On Finite Memory Approximations Constructed from Input/Output Snapshots

Danielle C. Tarraf and Luis A. Duffaut Espinosa

Abstract— The problem of approximating systems with finite input and output alphabets by finite memory systems for verification or certified control has received much deserved attention in the recent past. The present paper is a further step in that direction, building upon a robust control inspired notion of approximation we recently proposed. A constructive algorithm for deriving deterministic finite state machine (DFM) approximations of a given system over finite alphabets is proposed, based on a partitioning of its input/output behavior into equivalence classes of finite length snapshots. The algorithm is analyzed, and the resulting nominal models and corresponding approximation errors are shown to have desirable properties. An algorithm for conservatively quantifying the resulting approximation error in a manner consistent with the objective of control synthesis is also proposed. Several simple illustrative examples are presented to demonstrate the approach.

I. INTRODUCTION

Modern control systems are increasingly complex due to the confluence of nonlinearities [9], quantization effects [5], [6], [11], [21], [2], and interacting analog and discrete (hybrid) dynamics [4], [28]. This complexity, coupled with stringent performance objectives and the need for verifiably safe systems, gives rise to a new set of analysis and synthesis challenges. Lower complexity models that are more easily amenable to analysis and optimization, and that adequately abstract or approximate the complex systems in question, provide a potential means for addressing these new challenges. Accordingly, the problem of finding lower complexity models of hybrid systems with quantized inputs and outputs, particularly *finite state* models (which are tractable in principle though not always practically so!), has received much attention in the recent past [1], [34], [3], [35], [36].

One direction of active research in this area, inspired by the theory of formal verification in computer science [20], [17], aims to construct finite deterministic abstractions of the underlying hybrid system that are related to the original system by a simulation or bisimulation relation [8], or an approximation thereof [26], [27], [7], [22]. Since simulations and bisimulations fundamentally relate the underlying state-spaces of the plant and the abstraction, construction of the approximate model typically involves partitioning or otherwise covering the state-space of the original system. A second direction of active research in this area, inspired by

the behavioral approach to systems in control theory, aims to derive non-deterministic finite state models whose output or input/output behavior contains that of the original system [13], [10], [14], [15], [18], [19], [16].

In our past work, we proposed an approach for deriving and quantifying deterministic finite state machine (DFM) approximations for a class of hybrid systems, namely switched second order homogeneous systems under binary sensing constraints [31], [33]. We also developed a streamlined minimax optimization based synthesis approach, thus demonstrating the use of these approximations for synthesizing certified-by-design finite memory controllers for the switched plant. Construction of the approximate models in this setting involved constructing finite coverings of the state-space of the plant meeting certain criteria. More recently, we formally proposed a robust control inspired notion of control oriented finite state approximation for plants with discrete actuation and sensing [29], of which the work in [31], [33] is a specific instance. The objective of the present contribution is to explore an alternative procedure for deriving control oriented DFM approximations by explicitly partitioning the plant's input/output behavior over finite length time horizons.

The idea of using finite length snapshots of inputs and outputs to generate models of dynamical systems is of course widely used in system identification [12]. However, the underlying assumptions in the setting of interest to us here are very different from those in a typical system identification setup. Specifically: (i) the dynamics of the system are assumed to be exactly known, (ii) the input and output data is uncorrupted by noise, and (iii) complete data of length L can be generated for any choice of integer L.

In related work, an approach for constructing finite behavioral abstractions of systems by associating the states of the nominal models with finite length sequences of input/output behavior of the original system (assumed to have quantized inputs and outputs) was proposed in [23] and later refined in subsequent developments [19], [16]. The approach appears to have been inspired by Willems' approach for computing realizations of l-complete systems [37]. The resulting non-deterministic finite state automaton can then be used in conjunction with a synthesis procedure [18] inspired from the Ramadge/Wonham framework [24], [25] to synthesize supervisory controllers for the plant. While the idea of associating the states of the nominal model with finite length input/output snapshots is utilized in this paper, our contribution is fundamentally different from the above referenced work in the considered class of nominal models, the notion of approximation employed, as well as the proposed approach

D. C. Tarraf is with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA (dtarraf@jhu.edu).
L. A. Duffaut Espinosa was a postdoctoral fellow in the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA at the time of initial submisssion.

for describing and quantifying the quality of the approximate model. Likewise, our earlier work addressing a first order case study [30] fundamentally differs from the contributions of the present paper in the notion of approximation employed as well as the structures of the approximate models and corresponding approximation errors.

Organization of the Paper: We begin in Section II by reviewing some basic concepts and highlight the definition of approximation of interest to us. We formally state the problem considered and summarize our approach and contributions in Section III. We present the details of the proposed construction of the deterministic finite state machine (DFM) model in Section IV. We analyze the properties of the resulting approximate model and those of the corresponding approximation error in Section V. We propose a computationally efficient approach for computing an upper bound on the gain of the approximation error in Section VI. We end with simple illustrative examples in Section VII and we conclude with a brief discussion of directions for future work in VIII.

Notation: A word on notation: $\mathbb{Z}_+$ and $\mathbb{R}_+$ denote the set of non-negative integers and non-negative reals, respectively. Given a set $\mathcal{A}$, $\mathcal{A}^{\mathbb{Z}_+}$ denotes the set of all infinite sequences (indexed by $\mathbb{Z}_+$) over $\mathcal{A}$. An element of $\mathcal{A}$ is denoted by a while an element of $\mathcal{A}^{\mathbb{Z}_+}$ is denoted by (boldface) $\mathbf{a}$. For $\mathbf{a} \in \mathcal{A}^{\mathbb{Z}_+}$, $a(i)$ denotes the ith term. For $f : A \to B$ and subset C of B, $f^{-1}(C) = \{a \in A | f(a) \in C\}$. The cardinality of a finite set $\mathcal{A}$ is denoted by $|\mathcal{A}|$ while its power set is denoted by $\mathcal{A}^2$.

II. PRELIMINARIES

We begin by reviewing some background that will be needed in our development. Interested readers are referred to [32], [29] for an in-depth treatment of the various notions. A (discrete-time) signal is understood to be an infinite sequence over some prescribed "alphabet" set, while a (discrete-time) system S is simply a set of pairs of feasible signals, $S \subset \mathcal{U}^{\mathbb{Z}_+} \times \mathcal{Y}^{\mathbb{Z}_+}$, where $\mathcal{U}$ and $\mathcal{Y}$ are given input and output alphabets. System properties of interest are captured by means of 'integral' constraints on the feasible signals.

