A discrete-time Pontryagin maximum principle on matrix Lie groups☆

Karmvir Singh Phogat *, Debasish Chatterjee, Ravi N. Banavar

Systems & Control Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

ARTICLE INFO

ABSTRACT

In this article we derive a Pontryagin maximum principle (PMP) for discrete-time optimal control problems on matrix Lie groups. The PMP provides first order necessary conditions for optimality; these necessary conditions typically yield two point boundary value problems, and these boundary value problems can then be solved to extract optimal control trajectories. Constrained optimal control problems for mechanical systems, in general, can only be solved numerically, and this motivates the need to derive discrete-time models that are accurate and preserve the non-flat manifold structures of the underlying continuous-time controlled systems. The PMPs for discrete-time systems evolving on Euclidean spaces are not readily applicable to discrete-time models evolving on non-flat manifolds. In this article we bridge this gap and establish a discrete-time PMP on matrix Lie groups. Our discrete-time models are derived via discrete mechanics, (a structure preserving discretization scheme) leading to the preservation of the underlying manifold under the dynamics, thereby resulting in greater numerical accuracy of our technique. This PMP caters to a class of constrained optimal control problems that includes point-wise state and control action constraints, and encompasses a large class of control problems that arise in various field of engineering and the applied sciences.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The Pontryagin maximum principle (PMP) provides first order necessary conditions for a broad class of optimal control problems. These necessary conditions typically lead to two-point boundary value problems that characterize optimal control, and these problems may be solved to arrive at the optimal control functions. This approach is widely applied to solve optimal control problems for controlled dynamical systems that arise in various fields of engineering including robotics, aerospace (Agrachev & Sachkov, 2004; Brockett, 1973; Lee, Leok, & McClamroch, 2008a, b), and quantum mechanics (Bonnard & Sugny, 2012; Khaneja, Brockett, & Glaser, 2001).

Constrained optimal control problems for nonlinear continuous-time systems can, in general, be solved only numerically, and two technical issues inevitably arise. First, the accuracy guaranteed by a numerical technique largely depends on the discretization of the continuous-time system underlying the problem. For control systems evolving on complicated state spaces such as

manifolds, preserving the manifold structure of the state space under discretization is a nontrivial matter. For controlled mechanical systems evolving on manifolds, discrete-time models preferably are derived via discrete mechanics since this procedure respects certain system invariants such as momentum, kinetic energy, (unlike other discretization schemes derived from Euler's step) resulting in greater numerical accuracy (Marsden & West, 2001; Ober-Blöbaum, 2008; Ober-Blöbaum, Junge, & Marsden, 2011). Second, classical versions of the PMP are applicable only to optimal control problems in which the dynamics evolve on Euclidean spaces, and do not carry over directly to systems evolving on more complicated manifolds. Of course, the PMP, first established by Pontryagin and his students (Gamkrelidze, 1999; Pontryagin, 1987) for continuous-time controlled systems with smooth data, has, over the years, been greatly generalized, see e.g., Agrachev and Sachkov (2004), Barbero-Liñán and Muñoz Lecanda (2009), Clarke (2013), Clarke (1976), Dubovitskii and Milyutin (1968), Holtzman (1966), Milyutin and Osmolovskii (1998), Mordukhovich (1976), Sussmann (2008) and Warga (1972). However, there is still no PMP that is readily applicable to control systems with discrete-time dynamics evolving on manifolds. As is evident from the preceding discussion, numerical solutions to optimal control problems, via digital computational means, need a discrete-time PMP. Here we establish a PMP for a class of discrete-time controlled systems evolving on matrix Lie groups.

Optimal control problems on Lie groups are of great interest due to their wide applicability across the discipline of engineering: robotics (Bullo & Lynch, 2001), computer vision (Vemulapalli,

---

Arrate, & Chellappa, 2014), quantum dynamical systems (Bonnard & Sugny, 2012; Khaneja et al., 2001), and aerospace systems such as attitude maneuvers of a spacecraft (Kobilarov & Marsden, 2011; Lee et al., 2008b; Saccon, Hauser, & Aguiar, 2013). The conjunction of discrete mechanics and optimal control (DMOC) for solving constrained optimal control problems while preserving the geometric properties of the system has been explored in Ober-Blöbaum (2008). The aforementioned DMOC technique is a direct geometric optimal control technique that differs from our technique on the account that our technique is an indirect method (Trélat, 2012); consequently (Trélat, 2012), the proposed technique is likely to provide more accurate solutions than the DMOC technique. Another important feature of our PMP is that it can characterize abnormal extremals unlike DMOC and other direct methods. Early results on indirect methods for optimal control problems on Lie groups for discrete-time systems derived via discrete mechanics may be found in Kobilarov and Marsden (2011) and Lee et al. (2008a, b) . Another, such technique is to derive higher order variational integrators to solve optimal control problems (Colombo, Ferraro, & Martín de Diego, 2016; Colombo, Jiménez, & Martín de Diego, 2012). It is worth noting that simultaneous state and action constraints have not been considered in any of these formulations. The inclusion of state and action constraints in optimal control problems, while of crucial importance in all real-world problems, makes constrained optimal control problems technically challenging, and, moreover, classical variational analysis techniques are not applicable in deriving first order necessary conditions for such constrained problems (Pontryagin, 1987, p. 3). More precisely, the underlying assumption in calculus of variations that an extremal trajectory admits a neighborhood in the set of admissible trajectories does not necessarily hold for such problems due to the presence of the constraints. This article addresses a class of optimal control problems in which the discrete-time controlled system dynamics evolve on matrix Lie groups, and are subject to simultaneous state and action constraints. We derive first order necessary conditions bypassing techniques involving classical variational analysis. Discrete-time PMPs for various special cases are subsequently derived from the main result.

A discrete-time PMP is fundamentally different from a continuous-time PMP due to intrinsic technical differences between continuous and discrete-time systems (Bourdin & Trélat, 2016, p. 53). While a significant research effort has been devoted to developing and extending the PMP in the continuous-time setting, by far less attention has been given to the discrete-time versions. A few versions of discrete-time PMP can be found in Boltyanskii, Martini, and Soltan (1999), Dubovitskii (1978) and Holtzman (1966).[1] In particular, Boltyanskii developed the theory of tents using the notion of local convexity, and derived general discrete-time PMPs that address a wide class of optimal control problems in Euclidean spaces subject to simultaneous state and action constraints (Boltyanskii, 1975). This discrete-time PMP serves as a guiding principle in the development of our discrete-time PMP on matrix Lie groups even though it is not directly applicable in our problem; see Remark 12 ahead for details.

This article unfolds as follows: our main result, a discrete-time PMP for controlled dynamical systems on matrix Lie groups, and its applications to various special cases are derived in Section 2. Section 3 provides a detailed proof of our main result, and the proofs of the other auxiliary results and corollaries are collected in the Appendices.
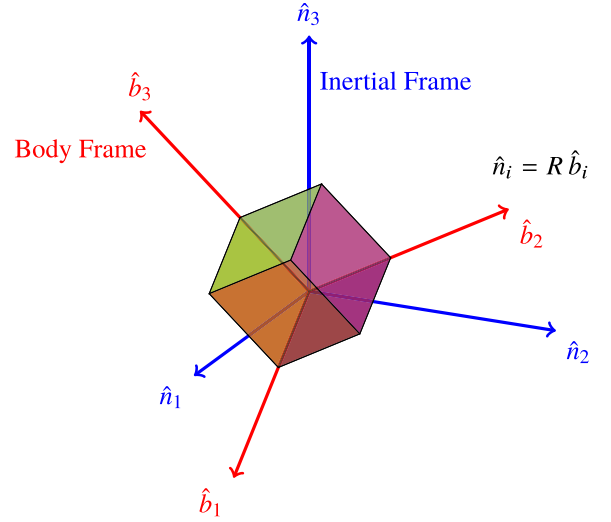

Fig. 1. Rigid body orientation.

2. Background and main results

This section contains an introduction to Lie group variational integrators that motivates a general form of discrete-time systems on Lie groups. Later in this section we establish a discrete-time PMP for optimal control problems associated with these discrete-time systems.

To illustrate the engineering motivation for our work, and ease understanding, we first consider an aerospace application. Let us first consider an example of control of spacecraft attitude dynamics in continuous time. The configuration space SO(3) (the set of 3×3 orthonormal matrices with real entries and determinant 1) of a spacecraft performing rotational maneuvers (Lee et al., 2008b), is a matrix Lie group with matrix multiplication as the group operation. Let $R \in$ SO(3) be the rotation matrix that relates coordinates in the spacecraft body frame to the inertial frame, (see Fig. 1,) let $\omega \in \mathbb{R}^3$ be the spacecraft momentum vector in the body frame, and let $u \in \mathbb{R}^3$ be the torque applied to the spacecraft in the body frame. The attitude dynamics in this setting is given in the spacecraft body frame (Lee et al., 2008b) as:

$$\dot{R} = R\hat{\omega}, \tag{1}$$

$$J\dot{\omega} = \hat{\omega}J\omega + u, \tag{2}$$

where J is the 3×3 moment of inertia matrix of the spacecraft in the body frame, $\hat{\omega} \in \mathfrak{so}(3)$ and $\mathfrak{so}(3)$ (the set of 3×3 skew-symmetric matrices with real entries) is the Lie algebra (Sachkov, 2009) corresponding to the Lie group SO(3). The first equation (1) describes the kinematic evolution and the second equation (2) describes the dynamics.

Let us, as a first step, uniformly discretize the continuous-time model (1)–(2) to arrive at an approximate discrete-time model. Fixing a step length $h > 0$, we have the discrete-time instances $t \in \{0\} \cup \mathbb{N}$ corresponding to the continuous-time instances $th \in \mathbb{R}$ in a standard way. Therefore, the system configurations at the discrete-time instances defined above are given by

$$R_t := R(th), \quad \omega_t := \omega(th) \quad \text{for all} \quad t \in \{0\} \cup \mathbb{N}.$$

If we assume that spacecraft body momentum is constant on the interval $[th, (t + 1)h[$, i.e., $\omega(s) = \omega(th)$ for $s \in [th, (t + 1)h[$, then the corresponding kinematic equations $\dot{R}(s) = R(s)\hat{\omega}_t$ for $s \in$

---

[1] Some early attempts in establishing discrete-time PMP in Euclidean spaces have been mathematically incorrect (Bourdin & Trélat, 2016, p. 53).

$[th, (t + 1)h[$ represent a linear system[2] in the time interval $[th, (t + 1)h[$. This linear system admits an analytical solution, and the discrete-time evolution of the continuous-time kinematic equation (1) is approximated as

$$R_{t+1} = R_t e^{h\hat{\omega}_t}, \tag{3}$$

where $e : \mathfrak{so}(3) \to SO(3)$ is the exponential map (Abraham & Marsden, 1978, p. 256) from the Lie algebra $\mathfrak{so}(3)$ to the Lie group $SO(3)$.[3] Similarly, the discrete-time system corresponding to the continuous-time dynamics (2) is approximated using Euler's step to be

$$J\omega_{t+1} = (I_{3\times3} + h\hat{\omega}_t)J\omega_t + hu_t. \tag{4}$$

It is worth noting here that the discrete integration step $e^{h\hat{\omega}_t}$ that describes the discrete evolution of the kinematic equation (3) can, in general, be a function of the configuration space variable $R_t \in SO(3)$ along with the spacecraft momentum vector $\hat{\omega}_t \in \mathfrak{so}(3)$. Similarly, the discrete-time evolution of the spacecraft momentum dynamics (4) can also, in general, depend on the orientation especially if the spacecraft is subjected to internal actuations via reaction wheels.

The above considerations lead to the following general form of the state dynamics:

$$\begin{cases} R_{t+1} = R_t s_t(R_t, \omega_t), \\ \omega_{t+1} = f_t(R_t, \omega_t, u_t), \end{cases} \tag{5}$$

where $f_t : SO(3) \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, $s_t : SO(3) \times \mathbb{R}^3 \to SO(3)$ are maps that define the discrete evolution of the system. Note that the exponential map $e : \mathfrak{so}(3) \to SO(3)$ is a diffeomorphism on a suitable neighborhood of $0 \in \mathfrak{so}(3)$. Let $\mathcal{O} \subset \mathfrak{so}(3)$ be a bounded open set; then there exists $\bar{h} > 0$ such that for all $h < \bar{h}$, $\mathcal{O} \ni \hat{a}_t \mapsto e^{\hat{a}_t h} \in e(\mathcal{O})$ is a diffeomorphism. The diffeomorphic property of the map $\mathcal{O} \ni \hat{a}_t \mapsto e^{\hat{a}_t h} \in e(\mathcal{O})$ restricted to a suitable open set is crucial for defining the local parametrization of the Lie group $SO(3)$ in terms of Lie algebraic elements, thus distilling a vector space structure to the discrete-time optimal control problem defined on the Lie group $SO(3)$.

We are now in a position to generalize the idea of discretization brought forth in the example of the attitude dynamics of a spacecraft to dynamical systems evolving on matrix Lie groups. To this end, let N be a natural number; in the sequel N will play the role of a planning or control horizon, and will be fixed throughout. Inspired by (5), we consider the dynamics split into two parts, the first of which occurs on a matrix Lie group G, and the second on a Euclidean space $\mathbb{R}^{n_x}$. The discrete-time evolution of our control system on the configuration space $G \times \mathbb{R}^{n_x}$ is defined by the recursion

$$\begin{cases} q_{t+1} = q_t s_t(q_t, x_t), \\ x_{t+1} = f_t(q_t, x_t, u_t), \end{cases} \quad t = 0, \dots, N-1, \tag{6}$$

with the following data:

(6-a) $q_t \in G, x_t \in \mathbb{R}^{n_x}$ are the states of the system,
(6-b) $s_t : G \times \mathbb{R}^{n_x} \to G$ depicts the dynamics on the matrix Lie group G,
(6-c) $f_t : G \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_x}$ captures the dynamics on $\mathbb{R}^{n_x}$,
(6-d) $u_t \in \mathbb{U}_t \subset \mathbb{R}^{n_u}$, where $\mathbb{U}_t$ is a nonempty set of feasible control actions at time t.

---

[2] This linear system can be written in the standard form, i.e., $\dot{x} = Ax + bu$, by embedding $SO(3)$ into $M(3, \mathbb{R})$ (the set of 3×3 real matrices) and identifying $M(3, \mathbb{R})$ with R^9.

[3] Let G be the Lie group with associated Lie algebra $\mathfrak{g}$. Then, for any $X \in \mathfrak{g}$, there exist a map $e^X(\cdot) : \mathbb{R} \to G$ such that: $e^X(0) = e \in G$, $\frac{d}{dt}\big|_{t=0} e^X(t) = X$, and $e^X(t + s) = e^X(s)e^X(t)$, where e is the group identity.

A control action u_t is applied to our system at the instant t to drive the system states from (q_t, x_t) to (q_{t+1}, x_{t+1}) governed by (6). The sequence $\{u_t\}_{t=0}^{N-1}$ is known as the *controller*, the sequence $\{(q_t, x_t)\}_{t=0}^{N}$ describing system states is called the *system trajectory* (Boltyanskii, 1975) under the controller $\{u_t\}_{t=0}^{N-1}$, with the pair $\left(\{(q_t, x_t)\}_{t=0}^{N}, \{u_t\}_{t=0}^{N-1}\right)$ referred to as a state–action trajectory.

We synthesize a controller for our system (6) by minimizing the performance index

$$\mathscr{J}(\mathbf{q}, \mathbf{x}, \mathbf{u}) := \sum_{t=0}^{N-1} c_t(q_t, x_t, u_t) + c_N(q_N, x_N) \tag{7}$$

with the following data:

(7-a) $c_t : G \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \to \mathbb{R}$ accounts for the cost-per-stage for each $t = 0, \dots, N-1$,
(7-b) $c_N : G \times \mathbb{R}^{n_x} \to \mathbb{R}$ accounts for the final cost.

In addition, we impose

(7-c) control constraints $u_t \in \mathbb{U}_t$ for each $t = 0, \dots, N-1$,
(7-d) state constraints $g_t(q_t, x_t) \leq 0$, for each $t = 1, \dots, N$, where $g_t : G \times \mathbb{R}^{n_x} \to \mathbb{R}^{n_t}$ is a given map, and
(7-e) initial conditions $(q_0, x_0) = (\bar{q}_0, \bar{x}_0)$, where $(\bar{q}_0, \bar{x}_0)$ is a given point.

The set

$$\mathcal{A} := \Big\{ \left(\{(q_t, x_t)\}_{t=0}^{N}, \{u_t\}_{t=0}^{N-1}\right) \,\Big|\, (q_0, x_0) = (\bar{q}_0, \bar{x}_0), $$
$$g_t(q_t, x_t) \leq 0, \text{ and } u_t \in \mathbb{U}_t \Big\}$$

is termed the set of *admissible state–action trajectories*.

Assumption 1. The following assumptions on the various maps in (6) and (7) are enforced throughout this article:

(A-i) The maps s_t, f_t, g_t, c_t, c_N are smooth.
(A-ii) There exists an open set $\mathcal{O} \subset \mathfrak{g}$ such that:

    (a) the exponential map $e : \mathcal{O} \to e(\mathcal{O}) \subset G$ is a diffeomorphism, and
    (b) the integration step s_t is in $e(\mathcal{O})$ for all t; see Fig. 3.

(A-iii) The set of feasible control actions $\mathbb{U}_t$ is convex for each $t = 0, \dots, N-1$.

