IEEE Control Systems Letters paper presented at
2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Scalable Reachset-Conformant Identification of Linear Systems

Laura Lützow and Matthias Althoff

***Abstract*— By monitoring the set of reachable outputs, safety can be verified. However, to compute the reachable set of real-world systems, we require models that are able to produce all possible system behaviors. These kinds of models are called reachset-conformant, and their identification is a promising new research direction. While many existing reachset-conformant identification techniques require the computation of the halfspace representation of the zonotopic reachable sets, we propose an approach that leads to the same optimal identification results using the more scalable generator representation. Thus, our approach offers greater efficiency for high-dimensional systems and long time horizons. The scalability and accuracy of both approaches are compared in numerical experiments with linear time-variant systems.**

I. Introduction

The synthesis and verification of controllers for real systems require conformant models that can capture all possible system behaviors [1]. Considering all real-world effects in mathematical modeling, such as friction or a temperature dependence of model parameters, would lead to complex models unsuitable for formal controller synthesis or verification. Thus, one often identifies simple models with bounded uncertainties from data.

With set-membership approaches, we can compute the set of parameters compatible with the model structure, the measurements, and the error bounds. The computations can be done using interval methods [2]–[4] or overapproximating the set of feasible parameters by ellipsoids [5]. For additive zonotopic uncertainties, the set of models consistent with the data can be computed as a matrix zonotope [6]. However, these approaches require that overapproximations of the uncertainty bounds are known, as we might not find a nonempty set of parameters otherwise.

Alternatively, we can determine the system model with reachset-conformant identification methods. The main idea is to directly identify the model and the uncertainty bounds from real-world data, such that all measured system trajectories are contained in the reachable set of the model, as illustrated in Fig. 1. Because this research area emerged recently, only a limited amount of previous work exists: In [7], a piecewise linear model of an analog circuit is identified, and reachset conformance is achieved by adapting the additive model and measurement errors. The estimation of reachset-conformant disturbance sets and system matrices of a linear system can

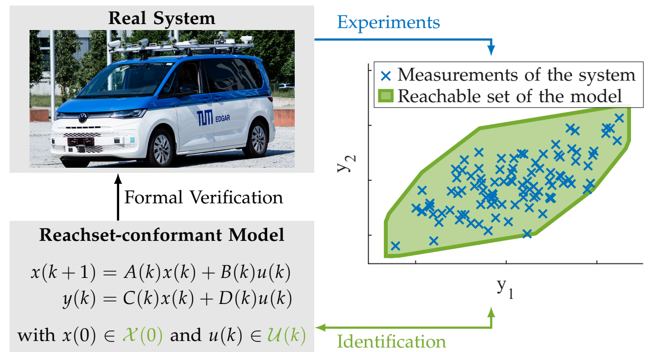


Fig. 1. Reachset-conformant identification of the uncertainty sets $\mathcal{X}(0)$ and $\mathcal{U}(k)$ of a linear time-variant model.

be formulated as a convex programming problem, which minimizes the volume of the disturbance set [8]. To obtain tight reachable sets, we can also directly minimize the size of the reachable set: By converting the reachable set represented with zonotopes to its halfspace representation, reachset-conformant disturbance sets for linear systems can be identified using linear programming [9], [10]. A nonlinear programming layer can be added to estimate unknown model parameters. This idea has been extended to simultaneously identify the initial state uncertainty set [11]. Depending on the chosen norm for evaluating the size of the reachable set, the problem can be solved with linear or quadratic programming. However, the above reachset-conformant identification methods are only valid for linear time-invariant systems and do not scale well with increasing system dimensions.

This paper presents the following contributions:

1) We generalize the state-of-the-art reachset-conformant identification approach [10], [11] to linear time-variant systems.
2) We propose a new reachset-conformant identification method for linear time-invariant or time-variant systems, which does not require the halfspace representation of zonotopes as in [10], [11] and, thus, significantly improves the scalability with respect to the system dimension and the time horizon.
3) We compare the computational complexity of both identification approaches.
4) We conduct numerical experiments to demonstrate the scalability and accuracy of the proposed approaches.

Sec. II presents fundamentals on reachability analysis and zonotopes and introduces the problem statement. In Sec. III, we describe two approaches to identify the uncertainty bounds of linear time-variant systems. Sec. IV compares the computational complexity of the proposed approaches.

This work was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1608 – under grant number 501798263 and in part by the European Commission project justITSELF under grant number 817629.

L. Lützow and M. Althoff are both with the School of Computation, Information and Technology, Technical University of Munich, Garching, Germany. {laura.luetzow@tum.de, althoff@in.tum.de}

In Sec. V, we demonstrate the scalability and accuracy of our approaches with experiments. Sec. VI concludes this work.

II. Preliminaries

A. Notation

We denote sets by calligraphic letters, matrices by uppercase letters, and vectors and scalars by lowercase letters. The symbols $\mathbf{1}$, $\mathbf{I}$, and $\mathbf{0}$ represent a vector filled with ones, the identity matrix, and a matrix of zeros of appropriate dimensions, respectively. The matrix dimensions can be explicitly mentioned in the subscript. We use $\mathrm{diag}(v)$ for a diagonal matrix whose diagonal are the elements of the vector v, $\mathrm{diag}(A_1, A_2, ..., A_n)$ is a blockdiagonal matrix where the matrices A_1, A_2, ..., A_n are concatenated diagonally, and $\mathrm{diag}_n(A_1, A_2, ..., A_n)$ represents a blockdiagonal matrix where n copies of the matrix $\mathrm{diag}(A_1, A_2, ..., A_n)$ are concatenated diagonally. The expression $\mathrm{vert}_n(A)$ is the vertical concatenation of n copies of the matrix A.

