstract setting: Let us define the metric spaces

$$\begin{aligned}
(4.2)\qquad \Xi &:= W_{1,1,D}^{n_x} \times L_1^{n_y} \times \mathcal{U} \times W_{1,1,0}^{n_x} \times L_\infty^{n_y}\,,\\
W_{1,1,D}^{n_x} &:= \left\{ x \in W_{1,1}^{n_x} \mid 0 = D\left(x(0) - x^0\right) \right\},\\
W_{1,1,0}^{n_x} &:= \left\{ \lambda \in W_{1,1}^{n_x} \mid 0 = \lambda(1) \right\},\\
(4.3)\qquad \Omega &:= L_1^{n_x} \times W_{1,1}^{n_y} \times L_1^{n_x} \times L_\infty^{n_y} \times L_\infty^{n_u}\,,
\end{aligned}$$

with elements $\xi = (x, y, u, \lambda, \mu) \in \Xi$, $\omega = \left(\omega_f, \omega_g, \omega_{\mathcal{H}_x}, \omega_{\mathcal{H}_y}, \omega_{\mathcal{H}_u}\right) \in \Omega$, and the metrics

$$\begin{aligned}
(4.4)\qquad d_\Xi\left(\xi^1, \xi^2\right) &:= \left\|x^1 - x^2\right\|_{1,1} + \left\|y^1 - y^2\right\|_1 + \left\|u^1 - u^2\right\|_1\\
&\quad + \left\|\lambda^1 - \lambda^2\right\|_{1,1} + \left\|\mu^1 - \mu^2\right\|_\infty\,,\\
(4.5)\qquad d_\Omega\left(\omega^1, \omega^2\right) &:= \left\|\omega_f^1 - \omega_f^2\right\|_1 + \left\|\omega_g^1 - \omega_g^2\right\|_{1,1} + \left\|\omega_{\mathcal{H}_x}^1 - \omega_{\mathcal{H}_x}^2\right\|_1\\
&\quad + \left\|\omega_{\mathcal{H}_y}^1 - \omega_{\mathcal{H}_y}^2\right\|_\infty + \left\|\omega_{\mathcal{H}_u}^1 - \omega_{\mathcal{H}_u}^2\right\|_\infty.
\end{aligned}$$

We denote the function $\mathbf{T}\colon \Xi \to \Omega$ and the set-valued mapping $\mathbf{F}\colon \Xi \rightrightarrows \Omega$ by

$$(4.6)\qquad \mathbf{T}(\xi) := \begin{pmatrix} \dot{x} - f(x, y, u) \\ g(x) \\ \dot{\lambda} + \nabla_x \mathcal{H}(\xi) \\ \nabla_y \mathcal{H}(\xi) \\ \nabla_u \mathcal{H}(\xi) \end{pmatrix}, \quad \mathbf{F}(\xi) := \begin{pmatrix} \{0\} \\ \{0\} \\ \{0\} \\ \{0\} \\ \mathcal{N}_\mathcal{U}(u) \end{pmatrix},$$

with the normal cone operator to $\mathcal{U}$ at u

$$\mathcal{N}_\mathcal{U}(u) := \begin{cases} \{w \in L_\infty^{n_u} \mid w(t) \in \mathcal{N}_U(u(t)) \text{ for a.e. } t \in [0,1]\}\,, & \text{if } u \in \mathcal{U}\,,\\ \emptyset\,, & \text{otherwise.} \end{cases}$$

Then, the continuous KKT-conditions can be written as the generalized equation

$$(4.7)\qquad 0 \in \mathbf{T}(\xi) + \mathbf{F}(\xi)\,.$$

In the main theorem of this section, we state that $\mathbf{T} + \mathbf{F}$ is SMsR at

$$(4.8)\qquad \hat{\xi} := \left(\hat{x}, \hat{y}, \hat{u}, \hat{\lambda}, \hat{\mu}\right) \in \Xi$$

for $0 \in \Omega$. For the reader's convenience we included the proof, where we use similar techniques as in [32, Theorem 3.1, Proposition 3.2], in Appendix A.1.

Theorem 4.2. *Let (A1)–(A4) be satisfied. Then, there exist $\alpha, \beta, \gamma > 0$ depending only on the constants $\mathbb{L}, \mathbb{M}$, and η (compare (A1), Remark 3.1(ii),* (3.13)*) such that $\mathbf{T} + \mathbf{F}$ is strongly metrically sub-regular at $\hat{\xi}$ for* 0 *with constants α, β, γ.*

5. Existence of a discrete Solution

We associate solutions (x, y, u, λ, μ) of the discrete problem (DOCP) and the alternative necessary conditions (2.22)–(2.25) with piecewise linear, continuous functions x_h, $\lambda_h \in W^{n_x}_{1,\infty,h}$, and piecewise constant functions $y_h \in L^{n_y}_{\infty,h}$, $u_h \in L^{n_u}_{\infty,h}$, $\mu_h \in L^{n_y}_{\infty,h}$ (compare (1.1) and (1.2)) defined by

$$
\begin{aligned}
&(5.1) \quad x_h(t) = x_h'(t_i)\,(t - t_{i-1}) + x_h(t_{i-1}), && t \in (t_{i-1}, t_i]\,,\ i = 1, \ldots, N\,,\\
&(5.2) \quad y_h(t) = y_h(t_i), && t \in (t_{i-1}, t_i]\,,\ i = 1, \ldots, N\,,\\
&(5.3) \quad u_h(t) = u_h(t_i), && t \in (t_{i-1}, t_i]\,,\ i = 1, \ldots, N\,,\\
&(5.4) \quad \lambda_h(t) = \lambda_h'(t_i)\,(t - t_{i-1}) + \lambda_h(t_{i-1}), && t \in (t_{i-1}, t_i]\,,\ i = 1, \ldots, N\,,\\
&(5.5) \quad \mu_h(t) = \mu_h(t_{i-1}), && t \in (t_{i-1}, t_i]\,,\ i = 1, \ldots, N\,.
\end{aligned}
$$

Note that we evaluate μ_h at t_{i-1} for $t \in (t_{i-1}, t_i]$ instead of t_i.

In this section, we aim to show that there exists a discrete solution

$$\hat{\xi}_h := \left(\hat{x}_h, \hat{y}_h, \hat{u}_h, \hat{\lambda}_h, \hat{\mu}_h\right) \tag{5.6}$$

of (DOCP) and the alternative necessary conditions (2.22)–(2.25) with

$$d_\Xi\left(\hat{\xi}_h, \hat{\xi}\right) \le \beta\,. \tag{5.7}$$

Herein, $\beta > 0$ is derived from Theorem 4.2 and can be arbitrarily small. In order to prove the main result of this section (Theorem 5.3 below), we consider the discrete problem (DOCP) with the extra constraints

$$\|x_h - \hat{x}\|_{1,1} \le \upsilon\,, \quad \|y_h - \hat{y}\|_1 \le \upsilon\,, \quad \|u_h - \hat{u}\|_1 \le \upsilon\,, \tag{5.8}$$

for an arbitrarily small constant $0 < \upsilon \le \epsilon$ (compare Remark 3.1(ii)), and show that for sufficiently small h this problem has a solution. To that end, we first prove that the feasible set of this problem is not empty. Let (A1)–(A3) be satisfied. Then, according to Lemma 3.2, we have $\hat{u} \in BV^{n_u}$. Set

$$\bar{u}_h(t) := \hat{u}(t_i)\,, \quad t \in (t_{i-1}, t_i]\,, \quad i = 1, \ldots, N\,,$$

which implies $\bar{u}_h \in \mathcal{U}$ and $\|\bar{u}_h - \hat{u}\|_1 \le \bigvee_0^1 h \le \upsilon$ for sufficiently small h (cf. [2, Lemma 3.1]). Thus, $\bar{u}_h$ is feasible. Next, we consider the system

$$
\begin{aligned}
x_h(t_i) &= f\left(x_h(t_i), y_h(t_i), \hat{u}(t_i)\right), && i = 1, \ldots, N\,, \quad x_h(t_0) = \hat{x}(0)\,,\\
0 &= g\left(x_h(t_i)\right), && i = 1, \ldots, N\,,\\
\|x_h - \hat{x}\|_{1,1} &\le \upsilon\,, \quad \|y_h - \hat{y}\|_1 \le \upsilon\,.
\end{aligned}
$$

We prove that this system has a solution $(\bar{x}_h, \bar{y}_h) \in W^{n_x}_{1,\infty,h} \times L^{n_y}_{\infty,h}$, which implies that the feasible set is not empty, by applying [13, Theorem 3.1]. Let us introduce the following abstract setting: We define the spaces

$$\mathbf{X}_h := \left\{z_h = (x_h, y_h) \in W^{n_x}_{1,\infty,h} \times L^{n_y}_{\infty,h} \mid x_h(t_0) = \hat{x}(0)\right\}, \quad \mathbf{Y}_h := L^{n_x}_{1,h} \times W^{n_y}_{1,1,h}\,,$$

equipped with the metric $d_{\mathbf{X}}\left(z_h^1, z_h^2\right) := \left\|x_h^1 - x_h^2\right\|_{1,1} + \left\|y_h^1 - y_h^2\right\|_1$ and the norm $\|(\omega_{f,h}, \omega_{g,h})\|_{\mathbf{Y}} := \|\omega_{f,h}\|_1 + \|\omega_{g,h}\|_{1,1}$. Additionally, with $\hat{z} := (\hat{x}, \hat{y})$ we define the function $\mathcal{T}\colon \mathbf{X}_h \to \mathbf{Y}_h$ and the linear mapping $\mathcal{L}\colon \mathbf{X}_h \to \mathbf{Y}_h$ by

$$\mathcal{T}\left(z_h(t_i)\right) := \begin{pmatrix} x_h'(t_i) - f\left(x_h(t_i), y_h(t_i), \hat{u}(t_i)\right) \\ g\left(x_h(t_i)\right) \end{pmatrix},$$

$$\mathcal{L}\left(z_h(t_i)\right) := \mathcal{T}'\left(\hat{z}(t_i)\right) z_h(t_i) = \begin{pmatrix} x_h'(t_i) - f_x'\left[t_i\right] x_h(t_i) - B\left[t_i\right] y_h(t_i) \\ g'\left[t_i\right] x_h(t_i) \end{pmatrix}$$

for $i = 1, \ldots, N$. Finally, we define the elements $\hat{\omega} := -\mathcal{T}\left(\hat{z}\right)$, $\hat{\pi} := \mathcal{T}\left(\hat{z}\right) - \mathcal{L}\left(\hat{z}\right)$, and the set $\Pi := \mathbf{B}_{\mathbf{Y}}\left(\hat{\pi}; \nu\right) \subset \mathbf{Y}_h$ for sufficiently small $\nu > 0$. In order to not disturb the reading flow, we moved the proof of the following lemma to Appendix A.2.

