Active Model Discrimination Using Partition-Based Output Feedback Input Design

Qiang Shen[1] and Sze Zheng Yong[2]

Abstract— **In this paper, we propose a partition-based output feedback active model discrimination approach that generates optimal output feedback inputs in a fixed time horizon for separating a set of discrete-time affine models subject to uncontrolled inputs, noises and uncertain initial conditions. Instead of computing the optimal input by solving a parametric mixed-integer linear program (MILP) at run time, we move this computationally demanding optimization task offline by partitioning the measurement domain and building a partition tree over the fixed time horizon. Since output measurements are available at each time instant during run time, we can update the separating input correspondingly and improve the model discrimination performance by reducing the input cost. The effectiveness of the proposed approach is demonstrated through simulations for identifying intention models of human-driven vehicles in a lane changing scenario.**

I. Introduction

The discrimination and identification of internal states (i.e., intention, fault or mode of operation) is essential for cyber-physical systems. For example, unknown malfunctions or external attacks during system operation can result in performance degradation, unsafe behaviors or even critical situations, although they often cannot be directly measured or observed by system outputs. Therefore, it is of great interest to develop model discrimination approaches.

Literature Review: The objective of model discrimination is to separate all models from each other from a set containing all possible models despite the presence of external disturbances, process and measurement noises. It has been widely studied in various research areas such as intention estimation, fault detection and model discernibility. The current approaches for model discrimination can be generally categorized into either passive or active methods. By checking whether the available input-output data is compatible with each potential model in real time, the passive approach ensures model discrimination without designing/perturbing the input applied to the system [1], [2], [3]. The passive approach is relatively easy to implement but is only applicable for problems with specific system properties [4]. Moreover, system uncertainties and the actions of the feedback controller could mask the influence of model differences on the output and reduce the capability of passive approaches. On the other hand, the active approach seeks a common input for all models that minimally intervenes with the system [5], [6], [7], [8], so that behaviors of different models under this input are guaranteed to be distinct. Primary techniques used for solving the active model discrimination problem are polyhedral projection [6] and mixed-integer program (MIP) [7], [8]. However, the above mentioned approaches are open-loop since the active input is computed offline and no modification is made at run time.

Closed-loop approaches for active model discrimination have been studied to achieve less conservative separating inputs. Instead of injecting the entire input sequence as in open-loop approaches, the closed-loop one refines the input at each time step by leveraging real-time measurements. In [9], the closed-loop active model discrimination problem was solved in a moving horizon framework by using a constrained zonotope that represents the polytopic uncertainties and a set-valued observer that incorporates the online measurements. In [10], to reduce the computational complexity in resolving the optimization problem at each time step, a multi-parametric program that expresses the obtained separating input as a function of parameters was employed to obtain parametrized separating input sequences. In [11], the input design problem for the closed-loop active fault diagnosis approach was proposed for stochastic linear systems in the presence of multiple fault models and stochastic disturbances and measurements, where open-loop inputs are designed using a receding horizon instead of employing output feedback. On the other hand, a partition-based parametric active model discrimination approach is recently presented in [12], where the model-independent parameter space of the affine time-invariant systems was partitioned to improve performance and reduce cost.

Contributions: This paper considers the output feedback input design problem for active model discrimination of discrete-time affine systems in the presence of uncontrolled inputs and measurement noises as well as unknown initial conditions. By leveraging available output measurements at each time instant, the output feedback active model discrimination problem can be formulated as a parametric MILP at each time instant. However, since solving a parametric MILP could often be computationally expensive, if not intractable, especially when there is a large number of binary variables, we partition the measurement space and build a partition tree. Then, we can compute separating inputs for each node of the partition tree offline, and at run time, we simply select the pre-computed input associated with the node that the current measurements belongs to at each time instant.

In contrast to previous open-loop active model discrimination approaches in [4], [8], the proposed approach leverages additional information from the output measurements, resulting in reduced separating input cost over a fixed time horizon. In addition, we relax a rather limiting assumption

[1]School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, P.R. China (e-mail: qiang.shen@sjtu.edu.cn)

[2]School for Engineering of Matter, Transport and Energy, Arizona State University, Tempe, AZ, USA (e-mail: szyong@asu.edu.

This work was done at Arizona State University and was supported in part by DARPA grant D18AP00073.

in [4], [8], [12] that the separating input does not affect the uncontrolled state constraints, and hence, the proposed approach is applicable to more general systems. Finally, we demonstrate the effectiveness of the proposed output feedback active model discrimination approach to separate intention models of human-driven vehicles for autonomous vehicles in a lane changing scenario.

II. Preliminaries

A. Notation and Definitions

Let $x \in \mathbb{R}^n$ denote a vector and $M \in \mathbb{R}^{n\times m}$ a matrix, with transpose $M^\top$ and $M \geq 0$ denotes element-wise non-negativity. The vector norm of x is denoted by $\|x\|_i$ with $i \in \{1, 2, \infty\}$, while $\mathbf{0}$, $\mathbb{1}$ and $\mathbb{I}$ represent the vector of zeros, the vector of ones and the identity matrix of appropriate dimensions. The set of positive integers up to n is denoted by $\mathbb{Z}_n^+$, and the set of non-negative integers up to n is denoted by $\mathbb{Z}_n^0$. In addition, the set of non-negative integers from j_1 to j_2 ($0 \leq j_1 \leq j_2$) is denoted by $\mathbb{Z}_{j_2}^{j_1}$.

Definition 1 (Partition)**.** *A partition of a polyhedral set $\mathcal{Z}$ is a collection of ℓ disjoint subsets $\mathcal{Z}_i$ such that $\bigcup_{i\in\mathbb{Z}_\ell^+} \mathcal{Z}_i = \mathcal{Z}$, where each partition $\mathcal{Z}_i$ is also a polyhedral set.*

B. Modeling Framework

Consider N discrete-time affine time-invariant models $\mathcal{G}_i = (A_i, B_i, B_{w,i}, C_i, D_{v,i}, f_i, g_i)$, each with states $\vec{x}_i \in \mathbb{R}^n$, outputs $z_i \in \mathbb{R}^{n_z}$, inputs $\vec{u}_i \in \mathbb{R}^m$, process noise $w_i \in \mathbb{R}^{m_w}$, measurement noise $v_i \in \mathbb{R}^{m_v}$. The models evolve according to the following state and output equations:

$$\vec{x}_i(k+1) = A_i\vec{x}_i(k) + B_i\vec{u}_i(k) + B_{w,i}w_i(k) + f_i, \tag{1}$$

$$z_i(k) = C_i\vec{x}_i(k) + D_{v,i}v_i(k) + g_i. \tag{2}$$

The initial condition for model i, denoted by $\vec{x}_i^0 = \vec{x}_i(0)$, is constrained to a polyhedral set with c_0 inequalities:

$$\vec{x}_i^0 \in \mathcal{X}_0 = \{\vec{x} \in \mathbb{R}^n : P_0\vec{x} \leq p_0\},\ \forall i \in \mathbb{Z}_N^+. \tag{3}$$