We further introduce three important set operations: A linear transformation of a set $\mathcal{S} \subset \mathbb{R}^n$ with a matrix $A \in \mathbb{R}^{m\times n}$ is defined as $A\mathcal{S} = \{As | s \in \mathcal{S}\}$. The Minkowski sum of two sets $\mathcal{S}_a,\ \mathcal{S}_b \subset \mathbb{R}^n$ is defined as $\mathcal{S}_a \oplus \mathcal{S}_b = \{s_a + s_b | s_a \in \mathcal{S}_a,\ s_b \in \mathcal{S}_b\}$, and the Cartesian product is defined as $\mathcal{S}_a \times \mathcal{S}_b = \{[s_a^\top\ s_b^\top]^\top \mid s_a \in \mathcal{S}_a,\ s_b \in \mathcal{S}_b\}$.

B. Reachability Analysis

We use reachability analysis to compute the set of outputs that is reachable for a system starting at an uncertain initial state and subject to uncertain inputs. Our data is given in the form of test cases:

Definition 1 (Test Case): *The m-th test case, $m = 1, ..., n_m$, consists of a nominal initial state $x_*^{(m)}(0) \in \mathbb{R}^{n_x}$, the nominal input trajectory $u_*^{(m)}(k) \in \mathbb{R}^{n_u}$, $k = 0, ..., n_k - 1$, and the measured output trajectory $y^{(m)}(k) \in \mathbb{R}^{n_y}$, $k = 0, ..., n_k - 1$.*

Due to real-world uncertainties, the actual initial state $x^{(m)}(0)$ and input trajectory $u^{(m)}(\cdot)$ deviate from the nominal signals $x_*^{(m)}(0)$ and $u_*^{(m)}(\cdot)$. We assume that these deviations can be described by the initial state uncertainty $\mathcal{X}_0$ and the constant input uncertainty $\mathcal{U}_c$ such that

$$
\begin{aligned}
x^{(m)}(0) &\in \mathcal{X}^{(m)}(0),\\
\forall k\colon\ u^{(m)}(k) &\in \mathcal{U}^{(m)}(k),
\end{aligned}
$$

where $\mathcal{X}^{(m)}(0) := x_*^{(m)}(0) \oplus \mathcal{X}_0,\ \mathcal{U}^{(m)}(k) := u_*^{(m)}(k) \oplus \mathcal{U}_c$.

Definition 2 (Reachable Set): *The reachable set $\mathcal{Y}^{(m)}(k)$ of the system S for the m-th test case is defined as*

$$
\begin{aligned}
\mathcal{Y}^{(m)}(k) = \big\{\tau(k; x(0), u(\cdot)) \;\big|\; & x(0) \in \mathcal{X}^{(m)}(0),\\
& u(\tilde{k}) \in \mathcal{U}^{(m)}(\tilde{k}),\ \tilde{k} = 0, ..., k\big\},
\end{aligned}
$$

where $\tau(k; x(0), u(\cdot))$ is the output of system S at time step k starting at the initial state $x(0)$ and applying the input trajectory $u(\cdot)$.

In this work, we consider linear time-variant (LTV), discrete-time systems S_{LTV} of the form

$$
x(k+1) = A(k)x(k) + B(k)u(k) \tag{1a}
$$
$$
y(k) = C(k)x(k) + D(k)u(k). \tag{1b}
$$

Proposition 1 (Reachability Analysis for LTV Systems): *The reachable set of the system S_{LTV} can be computed as*

$$
\mathcal{Y}^{(m)}(k) = y_*^{(m)}(k) \oplus \mathcal{Y}_a(k) \tag{2}
$$

with the nominal output

$$
y_*^{(m)}(k) = \bar{C}(k)x_*^{(m)}(0) + \sum_{j=0}^{k} \bar{D}_j(k)u_*^{(m)}(j),
$$

the set of possible deviations from the true output to the nominal output due to uncertainties

$$
\mathcal{Y}_a(k) = \bar{C}(k)\mathcal{X}_0 \oplus \bigoplus_{j=0}^{k} \bar{D}_j(k)\mathcal{U}_c, \tag{3}
$$

and the time-varying matrices

$$
\bar{C}(k) = C(k)\prod_{i=1}^{k} A(k-i),
$$
$$
\bar{D}_j(k) = \begin{cases} C(k)\left(\prod_{i=1}^{k-j-1} A(k-i)\right)B(j) & 0 \le j < k,\\ D(k) & j = k. \end{cases}
$$

Proof. By recursively applying (1a) and inserting the result in (1b), we obtain

$$
y(k) = \bar{C}(k)x(0) + \sum_{j=0}^{k} \bar{D}_j(k)u(k). \tag{4}
$$

A set-based evaluation of (4) with $x(0) \in x_*^{(m)}(0) \oplus \mathcal{X}_0$ and $u(k) \in u_*^{(m)}(k) \oplus \mathcal{U}_c$ results in (2). □

Please denote that we do not restrict the model space by only considering input and initial state uncertainties. Models with additive process disturbance $w \in \mathcal{W}$ and measurement noise $v \in \mathcal{V}$ can be written as in (1) by defining the input vector as $[u^\top\ w^\top\ v^\top]^\top \in \mathcal{U} \times \mathcal{W} \times \mathcal{V}$ and augmenting the input and feedthrough matrix to $[B\ \mathbf{I}\ \mathbf{0}]$ and $[D\ \mathbf{0}\ \mathbf{I}]$, respectively. The nominal signal corresponding to w and v can be set to zero.

