ESAIM: COCV 24 (2018) 1705–1734
https://doi.org/10.1051/cocv/2017070

ESAIM: Control, Optimisation and Calculus of Variations
www.esaim-cocv.org

SUFFICIENCY AND SENSITIVITY FOR NONLINEAR OPTIMAL CONTROL PROBLEMS ON TIME SCALES *VIA* COERCIVITY *

Roman Šimon Hilscher[1,**] and Vera Zeidan[2]

Abstract. The main focus of this paper is to develop a sufficiency criterion for optimality in nonlinear optimal control problems defined on time scales. In particular, it is shown that the coercivity of the second variation together with the controllability of the linearized dynamic system are sufficient for the weak local minimality. The method employed is based on a direct approach using the structure of this optimal control problem. The second aim pertains to the sensitivity analysis for parametric control problems defined on time scales with separately varying state endpoints. Assuming a slight strengthening of the sufficiency criterion at a base value of the parameter, the perturbed problem is shown to have a weak local minimum and the corresponding multipliers are shown to be continuously differentiable with respect to the parameter. A link is established between (i) a modification of the shooting method for solving the associated boundary value problem, and (ii) the sufficient conditions involving the coercivity of the accessory problem, as opposed to the Riccati equation, which is also used for this task. This link is new even for the continuous time setting.

Mathematics Subject Classification. 49K15, 49K40, 34N05, 34K35, 90C31, 39A12.

Received March 1, 2017. Accepted October 24, 2017.

1. Introduction

Consider the nonlinear time scale optimal control problem

$$\text{minimize} \quad J(x,u) := K(x(a),x(b)) + \int_a^b L(t,x(t),u(t))\,\Delta t \tag{C}$$

subject to $x \in \mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}$ and $u \in \mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$ such that

$$x^{\Delta}(t) = f(t,x(t),u(t)), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{1.1}$$

$$\psi(t,u(t)) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{1.2}$$

$$\varphi(x(a),x(b)) = 0, \tag{1.3}$$

Keywords and phrases. Optimal control problem on time scales, Weak Pontryagin maximum principle, Weak local minimum, Coercivity, Sufficient optimality condition, Sensitivity analysis, Second variation, Controllability.

* *This research was supported by the Czech Science Foundation under grant GA16–00611S.*

[1] Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Kotlářská 2, CZ-61137 Brno, Czech Republic.

[2] Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, USA.

**Corresponding author: hilscher@math.muni.cz

where $\mathbb{T}$ is a bounded time scale, *i.e.*, an arbitrary nonempty compact subset of $\mathbb{R}$. Here we set $a := \min \mathbb{T}$ and $b := \max \mathbb{T}$ and define the time scale intervals by $[c,d]_{\mathbb{T}} := [c,d] \cap \mathbb{T}$. We assume that $\operatorname{card} \mathbb{T} \geq 2$, *i.e.*, $a < b$. For basic theory of dynamic equations on time scales we refer the reader to [9,10,25], and for piecewise rd-continuous and piecewise rd-continuously differentiable functions to [28]. We assume that $n, m, k, r \in \mathbb{N}$ are given dimensions with $k \leq m$ and $r \leq 2n$,

$$\begin{aligned} &L : [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, && K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \\ &f : [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, && \varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^r, \\ &\psi : [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^m \to \mathbb{R}^k, \end{aligned}$$

the state $x : [a,b]_{\mathbb{T}} \to \mathbb{R}^n$ is piecewise rd-continuously Δ-differentiable ($\mathrm{C}^1_{\mathrm{prd}}$) and the control $u : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^m$ is piecewise rd-continuous ($\mathrm{C}_{\mathrm{prd}}$). The regularity of the data will be specified in Section 2, see assumptions (A1) and (A2). The Hamiltonian corresponding to problem (C) is defined by

$$H(t,x,u,p,\lambda,\lambda_0) := p^T f(t,x,u) + \lambda_0\, L(t,x,u) + \lambda^T \psi(t,u). \tag{1.4}$$

A pair (x,u) is *feasible* for problem (C) if $x \in \mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}$ and $u \in \mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$ and it satisfies (1.1)–(1.3). A feasible pair $(\hat{x},\hat{u})$ is a *weak local minimum* for (C) if there exists $\delta > 0$ such that for any feasible pair (x,u) with $\|(x-\hat{x}, u-\hat{u})\|_{\mathrm{C}_{\mathrm{prd}}} < \delta$ we have $J(\hat{x},\hat{u}) \leq J(x,u)$, where

$$\|z\|_{\mathrm{C}_{\mathrm{prd}}} := \sup_{t \in [a,\rho(b)]_{\mathbb{T}}} |z(t)|.$$

If the inequality $J(\hat{x},\hat{u}) < J(x,u)$ holds for all such feasible pairs $(x,u) \neq (\hat{x},\hat{u})$, then $(\hat{x},\hat{u})$ is a *strict weak local minimum* for (C).

The formulation of problem (C) encompasses as special cases both the continuous time optimal control problem for $[a,b]_{\mathbb{T}} = [a,b]$ and the discrete time optimal control problem for $[a,b]_{\mathbb{T}} = [0,N+1]_{\mathbb{Z}} = \{0,1,\dots,N+1\}$. From this point of view the study of optimal control problems on time scales provides not only a unification, but also a valuable insight into the connections between these two extreme time domains. In many cases we can obtain new results even for these two special domains. More importantly, in the context of general time scales we can study "hybrid" problems, which are neither purely continuous nor purely discrete, but a combination of both. For instance, this is the case of problems in quantum calculus [32,41], impulsive systems [29], or some economic and population models [3,7,8]. See also the references in the above quoted publications.

Regarding the question of optimality conditions for problem (C), the focus during the recent years was concentrated on deriving first and second order necessary conditions, such as the weak Pontryagin maximum principle and the accessory problem in [31,53] and the strong Pontryagin-type maximum principle in [16,17]. In the special case of the calculus of variations on time scales, that is, when $f(t,x,u) = u$ and $\psi(t,u) \equiv 0$, necessary as well as sufficient optimality conditions are obtained in [28] in terms of the Euler–Lagrange equation and the nonnegativity, respectively the coercivity, of the second variation. Furthermore, in [30] the authors derived other sufficient conditions for this problem in terms of the associated time scale Riccati matrix equation and the strengthened Legendre condition. In the optimal control setting on time scales, *i.e.*, for problem (C), sufficient optimality conditions are derived in [54] in terms of the Hamilton–Jacobi equation. For linear-quadratic optimal control problems on time scales, sufficient optimality conditions are known also in [54] and in [11–13] in terms of the Riccati equation. However, the question of the sufficiency criteria for the nonlinear problem (C) in terms of the second variation is completely open.

The first aim of this paper is to develop a sufficient condition for the weak local minimum of problem (C) in terms of the coercivity of the second variation (see Thm. 3.3), thus generalizing [28], (Thm. 2) from the calculus of variations to the optimal control setting. To obtain the sufficiency criterion in terms of the coercivity of the second variation we could not use the known results derived for the abstract setting of infinite dimensional optimization problems over a Banach space, see *e.g.* [22] with L^∞ controls, because the optimality in our problem (C) is over the piecewise rd-continuous controls, which do not form a Banach space. For this reason, the direct approach used in this paper is inevitable.

Optimal control problems with equality type control constraints have been well-studied for both discrete and continuous time problems (see *e.g.*, [27, 44, 57]), as well as for the time scales setting (see *e.g.*, [31]).

In the continuous time setting, sufficiency criteria for optimal control problems with piecewise continuous controls are derived *via* the Hamilton–Jacobi theory, the Riccati differential equation, or the method of quadratic approximations (coercivity of the second variation), see [2, 14, 15, 33, 45–47, 49]. In the present paper we follow the latter mentioned approach.

The second aim of this paper is to study the sensitivity of the weak local minimum of problem (C) with respect to parameters (see Thm. 4.3). Problem ($\mathrm{C}_{\hat\omega}$) in Section 4, corresponding to a fixed parameter $\hat\omega$, is considered to be the unperturbed problem and it is assumed that it has a weak local minimum $(\hat x, \hat u)$. An important problem in sensitivity analysis is the following: Find conditions on the unperturbed optimal solution $(\hat x, \hat u)$ such that the perturbed problem (C_ω) admits an optimal solution $\big(x(\cdot,\omega), u(\cdot,\omega)\big)$ near $(\hat x, \hat u)$ that is a continuously differentiable function of the parameter ω near $\hat\omega$. It is well documented that this type of sensitivity results in optimization and optimal control are strongly based on having second-order sufficient optimality condition that are stable under small perturbation.

In the case of *discrete* time setting, the optimal control can be formulated as a mathematical programming problem in finite dimension, and known sensitivity results in mathematical programming are adapted to them, see [23]. For the continuous time setting one approach to sensitivity analysis is to apply results obtained for abstract optimization problems over Banach or Hilbert spaces, see *e.g.* [36] and [22]. Since our optimal control problem is *not* over a Banach space, this approach cannot be used here. There is a second approach that directly deals with the second-order sensitivity analysis for nonlinear control problems. It mainly concerns with developing neighboring feedback schemes for perturbed solutions, which lead to having to solve a boundary value problem (BVP) *via* the shooting method. The main idea goes back to [18, 19], and was followed by many papers which suffer by being formal. A rigorous result in this direction was given in [42], which was generalized in [37]. However, the sufficient conditions used in [37, 42] involved the existence of a solution of a Riccati differential equation, which also served in [37] to prove that the iteration matrix of the shooting method is nonsingular.

In this paper we intend to use the approach of shooting method for the general *time scale* setting. Since our sufficiency criterion involves the coercivity of the second variation instead of the Riccati equation, a serious modification to this approach is necessary to be made in order to solve the time scale (BVP). Furthermore, unlike the problems in [37, 42] we allow *both* endpoints to vary, and hence our result here is more general than those in [37, 42], even when we specialize it to the continuous time setting.

The paper is organized as follows. In Section 2 we introduce notation and state main assumptions and known results about problem (C). In Section 3 we discuss the coercivity of the second variation of the functional J and its role in sufficiency criteria for problem (C). In Section 4 we perform the sensitivity analysis for problem (C) depending on a parameter and with separated endpoints. In Section 5 contains conclusions and remarks about the presented theory. Finally, in Appendix A we present the proofs of two important approximation lemmas (Lem. 3.4 and 3.6).

Remark 1.1. Optimization problems over time scales are often considered in the form with the shifted state $x^\sigma(t) := x(\sigma(t))$ instead of $x(t)$, where $\sigma(t)$ is the forward jump operator at t. This form is known in the calculus of variations setting [5, 6, 20, 28–30, 38–40] as well as in the optimal control setting [31, 53, 54]. In the latter case the shifted state $x^\sigma(t)$ appears in the Lagrangian L and the dynamics f. We denote such an optimal control problem by (C^σ). The form of our problem (C) with $x(t)$ only is however more natural, since the equation of motion at the instant t does not depend on the future values of the state, *i.e.*, on the values at $\sigma(t)$. Nevertheless, in view of ([54], Sect. 3) the optimal control problems (C) and (C^σ) are *equivalent* by using a transformation involving the implicit function theorem. In this case the resulting costate function $\hat p(t)$ in formulas (2.11), (2.12), and (2.16) below appears *without the shift*, see ([31], Sects. 6 and 7 and [54], Sect. 6). With this understanding the main results of this paper apply also to such a problem (C^σ).

2. Optimal control problem on time scales

In this section we recall basic notions and results about the optimal control problem (C). The objectives are to state and analyze the main assumptions and to recall known results used in the rest of this paper: the M-controllability, the weak Pontryagin maximum principle on time scales, and the second variation of problem (C).

Let $(\hat{x}, \hat{u})$ be a feasible pair for problem (C). Consider the matrices $\mathcal{A}(t) \in \mathbb{R}^{n\times n}$, $\mathcal{B}(t) \in \mathbb{R}^{n\times m}$, $M \in \mathbb{R}^{r\times 2n}$, and $N(t) \in \mathbb{R}^{k\times m}$ given by

$$\mathcal{A}(t) := f_x(t, \hat{x}(t), \hat{u}(t)), \quad \mathcal{B}(t) := f_u(t, \hat{x}(t), \hat{u}(t)), \tag{2.1}$$

$$M := \nabla\varphi(\hat{x}(a), \hat{x}(b)), \quad N(t) := \nabla_u\, \psi(t, \hat{u}(t)). \tag{2.2}$$

For $\varepsilon > 0$ we define the ε-tube about the function $(\hat{x}, \hat{u})$ as

$$T_\varepsilon(\hat{x}, \hat{u}) := \big\{(t, x, u) \in [a, \rho(b)]_{\mathbb{T}} \times \mathbb{R}^n \times \mathbb{R}^m \text{ such that } \big|(x, u) - (\hat{x}(t), \hat{u}(t))\big| < \varepsilon\big\}, \tag{2.3}$$

the projection of $T_\varepsilon(\hat{x}, \hat{u})$ onto $\mathbb{R}^n \times \mathbb{R}^m$ as

$$\mathcal{P}T_\varepsilon(\hat{x}, \hat{u}) := \big\{(x, u) \in \mathbb{R}^n \times \mathbb{R}^m \text{ such that } \exists\, t \in [a, \rho(b)]_{\mathbb{T}} : \big|(x, u) - (\hat{x}(t), \hat{u}(t))\big| < \varepsilon\big\},$$

and the ε-ball about the vector $\hat{y} \in \mathbb{R}^j$ by

$$B_\varepsilon(\hat{y}) := \big\{y \in \mathbb{R}^j \text{ such that } |y - \hat{y}| < \varepsilon\big\}.$$

If (x, u) is another feasible pair for problem (C), then the notation $(x, u) \in T_\varepsilon(\hat{x}, \hat{u})$ means that $(t, x(t), u(t)) \in T_\varepsilon(\hat{x}, \hat{u})$ for all $t \in [a, \rho(b)]_{\mathbb{T}}$. Here $\rho(t)$ is the backward jump operator at t and $\mu(t) := \sigma(t) - t$ is the graininess function at the point t.

The assumptions on the regularity of the data $h(t, x, u) := (L(t, x, u), f(t, x, u), \psi(t, u))$, $K(x, y)$, and $\varphi(x, y)$ of problem (C) near the feasible pair $(\hat{x}, \hat{u})$ are summarized in the following.

(A1) There exists $\varepsilon_1 > 0$ such that the functions $K(\cdot)$ and $\varphi(\cdot)$ are continuously differentiable on $B_{\varepsilon_1}\big(\hat{x}(a), \hat{x}(b)\big)$; the function $h(t, \cdot, \cdot)$ is differentiable in (x, u) on $B_{\varepsilon_1}\big(\hat{x}(t), \hat{u}(t)\big)$; the functions $h(t, \cdot, \cdot)$ and $\nabla_{(x,u)} h(t, \cdot, \cdot)$ are continuous at $(\hat{x}, \hat{u})$ uniformly in t; and for (x, u) in $\mathcal{P}T_{\varepsilon_1}(\hat{x}, \hat{u})$ the functions $h(\cdot, x, u)$ and $\nabla_{(x,u)} h(\cdot, x, u)$ are rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$; the $n \times n$ matrix $I + \mu(t)\, f_x(t, \hat{x}(t), \hat{u}(t))$ is invertible for all $t \in [a, \rho(b)]_{\mathbb{T}}$; the matrices M and $N(t)$ for $t \in [a, \rho(b)]_{\mathbb{T}}$ have full rank.

(A2) There exists $\varepsilon_1 > 0$ such that the functions $K(\cdot)$ and $\varphi(\cdot)$ are twice continuously differentiable on $B_{\varepsilon_1}\big(\hat{x}(a), \hat{x}(b)\big)$; the function $h(t, \cdot, \cdot)$ is twice differentiable in (x, u) on $B_{\varepsilon_1}\big(\hat{x}(t), \hat{u}(t)\big)$; the functions $h(t, \cdot, \cdot)$, $\nabla_{(x,u)} h(t, \cdot, \cdot)$, and $\nabla^2_{(x,u)} h(t, \cdot, \cdot)$ are continuous at $(\hat{x}, \hat{u})$ uniformly in t; and for (x, u) in $\mathcal{P}T_{\varepsilon_1}(\hat{x}, \hat{u})$ the functions $h(\cdot, x, u)$, $\nabla_{(x,u)} h(\cdot, x, u)$, and $\nabla^2_{(x,u)} h(\cdot, x, u)$ are rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$; the $n \times n$ matrix $I + \mu(t)\, f_x(t, \hat{x}(t), \hat{u}(t))$ is invertible for all $t \in [a, \rho(b)]_{\mathbb{T}}$; the matrices M and $N(t)$ for $t \in [a, \rho(b)]_{\mathbb{T}}$ have full rank.

Remark 2.1. We recall from Definition 2 of [28] that $h(t, \cdot, \cdot)$ is continuous in (x, u) at the feasible pair $(\hat{x}, \hat{u})$ uniformly in t on $[a, \rho(b)]_{\mathbb{T}}$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $t \in [a, \rho(b)]_{\mathbb{T}}$ and all $(x, u) \in \mathbb{R}^n \times \mathbb{R}^m$ with $0 < |(x, u) - (\hat{x}(t), \hat{u}(t))| < \delta$ we have $\big|h(t, x, u) - h(t, \hat{x}(t), \hat{u}(t))\big| < \varepsilon$.

Remark 2.2. The assumptions involving the function h in (A1) are automatically satisfied, whenever $h(\cdot, \cdot, \cdot)$ and $\nabla_{(x,u)} h(\cdot, \cdot, \cdot)$ are $\mathrm{C_{rd}} \times \mathrm{C} \times \mathrm{C}$ continuous on the tube $T_\varepsilon(\hat{x}, \hat{u})$, see ([28], Def. 3 and Cor. 1). Similarly, the assumptions involving the function h in (A2) are automatically satisfied, whenever $h(\cdot, \cdot, \cdot)$, $\nabla_{(x,u)} h(\cdot, \cdot, \cdot)$, and $\nabla^2_{(x,u)} h(\cdot, \cdot, \cdot)$ are $\mathrm{C_{rd}} \times \mathrm{C} \times \mathrm{C}$ continuous on the tube $T_\varepsilon(\hat{x}, \hat{u})$. In particular, if the function $h(\cdot, \cdot, \cdot)$ is autonomous (independent of t) and of class C^1, resp. C^2, on $\mathbb{R}^n \times \mathbb{R}^m$, then the regularity assumption (A1), resp. (A2), is satisfied.

Remark 2.3. The assumption of $N(t)$ invertible is customary in optimization in order $\hat{u}$ to be a regular point for the control constraints ([35], Chap. 10). Moreover, inspired by the notion of piecewise continuous invertible matrices in the continuous time setting, the assumption in (A1) and (A2) that $N(t)$ has full rank for all $t \in [a, \rho(b)]_{\mathbb{T}}$ also includes the full rank of the one-sided limits $N(t^{\pm})$ at left-dense and right dense points $t \in [a, \rho(b)]_{\mathbb{T}}$. As a consequence, we obtain that the inverse matrix function $[N(t)\, N^T(t)]^{-1}$ is also piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$.

Remark 2.4. We note that if the point b is left-scattered and $\varepsilon > 0$ and the tube $T_\varepsilon(\hat{x}, \hat{u})$ are given, then there exists $\delta \in (0, \varepsilon]$ such that for every feasible pair $(x, u) \in T_\varepsilon(\hat{x}, \hat{u})$ with $(x^\rho(b), u^\rho(b)) \in B_\delta(\hat{x}^\rho(b), \hat{u}^\rho(b))$ we have $x(b) \in B_{\delta+\varepsilon\mu(\rho(b))}(\hat{x}(b)) \subseteq B_{\varepsilon[1+\mu(\rho(b))]}(\hat{x}(b))$. This means that for $(x, u) \in T_\varepsilon(\hat{x}, \hat{u})$ the distance $x(b) - \hat{x}(b)$ is also controlled when the point b is left-scattered, even though the values $x(b)$ and $\hat{x}(b)$ are not in this case explicitly present in the expression for $T_\varepsilon(\hat{x}, \hat{u})$ in (2.3).

In Lemma 3 of [28], we established conditions which guarantee the composition of $h(t, \cdot, \cdot)$ with $(\hat{x}(\cdot), \hat{u}(\cdot))$ to be piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$.

Lemma 2.5. *Let $(\hat{x}, \hat{u})$ be feasible for (C) and h be defined on a tube $T_\varepsilon(\hat{x}, \hat{u})$. If $h(t, \cdot, \cdot)$ is continuous at $(\hat{x}, \hat{u})$ uniformly in t on $[a, \rho(b)]_{\mathbb{T}}$ and if for $(x, u) \in \mathcal{P}T_\varepsilon(\hat{x}, \hat{u})$ the function $h(\cdot, x, u)$ is rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$, then $h(\cdot, \hat{x}(\cdot), \hat{u}(\cdot))$ is piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$.*

Remark 2.6. A close look at the proof of Lemma 3 of [28] easily shows that in the above lemma the rd-continuity of $h(\cdot, x, u)$ for any $(x, u) \in \mathcal{P}T_\varepsilon(\hat{x}, \hat{u})$ can be weakened by only assuming

(i) for any right-dense point $t_r \in [a, b)_{\mathbb{T}}$ at which the control $\hat{u}$ is continuous, the function $h(\cdot, \hat{x}(t_r), \hat{u}(t_r))$ is continuous at t_r;
(ii) at any right-dense $t_r \in [a, b)_{\mathbb{T}}$ and any left-dense point $t_l \in (a, b]_{\mathbb{T}}$, the one-sided limits $\lim_{t \to t_r^+} h(t, \hat{x}(t_r), \hat{u}(t_r^+))$ and $\lim_{t \to t_l^-} h(t, \hat{x}(t_l), \hat{u}(t_l^-))$ exist and are finite.