Definition 1. *Consider a system $S \subset \mathcal{U}^{\mathbb{Z}_+} \times \mathcal{Y}^{\mathbb{Z}_+}$ and let $\rho : \mathcal{U} \to \mathbb{R}$ and $\mu : \mathcal{Y} \to \mathbb{R}$ be given functions. S is ρ/μ gain stable if there exists a finite non-negative constant γ such that the following inequality is satisfied for all $(\mathbf{u}, \mathbf{y})$ in S:*

$$\inf_{T \geq 0} \sum_{t=0}^{T} \gamma \rho(u(t)) - \mu(y(t)) > -\infty. \qquad (1)$$

In particular, when ρ and μ are non-negative (and not identically zero), a notion of 'gain' can be defined in the usual manner, meaning as the infimum of γ such that (1) is satisfied.

We are typically interested in discrete-time plants that interact with their controllers through fixed discrete alphabets (i.e. plants with finite-valued actuators and sensors), which we refer to as 'systems over finite alphabets'. More precisely,

a *system over finite alphabets* S is a discrete-time system $S \subset (\mathcal{U} \times \mathcal{R})^{\mathbb{Z}_+} \times (\mathcal{Y} \times \mathcal{V})^{\mathbb{Z}_+}$ whose alphabets $\mathcal{U}$ and $\mathcal{Y}$ are finite. $\mathbf{r} \in \mathcal{R}^{\mathbb{Z}_+}$, $\mathbf{u} \in \mathcal{U}^{\mathbb{Z}_+}$, $\mathbf{v} \in \mathcal{V}^{\mathbb{Z}_+}$, and $\mathbf{y} \in \mathcal{Y}^{\mathbb{Z}_+}$ represent the exogenous input, the control input, the performance output and the sensor output of the plant, respectively. No restrictions are placed a-priori on the internal dynamics of the plant: The underlying state-space may be analog, discrete or hybrid. Likewise, alphabets $\mathcal{R}$ and $\mathcal{V}$ may be finite, countably infinite or simply infinite in general. We will use the notation $S|_{\mathbf{u_o}, \mathbf{y_o}}$ to denote the subset of feasible signals of S defined by

$$S|_{\mathbf{u_o}, \mathbf{y_o}} = \Big\{ (\mathbf{u}, \mathbf{y}) \Big) \in S \Big| \mathbf{u} = \mathbf{u_o} \text{ and } \mathbf{y} = \mathbf{y_o} \Big\}.$$

Note that $S|_{\mathbf{u_o}, \mathbf{y_o}}$ may be an empty set or an infinite set for specific choices of $\mathbf{u_o}$ and $\mathbf{y_o}$.

The nominal models of interest to us are deterministic finite state machine (DFM) models, which are simply discrete-time systems with *finite* input and output alphabets and whose feasible signals are related by

$$\begin{aligned} q(t+1) &= f(q(t), u(t)) \\ y(t) &= g(q(t), u(t)) \end{aligned}$$

for some functions f and g and some *finite* state set $\mathcal{Q}$, $q(t) \in \mathcal{Q}$.

The underlying assumption generally is that the purpose of deriving a DFM approximation of a system over finite alphabets P is to simplify the process of synthesizing a controller K such that the closed loop system (P, K) is ρ/μ gain stable with $\gamma = 1$ for some given functions ρ and μ. The notion of approximation of interest to us is given by Definition 2.
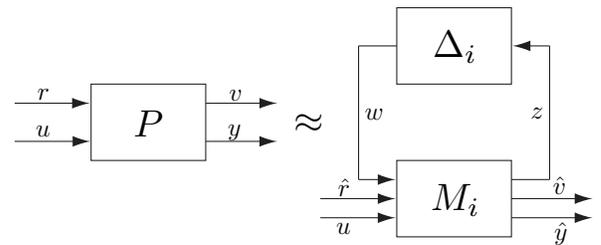
Fig. 1. A finite state approximation of P

Definition 2 (Notion of DFM Approximation [29]). *Consider a system over finite alphabets $P \subset (\mathcal{U} \times \mathcal{R})^{\mathbb{Z}_+} \times (\mathcal{Y} \times \mathcal{V})^{\mathbb{Z}_+}$ and a desired closed loop performance objective*

$$\inf_{T \geq 0} \sum_{t=0}^{T} \rho(r(t)) - \mu(v(t)) > -\infty \qquad (2)$$

for given functions $\rho : \mathcal{R} \to \mathbb{R}$ and $\mu : \mathcal{V} \to \mathbb{R}$. A sequence $\{M_i\}$ ($i \in \mathbb{Z}_+$) of deterministic finite state machines

$$M_i \subset (\mathcal{U} \times \hat{\mathcal{R}}_i \times \mathcal{W})^{\mathbb{Z}_+} \times (\hat{\mathcal{Y}} \times \hat{\mathcal{V}}_i \times \mathcal{Z})^{\mathbb{Z}_+}$$

*with $\hat{\mathcal{R}}_i \subset \mathcal{R}$ and $\hat{\mathcal{V}}_i \subset \mathcal{V}$ is a $\boldsymbol{\rho/\mu}$ **approximation** of P if there exists a corresponding sequence of systems $\{\Delta_i\}$,*

$\Delta_i \subset \mathcal{Z}^{\mathbb{Z}^+} \times \mathcal{W}^{\mathbb{Z}_+}$, and non-zero functions $\rho_\Delta : \mathcal{Z}^{\mathbb{Z}^+} \to \mathbb{R}_+$ and $\mu_\Delta : \mathcal{W}^{\mathbb{Z}_+} \to \mathbb{R}_+$, such that for every i:

1) *There exists a surjective map $\psi_i : P \to \hat{P}_i$ satisfying*

$$P|_{\mathbf{u_o}, \mathbf{y_o}} \subseteq \psi_i^{-1}\left(\hat{P}_i|_{\mathbf{u_o}, \mathbf{y_o}}\right)$$

for all $(\mathbf{u_o}, \mathbf{y_o}) \in \mathcal{U}^{\mathbb{Z}_+} \times \mathcal{Y}^{\mathbb{Z}_+}$, where

$$\hat{P}_i \subset (\mathcal{U} \times \hat{\mathcal{R}}_i)^{\mathbb{Z}_+} \times (\hat{\mathcal{Y}} \times \hat{\mathcal{V}}_i)^{\mathbb{Z}_+}$$

is the feedback interconnection of M_i and Δ_i as shown in Figure 1.