C. Zonotopes

In this paper, we represent reachable sets with zonotopes.

Definition 3 (Generator Representation of Zonotopes [12]): *A zonotope $\mathcal{Z} \subset \mathbb{R}^n$ can be described by*

$$
\mathcal{Z} = \left\{ c + \sum_{i=1}^{\eta} \lambda_i g_i \,\middle|\, \lambda_i \in [-1, 1] \right\} = \langle c, G \rangle,
$$

where $c = \mathrm{cen}(\mathcal{Z}) \in \mathbb{R}^n$ is the zonotope center, $G = \mathrm{gen}(\mathcal{Z}) = [g_1\ ...\ g_\eta] \in \mathbb{R}^{n\times\eta}$ is the generator matrix, and η is the number of generators.

Alternatively, a zonotope can be described by its halfspace representation:

Definition 4 (Halfspace Representation of Zonotopes [13]): *A zonotope can be represented as the intersection of a finite number of halfspaces:*

$$\mathcal{Z} = \{x \in \mathbb{R}^n | Nx \leq d\},$$

where the rows of $N \in \mathbb{R}^{h\times n}$ contain the normal vectors of each halfspace, the corresponding element in $d \in \mathbb{R}^h$ is the offset to the origin, and h is the number of halfspaces.

A zonotope in generator representation can be converted to its halfspace representation by computing the halfspace normal vectors N with [13, Thm. 7] and the offset as

$$d = Nc + |NG|\mathbf{1}. \tag{5}$$

The number of halfspaces $h = 2\binom{\eta}{n-1}$ grows exponentially with the dimension n [13], so the generator representation is often preferred. Furthermore, many set operations can be efficiently performed in generator representation [14]: Given two zonotopes $\mathcal{Z}_a = \langle c_a, G_a\rangle$ and $\mathcal{Z}_b = \langle c_b, G_b\rangle$, the Minkowski sum can be computed as $\mathcal{Z}_a \oplus \mathcal{Z}_b = \langle c_a + c_b, [G_a\ G_b]\rangle$, and the Cartesian product as $\mathcal{Z}_a \times \mathcal{Z}_b = \langle [c_a^\top\ c_b^\top]^\top, \mathrm{diag}(G_a, G_b)\rangle$. The linear transformation of $\mathcal{Z}_a$ with a matrix A can be computed as $A\mathcal{Z}_a = \langle Ac_a, AG_a\rangle$.

To evaluate the size of a zonotope, we will use the interval norm in this work:

Definition 5 (Interval Norm of Zonotopes [11]): *The interval norm for the zonotope $\mathcal{Z} = \langle c, G\rangle$ is defined as the absolute sum over all elements of G:*

$$\|\mathcal{Z}\|_I = \mathbf{1}^\top |G| \mathbf{1}.$$

Verifying whether a point $z \in \mathbb{R}^n$ is contained in a zonotope can be done for the halfspace representation by checking whether $Nz \leq d$ holds. Similarly, we can use the definition of the generator representation leading to the containment constraint

$$z \in \mathcal{Z} = \langle c, G\rangle \iff \exists \lambda \in \mathbb{R}^\eta \colon |\lambda| \leq \mathbf{1},\ z = c + G\lambda. \tag{6}$$

D. Problem Statement

In our work, the uncertainty sets $\mathcal{X}_0$ and $\mathcal{U}_c$ for the reachset-conformant model are represented by the zonotopes

$$\mathcal{X}_0 = \langle c_x, G_x \mathrm{diag}(\alpha_x)\rangle,\ \alpha_x \in \mathbb{R}^{\eta_x}_{\geq 0}, \tag{7a}$$
$$\mathcal{U}_c = \langle c_u, G_u \mathrm{diag}(\alpha_u)\rangle,\ \alpha_u \in \mathbb{R}^{\eta_u}_{\geq 0}. \tag{7b}$$

The generator templates $G_x \in \mathbb{R}^{n_x \times \eta_x}$ and $G_u \in \mathbb{R}^{n_u \times \eta_u}$ can be chosen arbitrarily. Initial states or inputs without uncertainty can be modeled by setting the corresponding rows in G_x or G_u to zero. To shorten the notation, we will write α for the stacked scaling factors $[\alpha_x^\top\ \alpha_u^\top]^\top$ and c for the stacked center vectors $[c_x^\top\ c_u^\top]^\top$. Furthermore, we will often refrain from explicitly mentioning the index m of the test case in the superscript.

