2022 American Control Conference (ACC)
Atlanta, USA, June 8-10, 2022

A Multi-Parametric Method for Active Model Discrimination of Nonlinear Systems with Temporal Logic-Constrained Switching

Ruochen Niu, Syed M. Hassaan and Sze Zheng Yong

Abstract— **In this paper, we consider the optimal input design problem for active model discrimination (AMD) among a set of switched nonlinear models that are constrained by metric/signal temporal logic specifications and affected by uncontrolled inputs and noise. To deal with nonlinear and non-convex constraints in the resulting bilevel optimization problem, we first over-approximate the nonlinear dynamics using piecewise affine abstractions. Then, we solve the relaxed inner problem of the bilevel AMD problem as parametric optimization problems and substitute the parametric solutions into the outer problem to obtain sufficient separating inputs for AMD. Moreover, since the parametric optimization problems are often computationally demanding, we propose several strategies to reduce the computational time, while preserving feasibility of the separating inputs for AMD. Finally, we demonstrate the effectiveness of our approach on several illustrative examples on fault detection and lane changing scenario.**

I. Introduction

Model estimation and fault detection are crucial for successful operation of cyber-physical systems (CPS) such as distributed robots, autonomous vehicles and medical devices, since CPS often interact with other systems whose operating modes are unknown or partially observed. Further, CPS often involve logical/discrete elements, e.g., temporal logic task or safety specifications. Thus, to guarantee and improve CPS safety, it is of great interest to develop approaches for model discrimination with temporal logic constraints.

Literature Review. Model discrimination, with applications in fault detection and intent estimation, can generally be grouped into passive and active approaches. The passive method seeks to identify the true model among a set of healthy/normal and faulty models given a sequence of input-output data [1], [2], while the active approach, also known as AMD, designs a minimally perturbing input sequence for the system such that the (output) behaviors of different models are guaranteed to be distinct [3]–[5].

This paper mainly focuses on AMD approaches, which have been mainly considered using polyhedral projection [4], ellipsoids [6] and mixed-integer optimization [5], [7]. Moreover, closed-loop active model discrimination using set-valued observers has been studied in [8], [9], where the latter also employed multi-parametric programming. On the other hand, partition-based AMD approaches have been proposed to utilize run-time revealed information [10] and measurement [11] to further minimize the separating (perturbation) input by building a partition tree. However, the above-mentioned papers are only applicable for linear or affine models, while [12] proposed an AMD method for nonlinear or uncertain affine models by constructing (single-region/non-switched) affine inclusion/abstraction models.

On the other hand, for systems subjected to temporal logic specifications, including for fault and anomaly detection problems [13], [14] showed that temporal logic specifications can be formulated using mixed-integer constraints via the construction of a nondeterministic finite state machine called a monitor, while [15] proposed an optimization-based method by directly converting signal/metric temporal logic specifications into mixed-integer constraints.

Contribution. In this paper, we propose an optimization-based approach for active model discrimination among switched nonlinear systems based on *piecewise affine abstraction* [16], [17], as an extension of [12], where only a single-domain/piece affine abstraction/over-approximation is considered that may result in overly conservative solutions. Moreover, we consider switched nonlinear models whose switching/mode sequence is constrained by metric/signal temporal logic (MTL/STL) specifications, as a counterpart to the passive model discrimination problem in [15].

In particular, to design a separating input sequence such that the trajectories of all models are distinct from each other in spite of noise and uncontrolled inputs, we formulate the AMD problem as a bilevel optimization problem with multiple inner problems that involve nonlinear, integral and non-convex constraints, which is generally computationally intractable. Thus, we leverage *piecewise affine abstractions* [16], [17] as affine over-approximations of the nonlinear models and solve the relaxed inner problems with mixed-integer linear constraints as parametric optimization problems. Then, we utilize the parametric inner problem solutions to solve the AMD problem using (tractable) linear programs.

Furthermore, since the parametric problems are also often computationally expensive or intractable for large problems, we propose several strategies based on horizon truncation and input domain partitioning to render them applicable to larger problems. Finally, we demonstrate the effectiveness of the approach on several illustrative examples on fault detection and intent estimation in a lane changing scenario.

II. Preliminaries

A. Notations

For vectors $v \in \mathbb{R}^n$ and a matrix $M \in \mathbb{R}^{p\times q}$, $\|v\| \triangleq \max_i |v_i|$ denotes its ∞-norm, $M^\top$ its transpose and $M \geq 0$

The authors are with School for Engineering of Matter, Transport and Energy, Arizona State Univ., Tempe, AZ 85287 (email: {rniu6,shassaan,szyong}@asu.edu). The work by S.M. Hassaan and S.Z. Yong was supported in part by DARPA grant D18AP00073 and NSF grant CNS-1943545.

element-wise non-negativity. The set of positive integers up to n is denoted by $\mathbb{Z}_n^+$, and the set of integers from l to n is denoted by $\mathbb{Z}_l^n$. The vec operator is defined for a collection of vectors v_k, $k \in \mathbb{Z}_l^n$ as $\text{vec}_{k=l}^n\{v_k\} = \begin{bmatrix} v_l^T & \dots & v_n^T \end{bmatrix}^T$.

We will also make use of Special Ordered Set of degree 1 (SOS-1) constraint[1] and set partitions, defined as follows:

Definition 1 (SOS-1 Constraint [18]). *A special ordered set of degree 1 (SOS-1) constraint is a set of integer, continuous or mixed-integer scalar variables for which at most one variable in the set may take a value other than zero, denoted as SOS-1: $\{v_1, \dots, v_N\}$. For instance, if $v_i \neq 0$, then this constraint imposes that $v_j = 0$ for all $j \neq i$.*

Definition 2 (Partition for Polyhedral Sets). *A partition of a polyhedral set $\mathcal{P}$ is a collection of $|S|$ disjoint subsets $\mathcal{P}_i$ such that $\bigcup_{i=1}^{|S|} \mathcal{P}_i = \mathcal{P}$, where each partition $\mathcal{P}_i$ is also a polyhedral set.*

B. Metric/Signal Temporal Logic (MTL/STL)

Definition 3 (Atomic Proposition). *An atomic proposition is an assertion whether a system variable or signal is either True (1 or $\top$) or False (0 or $\bot$).*

For a finite set of modes, denoted by Σ, the syntax of MTL/STL formulas over it is given by:

$$\varphi ::= p \mid \neg\varphi \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 \mathcal{U}_{[k_1,k_2]} \varphi_2, \tag{1}$$

where $p \in \Sigma$. The operations $\neg$, $\vee$, and $\mathcal{U}_{[k_1,k_2]}$ are the negation, disjunction, and time-constrained until operators, respectively, whereas $[k_1, k_2] \subset [0, \infty)$ is an integer interval. Using the grammar in (1), we can define next ($\bigcirc$) as $\bigcirc\varphi = \top\mathcal{U}_{[0,1]}\varphi$, conjunction ($\wedge$) as $\varphi_1 \wedge \varphi_2 = \neg(\neg\varphi_1 \vee \neg\varphi_2)$, implication ($\Rightarrow$) as $\varphi_1 \Rightarrow \varphi_2 = \neg\varphi_1 \vee \varphi_2$, eventually in $[k_1, k_2]$ ($\Diamond_{[k_1,k_2]}$) as $\Diamond_{[k_1,k_2]}\varphi = \top\mathcal{U}_{[k_1,k_2]}\varphi = \bigvee_{\tau=k_1}^{k_2} \bigcirc^\tau \varphi$, and always in $[k_1, k_2]$ ($\Box_{[k_1,k_2]}$) as $\Box_{[k_1,k_2]}\varphi = \neg\Diamond_{[k_1,k_2]}\neg\varphi = \bigwedge_{\tau=k_1}^{k_2} \bigcirc^\tau \varphi$. For brevity, we will denote $\mathcal{U}_{[0,\infty)}$, $\Diamond_{[0,\infty)}$, $\Box_{[0,\infty)}$ as $\mathcal{U}$, $\Diamond$, $\Box$ throughout the paper.

