

Data-Driven Nonconvex Reachability Analysis Using Exact Multiplication

Zhen Zhang¹, M. Umar B. Niazi², Michelle S. Chong³, Karl H. Johansson², and Amr Alanwar¹

Abstract—This paper addresses a fundamental challenge in data-driven reachability analysis: accurately representing and propagating non-convex reachable sets. We propose a novel approach using constrained polynomial zonotopes to describe reachable sets for unknown LTI systems. Unlike constrained zonotopes commonly used in existing literature, constrained polynomial zonotopes are closed under multiplication with constrained matrix zonotopes. We leverage this property to develop an exact multiplication method that preserves the non-convex geometry of reachable sets without resorting to approximations. We demonstrate that our approach provides tighter over-approximations of reachable sets for LTI systems compared to conventional methods.

I. INTRODUCTION

Reachability analysis provides the set of all possible states a system can transition to from a given initial state set, considering uncertainties within defined bounds or according to specified probabilistic models. It plays an essential role in providing formal guarantees required for safety-critical applications such as autonomous vehicles, robotics, and aerospace systems. Traditional model-based approaches for reachability require accurate mathematical models, which are often difficult to obtain for complex systems. Data-driven reachability analysis offers a promising alternative by leveraging measured data instead of relying on models.

Data-driven approaches for reachability can be categorized into deterministic frameworks with bounded uncertainties and stochastic frameworks with probabilistic uncertainty models. Within the deterministic framework, several set representations have been employed to obtain overapproximated reachable sets. For instance, [1] employs matrix zonotopes for representing the set of possible models consistent with noisy data, which is then used to approximate the reachable set as a constrained zonotope, a set representation introduced by [2]. To incorporate prior knowledge and reduce conservatism, [3] employs constrained matrix zonotopes and extends the method to polynomial systems and Lipschitz nonlinear systems. Similarly, [4] uses linear time-varying

approximations with bounded error terms for nonlinear systems and proposes an ellipsoidal over-approximation of the reachable set by solving a convex optimization problem. For hybrid or logical systems, [5] and [6] introduce hybrid zonotopes and logical polynomial zonotopes, respectively, for representing reachable sets.

Within the stochastic framework, [7], [8] use kernel distribution embeddings in reproducing kernel Hilbert spaces for data-driven reachability, while [9] apply empirical inverse Christoffel functions with probably approximately correct (PAC) guarantees, and [10] employ Gaussian process state space models for exact probability measures of trajectories. For human-in-the-loop systems, [11] and [12] use Gaussian mixture models and optimized disturbance bounds, respectively. Recent advances also leverage deep learning and statistical methods, e.g., [13] approximate stochastic reachability via neural networks, and [14] address distribution shift in stochastic cyber-physical systems.

Deterministic approaches to data-driven reachability are preferred when constraint violations are unacceptable, e.g., collision avoidance. This is why regulatory frameworks for safety-critical systems (e.g., DO-178C [15]: Software Considerations in Airborne Systems and Equipment Certification) often require deterministic guarantees rather than probabilistic ones. Furthermore, when data is limited or uncertainty distributions are unknown, deterministic approaches avoid potentially incorrect distributional assumptions, leading to robust guarantees even against worst-case scenarios.

Handling non-convex sets in deterministic reachability is challenging, e.g., obstacle-induced initial sets. While convex sets have a computational advantage, they sacrifice accuracy by overapproximating the reachable set [3]. Recent works [16]–[18] have addressed non-convex reachable sets to some degree. However, accurately representing and propagating non-convex sets without introducing excessive conservatism or computational complexity remains an open problem.

The challenge of non-convexity in data-driven reachability is particularly evident in linear systems where one obtains a reachable set by multiplying a set of system matrices consistent with data with a set of possible states. For instance, in [3], reachable sets and a set of possible system matrices consistent with data are described using constrained zonotopes (CZs) and constrained matrix zonotopes (CMZs), respectively. However, CZs are not closed under multiplication with CMZs. This is the reason why convex set-based approaches to reachability typically resort to approximations that overbound the resulting set, leading to conservatism.

¹ School of Computation, Information and Technology, Technical University of Munich, Germany. (Email: {zhenzhang.zhang, alanwar}@tum.de)

² Division of Decision and Control Systems, Digital Futures, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden (Email: {mubniazi, kallej}@kth.se)

³ Department of Mechanical Engineering, Eindhoven University of Technology, the Netherlands (Email: m.s.t.chong@tue.nl)

M. U. B. Niazi is supported by the European Union's Horizon Research and Innovation Programme under Marie Skłodowska-Curie grant agreement No. 101062523.

In this paper, we address this challenge in deterministic data-driven reachability by employing constrained polynomial zonotopes (CPZs) recently introduced by [19]. While [20] presents an exact multiplication for matrix zonotopes (MZs) and polynomial zonotopes (PZs), it does not incorporate constraints into the multiplication itself. Therefore, our key contribution is to add the constraints of the sets in the exact multiplication operation by leveraging the closure property of CPZs when multiplied with CMZs. This extension preserves the non-convex geometry without introducing conservatism, thereby enabling more accurate safety verification for systems that can be more precisely characterized as CMZs—as opposed to just MZs—when dealing with non-convex reachable sets.

To summarize, our contributions are twofold. First, we provide a refinement method that updates the set of possible models as new data becomes available, enabling more accurate iterative reachability analysis. For this, we provide an exact method for intersecting two CMZs. Second, we introduce constrained polynomial matrix zonotopes (CPMZs) as a novel representation and provide an exact multiplication operation between a CPMZ and a CPZ. This operation is crucial for propagating non-convex sets of states accurately without introducing the conservatism typically associated with approximation techniques. This addresses the fundamental challenge in data-driven reachability analysis of maintaining the precise geometry of non-convex reachable sets.

II. PRELIMINARIES AND PROBLEM FORMULATION

A. Notations

The sets of real and natural numbers are denoted by \mathbb{R} and \mathbb{N} , respectively, with $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We denote the matrix of zeros and ones of size $m \times n$ by $0_{m \times n}$ and $1_{m \times n}$, respectively, and the identity matrix by $I_n \in \mathbb{R}^{n \times n}$, where the subscripts are sometimes omitted to avoid clutter. For a matrix A , A^\top denotes the transpose and A^\dagger the Moore–Penrose pseudoinverse. By $A_{(i,j)}$, we denote the (i,j) -th entry and by $A_{(\cdot,j)}$, the j -th column. A slight abuse of notation is adopted for indexed matrices (e.g., A_n), where we use $A_n^{(i,j)}$ and $A_n^{(\cdot,j)}$. For a vector or list v , $v_{(i)}$ denotes the i -th element, and $v_{(p_1:p_2)} \triangleq (v_{(p_1)}, \dots, v_{(p_2)})$ its restriction. The vectorization of a matrix $A \in \mathbb{R}^{n \times m}$ is denoted by $\text{vec}(A) \in \mathbb{R}^{nm \times 1}$. For a set $\mathcal{H} = \{1, 2, \dots, |\mathcal{H}|\}$ with $|\mathcal{H}|$ denoting its cardinality, $A_{(\cdot, [1:|\mathcal{H}]])}$ denotes the matrix $[A_{(\cdot,1)}, \dots, A_{(\cdot,|\mathcal{H}]})]$. The notation $[\]$ denotes an empty matrix or vector. The operators \boxplus and \boxtimes denote exact addition and exact multiplication, respectively.

B. Problem Statement

We consider a linear time-invariant system in discrete-time $k \in \mathbb{N}_0$ with unknown system matrices Φ_{tr} and Γ_{tr} :

CZ	Constrained Zonotope
CPZ	Constrained Polynomial Zonotope
MZ	Matrix Zonotope
CMZ	Constrained Matrix Zonotope
CPMZ	Constrained Polynomial Matrix Zonotope

TABLE I: List of acronyms

$$x_{(k)} = \Phi_{\text{tr}}x_{(k-1)} + \Gamma_{\text{tr}}u_{(k-1)} + w_{(k)}, \quad (1)$$

where $x_{(k)} \in \mathbb{R}^{n_x}$ is the state at time k , $u_{(k)} \in \mathbb{R}^{n_u}$ is the control input, and $w_{(k)} \in \mathbb{R}^{n_x}$ is the unknown noise. Reachability analysis computes the set containing all possible states $x_{(k)}$ that can be reached from a compact set \mathcal{X}_0 containing initial states given a compact set $\mathcal{U}_k \ni u_{(k)}$ of feasible inputs and a compact set $\mathcal{Z}_w \ni w_{(k)}$ of possible noise.

