

# A SIMPLE PROOF OF THE DISCRETE TIME GEOMETRIC PONTRYAGIN MAXIMUM PRINCIPLE ON SMOOTH MANIFOLDS

MISHAL ASSIF P K

*Department of Mechanical Engineering  
IIT Bombay, Powai  
Mumbai 400076, India.  
<http://mishalassif.github.io>*

DEBASISH CHATTERJEE AND RAVI BANAVAR

*Systems & Control Engineering  
IIT Bombay, Powai  
Mumbai 400076, India.  
<http://www.sc.iitb.ac.in/~chatterjee>  
<http://www.sc.iitb.ac.in/~banavar>*

**ABSTRACT.** We establish a geometric Pontryagin maximum principle for discrete time optimal control problems on finite dimensional smooth manifolds under the following three types of constraints: a) constraints on the states pointwise in time, b) constraints on the control actions pointwise in time, c) constraints on the frequency spectrum of the optimal control trajectories. Our proof follows, in spirit, the path to establish geometric versions of the Pontryagin maximum principle on smooth manifolds indicated in [Cha11] in the context of continuous-time optimal control.

## §1. INTRODUCTION

The celebrated Pontryagin maximum principle (PMP) is a central tool in optimal control theory that provides first order necessary conditions for optimal controls. These necessary conditions can be used by algorithms to arrive at optimal control actions. The PMP was first introduced for continuous time control systems on  $\mathbb{R}^n$  by Pontryagin and his students in [PBG62] and alternate proofs for the PMP later appeared in [Bol71] and [LM67]. The discrete time Pontryagin maximum principle was developed primarily by Boltyanskii (see [Bol75, Bol78] and the references therein) and discrete time is the setting of our current work.

---

*E-mail addresses:* {mishal\_assif, dchatter, banavar}@iitb.ac.in.

*Key words and phrases.* optimal control, Pontryagin maximum principle, smooth manifolds.

The authors were supported in part by the grant 17ISROC001 from the Indian Space Research Organization.

While control systems evolving on  $\mathbb{R}^n$  are the most common, systems with non-flat manifolds as configuration spaces also appear in a variety of engineering disciplines including robotics, quantum mechanical systems, and aerospace systems. Justifiably so, the continuous time PMP was extended to control systems evolving on smooth manifolds in a sequence of works from [Sus98] through [AS04]; however the proofs given in these sources are quite complicated. The most recent proof of the geometric continuous time PMP appears in [Cha11]; it deserves special mention because of its sheer simplicity. This work serves as a source of inspiration for our current article. Assuming the validity of the PMP on Euclidean spaces, in [Cha11] the author derives the geometric version of the PMP by embedding the underlying manifold in a suitable Euclidean space and extending the optimal control problem on the manifold into an equivalent control problem on the Euclidean space, followed by appealing to the PMP on the Euclidean space, and finally translating the necessary conditions furnished by the PMP for the extended problem on the Euclidean space back to the manifold. This is the route that we follow here in the discrete time setting.

Almost all physical systems that are to be controlled naturally come with an array of constraints attached to them. In spite of this, there are few control techniques available that can actually compute constrained control actions in a tractable fashion. The continuous time PMP is no exception to this: numerical algorithms that seek to identify optimal controls from the necessary condition given by the PMP can handle control constraints rather efficiently. However, the necessary conditions given by the continuous time PMP for point-wise state constraints typically involve a measure, which is an infinite dimensional object, and numerical methods face grave difficulties in this setting. If one wants to include point-wise state constraints in the optimal control problem during the synthesis stage, it is better to perform some kind of discretization of the system first, and this is where the relevance of discrete time optimal control arises. A discrete time PMP on smooth manifolds can be employed by algorithms to solve state and control constrained control problems with relative ease.

In this article we address optimal control problems for discrete-time smooth control systems evolving on finite dimensional smooth manifolds in the presence of the following three important classes of constraints:

- (I) constraints on the states at each time instant,
- (II) constraints on the control magnitudes at each time instant, and
- (III) constraints on the frequency of the control functions.

We prove a discrete time PMP for control systems on smooth finite dimensional manifolds under the presence of the three classes of constraints of type mentioned above with the aid of three simple ingredients:

- (Step 1) The Whitney embedding theorem, which is employed for embedding the smooth manifold in a suitable Euclidean space.
- (Step 2) A few basic extension theorems for smooth functions defined on embedded sub-manifolds, employed here to extend the original optimal control problem to the Euclidean space given by Step 1.
- (Step 3) The discrete time PMP on  $\mathbb{R}^n$  under frequency constraints [PC19], employed to arrive at first order necessary conditions for optimality of the extended problem.

To our knowledge the only sources that discuss versions of the PMP for discrete time geometric optimal control problems are [PCB18] and [KG17]. The former establishes a

PMP for a class of smooth control systems evolving on Lie groups under mild structural assumptions on the system dynamics. In contrast, in the present article we remove all such assumptions and present a neater version of the PMP with broader applicability using very different and simple tools. [KG17] proves a PMP on smooth manifolds subject to the similar types of constraints that we consider here, but with the exception of the frequency constraints, and they do so under weaker assumptions on the smoothness of the cost, the constraints and the state transition maps. However the exposition in [KG17] heavily relies on nontrivial tools of nonsmooth analysis, and is nowhere nearly as simple as the proof we present here. The frequency constraints treated in this article first appeared in [PC19], but the exposition there was limited to systems evolving on finite dimensional Euclidean spaces, as opposed to non-flat smooth manifolds.

## §2. PRELIMINARIES

We employ standard notation throughout the article:  $\mathbb{N}$  denotes the non-negative integers,  $\mathbb{N}^*$  the positive integers,  $\mathbb{R}$  the real numbers. If  $k$  is a positive integer, we let  $[k] := \{1, \dots, k\}$ . The vector space  $\mathbb{R}^d$  is always assumed to be equipped with the standard inner product  $\langle v, v' \rangle := v^\top v'$  for every  $v, v' \in \mathbb{R}^d$ , and we denote by  $v^k$  the  $k^{\text{th}}$  component of  $v$ . In the theorem statements, we use  $(\mathbb{R}^d)^\star$  to denote the dual space of  $\mathbb{R}^d$  for the sake of precision; of course,  $(\mathbb{R}^d)^\star$  is isomorphic to  $\mathbb{R}^d$  in view of the Riesz representation theorem. It is also assumed that  $\mathbb{R}^d$  is endowed with the standard partial order  $\leq$ ; i.e., two vectors  $v, w \in \mathbb{R}^d$  are related by  $v \leq w$  if and only if  $v^i \leq w^i$  for all  $i = 1, \dots, d$ .