The first m_u components of $\vec{u}_i$ are controlled inputs (i.e., to be designed as separating inputs), denoted as $u \in \mathbb{R}^{m_u}$, which are the same for all $\vec{u}_i$, while the other $m_d = m - m_u$ components of $\vec{u}_i$, denoted as $d_i \in \mathbb{R}^{m_d}$, are uncontrolled inputs that are model-dependent. The controlled and uncontrolled inputs are constrained to the following polyhedral domains (for $k \in \mathbb{Z}_{T-1}^0$) with c_u and c_d inequalities:

$$u(k) \in \mathcal{U} = \{u \in \mathbb{R}^{m_u} : Q_u u \leq q_u\}, \tag{4}$$

$$d_i(k) \in \mathcal{D}_i = \{d \in \mathbb{R}^{m_{d_i}} : Q_{d,i}d \leq q_{d,i}\}. \tag{5}$$

Correspondingly, the states $\vec{x}_i$ are also divided into controlled state $x_i \in \mathbb{R}^{n_x}$ and uncontrolled state $y_i \in \mathbb{R}^{n_y}$, constrained to the following polyhedral domains (for $k \in \mathbb{Z}_{T-1}^0$) with c_x and c_y inequalities, respectively:

$$x_i(k) \in \mathcal{X}_{x,i} = \{x \in \mathbb{R}^{n_x} : P_{x,i}x \leq p_{x,i}\}, \tag{6}$$

$$y_i(k) \in \mathcal{X}_{y,i} = \{y \in \mathbb{R}^{n_y} : P_{y,i}y \leq p_{y,i}\}. \tag{7}$$

On the other hand, the process noise w_i and measurement noise v_i are also polyhedrally constrained with c_w and c_v inequalities, respectively:

$$w_i(k) \in \mathcal{W}_i = \{w \in \mathbb{R}^{m_w} : Q_{w,i}w \leq q_{w,i}\}, \tag{8}$$

$$v_i(k) \in \mathcal{V}_i = \{v \in \mathbb{R}^{m_v} : Q_{v,i}v \leq q_{v,i}\}. \tag{9}$$

Moreover, for the output measurement $z_m(k)$ in real time, we assume that it satisfies c_z polyhedral constraint:

$$z_m(k) \in \mathcal{Z} = \{z \in \mathbb{R}^{n_z} : P_z z \leq p_z\}. \tag{10}$$

Since $z_m(k)$ is revealed in real time, only output measurements obtained so far at the current time t, i.e., $z_m(k)$, $\forall k \in \mathbb{Z}_t^0$, are available for the active input design.

III. Problem Formulation

In this paper, we aim to design a sequence of causal active separating input vector $u_T^t(z_{m,t})$ as a function of the past and current output measurements $z_{m,t} = \mathrm{vec}_{i=0}^t\{z_m(i)\}$ for all $t \in \mathbb{Z}_{T-2}^0$, where $u_T^t(\cdot) = [u^t(0), \ldots, u^t(t), \ldots, u^t(T-1)]^\top$ denotes the concatenated separating input (i.e., the controlled input) over a fixed horizon T at the current time step t and $u_T^{-1}(\cdot) = \emptyset$. By leveraging the newly available output measurements, we can update the separating input at each time step, and hence obtain better model discrimination results (e.g., a smaller objective function value) when compared with open-loop methods [4], [8]. In addition, due to causality, we enforce that the input variable of the optimization problem at current time instant t must inherit the first $t-1$ values of optimal input sequence $u_T^{*,t-1}$, i.e., $u_T^t(k) = u_T^{*,t-1}(k)$ for all $k \in \mathbb{Z}_{t-1}^0$ and $t \in \mathbb{Z}_{T-2}^0$.

Formally, the problem of active model discrimination leveraging the output measurement is defined as follows:

Problem 1 (Output Feedback Active Model Discrimination)**.** *Consider N affine models $\mathcal{G}_i$, and state, input and noise constraints defined in* (3) *and* (4)-(9)*. For each time instant $t \in \mathbb{Z}_{T-1}^0$ (sequentially, starting from $t = 0$), with all past and current available measurement $z_m(k) \in \mathcal{Z}$, $\forall k \in \mathbb{Z}_t^0$ for output feedback, and given the optimal input sequence $u_T^{*,t-1}$ from the previous time instant $t-1$, find an optimal input sequence $u_T^{*,t}$ with fixed horizon T to minimize a given cost function $J(u_T^t) = \|u_T^t\|_1$ subject to $u_T^t(k) = u_T^{*,t-1}(k)$ for all $k \in \mathbb{Z}_{t-1}^0$ such that for all possible initial states $\boldsymbol{x}_0$, uncontrolled inputs d_T, unrevealed output measurement $z_{t+1:T-1}$, process noise w_T and measurement noise v_T, 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 instant of the horizon T.*

Formally, for each time instant t with all available measurement $z_m(k)$ for all $k \in \mathbb{Z}_t^0$ and all previously chosen inputs, $u_T^{*,t-1}$, the optimization problem can be formulated as a sequence of optimization problems as follows (for $t = 0, \ldots, T-2$):

$$u_T^{*,t} = \arg\min_{u_T^t} J(u_T^t)$$

$$\textit{s.t.}\quad \forall k \in \mathbb{Z}_{T-1}^0 : (4)\ \textit{holds}, \tag{11a}$$

$$\left.\begin{array}{r} \forall i \in \mathbb{Z}_{\mathrm{N}}^+, \forall k \in \mathbb{Z}_{T-1}^0, \\ \forall \vec{\boldsymbol{x}}_i^0, y_i(k), d_i(k), w_i(k), v_i(k) : \\ (1)\text{-}(3),\ (5),\ (7)\text{-}(9)\ \textit{hold}; \\ \forall k \in \mathbb{Z}_{T-1}^{t+1} : (10)\ \textit{holds}; \\ \forall i \in \mathbb{Z}_N^+, \forall k \in \mathbb{Z}_t^0 : z_i(k) = z_m(k); \\ \forall k \in \mathbb{Z}_{t-1}^0 : u_T^t(k) = u_T^{*,t-1}(k) \end{array}\right\} : \begin{array}{l} \{\forall k \in \mathbb{Z}_T^+ : \\ (6)\ \textit{holds}\} \wedge \\ \{\forall i, j \in \mathbb{Z}_N^+, i \neq j, \\ \exists k \in \mathbb{Z}_{T-1}^0 : \\ |z_i(k) - z_j(k)| \geq \epsilon\}. \end{array} \tag{11b}$$

Note that the constraint (11b) not only enforces that constraint of the controlled state is satisfied but also guarantees that all models are separated using the input $u_T^{*,t}$. Moreover, we only compute the sequence of u_T^t up until $t = T-2$

because we consider the output separation constraint (11b) up to time $T-1$ and hence, it is not meaningful to further improve $u(T-1)$ that can only affect the output at time T.

IV. Main Approach

In this section, we propose a partition-based approach to solve the output feedback active model discrimination defined in Problem 1. To solve Problem 1 offline, the optimization formulated in (11) would need to be solved with $z_m(k)$ for all $k \in \mathbb{Z}_t^0$ as yet unknown outputs. However, solving (11) could be computationally expensive using current multi-parametric optimization toolboxes, especially when there exists a large number of binary variables. To tackle this issue, we propose a more tractable approach by partitioning the measurement space and solving the active model discrimination problem for each partition. Although partitioning the measurement space leads to a slightly conservative solution and marginally reduces the benefit of using output measurement, we can move the computation offline without the need for multi-parametric optimization and make the process more practical and tractable.