Assumption 1 is crucial, as we shall see, in order to ensure the existence of multipliers that appear in the necessary optimality conditions for the optimal control problem. In particular, (A-i) ensures the existence of a convex approximation (known as a tent Boltyanskii, 1975) of the feasible region in a neighborhood of an optimal triple $(\dot{q}_t, \dot{x}_t, \dot{u}_t)$. (A-ii) gives the local representation of admissible trajectories $\{(q_t, x_t)\}_{t=0}^{N} \in \mathcal{A}$ in a Euclidean space. This assumption naturally holds in situations in which the discrete-time dynamics are derived from an underlying continuous-time system (6), thereby transforming our optimal control problem to a Euclidean space; first order necessary conditions for optimality are thereafter obtained using Boltyanskii's method of tents (Boltyanskii, 1975). These first order necessary conditions are interpreted in terms of the (global) configuration space variables. (A-iii) leads to a pointwise non-positive condition of the gradient of the Hamiltonian over the set of feasible control actions, which is explained in detail in Section 3.

Before defining the optimal control problem (7) formally, let us introduce the geometric notions that frequently arise in this article.

Definition 2 (*Marsden & Ratiu, 1999, p. 124*)**.** Let $f : M \to \mathbb{R}$ be a function defined on a manifold M. The derivative of the function f
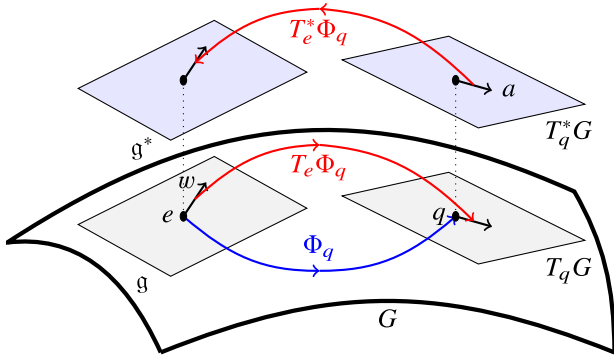
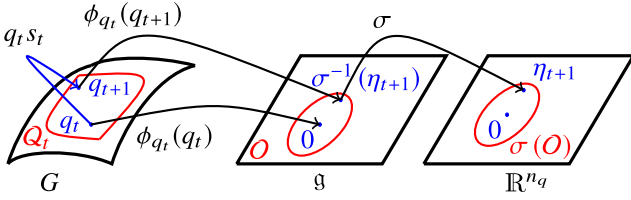**Fig. 2.** Pictorial representation of the cotangent lift of an action Φ on G.

Fig. 3. Local parametrization of q_{t+1} given q_t.

at a point $q_0 \in M$ is the map

$$T_{q_0}M \ni v \mapsto \mathcal{D}f(q_0)v := \left.\frac{d}{dt}\right|_{t=0} f(g(t)) \in \mathbb{R},$$

where $g(t)$ is a path in the manifold M with $g(0) = q_0$ and $\left.\frac{d}{dt}\right|_{t=0}g(t) = v$.

Definition 3 (*Marsden & Ratiu, 1999*, p. 173). Let $\Phi : G \times G \to G$ be a left action, so $\Phi_g : G \to G$ for all $g \in G$. The tangent lift of Φ, $T\Phi : G \times TG \to TG$ is the action

$$(g, (h, v)) \mapsto T\Phi_g(h, v) = \left(\Phi_g(h), T_h\Phi_g(v)\right), \quad (h, v) \in T_hG,$$

where $T_h\Phi_g(v) := \mathcal{D}\Phi_g(h)v$.

The cotangent lift of Φ, $T^*\Phi : G \times T^*G \to T^*G$ is the action

$$(g, (h, a)) \mapsto T^*\Phi_g(h, a) = \left(\Phi_g(h), T^*_{\Phi_g(h)}\Phi_{g^{-1}}(a)\right).$$

In particular, if we choose $g = q^{-1}$ and $h = q$ then for $a \in T_q^*G$,

$$\left\langle T_e^*\Phi_q(a), w \right\rangle := \left\langle a, T_e\Phi_q(w) \right\rangle \quad \text{for all } w \in \mathfrak{g}.$$

Note that the map $T_e^*\Phi_q : T_q^*G \to \mathfrak{g}^*$ is known as the cotangent left trivialization, see Fig. 2.

Definition 4 (*Marsden & Ratiu, 1999*, p. 311). The Adjoint action of G on $\mathfrak{g}$ is

$$G \times \mathfrak{g} \ni (g, \beta) \mapsto \mathrm{Ad}_g\beta := \left.\frac{d}{ds}\right|_{s=0} ge^{s\beta}g^{-1} \in \mathfrak{g}.$$

The Co-Adjoint action of $g \in G$ on $\mathfrak{g}^*$ is the dual of the adjoint action of g^{-1} on $\mathfrak{g}$, i.e.,

$$G \times \mathfrak{g}^* \ni (g, a) \mapsto \mathrm{Ad}^*(g, a) = \mathrm{Ad}^*_{g^{-1}}a \in \mathfrak{g}^*,$$

where

$$\left\langle \mathrm{Ad}^*_{g^{-1}}a, \beta \right\rangle = \left\langle a, \mathrm{Ad}_{g^{-1}}\beta \right\rangle$$

for all $\beta \in \mathfrak{g}, a \in \mathfrak{g}^*$.

Hereinafter we let $[N]$ denote the set of all integers from zero to N in increasing order. Now we shall proceed to define the optimal control problem (7) in a mathematical form, and derive the first order necessary conditions for optimality for the optimal control problem.

Collecting the definitions from above, our optimal control problem stands as:

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \mathscr{J}(\mathbf{q}, \mathbf{x}, \mathbf{u}) := \sum_{t=0}^{N-1} c_t(q_t, x_t, u_t) + c_N(q_N, x_N)$$

$$\text{subject to} \quad \begin{cases} \begin{cases} q_{t+1} = q_t s_t(q_t, x_t) \\ x_{t+1} = f_t(q_t, x_t, u_t) \quad \text{for each } t \in [N-1], \\ u_t \in \mathbb{U}_t \end{cases} \\ g_t(q_t, x_t) \leq 0 \quad \text{for each } t = 1, \ldots, N, \\ (q_0, x_0) = (\bar{q}_0, \bar{x}_0), \\ \text{Assumption 1.} \end{cases} \tag{8}$$

Our main result is the following:

Theorem 5 (*Discrete-time PMP*). Let $\{\mathring{u}_t\}_{t=0}^{N-1}$ be an optimal controller that solves the problem (8) with $\{(\mathring{q}_t, \mathring{x}_t)\}_{t=0}^{N}$ the corresponding state trajectory. Define the Hamiltonian

$$[N-1] \times \mathfrak{g}^* \times \left(\mathbb{R}^{n_x}\right)^* \times G \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \ni (\tau, \zeta, \xi, q, x, u) \mapsto$$

$$H^\nu(\tau, \zeta, \xi, q, x, u) := \nu c_\tau(q, x, u) + \left\langle \zeta, e^{-1}(s_\tau(q, x)) \right\rangle_\mathfrak{g}$$

$$+ \left\langle \xi, f_\tau(q, x, u) \right\rangle \in \mathbb{R}, \tag{9}$$

for $\nu \in \mathbb{R}$. There exist

- an adjoint trajectory $\left\{(\zeta^t, \xi^t)\right\}_{t=0}^{N-1} \subset \mathfrak{g}^* \times (\mathbb{R}^{n_x})^*$, covectors $\mu^t \in (\mathbb{R}^{n_t})^*$ for $t = 1, \ldots, N$, and
- a scalar $\nu \in \{-1, 0\}$

such that, with

$$\mathring{\gamma}_t := \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t\right)$$

and

$$\rho^t := \left(\mathcal{D}e^{-1}(\mathring{q}_t^{-1}\mathring{q}_{t+1}) \circ T_e\Phi_{\mathring{q}_t^{-1}\mathring{q}_{t+1}}\right)^*(\zeta^t),$$

the following hold:

(MP-i) *state and adjoint system dynamics*

$$\text{state} \begin{cases} \mathring{q}_{t+1} = \mathring{q}_t e^{\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}, \\ \mathring{x}_{t+1} = \mathcal{D}_\xi H^\nu(\mathring{\gamma}_t), \end{cases}$$

$$\text{adjoint} \begin{cases} \rho^{t-1} = \mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}}\rho^t \\ \qquad + T_e^*\Phi_{\mathring{q}_t}\left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t)\right) \\ \qquad + T_e^*\Phi_{\mathring{q}_t}\left(\mu^t \mathcal{D}_q g_t(\mathring{q}_t, \mathring{x}_t)\right), \\ \xi^{t-1} = \mathcal{D}_x H^\nu(\mathring{\gamma}_t) + \mu^t \mathcal{D}_x g_t(\mathring{q}_t, \mathring{x}_t), \end{cases}$$

(MP-ii) *transversality conditions*

$$\rho^{N-1} = T_e^*\Phi_{\mathring{q}_N}\left(\nu\mathcal{D}_q c_N(\mathring{q}_N, \mathring{x}_N)\right)$$

$$+ T_e^*\Phi_{\mathring{q}_N}\left(\mu^N \mathcal{D}_q g_N(\mathring{q}_N, \mathring{x}_N)\right),$$

$$\xi^{N-1} = \nu\mathcal{D}_x c_N(\mathring{q}_N, \mathring{x}_N) + \mu^N \mathcal{D}_x g_N(\mathring{q}_N, \mathring{x}_N),$$

(MP-iii) *Hamiltonian non-positive gradient condition*

$$\left\langle \mathcal{D}_u H^\nu(\mathring{\gamma}_t), w - \mathring{u}_t \right\rangle \leq 0 \quad \text{for all } w \in \mathbb{U}_t,$$

(MP-iv) *complementary slackness conditions*

$$\mu_j^t g_t^j(\mathring{q}_t, \mathring{x}_t) = 0 \quad \text{for all } j = 1, \ldots, n_t,$$

$$\text{and } t = 1, \ldots, N,$$

(MP-v) *non-positivity condition*

$$\mu^t \leq 0 \quad \text{for all} \quad t = 1, \ldots, N,$$

(MP-vi) *non-triviality condition*
adjoint variables $\left\{ \left(\zeta^t, \xi^t \right) \right\}_{t=0}^{N-1}$, *covectors* $\left\{ \mu^t \right\}_{t=1}^{N}$, *and the scalar ν do not simultaneously vanish.*

We present a proof of Theorem 5 in Section 3. This discrete-time PMP on matrix Lie groups is a generalization of the standard discrete-time PMP on Euclidean spaces since the variable q in the combined state (q, x) evolves on the Lie group G; consequently, the assertions of Theorem 5 appear different from the discrete-time PMP on Euclidean spaces (Boltyanskii, 1975). Let us highlight some of its features:

- The adjoint system of the discrete-time PMP on matrix Lie groups corresponding to the states q evolves on the dual $\mathfrak{g}^*$ of the Lie algebra $\mathfrak{g}$ despite the fact that the state dynamics (6) evolves on the Lie group G.
- The adjoint system is linear in the adjoint variables (ζ^t, ξ^t) because the maps $\mathfrak{g}^* \ni \zeta^t \mapsto \left(\mathcal{D}e^{-1}(\mathring{q}_t^{-1}\mathring{q}_{t+1}) \circ T_e \Phi_{\mathring{q}_t^{-1}\mathring{q}_{t+1}} \right)^* (\zeta^t) \in \mathfrak{g}$, $\mathfrak{g}^* \ni \rho^t \mapsto \mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}} \left(\rho^t \right) \in \mathfrak{g}^*$ and $T_q^* G \ni \mathcal{D}_q H^\nu(\mathring{\gamma}_t) \mapsto T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t) \right) \mathfrak{g}^*$ are linear for all t.
- The assertions the "Hamiltonian non-positive gradient condition", the "complementary slackness condition", the "non-positivity condition", and the "non-triviality condition" are identical to the discrete-time PMP on Euclidean spaces.

Remark 6.

(a) Assumption 1 is not the most general set of hypotheses for which we can derive a discrete-time PMP on matrix Lie groups. Continuous differentiability of the functions s_t, f_t, g_t, c_t, c_N suffices for Theorem 5 to hold.

(b) Due to certain fundamental differences between discrete-time and continuous-time optimal control problems, the standard Hamiltonian maximization condition in continuous time (Sussmann, 1998, Theorem MP on p. 14) does not carry over to the discrete-time version. For a detailed discussion, see e.g., Pshenichnyi (1971, p. 199). We do, however, get a weaker version contained in assertion (MP-iii) of Theorem 5. If the Hamiltonian is concave in control u over the set of feasible control actions $\mathbb{U}_t$ and $\mathbb{U}_t$ is compact for each t, in addition to (A-iii), the assertion can be strengthened to

$$H^\nu \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t \right) = \max_{w \in \mathbb{U}_t} H^\nu \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, w \right).$$

In the rest of this section we apply Theorem 5 to a class of optimal control problems on Lie groups that frequently arise in engineering applications, and derive the corresponding first order necessary optimality conditions.

2.1. Problem 1

Consider the version of (8) in which the final conditions (q_N, x_N) are constrained to take values in an immersed submanifold M_{fin} in $G \times \mathbb{R}^{n_x}$. Let us define the optimal control problem:

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \mathscr{J}(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{u}) := \sum_{t=0}^{N-1} c_t (q_t, x_t, u_t) + c_N (q_N, x_N)$$

subject to
$$\begin{cases} \begin{cases} q_{t+1} = q_t s_t (q_t, x_t) \\ x_{t+1} = f_t (q_t, x_t, u_t) \quad \text{for all } t \in [N-1], \\ u_t \in \mathbb{U}_t \end{cases} \\ g_t (q_t, x_t) \leq 0 \quad \text{for all} \quad t = 1, \ldots, N, \\ (q_0, x_0) \in (\bar{q}_0, \bar{x}_0), \\ (q_N, x_N) \in M_{\text{fin}}, \\ \text{Assumption 1.} \end{cases} \quad (10)$$

The first order necessary conditions for optimality for (10) are given by:

Corollary 7. *Let $\{\mathring{u}_t\}_{t=0}^{N-1}$ be an optimal controller that solves the problem* (10) *with $\{(\mathring{q}_t, \mathring{x}_t)\}_{t=0}^{N}$ the corresponding state trajectory. For the Hamiltonian defined in* (9), *there exist*

- *an adjoint trajectory* $\left\{ \left(\zeta^t, \xi^t \right) \right\}_{t=0}^{N-1} \subset \mathfrak{g}^* \times (\mathbb{R}^{n_x})^*$, *covectors* $\mu^t \in (\mathbb{R}^{n_t})^*$ *for $t = 1, \ldots, N$, and*
- *a scalar $\nu \in \{-1, 0\}$*

such that, with

$$\mathring{\gamma}_t := \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t \right)$$

and

$$\rho^t := \left(\mathcal{D}e^{-1}(\mathring{q}_t^{-1}\mathring{q}_{t+1}) \circ T_e \Phi_{\mathring{q}_t^{-1}\mathring{q}_{t+1}} \right)^* (\zeta^t),$$

the following hold:

(i) (MP-i) *holds,*
(ii) *transversality conditions*

$$\left\{ \left(T_{\mathring{q}_N}^* \Phi_{\mathring{q}_N^{-1}} \left(\rho^{N-1} \right), \xi^{N-1} \right) - \nu \mathcal{D}_{(q,x)} c_N \left(\mathring{q}_N, \mathring{x}_N \right) \right.$$
$$\left. - \mu^N \mathcal{D}_{(q,x)} g_N \left(\mathring{q}_N, \mathring{x}_N \right) \right\} \perp T_{(\mathring{q}_N, \mathring{x}_N)} M_{\text{fin}},$$

(iii) (MP-iii) *holds,*
(iv) (MP-iv) *holds,*
(v) (MP-v) *holds,*
(vi) (MP-vi) *holds.*

2.2. Problem 2

Consider the version of (8) in which the boundary conditions are given and fixed, and the final cost and the state inequality constraints are absent. In other words, we have the control problem

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \mathscr{J}(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{u}) := \sum_{t=0}^{N-1} c_t (q_t, x_t, u_t)$$

subject to
$$\begin{cases} \begin{cases} q_{t+1} = q_t s_t (q_t, x_t) \\ x_{t+1} = f_t (q_t, x_t, u_t) \quad \text{for all } t \in [N-1], \\ u_t \in \mathbb{U}_t \end{cases} \\ (q_0, x_0) = (\bar{q}_0, \bar{x}_0), \\ (q_N, x_N) = (\bar{q}_N, \bar{x}_N), \\ \text{Assumption 1.} \end{cases} \quad (11)$$

The first order necessary conditions for optimality for (11) are given by:

Corollary 8. *Let $\{\mathring{u}_t\}_{t=0}^{N-1}$ be an optimal controller that solves the problem* (11) *with $\{(\mathring{q}_t, \mathring{x}_t)\}_{t=0}^{N}$ the corresponding state trajectory. For the Hamiltonian defined in* (9), *there exist*

- *an adjoint trajectory* $\left\{ \left(\zeta^t, \xi^t \right) \right\}_{t=0}^{N-1} \subset \mathfrak{g}^* \times (\mathbb{R}^{n_x})^*$ *and*
- *a scalar $\nu \in \{-1, 0\}$*

such that, with

$$\mathring{\gamma}_t := \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t \right)$$

and

$$\rho^t := \left(\mathcal{D}e^{-1}(\mathring{q}_t^{-1}\mathring{q}_{t+1}) \circ T_e \Phi_{\mathring{q}_t^{-1}\mathring{q}_{t+1}} \right)^* (\zeta^t),$$

the following hold:

