Discrete Optimal Control: The Accessory Problem and Necessary Optimality Conditions

Roman Hilscher¹

Department of Mathematics, Faculty of Science, Masaryk University Brno, Janackovo nam. 2a, CZ-66295 Brno, Czech Republic
E-mail: houska@math.muni.cz

and

Vera Zeidan

Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027
E-mail: zeidan@math.msu.edu

Submitted by L. Berkovitz
Received February 9, 1999

In this work we derive the accessory problem for a general nonlinear discrete optimal control problem. We obtain necessary conditions for optimality—nonnegativity of the second variation. As the second variation is a discrete quadratic functional, we proceed by deriving necessary conditions for the nonnegativity of such functionals. A certain controllability (or normality) assumption is required.

Key Words: discrete maximum principle; discrete Hamiltonian system; discrete quadratic functional; accessory problem; optimality conditions; Jacobi system; conjugate interval; controllability.

1. INTRODUCTION

Let $n, m, N \in \mathbb{N}$ be given with $N \geq 2$. By the interval $[a, b]$ we always mean the interval of integers $(a, a + 1, \ldots, b - 1, b)$. Thus, denote $J := [0, N]$ and $J^* := [0, N + 1]$. By $I$, resp. 0, we denote the identity matrix, resp. the zero matrix or column (will be clear from the context), of the

¹ Research partially supported by the Department of Mathematics, Michigan State University, East Lansing, and partially by Grants GA 143 (FR VŠ), 201/98/0677 (GA ČR), Prague.
corresponding dimension. Consider the following discrete optimal control problem

\[
\begin{align*}
\minimize & \quad J(x, u) := K(x_{N+1}) + \sum_{k=0}^{N} g(k, x_{k+1}, u_k), \\
\Delta x_k &= f(k, x_{k+1}, u_k), \quad k \in J, \\
x_0 = a, \quad \varphi(x_{N+1}) = 0.
\end{align*}
\]

(P)

Here \(g, f, K, \varphi\) are given functions,

\[
\begin{align*}
x &\colon J^s \to \mathbb{R}^n, & g &\colon J \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, & \varphi &\colon \mathbb{R}^n \to \mathbb{R}^r, & r &\leq n, \\
u &\colon J \to \mathbb{R}^m, & f &\colon J \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, & K &\colon \mathbb{R}^n \to \mathbb{R},
\end{align*}
\]

and \(a \in \mathbb{R}^n\) is a given vector. We always consider the gradient \(\nabla F\) of a real valued function \(F\) to be the row of the corresponding size.

The difference equation in (P) is called the equation of motion and the pair \((x, u)\) satisfying this equation on \(J\) and the boundary conditions \(x_0 = a, \varphi(x_{N+1}) = 0\) is called feasible (on \(J^s\)). A feasible pair \((\hat{x}, \hat{u})\) is called a weak local minimum for (P) if for some \(\epsilon > 0\), \((\hat{x}, \hat{u})\) minimizes \(J(x, u)\) over all feasible pairs \((x, u)\) satisfying \(|x_k - \hat{x}_k| < \epsilon, k \in J^s\), and \(|u_k - \hat{u}_k| < \epsilon, k \in J\), where \(|\cdot|\) is any norm in \(\mathbb{R}^r\), \(s = n, m\).

One direction of this paper is to derive the accessory problem as well as the theory of conjugate points for general discrete optimal control problem (P) with the control set \(U := \mathbb{R}^m\). This subject evokes the linear quadratic problem and its linear self-adjoint Hamiltonian system

\[
\begin{align*}
\Delta \eta_k &= A_k \eta_{k+1} + B_k q_k, \\
\Delta q_k &= C_k \eta_{k+1} - A_k^T q_k, \quad k \in J.
\end{align*}
\]

(H)

As noted in [1, Section 3], in order to obtain a self-adjoint Hamiltonian system for the linear quadratic problem, one needs to consider shifting the state variable with respect to control; see also [2] for discrete variational problems. Generalizing this idea to a nonlinear setting, we have proposed the above-stated general discrete optimal control problem. Note that the dynamics considered is equivalent to the ones used by Lewis and Syrmos [15, Chapter 2] or by Boltyanskii [11, Chapter V]

\[
x_{k+1} = d(k, x_k, u_k), \quad k \in J,
\]

with \(d_k\) invertible. This invertibility condition is inevitable in order to obtain a Hamiltonian system out of the results in [15, p. 32]. Hence, the treatment in this paper of the nonlinear problem incorporates all the other forms.
Throughout the paper we will use the following assumptions:

\[
\begin{align*}
&\text{For all } k \in J, \, g(k, \cdot, \cdot) \text{ and } f(k, \cdot, \cdot) \text{ are continuously differentiable;} \\
&\text{for all } k \in J, \, I - f_s(k, \cdot, \cdot) \text{ is invertible; and } K, \varphi \text{ are continuously differentiable.} \\
&\text{(A1)}
\end{align*}
\]

\[
\begin{align*}
&\text{For all } k \in J, \, g(k, \cdot, \cdot) \text{ and } f(k, \cdot, \cdot) \text{ are twice continuously differentiable;} \\
&\text{for all } k \in J, \, I - f_s(k, \cdot, \cdot) \text{ is invertible; and } K, \varphi \text{ are twice continuously differentiable.} \\
&\text{(A2)}
\end{align*}
\]

For the special case of the discrete calculus of variations, that is, when the function \( f(k, x, u) = u \) on its domain, the second variation of the corresponding problem (P) was derived in [13] and the accessory problem was obtained. This latter is also a discrete calculus of variations problem but with quadratic functional and with minimum value equal to zero. However, the derivation of the accessory problem for the optimal control problem (P) is to our knowledge an open question.

Intensive studies have been conducted for the special case when the data \( K \) and \( g \) are quadratic and the function \( f \) is linear in \((x, u)\). In this case, it is known that a self-adjoint Hamiltonian system (H) is naturally associated to (P). In [6], by means of the \textit{generalized zeros} concept, Bohner defined the notion of \textit{disconjugate} Hamiltonian system. This notion alone characterizes the positivity of the quadratic functional with fixed endpoints state constraints (see [6, Theorem 2]). When one or more endpoints vary, it is shown in the same paper that disconjugacy alone is not enough but an extra condition is needed for this case to characterize the positivity of the quadratic functional. However, in the continuous-time setting, it is known (see [16, Theorem 2.1]) that by extending the definition of “generalized zeros to the variable final endpoint case, the positivity of the quadratic functional is completely characterized by this concept of “conjugate points.”

In [17, Theorem 6.1], it is shown that in this continuous-time setting, the nonexistence of such conjugate points is in fact necessary for the optimality in the nonlinear optimal control problem. This result was obtained by defining the concept of conjugate point for the associated accessory problem and by showing that the existence of such points contradicts that the minimum value of the accessory problem is zero. However, in the discrete-time case and when the endpoints of the state are not both fixed, a notion parallel to such concept of disconjugacy was never introduced, and second-order necessary conditions for optimality in the nonlinear problem (P) were never developed. This latter question is the same as
deriving necessary conditions for the nonnegativity of a quadratic functional. Such type of results that are known to be fundamental for the continuous-time optimal control theory (see, e.g., [12, 18]), would play an analog role if they could be derived for the discrete setting.

The purpose of this work is to positively answer the above-raised questions. The paper is divided as follows. In Section 2 we recall the first-order weak optimality condition obtained through the Pontryagin maximum principle, and we introduce the notion of $M$-controllability. In Section 3 we characterize the set of feasible directions which produce feasible families for the problem (P). The characterization in Section 3 is crucial for the derivation of the accessory problem obtained in Section 4. In Section 5 we derive several necessary conditions for optimality in (P) by means of the development of necessary conditions for the nonnegativity of a quadratic functional that alludes to the accessory problem. The first result is in terms of the positivity of the partial quadratic sum (see Theorem 4). Then, we introduce the notion of conjugate interval to $N + 1$ and we show that the nonexistence of such intervals in $[1, N + 1]$ is necessary for the nonnegativity of the quadratic functional (Theorem 5). A nother necessary condition is given in Theorem 6 and it is in terms of a conjoined basis with appropriate endpoint and rank conditions. We also show via a numerical example why a definition parallel to the one known in the continuous-time setting would be insufficient in our discrete-time setting. The result given by Theorem 7 presents necessary conditions expressed in terms of the Riccati equation with appropriate endpoint and rank conditions. The last necessary condition is presented in Theorem 8 and evokes the projection of a tridiagonal matrix on the controllability matrix. This result is closely related to the discrete form of the Legendre necessary condition (see Corollary 2).

2. PRELIMINARY RESULTS

In this section we state the weak version of the Pontryagin maximum principle corresponding to the nonlinear discrete optimal control problem (P). This result is obtained as a direct translation to our setting of theorem 43.1 of [11, p. 342]. Such translation requires a special care of the notation. We also introduce in this section the notion of $M$-controllability.

