# Convexifying State-Constrained Optimal Control Problem

Donggun Lee ©, Shankar A. Deka ©, *Member, IEEE*, and Claire J. Tomlin ©, *Fellow, IEEE*

*Abstract*—**This article presents a method that convexifies state-constrained optimal control problems in the control-input space. The proposed method enables convex programming methods to find the globally optimal solution even if costs and control constraints are nonconvex in control and convex in state, dynamics is nonaffine in control and convex in state, and state constraints are convex in state. Under the above conditions, generic methods do not guarantee to find optimal solutions, but the proposed method does. The proposed approach is demonstrated in a 16-D navigation example.**

*Index Terms*—**Nonlinear control systems, optimal control.**

## I. INTRODUCTION

This article focuses on state-constrained optimal control problems (OCPs) for high-dimensional systems. Table I shows popular methods in this field, and their theoretical assumptions for optimality, scability to deal with high-dimensional systems, and existence of numerical methods to guarantee to find optimal solutions. Multiple shooting [1] and collocation methods [2] are direct methods that discretize OCPs into finite-dimensional problems, and then apply numerical optimizers. They are computationally efficient and capable of handling high-dimensional systems. However, they are guaranteed to find optimal solutions only if the OCP's costs and constraints are convex and the dynamics is affine in the state and control. On the other hand, Pontryagin's minimum principle (PMP) [3] is an indirect method that provides a set of necessary conditions for optimality, and then applies numerical methods to find a solution that satisfies the PMP conditions. These PMP conditions are sufficient for optimality under particular conditions [3, Ch. 8], which are satisfied if the OCP's costs and constraints are convex in the state, dynamics is affine in the state, and the control constraint is convex in the control. Thus, the PMP optimality conditions are more general than the ones for the direct methods; however, they can be highly nonconvex and nonlinear, making it challenging for numerical methods to find a solution.

Other generic methods exist, but they are not appropriate for solving state-constrained OCPs for high-dimensional systems. HJ analysis [4] is induced by dynamic programming, so it guarantees optimality for general OCPs. However, its computational complexity grows exponentially in the state dimension; thus, it is typically used for low-dimensional systems. Koopman-based methods [5] lift nonlinear systems to higher dimensional linear systems, but the state-of-the-art Koopman-based methods still have not considered general state constraints. Branch-and-Lift algorithms [6] also provide an $\epsilon$-close approximation to the optimal solution, but they require the exact computation of reachable sets, for which one may need to employ computationally intensive methods based on HJ analysis.

This article aims to relax the optimality conditions in a way similar to the direct methods and PMP, while still guaranteeing to find optimal solutions using numerical solvers. We draw our motivation from convexification of the Hopf–Lax theory [7], which provides a unique viscosity solution to HJ PDEs. Notably, the Hopf–Lax formula [7] for a particular class of HJ PDEs is a convex optimization problem if the terminal function is convex in the state. We observe that some OCPs have nonconvex costs in the control-input space, but the Hopf–Lax formulae [7] for HJ PDEs relevant to those OCPs are always convex in the control-input space. This implies that the direct methods are guaranteed to find the solution for the Hopf–Lax formula, while not necessarily for the OCP. This article aims to extend the convexification of the Hopf–Lax formula to general state-constrained OCPs, which is a novel development given that Hopf–Lax theories have not yet dealt with HJ PDEs relevant to state-constrained problems.

In prior related works [8], [9], convexification methods have been developed under different assumptions. For example, the control-set convexification method [8] convexifies nonconvex control sets and assumes convex cost in the state and the control, linear dynamics, and convex state constraints. On the other hand, the successive convexification method [9] provides a locally optimal solution for nonlinear dynamics, convex cost, and convex state constraints. The work in [8] is listed in Table I as a method for global optimality.

Fig. 1 illustrates our framework to convexify the OCP in the control-input space. Given the OCP, we propose a convexified OCP (COCP) without approximation; thus, the optimal costs of the OCP and the COCP are the same and represented as $\vartheta$. The COCP applies the control signal transformation technique in [7], which converts the control signal variable $\alpha$ to a new variable $\beta$. Notably, the COCP is always convex in the new control signal space. Thus, if the OCP's functions and constraints are convex only in the state but not necessarily in the control input, direct methods are guaranteed to find the COCP's optimal cost ($\vartheta$), state trajectory ($x_*$), and control signal ($\beta_*$). In particular, we will prove that the COCP's optimal state trajectory is also optimal for the OCP. Additionally, we propose algorithms to find an optimal control signal $\alpha_*$ for the OCP from the COCP solution ($x_*, \beta_*$).

The contributions of this article are as follows.
1) We convexify general state-constrained OCPs in control. Our methods provide more general optimality conditions than any other methods that are scalable for high-dimensional systems, as in Table I.
2) We also propose a numerical algorithm that is guaranteed to find the OCP's optimal solution. This is not guaranteed by numerical solvers that implement PMP.

The rest of this article is organized as follows. Section II presents the problem description. Section III proposes the COCP and numerical algorithms to compute an optimal solution to the OCP. Next, Section IV presents convexity conditions for the COCP. Section V provides a numerical example, and finally, Section VI concludes this article.

TABLE I
A COMPARISON OF DIFFERENT OCP METHODS. $x$ AND $a$ REPRESENT THE STATE AND CONTROL, RESPECTIVELY

| methods | assumptions for optimality | | | | scalablility to deal with high-dimensional OCPs | guarantees to find solutions |
| --- | --- | --- | --- | --- | --- | --- |
| | cost | dynamics | state constraints | control contraints | | |
| multiple shooting [1] | convex in $(x, a)$ | affine in $(x, a)$ | convex in $x$ | convex in $a$ | scalable | yes |
| collocation [2] | convex in $(x, a)$ | affine in $(x, a)$ | convex in $x$ | convex in $a$ | scalable | yes |
| Pontryagin's principle [3] | convex in $x$ | affine in $x$ | convex in $x$ | convex in $a$ | scalable | no |
| Hamilton-Jacobi [4] | general | general | general | general | not scalable | yes |
| Koopman-based methods [5] | This method works for state-unconstrained problems. | | | | scalable | no |
| Branch-and-Lift [6] | general | general | general | convex in $a$ | not scalable | yes |
| Control-set convexification [8] | convex in $(x, a)$ | linear in $(x, a)$ | convex in $x$ | general | scalable | yes |
| COCP (ours) | convex in $x$ | affine in $x$ | convex in $x$ | general | scalable | yes |
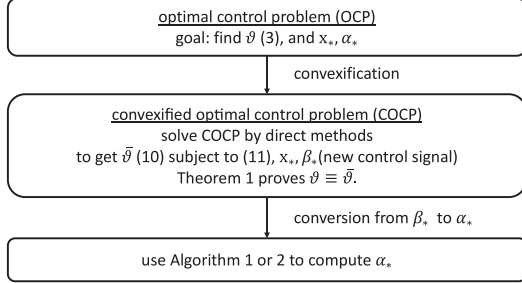


Fig. 1. Goal is to find the OCP's optimal value $\vartheta$, optimal state trajectory $x_*$, and optimal control signal $\alpha_*$. Our COCP formulation converts the control signal space $\alpha$ to $\beta$ and then convexifies the OCP in the control-input space without approximation. If the OCP's functions and constraints are convex only in the state space, the COCP's functions and constraints become convex in both state and control input spaces. Thus, direct methods provide the COCP's optimal solution $\vartheta$, $x_*$, and optimal control signal in the converted space $\beta_*$. Finally, our algorithm generates the optimal control signal $\alpha_*$ from $\beta_*$.

In terms of notation in this article, the subscript $_*$ denotes optimality, and the superscript $^*$ denotes the Legendre–Fenchel transformation. For example, $\alpha_*$ is an optimal control signal, and $H^*$ is the Legendre–Fenchel transformation of a function $H$. The double superscript $^{**}$ denotes the biconjugate, double Legendre–Fenchel transformation: $H^{**} = (H^*)^*$.

## II. STATE-CONSTRAINED OCP

Consider a state trajectory $x : [0, T] \to \mathbb{R}^n$ solving the ODE

$$\dot{x}(t) = f(t, x(t), \alpha(t)), \ t \in [0, T], \quad x(0) = x_0 \in \mathbb{R}^n, \quad (1)$$

where $f : [0, T] \times \mathbb{R}^n \times A \to \mathbb{R}^n$ is the system dynamics, $x_0$ is the initial state, $\alpha \in \mathcal{A}$ is the measurable control signal where $\mathcal{A}$ is the set of admissible control signals:

$$\mathcal{A} := \{\alpha : [0, T] \to A \mid \|\alpha\|_{L^\infty(0,T)} < \infty\} \quad (2)$$

and $A$ is a compact subset in $\mathbb{R}^m$. We would like to solve a finite-horizon, state-constrained OCP for the above dynamical system with respect to a cost $\vartheta$, which we describe next.