In the next lemma we show that the conclusion of Lemma 2.5 can be transfered into functions (x, u) sufficiently close to $(\hat{x}, \hat{u})$. This result then implies that the assumption on the integrability of $h(\cdot, x(\cdot), u(\cdot))$ and $\nabla_{(x,u)} h(\cdot, x(\cdot), u(\cdot))$ used in assumption (A1) and (A2) from [31] can be dropped.

Lemma 2.7. *Let $h(\cdot, \cdot, \cdot)$ be defined on $T_{\varepsilon_1}(\hat{x}, \hat{u})$. If $\nabla_{(x,u)} h(t, \cdot, \cdot)$ is continuous at $(\hat{x}, \hat{u})$ uniformly in t on $[a, \rho(b)]_{\mathbb{T}}$ and if for $(\xi, v) \in \mathcal{P}T_{\varepsilon_1}(\hat{x}, \hat{u})$ the function $\nabla_{(x,u)} h(\cdot, \xi, v)$ is rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$, then there exists $\delta_0 \in (0, \varepsilon_1)$ such that for every pair $(x, u) \in T_{\delta_0}(\hat{x}, \hat{u})$ and for $\ell := \|\nabla_{(x,u)} h(\cdot, \hat{x}(\cdot), \hat{u}(\cdot))\|_{\mathrm{C_{prd}}} + 1$, we have*

(i) $\|\nabla_{(x,u)} h(\cdot, x(\cdot), u(\cdot))\|_{\mathrm{C_{prd}}} \le \ell$,
(ii) *$h(t, \cdot, \cdot)$ is continuous at the pair (x, u) uniformly in t on $[a, \rho(b)]_{\mathbb{T}}$.*

If in addition we assume that the function $h(\cdot, \xi, v)$ is rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$ for any vector $(\xi, v) \in \mathcal{P}T_{\delta_0}(\hat{x}, \hat{u})$, then

(iii) *$h(\cdot, x(\cdot), u(\cdot))$ is piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$.*

Proof. We abbreviate $\hat{z}(\cdot) := (\hat{x}(\cdot), \hat{u}(\cdot))$ and $z(\cdot) := (x(\cdot), u(\cdot))$. Since $\hat{z}$ and $\nabla_z h$ satisfy for $\varepsilon = \varepsilon_1$ the assumptions of Lemma 2.5, it follows that $\nabla_z h(\cdot, \hat{z}(\cdot))$ is piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$. Therefore, the number $k := \|\nabla_z h(\cdot, \hat{z}(\cdot))\|_{\mathrm{C_{prd}}} = \ell - 1$ is finite. The continuity of $\nabla_z h(t, \cdot)$ at $\hat{z}$ uniformly in t implies that there exists $\delta \in (0, \varepsilon_1)$ such that

$$\text{for all } (t, w) \in [a, \rho(b)]_{\mathbb{T}} \times \mathbb{R}^{n+m} \text{ such that } |w - \hat{z}(t)| < \delta \text{ we have } |\nabla_z h(t, w)| \le \ell. \tag{2.4}$$

Hence, for any $\delta_0 \in (0, \delta]$ and for all $z \in T_{\delta_0}(\hat{z})$, we have $\|\nabla_z h(\cdot, z(\cdot))\|_{\mathrm{C_{prd}}} \le \ell$, and thus, part (i) is proven for any $\delta_0 \in (0, \delta]$. For part (ii), we set $\delta_0 := \delta/2$ and let z be any function in $T_{\delta_0}(\hat{z})$. To show the uniform continuity of $h(t,\cdot,\cdot)$ at z, take any $\varepsilon > 0$ and let $\delta_\varepsilon := \min\{\varepsilon/\ell, \delta_0\}$. Then, by the generalized mean value theorem, for every $(t,y) \in [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^{n+m}$ such that $|y - z(t)| < \delta_\varepsilon$ we have

$$\big|h(t,y) - h(t,z(t))\big| \le \sup\Big\{\big|\nabla_z h\big(t, z(t) + \tau\,[y - z(t)]\big)\big|,\ \tau \in [0,1]\Big\}\,|y - z(t)|. \tag{2.5}$$

Set $w_\tau(t) := z(t) + \tau(y - z(t))$. Then, for all $\tau \in [0,1]$ and for all $t \in [a,\rho(b)]_{\mathbb{T}}$ we have

$$|w_\tau(t) - \hat{z}(t)| \le |z(t) - \hat{z}(t)| + |y - z(t)| \le \delta_0 + \delta_\varepsilon \le 2\,\delta_0 = \delta.$$

Hence, applying (2.4) to $w_\tau(t)$ and using (2.5) it follows that

$$\big|h(t,y) - h(t,z(t))\big| \le \ell\,|y - z(t)| < \ell\,\delta_\varepsilon \le \varepsilon,$$

which proves (ii) for $z \in T_{\delta_0}(\hat{z})$. Finally, part (iii) now follows from part (ii) and Lemma 2.5, in which we take $\varepsilon := \delta_0$. □

Define the tangent spaces $T(t)$ and the space $\mathcal{T}$ of tangent functions by

$$\begin{aligned} T(t) &:= \{v \in \mathbb{R}^m,\ N(t)\,v = 0\},\\ \mathcal{T} &:= \{v(\cdot) \in \mathrm{C_{prd}}[a,\rho(b)]_{\mathbb{T}},\ v(t) \in T(t) \text{ for all } t \in [a,\rho(b)]_{\mathbb{T}}\}. \end{aligned}$$

The assumption that $N(t)$ has full rank k for all $t \in [a,\rho(b)]_{\mathbb{T}}$ implies that one can choose the function $Y : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^{m\times(m-k)}$, $Y \in \mathrm{C_{prd}}$, such that the columns of $Y(t)$ form an orthonormal basis for the space $T(t)$. That is, we have

$$Y^T(t)\,Y(t) = I, \quad \operatorname{Im} Y(t) = T(t) = \operatorname{Ker} N(t), \quad N(t)\,Y(t) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}. \tag{2.6}$$

According to Definition 9.1 of [31], the linear system

$$\eta^\Delta = \mathcal{A}(t)\,\eta + \mathcal{B}(t)\,v, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{2.7}$$

is said to be *M-controllable over $\mathcal{T}$* if for any vector $d \in \mathbb{R}^r$ there exists a vector $\alpha \in \mathbb{R}^n$ and a function $v \in \mathcal{T}$ such that the solution η of the initial value problem (2.7) with $\eta(a) = \alpha$ satisfies $M\begin{pmatrix}\eta(a)\\ \eta(b)\end{pmatrix} = d$.

Remark 2.8. Throughout the paper we denote by $\Phi(t)$ the fundamental matrix of system (2.7), that is, $\Phi(t)$ is the solution of the initial value problem $\Phi^\Delta = \mathcal{A}(t)\,\Phi$ on $[a,\rho(b)]_{\mathbb{T}}$ with $\Phi(a) = I$. Since the matrix $I + \mu(t)\,\mathcal{A}(t)$ is assumed to be invertible on $[a,\rho(b)]_{\mathbb{T}}$, the matrix $\mathcal{A}(t)$ is regressive on $[a,\rho(b)]_{\mathbb{T}}$ and it follows that $\Phi(t)$ is invertible on $[a,b]_{\mathbb{T}}$, see [9] (Thm. 5.21) and also [50] (Prop. 2.1). The solutions of (2.7) can be then expressed *via* the variation of constants formula ([9], Thm. 5.27) as

$$\eta(t) = \Phi(t)\,\eta(a) + \Phi(t)\int_a^t [\Phi^\sigma(\tau)]^{-1}\mathcal{B}(\tau)\,v(\tau)\,\Delta\tau, \quad t \in [a,b]_{\mathbb{T}}. \tag{2.8}$$

The following characterization of the M-controllability of (2.7) over $\mathcal{T}$ in terms of the associated Grammian matrix is proven in the same way as in [31], (Prop. 4.6).

Proposition 2.9. *Assume that the matrices $M = \begin{pmatrix}M_a & M_b\end{pmatrix}$ and $N(t)$ have full rank and that $I + \mu(t)\,\mathcal{A}(t)$ is invertible on $[a,\rho(b)]_{\mathbb{T}}$. Then the linear system (2.7) is M-controllable over $\mathcal{T}$ if and only if there exists $\varepsilon_0 > 0$ such that the $r \times r$ Grammian matrix*

$$\mathcal{Z} := \varepsilon_0^2\,\mathcal{D}\mathcal{D}^T + \int_a^b \mathcal{E}(t)\,\mathcal{E}^T(t)\,\Delta t > 0, \tag{2.9}$$

i.e., $\mathcal{Z}$ is positive definite, where $\mathcal{D} \in \mathbb{R}^{r\times n}$ and $\mathcal{E}(t) \in \mathbb{R}^{r\times(m-k)}$ are matrices defined by

$$\mathcal{D} := M \begin{pmatrix} I \\ \Phi(b) \end{pmatrix} = M_a + M_b\,\Phi(b), \quad \mathcal{E}(t) := M_b\,\Phi(b)\,[\Phi^\sigma(t)]^{-1}\mathcal{B}(t)\,Y(t). \tag{2.10}$$

The weak Pontryagin maximum principle on time scales for problem (C) is proven in [31], (Thm. 9.4).

Proposition 2.10. *Assume that $(\hat{x},\hat{u})$ is a weak local minimum for problem (C) such that the assumption* (A1) *holds. Then there exists a constant $\lambda_0 \geq 0$, a vector $\hat{\gamma} \in \mathbb{R}^r$, a function $\hat{\lambda} : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^k$, $\hat{\lambda} \in \mathrm{C}_{\mathrm{prd}}$, and a function $\hat{p} : [a,b]_{\mathbb{T}} \to \mathbb{R}^n$, $\hat{p} \in \mathrm{C}^1_{\mathrm{prd}}$, such that $\lambda_0 + \|\hat{p}\|_{\mathrm{C}} \neq 0$ and satisfying the conditions*

$$-\hat{p}^\Delta(t) = \nabla_x H^T(t,\hat{x}(t),\hat{u}(t),\hat{p}^\sigma(t),\hat{\lambda}(t),\lambda_0), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{2.11}$$

$$\nabla_u H^T(t,\hat{x}(t),\hat{u}(t),\hat{p}^\sigma(t),\hat{\lambda}(t),\lambda_0) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{2.12}$$

$$\begin{pmatrix} -\hat{p}(a) \\ \hat{p}(b) \end{pmatrix} = \lambda_0 \nabla K^T(\hat{x}(a),\hat{x}(b)) + M^T\hat{\gamma}. \tag{2.13}$$

Moreover, if the system (2.7) is M-controllable over $\mathcal{T}$, then we may take $\lambda_0 = 1$ and in this case the quantities $\hat{\gamma}$, $\hat{\lambda}(\cdot)$, and $\hat{p}(\cdot)$ are unique.

Let $(\hat{x},\hat{u})$ be a feasible pair satisfying assumption (A2). We recall from Section 9 of [31] the following notions. A pair (η,v) is called an *admissible direction* (or a tangent direction) if $\eta \in \mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}$, $v \in \mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$, and

$$\eta^\Delta(t) = \mathcal{A}(t)\,\eta(t) + \mathcal{B}(t)\,v(t), \quad N(t)\,v(t) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \quad M \begin{pmatrix} \eta(a) \\ \eta(b) \end{pmatrix} = 0. \tag{2.14}$$

The second variation of the functional J in problem (C) at $(\hat{x},\hat{u})$ in the admissible direction (η,v) is defined as the quantity

$$J''(\eta,v) := \frac{1}{2} \begin{pmatrix} \eta(a) \\ \eta(b) \end{pmatrix}^T \Gamma \begin{pmatrix} \eta(a) \\ \eta(b) \end{pmatrix} + \frac{1}{2}\int_a^b \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix}^T \nabla^2_{(x,u)}\,\hat{H}(t) \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix} \Delta t, \tag{2.15}$$

where the function $\nabla^2_{(x,u)}\,\hat{H}(t)$ on $[a,\rho(b)]_{\mathbb{T}}$ is given

$$\nabla^2_{(x,u)}\hat{H}(t) := \nabla^2_{(x,u)}H(t,\hat{x}(t),\hat{u}(t),\hat{p}^\sigma(t),\hat{\lambda}(t),1) \tag{2.16}$$

and the symmetric $2n \times 2n$ matrix Γ is defined by

$$\Gamma := \nabla^2 K(\hat{x}(a),\hat{x}(b)) + \hat{\gamma}^T\nabla^2\varphi(\hat{x}(a),\hat{x}(b)). \tag{2.17}$$

When (η,v) varies in the space of admissible directions, then J'' becomes a quadratic functional.

3. Sufficiency *via* coercivity

In this section we prove the first main result of this paper (Thm. 3.3 below), namely a sufficient condition for a strict weak local minimum in problem (C) expressed in terms of the coercivity of the second variation J''. It is known in [31], (Thm. 9.7) that the nonnegativity of the second variation is a necessary condition for a weak local minimum in (C). In this respect the gap between the necessary condition (the nonnegativity of J'') and the presented sufficient condition (the coercivity of J'') is as small as possible. The examples illustrating this fact are already known in the calculus of variations setting, see *e.g.* [28,30].

Definition 3.1. The functional J'' is *coercive* if there exists $\alpha > 0$ such that

$$J''(\eta, v) \geq \frac{\alpha}{2} \left\{ |\eta(a)|^2 + |\eta(b)|^2 + \int_a^b \left(|\eta(t)|^2 + |v(t)|^2 \right) \Delta t \right\} \tag{3.1}$$

for all admissible directions (η, v).

Remark 3.2. One can easily show by using the variation of constants formula (2.8) that the above definition of the coercivity is equivalent to removing the term $|\eta(t)|^2$ from the integrand. Indeed, assume that there exists $\beta > 0$ such that for all admissible directions (η, v) we have

$$J''(\eta, v) \geq \frac{\beta}{2} \left\{ |\eta(a)|^2 + |\eta(b)|^2 + \int_a^b |v(t)|^2 \, \Delta t \right\}. \tag{3.2}$$

Then we obtain from (2.8) by the Schwarz and arithmetic–geometric mean inequalities for some constant $k > 0$ the estimate

$$|\eta(t)|^2 \leq k \left\{ |\eta(a)|^2 + \int_a^b |v(t)|^2 \Delta t \right\},$$

which upon adding the term $k\,|\eta(b)|^2$ on the right–hand side and integration over $[a, b]_{\mathbb{T}}$ yields

$$\int_a^b |\eta(t)|^2 \Delta t \leq k(b-a) \left\{ |\eta(a)|^2 + |\eta(b)|^2 + \int_a^b |v(t)|^2 \Delta t \right\}. \tag{3.3}$$

Therefore, if (3.2) holds, then by (3.3) we get

$$\begin{aligned} J''(\eta, v) &\geq \frac{\beta}{4} \left\{ |\eta(a)|^2 + |\eta(b)|^2 + \int_a^b |v(t)|^2 \, \Delta t \right\} + \frac{\beta}{4k(b-a)} \int_a^b |\eta(t)|^2 \\ &\geq \frac{\alpha}{2} \left\{ |\eta(a)|^2 + |\eta(b)|^2 + \int_a^b \left(|\eta(t)|^2 + |v(t)|^2 \right) \Delta t \right\}, \quad \alpha := \min \left\{ \frac{\beta}{2}, \frac{\beta}{2k(b-a)} \right\}, \end{aligned}$$

showing that (3.1) is satisfied. The coercivity condition in (3.2) is used *e.g.* for the time scale calculus of variations problem in [28], (Sect. 6), where $v(t) = \eta^{\Delta}(t)$.

The following theorem is a direct generalization of Theorem 2 of [28] from the calculus of variations setting to the optimal control setting on time scales. This result also extends the sufficient condition for continuous time nonlinear optimal control problems in [55], (Thm. 2.3).

Theorem 3.3 (Sufficiency for problem (C))**.** *Let $(\hat{x}, \hat{u})$ be feasible for problem* (C)*, assumption* (A2) *holds, and system* (2.7) *is M-controllable over $\mathcal{T}$. Let $\hat{p}(\cdot)$, $\hat{\lambda}(\cdot)$, $\lambda_0 = 1$, and $\hat{\gamma}$ satisfy the weak Pontryagin maximum principle* (*Prop.* 2.10) *and assume that the second variation of the functional J at $(\hat{x}, \hat{u})$ is coercive, i.e., for some $\alpha > 0$ inequality* (3.1) *holds for all admissible directions (η, v). Then $(\hat{x}, \hat{u})$ is a strict weak local minimum for* (C)*. In addition, there exists $\delta_0 > 0$ such that for any feasible pair (x, u) with $\|(x - \hat{x}, u - \hat{u})\|_{\mathrm{C_{prd}}} < \delta_0$ we have*

$$J(x, u) - J(\hat{x}, \hat{u}) \geq \frac{\alpha}{16} \left\{ |x(a) - \hat{x}(a)|^2 + |x(b) - \hat{x}(b)|^2 + \int_a^b \left(|x(t) - \hat{x}(t)|^2 + |u(t) - \hat{u}(t)|^2 \right) \Delta t \right\}.$$

The proof of Theorem 3.3 is displayed below after important preparatory lemmas. The first lemma states that the difference $(x, u) - (\hat{x}, \hat{u})$ of two feasible pairs (x, u) and $(\hat{x}, \hat{u})$ can be approximated by an admissible pair (η, v). This statement generalizes Lemma 7 of [28] to the control setting. Its proof is displayed in Appendix A.

Lemma 3.4. *Let $(\hat{x}, \hat{u})$ be a feasible pair for* (C), *assumption* (A1) *holds, and system* (2.7) *is M-controllable over $\mathcal{T}$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any feasible pair (x, u) with $\|(x - \hat{x}, u - \hat{u})\|_{\mathrm{C_{prd}}} < \delta$ there exists an admissible direction (η, v), i.e., a solution of* (2.14)*, which satisfies for all $t \in [a, \rho(b)]_{\mathbb{T}}$ the estimate*

$$\begin{aligned} &\big|\big(x(t) - \hat{x}(t) - \eta(t), u(t) - \hat{u}(t) - v(t)\big)\big| \\ &\qquad \le \varepsilon \left\{ \big|\big(x(a) - \hat{x}(a), x(b) - \hat{x}(b)\big)\big| + \int_a^b \big|\big(x(t) - \hat{x}(t), u(t) - \hat{u}(t)\big)\big|\, \Delta t \right\}. \end{aligned} \tag{3.4}$$

In addition, if b is left-scattered, then inequality (3.4) *with $|x(b) - \hat{x}(b) - \eta(b)|$ on its left-hand side is also satisfied.*

Remark 3.5. In the calculus of variations setting, see *e.g.* [6, 28], we have $f(t, x, u) = u$ and no control constraints. In this case $m = k = n$, $\mathcal{A}(t) \equiv 0$, $\mathcal{B}(t) \equiv I$, $Y(t) \equiv I$, $N(t) \equiv 0$, $\Phi(t) \equiv I$ on $[a, \rho(b)]_{\mathbb{T}}$. This yields that

$$\mathcal{D} = M_a + M_b, \quad \mathcal{E}(t) \equiv M_b, \quad \mathcal{Z} = \varepsilon_0^2 \,(M_a + M_b)\,(M_a^T + M_b^T) + (b - a)\, M_b\, M_b^T.$$

The problem (C) is then always M-controllable, which can be also seen from the invertibility of the above matrix $\mathcal{Z}$. Indeed, $\mathcal{Z}$ is in this case positive semidefinite (for all $\varepsilon_0 > 0$), and if $\mathcal{Z}\omega = 0$ for some $\omega \in \mathbb{R}^r$, then $M_b^T \omega = 0$ as well as $M_a^T \omega = 0$, *i.e.*, $M^T \omega = 0$. But since M has full rank, it follows that $\omega = 0$. The construction of the pair $(\tilde{\eta}, \tilde{v})$ in (A.16) and (A.15) in the proof of Lemma 3.4 in Appendix A then reduces to $g(t) \equiv 0$, $\pi(t) \equiv 0$, $\nu(t) \equiv 0$, $\beta = 0$, and

$$\begin{aligned} \tilde{\alpha} &= -\varepsilon_0^2 \,(M_a^T + M_b^T)\, \mathcal{Z}^{-1} \gamma, \\ \tilde{v}(t) &= -M_b^T \mathcal{Z}^{-1} \gamma = \tilde{\eta}^{\Delta}(t), \quad t \in [a, \rho(b)]_{\mathbb{T}}, \\ \tilde{\eta}(t) &= -\varepsilon_0^2 \,(M_a^T + M_b^T)\, \mathcal{Z}^{-1} \gamma - (t - a)\, M_b^T \mathcal{Z}^{-1} \gamma, \quad t \in [a, b]_{\mathbb{T}}. \end{aligned}$$

In the proof of Lemma 7, page 160 from [28] we have used a slightly simpler construction of a constant function $\tilde{\eta}^{\Delta}(t)$ and a linear function $\tilde{\eta}(t)$ with the aid of the matrix $(MM^T)^{-1}$. In this respect it is quite surprising that such an explicit construction of $(\tilde{\eta}, \tilde{v})$ is also possible in the optimal control setting. In this construction we can appreciate the crucial role of the matrix $\mathcal{Z}$, whose invertibility characterizes the M-controllability of system (2.7) over $\mathcal{T}$, according to Proposition 2.9.

