

Trajectory Optimization for Nonlinear Multi-Agent Systems using Decentralized Learning Model Predictive Control*

Edward L. Zhu¹, Yvonne R. Stürz¹, Ugo Rosolia², and Francesco Borrelli¹

Abstract—We present a decentralized minimum-time trajectory optimization scheme based on learning model predictive control for multi-agent systems with nonlinear decoupled dynamics and coupled state constraints. By performing the same task iteratively, data from previous task executions is used to construct and improve local time-varying safe sets and an approximate value function. These are used in a decoupled MPC problem as terminal sets and terminal cost functions. Our framework results in a decentralized controller, which requires no communication between agents over each iteration of task execution, and guarantees persistent feasibility, finite-time closed-loop convergence, and non-decreasing performance of the global system over task iterations. Numerical experiments of a multi-vehicle collision avoidance scenario demonstrate the effectiveness of the proposed scheme.

I. INTRODUCTION

In this paper, we study the problem of decentralized Model Predictive Control (MPC) for dynamically decoupled multi-agent systems under the minimum-time cost and coupled state constraints. Multi-agent systems typically exhibit inter-agent coupling, which can be expressed as constraints on the global system. MPC is a well-studied approach to the control of such constrained systems and can be applied in a global manner for centralized control of multi-agent systems with a small number of agents. However, as the number of agents increases, centralized approaches typically become intractable in practice due to limitations in computational power and communication capacities [1]. This gives rise to decentralized and distributed MPC schemes, which leverage the inherent parallelizable structure of multi-agent systems to reduce the required computational effort. Feasibility and stability of MPC are typically obtained using a terminal cost function and terminal constraints in the MPC design [2], which we refer to as terminal components. However, synthesis of these terminal components for the control of nonlinear multi-agent systems is in general challenging. There is extensive literature on distributed [3]–[7] and decentralized MPC methods [8]–[11], as well as iterative approaches to multi-agent trajectory optimization [12], [13]. However a detailed survey goes beyond the scope of this paper.

In this paper, we propose a decentralized approach to trajectory optimization for nonlinear multi-agent systems. By

¹The authors are with the Department of Mechanical Engineering, University of California at Berkeley, Berkeley, CA 94701 USA
 edward.zhu@berkeley.edu

²The author is with the Department of Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA 91125 USA

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 846421. This work was partially funded by the grant ONR-N00014-18-1-2833.

performing the same task over multiple iterations, we collect data on the agents' closed-loop behavior to successively improve the construction of terminal components for the local controllers. In particular, as we improve an estimate of the terminal cost (an approximate value function) and expand the terminal constraint set (the domain of this function), we iteratively improve closed-loop performance of the multi-agent system, while maintaining feasibility and finite-time convergence guarantees.

The contribution of this work is twofold. We extend the method of Learning Model Predictive Control (LMPC) from [14] to the multi-agent case for dynamically decoupled agents under the minimum-time cost and coupled state constraints. In particular, we first propose a procedure for synthesizing the MPC terminal components using data from previous iterations of task execution. We then show that the resulting decentralized LMPC has the properties of persistent feasibility, finite-time closed-loop convergence to the goal state, and non-decreasing performance over iterations. We demonstrate the effectiveness of the decentralized method with a numerical example in the context of multi-vehicle collision avoidance, where we observe a significant reduction of computational effort compared to a centralized approach.

II. PRELIMINARIES

A. System Description

For a system of M agents, the set of indices $\{1, \dots, M\}$ is denoted as \mathcal{M} . ‘ \preceq ’ denotes an element-wise inequality. Consider the global nonlinear time-invariant discrete-time system composed of M agents

$$x_{t+1} = f(x_t, u_t), \quad (1)$$

where the global state and input vectors at sampling time $t \geq 0$ are formed by stacking those from each agent into a single column, i.e. $x_t = \text{col}_{i \in \mathcal{M}}(x_{i,t}) = [x_{1,t}^\top, \dots, x_{M,t}^\top]^\top \in \mathbb{R}^n$ and $u_t = \text{col}_{i \in \mathcal{M}}(u_{i,t}) \in \mathbb{R}^m$, where $x_{i,t} \in \mathbb{R}^{n_i}$ and $u_{i,t} \in \mathbb{R}^{m_i}$. Each agent in the global system is subject to local state and input constraints,

$$x_{i,t} \in \mathcal{X}_i \subseteq \mathbb{R}^{n_i}, \quad u_{i,t} \in \mathcal{U}_i \subseteq \mathbb{R}^{m_i}, \quad \forall t \geq 0. \quad (2)$$

These local constraint sets are assumed to be closed, compact, and include the goal states $x_{i,F}$ in their relative interiors. The global system is additionally subject to coupling constraints on the system state,

$$g(x_t) \preceq 0, \quad \forall t \geq 0. \quad (3)$$

We assume that the agents are dynamically decoupled with continuous dynamics and locally stabilizable, which means

that we can write (interchangeably with (1)), for all agents $i \in \mathcal{M}$, the local dynamics as

$$x_{i,t+1} = f_i(x_{i,t}, u_{i,t}). \quad (4)$$

B. Control Objective

The objective is to design a controller which drives the system state to a goal state x_F by solving the following optimal control problem for the global system