If  $\mathbb{M}_1$  and  $\mathbb{M}_2$  are smooth manifolds and  $f : \mathbb{M}_1 \rightarrow \mathbb{M}_2$  is a smooth map, then  $Tf : T\mathbb{M}_1 \rightarrow T\mathbb{M}_2$  denotes the tangent lift of the map  $f$  and  $T^*f : T^*\mathbb{M}_2 \rightarrow T^*\mathbb{M}_1$  denotes the cotangent lift of the map  $f$ .  $Tf(x_0) : T_{x_0}\mathbb{M}_1 \rightarrow T_{f(x_0)}\mathbb{M}_2$  will denote the tangent lift of the map  $f$  at  $x_0$ , and  $T^*f(x_0) : T_{f(x_0)}^*\mathbb{M}_2 \rightarrow T_{x_0}^*\mathbb{M}_1$  will denote the cotangent lift of the map  $f$  at  $x_0$ . Similarly, if  $f : \mathbb{M}_1 \rightarrow \mathbb{R}$  is a smooth function, then  $df : \mathbb{M}_1 \rightarrow T^*\mathbb{M}_1$  will denote the differential of the function  $f$ .

In the rest of this section we shall define the basic concepts regarding convex sets and tents which appear later in the statement of the main result. For the sake of brevity, we will omit all proofs in this section.

- Let  $d$  be a positive integer. Recall that a non-empty subset  $K \subset \mathbb{R}^d$  is a **cone** if for every  $y \in K$  and  $\alpha \geq 0$  we have  $\alpha y \in K$ . In particular,  $0 \in \mathbb{R}^d$  belongs to  $K$ . A non-empty subset  $C \subset \mathbb{R}^d$  is **convex** if for every  $y, y' \in C$  and  $\theta \in [0, 1]$  we have  $(1 - \theta)y + \theta y' \in C$ .
- A hyperplane  $\Gamma$  in  $\mathbb{R}^d$  is an  $(d - 1)$ -dimensional affine subset of  $\mathbb{R}^d$ . It can be viewed as the level set of a nontrivial linear function  $p : \mathbb{R}^d \rightarrow \mathbb{R}$ . If  $p$  is given by  $p(x) = \langle a, x \rangle$  for some  $a (\neq 0) \in \mathbb{R}^d$ , then

$$\Gamma := \{x \in \mathbb{R}^d \mid \langle a, x \rangle = \alpha\}.$$

- Let  $\Omega$  be a nonempty set in  $\mathbb{R}^d$ . By  $\text{aff } \Omega$  we denote the set of all affine combinations of points in  $\Omega$ . That is,

$$\text{aff } \Omega = \left\{ \sum_{i=1}^k \theta_i x_i \mid \sum_{i=1}^k \theta_i = 1, \quad x_i \in \Omega \quad \text{for } i = 1, \dots, k, \text{ and } k \in \mathbb{N}^* \right\}.$$

In other words,  $\text{aff } \Omega$  is also the smallest affine set containing  $\Omega$ . The relative interior  $\text{ri } \Omega$  of  $\Omega$  denotes the interior of  $\Omega$  relative to the affine space  $\text{aff } \Omega$ .

- Let  $M$  be a convex set and  $x_0 \in M$ . The union of all the rays emanating from  $x_0$  and passing through points of  $M$  other than  $x_0$  is a convex cone with vertex at  $x_0$ . The closure of this cone is called the **supporting cone** of  $M$  at  $x_0$ .

We will now provide some definitions associated with the method of tents. Although we will not be directly using the method of tents in the proof of the main result, tents do appear in our final result, and so one needs to be familiar at least with the basic definition of what a tent is.

**Definition 2.1.** Let  $\Omega$  be a subset of  $\mathbb{R}^d$  and let  $x_0 \in \Omega$ . A convex cone  $Q \subset \mathbb{R}^d$  with vertex  $x_0$  is a **tent** of  $\Omega$  at  $x_0$  if there exists a smooth map  $\rho$  defined in a neighbourhood of  $x_0$  such that:<sup>1</sup>

- (1)  $\rho(x) = x + o(x - x_0)$ ,<sup>2</sup> and
- (2) there exists  $\epsilon > 0$  such that  $\rho(x) \in \Omega$  for  $x \in Q \cap B_\epsilon(x_0)$ .

We say that a convex cone  $K \subset \mathbb{R}^d$  with vertex at  $x_0$  is a **local tent** of  $\Omega$  at  $x_0$  if for every  $x \in \text{ri } K$  there is a convex cone  $Q \subset K$  with vertex at  $x_0$  such that  $Q$  is a tent of  $\Omega$  at  $x_0$ ,  $x \in \text{ri } Q$ , and  $\text{aff } Q = \text{aff } K$ . Observe that if  $K$  is a tent of  $\Omega$  at  $x_0$ , then  $K$  is a local tent of  $\Omega$  at  $x_0$ .

A tent to a set at a point is just a linear approximation of the set about the point. Intuitively, it is the set of directions along which it is possible to enter the set from the point. This intuition is reinforced through the following theorems which characterize the tents of some sets which appear commonly in applications.

**Theorem 2.1** ([Bol75, Theorem 8 on p. 11]). *Let  $\Omega$  be a smooth manifold in  $\mathbb{R}^d$  and  $K$  the tangent plane to  $\Omega$  at  $x_0 \in \Omega$ . Then  $K$  is a tent of  $\Omega$  at  $x_0$ .*

**Theorem 2.2** ([Bol75, Theorem 9 on p. 12]). *Given a smooth function  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ , let  $x_0$  be such that  $D_x \varphi(x_0) \neq 0$ . Define sets  $\Omega, \Omega_0 \in \mathbb{R}^d$  as*

$$\Omega := \{x \in \mathbb{R}^d \mid \varphi(x) \leq \varphi(x_0)\}, \quad \Omega_0 := \{x_0\} \cup \{x \in \mathbb{R}^d \mid \varphi(x) < \varphi(x_0)\}.$$

*Then the half-space  $K$  given by the inequality  $\langle D_x \varphi(x_0), x - x_0 \rangle \leq 0$  is a tent of both  $\Omega$  and  $\Omega_0$  at  $x_0$ .*

**Theorem 2.3** ([Bol75, Theorem 10 on p. 12]). *Let  $\Omega \subset \mathbb{R}^d$  be a convex set and let  $K$  be its supporting cone at  $x_0 \in \Omega$ . Then  $K$  is a local tent of  $\Omega$  at  $x_0$ .*

We will also need the following two theorems regarding embedded submanifolds for the proof of our main result.