Lemma 5.1. *Let (A1)–(A3) be satisfied. Then, the equation*

$$0 = \mathcal{T}\left(z_h\right), \tag{5.9}$$

has a solution $\bar{z}_h = (\bar{x}_h, \bar{y}_h) \in \mathbf{X}_h$ *and there exists* $\Gamma \geq 0$ *independent of* h *such that*

$$\|\bar{x}_h - \hat{x}\|_{1,1} \leq \Gamma\, h\,, \quad \|\bar{y}_h - \hat{y}\|_1 \leq \Gamma\, h\,. \tag{5.10}$$

This proves that the feasible set of (DOCP) together with (5.8) is not empty for sufficiently small h. Additionally, the objective functional is continuous and the feasible set is compact. Hence, the problem has a solution

$$(\hat{x}_h, \hat{y}_h, \hat{u}_h) \in W_{1,\infty,h}^{n_x} \times L_{\infty,h}^{n_y} \times L_{\infty,h}^{n_u}\,. \tag{5.11}$$

This also yields a bound for the associated multipliers:

Lemma 5.2. *Let (A1)–(A3) be satisfied. Then, there exist unique multipliers associated with* (5.11)

$$\left(\hat{\lambda}_h, \hat{\mu}_h\right) \in W_{1,\infty,h}^{n_x} \times L_{\infty,h}^{n_y} \tag{5.12}$$

satisfying the alternative necessary conditions (2.22)–(2.25)*, and there exists a constant* $\Gamma \geq 0$ *independent of* h *such that*

$$\left\|\hat{\lambda}_h - \hat{\lambda}\right\|_{1,1} \leq \Gamma\left(\upsilon + h\right), \quad \|\hat{\mu}_h - \hat{\mu}\|_\infty \leq \Gamma\left(\upsilon + h\right), \tag{5.13}$$

where $\upsilon > 0$ *is given in* (5.8).

The proof of Lemma 5.2 can be found in Appendix A.3. Finally, we are able to prove the main result of this section:

Theorem 5.3. *Let (A1)–(A3) be satisfied. Then, for every* $\zeta > 0$ *and sufficiently small* h*, Problem* (DOCP) *has a solution* $(\hat{x}_h, \hat{y}_h, \hat{u}_h) \in W_{1,\infty,h}^{n_x} \times L_{\infty,h}^{n_y} \times L_{\infty,h}^{n_u}$ *associated with unique multipliers* $\left(\hat{\lambda}_h, \hat{\mu}_h\right) \in W_{1,\infty,h}^{n_x} \times L_{\infty,h}^{n_y}$ *solving the alternative necessary conditions* (2.22)–(2.25) *such that*

$$\|\hat{x}_h - \hat{x}\|_{1,1} + \|\hat{y}_h - \hat{y}\|_1 + \|\hat{u}_h - \hat{u}\|_1 + \left\|\hat{\lambda}_h - \hat{\lambda}\right\|_{1,1} + \|\hat{\mu}_h - \hat{\mu}\|_\infty \leq \zeta\,. \tag{5.14}$$

Proof. Let $\zeta > 0$ be given. Then, according to Lemma 5.1 and Lemma 5.2 for sufficiently small $\upsilon, h > 0$, Problem (DOCP) and the alternative necessary conditions (2.22)–(2.25) have a solution (5.6) satisfying the bound (5.14). ☐

6. Error Estimates for Local Solutions

With the results of the previous sections we are now able to prove the main result of this paper, i.e., with the requirements of Theorem 3.4 we show that

$$\|\hat{x}_h - \hat{x}\|_{1,1} + \|\hat{y}_h - \hat{y}\|_1 + \|\hat{u}_h - \hat{u}\|_1 + \left\|\hat{\lambda}_h - \hat{\lambda}\right\|_{1,1} + \|\hat{\mu}_h - \hat{\mu}\|_\infty \leq \Gamma\, h\,,$$

where the constant $\Gamma \geq 0$ is independent of the mesh size h.

Proof of Theorem 3.4. By Theorem 4.2, the set-valued mapping $\mathbf{T} + \mathbf{F}$ defined in (4.6) is strongly metrically sub-regular at $\hat{\xi} = \left(\hat{x}, \hat{y}, \hat{u}, \hat{\lambda}, \hat{\mu}\right) \in \Xi$ for $0 \in \Omega$ with constants $\alpha, \beta, \gamma > 0$. Hence, for any $\omega \in \mathbf{B}_\Omega(0;\alpha)$ and for any solution $\xi \in \mathbf{B}_\Xi\left(\hat{\xi};\beta\right)$ of $\omega \in \mathbf{T}(\xi) + \mathbf{F}(\xi)$ we have

$$d_\Xi\left(\xi, \hat{\xi}\right) \leq \gamma\, d_\Omega(\omega, 0)\,. \tag{6.1}$$

In addition, according to Theorem 5.3, for sufficiently small h there exists a solution $\hat{\xi}_h = \left(\hat{x}_h, \hat{y}_h, \hat{u}_h, \hat{\lambda}_h, \hat{\mu}_h\right) \in \Xi$ of (DOCP) and the alternative necessary conditions (2.22)–(2.25) in $\mathbf{B}_\Xi\left(\hat{\xi};\beta\right)$. With the residual

$$\hat{\omega}_h := \left(\hat{\omega}_{f,h}, \hat{\omega}_{g,h}, \hat{\omega}_{\mathcal{H}_x,h}, \hat{\omega}_{\mathcal{H}_y,h}, \hat{\omega}_{\mathcal{H}_u,h}\right) \in \Omega \tag{6.2}$$

defined for $t \in (t_{i-1}, t_i]$, $i = 1, \dots, N$ by

$$\begin{aligned}
\hat{\omega}_{f,h}(t) &:= \dot{\hat{x}}_h(t) - f\left(\hat{x}_h(t), \hat{y}_h(t), \hat{u}_h(t)\right), && (6.3)\\
\hat{\omega}_{g,h}(t) &:= g\left(\hat{x}_h(t)\right), && (6.4)\\
\hat{\omega}_{\mathcal{H}_x,h}(t) &:= \dot{\hat{\lambda}}_h(t) + \nabla_x \mathcal{H}\left(\hat{\xi}_h(t)\right), && (6.5)\\
\hat{\omega}_{\mathcal{H}_y,h}(t) &:= \nabla_y \mathcal{H}\left(\hat{\xi}_h(t)\right) && (6.6)\\
&\quad -\nabla_y \mathcal{H}_h\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i), \hat{\lambda}_h(t_{i-1}), \hat{\mu}_h(t_{i-1})\right), &&\\
\hat{\omega}_{\mathcal{H}_u,h}(t) &:= \nabla_u \mathcal{H}\left(\hat{\xi}_h(t)\right) && (6.7)\\
&\quad -\nabla_u \mathcal{H}_h\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i), \hat{\lambda}_h(t_{i-1}), \hat{\mu}_h(t_{i-1})\right), &&
\end{aligned}$$

we have $\hat{\omega}_h \in \mathbf{T}\left(\hat{\xi}_h\right) + \mathbf{F}\left(\hat{\xi}_h\right)$. Hence, it remains to show $d_\Omega(\hat{\omega}_h, 0) \leq \Gamma\, h \leq \alpha$ for $\Gamma \geq 0$ and sufficiently small h. Then, (6.1) yields an error estimate for the discrete solution $\hat{\xi}_h$.

For the first component of $\hat{\omega}_h$ we obtain

$$
\begin{aligned}
\|\hat{\omega}_{f,h}\|_1 &= \sum_{i=1}^{N} \int\limits_{t_{i-1}}^{t_i} \left| \dot{\hat{x}}_h(t) - f\left(\hat{x}_h(t), \hat{y}_h(t), \hat{u}_h(t)\right) \right| \mathrm{d}t \\
&= \sum_{i=1}^{N} \int\limits_{t_{i-1}}^{t_i} \left| f\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) - f\left(\hat{x}_h(t), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) \right| \mathrm{d}t \\
&\leq \sum_{i=1}^{N} \int\limits_{t_{i-1}}^{t_i} \mathbb{L} \left| \hat{x}_h(t_i) - \hat{x}_h(t) \right| \mathrm{d}t \\
&\leq \sum_{i=1}^{N} \int\limits_{t_{i-1}}^{t_i} \mathbb{L} \left| f\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) \right| h \, \mathrm{d}t \\
&\leq \mathbb{L}\, \mathbb{M}\, h\,.
\end{aligned}
$$

The second component of $\hat{\omega}_h$ can be estimated by

$$
\left|\hat{\omega}_{g,h}(t)\right| = \left|g\left(\hat{x}_h(t)\right)\right| = \left|g\left(\hat{x}_h(t)\right) - g\left(\hat{x}_h(t_i)\right)\right| \leq \mathbb{L}\, \mathbb{M}\, h
$$

for $t \in (t_{i-1}, t_i]\,,\ i = 1, \ldots, N$. For its time derivative, we derive

$$
\begin{aligned}
\left\|\dot{\hat{\omega}}_{g,h}\right\|_1 &= \int\limits_{0}^{1} \left| g'\left(\hat{x}_h(t)\right) \dot{\hat{x}}_h(t) \right| \mathrm{d}t \\
&= \sum_{i=1}^{N} \int\limits_{t_{i-1}}^{t_i} \left| g'\left(\hat{x}_h(t)\right) \hat{x}_h'(t_i) - \frac{g\left(\hat{x}_h(t_i)\right) - g\left(\hat{x}_h(t_{i-1})\right)}{h} \right| \mathrm{d}t \\
&= \sum_{i=1}^{N} \int\limits_{t_{i-1}}^{t_i} \left| \int\limits_{0}^{1} \left[g'\left(\hat{x}_h(t)\right) - g'\left(\hat{x}_h(t_{i-1}) + \theta h \hat{x}_h'(t_i)\right)\right] \hat{x}_h'(t_i)\, \mathrm{d}\theta \right| \mathrm{d}t \\
&\leq 2\, \mathbb{L}\, \mathbb{M}^2 h\,.
\end{aligned}
$$

For $\hat{\omega}_{\mathcal{H}_x,h}$ and $t \in (t_{i-1}, t_i]$, $i = 1, \ldots, N$ we have

$$\begin{aligned}
\hat{\omega}_{\mathcal{H}_x,h}(t) &= \dot{\hat{\lambda}}_h(t) + \nabla_x \mathcal{H}\left(\hat{\xi}_h(t)\right) = \hat{\lambda}'_h(t_i) + \nabla_x \mathcal{H}\left(\hat{\xi}_h(t)\right) \\
&= \nabla_x \mathcal{H}\left(\hat{\xi}_h(t)\right) - \nabla_x \mathcal{H}_h\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i), \hat{\lambda}_h(t_{i-1}), \hat{\mu}_h(t_{i-1})\right) \\
&= \nabla_x f_0\left(\hat{x}_h(t), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) - \nabla_x f_0\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) \\
&\quad + \left[f'_x\left(\hat{x}_h(t), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) - f'_x\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right)\right]^\top \hat{\lambda}_h(t) \\
&\quad + f'_x\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right)^\top \left(\hat{\lambda}_h(t) - \hat{\lambda}_h(t_{i-1})\right) \\
&\quad + \Bigg[g''\left(\hat{x}_h(t)\right) - \frac{g'\left(\hat{x}_h(t_i)\right) - g'\left(\hat{x}_h(t_{i-1})\right)}{h} + \\
&\quad + \left(g'\left(\hat{x}_h(t)\right) - g'\left(\hat{x}_h(t_{i-1})\right)\right) f'_x\left(\hat{x}_h(t), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) + g'\left(\hat{x}_h(t_{i-1})\right) \\
&\qquad \left(f'_x\left(\hat{x}_h(t), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) - f'_x\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right)\right)\Bigg]^\top \hat{\mu}_h(t_i)\,,
\end{aligned}$$

which implies $\|\hat{\omega}_{\mathcal{H}_x,h}\|_1 \leq \Gamma_1 h$ for $\Gamma_1 \geq 0$ independent of h. Then, for $\hat{\omega}_{\mathcal{H}_y,h}$ and $t \in (t_{i-1}, t_i]$, $i = 1, \ldots, N$ we get