(i) *state and adjoint system dynamics*

$$state \begin{cases} \mathring{q}_{t+1} = \mathring{q}_t e^{\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}, \\ \mathring{x}_{t+1} = \mathcal{D}_\xi H^\nu(\mathring{\gamma}_t), \end{cases}$$

$$adjoint \begin{cases} \rho^{t-1} = \mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}} \rho^t + T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t) \right), \\ \xi^{t-1} = \mathcal{D}_x H^\nu(\mathring{\gamma}_t), \end{cases}$$

(ii) (MP-iii) *holds,*

(iii) *non-triviality condition*
adjoint variables $\left\{ (\zeta^t, \xi^t) \right\}_{t=0}^{N-1}$, *and the scalar* ν *do not simultaneously vanish.*

2.3. Problem 3

Consider the version of (8) with fixed boundary conditions, without state inequality constraints, and without the final cost. Let us consider the case in which the integration step $s_t \in G$ of the discrete-time evolution is related to the states (q_t, x_t) by an implicit equation $v_t(s_t, q_t, x_t) = 0$ such that the map $v_t(\cdot, q_t, x_t) : O_e \to \mathbb{R}^{n_q}$ is a diffeomorphism for all admissible trajectories, i.e., $\{(q_t, x_t)\}_{t=0}^N \in \mathcal{A}$, where O_e is a neighborhood of the group identity e in G. The optimal control problem can be defined as follows:

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \mathscr{J}(\boldsymbol{s}, \boldsymbol{q}, \boldsymbol{x}, \boldsymbol{u}) := \sum_{t=0}^{N-1} c_t(s_t, q_t, x_t, u_t)$$

$$\text{subject to} \quad \begin{cases} \begin{cases} q_{t+1} = q_t s_t \\ v_t(s_t, q_t, x_t) = 0 \\ x_{t+1} = f_t(s_t, q_t, x_t, u_t) \\ u_t \in \mathbb{U}_t \end{cases} \text{for all } t \in [N-1], \quad (12) \\ (q_0, x_0) = (\bar{q}_0, \bar{x}_0), \\ (q_N, x_N) = (\bar{q}_N, \bar{x}_N), \\ \text{Assumption 1.} \end{cases}$$

The first order necessary conditions for optimality for (12) are given by:

Corollary 9. *Let* $\{\mathring{u}_t\}_{t=0}^{N-1}$ *be an optimal controller that solves the problem* (12) *with* $\{(\mathring{q}_t, \mathring{x}_t)\}_{t=0}^N$ *the corresponding state trajectory and* $\{\mathring{s}_t\}_{t=0}^{N-1} \subset G$ *such that* $v_t(\mathring{s}_t, \mathring{q}_t, \mathring{x}_t) = 0$ *for* $t \in [N-1]$. *Define the Hamiltonian function*

$$[N-1] \times \mathfrak{g}^* \times \left(\mathbb{R}^{n_x}\right)^* \times G \times G \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u}$$
$$\ni (\tau, \zeta, \xi, s, q, x, u) \mapsto H^\nu(\tau, \zeta, \xi, s, q, x, u)$$
$$:= \nu c_\tau(s, q, x, u) + \left\langle \zeta, e^{-1}(s) \right\rangle_{\mathfrak{g}} + \left\langle \xi, f_\tau(s, q, x, u) \right\rangle \in \mathbb{R} \quad (13)$$

for $\nu \in \mathbb{R}$. *There exist*

- *an adjoint trajectory* $\left\{ (\zeta^t, \xi^t) \right\}_{t=0}^{N-1} \subset \mathfrak{g}^* \times (\mathbb{R}^{n_x})^*$, *and*
- *a scalar* $\nu \in \{-1, 0\}$

such that, with

$$\mathring{\gamma}_t := \left(t, \zeta^t, \xi^t, \mathring{s}_t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t \right), \quad \mathring{v}_t := v_t(\mathring{s}_t, \mathring{q}_t, \mathring{x}_t)$$

and

$$\rho^t := \left(\mathcal{D}e^{-1}(\mathring{q}_t^{-1}\mathring{q}_{t+1}) \circ T_e \Phi_{\mathring{q}_t^{-1}\mathring{q}_{t+1}} \right)^*(\zeta^t),$$

the following hold:

(i) *state and adjoint system dynamics*

$$state \begin{cases} \mathring{q}_{t+1} = \mathring{q}_t e^{\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}, \quad 0 = v_t(\mathring{s}_t, \mathring{q}_t, \mathring{x}_t), \\ \mathring{x}_{t+1} = \mathcal{D}_\xi H^\nu(\mathring{\gamma}_t), \end{cases}$$

$$adjoint \begin{cases} \rho^{t-1} = \mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}} \rho^t + T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t) \right) \\ \qquad\quad - T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_s H^\nu(\mathring{\gamma}_t) \circ \mathcal{D}_s \mathring{v}_t^{-1} \circ \mathcal{D}_q \mathring{v}_t \right), \\ \xi^{t-1} = -\mathcal{D}_s H^\nu(\mathring{\gamma}_t) \circ \mathcal{D}_s \mathring{v}_t^{-1} \circ \mathcal{D}_x \mathring{v}_t, \\ \qquad\quad + \mathcal{D}_x H^\nu(\mathring{\gamma}_t), \end{cases}$$

(ii) *Hamiltonian non-positive gradient condition*

$$\langle \mathcal{D}_u H^\nu(\mathring{\gamma}_t), w - \mathring{u}_t \rangle \le 0 \quad \text{for all } w \in \mathbb{U}_t,$$

(iii) *non-triviality condition*
adjoint variables $\left\{ (\zeta^t, \xi^t) \right\}_{t=0}^{N-1}$, *and the scalar* ν *do not simultaneously vanish.*

3. Proof of the maximum principle (Theorem 5)

Sketch of proof: We present our proof via the steps below:

Step (I) We prove the existence of a local parametrization of the Lie group G and define the optimal control problem (8) in local coordinates.

Step (II) First order necessary conditions for the optimal control problem defined in local coordinates are derived using the method of tents (Boltyanskii et al., 1999).

Step (III) The first order necessary conditions derived in Step (II) are represented in configuration space variables.

Step (IV) We prove that the first order necessary conditions derived in Step (III) are independent of the choice of the coordinate system.

Henceforth $\left(\{(\mathring{q}_t, \mathring{x}_t)\}_{t=0}^N, \{\mathring{u}_t\}_{t=0}^{N-1} \right)$ denotes an optimal state–action trajectory.

3.1. Step (I). Local parametrization of the Lie group G

Let us define the following local parametrization of the Lie group G induced by the exponential map.

Fact 10. *If* (A-ii) *holds, then for* $\mathcal{Q}_t := \left\{ \Phi_{q_t}(s) \mid s \in e(\mathcal{O}) \right\}$ *and* $\Phi_{q_t}(s) := q_t s$ *for all* $s \in G$, *the map* $\phi_{q_t} := \left(\Phi_{q_t} \circ e \right)^{-1} : \mathcal{Q}_t \to \mathcal{O} \subset \mathfrak{g}$ *provides a unique representation of* $q_{t+1} \in G$ *on the Lie algebra* $\mathfrak{g}$ *for a given* $q_t \in G$ *for all* $t \in [N-1]$.

With the help of this local parametrization, we define the dynamics evolving on the Lie group G in local coordinates. The Lie algebra $\mathfrak{g}$ of the matrix Lie group G is a finite dimensional vector space (Lee, 2013, Theorem 8.37). Therefore, there exists a linear homeomorphism

$$\sigma : \mathfrak{g} \to \mathbb{R}^{n_q},$$

where n_q is the dimension of the Lie algebra. To compress the notation we define

$$\boldsymbol{q} := (q_0, \ldots, q_N) \in G^{N+1}, \quad \boldsymbol{x} := (x_0, \ldots, x_N) \in \left(\mathbb{R}^{n_x}\right)^{N+1},$$

and $\quad \boldsymbol{u} := (u_0, \ldots, u_{N-1}) \in \left(\mathbb{R}^{n_u}\right)^N$.

Let us define the product manifold

$$\mathcal{M} := G^{N+1} \times \left(\mathbb{R}^{n_x}\right)^{N+1} \times \left(\mathbb{R}^{n_u}\right)^N$$

such that the state–action trajectory is a point on $\mathcal{M}$, i.e.,

$$(\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{u}) = (q_0, \ldots, q_N, x_0, \ldots, x_N, u_0, \ldots, u_{N-1}) \in \mathcal{M}.$$

In order to translate the optimal control problem (8) to a Euclidean space, we need to define a diffeomorphism from an open subset of a Euclidean space to an open subset of the product manifold $\mathcal{M}$ such that the state–action trajectories lie in the image of that diffeomorphism.

For a fixed $\bar{q}_0 \in G$, $\boldsymbol{\beta} := (\beta_0, \ldots, \beta_N) \in \left(\mathbb{R}^{n_q}\right)^{N+1}$, $\Lambda := \left(\sigma(\mathcal{O})\right)^{N+1} \times \left(\mathbb{R}^{n_x}\right)^{N+1} \times \left(\mathbb{R}^{n_u}\right)^{N} \subset \left(\mathbb{R}^{n_q}\right)^{N+1} \times \left(\mathbb{R}^{n_x}\right)^{N+1} \times \left(\mathbb{R}^{n_u}\right)^{N}$ and $\psi_t(\boldsymbol{\beta}) := \bar{q}_0 e^{\sigma^{-1}(\beta_0)} \cdots e^{\sigma^{-1}(\beta_t)}$ for $t \in [N]$; let us define the map

$$\Lambda \ni (\beta_0, \ldots, \beta_N, \boldsymbol{x}, \boldsymbol{u}) \mapsto \Psi(\beta_0, \ldots, \beta_N, \boldsymbol{x}, \boldsymbol{u})$$
$$:= (\psi_0(\boldsymbol{\beta}), \ldots, \psi_N(\boldsymbol{\beta}), \boldsymbol{x}, \boldsymbol{u}) \in \Psi(\Lambda). \tag{14}$$

Observe that the map Ψ is a smooth bijection, and the inverse map is given by

$$\Psi(\Lambda) \ni (\alpha_0, \alpha_1, \ldots, \alpha_N, \boldsymbol{x}, \boldsymbol{u}) \mapsto \Psi^{-1}(\alpha_0, \alpha_1, \ldots, \alpha_N, \boldsymbol{x}, \boldsymbol{u})$$
$$= \left(\left(\sigma \circ e^{-1}\right)\left(\bar{q}_0^{-1}\alpha_0\right), \left(\sigma \circ e^{-1}\right)\left(\alpha_0^{-1}\alpha_1\right), \right.$$
$$\left. \ldots, \left(\sigma \circ e^{-1}\right)\left(\alpha_{N-1}^{-1}\alpha_N\right), \boldsymbol{x}, \boldsymbol{u} \right) \in \Lambda. \tag{15}$$

Since the inverse map Ψ^{-1} is also smooth, Ψ is a diffeomorphism.

Claim 11. *State–action trajectories corresponding to all admissible control actions and starting at $\bar{q}_0$ lie in the image of Ψ.*

Proof. See Phogat, Chatterjee, and Banavar (2018, Appendix B.1). □

We now employ the diffeomorphism Ψ to translate the optimal control problem (8) to the open subset Λ as: for $\boldsymbol{\eta} := (\eta_0, \ldots, \eta_N) \in \left(\mathbb{R}^{n_q}\right)^{N+1}$,

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \tilde{\mathscr{J}}(\boldsymbol{\eta}, \boldsymbol{x}, \boldsymbol{u}) = \sum_{t=0}^{N-1} c_t(\psi_t(\boldsymbol{\eta}), x_t, u_t)$$
$$+ c_N(\psi_N(\boldsymbol{\eta}), x_N)$$

$$\text{subject to} \quad \begin{cases} \psi_t(\boldsymbol{\eta}) := \bar{q}_0 e^{\sigma^{-1}(\eta_0)} \cdots e^{\sigma^{-1}(\eta_t)} \quad \text{for } t \in [N], \\ \begin{cases} \eta_{t+1} = \left(\sigma \circ e^{-1} \circ s_t\right)(\psi_t(\boldsymbol{\eta}), x_t) \\ x_{t+1} = f_t(\psi_t(\boldsymbol{\eta}), x_t, u_t) \\ u_t \in \mathbb{U}_t \end{cases} \\ \text{for } t \in [N-1], \\ g_t(\psi_t(\boldsymbol{\eta}), x_t) \leq 0 \quad \text{for } t = 1, \ldots, N, \\ (\eta_0, x_0) = (0, \bar{x}_0). \end{cases} \tag{16}$$

3.2. Step (II). Necessary optimality conditions in local coordinates

In Step (I) we have distilled the optimal control problem (16) from (8) such that (16) is defined on a Euclidean space. We apply first order necessary conditions for optimality for constrained optimal control problems on Euclidean spaces derived via method of tents in Boltyanskii (1975) to (16).

Remark 12. Note that even though (16) has been distilled from (8) and is a well-defined problem on a Euclidean space, the standard discrete-time PMP does not apply to (16) on account of the fact that the "system dynamics" has memory, i.e., $\eta_{t+1} = \left(\sigma \circ e^{-1} \circ s_t\right)(\psi_t(\boldsymbol{\eta}), x_t)$ for $t \in [N-1]$. In other words the system in (16) is in a non-standard form. It turns out that first order necessary conditions do lead to a PMP for (16) once we lift back the necessary conditions to the configuration space.

We convert the optimal control problem (16) into a relative extremum problem in a higher-dimensional Euclidean space. Let

$$\boldsymbol{z} := (\boldsymbol{\eta}, \boldsymbol{x}, \boldsymbol{u}) = \left(\eta_0^\top, \ldots, \eta_N^\top, x_0^\top, \ldots, x_N^\top, u_0^\top, \ldots, u_{N-1}^\top\right)^\top \tag{17}$$

be the stacked vector of states and controls corresponding to (16); clearly,

$$\boldsymbol{z} \in \mathbb{R}^m \quad \text{where} \quad m = (n_x + n_q)(N+1) + n_u N.$$

Let us define the admissible action set $\mathbb{U}_t$ in terms of $\boldsymbol{z} \in \mathbb{R}^m$ as

$$\Omega_u^t := \left(\mathbb{R}^{n_q}\right)^{N+1} \times \left(\mathbb{R}^{n_x}\right)^{N+1} \times \underbrace{\mathbb{R}^{n_u} \cdots \overbrace{\mathbb{U}_t}^{(t+1)\text{th factor}} \cdots \mathbb{R}^{n_u}}_{N \text{ factors}} \subset \mathbb{R}^m$$

for $t \in [N-1]$, and the set of initial conditions as

$$\Omega_B := \{0\} \times \left(\mathbb{R}^{n_q}\right)^N \times \{\bar{x}_0\} \times \left(\mathbb{R}^{n_x}\right)^N \times \left(\mathbb{R}^{n_u}\right)^N \subset \mathbb{R}^m.$$

We define the cost function, the dynamics, and the state constraints in terms of $\boldsymbol{z}$ as:

$$C(\boldsymbol{z}) := \sum_{t=0}^{N-1} c_t(\psi_t(\boldsymbol{\eta}), x_t, u_t) + c_N(\psi_N(\boldsymbol{\eta}), x_N),$$

$$S_t(\boldsymbol{z}) := -\eta_{t+1} + \left(\sigma \circ e^{-1}\right) \circ s_t(\psi_t(\boldsymbol{\eta}), x_t),$$

$$F_t(\boldsymbol{z}) := -x_{t+1} + f_t(\psi_t(\boldsymbol{\eta}), x_t, u_t),$$

$$G_t^j(\boldsymbol{z}) := g_t^j(\psi_t(\boldsymbol{\eta}), x_t) \quad \text{for } j = 1, \ldots, n_t.$$

The optimal control problem (16) can now be defined in the augmented space $\mathbb{R}^m$ as

$$\underset{\boldsymbol{z}}{\text{minimize}} \quad C(\boldsymbol{z})$$

$$\text{subject to} \quad \begin{cases} S_t^j(\boldsymbol{z}) = 0 \quad \text{for } t \in [N-1], \ j = 1, \ldots, n_q, \\ F_t^j(\boldsymbol{z}) = 0 \quad \text{for } t \in [N-1], \ j = 1, \ldots, n_x, \\ G_t^j(\boldsymbol{z}) \leq 0 \quad \text{for } t = 1, \ldots, N, \ j = 1, \ldots, n_t, \\ \boldsymbol{z} \in \left(\bigcap_{t=0}^{N-1} \Omega_u^t\right) \cap \Omega_B. \end{cases} \tag{18}$$

Let us define

$$\mathring{\boldsymbol{z}} := (\mathring{\boldsymbol{\eta}}, \mathring{\boldsymbol{x}}, \mathring{\boldsymbol{u}}) = \Psi(\mathring{\boldsymbol{q}}, \mathring{\boldsymbol{x}}, \mathring{\boldsymbol{u}}) \in \mathbb{R}^m,$$

where $(\mathring{\boldsymbol{q}}, \mathring{\boldsymbol{x}}, \mathring{\boldsymbol{u}}) \in \mathcal{M}$ is an optimal state–action trajectory. Then the point $\mathring{\boldsymbol{z}}$ is a solution of the optimization problem (18). The necessary conditions for optimality for an optimization problem are defined in terms of dual cones. Before deriving the necessary conditions for optimality for the optimization problem (18), we provide a quick revision to cones and their dual cones.