**Theorem 2.4** ([Lee13, Theorem 6.15 on p. 134]). *Every smooth  $n$ -manifold admits an embedding into  $\mathbb{R}^{2n+1}$  as a closed submanifold.*

**Theorem 2.5** ([Lee13, Lemma 5.34 on p. 115]). *Let  $\mathbb{M}$  be an  $n$ -dimensional smooth manifold and  $i : \mathbb{M} \rightarrow \mathbb{R}^N$  be a smooth embedding such that  $i(\mathbb{M})$  is a closed subset of  $\mathbb{R}^N$ . If  $f : \mathbb{M} \rightarrow \mathbb{R}$  is a smooth function, there exists a smooth function  $\tilde{f} : \mathbb{R}^N \rightarrow \mathbb{R}$  such that  $f = \tilde{f} \circ i$ .*

<sup>1</sup>The theory also works for  $\rho$  continuous.

<sup>2</sup>Recall the Landau notation  $\varphi(x) = o(x)$  that stands for a function  $\varphi(0) = 0$  and  $\lim_{x \rightarrow 0} \frac{|\varphi(x)|}{|x|} = 0$ .

## §3. PROBLEM SETUP

Consider a discrete time control system evolving on an  $n$  dimensional smooth manifold  $\mathbb{M}$  described by

$$(3.1) \quad x_{t+1} = f_t(x_t, u_t) \quad \text{for } t = 0, \dots, T-1,$$

where  $x_t \in \mathbb{M}$ ,  $u_t \in \mathbb{R}^m$ , and  $(f_t)_{t=0}^{T-1}$  is a family of maps such that  $\mathbb{M} \times \mathbb{R}^m \ni (\xi, \mu) \mapsto f_s(\xi, \mu) \in \mathbb{M}$  is continuously differentiable for each  $s = 0, \dots, T-1$ . We emphasize that the condition  $x_t \in \mathbb{M}$  is not being enforced as a constraint;  $\mathbb{M}$  is the natural state space of the control system (3.1). To wit, it is an intrinsic property of the family of the dynamics  $(f_t)_{t=0}^{T-1}$  that any trajectory of (3.1) starting on the manifold  $\mathbb{M}$  lies entirely on  $\mathbb{M}$ .

Let  $u^k := (u_t^k)_{t=0}^{T-1}$  denote the  $k^{\text{th}}$  control sequence, and  $\widehat{u}^k$  denote its discrete Fourier transform (DFT). The relationship between  $\widehat{u}^k$  and  $u^k$  is given by [SS03, Chapter 7]:

$$(3.2) \quad \widehat{u}^k := (\widehat{u}_\xi^k)_{\xi=0}^{T-1} = \left( \sum_{t=0}^{T-1} u_t^k e^{-i2\pi\xi t/T} \right)_{\xi=0}^{T-1} \quad \text{for } \xi = 0, \dots, T-1$$

and  $k = 1, \dots, m$ .

Let  $T \in \mathbb{N}^*$  be fixed. The objective of this article is to provide first-order necessary conditions of a finite horizon constrained optimal control problem with continuously differentiable stage cost, terminal cost, and inequality and equality constraints. We write our abstract optimal control problem as:

$$(3.3) \quad \begin{aligned} & \underset{(u_t)_{t=0}^{T-1}}{\text{minimize}} && \sum_{t=0}^{T-1} c_t(x_t, u_t) + c_T(x_T) \\ & \text{subject to} && \begin{cases} \text{dynamics (3.1),} \\ \text{state constraints at each stage } t = 0, \dots, T, \\ \text{control constraints at each stage } t = 0, \dots, T-1, \\ \text{constraints on frequency components of the control sequence.} \end{cases} \end{aligned}$$

where  $\mathbb{M} \ni \xi \mapsto c_T(\xi) \in \mathbb{R}$  and  $\mathbb{M} \times \mathbb{R}^m \ni (\xi, \mu) \mapsto c_t(\xi, \mu) \in \mathbb{R}$  are continuously differentiable functions representing the terminal cost and stage cost at time  $t$  respectively, for  $t = 0, \dots, T-1$ .

The three types of constraints considered in the optimal control problem (3.3) are as follows:

- (i) *State constraints:* Let  $(g_t)_{t=1}^T$  be a family of maps such that  $\mathbb{M} \ni \xi \mapsto g_s(\xi) \in \mathbb{R}^{n_s}$  is continuously differentiable for each  $s = 0, \dots, T$ . We restrict the trajectories of the states  $(x_t)_{t=0}^T$  to be such that

$$x_0 = q_0 \text{ and } g_t(x_t) \leq 0 \quad \text{for } t = 1, \dots, T.$$

- (ii) *Control constraints:*  $\mathbb{U}_t \subset \mathbb{R}^m$  is a given but otherwise arbitrary non-empty set for each  $t = 0, \dots, T$ . We impose the requirement that the control action  $u_t$  at stage  $t$  must lie in  $\mathbb{U}_t$ :

$$(3.4) \quad u_t \in \mathbb{U}_t \quad \text{for } t = 0, \dots, T-1.$$

- (iii) *Frequency constraints*: For the  $k^{\text{th}}$  component of the control sequence  $u^k$ , we define  $\mathbb{F}^k \subset \mathbb{C}^T$  to be the set of admissible frequency components of its discrete Fourier transform (DFT)  $\widehat{u}^k = (\widehat{u}^k_\xi)_{\xi=0}^{T-1}$ . For a vector  $v \in \mathbb{C}^T$  we define its support as

$$\text{supp}(v) := \{i \in \{1, \dots, T\} \mid v_i \neq 0\},$$

and stipulate that

$$(3.5) \quad \widehat{u}^k \in \mathbb{F}^k := \{v \in \mathbb{C}^T \mid \text{supp}(v) \subset W^k\},$$

where  $W^k \subset \{1, \dots, T\}$  represents the support for the admissible frequencies in the  $k^{\text{th}}$  control sequence. The sets  $(W^k)_{k=1}^m$  are assumed to be given as part of the problem specification. In effect the constraint (3.5) ensures that the frequency spectrum of the  $k^{\text{th}}$  component of the control sequence does not contain any non-zero entries lying outside the set  $W^k$ . Frequency constraints of the form (3.5) are required in applications where the designer is required to suppress certain undesirable frequency components in the control sequence. For instance, in satellites with flexible structures attached to them, damages to such structures may occur if their natural frequencies are excited in course of their motion. In such a situation it is essential to avoid the natural frequencies of the structures from the spectrum of the control trajectories, and such constraints are ensured precisely by constraints of the form (3.5). It can be shown that (3.5) can be recast into a more condensed form as

$$(3.6) \quad F(u_0, \dots, u_{T-1}) = \sum_{t=0}^{T-1} F_t u_t = 0 \quad , \text{where } F_t \in \mathbb{R}^{\ell \times m}, \text{ for some } \ell \in \mathbb{N}^*,$$

where the matrices  $(F_t)_{t=0}^{T-1}$  depend on the sets  $(W^k)_{k=1}^m$ . For a more detailed discussion on how this transformation can be done, we refer the reader to [PC19]. We shall refer to  $F$  as our *frequency constraint map*.