2) *For every feasible signal $\left((\mathbf{u_o}, \hat{\mathbf{r}}_\mathbf{o})(\hat{\mathbf{y}}_\mathbf{o}, \hat{\mathbf{v}}_\mathbf{o})\right) \in \hat{P}_i$, and every corresponding feasible signal $\left((\mathbf{u}, \mathbf{r}), (\mathbf{y}, \mathbf{v})\right) \in \psi_i^{-1}(\{\left((\mathbf{u_o}, \hat{\mathbf{r}}_\mathbf{o}), (\hat{\mathbf{y}}_o, \hat{\mathbf{v}}_\mathbf{o})\right)\})$, we have*

$$\rho(r(t)) - \mu(v(t)) \geq \rho(\hat{r}_o(t)) - \mu(\hat{v}_o(t)), \ \forall t \in \mathbb{Z}_+ \quad (3)$$

3) *Δ_i is ρ_Δ / μ_Δ gain stable, and moreover, the corresponding ρ_Δ / μ_Δ gains satisfy $\gamma_i \geq \gamma_{i+1}$.*

Intuitively, condition (1) of Definition 2 establishes a 1-1 correspondence between certain equivalence classes of input/output behaviors of the plant and the approximate model, which is needed for certified-by-design control synthesis. Condition (2) reflects our need to approximate the performance specifications in general, as the plant may have analog input while the nominal model may not. Condition (3) reflects our need to approximate the dynamics of the plant, and to quantify the quality of approximation in a manner consistent with control design. Interested readers are referred to [29] for the details of this notion of approximation and its significance.

III. SETUP OF THE PROBLEM

A. Problem Statement

In this paper, we consider a simplified setup in order to focus on the problem of *approximating the plant dynamics* while momentarily overlooking the problem of approximating the performance objectives. Specifically, we consider plants with a single input channel, namely the control input, and a single output channel, namely the sensor output. Moreover, we will assume that a state-space model of the plant P is available, given by

$$\begin{aligned} x(t+1) &= f(x(t), u(t)) \\ y(t) &= g(x(t)) \end{aligned}$$

where $t \in \mathbb{Z}_+$, $x(t) \in \mathbb{R}^n$, $u(t) \in \mathcal{U}$ and $y(t) \in \mathcal{Y}$, $\mathcal{U}$ and $\mathcal{Y}$ are given finite alphabets with cardinality m and p, respectively. Alphabets $\mathcal{U}$ and $\mathcal{Y}$ may represent quantized values of analog inputs and outputs, or may simply be symbolic inputs and outputs in general. Functions $f : \mathbb{R}^n \times \mathcal{U} \to \mathbb{R}^n$ and $g : \mathcal{R}^n \to \mathcal{Y}$ are given.

Our first objective is to construct a sequence of DFM models $\{M_i\}$ and a corresponding sequence of approximation errors $\{\Delta_i\}$ that approximate plant P, in the sense of satisfying properties (1) and (3) of Definition 2. Our second objective is to quantify the quality of approximation, by quantifying the gain of the approximation error.

B. Overview of Main Contributions

The contributions of the paper are as follows:

1) We propose a constructive procedure for generating the approximate models and the corresponding approximation error. The proposed approach associates the states of the approximate model at every time instant with the past inputs and outputs of the system over a finite time horizon of fixed length L. The structure of the corresponding approximation error is inspired from the 'observer-like' structure first proposed in [31], [33]. The details of this contribution are presented in Section IV.

2) We analyze the proposed algorithm to highlight some of its features, as well as the positive features of the resulting approximate models. The details of this analysis are presented in Section V.

3) We propose a tractable, LP based algorithm for verifying an upper bound on the gain of the resulting approximation error system. The details and development of this algorithm are presented in Section VI.

IV. CONSTRUCTION OF THE DFM APPROXIMATION

A. Internal Structure of M_i and Δ_i
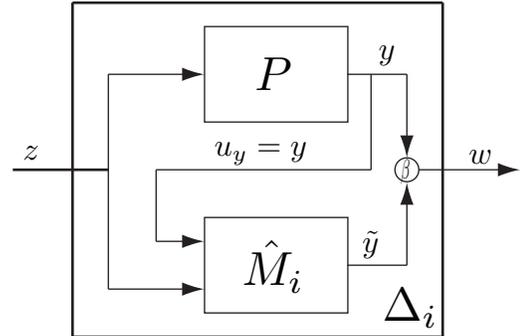
Fig. 2.   The internal structure of Δ_i

We will construct a sequence of DFM systems $\{M_i\}$ and a corresponding sequence of error systems $\{\Delta_i\}$ with the internal structures shown in Figures 3 and 2, respectively, for each index i. Alphabet $\mathcal{Z}$ in this construct simply equals the control input alphabet $\mathcal{U}$. Functions $\beta : \mathcal{Y} \times \mathcal{Y} \to \mathcal{W}$ and $\alpha : \mathcal{W} \times \mathcal{Y} \to \mathcal{Y}$ are chosen to satisfy

$$\alpha\left(\beta(y, \tilde{y}), \tilde{y}\right) = y \quad (4)$$

for all choices of y, $\tilde{y}$ in $\mathcal{Y}$. Note that since $\mathcal{Y}$ is a finite alphabet, construction of such maps is always possible provided $\mathcal{W}$ (which is a design parameter) is chosen to have sufficiently large cardinality. Also note that the approximation error Δ_i is an *infinite* memory system in general. Moreover, its internal structure is meaningful in a *control* setting as system $\hat{M}_i$ will typically be used an a finite memory *observer* for the plant as detailed in [33], and hence has access to both the control input and the sensor output. The only a priori constraint placed on DFM system $\hat{M}_i$ is that it is not allowed

to admit direct feedthrough from its input u_y to its output $\tilde{y}$, otherwise the proposed setup becomes meaningless.
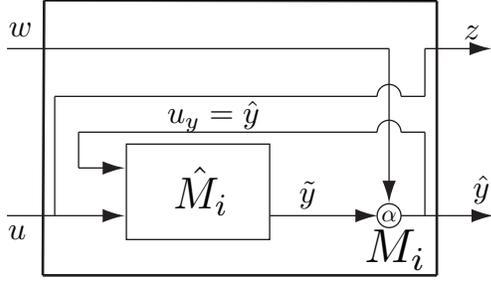
Fig. 3.   The internal structure of M_i