*State-constrained OCP:* Given the dynamical system (1)–(2), solve

$$\vartheta := \lim_{\epsilon \to 0+} \vartheta^\epsilon \quad (3)$$

where

$$\vartheta^\epsilon := \inf_{\alpha \in \mathcal{A}} \int_0^T L(t, x(t), \alpha(t))dt + g(x(T)) \quad (4)$$

$$\text{subject to } c(t, x(t)) \leq \epsilon, \quad t \in [0, T] \quad (5)$$

where x solves (1). Here, $L : [0, T] \times \mathbb{R}^n \times A \to \mathbb{R}$ is the stage cost, $g : \mathbb{R}^n \to \mathbb{R}$ is the terminal cost, and $c : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ is the state

constraint. For example, $c(t, \cdot)$ can represent unsafe regions in the state space over $t \in [0, T]$, so that $c(t, x) < 0$ for $x \in \mathbb{R}^n$ away from the unsafe regions, $c(t, x) = 0$ for $x$ on the boundary of the unsafe regions, and $c(t, x) > 0$ for $x$ in the unsafe regions.

In this article, we adopt the extended value setting [10, Ch. 3.1.2] for all optimization problems, which adds $\infty$ to the objective function if constraints are not satisfied. For example, $\vartheta^\epsilon$ can be written as

$$\vartheta^\epsilon = \inf_{\alpha \in \mathcal{A}} \int_0^T L(t, x(t), \alpha(t))dt + g(x(T)) + \max_{t \in [0,T]} \mathcal{I}(c(t, x(t)) \leq \epsilon) \quad (6)$$

where

$$\mathcal{I}(\text{``condition''}) = \begin{cases} 0, & \text{``condition'' is satisfied} \\ \infty, & \text{otherwise.} \end{cases} \quad (7)$$

In this extended value setting, the limit defined by $\vartheta$ in (3) exists. This is because $\vartheta^{\epsilon_1} \leq \vartheta^{\epsilon_2}$ for any $\epsilon_1 > \epsilon_2 > 0$. Thus, as $\epsilon$ converges to zero, the limit of $\vartheta^\epsilon$ exists in $\mathbb{R} \cup \{\infty\}$.

Also note that $\vartheta$ and $\vartheta^0$ are different because the set of state trajectories solving (1) is not compact. Consider a case in which $\max_{t \in [0,T]} c(t, x(t)) > 0$ for all $\alpha \in \mathcal{A}$, but there exists a sequence of $\alpha_k \in \mathcal{A}$ such that $\lim_{k \to \infty} \max_{t \in [0,T]} c(t, x_k(t)) = 0$, where $x_k$ solves (1) for $\alpha_k$. In this case, $\vartheta^0$ is $\infty$, but $\vartheta$ is finite. On the other hand, if $A$ is convex, the set of state trajectories solving (1) becomes compact [4]. In that case, $\vartheta^0 = \vartheta$.

In this article, we assume the following.
*Assumption 1 (Continuity and compactness):*
A1. the control set $A$ is compact;
A2. $f = f(t, x, a) : [0, T] \times \mathbb{R}^n \times A \to \mathbb{R}^n$ is Lipschitz continuous in $(t, x)$ for each $a \in A$, and uniformly continuous in $a$ for each $(t, x)$;
A3. the stage cost $L = L(t, x, a) : [0, T] \times \mathbb{R}^n \times A \to \mathbb{R}$ is continuous in $(t, x, a)$ and, for each $a$, is locally Lipschitz in $(t, x)$;
A4. the terminal cost $g = g(x) : \mathbb{R}^n \to \mathbb{R}$ is continuous in $x$ and locally Lipscthiz in $x$;
A5. the state constraint $c = c(t, x) : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ is continuous in $(t, x)$ and, for each $t$, is locally Lipschitz in $x$.

Assumptions A1 and A2 guarantee a unique state trajectory for the dynamics (1). Since we consider a bounded time interval, all state trajectories are also bounded. Thus, Assumptions A3 and A4 are enough to ensure that the integration of the cost in $\vartheta$ is bounded.

## III. COCPs

Section III-A presents the theorem for the COCP, and Section III-B builds up mathematical background to prove the theorem, which is done in Section III-C.

## A. COCP

Consider a multivalued map: for $(t, x) \in [0, T] \times \mathbb{R}^n$,

$$B(t, x) := \{b \,|\, b = -f(t, x, a), a \in A\} \subset \mathbb{R}^n \qquad (8)$$

where $f$ and $A$ are the dynamics and the control constraint of the system, respectively. $B(t, x)$ is compact for all $(t, x)$ since $A$ is compact and $f$ is uniformly continuous in $a$ as in Assumption 1. Consider a state trajectory solving

$$\dot{\mathrm{x}}(t) = -\beta(t), \; t \in [0, T], \quad \mathrm{x}(0) = x_0$$
$$\beta(t) \in \mathrm{co}(B(t, \mathrm{x})), \; t \in [0, T] \qquad (9)$$

where "co" represents a convex-hull operation. We assume that $\beta$ is a measurable control signal. With respect to this differential inclusion, we propose our convexifying method in the following theorem, which will be later proved in Section III-C.

*Theorem 1 (COCP):* Suppose Assumption 1 holds. Define

$$\bar{\vartheta} := \inf_\beta \int_0^T H^*(t, \mathrm{x}(t), \beta(t)) dt + g(\mathrm{x}(T)) \qquad (10)$$

$$\text{subject to } c(t, \mathrm{x}(t)) \leq 0, \quad t \in [0, T] \qquad (11)$$

where x and $\beta$ satisfy (9), $H : [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is defined as

$$H(t, x, p) := \max_{a \in A}[-p \cdot f(t, x, a) - L(t, x, a)] \qquad (12)$$

and $H^*$ is the Legendre–Fenchel transformation (convex conjugate) of $H$ with respect to the costate $p$:

$$H^*(t, x, b) := \sup_{p \in \mathbb{R}^n}[p \cdot b - H(t, x, p)]. \qquad (13)$$

Then, $\vartheta$ in (3) and $\bar{\vartheta}$ are the same

$$\vartheta = \bar{\vartheta}. \qquad (14)$$

Both $H$ in (12) and $H^*$ in (13) adopt the extended-value settings. However, for each $(t, x)$, $H$ is finite for all $p$ since $A$ is compact as in Assumption A1; on the other hand, $H^*$ is $\infty$ for some $b \in \mathbb{R}^n$. The Legendre–Fenchel transformation always provides a closed proper convex function; thus, the new stage cost $H^*$ is convex in the new control $b$.

## B. Mathematical Formulation and Optimal Control Analysis

In order to establish the relationship between the OCP and the COCP proposed in Theorem 1, we will first build the relationship between two control inputs $a$ in the OCP and $b$ in the COCP.

For $b \in \mathbb{R}^n$, define

$$L^b(t, x, b) := \min_{a \in A} L(t, x, a) \quad \text{subject to} \quad f(t, x, a) = -b. \qquad (15)$$

*Lemma 1:* Suppose Assumption 1 holds. Then, $L^b$ and $H$ defined in (15) and (12), respectively, have the following properties:

$$(L^b)^*(t, x, p) = H(t, x, p) \quad \text{in } [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \qquad (16)$$

$$(L^b)^{**}(t, x, b) = H^*(t, x, b) \quad \text{in } [0, T] \times \mathbb{R}^n \times \mathbb{R}^n \qquad (17)$$

$$\mathrm{Dom}(H^*(t, x, \cdot)) = \{b \,|\, H^*(t, x, b) < \infty\} = \mathrm{co}(B(t, x)). \qquad (18)$$

Here, $(L^b)^*$ and $H^*$ are the Legendre–Fenchel transformations (convex conjugate) of $L^b$ and $H$, respectively, with respect to $p$ for each $(t, x)$. The double Legendre–Fenchel transformation of $L^b$, i.e. $(L^b)^{**} = ((L^b)^*)^*$, is the biconjugate of $L^b$. $\mathrm{Dom}(H^*(t, x, \cdot))$ represents the domain of $H^*(t, x, \cdot)$, and $\mathrm{co}(B(t, x))$ represents the convex-hull of $B(t, x)$.
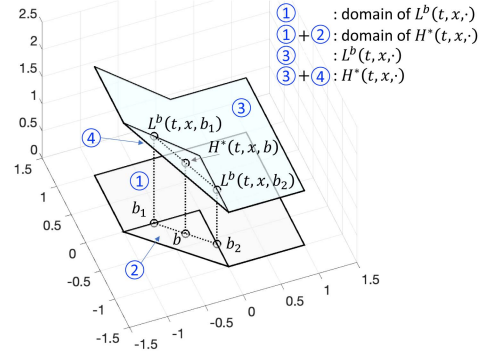


Fig. 2. Bottom 2-D space represents $b \in \mathbb{R}^2$, and the vertical axis represents function values. For fixed $(t, x)$, ①$(\subset \mathbb{R}^2)$ is $B(t, x)$, the domain of $L^b(t, x, \cdot)$. The union of ① and ②$(\subset \mathbb{R}^2)$ is the domain of $H^*(t, x, \cdot)$, which is a convex-hull of ①. The curvature ③ represents $L^b(t, x, \cdot)$, and the curvatures ③ and ④ represent $H^*(t, x, \cdot)$. In the illustrated example, for $b \in \mathrm{co}(B(t, x))$ and $b \notin B(t, x)$, there exists $b_1, b_2 \in B(t, x)$ such that the vector $[H^*(t, x, b), b]$ is a convex combination of vectors $[L^b(t, x, b_1), b_1]$ and $[L^b(t, x, b_2), b_2]$.