Recall that the dual cone K^+ of a cone[4] K in $\mathbb{R}^m$ is a convex cone such that every vector in K^+ makes an acute angle with every element of K, i.e.,

$$K^+ := \left\{ \rho \in \mathbb{R}^m \mid \langle \rho, \boldsymbol{z} \rangle \geq 0 \text{ for all } \boldsymbol{z} \in K \right\}.$$

A set $K(\mathring{\boldsymbol{z}})$ in $\mathbb{R}^m$ is a cone with apex at $\mathring{\boldsymbol{z}}$ if for every $\boldsymbol{z} \in K(\mathring{\boldsymbol{z}})$, $\mathring{\boldsymbol{z}} + \alpha(\boldsymbol{z} - \mathring{\boldsymbol{z}}) \in K(\mathring{\boldsymbol{z}})$ for all $\alpha \geq 0$. Subsequently, the dual cone $K^+(\mathring{\boldsymbol{z}})$ for the cone $K(\mathring{\boldsymbol{z}})$ is defined as (Boltyanskii, 1975, p. 8)

$$K^+(\mathring{\boldsymbol{z}}) := \left\{ a \in \mathbb{R}^m \mid \langle a, \boldsymbol{z} - \mathring{\boldsymbol{z}} \rangle \geq 0, \boldsymbol{z} \in K(\mathring{\boldsymbol{z}}) \right\}. \tag{19}$$

Clearly, $K^+(\mathring{\boldsymbol{z}})$ is a closed convex cone. Equipped with the notions of dual cones, we state a theorem of Boltyanskii that provides necessary conditions for optimality for an optimization problem. This theorem is a key result for deriving necessary conditions for optimality for the optimization problem (18).

Theorem 13 (*Boltyanskii, 1975, Theorem 18, p. 22*)**.** *Let $\Omega_1, \ldots, \Omega_s$ be convex subsets of $\mathbb{R}^m$, and let C be a real-valued smooth function whose domain contains the set*

$$\Sigma = \Omega_1 \cap \ldots \cap \Omega_s \cap \Omega^*,$$

---

[4] A set $K \subset \mathbb{R}^m$ is a cone if for every $\boldsymbol{z} \in K$, $\alpha \boldsymbol{z} \in K$ for all $\alpha \geq 0$.

where

$$\Omega^* := \big\{ \boldsymbol{z} \in \mathbb{R}^m \,|\, F_1(\boldsymbol{z}) = 0, \dots, F_{r_{eq}}(\boldsymbol{z}) = 0, \\ f_1(\boldsymbol{z}) \le 0, \dots, f_{r_{iq}}(\boldsymbol{z}) \le 0 \big\},$$

and F_i, f_j are real valued smooth functions for all i, j. Let $\mathring{\boldsymbol{z}} \in \Sigma$ and let $K_i(\mathring{\boldsymbol{z}})$ be the support cone[5] of Ω_i with apex at $\mathring{\boldsymbol{z}}$. If C attains its minimum at $\mathring{\boldsymbol{z}}$ relative to Σ, then there exist scalars $\nu, \eta_F^1, \dots, \eta_F^{r_{eq}}, \eta_f^1, \dots, \eta_f^{r_{iq}}$, and vectors $a_j \in K_j^+(\mathring{\boldsymbol{z}})$ for $j = 1, \dots, s$, satisfying the following conditions:

(i) $\nu \le 0$, and if $\nu = \eta_F^1 = \dots = \eta_F^{r_{eq}} = \eta_f^1 = \dots = \eta_f^{r_{iq}} = 0$, then at least one of the vectors $a_1, \dots, a_s$ is not zero;

(ii) $\nu \mathcal{D}_{\boldsymbol{z}} C(\mathring{\boldsymbol{z}}) + \sum_{i=1}^{r_{eq}} \eta_F^i \mathcal{D}_{\boldsymbol{z}} F_i(\mathring{\boldsymbol{z}}) + \sum_{j=1}^{r_{iq}} \eta_f^j \mathcal{D}_{\boldsymbol{z}} f_j(\mathring{\boldsymbol{z}}) + a_1 + \dots + a_s = 0$;

(iii) for every $j = 1, \dots, r_{iq}$, we have

$$\eta_f^j \le 0, \quad \text{and} \quad \eta_f^j f_j(\mathring{\boldsymbol{z}}) = 0.$$

To apply Theorem 13 to the optimization problem (18), we define support cones of the convex sets $\Omega_u^0, \dots, \Omega_u^{N-1}, \Omega_B$ at $\mathring{\boldsymbol{z}}$. Let us derive the support cone $K_u^t(\mathring{\boldsymbol{z}})$ of the set Ω_u^t with apex at $\mathring{\boldsymbol{z}} \in \Omega_u^t$ for $t \in [N-1]$ as

$$K_u^t(\mathring{\boldsymbol{z}}) := \text{closure}\Big(\bigcup_{\alpha > 0} \big\{ \mathring{\boldsymbol{z}} + \alpha(\boldsymbol{z} - \mathring{\boldsymbol{z}}) \mid \boldsymbol{z} \in \Omega_u^t \big\} \Big)$$

$$= \text{closure}\Big(\big(\mathbb{R}^{n_q}\big)^{N+1} \times \big(\mathbb{R}^{n_x}\big)^{N+1} \times \underbrace{\mathbb{R}^{n_u} \cdots \overbrace{\mathcal{U}_t}^{(t+1)\text{th factor}} \cdots \mathbb{R}^{n_u}}_{N \text{ factors}} \Big),$$

where $\mathcal{U}_t := \big\{ \bigcup_{\alpha > 0} \{ \mathring{u}_t + \alpha(u_t - \mathring{u}_t) \mid u_t \in \mathbb{U}_t \} \big\}$ is the collection of rays emanating from $\mathring{u}_t$ and passing through $u_t \in \mathbb{U}_t$. Consequently, the support cone $K_u^t(\mathring{\boldsymbol{z}})$ is defined by

$$K_u^t(\mathring{\boldsymbol{z}}) = \big(\mathbb{R}^{n_q}\big)^{N+1} \times \big(\mathbb{R}^{n_x}\big)^{N+1} \times \underbrace{\mathbb{R}^{n_u} \cdots \overbrace{Q_u^t(\mathring{u}_t)}^{(t+1)\text{th factor}} \cdots \mathbb{R}^{n_u}}_{N \text{ factors}},$$

where

$$Q_u^t(\mathring{u}_t) := \text{closure}\Big(\bigcup_{\alpha > 0} \{ \mathring{u}_t + \alpha(u_t - \mathring{u}_t) \mid u_t \in \mathbb{U}_t \} \Big)$$

is the support cone of the convex set $\mathbb{U}_t$ with apex at $\mathring{u}_t$. Note that the set Ω_B is an affine subspace passing through $\mathring{\boldsymbol{z}}$, and therefore the support cone $K_B(\mathring{\boldsymbol{z}})$ of the set Ω_B with apex at $\mathring{\boldsymbol{z}}$ is the set Ω_B itself, i.e.,

$$K_B(\mathring{\boldsymbol{z}}) := \Omega_B. \tag{20}$$

Subsequently, the dual cone of the support cone $K_B(\mathring{\boldsymbol{z}})$ is given by

$$K_B^+(\mathring{\boldsymbol{z}}) := \big\{ \rho \in \mathbb{R}^m \mid \langle \rho, \boldsymbol{z} - \mathring{\boldsymbol{z}} \rangle \ge 0 \text{ for all } \boldsymbol{z} \in K_B(\mathring{\boldsymbol{z}}) \big\}. \tag{21}$$

We know from (16) that $\big(\mathring{\eta}_0, \mathring{x}_0\big) = \big(0, \bar{x}_0\big)$, and therefore from (20), it is easy to conclude that

$$K_B(\mathring{\boldsymbol{z}}) = \mathring{\boldsymbol{z}} + K_B,$$

where

$$K_B := \{0\} \times \big(\mathbb{R}^{n_q}\big)^N \times \{0\} \times \big(\mathbb{R}^{n_x}\big)^N \times \big(\mathbb{R}^{n_u}\big)^N.$$

For $v := \boldsymbol{z} - \mathring{\boldsymbol{z}} \in \mathbb{R}^m$, the dual cone (21) is rewritten as

$$K_B^+(\mathring{\boldsymbol{z}}) = \big\{ \rho \in \mathbb{R}^m \mid \langle \rho, v \rangle \ge 0 \text{ for all } v \in K_B \big\}.$$

Note that K_B is a subspace of $\mathbb{R}^m$. Therefore, the dual cone $K_B^+(\mathring{\boldsymbol{z}})$ is the orthogonal complement of the subspace K_B, i.e.,

$$K_B^+(\mathring{\boldsymbol{z}}) = \mathbb{R}^{n_q} \times \big(\{0\}\big)^N \times \mathbb{R}^{n_x} \times \big(\{0\}\big)^{2N} \subset \mathbb{R}^m.$$

Now we are in the position to apply Theorem 13 to the optimization problem (18).

By Theorem 13, if the function C attains its minimum at the point $\mathring{\boldsymbol{z}}$ then there exist scalars μ_j^t for $j = 1, \dots, n_t$ and $t = 1, \dots, N$, λ_j^t for $j = 1, \dots, n_q$ and $t \in [N-1]$, ξ_j^t for $j = 1, \dots, n_x$ and $t \in [N-1]$, vectors $b \in K_B^+(\mathring{\boldsymbol{z}})$, $a_t \in \big(K_u^t\big)^+(\mathring{\boldsymbol{z}})$ for $t \in [N-1]$, and a scalar $\nu \in \mathbb{R}$ satisfying the following conditions:

(Opt-i) $\nu \le 0$, and if $\nu = 0, \mu_j^t = 0$ for $j = 1, \dots, n_t$ and $t = 1, \dots, N$, $\lambda_j^t = 0$ for $j = 1, \dots, n_q$ and $t \in [N-1]$, $\xi_j^t = 0$ for $j = 1, \dots, n_x$ and $t \in [N-1]$, then at least one of the vectors $a_1, \dots, a_{N-1}, b$ is not zero;

(Opt-ii) $\sum_{t=0}^{N-1} \sum_{j=1}^{n_q} \lambda_j^t \mathcal{D}_{\boldsymbol{z}} S_t^j(\mathring{\boldsymbol{z}}) + \sum_{t=0}^{N-1} \sum_{j=1}^{n_x} \xi_j^t \mathcal{D}_{\boldsymbol{z}} F_t^j(\mathring{\boldsymbol{z}}) + \sum_{t=1}^{N} \sum_{j=1}^{n_t} \mu_j^t \mathcal{D}_{\boldsymbol{z}} G_t^j(\mathring{\boldsymbol{z}}) + a_0 + \dots + a_{N-1} + b + \nu \mathcal{D}_{\boldsymbol{z}} C(\mathring{\boldsymbol{z}}) = 0$;

(Opt-iii) for every $j = 1, \dots, n_t$ we have

$$\mu_j^t \le 0, \quad \mu_j^t G_t^j(\mathring{\boldsymbol{z}}) = 0 \quad \text{for all } t = 1, \dots, N.$$

Claim 14. The family of cones $K_u^0(\mathring{\boldsymbol{z}}), \dots, K_u^{N-1}(\mathring{\boldsymbol{z}}), K_B(\mathring{\boldsymbol{z}})$ is inseparable in $\mathbb{R}^m$.[6]

Proof. See Phogat et al. (2018, Appendix B.2). ∎

So, using the inseparability of this family, we arrive at stricter conditions for optimality than (Opt-i)–(Opt-ii). To compress notation, let us define covectors (row vectors) as

$$\big(\mathbb{R}^{n_t}\big)^* \ni \mu^t := \big(\mu_1^t, \dots, \mu_{n_t}^t\big) \quad \text{for} \quad t = 1, \dots, N,$$

$$\big(\mathbb{R}^{n_q}\big)^* \ni \lambda^t := \big(\lambda_0^t, \dots, \lambda_{n_q}^t\big) \quad \text{for} \quad t \in [N-1],$$

$$\big(\mathbb{R}^{n_x}\big)^* \ni \xi^t := \big(\xi_0^t, \dots, \xi_{n_x}^t\big) \quad \text{for} \quad t \in [N-1].$$

Then the optimality conditions (Opt-i)–(Opt-ii) translate to the following:

(i) Suppose that the covectors $\mu^1, \dots, \mu^N, \lambda^0, \dots, \lambda^{N-1}, \xi^0, \dots, \xi^{N-1}$, and the scalar ν are all zero. Then the optimality conditions (Opt-i)–(Opt-ii) lead to the following: there exist vectors $a_0, \dots, a_{N-1}, b$ not all zero such that

$$a_0 + \dots + a_{N-1} + b = 0.$$

This contradicts inseparability of the family of cones $K_u^0(\mathring{\boldsymbol{z}}), \dots, K_u^{N-1}(\mathring{\boldsymbol{z}}), K_B(\mathring{\boldsymbol{z}})$, and therefore (Opt-i) leads to the following stronger non-triviality condition than (Opt-i):

$\nu \le 0$, and if $\nu = 0$, then at least one of the covectors

$$\mu^1, \dots, \mu^N, \lambda^0, \dots, \lambda^{N-1}, \xi^0, \dots, \xi^{N-1} \text{ is not zero.} \tag{22}$$

(ii) Note that $\tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_u(\mathring{\boldsymbol{z}}) := \Big(\bigcap_{t=0}^{N-1} K_u^t(\mathring{\boldsymbol{z}}) \Big) \cap K_B(\mathring{\boldsymbol{z}})$. Therefore $\tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_u^t(\mathring{\boldsymbol{z}})$ for $t \in [N-1]$ and $\tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_B$. Since, $a_t \in \big(K_u^t\big)^+(\mathring{\boldsymbol{z}})$ for $t \in [N-1]$ and $b \in (K_B)^+$, using the definition of dual cone leads to the following:

$$\langle a_t, \tilde{\boldsymbol{z}} \rangle \ge 0, \quad \langle b, \tilde{\boldsymbol{z}} \rangle \ge 0 \quad \text{for } \tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_u(\mathring{\boldsymbol{z}}).$$

Consequently, using the fact that

$$\big\langle a_0 + \dots + a_{N-1} + b, \tilde{\boldsymbol{z}} \big\rangle \ge 0 \quad \text{for all } \tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_u(\mathring{\boldsymbol{z}});$$

---

[5] The support cone $K(\mathring{\boldsymbol{z}})$ of a convex set $\Omega \subset \mathbb{R}^m$ with apex at $\mathring{\boldsymbol{z}} \in \Omega$ is defined as

$$K_\Omega(\mathring{\boldsymbol{z}}) := \text{closure}\Big(\bigcup_{\alpha > 0} \{ \mathring{\boldsymbol{z}} + \alpha(\boldsymbol{z} - \mathring{\boldsymbol{z}}) \mid \boldsymbol{z} \in \Omega \} \Big).$$

For more details see Boltyanskii et al. (1999, p. 29).

[6] A family of convex cones is said to be *separable* if there exists a hyperplane that separates one of them from the intersection of the others. If the family is not *separable* then it is called *inseparable*. For more details see Janković (2002, p. 2575).

the (Opt-ii) is rewritten as

$$\left\langle \sum_{i=0}^{N-1} \lambda^i \mathcal{D}_{\boldsymbol{z}} S_i(\mathring{\boldsymbol{z}}) + \sum_{k=0}^{N-1} \xi^k \mathcal{D}_{\boldsymbol{z}} F_k(\mathring{\boldsymbol{z}}) + \sum_{t=1}^{N} \mu^t \mathcal{D}_{\boldsymbol{z}} G_t(\mathring{\boldsymbol{z}}) \right.$$
$$\left. + \nu \mathcal{D}_{\boldsymbol{z}} C(\mathring{\boldsymbol{z}}), \tilde{\boldsymbol{z}} \right\rangle \leq 0 \quad \text{for all } \tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_u(\mathring{\boldsymbol{z}}). \tag{23}$$

The variational inequality (23) gives rise to "*state and adjoint system dynamics*" (MP-i), "*transversality conditions*" (MP-ii) and "*Hamiltonian non-positive gradient condition*" (MP-iii). The optimality condition (Opt-iii) establishes "*complementary slackness conditions*" (MP-iv) and "*non-positivity conditions*" (MP-v), and the "*non-triviality condition*" (MP-vi) comes from (22). Now we shall derive these conditions (MP-i)–(MP-vi) in the configuration variables.

3.3. Step (III). Representation of the necessary conditions for optimality in terms of the configuration variables

Let us define the Hamiltonian as

$$[N-1] \times \mathfrak{g}^* \times \left(\mathbb{R}^{n_x}\right)^* \times G \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \ni (\tau, \zeta, \xi, q, x, u) \mapsto$$
$$H^\nu(\tau, \zeta, \xi, q, x, u) := \nu c_\tau(q, x, u) + \left\langle \zeta, \mathrm{e}^{-1}\left(s_\tau(q, x)\right)\right\rangle_{\mathfrak{g}}$$
$$+ \langle \xi, f_\tau(q, x, u) \rangle \in \mathbb{R},$$

with $\nu \in \mathbb{R}$.