Definition 1: The exact reachable set

$$\mathcal{R}_N \triangleq \{x_{(N)} \in \mathbb{R}^{n_x} \mid x_{(k+1)} = \Phi_{\text{tr}}x_{(k)} + \Gamma_{\text{tr}}u_{(k)} + w_{(k)}, x_{(0)} \in \mathcal{X}_0, u_{(k)} \in \mathcal{U}_k, w_{(k)} \in \mathcal{Z}_w, k = 0, \dots, N-1\} \quad (2)$$

is the set of all states that can be reached after N time steps starting from initial set \mathcal{X}_0 subject to inputs $u_{(k)} \in \mathcal{U}_k$ and noise $w_{(k)} \in \mathcal{Z}_w$, for $k = 0, \dots, N-1$. \lrcorner

Our goal is to estimate the exact reachable set of the system (1) using input-state data, without explicit knowledge of the system matrices Φ_{tr} and Γ_{tr} . Moreover, we are interested in the case where the initial set \mathcal{X}_0 is non-convex, and we must compute a reachable set that preserves this non-convex geometry without resorting to convex relaxations.

C. Set Representations

For our data-driven reachability analysis, we define the required set representations below.

Definition 2: Given an offset $c \in \mathbb{R}^n$, generator $G \in \mathbb{R}^{n \times h}$, and constraints $A \in \mathbb{R}^{n_c \times h}$ and $b \in \mathbb{R}^{n_c}$, a *constrained zonotope (CZ)* is a compact set in \mathbb{R}^n defined as (see [2])

$$\mathcal{C} = \left\{ c + \sum_{k=1}^h \alpha_{(k)} G_{(\cdot,k)} \mid \sum_{k=1}^h \alpha_{(k)} A_{(\cdot,k)} = b, \alpha_{(k)} \in [-1, 1] \right\}. \quad (3)$$

Furthermore, for reasons that will become apparent later, we associate an identifier $\text{id} \in \mathbb{N}^{1 \times h}$ with \mathcal{C} for identifying the factors $\alpha_{(1)}, \dots, \alpha_{(h)}$ in (3). To describe a CZ, we use a shorthand notation $\mathcal{C} = \langle c, G, A, b, \text{id} \rangle_{\text{CZ}}$. \lrcorner

Note that zonotopes are a special case of CZs, where constraints are empty [19], i.e., $A = []$ and $B = []$.

Definition 3: Given an offset $c \in \mathbb{R}^n$, generator $G \in \mathbb{R}^{n \times h}$, exponent $E \in \mathbb{N}_0^{p \times h}$, and constraints $A \in \mathbb{R}^{n_c \times q}$, $b \in \mathbb{R}^{n_c}$, and $R \in \mathbb{N}_0^{p \times q}$, a *constrained polynomial zonotope (CPZ)* is a compact set in \mathbb{R}^n defined as (see [19])

$$\mathcal{P} = \left\{ c + \sum_{i=1}^h \left(\prod_{k=1}^p \alpha_{(k)}^{E_{(k,i)}} \right) G_{(\cdot,i)} \mid \sum_{i=1}^q \left(\prod_{k=1}^p \alpha_{(k)}^{R_{(k,i)}} \right) A_{(\cdot,i)} = b, \alpha_{(k)} \in [-1, 1] \right\}. \quad (4)$$

Similar to CZs, we associate an identifier $\text{id} \in \mathbb{N}^{1 \times p}$ with \mathcal{P} for identifying the factors $\alpha_{(1)}, \dots, \alpha_{(p)}$ in (4). We denote a CPZ as $\mathcal{P} = \langle c, G, E, A, b, R, \text{id} \rangle_{\text{CPZ}}$. \lrcorner

Example 1: To understand the role of id , consider

$$\mathcal{P}_1 = \left\langle \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 3 & 2 \\ 1 & 5 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 0 & 2 \end{bmatrix}, [1 \ 2] \right\rangle_{\text{CPZ}} \quad (5)$$

which describes the following CPZ

$$\mathcal{P}_1 = \left\{ \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} \alpha_{(1)}^4 + \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix} \alpha_{(1)} \alpha_{(2)}^2 \mid \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \alpha_{(1)}^4 + \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} \alpha_{(1)}^2 \alpha_{(2)}^2 = \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, \alpha_{(1)}, \alpha_{(2)} \in [-1, 1] \right\}. \quad (6)$$

Here, $\text{id} = [1 \ 2]$ identifies the dependent factor $\alpha_{(1)}$ with $\text{id}_{(1)} = 1$ and $\alpha_{(2)}$ with $\text{id}_{(2)} = 2$. \lrcorner

Definition 4: Given an offset $C \in \mathbb{R}^{m \times n}$, generators $G = [G_{(1)} \dots G_{(\gamma)}] \in \mathbb{R}^{m \times (n\gamma)}$, constraints $A = [A_{(1)} \dots A_{(\gamma)}] \in \mathbb{R}^{n_c \times (n_a\gamma)}$, and $B \in \mathbb{R}^{n_c \times n_a}$, a *constrained matrix zonotope* (CMZ) is a compact set in $\mathbb{R}^{m \times n}$ defined as (see [3])

$$\mathcal{N} = \left\{ C + \sum_{k=1}^{\gamma} \alpha_{(k)} G_{(k)} \mid \sum_{k=1}^{\gamma} \alpha_{(k)} A_{(k)} = B, \alpha_{(k)} \in [-1, 1] \right\}. \quad (7)$$

We associate an identifier $\text{id} \in \mathbb{N}^{1 \times p}$ with \mathcal{N} for identifying the factors $\alpha_{(1)}, \dots, \alpha_{(p)}$. Denote $\mathcal{N} = \langle C, G, A, B, \text{id} \rangle_{\text{CMZ}}$. \lrcorner

Note that matrix zonotopes are a special case of CMZs, denoted by $\mathcal{N} = \langle C, G, [], [], \text{id} \rangle_{\text{CMZ}}$.

Definition 5: Given an offset $C \in \mathbb{R}^{m \times n}$, generators $G = [G_{(1)} \dots G_{(\gamma)}] \in \mathbb{R}^{m \times (n\gamma)}$, exponent $E \in \mathbb{N}_0^{p \times \gamma}$, constraints $A = [A_{(1)} \dots A_{(\gamma)}] \in \mathbb{R}^{n_c \times (n_a\gamma)}$, $B \in \mathbb{R}^{n_c \times n_a}$, and $R \in \mathbb{N}_0^{p \times \gamma}$, a *constrained polynomial matrix zonotope* (CPMZ) is

$$\mathcal{Y} = \left\{ C + \sum_{i=1}^{\gamma} \left(\prod_{k=1}^p \alpha_{(k)}^{E_{(k,i)}} \right) G_{(i)} \mid \sum_{i=1}^{\gamma} \left(\prod_{k=1}^p \alpha_{(k)}^{R_{(k,i)}} \right) A_{(i)} = B, \alpha_{(k)} \in [-1, 1] \right\}. \quad (8)$$

We associate an identifier $\text{id} \in \mathbb{N}^{1 \times p}$ with \mathcal{Y} for identifying the factors $\alpha_{(1)}, \dots, \alpha_{(p)}$. Furthermore, we use the shorthand notation $\mathcal{Y} = \langle C, G, E, A, B, R, \text{id} \rangle_{\text{CPMZ}}$. \lrcorner

To perform operations between two CPZs, the mergeID operator is needed to make them compatible with each other.

Proposition 1 (mergeID [21]): Given two CPZs

$$\mathcal{P}_1 = \langle c_1, G_1, E_1, A_1, b_1, R_1, \text{id}_1 \rangle_{\text{CPZ}}$$

$$\mathcal{P}_2 = \langle c_2, G_2, E_2, A_2, b_2, R_2, \text{id}_2 \rangle_{\text{CPZ}}$$

the mergeID operator returns two adjusted CPZs that are equivalent to \mathcal{P}_1 and \mathcal{P}_2 :

$$\text{mergeID}(\mathcal{P}_1, \mathcal{P}_2) = \left\{ \langle c_1, G_1, \bar{E}_1, A_1, b_1, \bar{R}_1, \bar{\text{id}} \rangle_{\text{CPZ}}, \langle c_2, G_2, \bar{E}_2, A_2, b_2, \bar{R}_2, \bar{\text{id}} \rangle_{\text{CPZ}} \right\}$$

with $\bar{\text{id}} = [\text{id}_1 \ \text{id}_2^{(\cdot, \mathcal{H})}]$, $\mathcal{H} = \{i \mid \text{id}_2^{(i)} \notin \text{id}_1\}$, and

$$\bar{E}_1 = \begin{bmatrix} E_1 \\ 0_{|\mathcal{H}| \times h_1} \end{bmatrix} \in \mathbb{R}^{a \times h_1}, \quad \bar{R}_1 = \begin{bmatrix} R_1 \\ 0_{|\mathcal{H}| \times q_1} \end{bmatrix} \in \mathbb{R}^{a \times q_1}$$

$$\bar{E}_2^{(i, \cdot)} = \begin{cases} E_2^{(j, \cdot)}, & \text{if } \exists j \ \bar{\text{id}}_{(i)} = \text{id}_2^{(j)} \\ 0_{1 \times h_2}, & \text{otherwise} \end{cases}$$

$$\bar{R}_2^{(i, \cdot)} = \begin{cases} R_2^{(j, \cdot)}, & \text{if } \exists j \ \bar{\text{id}}_{(i)} = \text{id}_2^{(j)} \\ 0_{1 \times q_2}, & \text{otherwise} \end{cases}$$

where $i = 1, \dots, a$ with $a = |\mathcal{H}| + p_1$ for $\text{id}_1 \in \mathbb{N}^{1 \times p_1}$. \lrcorner