*Proof:* (i) Proof of (16) and (17)

Under the measurable assumptions on $\alpha$ and $\beta$, a state-unconstrained problem

$$V_1(t, x) := \inf_{\alpha \in \mathcal{A}} \int_t^T L(s, \mathrm{x}(s), \alpha(t)) ds + g(\mathrm{x}(T)) \qquad (19)$$

where x solves $\dot{\mathrm{x}}(s) = f(s, \mathrm{x}(s), \alpha(s))$, $\mathrm{x}(t) = x$, is equivalent to

$$V_2(t, x) := \inf_\beta \int_t^T L^b(s, \mathrm{x}(s), \beta(s)) ds + g(\mathrm{x}(T)) \qquad (20)$$

where x solves $\dot{\mathrm{x}}(s) = -\beta(s)$, $\beta(s) \in B(s, \mathrm{x}(s))$, where $B$ is defined in (8), and $\mathrm{x}(t) = x$. For a rigorous proof of this equivalence in the sense of differential inclusion, the reader is referred to [11, Ch. 4.1, Fillipov's Theorem]. Then, the corresponding HJ PDEs [12] for $V_1$ and $V_2$ are the same; thus, the Hamiltonians in the HJ PDEs are the same: $\max_{b \in B(t,x)}[p \cdot b - L^b(t, x, b)] = H(t, x, p)$. Hence, $H \equiv (L^b)^*$ and $H^* \equiv (L^b)^{**}$.

(ii) Proof of (18)

*Case 1:* $b \in \mathrm{co}(B(t, x))$

There exist $b_i \in B(t, x)$ and $\gamma_i \geq 0$ ($\sum_{i=1}^I \gamma_i = 1$) such that $b = \sum_{i=1}^I \gamma_i b_i$. Since $H^*(t, x, \cdot)$ is convex in $b$,

$$H^*(t, x, b) \leq \sum_{i=1}^I \gamma_i H^*(t, x, b_i) < \infty$$

since each of $H^*(t, x, b_i)$ is finite.

*Case 2:* $b \notin \mathrm{co}(B(t, x))$

For two closed convex sets $\{b\}$ and $\mathrm{co}(B(t, x))$, the separating hyperplane theorem [10, Ch. 2.5] implies that there exists a hyperplane $(P : \mathbb{R}^n \to \mathbb{R})$: $P(b') := p' \cdot b' + c$ such that $P(b')$ is positive if $b' \notin \mathrm{co}(B(t, x))$; otherwise, $P(b')$ is nonpositive. By picking $p = \gamma p'$ where $\gamma \in \mathbb{R}$,

$$H^*(t, x, b) = \sup_p \inf_{b'}[p \cdot (b - b') + L^b(t, x, b')]$$

$$\geq \sup_\gamma \inf_{b'}[\gamma p' \cdot (b - b') + L^b(t, x, b')] = \infty$$

since $p' \cdot (b - b') > 0$ for all $b' \in \mathrm{co}(B(t, x))$. ∎

Fig. 2 illustrates that the epigraph of $H^*(t, x, \cdot)$ is the convex-hull of the epigraph of $L^b(t, x, \cdot)$ for fixed $(t, x)$ [13, Ch. 4]. Thus, for any $b \in \mathrm{co}(B(t, x))$, there are a finite number ($I$) of weights $\gamma_i$ and

control inputs $b_i \in B(t, x)$ such that the pair of $(H^*(t, x, b), b)$ is a convex combination of $(L^b(t, x, b_i), b_i)$ with weights $\gamma_i$ ($\sum_{i=1}^I \gamma_i = 1$, $\gamma_i \geq 0$):

$$\begin{bmatrix} H^*(t, x, b) \\ b \end{bmatrix} = \sum_{i=1}^I \gamma_i \begin{bmatrix} L^b(t, x, b_i) \\ b_i \end{bmatrix} = \sum_{i=1}^I \gamma_i \begin{bmatrix} L(t, x, a_i) \\ -f(t, x, a_i) \end{bmatrix} \quad (21)$$

where $a_i$ satisfies $L^b(t, x, b_i) = L(t, x, a_i)$. Notably, $I$ depends on $(t, x, b)$. In general, $I$ is less than or equal to $n + 1$, but $I$ is 1 if $b \in B(t, x)$.

Equation (18) in Lemma 1 implies that the domain of $H^*(t, x, \cdot)$ contains the domain of $L^b(t, x, \cdot)$ for each $(t, x)$. Even though $\mathrm{co}(B(t, x))$ is convex in $b \in \mathbb{R}^n$, $\mathrm{Dom}(H^*(t, \cdot, \cdot)) = \{(x, b) \mid b \in \mathrm{co}(B(t, x))\}$ is generally nonconvex in $(x, b) \in \mathbb{R}^n \times \mathbb{R}^n$ for each $t \in [0, T]$.

Using the above analysis for two pointwise control inputs $a, b$, we now find the relationship between the OCP's and the COCP's control signals ($\alpha$ and $\beta$, respectively). Consider a measurable control signal ($\beta$) and state (x) trajectories solving the dynamics (9). Assume that $\beta$ is Riemann integrable in $(0, T)$. For some $\delta > 0$, consider a temporal discretization: $\{t_0 = 0, \ldots t_K = T\}$ such that $\Delta t_k := t_{k+1} - t_k < \delta \forall k = 0, \ldots, K - 1$. Note that the step size does not need to be fixed in this temporal discretization. We define a control signal $\alpha^\epsilon \in \mathcal{A}$: for $k = \{0, \ldots, K - 1\}$,

$$\alpha^\epsilon(t) = a_i^k, \quad t \in \left[ t_k + \sum_{j=1}^{i-1} \gamma_j^k \Delta t_k, t_k + \sum_{j=1}^i \gamma_j^k \Delta t_k \right) \quad (22)$$

where $a_i^k$ and $\gamma_i^k$ are, respectively, the $i$th control and coefficient in (21) for $t = t_k$, $x = \mathrm{x}(t_k)$, and $b = \beta(t_k)$, where $i = 1, \ldots, I_k$. Then, we will prove that $\mathrm{x}^\epsilon$ solving (1) for $\alpha^\epsilon$ is close to x, and the integration of the stage and terminal costs of the OCP induced by $\alpha^\epsilon$ is also close to the counterpart of the COCP induced by $\beta$.

*Theorem 2:* Suppose Assumption 1 holds. Consider any measurable control signal $\beta$ and state trajectory x solving (9). Assume that $\beta$ is Riemann integrable in $[0, T]$. Then, for any $\epsilon > 0$, there exists $\delta > 0$ such that, for any discretization $\{t_0 = 0, \ldots, t_K = T\}$ where $|\Delta t_k| < \delta$, $k = 0, \ldots, K - 1$, we have

$$\|\mathrm{x} - \mathrm{x}^\epsilon\|_{L^\infty(0, T)} < \epsilon \quad (23)$$

and

$$\left| \int_0^T H^*(t, \mathrm{x}(t), \beta(t)) dt + g(\mathrm{x}(T)) \right.$$
$$\left. - \int_0^T L(t, \mathrm{x}^\epsilon(t), \alpha^\epsilon(t)) dt - g(\mathrm{x}^\epsilon(T)) \right| < \epsilon \quad (24)$$

where $\mathrm{x}^\epsilon$ solves (1) for $\alpha^\epsilon$ in (22).