$$\min_{\mathbf{u}_T} \sum_{t=0}^T h(x_t, u_t) \text{ s.t. } \begin{cases} x_{t+1} = f(x_t, u_t), & \forall t \in \{0, \dots, T-1\} \\ x_0 = x_S, x_T = x_F \\ x_t \in \mathcal{X}, u_t \in \mathcal{U}, & \forall t \in \{0, \dots, T\} \\ g(x_t) \leq 0, & \forall t \in \{0, \dots, T\} \end{cases} \quad (5)$$

where the goal state $x_F = \text{col}_{i \in \mathcal{M}}(x_{i,F})$ is assumed to be a feasible equilibrium state of (1). $x_S = \text{col}_{i \in \mathcal{M}}(x_{i,S})$ is the initial condition of the system, and the state and input constraint sets in (5) are the Cartesian products of the local constraint sets in (2). In this work, we use the minimum-time cost, which is defined as follows:

$$h(x_t, u_t) = \mathbb{1}(x; x_F) = \begin{cases} 0 & \text{if } x = x_F \\ 1 & \text{otherwise.} \end{cases}$$

C. Learning Model Predictive Control (LMPC)

In this section, we briefly introduce and review the key concepts of LMPC [14], [15], which are extended upon in this work. LMPC was proposed as an iterative trajectory optimization method for single-agent nonlinear dynamical systems performing iterative tasks. In particular, the method provides a data-driven approach to terminal set and cost synthesis given that an initial feasible input sequence $\mathbf{u}^0 = \{u_0^0, u_1^0, \dots, u_{T^0-1}^0\}$ and closed-loop state trajectory $\mathbf{x}^0 = \{x_0^0, x_1^0, \dots, x_{T^0}^0\}$ exists for (1) and is available at iteration 0. T^q denotes the time at which the closed-loop system reaches the terminal state at iteration q , i.e. $x_{T^q}^q = x_F$. We note that for iterative tasks, the initial condition of the system is assumed to be the same over iterations, i.e. $x_0^q = x_S, \forall q \geq 0$.

LMPC solves the following finite horizon optimal control problem (FHOCUP), which approximates (5), in a receding horizon fashion.

$$\min_{\mathbf{u}_{t,N}^q} \sum_{k=0}^{N-1} \mathbb{1}(x_{k|t}^q; x_F) + V^{q-1}(x_{N|t}^q) \quad (6a)$$

$$\text{s.t. } x_{k+1|t}^q = f(x_{k|t}^q, u_{k|t}^q), \forall k \in \{0, \dots, N-1\} \quad (6b)$$

$$x_{0|t}^q = x \quad (6c)$$

$$x_{k|t}^q \in \mathcal{X}, u_{k|t}^q \in \mathcal{U}, \forall k \in \{0, \dots, N-1\} \quad (6d)$$

$$g(x_{k|t}^q) \leq 0, \forall k \in \{0, \dots, N-1\} \quad (6e)$$

$$x_{N|t}^q \in \mathcal{SS}^{q-1}, \quad (6f)$$

where $x_{k|t}^q$ and $u_{k|t}^q$ denote the decision variables of the predicted state and input at the sampling time $t+k$ of iteration q . The first input $u_{0|t}^q$ is then applied to the system (1).

The terminal set at iteration $q-1$ in (6f), called a sampled safe set, is defined as

$$\mathcal{SS}^{q-1} = \left\{ \bigcup_{p \in \mathcal{Q}^{q-1}} \bigcup_{t=0}^{T^p} x_t^p \right\}, \quad (7)$$

where $\mathcal{Q}^{q-1} = \{p \in \{0, \dots, q-1\} : x_{T^p}^p = x_F\}$ is the set of iteration indices where the goal state x_F was successfully reached. The sampled safe set collects the closed-loop state trajectories from previous successful iterations, which implies that at iteration q , $\forall x \in \mathcal{SS}^q$, there exists a known and feasible sequence of inputs, with $x_0 = x$, such that $f(x_t, u_t) \in \mathcal{SS}^q, \forall t \geq 0$. We note that by construction, \mathcal{SS}^q is a control invariant set.

For the terminal cost function in (6a), an approximate value function is constructed which returns the minimum cost-to-go (over iterations 0 to $q-1$) from each state in the safe set. The cost-to-go from the state at time t along the closed-loop trajectory \mathbf{x}^q with corresponding input sequence \mathbf{u}^q is defined as $J_{T^q}^q(x_t^q, t) = \sum_{k=t}^{T^q} \mathbb{1}(x_k^q; x_F)$.

The approximate value function at iteration $q-1$ is then

$$V^{q-1}(x) = \begin{cases} \min_{(p,t) \in \mathcal{F}^{q-1}(x)} J_{T^p}^p(x, t) & \text{if } x \in \mathcal{SS}^{q-1} \\ +\infty & \text{otherwise,} \end{cases} \quad (8)$$

where $\mathcal{F}^{q-1}(x) = \{(p, t) : x_t^p = x \text{ and } x_t^p \in \mathcal{SS}^{q-1}\}$ returns the set of iteration and time index pairs for the states in previous trajectories which are equal to x .

Using the properties of the constructed terminal components, the resulting scheme guarantees persistent feasibility of the FHOCUP and finite-time closed-loop convergence of the closed-loop system. Non-decreasing performance can also be shown over iterations of task execution.

III. PROBLEM FORMULATION

In this section, we extend the idea of LMPC to the multi-agent case. Specifically, our formulation leverages data from previous iterations of task execution to synthesize local controllers for each agent via decoupled FHOCUPs. This allows for an entirely decentralized control scheme for the multi-agent system at each iteration. In the following, we describe the synthesis of the local terminal components, namely the time-varying sampled safe sets and approximate value function. We then present the resulting decoupled FHOCUPs, which are solved in a receding horizon manner. Combining these elements, we arrive at a decentralized LMPC procedure for trajectory optimization of multi-agent systems.

