Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Approximate Bisimulations for Constrained Linear Systems

Antoine Girard and George J. Pappas

Abstract— **In this paper, inspired by exact notions of bisimulation equivalence for discrete-event and continuous-time systems, we establish approximate bi-simulation equivalence for linear systems with internal but bounded disturbances. This is achieved by developing a theory of approximation for transition systems with observation metrics, which require that the distance between system observations is and remains arbitrarily close in the presence of nondeterministic evolution. Our notion of approximate bisimulation naturally reduces to exact bisimulation when the distance between the observations is zero. Approximate bisimulation relations are then characterized by a class of Lyapunov-like functions which are called bisimulation functions. For the class of linear systems with constrained disturbances, we obtain computable characterizations of bisimulation functions in terms of linear matrix inequalities, set inclusions, and optimal values of static games. We illustrate our framework in the context of safety verification.**

I. INTRODUCTION

Complexity reduction and compositional reasoning in the verification of discrete systems have resulted in established notions of system refinement and equivalence, such as language inclusion, simulation and bisimulation relations [3]. Much more recently, simulation and bisimulation relations have been extended to continuous and hybrid state-spaces resulting in new equivalence notions for nondeterministic continuous and hybrid systems [10], [14], [16], [19].

These abstraction concepts are *exact* for both discrete and continuous systems, requiring external behavior of two systems to be identical. When interacting with the physical world, typically captured by continuous variables or dynamical systems with imprecise observations, exact refinement and equivalence notions are quite restrictive and not robust. Approximate versions of simulation and bisimulation relations seem much more appropriate in this context. This idea has recently been explored for quantitative [4], stochastic [5], [18] and metric transition systems [8], [9].

In [9], we developed a framework for (discrete and continuous) system approximation for general metric transition systems. Approximate simulation and bisimulation relations are defined based on a metric on the set of observation. Rather than requiring that the distance between system observations is and remains zero, we require that the distance between observations is and remains bounded. We showed that a class of functions called bisimulation functions allows to characterize approximate bisimulation relations in a computationally efficient manner.

This research is partially supported by the Région Rhône-Alpes (Projet CalCel) and the NSF Presidential Early CAREER (PECASE) Grant 0132716.

Antoine Girard and George J. Pappas are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 {agirard,pappasg}@seas.upenn.edu

In this paper, we extend our work by developing Lyapunov-like differential inequalities for bisimulation functions to a class of constrained linear systems. For a specific class of functions based on quadratic forms, these conditions can be interpreted in terms of linear matrix inequalities, set inclusions and optimal values of static games. In [8], the method is generalized to the class of metric transition systems generated by nonlinear but deterministic (autonomous) systems.

Compared to other approximation frameworks for linear systems such as traditional model reduction techniques [1], [2], [11], the reduction problem we consider is quite different and much more natural for safety verification for the following reasons. First, the systems we consider have constrained inputs which are internal (and hence they should be thought of as internal disturbances). Second, we do not assume that the systems are initially at the equilibrium: contrarily to the model reduction framework, the transient dynamics of the systems are not ignored during the approximation process. From the point of view of verification, the transient phase and the asymptotic phase of a trajectory are of equal importance. In fact, the quality of the approximation may critically depend on initial set of states. Finally, since our research has been motivated by the algorithmic verification of continuous and hybrid systems, the error bounds we compute are based on the L^∞ norm which is the only norm which makes sense for safety verification. In comparison, in [1], [2], the error bounds stand for the L^2 norm; in [11] the error bound is valid only on a time interval of finite length. We conclude this paper by illustrating this point in the context of safety verification for constrained linear systems.

II. APPROXIMATION OF TRANSITION SYSTEMS

In this section, we summarize the notion of approximate bisimulation of labeled transition systems as developed in [9]. Labeled transition systems can be seen as graphs, possibly with an infinite number of states or transitions.

Definition 2.1: A labeled transition system with observations is a tuple $T = (\mathcal{Q}, \Sigma, \rightarrow, \mathcal{Q}^0, \Pi, \langle\langle . \rangle\rangle)$ that consists of:

- a (possibly infinite) set $\mathcal{Q}$ of states,
- a (possibly infinite) set Σ of labels,
- a transition relation $\rightarrow \subseteq \mathcal{Q} \times \Sigma \times \mathcal{Q}$,
- a (possibly infinite) set $\mathcal{Q}^0 \subseteq \mathcal{Q}$ of initial states,
- a (possibly infinite) set Π of observations, and
- an observation map $\langle\langle . \rangle\rangle : \mathcal{Q} \rightarrow \Pi$.