*Proof:* We will first prove that x and $\mathrm{x}^\epsilon$ are close within an $\epsilon$-bound in the $L^\infty$-norm. We will define two more state trajectories $\mathrm{x}_0$ and $\mathrm{x}_1$, and then show x is close to $\mathrm{x}_0$, $\mathrm{x}_0$ is close to $\mathrm{x}_1$, and $\mathrm{x}_1$ is close to $\mathrm{x}^\epsilon$. Here, $\mathrm{x}_0$ solves (9) for a piecewise constant control generated from $\beta$ on $\{t_0, \ldots, t_K\}$:

$$\dot{\mathrm{x}}_0(t) = -\beta(t_k), t \in [t_k, t_{k+1}), \quad \mathrm{x}_0(0) = x_0. \quad (25)$$

By the control decomposition in (21), we have $I_k$ number of $(b_i^k, a_i^k, \gamma_i^l)$ pairs for $t = t_k, x = \mathrm{x}(t_k)$, and $b = \beta(t_k)$ such that $b_i^k \in B(t_k, \mathrm{x}(t_k))$, $a_i^k \in A$, $\gamma_i^k \geq 0$, $\sum_{i=1}^{I_k} \gamma_i^k = 1$ and

$$H^*(t_k, \mathrm{x}(t_k), \beta(t_k)) = \sum_{i=1}^{I_k} \gamma_i^k L^b(t_k, \mathrm{x}(t_k), b_i^k)$$
$$= \sum_{i=1}^{I_k} \gamma_i^k L(t_k, \mathrm{x}(t_k), a_i^k) \quad (26)$$

$$\beta(t_k) = \sum_{i=1}^{I_k} \gamma_i^k b_i^k, \quad b_i^k = -f(t_k, \mathrm{x}(t_k), a_i^k). \quad (27)$$

Consider partitioning each interval $[t_k, t_{k+1})$ into $I_k$ intervals $[t_{k,i}, t_{k,i+1}) := [t_k + \sum_{j=1}^{i-1} \gamma_j^k \Delta t_k, t_k + \sum_{j=1}^i \gamma_j^k \Delta t_k]$ as in (22), where $t_{k,0} = t_k$ and $t_{k,\bar{\imath}_k} = t_{k+1}$, and design a new piecewise constant control signal $\beta_1$

$$\beta_1(t) = b_{i+1}^k, \ t \in [t_{k,i}, t_{k,i+1}). \quad (28)$$

We define $\mathrm{x}_1$ as the solution of (9) corresponding to this input $\beta_1$.

(i) Show $\|\mathrm{x} - \mathrm{x}_0\|_{L^\infty(0, T)} < \epsilon$

Since $\beta$ is Riemann integrable, for all $\epsilon$, there exists $\delta_1 > 0$ such that $\|\mathrm{x}_0 - \mathrm{x}\|_{L^\infty(t, T)} < \epsilon$ if $\delta \leq \delta_1$.

(ii) Show $\|\mathrm{x}_0 - \mathrm{x}_1\|_{L^\infty(0, T)} < \epsilon$

From (27), $\mathrm{x}_0(t_k) = \mathrm{x}_1(t_k)$ for all $k = 0, \ldots, K$. For $t \in [t_k, t_{k+1})$, we have

$$\|\mathrm{x}_1(t) - \mathrm{x}_0(t)\| \leq \int_{t_k}^t \| - \beta_1(s) + \beta(t_k)\| ds. \quad (29)$$

Note that the set of states that can be reached

$$\mathcal{X} := \{\mathrm{x}(t) \mid \mathrm{x} \text{ solves } (1) \text{ for some } \alpha \in \mathcal{A}, t \in [0, T]\} \quad (30)$$

is bounded since Assumption A2 holds. Thus,

$$c_1 := \max_{t \in [0, T], x \in \mathrm{cl}(\mathcal{X}), a \in A} \|f(t, x, a)\| \quad (31)$$

is finite, where $\mathrm{cl}(\mathcal{X})$ is the closure of $\mathcal{X}$. Since $\| - \beta_1(s) + \beta(t_k)\|$ is less than $2c_1$, we conclude the right-hand side of (29) is less than or equal to $2c_1\delta$.

(iii) Show $\|\mathrm{x}_1 - \mathrm{x}^\epsilon\|_{L^\infty(0, T)} < \epsilon$

We will first prove $\mathrm{x}^\epsilon(t_{k,i+1})$ is close to $\mathrm{x}_1(t_{k,i+1})$ for all $k = 0, \ldots, K - 1, i = 1, \ldots, I_k$. We have,

$$\|\mathrm{x}^\epsilon(t_{k,i+1}) - \mathrm{x}_1(t_{k,i+1})\| \leq \|\mathrm{x}^\epsilon(t_{k,i}) - \mathrm{x}_1(t_{k,i})\|$$
$$+ \int_{t_{k,i}}^{t_{k,i+1}} \|f(s, \mathrm{x}^\epsilon(s), a_{i+1}^k) - f(t_k, \mathrm{x}(t_k), a_{i+1}^k)\| ds. \quad (32)$$

By Assumption A2, $\|f(s, \mathrm{x}^\epsilon(s), a_{i+1}^k) - f(t_k, \mathrm{x}(t_k), a_{i+1}^k)\|$ is less than or equal to $L_f(s - t_k + \|\mathrm{x}^\epsilon(s) - \mathrm{x}(t_k)\|)$. Moreover,

$$\|\mathrm{x}^\epsilon(s) - \mathrm{x}(t_k)\| \leq \|\mathrm{x}^\epsilon(t_{k,i}) - \mathrm{x}_1(t_{k,i})\| + 4c_1\delta + \epsilon \quad (33)$$

since $\|\mathrm{x}^\epsilon(s) - \mathrm{x}^\epsilon(t_{k,i})\| < c_1\delta, \|\mathrm{x}_1(t_{k,i}) - \mathrm{x}_1(t_k)\| < c_1\delta$, $\|\mathrm{x}_1(t_k) - \mathrm{x}_0(t_k)\| < 2c_1\delta$, and $\|\mathrm{x}_0(t_k) - \mathrm{x}(t_k)\| < \epsilon$, where $c_1$ is defined in (31). Hence, we get

$$\|\mathrm{x}^\epsilon(t_{k,i+1}) - \mathrm{x}_1(t_{k,i+1})\| + c_2\delta + \epsilon \leq$$
$$(1 + L_f(t_{k,i+1} - t_{k,i}))(\|\mathrm{x}^\epsilon(t_{k,i}) - \mathrm{x}_1(t_{k,i})\| + c_2\delta + \epsilon) \quad (34)$$

where $c_2 = 1 + 4c_1$. Furthermore,

$$\prod_{k,'i'}^{t_{k,'i'} \leq t_{k,i}} (1 + L_f(t_{k,'i'} - t_{k,'i'-1}))$$
$$\leq \prod_{k,'i'}^{t_{k,'i'} \leq T} (1 + L_f(t_{k,'i'} - t_{k,'i'-1})) \leq \left( 1 + \frac{L_f T}{\bar{m}} \right)^{\bar{m}} < e^{L_f T}. \quad (35)$$

The second inequality follows from Jensen's inequality for log. Thus, (34) and (35) imply

$$\|\mathrm{x}^\epsilon(t_{k,i+1}) - \mathrm{x}_1(t_{k,i+1})\| \leq (e^{L_f T} - 1)(c_2\delta + \epsilon). \quad (36)$$

For $t \in (t_{k,i}, t_{k,i+1})$, we can similarly show $\|\mathrm{x}^\epsilon(t) - \mathrm{x}_1(t)\| \leq (e^{L_f T} - 1)(c_3\delta + \epsilon)$ for some $c_3 > 0$.

To sum up (i)–(iii), we conclude that, for any small $\epsilon > 0$, we can find $\delta$ such that $\|\mathrm{x} - \mathrm{x}^\epsilon\|_{L^\infty(0, T)} < \epsilon$.