The abstract optimal control problem (3.3) can now be formally written as:

$$(3.7) \quad \begin{aligned} & \underset{(u_t)_{t=0}^{T-1}}{\text{minimize}} && \sum_{t=0}^{T-1} c_t(x_t, u_t) + c_T(x_T) \\ & \text{subject to} && \begin{cases} \text{dynamics (3.1),} \\ x_0 = q_0, \\ g_t(x_t) \leq 0 \quad \text{for } t = 1, \dots, T, \\ u_t \in \mathbb{U}_t \quad \text{for } t = 0, \dots, T-1, \\ \sum_{t=0}^{T-1} F_t u_t = 0 \end{cases} \end{aligned}$$

An optimal solution  $(u_t^\circ)_{t=0}^{T-1}$  of (3.7) is a sequence in  $\prod_{i=0}^{T-1} \mathbb{U}_i$ , and it generates its corresponding optimal state trajectory  $(x_t^\circ)_{t=0}^T$  according to (3.1). The pair  $((x_t^\circ)_{t=0}^T, (u_t^\circ)_{t=0}^{T-1})$  is called an *optimal state-action trajectory*.

At this point we make note of the following notational convention in effect throughout the sequel:  $T_x^* f_t(x_0, u_0)$  will denote the cotangent lift of the map  $f_t(\cdot, u_0) : \mathbb{M} \rightarrow \mathbb{M}$  at  $x_0$  and  $T_u^* f_t(x_0, u_0)$  will denote the cotangent lift of the map  $f_t(x_0, \cdot) : \mathbb{R}^m \rightarrow \mathbb{M}$  at  $u_0$ . Similarly,  $d_x c_t(x_0, u_0)$  will denote the differential of the map  $c_t(\cdot, u_0) : \mathbb{M} \rightarrow \mathbb{R}$  at  $x_0$  and  $d_u c_t(x_0, u_0)$  will denote the differential of the map  $c_t(x_0, \cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$  at  $u_0$ .

## §4. MAIN RESULT

The following theorem provides first order necessary conditions for optimal solutions of (3.7); it is the main result of this article.

**Theorem 4.1** (PMP on smooth manifolds). *Let  $((x_t^\circ)_{t=0}^T, (u_t^\circ)_{t=0}^{T-1})$  be an optimal state-action trajectory for (3.7). Then there exist*

- a trajectory  $(\eta_t^f)_{t=1}^T \subset T^*\mathbb{M}$  with  $\eta_t^f \in T_{x_t^\circ}^*\mathbb{M}$  for each  $t$  (the adjoint trajectory),
- a sequence  $(\eta_t^x)_{t=1}^T$  with  $\eta_t^x \in (\mathbb{R}^{n_t})^*$  for each  $t$  (the multipliers corresponding to the point-wise state constraints), and
- a pair  $(\eta^C, \widehat{\eta}^u) \in \mathbb{R} \times (\mathbb{R}^\ell)^*$  (the abnormal multiplier and the multiplier corresponding to the frequency constraints, respectively),

satisfying the following conditions:

(PMP-i) non-negativity:

$$\eta^C \geq 0, (\eta_t^x)_{t=1}^T \geq 0;$$

(PMP-ii) non-triviality:

the sequence  $(\eta_t^x)_{t=1}^T$  and the pair  $(\eta^C, \widehat{\eta}^u)$  do not simultaneously vanish;

(PMP-iii) state and adjoint system dynamics

$$x_{t+1}^\circ = f_t(x_t^\circ, u_t^\circ) \quad \text{for } t = 0, \dots, T-1,$$

$$\eta_t^f = T_x^* f_t(x_t^\circ, u_t^\circ) \eta_{t+1}^f - \eta^C d_x c_t(x_t^\circ, u_t^\circ) - T^* g_t(x_t^\circ) \eta_t^x \quad \text{for } t = 1, \dots, T-1;$$

(PMP-iv) transversality:

$$\eta_T^f = -\eta^C d c_T(x_T^\circ) - T^* g_T(x_T^\circ) \eta_T^x;$$

(PMP-v) Hamiltonian maximization, point-wise in time,

$$\left\langle T_u^* f_t(x_t^\circ, u_t^\circ) \eta_{t+1}^f - \eta^C d_u c_t(x_t^\circ, u_t^\circ) + F_t^T \widehat{\eta}^u, \widehat{u}_t \right\rangle \leq 0$$

whenever  $u_t^\circ + \widehat{u}_t \in q_t^u(u_t^\circ)$ , where  $q_t^u(u_t^\circ)$  is a local tent at  $u_t^\circ$  of the set  $\mathbb{U}_t$  of admissible actions;

(PMP-vi) complementary slackness:

$$(\eta_t^x)^j g_t^j(x_t^\circ) = 0 \quad \text{for all } j \in [n_t].$$

We present a complete but elementary proof of [Theorem 4.1](#) in [§5](#).

**Discussion.** The rest of this section is devoted to a scrutiny of various facets of [Theorem 4.1](#) over a sequence of remarks.

*Remark 4.1.* One of the points of departure of [Theorem 4.1](#) from the Euclidean version of the PMP given in [[PC19](#), Theorem 3.1] is (PMP-v). To wit, there appears to be no natural way of defining a Hamiltonian function analogous to the one given in [[PC19](#), Theorem 3.1] in the geometric framework. It is also worth noting that the absence of a natural Hamiltonian is peculiar to the discrete time setting since a Hamiltonian function arises naturally in the continuous time geometric PMP. Indeed, in continuous time a key element in the definition of the Hamiltonian is the duality product between the adjoint trajectory lying on the cotangent bundle and the tangent vector field along the optimal state trajectory lying on the tangent bundle. In the discrete time geometric setting, however, the

adjoint trajectory remains on the cotangent bundle of the manifold, but the tangent vector field is replaced by a discrete trajectory lying on the manifold itself. Since there is no natural product (pairing) between an element of the cotangent bundle and an element of the manifold, a natural definition of a Hamiltonian is difficult to arrive at.