Remark 1. The assumption that $I - f_s(k, x, u)$ be invertible for all $k \in J$ corresponds to the solvability of the equation of motion with respect to $x_{k+1}$. Thus, denote this solution by $w: J \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, i.e.,

$$\Delta x_k = f(k, x_{k+1}, u_k), \quad k \in J, \quad \text{iff} \quad x_{k+1} = w(k, x_k, u_k), \quad k \in J. \quad (1)$$
If (A1) holds, the function $w$ has gradients

$$w_a(k, x_k, u_k) = \left[ I - f_a(k, x_{k+1}, u_k) \right]^{-1}, \quad k \in J,$$

$$w_o(k, x_k, u_k) = \left[ I - f_o(k, x_{k+1}, u_k) \right]^{-1} f_o(k, x_{k+1}, u_k), \quad k \in J.$$  \hspace{1cm} (2)

**Theorem 1** (discrete weak Pontryagin maximum principle). Let (A1) hold. Let $(x, u)$ be any feasible pair on $J^*$, In order that $(x, u)$ be the optimal solution of (P), it is necessary that there exist a real number $\lambda_0 \geq 0$, a vector $\gamma \in \mathbb{R}^s$, and a parameter $p^* : J^* \to \mathbb{R}^n$ satisfying the following conditions:

(i) normality condition: if $\lambda_0 = 0$, then $|\gamma| + \|p\| \neq 0$, where $\|\cdot\|$ is any norm in the parameter space, e.g., $\|p\| := \sum_{k=0}^{n+1} |p_k|^2$, and $|\cdot|$ is the usual norm in $\mathbb{R}^l$, $s = n, r$,

(ii) the adjoint equation: $-\Delta p_k = \lambda_0 g^T_k(k, x_{k+1}, u_k) + f^T_k(k, x_{k+1}, u_k)p_k, \quad k \in J$,

(iii) the stationarity condition: $\lambda_0 g^T_n(k, x_{k+1}, u_k) + f^T_n(k, x_{k+1}, u_k)p_k = 0, \quad k \in J$,

(iv) the transversality condition: $p_{N+1} = M^T \gamma + \lambda_0 \nabla K^T(x_{N+1}),$ where

$$M := \nabla \psi(x_{N+1}).$$ \hspace{1cm} (3)

If $\lambda_0 \neq 0$, then we may take $\lambda_0 = 1$.

**Proof.** We reformulate our problem (P) in terms of the notation of [11, Chapter V]

$$\begin{align*}
\text{minimize} & \quad \mathcal{T}_g(x, v) = \sum_{k=1}^{N+1} h_0^a(k, x_{k-1}, u_k), \\
& \quad x_k = h(k, x_{k-1}, u_k), \quad k \in [1, N+1], \\
& \quad x_k \in \mathcal{M}, \quad k \in J^*,
\end{align*}$$

(P')

where

$$u_k := u_{k-1}, \quad k \in [1, N+1],$$

$$h^a_0(k, x, v) := g(k - 1, w(k - 1, x, v), v), \quad k \in [1, N],$$

$$h^a_0(N+1, x, v) := g(N, w(N, x, v), v) + K(w(N, x, v)), \hspace{1cm} (4)$$

$$h(k, x, v) := w(k - 1, x, v), \quad k \in [1, N+1],$$

$$\mathcal{M} := \text{Ker} W_k, \quad W_k : \mathbb{R}^n \to \mathbb{R}^{r_k}, \quad r_k \leq n, \quad k \in J^*$$

$$W_0(x) = x - a, \quad r_0 := n, \quad W_{N+1}(x) := \psi(x), \quad r_{N+1} := r,$$

$$W_k(x) = 0, \quad r_k := 0, \quad k \in [1, N].$$
All the other spaces and cones (like $U_k, \Omega_k, \Theta_k, \Pi_k, \Sigma_k, \Omega_k, L_k$) appearing in [11, Theorem 43.1] are equal to $\mathbb{R}^n$. Then for the process $(x, v)$ being optimal for $(P')$, it is necessary that there exist a real constant $\bar{\lambda}_0 \leq 0$ and vectors $\psi: [1, N + 1] \to \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $\bar{\gamma} \in \mathbb{R}^r$, such that for the Hamiltonian $H(k, x, v)$ defined by

$$H(k, x, v) := \bar{\lambda}_0 h^0(k, x, v) + \psi^T h(k, x, v)$$

$$= \bar{\lambda}_0 g_s(k - 1, w(k - 1, x, v), v) + \psi^T k w(k - 1, x, v),$$

$k \in [1, N + 1]$, the following conditions are satisfied:

(a) If $\bar{\lambda}_0 = 0$, then either $\psi_k$ for some $k \in [1, N + 1]$, or $\mu$, or $\bar{\gamma}$ is nonzero.

(b_1) For all $k \in [1, N]$, we have $\psi_k = H^T_s(k + 1, x_k, v_{k+1})$.

(b_2) $H^T_s(1, x_0, v_1) + \nabla w^T_0(x_0) \mu = 0$.

(b_3) $\psi_{N + 1} = \nabla w^T_{N + 1}(x_{N + 1}) \bar{\gamma}$.

(c) For all $k \in [1, N + 1]$ we have $H^T_s(k, x_{k-1}, v_k) = 0$.

To interpret the above equations, we calculate the gradients $H_s, H_v$: here $k \in [1, N],$

$$H_s(k, x, v) = \bar{\lambda}_0 h^0_s(k, x, v) + \psi^T h_s(k, x, v)$$

$$= \left[ \bar{\lambda}_0 g_s(k - 1, w(k - 1, x, v), v) + \psi^T k w(k - 1, x, v), \right.$$

$$\left. \times w_s(k - 1, x, v) \right],$$

$$H_s(N + 1, x, v) = \left[ \bar{\lambda}_0 g_s(N, w(N, x, v), v) + \psi^T_{N + 1} \right.$$

$$\left. + \bar{\lambda}_0 \nabla K(w(N, x, v)) \right] w_s(N, x, v),$$

$$H_v(k, x, v) = \bar{\lambda}_0 h^0_v(k, x, v) + \psi^T h_v(k, x, v)$$

$$= \left[ \bar{\lambda}_0 g_v(k - 1, w(k - 1, x, v), v) + \psi^T k w(k - 1, x, v) \right.$$

$$\left. \times w_v(k - 1, x, v) \right.$$

$$\left. + \bar{\lambda}_0 g_v(k - 1, w(k - 1, x, v), v) \right],$$

$$H_v(N + 1, x, v) = \left[ \bar{\lambda}_0 g_v(N, w(N, x, v), v) + \psi^T_{N + 1} \right.$$

$$\left. + \bar{\lambda}_0 \nabla K(w(N, x, v)) \right] w_v(N, x, v)$$

$$+ \bar{\lambda}_0 g_v(N, w(N, x, v), v).$$
Set $p_k := -\psi_k$, $k \in [1, N]$ and $p_{N+1} := -\psi_{N+1} - \bar{\lambda}_0 \nabla K^T(x_{N+1})$, $\lambda_0 := -\bar{\lambda}_0$, $\gamma := -\bar{\gamma}$. Then directly from (b), the equality $p_{N+1} = M^T \gamma + \lambda_0 \nabla K^T(x_{N+1})$ follows. Condition (b) then implies

$$p_k = w_t^T(k, x_k, u_{k+1}) \left[ \lambda_0 g_t^T(k, x_{k+1}, u_{k+1}) + p_{k+1} \right], \quad k \in [1, N].$$

(5)

Define $p_0 \in \mathbb{R}^n$ by the above formula at $k = 0$. Then (b) implies $p_0 = \mu$, so that (a) becomes (i) in the theorem. Moreover, using Eqs. (2) and (4), we obtain

$$-\Delta p_k = \lambda_0 g_t^T(k, x_{k+1}, u_k) + f_t^T(k, x_{k+1}, u_k) p_k, \quad k \in J.$$ 

Finally, to prove (iii), we have for $k \in J$, by Eqs. (2), (4), and (5),

$$0 = -H_t^T(k + 1, x_k, u_{k+1})$$

$$= w_t^T(k, x_k, u_{k+1}) \left[ \lambda_0 g_t^T(k, x_{k+1}, u_{k+1}) + p_{k+1} \right] + \lambda_0 g_t^T(k, x_{k+1}, u_{k+1})$$

$$= w_t^T(k, x_k, u_{k+1}) w_t^{-1}(k, x_k, u_{k+1}) p_k + \lambda_0 g_t^T(k, x_{k+1}, u_{k+1})$$

$$= \lambda_0 g_t^T(k, x_{k+1}, u_k) + f_t^T(k, x_{k+1}, u_k) p_k.$$ 

The proof is complete.

Remark 2. The Pontryagin Hamiltonian $\mathcal{H}$ corresponding to problem (P) is defined to be

$$\mathcal{H}(k, x, u, p, \lambda_0) := p^T f(k, x, u) + \lambda_0 g(k, x, u), \quad k \in J.$$ 

Conditions (ii) and (iii) of the above theorem, and the equation of motion, can be formulated in terms of $\mathcal{H}$ respectively as

$$-\Delta p_k = \mathcal{H}_k^T(k, x_{k+1}, u, p_k, \lambda_0), \quad \mathcal{H}_u^T(k, x_{k+1}, u, p_k, \lambda_0) = 0,$$

$$\Delta x_k = \mathcal{H}_p^T(k, x_{k+1}, u, p_k, \lambda_0),$$

for $k \in J$, where $\mathcal{H}_k, \mathcal{H}_u, \mathcal{H}_p$ denote the gradients of $\mathcal{H}$ with respect to the corresponding variable.