Example 2: Consider a CPZ

$$\mathcal{P}_2 = \left\langle \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 3 & 0 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ 3 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 2 & 3 \end{bmatrix}, [1 \ 3] \right\rangle_{\text{CPZ}}$$

describing the following set

$$\mathcal{P}_2 = \left\{ \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} \alpha_{(1)}^3 \alpha_{(2)}^3 + \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix} \alpha_{(1)}^2 \mid \alpha_{(1)}^2 \alpha_{(2)}^2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha_{(2)}^3 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \alpha_{(1)}, \alpha_{(2)} \in [-1, 1] \right\} \quad (9)$$

where $\text{id} = [1 \ 3]$ identifies the dependent factor $\alpha_{(1)}$ with $\text{id}_{(1)} = 1$ and $\alpha_{(2)}$ with $\text{id}_{(2)} = 3$. If we apply the operator $\text{mergeID}(\mathcal{P}_1, \mathcal{P}_2)$ where \mathcal{P}_1 is defined in (5), we get the following sets with common identifiers.

$$\bar{\mathcal{P}}_1 = \left\langle \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 5 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 2 \\ 0 & 0 \end{bmatrix}, [1 \ 2 \ 3] \right\rangle$$

$$\bar{\mathcal{P}}_2 = \left\langle \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 3 & 0 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 3 & 2 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 0 \\ 2 & 3 \end{bmatrix}, [1 \ 2 \ 3] \right\rangle. \quad \lrcorner$$

To preserve dependencies between generators during the addition of two CPZs and prevent over-approximation, we adopt the exact addition operation proposed by [21]. By retaining dependency information, the resulting sets are less conservative, improving the computational efficiency and reliability of the resulting CPZ.

Definition 6: Let $\mathcal{P}_1 = \langle c_1, G_1, E_1, A_1, b_1, R_1, \text{id}_1 \rangle_{\text{CPZ}} \subset \mathbb{R}^n$ and $\mathcal{P}_2 = \langle c_2, G_2, E_2, A_2, b_2, R_2, \text{id}_2 \rangle_{\text{CPZ}} \subset \mathbb{R}^n$. Then, the exact addition is given by

$$\mathcal{P}_1 \boxplus \mathcal{P}_2 = \left\langle \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \begin{bmatrix} G_1 & 0_{n \times h_2} \\ 0_{n \times h_1} & G_2 \end{bmatrix}, [\bar{E}_1 \ \bar{E}_2], \begin{bmatrix} A_1 & 0_{n_{c_1} \times q_2} \\ 0_{n_{c_2} \times q_1} & A_2 \end{bmatrix}, \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, [\bar{R}_1 \ \bar{R}_2], \text{id}_{12} \right\rangle_{\text{CPZ}} \quad (10)$$

where $\{\bar{E}_1, \bar{E}_2, \bar{R}_1, \bar{R}_2, \text{id}_{12}\} \leftarrow \text{mergeID}(\mathcal{P}_1, \mathcal{P}_2)$. \lrcorner

D. Assumptions

We assume access to $n_T \in \mathbb{N}$ input-state trajectories of length $T_i + 1$, with inputs $\{u_{(k)}\}_{k=0}^{T_i-1}$ and states $\{x_{(k)}\}_{k=0}^{T_i}$ for $i \in \{0, \dots, n_T - 1\}$. For simplicity, we consider the offline data from a single trajectory ($n_T = 1$) of length T_0 and organize the input and noisy state data into matrices:

$$X_0^+ = [x_{(1)} \ x_{(2)} \ \dots \ x_{(T_0)}],$$

$$X_0^- = [x_{(0)} \ x_{(1)} \ \dots \ x_{(T_0-1)}], \quad (11)$$

$$U_0^- = [u_{(0)} \ u_{(1)} \ \dots \ u_{(T_0-1)}].$$

Let $D_0^- = [X_0^{-\top} \ U_0^{-\top}]^\top$ and $D_0 = [X_0^{+\top} \ X_0^{-\top} \ U_0^{-\top}]^\top$.

Assumption 1: For all $k \in \mathbb{Z}_{\geq 0}$, the noise $w_{(k)}$ is bounded by a known zonotope \mathcal{Z}_w , which includes the origin. \lrcorner

Assumption 2: The data matrix D_0^- has full row rank, i.e., $\text{rank}(D_0^-) = n_x + n_u$. \lrcorner

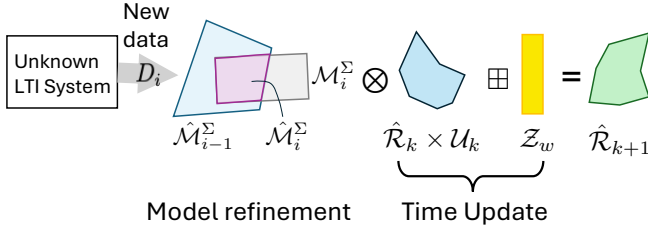


Fig. 1: Data-driven reachability analysis with model refinement. Legend: \otimes denotes the exact multiplication operation and \oplus denotes the exact addition operation.

We represent the sequence of unknown noise as $\{w(k)\}_{k=0}^{T_0}$. From Assumption 1, it follows that

$$W_0^- = [w(0) \ \cdots \ w(T_0-1)] \in \mathcal{M}_w = \langle C_{\mathcal{M}_w}, G_{\mathcal{M}_w} \rangle,$$

where $C_{\mathcal{M}_w} \in \mathbb{R}^{n_x \times n_T}$ and $G_{\mathcal{M}_w} \in \mathbb{R}^{n_x \times \gamma_{Z_w} n_T}$ with \mathcal{M}_w denoting the MZ resulting from the concatenation of multiple noise zonotopes Z_w [3]. Finally, let all system matrices $[\Phi \ \Gamma]$ of (1) consistent with the data D_- be

$$\mathcal{N}_\Sigma = \left\{ [\Phi \ \Gamma] \mid X_0^+ = \Phi X_0^- + \Gamma U_0^- + W_0^-, W_0^- \in \mathcal{M}_w \right\}. \quad (12)$$

By definition, $[\Phi_{\text{tr}} \ \Gamma_{\text{tr}}] \in \mathcal{N}_\Sigma$.

III. DATA-DRIVEN REACHABILITY ANALYSIS WITH NON-CONVEX SET REPRESENTATIONS

Given system (1) with unknown matrices Φ_{tr} and Γ_{tr} , the presence of noise implies multiple feasible matrices $[\Phi \ \Gamma]$ are consistent with the observed data. For the reachability analysis to be reliable, one must incorporate all such data-consistent models. To this end, we extend the offline approach proposed by [3] to continue refining the estimated set of models by leveraging online data.

Our proposed data-driven reachability analysis algorithm comprises two steps: first, determining a set of models consistent with observed data given bounded uncertainties; and second, using this model set to compute the region in state space containing all possible true states. Using offline data (11), we compute an initial set of models represented by a CMZ. Then, in the online phase, we refine this CMZ iteratively by collecting online input-state trajectories. For computing the reachable set as a CPZ in the online phase, we propose an exact multiplication method for the time update to multiply the CMZ representing a set of models with the CPZ representing the previous reachable set.

A. Initial Set of Models

Using the offline data (11), we compute the initial set of models \mathcal{M}_0^Σ that is consistent with the data as follows.

Lemma 1 ([3]): Consider the offline input-state data (11) such that Assumption 2 holds and consider the MZ

$$\mathcal{M}_0^\Sigma = (X_0^+ - \mathcal{M}_w) \begin{bmatrix} X_0^- \\ U_0^- \end{bmatrix}^\dagger. \quad (13)$$

Then, $[\Phi_{\text{tr}} \ \Gamma_{\text{tr}}] \in \mathcal{M}_0^\Sigma$ and $\mathcal{N}_\Sigma \subseteq \mathcal{M}_0^\Sigma$. \lrcorner

Remark 1: Requiring the full row rank in Assumption 2, i.e., $[X^{-\top} \ U^{-\top}]^\top = n_x + n_u$, implies that there exists a right-inverse of the matrix $[X^{-\top} \ U^{-\top}]^\top$. \lrcorner

B. Intersection Between CMZs

For the following result, we omit the identifier id of CMZs as it is irrelevant when intersecting two sets.