$$\begin{aligned}
\left|\hat{\omega}_{\mathcal{H}_y,h}(t)\right| &\leq \left|q\left(\hat{x}_h(t)\right) - q\left(\hat{x}_h(t_i)\right)\right| + \left|B\left(\hat{x}_h(t)\right) - B\left(\hat{x}_h(t_i)\right)\right| \left|\hat{\lambda}_h(t)\right| \\
&\quad + \left|B\left(\hat{x}_h(t_i)\right)\right| \left|\hat{\lambda}_h(t) - \hat{\lambda}_h(t_{i-1})\right| \\
&\quad + \left|B\left(\hat{x}_h(t)\right) - B\left(\hat{x}_h(t_i)\right)\right| \left|g\left(\hat{x}_h(t)\right)\right| \left|\hat{\mu}_h(t_{i-1})\right| \\
&\quad + \left|B\left(\hat{x}_h(t_i)\right)\right| \left|g\left(\hat{x}_h(t)\right) - g\left(\hat{x}_h(t_{i-1})\right)\right| \left|\hat{\mu}_h(t_{i-1})\right| \\
&\leq \Gamma_2 h\,,
\end{aligned}$$

and analogously for $\hat{\omega}_{\mathcal{H}_u,h}$ we obtain $\|\hat{\omega}_{\mathcal{H}_u,h}\|_\infty \leq \Gamma_3 h$ with $\Gamma_2, \Gamma_3 \geq 0$ independent of h. In conclusion, for sufficiently small h we have $d_\Omega\left(\hat{\omega}_h, 0\right) \leq \Gamma_4 h \leq \alpha$, where $\Gamma_4 \geq 0$ is independent of h. Thus, the inequality (6.1) yields

$$\begin{aligned}
&\|\hat{x}_h - \hat{x}\|_{1,1} + \|\hat{y}_h - \hat{y}\|_1 + \|\hat{u}_h - \hat{u}\|_1 + \left\|\hat{\lambda}_h - \hat{\lambda}\right\|_{1,1} + \|\hat{\mu}_h - \hat{\mu}\|_\infty \\
&\quad = d_\Xi\left(\hat{\xi}_h, \hat{\xi}\right) \leq \gamma d_\Omega\left(\hat{\omega}_h, 0\right) \leq \gamma\, \Gamma_4 h\,,
\end{aligned}$$

which completes the proof. □

7. Numerical Example

By the following illustrative example, we numerically confirm the error estimates of Theorem 3.4.

Example 7.1. We study the following modified minimum energy problem (cf. [18, Example 3.1.13]):

$$
\begin{aligned}
(\mathrm{P}_\omega) \qquad & \text{Minimize} && \int_0^1 \Big[x_1(t)^4 + (x_2(t)+1)^2\Big]\,\mathrm{d}t \\
& \text{subject to} && \dot{x}_1(t) = u(t) - y(t) && \text{a.e. in } [0,1]\,, \\
&&& \dot{x}_2(t) = \big[1 + \omega\, x_1(t)\big]u(t) && \text{a.e. in } [0,1]\,, \\
&&& \dot{x}_3(t) = -x_2(t) && \text{a.e. in } [0,1]\,, \\
&&& \;0 \quad = x_1(t) + x_3(t) && \text{in } [0,1]\,, \\
&&& x_1(0) = 0\,,\ x_2(0) = 1\,,\ x_3(0) = 0\,, \\
&&& u(t) \in [-3,-1] && \text{a.e. in } [0,1]\,.
\end{aligned}
$$

The system equation for $\dot{x}_2$ is control-affine and depends on some parameter $\omega \in \mathbb{R}$. Problem (P_ω) clearly is of type (OCP) with

$$
f_0\,(x,y,u) = x_1^4 + (x_2+1)^2\,, \quad f\,(x,y,u) = \begin{pmatrix} u-y \\ [1+\omega\, x_1]\,u \\ -x_2 \end{pmatrix}, \quad g\,(x) = x_1 + x_3\,.
$$

Since two differentiations of the algebraic constraint w.r.t. t are necessary to obtain a differential equation for the algebraic variable y, the DAE has index two. The Hamilton function (2.7) for Problem (P_ω) is now given by

$$
\begin{aligned}
\mathcal{H}\,(x,y,u,\lambda,\mu) = {} & x_1^4 + (x_2+1)^2 \\
& + \lambda_1\,(u-y) + \lambda_2\,(1+\omega\, x_1)\,u - \lambda_3 x_2 + \mu\,(u-y-x_2)\,.
\end{aligned}
$$

Therefore, we have

$$
\nabla_x \mathcal{H}\,(x,y,u,\lambda,\mu) = \begin{pmatrix} 4x_1^3 + \omega\,\lambda_2 u \\ 2\,(x_2+1) - \lambda_3 - \mu \\ 0 \end{pmatrix}
$$

and

$$
\nabla_{xx} \mathcal{H}\,(x,y,u,\lambda,\mu) = \begin{pmatrix} 12\,x_1^2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \nabla_{xu} \mathcal{H}\,(x,y,u,\lambda,\mu) = \begin{pmatrix} \omega\,\lambda_2 \\ 0 \\ 0 \end{pmatrix}.
$$

Since $\nabla_{xy}\mathcal{H}\,(x,y,u,\lambda,\mu) = 0$, the quadratic functional (3.7) in Assumption (A4) is given by

$$
\mathcal{P}\,(\delta x,\delta y,\delta u) := \int_0^1 \Big[6\,\hat{x}_1(t)^2\,(\delta x_1(t))^2 + (\delta x_2(t))^2 + \omega\,\delta x_1(t)\hat{\lambda}_2(t)\delta u(t)\Big]\,\mathrm{d}t\,.
$$

If ω is sufficiently small and $\hat{x}_1 \not\equiv 0$ on $[0,1]$, Assumption (A4) is fulfilled, i.e., there exists some constant η_2, such that

$$
2\mathcal{P}\,(\delta x,\delta y,\delta u) \geq -\eta_2\,\|\delta u\|_1^2
$$

for every $(\delta x,\delta y,\delta u) \in W_{1,1}^3 \times L_1 \times L_1$, especially for every $(\delta x,\delta y,\delta u)$ satisfying (3.9)–(3.12).

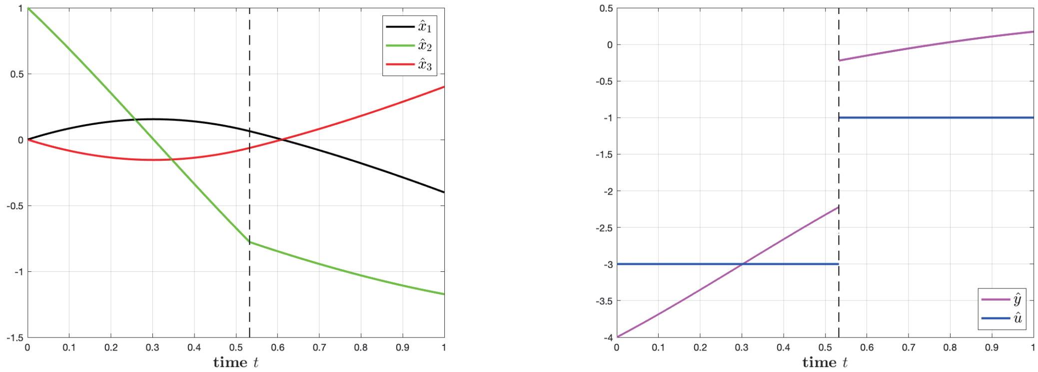

FIGURE 1. Optimal differential states $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$ (left), optimal algebraic state $\hat{y}$ and optimal control $\hat{u}$ (right) for $\omega = 1$, $N = 2048$.

TABLE 1. Discretization error for $\omega = 1$.

N	16	32	64	128	256	512
$\|\hat{u}_h - \hat{u}\|_1$	0.0813	0.0485	0.0240	0.0115	0.0051	0.0029
$\frac{\|\hat{u}_h - \hat{u}\|_1}{h}$	1.3000	1.5514	1.5389	1.4720	1.3173	1.4757

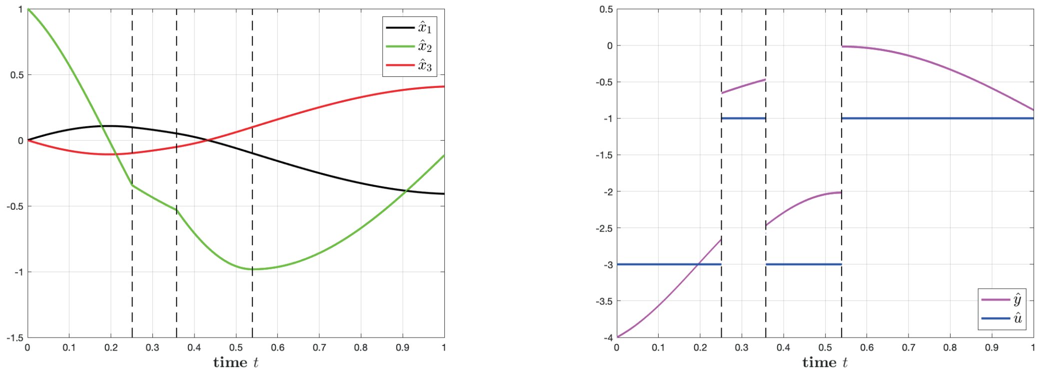

FIGURE 2. Optimal differential states $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_3$ (left), optimal algebraic state $\hat{y}$ and optimal control $\hat{u}$ (right) for $\omega = 10$, $N = 2048$.

We study the two cases where $\omega = 1$ and $\omega = 10$, respectively. The numerical solution of Problem (P_ω) is done by the interior point solver IPOPT, see [41]. We generated a fine-mesh reference solution for $N = 2^{11} = 2048$ discretization points. The optimal states and controls are shown in Figures 1 and 2. In both cases, $\hat{x}_1 \not\equiv 0$. Therefore, Assumption (A4) is fulfilled. The discretization errors for $N = 2^k$, $k = 4, \dots, 9$, depicted in Tables 1 and 2, indicate convergence of order one w.r.t. the mesh size $h = 1/N$ as predicted by Theorem 3.4.

Table 2. Discretization error for $\omega = 10$.

N	16	32	64	128	256	512
$\|\hat{u}_h - \hat{u}\|_1$	0.5109	0.3442	0.1915	0.0994	0.0483	0.0194
$\frac{\|\hat{u}_h - \hat{u}\|_1}{h}$	8.1746	11.0147	12.2559	12.7224	12.3754	9.9566

Appendix A. Proofs

A.1. **Proof of Theorem 4.2.** Let us introduce the abbreviations

$$\begin{aligned} &\hat{A}(\cdot) := f_x'[\cdot], && \hat{B}(\cdot) := B[\cdot], && \hat{C}(\cdot) := C[\cdot], \\ &\hat{G}(\cdot) := g'[\cdot], \\ &\hat{P}(\cdot) := \nabla^2_{xx}\mathcal{H}[\cdot], && \hat{Q}(\cdot) := \nabla^2_{xy}\mathcal{H}[\cdot], && \hat{R}(\cdot) := \nabla^2_{xu}\mathcal{H}[\cdot], \\ &\hat{p}(\cdot) := \nabla_x f_0[\cdot], && \hat{q}(\cdot) := q[\cdot], && \hat{r}(\cdot) := r[\cdot], \end{aligned} \tag{A.1}$$

and the alternative mappings

$$\mathbf{L}(\xi) := \begin{pmatrix} \left(\dot{x} - \dot{\hat{x}}\right) - \hat{A}(\cdot)(x - \hat{x}) - \hat{B}(\cdot)(y - \hat{y}) - \hat{C}(\cdot)(u - \hat{u}) \\ \hat{G}(\cdot)(x - \hat{x}) \\ \left(\dot{\lambda} - \dot{\hat{\lambda}}\right) + \nabla^2_{x\xi}\mathcal{H}[\cdot]\left(\xi - \hat{\xi}\right) \\ \nabla^2_{y\xi}\mathcal{H}[\cdot]\left(\xi - \hat{\xi}\right) \\ \nabla^2_{u\xi}\mathcal{H}[\cdot]\left(\xi - \hat{\xi}\right) + \nabla_u\mathcal{H}[\cdot] + \mathcal{N}_{\mathcal{U}}(u) \end{pmatrix} \tag{A.2}$$

$$\mathbf{S}(\xi) := \begin{pmatrix} \left(\dot{\hat{x}} - f(x,y,u)\right) + \hat{A}(\cdot)(x - \hat{x}) + \hat{B}(\cdot)(y - \hat{y}) + \hat{C}(\cdot)(u - \hat{u}) \\ (g(x) - g[\cdot]) - \hat{G}(\cdot)(x - \hat{x}) \\ (\nabla_x\mathcal{H}(\xi) - \nabla_x\mathcal{H}[\cdot]) - \nabla^2_{x\xi}\mathcal{H}[\cdot]\left(\xi - \hat{\xi}\right) \\ (\nabla_y\mathcal{H}(\xi) - \nabla_y\mathcal{H}[\cdot]) - \nabla^2_{y\xi}\mathcal{H}[\cdot]\left(\xi - \hat{\xi}\right) \\ (\nabla_u\mathcal{H}(\xi) - \nabla_u\mathcal{H}[\cdot]) - \nabla^2_{u\xi}\mathcal{H}[\cdot]\left(\xi - \hat{\xi}\right) \end{pmatrix}, \tag{A.3}$$

which satisfy

$$\mathbf{T}(\xi) + \mathbf{F}(\xi) = \mathbf{L}(\xi) + \mathbf{S}(\xi), \quad \xi \in \Xi. \tag{A.4}$$