Let a feasible pair $(\bar{x}, \bar{u})$ be given for (P). Set

$$\mathcal{A}_k := f_u(k, \bar{x}_{k+1}, \bar{u}_k), \quad \mathcal{B}_k := f_u(k, \bar{x}_{k+1}, \bar{u}_k), \quad k \in J,$$

(6)

and consider the variational system

$$\Delta \eta_k = \mathcal{A}_k \eta_{k+1} + \mathcal{B}_k u_k, \quad k \in J,$$

(7)

$$\eta_0 = 0, \quad M \eta_{N+1} = 0,$$

(8)
where \( \eta : J^* \to \mathbb{R}^n, \upsilon : J \to \mathbb{R}^m \), and \( M \) is defined by Eq. (3). A pair \( (\eta, \upsilon) \) satisfying Eq. (7) is called \textit{admissible on} \( J^* \).

For the purpose of the following definitions, the matrices \( \mathcal{A}_k, \mathcal{B}_k, k \in J \), resp. \( M \), are arbitrary real \( n \times n \), \( n \times m \)-matrices, resp. an \( r \times n \)-matrix.

**Definition 1 (\( M \)-controllability).** The pair \( (\mathcal{A}, \mathcal{B}) \) is called \textit{\( M \)-controllable on} \( J^* \) if \( M \) is of full rank and, for every \( b \in \mathbb{R}^n, c \in \mathbb{R}^r \), there exists an admissible pair \( (\eta, \upsilon) \) satisfying the boundary conditions

\[
\eta_0 = b, \quad M\eta_{N+1} = c. \tag{9}
\]

**Definition 2** (strong normality). The pair \( (\mathcal{A}, \mathcal{B}) \) is called \textit{strongly normal on} \( J^* \) if \( M \) is of full rank and, for some \( \gamma \in \mathbb{R}^r \), the system

\[ -\Delta p_k = \mathcal{A}_k^T p_k, \quad \mathcal{B}_k^T p_k = 0, \quad k \in J, \quad p_{N+1} = M^T \gamma, \]

possesses only the zero solution \( p_k \equiv 0 \) on \( J^* \).

**Remark 3.** It can be proven that \( (\mathcal{A}, \mathcal{B}) \) is strongly normal on \( J^* \) iff it is \( M \)-controllable on \( J^* \).

### 3. Feasible Directions and Feasible Families

In this section we show that each pair \( (\eta, \upsilon) \) satisfying the system (7) and (8) leads to a family of feasible pairs \( (x(e), u(e)) \). Such a pair \( (\eta, \upsilon) \) is called a \textit{feasible direction}.

**Theorem 2.** Let assumption (A2) hold and suppose that the variational system (7) is \( M \)-controllable on \( J^* \). Assume that \( (\eta, \upsilon) \) solves the variational system (7), (8), with \( \mathcal{A}_k, \mathcal{B}_k \) given by Eq. (6) for some feasible pair \( (\bar{x}, \bar{u}) \).

Then there exist \( \delta > 0 \) and functions \( x(e), u(e) \),

\[ x : J^* \times (-\delta, \delta) \to \mathbb{R}^n, \quad u : J \times (-\delta, \delta) \to \mathbb{R}^m, \]

which are twice continuously differentiable in \( e \), such that

(i) \( \Delta x_k(e) = f(k, x_{k+1}(e), u_k(e)), k \in J, \ e \in (-\delta, \delta), \)

(ii) \( x_0(e) = a, \varphi(x_{N+1}(e)) = 0, \ e \in (-\delta, \delta), \)

(iii) \( x_k(0) = \bar{x}_k \) and \( x'_k(0) = \eta_k, k \in J^* \),

(iv) \( u_k(0) = \bar{u}_k \) and \( u'_k(0) = v_k, k \in J \),

where \( \dot{x}_k \) and \( \dot{u}_k \) denote the derivatives of \( x_k \) and \( u_k \) with respect to \( e \).

**Proof.** The proof is similar to that for [17, Theorem 4.1]. Let \( \xi^0, \ldots, \xi^r \) be the vertices of an \((r + 1)\)-simplex in \( \mathbb{R}^r \) that contains 0 in its interior,
that is, 0 can be uniquely expressed as their convex combination, i.e., there exist unique $\beta^0, \ldots, \beta^r \in \mathbb{R}$ such that
\[ \sum_{i=0}^{r} \beta^i = 1 \quad \text{and} \quad \sum_{i=0}^{r} \beta^i \xi^i = 0. \]
By $M$-controllability of $(\mathcal{A}, \mathcal{B})$, for every $i \in [0, r]$ there exists a pair $(\eta^i, \nu^i)$ such that
\[ \Delta \eta^i_k = \mathcal{A}_k \eta^i_{k+1} + \mathcal{B}_k \nu^i_k, \quad k \in J, \]
\[ \eta^i_0 = 0, \quad M \eta^i_{r+1} = \xi^i. \]
Moreover, without loss of generality we may assume that $\sum_{i=0}^{r} \beta^i v^i = v$ [otherwise change $v^i$ to $(1/\beta^i)(v - \sum_{i=0}^{r-1} \beta^i v^i)$, so that for the solution $\eta^i$ of
\[ \Delta \eta^i_k = \mathcal{A}_k \eta^i_{k+1} + \mathcal{B}_k v^i_k, \quad k \in J, \]
\[ \eta^i_0 = 0, \]
we have $M \eta^i_{r+1} = \xi^i$ by linearity]. Define for $\alpha = (\alpha^0, \ldots, \alpha^r)^T \in \mathbb{R}^{r+1}$ the function $u: J \times \mathbb{R}^{r+1} \to \mathbb{R}^m$ by
\[ u_k(\alpha) := \overline{u}_k + \sum_{i=0}^{r} \alpha^i v^i_k, \quad k \in J. \]

Then $\partial u_k / \partial \alpha^i = v^i_k, i \in [0, r], k \in J$. Let the function $x: J^u \times \mathbb{R}^{r+1} \to \mathbb{R}^n$ defined by $(k, \alpha) \mapsto x_k(\alpha)$, where $x(\alpha)$ is the (unique) solution of
\[ \Delta x_k = f(k, x_{k+1}, u_k(\alpha)), \quad k \in J, \]
\[ x_0 = a. \]

Then it follows immediately that $(\partial x_k / \partial \alpha^i)(0)$ satisfies the equality
\[ \Delta \frac{\partial x_k}{\partial \alpha^i}(0) = \mathcal{A}_k \frac{\partial x_{k+1}}{\partial \alpha^i}(0) + \mathcal{B}_k v^i_k, \quad k \in J, \]
\[ \frac{\partial x_0}{\partial \alpha^i}(0) = 0. \]

Therefore, by the uniqueness of solutions of this initial value problem (observe $I - \mathcal{A}_k$ invertible), we obtain
\[ \frac{\partial x_k}{\partial \alpha^i}(0) = \eta^i_k, \quad k \in J^u, \quad i \in [0, r]. \]
We proceed now by defining the function $F$ by

$$F: \mathbb{R}^{r+1} \to \mathbb{R}^{r+1}, \quad F(\alpha) := \left( \sum_{i=0}^{r} \alpha^i \varphi(x_{N+1}(\alpha)) \right).$$

Then $F(0) = (0; \varphi(x_{N+1})) = (0; 0)$, and $F$ is twice continuously differentiable on some neighborhood of 0. Moreover, $(\partial F/\partial \alpha^i)(0) = [1; (M\eta_{N+1}^i)^T] = [1; (\xi^i)^T]$, so that the Jacobian matrix

$$\frac{\partial F}{\partial \alpha}(0) = \begin{pmatrix} 1 & \cdots & 1 \\ \xi^0 & \cdots & \xi^r \end{pmatrix}$$

is of full rank by the construction of $(\xi^i)$’s. By the inverse function theorem, there exists $\delta > 0$ such that $F$ is invertible on the $(r + 1)$-ball with radius $\delta$. Let

$$\alpha: (-\delta, \delta) \to \mathbb{R}^{r+1}, \quad \alpha(\epsilon) := F^{-1}(\epsilon; 0),$$

be the inverse of $F$ on $(-\delta, \delta)$. Then $\alpha$ is twice continuously differentiable in $\epsilon$, $\alpha(0) = 0$, and

$$\sum_{i=0}^{r} \alpha^i(\epsilon) = \epsilon, \quad \varphi(x_{N+1}(\alpha(\epsilon))) = 0, \quad \epsilon \in (-\delta, \delta). \quad (14)$$