A. Time-Varying Sampled Safe Sets and Global Constraint Decomposition

In this part, we address two challenges posed by a decentralized receding horizon approach. Namely 1) how global constraint satisfaction can be enforced through decentralized local constraint satisfaction, and 2) how feasibility over the entire task horizon can be maintained in a receding horizon implementation of (5), particularly when time-varying constraints are introduced.

To address 1), we introduce the following assumption. Notice that the original global time-invariant constraints are transformed into local time-varying constraints.

Assumption 1. At iteration q , there exists a time-varying local decomposition of the global constraints: $g_{i,t}^q(\cdot), \forall t \geq 0, i \in \mathcal{M}$, which can be constructed from feasible trajectories of the global system. At each time t , joint local constraint

satisfaction is sufficient for global constraint satisfaction, i.e. $g_{i,t}^q(x_{i,t}) \preceq 0, \forall i \in \mathcal{M} \implies g(x_t) \preceq 0$.

Remark 1. Given a feasible trajectory \mathbf{x}_i^q , we may always obtain the decomposition $g_{i,t}^q(x_i) = \|x_i - x_{i,t}^q\|$ for each trajectory point in \mathbf{x}_i^q , which satisfies the condition in Assumption 1. In Sec. V, we present a technique for constructing $g_{i,t}^q(\cdot)$ in a less conservative manner.

To address 2), constraint satisfaction of the FHOCPs over the entire task horizon is typically obtained via control invariant sets. This was done in [14] with (7), which, recall, are control invariant sets by definition. However, in (5) (and due to the constraint decomposition in Assumption 1), we include time-varying constraints for which the classical definition of set invariance does not apply. For this reason, we do not collect all recorded states of the global system, which converge to x_F , into a single invariant set as in (7). Instead, we interpret the states of agent i at time t of previous successful iterations as a sampled subset of the $(\hat{T}^{q-1} - t)$ -step reachable set to $x_{i,F}$, where $\hat{T}^q = \max_{p \in \{0, \dots, q\}} T_i^p$. As such, for each agent $i \in \mathcal{M}$ at iteration q , we propose to construct time-varying sampled safe sets $\mathcal{SS}_{i,t}^q$ using data from previous iterations of task execution, which are one-step reachable to $\mathcal{SS}_{i,t+1}^q$ (and therefore K -step reachable to $x_{i,F}$ for some $K \in \{1, \dots, \hat{T}^{q-1} - t\}$). Moreover, we require that the global coupling constraints are satisfied for all combinations of states in the constructed safe sets, i.e. $\forall t \geq 0$ and $\bar{x}_i \in \mathcal{SS}_{i,t}^q, g(\bar{x}) \preceq 0$, where $\bar{x} = \text{col}_{i \in \mathcal{M}}(\bar{x}_i)$. The following assumption allows us to start with non-empty safe sets at the first iteration.

Assumption 2. At iteration $q = 0$, an initial feasible input and state sequence, which converges to the goal state in T_i^0 steps, exists for each agent (4). Denote this input sequence and state trajectory for agent $i \in \mathcal{M}$ as \mathbf{u}_i^0 and \mathbf{x}_i^0 .

Remark 2. This assumption is reasonable as one may provide initial feasible trajectories through demonstration or by employing a conservative controller.

Let $(\mathcal{D}_x, \mathcal{D}_u)$ be the dataset which collects the state trajectories $\mathbf{x}_i^0, \dots, \mathbf{x}_i^q$ and input sequences $\mathbf{u}_i^0, \dots, \mathbf{u}_i^q$, and let $g_{i,t}^q(\cdot)$ be a constraint decomposition which satisfies Assumption 1. Then, initializing the recursive relationship with $\mathcal{SS}_{i,\hat{T}^{q-1}}^q = \{x_{i,F}\}$, we construct the time-varying sampled safe sets for agent i at time t of iteration q as

$$\mathcal{SS}_{i,t}^q = \left\{ x_{i,k}^p \in \mathcal{D}_x : g_{i,t}^q(x_{i,k}^p) \preceq 0 \text{ and } f_i(x_{i,k}^p, u_{i,k}^p) \in \mathcal{SS}_{i,t+1}^q \right\}. \quad (9)$$

The time-varying sampled safe sets $\mathcal{SS}_{i,t}^q$ collect the state trajectory points from previous successful iterations for which there exists a trajectory to the goal state that satisfies the global constraints. We show in Sec. IV that the sampled safe sets are a non-empty family of reachable sets to $x_{i,F}$.

Remark 3. For successful iterations, i.e. where $x_{i,T_i^q}^q = x_{i,F}$, it is straightforward to accommodate task iterations of different lengths when constructing $\mathcal{SS}_{i,t}^q$ for $t \in$

$\{0, \dots, \max_{p \in \mathcal{I}}(T_i^p)\}$. Since $x_{i,F}$ is an equilibrium of system i , we may trivially apply the zero input to obtain $\mathcal{SS}_{i,t}^q = \{x_{i,F}\}$ for all $t > T_i^q$.

Remark 4. How the constraint decomposition is constructed depends on the specific problem at hand. In the multi-vehicle collision avoidance example shown in Sec. V, we propose a procedure to construct hyperplanes which separate the position states of the sampled safe sets for pairs of agents.