A. Partition Tree

Recall the measurement domain $\mathcal{Z}$ (cf. (10) for its definition) of the measurement $z_m(k)$. At each time instant $t \in \mathbb{Z}_{T-1}^0$, let $\{\mathcal{Z}_{s(t)}\}_{s(t)\in\mathbb{N}_{\ell_t}^+}$ be a *partition* of $\mathcal{Z}$ (cf. Definition 1) with ℓ_t subregions. Each subregion $\mathcal{Z}_{s(t)}$ is a polyhedral set:

$$\mathcal{Z}_{s(t)} = \{z \in \mathbb{R}^{n_z} : P_{z,s(t)} z \le p_{z,s(t)}\}. \tag{12}$$

As a result, any newly revealed measurement $z_m(t) \in \mathcal{Z}_{s(t)}$ at time instant $t \in \mathbb{Z}_{T-2}^0$ can be over-approximated by $z_m(t) = z(t)$ with $z(t) \in \mathcal{Z}_{s(t)}$. Then, we use a partition tree to capture all possible combinations of partitions of the measurement space over the time interval of interest.

Definition 2 (Partition Tree)**.** *A partition tree $\mathcal{T}$ is a collection of nodes connected by directed edges, where each node contains a partition of the whole domain. The node at the top of the tree is called the root node. Every node (excluding the root node) is connected by a directed edge from exactly one other node with the direction "parent $\rightarrow$ children". The node which does not have any child node is called the leaf node. The depth γ of a node is the number of edges from the root to the node. The depth γ is associated with the time instant t, satisfying the relation $\gamma = t+1$. A trajectory/path $\mathcal{P}$ refers to the sequence of nodes (excluding the root node) along the edges of a tree from the root node to a leaf node.*

Example 1. *Fig. 1 shows the partition tree over time instant $t \in \mathbb{Z}_{T-2}^0$ with $T=3$, where the measurement space is partitioned into 2 subregions at each time step, i.e., $\ell_t = 2$, $s(t) \in \{1,2\}$ for all $t \in \mathbb{Z}_2^0$ and $\mathcal{Z}_{s(t)} \in \{\mathcal{Z}_1, \mathcal{Z}_2\}$. Since no output measurement is available before the initial time instant ($t=-1$), the root is the entire domain and the corresponding separating input $u_T^{-1}(\cdot) = \emptyset$. At time instant $t = 0$, the measurement domain $\mathcal{Z}$ is partitioned into two subregions $\mathcal{Z}_1$ and $\mathcal{Z}_2$. As a result, the root has two children, i.e., two nodes with depth $\gamma = 1$, each of which represents a subregion. Moving to the next time instant $t = 1$ of interest, we partition the measurement domain again into two subregions, which means that each node with depth $\gamma = 1$ has two children, resulting in 4 nodes with depth $\gamma = 2$ in total, which are leaf nodes as they do not have any child. Moreover, it is clear from Fig. 1 that there are 4 different trajectories/paths in this partition tree with $\mathcal{P}_1 = \{\mathcal{Z}_1, \mathcal{Z}_1\}$, $\mathcal{P}_2 = \{\mathcal{Z}_1, \mathcal{Z}_2\}$, $\mathcal{P}_3 = \{\mathcal{Z}_2, \mathcal{Z}_1\}$ and $\mathcal{P}_4 = \{\mathcal{Z}_2, \mathcal{Z}_2\}$.*

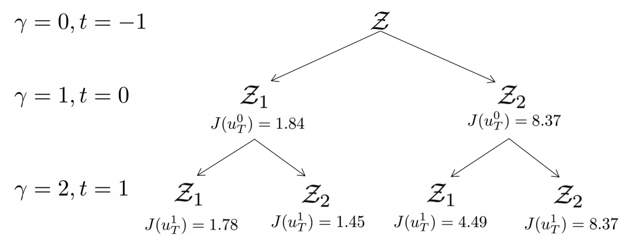


Fig. 1: Partition tree for the output measurement z_m over time instants $t = 0$ and $t = 1$.

Problem 2 (Partition-Based Output Feedback Model Discrimination)**.** *For each trajectory of the partition tree given by $\{\mathcal{Z}_{s(k)}\}_{k=0}^{T-1}$, the output feedback active model discrimination problem in Problem 1 can be reformulated as a sequence of optimization problems as follows (for $t = 0, \ldots, T-2$):*

$$u_T^{*,t} = \arg\min_{u_T^t} J(u_T^t)$$

$$\textit{s.t.} \quad \forall k \in \mathbb{Z}_{T-1}^0 : (4) \textit{ holds}, \tag{13a}$$

$$\left.\begin{array}{r} \forall i \in \mathbb{Z}_{\mathrm{N}}^+, \forall k \in \mathbb{Z}_{T-1}^0, \\ \forall \vec{x}_i^0, y_{i,T}, d_{i,T}, w_{i,T}, v_{i,T} : \\ \text{(1)-(3), (5), (7)-(9) } \textit{hold}; \\ \forall k \in \mathbb{Z}_{T-1}^{t+1} : \text{(10) } \textit{holds}; \\ \forall i \in \mathbb{Z}_N^+, \forall k \in \mathbb{Z}_t^0 : z_i(k) \in \mathcal{Z}_{s(k)}; \\ \forall k \in \mathbb{Z}_{t-1}^0 : u_T^t(k) = u_T^{*,t-1}(k) \end{array}\right\} : \begin{array}{l} \{\forall k \in \mathbb{Z}_T^+ : \\ \text{(6) } \textit{holds}\} \wedge \\ \{\forall i, j \in \mathbb{Z}_N^+, i \ne j, \\ \exists k \in \mathbb{Z}_{T-1}^0 : \\ |z_i(k) - z_j(k)| \ge \epsilon\}, \end{array} \tag{13b}$$

where $u_T^{*,t-1}$ is the optimal input sequence from the previous time instant $t-1$. Note that any pair of trajectories that share the same node at time instant t on the partition tree, their optimal input subsequences up to t must be the same. The total number of optimization problems (corresponding to the nodes excluding the root node) is $\sum_{i=0}^{T-2} \prod_{t=0}^{i} \ell_t$.

Comparing with Problem 1, the equality constraints with measurement $z_m(k)$ for all $k \in \mathbb{Z}_t^0$ are replaced by its over-approximation in Problem 2, which corresponds to the subregion that the revealed measurement $z_m(k)$ lies in. Moreover, since this enlarges the uncertainty set of Problem 1 and thus, shrinks its feasibility set, it is straightforward to see that the solution of Problem 2 also solves Problem 1, albeit with a slight loss of optimality.