Definition 4 (MTL/STL Semantics). *Let $\boldsymbol{\sigma}$ represent an ω-length word over Σ, i.e., $\boldsymbol{\sigma} \in \Sigma^\omega$, and let σ_k be k^{th} element of $\boldsymbol{\sigma}$. Then MTL/STL semantics are defined as follows:*

1) $(\boldsymbol{\sigma}, k) \models p \Leftrightarrow \sigma_k = p$,
2) $(\boldsymbol{\sigma}, k) \models \neg\varphi \Leftrightarrow (\boldsymbol{\sigma}, k) \not\models \varphi$,
3) $(\boldsymbol{\sigma}, k) \models \varphi_1 \vee \varphi_2 \Leftrightarrow (\boldsymbol{\sigma}, k) \models \varphi_1$ *or* $(\boldsymbol{\sigma}, k) \models \varphi_2$,
4) $(\boldsymbol{\sigma}, k) \models \varphi_1 \wedge \varphi_2 \Leftrightarrow (\boldsymbol{\sigma}, k) \models \varphi_1$ *and* $(\boldsymbol{\sigma}, k) \models \varphi_2$,
5) $(\boldsymbol{\sigma}, k) \models \varphi_1 \mathcal{U}_{[k_1,k_2]} \varphi_2 \Leftrightarrow \exists k' \in [k + k_1, k + k_2]$: $(\boldsymbol{\sigma}, k') \models \varphi_2$ *and* $\forall k'' \in [k, k'] : (\boldsymbol{\sigma}, k'') \models \varphi_1$,
6) $(\boldsymbol{\sigma}, k) \models \Diamond_{[k_1,k_2]}\varphi \Leftrightarrow \exists k' \in [k+k_1, k+k_2], (\boldsymbol{\sigma}, k') \models \varphi$,
7) $(\boldsymbol{\sigma}, k) \models \Box_{[k_1,k_2]}\varphi \Leftrightarrow \forall k' \in [k+k_1, k+k_2], (\boldsymbol{\sigma}, k') \models \varphi$.

Moreover, $\boldsymbol{\sigma} \models \varphi$ implies $(\boldsymbol{\sigma}, 0) \models \varphi$.

Additionally, the following useful definitions from [15] for (infinite-length) MTL/STL specifications are introduced.

Definition 5 (Valid Subtrace of an MTL/STL). *A length-T word $\boldsymbol{q} \in \Sigma^T$ is called a* valid subtrace *from k_0 of an MTL/STL formula φ, if there exist a k_0-length prefix $\boldsymbol{p} \in \Sigma^{k_0}$ and a suffix $\boldsymbol{r} \in \Sigma^\omega$ such that their ω-word concatenation $\boldsymbol{pqr}$ satisfies φ, i.e., $\boldsymbol{pqr} \models \varphi$. Further, the set of length-T valid subtraces from k_0 is denoted as $V_{k_0}^T(\varphi)$.*

Definition 6 (MTL/STL Bound of an MTL/STL). *The bound of an MTL/STL formula φ, denoted by b^φ, is the time length required to evaluate the satisfaction of φ and is recursively computed as follows: 1) $b^{\neg\varphi} = b^\varphi$; 2) $b^{\varphi_1 \wedge \varphi_2} = \max(b^{\varphi_1}, b^{\varphi_2})$; 3) $b^{\Diamond_{[t_1,t_2]}\varphi} = b^{\Box_{[t_1,t_2]}\varphi} = t_2 + b^\varphi$; and 4) $b^{\varphi_1 \mathcal{U}_{[t_1,t_2]} \varphi_2} = t_2 + \max(b^{\varphi_1}, b^{\varphi_2})$.*

To solve active model discrimination problems involving MTL/STL specifications, equivalent integer encodings of the specifications have been introduced in [15], which are provided in the Appendix for the sake of completeness.

C. Modeling Framework

Consider N discrete-time switched nonlinear system models $\{\mathcal{G}_l\}_{l=1}^N$, each with states $\boldsymbol{x}_l \in \mathbb{R}^n$, measurements/outputs $z_l \in \mathbb{R}^{n_z}$, inputs $\boldsymbol{u_l} \in \mathbb{R}^m$, process noise $w_l \in \mathbb{R}^{m_w}$, measurement noise $v_l \in \mathbb{R}^{m_v}$, σ_l is the (controlled or uncontrollable) discrete switching signal/mode, from a finite set Σ_l with cardinality $|\Sigma_l| = K$. The models are given by:

$$\boldsymbol{x}_l(k+1) = h_l^{\sigma_{l,k}}(\boldsymbol{x}_l(k), \boldsymbol{u}_l(k), w_l(k)), \tag{2}$$
$$z_l(k) = g_l^{\sigma_{l,k}}(\boldsymbol{x}_l(k), \boldsymbol{u}_l(k), v_l(k)), \tag{3}$$

where, $h_l^{\sigma_{l,k}}$ and $g_l^{\sigma_{l,k}}$ are continuous functions/mappings and $\boldsymbol{x}_l(k+1)$ denotes the state at the next time instant.

The states $\boldsymbol{x}_l$ can be divided into controlled states $x_l \in \mathbb{R}^{n_x}$ and uncontrolled states $y_l \in \mathbb{R}^{n_y}$ with $n_y = n - n_x$ accordingly. Similarly, the first m_u components of $\boldsymbol{u_l}$ are controlled inputs denoted as $u \in \mathbb{R}^{m_u}$, while the other $m_d = m - m_u$ components of $\boldsymbol{u}_l$, denoted as $d_l \in \mathbb{R}^{m_d}$, are uncontrolled inputs that are model-dependent. As a consequence, we have

$$\boldsymbol{x}_l(k) \triangleq \begin{bmatrix} x_l(k) \\ y_l(k) \end{bmatrix}, \boldsymbol{u}_l(k) \triangleq \begin{bmatrix} u(k) \\ d_l(k) \end{bmatrix}. \tag{4}$$

The initial condition (or the initial state of a moving horizon starting at time k_0) for model l, denoted by $\boldsymbol{x}_l^{k_0} = \boldsymbol{x}_l(k_0)$, is constrained to a polyhedral set with c_0 inequalities:

$$\boldsymbol{x}_l^{k_0} \in \mathcal{X}_{k_0} = \{\boldsymbol{x} \in \mathbb{R}^n : P_{k_0}\boldsymbol{x} \leq p_{k_0}\},\ \forall l \in \mathbb{Z}_N^+. \tag{5}$$

Moreover, the states x_l and y_l satisfy polyhedral state constraints with c_x and c_y inequalities:

$$x_l(k) \in \mathcal{X}_{x,l} = \{x \in \mathbb{R}^{n_x} : P_{x,l}x \leq p_{x,l}\}, \tag{6}$$
$$y_l(k) \in \mathcal{X}_{y,l} = \{y \in \mathbb{R}^{n_y} : P_{y,l}y \leq p_{y,l}\}, \tag{7}$$

On the other hand, the controlled and uncontrolled inputs u and d_l must also satisfy the following polyhedral constraints with c_u and c_d inequalities:

$$u(k) \in \mathcal{U} = \{u \in \mathbb{R}^{m_u} : Q_u u \leq q_u\}, \tag{8}$$
$$d_l(k) \in \mathcal{D}_l = \{d \in \mathbb{R}^{m_d} : Q_{d,l}d \leq q_{d,l}\}. \tag{9}$$

The process and measurement noises, w_l and v_l, are also polyhedrally constrained with c_w and c_y inequalities:

$$w_l(k) \in \mathcal{W}_l = \{w \in \mathbb{R}^{m_w} : Q_{w,l}w \leq q_{w,l}\}, \tag{10}$$
$$v_l(k) \in \mathcal{V}_l = \{v \in \mathbb{R}^{m_v} : Q_{v,l}v \leq q_{v,l}\}. \tag{11}$$

[1] Off-the-shelf solvers such as Gurobi and CPLEX [18], [19] can readily handle these constraints, which can significantly reduce the search space for integer variables in branch and bound algorithms.

Each given $\mathcal{G}_l$ also contains an MTL/STL formula φ_l that governs the set of allowed switching signals $\boldsymbol{\sigma_l} \triangleq \{\sigma_{l,k}\}_0^\infty \in \Sigma_l^\omega$, i.e., $\boldsymbol{\sigma_l} \models \varphi_l$ must hold. Moreover, we assume that formulas φ_l are of the (unbounded global/safety) form:

$$\varphi_l = \phi_{b,l} \wedge \Box \phi_{g,l}, \tag{12}$$

where $\phi_{b,l}$ and $\phi_{g,l}$ are bounded negation-free MTL/STL formulas with bounds $b^{\phi_{b,l}}$ and $b^{\phi_{g,l}}$ (cf. Definition 6). This assumption allows us to formulate the problem and the proposed algorithms using a finite number of variables.