The transition $(q, \sigma, q') \in \rightarrow$ is denoted $q \xrightarrow{\sigma} q'$. For all labels $\sigma \in \Sigma$, the σ-successor is defined as the set valued map given by

$$\forall q \in Q,\ \mathrm{Post}^{\sigma}(q) = \left\{ q' \in Q |\ q \xrightarrow{\sigma} q' \right\}.$$

We assume that the systems we consider are non-blocking. A state trajectory of T is an infinite sequence of transitions,

$$q^0 \xrightarrow{\sigma^0} q^1 \xrightarrow{\sigma^1} q^2 \xrightarrow{\sigma^2} \dots,\ \text{where } q^0 \in \mathcal{Q}^0.$$

The associated external trajectory $\pi^0 \xrightarrow{\sigma^0} \pi^1 \xrightarrow{\sigma^1} \pi^2 \xrightarrow{\sigma^2} \dots$ (where $\pi^i = \langle\langle q^i \rangle\rangle$ for all $i \in \mathbb{N}$) describes the evolution of the observations under the dynamics of the labeled transition system. The set of external trajectories of the labeled transition system T is called the language of T.

A. Approximate Bisimulations

Exact bisimulation between two labeled transition systems requires that their observations are (and remain) identical [3]. Approximate bisimulation is less strict since it only requires that the observations of both systems are (and remain) arbitrarily close. Let $T_1 = (\mathcal{Q}_1, \Sigma_1, \rightarrow_1, \mathcal{Q}_1^0, \Pi_1, \langle\langle . \rangle\rangle_1)$ and $T_2 = (\mathcal{Q}_2, \Sigma_2, \rightarrow_2, \mathcal{Q}_2^0, \Pi_2, \langle\langle . \rangle\rangle_2)$ be two labeled transition systems with the same set of labels ($\Sigma_1 = \Sigma_2 = \Sigma$) and the same set of observations ($\Pi_1 = \Pi_2 = \Pi$). Let us assume that the sets of states Q_1, Q_2 and the set of observations Π are metric spaces. We assume that the initial sets Q_1^0 and Q_2^0 as well as the sets $\mathrm{Post}_1^{\sigma}(q_1)$ and $\mathrm{Post}_2^{\sigma}(q_2)$ (for all $\sigma \in \Sigma$, $q_1 \in Q_1$, $q_2 \in Q_2$) are compact sets. Let us note by d_Π a metric on set of observations Π.

Definition 2.2: A relation $\mathcal{B}_\delta \subseteq \mathcal{Q}_1 \times \mathcal{Q}_2$ is a δ-approximate bisimulation between T_1 and T_2 if for all $(q_1, q_2) \in \mathcal{B}_\delta$:

1) $d_\Pi(\langle\langle q_1 \rangle\rangle_1, \langle\langle q_2 \rangle\rangle_2) \leq \delta$,
2) $\forall q_1 \xrightarrow{\sigma}_1 q_1'$, $\exists q_2 \xrightarrow{\sigma}_2 q_2'$ such that $(q_1', q_2') \in \mathcal{B}_\delta$,
3) $\forall q_2 \xrightarrow{\sigma}_2 q_2'$, $\exists q_1 \xrightarrow{\sigma}_1 q_1'$ such that $(q_1', q_2') \in \mathcal{B}_\delta$.

Note that for $\delta = 0$, we have the usual notion of exact bisimulation [3].

Definition 2.3: T_1 and T_2 are said to be approximately bisimilar with the precision δ (noted $T_1 \sim_\delta T_2$), if there exists $\mathcal{B}_\delta$, a δ-approximate bisimulation between T_1 and T_2 such that for all $q_1 \in \mathcal{Q}_1^0$, there exists $q_2 \in \mathcal{Q}_2^0$ such that $(q_1, q_2) \in \mathcal{B}_\delta$, and conversely.

Approximate bisimilarity of two systems guarantees that the distance between their language is bounded.

Theorem 2.4: [9] If T_1 and T_2 are approximately bisimilar with the precision δ then for all external trajectory of T_1 (respectively T_2), $\pi_1^0 \xrightarrow{\sigma^0} \pi_1^1 \xrightarrow{\sigma^1} \pi_1^2 \xrightarrow{\sigma^2} \dots$, there exists an external trajectory of T_2 (respectively T_1) with the same sequence of labels $\pi_2^0 \xrightarrow{\sigma^0} \pi_2^1 \xrightarrow{\sigma^1} \pi_2^2 \xrightarrow{\sigma^2} \dots$ such that for all $i \in \mathbb{N}$, $d_\Pi(\pi_1^i, \pi_2^i) \leq \delta$.