B. Time-Concatenated Model

Before proceeding with the main approach, we introduce some time-concatenated notations and write the considered N models in a time-concatenated form. The time-concatenated states and outputs are defined as

$$\vec{x}_{i,T} = \mathrm{vec}_{k=0}^{T}\{\vec{x}_i(k)\}, \quad x_{i,T} = \mathrm{vec}_{k=0}^{T}\{x_i(k)\},$$
$$y_{i,T} = \mathrm{vec}_{k=0}^{T}\{y_i(k)\}, \quad z_{i,T} = \mathrm{vec}_{k=0}^{T-1}\{z_i(k)\},$$

while the time-concatenated inputs and noises are defined as

$$\vec{u}_{i,T} = \mathrm{vec}_{k=0}^{T-1}\{\vec{u}_i(k)\}, u_T^t = \mathrm{vec}_{k=0}^{T-1}\{u^t(k)\},$$
$$d_{i,T} = \mathrm{vec}_{k=0}^{T-1}\{d_i(k)\}, w_{i,T} = \mathrm{vec}_{k=0}^{T-1}\{w_i(k)\},$$
$$v_{i,T} = \mathrm{vec}_{k=0}^{T}\{v_i(k)\}.$$

In addition, for $0 \leq j_1 \leq j_2 \leq T-1$, we also define the measurement sequence as $z_{m,j_1:j_2} = [z_m(j_1)^\top, \ldots, z_m(j_2)^\top]^\top \in \mathbb{R}^{(j_2-j_1+1)n_z}$. Therefore, at each time instant $t \in \mathbb{Z}_{T-1}^0$ with the past and current revealed output measurements $\{z_m(k)\}_{k=0}^t$ lying in subregions $\{\mathcal{Z}_{s(k)}\}_{k=0}^t$, we have revealed output measurements $z_{m,0:t} = \mathrm{vec}_{k=0}^t\{z_m(k)\}$, where $z_m(k) \in \mathcal{Z}_{s(k)}, \forall k \in \mathbb{Z}_t^0$, and unrevealed output measurements $z_{m,t+1:T-1} = \mathrm{vec}_{k=t+1}^{T-1}\{z_m(k)\}$, where $z_m(k) \in \mathcal{Z}, \forall k \in \mathbb{Z}_{T-1}^{t+1}$. The time-concatenated measurement vector over the entire horizon is further defined as $z_{m,T} = \begin{bmatrix} z_{m,0:t}^\top & z_{m,t+1:T-1}^\top \end{bmatrix}^\top$.

Given N discrete-time affine models, there are $I = \binom{N}{2}$ model pairs and let the mode $\iota \in \{1, \cdots, I\}$ denote the pair of models (i,j). Then, concatenating each model pair yields

$$\vec{x}_0^\iota = \mathrm{vec}_{i,j}\{\vec{x}_i^0\},\ \vec{x}_T^\iota = \mathrm{vec}_{i,j}\{\vec{x}_{i,T}\},\ \vec{u}_T^\iota = [u_T^\top, d_T^{\iota\top}]^\top,$$
$$x_T^\iota = \mathrm{vec}_{i,j}\{x_{i,T}\},\ y_T^\iota = \mathrm{vec}_{i,j}\{y_{i,T}\},\ z_T^\iota = \mathrm{vec}_{i,j}\{z_{i,T}\},$$
$$d_T^\iota = \mathrm{vec}_{i,j}\{d_{i,T}\},\ w_T^\iota = \mathrm{vec}_{i,j}\{w_{i,T}\},\ v_T^\iota = \mathrm{vec}_{i,j}\{v_{i,T}\}.$$

The states and outputs over the entire time horizon for each mode ι can be written as simple functions of the initial state $\vec{x}_0^\iota$, inputs u_T^ι, d_T^ι, parameter p_T and noises w_T^ι, v_T^ι:

$$x_T^\iota = M_x^\iota \vec{x}_0^\iota + \Gamma_{xu}^\iota u_T^t + \Gamma_{xd}^\iota d_T^\iota + \Gamma_{xw}^\iota w_T^\iota + \tilde{f}_x^\iota, \tag{14}$$
$$y_T^\iota = M_y^\iota \vec{x}_0^\iota + \Gamma_{yu}^\iota u_T^t + \Gamma_{yd}^\iota d_T^\iota + \Gamma_{yw}^\iota w_T^\iota + \tilde{f}_y^\iota, \tag{15}$$
$$\vec{x}_T^\iota = \bar{A}^\iota \vec{x}_0^\iota + \Gamma_u^\iota u_T^t + \Gamma_d^\iota d_T^\iota + \Gamma_w^\iota w_T^\iota + \tilde{f}^\iota, \tag{16}$$
$$z_T^\iota = \bar{C}^\iota \vec{x}_0^\iota + \bar{D}_u^\iota u_T^t + \bar{D}_d^\iota d_T^\iota + \bar{D}_v^\iota v_T^\iota + \tilde{g}^\iota. \tag{17}$$

The matrices and vectors $M_\star^\iota$, $\Gamma_{\star u}^\iota$, $\Gamma_{\star d}^\iota$, $\Gamma_{\star w}^\iota$ and $\tilde{f}_\star^\iota$ for $\star \in \{x,y\}$, and $\bar{A}^\iota$, Γ_u^ι, Γ_d^ι, Γ_ω^ι, $\bar{C}^\iota$, $\bar{D}_u^\iota$, $\bar{D}_d^\iota$, $\bar{D}_v^\iota$, $\tilde{f}\iota$, $\tilde{g}^\iota$ can be found in the Appendix of [8].

Moreover, the uncertainties for each mode ι are concatenated as $\bar{x}^\iota = \begin{bmatrix} \vec{x}_0^{\iota\top} & d_T^{\iota\top} & w_T^{\iota\top} & v_T^{\iota\top} \end{bmatrix}^\top$. We then concatenate state constraints in (6) and (7) and eliminate x_T and y_T in them and expressing them in terms of $\bar{x}^\iota$ and u_T. First, let

$$\bar{P}_x^\iota = \mathrm{diag}_{i,j}\, \mathrm{diag}_T\{P_{x,i}\}, \quad \bar{P}_y^\iota = \mathrm{diag}_{i,j}\, \mathrm{diag}_T\{P_{y,i}\},$$
$$\bar{p}_x^\iota = \mathrm{vec}_{i,j}\, \mathrm{vec}_T\{p_{x,i}\}, \quad \bar{p}_y^\iota = \mathrm{vec}_{i,j}\, \mathrm{vec}_T\{p_{y,i}\}.$$

Then, we can rewrite the polyhedral constraints as:

$$\bar{P}_\star^\iota x_T^\iota \leq \bar{p}_\star^\iota \Leftrightarrow H_\star^\iota \bar{x}^\iota \leq h_{\star,t}^\iota(u_T^t),\ \star \in \{x,y\}$$

where $H_\star^\iota = \bar{P}_\star^\iota \begin{bmatrix} M_\star^\iota & \Gamma_{\star d}^\iota & \Gamma_{\star w}^\iota & 0 \end{bmatrix}$ and $h_\star^\iota(u_T^t) = \bar{p}_\star^\iota - \bar{P}_\star^\iota \Gamma_{\star u}^\iota u_T^t - \bar{P}_\star^\iota \tilde{f}_\star^\iota$. Similarly, let

$$\bar{Q}_u = \mathrm{diag}_T\{Q_u\},\ \ \bar{Q}_\dagger^\iota = \mathrm{diag}_{i,j}\, \mathrm{diag}_T\{Q_{\dagger,i}\},$$
$$\bar{q}_u = \mathrm{vec}_T\{q_u\},\ \ \bar{q}_\dagger^\iota = \mathrm{vec}_{i,j}\, \mathrm{vec}_T\{q_{\dagger,i}\},\ \dagger \in \{d,w,v\}.$$