B. Construction of System $\hat{M}_i$

For each index i, system $\hat{M}_i$ is a DFM with two inputs, namely $u_y \in \mathcal{Y}$ and $u \in \mathcal{U}$, and a single output, namely $\tilde{y} \in \mathcal{Y}$. Letting $L = i$, the intention is to construct a deterministic transition system $\hat{M}_i$ whose state at time $t \geq L$ effectively keep track of the past L inputs and outputs of the plant, as well as the current output of $\hat{M}_i$ at time t, namely:

$$q(t) = \begin{pmatrix} \tilde{y}(t) \\ y(t-1) \\ \vdots \\ y(t-L) \\ u(t-1) \\ \vdots \\ u(t-L) \end{pmatrix}$$

Introduce the notation $f_u(x)$ as shorthand for $f(u, x)$, and $f_{u_1 \circ u_2}$ as shorthand for $f_{u_1} \circ f_{u_2}$, the construction of $\hat{M}_i$ proceeds as follows:

Step1 [Construction of the Set of Feasible States] By definition, a state

$$q = (\tilde{y}, y_1, \ldots, y_L, u_1, \ldots, u_L)' \in \mathcal{Y}^{L+1} \times \mathcal{U}^L$$

is feasible if there exists a state $x_o \in \mathbb{R}^n$ satisfying the following set of $L + 1$ equalities:

$$
\begin{aligned}
y_L &= g\big(x_o\big) \\
y_{L-1} &= g\big(f_{u_L}(x_o)\big) \\
y_{L-2} &= g\big(f_{u_{L-1} \circ u_L}(x_o)\big) \\
\vdots &= \vdots \\
y_1 &= g\big(f_{u_2 \circ \ldots \circ u_L}(x_o)\big)
\end{aligned}
\tag{5}
$$

and

$$\tilde{y} = g\big(f_{u_1 \circ \ldots \circ u_L}(x_o)\big) \tag{6}$$

Intuitively, a state q is a feasible state of the machine if there exists an initial state x_o of the plant with corresponding output y_L such that, under input sequence $\{u_L, u_{L-1}, \ldots, u_2, u_1\}$ we observe output sequence

$\{y_{L-1}, y_{L-2}, \ldots, y_1, \hat{y}\}$. We will denote the set of all feasible states by $\mathcal{Q}_f$.

Step 2 [Feasible Transitions] Consider the map

$$\overline{F} : \mathcal{Q}_f \times \mathcal{U} \times \mathcal{Y} \to \mathcal{Q}_f^2$$

(where $\mathcal{Q}_f^2$ denotes the power set of $\mathcal{Q}_f$) defined as follows: Letting $q_a = (\tilde{y}_a, y_{1,a}, \ldots, y_{L,a}, u_{1,a}, \ldots, u_{L,a})'$ and $q_b = (\tilde{y}_b, y_{1,b}, \ldots, y_{L,b}, u_{1,b}, \ldots, u_{L,b})'$ denote two states in $\mathcal{Q}_f$, $q_b \in \overline{F}(q_a, u, y)$ iff $u_{1,b} = u$, $y_{1,b} = y$, $u_{k,b} = u_{k-1,a}$ and $u_{k,b} = u_{k-1,a}$ for $k \in \{2, \ldots, L\}$. Note that for any choice of (q_a, u, y), the cardinality of $\overline{F}(q_a, u, y)$ ranges from 0 to p: Cardinality 0 corresponds to the case where the plant simply cannot exhibit a particular input/output signal pair of length $L + 1$. We therefore introduce a new state $q_\emptyset$ of the machine, we let $\mathcal{Q}_e = \mathcal{Q}_f \cup \{q_\emptyset\}$, and we construct a new map

$$F : \mathcal{Q}_e \times \mathcal{U} \times \mathcal{Y} \to \mathcal{Q}_f^2 \cup \{q_\emptyset\}$$

by defining

$$
\begin{aligned}
F(q_\emptyset, u, y) &= \{q_\emptyset\}, \\
F(q, u, y) &= \begin{cases} \{q_\emptyset\} & \text{when } |\overline{F}(q, u, y| = 0 \\ \overline{F}(q, u, y) & \text{otherwise} \end{cases}
\end{aligned}
$$

Step 3 [Initial States]: For any $t < L$, the plant can only memorize $t - 1$ past input sequences. As such, it is necessary to define an additional set of *initial* states. Considering equivalence relations $R_1, \ldots R_{L-1}$ on $\mathcal{Q}_f$ defined by

$$q_a R_k q_b \Leftrightarrow u_{i,a} = u_{i,b}, y_{i,a} = y_{i,b} \text{ for } 1 \leq i \leq k$$

we can construct equivalence classes $\mathcal{E}_k = \mathcal{Q}_f \setminus R_k$, with $\mathcal{E}_0$ denoting the equivalence class of all feasible states. We can now define a set $\mathcal{Q}_0$ of initial states to be the union of $\mathcal{E}_0, \ldots, \mathcal{E}_{L-1}$. We arbitrarily associate with every state $q \in \mathcal{Q}_0$ an output $\tilde{y} \in \mathcal{Y}$, since the choices we make here only influence the input/output behavior of the finite state model for the first $L - 1$ time steps, and hence are of limited impact. We extend the transition map F to this set of states in the (unique !) natural way.

Step 4 [Construction of the DFM Model]: We can now construct a DFM $\hat{M}_i$ whose states are given by

$$\hat{\mathcal{Q}}_i = \mathcal{Q}_0 \cup \mathcal{Q}_e$$

whose state transition function

$$\hat{f}_i : \mathcal{Q} \times \mathcal{U} \times \mathcal{Y} \to \mathcal{Q}$$

is arbitrarily defined provided

$$\hat{f}_i(q, u, y) \in F(q, u, y),$$

and whose output function

$$\hat{g}_i : \hat{\mathcal{Q}}_i \to \mathcal{Y}$$

is defined by

$$\hat{g}_i(q) = \tilde{y} \text{ when } q = (\tilde{y}, y_1, \ldots, y_L, u_1, \ldots, u_L)' \in \mathcal{Q}_f$$

and arbitrarily fixed otherwise.