B. Bisimulation Functions

The construction and precision of approximate bisimulations can be performed using a class of functions called bisimulation functions. Essentially, bisimulation functions are positive functions defined on $\mathcal{Q}_1 \times \mathcal{Q}_2$, bounding the distance between the observations associated to a couple (q_1, q_2) and non increasing under the dynamics of the systems.

Definition 2.5: A function $V_\mathcal{B} : \mathcal{Q}_1 \times \mathcal{Q}_2 \rightarrow \mathbb{R}^+$ is a bisimulation function between T_1 and T_2 if its level sets are closed sets, and for all $(q_1, q_2) \in \mathcal{Q}_1 \times \mathcal{Q}_2$ we have:

1) $V_\mathcal{B}(q_1, q_2) \geq d_\Pi(\langle\langle q_1 \rangle\rangle_1, \langle\langle q_2 \rangle\rangle_2)$,
2) $V_\mathcal{B}(q_1, q_2) \geq \max_{q_1 \xrightarrow{\sigma}_1 q_1'} \min_{q_2 \xrightarrow{\sigma}_2 q_2'} V_\mathcal{B}(q_1', q_2')$,
3) $V_\mathcal{B}(q_1, q_2) \geq \max_{q_2 \xrightarrow{\sigma}_2 q_2'} \min_{q_1 \xrightarrow{\sigma}_1 q_1'} V_\mathcal{B}(q_1', q_2')$.

The level sets of a bisimulation functions define approximate bisimulation relations.

Theorem 2.6: [9] Let $V_\mathcal{B}$ be a bisimulation function. Then, for all $\delta \geq 0$, the set

$$\mathcal{B}_\delta = \{(q_1, q_2) \in \mathcal{Q}_1 \times \mathcal{Q}_2, V_\mathcal{B}(q_1, q_2) \leq \delta\}$$

is a δ-approximate bisimulation between T_1 and T_2.

Let us remark that particularly, the zero set of a bisimulation function is an exact bisimulation between T_1 and T_2. The following corollary is straightforward from Theorem 2.6 and Definition 2.3.

Corollary 2.7: [9] Let $V_\mathcal{B}$ be a bisimulation function. Let δ be the value of the following static game:

$$\delta = \max\left(\max_{q_1 \in \mathcal{Q}_1^0} \min_{q_2 \in \mathcal{Q}_2^0} V_\mathcal{B}(q_1, q_2), \max_{q_2 \in \mathcal{Q}_2^0} \min_{q_1 \in \mathcal{Q}_1^0} V_\mathcal{B}(q_1, q_2) \right) \quad (1)$$

Then, T_1 and T_2 are approximately bisimilar with the precision δ.

Thus, the challenge consists in developing methods to compute bisimulation functions for several classes of transition systems. In the following, this is done for constrained linear systems.

III. Bisimulation functions for constrained linear Systems

We consider continuous-time linear dynamical systems of the form:

$$\Delta_i : \begin{cases} \dot{x}_i(t) &= A_i x_i(t) + B_i u_i(t), \\ y_i(t) &= C_i x_i(t) \end{cases},\ i = 1, 2$$

with $y_i(t) \in \mathbb{R}^{p_i}$, $x_i(t) \in \mathbb{R}^{n_i}$, $x_i(0) \in I_i$ where I_i is a compact subset of $\mathbb{R}^{n_i}$ and $u_i(t) \in U_i$ where U_i is a compact subset of $\mathbb{R}^{m_i}$. We assume that both systems have the same observation space (*i.e.* $\mathbb{R}^{p_1} = \mathbb{R}^{p_2} = \mathbb{R}^p$) which is equipped with the usual Euclidean distance.

As suggested in [14], Δ_i can be seen as a labeled transition system $T_i = (\mathcal{Q}_i, \Sigma_i, \rightarrow_i, \mathcal{Q}_i^0, \Pi_i, \langle\langle . \rangle\rangle_i)$, where:

- the set of states is $\mathcal{Q}_i = \mathbb{R}^{n_i}$,
- the set of labels is $\Sigma_i = \mathbb{R}_+$,
- the transition relation $\rightarrow_i$ is given by $x \xrightarrow{t}_i x'$ if and only if there exists a locally measurable function $u_i(.)$ such that $\forall s \in [0, t], u_i(s) \in U_i$ and

$$x' = e^{A_i t} x + \int_0^t e^{A_i(t-s)} B_i u_i(s) ds,$$

- the set of initial states is $\mathcal{Q}_i^0 = I_i$,
- the set of observations is $\Pi_i = \mathbb{R}^p$,
- the observation map is given by $\langle\langle x \rangle\rangle_i = C_i x$.