Proposition 2 (Intersection): Consider two CMZs

$$\begin{aligned} \mathcal{N}_1 &= \langle C_1, G_1, A_1, B_1 \rangle_{\text{CMZ}} \subset \mathbb{R}^{n_x \times n} \\ \mathcal{N}_2 &= \langle C_2, G_2, A_2, B_2 \rangle_{\text{CMZ}} \subset \mathbb{R}^{n_x \times n}. \end{aligned}$$

Then, their intersection is given by

$$\mathcal{N}_1 \cap \mathcal{N}_2 = \left\langle C_1, \begin{bmatrix} G_1, 0_{n_x \times (n\gamma_2)} \end{bmatrix}, \hat{A}, \hat{B} \right\rangle_{\text{CMZ}} \quad (14)$$

where

$$\hat{A} = \begin{bmatrix} \hat{A}_1 & 0_{n_{c_1} n_{a_1} \times \gamma_2} \\ 0_{n_{c_2} n_{a_2} \times \gamma_1} & \hat{A}_2 \\ \hat{G}_1 & -\hat{G}_2 \end{bmatrix}, \quad \hat{B} = \begin{bmatrix} \text{vec}(B_1) \\ \text{vec}(B_2) \\ \text{vec}(C_2 - C_1) \end{bmatrix}$$

with $\hat{A}_i = [\text{vec}(A_i^{(1)}), \dots, \text{vec}(A_i^{(\gamma_i)})]$ and $\hat{G}_i = [\text{vec}(G_i^{(1)}), \dots, \text{vec}(G_i^{(\gamma_i)})]$, for $i = 1, 2$. \lrcorner

Although the proof is non-trivial, it follows lines of arguments similar to those of [2, Proposition 1]. It is therefore omitted due to space restrictions.

Complexity: The computations of $\text{vec}(B_1)$, $\text{vec}(B_2)$, and $\text{vec}(C_2 - C_1)$ require $\mathcal{O}(n_{c_1} n_{a_1})$, $\mathcal{O}(n_{c_2} n_{a_2})$, and $\mathcal{O}(2n_x n)$ operations, respectively. The constructions of \hat{A}_1 , \hat{A}_2 , \hat{G}_1 , and \hat{G}_2 involve complexities $\mathcal{O}(n_{c_1} n_{a_1} \gamma_1)$, $\mathcal{O}(n_{c_2} n_{a_2} \gamma_2)$, $\mathcal{O}(n_x n \gamma_1)$, and $\mathcal{O}(n_x n \gamma_2)$, respectively. Thus, the overall computational complexity of intersection (2) is $\mathcal{O}(\gamma_1(n_{c_1} n_{a_1} + n_x n) + \gamma_2(n_{c_2} n_{a_2} + n_x n))$. \diamond

Remark 2: The intersection $\mathcal{N}_1 \cap \mathcal{N}_2 = \hat{\mathcal{N}}$ in (14) yields constraint matrices $\hat{A} \in \mathbb{R}^{\hat{m} \times (\gamma_1 + \gamma_2)}$ and $\hat{B} \in \mathbb{R}^{\hat{m} \times 1}$, with $\hat{m} = n_{c_1} n_{a_1} + n_{c_2} n_{a_2} + n_x n$. This corresponds to a special case of CMZs with $n_a = 1$, adopted to facilitate exact multiplications with CPZs in subsequent computations. When $\hat{m}/\hat{n}_c \in \mathbb{N}$, the vector \hat{B} can be reshaped as a matrix $\hat{B} \in \mathbb{R}^{\hat{n}_c \times (\hat{m}/\hat{n}_c)}$ by the reshaping operator defined as follows:

$$\text{Convert}(\hat{B} \in \mathbb{R}^{\hat{m} \times 1}, \hat{n}_c) = \hat{B} \in \mathbb{R}^{\hat{n}_c \times (\hat{m}/\hat{n}_c)}.$$

Accordingly, the reshaped constraint matrix \hat{A} lies in the space $\hat{A} \in \mathbb{R}^{\hat{n}_c \times ((\hat{m}/\hat{n}_c) \cdot (\gamma_1 + \gamma_2))}$, where $\hat{A} = [\text{Convert}(\hat{A}^{(1)}), \dots, \text{Convert}(\hat{A}^{(\gamma_1 + \gamma_2)})]$. \lrcorner

C. Iterative Refinement of Set of Models

Similar to Lemma 1, given any input-state data $D_i = (X_i^+, X_i^-, U_i^-)$ of system (1) over the interval $t \in [T_{i-1}, T_i]$, for $i \geq 1$, and that Assumption 2 holds, the resulting MZ \mathcal{M}_i^Σ contains all matrices $[\Phi \ \Gamma]$ consistent with the data and noise bound, i.e., $\mathcal{N}_\Sigma \subseteq \mathcal{M}_i^\Sigma$. At iteration i of the refinement phase ($i \geq 1$), we intersect the set of models \mathcal{M}_i^Σ derived from newly obtained input-state trajectories with the previous model set $\hat{\mathcal{M}}_{i-1}^\Sigma$ to get refined set $\hat{\mathcal{M}}_i^\Sigma$.

The initial set of models and refined sets are subsequently used to compute an over-approximation of the reachable sets of the unknown system, as illustrated in Fig. 1.

Proposition 3: Consider data $D_i = (X_i^+, X_i^-, U_i^-)$, for $i \geq 0$, with D_0 an initial data and D_1, D_2, \dots subsequent new data obtained from system (1), where each data D_i satisfies Assumption 2 and results in a set of models \mathcal{M}_i^Σ using Lemma 1. The refined set of models

$$\hat{\mathcal{M}}_i^\Sigma = \mathcal{M}_i^\Sigma \cap \hat{\mathcal{M}}_{i-1}^\Sigma, \quad i \geq 1, \quad \text{with} \quad \hat{\mathcal{M}}_0^\Sigma = \mathcal{M}_0^\Sigma \quad (15)$$

is a CMZ and contains the true model $[\Phi_{\text{tr}} \quad \Gamma_{\text{tr}}]$.

Proof: A straightforward consequence of Lemma 1. ■

D. Exact Multiplication for Non-Convex Reachability

At time k , as depicted in Fig. 1, we need to iteratively multiply $\hat{\mathcal{M}}_i^\Sigma$ with the cartesian product of reachable set $\hat{\mathcal{R}}_k$ and the input set \mathcal{U}_k to obtain the next reachable set $\hat{\mathcal{R}}_{k+1}$. Although $\hat{\mathcal{M}}_i^\Sigma$ is a CMZ, notice that CMZs are a special case CPMZ. Therefore, to be more general, we provide an exact multiplication method for CPMZ with CPZ.

Consider $\mathcal{Y} = \langle C_{\mathcal{Y}}, G_{\mathcal{Y}}, E_{\mathcal{Y}}, A_{\mathcal{Y}}, B_{\mathcal{Y}}, R_{\mathcal{Y}}, \text{id}_{\mathcal{Y}} \rangle_{\text{CPMZ}} \subset \mathbb{R}^{n_x \times n}$ and $\mathcal{P} = \langle c_{\mathcal{P}}, G_{\mathcal{P}}, E_{\mathcal{P}}, A_{\mathcal{P}}, b_{\mathcal{P}}, R_{\mathcal{P}}, \text{id}_{\mathcal{P}} \rangle_{\text{CPZ}} \subset \mathbb{R}^n$. Since both sets may contain shared dependent factors, consistent indexing of exponent matrices is required. To address this, we leverage the `mergeID` operator, which constructs a unified exponent representation to preserve these dependencies:

$$\text{mergeID}(\mathcal{Y}, \mathcal{P}) = \left\{ \underbrace{\langle C_{\mathcal{Y}}, G_{\mathcal{Y}}, \bar{E}_{\mathcal{Y}}, A_{\mathcal{Y}}, B_{\mathcal{Y}}, \bar{R}_{\mathcal{Y}}, \text{id}_{\mathcal{Y}\mathcal{P}} \rangle_{\text{CPMZ}}}_{\mathcal{Y}}, \underbrace{\langle c_{\mathcal{P}}, G_{\mathcal{P}}, \bar{E}_{\mathcal{P}}, A_{\mathcal{P}}, b_{\mathcal{P}}, \bar{R}_{\mathcal{P}}, \text{id}_{\mathcal{Y}\mathcal{P}} \rangle_{\text{CPZ}}}_{\mathcal{P}} \right\}. \quad (16)$$