By establishing reachset conformance, we can transfer safety properties from a simulation model to the real system despite uncertainties [1]. Thus, we want to identify optimal scaling factors α and center vectors c, such that all measurements of the system are contained in the reachable set of the model, while minimizing the interval norm of the reachable set:

Problem 1 (Reachset-conformant Identification). *Given test cases $m = 1, ..., n_m$ of the real system, the scaling factors α and center vectors c of a reachset-conformant model are identified by solving*

$$\underset{\alpha, c}{\arg\min} \sum_{m=1}^{n_m} \sum_{k=0}^{n_k - 1} w(k) \|\mathcal{Y}^{(m)}(k)\|_I \tag{8a}$$
$$\text{s.t. } \forall k, \forall m \colon y^{(m)}(k) \in \mathcal{Y}^{(m)}(k), \tag{8b}$$

where $\mathcal{Y}^{(m)}(k)$ is the reachable set of the model, $y^{(m)}(k)$ are measurements of the real system, and $w(k)$ is the weight for time step k.

Approaches to efficiently solve Prob. 1 for LTV models S_{LTV} are developed in the following section.

III. Reachset-conformant Identification of LTV Systems

In this section, we present two linear programs that solve Prob. 1. First, we derive a linear cost function that describes the interval norm of the reachable sets. Then, we describe two linear programs that use the derived cost function but different formulations of the conformance constraints (8b).

To simplify later computations, we briefly derive the generator representation of $\mathcal{Y}^{(m)}(k)$ from Prop. 1 using the uncertainty sets from (7): Insertion of (7) in (3) yields the zonotope

$$\mathcal{Y}^{(m)}(k) = \langle y_*^{(m)}(k) + \mathrm{cen}\big(\mathcal{Y}_a(k)\big),\quad \mathrm{gen}\big(\mathcal{Y}_a(k)\big)\rangle \tag{9}$$

with

$$\mathrm{cen}\big(\mathcal{Y}_a(k)\big) = \bar{C}(k)c_x + \sum_{j=0}^{k} \bar{D}_j(k)c_u, \tag{10a}$$
$$\mathrm{gen}\big(\mathcal{Y}_a(k)\big) = \mathrm{gen}'\big(\mathcal{Y}_a(k)\big)\mathrm{diag}\Big(\big[\alpha_x^\top \quad \alpha_u^\top \quad \cdots \quad \alpha_u^\top\big]^\top\Big), \tag{10b}$$

where

$$\mathrm{gen}'\big(\mathcal{Y}_a(k)\big) = \big[\bar{C}(k)G_x \quad \bar{D}_0(k)G_u \quad \cdots \quad \bar{D}_k(k)G_u\big].$$

A. Cost Function

A linear formulation of (8a) can be derived by adapting the approach from [10] to LTV systems:

Lemma 1 (Linear Cost Function): *The sum over the interval norms of the reachable sets can be expressed linearly in the scaling factors α:*

$$\sum_{k=0}^{n_k - 1} w(k) \|\mathcal{Y}(k)\|_I = \gamma\alpha, \tag{11}$$

where $\alpha \geq \mathbf{0}$ and

$$\gamma = \sum_{k=0}^{n_k - 1} w(k)\mathbf{1}^\top \Big[|\bar{C}(k)G_x| \quad \textstyle\sum_{j=0}^{k} |\bar{D}_j(k)G_u|\Big].$$

Proof. Applying the interval norm from Def. 5 on the reachable set $\mathcal{Y}(k)$ from (2) and using the assumption $\alpha \geq \mathbf{0}$ leads to

$$\begin{aligned}\|\mathcal{Y}(k)\|_I &\overset{(9)}{=} \mathbf{1}^\top \left|\mathrm{gen}\big(\mathcal{Y}_a(k)\big)\right| \mathbf{1} \\ &\overset{(10b)}{=} \mathbf{1}^\top \left| \begin{bmatrix}\bar{C}(k)G_x & \bar{D}_0(k)G_u & \cdots & \bar{D}_k(k)G_u\end{bmatrix} \right| \\ &\qquad \mathrm{diag}\Big(\begin{bmatrix}\alpha_x^\top & \alpha_u^\top & \cdots & \alpha_u^\top\end{bmatrix}^\top\Big)\mathbf{1} \\ &= \mathbf{1}^T \left[|\bar{C}(k)G_x| \quad \sum_{j=0}^{k} |\bar{D}_j(k)G_u|\right] \begin{bmatrix}\alpha_x \\ \alpha_u\end{bmatrix}.\end{aligned}$$

Multiplying with the weight $w(k)$ and summing over k results in (11). □

B. Halfspace Constraints

We can use the halfspace representation of the reachable set $\mathcal{Y}(k)$ to formulate the containment constraints (8b) as linear inequalities [10]. For LTV systems, we obtain the following identification approach:

Theorem 1 (Identification using Halfspace Constraints): *The scaling factors α^* and the center vectors c^* solving Prob. 1 for S_{LTV} can be computed by the following linear program:*

$$p^* = \arg\min_p \begin{bmatrix}\gamma & \mathbf{0}\end{bmatrix} p \tag{12a}$$

$$\textit{s.t. } \forall k\colon\ \max_m\big(N(k)y_a^{(m)}(k)\big)\big) \leq \begin{bmatrix}P_\alpha(k) & P_c(k)\end{bmatrix} p, \tag{12b}$$

$$0 \leq \begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix} p, \tag{12c}$$

where γ can be computed with Lem. 1, $p = [\alpha^\top \ c^\top]^\top$, $y_a^{(m)}(k) = y^{(m)}(k) - y_*^{(m)}(k)$, and

$$P_\alpha(k) = \left[|N(k)\bar{C}(k)G_x| \quad \textstyle\sum_{j=0}^{k} |N(k)\bar{D}_j(k)G_u|\right],$$

$$P_c(k) = \left[N(k)\bar{C}(k) \quad N(k)\textstyle\sum_{j=0}^{k} \bar{D}_j(k)\right].$$

The rows of $N(k)$ are the normal vectors of the halfspace representation of $\mathcal{Y}_a(k)$, which can be obtained using [13, Thm. 7].