Then, the uncertainties and input constraints in (4)-(5) and (8)-(9) over the entire horizon are written as $\bar{Q}_u u_T^t \leq \bar{q}_u$ and $\bar{Q}_\dagger^\iota \dagger_T^\iota \leq \bar{q}_\dagger^\iota$. As the available measurements $\{z_m(k)\}_{k=0}^t$ are located in subregions $\{\mathcal{Z}_{s(k)}\}_{k=0}^t$, $z_{i,0:t}$ satisfies

$$\bar{P}_{z,s(0:t)} z_{i,0:t} \leq \bar{p}_{z,s(0:t)},$$

where $\bar{P}_{z,s(0:t)} = \mathrm{diag}_{k=0}^t\{P_{z,s(k)}\}$ and $\bar{p}_{z,s(0:t)} = \mathrm{vec}_{k=0}^t\{p_{z,s(k)}\}$. Due to the fact that outputs $z_{i,t+1:T-1}$ are unrevealed/unmeasured at the time instant t, we also have

$$\bar{P}_{z,t+1:T-1} z_{i,t+1:T-1} \leq \bar{p}_{z,t+1:T-1}, \tag{18}$$

where $\bar{P}_{z,t+1:T-1} = \mathrm{diag}_{T-t-1}\{P_z\}$ and $\bar{p}_{z,t+1:T-1} = \mathrm{vec}_{T-t-1}\{p_z\}$. Then, the constraint on $z_{i,T}$ is obtained as

$$\bar{P}_z^t z_{i,T} \leq \bar{p}_z^t,$$

with $\bar{P}_z^t = \begin{bmatrix} \bar{P}_{z,s(0:t)} & 0 \\ 0 & \bar{P}_{z,t+1:T-1} \end{bmatrix}$ and $\bar{p}_z^t = \begin{bmatrix} \bar{p}_{z,s(0:t)} \\ \bar{p}_{z,t+1:T-1} \end{bmatrix}$. Then, with $\bar{P}_z^{t,\iota} = \mathrm{diag}_2\{\bar{P}_z^t\}$, $\bar{p}_z^{t,\iota} = \mathrm{vec}_2\{\bar{p}_z^t\}$ and $z_{i,T}^\iota = \mathrm{vec}_{i,j}\{z_{i,T}\}$, we concatenate the above constraint on the output measurement and express it in terms of $\bar{x}^\iota$, leading to

$$\bar{P}_z^{t,\iota} z_T^\iota \leq \bar{p}_z^{t,\iota} \Leftrightarrow H_z^{t,\iota} \bar{x}^\iota \leq h_z^{t,\iota}(u_T^t), \tag{19}$$

where $H_z^{t,\iota} = \bar{P}_z^{t,\iota} \begin{bmatrix} \bar{C}^{\iota\top} & \bar{D}_d^{\iota\top} & 0 & \bar{D}_v^{\iota\top} \end{bmatrix}^\top$ and $h_z^{t,\iota}(u_T^t) = \bar{p}_z^{t,\iota} - \bar{P}_z^{t,\iota} \bar{D}_u^\iota u_T^t - \bar{P}_z^{t,\iota} \tilde{g}^\iota$.

Moreover, we concatenate the initial state constraint in (3):

$$\bar{P}_0^\iota = \mathrm{diag}_2\{P_0\}, \quad \bar{p}_0^\iota = \mathrm{vec}_2\{p_0\}.$$

Hence, in terms of $\bar{x}^\iota$, we have a polyhedral constraint of the form $H_{\bar{x}}^\iota \bar{x}^\iota \leq h_{\bar{x}}^\iota$ for each time $t \in \mathbb{Z}_{T-1}^0$, with $H_{\bar{x}}^\iota = \mathrm{diag}\{\bar{P}_0, \bar{Q}_d^\iota, \bar{Q}_w^\iota, \bar{Q}_v^\iota\}$ and $h_{\bar{x}}^\iota = \begin{bmatrix} \bar{p}_0^{\iota\top} & \bar{q}_d^{\iota\top} & \bar{q}_w^{\iota\top} & \bar{q}_v^{\iota\top} \end{bmatrix}^\top$.

C. Active Model Discrimination Approach

Next, having introdcued the partition tree and some useful matrix definitions, we are ready to solve Problem 2 (and hence, Problem 1). We begin by deriving the following lemma that enables us to relax a rather limiting assumption in [4], [8], [12] that the uncontrolled state constraints are not affected by the separating input.

Lemma 1 (Semi-Infinite Constraint Reformulation)**.** *The following semi-infinite constraint*

$$\mathcal{A}\chi \leq \mathcal{b} + \mathcal{C}y, \forall \chi \in \mathcal{X} \triangleq \{\chi : \mathcal{D}\chi \leq \mathcal{e} + \mathcal{F}y, \mathcal{G}\chi \leq \mathcal{h}\}$$

with variables χ and y is equivalent to

$$p_a \leq 0,\ \nu_{1,a} \geq 0,\ \nu_{2,a} \geq 0, \tag{20a}$$
$$\forall b: 0 = \textstyle\sum_i (\nu_{1,a})_{(i)} \mathcal{D}(i,b) + \sum_j (\nu_{2,a})_{(j)} \mathcal{G}(j,b) - \mathcal{A}(a,b), \tag{20b}$$
$$\begin{aligned} &p_a = \mathcal{A}_{(a)}\chi_a - \mathcal{b}_{(a)} - \mathcal{C}_{(a)}y, \\ &\mathcal{D}\chi_a \leq \mathcal{e} + \mathcal{F}y,\ \ \mathcal{G}\chi_a \leq \mathcal{h}, \end{aligned} \tag{20c}$$
$$\begin{aligned} &\forall i: \ \textit{SOS-1}: \{(\nu_{1,a})_{(i)}, \mathcal{D}_{(i)}\chi_a - \mathcal{e}_{(i)} - \mathcal{F}_{(i)}y\}, \\ &\forall j: \ \textit{SOS-1}: \{(\nu_{2,a})_{(j)}, \mathcal{G}_{(j)}\chi_a - \mathcal{h}_{(j)}\}, \end{aligned} \tag{20d}$$

for all $a \in \mathbb{Z}_{n_a}^+$, where n_a is the number of rows of $\mathcal{A}$. We denote as $\mathcal{A}_{(a)}$, $\mathcal{b}_{(a)}$ and $\mathcal{C}_{(a)}$ the a-th rows of $\mathcal{A}$, $\mathcal{b}$ and $\mathcal{C}$, respectively, while χ_a, $\nu_{1,a}$ and $\nu_{2,a}$ are additional slack variables for each a.