(iv) Prove (24). Define a discrete sum of $H^*$:

$$S := \sum_{k=0}^{K-1} H^*(t_k, \mathrm{x}(t_k), \beta(t_k))(t_{k+1} - t_k) + g(\mathrm{x}(T)).$$

Since $H^*$ is proper and $\beta$ is Riemann integrable, $H^*(t, \mathrm{x}(t), \beta(t))$ is also Riemann integrable. Thus, for any $\epsilon_1$, there exists $\delta$ such that

$$\left| \int_0^T H^*(t, \mathrm{x}(t), \beta(t))dt + g(\mathrm{x}(T)) - S \right| < \epsilon_1. \qquad (37)$$

We can show $S$ is close to $\int_0^T L(t, \mathrm{x}^\epsilon(t), \alpha^\epsilon(t))dt + g(\mathrm{x}^\epsilon(T))$ by utilizing similar techniques used to prove $\|\mathrm{x} - \mathrm{x}^\epsilon\|_{L^\infty(0,T)} < \epsilon$. In the proof, it is also necessary to use the fact that $L$ is locally Lipschitz in $(t, x) \in [0, T] \times \mathrm{cl}(\mathcal{X})$ and $g$ is locally Lipschitz in $x \in \mathrm{cl}(\mathcal{X})$, where $\mathcal{X}$ is defined in (30). ∎

Consider the case where the COCP is feasible and has optimal control signal $\beta_*$ and state trajectory $\mathrm{x}_*$. Then, Theorem 2 provides the corresponding $(\mathrm{x}_*^\epsilon, \alpha_*^\epsilon)$ whose OCP cost without state constraints is close to the COCP's optimal cost, and $\mathrm{x}_*^\epsilon$ is close to the COCP's optimal state trajectory within $\epsilon$-distance in $L^\infty$-norm. As the time-step variable $\delta$ converges to 0, we can force $\epsilon$ to converge to 0. This implies that $\mathrm{x}_*^\epsilon$ converges to the COCP's optimal state trajectory $\mathrm{x}_*$. Thus, the OCP's state constraint would be satisfied at the limit of $\epsilon$ at 0: $\lim_{\epsilon \to 0} \max_{t \in [0,T]} c(t, \mathrm{x}_*^\epsilon(t)) \le 0$ although the limit point of $\alpha_*^\epsilon$ might not exist. However, we need more careful analysis as to whether the OCP's optimal value is finite. On the other hand, where the COCP is infeasible, Theorem 2 does not help us analyze if the OCP's optimal value is finite or not. In the following section, we rule out these mathematical issues and complete the proof of Theorem 1, which states that the OCP and the COCP are equivalent.

### C. Proof for the COCP

This section proves Theorem 1 using mathematical background from Section III-B.

*Proof for Theorem 1:* We will prove that the two epigraphs of $\vartheta$ and $\bar{\vartheta}$ are the same. We first define a function $J = J(z, \alpha)$

$$J(z, \alpha) := \max \left\{ \max_{t \in [0,T]} c(t, \mathrm{x}(t)), \right.$$
$$\left. \int_0^T L(t, \mathrm{x}(t), \alpha(t))dt + g(\mathrm{x}(T)) - z \right\} \qquad (38)$$

where $\mathrm{x}$ solves (1) for $\alpha$, and $z$ represents the value-axis variable. The authors in [4] proved that the zero sublevel set of $\inf_\alpha J(z, \alpha)$ is the epigraph of $\vartheta^0$ if $A$ is convex. In contrast, we do not assume that $A$ is convex. However, the proof in [4, Th. 3.1] can be easily modified to show the zero sublevel set of $J(z, \alpha)$ is the epigraph of $\vartheta$:

$$\vartheta = \min z \text{ subject to } \inf_{\alpha \in \mathcal{A}} J(z, \alpha) \le 0. \qquad (39)$$

Similarly, we define a function $\bar{J} = \bar{J}(z, \beta)$

$$\bar{J}(z, \beta) := \max \left\{ \max_{t \in [0,T]} c(t, \mathrm{x}(t)), \right.$$
$$\left. \int_0^T H^*(t, \mathrm{x}(t), \beta(t))dt + g(\mathrm{x}(T)) - z \right\} \qquad (40)$$

where $\mathrm{x}$ solves (9) for $\beta$, so that the zero sublevel set of $\inf_\beta \bar{J}(z, \beta)$ represents the epigraph of $\bar{\vartheta}$:

$$\bar{\vartheta} = \min z \text{ subject to } \inf_\beta \bar{J}(z, \beta) \le 0 \qquad (41)$$

where $\beta(t) \in \mathrm{co}(B(t, \mathrm{x}(t)))$ for all $t \in [0, T]$.

---

**Algorithm 1:** Computing the OCP's Optimal State Trajectory and Control Signal.

1: **Output:** OCP's optimal state trajectory $\mathrm{x}_*^\epsilon$ and control signal $\alpha_*^\epsilon$
2: Solve $\bar{\vartheta}$ (COCP) in (10) subject to (11), and get $\mathrm{x}_*, \beta_*$ by multiple shooting or collocation methods
3: On a temporal discretization: $\{t_0 = 0, \ldots, t_K = T\}$, for each $k = 0, \ldots, K - 1$, find $I_k$-number of $(a_i^k, b_i^k, \gamma_i^k)$ solving (21) for $t = t_k$, $x = \mathrm{x}_*(t_k)$ and $b = \beta_*(t_k)$
4: Additionally discretize each temporal interval:

$$[t_k, t_{k+1}) = \bigcup_{i=1}^{I_k} \left[ t_k + \sum_{j=1}^{i-1} \gamma_j^k \Delta t_k, t_k + \sum_{j=1}^{i} \gamma_j^k \Delta t_k \right) \quad (45)$$

5: Design $\alpha_*^\epsilon$ using $a_i^k$ by (22) and compute $\mathrm{x}_*^\epsilon$ solving (1)

---

(i) $\inf_\alpha J(z, \alpha) \ge \inf_\beta \bar{J}(z, \beta)$

For any state ($\mathrm{x}$) and control ($\alpha$) trajectories solving (1), define a control signal ($\beta$):

$$\beta(t) = -f(t, \mathrm{x}(t), \alpha(t)) \in B(t, \mathrm{x}(t)), \quad t \in [0, T].$$

Then, $\mathrm{x}$ and $\beta$ solve $\dot{\mathrm{x}}(t) = -\beta(t)$ for $t \in [0, T]$ and $\mathrm{x}(0) = x_0$. By Lemma 1, $L(t, \mathrm{x}(t), \alpha(t)) = L^b(t, \mathrm{x}(t), \beta(t)) \ge H^*(t, \mathrm{x}(t), \beta(t))$ for all $t \in [0, T]$. This implies that

$$J(z, \alpha) \ge \bar{J}(z, \beta) \ge \inf_\beta \bar{J}(z, \beta).$$

Since the above inequality holds for all $\mathrm{x}$ and $\alpha$, we conclude $\inf_\alpha J(z, \alpha) \ge \inf_\beta \bar{J}(z, \beta)$.

(ii) $\inf_\alpha J(z, \alpha) \le \inf_\beta \bar{J}(z, \beta)$

By Assumption A2, the set of all feasible states $\mathcal{X} := \{\mathrm{x}(t) \mid \mathrm{x} \text{ solves (1) for some } \alpha \in \mathcal{A}, t \in [0, T]\}$ is bounded. Since $\mathrm{cl}(\mathcal{X})$ is compact, Assumption A5 implies that $c(t, \cdot)$ is locally Lipschitz in $x \in \mathrm{cl}(\mathcal{X})$ for some constant $L_c > 0$. For any state ($\mathrm{x}$) and control ($\beta$) trajectories solving (9), by Theorem 2, there exists $\mathrm{x}^\epsilon$ and $\alpha^\epsilon$ that solve (1) and satisfy (23) and (24) for any $\epsilon > 0$. Then, $\bar{J}(z, \beta) \ge J(z, \alpha^\epsilon) - \max\{1, L_c\}\epsilon \ge \inf_\alpha J(z, \alpha) - \max\{1, L_c\}\epsilon$. Since this inequality holds for any $\mathrm{x}, \beta$ solving (9) and $\epsilon > 0$, $\inf_\beta \bar{J}(z, \beta) \ge \inf_\alpha J(z, \alpha)$.

(iii) Since $\inf_\alpha J(z, \alpha)$ and $\inf_\beta \bar{J}(z, \beta)$ are the same, the epigraphs of $\vartheta$ and $\bar{\vartheta}$ are the same, and (39) and (41) conclude $\vartheta \equiv \bar{\vartheta}$. ∎

*Remark 1:*

1) The optimal costs for the OCP and the COCP are the same. Also, the COCP is feasible if and only if the OCP for $\vartheta^\epsilon$ is feasible for all $\epsilon > 0$.

2) For a feasible COCP whose optimal control signal is $\beta_*$ and state trajectory is $\mathrm{x}_*$, we can construct $\alpha_*^\epsilon$ as in (22) with a temporal discretization $\{t_0 = 0, \ldots, t_K = T\}$. Then, Theorems 1 and 2 imply uniform convergence of the following three terms as $\epsilon$ converges to 0:

$$\mathrm{x}_*^\epsilon \to \mathrm{x}_* \qquad (42)$$

$$\int_0^T L(t, \mathrm{x}_*^\epsilon(t), \alpha_*^\epsilon(t))dt + g(\mathrm{x}_*^\epsilon(T)) \to \vartheta \qquad (43)$$

$$c(t, \mathrm{x}_*^\epsilon(t)) \to c(t, \mathrm{x}_*(t)) \quad \forall t \in [0, T], \qquad (44)$$

where $\mathrm{x}_*^\epsilon$ solves (1) for $\alpha_*^\epsilon$, and $c(t, \mathrm{x}_*(t)) \le 0$ for all $t \in [0, T]$.