B. Value Function Approximation

For an input sequence \mathbf{u}_i^q and closed-loop state trajectory \mathbf{x}_i^q with length T_i^q , we define the cost-to-go from the state $x_{i,t}^q$ at time t along the closed-loop trajectory to be

$$J_{i,T_i^q}^q(x_{i,t}^q, t) = \sum_{k=t}^{T_i^q} \mathbf{1}(x_{i,k}^q, x_{i,F}) = T_i^q - t. \quad (10)$$

The iteration cost for agent i at iteration q can then be written as $J_{i,T_i^q}^q(x_{i,S}, 0)$, recalling that the system is initialized at the same state x_S at each iteration. This leads to the definition of the approximate value function $V_i^q(\cdot, t)$ at time t over the sampled safe set $\mathcal{SS}_{i,t}^q$ as

$$V_i^q(x_i, t) = \begin{cases} \min_{(p,k) \in \mathcal{F}_{i,t}^q(x_i)} J_{i,T_i^p}^p(x_i, k), & \text{if } x_i \in \mathcal{SS}_{i,t}^q \\ +\infty, & \text{otherwise,} \end{cases} \quad (11)$$

where $\mathcal{F}_{i,t}^q(x_i)$ is defined in the same way as in (8). Thus, for a state x_i whose value is equal to some state in $\mathcal{SS}_{i,t}^q$, $V_i^q(x_i, t)$ returns the minimum cost-to-go over trajectories which pass through x_i in $\mathcal{SS}_{i,k}^q$ for $k \geq t$.

C. The Finite Horizon Optimal Control Problem

Synthesis of the terminal components is achieved using the function `synthesizeFHOCP` in Alg. 1, which acts as a global coordinator between iterations of task execution in our decentralized framework. In this algorithm, we introduce the design parameters $\bar{t}, \underline{t} \geq 0$ and $q \in \{0, \dots, q\}$ to reduce the computational burden of safe set construction and a procedure which guarantees satisfaction of one-step reachability from $\mathcal{SS}_{i,t}^q$ to $\mathcal{SS}_{i,t+1}^q$. In particular, we construct the set \mathcal{I} which contains the iteration indices of the q most recent successful iterations and the sliding time range $\mathcal{T}_t = \{\max(t - \bar{t}, 0), \dots, t, \dots, t + \bar{t}\}$ and obtain the *candidate* safe sets for time t (line 6). After a constraint decomposition is constructed (line 8, see Rem. 4 and Sec. V), the candidate safe sets are checked for constraint satisfaction (line 10). In the case where any constraint is violated, the design parameters are updated (lines 11-15) and new candidate safe sets are constructed (line 16). This procedure is repeated until safe sets satisfying (9) are found. We will show in Sec. IV that the procedure described in Alg. 1 results in non-empty safe sets, which satisfy the coupled constraints and are reachable to $x_{i,F}$. As such, after performing the task at iteration $q - 1$, we obtain for agent i the decomposition of the global constraint $g_{i,t}^{q-1}(\cdot)$, the sampled safe sets $\mathcal{SS}_{i,t}^{q-1}$, and the approximate value function $V_i^{q-1}(\cdot, t)$, which are constructed using data from iterations 0 to $q - 1$.

Algorithm 1: synthesizeFHOCP

Input: $\mathcal{D}_x, \mathcal{D}_u, \mathbf{t}, \theta = \{\underline{q}, \underline{t}, \bar{t}\}$

- 1 $\check{q} \leftarrow \underline{q}$;
- 2 $\mathcal{I} \leftarrow \{\max(q - \underline{q}), \dots, q\}$;
- 3 **for** $t \in \{0, \dots, \max_{p \in \mathcal{I}}(T_i^p)\}$ **do**
- 4 **for** $i \in \mathcal{M}$ **do**
- 5 $\mathcal{T}_t \leftarrow \{\max(t - \underline{t}, 0), \dots, t, \dots, t + \bar{t}\}$;
- 6 $\mathcal{SS}_{i,t}^q \leftarrow \bigcup_{p \in \mathcal{I}} \{x_{i,k}^p \in \mathcal{D}_x : k \in \mathcal{T}_t\}$;
- 7 **end**
- 8 $\{g_{i,t}^q(\cdot)\}_{i \in \mathcal{M}} \leftarrow$ obtain decomposition
(Assumption 1);
- 9 **end**
- 10 **while** any constraint in $g_{i,t}^q(x)$ is violated for any
 $x \in \mathcal{SS}_{i,t}^q, i \in \mathcal{M}, t \in \{0, \dots, \max_{p \in \mathcal{I}}(T_i^p)\}$ **do**
- 11 $\underline{q} \leftarrow \underline{q} - 1$;
- 12 **if** $\underline{q} = 0$ **then**
- 13 $\bar{t} \leftarrow \max(\bar{t} - 1, 0), \underline{t} \leftarrow \max(\underline{t} - 1, 0)$;
- 14 $\underline{q} \leftarrow \check{q}$;
- 15 **end**
- 16 Repeat lines 2-9 with updated \underline{t}, \bar{t} and \underline{q} ;
- 17 **end**
- 18 $V_i^q(\cdot, t) \leftarrow$ compute as in (10) and (11) $\forall x_i^q \in \mathcal{SS}_{i,t}^q$;

Output: $\{\mathcal{SS}_{i,t}^q, V_i^q(\cdot, t), g_{i,t}^q(\cdot)\}$

To obtain the control action for agent i at sampling time t of iteration q , we solve the following decoupled FHOCP with horizon length N and initial condition x_i .