Proof. To obtain the result in Lemma 1, we introduce a slack variable p_a for each row of $\mathcal{A}_{(a)}\chi \leq \mathcal{b}_{(a)} + \mathcal{C}_{(a)}y$ and convert the semi-infinite constraint into $p_a^* \leq 0$, where p_a^* is the maximum p_a satisfying the constraints in $\mathcal{X}$, i.e.,

$$\begin{aligned} p_a^* = \arg &\min_{p_a, \chi_a} -p_a \\ &\text{s.t. } \mathcal{D}\chi_a \leq \mathcal{e} + \mathcal{F}y, \mathcal{G}\chi_a \leq \mathcal{h}, \\ &\quad\ \ p_a = \mathcal{A}_{(a)}\chi_a - \mathcal{b}_{(a)} - \mathcal{C}_{(a)}y. \end{aligned}$$

Then, applying KKT conditions and rewriting the complementary slackness constraints as SOS-1 constraints, we obtain (20a)-(20d). □

By leveraging Lemma 1, we recast the optimization in Problem 2 to an MILP with SOS-1 constraints, which can readily be solved using off-the-shelf optimization software tools, e.g., Gurobi and CPLEX [13], [14].

Theorem 1 (Partition-Based Output Feedback Discriminating Input Design as a Sequence of MILP)**.** *For each time instant $t \in \mathbb{Z}_{T-1}^0$, given a separability index ϵ and a trajectory on the partition tree corresponding to subregions $\{\mathcal{Z}_{s(k)}\}_{k=0}^{T-1}$, the partition-based output feedback active* *model discrimination problem (Problem 2) is equivalent to a sequence of MILP problems (for $t = 0, \dots, T-2$):*

$$u_T^{*,t} = \arg \min_{u_T^t, \delta^\iota, \bar{x}^\iota, \mu_1^\iota, \mu_2^\iota, \nu_{1,a}^\iota, \nu_{2,a}^\iota} J(u_T^t) \qquad (P_{POFDID})$$

$$\text{s.t.} \quad \bar{Q}_u u_T^t \leq \bar{q}_u, \tag{21a}$$

$$\forall k \in \mathbb{Z}_{t-1}^0 : \; u_T^t(k) = u_T^{*,t-1}(k), \tag{21b}$$

$$\forall \iota \in \mathbb{Z}_I^+, \; \forall a \in \mathbb{Z}_{n_\mathcal{A}}^+ :$$

$$\begin{cases} p_a^\iota \leq 0, \; \nu_{1,a}^\iota \geq 0, \; \nu_{2,a}^\iota \geq 0, \\ \forall b \in \mathbb{Z}_{m_\mathcal{A}}^+ : 0 = \sum_{i=1}^{n_\mathcal{D}} (\nu_{1,a}^\iota)_{(i)} \Psi_t^\iota(i,b) \\ \qquad\qquad + \sum_{j=1}^{n_\mathcal{G}} (\nu_{2,a}^\iota)_{(j)} \Psi_t^\iota(j+n_\mathcal{D}, b) - (H_x^\iota)(a,b), \\ p_a^\iota = (H_x^\iota)_{(a)} \bar{x}_a^\iota - (h_x^\iota(u_T^t))_{(a)}, \\ \Psi_t^\iota \bar{x}^\iota \leq \psi_t^\iota(u_T^t), \\ \forall i \in \mathbb{Z}_{n_\mathcal{D}}^+ : \textit{SOS-1} : \{(\nu_{1,a}^\iota)_{(i)}, (\Psi_t^\iota)_{(i)} \bar{x}_a^\iota - (\psi_t^\iota)_{(i)}\}, \\ \forall j \in \mathbb{Z}_{n_\mathcal{G}}^+ : \textit{SOS-1} : \{(\nu_{2,a}^\iota)_{(j)}, (\Psi_t^\iota)_{(j+n_\mathcal{D})} \bar{x}_a^\iota - (\psi_t^\iota)_{(j+n_\mathcal{D})}\}, \end{cases} \tag{21c}$$

$$\forall \iota \in \mathbb{Z}_I^+ :$$

$$\begin{cases} \delta^\iota \geq \epsilon, \; \mu_1^\iota \geq 0, \; \mu_2^\iota \geq 0, \; 0 = 1 - {\mu_1^\iota}^\top \mathbb{1}, \\ \forall \kappa \in \mathbb{Z}_\eta^+ : 0 = \sum_{i=1}^{\xi} (\mu_1^\iota)_{(i)} \Phi_t^\iota(i,\kappa) \\ \qquad\qquad + \sum_{j=1}^{\rho} (\mu_2^\iota)_{(j)} \Phi_t^\iota(j+\xi, \kappa), \\ \forall i \in \mathbb{Z}_\xi^+ : (\Phi_t^\iota)_{(i)} \bar{x}^\iota - \delta^\iota - (\phi_t^\iota(u_T^t))_{(i)} \leq 0, \\ \forall j \in \mathbb{Z}_\rho^+ : (\Phi_t^\iota)_{(\xi+j)} \bar{x}^\iota - (\phi_t^\iota(u_T^t))_{(\xi+j)} \leq 0, \\ \forall i \in \mathbb{Z}_\xi^+ : \; \textit{SOS-1} : \{\mu_{1,i}^\iota, (\Phi_t^\iota)_{(i)} \bar{x}^\iota - \delta^\iota - (\phi_t^\iota(u_T^t))_{(i)}\}, \\ \forall j \in \mathbb{Z}_\rho^+ : \; \textit{SOS-1} : \{\mu_{2,j}^\iota, (\Phi_t^\iota)_{(\xi+j)} \bar{x}^\iota - (\phi_t^\iota(u_T^t))_{(\xi+j)}\}, \end{cases} \tag{21d}$$

where $\nu_{1,a}^\iota$, $\nu_{2,a}^\iota$, μ_1^ι, μ_2^ι are slack variables, $(\bullet)_{(i)}$ denotes the i-th row of a matrix/vector, $(\bullet)(i,j)$ denotes (i,j)-th entry of a matrix, $u_T^{*,t-1}$ is the optimal input that has been determined/computed from time $t-1$, and all the remaining matrices and constants will be defined in the proof below.

Proof. Since the universal quantifier distributes over conjunction [15, pp. 45–46], we can separate the constraint (13b) of Problem 2 into two independent constraints for all possible values of uncertain variables at time instant $t \in \mathbb{Z}_{T-1}^0$, i.e., the controlled state constraint and the separability condition, respectively. For the controlled state constraint (6) for each time $t \in \mathbb{Z}_{T-1}^0$ can be equivalently written as:

$$\forall k \in \mathbb{Z}_{t-1}^0 : \; u_T^t(k) = u_T^{*,t-1}(k), \tag{22}$$

$$\forall \bar{x}^\iota \in \left\{ \bar{x}^\iota \,\middle|\, \Psi_t^\iota \bar{x}^\iota \leq \psi_t^\iota(u_T^t) \right\} : H_x^\iota \bar{x}^\iota \leq h_{x,t}^\iota(u_T^t), \tag{23}$$

where

$$\Psi_t^\iota = \begin{bmatrix} H_{\bar{x}}^\iota \\ H_y^\iota \\ H_{z,t}^\iota \end{bmatrix}, \; \psi_t^\iota(u_T^t) = \begin{bmatrix} h_{\bar{x}}^\iota \\ \bar{p}_y^\iota - \bar{P}_y^\iota \tilde{f}^\iota - \bar{P}_y^\iota \Gamma_{y\underline{u}}^\iota u_T^t \\ \bar{p}_z^{t,\iota} - \bar{P}_z^{t,\iota} \tilde{g}^\iota - \bar{P}_z^{t,\iota} \bar{D}_u^\iota u_T^t \end{bmatrix}.$$