### D. Numerical Algorithm

Algorithm 1 presents a numerical algorithm to compute an optimal state trajectory ($\mathrm{x}$) and a control signal ($\alpha$) for the OCP using the COCP in Theorem 1.

---

**Algorithm 2:** Computing the OCP's Optimal State Trajectory and Control Signal to Mitigate Chattering Behavior.

1: **Output:** OCP's approximate optimal state trajectory $\tilde{\mathrm{x}}$ and control signal $\tilde{\alpha}$
2: Solve $\bar{\vartheta}$ (COCP) in (10) subject to (11), and get $\mathrm{x}_*, \beta_*$ by multiple shooting or collocation methods
3: Solve the following problem whose optimal cost is 0:

$$\inf_{\tilde{\alpha}} \int_0^T \|\tilde{\mathrm{x}}(t) - \mathrm{x}_*(t)\| dt, \tag{46}$$

where $\tilde{\mathrm{x}}$ solves $\dot{\tilde{\mathrm{x}}}(t) = f(t, \tilde{\mathrm{x}}(t), \tilde{\alpha})$ and $\tilde{\mathrm{x}}(0) = x_0$.

---

We first numerically compute an optimal state trajectory ($\mathrm{x}_*$) and a control signal ($\beta_*$) for the COCP in Theorem 1. The choice of method to solve $\bar{\vartheta}(=\vartheta)$ on line 2 of Algorithm 1 is open to the user, for example, direct methods with various optimizers [14] could be used. The rest of the algorithm is designed to find the numerical-optimal control signal $\alpha_*^\epsilon$ and state trajectory $\mathrm{x}_*^\epsilon$. On line 3 of Algorithm 1, we find a pair of $(a_i^k, b_i^k, \gamma_i^k)$ satisfying (21) for $t = t_k$, $x = \mathrm{x}_*(t_k)$, and $b = \beta_*(t_k)$. As explained in Section III-B and Fig. 2, these pairs can be analytically found using the biconjugate's geometrical property: $(\beta_*(t_k), H^*(t_k, \mathrm{x}_*(t_k), \beta_*(t_k)))$ is a convex combination of $(b_i^k, L^b(t_k, \mathrm{x}_*(t_k), b_i^k))$s. At last, on lines 4 and 5, we design $\alpha_*^\epsilon$ and compute $\mathrm{x}_*^\epsilon$ by solving (1). Note that the additional temporal discretization on line 4 could result in a frequent control-switching behavior unless $\beta_*(t) \in B(t, \mathrm{x}_*(t))$ for all $t$.

To resolve this chattering issue, we propose another algorithm if the stage cost does not depend on the control input. As in Remark 1, the OCP's state trajectory solving (1) for the designed control signal $\alpha_*^\epsilon$ (22) uniformly converges to the COCP's optimal state trajectory. Algorithm 2 utilizes this fact to avoid the control-switching behavior by removing the additional temporal discretization. Instead, Algorithm 2 has one additional optimal-control solving step as in line 3. For the OCP (46), direct methods result in nonconvex problems, but we know the optimal cost is always 0. Thus, if iterative algorithms to solve (46) converge to a local minimum with nonzero cost, then we can reinitialize a solution candidate and iterate again until the cost is close to 0. This is a big difference compared to direct methods, since it is hard to get any sense of how close locally optimal solutions by direct methods are to the optimal solution.

## IV. CONVEXITY ANALYSIS FOR THE COCP

This section investigates conditions under which the COCP has more general convexity conditions than the OCP using direct methods, including multiple shooting and collocation methods. The direct methods discretize the OCP on a temporally $\{t_0 = 0, \ldots, t_K = T\}$ and implicitly contain numerical methods, such as Euler or Runge–Kutta methods, to solve ordinary differential equations or integral equations.

Denote $\mathrm{x}[k] = \mathrm{x}(t_k)$ and $\alpha[k] = \alpha(t_k)$. x denotes both state trajectories $\mathrm{x}(\cdot)$ in continuous time and sequences $\mathrm{x}[\cdot]$ in discrete time, and the same notation rule is applied to $\alpha$ as well.

The direct methods result in finite-dimensional convex optimization problems in $\mathrm{x}[0], \alpha[0], \ldots, \mathrm{x}[K], \alpha[K] \in \mathbb{R}^{(n+m) \times (K+1)}$ if the following convexity conditions hold: for each $t$, the stage cost $L$ is convex in $(x, a) \in \mathbb{R}^n \times \mathbb{R}^m$, the terminal cost $g$ is convex in $x \in \mathbb{R}^n$, the control constraint set $A$ is convex in $\mathbb{R}^m$, the dynamics $f$ is affine in $(x, a) \in \mathbb{R}^n \times \mathbb{R}^m$, and the state constraint function $c$ is convex in $x \in \mathbb{R}^n$. Then, numerical methods for convex programming guarantee optimality. The above argument works for any different numerical integration methods, including the midpoint, trapezoidal, and Simpson's rule, and numerical ODE solving methods, including linear multistep and Runge–Kutta methods.

In this section, we analyze convexity conditions for the COCP's cost, constraint functions and dynamics, and verify the benefit of solving the COCP reformulation by direct methods over solving the original OCP formulation.

Suppose the stage cost $L$ can be decomposed into the state- and control-dependent parts:

$$L(t, x, a) = L^x(t, x) + L^a(t, a). \tag{47}$$

We first analyze the convexity conditions for the COCP's stage cost ($H^*$).

*Lemma 2:* Suppose (47) holds. Then,

$$H^*(t, x, b) = L^x(t, x) + (H^a)^*(t, x, b) \tag{48}$$

where $H^*$ is defined in (13),

$$H^a(t, x, p) := \max_{a \in A}[-p \cdot f(t, x, a) - L^a(t, a)] \tag{49}$$

$$(H^a)^*(t, x, b) := \sup_p[p \cdot b - H^a(t, x, p)]. \tag{50}$$

*Proof:* By (15), $L^b(t, x, b) = L^x(t, x) + (L^a)^b(t, x, b)$, where $(L^a)^b(t, x, b) = \min_a L^a(t, a)$ subject to $f(t, x, a) = -b$. From the definition of $H$ in (12) and the fact that $(L^b)^* \equiv H$ (see Lemma 1),

$$H(t, x, p) = -L^x(t, x) + H^a(t, x, p)$$

where $H^a(t, x, p) = ((L^a)^b)^*(t, x, p)$. Thus,

$$H^*(t, x, b) = L^x(t, x) + (H^a)^*(t, x, b).$$
∎

This lemma shows that the COCP's stage cost ($H^*$) is decomposed into the control-independent ($L^x$) and control-dependent ($(H^a)^*$) parts. One may observe that the COCP's control-independent stage cost ($L^x$) is exactly the same as the OCP's counterpart ($L^x$). Thus, if the OCP's stage cost does not depend on the control, the COCP's stage cost is the same as the OCP's stage cost.

*Corollary 1:* If $L(t, x, a) = L^x(t, x)$, then $H^*(t, x, b) = L(t, x)$.

Lemma 2 implies that the convexity of the COCP's stage cost $H^*$ is also affected by the OCP's dynamics $f$.

*Lemma 3 (Convexity of the stage cost):* Suppose Assumption 1 and (47) hold. If $L^x(t, \cdot)$ is convex in $x \in \mathbb{R}^n$ for each $t \in [0, T]$ and the dynamics $f$ is in the following form:

$$f(t, x, a) = M(t)x + \varphi(t, a) \tag{51}$$

where $M(\cdot)$ is a time-varying linear matrix and $\varphi$ is a general nonlinear function, then

$$H^*(t, x, b) = L^x(t, x) + (\bar{H}^a)^*(t, b + M(t)x) \tag{52}$$

where

$$\bar{H}^a(t, p) := \max_{a \in A}[-p \cdot \varphi(t, a) - L^a(t, a)] \tag{53}$$

and $H^*(t, \cdot, \cdot)$ is convex in $(x, b) \in \mathbb{R}^n \times \mathbb{R}^n$ for each $t \in [0, T]$.

Note that $\bar{H}^a(t, p)$ is independent of $x$ and convex in $p$ for each $t$.

*Proof:* By (49) and (50),

$$H^a(t, x, p) = -p \cdot (M(t)x) + \bar{H}^a(t, p)$$

$$(H^a)^*(t, x, b) = \sup_{b \in \mathbb{R}^n} p \cdot (b + M(t)x) - \bar{H}^a(t, p)$$

$$= (\bar{H}^a)^*(t, b + M(t)x).$$

By Lemma 2,

$$H^*(t, x, b) = L^x(t, x) + (\bar{H}^a)^*(t, b + M(t)x).$$

Since $(\bar{H}^a)^*(t, \cdot)$ is convex in $b$ and $b + M(t)x$ is affine in $(x, b)$, $(H^a)^*(t, \cdot, \cdot)$ is convex in $(x, b)$. Therefore, $H^*(t, \cdot, \cdot)$ is convex in $(x, b)$ for each $t \in [0, T]$ if $L^x(t, \cdot)$ is convex in $x$ for each $t \in [0, T]$.