C. Remarks on the Proposed Algorithm

Remark 1. *Notice that for any choice of index i, the algorithm described in Section IV-B leads to a finite number of possible systems $\hat{M}_i$, usually more than one. As such, the algorithm does not result in a unique sequence of $\{M_i\}$, but rather an infinite number of such choices.*

Proposition 1. *Let N be the cardinality of the state space of $\hat{M}_i$. N is bounded by:*

$$1 + p\left[\frac{m - m^{L+1}}{1 - m}\right] \leq N \leq 1 + \frac{pm - p^L m^L}{1 - pm} + p^{L+1} m^L$$

Proof: The state space of $\hat{M}_i$ is the union of $\{q_\emptyset\}$, the feasible states $\mathcal{Q}_f$, and the equivalence classes $\mathcal{E}_0, ..., \mathcal{E}_{L-1}$ encoding the initial states. A lower bound can be derived for $\mathcal{Q}_f$ by observing that each control sequence of length L and each initial state, hence initial output, have to appear at least once. We thus conclude that $pm^L \leq |\mathcal{Q}_f| \leq p^{L+1} m^L$. Similar observations allow us to conclude that $m^k p \leq |\mathcal{E}_k| \leq m^k p^k$ for $k \in \{2, \ldots, L-1\}$. Finally, noting that $|\mathcal{E}_0| = |\{q_\emptyset\}| = 1$ and combining these bounds, we get the formula presented. $\square$

This observation, particularly the lower bound, suggest that this approach may only be computationally efficient when the cardinality of the input set is relatively small. However, this statement is tempered by the caveat that it may be possible to aggregate the states of the approximate model at a later stage.

V. PROPERTIES OF THE APPROXIMATE SYSTEMS

Now associate with each state $q \in \hat{\mathcal{Q}}_i$ two subsets of $\mathbb{R}^n$, X_q^o and X_q^+, defined as follows:

- When $q \in \mathcal{Q}_f$, $q = (\tilde{y}, y_1, \ldots, y_L, u_1, \ldots, u_l)$, define:

$$X_q^o = \left\{x \in \mathbb{R}^n | x \text{ satisfies (5)}\right\}$$

and

$$X_q^+ = f_{u_1 \circ \ldots \circ u_L} X_q^o$$

- When $q \in \mathcal{E}_1, ..., \mathcal{E}_k$, the sets X_q^o and X_q^+ can be similarly defined by simply considering the appropriate subset of equality constraints and appropriately truncated input sequences, respectively.
- When $q \in \mathcal{E}_o$, set $X_q^o = X_q^+ = \mathbb{R}^n$
- When $q = q_\emptyset$, $X_q^o = X_q^+ = \emptyset$

Proposition 2. *Consider the interconnection of $\hat{M}_i$ and P as in Figure 2, and let $x(t)$ and $q(t)$ denote the state of P and $\hat{M}_i$, respectively, at time t. We have $x(t) \in X_{q(t)}^+$, for all $t \geq 0$.*

Proof: By construction, $X_{q(0)}^+ = \mathbb{R}^n$ and hence $x(0) \in X_{q(0)}^+$. Moreover by construction, if $x(t) \in X_{q(t)}^+ \Rightarrow x(t+1) \in X_{q(t+1)}^+$. The proof thus follows by induction on t. $\square$

This simple observation substantiates our previous statement that $\hat{M}_i$ may be used as a finite memory (set-valued) observer for the plant.

Theorem 1. *For any DFM M_i and corresponding approximation error Δ_i constructed as described in Section IV there exists a $\psi_i : P \to \hat{P}_i$, where $\hat{P}_i$ denotes the feedback interconnection of M_i and Δ_i, such that ψ_i is surjective and $\psi_i^{-1}(\hat{P}_i|_{\mathbf{u},\mathbf{y}}) \supseteq P|_{\mathbf{u},\mathbf{y}}$.*

Proof: The proof is constructive. Fix index i. Consider $\psi_1 : P \to \hat{M}_i$ defined by $\psi_1(\mathbf{u}, \mathbf{y}) = ((\mathbf{u}, \mathbf{y}), \tilde{\mathbf{y}}) \in \hat{M}_i$, where $\tilde{\mathbf{y}}$ is the *unique* (initial conditions is fixed!) output of $\hat{M}_i$ to inputs $\mathbf{u}$ and $\mathbf{u_y} = \mathbf{y}$.

Also consider $\psi_2 : \psi_1(P) \to \hat{P}_i$ defined by

$$\psi_2\Big(((\mathbf{u}, \mathbf{y}), \tilde{\mathbf{y}}\Big) = (\mathbf{u}, \mathbf{y})$$

Let $\psi_i = \psi_2 \circ \psi_1$. ψ_i is surjective, since ψ_2 is surjective and $\psi_2^{-1}(\hat{P}_i) = \psi_1(P)$ by definition. Moreover, $\psi_i(P|_{\mathbf{u},\mathbf{y}}) \subseteq \hat{P}_i|_{\mathbf{u},\mathbf{y}}$:

$$\psi_i(P|_{\mathbf{u},\mathbf{y}}) = \psi_2(\psi_1(P|_{\mathbf{u},\mathbf{y}})) \subseteq \psi_2(\psi_1(P)|_{\mathbf{u},\mathbf{y}}) \subseteq \hat{P}_i|_{\mathbf{u},\mathbf{y}}$$

Hence, $P_{\mathbf{u},\mathbf{y}} \subseteq \psi_i^{-1}(\hat{P}_i|_{\mathbf{u},\mathbf{y}})$. $\square$

VI. QUANTIFYING THE APPROXIMATION ERROR

Ideally, we would like to be able to derive a priori bounds on the gain of the error system, much like what is done in traditional model order reduction schemes such a balanced truncation or Hankel reduction. What we provide in this section is an approach for computing upper bounds on the gain of the error system Δ_i, for any given choice of functions ρ_Δ and μ_Δ. The relevance of this result stems from the fact that Δ_i is generally an *infinite* memory system with a hybrid (analog-discrete) state space. Hence direct computation of the gain is difficult in general.