$$\mathcal{P}_{i,N}^q(x_i, t) = \min_{\mathbf{u}_{i,t}^q} \sum_{k=0}^{N-1} \mathbb{1}(x_{i,k|t}^q; x_{i,F}) + V_i^{q-1}(x_{i,N|t}^q, t+N)$$

s.t. $x_{i,k+1|t}^q = f_i(x_{i,k|t}^q, u_{i,k|t}^q), \quad (12a)$

$$x_{i,0|t}^q = x_i \quad (12b)$$

$$x_{i,k|t}^q \in \mathcal{X}_i, u_{i,k|t}^q \in \mathcal{U}_i, \quad (12c)$$

$$g_{i,t+k}^{q-1}(x_{i,k|t}^q) \preceq 0, \quad (12d)$$

$$x_{i,N|t}^q \in \mathcal{SS}_{i,t+N}^{q-1} \quad (12e)$$

$$\forall k \in \{0, \dots, N-1\}$$

where (12a) and (12b) represent the dynamics and initial condition, respectively. The local state and input constraints are given in (12c). (12d) enforces satisfaction of the decomposed constraint for each agent, which is sufficient for global constraint satisfaction. Finally, (12e) ensures that the terminal state is an element of the time-varying sampled safe set $\mathcal{SS}_{i,t+N}^{q-1}$. We denote the locally optimal value of the FHOCP cost in (12) as $J_{i,N}^{*,q}(x_i, t)$.

Let $\mathbf{u}_{i,t}^{*,q}(x_i) = \{u_{i,0|t}^{*,q}(x_i), \dots, u_{i,N-1|t}^{*,q}(x_i)\}$ denote the input sequence which minimizes (12) for initial state x_i at sampling time t , and $\mathbf{x}_{i,t}^{*,q} = \{x_{i,0|t}^{*,q}, \dots, x_{i,N|t}^{*,q}\}$ be the corresponding state trajectory beginning at $x_{i,0|t}^{*,q} = x_i$. In typical *receding horizon* fashion, for each agent $i \in \mathcal{M}$, the first element of $\mathbf{u}_{i,t}^{*,q}(x_i)$ is applied to system (4), which defines the state feedback policy

$$u_{i,t}^q = \kappa_i^q(x_i, t) \doteq u_{i,0|t}^{*,q}(x_i), \quad (13)$$

Algorithm 2: Decentralized Learning Model Predictive Control

Input: $Z, \mathbf{x}^0, \mathbf{u}^0, \{T_i^0\}_{i \in \mathcal{M}}, \theta$

- 1 $\mathcal{D}_x \leftarrow \{\mathbf{x}^0\}, \mathcal{D}_u \leftarrow \{\mathbf{u}^0\}, \mathbf{t} \leftarrow \{T_i^0\}_{i \in \mathcal{M}}$;
- 2 $\{\mathcal{SS}_{i,0}^0, V_i^0(\cdot, t), g_{i,0}^0(\cdot)\} \leftarrow$ synthesizeFHOCP($\mathcal{D}_x, \mathcal{D}_u, \mathbf{t}, \theta$);
- 3 **for** $q \in \{1, \dots, Z\}$ **do**
- 4 **for** $i \in \mathcal{M}$ *in parallel* **do**
- 5 $t \leftarrow 0, x_{i,0}^q \leftarrow x_{i,S}, \mathbf{x}_i^q \leftarrow \{x_{i,S}\}, \mathbf{u}_i^q \leftarrow \emptyset$;
- 6 **while** $x_{i,t}^q \neq x_{i,F}$ **do**
- 7 $\kappa_i^q(\cdot, t) \leftarrow$ solve $\mathcal{P}_{i,N}^q(x_{i,t}^q, t)$;
- 8 $x_{i,t+1}^q \leftarrow$ apply $u_{i,t}^q = \kappa_i^q(x_{i,t}^q, t)$ and
measure state;
- 9 $\mathbf{x}_i^q \leftarrow \mathbf{x}_i^q \cup \{x_{i,t+1}^q\}, \mathbf{u}_i^q \leftarrow \mathbf{u}_i^q \cup \{u_{i,t}^q\}$;
- 10 $t \leftarrow t + 1$;
- 11 **end**
- 12 $\mathcal{D}_x \leftarrow \mathcal{D}_x \cup \{\mathbf{x}_i^q\}, \mathcal{D}_u \leftarrow \mathcal{D}_u \cup \{\mathbf{u}_i^q\}$;
- 13 $T_i^q \leftarrow t$
- 14 **end**
- 15 $\mathbf{t} \leftarrow \mathbf{t} \cup \{T_i^q\}_{i \in \mathcal{M}}$;
- 16 $\{\mathcal{SS}_{i,0}^q, V_i^q(\cdot, t), g_{i,0}^q(\cdot)\} \leftarrow$ synthesizeFHOCP($\mathcal{D}_x, \mathcal{D}_u, \mathbf{t}, \theta$);
- 17 **end**

with $x_i = x_{i,t}^q$. This results in the closed-loop state trajectory $\mathbf{x}_i^q = \{x_{i,0}^q, x_{i,1}^q, \dots, x_{i,t}^q, \dots\}$ and input sequence $\mathbf{u}_i^q = \{u_{i,0}^q, u_{i,1}^q, \dots, u_{i,t}^q, \dots\}$ for agent i at iteration q . We will show that the closed-loop state trajectory under (13) converges to the goal state $x_{i,F}$ in finite-time.