Proposition 4 (Exact Multiplication): Given a CPMZ $\mathcal{Y} = \langle C_{\mathcal{Y}}, G_{\mathcal{Y}}, E_{\mathcal{Y}}, A_{\mathcal{Y}}, B_{\mathcal{Y}}, R_{\mathcal{Y}}, \text{id}_{\mathcal{Y}} \rangle_{\text{CPMZ}} \subset \mathbb{R}^{n_x \times n}$ and a CPZ $\mathcal{P} = \langle c_{\mathcal{P}}, G_{\mathcal{P}}, E_{\mathcal{P}}, A_{\mathcal{P}}, b_{\mathcal{P}}, R_{\mathcal{P}}, \text{id}_{\mathcal{P}} \rangle_{\text{CPZ}} \subset \mathbb{R}^n$, the following identity holds

$$\mathcal{Y} \otimes \mathcal{P} = \left\langle C_{\mathcal{Y}}c_{\mathcal{P}}, [G_{\mathcal{Y}}c_{\mathcal{P}} \quad C_{\mathcal{Y}}G_{\mathcal{P}} \quad G_f], E_{\mathcal{Y}\mathcal{P}}, A_{\mathcal{Y}\mathcal{P}}, B_{\mathcal{Y}\mathcal{P}}, R_{\mathcal{Y}\mathcal{P}}, \text{id}_{\mathcal{Y}\mathcal{P}} \right\rangle_{\text{CPZ}}, \quad (17)$$

where $\mathcal{Y} \otimes \mathcal{P} \subset \mathbb{R}^{n_x}$ and

$$A_{\mathcal{Y}\mathcal{P}} = \begin{bmatrix} \text{vec}(A_{\mathcal{Y}}^{(1)}) & \dots & \text{vec}(A_{\mathcal{Y}}^{(\gamma)}) & 0_{n_c n_a \times q_{\mathcal{P}}} \\ 0_{m_{\mathcal{P}} \times 1} & \dots & 0_{m_{\mathcal{P}} \times 1} & A_{\mathcal{P}} \end{bmatrix}$$

$$B_{\mathcal{Y}\mathcal{P}} = \begin{bmatrix} \text{vec}(B_{\mathcal{Y}}) \\ b_{\mathcal{P}} \end{bmatrix}$$

$$E_{\mathcal{Y}\mathcal{P}} = \left[\bar{E}_{\mathcal{Y}}, \bar{E}_{\mathcal{P}}, [\bar{E}_{\mathcal{Y}}^{(:,1)} + \bar{E}_{\mathcal{P}}^{(:,1)}], \dots, [\bar{E}_{\mathcal{Y}}^{(:,1)} + \bar{E}_{\mathcal{P}}^{(:,h_{\mathcal{P}})}], \dots, [\bar{E}_{\mathcal{Y}}^{(:,\gamma)} + \bar{E}_{\mathcal{P}}^{(:,1)}], \dots, [\bar{E}_{\mathcal{Y}}^{(:,\gamma)} + \bar{E}_{\mathcal{P}}^{(:,h_{\mathcal{P}})}] \right]$$

$$R_{\mathcal{Y}\mathcal{P}} = [\bar{R}_{\mathcal{Y}} \quad \bar{R}_{\mathcal{P}}]$$

$$G_f = \begin{bmatrix} g_f^{(1)} & \dots & g_f^{(h_{\mathcal{P}}\gamma)} \end{bmatrix}$$

with $\bar{E}_{\mathcal{Y}}, \bar{E}_{\mathcal{P}}, \bar{R}_{\mathcal{Y}}, \bar{R}_{\mathcal{P}}, \text{id}_{\mathcal{Y}\mathcal{P}}$ obtained from (16) and

$$g_f^{(k)} = G_{\mathcal{Y}}^{(i)} G_{\mathcal{P}}^{(:,j)}, \quad k = h_{\mathcal{P}}(i-1) + j,$$

for $i = 1, \dots, \gamma, j = 1, \dots, h_{\mathcal{P}}$.

Proof: Let $\hat{\mathcal{P}}$ be the right-hand side of (17) and let $Y \in \mathcal{Y}$ and $p \in \mathcal{P}$. We will prove that $\mathcal{Y}\mathcal{P} \subseteq \hat{\mathcal{P}}$ and $\hat{\mathcal{P}} \subseteq \mathcal{Y}\mathcal{P}$, for all $Y \in \mathcal{Y}$ and $p \in \mathcal{P}$. Note that with the implementation of `mergeID`(\mathcal{Y}, \mathcal{P}), Y and p can be written as:

$$\exists \hat{\alpha}_{([1:a])} : \quad Y = C_{\mathcal{Y}} + \sum_{i=1}^{\gamma} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,i)})} \right) G_{\mathcal{Y}}^{(i)}$$

$$\exists \hat{\alpha}_{([1:a])} : \quad p = c_{\mathcal{P}} + \sum_{i=1}^{h_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{P}}^{(\bar{E}_{\mathcal{P}}^{(k,i)})} \right) G_{\mathcal{P}}^{(:,i)} \quad (18)$$

where $a = |\mathcal{H}| + \gamma$, $\mathcal{H} = \{i \mid \text{id}_{\mathcal{P}(i)} \notin \text{id}_{\mathcal{Y}}\}$. Thus, we have

$$Yp = C_{\mathcal{Y}}c_{\mathcal{P}} + \sum_{i=1}^{\gamma} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,i)})} \right) G_{\mathcal{Y}}^{(i)} c_{\mathcal{P}}$$

$$+ C_{\mathcal{Y}} \sum_{i=1}^{h_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{P}}^{(\bar{E}_{\mathcal{P}}^{(k,i)})} \right) G_{\mathcal{P}}^{(:,i)}$$

$$+ \sum_{i=1}^{\gamma} \sum_{j=1}^{h_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,i)})} \right) \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{P}}^{(\bar{E}_{\mathcal{P}}^{(k,j)})} \right) G_{\mathcal{Y}}^{(i)} G_{\mathcal{P}}^{(:,j)}. \quad (19)$$

For the second and third terms on the right-hand side of (19), we have two sets of factors, each containing γ and $h_{\mathcal{P}}$ elements. Consequently, the first $\gamma + h_{\mathcal{P}}$ entries of $\hat{\alpha}$ and the columns of $E_{\mathcal{Y}\mathcal{P}}$ are defined accordingly as follows:

$$\hat{\alpha}_{([1:\gamma+h_{\mathcal{P}}])} = \left[\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,1)})}, \dots, \prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,\gamma)})}, \prod_{k=1}^a \hat{\alpha}_{\mathcal{P}}^{(\bar{E}_{\mathcal{P}}^{(k,1)})}, \dots, \prod_{k=1}^a \hat{\alpha}_{\mathcal{P}}^{(\bar{E}_{\mathcal{P}}^{(k,h_{\mathcal{P}})})} \right] \quad (20)$$

$$E_{\mathcal{Y}\mathcal{P}}^{([1:\gamma+h_{\mathcal{P}}])} = [\bar{E}_{\mathcal{Y}}, \bar{E}_{\mathcal{P}}]. \quad (21)$$

Because of `mergeID`(\mathcal{Y}, \mathcal{P}) in (16), the fourth term on the right-hand side of (19) can be expressed as:

$$\sum_{i=1}^{\gamma} \sum_{j=1}^{h_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,i)})} \hat{\alpha}_{\mathcal{P}}^{(\bar{E}_{\mathcal{P}}^{(k,j)})} \right) G_{\mathcal{Y}}^{(i)} G_{\mathcal{P}}^{(:,j)} = \sum_{i=1}^{\gamma} \sum_{j=1}^{h_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,i)} + \bar{E}_{\mathcal{P}}^{(k,j)})} \right) G_{\mathcal{Y}}^{(i)} G_{\mathcal{P}}^{(:,j)}. \quad (22)$$