We will consider the scenario where we are interested in computing an upper bound for the ρ_Δ/μ_Δ gain of the approximation error system Δ_i for some choice of functions $\rho_\Delta : \mathcal{U} \to \mathbb{R}_+$ and $\mu_\Delta : \mathcal{Y} \to \mathbb{R}_+$ (not identically zero). We will mark a special element of $\mathcal{W}$, namely w_o. Without loss of generality, we will assume that $\beta(y, y') = w_o$ iff $y = y'$. Note that this assumption does not prevent us from defining a pair of memoryless function α and β satisfying (4). Further, we will assume that $\mu_\Delta(w) = 0$ iff $w = w_o$. The rationale is that in quantifying Δ_i, we are typically interested in penalizing the mismatched outputs of $\hat{M}_i$ and P, and only the mismatched outputs.

Under these assumptions, consider the map

$$d : \hat{\mathcal{Q}}_i \times \mathcal{U} \times \mathcal{Y} \to \mathbb{R}_+$$

defined by

$$\delta(q, u, y) = \max_{q^+ \in F(q,u,y)} \mu_\Delta\left[\beta\Big(\tilde{y}(q^+), \tilde{y}(\hat{f}_i(q, u, y))\Big)\right]$$

Note that by definition, $\delta(q, u, y) = 0$ whenever $|F(q, u, y)| = 1$, and in particular when $q = q_\emptyset$ or $q \in \mathcal{E}_k$ for some k.

Theorem 2. *If $\hat{M}_i$ satisfies the inequality constraint:*

$$\inf_{T \geq 0} \sum_{t=0}^{T} \gamma \rho_\Delta(u(t)) - \delta(q(t), u(t), y(t)) > -\infty \quad (7)$$

for some $\gamma > 0$, then the corresponding Δ_i is ρ_Δ/μ_Δ gain stable with gain bound not exceeding γ.

Proof: Let $x(t)$ and $y(t)$ denote the state and output trajectories, respectively, of P under input $u(t)$, and let $q(t)$ be the corresponding state trajectory of $\hat{M}_i$, interconnected with P as shown in Figure 2. By Proposition 2, $x(t) \in X^+_{q(t)}$ for all $t \geq 0$. Now, when $t \geq L$: If $|F(q(t), u(t), y(t)| = 1$, there is no ambiguity in the transition or outputs. Hence the outputs of P and $\hat{M}_i$ must match, and

$$\mu_\Delta(w(t+1)) = 0 = d(q(t), u(t), y(t))$$

Otherwise, $y(t + 1) = \tilde{y}(x_o)$ for some $x_o \in F(q(t), u(t), y(t))$, and we thus have

$$\begin{aligned}
&\mu_\Delta(w(t+1)) \\
= \ &\mu_\Delta\Big[\beta(\tilde{y}(x_o), \tilde{y}(\hat{f}_i q(t), u(t), y(t))\Big] \\
\leq \ &\max_{q^+ \in F(q,u,y)} \mu_\Delta\Big[\beta\Big(\tilde{y}(q^+), \tilde{y}(\hat{f}_i(q(t), u(t), y(t)))\Big)\Big] \\
= \ &d(q(t), u(t), y(t))
\end{aligned}$$

Summing up along state trajectories from $t = 0$ to T and taking the infimum, we thus get

$$\inf_{T \geq 0} \sum_{t=0}^{T} \gamma\rho_\Delta(u(t)) - \mu_\Delta(w(t))$$

$$\geq C + \inf_{T \geq 0} \sum_{t=0}^{T} \gamma\rho_\Delta(u(t)) - \delta(q(t), u(t), w(t))$$

where C is a *finite* constant that accounts for transient effects ($t < L$) and for the 1 step time shift in the summation. Hence, if $\hat{M}$ satisfies equation (7), then Δ is ρ_Δ/μ_Δ gain stable with gain at most equal to γ. $\square$

The problem of verifying an 'integral constraint' for a deterministic finite state machine was shown to be tractable in [32]: The problem reduces to the search for an appropriate storage function, which in turn can be formulated as linear program. Computing an upper bound for the gain for the gain of Δ_i thus reduces to solving a linear program in $N + 1$ decision variables, where N is the number of states of the approximate model M_i. In practice, the computational burden may be alleviated by first appropriately aggregating the states of $\hat{M}_i$ prior to running the linear program.

VII. ILLUSTRATIVE EXAMPLES

A. LTI System with Input and Output Quantizers

We begin by revisiting the simple case study we addressed in [30]. Consider a discrete-time, analog-state system P consisting of a stable first order system described by

$$\begin{cases} x(t+1) = ax(t) + u(t) \\ y(t) = Q(x(t)) \end{cases} \quad (8)$$

The input of the system is restricted to take on integer values, $u(t) \in \mathcal{U} = \{-1, 0, 1\}$. Q is a given uniform quantizer with

quantization step $l = 2$:

$$Q : y_{(t)} = \begin{cases} \vdots \\ -l & \text{if } \frac{-3l}{2} < x_{(t)} \leq \frac{-l}{2} \\ 0 & \text{if } -\frac{l}{2} < x_{(t)} < \frac{l}{2} \\ l & \text{if } \frac{l}{2} \leq x_{(t)} < \frac{3l}{2} \\ \vdots \end{cases} \quad (9)$$

The initial state of the system is assumed to lie in compact set $I = [\frac{-1}{1-|a|}, \frac{1}{1-|a|}]$, which is an invariant subset of the state space. The output $y(t)$ takes on values in $\mathcal{Y} \subset \{-2, 0, 2\}$.

Simulation results for three values of a are shown in Table I. Three definitions for ρ_Δ and μ_Δ were considered: "γ" correspond to $\rho_\Delta(u) = |u|$ and $\mu_\Delta(w) = \mu_\Delta(\beta(y, \tilde{y})) = |y - \tilde{y}|$. "$\gamma$" correspond to $\rho_\Delta(u) = |u|$, $\mu_\Delta(w) = \mu_\Delta(\beta(y, \tilde{y})) = d(y, \tilde{y})$, where d denotes the discrete metric, and "γ_{dd}" corresponds to the choice $\rho_\Delta(u) = 1$ for all u, $\mu_\Delta(w) = \mu_\Delta(\beta(y, \tilde{y})) = d(y, \tilde{y})$. The decreasing gain bounds for γ_{dd} seem to capture the intuition that the "average number" of mismatches decreases as L increases.