We first show that for arbitrary $\bar{\alpha}, \bar{\beta} > 0$ there exist $\bar{\gamma} > 0$ depending only on $\mathbb{L}, \mathbb{M}, \eta$ such that $\mathbf{L}$ defined in (A.2) is SMsR at $\hat{\xi} \in \Xi$ for $0 \in \Omega$ with constants $\bar{\alpha}, \bar{\beta}, \bar{\gamma}$. To that end, let $\bar{\alpha}, \bar{\beta} > 0$ be arbitrary, $\omega \in \mathbf{B}_\Omega(0; \bar{\alpha})$, and let $\xi \in \mathbf{B}_\Xi\left(\hat{\xi}; \bar{\beta}\right)$ be a solution of $\omega \in \mathbf{L}(\xi)$. With the notation

$$\Delta\xi := \xi - \hat{\xi} = (\Delta x, \Delta y, \Delta u, \Delta\lambda, \Delta\mu), \quad \Delta\dot{x} = \dot{x} - \dot{\hat{x}}, \quad \Delta\dot{\lambda} = \dot{\lambda} - \dot{\hat{\lambda}} \tag{A.5}$$

the generalized equation $\omega \in \mathbf{L}(\xi)$ yields the perturbed system

$$\text{(A.6)} \quad \Delta\dot{x}(t) = \hat{A}(t)\Delta x(t) + \hat{B}(t)\Delta y(t) + \hat{C}(t)\Delta u(t) + \omega_f(t), \qquad \text{a.e. in } [0,1],$$

$$\text{(A.7)} \quad 0 = \hat{G}(t)\Delta x(t) - \omega_g(t), \qquad \text{in } [0,1],$$

$$\text{(A.8)} \quad 0 = D\Delta x(0),$$

$$\text{(A.9)} \quad \Delta\dot{\lambda}(t) = -\hat{P}(t)\Delta x(t) - \hat{Q}(t)\Delta y(t) - \hat{R}(t)\Delta u(t) - \hat{A}(t)^\top \Delta\lambda(t) - \left(\dot{\hat{G}}(t) + \hat{G}(t)\hat{A}(t)\right)^\top \Delta\mu(t) + \omega_{\mathcal{H}_x}(t), \qquad \text{a.e. in } [0,1],$$

$$\text{(A.10)} \quad 0 = \hat{Q}(t)^\top \Delta x(t) + \hat{B}(t)^\top \Delta\lambda(t) + \left(\hat{G}(t)\hat{B}(t)\right)^\top \Delta\mu(t) - \omega_{\mathcal{H}_y}(t), \qquad \text{a.e. in } [0,1],$$

$$\text{(A.11)} \quad 0 = \Delta\lambda(1),$$

$$\text{(A.12)} \quad 0 \in \hat{R}(t)^\top \Delta x(t) + \hat{C}(t)^\top \Delta\lambda(t) + \left(\hat{G}(t)\hat{C}(t)\right)^\top \Delta\mu(t) + \sigma(t) - \omega_{\mathcal{H}_u}(t) + \mathcal{N}_U(u(t)), \qquad \text{a.e. in } [0,1]$$

with the switching function σ defined in (3.3). Multiplying (A.10) from the left by $\Delta y(t)$ and (A.12) give us

$$\text{(A.13)} \quad 0 = \left\langle \Delta y(t), \hat{Q}(t)^\top \Delta x(t) + \hat{B}(t)^\top \Delta\lambda(t) + \left(\hat{G}(t)\hat{B}(t)\right)^\top \Delta\mu(t) - \omega_{\mathcal{H}_y}(t) \right\rangle,$$

$$\text{(A.14)} \quad 0 \geq \left\langle \Delta u(t), \hat{R}(t)^\top \Delta x(t) + \hat{C}(t)^\top \Delta\lambda(t) + \left(\hat{G}(t)\hat{C}(t)\right)^\top \Delta\mu(t) + \sigma(t) - \omega_{\mathcal{H}_u}(t) \right\rangle$$

for almost every $t \in [0,1]$. Furthermore, (A.7) at $t = 0$ and (A.8) together with (A2) imply

$$\text{(A.15)} \quad \Delta x(0) = \begin{pmatrix} \hat{G}(0) \\ D \end{pmatrix}^{-1} \begin{pmatrix} \omega_g(0) \\ 0 \end{pmatrix} =: \tilde{\omega}_{gD} \in \mathbb{R}^{n_x}.$$

Differentiating the algebraic equation (A.7) with respect to t yields

$$\text{(A.16)} \quad 0 = \left(\dot{\hat{G}}(t) + \hat{G}(t)\hat{A}(t)\right)\Delta x(t) + \hat{G}(t)\hat{B}(t)\Delta y(t) + \hat{G}(t)\hat{C}(t)\Delta u(t) - \dot{\omega}_g(t)$$

for almost every $t \in [0,1]$. Assumption (A2) allows us to solve (A.16) and (A.10) for Δy and $\Delta\mu$, respectively, to obtain

$$\text{(A.17)} \quad \Delta y(\cdot) = -\left(\hat{G}(\cdot)\hat{B}(\cdot)\right)^{-1}\left[\left(\dot{\hat{G}}(\cdot) + \hat{G}(\cdot)\hat{A}(\cdot)\right)\Delta x(\cdot) + \hat{G}(\cdot)\hat{C}(\cdot)\Delta u(\cdot) + \hat{G}(\cdot)\omega_f(\cdot) - \dot{\omega}_g(\cdot)\right],$$

$$\text{(A.18)} \quad \Delta\mu(\cdot) = -\left[\left(\hat{G}(\cdot)\hat{B}(\cdot)\right)^{-1}\right]^\top \left[\hat{Q}(\cdot)^\top \Delta x(\cdot) + \hat{B}(\cdot)^\top \Delta\lambda(\cdot) - \omega_{\mathcal{H}_y}(\cdot)\right].$$

Hence, for $\Delta\xi$ and the solution $(\delta x, \delta y, \delta u)$ of the unperturbed, linearized system (3.9)–(3.11) with $\delta u = \Delta u \in \mathcal{U} - \hat{u}$ we get

$$\begin{array}{ll} \|\Delta x - \delta x\|_\infty \le \Gamma_1 d_\Omega(\omega, 0), & \|\Delta y - \delta y\|_1 \le \Gamma_1 d_\Omega(\omega, 0), \\ \|\Delta x\|_\infty \le \Gamma_1 (\|\delta u\|_1 + d_\Omega(\omega, 0)), & \|\Delta y\|_1 \le \Gamma_1 (\|\delta u\|_1 + d_\Omega(\omega, 0)), \\ \|\Delta \lambda\|_\infty \le \Gamma_1 (\|\delta u\|_1 + d_\Omega(\omega, 0)), & \|\Delta \mu\|_\infty \le \Gamma_1 (\|\delta u\|_1 + d_\Omega(\omega, 0)) \end{array} \tag{A.19}$$

with $\Gamma_1 \ge 0$ depending only on $\mathbb{L}, \mathbb{M}$. Now, using (A.6), (A.9), (A.11), and (A.15) we have

$$\begin{aligned} -\langle \tilde{\omega}_{gD}, \Delta\lambda(0)\rangle &= \langle \Delta x(1), \Delta\lambda(1)\rangle - \langle \Delta x(0), \Delta\lambda(0)\rangle \\ &= \int_0^1 \frac{d}{dt}\langle \Delta x(t), \Delta\lambda(t)\rangle \,\mathrm{d}t \\ &= \int_0^1 \langle \Delta\dot{x}(t), \Delta\lambda(t)\rangle + \left\langle \Delta x(t), \Delta\dot{\lambda}(t)\right\rangle \mathrm{d}t \\ &= \int_0^1 \left\langle \hat{A}(t)\Delta x(t) + \hat{B}(t)\Delta y(t) + \hat{C}(t)\Delta u(t) + \omega_f(t), \Delta\lambda(t)\right\rangle \\ &\quad -\Big\langle \Delta x(t), \hat{P}(t)\Delta x(t) + \hat{Q}(t)\Delta y(t) + \hat{R}(t)\Delta u(t) + \hat{A}(t)^\top \Delta\lambda(t) \\ &\quad + \left(\dot{\hat{G}}(t) + \hat{G}(t)\hat{A}(t)\right)^\top \Delta\mu(t) - \omega_{\mathcal{H}_x}(t)\Big\rangle \,\mathrm{d}t \\ &= \int_0^1 \left\langle \hat{B}(t)\Delta y(t) + \hat{C}(t)\Delta u(t) + \omega_f(t), \Delta\lambda(t)\right\rangle \\ &\quad -\Big\langle \Delta x(t), \hat{P}(t)\Delta x(t) + \hat{Q}(t)\Delta y(t) + \hat{R}(t)\Delta u(t) \\ &\quad + \left(\dot{\hat{G}}(t) + \hat{G}(t)\hat{A}(t)\right)^\top \Delta\mu(t) - \omega_{\mathcal{H}_x}(t)\Big\rangle \mathrm{d}t\,. \end{aligned} \tag{A.20}$$

Integrating and subtracting (A.13), (A.14) from (A.20), and exploiting (A.16) yields

$$\begin{aligned} -\langle \tilde{\omega}_{gD}, \Delta\lambda(0)\rangle \le\ & \int_0^1 \langle \omega_f(t), \Delta\lambda(t)\rangle + \langle \dot{\omega}_g(t), \Delta\mu(t)\rangle \\ & + \langle \omega_{\mathcal{H}_x}(t), \Delta x(t)\rangle + \left\langle \omega_{\mathcal{H}_y}(t), \Delta y(t)\right\rangle + \langle \omega_{\mathcal{H}_u}(t), \Delta u(t)\rangle \,\mathrm{d}t \\ & - \int_0^1 \langle \sigma(t), \Delta u(t)\rangle \,\mathrm{d}t - 2\mathcal{P}(\Delta x, \Delta y, \Delta u)\,. \end{aligned}$$