Proof. The containment constraint (8b) can be equivalently formulated as

$$\begin{aligned} y^{(m)}(k) \in \mathcal{Y}^{(m)}(k) &\overset{\text{Prop. 1}}{\Leftrightarrow} y_a^{(m)}(k) \in \mathcal{Y}_a(k) \\ &\overset{\text{Def. 4}}{\Leftrightarrow} N(k)y_a^{(m)}(k) \leq d(k).\end{aligned}$$

The halfspace normal vectors $N(k)$ of $\mathcal{Y}_a(k)$ can be computed with [13, Thm. 7] and, as shown in [9, Lemma 3], are independent of the scaling factor α. The vector $d(k)$ can be computed as

$$\begin{aligned} d(k) &\overset{(5)}{=} N(k)\mathrm{cen}\big(\mathcal{Y}_a(k)\big) + |N(k)\mathrm{gen}\big(\mathcal{Y}_a(k)\big)|\mathbf{1} \\ &\overset{(10)}{=} N(k)\Big(\bar{C}(k)c_x + \sum_{j=0}^{k} \bar{D}_j(k)c_u\Big) \\ &\quad + \Big|N(k)\begin{bmatrix}\bar{C}(k)G_x & \bar{D}_0(k)G_u & \cdots & \bar{D}_k(k)G_u\end{bmatrix}\Big| \\ &\qquad \mathrm{diag}\Big(\begin{bmatrix}\alpha_x^\top & \alpha_u^\top & \cdots & \alpha_u^\top\end{bmatrix}^\top\Big)\mathbf{1} \\ &= P_c(k)\begin{bmatrix}c_x^\top & c_u^\top\end{bmatrix}^\top + P_\alpha(k)\begin{bmatrix}\alpha_x^\top & \alpha_u^\top\end{bmatrix}^\top,\end{aligned}$$

where we used the constraint $\alpha \geq \mathbf{0}$ in the first step, which is enforced by (12c). As $d(k)$ does not depend on the test case m, we can reduce the computational complexity by considering only the measurement closest to the halfspaces $N(k)$, i.e., we enforce $\max_m\big(N(k)y_a^{(m)}(k)\big) \leq d(k)$, which results in (12b). □

C. Generator Constraints

As the number of halfspaces increases exponentially with the measurement dimension, we propose a novel approach that enforces the containment constraints via the generator representation of the reachable sets:

Theorem 2 (Identification using Generator Constraints): *The scaling factors α^* and the center vectors c^* solving Prob. 1 for S_{LTV} can be computed by the following linear program:*

$$p^* = \arg\min_p \begin{bmatrix}\gamma & \mathbf{0} & \mathbf{0}\end{bmatrix} p \tag{13a}$$

$$\textit{s.t. } \tilde{y}_a = \begin{bmatrix}\mathbf{0} & Q_c & Q_\beta\end{bmatrix} p, \tag{13b}$$

$$0 \leq \begin{bmatrix}R_\alpha & \mathbf{0} & \mathbf{I} \\ R_\alpha & \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} & \mathbf{0}\end{bmatrix} p, \tag{13c}$$

where γ can be computed with Lem. 1, $p = [\alpha^\top \ c^\top \ \beta^\top]^\top$, and β is a vector of additional optimization variables to enforce the containment constraint. The vector of stacked measurement deviations $\tilde{y}_a$ can be obtained with

$$\tilde{y}_a = \begin{bmatrix}\tilde{y}_a^{(1)} \\ \vdots \\ \tilde{y}_a^{(n_m)}\end{bmatrix} \textit{ with } \quad \tilde{y}_a^{(m)} = \begin{bmatrix} y_a^{(m)}(0) \\ \vdots \\ y_a^{(m)}(n_k - 1)\end{bmatrix}. \tag{14}$$

The constraint matrices are

$$Q_c = \mathrm{vert}_{n_m}\left(\begin{bmatrix}\bar{C}(0) & \sum_{j=0}^{0}\bar{D}_j(0) \\ \vdots & \vdots \\ \bar{C}(n_k-1) & \sum_{j=0}^{n_k-1}\bar{D}_j(n_k-1)\end{bmatrix}\right),$$

$$Q_\beta = \mathrm{diag}_{n_m}\left(\mathrm{gen}'\big(\mathcal{Y}_a(0)\big), \cdots, \mathrm{gen}'\big(\mathcal{Y}_a(n_k-1)\big)\right),$$

$$R_\alpha = \mathrm{vert}_{n_m}\left(\begin{bmatrix}R_{\alpha,0} \\ \vdots \\ R_{\alpha,n_k-1}\end{bmatrix}\right), \tag{15a}$$

$$\textit{with } R_{\alpha,k} = \begin{bmatrix}\mathbf{I}_{\eta_x} & \mathbf{0}_{\eta_x \times \eta_u} \\ \mathbf{0}_{(k+1)\eta_u \times \eta_x} & \mathrm{vert}_{k+1}(\mathbf{I}_{\eta_u})\end{bmatrix}. \tag{15b}$$