*Remark 4.2.* It is not entirely appropriate to use the term ‘‘Hamiltonian maximization condition’’ for (PMP-v); we have not even defined a Hamiltonian function here, let alone derive a maximization condition. We still use this name for the condition because it is analogous to the actual Hamiltonian maximization condition in the continuous time counterpart of the PMP. However, such a maximization condition does hold under additional structural assumptions on the sets of admissible actions and regularity assumptions on the cost and transition maps. We refer the reader to [KG17, §3.1] for a detailed exposition on this.

*Remark 4.3.* The non-triviality condition (PMP-ii) stated here is somewhat non-standard. The non-triviality condition is usually stated as the adjoint trajectory  $(\eta_t^f)_{t=1}^T$  and the pair  $(\eta^C, \widehat{\eta}^u)$  do not simultaneously vanish. The condition given in (PMP-ii) is slightly weaker than the standard non-triviality condition; if  $((\eta_t^f)_{t=1}^T, \eta^C, \widehat{\eta}^u)$  could not simultaneously vanish, then  $((\eta_t^x)_{t=1}^T, \eta^C, \widehat{\eta}^u)$  would not vanish simultaneously either, since if it did, by (PMP-iii) and (PMP-iv)  $((\eta_t^f)_{t=1}^T, \eta^C, \widehat{\eta}^u)$  would also vanish simultaneously.

However, under the additional assumption of the constraints  $g_t$  being regular (as defined in Definition 5.1) at  $x_t^\circ$ , the condition stated in (PMP-ii) is equivalent to the standard non-triviality condition. Suppose  $((\eta_t^f)_{t=1}^T, \eta^C, \widehat{\eta}^u)$  did vanish simultaneously, then by (PMP-iii) and (PMP-iv), we get that

$$T^* g_t(x_t^\circ) \eta_t^x = 0.$$

Also, by (PMP-i) and (PMP-vi), we have

$$\eta_t^x \geq 0 \quad \text{and} \quad (\eta_t^x)^j g_t^j(x_t^\circ) = 0 \quad \text{for all } j \in [n_t].$$

If the constraints  $g_t$  are regular at  $x_t^\circ$ , the only  $\eta_t^x$  satisfying these three conditions will be  $\eta_t^x = 0$ . Therefore,  $((\eta_t^x)_{t=1}^T, \eta^C, \widehat{\eta}^u)$  would also vanish simultaneously, contradicting (PMP-ii).

*Remark 4.4.* First order necessary conditions for locally optimal solutions of finite dimensional constrained optimization problems (such as the KKT conditions) usually accompany a ‘‘constraint qualification’’ condition which at first glance is completely absent in our discussion. The difference between conditions (PMP-i) - (PMP-vi) and the standard KKT conditions is the presence of the abnormal multiplier  $\eta^C$ . Observe that (PMP-i) only guarantees that  $\eta^C \geq 0$ , it is still possible that  $\eta^C = 0$ . When  $\eta^C = 0$  we arrive at an ‘‘abnormal’’ situation where the necessary conditions (PMP-i) - (PMP-vi) no longer depend on either the stage costs or the terminal cost; this situation arises typically when the constraints are so tight that the cost functions play no key role in the determination of the optimizer(s). In the context of the PMP, constraint qualification conditions serve the purpose of strengthening the conditions of Theorem 4.1 by guaranteeing that  $\eta^C$  is non-zero thereby precluding the aforementioned abnormal situation. Due to the presence of the abnormal multiplier, the PMP as presented in Theorem 4.1 holds regardless of any such constraint qualification conditions.

*Remark 4.5.* The conditions (PMP-i) - (PMP-vi) together constitute a well-defined two point boundary value problem, with (PMP-iv) along with the initial condition  $x_0 = q_0$  giving

the entire set of boundary conditions. Algorithms based on Newton step methods may be employed to solve this (algebraic) two point boundary value problem; see, eg., [Tré12, §2.4] for an illuminating discussion in the context of continuous-time problems. Fast solution techniques for two point boundary value problems is an active field of research.

## §5. PROOF OF THE MAIN RESULT

We present a proof of [Theorem 4.1](#) through the following steps:

- **Step 1:** The configuration manifold is embedded in a Euclidean space and we convert (3.7) into an equivalent optimal control problem on this Euclidean space.
- **Step 2:** First order necessary conditions for the equivalent problem on the Euclidean space are applied to the problem in Step 1.
- **Step 3:** The necessary conditions in Step 2 are lifted back to the original manifold.

**§ 5.1. Step 1.** By [Theorem 2.4](#), one can find a smooth embedding of  $\mathbb{M}$  in  $\mathbb{R}^N$ , where  $N = 2n + 1$ , such that the image of the embedding is a *closed* subset of  $\mathbb{R}^N$ . Let  $i : \mathbb{M} \rightarrow \mathbb{R}^N$  denote such a smooth embedding.

We observe that  $i(\mathbb{M}) \times \mathbb{U} \ni (\xi, \mu) \mapsto i \circ f_t(i^{-1}(\xi), \mu) \in \mathbb{R}^N$  is a smooth map from a closed subset of  $\mathbb{R}^N \times \mathbb{R}^m$  to  $\mathbb{R}^N$ . Hence, it can be extended to a smooth map  $\tilde{f}_t : \mathbb{R}^N \times \mathbb{R}^m \rightarrow \mathbb{R}^N$  on the whole of  $\mathbb{R}^N$ . Similarly,  $i(\mathbb{M}) \ni \xi \mapsto g_t \circ i^{-1}(\xi) \in \mathbb{R}^{n_r}$  and  $i(\mathbb{M}) \ni \xi \mapsto c_t \circ i^{-1}(\xi) \in \mathbb{R}$  are smooth maps from a closed subset of  $\mathbb{R}^N$  to  $\mathbb{R}^{n_r}$  and  $\mathbb{R}$ , and hence they can be extended to corresponding smooth maps  $\tilde{g}_t : \mathbb{R}^N \rightarrow \mathbb{R}^{n_r}$  and  $\tilde{c}_t : \mathbb{R}^N \rightarrow \mathbb{R}$ .