Thus, by leveraging Lemma 1, (23) can be converted to (21b) and (21c) with the following corresponding matrices: $\mathcal{A} \triangleq H_x^\iota$, $\mathcal{X} \triangleq \bar{x}^\iota$, $\mathcal{Y} \triangleq u_T^t$, $\mathcal{B} = \bar{p}_x^\iota - \bar{P}_x^\iota \tilde{f}^\iota$, $\mathcal{C} = -\bar{P}_x^\iota \Gamma_{xu}^\iota$, $\mathcal{D} = \begin{bmatrix} H_y^{\iota\top} & H_{z,t}^{\iota\top} \end{bmatrix}^\top$, $\mathcal{E} \triangleq \begin{bmatrix} (\bar{p}_y^\iota - \bar{P}_y^\iota \tilde{f}^\iota)^\top & (\bar{p}_z^{t,\iota} - \bar{P}_z^{t,\iota} \tilde{g}^\iota)^\top \end{bmatrix}^\top$, $\mathcal{F} \triangleq \begin{bmatrix} -(\bar{P}_y^\iota \Gamma_{yu}^\iota)^\top & -(\bar{P}_z^{t,\iota} \bar{D}_u^\iota)^\top \end{bmatrix}^\top$, $\mathcal{G} = H_{\bar{x}}^\iota$ and $\mathcal{h} = h_{\bar{x}}^\iota$. In addition, in (21c), $n_\mathcal{A}$, $n_\mathcal{D}$ and $n_\mathcal{G}$ are the number of rows of $\mathcal{A}$, $\mathcal{D}$ and $\mathcal{G}$, $m_\mathcal{A}$ is the number of columns of $\mathcal{A}$.

Moreover, concatenating the separation condition with time for each model pair $\iota \in \mathbb{Z}_I^+$ at time instant $t \in \mathbb{Z}_{T-1}^0$ and considering its double negation equivalence [8, Lemma 1], the separation condition in Problem 2 is recast as

$$\delta^{\iota *}(u_T^t) \geq \epsilon, \tag{24a}$$

$$\delta^{\iota *}(u_T^t) = \min_{\delta^\iota, \bar{x}^\iota} \delta^\iota$$

$$\text{s.t.} \quad \Phi_t^\iota \bar{x}^\iota - \begin{bmatrix} \mathbb{1} \\ 0 \end{bmatrix} \delta^\iota \leq \phi_t^\iota(u_T^t), \tag{24b}$$

$$\forall k \in \mathbb{Z}_{t-1}^0 : \; u_T^t(k) = u_T^{*,t-1}(k), \tag{24c}$$

where

$$\Phi_t^\iota = \begin{bmatrix} \Lambda^\iota \\ H_{\bar{x}}^\iota \\ H_y^\iota \\ H_{z,t}^\iota \end{bmatrix}, \; \phi_t^\iota(u_T^t) = \begin{bmatrix} -\bar{E}^\iota \tilde{f}^\iota - (\bar{E}^\iota \Gamma_u^\iota + \overline{F}_u^\iota) u_T^t \\ h_{\bar{x}}^\iota \\ \bar{p}_y^\iota - \bar{P}_y^\iota \tilde{f}^\iota - \bar{P}_y^\iota \Gamma_{y\underline{u}}^\iota u_T^t \\ \bar{p}_z^{t,\iota} - \bar{P}_z^{t,\iota} \tilde{g}^\iota - \bar{P}_z^{t,\iota} \bar{D}_u^\iota u_T^t \end{bmatrix},$$

and matrices Λ^ι, $\bar{E}^\iota$, $\tilde{f}^\iota$, Γ_u^ι and $\overline{F}_u^\iota$ can be found in the Appendix of [8]. We also denote η as the number of columns of Φ_t^ι, ξ as the number of rows of $\mathbb{1}$ (i.e., constraints that depend on δ^ι) in (24b) and ρ as the number of rows of 0 (i.e., constraints that are independent of δ^ι) in (24b).

Then, applying KKT conditions to the above minimization and rewriting the complementary slackness constraints using SOS-1, we obtain the constraints (21d) in Problem (P_{POFDID}). This completes the proof. □

Remark 1. *To identify the true model from a set of potential models actively, two components, i.e., active model discrimination and model invalidation, are required (e.g., [8], [16]), which are implemented offline and online, respectively. Since the proposed active model discrimination approach can also take output measurements into account when selecting the optimal separating output feedback inputs at run time, we can concurrently implement the output feedback active model discrimination and model invalidation algorithms at run time for enhancing model discrimination. Specifically, we can precompute the output feedback active model discrimination offline where the partition tree is augmented to include branches that corresponding to both the power set of yet invalidated models and the partitions of the output space. Then, at each time instant t, we can first implement model invalidation using all available measurements to narrow down the model set, before selecting the optimal separating input u_T^t from the augmented partition tree based on the reduced model set and the partition corresponding to the current output measurement.*

V. Simulation Example

The proposed active model discrimination approach is applied to the highway lane-changing scenario to identify the intention of other road participants. As in [8], a two-car system consisting of an ego car and an human-driven car is considered, whose discrete-time equations of motion are:

$$\begin{aligned} x_e(k+1) &= x_e(k) + v_{x,e}(k)\delta t, \\ v_{x,e}(k+1) &= v_{x,e}(k) + u_{x,e}(k)\delta t + w_{x,e}(k)\delta t, \\ y_e(k+1) &= y_e(k) + v_{y,e}(k)\delta t + w_{y,e}(k)\delta t, \\ x_o(k+1) &= x_o(k) + v_{x,o}(k)\delta t, \\ v_{x,o}(k+1) &= v_{x,o}(k) + d_i(k)\delta t + w_{x,o}(k)\delta t, \end{aligned}$$

where, respectively, x_e and y_e, and $v_{x,e}$ and $v_{y,e}$ are the ego car's longitudinal and lateral positions, and the ego car's longitudinal and lateral velocities, x_o and $v_{x,o}$ are the human-driven car's longitudinal position and longitudinal velocity, while $u_{x,e}$ and d_i are ego car and human-driven car's acceleration inputs, $w_{x,e}$, $w_{x,e}$ and $w_{x,e}$ are process noise signals and δt is the sampling time. In addition, the system's output is the longitudinal velocity of the human-driven car in the form of $z(k) = v_{x,o}(k) + v(k)$, where $v(k)$ is the output noise. In addition, we assume that the initial position of the ego car is 0, and the initial position of the other car is constrained by their initial relative distance. The initial velocities of the cars are also constrained to match typical speed limits of the highway, and at the beginning, both cars are close to the center of the lanes. In this case, the initial conditions are:

$$v_{x,e}(0) \in [28, 30]\ (\tfrac{m}{s}), \quad y_e(0) \in [1.1, 1.8]\ (m),$$
$$v_{x,o}(0) \in [28, 32]\ (\tfrac{m}{s}), \quad x_o(0) \in [7, 12]\ (m).$$