D. Decentralized Learning Model Predictive Control

The resulting iterative LMPC scheme for the multi-agent system is described in Alg. 2, where the for loop beginning at line 3 corresponds to the iterations of the decentralized LMPC, the for loop over agents from lines 4 to 14 may be executed in an entirely decentralized manner with no communication between agents, and the while loop from lines 6 to 11 iterates over time steps. Note that we assume that there is no mismatch between the model used in (12a) and the system on which the policy is applied (line 8). The inputs to Alg. 2 are $Z \geq 1$: the number of LMPC iterations, \mathbf{x}^0 and \mathbf{u}^0 : the initial feasible state and input sequences, $\{T_i^0\}_{i \in \mathcal{M}}$: the lengths of the initial trajectories, and θ which contains the design parameters for synthesizeFHOCP. It is implicitly assumed that each iteration is successful. We will show that this is true in Sec. IV.

IV. PROPERTIES OF THE DECENTRALIZED LMPC

In this section, we show that for each iteration q , the sampled safe sets as constructed in (9) have the property of reachability to $x_{i,F}$, which is sufficient for persistent feasibility of the decentralized LMPC over the entire task horizon for all iterations q . Furthermore, we show that the closed-loop system converges to the goal state in finite time, and that task performance is non-decreasing over iterations. Due to space limitations, we omit the proofs and refer the reader to [16] for details.

Proposition 1. Let Assumption 2 hold, then for agent i at iteration q , the sampled safe sets as constructed in (9) using Alg. 1 and the dataset $(\mathcal{D}_x, \mathcal{D}_u)$ are reachable to $x_{i,F}$ and satisfy the decomposed constraints $g_{i,t}^q(x) \succeq 0, \forall x \in \mathcal{SS}_{i,t}^q$.

Proposition 2. Let Assumption 2 hold and assume that $\mathcal{P}_{i,N}^q(x_{i,t}^q, t)$ defined in (12) is feasible for agent i at time t and iteration q for some $x_{i,t}^q \in \mathcal{X}_i$. If $x_{i,F} \in \mathcal{SS}_{i,t+N}^{q-1}$, then the system (4) in closed-loop with (12) and (13) converges in at most $t + \bar{T}$ steps to $x_{i,F}$, where $\bar{T} = N + T_i^{p^*} - k^*$ and $(k^*, p^*) = \arg \min_{k,p} T_i^p$ subject to $x_{i,N|t}^{*,q} = x_{i,k}^p$.

Theorem 1. Consider the system in (1) in closed loop with the decentralized LMPC (12) and (13). Let Assumption 1 and 2 hold. Then at each iteration q for every agent $i \in \mathcal{M}$: 1) (12) is feasible for all time steps $t \geq 0$. 2) The system (4) in closed-loop with (12) and (13) converges to the equilibrium point $x_{i,F}$ in finite time. 3) The sampled safe sets $\mathcal{SS}_{i,t}^q$ are non-empty for all sample times t .

Corollary 1. Consider the system (1) in closed loop with the decentralized LMPC (12) and (13). Let Assumption 2 hold. Then for every agent $i \in \mathcal{M}$, the task completion time does not increase with the iteration index q , i.e. $T_i^q \leq T_i^{q-1}$.

V. MULTI-VEHICLE COLLISION AVOIDANCE

In this section, we present a numerical example of the proposed method in the context of multi-vehicle collision avoidance. The control objective is for $M = 3$ vehicles to reach the goal equilibrium points from their respective initial states in minimum time (the code is available online^{1,2}). We model the vehicles using the kinematic bicycle model with state $\zeta_{i,t}$ and input $u_{i,t}$, which is discretized using forward Euler integration with a time step of 0.1s as in [17]. The vehicles are coupled via collision avoidance constraints where we define a circular collision buffer about the geometric center of the vehicle with radius r_i and require that for all sample times,

$$\|\zeta_{i,t}(1:2) - \zeta_{j,t}(1:2)\|_2 \geq r_i + r_j, \quad \forall i, j \in \mathcal{M}, i \neq j. \quad (14)$$

A. Decoupled FHOCP Formulation

The decoupled FHOCP is formulated as in (12). Details on the local state and input constraints can be found in [16]. In (12d), we decompose the global constraints into time-varying hyperplane constraints on the position states of each vehicle, i.e. $g_{i,t+k}^{q-1}(\zeta_{i,k|t}^q) = H_{i,t+k}^{q-1} \zeta_{i,k|t}^q + h_{i,t+k}^{q-1}$, with $H_{i,t+k}^{q-1} \in \mathbb{R}^{M-1 \times 4}$ and $h_{i,t+k}^{q-1} \in \mathbb{R}^{M-1}$ for all $k \in \{0, \dots, N\}$. This implementation of line 8 of Alg. 1 is achieved by solving a hard margin support vector machine (SVM) for all agent pairs $(i, j) \in \mathcal{M}, i \neq j$, where maximum margin separating hyperplanes can be found for time t if the position states in the candidate sampled safe sets $\mathcal{SS}_{i,t}^{q-1}$ and $\mathcal{SS}_{j,t}^{q-1}$ are linearly separable. If all $\binom{M}{2}$ pairwise SVM problems are feasible for all time t and the distance between the hyperplanes is no less than $2r$ in all cases, then we construct

¹<https://github.com/zhu-edward/multi-agent-LMPC>

²A video of the results may be found at <https://youtu.be/cB9zckRm5j8>

the constraint decomposition for agent i at time t by stacking the solution H_{ij} and h_i from all SVM problems involving agent i into the matrix $H_{i,t}^{q-1}$ and vector $h_{i,t}^{q-1}$ respectively. Otherwise, following Alg. 1, we update the sets \mathcal{T}_t and \mathcal{I} and retry the SVM problem.