Now let us define an extended optimal control problem

$$(5.1) \quad \begin{aligned} & \underset{(u_t)_{t=0}^{T-1}}{\text{minimize}} && \sum_{t=0}^{T-1} \tilde{c}_t(\tilde{x}_t, u_t) + \tilde{c}_T(\tilde{x}_T) \\ & \text{subject to} && \begin{cases} \tilde{x}_{t+1} = \tilde{f}_t(\tilde{x}_t, u_t) & \text{for } t = 0, \dots, T-1, \\ \tilde{x}_0 = \tilde{q}_0 = i(q_0), \\ \tilde{g}_t(\tilde{x}_t) \leq 0 & \text{for } t = 1, \dots, T, \\ u_t \in \mathbb{U}_t & \text{for } t = 0, \dots, T-1, \\ \sum_{t=0}^{T-1} F_t u_t = 0. \end{cases} \end{aligned}$$

If  $((x_t)_{t=0}^T, (u_t)_{t=0}^{T-1})$  is a feasible state-action trajectory of (3.7), then  $((i(x_t))_{t=0}^T, (u_t)_{t=0}^{T-1})$  is, clearly, a feasible state-action trajectory of (5.1). If  $((\tilde{x}_t)_{t=0}^T, (u_t)_{t=0}^{T-1})$  is a feasible state-action trajectory of (3.7), then  $(\tilde{x}_t)_{t=0}^T \subset i(\mathbb{M})$ , since  $\tilde{x}_0 = i(q_0) \in i(\mathbb{M})$  and  $\tilde{f}_t$  is an extension of  $i \circ f_t$ . So, the state-action trajectory  $((x_t)_{t=0}^T, (u_t)_{t=0}^{T-1})$  is a feasible solution of (3.7) if and only if  $((i(x_t))_{t=0}^T, (u_t)_{t=0}^{T-1})$  is a feasible solution of (5.1). It is also straightforward to see that the cost incurred by the trajectory  $((x_t)_{t=0}^T, (u_t)_{t=0}^{T-1})$  is the same as that incurred by  $((i(x_t))_{t=0}^T, (u_t)_{t=0}^{T-1})$ . Therefore, the state-action trajectory  $((x_t^\circ)_{t=0}^T, (u_t^\circ)_{t=0}^{T-1})$  is an optimal solution of (3.7) if and only if  $((i(x_t^\circ))_{t=0}^T, (u_t^\circ)_{t=0}^{T-1})$  is an optimal solution of (5.1).

**§5.2. Step 2.** In this step we find first order necessary conditions satisfied by a solution of 5.1. To this end, we define the set  $\Omega_t := \{x \in \mathbb{R}^N \mid \tilde{g}_t(x) \leq 0\}$ . For  $x \in \Omega_t$  we define the *active set* of indices  $A_t(x) := \{i \in [n_t] \mid \tilde{g}_t^i(x) = 0\}$ .

*Definition 5.1.* Let  $\varphi : \mathbb{M} \rightarrow \mathbb{R}^d$  be a smooth map from a smooth manifold  $\mathbb{M}$  to  $\mathbb{R}^d$ . We say that  $\varphi$  is **regular** at  $x_0 \in \mathbb{M}$  if the only  $\mu \in T_{\varphi(x_0)}^* \mathbb{R}^d$  satisfying the three conditions

- (i)  $T^* \varphi(x_0) \mu = 0$ ,
- (ii)  $\mu \geq 0$ , and
- (iii)  $\mu^i \varphi^i(x_0) = 0$  for all  $i \in [d]$ ,

is  $\mu = 0$ .

**Proposition 5.1.** *If  $\tilde{g}_t$  is regular at  $x_0$ , then the closed convex cone*

$$K_t(x_0) := \{x \in \mathbb{R}^N \mid T\tilde{g}_t^i(x_0)(x - x_0) \leq 0 \text{ for all } i \in A_t(x_0)\}$$

*is a tent of  $\Omega_t$  at  $x_0$ . Moreover, the closed convex cone*

$$K_t(x_0)^* := \{T^* \tilde{g}_t(x_0) \mu \mid \mu \geq 0, \mu^i \tilde{g}_t^i(x_0) = 0\}$$

*is the dual cone of  $K_t(x_0)$ .*

*Proof.* Define  $\Omega_t^i := \{x \in \mathbb{R}^N \mid \tilde{g}_t^i(x) \leq 0\}$ . Two cases arise. If  $i \in A_t(x_0)$ , then the closed convex cone

$$K_t(x_0)^i := \{x \in \mathbb{R}^N \mid T\tilde{g}_t^i(x_0)(x - x_0) \leq 0\} \text{ is a tent of } \Omega_t^i \text{ at } x_0.$$

If  $i \notin A_t(x_0)$ , then  $x_0$  lies in the interior of  $\Omega_t^i$ , and therefore  $K_t(x_0)^i = \mathbb{R}^N$  is a tent of  $\Omega_t^i$  at  $x_0$ . The condition that  $\tilde{g}_t$  is regular at  $x_0$  is equivalent to [Bol75, Theorem 2] the cones  $K_t(x_0)^i$  being inseparable. Since  $\Omega_t = \bigcap_{i=1}^{n_t} \Omega_t^i$ , it follows from [Bol75, Theorem 11] that  $K_t(x_0) = \bigcap_{i=1}^{n_t} K_t(x_0)^i$  is a tent of  $\Omega_t$  at  $x_0$ , proving the first part of the claim.

The fact that  $K_t(x_0)^*$  is closed and that it is the dual cone of  $K_t(x_0)$  follows from Farkas' lemma as given in [Ber09, Proposition 2.3.1].  $\square$

The notational conventions mentioned earlier will be used in this section also.  $T_{\tilde{x}}^* \tilde{f}_t(x_0, u_0)$  will denote the cotangent lift of the map  $\tilde{f}_t(\cdot, u_0) : \mathbb{R}^N \rightarrow \mathbb{R}^N$  at  $x_0$  and  $T_u^* \tilde{f}_t(x_0, u_0)$  will denote the cotangent lift of the map  $\tilde{f}_t(x_0, \cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^N$  at  $u_0$ . Similarly,  $d_{\tilde{x}} \tilde{c}_t(x_0, u_0)$  will denote the differential of the map  $\tilde{c}_t(\cdot, u_0) : \mathbb{R}^N \rightarrow \mathbb{R}$  at  $x_0$  and  $d_u \tilde{c}_t(x_0, u_0)$  will denote the differential of the map  $\tilde{c}_t(x_0, \cdot) : \mathbb{R}^m \rightarrow \mathbb{R}$  at  $u_0$ .