Hence, differentiating Eq. (14) at $\epsilon = 0$, we get $\sum_{i=0}^{r} \dot{\alpha}^i(0) = 1$ and

$$0 = \sum_{i=0}^{r} \nabla \varphi(x_{N+1}) \frac{\partial x_{N+1}}{\partial \alpha^i}(0) \dot{\alpha}^i(0) = \sum_{i=0}^{r} M\eta_{N+1}^i \dot{\alpha}^i(0) = \sum_{i=1}^{r} \xi^i \dot{\alpha}^i(0),$$

where we used Eq. (13). The properties of $(\xi^i)$’s hence imply

$$\dot{\alpha}^i(0) = \beta^i, \quad i \in [0, r]. \quad (15)$$

Finally, for $\epsilon \in (-\delta, \delta)$, set

$$u_k(\epsilon) := \bar{u}_k + \sum_{i=0}^{r} \alpha^i(\epsilon) \upsilon^i_k, \quad k \in J, \quad x_k(\epsilon) \text{ the corresponding solution of Eq. (12), } \quad k \in J^e. \quad (16)$$

Then (i) and (ii) follows from the construction of $(x(\epsilon), u(\epsilon))$, i.e., from Eqs. (14) and (12). Clearly, $u_k(0) = \bar{u}_k$, $k \in J$, and hence $x_k(0) = \bar{x}_k$, $k \in J^e$, from the uniqueness theorem. It follows that $u_k(\cdot)$ is twice contin-
ulously differentiable, and hence \( x_k(\cdot) \) is so. We have for all \( k \in J \)
\[
\dot{u}_k(0) = \sum_{i=0}^{r} \frac{\partial u_k}{\partial \alpha} \left( \alpha(\epsilon) \right) \dot{\alpha}(\epsilon)|_{\epsilon=0} = \sum_{i=0}^{r} \beta^i(0) v_k^i = \sum_{i=0}^{r} \beta^i v_k^i = v_k,
\]
\[
\dot{x}_k(0) = \sum_{i=0}^{r} \frac{\partial x_k}{\partial \alpha} \left( \alpha(\epsilon) \right) \dot{\alpha}(\epsilon)|_{\epsilon=0} = \sum_{i=0}^{r} \alpha^i(0) \dot{\alpha}_k^i = \sum_{i=0}^{r} \beta^i \eta_k^i = \eta_k,
\]
where we used Eq. (15). The proof is complete. 

4. THE ACCESSORY PROBLEM

Define the set of feasible directions
\[
\mathcal{D} := \{ (\eta, v) \text{ satisfying Eqs. (7), (8)} \}.
\]

**Theorem 3 (second variation).** (i) Let (A1) hold. Let \((\hat{x}, \hat{u})\) be a weak local minimum for \((P)\) and suppose that in the corresponding variational system (7) the pair \((\mathcal{A}, \mathcal{B})\) is \(M\)-controllable on \(J^*\), where \(M\) is defined by Eq. (3). Then there exists a unique triple \((\lambda_0, \gamma, p)\) of parameters, \(\lambda_0 = 1, \gamma \in \mathbb{R}^r, \) and \(p: J^* \to \mathbb{R}^n, \) such that the conditions (ii)–(iv) of Theorem 1 hold with \((x, u) := (\hat{x}, \hat{u})\).

(ii) If in addition, assumption (A2) is met, then the second variation \(\mathcal{F}_2\) of \(\mathcal{F}\) is nonnegative definite; we write \(\mathcal{F}_2 \geq 0\), i.e.,
\[
\mathcal{F}_2(\eta, v) \geq 0 \quad \text{for all } (\eta, v) \in \mathcal{D},
\]
where \(\mathcal{H}(k, x, u, p, 1) = g(k, x, u) + p^T f(k, x, u), k \in J, \) and
\[
\mathcal{F}_2(\eta, v) := \frac{1}{2} \eta_{k+1}^T \Gamma_{N+1} \eta_{N+1} + \frac{1}{2} \sum_{k=0}^{N} \left( \eta_{k+1}^T v_k^T \begin{bmatrix} \mathcal{H}_{x_k} & \mathcal{H}_{xu} \end{bmatrix} \right) \begin{bmatrix} \eta_{k+1} \nu_k \end{bmatrix},
\]
\[
\Gamma := \nabla^2 K(\hat{x}_{N+1}) + \nabla^2 \psi(\hat{x}_{N+1}) \gamma,
\]
and where the second partial derivatives of \(\mathcal{H}\) are evaluated at \((k, \hat{x}_{k+1}, \hat{u}_k, p_k, 1)\).

**Proof.** Part (i) follows directly from Theorem 1, since Remark 3 and the \(M\)-controllability of \((\mathcal{A}, \mathcal{B})\) imply \(\lambda_0 \neq 0\) and \((1, \gamma, p)\) is unique. Thus, let us prove (ii). Let \((\eta, v)\) be an admissible pair satisfying \(\eta_0 = 0, M\eta_{N+1} = 0, \) i.e., \((\eta, v) \in \mathcal{D}, \) and let \((x(\epsilon), u(\epsilon))\) be the feasible family from Theorem 2 corresponding to \((\eta, v)\). Set
\[
F(\epsilon) := \mathcal{F}(x(\epsilon), u(\epsilon))
\]
\[
= K(x_{N+1}(\epsilon)) + \psi(x_{N+1}(\epsilon)) \gamma
\]
\[
+ \sum_{k=0}^{N} \left\{ \mathcal{H}(k, x_{k+1}(\epsilon), u_{k}(\epsilon), p_k, 1) - p^T f(k, x_{k+1}(\epsilon), u_{k}(\epsilon)) \right\},
\]

where
in fact, \( \varphi^T(x_{N+1}(e))\gamma = 0 \), by Theorem 2(ii). Then \( F(e) \) is twice continuously differentiable and attains its local minimum at \( e = 0 \), so that \( F'(0) = 0 \) and \( F''(0) \geq 0 \). Observe that the equality \( F'(0) = 0 \) comes also from part (i) of this theorem. We have

\[
F'(e) = \dot{x}_{N+1}^T(e) \left\{ \nabla K(x_{N+1}(e)) + \nabla \varphi^T(x_{N+1}(e))\gamma - p_{N+1} \right\} + \sum_{k=0}^{N} \left\{ (\Delta p_k^T + \mathcal{H}_k) \dot{x}_{k+1} + \mathcal{R}_k \dot{u}_k \right\},
\]

\[
F''(e) = \ddot{x}_{N+1}^T(e) \left\{ \nabla K(x_{N+1}(e)) + \nabla \varphi^T(x_{N+1}(e))\gamma - p_{N+1} \right\} + \dot{x}_{N+1}^T(e) \left\{ \nabla^2 K(x_{N+1}(e)) + \nabla^2 \varphi^T(x_{N+1}(e))\gamma \right\} \dot{x}_{N+1}(e) + \sum_{k=0}^{N} \left\{ \left[ \dot{x}_{k+1}^T(e) \mathcal{H}_{xu} + \dot{u}_k^T(e) \mathcal{H}_{uu} \right] \dot{u}_k + \mathcal{R}_k \dot{u}_k \right\},
\]

where the first- and second-order partial derivatives of \( \mathcal{H} \) are evaluated at \((k, x_{k+1}, u_k, p_k, 1)\). Now use (ii)–(iv) of Theorem 2 and Remark 2 to end up at \( e = 0 \) with \( F'(0) = 0 \) and \( F''(0) \geq 0 \).

**Remark 4.** The expression \( F'(\dot{x}, \dot{u}; \eta, v) := F'(0) \) is called the *first variation* of \( F \) at \((\dot{x}, \dot{u})\) along \((v, u)\).

Next, we derive the Jacobi system for the problem (P). This is done by applying Theorem 1 to the functional \( F \) of the second variation—the so-called *accessory problem*. Suppose that assumption (A2) holds and denote

\[
\mathcal{A}_k := f_k(x_k, \dot{x}_{k+1}, \dot{u}_k) \in \mathbb{R}^{n \times n},
\]

\[
\mathcal{B}_k := f_k(x_k, \dot{x}_{k+1}, \dot{u}_k) \in \mathbb{R}^{n \times m},
\]

\[
P_k := \mathcal{H}_{xu}(x_k, \dot{x}_{k+1}, \dot{u}_k, p_k, 1) \in \mathbb{R}^{n \times n},
\]

\[
Q_k := \mathcal{H}_{uu}(x_k, \dot{x}_{k+1}, \dot{u}_k, p_k, 1) \in \mathbb{R}^{n \times m},
\]

\[
R_k := \mathcal{H}_{uu}(x_k, \dot{x}_{k+1}, \dot{u}_k, p_k, 1) \in \mathbb{R}^{m \times n}.
\]

Define the \( n \times n \)-matrices \( A, B, C : J \to \mathbb{R}^{n \times n} \) by

\[
A_k := \mathcal{A}_k - \mathcal{B} R_k^{-1} Q_k^T, \quad B_k := \mathcal{B} R_k^{-1} \mathcal{B}^T, \quad C_k := P_k - Q_k R_k^{-1} Q_k^T,
\]

\( k \in J \). (20)
Observe that $B_k$ and $C_k$ are symmetric. We will always suppose the following:

For all $k \in J$, the matrices $R_k$, $B_k^T B_k$, and $I - A_k$ defined by Eqs. (19), (20) are invertible. \hfill (A3)

**Remark 5.** Under assumption (A3), it is easy to see that $(\mathcal{A}, \mathcal{B})$ is $M$-controllable on $J^*$ iff $(A, B)$ is $M$-controllable on $J^*$.