Proof. As in Thm. 1, we use the cost function from Lem. 1 but formulate the containment constraints with the generator representation as

$$\begin{aligned} & y^{(m)}(k) \in \mathcal{Y}^{(m)}(k) \\ &\overset{\text{Prop. 1}}{\Leftrightarrow} y_a^{(m)}(k) \in \mathcal{Y}_a(k) \\ &\overset{(6)}{\Leftrightarrow} \exists \lambda_k^{(m)} \in \mathbb{R}^{\eta_x + (k+1)\eta_u}\colon \\ &\qquad |\lambda_k^{(m)}| \leq \mathbf{1}, \\ &\qquad y_a^{(m)}(k) = \mathrm{cen}\big(\mathcal{Y}_a(k)\big) + \mathrm{gen}\big(\mathcal{Y}_a(k)\big)\lambda_k^{(m)}\end{aligned}$$

TABLE I

NUMBER OF OPTIMIZATION VARIABLES (OV), EQUALITY CONSTRAINTS (EC), AND INEQUALITY CONSTRAINTS (IC) IN THM. 1 AND IN THM. 2.

	# OV	# EC	# IC
Thm. 1	$\mathcal{O}(n_y)$	-	$\mathcal{O}(n_k^{n_y})$
Thm. 2	$\mathcal{O}(n_k^2 n_m n_y)$	$\mathcal{O}(n_k n_m n_y)$	$\mathcal{O}(n_k^2 n_m n_y)$

$$\overset{(10)}{\Leftrightarrow} \exists \beta_k^{(m)} \in \mathbb{R}^{\eta_x+(k+1)\eta_u}:$$
$$|\beta_k^{(m)}| \leq \begin{bmatrix}\alpha_x^\top & \alpha_u^\top & \cdots & \alpha_u^\top\end{bmatrix}^\top, \quad (16a)$$
$$y_a^{(m)}(k) = \begin{bmatrix}\bar{C}(k) & \sum_{j=0}^{k} \bar{D}_j(k) & \mathrm{gen}'(\mathcal{Y}_a(k))\end{bmatrix} \begin{bmatrix}c_x^\top & c_u^\top & \beta_k^{(m)\top}\end{bmatrix}^\top, \quad (16b)$$

where we define $\beta_k^{(m)} := \mathrm{diag}([\alpha_x^\top \;\; \alpha_u^\top \;\; \cdots \;\; \alpha_u^\top]^\top)\lambda_k^{(m)}$ in the last step. By stacking the vectors $\beta_k^{(m)}$ for all measurements and time steps, i.e.,

$$\beta = \begin{bmatrix}\beta^{(1)} \\ \vdots \\ \beta^{(n_m)}\end{bmatrix} \text{ with } \beta^{(m)} = \begin{bmatrix}\beta_0^{(m)} \\ \vdots \\ \beta_{n_k-1}^{(m)}\end{bmatrix},$$

we obtain the equality constraint (13b) from (16b) and the inequality constraint (13c) from (16a) with $\alpha \geq \mathbf{0}$. □

IV. COMPUTATIONAL COMPLEXITY

Next, we analyze the computational complexity of the proposed linear programs, assuming a linear dependence of η_x, η_u, n_x, and n_u on the output dimension n_y.

Proposition 2 (Computational Complexity): *The number of optimization variables, equality constraints, and inequality constraints in Thm. 1 and in Thm. 2 are as displayed in Tab. I.*

Proof. In Thm. 1, we have the optimization variables $p = [\alpha^\top \;\; c^\top]^\top \in \mathbb{R}^{\eta_x+\eta_u+n_x+n_u}$. As we assume that η_x, η_u, n_x, and n_u depend linearly on n_y, the number of optimization variables is $\mathcal{O}(n_y)$. We do not have equality constraints, but the number of inequality constraints in (12b) equals the number of halfspaces summed over all time steps. At time step k, the number of halfspaces is $2\binom{\eta_{y(k)}}{n_y-1}$ [13], where $\eta_{y(k)} = \eta_x + (k+1)\eta_u$ is the number of generators of $\mathcal{Y}(k)$. Using Stirling's formula $(n_y-1)! \geq ((n_y-1)/e)^{n_y-1}$ [15], the binomial coefficient can be upper-bounded by $\binom{\eta_{y(k)}}{n_y-1} \leq \left(\eta_{y(k)}e/(n_y-1)\right)^{n_y-1}$. As $\eta_{y(k)}$ is proportional to $(k+1)n_y$, we obtain $\mathcal{O}((k+1)^{n_y-1})$ for the number of halfspaces at time step k. Using the Faulhabersche formula [16, p. 106], the number of inequality constraints for $k = 0, ..., n_k-1$ is $\mathcal{O}(\sum_{k=0}^{n_k-1}(k+1)^{n_y-1}) = \mathcal{O}(n_k^{n_y})$.