Further, given a time horizon of length T, we define:

$$\begin{gathered}\boldsymbol{x}_{l,T} = \text{vec}_{k=k_0}^{k_0+T}\{\boldsymbol{x}_l(k)\}, d_{l,T} = \text{vec}_{k=k_0}^{k_0+T-1}\{d_l(k)\},\\ w_{l,T} = \text{vec}_{k=k_0}^{k_0+T-1}\{w_l(k)\}, v_{l,T} = \text{vec}_{k=k_0}^{k_0+T-1}\{v_l(k)\},\\ u_T = \text{vec}_{k=k_0}^{k_0+T-1}\{u(k)\}, z_{l,T} = \text{vec}_{k=k_0}^{k_0+T-1}\{z_l(k)\},\\ \sigma_{l,T} = \{\sigma_{l,k}\}_{k=\underline{k}}^{\overline{k}},\end{gathered}$$

for model $l \in \mathbb{Z}_N^+$, as well as the subtrace set $V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi_l)$, where $\underline{k} = k_0 - \max_l\{b^{\phi_{g,l}}\}$ and $\overline{k} = k_0 + T - 1 + \max_l\{b^{\phi_{g,l}}\}$ if $k_0 > \max_l\{b^{\phi_{b,l}}\}$, otherwise, with $\underline{k} = k_0$ and $\overline{k} = \max\{k_0 + T - 1 + \max_l\{b^{\phi_{g,l}}\}, \max_l\{b^{\phi_{b,l}}\}\}$.

1) Abstraction of Nonlinear Functions: To deal with nonlinearities in $\mathcal{G}_l$ in active model discrimination problems, i.e., the functions $h_l^{\sigma_l}$ and $g_l^{\sigma_l}$, an affine abstraction/over-approximation method is considered in [12], however, it can only considers a single domain. Here, we propose to leverage the piecewise affine abstraction approaches in [16], [17] to over-approximate the nonlinearities with piecewise affine inclusions, where the precision of the abstraction can be improved with more and better chosen *partitions*, as defined below, although it may lead to longer computation times and even intractability due to more integer variables in our solutions in Section IV (cf. [17] for details).

Definition 7 (Partition for $h_l^{\sigma_l}$). *For each function $h_l^{\sigma_l}$, a partition $\mathcal{I}_l^{h,\sigma_l}$ of the closed bounded region $\mathcal{X} \times \mathcal{U} \times \mathcal{W} \subseteq \mathbb{R}^{n+m+m_w}$ is a collection of q_l^{h,σ_l} subregions $\mathcal{I}_l^{h,\sigma_l} = \{I_{l,i}^{h,\sigma_l} | i \in \mathbb{Z}_1^{q_l^{h,\sigma_l}}\}$ such that $\mathcal{X} \times \mathcal{U} \times \mathcal{W} \subseteq \bigcup_{i=1}^{q_l^{h_l,\sigma_l}} I_{l,i}^{h,\sigma_l}$ and $I_{l,i}^{h,\sigma_l} \cap I_{l,j}^{h,\sigma_l} = \partial I_{l,i}^{h,\sigma_l} \cap \partial I_{l,j}^{h,\sigma_l}$, $\forall i \neq j \in \mathbb{Z}_1^q$, where $\partial I_{l,i}^{h,\sigma_l}$ is the boundary of set $I_{l,i}^{h,\sigma_l}$. Similarly, a partition $\mathcal{I}_l^{g,\sigma_l}$ of the closed bounded region $\mathcal{X} \times \mathcal{U} \times \mathcal{V} \subseteq \mathbb{R}^{n+m+n_v}$ for each function $g_l^{\sigma_l}$ can be defined.*

We assume the partitions to be polytopic. Then, for each polytopic subregion $I_{l,i}^{h,\sigma_l} \in \mathcal{I}_l^{h,\sigma_l}$ (or $I_{l,j}^{g,\sigma_l} \in \mathcal{I}_l^{g,\sigma_l}$) that partitions the domain of interest, the nonlinear function $h_l^{\sigma_l}$ (or $g_l^{\sigma_l}$) can be over-approximated/abstracted by a pair of affine functions $\underline{h}_{l,i}^{\sigma_l}$, $\overline{h}_{l,i}^{\sigma_l}$ (or $\underline{g}_{l,j}^{\sigma_l}$, $\overline{g}_{l,j}^{\sigma_l}$) via the abstraction algorithms in [16], [17]. As a result, for all $(\boldsymbol{x}, \boldsymbol{u}, w) \in I_{l,i}^{h,\sigma_l}$ (or $(\boldsymbol{x}, \boldsymbol{u}, v) \in I_{l,j}^{g,\sigma_l}$), the function $h_{l,i}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, w)$ (or $g_{l,j}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, v)$) is sandwiched/framed by a pair of affine functions, i.e., $\underline{h}_{l,i}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, w) \leq h_{l,i}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, w) \leq \overline{h}_{l,i}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, w)$ (or $\underline{g}_{l,j}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, v) \leq g_{l,j}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, v) \leq \overline{g}_{l,j}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, v)$) with

$$\begin{aligned}\underline{h}_{l,i}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, w) &= \underline{A}_{l,i}^{\sigma_l}\boldsymbol{x} + \underline{B}_{l,i}^{\sigma_l}\boldsymbol{u} + \underline{W}_{l,i}^{\sigma_l}w + \underline{f}_{l,i}^{\sigma_l},\\ \overline{h}_{l,i}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, w) &= \overline{A}_{l,i}^{\sigma_l}\boldsymbol{x} + \overline{B}_{l,i}^{\sigma_l}\boldsymbol{u} + \overline{W}_{l,i}^{\sigma_l}w + \overline{f}_{l,i}^{\sigma_l},\\ \underline{g}_{l,j}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, v) &= \underline{C}_{l,j}^{\sigma_l}\boldsymbol{x} + \underline{D}_{l,j}^{\sigma_l}\boldsymbol{u} + \underline{V}_{l,j}^{\sigma_l}v + \underline{o}_{l,j}^{\sigma_l},\\ \overline{g}_{l,j}^{\sigma_l}(\boldsymbol{x}, \boldsymbol{u}, v) &= \overline{C}_{l,j}^{\sigma_l}\boldsymbol{x} + \overline{D}_{l,j}^{\sigma_l}\boldsymbol{u} + \overline{V}_{l,j}^{\sigma_l}v + \overline{o}_{l,j}^{\sigma_l},\end{aligned}$$

where $\underline{A}_{l,i}^{\sigma_l}$, $\overline{A}_{l,i}^{\sigma_l}$, $\underline{B}_{l,i}^{\sigma_l}$, $\overline{B}_{l,i}^{\sigma_l}$, $\underline{C}_{l,j}^{\sigma_l}$, $\overline{C}_{l,j}^{\sigma_l}$, $\underline{D}_{l,j}^{\sigma_l}$, $\overline{D}_{l,j}^{\sigma_l}$, $\underline{W}_{l,i}^{\sigma_l}$, $\overline{W}_{l,i}^{\sigma_l}$, $\underline{V}_{l,j}^{\sigma_l}$, $\overline{V}_{l,j}^{\sigma_l}$, $\underline{f}_{l,i}^{\sigma_l}$, $\overline{f}_{l,i}^{\sigma_l}$, $\underline{o}_{l,j}^{\sigma_l}$ and $\overline{o}_{l,j}^{\sigma_l}$ are of appropriate dimensions and are constants that are determined by the abstraction algorithms in [16], [17]. The abstracted piecewise affine inclusion models $\mathcal{H}_l$ is then given by:

$$\begin{gathered}\begin{pmatrix}\underline{A}_{l,i}^{\sigma_{l,k}}\boldsymbol{x}_l(k) + \underline{B}_{l,i}^{\sigma_{l,k}}\boldsymbol{u}(k)\\ +\underline{f}_{l,i}^{\sigma_{l,k}} + \underline{W}_{l,i}^{\sigma_{l,k}}w_l(k)\end{pmatrix} \leq \boldsymbol{x}_l(k+1) \leq \begin{pmatrix}\overline{A}_{l,i}^{\sigma_{l,k}}\boldsymbol{x}_l(k) + \overline{B}_{l,i}^{\sigma_{l,l,k}}\boldsymbol{u}(k)\\ +\overline{f}_{l,i}^{\sigma_{l,k}} + \overline{W}_{l,i}^{\sigma_k}w_l(k)\end{pmatrix},\\ \begin{pmatrix}\underline{C}_{l,j}^{\sigma_{l,k}}\boldsymbol{x}_l(k) + \underline{D}_{l,j}^{\sigma_{l,k}}\boldsymbol{u}(k)\\ +\underline{o}_{l,j}^{\sigma_{l,k}} + \underline{V}_{l,j}^{\sigma_{l,k}}v_l(k)\end{pmatrix} \leq z_l(k) \leq \begin{pmatrix}\overline{C}_{l,j}^{\sigma_{l,k}}\boldsymbol{x}_l(k) + \overline{D}_{l,j}^{\sigma_{l,k}}\boldsymbol{u}(k)\\ +\overline{o}_{l,j}^{\sigma_{l,k}} + \overline{V}_{l,j}^{\sigma_{l,k}}v_l(k)\end{pmatrix},\end{gathered} \tag{13}$$

where their corresponding polytopic subregions $I_{l,i}^{h,\sigma_l}$ and $I_{l,j}^{g,\sigma_l}$ are given by the following linear constraints:

$$\begin{aligned}S_{l,i}^{x,\sigma_{l,k}}\boldsymbol{x}_l(k) + S_{l,i}^{u,\sigma_{l,k}}\boldsymbol{u}(k) + S_{l,i}^{w,\sigma_{l,k}}w_l(k) \leq \beta_{l,i}^{\sigma_{l,k}},\\ M_{l,j}^{x,\sigma_k}\boldsymbol{x}_l(k) + M_{l,j}^{u,\sigma_{l,k}}\boldsymbol{u}(k) + M_{l,j}^{v,\sigma_{l,k}}v_l(k) \leq \alpha_{l,j}^{\sigma_{l,k}},\end{aligned} \tag{14}$$

respectively, with $S_{l,i}^{x,\sigma_{l,k}}$, $S_{l,i}^{u,\sigma_{l,k}}$, $S_{l,i}^{w,\sigma_{l,k}}$, $M_{l,j}^{x,\sigma_{l,k}}$, $M_{l,j}^{u,\sigma_{l,k}}$, $M_{l,j}^{v,\sigma_{l,k}}$, $\beta_{l,i}^{\sigma_{l,k}}$ and $\alpha_{l,j}^{\sigma_{l,k}}$ of appropriate dimensions.

III. Problem Formulation

In this paper, we aim to design a separating input sequence such that the (output) trajectories of all the models diverge, even temporarily, regardless of any realization of uncertainties. Formally, the problem of AMD is defined as follows:

Problem 0 (Active Model Discrimination for $\{\mathcal{G}_l\}_{l=1}^N$). *Given N nonlinear models $\{\mathcal{G}_l\}_{l=1}^N$, and state, input and noise constraints, i.e.,* (5)–(11), *find an optimal input sequence u_T^* to minimize a given cost function $J(u_T)$ such that for all possible initial states $\boldsymbol{x}_l^{k_0}$, uncontrolled inputs $d_{l,T}$, process noise $w_{l,T}$, measurement noise $v_{l,T}$ and switching signals $\sigma_{l,k}$, only one model is valid, i.e., the output trajectories of any pair of models have to differ by a threshold ϵ in at least one time instance. The optimization problem can be formally stated as follows:*

$$u_T^\star = \arg\min_{u_T, x_T, z_T} J(u_T)$$

$$\text{s.t. } \forall k \in \mathbb{Z}_{k_0}^{k_0+T-1}: (8) \textit{ holds}, \tag{15a}$$

$$\left.\begin{array}{l}\forall l \in \mathbb{Z}_N^+, \forall k \in \mathbb{Z}_{k_0}^{k_0+T-1},\\ \forall \boldsymbol{x}_l(k_0), d_{l,T}, w_{l,T}, \sigma_{l,T}:\\ (2),(5),(7),(9),(10) \textit{ and}\\ \sigma_{l,T} \in V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi_l) \textit{ hold}\end{array}\right\} : (6) \textit{ holds}, \tag{15b}$$

$$\left.\begin{array}{l}\forall i, j \in \mathbb{Z}_N^+, i < j, \forall k \in \mathbb{Z}_{k_0}^{k_0+T-1},\\ \forall \boldsymbol{x}_l(k_0), d_{l,T}, w_{l,T}, v_{l,T}, \sigma_{l,T},\\ l \in \{i,j\} : (2),(3),(5),(7),(9)\text{-}(11)\\ \textit{and } \sigma_{l,T} \in V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi_l) \textit{ hold}\end{array}\right\} : \begin{array}{l}\{\exists k' \in \mathbb{Z}_{k_0}^{k_0+T-1}:\\ \|z_i(k') - z_j(k')\| \geq \epsilon\}.\end{array} \tag{15c}$$

where the subtrace set $V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi_l)$ is constructed recursively using mixed-integer linear program (MILP) formulations of MTL/STL specifications given in the Appendix.

In the above, constraint (15b) means that for all possible realizations of initial states, uncontrolled states, uncontrolled inputs, process noise and possible switching sequence (on the left side of the brace), the controlled state constraint (6) holds over the entire horizon for each model. Meanwhile, constraint (15c) defines the model separability condition, which means the output trajectories of all pairs of models have to differ by a threshold ϵ in at least one time instant.

However, since the original models $\mathcal{G}_l$ are nonlinear and thus, "hard" to directly compute with since it would result in an intractable mixed-integer nonlinear program, we propose to solve a "simplified" problem by leveraging the piecewise affine abstraction tools described in the previous section to obtain a more tractable MILP, in lieu of solving Problem 0. Since the behavior of the abstracted models $\mathcal{H}_l$ contains the behavior of the original models $\mathcal{G}_l$ (c.f. [15] for details), the following problem provides a sufficient solution that still guarantees model discrimination:

Problem 1 (Active Model Discrimination for $\{\mathcal{H}_l\}_{l=1}^N$)**.** *Given N abstracted piecewise affine inclusion models $\{\mathcal{H}_l\}_{l=1}^N$, and state, input and noise constraints, i.e.,* (5)–(11)*, find an optimal input sequence u_T^* to minimize a given cost function $J(u_T)$ such that for all possible initial states $\boldsymbol{x}_l^{k_0}$, uncontrolled inputs $d_{l,T}$, process noise $w_{l,T}$ and measurement noise $v_{l,T}$, switching signals $\sigma_{l,k}$, only one model is valid, i.e., the output trajectories of any model pair have to differ by a threshold ϵ in at least one time instance. The optimization problem can be formally stated similar to* (15) *with the slight modification of replacing* (2) *and* (3) *with* (13) *and* (14)*.*

As we will see in Section V, the resulting optimization problem for solving Problem 1 is generally still computationally demanding for even relatively small problems. This is due to potentially large input domains and the involvement of binary variables to denote the active partition of the piecewise abstraction model and to represent the switching mode. A secondary goal of this paper is thus to tackle the issue of computational efficiency.

Problem 2 (Strategies for Tractable AMD)**.** *Design strategies to make the AMD problem in Problem 1 more computationally efficient while guaranteeing model separation.*

IV. Main Result

In this section, we first present optimization-based approaches to solve Problem 1 (and, in turn, Problem 0) by formulating the constraints in Problem 1 as optimization problems in Proposition 1 and 2, leading to a bilevel optimization problem. Then, we propose to parametrically solve the inner (optimization) problems given in Proposition 1 and 2 using a multi-parametric programming solver, e.g., MPT 3.0 [20], whose solutions are then incorporated into the original Problem 1 in Theorem 1. As the parametric problems (Proposition 1 and 2) are often computationally demanding or even intractable, we also propose several strategies/approaches to solve them efficiently.

A. Active Model Discrimination for $\{\mathcal{H}_l\}_{l=1}^N$

We provide the following propositions for reformulating the controlled states constraints in (15b), model separability condition in (15c) of Problem 1 (with (13) and (14)). Note that in each of the following propositions, we assume that the other constraints in Problem 1 are satisfied.