TABLE II
CONVEXITY CONDITIONS FOR THE OCP AND COCP

| | OCP (4) s.t. (5) | COCP (10) s.t. (11) |
|---|---|---|
| stage cost $(L = L^x + L^a)$ | $L^x(t, \cdot)$ is convex in $x$ $L^a(t, \cdot)$ is convex in $a$ | $L^x(t, \cdot)$ is convex in $x$ See Lemma 3 |
| terminal cost $(g)$ | convex in $x$ | convex in $x$ |
| control constraint $(A)$ | convex | no condition See Lemma 4 |
| dynamics $(f)$ | $f = M(t)x + N(t)a + C(t)$ | $f = M(t)x + \varphi(t, a)$ See Lemma 3 and 4 |
| state constraint $(c)$ | $c(t, \cdot)$ is convex in $x$ | $c(t, \cdot)$ is convex in $x$ |

We define the control constraint of the COCP in $(x, b)$: for $t \in [0, T]$,

$$\bar{B}(t) := \{(x, b) \in \mathbb{R}^n \times \mathbb{R}^n \mid b \in \text{co}(B(t, x))\} \quad (54)$$

where $B(t, x)$ is defined in (8).

*Lemma 4 (Convexity of the control constraint):* Suppose Assumption 1 and (51) hold. Then, $\bar{B}(t)$ is convex in $(x, b) \in \mathbb{R}^n \times \mathbb{R}^n$ for each $t \in [0, T]$.

*Proof:* Consider $(x_1, b_1)$, $(x_2, b_2) \in \bar{B}(t)$, and $d \in [0, 1]$. Since $b_j \in \text{co}(B(t, x_j))$ for $j = 1, 2$, there exist a finite number of $a_{jl}$ and $\gamma_{jl} \in [0, 1]$ ($\sum_{l=1}^{I_j} \gamma_{jl} = 1$ for each $j$) such that

$$b_j = -M(t)x_j - \sum_{l=1}^{I_j} \gamma_{jl} \varphi(t, a_{jl}).$$

Using this, we have

$$db_1 + (1 - d)b_2 + M(t)(dx_1 + (1 - d)x_2) =$$

$$-\sum_{l=1}^{I_1} d\gamma_{1,l} \varphi(t, a_{1,l}) - \sum_{l=1}^{I_2} (1 - d)\gamma_{2,l} \varphi(t, a_{2,l}). \quad (55)$$

Since $\text{co}(\{-\varphi(t, a) \mid a \in A\}$ is a convex set, both the left- and right-hand sides of (55) belong to $\text{co}(\{-\varphi(t, a) \mid a \in A\}$. Thus, $db_1 + (1 - d)b_2$ belongs to $\text{co}(B(t, dx_1 + (1 - d)x_2))$. ∎

Table II summarizes the convexity conditions for the OCP and the COCP. If all the convexity conditions for the OCP are satisfied, the ones for the COCP are also satisfied. However, the opposite is not true.

*Remark 2 (Benefits of the COCP):* The COCP converts the OCP's nonconvexity in the control space to convexity. The convexity conditions for the COCP *do not* require that

1) the control constraint $A$ is convex;
2) the control-dependent stage $L^a(t, a)$ in (47) is convex in $a \in A \subset \mathbb{R}^m$ for each $t \in [0, T]$;
3) the control-dependent dynamics $\varphi(t, a)$ in (51) is affine in $a \in A$ for each $t \in [0, T]$.

## V. EXAMPLES AND DEMONSTRATIONS

We now introduce a 16-D example that illustrates the benefits of our COCP approach. For numerical computation, a computer with a 2.8-GHz Quad-Core i7 CPU and 16-GB RAM was used.

We consider four 4-D nonlinear vehicles (16-D): $\text{x}(t) = [\text{x}^1(t); \text{x}^2(t); \text{x}^3(t); \text{x}^4(t)] \in \mathbb{R}^{16}$, $\text{x}^l(t) = [\text{x}^l(t, 1); \text{x}^l(t, 2); \text{x}^l(t, 3); \text{x}^l(t, 4)] \in \mathbb{R}^4$, and

$$\dot{\text{x}}^l(t) = f(t, \text{x}^l(t), \alpha^l(t)) = [\text{x}^l(t, 2); \alpha^l(t, 1)\cos(\alpha^l(t, 2))$$
$$\text{x}^l(t, 4); \alpha^l(t, 1)\sin(\alpha^l(t, 2))] \quad (56)$$

where $l \in \{1, 2, 3, 4\}$ is the agent index. $\text{x}^l(t, 1)$ and $\text{x}^l(t, 2)$ are horizontal position and velocity in the 2-D space, $\text{x}^l(t, 3)$ and $\text{x}^l(t, 4)$ are vertical position and velocity in the 2-D space, and $\alpha^l(t, 1)$ and $\alpha^l(t, 2)$ are the magnitude of the acceleration and the angle of agent $l$, respectively, at time $t \in [0, 10]$.
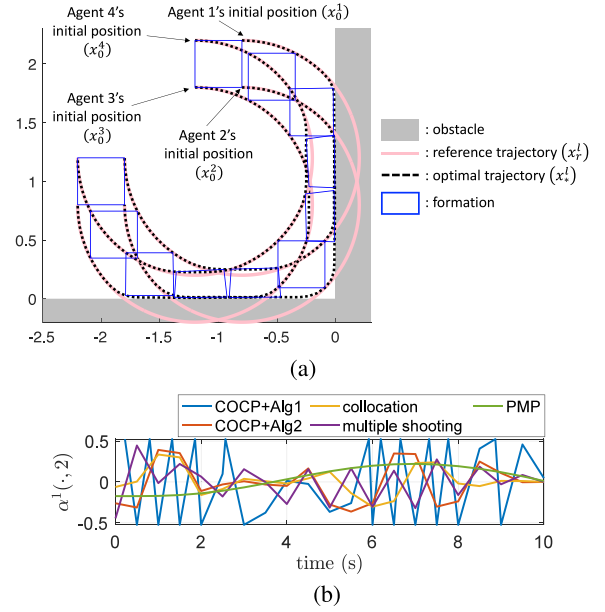


Fig. 3. (a) 16-D nonlinear system with four vehicles attempting to track their reference trajectories and maintain formation. The optimal state trajectories are computed by our COCP and Algorithm 1. The four vehicles start with a perfect square formation but compress their formation to avoid obstacles. (b) Five numericals $\alpha^1(\cdot, 2)$ computed by the five methods. Our COCP with Algorithm 1 shows control switching behavior. Algorithm 2 mitigates this, demonstrating a similar value of the sum of numerical derivatives of control signals with the other methods, as shown in Table III.

We define an OCP where four agents attempt to track their own reference trajectories while maintaining formation:

$$\inf_\alpha \int_0^{10} \sum_{l=1}^4 \text{dist}(\text{x}^l(t), \text{x}_r^l(t))dt \quad (57)$$

$$\text{subject to} \begin{cases} (56), \text{x}^l(0) = x_0^l \\ \alpha^l(t, 1) \in [-2, 2], \alpha^l(t, 2) \in [-\frac{\pi}{6}, \frac{\pi}{6}] \\ c(\text{x}(t)) \leq 0, \quad t \in [0, 10], l = 1, \ldots, 4 \end{cases} \quad (58)$$

where $\text{dist}(\text{x}^l(t), \text{x}_r^l(t))$ is the position distance of the agent $l$ from its reference trajectory in $l_2$ norm, the four reference trajectories $\text{x}_r^l$ are in a square formation at each time, $x_0^l$ is an initial state for each agent $l$, as shown in Fig. 3, and the state constraint function $c$ includes obstacles and agent-collision avoidance constraints ($c_{\text{obs}}$ and $c_{\text{col}}$, respectively): $c(\text{x}(t)) = \max\{c_{\text{obs}}(\text{x}(t)), c_{\text{col}}(\text{x}(t))\}$, where $c_{\text{obs}}(\text{x}(t)) = \max_{l=1,2,3,4} \max\{\text{x}^l(t, 1), -\text{x}^l(t, 3)\}$, $c_{\text{col}}(\text{x}(t)) = \max_{i<j}\{c_{\text{col}}^{i,j}\}$, and $c_{\text{col}}^{i,j}$ encodes an affine constraint to prevent collision between agent $i$ and $j$, for example, $c_{\text{col}}^{1,2} = -\text{x}^1(t, 3) + \text{x}^2(t, 3) + 0.2$.