The second lemma provides an estimate for the second variation of J evaluated at the difference between a feasible pair (x, u) and the reference pair $(\hat{x}, \hat{u})$. Its proof is also displayed in Appendix A.

Lemma 3.6. *Let $(\hat{x}, \hat{u})$ be feasible for problem* (C), *assumption* (A2) *holds, and system* (2.7) *is M-controllable over $\mathcal{T}$. If the functional J'' at $(\hat{x}, \hat{u})$ is coercive for some $\alpha > 0$, then there exists $\delta_0 > 0$ such that for all feasible pairs (x, u) with $\|(x - \hat{x}, u - \hat{u})\|_{\mathrm{C_{prd}}} < \delta_0$ we have*

$$J''(x - \hat{x}, u - \hat{u}) \ge \frac{\alpha}{8} \left\{ |x(a) - \hat{x}(a)|^2 + |x(b) - \hat{x}(b)|^2 + \int_a^b \big(\, |x(t) - \hat{x}(t)|^2 + |u(t) - \hat{u}(t)|^2 \,\big)\, \Delta t \right\}.$$

We are now ready to present the proof of Theorem 3.3.

Proof of Theorem 3.3. First we recall the definition of the Hamiltonian $H(t, x, u, p, \lambda, \lambda_0)$ with $\lambda_0 = 1$ in (1.4). Let $(\hat{x}, \hat{u})$ be feasible and let $\hat{p}(\cdot)$, $\hat{\lambda}(\cdot)$, $\lambda_0 = 1$, and $\hat{\gamma}$ be the associated functions satisfying the weak Pontryagin maximum principle (Prop. 2.10). Moreover, let the second variation of J at $(\hat{x}, \hat{u})$ be coercive with a constant $\alpha > 0$ as in (3.1). As in (2.16), we use the notation $\hat{H}(t) := H(t, \hat{x}(t), \hat{u}(t), \hat{p}^{\sigma}(t), \hat{\lambda}(t), 1)$, and if (x, u) is another feasible pair, then we set $H(t) := H(t, x(t), u(t), \hat{p}^{\sigma}(t), \hat{\lambda}(t), 1)$. Similar notation will be used for the gradient and Hessian matrix of $\hat{H}(t)$ and $H(t)$ with respect to the second and third variables. We adopt the notation

$$\mathrm{d}\hat{x}(t) := x(t) - \hat{x}(t), \quad t \in [a, b]_{\mathbb{T}}, \quad \mathrm{d}\hat{u}(t) := u(t) - \hat{u}(t), \quad t \in [a, \rho(b)]_{\mathbb{T}}, \tag{3.5}$$

and set $c := (\mathrm{d}\hat{x}^T(a), \mathrm{d}\hat{x}^T(b))^T$ and $\mathrm{d}\hat{y}(t) := (\mathrm{d}\hat{x}^T(t), \mathrm{d}\hat{u}^T(t))^T$. By the second order Taylor expansion and the integration by parts, we calculate the difference

$$\begin{aligned} J(x,u) - J(\hat{x},\hat{u}) &= (K + \hat{\gamma}^T\varphi)(x(a),x(b)) - (K+\hat{\gamma}^T\varphi)(\hat{x}(a),\hat{x}(b)) \\ &\quad + \int_a^b \left\{ H(t) - \hat{H}(t) - [\hat{p}^\sigma(t)]^T \mathrm{d}\hat{x}^\Delta(t) \right\} \Delta t \\ &= \left[\nabla K(\hat{x}(a),\hat{x}(b)) + \hat{\gamma}^T M + \begin{pmatrix} \hat{p}(a) \\ -\hat{p}(b) \end{pmatrix}^T \right] c + \frac{1}{2}\, c^T\, \nabla^2 \big(K^T + \hat{\gamma}^T \nabla^2 \varphi^T\big)(\omega_0,\zeta_0)\, c \\ &\quad + \int_a^b \Big\{ \big(\nabla_x \hat{H}(t) + [\hat{p}^\Delta(t)]^T\big)\, \mathrm{d}\hat{x}(t) + \nabla_u \hat{H}(t)\, \mathrm{d}\hat{u}(t) \\ &\quad + \frac{1}{2}\, \mathrm{d}\hat{y}^T(t)\, \nabla^2_{(x,u)} H(t,\xi_0(t),\nu_0(t),\hat{p}^\sigma(t),\hat{\lambda}(t),1)\, \mathrm{d}\hat{y}(t) \Big\}\, \Delta t, \end{aligned} \tag{3.6}$$

where

$$\begin{pmatrix} \omega_0 \\ \zeta_0 \end{pmatrix} = \begin{pmatrix} \hat{x}(a) \\ \hat{x}(b) \end{pmatrix} + \theta_1\, c, \quad \begin{pmatrix} \xi_0(t) \\ \nu_0(t) \end{pmatrix} = \begin{pmatrix} \hat{x}(t) \\ \hat{u}(t) \end{pmatrix} + \theta_2(t)\, \mathrm{d}\hat{y}(t), \quad t \in [a,\rho(b)]_{\mathbb{T}},$$

with $\theta_1 \in (0,1)$ and $\theta_2(t) \in (0,1)$ for all $t \in [a,\rho(b)]_{\mathbb{T}}$. We now apply the first order conditions from the weak Pontryagin maximum principle (Prop. 2.10) and subtract and add the term $J''(\mathrm{d}\hat{x},\mathrm{d}\hat{u})$ in (3.6) to get

$$\begin{aligned} J(x,u) - J(\hat{x},\hat{u}) &= J''(\mathrm{d}\hat{x},\mathrm{d}\hat{u}) + \frac{1}{2}\, c^T \big[\, \nabla^2 (K + \hat{\gamma}^T\varphi)(\omega_0,\zeta_0) - \Gamma \,\big]\, c \\ &\quad + \frac{1}{2} \int_a^b \Big\{ \mathrm{d}\hat{y}^T(t) \big[\, \nabla^2_{(x,u)} H(t,\xi_0(t),\nu_0(t),\hat{p}^\sigma(t),\hat{\lambda}(t),1) - \nabla^2_{(x,u)} \hat{H}(t) \,\big]\, \mathrm{d}\hat{y}(t) \Big\}\, \Delta t, \end{aligned}$$

where Γ is given in (2.17). By the continuity of $\nabla^2(K + \hat{\gamma}^T\varphi)(\cdot,\cdot)$ and by the continuity of $\nabla^2_{(x,u)} H(t,\cdot,\cdot,\hat{p}^\sigma(t),\hat{\lambda}(t),1)$ at $(\hat{x},\hat{u})$ uniformly in t on $[a,\rho(b)]_{\mathbb{T}}$, for every $\beta > 0$ there exists $\delta_1 > 0$ (with $\delta_1 < \varepsilon_1$) such that for any $(\omega,\zeta) \in B_{\delta_1}(\hat{x}(a),\hat{x}(b))$ and for any $(\xi,\nu) \in \mathbb{R}^n \times \mathbb{R}^m$ and $t \in [a,\rho(b)]_{\mathbb{T}}$ with $(t,\xi,\nu) \in T_{\delta_1}(\hat{x},\hat{u})$ we have

$$\big|\, \nabla^2(K + \hat{\gamma}^T\varphi)(\omega,\zeta) - \Gamma \,\big| < \beta, \quad \big|\, \nabla^2_{(x,u)} H(t,\xi,\nu,\hat{p}^\sigma(t),\hat{\lambda}(t),1) - \nabla^2_{(x,u)} \hat{H}(t) \,\big| < \beta. \tag{3.7}$$

Put $\delta := \min\{\delta_0,\delta_1,\varepsilon_1\}$, where δ_0 is the number from Lemma 3.6 and δ_1 is the number from above corresponding to $\beta := \alpha/8$. Then by Lemma 3.6 and inequality (3.7) for any feasible pair (x,u) with $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\|_{\mathrm{C_{prd}}} < \delta$ we have

$$\begin{aligned} J(x,u) - J(\hat{x},\hat{u}) &\geq \left(\frac{\alpha}{8} - \frac{\beta}{2} \right) \left\{ |c|^2 + \int_a^b |\mathrm{d}\hat{y}(t)|^2\, \Delta t \right\} \\ &= \frac{\alpha}{16} \left\{ |\mathrm{d}\hat{x}(a)|^2 + |\mathrm{d}\hat{x}(b)|^2 + \int_a^b \big(\big|\mathrm{d}\hat{x}(t)\big|^2 + |\mathrm{d}\hat{u}(t)|^2 \big)\, \Delta t \right\}. \end{aligned}$$

Therefore, $(\hat{x},\hat{u})$ is a strict weak local minimum for problem (C). We note that the term $(1/2)\, \mathrm{d}\hat{y}^T(t)\, \nabla^2_{(x,u)} H(t,\xi_0(t),\nu_0(t),\hat{p}^\sigma(t),\hat{\lambda}(t),1)\, \mathrm{d}\hat{y}(t)$ used in the above calculations is indeed integrable on $[a,b]_{\mathbb{T}}$, since it is equal to $H(t) - \hat{H}(t) - \nabla_{(x,u)}\hat{H}(t)\, \mathrm{d}\hat{y}(t)$ on $[a,\rho(b)]_{\mathbb{T}}$, which is an rd-continuous function by Lemma 2.7 (iii). The proof is now complete. ∎

We illustrate the applicability of Theorem 3.3 by the following example, which was analyzed in the continuous time setting in [45], Example 3.1.

Example 3.7. Consider the time scale $\mathbb{T} := [0,1] \cup [2,3]$, which is denoted by $[0,3]_{\mathbb{T}}$. Define the following optimal control problem on $[0,3]_{\mathbb{T}}$ by

$$\text{minimize} \quad J(x,u) := \int_0^3 \left\{ u_1^3(t) - \frac{1}{2}\, u_2^2(t) - \frac{1}{8}\, x(t)\, u_1(t) \right\} \Delta t \tag{3.8}$$

subject to $x \in \mathrm{C}^1_{\mathrm{prd}}[0,3]_{\mathbb{T}}$ and $u \in \mathrm{C}_{\mathrm{prd}}[0,3]_{\mathbb{T}}$ such that

$$x^{\Delta}(t) = -\frac{1}{8}\, x(t) + u_1(t), \quad u_1^2(t) + u_2^2(t) = 4, \quad t \in [0,3]_{\mathbb{T}}, \quad x(0) = 0 = x(3).$$

Here $n = 1$, $m = 2$, $k = 1$, $r = 2$, and $u = (u_1,\, u_2)^T$. The goal is to use Theorem 3.3 to show that the feasible pair $(\hat{x}, \hat{u})$ with $\hat{x}(t) \equiv 0$ and $\hat{u}(t) \equiv (0,\, 2)^T$ on $[0,3]_{\mathbb{T}}$ is a weak local minimum for problem (3.8). It is immediate that assumption (A2) holds, where $M = I_{2\times 2}$ and $N(t) \equiv (0,\, 4)$ are of full rank, and $I + \mu(t)\, f_x\big(t, \hat{x}(t), \hat{u}(t)\big) = 1 - \frac{1}{8}\, \mu(t) \geq 1 - \frac{1}{8} > 0$ on $[0,3]_{\mathbb{T}}$. System (2.14) with $v = (v_1,\, v_2)^T$ is then of the form

$$\eta^{\Delta}(t) = -\frac{1}{8}\, \eta(t) + v_1(t), \quad v_2(t) \equiv 0, \quad t \in [0,3]_{\mathbb{T}}, \quad \eta(0) = 0 = \eta(3). \tag{3.9}$$

That is, $\mathcal{A}(t) \equiv -\frac{1}{8}$ and $\mathcal{B}(t) \equiv (1,\, 0)$ on $[0,3]_{\mathbb{T}}$. Hence, the system in (3.9) is in this case M-controllable over the set $\mathcal{T} = \{v(\cdot) = (v_1(\cdot),\, 0)^T,\ v_1 \in \mathrm{C}_{\mathrm{prd}}[0,3]_{\mathbb{T}}\}$, as it can be seen directly or *via* Theorem 9.3 of [31].

Set $\hat{p}(t) \equiv 0$, $\hat{\lambda}(t) \equiv \frac{1}{2}$, $\lambda_0 = 1$, and take any $\hat{\gamma} \in \mathbb{R}^2$. Then clearly $(\hat{x}, \hat{u}, \hat{p}, \lambda, 1, \hat{\gamma})$ satisfies the weak Pontryagin maximum principle (Prop. 2.10). According to (2.15) and taking into account that $v_2(t) \equiv 0$, the second variation of (3.8) at $(\hat{x}, \hat{u})$ is

$$J''(\eta, v_1) := \frac{1}{2} \int_0^3 \left\{ v_1^2(t) - \frac{1}{4}\, \eta(t)\, v_1(t) \right\} \Delta t, \tag{3.10}$$

where the functions $\eta \in \mathrm{C}^1_{\mathrm{prd}}[0,3]_{\mathbb{T}}$ and $v_1 \in \mathrm{C}_{\mathrm{prd}}[0,3]_{\mathbb{T}}$ satisfy (3.9). We first prove the positivity of (3.10) *via* the Picone formula in Proposition 4.11 from [54]. We set

$$W(t) := \begin{cases} \dfrac{1}{8}\, \tan\left(-\dfrac{1}{8}\, t + \dfrac{1}{8} + \arctan \dfrac{71}{56} \right), & t \in [0,1], \\[2ex] \dfrac{1}{8}\, \tan\left(-\dfrac{1}{8}\, t + \dfrac{1}{4} + \dfrac{\pi}{4} \right), & t \in [2,3]. \end{cases}$$

Then $W(t)$ is a solution of the Riccati dynamic equation

$$W^{\Delta} + \frac{\left(W^{\sigma} + \dfrac{1}{8} \right)^2}{1 - \mu(t)\, W^{\sigma}} + \frac{1}{64} = 0, \quad t \in [0,3]_{\mathbb{T}},$$

with $K(t) := 1 - \mu(t)\, W^{\sigma}(t) \geq 1 - W(2) = \frac{7}{8} > 0$ on $[0,3]_{\mathbb{T}}$. Thus, by Proposition 4.11 of [54], it follows that for any pair (η, v_1) admissible for (3.10) we have

$$J''(\eta, v_1) = \frac{1}{2} \int_0^3 \frac{z^2(t)}{K(t)}\, \Delta t \geq 0, \tag{3.11}$$

where $z(t) := K(t)\, v_1(t) - M(t)\, \eta(t)$ and $M(t) := \frac{1}{8}\, K(t) + W^{\sigma}(t)$ on $[0,3]_{\mathbb{T}}$. In addition, if $J''(\eta, v_1) = 0$ for such a pair (η, v_1), then (3.11) yields that $z(t) \equiv 0$, which by (3.9) means that η satisfies the linear dynamic equation $\eta^{\Delta}(t) = [W^{\sigma}(t)/K(t)]\, \eta(t)$ on $[0,3]_{\mathbb{T}}$ with $\eta(0) = 0 = \eta(3)$. This implies by the uniqueness of solutions that $\eta(t) \equiv 0$, and hence by (3.9) that also $v_1(t) \equiv 0$ on $[0,3]_{\mathbb{T}}$. Therefore, the functional $J''(\eta, v_1)$ is positive.

To prove that $J''(\eta, v_1)$ is coercive we notice from (3.9) that $v_1(t) = \eta^{\Delta}(t) + \frac{1}{8}\,\eta(t)$ on $[0,3]_{\mathbb{T}}$, and hence

$$J''(\eta, v_1) = \mathcal{I}(\eta) := \frac{1}{2}\int_0^3 \left\{ [\eta^{\Delta}(t)]^2 - \frac{1}{64}\,\eta^2(t) \right\} \Delta t,$$

where $\eta \in \mathrm{C}^1_{\mathrm{prd}}[0,3]_{\mathbb{T}}$ with $\eta(0) = 0 = \eta(3)$. Note that the functional $\mathcal{I}(\eta)$ is positive and the corresponding strengthened Legendre condition $R(t^{\pm}) \equiv 1 > 0$ at dense points of $[0,3]_{\mathbb{T}}$ holds. Thus, the functional $\mathcal{I}(\eta)$ is coercive in η^{Δ} by applying the version of Theorem 4.1 from [30] for quadratic forms without shift in η, see also Proposition 3 of [51] and Remark 1.1. It follows that there exists $\beta > 0$ such that for any $\eta \in \mathrm{C}^1_{\mathrm{prd}}[0,3]_{\mathbb{T}}$ with $\eta(0) = 0 = \eta(3)$ we have

$$\mathcal{I}(\eta) \geq \frac{\beta}{2}\int_0^3 \left|\eta^{\Delta}(t)\right|^2 \Delta t \geq \frac{\alpha}{2}\int_0^3 v_1^2(t)\,\Delta t = \frac{\alpha}{2}\int_0^3 v^2(t)\,\Delta t, \tag{3.12}$$

where $\alpha := 7\beta/128 > 0$ and where the last inequality in (3.12) is obtained by using (3.9) and the Cauchy–Schwarz inequality. Whence, we proved that the second variation $J''(\eta, v_1)$ in (3.10) is coercive. Therefore, the feasible pair $\hat{x}(t) \equiv 0$, $\hat{u}(t) \equiv (0,\, 2)^T$ together with $\lambda_0 = 1$, $\hat{p}(t) \equiv 0$, $\hat{\lambda}(t) \equiv \frac{1}{2}$, and $\hat{\gamma} \in \mathbb{R}^2$ satisfy the assumptions of Theorem 3.3. Hence, by this theorem, $(\hat{x}, \hat{u})$ is a strict weak local minimum for problem (3.8).

4. Application to sensitivity analysis

In this section we provide a sensitivity analysis for the nonlinear optimal control problem (C) with separated endpoints. In particular, we will show that given a problem (C_ω) depending on a parameter ω, then a weak local minimum $(\hat{x}, \hat{u})$ for ($\mathrm{C}_{\hat{\omega}}$) with corresponding multipliers $(\hat{p}, \hat{\lambda}, \hat{\gamma}_a, \hat{\gamma}_b)$ can be embedded into a family of optimal solutions $\big(x(\cdot,\omega), u(\cdot,\omega)\big)$ and corresponding multipliers $\big(p(\cdot,\omega), \lambda(\cdot,\omega), \gamma_a(\omega), \gamma_b(\omega)\big)$ for sufficiently small perturbations of the parameter ω near $\hat{\omega}$. This result is based on a suitable application of the implicit function theorem and the new sufficiency criterion in Theorem 3.3.

Consider the parametric nonlinear optimal control problem

$$\text{minimize} \quad J(x,u,\omega) := K_a(x(a),\omega) + K_b(x(b),\omega) + \int_a^b L(t, x(t), u(t), \omega)\,\Delta t \tag{C_ω}$$

subject to $x \in \mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}$ and $u \in \mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$ such that

$$x^{\Delta}(t) = f(t, x(t), u(t), \omega), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{4.1}$$

$$\psi(t, u(t), \omega) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{4.2}$$

$$\varphi_a(x(a),\omega) = 0, \quad \varphi_b(x(b),\omega) = 0. \tag{4.3}$$

Similarly to problem (C) in Section 1 we assume that $n, m, k, r_a, r_b, d \in \mathbb{N}$ are given dimensions with $k \leq m \leq n$ and $r_a \leq n$, $r_b \leq n$, the state $x : [a,b]_{\mathbb{T}} \to \mathbb{R}^n$, the control $u : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^m$, the parameter $\omega \in \mathbb{R}^d$, and the data satisfy

$$\begin{aligned}
&L : [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}, && K_a : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}, && K_b : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R},\\
&f : [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^n, && \varphi_a : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^{r_a}, && \varphi_b : \mathbb{R}^n \times \mathbb{R}^d \to \mathbb{R}^{r_b},\\
&\psi : [a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^k.
\end{aligned}$$

The Hamiltonian corresponding to problem (C_ω) is defined by

$$H(t, x, u, p, \lambda, \omega) := p^T f(t, x, u, \omega) + L(t, x, u, \omega) + \lambda^T \psi(t, u, \omega). \tag{4.4}$$

Since we now consider the problem (C_ω) with separated endpoints, we have $r = r_a + r_b$ in the context of problem (C) and

$$\varphi(x,y,\omega) = \begin{pmatrix} \varphi_a(x,\omega) \\ \varphi_b(y,\omega) \end{pmatrix}, \quad \nabla_{(x,y)}\, \varphi(x,y,\omega) = \begin{pmatrix} \nabla_x\, \varphi_a(x,\omega) & 0 \\ 0 & \nabla_y\, \varphi_b(y,\omega) \end{pmatrix}.$$

Throughout this section we fix $\hat{\omega} \in \mathbb{R}^d$ and the feasible pair $(\hat{x}, \hat{u})$ for problem ($\mathrm{C}_{\hat{\omega}}$). We assume the following hypothesis (H1) to hold on the data, where

$$h(t,x,u,\omega) := \big(L(t,x,u,\omega), f(t,x,u,\omega), \psi(t,u,\omega)\big).$$

(H1) There exists $\varepsilon_1 > 0$ such that the functions $(K_a(\cdot,\cdot), \varphi_a(\cdot,\cdot))$, and $(K_b(\cdot,\cdot), \varphi_b(\cdot,\cdot))$ are twice continuously differentiable respectively on $B_{\varepsilon_1}\big(\hat{x}(a), \hat{\omega}\big)$ and $B_{\varepsilon_1}\big(\hat{x}(b), \hat{\omega}\big)$; the function $h(t,\cdot,\cdot,\cdot)$ is twice differentiable in (x,u,ω) on $B_{\varepsilon_1}\big(\hat{x}(t), \hat{u}(t)\big) \times B_{\varepsilon_1}(\hat{\omega})$; the functions $h(t,\cdot,\cdot,\cdot)$, $\nabla_{(x,u)} h(t,\cdot,\cdot,\cdot)$, and $\nabla^2_{(x,u)} h(t,\cdot,\cdot,\cdot)$ are continuous at $(\hat{x}, \hat{u}, \hat{\omega})$ uniformly in t; and for (x,u,ω) in $\mathcal{P}T_{\varepsilon_1}(\hat{x},\hat{u}) \times B_{\varepsilon_1}(\hat{\omega})$ the functions $h(\cdot,x,u,\omega)$, $\nabla_{(x,u)} h(\cdot,x,u,\omega)$, and $\nabla^2_{(x,u)} h(\cdot,x,u,\omega)$ are rd-continuous on $[a,\rho(b)]_{\mathbb{T}}$.