Thus, using $|\tilde{\omega}_{gD}| \leq \Gamma_2 \|\omega_g\|_{1,1}$ and (A.19) we obtain

$$\begin{aligned}
&\int_0^1 \langle \sigma(t), \Delta u(t)\rangle \,\mathrm{d}t + 2\mathcal{P}(\Delta x, \Delta y, \Delta u) \\
&\leq \Gamma_3 \Big[\|\omega_g\|_{1,1} \|\Delta\lambda\|_\infty + \|\omega_f\|_1 \|\Delta\lambda\|_\infty + \|\omega_g\|_{1,1} \|\Delta\mu\|_\infty \\
&\quad + \|\omega_{\mathcal{H}_x}\|_1 \|\Delta x\|_\infty + \left\|\omega_{\mathcal{H}_y}\right\|_\infty \|\Delta y\|_1 + \|\omega_{\mathcal{H}_u}\|_\infty \|\Delta u\|_1\Big] \\
&\leq \Gamma_4 d_\Omega(\omega, 0)\left(\|\delta u\|_1 + d_\Omega(\omega, 0)\right)
\end{aligned} \tag{A.21}$$

for $\Gamma_2, \Gamma_3, \Gamma_4 \geq 0$ depending only on $\mathbb{L}, \mathbb{M}$. Additionally, the bounds (A.19) and $\delta u = \Delta u$ give us

$$\begin{aligned}
&|\mathcal{P}(\Delta x, \Delta y, \Delta u) - \mathcal{P}(\delta x, \delta y, \delta u)| \\
&\leq \frac{1}{2}\int_0^1 \Big|\Big\langle \Delta x(t), \hat{P}(t)\Delta x(t) + 2\hat{Q}(t)\Delta y(t) + 2\hat{R}(t)\Delta u(t)\Big\rangle \\
&\quad - \Big\langle \delta x(t), \hat{P}(t)\delta x(t) + 2\hat{Q}(t)\delta y(t) + 2\hat{R}(t)\delta u(t)\Big\rangle\Big| \,\mathrm{d}t \\
&= \frac{1}{2}\int_0^1 \Big|\Big\langle \Delta x(t) - \delta x(t), \hat{P}(t)\left(\Delta x(t) + \delta x(t)\right)\Big\rangle \\
&\quad +2\Big\langle \Delta x(t) - \delta x(t), \hat{Q}(t)\Delta y(t)\Big\rangle + 2\Big\langle \delta x(t), \hat{Q}(t)\left(\Delta y(t) - \delta y(t)\right)\Big\rangle \\
&\quad +2\Big\langle \Delta x(t) - \delta x(t), \hat{R}(t)\delta u(t)\Big\rangle\Big| \,\mathrm{d}t \\
&\leq \Gamma_5 d_\Omega(\omega, 0)\left(\|\delta u\|_1 + d_\Omega(\omega, 0)\right)
\end{aligned} \tag{A.22}$$

with $\Gamma_5 \geq 0$ depending only on $\mathbb{L}, \mathbb{M}$. Then, with (3.13), (A.21), and (A.22) we get

$$\begin{aligned}
\eta \|\delta u\|_1^2 &\leq \int_0^1 \langle \sigma(t), \delta u(t)\rangle \,\mathrm{d}t + 2\mathcal{P}(\delta x, \delta y, \delta u) \\
&\leq \int_0^1 \langle \sigma(t), \Delta u(t)\rangle \,\mathrm{d}t + 2\mathcal{P}(\Delta x, \Delta y, \Delta u) \\
&\quad +2\left|\mathcal{P}(\Delta x, \Delta y, \Delta u) - \mathcal{P}(\delta x, \delta y, \delta u)\right| \\
&\leq \Gamma_6 d_\Omega(\omega, 0)\left(\|\delta u\|_1 + d_\Omega(\omega, 0)\right)
\end{aligned}$$

with $\Gamma_6 \geq 0$ depending only on $\mathbb{L}, \mathbb{M}$. This yields the bound

$$\|\Delta u\|_1 = \|\delta u\|_1 \leq \frac{\Gamma_6 + \sqrt{\Gamma_6^2 + 4\Gamma_6}}{2\eta} d_\Omega(\omega, 0). \tag{A.23}$$

Hence, using (A.19) and the differential equations (A.6) and (A.9) gives us a constant $\bar{\gamma} > 0$ depending only on $\mathbb{L}, \mathbb{M}, \eta$ such that

$$d_\Xi\left(\xi, \hat{\xi}\right) \leq \bar{\gamma} d_\Omega(\omega, 0), \tag{A.24}$$

which proves the strong metric sub-regularity of $\mathbf{L}$.
Next, we aim to show that $\mathbf{L}+\mathbf{S}$ is SMsR at $\hat{\xi} \in \Xi$ for $0 \in \Omega$ using [32, Proposition 3.1], i.e., we need to find constants $\alpha, \beta, \gamma, \vartheta > 0$ such that

$$\vartheta\bar{\gamma} < 1, \quad \beta \le \bar{\beta}, \quad \alpha + \vartheta\beta \le \bar{\alpha}, \quad \gamma \ge \frac{\bar{\gamma}}{1-\vartheta\bar{\gamma}}, \tag{A.25}$$

and prove that the function $\mathbf{S}$ defined in (A.3) satisfies

$$\mathbf{S}\left(\hat{\xi}\right) = 0, \quad d_\Omega\left(\mathbf{S}\left(\xi\right), \mathbf{S}\left(\hat{\xi}\right)\right) \le \vartheta d_\Xi\left(\xi, \hat{\xi}\right) \quad \text{for all } \xi \in \mathbf{B}_\Xi\left(\hat{\xi}; \beta\right). \tag{A.26}$$

To that end, let $0 < \beta < \epsilon$ be sufficiently small with ϵ defined in Remark 3.1. Obviously, we have $\mathbf{S}(\hat{\xi}) = 0$. Let us consider the first component

$$\mathbf{S}_1\left(\xi\right) := -\left(f\left(x,y,u\right) - f\left(\hat{x},\hat{y},\hat{u}\right)\right) + \hat{A}(\cdot)\Delta x + \hat{B}(\cdot)\Delta y + \hat{C}(\cdot)\Delta u$$

of $\mathbf{S}\left(\xi\right)$. Using the mean-value theorem we get

$$\begin{aligned}\mathbf{S}_1\left(\xi\right) = -\int_0^1 \Big[&\left(f_x'\left(\left(\hat{x},\hat{y},\hat{u}\right) + \theta\left(\Delta x, \Delta y, \Delta u\right)\right) - f_x'\left(\hat{x},\hat{y},\hat{u}\right)\right)\Delta x \\ &+ \left(B\left(\hat{x}+\theta\Delta x\right) - B\left(\hat{x}\right)\right)\Delta y + \left(C\left(\hat{x}+\theta\Delta x\right) - C\left(\hat{x}\right)\right)\Delta u\Big] \mathrm{d}\theta.\end{aligned}$$

Hence, we obtain

$$\begin{aligned}\left\|\mathbf{S}_1\left(\xi\right)\right\|_1 &\le \int_0^1 \mathbb{L}\theta\left(\left\|\Delta x\right\|_{1,1} + \left\|\Delta y\right\|_1 + \left\|\Delta u\right\|_1\right)\left\|\Delta x\right\|_{1,1} \\ &\quad + \mathbb{L}\theta\left\|\Delta x\right\|_{1,1}\left\|\Delta y\right\|_1 + \mathbb{L}\theta\left\|\Delta x\right\|_{1,1}\left\|\Delta u\right\|_1 \mathrm{d}\theta \\ &\le \frac{\mathbb{L}}{2}\beta d_\Xi\left(\xi, \hat{\xi}\right).\end{aligned}$$

For the second component

$$\mathbf{S}_2\left(\xi\right) := g\left(x\right) - g\left(\hat{x}\right) - \hat{G}(\cdot)\Delta x$$

we have

$$\left\|\mathbf{S}_2\left(\xi\right)\right\|_1 = \int_0^1 \left|\int_0^1 \left(g'\left(\hat{x}(t) + \theta\Delta x(t)\right) - g'\left(\hat{x}(t)\right)\right)\Delta x(t)\mathrm{d}\theta\right| \mathrm{d}t \le \frac{\mathbb{L}}{2}\beta d_\Xi\left(\xi, \hat{\xi}\right).$$

Furthermore, we get the time derivative

$$\begin{aligned}\frac{d}{dt}\mathbf{S}_2(\xi)(t) &= g'(x(t))\dot{x}(t) - g'(\hat{x}(t))\dot{\hat{x}}(t) - g''(\hat{x}(t))\dot{\hat{x}}(t)\Delta x(t) - g'(\hat{x}(t))\Delta\dot{x}(t)\\ &= \left[g'(x(t)) - g'(\hat{x}(t))\right]\dot{x}(t) - g''(\hat{x}(t))\dot{\hat{x}}(t)\Delta x(t)\\ &= \left[g'(x(t)) - g'(\hat{x}(t))\right]\Delta\dot{x}(t) + \left[g'(x(t)) - g'(\hat{x}(t))\right]\dot{\hat{x}}(t)\\ &\quad - g''(\hat{x}(t))\dot{\hat{x}}(t)\Delta x(t)\\ &= \left[g'(x(t)) - g'(\hat{x}(t))\right]\Delta\dot{x}(t)\\ &\quad + \int_0^1 \left[g''(\hat{x}(t)+\theta\Delta x(t)) - g''(\hat{x}(t))\right]\dot{\hat{x}}(t)\Delta x(t)d\theta\end{aligned}$$

for almost every $t \in [0,1]$. Thus, we conclude

$$\left\|\frac{d}{dt}\mathbf{S}_2(\xi)\right\|_1 \leq \mathbb{L}\|\Delta x\|_{1,1}^2 + \frac{\mathbb{LM}}{2}\|\Delta x\|_{1,1}^2 \leq \frac{2\mathbb{L}+\mathbb{LM}}{2}\beta d_\Xi\left(\xi,\hat{\xi}\right),$$

and therefore

$$\|\mathbf{S}_2(\xi)\|_{1,1} \leq \frac{3\mathbb{L}+\mathbb{LM}}{2}\beta d_\Xi\left(\xi,\hat{\xi}\right).$$

Similar bounds can be found for the remaining components of $\mathbf{S}(\xi)$. Hence, we obtain

$$d_\Omega\left(\mathbf{S}(\xi),\mathbf{S}\left(\hat{\xi}\right)\right) = d_\Omega(\mathbf{S}(\xi),0) \leq \Gamma_7\beta d_\Xi\left(\xi,\hat{\xi}\right)$$

with Γ_7 depending only on $\mathbb{L}$, $\mathbb{M}$. Now, we choose $0 < \beta \leq \bar{\beta}$ sufficiently small such that for $\vartheta := \Gamma_7\beta$ we have

$$\vartheta\bar{\gamma} < 1, \quad \alpha := \bar{\alpha} - \vartheta\beta > 0, \quad \gamma := \frac{\bar{\gamma}}{1-\vartheta\bar{\gamma}} > 0,$$

i.e., conditions (A.25) and (A.26) are fulfilled. In conclusion, according to [32, Proposition 3.1], $\mathbf{T}+\mathbf{F} = \mathbf{L}+\mathbf{S}$ is SMsR at $\hat{\xi}$ for 0 with constants $\alpha,\beta,\gamma > 0$. □

A.2. **Proof of Lemma 5.1:** We need to verify conditions (P1)–(P4) in [13, Theorem 3.1]:

- **(P1)** $0 = \mathcal{T}(\hat{z}) + \hat{\omega}$ and $(\mathcal{T}-\mathcal{L})(\hat{z}) + \hat{\omega} \in \Pi$.
- **(P2)** For arbitrary $\vartheta > 0$ exists $0 < \tilde{\epsilon} \leq \epsilon$ such that for all $z_h^1, z_h^2 \in \mathbf{B_X}(\hat{z};\tilde{\epsilon})$
$$\left\|(\mathcal{T}-\mathcal{L})\left(z_h^1\right) - (\mathcal{T}-\mathcal{L})\left(z_h^2\right)\right\|_{\mathbf{Y}} \leq \vartheta d_{\mathbf{X}}\left(z_h^1, z_h^2\right).$$
- **(P3)** The map $\mathcal{L}^{-1}\colon \Pi \to \mathbf{X}_h$ is single-valued and Lipschitz continuous with constant $\kappa \geq 0$.
- **(P4)** $(\mathcal{T}-\mathcal{L})(\mathbf{B_X}(\hat{z};\tilde{\epsilon})) \subset \Pi$.