- **Adjoint system dynamics and transversality conditions** (MP-i)–(MP-ii): In order to derive the adjoint system dynamics and the transversality conditions corresponding to the state variable x, we restrict the choice of the variable $\tilde{\boldsymbol{z}}$ in (23) to

$$K_x^t(\mathring{\boldsymbol{z}}) := \{\mathring{\boldsymbol{\eta}}\} \times \overbrace{\{\mathring{x}_0\} \cdots \times \underbrace{\mathbb{R}^{n_x}}_{(t+1)\text{th factor}} \times \cdots \{\mathring{x}_N\}}^{(N+1) \text{ factors}} \times \{\mathring{\boldsymbol{u}}\} \tag{24}$$

for $t = 1, \ldots, N$. For a fixed $t = 1, \ldots, N$, choosing the collection of $\tilde{\boldsymbol{z}}$ in (23) such that $\tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_x^t(\mathring{\boldsymbol{z}})$, leads to the following set of conditions: for

$$\zeta^t := \sigma^*\left(\lambda^t\right) \quad \text{and} \quad \mathring{\gamma}_t := \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t\right),$$

  ○ adjoint equations corresponding to the state variables x :

$$\xi^{t-1} = \mathcal{D}_x H^\nu(\mathring{\gamma}_t) + \mu^t \mathcal{D}_x g_t(\mathring{q}_t, \mathring{x}_t)$$

  for $t = 1, \ldots, N-1$,
  ○ transversality conditions corresponding to the state variables x :

$$\xi^{N-1} = \nu \mathcal{D}_x c_N(\mathring{q}_N, \mathring{x}_N) + \mu^N \mathcal{D}_x g_N(\mathring{q}_N, \mathring{x}_N).$$

Similarly, to derive the adjoint and transversality conditions corresponding to the state variable q, we restrict the choice of the variable $\tilde{\boldsymbol{z}}$ in (23) such that

$$\tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_q(\mathring{\boldsymbol{z}}),$$

where

$$K_q(\mathring{\boldsymbol{z}}) := \{0\} \times \left(\mathbb{R}^{n_q}\right)^N \times \{\mathring{\boldsymbol{x}}\} \times \{\mathring{\boldsymbol{u}}\}. \tag{25}$$

The choice of $\tilde{\boldsymbol{z}}$ in (23) such that $\tilde{\boldsymbol{z}} + \mathring{\boldsymbol{z}} \in K_q(\mathring{\boldsymbol{z}})$, leads to the following condition:

$$+ \sum_{t=1}^{N} \left\langle \mu^t \mathcal{D}_q g_t(\mathring{q}_t, \mathring{x}_t), \mathcal{D}_\eta \psi_t(\mathring{\eta}) \tilde{\boldsymbol{\eta}} \right\rangle$$
$$+ \sum_{t=0}^{N-1} \left\langle \mathcal{D}_q H^\nu(\mathring{\gamma}_t), \mathcal{D}_\eta \psi_t(\mathring{\eta}) \tilde{\boldsymbol{\eta}} \right\rangle - \left\langle \lambda^t, \tilde{\eta}_{t+1} \right\rangle \tag{26}$$
$$+ \left\langle \nu \mathcal{D}_q c_N(\mathring{q}_N, \mathring{x}_N), \mathcal{D}_\eta \psi_N(\mathring{\eta}) \tilde{\boldsymbol{\eta}} \right\rangle = 0 \quad \text{for all } \tilde{\boldsymbol{\eta}} \in K_\eta,$$

where

$$\tilde{\boldsymbol{\eta}} := (\tilde{\eta}_0, \tilde{\eta}_1, \ldots, \tilde{\eta}_N) \text{ and } K_\eta := \{0\} \times \left(\mathbb{R}^{n_q}\right)^N.$$

In order to represent (26) in configuration variable, let us establish an association between $\tilde{\boldsymbol{\eta}}$ and tangent vectors $\tilde{q}_t \in T_{\mathring{q}_t} G$ for $t \in [N]$. It is evident from the diffeomorphism (14) that

$$\tilde{q}_t := \mathcal{D}_\eta \psi_t(\mathring{\eta}) \tilde{\boldsymbol{\eta}} \in T_{\mathring{q}_t} G \quad \text{for all } t \in [N]. \tag{27}$$

Furthermore, we derive $\tilde{\boldsymbol{\eta}}$ in terms of $\tilde{\boldsymbol{q}} := \left(\tilde{q}_0, \ldots, \tilde{q}_N\right)$ via (15) as follows

$$\tilde{\eta}_t := \left.\frac{d}{ds}\right|_{s=0} \eta_t(s) := \left.\frac{d}{ds}\right|_{s=0} \left(\sigma \circ \mathrm{e}^{-1}\right)\left(q_{t-1}(s)^{-1} q_t(s)\right) \tag{28}$$

for $t \in [N]$, where $q_{-1}(s) \equiv \mathring{q}_0$, $\eta_t(0) = \mathring{\eta}_t$, and with $\chi_t := T_{\mathring{q}_t} \Phi_{\mathring{q}_t^{-1}}\left(\tilde{q}_t\right) \in \mathfrak{g}$ for $t \in [N]$ and the map

$$\varphi_{t-1} := \mathcal{D}\mathrm{e}^{-1}\left(\mathring{q}_{t-1}^{-1} \mathring{q}_t\right) \circ T_e \Phi_{\mathring{q}_{t-1}^{-1} \mathring{q}_t}^{-1},$$

(28) simplifies to $\tilde{\eta}_0 = \sigma(\chi_0)$ and

$$\tilde{\eta}_t = -\sigma \circ \varphi_{t-1}\left(\mathrm{Ad}_{\mathring{q}_t^{-1} \mathring{q}_{t-1}}\left(\chi_{t-1}\right) - \chi_t\right) \tag{29}$$

for $t = 1, \ldots, N$. Using the fact that $\tilde{q}_t = T_e \Phi_{\mathring{q}_t}(\chi_t)$ for $t \in [N]$, and with (27) and (29), the necessary conditions (26) lead to the following conditions on the Lie algebra $\mathfrak{g}$ as

$$\sum_{t=1}^{N} \left\langle \sigma^*\left(\lambda^{t-1}\right), \varphi_{t-1}\left(\mathrm{Ad}_{\mathring{q}_t^{-1} \mathring{q}_{t-1}}\left(\chi_{t-1}\right) - \chi_t\right)\right\rangle$$
$$+ \sum_{t=0}^{N-1} \left\langle T_e^* \Phi_{\mathring{q}_t}\left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t)\right), \chi_t \right\rangle \tag{30}$$
$$+ \sum_{t=1}^{N} \left\langle T_e^* \Phi_{\mathring{q}_t}\left(\mu^t \mathcal{D}_q g_t(\mathring{q}_t, \mathring{x}_t)\right), \chi_t \right\rangle$$
$$+ \left\langle T_e^* \Phi_{\mathring{q}_N}\left(\nu \mathcal{D}_q c_N(\mathring{q}_N, \mathring{x}_N)\right), \chi_N \right\rangle = 0 \quad \text{for all } \boldsymbol{\chi} \in K_\chi,$$

where

$$\boldsymbol{\chi} := (\chi_0, \chi_1, \ldots, \chi_N) \quad \text{and} \quad K_\chi := \{0\} \times \mathfrak{g}^N.$$

Let us now restrict the choice of the variable $\boldsymbol{\chi}$ in (30) to

$$K_\chi^t := \overbrace{\{0\} \cdots \times \underbrace{\mathfrak{g}}_{(t+1)\text{th factor}} \times \cdots \{0\}}^{(N+1) \text{ factors}} \subset K_\chi$$

  for $t = 1, \ldots, N$,

to derive the adjoint dynamics and the transversality conditions corresponding to the states $q \in G$. For a fixed $t = 1, \ldots, N$, choosing $\chi_t \in K_\chi^t$ in (30) leads to the following set of conditions:

  ○ adjoint system corresponding to the states q :

$$\varphi_{t-1}^*(\zeta^{t-1}) = \mathrm{Ad}_{\mathrm{e}^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}}^* \circ \varphi_t^*(\zeta^t)$$
$$+ T_e^* \Phi_{\mathring{q}_t}\left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t)\right)$$
$$+ T_e^* \Phi_{\mathring{q}_t}\left(\mu^t \mathcal{D}_q g_t(\mathring{q}_t, \mathring{x}_t)\right) \tag{31}$$

  for $t = 1, \ldots, N-1$,
  ○ transversality conditions corresponding to the states q :

$$\varphi_{N-1}^*(\zeta^{N-1}) = T_e^* \Phi_{\mathring{q}_N}\left(\nu \mathcal{D}_q c_N(\mathring{q}_N, \mathring{x}_N)\right)$$
$$+ T_e^* \Phi_{\mathring{q}_N}\left(\mu^N \mathcal{D}_q g_N(\mathring{q}_N, \mathring{x}_N)\right). \tag{32}$$

- **Hamiltonian non-positive gradient condition** (MP-iii): Hamiltonian non-positive gradient condition is derived via restricting the choice of the variable $\tilde{z}$ in (23) such that

$$\tilde{z} + \mathring{z} \in \tilde{K}_u^t(\mathring{z}),$$

where

$$\tilde{K}_u^t(\mathring{z}) := \{\mathring{\eta}\} \times \{\mathring{x}\} \times \underbrace{\{\mathring{u}_0\} \cdots \times \overbrace{Q_u^t(\mathring{u}_t)}^{(t+1)\text{th factor}} \times \cdots \{\mathring{u}_{N-1}\}}_{N \text{ factors}}$$

for $t \in [N-1]$. For a fixed $t \in [N-1]$, choosing the collection of $\tilde{z}$ in (23) such that $\tilde{z} + \mathring{z} \in \tilde{K}_u^t(\mathring{z})$, leads to the following set of conditions:

$$\left\langle \mathcal{D}_u H^\nu(\mathring{\gamma}_t), \tilde{u}_t \right\rangle \leq 0 \quad \text{for all} \quad \mathring{u}_t + \tilde{u}_t \in Q_u^t(\mathring{u}_t) \tag{33}$$

for $t \in [N-1]$. Since, $\mathbb{U}_t \subset Q_u^t(\mathring{u}_t)$, (33) simplifies to the following

$$\left\langle \mathcal{D}_u H^\nu(\mathring{\gamma}_t), w - \mathring{u}_t \right\rangle \leq 0 \quad \text{for all} \quad w \in \mathbb{U}_t.$$

- **Complementary slackness conditions** (MP-iv): The complementary slackness conditions in the configuration variables follows from (Opt-iii) as:

$$\mu_j^t g_t^j(\mathring{q}_t, \mathring{x}_t) = 0 \text{ for all } j = 1, \ldots, n_t \text{ and } t = 1, \ldots, N.$$

- **Non-positivity condition** (MP-v): The non-positivity condition follows from (Opt-iii) as:

$$\mu^t \leq 0 \quad \text{for all} \quad t = 1, \ldots, N.$$

- **Non-triviality condition** (MP-vi): The variable λ^t in the non-triviality condition (Opt-i) is zero if and only if $\zeta^t := \sigma^*(\lambda^t)$ is zero because $\sigma^* : (\mathbb{R}^{n_q})^* \to \mathfrak{g}^*$ is a vector space homeomorphism. The adjoint variables $\left\{(\zeta^t, \xi^t)\right\}_{t=0}^{N-1}$, covectors $\left\{\mu^t\right\}_{t=1}^N$ and the scalar ν do not simultaneously vanish.

- **State dynamics** (MP-i): The system dynamics (6) in terms of the Hamiltonian is

$$\mathring{q}_{t+1} = \mathring{q}_t e^{\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}, \quad \mathring{x}_{t+1} = \mathcal{D}_\xi H^\nu(\mathring{\gamma}_t). \tag{34}$$

3.4. *Step (IV). The representation of the necessary conditions for optimality is coordinate free*

Let us establish that the necessary conditions derived in Section 3.3 are independent of the choice of the coordinate system. Suppose we have a curve

$$I \ni s \mapsto \mathbf{q}_t(s) \in G \quad \text{with} \quad \mathbf{q}_t(0) = \mathring{q}_t,$$

and the two different coordinate charts

$$\mathcal{X} : D^\mathcal{X} \subset G \to R^\mathcal{X} \subset \mathbb{R}^{n_q}, \quad \mathcal{Y} : D^\mathcal{Y} \subset G \to R^\mathcal{Y} \subset \mathbb{R}^{n_q}.$$

such that our curve $\mathbf{q}_t$ is contained in both domains, that is

$$\mathbf{q}_t(s) \in D^\mathcal{X} \cap D^\mathcal{Y} \quad \text{for all} \quad s \in I.$$

Suppose our two coordinate charts are $\mathcal{C}^1$ related, in the following sense:

Definition 15 (*Sussmann, 2000, Definition 2.3.1, p. 8*)**.** Let G be a manifold, and let $\mathcal{X}, \mathcal{Y}$, be two n_q- dimensional charts on G. Let k be a nonnegative integer. We say that $\mathcal{X}$ and $\mathcal{Y}$ are $\mathcal{C}^k$ related if

(1) the images $\mathcal{X}(D^\mathcal{X} \cap D^\mathcal{Y})$, $\mathcal{Y}(D^\mathcal{X} \cap D^\mathcal{Y})$ of the "overlap set" $D^\mathcal{X} \cap D^\mathcal{Y}$ under the coordinate maps $\mathcal{X}, \mathcal{Y}$, are open in $R^\mathcal{X}$, $R^\mathcal{Y}$, respectively;

(2) the "change of coordinates" maps

$$\Phi : \mathcal{X}(D^\mathcal{X} \cap D^\mathcal{Y}) \to \mathcal{Y}(D^\mathcal{X} \cap D^\mathcal{Y}),$$

$$\Psi : \mathcal{Y}(D^\mathcal{X} \cap D^\mathcal{Y}) \to \mathcal{X}(D^\mathcal{X} \cap D^\mathcal{Y}),$$

defined by the conditions

$$\begin{aligned}\Phi(X) = Y \quad &\text{whenever} \quad X = \mathcal{X}(q), Y = \mathcal{Y}(q) \\ &\text{for some} \quad q \in D^\mathcal{X} \cap D^\mathcal{Y}, \\ \Psi(Y) = X \quad &\text{whenever} \quad X = \mathcal{X}(q), Y = \mathcal{Y}(q) \\ &\text{for some} \quad q \in D^\mathcal{X} \cap D^\mathcal{Y},\end{aligned}$$

are of class $\mathcal{C}^k$.

We know that X, Y are coordinate representations of the configuration $q \in D^\mathcal{X} \cap D^\mathcal{Y}$ and $\dot{X}, \dot{Y}$ are coordinate representations of the tangent vectors $\dot{q} \in T_q G$ in the coordinate charts $\mathcal{X}, \mathcal{Y}$ respectively. Then we can express the tangent vector $\dot{Y}$(or $\dot{X}$) in one coordinate chart as a function of the configuration X(or Y) and the tangent vector $\dot{X}$(or $\dot{Y}$) in the other coordinate chart. The *transformation rules for tangent vectors* is given by the following formulae:

$$\dot{X} = \frac{\partial \Psi}{\partial Y}(Y)\dot{Y}, \quad \dot{Y} = \frac{\partial \Phi}{\partial X}(X)\dot{X}, \tag{35}$$

where $\left[\frac{\partial \Psi}{\partial Y}(Y)\right]_{ij} := \frac{\partial \Psi^{\mathbf{i}}}{\partial Y_j}(Y)$ is the Jacobian matrix.

In order to show the coordinate invariance of the adjoint system and the transversality conditions, we need to establish the transformation rules for the covectors ρ^{t-1}, $\mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}}\rho^t$ and $T_e^*\Phi_{\mathring{q}_t}\left(\mathcal{D}_q H^\nu(\mathring{\gamma}_t)\right) \in \mathfrak{g}^*$, and under that transformation rule these covectors admit unique representations in the coordinate charts $\mathcal{X}, \mathcal{Y}$. This task is accomplished in the following steps:

(1) We prove that the local representation $\dot{X}, \dot{Y}$ of $\dot{q} \in T_q G$ in the coordinate charts $\mathcal{X}, \mathcal{Y}$ respectively, has a unique representation $\chi^X, \chi^Y \in \mathfrak{g}$, the Lie algebra of the Lie group G via the tangent lift of the left action. Further, we establish the transformation rule for the vectors χ^X, χ^Y that is derived via the transformation rule of vectors $\dot{X}, \dot{Y}$ given by (35).

(2) Through the duality property, we derive the transformation rule for a local representation ρ_X, ρ_Y in the coordinate charts $\mathcal{X}, \mathcal{Y}$ respectively, of the covector $\rho \in \mathfrak{g}^*$ and establish the invariance of the adjoint equations and transversality conditions.