Following the notation in (2.1) and (2.2), we define the matrices $\hat{\mathcal{A}}(t) \in \mathbb{R}^{n\times n}$, $\hat{\mathcal{B}}(t) \in \mathbb{R}^{n\times m}$, $\hat{M}_a \in \mathbb{R}^{r_a\times n}$, $\hat{M}_b \in \mathbb{R}^{r_b\times n}$, $\hat{N}(t) \in \mathbb{R}^{k\times m}$ by

$$\hat{\mathcal{A}}(t) := f_x(t, \hat{x}(t), \hat{u}(t), \hat{\omega}), \quad \hat{\mathcal{B}}(t) := f_u(t, \hat{x}(t), \hat{u}(t), \hat{\omega}), \tag{4.5}$$

$$\hat{M}_a := \nabla_x\, \varphi_a(\hat{x}(a), \hat{\omega}), \quad \hat{M}_b := \nabla_x\, \varphi_b(\hat{x}(b), \hat{\omega}), \quad \hat{N}(t) := \nabla_u\, \psi(t, \hat{u}(t), \hat{\omega}). \tag{4.6}$$

For brevity, let $\hat{Y} : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^{m\times(m-k)}$, $\hat{Y}(\cdot) \in \mathrm{C}_{\mathrm{prd}}$, be the matrix whose columns form an orthonormal basis for $\operatorname{Ker} \hat{N}(t)$, *i.e.*, we have $\hat{N}(t)\, \hat{Y}(t) = 0$ and $\hat{Y}^T(t)\, \hat{Y}(t) = I_{m-k}$ on $[a,\rho(b)]_{\mathbb{T}} \times \mathbb{R}^d$.

For given multipliers $(\hat{p}, \hat{\lambda})$ in $\mathrm{C}^1_{\mathrm{prd}} \times \mathrm{C}_{\mathrm{prd}}$, we also define the function $\hat{S}(t) \in \mathbb{R}^{m\times m}$ by

$$\hat{S}(t) := \hat{H}_{uu}(t) - \mu(t)\, \hat{\mathcal{B}}^T(t)\, [I + \mu(t)\, \hat{\mathcal{A}}^T(t)]^{-1} \hat{H}_{xu}(t) \tag{4.7}$$

where the inverse in the above formula exists by hypothesis (H1), and

$$\hat{H}(t) := H(t, \hat{x}(t), \hat{u}(t), \hat{p}^\sigma(t), \hat{\lambda}(t), \hat{\omega}). \tag{4.8}$$

A pair $(\eta, v) \in \mathrm{C}^1_{\mathrm{prd}} \times \mathrm{C}_{\mathrm{prd}}$ is called *admissible* for problem (C_ω) at the feasible pair (x,u) if

$$\left.\begin{aligned} \eta^\Delta(t) = f_x(t,x(t),u(t),\omega)\, \eta(t) + f_u(t,x(t),u(t),\omega)\, v(t), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \\ \nabla_u\, \psi(t,u(t),\omega)\, v(t) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \\ \nabla_x\, \varphi_a(x(a),\omega)\, \eta(a) = 0, \quad \nabla_x\, \varphi_b(x(b),\omega)\, \eta(b) = 0. \end{aligned}\right\} \tag{4.9}$$

The second variation of the functional $J(\cdot,\omega)$ in problem (C_ω) at a feasible pair (x,u) and multipliers $(p, \lambda, \gamma_a, \gamma_b)$ in the admissible direction (η, v) is defined as

$$J''(\eta, v; x, u, \omega) := \frac{1}{2}\, \eta^T(a)\, \Gamma_a\, \eta(a) + \frac{1}{2}\, \eta^T(b)\, \Gamma_b\, \eta(b) + \frac{1}{2} \int_a^b \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix}^T \nabla^2_{(x,u)}\, H(t) \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix} \Delta t,$$

where the function $\nabla^2_{(x,u)}\, H(t)$ on $[a,\rho(b)]_{\mathbb{T}}$ is given by

$$\nabla^2_{(x,u)}\, H(t) = \begin{pmatrix} H_{xx}(t) & H_{xu}(t) \\ H^T_{xu}(t) & H_{uu}(t) \end{pmatrix} := \nabla^2_{(x,u)} H(t, x(t), u(t), p^\sigma(t), \lambda(t), \omega)$$

and the symmetric $n \times n$ matrices Γ_a and Γ_b are defined by

$$\left.\begin{aligned} \Gamma_a &:= \nabla^2_{xx} K_a(x(a),\omega) + {\gamma_a}^T \nabla^2_{xx} \varphi_a(x(a),\omega), \\ \Gamma_b &:= \nabla^2_{xx} K_b(x(b),\omega) + {\gamma_b}^T \nabla^2_{xx} \varphi_b(x(b),\omega). \end{aligned}\right\} \tag{4.10}$$

Here we have $p : [a,b]_{\mathbb{T}} \to \mathbb{R}^n$, $\lambda : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^k$ and $\gamma_a \in \mathbb{R}^{r_a}$, $\gamma_b \in \mathbb{R}^{r_b}$. We use the notation $J''(\eta,v;\hat\omega)$ and $\hat\Gamma_a$, $\hat\Gamma_b$ to designate respectively $J''(\eta,v;\hat x,\hat u,\hat\omega)$ and Γ_a, Γ_b in which the multipliers are $(\hat p,\hat\lambda,\hat\gamma_a,\hat\gamma_b)$.

In addition to hypothesis (H1) displayed above, we impose the following hypotheses on the base problem $(\mathrm{C}_{\hat\omega})$.

(H2) There are multipliers $(\lambda_0 = 1, \hat p, \hat\lambda, \hat\gamma_a, \hat\gamma_b)$ satisfying with $(\hat x,\hat u)$ the weak maximum principle given by Proposition 2.10 for the problem $(\mathrm{C}_{\hat\omega})$.

(H3) The $n \times n$ matrix $I + \mu(t)\,\hat{\mathcal{A}}(t)$ is invertible for all $t \in [a,\rho(b)]_{\mathbb{T}}$; the matrices $\hat M_a$, $\hat M_b$ and $\hat N(t)$ for $t \in [a,\rho(b)]_{\mathbb{T}}$ in (4.6) have full rank; and the pair $(\hat{\mathcal{A}}(\cdot),\hat{\mathcal{B}}(\cdot))$ is controllable in the sense that for every $\alpha_a \in \mathbb{R}^{r_a}$ and $\alpha_b \in \mathbb{R}^{r_b}$ there exist functions $v \in \mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$ and $\eta \in \mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}$ such that

$$\eta^{\Delta}(t) = \hat{\mathcal{A}}(t)\,\eta(t) + \hat{\mathcal{B}}(t)\,v(t), \quad \hat N(t)\,v(t) = 0, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{4.11}$$

$$\hat M_a\,\eta(a) = \alpha_a, \quad \hat M_b\,\eta(a) = \alpha_b. \tag{4.12}$$

(H4) The second variation $J''(\cdot;\hat\omega)$ is coercive on the space $W^{1,2} \times L^2$, *i.e.*, these exists $\alpha > 0$ such that

$$J''(\eta,v;\hat\omega) \geq \frac{\alpha}{2}\left\{|\eta(a)|^2 + |\eta(b)|^2 + \int_a^b \big(\,|\eta(t)|^2 + |v(t)|^2\,\big)\,\Delta t\right\}$$

for all functions $\eta \in W^{1,2}[a,b]_{\mathbb{T}}$ and $v \in L^2[a,\rho(b)]_{\mathbb{T}}$ satisfying equation (4.11) for almost all $t \in [a,\rho(b)]_{\mathbb{T}}$ and $\hat M_a\,\eta(a) = 0$, $\hat M_b\,\eta(b) = 0$.

(H5) The $(m-k)\times(m-k)$ matrix-valued function $\hat Y^T(t)\,\hat S(t)\,\hat Y(t)$ is invertible on $[a,\rho(b)]_{\mathbb{T}}$, where $\hat S(t)$ is given in (4.7).

Remark 4.1. As in Proposition 2.9, the controllability assumption in (H3) can be equivalently formulated in terms of the corresponding $(r_a + r_b)\times(r_a + r_b)$ Grammian matrix

$$\begin{pmatrix} \varepsilon_0^2\,\hat M_a\,\hat M_a^T & \varepsilon_0^2\,\hat M_a\,\hat{\mathcal{D}}_b^T \\ \varepsilon_0^2\,\hat{\mathcal{D}}_b\,\hat M_a^T & \hat{\mathcal{Z}}_b \end{pmatrix} > 0$$

for some $\varepsilon_0 > 0$, where $\hat{\mathcal{D}}_b \in \mathbb{R}^{r_b\times n}$, $\hat{\mathcal{Z}}_b \in \mathbb{R}^{r_b\times r_b}$, and $\hat{\mathcal{E}}_b(t) \in \mathbb{R}^{r_b\times(m-k)}$ are defined by

$$\hat{\mathcal{D}}_b := \hat M_b\,\hat\Phi(b), \quad \hat{\mathcal{Z}}_b := \varepsilon_0^2\,\hat{\mathcal{D}}_b\,\hat{\mathcal{D}}_b^T + \int_a^b \hat{\mathcal{E}}_b(t)\,\hat{\mathcal{E}}_b^T(t)\,\hat\Delta t,$$

$$\hat{\mathcal{E}}_b(t) := \hat M_b\,\hat\Phi(b)\,[\hat\Phi^{\sigma}(t)]^{-1}\hat{\mathcal{B}}(t)\,\hat Y(t).$$

Here $\hat\Phi(t)$ is the fundamental matrix of $\hat\Phi^{\Delta} = \hat{\mathcal{A}}(t)\,\hat\Phi$ on $[a,\rho(b)]_{\mathbb{T}}$ with $\hat\Phi(a) = I$.

Remark 4.2.

(i) Hypotheses (H1)–(H4) imply through Theorem 3.3 that $(\hat x,\hat u)$ is a weak local minimum for $(\mathrm{C}_{\hat\omega})$.

(ii) The invertibility of $I + \mu(t)\,\hat{\mathcal{A}}(t)$ on $[a,\rho(b)]_{\mathbb{T}}$, the full rank property of $\hat N(t)$ on $[a,\rho(b)]_{\mathbb{T}}$, and Remark 2.3 yield the existence of $\varepsilon_2 \in (0,\varepsilon_1)$ such that if $t \in [a,\rho(b)]_{\mathbb{T}}$ and (x,u,ω) satisfy $(x,u,\omega) \in B_{\varepsilon_2}(\hat x(t)) \times B_{\varepsilon_2}(\hat u(t)) \times B_{\varepsilon_2}(\hat\omega)$, then $\nabla_u\,\psi(t,u,\omega)$ is of full rank and the matrix $I + \mu(t)\,f_x(t,x,u,\omega)$ is invertible.

(iii) By using Remark 4.1 and part (ii) of this remark, hypothesis (H3) implies that for some $\varepsilon_3 \in (0,\varepsilon_2)$ the system (4.9) is controllable for all $(x,u) \in \mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}\times\mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$ with $\|x-\hat x\|_{\mathrm{C}_{\mathrm{prd}}} < \varepsilon_3$, $\|u-\hat u\|_{\mathrm{C}_{\mathrm{prd}}} < \varepsilon_3$ and $\omega \in B_{\varepsilon_3}(\hat\omega)$. Note that a direct proof of this fact can be easily made along the one given for the continuous time setting in ([22], p. 319).

(iv) We recall that the space L^2 consists of functions y which are Lebesgue measurable on $[a, \rho(b)]_{\mathbb{T}}$ and $\int_a^b |y(t)|^2 \Delta t$ is finite, and the space $W^{1,2}$ consists of functions y, such that y is Lebesgue measurable on $[a, b]_{\mathbb{T}}$, y^{Δ} exists almost everywhere and is Lebesgue measurable on $[a, \rho(b)]_{\mathbb{T}}$, and $y, y^{\Delta} \in L^2$. We refer to [1, 4, 24, 48] for properties of the Sobolev and Lebesgue spaces on time scales. We also note that the coercivity assumption in (H4) above is slightly stronger than in Definition 3.1, since the considered spaces are different, namely $\mathrm{C}^1_{\mathrm{prd}} \times \mathrm{C}_{\mathrm{prd}} \subseteq W^{1,2} \times L^2$.

(v) The controllability assumption in hypothesis (H3) does not only produce a surjective map from $\{(\eta, v) \in \mathrm{C}^1_{\mathrm{prd}} \times \mathrm{C}_{\mathrm{prd}}\}$ onto $\mathrm{C}_{\mathrm{prd}} \times \mathrm{C}_{\mathrm{prd}} \times \mathbb{R}^{r_a + r_b}$, but, similarly to the continuous time setting (see [22], (Lem. 1)), it also yields the surjectivity of the linearized constraints from $W^{1,2} \times L^2$ onto $L^2 \times L^2 \times \mathbb{R}^{r_a + r_b}$. That is, for any $(d, c, (\alpha_a, \alpha_b)) \in L^2 \times L^2 \times \mathbb{R}^{r_a + r_b}$ there exists $(\eta, v) \in W^{1,2} \times L^2$ satisfying (4.12) and

$$\eta^{\Delta}(t) = \hat{\mathcal{A}}(t)\, \eta(t) + \hat{\mathcal{B}}(t)\, v(t) + d(t), \quad \hat{N}(t)\, v(t) = c(t), \quad t \in [a, \rho(b)]_{\mathbb{T}}.$$

(vi) Hypotheses (H1) and (H4), and part (v) of this remark imply by [21], (Lem. 6) that there is $\varepsilon_4 \in (0, \varepsilon_3)$ such that for all $(\omega, \gamma_a, \gamma_b, x, u, p, \lambda)$ such that

$$\max \left\{ |\omega - \hat{\omega}|,\, |\gamma_a - \hat{\gamma}_a|,\, |\gamma_b - \hat{\gamma}_b|,\, \|(x, p) - (\hat{x}, \hat{p})\|_{\mathrm{C}_{\mathrm{prd}}},\, \|(u, \lambda) - (\hat{u}, \hat{\lambda})\|_{\mathrm{C}_{\mathrm{prd}}} \right\} < \varepsilon_4$$

the functional $J''(\eta, v; x, u, \omega)$ is coercive for all (η, v) which are admissible for (C_{ω}), *i.e.*, satisfying (4.9).

(vii) Assumption (H5) reduces to $\hat{S}(t)$ being invertible on $[a, \rho(b)]_{\mathbb{T}}$ when no control constraints are present $(k = 0)$, while it is vacuous when the maximal number of control constraints is imposed $(k = m)$.

The main result of this section is formulated as follows. The positive numbers ε_2 and ε_4 in the statement and its proof refer to parts (ii) and (vi) of Remark 4.2.

Theorem 4.3 (Sensitivity analysis)**.** *Under hypotheses* (H1)–(H5) *there exists* $\varepsilon \in (0, \varepsilon_4)$ *such that for all* $\omega \in \mathbb{R}^d$ *with* $|\omega - \hat{\omega}| < \varepsilon$ *the problem* (C_{ω}) *has a strict weak local minimum* $(x(\cdot, \omega), u(\cdot, \omega))$ *with multipliers* $p(\cdot, \omega) : [a, b]_{\mathbb{T}} \to \mathbb{R}^n$, $p(\cdot, \omega) \in \mathrm{C}^1_{\mathrm{prd}}$, *and* $\lambda(\cdot, \omega) : [a, \rho(b)]_{\mathbb{T}} \to \mathbb{R}^k$, $\lambda(\cdot, \omega) \in \mathrm{C}_{\mathrm{prd}}$, *and vectors* $\gamma_a(\omega) \in \mathbb{R}^{r_a}$, $\gamma_b(\omega) \in \mathbb{R}^{r_b}$, *which are* C^1 *in the argument* ω *and satisfy equations* (4.1)–(4.3) *and*

$$-p^{\Delta}(t, \omega) = H_x^T\big(t, x(t, \omega), u(t, \omega), p^{\sigma}(t, \omega), \lambda(t, \omega), \omega\big), \quad t \in [a, \rho(b)]_{\mathbb{T}}, \tag{4.13}$$

$$H_u^T\big(t, x(t, \omega), u(t, \omega), p^{\sigma}(t, \omega), \lambda(t, \omega), \omega\big) = 0, \quad t \in [a, \rho(b)]_{\mathbb{T}}, \tag{4.14}$$

$$-p(a, \omega) = \nabla_x K^T\big(x(a, \omega), \omega\big) + \nabla_x \varphi_a^T\big(x(a, \omega), \omega\big)\, \gamma_a(\omega), \tag{4.15}$$

$$p(b, \omega) = \nabla_x K^T\big(x(b, \omega), \omega\big) + \nabla_x \varphi_b^T\big(x(b, \omega), \omega\big)\, \gamma_b(\omega). \tag{4.16}$$

Proof. The proof aims to construct, for ω near $\hat{\omega}$, a family of pairs $(x(t, \omega), u(t, \omega))$ which are feasible for (C_{ω}), *i.e.*, satisfying equations (4.1)–(4.3), and multipliers $(p(t, \omega), \lambda(t, \omega), \gamma_a(\omega), \gamma_b(\omega))$ satisfying with $(x(t, \omega), u(t, \omega))$ the weak Pontryagin maximum principle, *i.e.*, equations (4.13)–(4.16). Then, by means of Theorem 4.3 and parts (ii), (iii), and (vi) of Remark 4.2, the strict weak local minimality of $(x(t, \omega), u(t, \omega))$ for (C_{ω}) follows.