Proposition 1. *(State Constraint Reformulation) The polytopic state constraint for each model $l \in \mathbb{Z}_N^+$ (subscript l is omitted for clarity), i.e., the controlled state constraint* (6)*, in* (15b) *of Problem 1 (with* (13) *and* (14)*) is equivalent to*

$$\rho \geq 0,, \tag{16}$$

where ρ (implicitly dependent on the decision variable u_T) is the solution to:

$$\begin{aligned}
&\rho = \arg \min_{\substack{\boldsymbol{x}_T, w_T, v_T, z_T, s_*(k),\\ \tilde{s}_\dagger(k), a_*(k), \tilde{a}_\dagger(k), c^\sigma(k), r^\sigma(k)}} \gamma && \text{(17a)}\\
&\textit{s.t. } \forall * \in \mathbb{Z}_{q^h}^+, \dagger \in \mathbb{Z}_{q^g}^+, \phi \in \mathbb{Z}_{n_x}^+, k \in \mathbb{Z}_{k_0}^{k_0+T-1}: && \text{(17b)}\\
&\boldsymbol{x}(k+1) \leq \overline{A}_*^{\sigma_k} \boldsymbol{x}(k) + \overline{B}_*^{\sigma_k} \boldsymbol{u}(k) + \overline{W}_*^{\sigma_k} w(k)\\
&\qquad + \overline{f}_*^{\sigma_k} + (s_*(k) + c^\sigma(k))\mathbb{1}, && \text{(17c)}\\
&\boldsymbol{x}(k+1) \geq \underline{A}_*^{\sigma_k} \boldsymbol{x}(k) + \underline{B}_*^{\sigma_k} \boldsymbol{u}(k) + \underline{W}_*^{\sigma_k} w(k)\\
&\qquad + \underline{f}_*^{\sigma_k} + (s_*(k) + c^\sigma(k))\mathbb{1}, && \text{(17d)}\\
&S_*^{x,\sigma_k} \boldsymbol{x}(k) + S_*^{u,\sigma_k} \boldsymbol{u}(k) + S_*^{w,\sigma_k} w(k) \leq \beta_*^{\sigma_k}\\
&\qquad + (s_*(k) + c^\sigma(k))\mathbb{1}, && \text{(17e)}\\
&z(k) \leq \overline{C}_\dagger^{\sigma_k} \boldsymbol{x}(k) + \overline{D}_\dagger^{\sigma_k} \boldsymbol{u}(k) + \overline{V}_\dagger^{\sigma_k} v(k)\\
&\qquad + \overline{o}_\dagger^{\sigma_k} + (\tilde{s}_\dagger(k) + c^\sigma(k))\mathbb{1}, && \text{(17f)}\\
&z(k) \geq \underline{C}_\dagger^{\sigma_k} \boldsymbol{x}(k) + \underline{D}_\dagger^{\sigma_k} \boldsymbol{u}(k) + \underline{V}_\dagger^{\sigma_k} v(k)\\
&\qquad + \underline{o}_\dagger^{\sigma_k} + (\tilde{s}_\dagger(k) + c^\sigma(k))\mathbb{1}, && \text{(17g)}\\
&M_\dagger^{x,\sigma_k} \boldsymbol{x}(k) + M_\dagger^{u,\sigma_k} \boldsymbol{u}(k) + M_\dagger^{v,\sigma_k} v(k) \leq \alpha_\dagger^{\sigma_k}\\
&\qquad + (\tilde{s}_\dagger(k) + c^\sigma(k))\mathbb{1}, && \text{(17h)}\\
&\tilde{\gamma_\phi} = p_{x,\phi} - P_{x,\phi} x_(k), P_y y(k) \leq p_y, \gamma = \min \tilde{\gamma_\phi}, && \text{(17i)}\\
&a_*(k) \in \{0,1\}, \tilde{a}_\dagger(k) \in \{0,1\}, r^\sigma(k) \in \{0,1\}, && \text{(17j)}\\
&\textit{SOS-1:}(a_*(k), s_*(k)), \textit{SOS-1:}(\tilde{a}_\dagger(k), \tilde{s}_\dagger(k)),\\
&\textit{SOS-1:}(c^\sigma(k), r^\sigma(k)), && \text{(17k)}\\
&\textstyle\sum_{\xi=1}^{q^{h,\sigma}} a_\xi(k) = 1, \sum_{\xi=1}^{q^{g,\sigma}} \tilde{a}_\xi(k) = 1, \sum_{\xi=1}^{|\Sigma|} r^\xi(k) = 1, && \text{(17l)}\\
&w(k) \in \mathcal{W}, v(k) \in \mathcal{V}, u(k) \in \mathcal{U}, && \text{(17m)}\\
&\{\sigma_k\}_{k=\underline{k}}^{\overline{k}} \in V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi), && \text{(17n)}
\end{aligned}$$

where $s_*(k)$, $\tilde{s}_\dagger(k)$ and $c^\sigma(k)$ are slack variables, $r^\sigma(k) = 1$ corresponds to $\sigma_k = \sigma$, and the subtrace set $V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi)$ is constructed recursively using MILP formulations of MTL/STL specifications given in the Appendix.

Proof. $a_*(k) = 1$ and $r^\sigma(k) = 1$ in (17c)–(17e) imply that the former inequalities in (13)–(14) hold, since the SOS-1 constraints in (17k) ensure that $s_*(k) = 0$ and $c^\sigma(k) = 0$. On the contrary, if $a_*(k) = 0$ and/or $r^\sigma(k) = 0$, it means that $s_*(k)$ and $c^\sigma(k)$ are free and then (17c)–(17e) hold trivially. Similarly, $\tilde{a}_\dagger(k) = 1$ and $r^\sigma(k) = 1$ in (17f)–(17h) imply that the latter inequalities in (13)–(14) hold. In addition, (17l) ensures that, at each time step k, only one partition is valid for each of the state and output equations, and only one switching signal/mode is valid. Moreover, to completely express the constraints induced by φ on σ_k and $r^\sigma(k)$ as MILP reformulation, the horizon of σ_k and $r^\sigma(k)$ is extended both backwards and forwards by a factor of $\max(\phi_g^i, \phi_g^j)$ via the definition of $\underline{k}$ and $\overline{k}$. Finally, (16) (also see (17i)) guarantees that the state constraint (15b) holds. □

Proposition 2. *(Separability Condition Reformulation) The separability condition in (15c) (with* (13) *and* (14)*) for each model pair $i, j \in \mathbb{Z}_N^+, i < j$ of Problem 1 is equivalent to*

$$\delta^{(i,j)} \geq \epsilon, \tag{18}$$

where $\delta^{(i,j)}$ (implicitly dependent on u_T) is the solution to:

$$\delta^{(i,j)} = \arg \min_{\substack{\boldsymbol{x}_T, w_{l,T}, v_{l,T}, z_T, s_{\star,*}(k), \\ \tilde{s}_{\star,\dagger}(k), a_{\star,*}(k), \tilde{a}_{\star,\dagger}(k), c^{\sigma_\star}(k), r^{\sigma_\star}(k)}} \eta \tag{19a}$$

$$\text{s.t. } \star \in \{i, j\}, \forall i, j \in \mathbb{Z}_N^+, i \neq j, \forall * \in \mathbb{Z}_1^{q^{h,\sigma_\star}}, \dagger \in \mathbb{Z}_1^{q^{g,\sigma_\star}}, k \in \mathbb{Z}_{k_0}^{k_0+T-1}: \tag{19b}$$

$$\boldsymbol{x}_\star(k+1) \leq \overline{A}_{\star,*}^{\sigma_\star,k} \boldsymbol{x}_\star(k) + \overline{B}_{\star,*}^{\sigma_\star,k} \boldsymbol{u}(k) + \overline{W}_{\star,*}^{\sigma_\star,k} w_\star(k) + \overline{f}_{\star,*}^{\sigma_\star,k} + (s_{\star,*}(k) + c^{\sigma_\star}(k))\mathbb{1}, \tag{19c}$$

$$\boldsymbol{x}_\star(k+1) \geq \underline{A}_{\star,*}^{\sigma_\star,k} \boldsymbol{x}_\star(k) + \underline{B}_{\star,*}^{\sigma_\star,k} \boldsymbol{u}(k) + \underline{W}_{\star,*}^{\sigma_\star,k} w_\star(k) + \underline{f}_{\star,*}^{\sigma_\star,k} + (s_{\star,*}(k) + c^{\sigma_\star}(k))\mathbb{1}, \tag{19d}$$