In the case of a Lie group G, the tangent space at a point $\mathring{q}_t \in G$ is characterized by the left (right) invariant vector fields (Sachkov, 2009), i.e.,

$$T_{\mathring{q}_t} G = \left\{ T_e \Phi_{\mathring{q}_t}(\chi) \mid \chi \in \mathfrak{g} \right\}.$$

Let χ^X, χ^Y be the coordinate representation of $\chi \in \mathfrak{g}$ in the coordinate charts $\mathcal{X}$ and $\mathcal{Y}$ respectively and therefore from (36), we know

$$T_e \Phi_{\mathring{q}_t}(\chi^X) = \frac{\partial \Psi}{\partial Y}(Y) \, T_e \Phi_{\mathring{q}_t}(\chi^Y). \tag{36}$$

Then the transformation rule for the vectors $\chi^X, \chi^Y \in \mathfrak{g}$ is derived from (35) as

$$\chi^X = \Gamma_t(\chi^Y), \quad \chi^Y = \Delta_t(\chi^X) \tag{37}$$

where

$$\Gamma_t := \left(\left(T_e \Phi_{\mathring{q}_t} \right)^{-1} \circ \frac{\partial \Psi}{\partial Y}(Y) \circ T_e \Phi_{\mathring{q}_t} \right) \text{ and } \Delta_t := \Gamma_t^{-1}$$

are linear transformations. These transformation rules of vectors in the Lie algebra $\mathfrak{g}$ induces the transformation rules for the corresponding covectors, i.e., for $\zeta_X, \zeta_Y \in \mathfrak{g}^*$, if

$$\langle \zeta_X, \chi^X \rangle = \langle \zeta_Y, \Delta_t(\chi^X) \rangle \quad \text{for all } \chi^X \in \mathfrak{g},$$

then

$$\zeta_X = \Delta_t^*(\zeta_Y) \text{ and } \zeta_Y = \Gamma_t^*(\zeta_X). \tag{38}$$

Therefore coordinate representations of the covector $T_e^* \Phi_{\dot{q}_t} \left(\mathcal{D}_q(H^\nu(\dot{\gamma}_t) + \mu^t g_t(\dot{q}_t, \dot{x}_t)) \right) \in \mathfrak{g}^*$ in chart $\mathcal{X}$ as $T_e^* \Phi_{\dot{q}_t^X} \left(\mathcal{D}_{q^X}(H^\nu(\dot{\gamma}_t) + \mu^t g_t(\dot{q}_t, \dot{x}_t)) \right)$ and in charts $\mathcal{Y}$ as $T_e^* \Phi_{\dot{q}_t^Y} \left(\mathcal{D}_{q^Y}(H^\nu(\dot{\gamma}_t) + \mu^t g_t(\dot{q}_t, \dot{x}_t)) \right)$ are related via the transformation rule (38) as

$$T_e^* \Phi_{\dot{q}_t^X} \left(\mathcal{D}_{q^X}(H^\nu(\dot{\gamma}_t) + \mu^t g_t(\dot{q}_t, \dot{x}_t)) \right)$$
$$= \Delta_t^* \left(T_e^* \Phi_{\dot{q}_t^Y} \left(\mathcal{D}_{q^Y}(H^\nu(\dot{\gamma}_t) + \mu^t g_t(\dot{q}_t, \dot{x}_t)) \right) \right). \tag{39}$$

In the similar manner, it can be established that

$$\mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_X} H^\nu(\dot{\gamma}_t)}} \rho_X^t = \Delta_t^* \left(\mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_Y} H^\nu(\dot{\gamma}_t)}} \rho_Y^t \right) \tag{40}$$

where

$$\left\langle \mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_X} H^\nu(\dot{\gamma}_t)}} \rho_X^t, \chi^X \right\rangle = \left\langle \mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_Y} H^\nu(\dot{\gamma}_t)}} \rho_Y^t, \chi^Y \right\rangle.$$

Furthermore, we derive the transformation rules for the covectors $\rho_Y^t, \rho_X^t \in \mathfrak{g}$ as

$$\left\langle \rho_X^t, \mathrm{Ad}_{e^{-\mathcal{D}_{\zeta_X} H^\nu(\dot{\gamma}_t)}} \chi^X \right\rangle = \left\langle \rho_Y^t, \mathrm{Ad}_{\dot{q}_{t-1}^{-1} \dot{q}_t} \Delta_t(\chi^X) \right\rangle$$
$$= \left\langle \rho_Y^t, \Delta_{t+1} \left(\mathrm{Ad}_{\dot{q}_{t-1}^{-1} \dot{q}_t} \chi^X \right) \right\rangle$$
$$= \left\langle \Delta_{t+1}^*(\rho_Y^t), \mathrm{Ad}_{e^{-\mathcal{D}_{\zeta_X} H^\nu(\dot{\gamma}_t)}} \chi^X \right\rangle$$

for all $\chi^X \in \mathfrak{g}$, which leads to

$$\rho_X^t = \Delta_{t+1}^*(\rho_Y^t). \tag{41}$$

Using (39)–(41), we conclude that the adjoint system (31) in coordinate charts $\mathcal{X}$ transform naturally to the adjoint system (31) in coordinate charts $\mathcal{Y}$, i.e.,

$$\rho_X^{t-1} - \mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_X} H^\nu(\dot{\gamma}_t)}} \rho_X^t$$
$$\quad - T_e^* \Phi_{\dot{q}_t} \left(\mathcal{D}_{q^X} H^\nu(\dot{\gamma}_t) + \mu^t \mathcal{D}_{q^X} g_t(\dot{q}_t, \dot{x}_t) \right)$$
$$= \Delta_t^* \left(\rho_Y^{t-1} - \mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_Y} H^\nu(\dot{\gamma}_t)}} \rho_Y^t \right.$$
$$\left. \quad - T_e^* \Phi_{\dot{q}_t} \left(\mathcal{D}_{q^Y} H^\nu(\dot{\gamma}_t) + \mu^t \mathcal{D}_{q^Y} g_t(\dot{q}_t, \dot{x}_t) \right) \right).$$

In other words,

$$\rho_X^{t-1} - \mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_X} H^\nu(\dot{\gamma}_t)}} \rho_X^t$$
$$\quad - T_e^* \Phi_{\dot{q}_t} \left(\mathcal{D}_{q^X} H^\nu(\dot{\gamma}_t) + \mu^t \mathcal{D}_{q^X} g_t(\dot{q}_t, \dot{x}_t) \right) = 0$$

if and only if

$$\rho_Y^{t-1} - \mathrm{Ad}^*_{e^{-\mathcal{D}_{\zeta_Y} H^\nu(\dot{\gamma}_t)}} \rho_Y^t$$
$$\quad - T_e^* \Phi_{\dot{q}_t} \left(\mathcal{D}_{q^Y} H^\nu(\dot{\gamma}_t) + \mu^t \mathcal{D}_{q^Y} g_t(\dot{q}_t, \dot{x}_t) \right) = 0$$

because Δ_t^* is an invertible linear transformation. Similarly, we can conclude that the transversality conditions (32) in coordinate charts $\mathcal{X}$ transform naturally to the transversality conditions (32) in coordinate charts $\mathcal{Y}$. This completes our proof.

4. Example

For illustration of our results we pick an example of energy optimal single axis maneuvers of a spacecraft. In effect, the state-space becomes $\mathbb{R} \times SO(2)$, which is isomorphic to $\mathbb{R} \times S^1$. This is a considerably elementary situation compared to general rigid body dynamics on SO(3), but it is easier to visualize and represent trajectories with figures. We adhere to this simpler setting in order not to blur the message of this article while retaining the coordinate-free nature of the problem. Let $h > 0$ be a step length, $R_t, F_t \in SO(2)$, (the set of 2×2 orthonormal matrices,) be the orientation of the spacecraft performing single axis maneuvers and the integration step at the discrete-time instant t respectively. Let $\omega_t \in \mathbb{R}$ be the momentum of the spacecraft about the axis of rotation and u_t be the control applied to the spacecraft about the rotation axis. Consider the discrete-time model of a spacecraft performing single axis maneuvers given by

$$\begin{cases} R_{t+1} = R_t F(\omega_t), \\ \omega_{t+1} = \omega_t + h u_t, \end{cases} \tag{42}$$

where

$$F(\omega) := \begin{pmatrix} \sqrt{1 - h^2 \omega^2} & -h\omega \\ h\omega & \sqrt{1 - h^2 \omega^2} \end{pmatrix}.$$

The optimal control problem is to maneuver a spacecraft from a fixed initial configuration (R_i, ω_i) to a fixed final configuration (R_f, ω_f) via a minimum energy path obeying state and action constraints simultaneously, i.e., $|\omega_t| \le d, |u_t| \le c$ for all t. The optimal control problem in discrete-time can be defined as

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \mathcal{J}(\boldsymbol{u}) := \sum_{t=0}^{N-1} \frac{u_t^2}{2}$$

$$\text{subject to} \quad \begin{cases} \begin{cases} R_{t+1} = R_t F(\omega_t) \\ \omega_{t+1} = \omega_t + h u_t \quad \text{for } t \in [N-1], \\ |u_t| \le c \end{cases} \\ \frac{1}{2} \left(\omega_t^2 - d^2 \right) \le 0 \quad \text{for } t = 1, \dots, N-1, \\ (R_0, \omega_0) = (R_i, \omega_i), \\ (R_N, \omega_N) = (R_f, \omega_f). \end{cases} \tag{43}$$

A set of first order necessary conditions for optimality for (43) is given by Corollary 7 as follows:

4.1. First order necessary conditions of optimality

Let $\hat{(\cdot)} : \mathbb{R} \to \mathfrak{so}(2)^*$ be a vector space homeomorphism. Define the Hamiltonian for the optimal control problem (43) as

$$\mathfrak{so}(2)^* \times \mathbb{R} \times SO(2) \times \mathbb{R} \times \mathbb{R} \ni \left(\hat{\zeta}, \xi, R, \omega, u \right) \mapsto \tag{44}$$

$$H^\nu(\hat{\zeta}, \xi, R, \omega, u) := \nu \frac{u^2}{2} + \left\langle \hat{\zeta}, e^{-1}(F(\omega)) \right\rangle_{\mathfrak{g}} + \xi(\omega + hu)$$
$$= \nu \frac{u^2}{2} + \zeta \sin^{-1}(h\omega) + \xi(\omega + hu) \in \mathbb{R}$$

Let $\{\mathring{u}_t\}_{t=0}^{N-1}$ be an optimal control that solves the problem (43). Then there exist a state-adjoint trajectory $\left(\left\{ \left(\hat{\zeta}^t, \xi^t \right) \right\}_{t=0}^{N-1}, \left\{ \left(\mathring{R}_t, \mathring{\omega}_t \right) \right\}_{t=0}^N \right)$ on the cotangent bundle $\mathfrak{so}(2)^* \times \mathbb{R}^* \times SO(2) \times \mathbb{R}$, covectors $\{\mu^t\}_{t=1}^N \subset \mathbb{R}^*$ and a scalar $\nu \in \{-1, 0\}$, not all zero, such that the following hold:

(i) State and adjoint system dynamics given by

$$\hat{\zeta}^{t-1} = \mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu\left(\hat{\zeta}^t, \xi^t, \mathring{R}_t, \mathring{\omega}_t, \mathring{u}_t\right)}} \hat{\zeta}^t$$

$$= F\left(\mathring{\omega}_t\right) \hat{\zeta}^t F\left(\mathring{\omega}_t\right)^\top = \hat{\zeta}^t,$$

$$\xi^{t-1} = \mathcal{D}_\omega H^\nu\left(\hat{\zeta}^t, \xi^t, \mathring{R}_t, \mathring{\omega}_t, \mathring{u}_t\right) + \mu^t \mathcal{D}_\omega g_t\left(\mathring{\omega}_t\right)$$

$$= \frac{h\zeta^t}{\sqrt{1 - h^2 \mathring{\omega}_t^2}} + \xi^t + \mu^t \mathring{\omega}_t.$$

The state and adjoint system can be written as

$$\text{adjoint} \begin{cases} \zeta^{t-1} = \zeta^t, \\ \xi^{t-1} = \dfrac{h\zeta^t}{\sqrt{1 - h^2 \mathring{\omega}_t^2}} + \xi^t + \mu^t \mathring{\omega}_t, \end{cases} \tag{45}$$

$$\text{state} \begin{cases} \mathring{R}_{t+1} = \mathring{R}_t F\left(\mathring{\omega}_t\right), \\ \mathring{\omega}_{t+1} = \mathring{\omega}_t + h\mathring{u}_t, \end{cases} \tag{46}$$

(ii) Complementary slackness conditions given by:

$$\mu^t\left(\mathring{\omega}_t^2 - d^2\right) = 0 \quad \text{for all} \quad t = 1, \ldots, N-1, \tag{47}$$

(iii) Non-positivity condition given by:

$$\mu^t \leq 0 \quad \text{for all} \quad t = 1, \ldots, N-1,$$

(iv) Hamiltonian maximization pointwise in time, given by:

$$H^\nu\left(\hat{\zeta}^t, \xi^t, \mathring{R}_t, \mathring{\omega}_t, \mathring{u}_t\right) := \max_{w \in [-c,c]} H^\nu\left(\hat{\zeta}^t, \xi^t, \mathring{R}_t, \mathring{\omega}_t, w\right).$$

Remark 16. Note that, since the Hamiltonian is concave in u the non-positive gradient condition, in this case, leads to the maximization of the Hamiltonian pointwise in time.

It follows that

$$\mathring{u}_t = \arg \max_{w \in [-c,c]} H^\nu\left(\hat{\zeta}^t, \xi^t, \mathring{R}_t, \mathring{\omega}_t, w\right),$$

$$= \arg \max_{w \in [-c,c]} \left(\nu \frac{w^2}{2} + \xi^t h w\right).$$

If $\nu = -1$ then

$$\mathring{u}_t = \begin{cases} c & \text{if } h\xi^t \geq c, \\ -c & \text{if } h\xi^t \leq -c, \\ h\xi^t & \text{elsewhere.} \end{cases} \tag{48}$$

In the case of an abnormal extremal, i.e., $\nu = 0$,

$$\mathring{u}_t = \begin{cases} c & \text{if } \xi^t > 0, \\ -c & \text{if } \xi^t < 0. \end{cases} \tag{49}$$

Remark 17. Note that the optimal control $\mathring{u}_t$ is the saturation function of the co-state $h\xi^t$ when $\nu = -1$, and the control is bang–bang in the case of an abnormal extremal, i.e., $\nu = 0$.

The constrained boundary value problem (45)–(46) subject to boundary conditions $(R_0, \omega_0) = (R_i, \omega_i)$, $(R_N, \omega_N) = \left(R_f, \omega_f\right)$, the complementary slackness conditions (47) and the state constraints $|\omega_t| \leq d$ for $t = 1, \ldots, N-1$, is solved using multiple shooting methods (Phogat, Chatterjee, & Banavar, 2017).

Assume that a satellite has an inertia $I = 800$ kg m^2 about the axis of rotation, and is fitted with an actuation device capable of producing torque of the magnitude 20 N m. The attitude maneuvers of the satellite are subject to a maximum permissible magnitude of the momentum of 70 N m s. The model of the satellite scaled to unit inertia has been considered for the simulations with the following data:

- sampling time $(T) = 0.05$ s,
- maximum torque or control bound $(c) = 25$ mN m,
- maximum momentum $(d) = 87.5$ mN m s,
- time duration $(t_{\max})$ can range between 0 s and 150 s.

For ease of representation of the initial and final orientations in figures, we denote the initial and final orientations with the rotation angle θ such that

$$R_t := \begin{pmatrix} \cos(\theta_t) & \sin(\theta_t) \\ -\sin(\theta_t) & \cos(\theta_t) \end{pmatrix} \quad \text{for } \theta_t \in [0, 2\pi[.$$

So, the initial and final configurations for the trajectories are defined by (θ_i, ω_i) and (θ_f, ω_f) respectively.

Three maneuvers with different initial and final conditions have been simulated:

- $\mathcal{T}_1 : (\theta_i, \omega_i) = (0°, 0 \text{ N m s})$, $(\theta_f, \omega_f) = (90°, 80 \text{ mN m s})$, $t_f = 100$ s,
- $\mathcal{T}_2 : (\theta_i, \omega_i) = (0°, 0 \text{ N m s})$, $(\theta_f, \omega_f) = (75°, 0 \text{ N m s})$, $t_f = 19$ s,
- $\mathcal{T}_3 : (\theta_i, \omega_i) = (90°, 0 \text{ N m s})$, $(\theta_f, \omega_f) = (265°, 0 \text{ N m s})$, $t_f = 40$ s.

The distinguishing feature of this approach is that the system dynamics is defined in the configuration space in contrast to the local representation that enables one to find optimal trajectories which need more than one chart for local representation, as shown in Fig. 4. The image of the trajectory $\mathcal{T}_1$ projected on SO(2) is SO(2), and this image cannot be represented on a single chart. So, the local representation of the system dynamics cannot characterize such optimal trajectories.

The trajectories $\mathcal{T}_2, \mathcal{T}_3$ are shown in Figs. 5a–6a, and their corresponding optimal control profiles are shown in Figs. 5b–6b. It is important to note that whenever the state constraints are active i.e. $|\omega_t| = d$, where $t = 1, \ldots, N-1$, the control actions at such time instances will be zero. The optimal control corresponding to the maneuver $\mathcal{T}_2$ saturates at the end points in order to achieve the maneuver in a specified time, see Fig. 5b. On the other hand, the optimal control corresponding to trajectories $\mathcal{T}_1, \mathcal{T}_3$ does not saturate because the time duration of these maneuvers is higher then the minimum time needed for such maneuvers without actuator saturation as shown in Fig. 6b.