Step 1. In this step we construct for $\xi := (x,p,\omega)$ that is uniformly near $\hat{\xi}(t) := (\hat{x}(t),\hat{p}(t),\hat{\omega})$, a family $(u(t,\xi),\lambda(t,\xi))$ that satisfies the control constraint (4.2) and the stationarity condition (4.14). By hypothesis (H2), we have $(\hat{x},\hat{u},\hat{\omega})$, $\hat{p}:[a,b]_{\mathbb{T}}\to\mathbb{R}^n$, $\hat{p}\in \mathrm{C}^1_{\mathrm{prd}}$, $\hat{\lambda}:[a,\rho(b)]_{\mathbb{T}}\to\mathbb{R}^k$, $\hat{\lambda}\in\mathrm{C}_{\mathrm{prd}}$, and $\hat{\gamma}_a\in\mathbb{R}^{r_a}$, $\hat{\gamma}_b\in\mathbb{R}^{r_b}$, which satisfy

$$\hat{x}^{\Delta}(t) = f(t,\hat{x}(t),\hat{u}(t),\hat{\omega}), \quad t\in[a,\rho(b)]_{\mathbb{T}}, \tag{4.17}$$

$$\psi(t,\hat{u}(t),\hat{\omega}) = 0, \quad t\in[a,\rho(b)]_{\mathbb{T}}, \tag{4.18}$$

$$\varphi_a(\hat{x}(a),\hat{\omega}) = 0, \quad \varphi_b(\hat{x}(b),\hat{\omega}) = 0, \tag{4.19}$$

$$-\hat{p}^{\Delta}(t) = \hat{H}_x^T(t) = \hat{\mathcal{A}}^T(t)\,\hat{p}^{\sigma}(t) + \hat{L}_x^T(t), \quad t\in[a,\rho(b)]_{\mathbb{T}}, \tag{4.20}$$

$$\hat{H}_u^T(t) = \hat{\mathcal{B}}^T(t)\,\hat{p}^{\sigma}(t) + \hat{L}_u^T(t) + \hat{N}^T(t)\,\hat{\lambda}(t) = 0, \quad t\in[a,\rho(b)]_{\mathbb{T}}, \tag{4.21}$$

$$-\hat{p}(a) = \nabla_x K_a^T(\hat{x}(a),\hat{\omega}) + \hat{M}_a^T\hat{\gamma}_a, \quad \hat{p}(b) = \nabla_x K_b^T(\hat{x}(b),\hat{\omega}) + \hat{M}_b^T\hat{\gamma}_b. \tag{4.22}$$

Here the Hamiltonian is defined in (4.4), and we use the abbreviated notation of (4.5)–(4.8) at $\omega=\hat{\omega}$, and

$$\hat{L}(t) := L(t,\hat{x}(t),\hat{u}(t),\hat{\omega}).$$

The same notation is applied with the first and second order gradients of the functions L, H, and ψ. From equations (4.20) and (4.21) we then obtain on $[a,\rho(b)]_{\mathbb{T}}$ the expressions (suppressing the argument t)

$$\hat{p}^{\sigma} = (I+\mu\hat{\mathcal{A}}^T)^{-1}(\hat{p}-\mu\hat{L}_x^T), \tag{4.23}$$

$$\hat{\mathcal{B}}^T(I+\mu\hat{\mathcal{A}}^T)^{-1}(\hat{p}-\mu\hat{L}_x^T) + \hat{L}_u^T + \hat{N}^T\hat{\lambda} = 0. \tag{4.24}$$

By Remark 4.2(ii), the matrix $I+\mu(t)\,f_x(t,x,u,\omega)$ is invertible for all $(x,u,\omega)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^d$ such that $|(x,u,\omega)-(\hat{x}(t),\hat{u}(t),\hat{\omega})|<\varepsilon_2$ for all $t\in[a,\rho(b)]_{\mathbb{T}}$. Note that, by the formula $(A^{-1})' = -A^{-1}A'A^{-1}$ for the derivative of the inverse of the matrix function A, we have

$$\nabla_u\,(I+\mu f_x^T)^{-1} = -\mu\,\tilde{f}_x^T f_{xu}^T\,\tilde{f}_x^T, \quad \text{where} \quad \tilde{f}_x := (I+\mu f_x^T)^{-1}. \tag{4.25}$$

Consider now the mapping $F:[a,\rho(b)]_{\mathbb{T}}\times\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}^d\to\mathbb{R}^{m+k}$,

$$F(t,x,u,p,\lambda,\omega) := \begin{pmatrix} G(t,x,u,p,\lambda,\omega) \\ \psi(t,u,\omega)\end{pmatrix},$$

where the function G is defined by

$$\begin{aligned} G(t,x,u,p,\lambda,\omega) := {} & f_u^T(t,x,u,\omega)\,[I+\mu(t)\,f_x^T(t,x,u,\omega)]^{-1}\,[p-\mu(t)\,L_x^T(t,x,u,\omega)] \\ & + L_u^T(t,x,u,\omega) + \lambda^T\psi_u(t,u,\omega). \end{aligned} \tag{4.26}$$

We aim to solve the equation $F(t,x,u,p,\lambda,\omega)=0$ for (u,λ) in terms of (t,x,p,ω) near $(t,\hat{x}(t),\hat{u}(t),\hat{p}(t),\hat{\lambda}(t),\hat{\omega})$ on $[a,\rho(b)]_{\mathbb{T}}$ by the implicit function theorem. By (4.25), the partial derivatives of G are given by (suppressing the arguments of the functions)

$$G_x = f_{ux}^T\,\tilde{f}_x^T(p-\mu L_x^T) - \mu\,f_u^T\,\tilde{f}_x^T\,[f_{xx}^T\,\tilde{f}_x^T(p-\mu L_x^T) + L_{xx}^T] + L_{ux}^T,$$

$$G_u = f_{uu}^T\,\tilde{f}_x^T(p-\mu L_x^T) - \mu\,f_u^T\,\tilde{f}_x^T\,[f_{xu}^T\,\tilde{f}_x^T(p-\mu L_x^T) + L_{xu}^T] + L_{uu}^T + \lambda^T\psi_{uu},$$

$$G_p = f_u^T\,\tilde{f}_x^T,$$

$$G_\omega = f_{u\omega}^T\,\tilde{f}_x^T(p-\mu L_x^T) - \mu\,f_u^T\,\tilde{f}_x^T\,[f_{x\omega}^T\,\tilde{f}_x^T(p-\mu L_x^T) + L_{x\omega}^T] + L_{u\omega}^T + \lambda^T\psi_{u\omega}.$$

Evaluating at $(t, \hat{x}(t), \hat{u}(t), \hat{p}(t), \hat{\lambda}(t), \hat{\omega})$ and using formulas (4.23) and (4.7) we obtain

$$\left.\begin{aligned} \hat{G}_x &= \hat{f}_{ux}^T \hat{p}^\sigma - \mu \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} (\hat{f}_{xx}^T \hat{p}^\sigma + \hat{L}_{xx}^T) + \hat{L}_{ux}^T \\ &= \hat{H}_{ux} - \mu \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} \hat{H}_{xx}, \\ \hat{G}_u &= \hat{f}_{uu}^T \hat{p}^\sigma - \mu \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} (\hat{f}_{xu}^T \hat{p}^\sigma + \hat{L}_{xu}^T) + \hat{L}_{uu}^T + \lambda^T \hat{\psi}_{uu} \\ &= \hat{H}_{uu} - \mu \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} \hat{H}_{xu} = \hat{S}, \\ \hat{G}_p &= \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1}, \\ \hat{G}_\omega &= \hat{f}_{u\omega}^T \hat{p}^\sigma - \mu \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} (\hat{f}_{x\omega}^T \hat{p}^\sigma + \hat{L}_{x\omega}^T) + \hat{L}_{u\omega}^T + \lambda^T \hat{\psi}_{u\omega} \\ &= \hat{H}_{u\omega} - \mu \hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} \hat{H}_{x\omega}. \end{aligned}\right\} \tag{4.27}$$

Therefore, the $(m+k) \times (m+k)$ Jacobi matrix of the mapping F with respect to (u, λ) is

$$\hat{K}(t) := \nabla_{(u,\lambda)} \hat{F}(t) = \begin{pmatrix} \hat{G}_u(t) & \hat{\psi}_u^T(t) \\ \hat{\psi}_u(t) & 0 \end{pmatrix} = \begin{pmatrix} \hat{S}(t) & \hat{N}^T(t) \\ \hat{N}(t) & 0 \end{pmatrix}.$$

We will show that the matrix $\hat{K}(t)$ is invertible. In fact, by assumptions (H3) and (H5) the matrices $\hat{N}(t)\, \hat{N}^T(t)$ and $\hat{Y}^T(t)\, \hat{S}(t)\, \hat{Y}(t)$ are invertible and by Remark 2.3, their inverses are piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$. One can verify that (suppressing the argument t)

$$\hat{K}^{-1} = \begin{pmatrix} \hat{Y}(\hat{Y}^T \hat{S} \hat{Y})^{-1} \hat{Y}^T & [I - \hat{Y}(\hat{Y}^T \hat{S} \hat{Y})^{-1} \hat{Y}^T \hat{S}]\, \hat{N}^\dagger \\ \hat{N}^{\dagger T} [I - \hat{S} \hat{Y}(\hat{Y}^T \hat{S} \hat{Y})^{-1} \hat{Y}^T] & -\hat{N}^{\dagger T} \hat{S}\, [I - \hat{Y}(\hat{Y}^T \hat{S} \hat{Y})^{-1} \hat{Y}^T \hat{S}]\, \hat{N}^\dagger \end{pmatrix}, \tag{4.28}$$

where $\hat{N}^\dagger = \hat{N}^T (\hat{N} \hat{N}^T)^{-1} \in \mathbb{R}^{m \times k}$ is the Moore–Penrose pseudoinverse of $\hat{N}$. Indeed, by using the identity $\hat{Y} \hat{Y}^T = I - \hat{N}^\dagger \hat{N}$ we calculate that $\hat{K} \hat{K}^{-1} = I_{m+k}$. Moreover, $\hat{K}^{-1}(t)$ is also piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$.

Given that $\hat{K}(t)$ is invertible and $\hat{F}(t) = F(t, \hat{x}(t), \hat{u}(t), \hat{p}(t), \hat{\lambda}(t), \hat{\omega}) = 0$ on $[a, \rho(b)]_{\mathbb{T}}$, by the implicit function theorem there exist $\varepsilon_5 \in (0, \varepsilon_4)$ and a unique pair of functions $(u, \lambda) : D \to \mathbb{R}^m \times \mathbb{R}^k$ such that $\big|(u(t,\xi), \lambda(t,\xi)) - (\hat{u}(t), \hat{\lambda}(t))\big| < \varepsilon_4$ and

$$F\big(t, x, u(t,\xi), p, \lambda(t,\xi), \omega\big) = \begin{pmatrix} G\big(t, x, u(t,\xi), p, \lambda(t,\xi), \omega\big) \\ \psi\big(t, u(t,\xi), \omega\big) \end{pmatrix} = 0 \tag{4.29}$$

on D, where $\xi := (x, p, \omega)$ and $\hat{\xi}(t) = (\hat{x}(t), \hat{p}(t), \hat{\omega})$ as above and

$$D := \Big\{(t, \xi) \in [a, \rho(b)]_{\mathbb{T}} \times \mathbb{R}^{2n} \times \mathbb{R}^d,\ \big|(x, p) - (\hat{x}(t), \hat{p}(t))\big| < \varepsilon_5,\ |\omega - \hat{\omega}| < \varepsilon_5\Big\}. \tag{4.30}$$

That is, by (4.26) we have for $(t, \xi) \in D$ the equations

$$\left.\begin{aligned} f_u^T(t, x, u(t,\xi), \omega)\ \ & \big[I + \mu(t)\, f_x^T(t, x, u(t,\xi), \omega)\big]^{-1} \big[p - \mu(t)\, L_x^T(t, x, u(t,\xi), \omega)\big] \\ & -L_u^T(t, x, u(t,\xi), \omega) + \lambda^T(t,\xi)\, \psi_u(t, u(t,\xi), \omega) = 0, \\ \psi(t, u(t,\xi), \omega) &= 0. \end{aligned}\right\} \tag{4.31}$$

The functions $u(t, \cdot)$ and $\lambda(t, \cdot)$ are C^1 in the argument $\xi = (x, p, \omega)$ uniformly in t, the functions $u(\cdot, \xi)$ and $\lambda(\cdot, \xi)$ are $\mathrm{C}_{\mathrm{prd}}$ in the argument t. By differentiating (4.29) with respect to ξ we then obtain the formula

$$\begin{pmatrix} \hat{u}_\xi \\ \hat{\lambda}_\xi \end{pmatrix} = \begin{pmatrix} \hat{u}_x & \hat{u}_p & \hat{u}_\omega \\ \hat{\lambda}_x & \hat{\lambda}_p & \hat{\lambda}_\omega \end{pmatrix} = -\hat{K}^{-1} \begin{pmatrix} \hat{G}_x & \hat{G}_p & \hat{G}_\omega \\ 0 & 0 & \hat{\psi}_\omega \end{pmatrix}. \tag{4.32}$$

The explicit formulas for the partial derivatives $\hat{u}_x$, $\hat{u}_p$, $\hat{u}_\omega$ and $\hat{\lambda}_x$, $\hat{\lambda}_p$, $\hat{\lambda}_\omega$ now follow by substituting the expressions in (4.28) and (4.27) into equation (4.32).

In the remaining steps of this proof we construct for ω near $\hat{\omega}$ a solution to the boundary value problem (BVP) in (x,p) obtained by inserting the functions $u(t,\xi)$ and $\lambda(t,\xi)$ found in Step 1 into equations (4.1) and (4.13) with boundary conditions given by (4.3), (4.15), and (4.16). The method we use is a modification and generalization of the "shooting method" used for the continuous time setting, see *e.g.* [37].

Step 2. We will perturb the initial state $\hat{x}(a)$ and the multiplier $\hat{\gamma}_a$ in a special way in terms of the parameter ω and a "shooting parameter" s, while preserving the initial condition $\varphi_a(x(a,s,\omega),\omega)=0$. Consider the equation

$$g(s,\gamma,\omega)=0, \quad g(s,\gamma,\omega):=\varphi_a\big(s-\hat{M}_a^T(\gamma-\hat{\gamma}_a),\omega\big),$$

where $s\in\mathbb{R}^n$ and $\gamma\in\mathbb{R}^{r_a}$. Then $g(\hat{x}(a),\hat{\gamma}_a,\hat{\omega})=\varphi_a(\hat{x}(a),\hat{\omega})=0$ and the partial derivatives

$$g_s(s,\gamma,\omega)=\nabla_x\,\varphi_a\big(s-\hat{M}_a^T(\gamma-\hat{\gamma}_a),\omega\big), \quad g_\gamma(s,\gamma,\omega)=-\nabla_x\,\varphi_a\big(s-\hat{M}_a^T(\gamma-\hat{\gamma}_a),\omega\big)\,\hat{M}_a^T.$$

By evaluating at the point $(s,\gamma,\omega)=(\hat{x}(a),\hat{\gamma}_a,\hat{\omega})$ we obtain that $g_s(\hat{x}(a),\hat{\gamma}_a,\hat{\omega})=\hat{M}_a$ and $g_\gamma(\hat{x}(a),\hat{\gamma}_a,\hat{\omega})=-\hat{M}_a\hat{M}_a^T$, which is an invertible $r_a\times r_a$ matrix. Hence, by the implicit function theorem there exists $\varepsilon_6\in(0,\varepsilon_5)$ and a unique C^1 function $\gamma_a(s,\omega)$ such that for all $(s,\omega)\in B_{\varepsilon_6}(\hat{x}(a))\times B_{\varepsilon_6}(\hat{\omega})$ we have

$$|\gamma_a(s,\omega)-\hat{\gamma}_a|<\varepsilon_5, \quad \gamma_a(\hat{x}(a),\hat{\omega})=\hat{\gamma}_a, \quad \varphi_a\big(s-\hat{M}_a^T(\gamma_a(s,\omega)-\hat{\gamma}_a),\omega\big)=0, \tag{4.33}$$

$$\nabla_s\,\gamma_a(s,\omega)=-g_\gamma^{-1}(s,\gamma_a(s,\omega),\omega)\,g_s(s,\gamma_a(s,\omega),\omega), \quad \nabla_s\,\gamma_a(\hat{x}(a),\hat{\omega})=(\hat{M}_a\hat{M}_a^T)^{-1}\hat{M}_a. \tag{4.34}$$

Step 3. On the interval $[a,\rho(b)]_{\mathbb{T}}$ we now consider the time scale differential system

$$\left.\begin{aligned} x^\Delta &= f(t,x,u(t,\xi),\omega),\\ -p^\Delta &= f_x^T(t,x,u(t,\xi),\omega)\,\big[I+\mu(t)\,f_x^T(t,x,u(t,\xi),\omega)\big]^{-1}p\\ &\quad +\big[I+\mu(t)\,f_x^T(t,x,u(t,\xi),\omega)\big]^{-1}L_x^T(t,x,u(t,\xi),\omega) \end{aligned}\right\} \tag{4.35}$$

together with the initial conditions

$$\left.\begin{aligned} x(a) &= s-\hat{M}_a^T(\gamma_a(s,\omega)-\hat{\gamma}_a),\\ p(a) &= -\nabla_x\,K_a^T\big(s-\hat{M}_a^T(\gamma_a(s,\omega)-\hat{\gamma}_a),\omega\big)\\ &\quad -\nabla_x\,\varphi_a^T\big(s-\hat{M}_a^T(\gamma_a(s,\omega)-\hat{\gamma}_a),\omega\big)\,\gamma_a(s,\omega), \end{aligned}\right\} \tag{4.36}$$

where $\xi=(x,p,\omega)$ as before, $u(t,\xi)$ is the function from Step 1 satisfying (4.31), and $\gamma_a(s,\omega)$ is the function from Step 2 satisfying (4.33). We know that the functions $\hat{x}(t)$ and $\hat{p}(t)$ solve system (4.35) at $(s,\omega)=(\hat{x}(a),\hat{\omega})$ and that the function $\gamma_a(s,\omega)$ is C^1 for $(s,\omega)\in B_{\varepsilon_6}(\hat{x}(a))\times B_{\varepsilon_6}(\hat{\omega})$. From the properties of $u(t,\xi)$ and the data we have that the right-hand side of (4.35) satisfies the assumptions of the embedding theorem [26], (Thm. 3.1) and hence, there exists $\varepsilon_7\in(0,\varepsilon_6)$ such that for every $(s,\omega)\in B_{\varepsilon_7}(\hat{x}(a))\times B_{\varepsilon_7}(\hat{\omega})$ there is a unique solution $x(t,s,\omega)$ and $p(t,s,\omega)$ of (4.35) and (4.36) with the following properties:

(i) The functions $x(t,\hat{x}(a),\hat{\omega})=\hat{x}(t)$ and $p(t,\hat{x}(a),\hat{\omega})=\hat{p}(t)$ on $[a,b]_{\mathbb{T}}$.
(ii) The functions $x(\cdot,\cdot,\cdot)$ and $p(\cdot,\cdot,\cdot)$ are continuous on $[a,b]_{\mathbb{T}}\times B_{\varepsilon_7}(\hat{x}(a))\times B_{\varepsilon_7}(\hat{\omega})$.
(iii) The pair $(x(t,s,\omega),p(t,s,\omega))\in B_{\varepsilon_5}(\hat{x}(t),\hat{p}(t))$ for all $t\in[a,b]_{\mathbb{T}}$ and $(s,\omega)\in B_{\varepsilon_7}(\hat{x}(a))\times B_{\varepsilon_7}(\hat{\omega})$ and hence, the equalities in (4.31) hold with $x=x(t,s,\omega)$ and $p=p(t,s,\omega)$, as $(t,(x,p),\omega)\in D$, defined by (4.30). That is,

$$\left.\begin{aligned} &G\Big(t,x(t,s,\omega),u\big(t,x(t,s,\omega),p(t,s,\omega),\omega\big),p(t,s,\omega),\\ &\qquad\qquad\qquad\qquad \lambda\big(t,x(t,s,\omega),p(t,s,\omega),\omega\big),\omega\Big)=0,\\ &\psi\Big(t,u\big(t,x(t,s,\omega),p(t,s,\omega),\omega\big),\omega\Big)=0, \end{aligned}\right\} \tag{4.37}$$

hold on $[a,\rho(b)]_{\mathbb{T}}\times B_{\varepsilon_7}(\hat{x}(a))\times B_{\varepsilon_7}(\hat{\omega})$.

(iv) The functions $x(t,\cdot,\cdot)$ and $p(t,\cdot,\cdot)$ are C^1 and the time scale derivatives $x^{\Delta}(t,\cdot,\cdot)$ and $p^{\Delta}(t,\cdot,\cdot)$ satisfy on $[a,\rho(b)]_{\mathbb{T}}$ the identity

$$\nabla_{(s,\omega)} \begin{pmatrix} x^{\Delta}(t,s,\omega) \\ p^{\Delta}(t,s,\omega) \end{pmatrix} \Bigg|_{(s,\omega)=(\hat{x}(a),\hat{\omega})} = \begin{pmatrix} \nabla_{(s,\omega)}\, x(t,\hat{x}(a),\hat{\omega}) \\ \nabla_{(s,\omega)}\, p(t,\hat{x}(a),\hat{\omega}) \end{pmatrix}^{\Delta}.$$

(v) The pair of $n\times n$ matrix-valued functions

$$\big(Z(t),\Lambda(t)\big) := \big(\nabla_s\, x(t,\hat{x}(a),\hat{\omega}), \nabla_s\, p(t,\hat{x}(a),\hat{\omega})\big), \quad t\in[a,b]_{\mathbb{T}}, \tag{4.38}$$

satisfies the linearized system of (4.35) and (4.36) with respect to s at $(\hat{x}(a),\hat{\omega})$. That is, by using (4.32) and (4.27) we have

$$Z^{\Delta} = \mathbb{A}(t)\, Z - \mathbb{B}(t)\, \Lambda, \quad -\Lambda^{\Delta} = \mathbb{C}(t)\, Z - \mathbb{D}(t)\, \Lambda, \quad t\in[a,\rho(b)]_{\mathbb{T}}, \tag{4.39}$$

$$Z(a) = I - \mathcal{M}_a, \quad -\Lambda(a) = \hat{\Gamma}_a(I-\mathcal{M}_a) + \mathcal{M}_a, \tag{4.40}$$

where the $n\times n$ coefficients $\mathbb{A}(t)$, $\mathbb{B}(t)$, $\mathbb{C}(t)$, $\mathbb{D}(t)$ are (suppressing the argument t)

$$\left.\begin{aligned}
\mathbb{A} &:= \hat{\mathcal{A}} - \hat{\mathcal{B}}\hat{Y}(\hat{Y}^T\hat{S}\hat{Y})^{-1}\hat{Y}^T\big[\hat{H}_{ux} - \mu\,\hat{\mathcal{B}}^T(I+\mu\hat{\mathcal{A}}^T)^{-1}\hat{H}_{xx}\big],\\
\mathbb{B} &:= \hat{\mathcal{B}}\hat{Y}(\hat{Y}^T\hat{S}\hat{Y})^{-1}\hat{Y}^T\hat{\mathcal{B}}^T(I+\mu\hat{\mathcal{A}}^T)^{-1},\\
\mathbb{C} &:= (I+\mu\hat{\mathcal{A}}^T)^{-1}\big[\hat{H}_{xx} - \hat{H}_{xu}\,\hat{Y}(\hat{Y}^T\hat{S}\hat{Y})^{-1}\hat{Y}^T\big(\hat{H}_{ux} - \mu\,\hat{\mathcal{B}}^T(I+\mu\hat{\mathcal{A}}^T)^{-1}\hat{H}_{xx}\big)\big],\\
\mathbb{D} &:= (I+\mu\hat{\mathcal{A}}^T)^{-1}\big[\hat{H}_{xu}\,\hat{Y}(\hat{Y}^T\hat{S}\hat{Y})^{-1}\hat{Y}^T\hat{\mathcal{B}}^T(I+\mu\hat{\mathcal{A}}^T)^{-1} - \hat{\mathcal{A}}^T\big],
\end{aligned}\right\} \tag{4.41}$$

and the symmetric $n\times n$ matrix $\mathcal{M}_a$ is defined by

$$\mathcal{M}_a := \hat{M}_a^T(\hat{M}_a\hat{M}_a^T)^{-1}\hat{M}_a. \tag{4.42}$$

We note that system (4.39) is a time scale symplectic system according to the terminology of [52], (Thm. 4.8), whose coefficients $\mathbb{A}(t)$, $\mathbb{B}(t)$, $\mathbb{C}(t)$, $\mathbb{D}(t)$ are piecewise rd-continuous on $[a,\rho(b)]_{\mathbb{T}}$ by (4.41).