**Proposition 5.2.** *Let  $((\tilde{x}_t^\circ)_{t=0}^T, (u_t^\circ)_{t=0}^{T-1})$  be an optimal state-action trajectory for (5.1). Then there exist*

- a trajectory  $(\tilde{\eta}_t^f)_{t=1}^T \subset T^* \mathbb{R}^N$  such that  $\tilde{\eta}_t^f \in T_{\tilde{x}_t^\circ}^* \mathbb{R}^N$ ,
- a sequence  $(\tilde{\eta}_t^x)_{t=1}^T$  such that  $\tilde{\eta}_t^x \in (\mathbb{R}^{n_t})^*$ , and
- a pair  $(\tilde{\eta}^C, \tilde{\eta}^u) \in \mathbb{R} \times (\mathbb{R}^\ell)^*$ ,

*satisfying the following conditions:*

(EPMP-i) *non-negativity condition*

$$\tilde{\eta}^C \geq 0, (\tilde{\eta}_t^x)_{t=1}^T \geq 0;$$

(EPMP-ii) *non-triviality condition*

*the multipliers  $(\tilde{\eta}_t^x)_{t=1}^T$  and the pair  $(\tilde{\eta}^C, \tilde{\eta}^u)$  do not simultaneously vanish;*

(EPMP-iii) *state and adjoint system dynamics*

$$\tilde{x}_{t+1}^\circ = \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \quad \text{for } t = 0, \dots, T-1,$$

$$\tilde{\eta}_t^f = T_{\tilde{x}}^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f - \tilde{\eta}^C d_{\tilde{x}} \tilde{c}_t(\tilde{x}_t^\circ, u_t^\circ) - T^* \tilde{g}_t(\tilde{x}_t^\circ) \tilde{\eta}_t^x \quad \text{for } t = 1, \dots, T-1;$$

(EPMP-iv) *transversality conditions*

$$\tilde{\eta}_T^f = -\tilde{\eta}^C d\tilde{c}_T(\tilde{x}_T^\circ) - T^* \tilde{g}_T(\tilde{x}_T^\circ) \tilde{\eta}_T^x;$$

(EPMP-v) *Hamiltonian maximization condition, point-wise in time,*

$$\left\langle T_u^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f - \eta^C d_u \tilde{c}_t(\tilde{x}_t^\circ, u_t^\circ) + F_t^T \tilde{\eta}^u, \hat{u}_t \right\rangle \leq 0$$

whenever  $u_t^\circ + \hat{u}_t \in q_t^u(u_t^\circ)$ , where  $q_t^u(u_t^\circ)$  is a local tent at  $u_t^\circ$  of the set  $\mathbb{U}_t$  of admissible actions;

(EPMP-vi) *complementary slackness*

$$(\tilde{\eta}_t^x)^j \tilde{g}_t^j(\tilde{x}_t^\circ) = 0 \quad \text{for all } j \in 1, 2, \dots, n_t.$$

*Proof.* Suppose  $\tilde{g}_s$  is not regular at  $\tilde{x}_s^\circ$  for some  $s$ . Then there exists  $\mu \in (\mathbb{R}^{n_s})^*$  such that  $\mu \neq 0$ ,  $T^* \tilde{g}_s(\tilde{x}_s^\circ) \mu = 0$ ,  $\mu \geq 0$ ,  $\mu^i \tilde{g}_s^i(\tilde{x}_s^\circ) = 0$ . We can now take  $\tilde{\eta}_s^x = \mu$ ,  $\tilde{\eta}_t^x = 0$  for all  $t \neq s$ ,  $\tilde{\eta}^C = 0$ ,  $\tilde{\eta}^u = 0$ ,  $\tilde{\eta}_t^f = 0$  and the conditions of Proposition 5.2 hold trivially.

If not, we can say that for all  $t \in [T]$ ,  $\tilde{g}_t$  is regular at  $\tilde{x}_t^\circ$ . From Proposition 5.1 we know that the set

$$K_t(x_0)^* := \{T^* \tilde{g}_t(x_0) \mu_t \mid \mu_t^i \geq 0, \mu_t^i \tilde{g}_t^i(x_0) = 0\}$$

is the dual cone of a tent to the set  $\Omega_t = \{x \in \mathbb{R}^N \mid \tilde{g}_t(x) \leq 0\}$ . It follows now that Proposition 5.2 is just a restatement of [PC19, Proposition C.6] except for the condition (EPMP-ii). Suppose the multipliers  $(\tilde{\eta}_t^x)_{t=1}^T$  and the pair  $(\tilde{\eta}^C, \tilde{\eta}^u)$  vanish simultaneously, then from the transversality condition (EPMP-iv),

$$\tilde{\eta}_T^f = -\tilde{\eta}^C d\tilde{c}_T(\tilde{x}_T^\circ) - T^* \tilde{g}_T(\tilde{x}_T^\circ) \tilde{\eta}_T^x = 0,$$

and from the adjoint dynamics (EPMP-iii),

$$\begin{aligned} \tilde{\eta}_t^f &= T_{\tilde{x}}^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f - \tilde{\eta}^C d_{\tilde{x}} \tilde{c}_t(\tilde{x}_t^\circ, u_t^\circ) - T^* \tilde{g}_t(\tilde{x}_t^\circ) \tilde{\eta}_t^x \\ &= T_{\tilde{x}}^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f \quad \text{for } t = 1, \dots, T-1. \end{aligned}$$

It follows that the trajectory  $(\tilde{\eta}_t^f)_{t=1}^T$  also vanishes. This contradicts the non-triviality condition given in [PC19, Proposition C.6], and proves (EPMP-ii).  $\square$

**§ 5.3. Step 3.** The necessary conditions we arrived at in Proposition 5.2 depends on both the particular embedding of  $i : \mathbb{M} \mapsto \mathbb{R}^N$  and the extensions  $(\tilde{g}_t)_{t=1}^T, (\tilde{f}_t)_{t=0}^{T-1}, (\tilde{c}_t)_{t=0}^T$ ; this isn't desirable. In this step we finally arrive at the conditions in Theorem 4.1 from the conditions in Proposition 5.2.