(P1): The first part is satisfied by definition of $\hat{\omega}$. For the second part, we have $(\mathcal{T}-\mathcal{L})(\hat{z}) + \hat{\omega} = \hat{\pi} + \hat{\omega}$ and $\Pi = \mathbf{B_Y}(\hat{\pi};\nu)$. Thus, we have to show that $\|\hat{\omega}\|_{\mathbf{Y}} \leq \nu$ for sufficiently small $h > 0$. To this end, we consider the first component $\hat{\omega}_1$ of $\hat{\omega}$ and exploit the mean value theorem, which yields

$$-\frac{\hat{x}(t_i) - \hat{x}(t_{i-1})}{h} + f[t_i] = \int_0^1 -f[t_{i-1}+\theta h] + f[t_i]\,\mathrm{d}\theta.$$

Since $(\hat{x},\hat{y},\hat{u}) \in BV_1^{n_x} \times BV^{n_y} \times BV^{n_u}$, we get $f[\cdot] \in BV^{n_x}$. This implies

$$|f[t_i] - f[t_{i-1}+\theta h]| \le |f[t_i] - f[t_{i-1}+\theta h]| + |f[t_{i-1}+\theta h] - f[t_{i-1}]| \le \bigvee_{t_{i-1}}^{t_i} f[\cdot]$$

for all $\theta \in (0,1)$ and $i = 1,\dots,N$. Thus, we obtain

$$\begin{aligned}\|\hat{\omega}_1\|_1 &= h\sum_{i=1}^{N}\left|\frac{\hat{x}(t_i)-\hat{x}(t_{i-1})}{h} - f[t_i]\right| \le h\sum_{i=1}^{N}\int_0^1 |f[t_{i-1}+\theta h] - f[t_i]|\,\mathrm{d}\theta \\ &\le h\sum_{i=1}^{N}\bigvee_{t_{i-1}}^{t_i} f[\cdot] \le h\bigvee_0^1 f[\cdot].\end{aligned}$$

Now, we consider the second component $\hat{\omega}_2$. We have

$$g(\hat{x}(t_i)) = 0, \quad i = 0,1,\dots,N, \qquad \frac{g(\hat{x}(t_i)) - g(\hat{x}(t_{i-1}))}{h} = 0, \quad i = 1,\dots,N,$$

and therefore $\|\hat{\omega}_2\|_{1,1} = 0$. Hence, we can choose h sufficiently small such that $\|\hat{\omega}\|_{\mathbf{Y}} \le h\bigvee_0^1 f[\cdot] \le \nu$, which verifies (P1).

(P2): Let $\vartheta > 0$ be given and $z_h^1 = \left(x_h^1, y_h^1\right), z_h^2 = \left(x_h^2, y_h^2\right) \in \mathbf{B_X}(\hat{z};\tilde{\epsilon})$ for $\tilde{\epsilon} > 0$ sufficiently small. Then, we have

$$\begin{aligned}&f\left(z_h^1(t_i),\hat{u}(t_i)\right) - f\left(z_h^2(t_i),\hat{u}(t_i)\right) \\ &\quad= \int_0^1 f'\left(z_h^2(t_i) + \theta\left(z_h^1(t_i) - z_h^2(t_i)\right), \hat{u}(t_i)\right)\left(z_h^1(t_i) - z_h^2(t_i)\right)\mathrm{d}\theta \\ &\quad= \int_0^1 f_x'\left(z_h^2(t_i) + \theta\left(z_h^1(t_i) - z_h^2(t_i)\right), \hat{u}(t_i)\right)\left(x_h^1(t_i) - x_h^2(t_i)\right)\mathrm{d}\theta \\ &\qquad+ \int_0^1 B\left(x_h^2(t_i) + \theta\left(x_h^1(t_i) - x_h^2(t_i)\right)\right)\left(y_h^1(t_i) - y_h^2(t_i)\right)\mathrm{d}\theta.\end{aligned}$$

Thus, for the first component of $(\mathcal{T}-\mathcal{L})\left(z_h^1\right)-(\mathcal{T}-\mathcal{L})\left(z_h^2\right)$ we obtain

$$\begin{aligned}
&h\sum_{i=1}^{N}\left|f\left(z_h^1(t_i),\hat{u}(t_i)\right)-f'\left[t_i\right]z_h^1(t_i)-\left(f\left(z_h^2(t_i),\hat{u}(t_i)\right)-f'\left[t_i\right]z_h^2(t_i)\right)\right|\\
&\quad\leq h\sum_{i=1}^{N}\int_0^1\left|f_x'\left(z_h^2(t_i)+\theta\left(z_h^1(t_i)-z_h^2(t_i)\right),\hat{u}(t_i)\right)-\hat{A}(t_i)\right|\left|x_h^1(t_i)-x_h^2(t_i)\right|\mathrm{d}\theta\\
&\qquad+h\sum_{i=1}^{N}\int_0^1\left|B\left(x_h^2(t_i)+\theta\left(x_h^1(t_i)-x_h^2(t_i)\right)\right)-\hat{B}\left(t_i\right)\right|\left|y_h^1(t_i)-y_h^2(t_i)\right|\mathrm{d}\theta\\
&\quad\leq\mathbb{L}\tilde{\epsilon}\left\|x_h^1-x_h^2\right\|_\infty+\mathbb{L}\tilde{\epsilon}\left\|y_h^1-y_h^2\right\|_1\leq\Gamma_1\tilde{\epsilon}d_{\mathbf{X}}\left(z_h^1,z_h^2\right)\leq\vartheta d_{\mathbf{X}}\left(z_h^1,z_h^2\right)
\end{aligned}$$

for $\Gamma_1\geq 0$ independent of h and sufficiently small $\tilde{\epsilon}>0$. For the second component we have

$$\begin{aligned}
&g\left(x_h^1(t_i)\right)-g\left(x_h^2(t_i)\right)-\hat{G}(t_i)\left(x_h^1(t_i)-x_h^2(t_i)\right)\\
&\quad=\int_0^1\left[g'\left(x_h^2(t_i)+\theta\left(x_h^1(t_i)-x_h^2(t_i)\right)\right)-g'\left(\hat{x}(t_i)\right)\right]\left(x_h^1(t_i)-x_h^2(t_i)\right)\mathrm{d}\theta\,,
\end{aligned}$$

for $i=0,\ldots,N$, which implies

$$\left|g\left(x_h^1(t_i)\right)-g\left(x_h^2(t_i)\right)-\hat{G}(t_i)\left(x_h^1(t_i)-x_h^2(t_i)\right)\right|\leq\frac{\mathbb{L}}{2}\tilde{\epsilon}d_{\mathbf{X}}\left(z_h^1,z_h^2\right).$$

We recall that for $t\in(t_{i-1},t_i]$, $i=1,\ldots,N$ we get $x_h(t)=x_h'(t_i)\left(t-t_{i-1}\right)+x_h(t_{i-1})$ and therefore $\dot{x}_h(t)=x_h'(t_i)$. Then, for the discrete derivative of the second component we obtain

$$
\begin{aligned}
&\Big(g\left(x_h^1(t_i)\right) - g\left(x_h^2(t_i)\right) - \hat{G}(t_i)\left(x_h^1(t_i) - x_h^2(t_i)\right)\Big)' \\
&\quad = \frac{1}{h}\int\limits_{t_{i-1}}^{t_i} \frac{\mathrm{d}}{\mathrm{d}t}\int\limits_0^1 \Big[g'\left(x_h^2(t) + \theta\left(x_h^1(t) - x_h^2(t)\right)\right) - \hat{G}(t)\Big]\left(x_h^1(t) - x_h^2(t)\right)\mathrm{d}\theta\mathrm{d}t \\
&\quad = \frac{1}{h}\int\limits_{t_{i-1}}^{t_i}\int\limits_0^1 \big[g''\left(x_h^2(t) + \theta\left(x_h^1(t) - x_h^2(t)\right)\right)\left(\dot{x}_h^2(t) + \theta\left(\dot{x}_h^1(t) - \dot{x}_h^2(t)\right)\right) \\
&\quad\quad - g''[t]\,\dot{\hat{x}}(t)\Big]\left(x_h^1(t) - x_h^2(t)\right) \\
&\quad\quad + \Big[g'\left(x_h^2(t) + \theta\left(x_h^1(t) - x_h^2(t)\right)\right) - \hat{G}(t)\Big]\left(\dot{x}_h^1(t) - \dot{x}_h^2(t)\right)\mathrm{d}\theta\mathrm{d}t \\
&\quad = \frac{1}{h}\int\limits_{t_{i-1}}^{t_i}\int\limits_0^1 \Big[g''\left(x_h^2(t) + \theta\left(x_h^1(t) - x_h^2(t)\right)\right)\Big(\dot{x}_h^2(t) + \theta\left(\dot{x}_h^1(t) - \dot{x}_h^2(t)\right) - \dot{\hat{x}}(t)\Big) \\
&\quad\quad + \left(g''\left(x_h^2(t) + \theta\left(x_h^1(t) - x_h^2(t)\right)\right) - g''[t]\right)\dot{\hat{x}}(t)\Big]\left(x_h^1(t) - x_h^2(t)\right) \\
&\quad\quad + \Big[g'\left(x_h^2(t) + \theta\left(x_h^1(t) - x_h^2(t)\right)\right) - \hat{G}(t)\Big]\left(\dot{x}_h^1(t) - \dot{x}_h^2(t)\right)\mathrm{d}\theta\mathrm{d}t\,,
\end{aligned}
$$

and thus

$$
\left\|\Big[g\left(x_h^1\right) - g\left(x_h^2\right) - \hat{G}(\cdot)]\left(x_h^1 - x_h^2\right)\Big]'\right\|_1 \leq (\mathbb{M} + \mathbb{L}\mathbb{M} + \mathbb{L})\,\tilde{\epsilon}\left\|x_h^1 - x_h^2\right\|_{1,1}.
$$

Hence, we have

$$
\begin{aligned}
&\left\|g\left(x_h^1\right) - g\left(x_h^2\right) - \hat{G}(\cdot)\left(x_h^1 - x_h^2\right)\right\|_1 \\
&\quad = \sum_{i=1}^{N}\int\limits_{t_{i-1}}^{t_i}\Big|\Big(g\left(x_h^1(t_i)\right) - g\left(x_h^2(t_i)\right) - \hat{G}(t_i)\left(x_h^1(t_i) - x_h^2(t_i)\right)\Big)'(t - t_{i-1}) \\
&\quad\quad + \Big(g\left(x_h^1(t_{i-1})\right) - g\left(x_h^2(t_{i-1})\right) - \hat{G}(t_{i-1})\left(x_h^1(t_{i-1}) - x_h^2(t_{i-1})\right)\Big)\Big|\mathrm{d}t \\
&\quad \leq \frac{(\mathbb{M} + \mathbb{L}\mathbb{M} + \mathbb{L})\,h + \mathbb{L}}{2}\tilde{\epsilon}d_{\mathbf{X}}\left(z_h^1, z_h^2\right).
\end{aligned}
$$

Summarizing, we obtain

$$
\left\|(\mathcal{T} - \mathcal{L})\left(z_h^1\right) - (\mathcal{T} - \mathcal{L})\left(z_h^2\right)\right\|_{\mathbf{Y}} \leq \vartheta d_{\mathbf{X}}\left(z_h^1, z_h^2\right)
$$

for sufficiently small $h, \tilde{\epsilon} > 0$.