Step 4. In this step we show that there exists $\varepsilon_8\in(0,\varepsilon_7)$ and continuously differentiable functions $s: B_{\varepsilon_8}(\hat{\omega}) \to B_{\varepsilon_7}(\hat{x}(a))$ and $\gamma_b: B_{\varepsilon_8}(\hat{\omega}) \to B_{\varepsilon_7}(\hat{\gamma}_b)$ such that the solution $\big(x(t,s(\omega),\omega), p(t,s(\omega),\omega)\big)$ of (4.35) and (4.36) also satisfies for $\omega\in B_{\varepsilon_8}(\hat{\omega})$ the equations

$$\left.\begin{aligned}
\varphi_b\big(x(b,s(\omega),\omega),\omega\big) &= 0,\\
p(b,s(\omega),\omega) &= \nabla_x K_b^T\big(x(b,s(\omega),\omega),\omega\big) + \nabla_x\,\varphi_b^T\big(x(b,s(\omega),\omega),\omega\big)\,\gamma_b(\omega).
\end{aligned}\right\} \tag{4.43}$$

To prove this we will apply the implicit function theorem to the r_b+n equations (4.43). Let

$$Q(s,\gamma,\omega) = \begin{pmatrix} \varphi_b\big(x(b,s,\omega),\omega\big) \\ p(b,s,\omega) - \nabla_x K_b^T\big(x(b,s,\omega),\omega\big) - \nabla_x\,\varphi_b^T\big(x(b,s,\omega),\omega\big)\,\gamma \end{pmatrix}. \tag{4.44}$$

Then we have $Q(\hat{x}(b),\hat{\gamma}_b,\hat{\omega})=0$. With the $r_b\times n$ matrix $\hat{M}_b$ given in (4.6) and the symmetric $n\times n$ matrix $\hat{\Gamma}_b$ defined in (4.10), where $x(b)=\hat{x}(b)$, $\gamma_b=\hat{\gamma}_b$, $\omega=\hat{\omega}$, and with the notation (4.38) we calculate

$$\begin{aligned}
T_b := \nabla_{(s,\gamma)}\, Q\big(\hat{x}(b),\hat{\gamma}_b,\hat{\omega}\big) &= \begin{pmatrix} \hat{M}_b\, x_s(b,\hat{x}(a),\hat{\omega}) & 0 \\ p_s(b,\hat{x}(a),\hat{\omega}) - \hat{\Gamma}_b\, x_s(b,\hat{x}(a),\hat{\omega}) & -\hat{M}_b^T \end{pmatrix}\\
&= \begin{pmatrix} \hat{M}_b\, Z(b) & 0 \\ \hat{\Lambda}(b) - \hat{\Gamma}_b\, Z(b) & -\hat{M}_b^T \end{pmatrix}.
\end{aligned} \tag{4.45}$$

We note that this Jacobian matrix has dimensions $(r_b + n) \times (n + r_b)$. To verify that the matrix T_b in (4.45) is invertible, we show that for every pair $(\alpha, \beta) \in \mathbb{R}^{r_b} \times \mathbb{R}^n$ there exists a unique pair $(c, d) \in \mathbb{R}^n \times \mathbb{R}^{r_b}$, which solves the linear algebraic system

$$\hat{M}_b \, Z(b) \, c = \alpha, \quad [\Lambda(b) - \hat{\Gamma}_b \, Z(b)] \, c - \hat{M}_b^T d = \beta. \tag{4.46}$$

For fixed vectors $\alpha \in \mathbb{R}^{r_b}$ and $\beta \in \mathbb{R}^n$ we consider the variational problem to minimize

$$J(\eta, v) := \beta^T \eta(b) + \frac{1}{2} \left[\eta^T(a) \, \hat{\Gamma}_a \, \eta(a) + \eta^T(b) \, \hat{\Gamma}_b \, \eta(b) \right] + \frac{1}{2} \int_a^b \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix}^T \nabla_{(x,u)} \hat{H}(t) \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix} \Delta t \tag{4.47}$$

subject to $\eta \in W^{1,2}[a, b]_{\mathbb{T}}$ and $v \in L^2[a, \rho(b)]_{\mathbb{T}}$ satisfying

$$\left. \begin{aligned} \eta^{\Delta}(t) = \hat{\mathcal{A}}(t) \, \eta(t) + \hat{\mathcal{B}}(t) \, v(t), \quad \hat{N}(t) \, v(t) = 0, \quad t \in [a, \rho(b)]_{\mathbb{T}} \quad \text{a.e.,} \\ \mathcal{M}_a \, \eta(a) = 0, \quad \mathcal{M}_b \, \eta(b) = \hat{M}_b^T (\hat{M}_b \hat{M}_b^T)^{-1} \alpha, \end{aligned} \right\} \tag{4.48}$$

where $\hat{\Gamma}_a$ and $\hat{\Gamma}_b$ are given by (4.10) at the "hat" quantities, $\hat{H}(t)$ is defined in (4.8), and the matrices $\mathcal{M}_a$ and $\mathcal{M}_b$ are given by (4.42) and

$$\mathcal{M}_b := \hat{M}_b^T (\hat{M}_b \hat{M}_b^T)^{-1} \hat{M}_b. \tag{4.49}$$

From assumption (H3) we know that the set of admissible directions (η, v) for problem (4.47)–(4.48) is not empty, and it is closed and convex. Also as noted in part (v) of Remark 4.2, the controllability assumption in (H3) yields the surjectivity of the constraints given by (4.48). Hence, by the coercivity assumption (H4), it follows from [34], (Thm. 1.1) that problem (4.47) has a unique solution $(\bar{\eta}, \bar{v}) \in W^{1,2} \times L^2$. Moreover, by [56], (Cor. 3.11) the weak Pontryagin maximum principle (Prop. 2.10) with $\lambda_0 = 1$ holds for (4.47), that is, there exist functions $\bar{q} : [a, b]_{\mathbb{T}} \to \mathbb{R}^n$ with $\bar{q} \in W^{1,2}$ and $\bar{\lambda} : [a, \rho(b)]_{\mathbb{T}} \to \mathbb{R}^k$ with $\bar{\lambda} \in L^2$, and vectors $\bar{\gamma}_a \in \mathbb{R}^n$ and $\bar{\gamma}_b \in \mathbb{R}^n$ such that

$$-\bar{q}^{\Delta}(t) = \hat{\mathcal{A}}^T(t) \, \bar{q}^{\sigma}(t) + \hat{H}_{xx}(t) \, \bar{\eta}(t) + \hat{H}_{xu}(t) \, \bar{v}(t), \quad t \in [a, \rho(b)]_{\mathbb{T}} \text{ a.e.,} \tag{4.50}$$

$$\hat{\mathcal{B}}^T(t) \, \bar{q}^{\sigma}(t) + \hat{N}^T(t) \, \bar{\lambda}(t) + \hat{H}_{ux}(t) \, \bar{\eta}(t) + \hat{H}_{uu}(t) \, \bar{v}(t), \quad t \in [a, \rho(b)]_{\mathbb{T}} \text{ a.e.,} \tag{4.51}$$

$$-\bar{q}(a) = \hat{\Gamma}_a \, \bar{\eta}(a) + \mathcal{M}_a \, \bar{\gamma}_a, \quad \bar{q}(b) = \hat{\Gamma}_b \, \bar{\eta}(b) + \mathcal{M}_b \, \bar{\gamma}_b + \beta. \tag{4.52}$$

In turn, this implies that the quadruple $(\bar{\eta}, \bar{q}, \bar{\gamma}_a, \bar{\gamma}_b)$ satisfies the vector form of the linear system (4.39) with (4.40), that is,

$$\bar{\eta}^{\Delta}(t) = \mathbb{A}(t) \, \bar{\eta}(t) - \mathbb{B}(t) \, \bar{q}(t), \quad \bar{q}^{\Delta}(t) = \mathbb{C}(t) \, \bar{\eta}(t) - \mathbb{D}(t) \, \bar{q}(t), \quad t \in [a, \rho(b)]_{\mathbb{T}} \quad \text{a.e.,} \tag{4.53}$$

together with (4.52) and

$$\mathcal{M}_a \, \bar{\eta}(a) = 0, \quad \mathcal{M}_b \, \bar{\eta}(b) = \hat{M}_b^T (\hat{M}_b \hat{M}_b^T)^{-1} \alpha. \tag{4.54}$$

Moreover, the function $\bar{v}(t)$ is given by [56], (Formula (3.45)) by (suppressing the argument t)

$$\bar{v} = \hat{Y} (\hat{Y}^T \hat{S} \hat{Y})^{-1} \hat{Y}^T \left[\hat{\mathcal{B}}^T (I + \mu \hat{\mathcal{A}}^T)^{-1} (\hat{H}_{xx} \, \bar{\eta} - \bar{q}) - \hat{H}_{ux} \, \bar{\eta} \right], \quad \text{a.e. on } [a, \rho(b)]_{\mathbb{T}}. \tag{4.55}$$

Since the functions $\bar{\eta}(t)$ and $\bar{q}(t)$ are continuous on $[a, b]_{\mathbb{T}}$ and the coefficients in (4.55) belong to C_{prd}, it follows that $\bar{v} \in \mathrm{C}_{\text{prd}}[a, \rho(b)]_{\mathbb{T}}$ as well. Given that $(\bar{\eta}, \bar{v})$ solves system (4.48) and $\bar{v} \in \mathrm{C}_{\text{prd}}[a, \rho(b)]_{\mathbb{T}}$, we obtain that $\bar{\eta} \in \mathrm{C}^1_{\text{prd}}[a, b]_{\mathbb{T}}$. In fact, from (4.53) and the piecewise rd-continuity of $\mathbb{A}(t)$, $\mathbb{B}(t)$, $\mathbb{C}(t)$, $\mathbb{D}(t)$ in (4.41) it follows that $(\bar{\eta}, \bar{q}) \in \mathrm{C}^1_{\text{prd}}[a, b]_{\mathbb{T}}$. Now we define the vectors $c \in \mathbb{R}^n$ and $d \in \mathbb{R}^{r_b}$ by

$$c := \bar{\eta}(a) + \mathcal{M}_a \, \bar{\gamma}_a, \quad d := (\hat{M}_b \hat{M}_b^T)^{-1} \hat{M}_b \, \bar{\gamma}_b. \tag{4.56}$$

Then by using $(I-\mathcal{M}_a)\,\hat{M}_a=0$ and $\mathcal{M}_a\bar{\gamma}_a=\mathcal{M}_a c$ we obtain

$$\bar{\eta}(a)=(I-\mathcal{M}_a)\,\bar{\eta}(a)=(I-\mathcal{M}_a)\,c,$$
$$-\bar{q}(a)=\hat{\Gamma}_a\,\bar{\eta}(a)+\mathcal{M}_a\,\bar{\gamma}_a=\hat{\Gamma}_a\,(I-\mathcal{M}_a)\,c+\mathcal{M}_a\,c.$$

Therefore, we get by the uniqueness of solutions of the initial value problem (4.39)–(4.40) that $(\bar{\eta}(t),\bar{q}(t))=(Z(t)\,c,\Lambda(t)\,c)$ on $[a,b]_{\mathbb{T}}$. Given that $\bar{\eta}(t)$ also satisfies the second condition in (4.54) and $\mathcal{M}_b$ is defined in (4.49), it results that the vectors c and d from (4.56) satisfy

$$\hat{M}_b\,Z(b)\,c=\hat{M}_b\,\bar{\eta}(b)=\alpha,$$
$$[\Lambda(b)-\hat{\Gamma}_b\,Z(b)]\,c-\hat{M}_b^T d=\bar{q}(b)-\hat{\Gamma}_b\,\bar{\eta}(b)-\mathcal{M}_b\,\bar{\gamma}_b\overset{(4.52)}{=}\beta.$$

This shows that the pair (c,d) solves the linear system (4.46) and hence, the matrix T_b defined in (4.45) is invertible. By the implicit function theorem applied to system (4.43), *i.e.*, to equation $Q(s,\gamma,\omega)=0$ with $Q(s,\gamma,\omega)$ given in (4.44), there exists $\varepsilon_8\in(0,\varepsilon_7)$ and a pair of C^1 functions $(s(\omega),\gamma_b(\omega))$ from $B_{\varepsilon_8}(\hat{\omega})$ to $B_{\varepsilon_7}(\hat{\omega})\times B_{\varepsilon_7}(\hat{\gamma}_b)$ such that $Q(s(\omega),\gamma_b(\omega),\omega)=0$ on $B_{\varepsilon_8}(\hat{\omega})$.

Set now $\varepsilon:=\varepsilon_8$. For $\omega\in B_\varepsilon(\hat{\omega})$ we now define the functions

$$x(t,\omega):=x(t,s(\omega),\omega),\quad p(t,\omega):=p(t,s(\omega),\omega),\quad t\in[a,b]_{\mathbb{T}},$$

where $x(t,s,\omega)$ and $p(t,s,\omega)$ are from Step 3, and the functions

$$u(t,\omega):=u(t,x(t,\omega),p(t,\omega),\omega),\quad \lambda(t,\omega):=\lambda(t,x(t,\omega),p(t,\omega),\omega),\quad t\in[a,\rho(b)]_{\mathbb{T}},$$

where $u(t,x,p,\omega)$ and $\lambda(t,x,p,\omega)$ are from Step 1, the function

$$\gamma_a(\omega):=\gamma_a(s(\omega),\omega),$$

where $\gamma_a(s,\omega)$ is from Step 2, and the function $\gamma_b(\omega)$ from this Step 4. Then the functions $x(t,\omega)$, $u(t,\omega)$, $p(t,\omega)$, $\lambda(t,\omega)$, $\gamma_a(\omega)$, and $\gamma_b(\omega)$ are C^1 in ω and satisfy for $\omega\in B_\varepsilon(\hat{\omega})$ the following relations. By (4.33) we have

$$\varphi_a\big(s(\omega)-\hat{M}_a^T(\gamma_a(\omega)-\hat{\gamma}_a),\omega\big)=0, \tag{4.57}$$

while (4.36) yields

$$\left.\begin{aligned} x(a,\omega)&=s(\omega)-\hat{M}_a^T(\gamma_a(\omega)-\hat{\gamma}_a),\\ -p(a,\omega)&=\nabla_x K_a^T\big(s(\omega)-\hat{M}_a^T(\gamma_a(\omega)-\hat{\gamma}_a),\omega\big)\\ &\quad+\nabla_x\varphi_a^T\big(s(\omega)-\hat{M}_a^T(\gamma_a(\omega)-\hat{\gamma}_a),\omega\big)\,\gamma_a(\omega),\end{aligned}\right\} \tag{4.58}$$

In conclusion, by (4.35) we obtain

$$\left.\begin{aligned} x^\Delta(t,\omega)&=f\big(t,x(t,\omega),u(t,\omega),\omega\big),\\ -p^\Delta(t,\omega)&=H_x^T\big(t,x(t,\omega),u(t,\omega),p^\sigma(t,\omega),\lambda(t,\omega),\omega\big),\end{aligned}\right\}\quad t\in[a,\rho(b)]_{\mathbb{T}}, \tag{4.59}$$

and by (4.37) we get

$$\left.\begin{aligned} H_u^T\big(t,x(t,\omega),u(t,\omega),p^\sigma(t,\omega),\lambda(t,\omega),\omega\big)&=0,\\ \psi\big(t,u(t,\omega),\omega\big)&=0,\end{aligned}\right\}\quad t\in[a,\rho(b)]_{\mathbb{T}}. \tag{4.60}$$

By combining (4.57) with (4.58) we obtain

$$\left.\begin{aligned} \varphi_a\big(x(a,\omega),\omega\big)&=0,\\ -p(a,\omega)&=\nabla_x K_a^T\big(x(a,\omega),\omega\big)+\nabla_x\varphi_a^T\big(x(a,\omega),\omega\big)\,\gamma_a(\omega),\end{aligned}\right\} \tag{4.61}$$

while by (4.43) we have

$$\left.\begin{aligned} \varphi_b\big(x(b,\omega),\omega\big) &= 0, \\ p(b,\omega) = \nabla_x\, K_b^T\big(x(b,\omega),\omega\big) &+ \nabla_x\, \varphi_b^T\big(x(b,\omega),\omega\big)\,\gamma_b(\omega). \end{aligned}\right\} \tag{4.62}$$

For each $\omega \in B_\varepsilon(\hat{\omega})$ consider the problem (C_ω). Then, assumptions (H1)–(H4) and parts (i)– (iii) and (vi) of Remark 4.2 are still satisfied for $\omega \in B_\varepsilon(\hat{\omega})$. It then follows from equations (4.59)–(4.62) that the functions $x(t,\omega)$, $u(t,\omega)$, $p(t,\omega)$, $\lambda(t,\omega)$ and vectors $\gamma_a(\omega)$, $\gamma_b(\omega)$ satisfy for $\omega \in B_\varepsilon(\hat{\omega})$ the assumptions of Theorem 3.3 for problem (C_ω). Therefore, by this theorem, for all $\omega \in B_\varepsilon(\hat{\omega})$ the pair $(x(\cdot,\omega), u(\cdot,\omega)$ is a strict weak local minimum for problem (C_ω). The proof is complete. □

5. Conclusions

In this paper we developed a sufficiency criterion for the weak local minimum in the nonlinear optimal control problem on time scales (C). Our setting includes piecewise rd-continuous control functions, arbitrary state-endpoints constraints, and pointwise equality control constraints. We employed a direct approach *via* the coercivity of the second variation. The presented criterion and the method of proof extends the corresponding result known for the calculus of variations on time scales in [28], (Thm. 2).

We also applied our new sufficiency criterion to obtain a sensitivity result for a separable endpoints problem (C_ω), which is a perturbation of the problem (C) by a parameter present in all the data. By assuming the sufficient conditions to be satisfied at the base problem ($\mathrm{C}_{\hat{\omega}}$), for ω near $\hat{\omega}$ we constructed a feasible solution for (C_ω) and a set of multipliers that are continuously differentiable in the parameter ω and which satisfy the sufficiency theorem developed in the first part of the paper. In this way the strict weak local optimality of the feasible pair for the perturbed problem (C_ω) is deduced. The approach employed is a modification and a generalization to the variable endpoints setting and to time scales of the "shooting method" previously used, for instance in [37, 42, 43], in the continuous time case with fixed initial state constraints. The way we parametrize the initial condition in Step 2 of the proof of Theorem 4.3 can be utilized to extend the results in [37, 42, 43] to the separable endpoints. Furthermore, unlike those references, we do not rely on the Riccati equation to construct a solution to the boundary value problem, but instead we use the coercivity condition used in the sufficiency theorem.

Note that the equality control constraints in (1.2) do not encompass the situation, when the constraints are of the type $u(t) \in U$ for all $t \in [a, \rho(b)]_{\mathbb{T}}$, where $U \subseteq \mathbb{R}^m$ is a given set. The extension of the results of this paper for this and more general constraints is under investigation.

A. Proofs of auxiliary lemmas

In this section we present the proofs of the two crucial approximation lemmas from Section 3.

Proof of Lemma 3.4. Let $\varepsilon > 0$ be fixed and let a feasible pair $(\hat{x}, \hat{u})$ satisfy (A1). If (x, u) is also a feasible pair, then we use the notation $(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})$ introduced in (3.5) for the difference of (x, u) and $(\hat{x}, \hat{u})$. Let $\Phi(t)$ be the fundamental matrix of the linear system in (2.7) or (2.14). Then $\Phi(t)$ is invertible on $[a, b]_{\mathbb{T}}$, by Remark 2.8. In this proof we abbreviate the inverse of Φ^σ by $\Phi^{\sigma-1}$ and the norm $\|\cdot\|_{\mathrm{C_{prd}}}$ by $\|\cdot\|$.