B. Tank with Bimodal Pump

TABLE III

SIMULATION RESULTS FOR EXAMPLE 2: THE SAMPLING TIMES ARE T=45 AND 60 SECONDS

		T=45				T=60		
L	N_f	N_0	γ	γ_i	N_f	N_0	γ	γ_i
1	6	1	1	1	6	1	1	1
2	14	5	1	1	12	5	0	0
3	30	17	0.3333	0.3333	24	17	0	0
4	62	42	0.3333	0.25	48	41	0	0
5	126	93	0.2	0.1667	96	89	0	0
6	254	192	0.2	0.2	192	185	0	0
7	510	385	0.1429	0.125	384	377	0	0
8	1022	757	0.1429	0.1429	768	761	0	0
9	2046	1474	0.1111	0.1	1536	1521	0	0
10	4094	2864	0.1111	0.1111	3072	3065	0	0

Consider a tank with cross-sectional area $A = 100$ sq.cm. and height $h = 30$ cm. The level in the water tank is controlled by a pump with two operating modes: Mode '+' in which the pump feeds water at a rate of 1 liter/minute and mode '-' in which the pump removes water at a rate of 1 liter/minute. The available sensor simply monitors whether the water level has crossed the 15 cm mark.

The dynamics of this simple system, sampled at intervals of T seconds, are described by

$$x(t+1) = \begin{cases} \min\left\{ x(t) + \frac{1}{6}T, 30 \right\}, & u(t) = + \\ \max\left\{ x(t) - \frac{1}{6}T, 0 \right\}, & u(t) = - \end{cases}$$

$$y(t) = \begin{cases} F, & x(t) \in (15, 30] \\ E, & x(t) \in [0, 15] \end{cases}$$

TABLE I

SIMULATION RESULTS FOR QUANTIZED LTI SYSTEM

L	a=0.551					a=0.612					a=0.671				
	N_f	N_0	γ	γ_d	γ_{dd}	N_f	N_0	γ	γ_d	γ_{dd}	N_f	N_0	γ	γ_d	γ_{dd}
2	47	10	2	1	0.5	47	10	2	1	0.5	47	10	2	1	1
3	153	49	2	1	0.4	153	49	2	1	0.5	47	10	2	1	0.67
4	477	180	2	1	0.3333	491	180	2	1	0.5	47	10	2	1	0.5
5	1459	639	2	1	0.3333	1523	645	2	1	0.5	47	10	2	1	0.375
6	4421	2112	2	1	0.3333	4665	2129	2	1	0.5	47	10	2	1	0.375

TABLE II

SIMULATION RESULTS FOR EXAMPLE 2: THE SAMPLING TIMES ARE T=15, 30 AND 40 SECONDS

L	T=15				T=30				T=40			
	N_f	N_0	γ	γ_i	N_f	N_0	γ	γ_i	N_f	N_0	γ	γ_i
1	6	1	1	1	6	1	1	1	6	1	1	1
2	14	5	1	1	14	5	1	1	14	5	1	1
3	32	17	1	1	32	17	1	1	31	17	0.3333	0.3333
4	70	45	1	1	70	45	0.5	0.3333	64	45	0.3333	0.3333
5	152	109	1	1	152	106	0.5	0.5	133	95	0.2	0.2
6	323	249	1	1	324	234	0.3333	0.22	268	219	0.2	0.2
7	682	553	0.7143	0.6667	686	497	0.3333	0.3333	545	471	0.1429	0.1429
8	1424	998	0.7143	0.7143	1438	1033	0.25	0.18	1091	979	0.1429	0.1429
9	2957	2362	0.5556	0.5	2998	2115	0.25	0.25	2199	1999	0.1111	0.1111
10	6103	5204	0.5556	0.5556	6212	4297	0.2	0.1538	4394	4043	0.1111	0.1111

Simulation results are provided in Tables II and III for various values of T. The system is interesting as for T sufficiently large, it input/output behavior is *l-complete*. As expected, our simulations shown that for $T = 60$, the gain is non-zero for $L = 1$ but 0 for larger values of L: The *deterministic* finite state machine constructed following our proposed algorithm thus faithfully reproduces the input/ouput behavior of the tank.

As expected, the computed gain bounds are improved at the expense of larger approximate models. The column labelled "γ_i" corresponds to the gain bounds of a "refined" approximate model: In this case, we explicitly used the current state (set-valued) estimate X_q^+ to refine the transitions of the system. The improvement sometimes seen in the computation of the upper bounds comes at the expense of the much larger computational effort expended in constructing $\hat{M}_i$. It is unclear at this point whether the improvement has significant repercussions, particularly in the control setting.

VIII. CONCLUSIONS AND FUTURE WORK

We proposed a constructive algorithm for deriving deterministic finite state machine (DFM) approximations of a given system over finite alphabets, based on a partitioning of its input/output behavior into equivalence classes of finite length snapshots. We analyzed the algorithm to highlight the desirable properties of the resulting nominal models and corresponding approximation error systems. Finally, we proposed an algorithm for conservatively quantifying the resulting approximation error in a manner consistent with the objective of control synthesis.

Future work will focus on further investigating this approach by considering a more general setup where both the plant dynamics and performance specifications are approx-imated, allowing us to synthesize closed loop controllers based on the approximate finite state models.

IX. ACKNOWLEDGMENTS

REFERENCES

[1] R. Alur, T. Henzinger, G. Lafferriere, and G. Pappas. Discrete abstractions of hybrid systems. *Proceedings of the IEEE*, 88(2):971–984, 2000.

[2] S. Azuma and T. Sugie. Synthesis of optimal dynamic quantizers for discrete-valued input control. *IEEE Transactions on Automatic Control*, 53(9):2064–2075, 2008.

[3] C. Belta, A. Bicchi, M. Egerstedt, E. Frazzoli, E. Klavins, and G. J. Pappas. Symbolic planning and control of robot motion: State of the art and grand challenges. *IEEE Robotics and Automation Magazine*, 14(1):61–70, March 2007.