(P3): We show that for sufficiently small h the linear system

$$
\mathcal{L}\left(z_h\right) = \pi
$$

has a unique solution $z_h(\pi)$ for all $\pi \in \Pi$, which satisfies

$$
d_{\mathbf{X}}\left(z_h(\pi^1), z_h(\pi^2)\right) \leq \kappa\left\|\pi^1 - \pi^2\right\|_{\mathbf{Y}}
$$

for all $\pi^1, \pi^2 \in \Pi$ and $\kappa \geq 0$ independent of h. To that end, for an arbitrary perturbation $\pi = (\pi_f, \pi_g) \in \Pi$ we consider the inhomogeneous system

$$x_h'(t_i) - \hat{A}(t_i)x_h(t_i) - \hat{B}\,(t_i)\,y_h(t_i) = \pi_f(t_i), \quad i = 1, \ldots, N, \quad x_h(t_0) = \hat{x}(0),$$
$$\hat{G}(t_i)x_h(t_i) = \pi_g(t_i), \quad i = 0, 1, \ldots, N.$$

For $i = 1, \ldots, N$ we have

$$\begin{aligned} \pi_g'(t_i) &= \frac{1}{h}\left(\hat{G}(t_i)x_h(t_i) - \hat{G}(t_{i-1})x_h(t_{i-1})\right) \\ &= \hat{G}(t_{i-1})x_h'(t_i) + \frac{\hat{G}(t_i) - \hat{G}(t_{i-1})}{h}x_h(t_i) \qquad \text{(A.27)} \\ &= \hat{G}(t_{i-1})\left(\hat{A}(t_i)x_h(t_i) + \hat{B}\,(t_i)\,y_h(t_i) + \pi_f(t_i)\right) + \frac{\hat{G}(t_i) - \hat{G}(t_{i-1})}{h}x_h(t_i). \end{aligned}$$

In addition, by (A2), for sufficiently small $h > 0$ the matrix

$$\hat{G}(t_{i-1})\hat{B}\,(t_i) = \hat{G}(t_i)\hat{B}\,(t_i) - \left(\hat{G}(t_i) - \hat{G}(t_{i-1})\right)B\,(t_i)$$

is non-singular for all $i = 1, \ldots, N$, since $\left|\hat{G}(t_i) - \hat{G}(t_{i-1})\right| \leq \mathbb{LM}h$. Hence, we can solve (A.27) for $y_h(t_i)$ and obtain the reduced difference equation

$$x_h'(t_i) = \tilde{A}_h(t_i)x_h(t_i) + \tilde{\pi}_f(t_i), \quad i = 1, \ldots, N, \quad x_h(t_0) = \hat{x}(0) \qquad \text{(A.28)}$$

with

$$\tilde{A}_h(t_i) = \hat{A}(t_i) - \hat{B}\,(t_i)\left(\hat{G}(t_{i-1})\hat{B}\,(t_i)\right)^{-1}\left(\frac{\hat{G}(t_i) - \hat{G}(t_{i-1})}{h} + \hat{G}(t_{i-1})\hat{A}(t_i)\right),$$
$$\tilde{\pi}_f(t_i) = \pi_f(t_i) - \hat{B}\,(t_i)\left(\hat{G}(t_{i-1})\hat{B}\,(t_i)\right)^{-1}\left(\hat{G}(t_{i-1})\pi_f(t_i) - \pi_g'(t_i)\right)$$

for $i = 1, \ldots, N$. Thus, for $\pi^1 = \left(\pi_f^1, \pi_g^1\right), \pi^2 = \left(\pi_f^2, \pi_g^2\right) \in \Pi$ we obtain

$$\left|x_h(\pi^1)(t_i) - x_h(\pi^2)(t_i)\right| \leq \Gamma_2 \left\|\pi^1 - \pi^2\right\|_{\mathbf{Y}}, \quad i = 0, \ldots, N,$$

which implies

$$\left\|x_h(\pi^1) - x_h(\pi^2)\right\|_1 \leq 3\left\|x_h(\pi^1) - x_h(\pi^2)\right\|_\infty \leq 3\Gamma_2 \left\|\pi^1 - \pi^2\right\|_{\mathbf{Y}}$$

for $\Gamma_2 \geq 0$ independent of h. Using the difference equation (A.28) yields

$$\left\|x_h'(\pi^1) - x_h'(\pi^2)\right\|_1 \leq \Gamma_3 \left\|\pi^1 - \pi^2\right\|_{\mathbf{Y}},$$

and therefore

$$\left\|x_h(\pi^1) - x_h(\pi^2)\right\|_{1,1} \leq (3\Gamma_2 + \Gamma_3)\left\|\pi^1 - \pi^2\right\|_{\mathbf{Y}},$$

where $\Gamma_3 \geq 0$ is independent of h. Exploiting

$$\begin{aligned} y_h(t_i) = -\left(\hat{G}(t_{i-1})\hat{B}\,(t_i)\right)^{-1}\Bigg[&\left(\frac{\hat{G}(t_i) - \hat{G}(t_{i-1})}{h} + \hat{G}(t_{i-1})\hat{A}(t_i)\right)x_h(t_i) \\ &+ \hat{G}(t_{i-1})\pi_f(t_i) - \pi_g'(t_i)\Bigg], \end{aligned}$$

for $i = 1, \dots, N$ gives us a constant $\Gamma_4 \geq 0$ independent of h such that

$$\left\| y_h(\pi^1) - y_h(\pi^2) \right\|_1 \leq \Gamma_4 \left\| \pi^1 - \pi^2 \right\|_{\mathbf{Y}}.$$

Summarizing, for $\kappa := 3\Gamma_2 + \Gamma_3 + \Gamma_4$ we have

$$d_{\mathbf{X}}\left(z_h(\pi^1), z_h(\pi^2)\right) = \left\| x_h(\pi^1) - x_h(\pi^2) \right\|_{1,1} + \left\| y_h(\pi^1) - y_h(\pi^2) \right\|_1 \leq \kappa \left\| \pi^1 - \pi^2 \right\|.$$

(P4): We recall $\hat{\pi} = (\mathcal{T} - \mathcal{L})(\hat{z})$ and $\Pi = \mathbf{B}_{\mathbf{Y}}(\hat{\pi}; \nu)$. Thus, using (P2) we get

$$\|(\mathcal{T} - \mathcal{L})(z_h) - \hat{\pi}\|_{\mathbf{Y}} = \|(\mathcal{T} - \mathcal{L})(z_h) - (\mathcal{T} - \mathcal{L})(\hat{z})\|_{\mathbf{Y}} \leq \vartheta d_{\mathbf{X}}(z_h, \hat{z}) \leq \vartheta \tilde{\epsilon} \leq \nu$$

for all $z_h \in \mathbf{B}_{\mathbf{X}}(\hat{z}; \tilde{\epsilon})$, if $\tilde{\epsilon} > 0$ is sufficiently small.

Now, applying [13, Theorem 3.1] yields a solution $\bar{z}_h = (\bar{x}_h, \bar{y}_h)$ solving (5.9) and satisfying the bound

$$d_{\mathbf{X}(\bar{z}_h, \hat{z})} \leq \frac{\kappa \|\hat{\omega}\|_{\mathbf{Y}}}{1 - \kappa\vartheta}$$

with $\vartheta > 0$ chosen such that $\kappa\vartheta < 1$. Exploiting $\|\hat{\omega}\|_{\mathbf{Y}} \leq \Gamma_5 h$ with $\Gamma_5 \geq 0$ independent of h (compare the proof of (P1)), gives us

$$\|\bar{x}_h - \hat{x}\|_{1,1} + \|\bar{y}_h - \hat{y}\|_1 \leq \frac{\kappa\Gamma_5}{1 - \kappa\vartheta} h.$$

Hence, the assertion follows for $\Gamma := \frac{\kappa\Gamma_5}{1-\kappa\vartheta}$. □

A.3. **Proof of Lemma 5.2.** By (A2) and Lemma 5.1, the alternative necessary conditions (2.22)–(2.25) have a unique solution $\left(\hat{\lambda}_h, \hat{\mu}_h\right) \in W^{n_x}_{1,\infty,h} \times L^{n_y}_{\infty,h}$ associated with $(\hat{x}_h, \hat{y}_h, \hat{u}_h)$. Moreover, we are able to solve the algebraic equations (2.9) and (2.23) for $\hat{\mu}$ and $\hat{\mu}_h$, respectively, to obtain

$$\hat{\mu}(t) = -\left(\left(\hat{G}(t)\hat{B}(t)\right)^{-1}\right)^{\top} \left[\hat{B}(t)^{\top}\hat{\lambda}(t) + \hat{q}(t)\right], \qquad t \in [0,1],$$

$$\text{(A.29)} \qquad \hat{\mu}_h(t_{i-1}) = -\left(\left(g'\left(\hat{x}_h(t_{i-1})\right) B\left(\hat{x}_h(t_i)\right)\right)^{-1}\right)^{\top} \left[B\left(\hat{x}_h(t_i)\right)^{\top} \hat{\lambda}_h(t_{i-1}) + q\left(\hat{x}_h(t_i)\right)\right], \qquad i = 1, \dots, N.$$

Inserting this into the respective differential and difference equations yields

$$\text{(A.30)} \qquad \begin{aligned} \dot{\hat{\lambda}}(t) &= -\bar{A}(t)^{\top}\hat{\lambda}(t) - \bar{\chi}(t), && \text{a.e. in } [0,1], && \hat{\lambda}(1) = 0, \\ \hat{\lambda}'_h(t_i) &= -\bar{A}_h(t_i)^{\top}\hat{\lambda}_h(t_{i-1}) - \bar{\chi}_h(t_i), && i = 1, \dots, N, && \hat{\lambda}_h(t_N) = 0, \end{aligned}$$

with

$$\begin{aligned} \bar{A}(\cdot) &= \hat{A}(\cdot) - \hat{B}(\cdot)\left(\hat{G}(\cdot)\hat{B}(\cdot)\right)^{-1}\left[\dot{\hat{G}}(\cdot) + \hat{G}(\cdot)\hat{A}(\cdot)\right], \\ \bar{\chi}(\cdot) &= \hat{p}(\cdot) - \left[\left(\hat{G}(\cdot)\hat{B}(\cdot)\right)^{-1}\left(\dot{\hat{G}}(\cdot) + \hat{G}(\cdot)\hat{A}(\cdot)\right)\right]^{\top} \hat{q}(\cdot), \\ \bar{A}_h(t_i) &= f'_x\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) \\ &\quad - B\left(\hat{x}_h(t_i)\right)\left(g'\left(\hat{x}_h(t_{i-1})\right) B\left(\hat{x}_h(t_i)\right)\right)^{-1} \tilde{g}'_{h,x}\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right), \\ \bar{\chi}_h(t_i) &= \nabla_x f_0\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right) \\ &\quad - \left[\left(g'\left(\hat{x}_h(t_{i-1})\right) B\left(\hat{x}_h(t_i)\right)\right)^{-1} \tilde{g}'_{h,x}\left(\hat{x}_h(t_i), \hat{y}_h(t_i), \hat{u}_h(t_i)\right)\right]^{\top} q\left(\hat{x}_h(t_i)\right) \end{aligned}$$

for $i = 1, \ldots, N$. Then, by Lemma 3.2 and Lemma 5.1, there exists $\Gamma_1 \geq 0$ independent of h such that

$$\left\|\bar{A}_h - \bar{A}\right\|_1 \leq \Gamma_1 \upsilon, \quad \left\|\bar{\chi}_h - \bar{\chi}\right\|_1 \leq \Gamma_1 \upsilon,$$
$$\left\|\hat{\lambda}_h\right\|_{1,1} \leq \Gamma_1, \quad \left\|\bar{A}\right\|_\infty \leq \Gamma_1, \quad \left\|\bar{A}_h\right\|_\infty \leq \Gamma_1, \quad \left\|\bar{\chi}\right\|_\infty \leq \Gamma_1, \quad \left\|\bar{\chi}_h\right\|_\infty \leq \Gamma_1.$$