Step 1. Define the function $F(t, x, u, w) := f(t, x, u) - w$, so that $\nabla_{(x,u,w)} F = (f_x, f_u, -I)$. We apply the Taylor theorem to the function $\varphi(\cdot,\cdot)$ near $(\hat{x}(a), \hat{x}(b))$ and to the functions $F(t,\cdot,\cdot,\cdot)$ and $\psi(t,\cdot)$ near $(\hat{x}(t), \hat{u}(t), \hat{x}^\Delta(t))$ and $\hat{u}(t)$, respectively, and use the compactness property of the time scale $[a, \rho(b)]_{\mathbb{T}}$. Then there exists $\delta_1 \in (0, \varepsilon_1)$, where ε_1 is from assumption (A1), such that for any feasible pair (x, u) with $|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))| < \delta_1$ and

$\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u}, \mathrm{d}\hat{x}^{\Delta})\| < \delta_1$ we have

$$\varphi(x(a), x(b)) = \varphi(\hat{x}(a), \hat{x}(b)) + M \begin{pmatrix} \mathrm{d}\hat{x}(a) \\ \mathrm{d}\hat{x}(b) \end{pmatrix} + \gamma, \tag{A.1}$$

$$F(t, x(t), u(t), x^{\Delta}(t)) = F(t, \hat{x}(t), \hat{u}(t), \hat{x}^{\Delta}(t)) + \big(\mathcal{A}(t),\ \mathcal{B}(t),\ -I\big) \begin{pmatrix} \mathrm{d}\hat{x}(t) \\ \mathrm{d}\hat{u}(t) \\ \mathrm{d}\hat{x}^{\Delta}(t) \end{pmatrix} + g(t), \tag{A.2}$$

$$\psi(t, u(t)) = \psi(t, \hat{u}(t) + N(t)\, \mathrm{d}\hat{u}(t) + \pi(t), \tag{A.3}$$

where $t \in [a, \rho(b)]_{\mathbb{T}}$ and where γ, $g(t)$, $\pi(t)$ are the Taylor remainders in (A.1)–(A.3), *i.e.*,

$$\gamma := \mathcal{R}_{\varphi}(x(a), x(b)), \quad g(t) := \mathcal{R}_F(t, x(t), u(t), x^{\Delta}(t)), \quad \pi(t) := \mathcal{R}_{\psi}(t, u(t)). \tag{A.4}$$

The Taylor remainders in (A.4) moreover satisfy the growth estimates

$$\left.\begin{aligned} |\gamma| \le \frac{\varepsilon}{2K_1} \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|, \quad |g(t)| \le \frac{\varepsilon}{8K_2^2} \big|(\mathrm{d}\hat{x}(t), \mathrm{d}\hat{u}(t), \mathrm{d}\hat{x}^{\Delta}(t))\big|, \\ |\pi(t)| \le \frac{\varepsilon}{4K_3} \big|\mathrm{d}\hat{u}(t)\big|, \end{aligned}\right\} \tag{A.5}$$

where K_1, K_2, K_3 are positive constants defined by

$$\begin{aligned} &K_1 := |\mathcal{Z}^{-1}|\, \|\Phi\| \times C \\ &K_2 := \max\Big\{D,\ |M|^2\, |\mathcal{Z}^{-1}|\, \|Y\|^2\, \|\mathcal{B}\|\, \|\Phi\|^2\, \|\Phi^{\sigma-1}\|^2,\ \|\Phi\|\, \|\Phi^{\sigma-1}\|\, \big(1 + |M|\, |\mathcal{Z}^{-1}|\, \|\Phi\| \times E\big)\Big\}, \\ &K_3 := |M|\, |\mathcal{Z}^{-1}|\, \|\mathcal{B}\|\, \|\Phi\|^2\, \|\Phi^{\sigma-1}\|\, \|N\|\, \|(NN^T)^{-1}\| \times C, \end{aligned}$$

and where C, D, E are given by

$$\left.\begin{aligned} &C := \max\big\{\, |M|\, \|Y\|^2\, \|\mathcal{B}\|\, \|\Phi^{\sigma-1}\|,\ E\,\big\}, \\ &D := 1 + \|\mathcal{A}\| + \|\mathcal{B}\|, \\ &E := \varepsilon_0^2\, |\mathcal{D}| + (b-a)\, |M|\, \|Y\|^2\, \|\mathcal{B}\|^2\, \|\Phi\|\, \|\Phi^{\sigma-1}\|^2. \end{aligned}\right\} \tag{A.6}$$

Recall that the matrix $Y(t)$ is chosen according to (2.6), the matrices $\mathcal{D}$ and $\mathcal{Z}$ and the number ε_0 are defined in (2.10) and (2.9) in Proposition 2.9, and the matrices $\mathcal{A}(t)$, $\mathcal{B}(t)$, M, $N(t)$ are defined in (2.1) and (2.2). From assumption (A1) and Lemma 2.5 it results that the norms $\|\cdot\| := \|\cdot\|_{\mathrm{C_{prd}}}$ occurring in the constants K_1, K_2, and K_3 are in fact finite. Observe also that the feasibility of (x, u) and equations (A.1) and (A.2) imply that on $[a, \rho(b)]_{\mathbb{T}}$ we have

$$\gamma = -M \begin{pmatrix} \mathrm{d}\hat{x}(a) \\ \mathrm{d}\hat{x}(b) \end{pmatrix}, \quad g(t) = \mathrm{d}\hat{x}^{\Delta}(t) - \mathcal{A}(t)\, \mathrm{d}\hat{x}(t) - \mathcal{B}(t)\, \mathrm{d}\hat{u}(t), \quad \pi(t) = -N(t)\, \mathrm{d}\hat{u}(t), \tag{A.7}$$

which yields that the functions $g(\cdot)$ and $\pi(\cdot)$ are piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$.

Step 2. We show that without loss of generality the term $\mathrm{d}\hat{x}^{\Delta}(t)$ in the second inequality in (A.5) can be deleted, that is, there exists $\delta_2 \in (0, \delta_1)$ such that for any feasible pair (x, u) with $|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))| < \delta_2$ and $\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\| < \delta_2$ we have on $[a, \rho(b)]_{\mathbb{T}}$

$$\left.\begin{aligned} |\gamma| \le \frac{\varepsilon}{2K_1} \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|, \quad |g(t)| \le \frac{\varepsilon}{4K_2} \big|(\mathrm{d}\hat{x}(t), \mathrm{d}\hat{u}(t))\big|, \\ |\pi(t)| \le \frac{\varepsilon}{4K_3} \big|(\mathrm{d}\hat{x}(t), \mathrm{d}\hat{u}(t))\big|. \end{aligned}\right\} \tag{A.8}$$

By the continuity of $f(t,\cdot,\cdot)$ at $(\hat{x},\hat{u})$ uniformly in t on $[a,\rho(b)]_{\mathbb{T}}$ there exists $\delta_3 \in (0,\delta_1/4)$ such that for any feasible pair (x,u) with $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\| < \delta_3$ and any $t \in [a,\rho(b)]_{\mathbb{T}}$ we have $|\mathrm{d}\hat{x}^{\Delta}(t)| = \big|f(t,x(t),u(t)) - f(t,\hat{x}(t),\hat{u}(t))\big| < \delta_1/2$. Hence, for such an (x,u) with $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\| < \delta_3$ we get

$$\big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t),\mathrm{d}\hat{x}^{\Delta}(t))\big| \le \big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))\big| + |\mathrm{d}\hat{x}^{\Delta}(t)| < \delta_3 + \delta_1/2 < \delta_1, \quad t \in [a,\rho(b)]_{\mathbb{T}}.$$

Therefore, $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u},\mathrm{d}\hat{x}^{\Delta})\| < \delta_1$ holds, which in turn implies that (A.5) is satisfied as well. In particular, the first inequality in (A.8) is established. Now, using assumption (A1) we can apply Lemma 2.7(i) to $\nabla_{(x,u)}f$ at $(\hat{x},\hat{u})$ to obtain $\delta \in (0,\delta_3)$ such that for any pair (x,u) with $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\| < \delta$ we have the inequality

$$\big|\nabla_{(x,u)}f(t,x(t),u(t))\big| \le 1 + \big\|\nabla_{(x,u)}f(\cdot,\hat{x}(\cdot),\hat{u}(\cdot))\big\| = 1 + \|(\mathcal{A},\mathcal{B})\| \le D, \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{A.9}$$

where D is defined in (A.6). Let (x,u) be any feasible pair with $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\| < \delta$. Then (A.9) and the generalized mean value theorem imply

$$|\mathrm{d}\hat{x}^{\Delta}(t)| = \big|f(t,x(t),u(t)) - f(t,\hat{x}(t),\hat{u}(t))\big| \le D\,|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))|, \quad t \in [a,\rho(b)]_{\mathbb{T}}. \tag{A.10}$$

Thus, using $\delta < \delta_3$, (A.5), (A.10), and $K_2 \ge D \ge 1$ (see the definition of K_2 above) we obtain

$$\begin{aligned} |g(t)| &\overset{(A.5)}{\le} \frac{\varepsilon}{8K_2^2}\,\big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))\big| + \frac{\varepsilon}{8K_2^2}\,\big|(\mathrm{d}\hat{x}^{\Delta}(t))\big| \overset{(A.10)}{\le} \frac{\varepsilon\,(1+D)}{8K_2^2}\,\big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))\big| \\ &\le \frac{2\varepsilon K_2}{8K_2^2}\,\big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))\big| \le \frac{\varepsilon}{4K_2}\,\big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))\big|, \quad t \in [a,\rho(b)]_{\mathbb{T}}. \end{aligned}$$

This shows the second inequality in (A.8). Moreover, by (A.5) we also get

$$|\pi(t)| \le \frac{\varepsilon}{4K_3}\,|\mathrm{d}\hat{u}(t)| \le \frac{\varepsilon}{4K_3}\,\big|(\mathrm{d}\hat{x}(t),\mathrm{d}\hat{u}(t))\big|, \quad t \in [a,\rho(b)]_{\mathbb{T}},$$

showing the third estimate in (A.8). The proof of Step 2 is finished.

Step 3. Consider now the nonhomogeneous linear system

$$\tilde{\eta}^{\Delta} = \mathcal{A}(t)\,\tilde{\eta} + \mathcal{B}(t)\,\tilde{v} + g(t), \quad N(t)\,\tilde{v}(t) = -\pi(t), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \quad M\begin{pmatrix}\tilde{\eta}(a)\\ \tilde{\eta}(b)\end{pmatrix} = -\gamma. \tag{A.11}$$

We define the auxiliary piecewise rd-continuous function $\nu : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^m$ and the vectors $\beta \in \mathbb{R}^r$ and $\tilde{\alpha} \in \mathbb{R}^n$ by

$$\nu(t) := -N^T(t)\,[N(t)\,N^T(t)]^{-1}\pi(t), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{A.12}$$

$$\beta := M_b\,\Phi(b)\int_a^b \Phi^{\sigma-1}(s)\,[\,\mathcal{B}(s)\,\nu(s) + g(s)\,]\,\Delta s, \tag{A.13}$$

$$\tilde{\alpha} := -\varepsilon_0^2\,\mathcal{D}^T\mathcal{Z}^{-1}(\gamma+\beta). \tag{A.14}$$

Moreover, with the matrix $\mathcal{E}(t)$ given in (2.10) we define the functions $\tilde{v} : [a,\rho(b)]_{\mathbb{T}} \to \mathbb{R}^m$ and $\tilde{\eta} : [a,b]_{\mathbb{T}} \to \mathbb{R}^n$ by

$$\tilde{v}(t) := -Y(t)\,\mathcal{E}^T(t)\,\mathcal{Z}^{-1}(\gamma+\beta) + \nu(t), \quad t \in [a,\rho(b)]_{\mathbb{T}}, \tag{A.15}$$

$$\tilde{\eta}(t) := \Phi(t)\,\tilde{\alpha} + \Phi(t)\int_a^t \Phi^{\sigma-1}(s)\,[\,\mathcal{B}(s)\,\tilde{v}(s) + g(s)\,]\,\Delta s, \quad t \in [a,b]_{\mathbb{T}}. \tag{A.16}$$

Then $\tilde{v}(\cdot)$ belongs to $\mathrm{C}_{\mathrm{prd}}[a,\rho(b)]_{\mathbb{T}}$ and $\tilde{\eta}(\cdot)$ belongs to $\mathrm{C}^1_{\mathrm{prd}}[a,b]_{\mathbb{T}}$, and the pair $(\tilde{\eta},\tilde{v})$ satisfies (A.11). Indeed, by using $\Phi^{\Delta}(t) = \mathcal{A}(t)\,\Phi(t)$ and the product rule $(fg)^{\Delta} = f^{\Delta}g + f^{\sigma}g^{\Delta}$ it is straightforward to verify that the Δ-derivative of the right-hand side of (A.16) is equal to $\mathcal{A}(t)\,\tilde{\eta}(t) + \mathcal{B}(t)\,\tilde{v}(t) + g(t)$, *i.e.*, the first equality in (A.11) holds. Next, we have

$$N(t)\,\tilde{v}(t) = -N(t)\,Y(t)\,\mathcal{E}^T(t)\,\mathcal{Z}^{-1}(\gamma+\beta) - \pi(t) \overset{(2.6)}{=} -\pi(t),$$

showing the second equality in (A.11). Finally, by using $\tilde{\eta}(a) = \tilde{\alpha}$ and evaluating (A.16) at $t = b$ together with the definition of $\mathcal{D}$ and $\mathcal{E}(t)$ in (2.10) we obtain

$$\begin{aligned} M\begin{pmatrix}\tilde{\eta}(a)\\ \tilde{\eta}(b)\end{pmatrix} &\overset{(A.16)}{=} M_a\,\tilde{\alpha} + M_b\left(\Phi(b)\,\tilde{\alpha} + \Phi(b)\int_a^b [\Phi^\sigma(s)]^{-1}\,[\,\mathcal{B}(s)\,\tilde{v}(s) + g(s)\,]\,\Delta s\right) \\ &\overset{(A.15),\,(A.14)}{=} -\varepsilon_0^2\,\mathcal{D}\mathcal{D}^T\mathcal{Z}^{-1}(\gamma+\beta) - \int_a^b \mathcal{E}(s)\,\mathcal{E}^T(s)\,\mathcal{Z}^{-1}(\gamma+\beta)\,\Delta s\, + \beta \\ &\overset{(2.9)}{=} -\mathcal{Z}\mathcal{Z}^{-1}(\gamma+\beta) + \beta = -\gamma. \end{aligned}$$

Therefore, the third equality in (A.11) is also satisfied, which completes the proof of Step 3.

Step 4. We claim that the pair $(\tilde{\eta}, \tilde{v})$ defined in (A.15)–(A.16) satisfies the estimate

$$\big|(\tilde{\eta}(t), \tilde{v}(t))\big| \le \varepsilon\left\{\big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big| + \int_a^b \big|(\mathrm{d}\hat{x}(s), \mathrm{d}\hat{u}(s))\big|\,\Delta s\right\}, \quad t \in [a, \rho(b)]_{\mathbb{T}}. \tag{A.17}$$

In addition, inequality (A.17) with $|\tilde{\eta}(b)|$ on the left–hand side also holds when b is left-scattered. For the proof of (A.17) we use the estimates in (A.8) with the properly chosen constants K_1, K_2, K_3 from Step 1. From (2.10) and (A.12)–(A.14) we get

$$\left.\begin{aligned} &\|\mathcal{E}\| \le |M|\,\|Y\|\,\|\mathcal{B}\|\,\|\Phi\|\,\|\Phi^{\sigma-1}\|, \quad |\nu(t)| \le \|N\|\,\|(NN^T)^{-1}\|\,|\pi(t)|, \quad t \in [a, \rho(b)]_{\mathbb{T}}, \\ &|\beta| \le |M|\,\|\Phi\|\,\|\Phi^{\sigma-1}\| \int_a^b \big(\,\|\mathcal{B}\|\,|\nu(s)| + |g(s)|\,\big)\,\Delta s, \quad |\tilde{\alpha}| \le \varepsilon_0^2\,|\mathcal{D}|\,|\mathcal{Z}^{-1}|\,\big(\,|\gamma| + |\beta|\,\big). \end{aligned}\right\} \tag{A.18}$$

Then it follows from (A.15) and the estimates in (A.18) that

$$|\tilde{v}(t)| \le |\mathcal{Z}^{-1}|\,\|Y\|\,\|\mathcal{E}\|\,\big(\,|\gamma| + |\beta|\,\big) + |\nu(t)| \tag{A.19}$$

$$\le K_1\,|\gamma| + K_2 \int_a^b |g(s)|\,\Delta s\, + K_3 \int_a^b |\pi(s)|\,\Delta s, \quad t \in [a, \rho(b)]_{\mathbb{T}}, \tag{A.20}$$

while from (A.16), (A.18), and (A.19) we obtain

$$\begin{aligned} |\tilde{\eta}(t)| &\le \|\Phi\|\,|\tilde{\alpha}| + \|\Phi\|\,\|\Phi^{\sigma-1}\|\left(\|\mathcal{B}\| \int_a^b |\tilde{v}(s)|\,\Delta s\, + \int_a^b |g(s)|\,\Delta s\right) \\ &\le K_1\,|\gamma| + K_2 \int_a^b |g(s)|\,\Delta s\, + K_3 \int_a^b |\pi(s)|\,\Delta s, \quad t \in [a, b]_{\mathbb{T}}. \end{aligned} \tag{A.21}$$

Therefore, combining (A.20) and (A.21) with (A.8) we get

$$\begin{aligned} \big|(\tilde{\eta}(t), \tilde{v}(t))\big| &\le |\tilde{\eta}(t)| + |\tilde{v}(t)| \le 2K_1\,|\gamma| + 2K_2 \int_a^b |g(s)|\,\Delta s\, + 2K_3 \int_a^b |\pi(s)|\,\Delta s \\ &\overset{(A.8)}{\le} \varepsilon\,\big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big| + \varepsilon \int_a^b \big|(\mathrm{d}\hat{x}(s), \mathrm{d}\hat{u}(s))\big|\,\Delta s \end{aligned} \tag{A.22}$$

for all $t \in [a, \rho(b)]_{\mathbb{T}}$, as we claim in (A.17). In addition, if b is left-scattered, then it follows from (A.21) at $t = b$ and (A.8) that the inequality in (A.22) with $|\tilde{\eta}(b)|$ on the left-hand side is also satisfied. The proof of Step 4 is complete.