*Proof of Theorem 4.1.* Define  $\eta_t^f := T^* i(x_t^\circ) \tilde{\eta}_t^f$ ,  $\eta_t^x := \tilde{\eta}_t^x$ ,  $\eta^C = \tilde{\eta}^C$ . Then

$$\begin{aligned} \eta_t^f &= T^* i(x_t^\circ) \tilde{\eta}_t^f \\ &= T^* i(x_t^\circ) (T_{\tilde{x}}^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f - \tilde{\eta}^C d_{\tilde{x}} \tilde{c}_t(\tilde{x}_t^\circ, u_t^\circ) - T^* \tilde{g}_t(\tilde{x}_t^\circ) \tilde{\eta}_t^x) \\ &= T_x^* (\tilde{f}_t \circ i)(x_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f - \eta^C d_x (\tilde{c}_t \circ i)(\tilde{x}_t^\circ, u_t^\circ) - T^* (\tilde{g}_t \circ i)(x_t^\circ) \eta_t^x. \end{aligned}$$

Since  $\tilde{c}_t$  and  $\tilde{g}_t$  are extensions of  $c_t$  and  $g_t$  respectively,  $\tilde{c}_t \circ i = c_t$ ,  $\tilde{g}_t \circ i = g_t$ ,  $\tilde{f}_t \circ i = i \circ f_t$ . Also,  $\eta_t^x = \tilde{\eta}_t^x$ ,  $\eta^C = \tilde{\eta}^C$  by definition. Therefore,

$$\begin{aligned} \eta_t^f &= T_x^* (i \circ f_t)(x_t^\circ, u_t^\circ) \tilde{\eta}_{t+1}^f - \eta^C d_x c_t(\tilde{x}_t^\circ, u_t^\circ) - T^* g_t(x_t^\circ) \eta_t^x \\ &= T_x^* f_t(x_t^\circ, u_t^\circ) T_x^* i(x_{t+1}^\circ) \tilde{\eta}_{t+1}^f - \eta^C d_x c_t(\tilde{x}_t^\circ, u_t^\circ) - T^* g_t(x_t^\circ) \eta_t^x. \end{aligned}$$

We now conclude that

$$\begin{aligned}\eta_t^f &= T_x^* f_t(x_t^\circ, u_t^\circ) \eta_{t+1}^f - \eta^C d_x c_t(x_t^\circ, u_t^\circ) - T^* g_t(x_t^\circ) \eta_t^x, \\ \eta_T^f &= -\eta^C d c_T(x_T^\circ, u_T^\circ) - T^* g_T(x_T^\circ) \eta_T^x.\end{aligned}$$

This proves (PMP-iii) and (PMP-iv). Since  $\tilde{c}_t(\tilde{x}_t^\circ, u) = c_t(x_t^\circ, u)$  for all  $u \in \mathbb{R}^m$ ,

$$d_u \tilde{c}_t(\tilde{x}_t^\circ, u) = d_u c_t(x_t^\circ, u).$$

Since  $i \circ f_t(x_t^\circ, u) = \tilde{f}_t(\tilde{x}_t^\circ, u)$  for all  $u \in \mathbb{R}^m$ ,

$$\begin{aligned}T_u^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_t^f &= T_u^*(i \circ f_t)(x_t^\circ, u_t^\circ) \tilde{\eta}_t^f \\ &= T_u^* f_t(x_t^\circ, u_t^\circ) T_x^* i(x_t^\circ) \tilde{\eta}_t^f \\ &= T_u^* f_t(x_t^\circ, u_t^\circ) \eta_t^f.\end{aligned}$$

Therefore,

$$\begin{aligned}\left\langle T_u^* f_t(x_t^\circ, u_t^\circ) \eta_t^f - \eta^C d_u c_t(x_t^\circ, u_t^\circ) + F_t^T \hat{\eta}^u, \hat{u} \right\rangle = \\ \left\langle T_u^* \tilde{f}_t(\tilde{x}_t^\circ, u_t^\circ) \tilde{\eta}_t^f - \tilde{\eta}^C d_u \tilde{c}_t(\tilde{x}_t^\circ, u_t^\circ) + F_t^T \hat{\eta}^u, \hat{u} \right\rangle \leq 0\end{aligned}$$

whenever  $u_t^\circ + \hat{u}_t \in q_t^u(u_t^\circ)$ , where  $q_t^u(u_t^\circ)$  is a local tent at  $u_t^\circ$  of the set  $\mathbb{U}_t$  of admissible actions. This proves (PMP-v). (PMP-i), (PMP-ii), and (PMP-vi) are just restatements of (EPMP-i), (EPMP-ii), and (EPMP-vi).  $\square$

## REFERENCES

- [AS04] A. Agrachev and Y. L. Sachkov, *Control Theory from the Geometric Viewpoint*, Springer-Verlag, Berlin, 2004.
- [Ber09] D. P. Bertsekas, *Convex Optimization Theory*, Athena Scientific, Belmont, 2009.
- [Bol71] V. G. Boltyanskii, *Mathematical Methods of Optimal Control*, Holt, Rinehart and Winston, Inc., New York-Montreal, Que.-London, 1971, Translated from the Russian by K. N. Tirogoff. Edited by Ivin Tarnove, Baskrishnan-Neustadt Series.
- [Bol75] ———, *The method of tents in the theory of extremal problems*, Russian Mathematical Surveys **30** (1975), no. 3, 1–54.
- [Bol78] ———, *Optimal Control of Discrete Systems*, John Wiley & Sons, 1978.
- [Cha11] D. E. Chang, *A simple proof of the Pontryagin maximum principle on manifolds*, Automatica **47** (2011), 630–633.
- [KG17] Robert Kipka and Rohit Gupta, *The discrete-time geometric maximum principle*, <https://arxiv.org/abs/1707.03873>, 2017.
- [Lee13] J. M. Lee, *Introduction to Smooth Manifolds*, 2nd ed., Springer, New York, 2013.
- [LM67] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, John Wiley & Sons, 1967.
- [PBG62] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, John Wiley & Sons, 1962.
- [PC19] P. Paruchuri and D. Chatterjee, *Discrete time Pontryagin maximum principle for optimal control problems under state-action-frequency constraints*, IEEE Transactions on Automatic Control **64** (2019), no. 9, doi: <https://doi.org/10.1109/TAC.2019.2893160>.
- [PCB18] K. S. Phogat, D. Chatterjee, and R. Banavar, *Discrete-time Pontryagin maximum principle on matrix Lie groups*, Automatica **97** (2018), 376–391.
- [SS03] E. M. Stein and R. Shakarchi, *Fourier Analysis, an Introduction*, Princeton University Press, Princeton, NJ, 2003.
- [Sus98] H. J. Sussmann, *An introduction to the coordinate-free maximum principle*, Geometry of Feedback and Optimal Control, Monographs and Textbooks in Pure and Applied Mathematics, vol. 207, Dekker, New York, 1998, pp. 463–557.
- [Tré12] E. Trélat, *Optimal control and applications to aerospace: some results and challenges*, Journal of Optimization Theory and Applications **154** (2012), no. 3, 713–758.