We compute the initial feasible state and input sequence (ζ_i^0, u_i^0) using a linear time-varying MPC controller. At each time step, the decoupled FHOCPs are solved using IPOPT [18] by constructing a set of $|\mathcal{SS}_{i,t+N}^{q-1}|$ problems where (12e) are formulated as equality constraints for each element in the safe set. Corollary 1 is useful here as it allows us to prune the safe sets using an upper bound on the iteration cost, which can be computed for each point in the safe sets without solving the FHOCP. Specifically, at each time t , we have the cost-to-come for the current state $x_{i,t}$, we know that the sum of the stage costs over the optimization horizon is upper bounded by the horizon length N , and the cost-to-go from each point in the safe set is known. If for any safe set point, the sum of these three quantities is greater than T_i^{q-1} , we may discard that point. This helps to manage the computational burden induced by the discrete safe sets. The parameters used for this experiment can be found in [16].

B. Results and Discussion

As seen in Table I, the decentralized LMPC converges to a steady state solution where the optimal cost of each iteration is non-increasing. We additionally implemented a centralized LMPC, with the same parameters and solver, for the global system subject to the original constraint (14). This approach achieved a steady state cost of 48, which is only about a 4% difference with respect to the decentralized case. We also compare the computation time for solving a single FTOCP in the two cases. This is summarized in Table II. For the former, we record the maximum solve time over agents at each sampling time. For the latter, we record the solve time of the centralized LMPC at each sampling time. We obtain that over all iterations, computation time of the decentralized case is lower by a factor of 4.6x to 24x.

In Figures 2a) and 2b), we compare the initial feasible trajectory to the steady state trajectory at convergence. In the initial feasible trajectory, the agents' movements are intentionally staggered in time to guarantee safety. This can be clearly seen in the velocity profile at iteration 0 in Figure 2 b). At convergence, all three agents begin moving simultaneously and steer to avoid collisions around the intersection point at the origin. We notice that in the steady state input sequence, the acceleration input either saturates or is close to saturating the imposed constraint and resembles a bang-bang controller [19] which switches between acceleration and deceleration at the midpoint of the trajectory.

In Figure 1, we look closely at the steady state trajectory about the intersection point and see that the collision avoidance constraints are satisfied and are almost active for agents 1 and 3 at 2.4s and agents 2 and 3 at 2.8s. This is clearly reflected in Figure 2c) which plots the minimum pairwise distance between the three agents over iterations of decentralized LMPC.

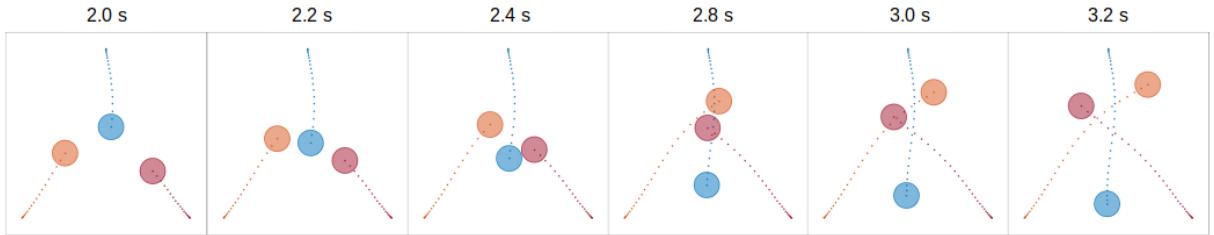


Fig. 1. Snapshots of the decentralized LMPC trajectory at convergence. Circles represent the collision buffers centered at the position of agent 1 (blue), 2 (orange), and 3 (red).

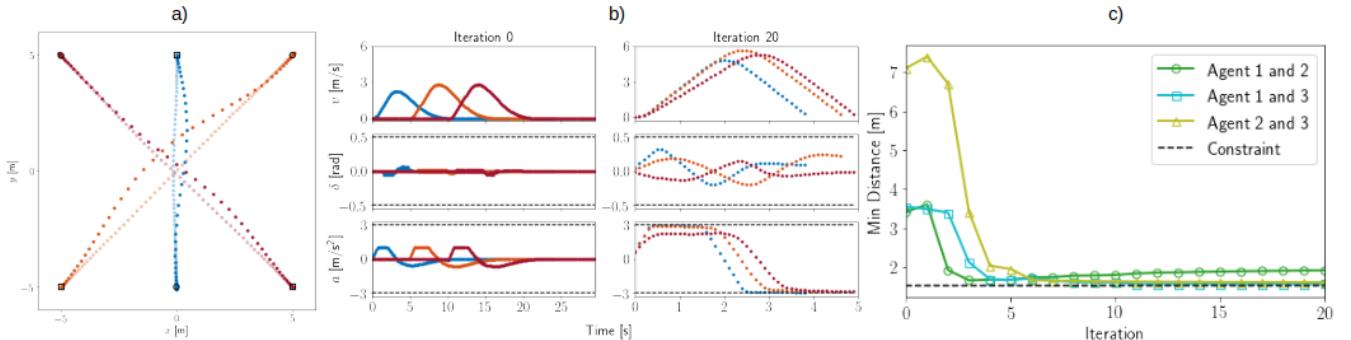


Fig. 2. a) Initial (light) vs. steady state (dark) position trajectory (top). The starting and goal states are denoted as the square and circle markers respectively. b) Velocity profile and input sequence (bottom) for agent 1 (blue), 2 (orange), and 3 (red) at the first and last iterations. The black dashed lines correspond to the input box constraints. c) Minimum distance between agents at each iteration.