Concatenating the factors in (22), we have

$$\hat{\alpha}_{([\gamma+h_{\mathcal{P}}+1:\gamma+h_{\mathcal{P}}+\gamma h_{\mathcal{P}}])}$$

$$= \left[\prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,1)} + \bar{E}_{\mathcal{P}}^{(k,1)})}, \dots, \prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,1)} + \bar{E}_{\mathcal{P}}^{(k,h_{\mathcal{P}})})}, \dots, \prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,\gamma)} + \bar{E}_{\mathcal{P}}^{(k,1)})}, \dots, \prod_{k=1}^a \hat{\alpha}_{\mathcal{Y}}^{(\bar{E}_{\mathcal{Y}}^{(k,\gamma)} + \bar{E}_{\mathcal{P}}^{(k,h_{\mathcal{P}})})} \right],$$

which results in $E_{\mathcal{Y}\mathcal{P}}^{([\gamma+h\mathcal{P}+1:\gamma+h\mathcal{P}+\gamma h\mathcal{P}]})$ and G_f as follows:

$$E_{\mathcal{Y}\mathcal{P}}^{([\gamma+h\mathcal{P}+1:\gamma+h\mathcal{P}+\gamma h\mathcal{P}]}) = \left[\left[\overline{E}_{\mathcal{Y}}^{(\cdot,1)} + \overline{E}_{\mathcal{P}}^{(\cdot,1)} \right], \dots, \left[\overline{E}_{\mathcal{Y}}^{(\cdot,1)} + \overline{E}_{\mathcal{P}}^{(\cdot,h\mathcal{P})} \right], \dots, \left[\overline{E}_{\mathcal{Y}}^{(\cdot,\gamma)} + \overline{E}_{\mathcal{P}}^{(\cdot,1)} \right], \dots, \left[\overline{E}_{\mathcal{Y}}^{(\cdot,\gamma)} + \overline{E}_{\mathcal{P}}^{(\cdot,h\mathcal{P})} \right] \right]$$

$$G_f = \left[G_{\mathcal{Y}}^{(1)} G_{\mathcal{P}}^{(\cdot,1)}, \dots, G_{\mathcal{Y}}^{(1)} G_{\mathcal{P}}^{(\cdot,h\mathcal{P})}, \dots, G_{\mathcal{Y}}^{(\gamma)} G_{\mathcal{P}}^{(\cdot,1)}, \dots, G_{\mathcal{Y}}^{(\gamma)} G_{\mathcal{P}}^{(\cdot,h\mathcal{P})} \right].$$

Secondly, we find the constraints on $\hat{\alpha}_{([1:a])}$ in (18). For $Y \in \mathcal{Y}$ and $p \in \mathcal{P}$, Yp should satisfy the constraints simultaneously. From (16), we have

$$\sum_{i=1}^{\gamma} \left(\prod_{k=1}^{\gamma} (\alpha_{\mathcal{Y}}^{(k)}) R_{\mathcal{Y}}^{(k,i)} \right) \text{vec}(A_{\mathcal{Y}}^{(i)}) = \sum_{i=1}^{\gamma} \left(\prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{R}_{\mathcal{Y}}^{(k,i)}} \right) \text{vec}(A_{\mathcal{Y}}^{(i)}) = \text{vec}(B_{\mathcal{Y}}) \quad (23)$$

and

$$\sum_{i=1}^{q_{\mathcal{P}}} \left(\prod_{k=1}^{h_{\mathcal{P}}} (\alpha_{\mathcal{P}}^{(k)}) R_{\mathcal{P}}^{(k,i)} \right) A_{\mathcal{P}}^{(\cdot,i)} = \sum_{i=1}^{q_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{R}_{\mathcal{P}}^{(k,i)}} \right) A_{\mathcal{P}}^{(\cdot,i)} = b_{\mathcal{P}}. \quad (24)$$

Combining (23) and (24), the following holds

$$\sum_{i=1}^{\gamma} \left(\prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{R}_{\mathcal{Y}}^{(k,i)}} \right) \begin{bmatrix} \text{vec}(A_{\mathcal{Y}}^{(\cdot,i)}) \\ 0_{m_{\mathcal{P}} \times 1} \end{bmatrix} + \sum_{i=1}^{q_{\mathcal{P}}} \left(\prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{R}_{\mathcal{P}}^{(k,i)}} \right) \begin{bmatrix} 0_{n_c n_a \times 1} \\ A_{\mathcal{P}}^{(\cdot,i)} \end{bmatrix} = \begin{bmatrix} \text{vec}(B_{\mathcal{Y}}) \\ b_{\mathcal{P}} \end{bmatrix} \quad (25)$$

which results in $A_{\mathcal{Y}\mathcal{P}}$ and $B_{\mathcal{Y}\mathcal{P}}$. Thus, $Yp \in \hat{\mathcal{P}}$ and therefore $\mathcal{Y}\mathcal{P} \subseteq \hat{\mathcal{P}}$. Conversely, let $\hat{p} \in \hat{\mathcal{P}}$, then

$$\exists \hat{\alpha}_{([1:\gamma+h\mathcal{P}+\gamma h\mathcal{P}])}: \hat{p} = \hat{c} + \sum_{i=1}^{\gamma+h\mathcal{P}+\gamma h\mathcal{P}} \left(\prod_{k=1}^a \hat{\alpha}_{(k)}^{E_{\mathcal{Y}\mathcal{P}}^{(k,i)}} \right) G_f^{(i)}.$$

By partitioning

$$\hat{\alpha}_{([1:\gamma+h\mathcal{P}+\gamma h\mathcal{P}])} = \left[\prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{E}_{\mathcal{Y}}^{(k,[1:\gamma])}}, \prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{E}_{\mathcal{P}}^{(k,[1:h\mathcal{P}])}}, \prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{E}_{\mathcal{Y}}^{(k,1)} + \overline{E}_{\mathcal{P}}^{(k,1)}}, \dots, \prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{E}_{\mathcal{Y}}^{(k,1)} + \overline{E}_{\mathcal{P}}^{(k,h\mathcal{P})}}, \dots, \prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{E}_{\mathcal{Y}}^{(k,\gamma)} + \overline{E}_{\mathcal{P}}^{(k,1)}}, \dots, \prod_{k=1}^a \hat{\alpha}_{(k)}^{\overline{E}_{\mathcal{Y}}^{(k,\gamma)} + \overline{E}_{\mathcal{P}}^{(k,h\mathcal{P})}} \right]$$

it follows that there exist $Y \in \mathcal{Y}$ and $p \in \mathcal{P}$ such that $\hat{p} = Yp$. Meanwhile, since $\hat{p} \in \hat{\mathcal{P}}$, it holds that

$$\sum_{i=1}^{q_{\mathcal{P}}+\gamma} \left(\prod_{k=1}^a \hat{\alpha}_{(k)}^{R_{\mathcal{Y}\mathcal{P}}^{(k,i)}} \right) A_{\mathcal{Y}\mathcal{P}}^{(\cdot,i)} = \begin{bmatrix} \text{vec}(B_{\mathcal{Y}\mathcal{P}}) \\ b_{\mathcal{P}} \end{bmatrix} \quad (26)$$

which satisfies the constraints in (25). Therefore, $\hat{p} \in \mathcal{Y}\mathcal{P}$ and thus $\hat{\mathcal{P}} \subseteq \mathcal{Y}\mathcal{P}$. ■

As noted before, $\mathcal{N} = \langle C_{\mathcal{N}}, G_{\mathcal{N}}, A_{\mathcal{N}}, B_{\mathcal{N}}, \text{id}_{\mathcal{N}} \rangle_{\text{CPMZ}}$ is a special case of CPMZ and thus can be reformulated as

$\mathcal{Y} = \langle C_{\mathcal{N}}, G_{\mathcal{N}}, E_{\mathcal{N}}, A_{\mathcal{N}}, B_{\mathcal{N}}, R_{\mathcal{N}}, \text{id}_{\mathcal{N}} \rangle_{\text{CPMZ}}$ with $E_{\mathcal{N}} = R_{\mathcal{N}} = I_{\gamma_{\mathcal{N}}}$. Then, according to Proposition 4, the CPMZ is multiplied exactly with a CPZ to yield a new CPZ, as illustrated by Fig. 2.

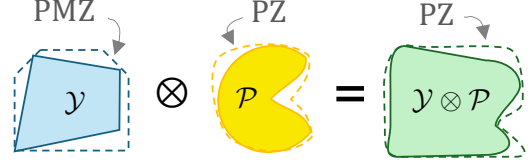


Fig. 2: Illustration of the proof of Proposition 4, where the dashed lines indicate the boundary of PMZ and PZs, respectively. Let $c(\mathcal{Y})$ and $c(\mathcal{P})$ denote the constraints associated with CPMZ \mathcal{Y} and CPZ \mathcal{P} , respectively. The exact multiplication of $\mathcal{Y} := \{\text{PMZ} : c(\mathcal{Y}) \text{ hold}\}$ with $\mathcal{P} := \{\text{PZ} : c(\mathcal{P}) \text{ hold}\}$ is a CPZ $\mathcal{Y} \otimes \mathcal{P} = \{\text{PZ} : c(\mathcal{Y} \otimes \mathcal{P}) \text{ hold}\}$, with constraints $c(\mathcal{Y} \otimes \mathcal{P}) := \text{Proj}_{\mathbb{R}^n} c(\mathcal{Y})$ and $c(\mathcal{P})$, where $\text{Proj}_{\mathbb{R}^n} c(\mathcal{Y})$ is the projection of $c(\mathcal{Y})$ from $\mathbb{R}^{n_x \times n}$ to \mathbb{R}^n .