[4] M. S. Branicky, M. M. Curtiss, J. Levine, and S. Morgan. Sampling-based planning, control and verification of hybrid systems. *IEE Proceedings Control Theory and Applications*, 153(5):575–590, September 2006.

[5] R. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45(7):1279–1289, July 2000.

[6] N. Elia and S. K. Mitter. Stabilization of linear systems with limited information. *IEEE Transactions on Automatic Control*, 46(9):1384–1400, 2001.

[7] A. Girard and G. J. Pappas. Approximation metrics for discrete and continuous systems. *IEEE Transactions on Automatic Control*, 52(5):782–798, 2007.

[8] E. Haghverdi, P. Tabuada, and G. J. Pappas. Bisimulation relations for dynamical, control, and hybrid systems. *Theoretical Computer Science*, 342(2-3):229–261, September 2005.

[9] D. Liberzon and J. P. Hespanha. Stabilization of nonlinear systems with limited information feedback. *IEEE Transactions on Automatic Control*, 50:910–915, 2005.

[10] G. Lichtenberg and J. Lunze. Observation of qualitative states by means of a qualitative model. *International Journal of Control*, 66:885–903, 1997.

[11] Q. D. Ling and M. D. Lemmon. Stability of quantized control systems under dynamic bit assignmnet. *IEEE Transactions on Automatic Control*, 50(5):734–740, 2005.

[12] L. Ljung. *System Identification: Theory for the User*. Information and System Sciences Series. Prentice Hall, second edition edition, 1999.

[13] J. Lunze. Qualitative modeling of linear dynamical systems with quantized state measurements. *Automatica*, 30:417–431, 1994.

[14] J. Lunze. Qualitative model of dynamical systems - motivation, methods and prospective applications. *Mathematics and Computers in Simulation*, 46(5-6):465–483, 1998.

[15] J. Lunze, B. Nixdorf, and J. Schröder. Deterministic discrete-event representation of continuous-variable systems. *Automatica*, 35:395–406, March 1999.

[16] J. Lunze and J. Raisch. Discrete models for hybrid systems. In *Modeling, Analysis and Design of Hybrid Systems*, volume 279 of *Lecture Notes in Control and Information Sciences*, pages 67–80. Springer, Berlin/Heidelberg, 2002.

[17] R. Milner. *Communication and Concurrency*. Prentice Hall International Series in Computer Science, New York, 1989.

[18] T. Moor and J. Raisch. Supervisory control of hybrid systems whithin a behavioral framework. *Systems & Control Letters, Special Issue on Hybrid Control Systems*, 38:157–166, 1999.

[19] T. Moor, J. Raisch, and S. O'Young. Discrete supervisory control of hybrid systems by l-complete approximations. *Journal of Discrete Event Dynamic Systems*, 12:83–107, 2002.

[20] D. Park. Concurrency and automata on infinite sequences. In *Proceedings of the Fifth GI Conference on Theoretical Computer Science*, volume 104 of *Lecture Notes in Computer Science*, pages 167–183. Springer-Verlag, 1981.

[21] B. Picasso and A. Bicchi. On the stabilization of linear systems under assigned i/o quantization. *IEEE Transactions on Automatic Control*, 52(10):1994–2000, 2007.

[22] G. Pola, A. Girard, and P. Tabuada. Approximately bisimilar symbolic models for nonlinear control systems. *Automatica*, 44(10):2508–2516, October 2008.

[23] J. Raisch and S. D. O'Young. Discrete approximation and supervisory control of continuous systems. *IEEE Transactions on Automatic Control*, 43(4):569–573, April 1998.

[24] P. J. Ramadge and W. M Wonham. Supervisory control of a class of discrete event processes. *SIAM Journal on Control and Optimization*, 25(1):206–230, 1987.

[25] P. J. Ramadge and W. M. Wonham. The control of discrete event systems. *Proceedings of the IEEE*, 77(1):81–98, January 1989.

[26] P. Tabuada. An approximate simulation approach to symbolic control. *IEEE Transactions on Automatic Control*, 53(6):1406–1418, July 2008.

[27] P. Tabuada. *Verification and Control of Hybrid Systems: A Symbolic Approach*. Springer, 2009.

[28] P. Tabuada, A. Ames, A. A. Julius, and G. Pappas. Approximate reduction of dynamical systems. *Systems & Control Letters*, 7(57):538–545, 2008.

[29] D. C. Tarraf. A control-oriented notion of finite state approximation. Submitted for publication, manuscript available at http://arxiv.org/abs/1105.3788, 2011.

[30] D. C. Tarraf, A. Megretski, and M. A. Dahleh. Finite automata approximations with error bounds for systems with quantized actuation and measurement: A case study. In *43rd IEEE Conference on Decision and Control*, volume 2, pages 1436–1441, December 2004.

[31] D. C. Tarraf, A. Megretski, and M. A. Dahleh. Finite state controllers for stabilizing switched systems with binary sensors. In A. Bemporad, A. Bicchi, and G. Buttazzo, editors, *Hybrid Systems: Computation and Control*, volume 4416 of *Lecture Notes in Computer Science*, pages 543–556. Springer, April 2007.

[32] D. C. Tarraf, A. Megretski, and M. A. Dahleh. A framework for robust stability of systems over finite alphabets. *IEEE Transactions on Automatic Control*, 53(5):1133–1146, June 2008.

[33] D. C. Tarraf, A. Megretski, and M. A. Dahleh. Finite approximations of switched homogeneous systems for controller synthesis. *IEEE Transactions on Automatic Control*, 2011. (May 2011, in press).

[34] A. Tiwari and G. Khanna. Series of abstractions for hybrid automata. In *Hybrid Systems: Computation and Control*, volume 2289 of *Lecture Notes in Computer Science*. Springer, 2002.

[35] K. Tsumura. Approximattion of discrete time linear systems via bit length of memory. In *Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems*, pages 1066–1072, Kyoto, Japan, July 2006.

[36] K. Tsumura. A lower bound of the necessary bit length of memory for approximation of discrete time systems. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 2241–2246, New Orleans, LA, U.S.A., December 2007.

[37] J. C. Willems. Models for dynamics. *Dynamics Reported*, 2:171–269, 1989.