TABLE I

OPTIMAL COST OF THE DECENTRALIZED LMPC AT EACH ITERATION

Iteration	Iteration Cost	Iteration	Iteration Cost
$q = 0$	296	$q = 5$	51
$q = 1$	122	$q = 6$	50
$q = 2$	79	$q = 7$	50
$q = 3$	55	$q = 8$	50
$q = 4$	51		

TABLE II

	Max Time [s]	Min Time [s]	Avg. Time [s]
Decentralized	10.2	1.97	3.35
Centralized	48.3	15.8	20.5

VI. CONCLUSION

In this paper, we presented a decentralized LMPC framework for dynamically decoupled multi-agent systems performing iterative tasks. In particular, we proposed a procedure for decomposing global constraints and synthesizing terminal sets and terminal cost functions for the decoupled FHOCP using data from previous iterations of task execution. We showed that the resulting scheme has the properties of persistent feasibility, finite-time convergence to the goal state, and non-decreasing performance over iterations.

In the multi-vehicle collision avoidance example, due to the parallelization opportunities afforded by the decentralized implementation, we observe a significant improvement in computation time compared to a centralized approach with only a 4% increase in cost. In fact, the steady state solution from the decentralized approach saw saturation of the coupling collision avoidance constraint.

REFERENCES

- [1] R. R. Negenborn, B. De Schutter, and J. Hellendoorn, “Multi-agent model predictive control: A survey,” *arXiv:0908.1076*, 2009.
- [2] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert, “Constrained model predictive control: Stability and optimality,” *Automatica*, vol. 36, no. 6, pp. 789–814, 2000.
- [3] A. Bemporad, M. Heemels, M. Johansson, *et al.*, *Networked control systems*. Springer, 2010, vol. 406.
- [4] L. Ferranti, R. R. Negenborn, T. Keviczky, and J. Alonso-Mora, “Coordination of multiple vessels via distributed nonlinear model predictive control,” in *ECC*, 2018, pp. 2523–2528.
- [5] C. E. Luis and A. P. Schoellig, “Trajectory generation for multiagent point-to-point transitions via distributed model predictive control,” *IEEE Robot. Autom. Letters*, vol. 4, no. 2, pp. 375–382, 2019.
- [6] J. Liu, D. M. de la Peña, and P. D. Christofides, “Distributed model predictive control of nonlinear systems subject to asynchronous and delayed measurements,” *Automatica*, vol. 46, no. 1, pp. 52–61, 2010.
- [7] Y. R. Stürz, E. L. Zhu, U. Rosolia, K. H. Johansson, and F. Borrelli, “Distributed learning model predictive control for linear systems,” *ArXiv:2006.13406*, 2020.
- [8] A. Alessio and A. Bemporad, “Decentralized model predictive control of constrained linear systems,” in *ECC*, 2007, pp. 2813–2818.
- [9] A. Alessio, D. Barcelli, and A. Bemporad, “Decentralized model predictive control of dynamically coupled linear systems,” *Journal of Process Control*, vol. 21, no. 5, pp. 705–714, 2011.
- [10] T. Keviczky, F. Borrelli, and G. J. Balas, “Decentralized receding horizon control for large scale dynamically decoupled systems,” *Automatica*, vol. 42, no. 12, pp. 2105–2115, 2006.
- [11] A. Filotheou, A. Nikou, and D. V. Dimarogonas, “Robust decentralised navigation of multi-agent systems with collision avoidance and connectivity maintenance using model predictive controllers,” *International Journal of Control*, vol. 93, no. 6, pp. 1470–1484, 2020.
- [12] F. Augugliaro, A. P. Schoellig, and R. D’Andrea, “Generation of collision-free trajectories for a quadrocopter fleet: A sequential convex programming approach,” in *IEEE/RSJ Int. Conf. Intell. Robots Syst.* IEEE, 2012, pp. 1917–1922.
- [13] Y. Chen, M. Cutler, and J. P. How, “Decoupled multiagent path planning via incremental sequential convex programming,” in *IEEE Int. Conf. Robot. Autom.*, 2015, pp. 5954–5961.
- [14] U. Rosolia and F. Borrelli, “Learning model predictive control for iterative tasks: a data-driven control framework,” *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 1883–1896, 2017.
- [15] ———, “Minimum time learning model predictive control,” *arXiv preprint arXiv:1911.09239*, 2019.
- [16] E. L. Zhu, Y. R. Stürz, U. Rosolia, and F. Borrelli, “Trajectory optimization for nonlinear multi-agent systems using decentralized learning model predictive control,” *arXiv preprint arXiv:2004.01298*, 2020.
- [17] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, “Kinematic and dynamic vehicle models for autonomous driving control design,” in *Intell. Veh. Symp.* IEEE, 2015, pp. 1094–1099.
- [18] A. Wächter and L. T. Biegler, “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming,” *Mathematical programming*, vol. 106, no. 1, pp. 25–57, 2006.
- [19] D. Liberzon, *Calculus of variations and optimal control theory: a concise introduction*. Princeton University Press, 2011.