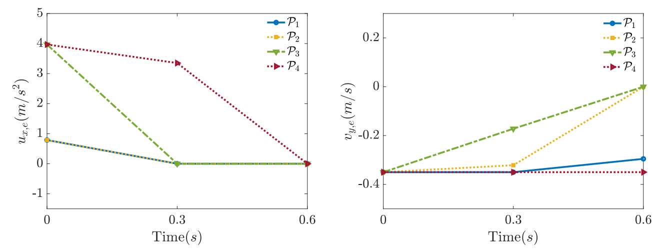
In the simulation, we consider three driver intentions [8]: **Inattentive Driver** ($i = \mathrm{I}$), who fails to notice the ego car and maintains his driving speed, thus proceeding with an acceleration input in a small range $d_I(k) \in \mathcal{D}_I \equiv 10\% \cdot \mathcal{U}$. **Cautious Driver** ($i = \mathrm{C}$), who tends to yield the lane to the ego car with the input equal to $K_{d,C}(v_{x,e}(k) - v_{x,o}(k)) - L_{p,C}(\bar{y} - y_e(k)) + L_{d,C} v_{y,e}(k) + d_C(k)$, where $K_{d,C} = -0.9$, $L_{p,C} = 2.5$ and $L_{d,C} = 8.9$ are PD controller parameters, $\bar{y} = 2$ and the input uncertainty is $d_C(k) \in \mathcal{D}_C \equiv 5\% \cdot \mathcal{U}$. **Malicious Driver** ($i = \mathrm{M}$), who does not want to yield the lane and attempts to cause a collision with input equal to $K_{d,M}(v_{x,e}(k) - v_{x,o}(k)) - L_{p,M}(\bar{y} - y_e(k)) + L_{d,M} v_{y,e}(k) + d_M(k)$, if provoked, where $K_{d,M} = 1.1$, $L_{p,M} = -2.3$ and $-L_{d,M} = 8.7$ are PD controller parameters, $\bar{y} = 2$ and the input uncertainty satisfies $d_M(k) \in \mathcal{D}_M \equiv 5\% \cdot \mathcal{U}$.

In this example, the time horizon is $T = 3$ with the sampling time $\delta t = 0.3\ (s)$. The controlled inputs are $u_{x,e}(k) \in \mathcal{U}_x \equiv [-7.85, 3.97]\ (\frac{m}{s^2})$ and $v_{y,e}(k) \in \mathcal{U}_y \equiv [-0.35, 0]\ (\frac{m}{s})$ (where y is in the direction away from the other lane), and process and output noises are bounded within $[-0.01, 0.01]$. We also constrain the ego car's longitudinal velocity as $x_e(k) \in [27, 35]\ (\frac{m}{s})$, while its lateral position is constrained as $y_e(k) \in [0.5, 2]\ (m)$. The threshold in the separability condition is chosen to $\epsilon = 0.3\ (\frac{m}{s})$. Note that there is a trade-off between the separation threshold and the magnitude of the computed separating input for a prescribed time. A larger threshold in the separability condition results in better discrimination capability (i.e., more obvious model separation) but a larger separating input (i.e., more perturbations to the original system behavior).

Moreover, we assume that the measurement domain over the entire time horizon is $\mathcal{Z} = [28, 32]\ (\frac{m}{s})$. At each time instant, we partition this domain into two subregions, i.e., $\mathcal{Z}_1 = [28, 30]$ and $\mathcal{Z}_2 = [30, 32]$. As a result, the partition tree for this example is the same as that shown in Fig. 1, containing 4 different trajectories. It is observed from Fig. 1 that the associated cost at each node (excluding the root node) decreases monotonically along any particular trajectory. This shows that incorporating output measurements into active model discrimination can improve its performance. Finally, the separating input sequences under the proposed approach for all 4 trajectories in the partition tree are shown in Fig. 2.

VI. Conclusion

In this paper, we considered the output feedback input design problem for discriminating among a finite set of discrete-time affine models subject to uncontrolled inputs, noises and uncertain initial conditions over a fixed time horizon. We take advantage of output measurements at run time to refine the computed separating input with a monotonically reduced cost as time goes on. Since obtaining the separating input at each time instant as a function of yet unknown outputs can be computationally expensive, we partitioned the measurement domain and constructed a partition tree. Then, for each node (excluding the root) in the partition tree, we computed an input that is guaranteed to separate all models from each other within the fixed time horizon. As a result, we can move the optimization computation offline and only determine which node the measurement belongs to at each time instant. Finally, we applied the proposed approach to infer intentions of vehicles in a lane changing scenario.



(a) Input $u_{x,e}$ (b) Input $v_{y,e}$

Fig. 2: u_T^t at the leaf node of different trajectories $\mathcal{P}_1$-$\mathcal{P}_4$.

References

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri. A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Comp. & Chem. Eng.*, 27(3):293–311, 2003.

[2] L. H. Chiang, E. L. Russell, and R. D. Braatz. *Fault Detection and Diagnosis in Industrial Systems*. Springer-Verlag London, 2001.

[3] H. Lou and P. Si. The distinguishability of linear control systems. *Nonlinear Analysis: Hybrid System*, 3(1):21–39, Nov. 2009.

[4] F. Harirchi, S. Z. Yong, E. Jacobsen, and N. Ozay. Active model discrimination with applications to fraud detection in smart buildings. *IFAC-PapersOnLine*, 50(1):9527–9534, 2017.

[5] S. Cheong and I. R. Manchester. Input design for discrimination between classes of LTI models. *Automatica*, 53:103–110, Mar. 2015.

[6] R. Nikoukhah and S. Campbell. Auxiliary signal design for active failure detection in uncertain linear systems with a priori information. *Automatica*, 42(2):219–228, Feb. 2006.

[7] J. K. Scott, R. Findeison, R. D. Braatz, and D. M. Raimondo. Input design for guaranteed fault diagnosis using zonotopes. *Automatica*, 50(6):1580–1589, Jun. 2014.

[8] 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 (ICCPS)*, pages 5161–5167, 2018.

[9] 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*, 74:107–117, Dec. 2016.

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

[11] J. A. Paulson, T. A. N. Heirung, R. D. Braatz, and A. Mesbah. Closed-loop active fault diagnosis for stochastic linear systems. In *American Control Conference (ACC)*, pages 735–741. IEEE, 2018.

[12] R. Niu, Q. Shen, and S. Z. Yong. Partition-based parametric active model discrimination with applications to driver intention estimation. In *European Control Conference (ECC)*, pages 3880–3885. IEEE, 2019.

[13] Gurobi Optimization, Inc. Gurobi optimizer reference manual, 2015.

[14] IBM ILOG CPLEX. V12. 1: User's manual for CPLEX. *International Business Machines Corporation*, 46(53):157, 2009.

[15] K. H. Rosen. *Discrete Mathematics and Its Applications*. New York: McGraw-Hill, 2011.

[16] V. S. Chipade, Q. Shen, L. Huang, N. Ozay, S. Z. Yong, and D. Panagou. Safe autonomous overtaking with intention estimation. In *European Control Conference (ECC)*, pages 2050–2057. IEEE, 2019.