Step 5. Define the pair (η, v) by $\eta(t) := \mathrm{d}\hat{x}(t) - \tilde{\eta}(t)$ on $[a, b]_{\mathbb{T}}$ and $v(t) := \mathrm{d}\hat{u}(t) - \tilde{v}(t)$ on $[a, \rho(b)]_{\mathbb{T}}$. Then $\eta \in \mathrm{C}^1_{\mathrm{prd}}[a, b]_{\mathbb{T}}$, $v \in \mathrm{C}_{\mathrm{prd}}[a, \rho(b)]_{\mathbb{T}}$, and by (A.7) and (A.11) we get

$$M \begin{pmatrix} \eta(a) \\ \eta(b) \end{pmatrix} = M \begin{pmatrix} \mathrm{d}\hat{x}(a) \\ \mathrm{d}\hat{x}(b) \end{pmatrix} - M \begin{pmatrix} \tilde{\eta}(a) \\ \tilde{\eta}(b) \end{pmatrix} = -\gamma + \gamma = 0.$$

Furthermore, by (A.7) and (A.11) we also have for $t \in [a, \rho(b)]_{\mathbb{T}}$

$$\eta^{\Delta}(t) = \mathrm{d}\hat{x}^{\Delta}(t) - \tilde{\eta}^{\Delta}(t) = \mathcal{A}(t)\,[\mathrm{d}\hat{x}(t) - \tilde{\eta}(t)] + \mathcal{B}(t)\,[\mathrm{d}\hat{u}(t) - \tilde{v}(t)] = \mathcal{A}(t)\,\eta(t) + \mathcal{B}(t)\,v(t),$$

and similarly

$$N(t)\,v(t) = N(t)\,\mathrm{d}\hat{u}(t) - N(t)\,\tilde{v}(t) = -\pi(t) + \pi(t) = 0.$$

This shows that (η, v) is admissible. Finally, since $(\mathrm{d}\hat{x} - \eta, \mathrm{d}\hat{u} - v) = (\tilde{\eta}, \tilde{v})$, the estimate in (3.4) is a direct consequence of (A.17). The proof of this lemma is complete. $\square$

Proof of Lemma 3.6. In this proof we will again utilize the notation in (3.5). Fix $\varepsilon > 0$. By Lemma 3.4 there exists $\delta > 0$ such that for every feasible pair (x, u) with $\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\|_{\mathrm{C}_{\mathrm{prd}}} < \delta$ there exists an admissible pair (η, v) satisfying (3.4). Take $\delta_0 := \delta$, let (x, u) be feasible with $\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\|_{\mathrm{C}_{\mathrm{prd}}} < \delta_0$, and let (η, v) be the associated admissible pair from Lemma 3.4. Define $(\tilde{\eta}, \tilde{v}) := (\mathrm{d}\hat{x} - \eta, \mathrm{d}\hat{u} - v)$, so that $(\mathrm{d}\hat{x}, \mathrm{d}\hat{u}) = (\eta, v) + (\tilde{\eta}, \tilde{v})$. Then

$$\begin{aligned} J''(\mathrm{d}\hat{x}, \mathrm{d}\hat{u}) &= J''(\eta + \tilde{\eta}, v + \tilde{v}) \\ &= J''(\eta, v) + J''(\tilde{\eta}, \tilde{v}) + \begin{pmatrix} \eta(a) \\ \eta(b) \end{pmatrix}^T \Gamma \begin{pmatrix} \tilde{\eta}(a) \\ \tilde{\eta}(b) \end{pmatrix} + \int_a^b \begin{pmatrix} \eta(t) \\ v(t) \end{pmatrix}^T \nabla_{(x,u)} \hat{H}(t) \begin{pmatrix} \tilde{\eta}(t) \\ \tilde{v}(t) \end{pmatrix} \Delta t, \end{aligned}$$

where Γ and $\nabla_{(x,u)} \hat{H}(t)$ are given in (2.17) and (2.16). Since $\nabla_{(x,u)} \hat{H}(t)$ is piecewise rd-continuous on $[a, \rho(b)]_{\mathbb{T}}$, there exists $C > 0$ such that $\|\nabla_{(x,u)} \hat{H}(t)\|_{\mathrm{C}_{\mathrm{prd}}} \leq C$ and $|\Gamma| \leq C$. Using the coercivity assumption on the term $J''(\eta, v)$ we then obtain

$$\begin{aligned} 2\,J''(\mathrm{d}\hat{x}, \mathrm{d}\hat{u}) \geq \alpha \left\{ |\eta(a)|^2 + |\eta(b)|^2 + \int_a^b \left(|\eta(t)|^2 + |v(t)|^2 \right) \Delta t \right\} - C \left\{ |\tilde{\eta}(a)|^2 + |\tilde{\eta}(b)|^2 \right\} \\ - C \int_a^b \left(|\tilde{\eta}(t)|^2 + |\tilde{v}(t)|^2 \right) \Delta t - 2\,C \left\{ |\eta(a)|\,|\tilde{\eta}(a)| + |\eta(b)|\,|\tilde{\eta}(b)| \right\} \\ - 2\,C \int_a^b \left(|\eta(t)|\,|\tilde{\eta}(t)| + |v(t)|\,|\tilde{v}(t)| \right) \Delta t. \end{aligned} \tag{A.23}$$

We will find a lower bound for the first term in (A.23) and upper bounds for all the others in terms of the quantities $\big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 = |\mathrm{d}\hat{x}(a)|^2 + |\mathrm{d}\hat{x}(b)|^2$ and

$$\big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 = \int_a^b \left(|\mathrm{d}\hat{x}(t)|^2 + |\mathrm{d}\hat{u}(t)|^2 \right) \Delta t.$$

In the following calculations we will also use the expression $\big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^1} = \int_a^b \big|(\mathrm{d}\hat{x}(t), \mathrm{d}\hat{u}(t))\big|\,\Delta t$ appearing in (3.4), as well as the estimate

$$\big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^1}^2 \leq (b - a)\,\big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2, \tag{A.24}$$

which follows from the Cauchy–Schwarz inequality. We also recall that

$$\eta = \mathrm{d}\hat{x} - \tilde{\eta}, \quad v = \mathrm{d}\hat{u} - \tilde{v}, \quad \eta^{\Delta} = \mathrm{d}\hat{x}^{\Delta} - \tilde{\eta}^{\Delta}. \tag{A.25}$$

By the arithmetic-geometric mean inequality ($2xy \le x^2 + y^2$) and (A.24), we get from (3.4) for all $t \in [a, \rho(b)]_{\mathbb{T}}$

$$|\tilde{\eta}(t)|^2 \le |\tilde{\eta}(t)|^2 + |\tilde{v}(t)|^2 \le 2\varepsilon^2 \Big\{ \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 \Big\}. \tag{A.26}$$

The first inequality in (A.26) holds of course also for $t = b$. Inequality (A.26) implies that

$$|\tilde{\eta}(a)|^2 + |\tilde{\eta}(b)|^2 \le 4\varepsilon^2 \Big\{ \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 \Big\} \tag{A.27}$$

and

$$\int_a^b \big(|\tilde{\eta}(t)|^2 + |\tilde{v}(t)|^2 \big)\, \Delta t \le 2\varepsilon^2 (b-a) \Big\{ \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 \Big\}. \tag{A.28}$$

Since by (A.25) we have $|\eta(t)| \le |\mathrm{d}\hat{x}(t)| + |\tilde{\eta}(t)|$ on $[a, b]_{\mathbb{T}}$ and $|v(t)| \le |\mathrm{d}\hat{u}(t)| + |\tilde{v}(t)|$ on $[a, \rho(b)]_{\mathbb{T}}$, it follows by the arithmetic-geometric mean inequality and (A.24) that for all $t \in [a, b]_{\mathbb{T}}$

$$|\eta(t)|\, |\tilde{\eta}(t)| \le \varepsilon\, |\mathrm{d}\hat{x}(t)|^2 + (\varepsilon + 2\varepsilon^2) \Big\{ \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 \Big\}, \tag{A.29}$$

and for all $t \in [a, \rho(b)]_{\mathbb{T}}$

$$|v(t)|\, |\tilde{v}(t)| \le \varepsilon\, |\mathrm{d}\hat{u}(t)|^2 + (\varepsilon + 2\varepsilon^2) \Big\{ \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 \Big\}. \tag{A.30}$$

Therefore, by taking $t = a$ and $t = b$ in (A.29) we obtain

$$|\eta(a)|\, |\tilde{\eta}(a)| + |\eta(b)|\, |\tilde{\eta}(b)| \le (3\varepsilon + 4\varepsilon^2) \Big\{ \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2 \Big\}, \tag{A.31}$$

and integrating the sum of (A.29) and (A.30) over $[a, b]_{\mathbb{T}}$ we get

$$\begin{aligned} \int_a^b \big(|\eta(t)|\, |\tilde{\eta}(t)| + |v(t)|\, |\tilde{v}(t)| \big)\, \Delta t &\le (2\varepsilon + 4\varepsilon^2)(b-a)\, \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 \\ &\quad + [\varepsilon + 2\varepsilon(b-a) + 4\varepsilon^2 (b-a)^2]\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2. \end{aligned} \tag{A.32}$$

Now by (A.25) we have $|\mathrm{d}\hat{x}(t)| \le |\eta(t)| + |\tilde{\eta}(t)|$ on $[a, b]_{\mathbb{T}}$ and $|\mathrm{d}\hat{u}(t)| \le |v(t)| + |\tilde{v}(t)|$ on $[a, \rho(b)]_{\mathbb{T}}$. Then, by taking the square on both sides we obtain $|\mathrm{d}\hat{x}(t)|^2 \le 2\, \big(|\eta(t)|^2 + |\tilde{\eta}(t)|^2 \big)$ on $[a, b]_{\mathbb{T}}$ and $|\mathrm{d}\hat{u}(t)|^2 \le 2\, \big(|v(t)|^2 + 2\, |\tilde{v}(t)|^2 \big)$ on $[a, \rho(b)]_{\mathbb{T}}$. This implies that for $t \in [a, \rho(b)]_{\mathbb{T}}$ we have

$$2\, \big(|\eta(t)|^2 + |v(t)|^2 \big) \ge |\mathrm{d}\hat{x}(t)|^2 + |\mathrm{d}\hat{u}(t)|^2 - 2\, \big(|\tilde{\eta}(t)|^2 + |\tilde{v}(t)|^2 \big).$$

By integrating over $[a, b]_{\mathbb{T}}$ and using (A.28) we then obtain

$$\int_a^b \big(|\eta(t)|^2 + |v(t)|^2 \big)\, \Delta t \ge -2\varepsilon^2 (b-a)\, \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 + [\tfrac{1}{2} - 2\varepsilon^2 (b-a)^2]\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2. \tag{A.33}$$

In a similar way we get by (A.27) that

$$\begin{aligned} |\eta(a)|^2 + |\eta(b)|^2 &\ge \tfrac{1}{2}\, \big(|\mathrm{d}\hat{x}(a)|^2 + |\mathrm{d}\hat{x}(b)|^2 \big) - \big(|\tilde{\eta}(a)|^2 + |\tilde{\eta}(b)|^2 \big) \\ &\ge (\tfrac{1}{2} - 4\varepsilon^2)\, \big|(\mathrm{d}\hat{x}(a), \mathrm{d}\hat{x}(b))\big|^2 - 4\varepsilon^2 (b-a)\, \big\|(\mathrm{d}\hat{x}, \mathrm{d}\hat{u})\big\|_{L^2}^2. \end{aligned} \tag{A.34}$$

Now we insert the estimates in (A.34), (A.33), (A.27), (A.28), (A.31), and (A.32) into inequality (A.23) and collect the terms with the same powers of ε to get

$$2\,J''(\mathrm{d}\hat{x},\mathrm{d}\hat{u}) \geq (\alpha/2 - D\varepsilon - E\varepsilon^2)\,\big|(\mathrm{d}\hat{x}(a),\mathrm{d}\hat{x}(b))\big|^2 + (\alpha/2 - F\varepsilon - G\varepsilon^2)\,\big\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\big\|_{L^2}^2,$$

where

$$D := 2C\,[3 + 2(b-a)], \qquad E := 2\,[6C + (\alpha + 5C)\,(b-a)]$$
$$F := 2C\,[1 + 5(b-a)], \qquad G := 2\,(b-a)\,[2\alpha + 6C + (\alpha+5C)\,(b-a)].$$

Choose $\varepsilon > 0$ small enough so that $\max\{D\varepsilon, E\varepsilon^2, F\varepsilon, G\varepsilon^2\} < \alpha/8$. Then with this choice of ε we have by the beginning of this proof (*i.e.*, by Lem. 3.4) the corresponding $\delta_0 := \delta$ such that for every feasible pair (x,u) with $\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\|_{\mathrm{C_{prd}}} < \delta_0$ the inequality

$$\begin{aligned} 2\,J''(\mathrm{d}\hat{x},\mathrm{d}\hat{u}) &\geq (\alpha/2 - \alpha/8 - \alpha/8)\,\big|(\mathrm{d}\hat{x}(a),\mathrm{d}\hat{x}(b))\big|^2 + (\alpha/2 - \alpha/8 - \alpha/8)\,\big\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\big\|_{L^2}^2 \\ &= (\alpha/4)\,\Big\{\big|(\mathrm{d}\hat{x}(a),\mathrm{d}\hat{x}(b))\big|^2 + \big\|(\mathrm{d}\hat{x},\mathrm{d}\hat{u})\big\|_{L^2}^2\Big\}. \end{aligned}$$

The proof of this lemma is complete. □

Acknowledgements. The second author wishes to thank the Department of Mathematics and Statistics of the Faculty of Science, Masaryk University, for hospitality while conducting the first part of this research project. The authors are also grateful to anonymous referees for their insightful comments and suggestions, which helped improve the organization and presentation of the results.

References

[1] R.P. Agarwal, V. Otero–Espinar and K. Perera, Basic properties of Sobolev's spaces on time scales. *Adv. Differ. Equ.* **2006** (2006) 38121.

[2] M.S. Aronna, J.F. Bonnans, A.V. Dmitruk and P.A. Lotito, Quadratic order conditions for bang-singular extremals. *Numer. Algebra Control Optim.* **2** (2012) 511–546.

[3] F.M. Atici, D.C. Biles and A. Lebedinsky, A utility maximisation problem on multiple time scales. *Int. J. Dyn. Syst. Differ. Equ.* **3** (2011) 38–47.

[4] B. Aulbach and L. Neidhart, Integration on measure chains. In: "New Progress in Difference Equations", *Proc. of the Sixth International Conference on Difference Equations (Augsburg, 2001)*. Edited by B. Aulbach, S. Elaydi, and G. Ladas. Chapman & Hall/CRC, Boca Raton, FL (2004) 239–252.

[5] Z. Bartosiewicz and D.F.M. Torres, Noether's theorem on time scales. *J. Math. Anal. Appl.* **342** (2008) 1220–1226.

[6] M. Bohner, Calculus of variations on time scales. *Dynam. Sys. Appl.* **13** (2004) 339–349.

[7] M. Bohner, M. Fan and J. Zhang, Existence of periodic solutions in predator-prey and competition dynamic systems. *Nonlinear Anal. Real World Appl.* **7** (2006) 1193–1204.

[8] M. Bohner, M. Fan and J. Zhang, Periodicity of scalar dynamic equations and applications to population models. *J. Math. Anal. Appl.* **330** (2007) 1–9.

[9] M. Bohner and A. Peterson, Dynamic Equations on Time Scales. An Introduction with Applications. Birkhäuser, Boston (2001).

[10] M. Bohner, A. Peterson and editors, Advances in Dynamic Equations on Time Scales. Birkhäuser, Boston (2003).

[11] M. Bohner and N. Wintz, The linear quadratic regulator on time scales. *Int. J. Differ. Equ.* **5** (2010) 149–174.

[12] M. Bohner and N. Wintz, The linear quadratic tracker on time scales. *Int. J. Dyn. Syst. Differ. Equ.* **3** (2011) 423–447.

[13] M. Bohner and N. Wintz, The Kalman filter for linear systems on time scales. *J. Math. Anal. Appl.* **46** (2013) 419–436.

[14] V.G. Boltyanskii, Sufficient conditions for optimality and the justification of the dynamic programming method. *SIAM J. Control Optim.* **4** (1966) 326–361.

[15] J.F. Bonnans, X. Dupuis and L. Pfeiffer, Second-order necessary conditions in Pontryagin form for optimal control problems. *SIAM J. Control Optim.* **52** (2014) 3887–3916.

[16] L. Bourdin and E. Trélat, Pontryagin maximum principle for finite dimensional nonlinear optimal control problems on time scales. *SIAM J. Control Optim.* **51** (2013) 3781–3813.

[17] L. Bourdin and E. Trélat, Optimal sampled-data control, and generalizations on time scales. *Math. Control. Relat. Fields* **6** (2016) 53–94.

[18] J.V. Breakwell and Y.C. Ho, On the conjugate point condition for the control problem. *Internat. J. Engrg. Sci.* **2** (1964/1965) 565–579.

[19] J.V. Breakwell, J.L. Speyer and A.E. Bryson, Optimization and control of nonlinear systems using the second variation. *SIAM J. Control Optim.* **1** (1963) 193–223.

[20] P.P. Cai, J.L. Fu and Y.X. Guo, Noether symmetries of the nonconservative and nonholonomic systems on time scales. *Science China – Physics Mechanics & Astronomy* **56** (2013) 1017–1028.

[21] A.L. Donchev and W.W. Hager, Lipschitzian stability in nonlinear control and optimization. *SIAM J. Control Optim.* **31** (1993) 569–603.

[22] A.L. Donchev, W.W. Hager, A.B. Poore and B. Yang, Optimality, stability, and convergence in nonlinear control. *Appl. Math. Optim.* **31** (1995) 297–326.

[23] A.V. Fiacco, Introduction to Sensitivity and Stability Analysis in Nonlinear Programming. Vol. 165 of *Mathematics in Science and Engineering*. Academic Press, Orlando, FL (1983).

[24] G.S. Guseinov, Integration on time scales. *J. Math. Anal. Appl.* **285** (2003) 107–127.

[25] S. Hilger, Analysis on measure chains – a unified approach to continuous and discrete calculus. *Results Math.* **18** (1990) 18–56.

[26] R. Hilscher, W. Kratz and V. Zeidan, Differentiation of solutions of dynamic equations on time scales with respect to parameters. *Adv. Dyn. Syst. Appl.* **4** (2009) 35–54.

[27] R. Hilscher and V. Zeidan, Second order sufficiency criteria for a discrete optimal control problem. *J. Differ. Equ. Appl.* **8** (2002) 573–602.

[28] R. Hilscher and V. Zeidan, Calculus of variations on time scales: weak local piecewise $\mathrm{C}^1_{\mathrm{rd}}$ solutions with variable endpoints. *J. Math. Anal. Appl.* **289** (2004) 143–166.

[29] R. Hilscher and V. Zeidan, Legendre, Jacobi, and Riccati type conditions for time scale variational problem with application. *Dynam. Systems Appl.* **16** (2007) 451–480.

[30] R. Hilscher and V. Zeidan, Time scale embedding theorem and coercivity of quadratic functionals. *Analysis (Munich)* **28** (2008) 1–28.

[31] R. Hilscher and V. Zeidan, Weak maximum principle and accessory problem for control problems on time scales. *Nonlinear Anal.* **70** (2009) 3209–3226.

[32] V. Kac and P. Cheung, Quantum Calculus. Springer-Verlag, New York, 2002.

[33] R.E. Kalman, The theory of optimal control and the calculus of variations. In *Symposium on Mathematical Optimization Techniques (Santa Monica, CA, 1960)*. Univ. California Press, Berkeley, CA. Mathematical Optimization Techniques. Edited by R. Bellman (1963) 309–331.

[34] J.-L. Lions, Optimal control of systems governed by partial differential equations. Translated from the French by S.K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin (1971).

[35] D.G. Luenberger, Linear and Nonlinear Programming. 2nd ed. Addison-Wesley, Reading, MA, 1984.

[36] K. Malanowski, Sensitivity analysis of optimization problems in Hilbert space with application to optimal control. *Appl. Math. Optim.* **21** (1990) 1–20.

[37] K. Malanowski and H. Maurer, Sensitivity analysis for parametric control problems with control-state constraints. *Comput. Optim. Appl.* **5** (1996) 253–283.

[38] A.B. Malinowska and D.F.M. Torres, Necessary and sufficient conditions for local Pareto optimality on time scales. *J. Math. Sci. (N.Y.)* **161** (2009) 803–810.

[39] A.B. Malinowska and D.F.M. Torres, Leitmann's direct method of optimization for absolute extrema of certain problems of the calculus of variations on time scales. *Appl. Math. Comput.* **217** (2010) 1158–1162.

[40] A.B. Malinowska and D.F.M. Torres, Euler–Lagrange equations for composition functionals in calculus of variations on time scales. *Discrete Contin. Dyn. Syst. Ser. A* **29** (2011) 577–593.

[41] A.B. Malinowska and D.F.M. Torres, Quantum Variational Calculus. *Springer Briefs in Electrical and Computer Engineering.* Springer, Cham (2014).

[42] H. Maurer and H.J. Pesch, Solution differentiability for nonlinear parametric control problem. *SIAM J. Control Optim.* **32** (1994) 1542–1554.

[43] H. Maurer and H.J. Pesch, Solution differentiability for parametric nonlinear control problems with control-state constraints. *J. Optim. Theory Appl.* **86**, (1995) 285–309.

[44] A.A. Milyutin and N.P. Osmolovskii, Calculus of variations and optimal control. Translated from the Russian manuscript by Dimitrii Chibisov. Translations of Vol. 180 of *Mathematical Monographs.* American Mathematical Society, Providence, RI (1998).

[45] D. Orrell and V. Zeidan, Another Jacobi sufficiency criterion for optimal control with smooth constraints. *J. Optim. Theory Appl.* **58** (1988) 283–300.

[46] N.P. Osmolovskii, Sufficient quadratic conditions of extremum for discontinuous controls in optimal control problems with mixed constraints. *J. Math. Sci.* **173** (2011) 1–106.

[47] L. Poggiolini and G. Stefani, Bang-singular-bang extremals: sufficient optimality conditions. *J. Dyn. Control Syst.* **17** (2011) 469–514.

[48] B.P. Rynne, L^2 spaces and boundary value problems on time-scales. *J. Math. Anal. Appl.* **328** (2007) 1217–1236.

[49] J.L. Speyer and D.H. Jacobson, Primer on Optimal Control Theory. In Vol. 20 of *Advances in Design and Control.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2010).

[50] P. Šepitka and R. Šimon Hilscher, Principal solutions at infinity for time scale symplectic systems without controllability condition. *J. Math. Anal. Appl.* **444** (2016) 852–880.

[51] R. Šimon Hilscher, A note on the time scale calculus of variations problems. In Vol. 14 of *Ulmer Seminare über Funktionalanalysis und Differentialgleichungen*. University of Ulm, Ulm (2009) 223–230.

[52] R. Šimon Hilscher and V. Zeidan, Symplectic structure of Jacobi systems on time scales. *Int. J. Differ. Equ.* **5** (2010) 55–81.

[53] R. Šimon Hilscher and V. Zeidan, First order conditions for generalized variational problems over time scales. *Comput. Math. Appl.* **62** (2011) 3490–3503.

[54] R. Šimon Hilscher and V. Zeidan, Hamilton–Jacobi theory over time scales and applications to linear-quadratic problems. *Nonlinear Anal.* **75** (2012) 932–950.

[55] V. Zeidan, Continuous versus discrete nonlinear optimal control problems. In: *Proc. of the 14th International Conference on Difference Equations and Applications (Istanbul, 2008)*. Edited by M. Bohner, Z. Došlá, G. Ladas, M. Ünal, and A. Zafer. Uğur-Bahçeşehir University Publishing Company, Istanbul (2009) 73–93.

[56] V. Zeidan, Constrained linear-quadratic control problems on time scales and weak normality. In Vol. 26 of Dynamic Systems and Applications (2017) 627–662.

[57] V. Zeidan and P. Zezza, The conjugate point condition for smooth control sets. *J. Math. Anal. Appl.* **132** (1988) 572–589.