The Hamiltonian of the accessory problem is taking $q_k$ as

$$
\mathcal{H}_{AP}(k, \eta, v, q) := \frac{1}{2} \eta^T P_k \eta + \eta^T Q_k v + \frac{1}{2} v^T R_k v - q^T (A_k \eta + B_k v).
$$

Then the gradients are

$$
\begin{align*}
\mathcal{H}_{\eta;AP}(k, \eta, v, q, 1) & = \eta^T P_k + v^T Q_k^T - q^T A_k, \\
\mathcal{H}_{v;AP}(k, \eta, v, q, 1) & = \eta^T Q_k + v^T R_k - q^T B_k.
\end{align*}
$$

Then Theorem 1 yields the adjoint equation, the stationarity equation, and the transversality condition in the form

$$
\begin{align*}
\Delta q_k & = P_k \eta_{k+1} + Q_k v_k - \mathcal{A}_k^T q_k, \quad k \in J, \\
Q_k^T \eta_{k+1} + R_k v_k - \mathcal{B}_k^T q_k & = 0, \quad k \in J, \\
q_{N+1} + M^T \gamma + \Gamma \eta_{N+1} & = 0, \quad \gamma \in \mathbb{R}^r.
\end{align*}
$$

If $R_k$ is invertible, we may solve Eq. (22) for $v_k$, i.e.,

$$
v_k = R_k^{-1} (\mathcal{B}_k^T q_k - Q_k^T \eta_{k+1}), \quad k \in J,
$$

and plug into the equation of motion and Eq. (21) to obtain

$$
\begin{align*}
\Delta \eta_k & = (\mathcal{A}_k - \mathcal{B}_k R_k^{-1} Q_k^T) \eta_{k+1} + \mathcal{B}_k R_k^{-1} \mathcal{B}_k^T q_k, \\
\Delta q_k & = (P_k - Q_k R_k^{-1} Q_k^T) \eta_{k+1} - (\mathcal{A}_k^T - Q_k R_k^{-1} \mathcal{B}_k^T) q_k.
\end{align*}
$$

In other words, we obtain the Jacobi system for (P)—the discrete linear Hamiltonian system

$$
\begin{align*}
\Delta \eta_k & = A_k \eta_{k+1} + B_k q_k, \\
\Delta q_k & = C_k \eta_{k+1} - A_k^T q_k, \quad k \in J.
\end{align*}
$$

**Lemma 1** [7, Lemma 1]. **Under assumption (A3), set**

$$
q_k := \mathcal{B}_k (\mathcal{B}_k^T \mathcal{B}_k)^{-1} (Q_k^T \eta_{k+1} + R_k v_k), \quad k \in J.
$$
Then for all \( k \in J \), we have

\[
\begin{pmatrix}
\eta_{k+1}^T & v_k^T
\end{pmatrix}
\begin{pmatrix}
P_k & Q_k \\
Q_k^T & R_k
\end{pmatrix}
\begin{pmatrix}
\eta_{k+1} \\
v_k
\end{pmatrix} = \eta_{k+1}^T C_k \eta_{k+1} + q_k^T B_k q_k,
\]

\[\mathcal{A}_k \eta_{k+1} + \mathcal{B}_k v_k = A_k \eta_{k+1} + B_k q_k.\]

Consequently, the accessory problem takes the form

\[
\begin{aligned}
\text{minimize} & \quad \mathcal{J}(\eta, q) = \frac{1}{2} \eta_{K+1}^T \hat{H} \eta_{N+1} + \frac{1}{2} \sum_{k=0}^{N} \{ \eta_{k+1}^T C_k \eta_{k+1} + q_k^T B_k q_k \}, \\
\Delta \eta_k & = A_k \eta_{k+1} + B_k q_k, \quad k \in J, \\
\eta_0 & = 0, \quad \mathcal{M} \eta_{N+1} = 0,
\end{aligned}
\]

(A P)

where we assumed without loss of generality

\[\mathcal{M} := M^T (MM^T)^{-1} M \quad \text{and} \quad \hat{H} := (I - \mathcal{M}) \Gamma (I - \mathcal{M}) \quad (24)\]

instead of \( M \) and \( \Gamma \).

Remark 6. One can easily see that the nonnegativity of the original \( \mathcal{J}(\eta, v) \) is equivalent to the nonnegativity of the transformed \( \mathcal{J}_2(\eta, q) \) in problem (A P), simply because their values are equal, by Lemma 1. However, condition (A 3) yields the equivalence of their positivity in both arguments. Suppose that \( \mathcal{J}_2 \) from Eq. (17) is positive definite and let \((\eta, q)\) be any pair admissible for problem (A P). Then \( \mathcal{J}_2(\eta, q) \geq 0 \). If \( \mathcal{J}_2(\eta, q) = 0 \), then also \( \mathcal{J}_2(\eta, v) = 0 \) and hence \( \eta = 0 \), i.e., \( v = 0 \) because of assumption (A 3). From Eq. (22), it follows that \( q = 0 \). Similarly for the converse.

The transversality condition (23) together with the boundary conditions on \( \eta \) may be interpreted as the boundary condition for system (H); we write \((\eta, q) \in \mathcal{R}) \), which means

\[
(R S + S^*) \begin{pmatrix} -\eta_0 \\ \eta_{N+1} \end{pmatrix} + R \begin{pmatrix} q_0 \\ q_{N+1} \end{pmatrix} = 0,
\]

where the matrices \( R, S, S^* \in \mathbb{R}^{2n \times 2n} \) are defined by

\[
R := \begin{pmatrix} 0 & 0 \\ 0 & I - \mathcal{M} \end{pmatrix}, \quad S := \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}, \quad S^* := \begin{pmatrix} -I & 0 \\ 0 & \mathcal{M} \end{pmatrix},
\]

and where \( \text{rank}(R, S^*) = 2n \) and \( \text{Im} R^T = \ker S^* \). See [7, 10] for more details, or [14, Section 2.3] for the continuous counterpart.
5. DISCRETE QUADRATIC FUNCTIONALS

In this section we derive necessary conditions for the nonnegativity of a quadratic functional. They are phrased in terms of

(i) the positivity of certain partial quadratic functionals,
(ii) the nonexistence of conjugate intervals to \( N + 1 \) in \([1, N + 1] \),
(iii) the existence of an \( n \times n \)-matrix solution of the matrix form of the Hamiltonian system \((H)\) satisfying certain boundary and rank conditions,
(iv) the existence of a solution of the corresponding discrete matrix Riccati equation with some special property.

Consider the discrete quadratic functional of the form

\[
\mathcal{J}(\eta, q) := \frac{1}{2} \eta \mathbf{T}_{N+1} \hat{\mathbf{I}} \eta_{N+1} + \frac{1}{2} \sum_{k=0}^{N} \{ \eta_{k+1}^T C_k \eta_{k+1} + q_k^T B_k q_k \},
\]

where \((\eta, q)\) satisfies \( \Delta \eta_k = A_k \eta_{k+1} + B_k q_k, k \in J, \) and the boundary conditions \( \eta_0 = 0, \mathbf{M} \eta_{N+1} = 0 \). Here \( A, B, C: J \rightarrow \mathbb{R}^{n \times n} \) are sequences of real matrices, \( B_k \) and \( C_k \) are symmetric, \( I - A_k \) nonsingular, \( \Gamma \in \mathbb{R}^{n \times n}, \Gamma \in \mathbb{R}^{r \times n} \) is full rank, \( \eta: J^* \rightarrow \mathbb{R}^n \) and \( q: J \rightarrow \mathbb{R}^n \) are sequences of real vectors, and the matrices \( \mathbf{M} \) and \( \mathbf{I} \) are defined in terms of \( M \) and \( \Gamma \) by Eq. (24). We denote \( A_k := (I - A_k)^{-1}, k \in J \). It is clear to see that the second variation of this quadratic functional \( \mathcal{J} \) is in fact itself. Hence, it will also be denoted by \( \mathcal{J} \).

Let us define, for \( m \in J \), the partial quadratic functionals \( \mathcal{J}_m \) to be

\[
\mathcal{J}_m(\eta, q) := \frac{1}{2} \eta \mathbf{T}_{N+1} \hat{\mathbf{I}} \eta_{N+1} + \frac{1}{2} \sum_{k=m}^{N} \{ \eta_{k+1}^T C_k \eta_{k+1} + q_k^T B_k q_k \}.
\]

Of course, with this notation we have \( \mathcal{J}_0 = \mathcal{J} \). Accordingly, the pair \((\eta, q)\) is called admissible (on \([m, N + 1]\)), provided \( \Delta \eta_k = A_k \eta_{k+1} + B_k q_k \) for all \( k \in [m, N] \). The pair \((\eta, v)\) is feasible if it is admissible and satisfies the boundary conditions.

**Definition 3 (positive (semi)definiteness).** Let \( m \in J \). The quadratic functional \( \mathcal{J}_m \) is called positive semidefinite (or nonnegative), we write \( \mathcal{J}_m \geq 0 \), if \( \mathcal{J}_m(\eta, q) \geq 0 \) for all pairs \((\eta, q)\) which are admissible on \([m, N + 1]\) and satisfy \( \eta_m = 0, \mathbf{M} \eta_{N+1} = 0 \).