We recall $\hat{\lambda}_h(t) = \hat{\lambda}_h'(t_k)(t - t_{k-1}) + \hat{\lambda}_h(t_{k-1})$ for $t \in (t_{k-1}, t_k]$, $k = 1, \ldots, N$. Thus, using $0 = \hat{\lambda}(1) = \hat{\lambda}_h(t_N)$ we obtain for $t \in (t_{i-1}, t_i]$

$$\begin{aligned}
\left|\hat{\lambda}(t) - \hat{\lambda}_h(t)\right| \leq & \int_t^1 \left|\bar{A}(\tau)\right| \left|\hat{\lambda}(\tau) - \hat{\lambda}_h(\tau)\right| \mathrm{d}\tau \\
& + \int_{t_{i-1}}^{t} \left|\bar{A}(\tau) - \bar{A}_h(t_i)\right| \left|\hat{\lambda}_h(t_{i-1})\right| + \left|\bar{A}(\tau)\right| \left|\hat{\lambda}_h(\tau) - \hat{\lambda}_h(t_{i-1})\right| \mathrm{d}\tau \\
& + \sum_{k=i}^{N} \int_{t_{k-1}}^{t_k} \left|\bar{A}(\tau) - \bar{A}_h(t_k)\right| \left|\hat{\lambda}_h(t_{k-1})\right| + \left|\bar{A}(\tau)\right| \left|\hat{\lambda}_h(\tau) - \hat{\lambda}_h(t_{k-1})\right| \mathrm{d}\tau \\
& + \int_{t_{i-1}}^{t} \left|\bar{\chi}(\tau) - \bar{\chi}_h(t_i)\right| \mathrm{d}\tau + \sum_{k=i}^{N} \int_{t_{k-1}}^{t_k} \left|\bar{\chi}(\tau) - \bar{\chi}_h(t_k)\right| \mathrm{d}\tau \\
\leq & \int_t^1 \left\|\bar{A}\right\|_\infty \left|\hat{\lambda}(\tau) - \hat{\lambda}_h(\tau)\right| \mathrm{d}\tau + 2\Gamma_1^2 (\upsilon + h) + 2\Gamma_1 \upsilon\,.
\end{aligned}$$

Hence, using the Gronwall Lemma yields

$$\left|\hat{\lambda}(t) - \hat{\lambda}_h(t)\right| \leq \Gamma_2 (\upsilon + h) \exp\left(\left\|\bar{A}\right\|_\infty (1 - t)\right) \leq \Gamma_3 (\upsilon + h)$$

for $\Gamma_2, \Gamma_3 \geq 0$ independent of h. Finally, exploiting (A.29) and (A.30) we get

$$\left\|\hat{\lambda} - \hat{\lambda}_h\right\|_{1,1} \leq \Gamma_4 (\upsilon + h)\,, \quad \left\|\hat{\mu} - \hat{\mu}_h\right\|_\infty \leq \Gamma_4 (\upsilon + h)$$

for $\Gamma_4 \geq 0$ independent of h, which proves the assertion. □

References

[1] A. A. Agrachev, G. Stefani,and P. Zezza, *Strong optimality for a bang-bang trajectory*, SIAM Journal on Control and Optimization **41** (2002), 991–1014.

[2] W. Alt, R. Baier, M. Gerdts and F. Lempio, *Error bounds for euler approximation of linear-quadratic control problems with bang-bang solutions*, Numerical Algebra, Control and Optimization **2** (2012), 547–570.

[3] W. Alt, U. Felgenhauer and M. Seydenschwanz, *Euler discretization for a class of nonlinear optimal control problems with control appearing linearly*, Computational Optimization and Applications **69** (2018), 825–856.

[4] W. Alt and U. Mackenroth, *Convergence of finite element approximations to state constrained convex parabolic boundary control problems*, SIAM Journal on Control and Optimization, **27** (1989), 718–736.

[5] W. Alt and C. Schneider, *Linear-quadratic control problems with L^1-control cost*, Optimal Control Applications and Methods **36** (2015), 512–534.
[6] W. Alt, C. Schneider and M. Seydenschwanz, *Regularization and implicit Euler discretization of linear-quadratic optimal control problems with bang-bang solutions*, Applied Mathematics and Computation **287–288** (2016), 104–124.
[7] W. Alt and M. Seydenschwanz, *An implicit discretization scheme for linear-quadratic control problems with bang-bang solutions*, Optimization Methods and Software **29** (2014), 535–560.
[8] A. Backes, *Extremalbedingungen für Optimierungs-Probleme mit Algebro-Differentialgleichungen*. PhD thesis, Humboldt-Universität zu Berlin, 2006.
[9] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. Advances in Design and Control. SIAM, 2nd edition, 2010.
[10] K. E. Brenan, S. L. Campell and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, volume 14 of *Classics In Applied Mathematics*. SIAM, 1996.
[11] M. Burger and M. Gerdts, *A survey on numerical methods for the simulation of initial value problems with sdaes*, in: Surveys in Differential-Algebraic Equations IV, Differential-Algebraic Equations Forum, A. Ilchmann and T. Reis (eds), Springer, 2017.
[12] A. L. Dontchev and W. W. Hager, *Lipschitzian stability in nonlinear control and optimization*, SIAM Journal on Control and Optimization **31** (1993), 569–603.
[13] A. L. Dontchev and W. W. Hager, *The Euler approximation in state constrained optimal control*, Mathematics of Computation **70** (2001), 173–203.
[14] A. L. Dontchev, W. W. Hager and K. Malanowski, *Error bounds for euler approximation of a state and control constrained optimal control problem*, Numerical Functional Analysis and Optimization **21** (2000), 653–682.
[15] A. L. Dontchev, W. W. Hager and V. M. Veliov, *Second-order Runge-Kutta approximations in control constrained optimal control*, SIAM Journal on Numerical Analysis **38** (2000), 202–226.
[16] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer, 2014.
[17] U. Felgenhauer, *On stability of bang-bang type Controls*, SIAM Journal on Control and Optimization **41** (2003), 1843–1867.
[18] M. Gerdts, *Optimal Control of ODEs and DAEs*, De Gruyter, 2012.
[19] E. Hairer and G. Wanner, *Solving ordinary differential equations II: Stiff and differential-algebraic problems*, volume 14 of *Springer Series in Computational Mathematics*, Springer, 2nd edition, 1996.
[20] J. L. Haunschmied, A. Pietrus and V. M. Veliov, *The Euler method for linear control systems revisited*, in: Proceedings of the 9th International Conference on Large-Scale Scientific Computations, Sozopol, 2013, pp. 90–97.
[21] D. Kraft, *FORTRAN-Programme zur numerischen Lösung optimaler Steuerungsprobleme*, volume 80 of *DFVLR-Mitteilung*, DFVLR, 1980.
[22] P. Kunkel and V. Mehrmann, *Differential-Algebraic Equations: Analysis and Numerical Solution*, EMS Textbooks in Mathematics. European Mathematical Society, 2006.
[23] K. Malanowski, C. Büskens and H. Maurer, *Convergence of approximations to nonlinear optimal control problems*, in: Mathematical Programming with Data Perturbations, Lecture Notes in Pure and Applied Mathematics, A. V. Fiacco (ed), CRC Press, 1997.
[24] B. Martens, *Necessary Conditions, Sufficient Conditions, and Convergence Analysis for Optimal Control Problems with Differential-Algebraic Equations*, PhD thesis, Universität der Bundeswehr München, 2019.
[25] B. Martens and M. Gerdts, *Convergence analysis of the implicit Euler-discretization and sufficient conditions for optimal control problems subject to index-one differential-algebraic equations*, Set-Valued and Variational Analysis **27** (2019), 405–431.
[26] B. Martens and M. Gerdts, *Convergence analysis for approximations of optimal control problems subject to higher index differential-algebraic equations and mixed control-state constraints*, SIAM Journal on Control and Optimization **58** (2000), 1–33.

[27] B. Martens and M. Gerdts, *Error analysis for the implicit Euler discretization of linear-quadratic control problems with higher index DAEs and bang-bang solutions*, in: Progress in Differential-Algebraic Equations II, Differential-Algebraic Equations Forum, T. Reis, S. Grundel, and S. Schöps (eds), Springer, 2020.

[28] B. Martens and M. Gerdts, *Convergence analysis for approximations of optimal control problems subject to higher index differential-algebraic equations and pure state constraints*, SIAM Journal on Control and Optimization **59** (2021),1903–1926, 2021.

[29] N. P. Osmolovskii and H. Maurer, *Equivalence of second order optimality conditions for bang-bang control problems. part 1: main results*, Control and Cybernetics **34** (2005).

[30] N. P. Osmolovskii and H. Maurer, *Equivalence of second order optimality conditions for bang-bang control problems. part 2: proofs, variational derivatives and representations*, Control and Cybernetics **36** (2007).

[31] N. P. Osmolovskii and H. Maurer, *Applications to Regular and Bang-Bang Control: Second-Order Necessary and Sufficient Optimality Conditions in Calculus of Variations and Optimal Control*, SIAM, 2012.

[32] N. P. Osmolovskii and V. M. Veliov, Metric sub-regularity in optimal control of affine problems with free end state, *ESAIM: Control, Optimisation and Calculus of Variations*, 2019.

[33] A. Pietrus, T. Scarinci, and V. M. Veliov, *High order discrete approximations to Mayer's problems for linear systems*, SIAM Journal on Control and Optimization **56** (2018), 102–119.

[34] T. Scarinci and V. M. Veliov, *Higher-order numerical scheme for linear-quadratic problems with bang-bang controls*, Computational Optimization and Applications **69** (2018), 403–422.

[35] C. Schneider and G. Wachsmuth, *Regularization and discretization error estimates for optimal control of ODEs with group sparsity*, ESAIM: Control, Optimisation and Calculus of Variations **24** (2018), 811–834.

[36] M. Seydenschwanz, *Convergence results for the discrete regularization of linear-quadratic control problems with bang–bang solutions*, Computational Optimization and Applications **61** (2015), 731–760.

[37] F. Tröltzsch, *Optimal Control of Partial Differential Equations*, volume 112 of *Graduate Studies in Mathematics*, American Mathematical Society, 2010.

[38] V. M. Veliov, *On the time-discretization of control systems*, SIAM Journal on Control and Optimization **35** (1997), 1470–1486.

[39] V. M. Veliov, *Error analysis of discrete approximations to bang-bang optimal control problems: the linear case*, Control and Cybernetics **34** (2005), 967–982.

[40] O. von Stryk, *Numerische Lösung optimaler Steuerungsprobleme: Diskretisierung, Parameteroptimierung und Berechnung der adjungierten Variablen*, PhD thesis, Technische Universität München, 1994.

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

Manuscript received August 28 2020
revised January 28 2021

B. Martens
Institute of Applied Mathematics and Scientific Computing, Department of Aerospace Engineering, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577 Neubiberg/Munich, Germany
E-mail address: `bjoern.martens@unibw.de`

C. Schneider
Ernst-Abbe-Hochschule Jena, University of Applied Sciences, Fachbereich Grundlagenwissenschaften, Carl-Zeiss-Promenade 2, 07745 Jena, Germany
E-mail address: `christopher.schneider@eah-jena.de`