Acknowledgments

Appendix. Proofs of Corollaries 7–9

A.1. Proof of Corollary 7

Proof. In order to apply Theorem 5 to optimal control problem (10), let us define the manifold M_{fin} in a neighborhood of $(\mathring{q}_N, \mathring{x}_N) \in M_{\mathrm{fin}}$ as a zero level set of a smooth submersion. In other words, there exist an open set $\mathcal{O}_{(\mathring{q}_N, \mathring{x}_N)} \subset G \times \mathbb{R}^{n_x}$ containing $(\mathring{q}_N, \mathring{x}_N)$ and a smooth submersion

$$\mathcal{O}_{(\mathring{q}_N, \mathring{x}_N)} \ni (q_N, x_N) \mapsto b_{\mathrm{fin}}(q_N, x_N) \in \mathbb{R}^{n_q + n_x - m_f}$$
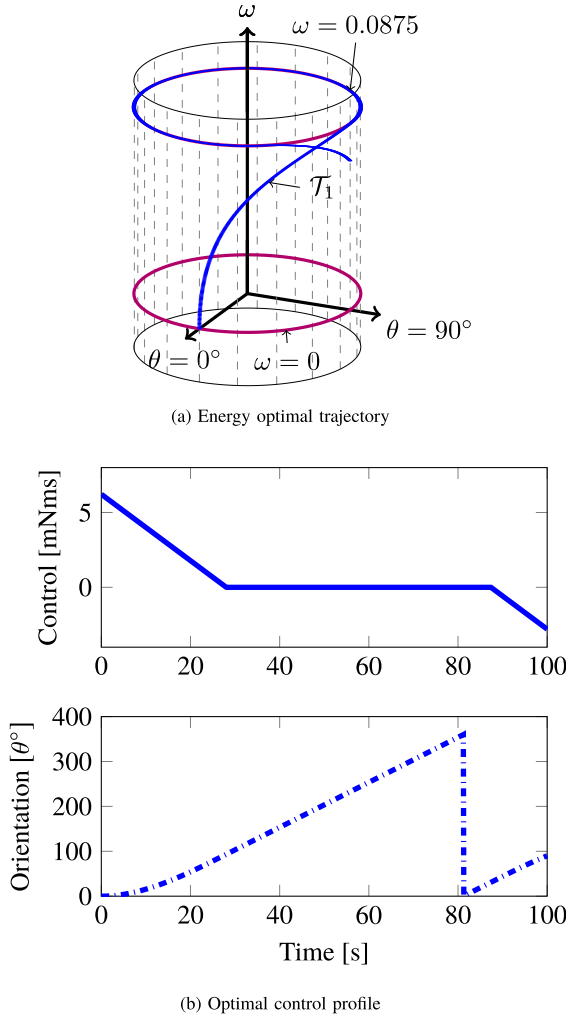
such that

$$b_{\mathrm{fin}}^{-1}(0) = \mathcal{O}_{(\mathring{q}_N, \mathring{x}_N)} \cap M_{\mathrm{fin}},$$

where m_f is the dimension of the manifold M_{fin}. So, the end point constraint is represented in the neighborhood $\mathcal{O}_{(\mathring{q}_N, \mathring{x}_N)}$ of the optimal point $(\mathring{q}_N, \mathring{x}_N)$ as

$$b_{\mathrm{fin}}(q_N, x_N) = 0. \tag{A.1}$$

(a) Energy optimal trajectory

(b) Optimal control profile

Fig. 4. The state trajectory and the corresponding optimal control for the maneuver $\mathcal{T}_1$.

(a) Energy optimal trajectory

(b) Optimal control profile

Fig. 5. The state trajectory and the corresponding optimal control for the maneuver $\mathcal{T}_2$.

Let us define the equality constraint (A.1) in the form of inequality constraints

$$\bar{g}_N (q_N, x_N) \leq 0, \tag{A.2}$$

where

$$\bar{g}_N (q_N, x_N) := \begin{pmatrix} b_{\text{fin}} (q_N, x_N) \\ -b_{\text{fin}} (q_N, x_N) \end{pmatrix} \in \mathbb{R}^{2(n_q + n_x - m_f)}.$$

It is evident from (A.2) that $\bar{g}_N (q_N, x_N) \leq 0$ if and only if $b_{\text{fin}} (q_N, x_N) = 0$. Now applying Theorem 5 to optimal control problem (10) with end point conditions defined by (A.2) in addition to state inequality constraints on the end points leads to the following set of conditions:

(i) (MP-i) holds,
(ii) The transversality conditions (MP-ii) can then be rewritten as d

$$\begin{aligned} & T^*_{\mathring{q}_N} \Phi_{\mathring{q}_N^{-1}} \left(\rho^{N-1} \right) - \nu \mathcal{D}_q c_N \left(\mathring{q}_N, \mathring{x}_N \right) \\ & \quad - \mu^N \mathcal{D}_q g_N \left(\mathring{q}_N, \mathring{x}_N \right) - \sigma \mathcal{D}_q \bar{g}_N \left(\mathring{q}_N, \mathring{x}_N \right) = 0, \\ & \xi^{N-1} - \nu \mathcal{D}_x c_N \left(\mathring{q}_N, \mathring{x}_N \right) \\ & \quad - \mu^N \mathcal{D}_x g_N \left(\mathring{q}_N, \mathring{x}_N \right) - \sigma \mathcal{D}_x \bar{g}_N \left(\mathring{q}_N, \mathring{x}_N \right) = 0, \end{aligned} \tag{A.3}$$

where $(\mathbb{R}^{n_N})^* \ni \sigma \leq 0$, and $n_N := 2 \left(n_q + n_x - m_f \right)$. Using (A.2), the transversality conditions (A.3) further simplify and
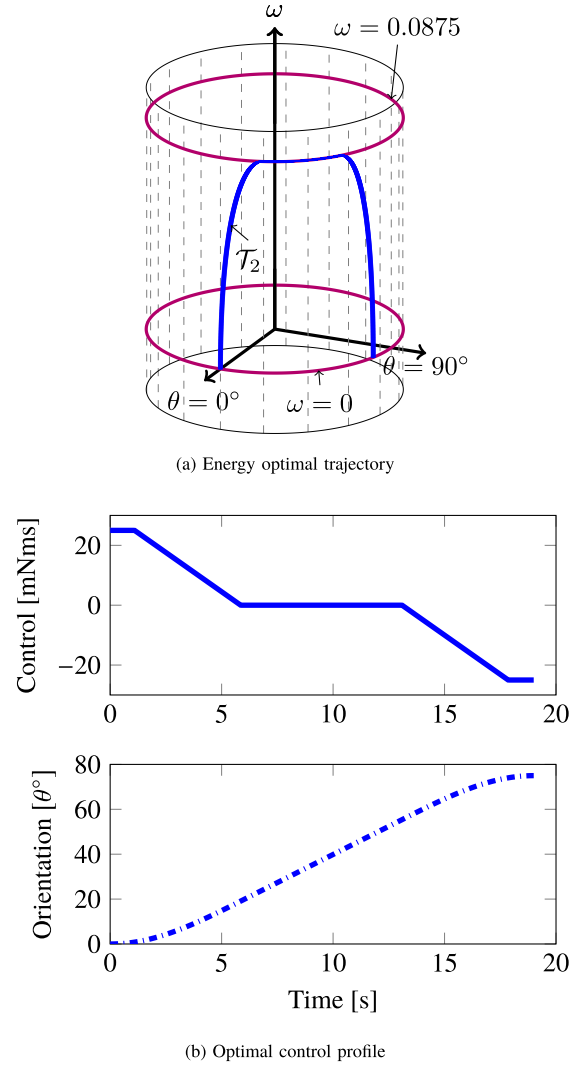
written in compressed form as

$$\left(T^*_{\mathring{q}_N} \Phi_{\mathring{q}_N^{-1}} \left(\rho^{N-1} \right), \xi^{N-1} \right) - \nu \mathcal{D}_{(q,x)} c_N \left(\mathring{q}_N, \mathring{x}_N \right) \tag{A.4}$$
$$- \mu^N \mathcal{D}_{(q,x)} g_N \left(\mathring{q}_N, \mathring{x}_N \right) = (\sigma_+ - \sigma_-) \mathcal{D}_{(q,x)} b_{\text{fin}} \left(\mathring{q}_N, \mathring{x}_N \right),$$

where

$$\sigma_- := \left(\sigma_1, \ldots, \sigma_{n_N/2} \right) \text{ and } \sigma_+ := \left(\sigma_{n_N/2+1}, \ldots, \sigma_{n_N} \right).$$

Note that the submanifold M_{fin} is locally embedded in $G \times \mathbb{R}^{n_x}$ via the embedding b_{fin}. Hence, it follows that

$$\mathcal{D}_{(q,x)} b_{\text{fin}} \left(\mathring{q}_N, \mathring{x}_N \right) v = 0 \quad \text{for all} \quad v \in T_{(\mathring{q}_N, \mathring{x}_N)} M_{\text{fin}},$$

and hence for a covector $\alpha \in \left(\mathbb{R}^{n_q + n_x - m_f} \right)^*$, we have

$$\alpha \mathcal{D}_{(q,x)} b_{\text{fin}} \left(\mathring{q}_N, \mathring{x}_N \right) v = 0 \quad \text{for all } v \in T_{(\mathring{q}_0, \mathring{x}_0)} M_{\text{init}},$$

which is equivalent to the following:

$$T^*_{(\mathring{q}_N, \mathring{x}_N)} \left(G \times \mathbb{R}^{n_x} \right) \ni \alpha \mathcal{D}_{(q,x)} b_{\text{fin}} \left(\mathring{q}_N, \mathring{x}_N \right)$$
$$\perp T_{(\mathring{q}_N, \mathring{x}_N)} M_{\text{fin}}. \tag{A.5}$$

So, the transversality conditions (A.4) with (A.5) leads to

$$\left\{ \left(T^*_{\mathring{q}_N} \Phi_{\mathring{q}_N^{-1}} (\rho^{N-1}), \xi^{N-1} \right) - \nu \mathcal{D}_{(q,x)} c_N \left(\mathring{q}_N, \mathring{x}_N \right) \right.$$
$$\left. - \mu^N \mathcal{D}_{(q,x)} g_N \left(\mathring{q}_N, \mathring{x}_N \right) \right\} \perp T_{(\mathring{q}_N, \mathring{x}_N)} M_{\text{fin}}.$$

(a) Energy optimal trajectory

(b) Optimal control profile
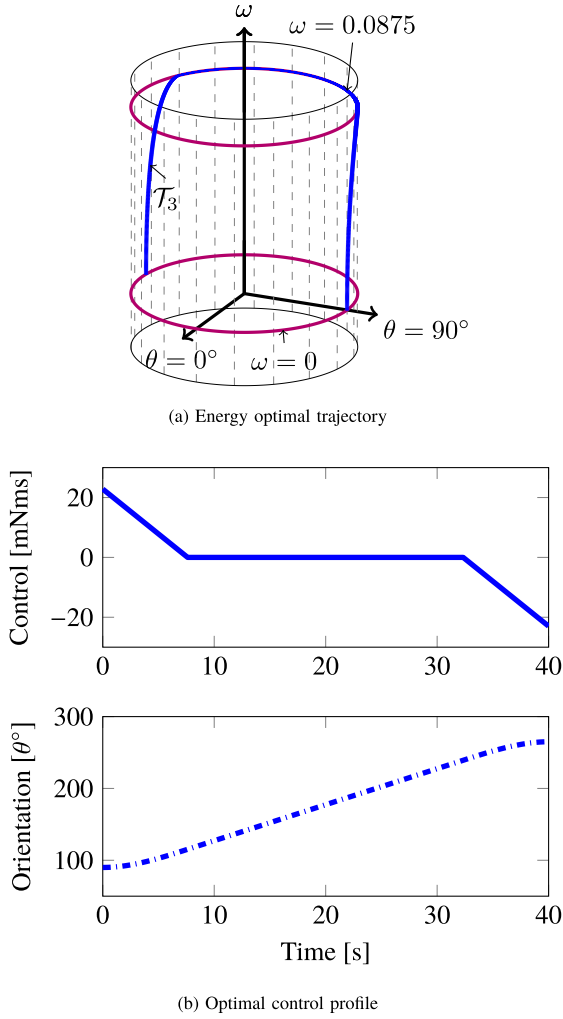
Fig. 6. The state trajectory and the corresponding optimal control for the maneuver $\mathcal{T}_3$.

(iii) (MP-iii) holds,
(iv) (MP-iv) holds,
(v) (MP-v) holds,
(vi) Suppose the adjoint variables $\left\{\left(\zeta^t, \xi^t\right)\right\}_{t=0}^{N-1}$, covectors $\left\{\mu^t\right\}_{t=1}^N$ and the scalar ν are all zero. Then (A.4) modifies to

$$\left(\sigma_+ - \sigma_-\right) \mathcal{D}_{(q,x)} b_{\mathrm{fin}} \left(\mathring{q}_N, \mathring{x}_N\right) = 0. \tag{A.6}$$

Since the map b_{fin} is a smooth submersion at $\left(\mathring{q}_N, \mathring{x}_N\right)$, $\mathcal{D}_{(q,x)} b_N$ $\left(\mathring{q}_N, \mathring{x}_N\right)$ is full rank, and therefore

$$\left(\sigma_+ - \sigma_-\right) = 0. \tag{A.7}$$

In the case $\sigma_+ = \sigma_- \neq 0$, (A.6) leads to triviality, i.e., the half spaces

$$H^- := \left\{ v \in T_{\left(\mathring{q}_N, \mathring{x}_N\right)} \left(G \times \mathbb{R}^{n_x}\right) \mid \right.$$

$$\left. \left\langle \sigma_+ \mathcal{D}_{(q,x)} b_{\mathrm{fin}} \left(\mathring{q}_N, \mathring{x}_N\right), v \right\rangle \leq 0 \right\}$$

and

$$H^+ := \left\{ v \in T_{\left(\mathring{q}_N, \mathring{x}_N\right)} \left(G \times \mathbb{R}^{n_x}\right) \mid \right.$$

$$\left. \left\langle \sigma_+ \mathcal{D}_{(q,x)} b_{\mathrm{fin}} \left(\mathring{q}_N, \mathring{x}_N\right), v \right\rangle \geq 0 \right\}$$

are separable. Otherwise, contradicts non-triviality condition (MP-vi). So, the non-triviality condition translates to the following:
the adjoint variables $\left\{\left(\zeta^t, \xi^t\right)\right\}_{t=0}^{N-1}$, covectors $\left\{\mu^t\right\}_{t=1}^N$ and the scalar ν are all not zero,

which completes our proof. $\qquad \blacksquare$

A.2. Proof of Corollary 8

Proof. Assume that the inequality constraints in Corollary 7 are given by

$$G \times \mathbb{R}^{n_x} \ni (q, x) \mapsto g_t (q, x) := -1 \in \mathbb{R} \tag{A.8}$$

for $t = 1, \ldots, N$, the final cost is given by

$$G \times \mathbb{R}^{n_x} \ni (q, x) \mapsto c_N (q, x) := 0 \in \mathbb{R},$$

and the fixed boundary point is an immersed manifold, i.e.,

$$M_{\mathrm{fin}} := \left(\bar{q}_N, \bar{x}_N\right) \subset G \times \mathbb{R}^{n_x}$$

Under these assumptions, applying Corollary 7 to optimal control problem (11) leads to the following set of conditions:

(i) Note that $\mathcal{D}_{(q,x)} g_t \left(\mathring{q}_t, \mathring{x}_t\right) = 0$ for $t = 1, \ldots, N$, and that simplifies the state and adjoint dynamics to

$$\text{state} \begin{cases} q_{t+1} = q_t e^{\mathcal{D}_\zeta H^\nu (\mathring{\gamma}_t)}, \\ x_{t+1} = \mathcal{D}_\xi H^\nu (\mathring{\gamma}_t), \end{cases}$$

$$\text{adjoint} \begin{cases} \rho^{t-1} = \mathrm{Ad}^*_{e^{-\mathcal{D}_\zeta H^\nu (\mathring{\gamma}_t)}} \rho^t + T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_q H^\nu (\mathring{\gamma}_t)\right), \\ \xi^{t-1} = \mathcal{D}_x H^\nu (\mathring{\gamma}_t), \end{cases}$$

(ii) The manifold M_{fin} is a singleton. Hence

$$T_{\left(\mathring{q}_N, \mathring{x}_N\right)} M_{\mathrm{fin}} = 0,$$

and therefore the annihilator of the tangent space $T_{\left(\mathring{q}_N, \mathring{x}_N\right)} M_{\mathrm{fin}}$ in $T^*_{\left(\mathring{q}_N, \mathring{x}_N\right)} \left(G \times \mathbb{R}^{n_x}\right)$ is $T^*_{\left(\mathring{q}_N, \mathring{x}_N\right)} \left(G \times \mathbb{R}^{n_x}\right)$. So, the transversality conditions of Corollary 7 are trivially satisfied.

(iii) Hamiltonian non-positive gradient condition:

$$\left\langle \mathcal{D}_u H^\nu \left(t, \zeta^t, \xi^t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t\right), w - \mathring{u}_t \right\rangle \leq 0 \quad \text{for all} \quad w \in \mathbb{U}_t.$$

(iv) In view of (A.8) and the complementary slackness conditions of Corollary 7, we obtain

$$\mu^t = 0 \quad \text{for all} \quad t = 1, \ldots, N.$$

The non-positivity condition of Corollary 7 is trivially satisfied and the non-triviality condition translates to the following:
adjoint variables $\left\{\zeta^t, \xi^t\right\}_{t=0}^{N-1}$ and the scalar ν do not simultaneously vanish. $\qquad \blacksquare$

A.3. Proof of Corollary 9

Proof. We first prove that there exist a unique representation of s_t in the neighborhood of $e \in G$ satisfying $v_t (s_t, q_t, x_t) = 0$. Then the optimal control problem is posed in the standard form (11) and necessary conditions are derived by applying Corollary 8.