The quadratic functional \( \mathcal{J}_m \) is called positive definite, we write \( \mathcal{J}_m > 0 \), provided \( \mathcal{J}_m(\eta, q) > 0 \) for all pairs \((\eta, q)\) which are admissible on \([m, N + 1]\) and satisfy \( \eta_m = 0, \mathbf{M} \eta_{N+1} = 0, \eta \neq 0 \).
The concept of strong normality (or controllability) plays an important role in our considerations. We recommend the reader to compare the parallel notion [6, Definition 5] or [5] of Bohner, which corresponds to "complete controllability on \( J^* \)," when considering the continuous-time systems.

**Definition 4 (strong normality).** The pair \((A, B)\) is called strongly normal on any interval of the form \([m, N + 1]\) if for all \( m \in J \), the system

\[
\begin{align*}
\Delta q_k &= -A_k^T q_k, & B_k q_k &= 0, & k \in [m, N], \\
q_{N+1} + M\gamma &= 0, & \gamma \in \mathbb{R}^n,
\end{align*}
\]

has only the zero solution \( q = 0 \) on \([m, N + 1]\).

The pair \((A, B)\) is called strongly normal on any interval of the form \([0, m + 1]\) if for all \( m \in [0, N - 1] \), the system

\[
\begin{align*}
\Delta q_k &= -A_k^T q_k, & B_k q_k &= 0, & k \in [0, m], \\
q_{N+1} + M\gamma &= 0, & \gamma \in \mathbb{R}^n,
\end{align*}
\]

has only the zero solution \( q = 0 \) on \([0, m + 1]\).

Observe that the strong normality of \((A, B)\) on intervals of the form \([m, N + 1]\) includes \( M \)-controllability of \((A, B)\) on \( J^* \) by simply taking the value zero for \( m \). The assumption of strong normality of both types is denoted by the following:

The pair \((A, B)\) is strongly normal on any interval of the form \([0, m + 1]\) and \([m, N + 1]\). (A4)

**Lemma 2.** Suppose that the pair \((A, B)\) is \( M \)-controllable on \( J^* \) and strongly normal on any interval of the form \([0, m + 1]\). Then \( \mathcal{A} \geq 0 \) implies that for all \( m \in [1, N] \), there is no pair \((\eta, q)\) admissible on \([m, N + 1]\) with \( \eta_m = 0, M\eta_{N+1} = 0, \eta \neq 0 \) on \([m, N + 1]\), satisfying \( \mathcal{J}_m(\eta, q) \leq 0 \).

**Proof.** Suppose that there exist \( m \in [1, N] \) and a pair \((\eta, q)\) admissible on \([m, N + 1]\) with \( \eta_m = 0, M\eta_{N+1} = 0, \eta \neq 0 \) on \([m, N + 1]\), and \( \mathcal{J}_m(\eta, q) \leq 0 \). We define

\[
(\bar{\eta}_k, \bar{q}_k) := \begin{cases} 
(\eta_k, q_k) & \text{for } k \in [m, N + 1], \\
(0, 0) & \text{for } k \in [0, m - 1].
\end{cases}
\]

Then \((\bar{\eta}, \bar{q})\) satisfies the equation of motion on \([0, m - 2]\) and on \([m, N]\). Moreover, the computation

\[
\tilde{A}_{m-1}\bar{\eta}_{m-1} + \tilde{A}_{m-1}B_{m-1}\bar{q}_{m-1} = 0 = \eta_m = \bar{\eta}_m
\]

(25)
shows that \((\bar{\eta}, \bar{q})\) is admissible on the whole interval \(J^*\). Since also \(\bar{\eta}_0 = 0, \bar{\eta}_{N+1} = 0\), then \((\bar{\eta}, \bar{q})\) is feasible, and hence, the assumption of the lemma implies \(\mathcal{J}(\bar{\eta}, \bar{q}) \geq 0\). However,

\[
\mathcal{J}(\bar{\eta}, \bar{q}) = \mathcal{J}_m(\bar{\eta}, \bar{q}) = \mathcal{J}_m(\eta, q) \leq 0.
\]

The strict inequality in the above formula contradicts \(\mathcal{J} \geq 0\). If in the other case \(\mathcal{J}(\bar{\eta}, \bar{q}) = 0\), then the pair \((\eta, q)\) is optimal for problem (AP). Then Theorem 1 ensures the existence of the number \(\lambda_0 \geq 0\) and \(n\)-vector parameters \(\gamma, p = (p_k)_{k=0}^{N+1}\) satisfying

\[
-\Delta p_k = \lambda_0 C_k \bar{\eta}_{k+1} + A_k^T p_k, \quad \lambda_0 B_k \bar{q}_k + B_k p_k = 0, \quad k \in J,
\]

\[
p_{N+1} = \mathcal{J}_m + \lambda_0 \nabla K^T(\bar{\eta}_{N+1}).
\]

Set \(w_k = -p_k\) on \(J^*\). If \(\lambda_0 = 0\), then we have \(\Delta w_k = -A_k^T w_k\) and \(B_k w_k = 0, k \in J\), with \(w_{N+1} + \mathcal{J}_m = 0\). Due to the \(M\)-controllability on \(J^*\), Remark 3 implies \(w = 0\) on \(J^*\). Hence, we have a contradiction with (i) of Theorem 1.

Thus \(\lambda_0 \neq 0\), and we may take actually \(\lambda_0 = 1\). In this case, \((\eta, w)\) solves system (H) on \(J\). Now since \(\bar{\eta}_k = 0\) on \([0, m]\), the strong normality on \([0, m]\) implies that \(w = 0\) on \([0, m]\) and hence \((\eta, w) = (0, 0)\) on \(J^*\). This contradicts the fact \(\eta = \bar{\eta} \neq 0\) on \([m, N+1]\) and the proof is complete. 

**Theorem 4** (necessary condition for \(\mathcal{J} \geq 0\)). Suppose that \((A, B)\) is \(M\)-controllable on \(J^*\) and strongly normal on any interval of the form \([0, m+1]\). Then

\[
\mathcal{J} \geq 0 \quad \text{implies} \quad \mathcal{J}_m > 0 \quad \text{for all} \quad m \in [1, N].
\]

**Remark 7.** For each \(m \in [1, N]\), the condition \(\mathcal{J}_m > 0\) can be expressed in view of [6, Theorem 2] in terms of any of the following equivalent conditions:

1. disconjugacy of system (H) on \([m, N+1]\),
2. the nonexistence of focal points in \((m, N+1]\) of the principal solution of system (H) at \(m\) \((X_m = 0, U_m = I)\),
3. the existence of a symmetric solution to an implicit (as opposed to the explicit form in Theorem 7) Riccati matrix equation.

Other equivalent formulations may be derived for the “time-reversed” system; see [8, Theorem 1].

**Proof.** Let \(m \in [1, N]\) and suppose \(\mathcal{J} \geq 0\). Take any pair \((\eta, q)\) admissible on \([m, N+1]\) with \(\eta_m = 0, \mathcal{J}_{\eta_{N+1}} = 0, \eta \neq 0\). Then extend \((\eta, q)\) by zero to the whole interval \(J^*\) to obtain a nontrivial pair \((\eta, q)\) which is
admissible on \( J^* \); see Eq. (25). Then \( \mathcal{I}_m(\eta, q) = \mathcal{I}(\eta, q) \geq 0 \) by the assumption. However, the equality \( \mathcal{I}_m(\eta, q) = 0 \) contradicts Lemma 2 and the desired result is proven. 

Parallel to the notion introduced in [6, Definition 4], we say that the interval \([m, m+1]\) is a generalized zero of a solution \((\eta, q)\) of system \((H)\), provided

\[
\eta_{m+1} \neq 0, \quad \eta_{m+1} \in \text{Im} \tilde{A}_m B_m, \quad \eta_m^T B_m^T (I - A_m) \eta_{m+1} \leq 0.
\]

This formulation is equivalent to saying that for some \( c \in \mathbb{R}^n \),

\[
\eta_{m+1} \neq 0, \quad \eta_{m+1} = \tilde{A}_m B_m c, \quad \eta_m^T c \leq 0. \tag{26}
\]

**Definition 5 (conjugate interval).** Let \( m \in J \). The interval \([m, m+1]\) is said to be conjugate to \( N + 1 \) if there exists a solution \((\eta, q)\) of Eq. (H) having \([m, m+1]\) as a generalized zero and satisfying

\[
\mathcal{A} \eta_{N+1} = 0, \quad \eta_{N+1}^T (q_{N+1} + \hat{\eta} \eta_{N+1}) = 0. \tag{27}
\]

**Remark 8.** Note that in case of fixed endpoints, i.e., for \( \mathcal{A} = I \), the above definition excludes \([N, N + 1]\) being conjugate to \( N + 1 \).

**Theorem 5 (necessary condition for \( \mathcal{I} \geq 0 \)).** Suppose the conditions of Lemma 2 hold. Then \( \mathcal{I} \geq 0 \) implies that there is no interval \([m, m+1] \subseteq [1, N + 1]\) conjugate to \( N + 1 \).