$$S_{\star,*}^{x,\sigma_\star,k} \boldsymbol{x}_\star(k) + S_{\star,*}^{u,\sigma_\star,k} \boldsymbol{u}(k) + S_{\star,*}^{w,\sigma_\star,k} w_\star(k) \leq \beta_{\star,*}^{\sigma_\star,k} + (s_{\star,*}(k) + c^{\sigma_\star}(k))\mathbb{1}, \tag{19e}$$

$$z_\star(k) \leq \overline{C}_{\star,\dagger}^{\sigma_\star,k} \boldsymbol{x}_\star(k) + \overline{D}_{\star,\dagger}^{\sigma_\star,k} \boldsymbol{u}(k) + \overline{V}_{\star,\dagger}^{\sigma_\star,k} v_\star(k) + \overline{o}_{\star,\dagger}^{\sigma_\star,k} + (\tilde{s}_{\star,\dagger}(k) + c^{\sigma_\star}(k))\mathbb{1}, \tag{19f}$$

$$z_\star(k) \geq \underline{C}_{\star,\dagger}^{\sigma_\star,k} \boldsymbol{x}_\star(k) + \underline{D}_{\star,\dagger}^{\sigma_\star,k} \boldsymbol{u}(k) + \underline{V}_{\star,\dagger}^{\sigma_\star,k} v_\star(k) + \underline{o}_{\star,\dagger}^{\sigma_\star,k} + (\tilde{s}_{\star,\dagger}(k) + c^{\sigma_\star}(k))\mathbb{1}, \tag{19g}$$

$$M_{\star,\dagger}^{x,\sigma_\star,k} \boldsymbol{x}_\star(k) + M_{\star,\dagger}^{u,\sigma_\star,k} \boldsymbol{u}(k) + M_{\star,\dagger}^{v,\sigma_\star,k} v_\star(k) \leq \alpha_{\star,\dagger}^{\sigma_\star,k} + (\tilde{s}_{\star,\dagger}(k) + c^{\sigma_\star}(k))\mathbb{1}, \tag{19h}$$

$$\|z_i(k) - z_j(k)\| \leq \eta, P_{y,\star} y_\star(k) \leq p_{y,\star}, \tag{19i}$$

$$a_{\star,*}(k) \in \{0,1\}, \tilde{a}_{\star,\dagger}(k) \in \{0,1\}, r^{\sigma_\star}(k) \in \{0,1\}, \tag{19j}$$

$$\text{SOS-1:}(a_{\star,*}(k), s_{\star,*}(k)), \text{SOS-1:}(\tilde{a}_{\star,\dagger}(k), \tilde{s}_{\star,\dagger}(k)), \text{SOS-1:}(c^{\sigma_\star}(k), r^{\sigma_\star}(k)), \tag{19k}$$

$$\textstyle\sum_{\xi=1}^{q_\star^{h,\sigma}} a_{\star,\xi}(k) = 1, \sum_{\xi=1}^{q_\star^{g,\sigma}} \tilde{a}_{\star,\xi}(k) = 1, \sum_{\xi=1}^{|\Sigma|} r^\xi(k) = 1, \tag{19l}$$

$$w_\star(k) \in \mathcal{W}, v_\star(k) \in \mathcal{V}, u(k) \in \mathcal{U}, \tag{19m}$$

$$\{\sigma_{\star,k}\}_{k=\underline{k}}^{\overline{k}} \in V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi_\star), \tag{19n}$$

where $s_{\star,*}(k)$, $\tilde{s}_{\star,\dagger}(k)$ and $c^{\sigma_\star}(k)$ are slack variables, $r^{\sigma_\star}(k) = 1$ corresponds to $\sigma_{\star,k} = \sigma$, and the subtrace set $V_{\underline{k}}^{\overline{k}-\underline{k}+1}(\varphi_\star)$ is constructed recursively using MILP formulations of MTL/STL specifications given in the Appendix.

Proof. The construction follows similar steps to Proposition 1, but with two different models i, j and with the separation condition in (19i). □

Using Propositions 1, 2, it is straightforward to show that the active model discrimination problem in Problem 1 can be recast as the following bilevel optimization problem:

Theorem 1 (Bilevel AMD)**.** *Given a separability index ϵ, Problem 1 is equivalent to*

$$u_T^* = \arg \min_{u_T} J(u_T) \tag{20a}$$

$$\text{s.t. } (16) \text{ and } (18) \text{ hold,} \tag{20b}$$

where ρ and $\delta^{(i,j)}$ are solutions to the inner problems given by (17) *and* (19)*, respectively.*

Since the inner problems (17) and (19) contain integer variables, standard reformulation methods using KKT or robust optimization, which was employed in [7], [12], do not apply. Thus, we propose to parametrically solve the inner (optimization) problems given in Proposition 1 and 2 using a multi-parametric programming solver, e.g., MPT 3.0 [20]. Their solutions are then incorporated into the outer problem in Theorem 1 to find the minimal separating input.

Then, using the computed optimal input and the observed measurements (corresponding to the unknown true model), we can eliminate the false models with the model invalidation algorithm proposed in [15].

B. Strategies for Tractable AMD

However, since even small parametric optimization problems can often be computationally demanding, we propose the following two strategies to solve Problem 2.

1) Horizon Truncation: Another way to make the multi-parametric optimization problems in Propositions 1 and 2 more tractable is by reducing the number of binary variables involved in some way, as they are the major reasons leading to intractable problems. The following horizon truncation strategy is one such solution that can be applied to obtain a potentially sub-optimal but tractable problem.

Proposition 3 (Tractable AMD with Truncated Horizon)**.** *In Propositions 1 and 2, a tractable* sufficient *problem can be obtained by reducing the extended time horizon for the constraints* (17n) *and* (19n)*, by setting the upper limit to $\overline{k} = k_0 + T - 1$ when $k_0 > \max_l\{b^{\phi_{b,l}}\}$, or $\overline{k} = \max\{k_0 + T - 1, \max_l\{b^{\phi_{b,l}}\}\}$ otherwise.*

Proof. By eliminating the binary variables $\sigma_{\star,k}$ and $r_k^{\sigma_\star}$ for all $k > k_0 + T - 1$ when $k_0 > \max_l\{b^{\phi_{b,l}}\}$, or $\overline{k} = \max\{k_0 + T - 1, \max_l\{b^{\phi_{b,l}}\}\}$ otherwise, the resulting optimization problem will have relaxed constraints (and also fewer decision variables). Thus, when the relaxed problem is solved, we obtain a smaller ρ, denoted as $\underline{\rho}$. If the solution to the relaxed problem exists, it provides a lower bound to the cost of the original inner problem in Proposition 1 as $\rho \geq \underline{\rho}$, hence the separating input that satisfies $\underline{\rho} \geq 0$ also satisfies $\rho \geq 0$. A similar argument applies to its sufficiency for Proposition 2. □

Although the strategy proposed in Proposition 3 leads to a sub-optimal solution in general, there are specific cases where applying the strategy will not result in the loss of optimality. One such case (cf. example in Section V-A) arises when the given MTL/STL specifications are such that the MTL/STL formulas only affect the switching modes and the future truth conditions for $\sigma_{l,i}$, $\forall i > k_0 + T - 1$ are independent of the choice of the separating input sequence.

2) Input Domain Partitioning: When the input domain is large, existing multi-parametric tools can have difficulties dealing with a large number of critical regions (e.g., in the order of thousands). One way to tackle this issue is to shrink/reduce the input region by partitioning the input set (cf. Definition 2), and solve each individual problem with the corresponding input subset separately (or in parallel, if desired). In other words, we still consider the same problem but we instead solve a finite number of "smaller" tractable problems. Finally, within the results from these tractable problems, we choose the most optimal feasible solution.

Remark 1 (Tractable AMD with Input Domain Partitioning)**.** *Using Definition 2, the input domain $\mathcal{U}$ in* (17m) *and* (19m) *in Propositions 1 and 2 can be partitioned into $\mathcal{U}_i$, where $i \in \mathbb{Z}^+_{|U|}$. When using the partitioned input region $\mathcal{U}_i$ in lieu of $\mathcal{U}$, the multi-parametric optimization problems in Propositions 1 and 2 typically result in much fewer critical regions, thus making the problems tractable.*

V. Simulation Examples

In this section, we apply our proposed approach to a fault detection example and an intent estimation example in a highway lane changing scenario. The fault detection simulation is implemented in MATLAB 2020b with Gurobi v9.0.3 [18] and MPT 3.0 [20] on a 2.6GHz hexa-core machine with 32 GB RAM, whereas the intent estimation example is implemented in MATLAB 2019b with Gurobi v8.1 [18] and MPT 3.0 [20] on a 1.3 GHz dual-core machine with 16 GB RAM.