Thm. 2 optimizes $p = [\alpha^\top \;\; c^\top \;\; \beta^\top]^\top$, where β has $n_m \sum_{k=0}^{n_k-1}(\eta_x + (k+1)\eta_u) = n_m n_k\left(\eta_x + \eta_u(n_k+1)/2\right)$ elements [16, p. 34]. Thus, the number of optimization variables is $\mathcal{O}(n_k^2 n_m n_y)$. As Q_c has $n_k n_m n_y$ rows, the number of equality constraints is $\mathcal{O}(n_k n_m n_y)$. Furthermore, we have two inequality constraints for each element of β and one additional inequality for each element of α. Thus, the number of inequality constraints is $\mathcal{O}(n_k^2 n_m n_y)$. □

The number of optimization variables and constraints of the approach using halfspace constraints (Thm. 1) is independent of the number of test cases n_m and, hence, leads to low computational complexity for a large number of test cases. However, the linear program using generator constraints, as in Thm. 2, is more efficient for large measurement dimensions n_y or large time horizons n_k.

V. NUMERICAL EXPERIMENTS

In this section, we evaluate the performance of the proposed approaches for the reachset-conformant identification of LTV systems. All computations are carried out in MATLAB on an i9-12900HK processor (2.5GHz) with 64GB memory. The code to reproduce the results will be integrated into the next release of the toolbox CORA [17]. Since the identification results of Thm. 1 and Thm. 2 are identical, we focus on the scalability of both approaches.

In our experiments, we use the vehicle platooning system from [18] with uncertain inputs. Assuming a constant velocity of the leader vehicle, the dynamics of the follower vehicle i, $i = 1, ..., n_v$, can be described by

$$\dot{x}_{3i-2} = -\xi_i(t)x_{3i-2} + \xi_i(t)u_i$$
$$\dot{x}_{3i-1} = x_{3i}$$
$$\dot{x}_{3i} = \begin{cases} -x_{3i-2} & \text{if } i = 1, \\ x_{3i-5} - x_{3i-2} & \text{else,} \end{cases}$$

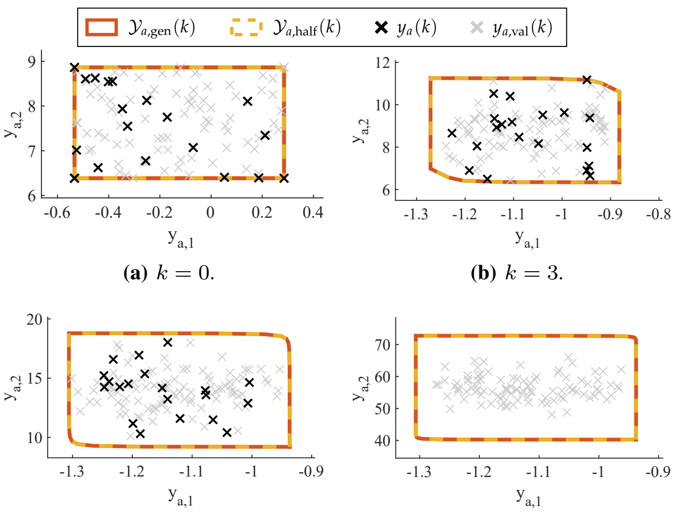
where n_v is the number of follower vehicles. The state x_{3i-2} represents the acceleration of vehicle i, x_{3i-1} is the distance between vehicle i and vehicle $i-1$, and x_{3i} is the first derivative of this distance. The input u_i describes the control signal for vehicle i. The parameters $\xi_i(t)$ represent the drivetrain dynamics and are assumed to be time-variant. We assume we can measure all states, i.e., $n_y = 3n_v$. The dynamics are discretized with the forward Euler method. The true uncertainty sets, used for creating the test cases via the simulation of the system dynamics, consist of a random center vector and a random diagonal generator matrix. For our identification, we assume that the generator templates G_x and G_u are identity matrices, as this assumption usually leads to good results.

The computation times of Thm. 1 and Thm. 2 for varying time horizons n_k, a varying number of test cases n_m, and a varying measurement dimension n_y, via changing the number of vehicles n_v, are shown in Tab. II. Increasing the number of test cases n_m influences the computational complexity of Thm. 1 only slightly since we consider just the sample $y_a^{(m)}$ closest to each halfspace. However, as the halfspace conversion scales exponentially with the dimension of the reachable set, Thm. 1 can only be used for systems with small measurement dimension n_y and small time horizons n_k. In contrast, Thm. 2 using the generator constraints is efficient for high-dimensional systems and long time horizons. Although its computational complexity depends substantially

TABLE II

COMPUTATION TIMES FOR IDENTIFYING A REACHSET-CONFORMANT MODEL OF A VEHICLE PLATOON CONSISTING OF $n_y/3$ VEHICLES, USING n_m TEST CASES WITH THE TIME HORIZON n_k.

n_k	n_m	n_y	**Computation time [s]** Thm. 1	Thm. 2
6	20	3	**0.02**	0.03
6	20	6	0.11	**0.06**
6	20	9	29.15	**0.09**
6	20	90	>120.00	**6.40**
6	10	6	0.12	**0.04**
6	100	6	**0.12**	0.48
6	1000	6	**0.21**	47.19
6	10000	6	**1.02**	>120.00
4	20	6	0.04	**0.03**
8	20	6	0.46	**0.09**
12	20	6	13.48	**0.22**
48	20	6	>120.00	**13.20**



(a) $k = 0$.
(b) $k = 3$.
(c) $k = 7$.
(d) $k = 19$.