**Proof.** Assume to the contrary that there exists \( m \in [1, N] \) such that \([m, m+1]\) is conjugate to \( N + 1 \). According to Remark 8, we may suppose that either \( \mathcal{A} \neq I \), or if \( \mathcal{A} = I \) then \( m < N \). That means there is a solution \((\eta, q)\) of system \((H)\) satisfying Eqs. (26) and (27). Set

\[
\tilde{\eta}_k := \begin{cases} \eta_k & \text{for } k \in [m + 1, N + 1], \\ 0 & \text{otherwise,} \end{cases} 
\]

\[
\tilde{q}_k := \begin{cases} c & \text{for } k = m, \\ q_k & \text{for } k \in [m + 1, N + 1], \\ 0 & \text{otherwise.} \end{cases}
\]

Then \( \tilde{\eta}_0 = 0, \mathcal{A} \tilde{\eta}_{N+1} = \mathcal{A} \eta_{N+1} = 0 \), and as in Eq. (25), we have

\[
\tilde{A}_m \tilde{\eta}_m + \tilde{A}_m B_m \tilde{q}_m = \tilde{A}_m B_m c = \eta_{m+1} = \tilde{\eta}_{m+1}.
\]
so that \( (\tilde{\eta}, \tilde{q}) \) is feasible on \( J^*. \) Then

\[
\mathcal{J}(\tilde{\eta}, \tilde{q}) = \frac{1}{2} \tilde{\eta}_{N+1}^T \hat{\Gamma} \tilde{\eta}_{N+1} + \frac{1}{2} \sum_{k=m}^{N} \left\{ \tilde{\eta}_{k+1}^T C_k \tilde{\eta}_{k+1} + \tilde{q}_{k}^T B_k \tilde{q}_k \right\}
\]

\[
= \frac{1}{2} \tilde{\eta}_{N+1}^T \hat{\Gamma} \tilde{\eta}_{N+1} + \frac{1}{2} \tilde{\eta}_{m+1}^T \hat{\Gamma} \tilde{\eta}_{m+1} - \frac{1}{2} \tilde{\eta}_{m}^T \tilde{q}_m + \frac{1}{2} \sum_{k=m}^{N} \left\{ \tilde{\eta}_{k+1}^T C_k \tilde{\eta}_{k+1} - A_k^T \tilde{q}_k - \Delta \tilde{q}_k \right\}
\]

\[
= \frac{1}{2} \tilde{\eta}_{N+1}^T \left( \hat{\Gamma}_N \eta_{N+1} + q_{N+1} \right) + \frac{1}{2} \tilde{\eta}_{m+1}^T \left( C_m \tilde{\eta}_{m+1} - A_m^T \tilde{q}_m - \Delta \tilde{q}_m \right)
\]

\[
= \frac{1}{2} \tilde{\eta}_{m+1}^T \left( C_m \tilde{\eta}_{m+1} - A_m^T \tilde{q}_m - \Delta \tilde{q}_m \right)
\]

\[
= \frac{1}{2} \left( \tilde{\eta}_{m}^T + q_m^T B_m \right) (c - q_m)
\]

\[
= \frac{1}{2} \eta_m^T c \leq 0.
\]

Since \( \mathcal{J} \geq 0 \), it follows that \( \mathcal{J}(\tilde{\eta}, \tilde{q}) < 0 \) cannot hold. Thus, \( \mathcal{J}(\tilde{\eta}, \tilde{q}) = 0 \). Since \( \tilde{\eta}_{m+1} \neq 0 \), then \( \tilde{\eta} \neq 0 \) on \( [m, N+1] \) and hence, Lemma 2 gives a contradiction. Therefore, the proof is complete.

We say that an \( n \times n \)-matrix solution \((X, U)\) of the matrix form of system (H) is a conjoined basis if \( X^T U_k \) is symmetric and rank \( (\frac{\eta}{\xi}) = n \) at some (and hence at any) \( k \in J^* \).

**Theorem 6** (necessary condition for \( \mathcal{J} \geq 0 \)). Suppose Eq. (A4) holds. Then the nonexistence of intervals \([m, m+1] \subseteq [1, N+1]\) implies the existence of a conjoined basis \((X, U)\) of the matrix form of system (H) satisfying

\[
X_k \text{ invertible for all } k \in [1, N],
\]

\[
\mathcal{M} X_{N+1} = 0, \quad X_{N+1}^T \left( U_{N+1} + \hat{\Gamma} X_{N+1} \right) = 0.
\]

In fact, we may take \((X, U)\) to be the solution of system (H) satisfying the boundary conditions

\[
X_{N+1} = I - \mathcal{M}, \quad U_{N+1} = -\hat{\Gamma} - \mathcal{M}.
\]

**Proof.** Let \((X, U)\) be the solution of system \((H)\) from the theorem and suppose on the contrary that there exist \( m \in [1, N] \) and \( \alpha \in \mathbb{R}^n, \alpha \neq 0, \) such that \( X_m \alpha = 0 \). Set \((\eta_m) := (\frac{\eta}{\xi}) \alpha\) on \( J^* \). Then \( \eta_m = 0 \). Also, \( \eta \neq 0 \) on \([m, N+1]\). In fact, if \( \eta = 0 \) on \([m, N+1]\), by the controllability on
Using the endpoints of $X_{N+1}$ and $U_{N+1}$, it results that $\alpha = 0$, which is a contradiction. Let $j$ be the first index in $[m, N + 1]$ such that $\eta_j = 0$ and $\eta_{j+1} \neq 0$. Then, for $c := q_j \in \mathbb{R}^n$, we have $\eta_{j+1} = A_j B_j c$ and $\eta_j^T c \leq 0$. Hence, $[j, j + 1]$ is conjugate to $N + 1$ and a contradiction is obtained.

Thus the result of the theorem holds true.

The proof of the above theorem yields the following result.

**Corollary 1.** The assumptions of Theorem 6 imply that every nonzero solution $(\eta, q)$ of system (H) with Eq. (27) satisfies $\eta_k \neq 0$ for all $k \in [1, N]$.

**Remark 9.** Here we present an example that our conjugate interval definition is the right one over the natural extension from the continuous control theory. We see that the condition $\eta_k \neq 0$, which by the corollary above is necessary for $\Theta(\eta, q) \leq 0$, is too weak to provide a satisfactory concept of “conjugate points” in the discrete case. In the example we will construct a discrete linear Hamiltonian system, for which we find a solution $(\eta, q)$ satisfying Eq. (27) and $\eta_k \neq 0$ on $J^*$. However, such a solution yields that $(1,2)$ is conjugate to 3 and that there is a feasible pair $(\tilde{\eta}, \tilde{q})$ satisfying $\Theta(\tilde{\eta}, \tilde{q}) < 0$. Take $n = 2$, $N = 2$, i.e., $J = [0, 2]$. Set $A_0 = 0, C_0 = 0, B_k = (\frac{3}{2})$ on $J, \Theta = 0, \Gamma = (-\frac{1}{2}, \frac{1}{2}), c = (-\frac{1}{2}), q = (-\frac{1}{2}), \eta_0 = (\frac{1}{2})$. Then system (H) is of the form

\[
\begin{align*}
\Delta \eta_k &= B_k q_k, \\
\Delta q_k &= 0,
\end{align*}
\]

or equivalently

\[
\begin{align*}
\eta_{k+1} &= \eta_k + B_k q_k, \\
q_{k+1} &= q_k (=: q),
\end{align*}
\]

$k = 0, 1, 2$.

Then Eq. (27) is satisfied and $\eta_k \neq 0$ everywhere. But $[1, 2)$ is conjugated to 3 because $\eta_2 = B_2 c \neq 0$ and $\eta_1^T c = -1 \leq 0$. Note that the pair $(\tilde{\eta}, \tilde{q})$ defined by $\tilde{\eta} := \eta_k$ on $[2, 3], \tilde{\eta}_0 := 0$ on $[0, 1], \tilde{q}_0 := 0, \tilde{q}_1 := c, \tilde{q}_2 = \tilde{q}_3 := q$, is feasible on $J^*$ and satisfies $\Theta(\tilde{\eta}, \tilde{q}) = -\frac{1}{2}$.

We associate with system (H) the discrete matrix Riccati equation

\[
\begin{align*}
R[W]_k &= \Delta W_k - C_k - A_k^T W_k + (W_{k+1} - C_k) \tilde{A}_k (A_k + B_k W_k) \\
&= (W_{k+1} - C_k) \tilde{A}_k (I + B_k W_k) - \tilde{A}_k^{-1} W_k \\
&= 0,
\end{align*}
\]

which is connected to the matrix form of system (H) via the formula $W_k = U_k X_k^{-1}$. In the following theorem we show that a necessary condition for the nonnegativity of the quadratic functional $\Theta$ is that the explicit matrix Riccati equation (R) possesses a solution with an appropriate boundary condition at $N + 1$. 