The COCP formulation from Theorem 1 is

$$\inf_\beta \int_0^{10} \sum_{l=1}^4 \text{dist}(\text{x}^l(t), \text{x}_r^l(t))dt \quad (59)$$

$$\text{subject to} \begin{cases} \dot{\text{x}}^l(t) = -\beta^l(t), \text{x}^l(0) = x_0^l \\ \beta^l(t, 1) = -\text{x}^l(t, 2), \beta^l(t, 3) = -\text{x}^l(t, 4) \\ (\beta^l(t, 2))^2 + (\beta^l(t, 4))^2 \leq 4, |\beta^l(t, 4)| \leq 1 \\ c(\text{x}(t)) \leq 0, \quad t \in [0, 10], l = 1, \ldots, 4 \end{cases} \quad (60)$$

where $\beta(t) = [\beta^1(t); \beta^2(t); \beta^3(t); \beta^4(t)] \in \mathbb{R}^{16}$ and $\beta^l(t) = [\beta^l(t, 1); \beta^l(t, 2); \beta^l(t, 3); \beta^l(t, 4)] \in \mathbb{R}^4$, $l = 1, \ldots, 4$. The COCP's stage cost ($H^*$) is the same as the OCP's ($L$) by Corollary 1. The

TABLE III
PERFORMANCE COMPARISON WITH VARIOUS METHODS

| | COCP & Alg 1 | COCP & Alg 2 | multiple shooting | collocation | PMP |
|---|---|---|---|---|---|
| cost | 1.85 (0.35) | 2.65 (0.53) | 5.77 (6.05) | 6.35 (6.59) | 7.97 (3.03) |
| min cost | 1.60 | 2.15 | 1.71 | 1.78 | 6.94 |
| max cost | 2.30 | 3.55 | 15.48 | 16.72 | 14.25 |
| computation time (s) | 439.26 (161.60) | 913.46 (387.57) | 1,221.65 (866.25) | 958.72 (658.40) | 136.18 (80.22) |
| $L_2$-norm of numerical $\dot\alpha$ | 44.78 (24.12) | 13.51 (1.95) | 7.36 (0.63) | 11.98 (2.24) | 3.05 (5.17) |
| success rate of convergence | 1 | 1 | 1 | 1 | 0.3 |

TABLE IV
COMPUTATION TIME AND COST OF COCP-MPC

| receding horizon | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| computation time | 0.22 (0.21) | 0.58 (0.05) | 1.25 (0.12) | 2.49 (0.29) |
| cost | 152.23 (128.24) | 1.79 (0.03) | 1.78 (0.23) | 1.66 (0.04) |

second and third lines in (60) are explicit expression for $\mathrm{co}(B(t, \mathrm{x}(t)))$. We apply the Hermite–Simpson collocation method to the COCP with Algorithms 1 and 2, for which the interior-point method [10, Ch. 11] is utilized.

Table III compares performance of various methods in terms of cost, computation time, $L_2$-norm of the numerical derivative of control signal $\dot\alpha$ (which measures the degree of control-switching behavior), and finally, the success rate of convergence from a random initial guess. We do not compare with the HJ methods [4], Branch-and-Lift [6], and control-set convexification methods [8] because the HJ method and Branch-and-Lift are intractable to handle this 16-D system, and control-set convexification method [8] assumes convex cost, linear dynamics, and convex state constraints. The three previous methods in Table III are directly applied to the OCP. For all five methods in Table III, we utilize interior-point methods [10, Ch. 11]. For each method, we run numerical solvers with ten different solution candidates and provide averages and standard deviations (in parentheses) for each performance measure, if applicable.

Algorithm 1 can been seen to outperform the other methods in terms of optimal cost, and has a much smaller standard deviation. This is because Algorithm 1 is guaranteed to find a solution for any initial solution, whereas the other methods do not. Also, Algorithm 1 uses less computation time than the other methods, with the exception of PMP. This is because the dimension of decision variables for PMP is *significantly lower* than the other four methods: unlike the rest, the decision variables for PMP do not depend on the number of temporal discretization points but rather just on the costate dimensions. However, it is important to note that faster computation time for PMP comes at the expense of solution success rate. The interior-point method fails to find solutions to the PMP optimality conditions for 70% of initial solution candidates; by contrast, the interior-point solver guarantees convergence to the globally optimal solution for Algorithm 1, and convergence to a locally optimal solution for the direct methods for any initial solution candidate.

We propose further improvements to our method from a practical standpoint, by addressing chattering behavior, which is caused by the subdiscretization in Algorithm 1's line 4. As shown in Fig. 3(b) and Table III, our Algorithm 2 improves upon this aspect. In addition to smoother control than Algorithm 1, COCP with Algorithm 2 continues to have a better average and standard deviation in cost than the remaining three methods, and less computation time than multiple shooting and collocations.

For robustness of controls and real-time computation, our approach can be combined with model predictive control (MPC) [15]. We demonstrate our preliminary results in Table IV, which shows the computation time of Algorithm 1 for each receding horizon, and the accumulated cost in the full time horizon [0,10]. We run numerical solvers with ten different solution candidates, and Table IV provides averages (without parentheses) and standard deviations (with parentheses) for each measure. Smaller receding horizon results in shorter computation time in each horizon but worse cost. Typically, longer receding horizons generally help satisfy constraints and find better cost.

## VI. CONCLUSION

This article proposes a convexification formulation (COCP) for state-constrained OCP and presents two algorithms to compute optimal solutions. It has been proved that our COCP provides the OCP's optimal cost and state trajectory without approximation. Also, our first algorithm is guaranteed to find optimal solutions if the OCP's functions and constraints are convex in the state, and not necessarily in the control. This condition is more general than multiple shooting, collocation, and PMP's optimality conditions. Our simulation supports these results. The frequent control-switching behavior induced by our method is mitigated by our second algorithm that adds one more optimal-control-solving step. Although the additional OCP induces a nonconvex problem, its optimal cost is known to be zero. Using this fact, we can run optimizers with multiple solution candidates to find optimal solutions. Our simulation shows that the second algorithm has a similar level of control-switching behavior compared to collocation. Lastly, we demonstrate receding-horizon techniques to COCP for real-time computation.

## REFERENCES

[1] H. G. Bock and K.-J. Plitt, "A multiple shooting algorithm for direct solution of optimal control problems," *IFAC Proc. Vol.*, vol. 17, no. 2, pp. 1603–1608, 1984.

[2] M. Kelly, "An introduction to trajectory optimization: How to do your own direct collocation," *SIAM Rev.*, vol. 59, no. 4, pp. 849–904, 2017.

[3] R. F. Hartl, S. P. Sethi, and R. G. Vickson, "A survey of the maximum principles for optimal control problems with state constraints," *SIAM Rev.*, vol. 37, no. 2, pp. 181–218, 1995.

[4] A. Altarovici, O. Bokanowski, and H. Zidani, "A general Hamilton–Jacobi framework for non-linear state-constrained control problems," *ESAIM: Control, Optimisation Calculus Variations*, vol. 19, no. 2, pp. 337–357, 2013.

[5] M. E. Villanueva, C. N. Jones, and B. Houska, "Towards global optimal control via Koopman lifts," *Automatica*, vol. 132, 2021, Art. no. 109610.

[6] B. Houska and B. Chachuat, "Global optimization in Hilbert space," *Math. Program.*, vol. 173, no. 1, pp. 221–249, 2019.

[7] M. Bardi and L. C. Evans, "On Hopf's formulas for solutions of Hamilton-Jacobi equations," *Nonlinear Anal., Theory, Methods Appl.*, vol. 8, no. 11, pp. 1373–1381, 1984.

[8] B. Açıkmeşe and L. Blackmore, "Lossless convexification of a class of optimal control problems with non-convex control constraints," *Automatica*, vol. 47, no. 2, pp. 341–347, 2011.

[9] Y. Mao, M. Szmuk, and B. Açıkmeşe, "Successive convexification of nonconvex optimal control problems and its convergence properties," in *Proc. IEEE 55th Conf. Decis. Control*, 2016, pp. 3636–3641.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[11] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*. Berlin, Germany: Springer, 2008, vol. 178.

[12] L. C. Evans and P. E. Souganidis, "Differential games and representation formulas for solutions of Hamilton–Jacobi–Isaacs equations," Wisconsin Univ-Madison Mathematics Research Center, Tech. Rep., 1983.

[13] J. Borwein and A. Lewis, *Convex Analysis*. Berlin, Germany: Springer, 2006.

[14] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.

[15] J. Guanetti, Y. Kim, and F. Borrelli, "Control of connected and automated vehicles: State of the art and future challenges," *Annu. Rev. Control*, vol. 45, pp. 18–40, 2018.