Complexity: Computation of $\text{vec}(B_{\mathcal{Y}})$, $A_{\mathcal{Y}\mathcal{P}}$, $E_{\mathcal{Y}\mathcal{P}}$, $C_{\mathcal{Y}\mathcal{P}}$, $G_{\mathcal{Y}\mathcal{P}}$, $C_{\mathcal{Y}}G_{\mathcal{P}}$, and G_f has complexity $\mathcal{O}(n_c n_a)$, $\mathcal{O}(n_c n_a \gamma)$, $\mathcal{O}((|\mathcal{H}| + \gamma)\gamma h_{\mathcal{P}})$, $\mathcal{O}(n_x n)$, $\mathcal{O}(\gamma n_x n)$, $\mathcal{O}(n_x n h_{\mathcal{P}})$, and $\mathcal{O}(\gamma n_x n h_{\mathcal{P}})$, respectively. Then, we obtain the overall complexity of computing exact multiplication (17) to be $\mathcal{O}(n_c n_a \gamma + (|\mathcal{H}| + \gamma)\gamma h_{\mathcal{P}} + \gamma n_x n h_{\mathcal{P}})$. ◊

Algorithm 1 Data-driven Non-Convex Reachability Analysis

Input: Initial input-state data $D_0 = (X_0^+, X_0^-, U_0^-)$, initial set \mathcal{X}_0 , noise zonotope \mathcal{Z}_w , input zonotopes \mathcal{U}_k ;

Output: Reachable sets $\hat{\mathcal{R}}_k$ for $k \in \mathbb{N}$.

Offline: Initialization

- 1: $\hat{\mathcal{R}}_0 \leftarrow \mathcal{X}_0$
- 2: $\mathcal{M}_0^\Sigma \leftarrow (X_0^+ - \mathcal{M}_w) \begin{bmatrix} X_0^- \\ U_0^- \end{bmatrix}^\dagger$
- 3: Set $\hat{\mathcal{M}}^\Sigma \leftarrow \mathcal{M}_0^\Sigma$
- 4: Set $X^+ \leftarrow []$, $X^- \leftarrow []$, $U^- \leftarrow []$

Online: Set Refinement and Reachability Analysis

- 1: **while** $k \geq 1$ **do**
- 2: Construct online data matrices $X^+ = [X^+ \ x_{(k)}]$, $X^- = [X^- \ x_{(k-1)}]$, $U^- = [U^- \ u_{(k-1)}]$
- 3: **if** $\text{rank} \left(\begin{bmatrix} X^- \\ U^- \end{bmatrix} \right) = n_x + n_u$ **then**
- 4: $\mathcal{M}^\Sigma \leftarrow (X^+ - \mathcal{M}_w) \begin{bmatrix} X^- \\ U^- \end{bmatrix}^\dagger$
- 5: $\hat{\mathcal{M}}^\Sigma \leftarrow \mathcal{M}^\Sigma \cap \hat{\mathcal{M}}^\Sigma$
- 6: $X^+ \leftarrow [X^+, X^-]$, $X^- \leftarrow [X^-, X^-]$, $U^- \leftarrow [U^-, U^-]$
- 7: **end if**
- 8: $\hat{\mathcal{R}}_{k+1} \leftarrow \hat{\mathcal{M}}^\Sigma \otimes (\hat{\mathcal{R}}_k \times \mathcal{U}_k) \boxplus \mathcal{Z}_w$
- 9: $k \leftarrow k + 1$.
- 10: **end while**

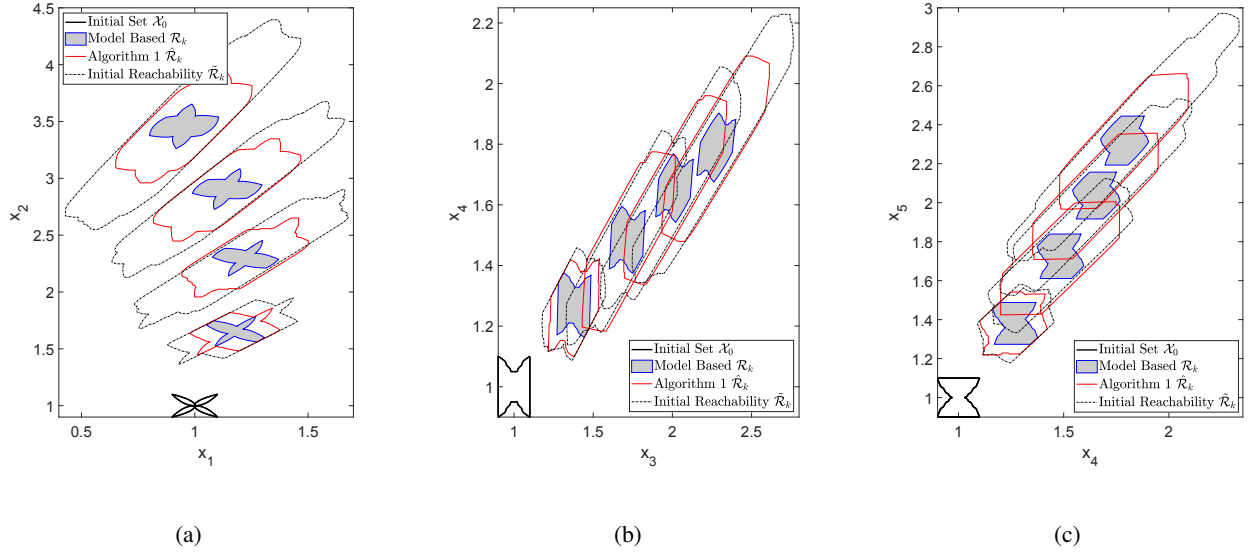


Fig. 3: Projection of Reachable Sets Computed from Input-State Data Using Algorithm 1 with a Non-Convex Initial Set

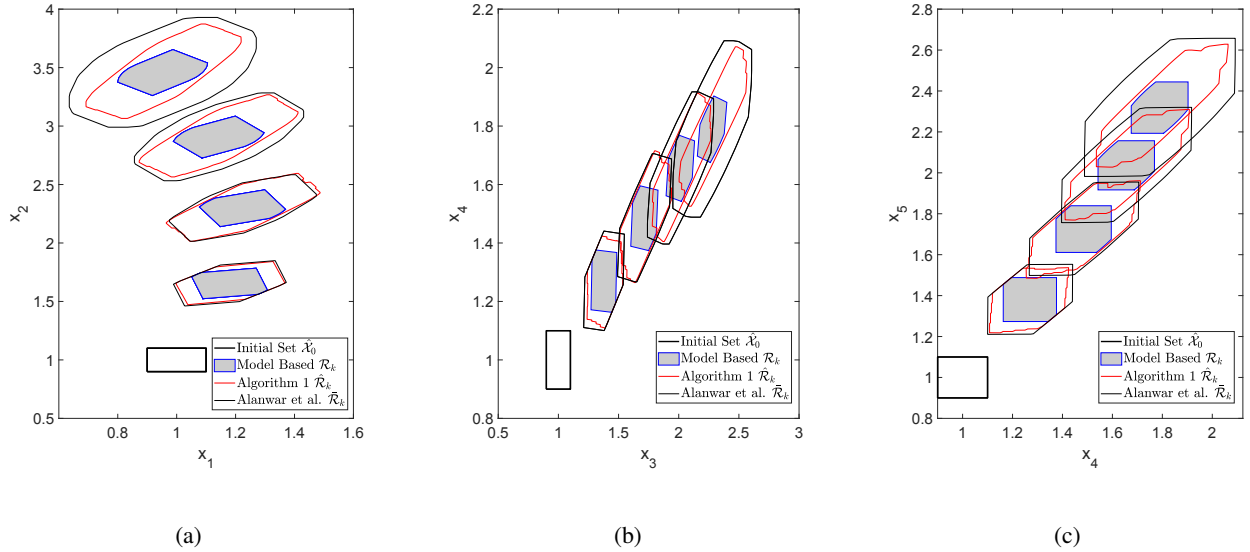


Fig. 4: Comparative Projection of Reachable Sets from Input-State Data Using Algorithm 1 and Alanwar et al. [3]

IV. NUMERICAL SIMULATIONS

Consider a five-dimensional system which is a discretization of the system used in [3] with sampling time 0.05 sec. All used code to reproduce our results are publicly available¹. The system has the following matrices.

$$\Phi_{tr} = \begin{bmatrix} 0.9323 & -0.1890 & 0 & 0 & 0 \\ 0.1890 & 0.9323 & 0 & 0 & 0 \\ 0 & 0 & 0.8596 & 0.0430 & 0 \\ 0 & 0 & -0.0430 & 0.8596 & 0 \\ 0 & 0 & 0 & 0 & 0.9048 \end{bmatrix},$$

$$\Gamma_{tr} = [0.0436 \quad 0.0533 \quad 0.0475 \quad 0.0453 \quad 0.0476]^\top.$$

¹<https://github.com/TUM-CPS-HN/Data-Driven-Nonconvex-Reachability-Analysis>

In the first experiment, the initial non-convex set is chosen to be $\mathcal{X}_0 = \langle c_0, G_0, E_0, [], [], [], \text{id}_0 \rangle_{\mathcal{P}}$, where

$$c_0 = 1_{5 \times 1}, \quad G_0 = 0.1I_5, \quad E_0 = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

For the second experiment, a convex initial set $\hat{\mathcal{X}}_0 = \langle 1_{5 \times 1}, 0.1I_5 \rangle$ is used to compare reachable sets computed via Algorithm 1 and the method of Alanwar et al. [3]. In both experiments, the input set is defined as $\mathcal{U}_k = \langle 10, 0.25 \rangle$, and random noise is sampled from the zonotope $\mathcal{Z}_w = \langle 0, [0.005, \dots, 0.005]^\top \rangle$.