Let \((X, U), (\tilde{X}, \tilde{U})\) be the conjoined bases of system (H) given by the final conditions

\[
X_{N+1} = I - \mathcal{M}, \quad \tilde{X}_{N+1} = (\tilde{\Gamma} + \mathcal{M})(I + \tilde{\Gamma}^2)^{-1},
\]

\[
U_{N+1} = -\tilde{\Gamma} - \mathcal{M}, \quad \tilde{U}_{N+1} = (I - \mathcal{M})(I + \tilde{\Gamma}^2)^{-1}.
\]

Observe that \(I + \tilde{\Gamma}^2 = X^TX + U^TU\) is invertible because of the rank condition on \((X, U)\). Note also that \((X, U), (\tilde{X}, \tilde{U})\) are normalized, i.e., \(X^TX - U^TU = I\).

**Theorem 7 (necessary condition for \(J \geq 0\)).** Suppose that there exists a conjoined basis \((X, U)\) satisfying the conclusion of the previous theorem. Then there exists a symmetric function \(W_k\) defined on \([1, N - 1]\), and

\[
I + B_kW_k \text{ invertible on } [1, N - 1],
\]

\[
(I - \mathcal{M})R[W]_N = 0,
\]

\[
(I - \mathcal{M})W_{N+1} + \tilde{\Gamma} = 0.
\]

In fact, we may take on \(J^*\)

\[
W = UX^T + (UX^T\tilde{X} - \tilde{U})(I - X^TX)U^T.
\]

**Proof.** Let \(W\) be as above. Then \(X_k\) is invertible on \([1, N]\), by Theorem 6, so that Eq. (R) is satisfied on \([1, N - 1]\). The first equation of (H) reads as \(X_{N+1} = A_N(I + B_kW_k)X_k^{-1}\) on \([1, N - 1]\), and hence Eq. (28) follows. Formula (30) is obtained from the computation

\[
W_{N+1} = -\tilde{\Gamma} + \mathcal{M} - \mathcal{M}(I + \tilde{\Gamma}^2)^{-1} \mathcal{M},
\]

where we used the defining properties of \((I - \mathcal{M})^T\) and the identities \(\hat{\Gamma}(I - \mathcal{M})^T = \hat{\Gamma}\) and \(\mathcal{M}(I - \mathcal{M})^T = 0\). Finally,

\[
R[W]_N = (W_{N+1} - C_N)A_N(I + B_kW_k) - A_N^T W_N
\]

\[
= (W_{N+1} - C_N)A_N(X_N + B_NU_N)X_N^{-1} - A_N^T U_N X_N^{-1}
\]

\[
= (W_{N+1} - C_N)X_N^{-1} - (I - A_N^T)U_N X_N^{-1}
\]

\[
= (W_{N+1}(I - \mathcal{M}) - (C_NX_{N+1} - A_N^T U_N + U_N))X_N^{-1}
\]

\[
= (-\tilde{\Gamma} - U_{N+1})X_N^{-1} = \mathcal{M}X_N^{-1}
\]

shows that Eq. (29) holds and the proof is finished. \[\blacksquare\]
Remark 10. Note that the proof reveals in fact more specific boundary condition for $W_{N+1}$ than Eq. (30), and more precise statement on $R[W]_N$ than Eq. (29).

Next we derive a necessary condition for the nonnegativity of $\mathcal{J}$, which requires no controllability (normality) assumption. In a sense, it may be viewed as Legendre necessary condition for $\mathcal{J} \geq 0$; see, e.g., [4, Section 4.3]. This condition is based on the discreteness of the problem considered here, so it has no natural equivalent in the continuous theory. Let $\mathcal{J}$ be the tridiagonal matrix associated with the quadratic functional $\mathcal{J}$, i.e.,

$$\mathcal{J} = \begin{pmatrix} T_0 & V_1 \\ V_1^T & T_1 & \ddots \\ & \ddots & \ddots & V_N \\ V_N^T & & & T_N \end{pmatrix},$$

where the matrices $T_k$ and $V_k$ are defined by

$$T_k := C_k + \tilde{A}_k^{-1} B_k \tilde{A}_k^{-1} + B_k^{T}, \quad k \in [0, N - 1],$$

$$T_N := C_N + \tilde{A}_N^{-1} B_N \tilde{A}_N^{-1} + \tilde{\Gamma},$$

$$V_k := -B_k \tilde{A}_k^{-1}, \quad k \in J.$$

This appeared in the literature already in several forms; see, e.g., [9] or the above reference. As in [6, Remark 3(ii)], set $G_0 := 0$ and for $k \in [1, N + 1],

$$G_k := \begin{pmatrix} \tilde{A}_{k-1} \tilde{A}_{k-2} \cdots \tilde{A}_0 & \tilde{A}_{k-1} & \cdots & \tilde{A}_1 B_1 & \cdots & \tilde{A}_k B_k \end{pmatrix}.$$

Define the $(N + 1)n \times (N + 1)n$ controllability matrix $\mathcal{G}$ by

$$\mathcal{G} := \begin{pmatrix} G_1^T & G_2^T & \cdots & G_{N+1}^T \end{pmatrix}^T,$$

i.e., $\mathcal{G}$ is the lower triangular matrix with the only nonzero $n \times n$ block entries

$$\mathcal{G}_{i,j} = \tilde{A}_{i-1} \tilde{A}_{i-2} \cdots \tilde{A}_{j-1} B_{j-1}, \quad i, j \in [1, N + 1], \quad i \geq j.$$

Theorem 8. The quadratic functional $\mathcal{J}$ is nonnegative if and only if $\mathcal{G}^T \mathcal{G} \geq 0$ on $\text{Ker} \, \mathcal{M} G_{N+1}$.

Proof. Let $(\eta, q)$ be admissible on $J^*$ with $\eta_0 = 0$ and $\mathcal{M} \eta_{N+1} = 0$. Then $\eta_k = G_k \omega_k$, $k \in J^*$, where we denote for $k \in J$

$$\omega_k := \begin{pmatrix} q_0 \\ \vdots \\ q_{k-1} \end{pmatrix} \quad \text{and} \quad \psi := \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_{N+1} \end{pmatrix}.$$
Then it follows that
\[ \mathcal{I}(\eta, q) = \frac{1}{2} \psi^T \mathcal{K} \psi = \frac{1}{2} \omega_{N+1}^T \mathcal{G} \mathcal{K} \omega_{N+1}. \]

The result follows immediately from \( \omega_{N+1} = \mathcal{G} \omega_{N+1} \), since the vector \( \omega_{N+1} \in \mathbb{R}^{(N+1)n} \) may be chosen arbitrarily.

By considering the above theorem over different subspaces of \( \omega_{N+1} \)'s, we may obtain, as corollaries, several necessary conditions for \( \mathcal{I} \geq 0 \).

A slightly different approach—using the vector \( \psi \in \mathbb{R}^{(N+1)n} \)—is used to derive the Legendre condition for discrete variational problems, i.e., when \( B_i \)'s are invertible, in [4, Theorem 4.16]; see also [3, Section 4]. The idea of isolating the diagonal elements of \( \mathcal{K} \) may be used also for problems with singular \( B_i \)'s and a variable endpoint. The result in [4, Theorem 4.16] is then a special case of the following corollary for invertible \( B_i \)'s and \( \mathcal{M} = I \).

For fixed \( m \in J \), denote by \( \mathcal{V}_{m+1} \) the space of those vectors \( c \in \mathbb{R}^n \) such that the equations \( B_m q_m = A_m^{-1} c \) and \( B_{m+1} q_{m+1} = -c \) are solvable with respect to \( q_m, q_{m+1} \), i.e.,
\[ \mathcal{V}_{m+1} = \{ c \in \mathbb{R}^n; \eta = (0, \ldots, 0, c^T, 0, \ldots, 0)^T \text{ is admissible on } J^* \}, \]

where the only nonzero entry is \( \eta_{m+1} = c \). The sequence \( \eta = (\eta_k)_{k=0}^{N+1} \) is called admissible on \( J^* \) if there exists \( q = (q_k)_{k=0}^{N} \) such that \( (\eta, q) \) is admissible on \( J^* \). In the case when \( (A, B) \) is strongly normal on \([0, m]\) and \([m, N+1]\), e.g., when \( B_i \)'s are invertible, we have \( \mathcal{V}_{m+1} = \mathbb{R}^n \).

Corollary 2 (discrete Legendre condition). If \( \mathcal{I} \geq 0 \), then for all \( k \in [0, N-1] \),
\[ T_k \geq 0 \text{ on } \mathcal{V}_{k+1}, \quad \text{and} \quad T_N \geq 0 \text{ on } \mathcal{V}_{N+1} \cap \text{Ker } \mathcal{M} \tilde{A}_N B_N. \]

6. Optimality Conditions for Problem (P)

We consider the nonlinear discrete optimal control problem (P) stated in Section 2. It has been shown in Section 4 that the accessory problem (AP) associated with (P) is nonnegative. Hence, by applying all the results of Section 5 to the second variation \( \mathcal{J}_2 \), we obtain the following necessary conditions for optimality in problem (P). The matrices \( A, B, C, \tilde{A}, \) and \( \mathcal{M} \) in Section 5 are now defined in terms of the data of the problem (P) through Eqs. (19), (20), (18), (3), and (24).

Theorem 9 (necessary optimality conditions). Assume that \((\hat{x}, \hat{u})\) is a weak local minimum for problem (P) and suppose that assumptions (A2)–(A4) are satisfied. Then the conclusions of Theorems 4–8 and Corollaries 1–2 hold true.
REFERENCES