A. Fault Detection Example

For the first example, we consider fault detection of a 2D affine system $\{\mathcal{G}_l\}_{l\in\{\mathtt{h},\mathtt{f}\}}$. Here, the model $\mathcal{G}_{l=h}$ denotes the healthy system, whereas $\mathcal{G}_{l=f}$ represents the faulty system. The system $\mathcal{G}_l$ is considered as $x_l(k+1) = A_l^{\sigma_{l,k}} x_l(k) + B_l^{\sigma_{l,k}} u(k) + h_l^{\sigma_{l,k}} + w_l(k)$, $z_l(k) = x_l(k) + v_l(k)$, with $\sigma_{l,k} \in \{1,2\}$ acting as a controlled switching signal, with disturbance $w_l(k) \in [-0.1, 0.1] \times [-0.1, 0.1]$, and noise $v_l(k) \in [-0.01, 0.01] \times [-0.01, 0.01]$. The states $x_l(k)$ lie within $X = [-5, 5] \times [-5, 5]$ and the initial state set is $X_l^0 = [1.5, 2.5] \times [0.5, 1.5]$. For the healthy system, its state-space matrices are:

$$A^1_{\mathtt{h}}=\begin{bmatrix}0.794 & 0.723\\ -0.260 & 0.794\end{bmatrix}, A^2_{\mathtt{h}}=\begin{bmatrix}0.794 & 0.434\\ -0.434 & 0.794\end{bmatrix},$$
$$B^1_{\mathtt{h}}=\begin{bmatrix}1\\0\end{bmatrix}, h^1_{\mathtt{h}}=\begin{bmatrix}0\\0\end{bmatrix}, B^2_{\mathtt{h}}=\begin{bmatrix}1\\0\end{bmatrix}, h^2_{\mathtt{h}}=\begin{bmatrix}0.115\\0.457\end{bmatrix}.$$

For the faulty system, a fault changes the dynamics of the mode $\sigma_{\mathtt{f},k} = 2$ to be of the following form:

$$A^2_{\mathtt{f}}=0.8\times\begin{bmatrix}0.635 & 0.578\\ -0.208 & 0.635\end{bmatrix}, B^2_{\mathtt{f}}=\begin{bmatrix}0\\0\end{bmatrix}, h^2_{\mathtt{f}}=\begin{bmatrix}0\\0\end{bmatrix}.$$

The switching signal for the healthy system is required to satisfy the MTL/STL specification $\varphi_{\mathtt{h}} = \phi^1_{\mathtt{h}} \wedge \phi^2_{\mathtt{h}}$, where

$$\begin{aligned}\phi^1_{\mathtt{h}} &= \Box\big(\Box_{[0,2]}(\sigma=1) \Rightarrow \bigcirc^3\neg(\sigma=1)\big),\\ \phi^2_{\mathtt{h}} &= \Box\big(\Box_{[0,2]}(\sigma=2) \Rightarrow \bigcirc^3\neg(\sigma=2)\big),\end{aligned} \tag{21}$$

whereas the faulty model does not have its switching signal governed by any specification. Since the system considered is affine, its abstraction will be the same as the system itself. Using Theorem 1 for fault detection with the input confined to $u(k) \in [-0.1, 0.1]$, the permanent fault in the system can be detected in 5 time steps with $\{u(k)\}_{k=0}^4 = \{0.0179, -0.0179, -0.0179, -0.0179, -0.0179\}$. On the other hand, without any separating control input, the system is 6-detectable when T-distinguishability solutions from [15] are applied. Note that due to the presence of binary variables in the form of switching modes, the resource requirement by MPT toolbox scales exponentially with the total number of binary variables involved in the problem. To observe the effect of the tractable strategy discussed in Proposition 3, the problem of fault detection is run with and without the strategy, and a comparison is shown in Table I. For this problem, as the horizon is extended backwards and forwards by 3 based on $\varphi_{\mathtt{h}}$, ignoring just 3 binary variables by truncating the forward horizon extension (due to the MTL/STL formulas) resulted in significant reduction in binary nodes involved in finding the solution, resulting in smaller solving time. Moreover, as the STL formula in (21) is such that the truth value of a σ_k only depends on switching modes and not the separating inputs, the optimal solutions obtained with and without the strategy in Proposition 3 are the same, as previously discussed.

TABLE I: Comparison of computation times in the fault detection example.

Strategy	Binary Nodes Involved	Solving Time (s)	$\|u\|_\infty$
Without Prop. 3	1854	47252.28	0.0179
With Prop. 3	298	6986.39	0.0179

B. Intent Estimation Example

Similar to [12], we consider an intent estimation problem in a lane changing scenario with two intent models for the other car $l \in \{C, M\}$ (see [12] for details): the *Cautious* driver tends to yield the lane to ego car or the *Malicious* driver does not want to yield the lane to the ego car. The piecewise affine abstraction/inclusion model with two pieces/subregions is obtained from the affine abstraction approach in [17] and is given by:

$$\begin{aligned}
&A_1 = [(1,3):0.9048;(2,4):21.781]_{6\times6};\\
&A_2 = [(1,3):0.9048;(2,4):26.6212]_{6\times6};\\
&B_1 = B_2 = W_1 = W_2 = B = [(3,1):1;(4,2):1;(6,3):1]_{6\times3};\\
&F_1 = [\underline{f}_1, \overline{f}_1] = [(1,1):[-3.4680, 5.8193];\\
&\qquad\qquad (1,2):[-9.2068, 9.2068]]_{6\times1};\\
&F_2 = [\underline{f}_2, \overline{f}_2] = [(1,1):[-4.1616, 6.9832];\\
&\qquad\qquad (1,2):[-10.8352, 10.8352]]_{6\times1};
\end{aligned}$$

with a sparse matrix notation with the size indicted in the subscript. Combining the abstraction with the intention models and using Euler method for time discretization with sampling time $\delta t = 0.4s$, we have the following intention models $\mathcal{H}_l$, $l \in \{C, M\}$ (with C and M representing the *Cautious* and *Malicious* driver models, respectively):

$$\begin{aligned}
&\tilde{A}_l = [(6,3):-K_{d,l};(6,4):L_{p,l};(6,6):K_{d,l}]_{6\times6},\\
&\tilde{B}_l = [(6,2):L_{d,l}]_{6\times3},\\
&\underline{A}_{l,\ddagger} = \overline{A}_{l,\ddagger} = \mathbb{I} + \delta t(A_\ddagger + \tilde{A}_l),\\
&\underline{B}_{l,\ddagger} = \overline{B}_{l,\ddagger} = \underline{W}_{l,\ddagger} = \overline{W}_{l,\ddagger} = \mathbb{I} + \delta t(B_\ddagger + \tilde{B}_l),\\
&\underline{C}_l = \overline{C}_l = [(1,6):1]_{1\times6}, \underline{D}_l = \overline{D}_l = 0, \underline{V}_l = \overline{V}_l = 1,\\
&\underline{f}_{l,\ddagger} = \delta t \underline{f}_\ddagger, \overline{f}_{l,\ddagger} = \delta t \overline{f}_\ddagger, S^x_{l,\ddagger} = [(1,3):1;(2,3):-1]_{2\times6},\\
&\beta_{l,1} = [25;-20], \beta_{l,2} = [30;-25],
\end{aligned}$$

where $\ddagger \in \{1,2\}$, with $K_{d,C} = 1$, $L_{p,C} = 12$ $L_{d,C} = 14$, $K_{d,M} = -0.9$, $L_{p,M} = -12$, and $L_{d,M} = -14$.