Firstly, the reachable sets obtained from a non-convex ini-

tial set using both the nominal system model and Algorithm 1 are illustrated in Fig. 3. The reachable set $\hat{\mathcal{R}}_k$ corresponds to the one computed using the model set $\mathcal{M}_0^\triangleright$, without applying any set refinement. $\hat{\mathcal{R}}_k$ represents the reachable set computed after performing the first online set refinement using new incoming data through Algorithm 1. Owing to the exact set multiplication employed in the algorithm, the non-convexity of the initial set is preserved throughout the reachability propagation process. Furthermore, Proposition 4 guarantees that the initial reachability computation remains applicable to non-convex initial sets by employing exact multiplication, thus addressing the limitation of the method proposed by Alanwar et al. [3], which is confined to convex cases.

Secondly, to evaluate the performance of the proposed algorithm, a comparative analysis was conducted under two scenarios. First, trajectory data collected over the interval $[0, T_1]$ with N_1 trajectories were used to perform the offline initialization step of Algorithm 1. At time T_1 , new data over $[T_1, T_2]$ containing N_2 trajectories became available, triggering the online set refinement step of Algorithm 1 and yielding the updated reachable set $\hat{\mathcal{R}}_k$. For consistency and fair comparison with the method of Alanwar et al. [3], we set $N_1 = N_2$. The reachable set $\bar{\mathcal{R}}_k$ computed using the method of Alanwar et al. [3] is based on the combined dataset D_a over the full interval $[0, T_2]$.

As illustrated in Fig. 4, the reachable set $\hat{\mathcal{R}}_k$ computed using the Algorithm 1 is less conservative than $\bar{\mathcal{R}}_k$ obtained using the method of Alanwar et al. [3]. This improvement stems from the incremental refinement of the model set, where newly acquired data are used to further constrain the previously computed set of models. As a result, a more accurate representation of system behavior is achieved in the form of a CMZ, rather than a MZ. The use of exact set multiplication between CMZ and CPZ in Algorithm 1 ensures exact set propagation, thereby further reducing conservativeness. These results highlight the advantages of the proposed incremental approach over direct one-shot data-driven methods, demonstrating its improved ability to capture system dynamics and perform set propagation.

V. CONCLUSION

We presented a data-driven framework for reachability analysis using non-convex set representations. We introduced a set refinement strategy that incrementally incorporates newly collected input-state data, enabling iterative model set updates. This refinement process is supported by an intersection operation for CMZs, allowing system dynamics to be represented more accurately over time. Furthermore, we introduced the CPMZ as a new set representation that enables exact multiplication with CPZs. Since CMZs are a special case of CPMZs, the proposed method seamlessly applies to the exact multiplication between CMZs and CPZs. This enables the exact propagation of non-convex reachable sets by avoiding the over-approximation errors inherent in the non-exact arithmetic of zonotope operations. The proposed approach thus achieves improved accuracy and reduced conservativeness in safety verification. Future work will

focus on extending the framework to systems with nonlinear dynamics.

REFERENCES

- [1] A. Alanwar, A. Koch, F. Allgöwer, and K. H. Johansson, "Data-driven reachability analysis using matrix zonotopes," in *Learning for Dynamics and Control*, 2021, pp. 163–175.
- [2] J. K. Scott, D. M. Raimondo, G. R. Marseglia, and R. D. Braatz, "Constrained zonotopes: A new tool for set-based estimation and fault detection," *Automatica*, vol. 69, pp. 126–136, 2016.
- [3] A. Alanwar, A. Koch, F. Allgöwer, and K. H. Johansson, "Data-driven reachability analysis from noisy data," *IEEE Transactions on Automatic Control*, vol. 68, no. 5, pp. 3054–3069, 2023.
- [4] H. Park, V. Vijay, and I. Hwang, "Data-driven reachability analysis for nonlinear systems," *IEEE Control Systems Letters*, 2024.
- [5] J. A. Siefert, T. J. Bird, J. P. Koeln, N. Jain, and H. C. Pangborn, "Robust successor and precursor sets of hybrid systems using hybrid zonotopes," *IEEE Control Systems Letters*, vol. 7, pp. 355–360, 2022.
- [6] A. Alanwar, F. J. Jiang, and K. H. Johansson, "Polynomial logical zonotope: A set representation for reachability analysis of logical systems," *Automatica*, vol. 171, p. 111896, 2025.
- [7] A. J. Thorpe and M. M. Oishi, "Model-free stochastic reachability using kernel distribution embeddings," *IEEE Control Systems Letters*, vol. 4, no. 2, pp. 512–517, 2019.
- [8] A. J. Thorpe, K. R. Ortiz, and M. M. Oishi, "SReachTools kernel module: Data-driven stochastic reachability using Hilbert space embeddings of distributions," in *60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 5073–5079.
- [9] A. Devonport, F. Yang, L. El Ghaoui, and M. Arcak, "Data-driven reachability and support estimation with Christoffel functions," *IEEE Transactions on Automatic Control*, vol. 68, no. 9, pp. 5216–5229, 2023.
- [10] P. Griffioen and M. Arcak, "Data-driven reachability analysis for gaussian process state space models," in *62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 4100–4105.
- [11] J. Choi, S. Byeon, and I. Hwang, "Data-driven forward stochastic reachability analysis for human-in-the-loop systems," in *62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 1730–1735.
- [12] V. Govindarajan, K. Driggs-Campbell, and R. Bajcsy, "Data-driven reachability analysis for human-in-the-loop systems," in *56th IEEE Conference on Decision and Control (CDC)*, 2017, pp. 2617–2622.
- [13] K. Sivaramakrishnan, V. Sivaramakrishnan, R. A. Devonport, and M. M. Oishi, "Stochastic reachability of uncontrolled systems via probability measures: Approximation via deep neural networks," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE, 2024, pp. 7534–7541.
- [14] N. Hashemi, L. Lindemann, and J. V. Deshmukh, "Statistical reachability analysis of stochastic cyber-physical systems under distribution shift," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 43, no. 11, pp. 4250–4261, 2024.
- [15] L. Rierison, *Developing safety-critical software: a practical guide for aviation software and DO-178C compliance*. CRC Press, 2017.
- [16] E. Dietrich, A. Devonport, and M. Arcak, "Nonconvex scenario optimization for data-driven reachability," in *6th Annual Learning for Dynamics & Control Conference*, 2024, pp. 514–527.
- [17] Z. Wang, B. Chen, R. M. Jungers, and L. Yu, "Data-driven reachability analysis of lipschitz nonlinear systems via support vector data description," in *62nd IEEE Conference on Decision and Control (CDC)*, 2023, pp. 7043–7048.
- [18] F. Djeumou, A. P. Vinod, E. Goubault, S. Putot, and U. Topcu, "On-the-fly control of unknown systems: From side information to performance guarantees through reachability," *IEEE Transactions on Automatic Control*, vol. 68, no. 8, pp. 4857–4872, 2022.
- [19] N. Kochdumper and M. Althoff, "Constrained polynomial zonotopes," *Acta Informatica*, vol. 60, no. 3, pp. 279–316, 2023.
- [20] E. Luo, N. Kochdumper, and S. Bak, "Reachability analysis for linear systems with uncertain parameters using polynomial zonotopes," in *Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control*, 2023, pp. 1–12.
- [21] N. Kochdumper and M. Althoff, "Sparse polynomial zonotopes: A novel set representation for reachability analysis," *IEEE Transactions on Automatic Control*, vol. 66, no. 9, pp. 4043–4058, 2020.