Fig. 2. Identification results of the setup $(n_k = 8, n_m = 20, n_y = 6)$ for different time steps k: Measurement deviation sets $\mathcal{Y}_{a,\text{half}}(k)$ and $\mathcal{Y}_{a,\text{gen}}(k)$ that were predicted with the identified uncertainty sets from Thm. 1 and Thm. 2, respectively, measurement deviations $y_a(k)$, $k = 0, ..., 7$, of the 20 test cases that were used in the identification, and measurement deviations $y_{a,\text{val}}(k)$, $k = 0, ..., 19$, of 100 test cases that were not used in the identification.

on the number of test cases n_m, this is not a significant drawback: The containment constraints are usually just active for a small number of test cases (particularly relevant test cases), and removing all the other test cases would not change the identification results. Thus, by applying test case selection strategies, which automatically find the particularly relevant test cases, before starting the uncertainty identification, we generally end up with a small number of test cases n_m.

Furthermore, we visualize the results for the first two output dimensions of the setup $(n_k = 8, n_m = 20, n_y = 6)$ in Fig. 2. The figure shows that the identified uncertainty sets lead to tight reachable sets, which also contain unseen test cases, even for time steps $k \geq n_k$.

VI. CONCLUSION

In this work, we present two novel reachset-conformant identification approaches for linear time-variant systems. One approach is based on the halfspace representation of zonotopes and scales well with regard to the number of measurements. The second approach uses the generator representation and, thus, is more efficient for high-dimensional systems and long time horizons. As demonstrated by numerical experiments, both approaches are able to estimate conformant disturbance sets that lead to tight reachable sets. By ensuring that the identified model accurately captures the behavior of the real-world system, safety properties can be transferred from the model domain to the actual system. Therefore, this work paves the way for the safe application of formal methods to real-world systems.

REFERENCES

[1] H. Roehm, J. Oehlerking, M. Woehrle, and M. Althoff, "Model conformance for cyber-physical systems: A survey," *ACM Transactions on Cyber-Physical Systems*, vol. 3, no. 3, p. Article 30, 2019.

[2] M. Casini, A. Garulli, and A. Vicino, "Feasible parameter set approximation for linear models with bounded uncertain regressors," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 2910–2920, 2014.

[3] M. Kieffer and E. Walter, "Guaranteed estimation of the parameters of nonlinear continuous-time models: Contributions of interval analysis," *International Journal of Adaptive Control and Signal Processing*, vol. 25, no. 3, pp. 191–207, 2011.

[4] A. Rauh, J. Kersten, and H. Aschemann, "Interval methods and contractor-based branch-and-bound procedures for verified parameter identification of quasi-linear cooperative system models," *Journal of Computational and Applied Mathematics*, vol. 367, pp. 1–10, 2019.

[5] B. T. Polyak, S. A. Nazin, C. Durieu, and E. Walter, "Ellipsoidal parameter or state estimation under model uncertainty," *Automatica*, vol. 40, no. 7, p. 1171–1179, 2004.

[6] A. Alanwar, A. Koch, F. Allgöwer, and K. H. Johansson, "Data-driven reachability analysis from noisy data," *IEEE Transactions on Automatic Control*, pp. 1–16, 2023.

[7] N. Kochdumper, A. Tarraf, M. Rechmal, M. Olbrich, L. Hedrich, and M. Althoff, "Establishing reachset conformance for the formal analysis of analog circuits," in *Proc. of the 25th Asia and South Pacific Design Automation Conference*, pp. 199–204, 2020.

[8] F. Gruber and M. Althoff, "Scalable robust safety filter with unknown disturbance set," *IEEE Transactions on Automatic Control*, pp. 1–15, 2023.

[9] S. B. Liu and M. Althoff, "Online verification of impact-force-limiting control for physical human-robot interaction," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 777–783, 2021.

[10] S. B. Liu, B. Schürmann, and M. Althoff, "Guarantees for real robotic systems: Unifying formal controller synthesis and reachset-conformant identification," *IEEE Transactions on Robotics*, pp. 1–15, 2023.

[11] M. Althoff, "Checking and establishing reachset conformance in CORA 2023," in *Proc. of the Workshop on Applied Verification of Continuous and Hybrid Systems*, pp. 9–33, 2023.

[12] W. Kühn, "Rigorously computed orbits of dynamical systems without the wrapping effect," *Computing*, vol. 61, no. 1, pp. 47–67, 1998.

[13] M. Althoff, O. Stursberg, and M. Buss, "Computing reachable sets of hybrid systems using a combination of zonotopes and polytopes," *Nonlinear Analysis: Hybrid Systems*, vol. 4, no. 2, p. 233–249, 2010.

[14] A. Girard, C. Le Guernic, and O. Maler, "Efficient computation of reachable sets of linear time-invariant systems with inputs," in *Proc. of the 9th International Conference on Hybrid Systems: Computation and Control*, pp. 257–271, 2006.

[15] G. Marsaglia and J. C. W. Marsaglia, "A new derivation of Stirling's approximation to n!," *The American Mathematical Monthly*, vol. 97, no. 9, pp. 826–829, 1990.

[16] J. H. Conway and R. K. Guy, *The Book of Numbers*. Springer, 1996.

[17] M. Althoff, "An introduction to CORA 2015," in *Proc. of the Workshop on Applied Verification for Continuous and Hybrid Systems*, pp. 120–151, 2015.

[18] J. P. Maschuw and D. Abel, "Longitudinal vehicle guidance in networks with changing communication topology," *Proc. of the IFAC Symposium on Advances in Automotive Control*, vol. 43, no. 7, pp. 785–790, 2010.