When applying Theorem 1 with $\|u_T\|$ as the cost function based on multi-parametric optimization to this intent estima-

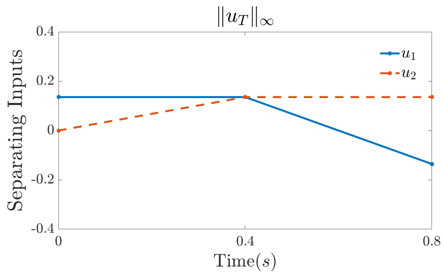

Fig. 1: Optimal inputs obtained by solving the proposed active model discrimination for intents, $l \in \{C, M\}$.

tion example with the above intent models, the problem was found to be computationally intractable. Thus, we employed the input domain partitioning strategy in Section IV-B, and with this strategy, we were able to obtain/compute a separating input sequence, as shown in Figure 1, where the ego car has to speed up at the first time step due to the constraint that the ego car can only move forwards. Further, the cost with the proposed approach was found to be less than the result from [12] by around 32%, which means that our approach that uses a piecewise affine abstraction/inclusion model to over-approximate nonlinear dynamics has better performance than the approach in [12] with (single-domain) affine abstraction. The computation time is longer than the previous method in [12] but this is less important since the AMD problems are meant to be solved only once offline and not at run time.

VI. Conclusions

This paper considered the AMD problem which distinguishes among a set of noisy switched nonlinear models constrained by metric/signal temporal logic specifications. To deal with nonlinear, integral and non-convex constraints in the problem, we leveraged piecewise affine abstraction tools and parametric optimization to tractably solve the resulting bilevel optimization problem. Furthermore, since these parametric optimization problems are often computationally demanding, we propose several strategies to reduce computational time and make the problem more tractable. Finally, the effectiveness of our approach is demonstrated via illustrative examples on fault detection and intent estimation.

References

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part i: Quantitative model-based methods," *Computers & Chemical Engineering*, vol. 27, no. 3, pp. 293–311, 2003.

[2] F. Harirchi, S. Z. Yong, and N. Ozay, "Guaranteed fault detection and isolation for switched affine models," in *IEEE Conference on Decision and Control (CDC)*, 2017, pp. 5161–5167.

[3] S. Cheong and I. Manchester, "Input design for discrimination between classes of LTI models," *Automatica*, vol. 53, pp. 103–110, 2015.

[4] R. Nikoukhah and S. Campbell, "Auxiliary signal design for active failure detection in uncertain linear systems with a priori information," *Automatica*, vol. 42, no. 2, pp. 219–228, Feb. 2006.

[5] J. K. Scott, R. Findeison, R. D. Braatz, and D. M. Raimondo, "Input design for guaranteed fault diagnosis using zonotopes," *Automatica*, vol. 50, no. 6, pp. 1580–1589, Jun. 2014.

[6] S. Zhai, W. Wang, and H. Ye, "Auxiliary signal design for active fault detection based on set-membership," *IFAC PapersOnLine*, vol. 48, no. 21, pp. 452–457, 2015.

[7] Y. Ding, F. Harirchi, S. Z. Yong, E. Jacobsen, and N. Ozay, "Optimal input design for affine model discrimination with applications in intention-aware vehicles," in *ACM/IEEE International Conference on Cyber-Physical Systems*, 2018, available from: arXiv:1702.01112.

[8] D. M. Raimondo, G. R. Marseglia, R. D. Braatz, and J. K. Scott, "Closed-loop input design for guaranteed fault diagnosis using set-valued observers," *Automatica*, vol. 74, pp. 107–117, Dec. 2016.

[9] G. R. Marseglia and D. M. Raimondo, "Active fault diagnosis: A multi-parametric approach," *Automatica*, vol. 79, pp. 223–230, 2017.

[10] R. Niu, Q. Shen, and S. Z. Yong, "Partition-based parametric active model discrimination design with applications to driver intention estimation," in *ECC*, 2019, pp. 3880–3885.

[11] Q. Shen and S. Z. Yong, "Active model discrimination using partition-based output feedback," in *European Control Conference*, 2020, pp. 712–717.

[12] K. Singh, Y. Ding, N. Ozay, and S. Z. Yong, "Input design for nonlinear model discrimination via affine abstraction," *IFAC PapersOnLine*, vol. 51, no. 16, pp. 175–880, 2018.

[13] E. Asarin, A. Donzé, O. Maler, and D. Nickovic, "Parametric identification of temporal properties," in *International Conference on Runtime Verification*. Springer, 2011, pp. 147–160.

[14] L. Yang and N. Ozay, "Fault detectability analysis of switched affine systems with linear temporal logic constraints," in *IEEE Conference on Decision and Control (CDC)*, 2019, pp. 5779–5786.

[15] R. Niu, S. M. Hassaan, L. Yang, Z. Jin, and S. Z. Yong, "Model discrimination of switched nonlinear systems with temporal logic-constrained switching," *IEEE Control Systems Letters*, vol. 6, pp. 151–156, 2021.

[16] K. Singh, Q. Shen, and S. Z. Yong, "Mesh-based affine abstraction of nonlinear systems with tighter bounds," in *IEEE Conference on Decision and Control*, 2018, pp. 3056–3061.

[17] Z. Jin, Q. Shen, and S. Z. Yong, "Mesh-based piecewise affine abstraction with polytopic partitions for nonlinear systems," *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1543–1548, 2020.

[18] Gurobi Optimization, Inc., "Gurobi optimizer reference manual," 2015. [Online]. Available: http://www.gurobi.com

[19] I. I. CPLEX, "V12. 1: User's manual for CPLEX," *International Business Machines Corporation*, vol. 46, no. 53, p. 157, 2009.

[20] M. Herceg, M. Kvasnica, C. N. Jones, and M. Morari, "Multi-parametric toolbox 3.0," in *the European Control Conference*, 2013, pp. 502–510.

Appendix

As discussed in [15], the MTL/STL formulas can be directly encoded into integer constraints for use in the proposed model discrimination problem. For brevity, we will now only present constraints for the satisfaction of each operator of the MTL/STL semantics, i.e., $(\boldsymbol{\sigma}, t) \models \varphi$ for the following operators, where p, q and p_i are atomic propositions, and P_φ^t is the truth value of formula φ at time t:

Negation: The formula $\varphi = \neg p$ can be modeled as:

$$P_\varphi^t = (1 - P_p^t).$$

Disjunction: The formula $\varphi = \bigvee_{i=1}^k p_i$ can be modeled as:

$$P_\varphi^t \leq \Sigma_{i=1}^k P_{p_i}^t; \quad P_\varphi^t \geq P_{p_i}^t, i \in \mathbb{Z}_1^k.$$

Conjunction: The formula $\varphi = \bigwedge_{i=1}^k p_i$ can be modeled as:

$$P_\varphi^t \geq \Sigma_{i=1}^k P_{p_i}^t - (k-1); \quad P_\varphi^t \leq P_{p_i}^t, i \in \mathbb{Z}_1^k.$$

Next: The formula $\varphi = \bigcirc p$ can be modeled as:

$$P_\varphi^t = P_p^{t+1}.$$

Until: The formula $\varphi = p\,\mathcal{U}_{[t_1,t_2]}q$ can be modeled as:

$$\begin{gathered} \alpha_{tj} \geq P_q^j + \Sigma_{\tau=t}^{j-1} P_p^\tau - (j-t), j \in \mathbb{Z}_{t+t_1}^{t+t_2}; \\ \alpha_{tj} \leq P_q^j, \ \alpha_{tj} \leq P_p^\tau, j \in \mathbb{Z}_{t+t_1}^{t+t_2}, \tau \in \mathbb{Z}_t^{j-1}; \\ P_\varphi^t \leq \Sigma_{j=t+t_1}^{t+t_2} \alpha_{tj}, P_\varphi^t \geq \alpha_{tj}, j \in \mathbb{Z}_{t+t_1}^{t+t_2}. \end{gathered}$$

Eventually: The formula $\varphi = \Diamond_{[t_1,t_2]}p$ can be modeled as:

$$P_\varphi^t \leq \Sigma_{\tau=t+t_1}^{t+t_2} P_p^\tau; \quad P_\varphi^t \geq P_p^\tau, \tau \in \mathbb{Z}_{t+t_1}^{t+t_2}.$$

Always: The formula $\varphi = \Box_{[t_1,t_2]}p$ can be modeled as:

$$P_\varphi^t \geq \Sigma_{\tau=t+t_1}^{t+t_2} P_p^\tau - (t_2 - t_1); \quad P_\varphi^t \leq P_p^\tau, \tau \in \mathbb{Z}_{t+t_1}^{t+t_2}.$$