Lemma 18. *If the map* $v_t (\cdot, q_t, x_t) : \mathcal{O}_e \subset G \to \mathbb{R}^{n_q}$ *is a diffeomorphism onto its image for all feasible pairs*

$$\mathcal{A}_t := \left\{ (q_t, x_t) \in G \times \mathbb{R}^{n_x} \mid \{(q_t, x_t)\}_{t=0}^N \text{ satisfying (6)} \right\},$$

where $\mathcal{O}_e$ *is an open set in* G *containing* $e \in G$, *then for a fixed* $\left(\left(\bar{q}_t, \bar{x}_t\right), \bar{s}_t\right) \in \mathcal{A}_t \times \mathcal{O}_e$, *there exist open neighborhoods* $\mathcal{N}_t$ *of*

$(\bar{q}_t, \bar{x}_t,)$ in $G \times \mathbb{R}^{n_x}$, $\mathcal{R}_t$ of $\bar{s}_t$ in G, and a continuously differentiable map $\kappa_t : \mathcal{N}_t \to \mathcal{R}_t$ such that

$$s_t = \kappa_t (q_t, x_t), \quad \text{satisfies} \quad v_t (\kappa_t (q_t, x_t), q_t, x_t) = 0.$$

Furthermore, the derivative $D\kappa_t (\bar{q}_t, \bar{x}_t) : T_{\bar{q}_t} G \times \mathbb{R}^{n_x} \to T_{\bar{s}_t} G$ is given by

$$\mathcal{D}_q \kappa_t (\bar{q}_t, \bar{x}_t) = -\mathcal{D}_s v_t (\bar{s}_t, \bar{q}_t, \bar{x}_t)^{-1} \circ \mathcal{D}_q v_t (\bar{s}_t, \bar{q}_t, \bar{x}_t)$$

and

$$\mathcal{D}_x \kappa_t (\bar{q}_t, \bar{x}_t) = -\mathcal{D}_s v_t (\bar{s}_t, \bar{q}_t, \bar{x}_t)^{-1} \circ \mathcal{D}_x v_t (\bar{s}_t, \bar{q}_t, \bar{x}_t).$$

Proof. See Phogat et al. (2018, Appendix B.3).

Let $\left(\{\mathring{q}_t, \mathring{x}_t\}_{t=0}^N, \{\mathring{u}_t\}_{t=0}^{N-1} \right)$ be an optimal state–action trajectory with $\{\mathring{s}_t\}_{t=0}^{N-1} \subset \mathcal{O}_e$ such that $v_t (\mathring{s}_t, \mathring{q}_t, \mathring{x}_t) = 0$ for $t = 0, \ldots, N-1$. Using Lemma 18: there exist open neighborhoods $\mathcal{N}_t \subset G \times \mathbb{R}^{n_x}$ of $(\mathring{q}_t, \mathring{x}_t)$, $\mathcal{R}_t \subset G$ of $\mathring{s}_t \in \mathcal{O}_e$, and the maps $\kappa_t : \mathcal{N}_t \to \mathcal{R}_t$ such that

$$s_t = \kappa_t (q_t, x_t), \quad \text{satisfies} \quad v_t (\kappa_t (q_t, x_t), q_t, x_t) = 0.$$

Let us define

$$\mathcal{N} := \mathcal{N}_0 \times \cdots \times \mathcal{N}_N,$$

where $\mathcal{N}_N \subset G \times \mathbb{R}^{n_x}$ is a neighborhood containing $(\mathring{q}_N, \mathring{x}_N)$. Note that the system dynamics in (12) is smooth in the control variables $\{u_t\}_{t=0}^{N-1}$. Therefore there exists a number $r > 0$ such that state trajectories starting at $(\bar{q}_0, \bar{x}_0)$ remains in $\mathcal{N}$ under all control actions

$$(u_0, \ldots, u_{N-1}) \in \tilde{U} := \tilde{U}_0 \times \cdots \times \tilde{U}_{N-1},$$

where $\tilde{U}_t := (\mathbb{U}_t \cap \mathcal{B}_r (\mathring{u}_t)) \subset \mathbb{R}^{n_u}$, and $\mathcal{B}_r (\mathring{u}_t)$ is an open neighborhood of radius r containing $\mathring{u}_t$.

This leads to the representation of the optimal control problem (12) in the neighborhood $\mathcal{N} \times \tilde{U}$ of the optimal state–action trajectory as :

$$\underset{\{u_t\}_{t=0}^{N-1}}{\text{minimize}} \quad \mathscr{J} (\boldsymbol{q}, \boldsymbol{x}, \boldsymbol{u}) = \sum_{t=0}^{N-1} c_t (\kappa_t (q_t, x_t), q_t, x_t, u_t)$$

$$\text{subject to} \quad \begin{cases} q_{t+1} = q_t \kappa_t(q_t, x_t) \\ x_{t+1} = f_t (\kappa_t (q_t, x_t), q_t, x_t, u_t) \\ u_t \in \tilde{U}_t \\ \text{for all } t \in [N-1], \\ (q_0, x_0) = (\bar{q}_0, \bar{x}_0), \\ (q_N, x_N) = (\bar{q}_N, \bar{x}_N), \\ \text{Assumption 1.} \end{cases} \quad \text{(A.9)}$$

Let us define the Hamiltonian function as in (13). Applying Corollary 8 to (A.9) immediately leads to the following conditions: for

$$\mathring{\gamma}_t := \left(t, \zeta^t, \xi^t, \mathring{s}_t, \mathring{q}_t, \mathring{x}_t, \mathring{u}_t \right), \quad \mathring{v}_t := v (\mathring{s}_t, \mathring{q}_t, \mathring{x}_t),$$

and

$$\rho^t := \left(\mathcal{D}e^{-1}(\mathring{q}_t^{-1}\mathring{q}_{t+1}) \circ T_e \Phi_{\mathring{q}_t^{-1}\mathring{q}_{t+1}} \right)^* (\zeta^t),$$

(i) State and adjoint system dynamics given by

$$\text{state} \quad \begin{cases} \mathring{q}_{t+1} = \mathring{q}_t e^{\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}, \\ v_t (\mathring{s}_t, \mathring{q}_t, \mathring{x}_t) = 0, \\ \mathring{x}_{t+1} = \mathcal{D}_\xi H^\nu (\mathring{\gamma}_t), \end{cases}$$

$$\text{adjoint} \quad \begin{cases} \rho^{t-1} = -T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_s H^\nu (\mathring{\gamma}_t) \circ \mathcal{D}_s \mathring{v}_t^{-1} \circ \mathcal{D}_q \mathring{v}_t \right) \\ \quad + T_e^* \Phi_{\mathring{q}_t} \left(\mathcal{D}_q H^\nu (\mathring{\gamma}_t) \right) + \text{Ad}_{e^{-\mathcal{D}_\zeta H^\nu(\mathring{\gamma}_t)}}^* \rho^t, \\ \xi^{t-1} = -\mathcal{D}_s H^\nu (\mathring{\gamma}_t) \circ \mathcal{D}_s \mathring{v}_t^{-1} \circ \mathcal{D}_x \mathring{v}_t \\ \quad + \mathcal{D}_x H^\nu (\mathring{\gamma}_t). \end{cases}$$

(ii) Hamiltonian non-positive gradient condition given by

$$\langle \mathcal{D}_u H^\nu (\mathring{\gamma}_t), \tilde{w} - \mathring{u}_t \rangle \leq 0 \quad \text{for all} \quad \tilde{w} \in \tilde{U}_t.$$

Since the set $\mathbb{U}_t$ is convex,

$$\tilde{w}_\alpha := \mathring{u}_t + \alpha (w - \mathring{u}_t) \in \mathbb{U}_t \text{ for any } \alpha \in [0, 1] \text{ and } w \in \mathbb{U}_t.$$

In particular, choosing

$$\bar{\alpha} := \begin{cases} \min \left\{ \dfrac{r}{2 \|w - \mathring{u}_t\|}, 1 \right\} & \text{for} \quad w \neq \mathring{u}_t, \\ 0 & \text{otherwise,} \end{cases}$$

ensures that $\tilde{w}_{\bar{\alpha}} \in \tilde{U}_t$, and consequently,

$$\langle \mathcal{D}_u H^\nu (\mathring{\gamma}_t), \bar{\alpha} (w - \mathring{u}_t) \rangle \leq 0 \quad \text{for all} \quad w \in \mathbb{U}_t.$$

Since $\bar{\alpha} \geq 0$, it follows that

$$\langle \mathcal{D}_u H^\nu (\mathring{\gamma}_t), w - \mathring{u}_t \rangle \leq 0 \quad \text{for all} \quad w \in \mathbb{U}_t,$$

and this proves the assertion.

(iii) Non-triviality condition given by:
adjoint variables $\{\zeta^t, \xi^t\}_{t=0}^{N-1}$ and the scalar ν do not simultaneously vanish.

References

Abraham, R., & Marsden, J. E. (1978). *Foundations of Mechanics* (2nd ed.). (pp. xxii+m–xvi+806). Benjamin/Cummings Publishing Co., Inc., Advanced Book Program, Reading, Mass., revised and enlarged, With the assistance of Tudor Raţiu and Richard Cushman.

Agrachev, A. A., & Sachkov, Y. L. (2004). *Encyclopaedia of mathematical sciences*: vol. 87. Control theory from the geometric viewpoint (p. xiv+412). Berlin: Springer-Verlag, Control Theory and Optimization, II.

Barbero-Liñán, M., & Muñoz Lecanda, M. C. (2009). Geometric approach to Pontryagin's maximum principle. *Acta Applicandae Mathematicae*, 108(2), 429–485.

Boltyanskii, V. G. (1975). The method of tents in the theory of extremal problems. *Russian Mathematical Surveys*, 30(3), 1–54.

Boltyanskii, V. G., Martini, H., & Soltan, V. (1999). *Combinatorial optimization*: vol. 4. Geometric methods and optimization problems (p. viii+429). Dordrecht: Kluwer Academic Publishers.

Bonnard, B., & Sugny, D. (2012). *AIMS series on applied mathematics*: vol. 5. Optimal control with applications in space and quantum dynamics (p. xvi+283). Springfield, MO: American Institute of Mathematical Sciences (AIMS).

Bourdin, L. c., & Trélat, E. (2016). Optimal sampled-data control, and generalizations on time scales. *Mathematical Control and Related Fields*, 6(1), 53–94.

Brockett, R. W. (1973). Lie theory and control systems defined on spheres. *SIAM Journal on Applied Mathematics*, 25, 213–225, Lie algebras: applications and computational methods (Conf., Drexel Univ., Philadelphia, Pa., 1972).

Bullo, F., & Lynch, K. M. (2001). Kinematic controllability and decoupled trajectory planning for underactuated mechanical systems. In *IEEE international conference on robotics and automation, 2001. Proceedings 2001 ICRA, Vol. 4* (pp. 3300–3307). IEEE.

Clarke, F. (2013). *Graduate texts in mathematics*: vol. 264. Functional analysis, calculus of variations and optimal control (p. xiv+591). London: Springer.

Clarke, F. H. (1976). The maximum principle under minimal hypotheses. *SIAM Journal on Control and Optimization*, 14(6), 1078–1091.

Colombo, L., Ferraro, S., & Martín de Diego, D. (2016). Geometric integrators for higher-order variational systems and their application to optimal control. *Journal of Nonlinear Science*, 26(6), 1615–1650.

Colombo, L., Jiménez, F., & Martín de Diego, D. (2012). Discrete second-order Euler-Poincaré equations. Applications to optimal control. *International Journal of Geometric Methods in Modern Physics*, 9(4), 1250037, 20.

Dubovitskii, A. Y. (1978). The discrete principle of the maximum. *Avtomatika I Telemekhanika*, (10), 55–71.

Dubovitskii, A. Y., & Milyutin, A. A. (1968). Necessary conditions for a weak extremum in optimal control problems with mixed constraints of the inequality. *USSR Computational Mathematics and Mathematical Physics*, 8(4), 24–98.

Gamkrelidze, R. V. (1999). Discovery of the maximum principle. *Journal of Dynamical and Control Systems*, 5(4), 437–451.

Holtzman, J. M. (1966). Convexity and the maximum principle for discrete systems. *IEEE Transactions on Automatic Control*, AC-11, 30–35.

Janković, V. (2002). Tent theory. *Journal of Mathematical Sciences (New York)*, 110(2), 2572–2582, Dynamical systems, 10.

Khaneja, N., Brockett, R., & Glaser, S. J. (2001). Time optimal control in spin systems. *Physical Review A*, 63(3), 032308.

Kobilarov, M. B., & Marsden, J. E. (2011). Discrete geometric optimal control on Lie groups. *IEEE Transactions on Robotics*, 27(4), 641–655.

Lee, J. M. (2013). *Graduate texts in mathematics*: *vol. 218. Introduction to smooth manifolds*. (2nd ed.). (p. xvi+708). New York: Springer.

Lee, T., Leok, M., & McClamroch, N. H. (2008a). Optimal attitude control of a rigid body using geometrically exact computations on SO(3). *Journal of Dynamical and Control Systems*, 14(4), 465–487.

Lee, T., Leok, M., & McClamroch, N. H. (2008b). Time optimal attitude control for a rigid body. In *American control conference, 2008* (pp. 5210–5215). IEEE.

Marsden, J. E., & Ratiu, T. S. (1999). *Texts in applied mathematics*: *vol. 17. Introduction to mechanics and symmetry: A basic exposition of classical mechanical systems*. (2nd ed.). (p. xviii+582). New York: Springer-Verlag.

Marsden, J. E., & West, M. (2001). Discrete mechanics and variational integrators. *Acta Numerica*, 10, 357–514.

Milyutin, A. A., & Osmolovskii, N. P. (1998). *Translations of mathematical monographs*: *vol. 180. Calculus of variations and optimal control* (p. xii+372). Providence, RI: American Mathematical Society, Translated from the Russian manuscript by Dimitrii Chibisov.

Mordukhovich, B. S. (1976). Maximum principle in the problem of time optimal response with nonsmooth constraints. *Journal of Applied Mathematics and Mechanics*, 40(6), 960–969 (1977).

Ober-Blöbaum, S. (2008). *Discrete mechanics and optimal control* (Ph.D. thesis), Paderborn: University of Paderborn.

Ober-Blöbaum, S., Junge, O., & Marsden, J. E. (2011). Discrete mechanics and optimal control: an analysis. *ESAIM. Control, Optimisation and Calculus of Variations*, 17(2), 322–352.

Phogat, K. S., Chatterjee, D., & Banavar, R. (2017). Discrete-time optimal attitude control of a spacecraft with momentum and control constraints. *Journal of Guidance, Control, and Dynamics*, 41(1), 199–211.

Phogat, K. S., Chatterjee, D., & Banavar, R. N. (2018). A discrete-time Pontryagin maximum principle on matrix Lie groups. https://arxiv.org/abs/1612.08022, 8.

Pontryagin, L. S. (1987). *Mathematical theory of optimal processes*. CRC Press.

Pshenichnyi, B. N. (1971). *Necessary conditions for an extremum*. New York: Marcel Dekker Inc.

Saccon, A., Hauser, J., & Aguiar, A. P. (2013). Optimal control on Lie groups: The projection operator approach. *IEEE Transactions on Automatic Control*, 58(9), 2230–2245.

Sachkov, Y. L. (2009). Control theory on Lie groups. *Journal of Mathematical Sciences*, 156(3), 381–439.

Sussmann, H. J. (1998). An introduction to the coordinate-free maximum principle. In *Monograph textbooks pure & applied math*: *vol. 207. Geometry of feedback and optimal control* (pp. 463–557).

Sussmann, H. J. (2000). The Euler-Lagrange equation and its invariance properties, Lecture notes.

Sussmann, H. J. (2008). A Pontryagin maximum principle for systems of flows. In *Lecture notes in control and information sciences*: *vol. 371. Recent advances in learning and control* (pp. 219–232). London: Springer.

Trélat, E. (2012). Optimal control and applications to aerospace: some results and challenges. *Journal of Optimization Theory and Applications*, 154(3), 713–758.

Vemulapalli, R., Arrate, F., & Chellappa, R. (2014). Human action recognition by representing 3d skeletons as points in a Lie group. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 588–595).

Warga, J. (1972). *Optimal control of differential and functional equations*. (p. xiii+531). New York-London: Academic Press.

Karmvir Singh Phogat received his M.Sc. in Applied Mathematics from IIT Roorkee in 2012, and Ph.D. in Systems and Control Engineering from IIT Bombay in 2018. He is currently a Postdoctoral researcher at KAIST, South Korea. His research interests include geometric optimal control and its applications in electrical and aerospace engineering.

Debasish Chatterjee received his Ph.D. in Electrical & Computer Engineering from the University of Illinois at Urbana–Champaign in 2007. He worked at ETH Zurich as a postdoc before joining IIT Bombay in 2011. His research interests lie in constrained control with emphasis on computational tractability, geometric techniques in control, and applied probability. He serves as an Associate Editor of Automatica and an Editor of the International Journal of Robust and Nonlinear Control.

Ravi N. Banavar received his B.Tech. in Mechanical Engineering from IIT Madras (1986), his Masters (Mechanical, 1988) and Ph.D. (Aerospace, 1992) degrees from Clemson University and the University of Texas at Austin, respectively. He had a brief teaching stint at UCLA in 1991–92, soon after which he joined the Systems and Control Engineering group at IIT Bombay in early 1993. He is currently a professor in this group and has spent a few sabbatical breaks during the years at UCLA (Los Angeles), IISc (Bangalore) and LSS (Supelec, France.) His research interests are broadly in the field of geometric mechanics and nonlinear control, with applications in electromechanical and aerospace engineering problems.