Let us remark that the systems are nondeterministic, since there are many possible evolutions from a state for a given t. We define the following notations:

$$x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix},\ A=\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},\ C=\begin{bmatrix} C_1 \mid -C_2 \end{bmatrix},$$
$$\overline{B}_1=\begin{bmatrix} B_1 \\ 0 \end{bmatrix},\ \overline{B}_2=\begin{bmatrix} 0 \\ B_2 \end{bmatrix}.$$

We consider the problem of computing a bisimulation function between the two constrained linear systems. It is not straightforward to derive computational methods from the characterization given by Definition 2.5. The following proposition provides a more tractable characterization of bisimulation function. Due to the lack of space, the proof is not stated here.

Proposition 3.1: Let $q:\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}\to\mathbb{R}^+$ be differentiable and let ∇q denote its gradient. If for all $x\in\mathbb{R}^{n_1+n_2}$,

$$q(x)\geq x^TC^TCx \tag{2}$$
$$\max_{u_1\in U_1}\min_{u_2\in U_2}\nabla q(x)^T(Ax+\overline{B}_1u_1+\overline{B}_2u_2)\leq 0 \tag{3}$$
$$\max_{u_2\in U_2}\min_{u_1\in U_1}\nabla q(x)^T(Ax+\overline{B}_1u_1+\overline{B}_2u_2)\leq 0 \tag{4}$$

then $V_{\mathcal{B}}(x)=\sqrt{q(x)}$ is a bisimulation function.

Remark 3.2: There are similarities between the notions of bisimulation function and robust control Lyapunov function [6], [13] as well as are some significant conceptual differences. Indeed, let us consider the input u_1 as a disturbance and the input u_2 as a control variable in equation (3). Then, the interpretation of this inequality is that for all disturbances there exists a control such that the bisimulation function decreases during the evolution of the system. This means that the choice of u_2 can be made with the knowledge of u_1. In comparison, a robust control Lyapunov function requires that there exists a control u_1 such that for all disturbances u_2, the function decreases during the evolution of the system. Thus, it appears that robust control Lyapunov functions require stronger conditions than bisimulation functions.

In the following, we show that for specific classes of bisimulation functions, we can derive from Proposition 3.1 computationally effective characterizations.

A. Bisimulation Functions for Stable Systems

Let us assume that Δ_1 and Δ_2 are asymptotically stable (*i.e.* the real part of all eigenvalues of A_1 and A_2 is strictly negative).

1) Autonomous systems: Let $B_1=0$, $B_2=0$. Then, equations (3) and (4) become equivalent and reduce to a Lyapunov-like condition. For linear systems, it is well known that the class of quadratic functions provides universal and computationally effective Lyapunov functions. Therefore, let us search for bisimulation functions of the form:

$$V_{\mathcal{B}}(x)=\sqrt{x^TMx}. \tag{5}$$

where M is a symmetric positive semidefinite matrix. Then, the characterization given by proposition 3.1 reduces to the following set of linear matrix inequalities:

$$M\geq C^TC \tag{6}$$
$$A^TM+MA\leq 0. \tag{7}$$

These equations provide tractable conditions for bisimulation functions since linear matrix inequalities can be solved efficiently using semidefinite programming [15], [17]. Moreover this class of bisimulation functions is universal for autonomous stable linear systems.

Proposition 3.3: Let Δ_1 and Δ_2 be autonomous asymptotically stable linear systems. Then, there exists a bisimulation function of the form (5) between Δ_1 and Δ_2.

Proof: Equation (6) implies that $M=C^TC+N$ where N is symmetric positive semidefinite. Then equation (6) becomes

$$A^TN+NA\leq -A^TC^TC+C^TCA. \tag{8}$$

Let Q be a symmetric positive semidefinite matrix such that $A^TC^TC+C^TCA\leq Q$. Then, since Δ_1 and Δ_2 are asymptotically stable, the Lyapunov equation

$$A^TN+NA=-Q \tag{9}$$

has a unique solution which is symmetric positive semidefinite. Moreover it is clear that this solution satisfies (8). ■

We assumed that the initial sets of Δ_1 and Δ_2 are compact and thus bounded. Hence, the value of the game (1) is necessarily finite. Then, any two autonomous asymptotically stable linear systems are approximately bisimilar.

2) Systems with inputs: We now consider systems with constrained inputs. For such systems, the class of quadratic functions is often too restrictive to find a bisimulation function. Indeed, the value of such functions at $x=0$ is always 0. Particularly, this means that if Δ_1 and Δ_2 start from 0, the outputs of both systems will be identical. Equivalently, this means that Δ_1 and Δ_2 have identical asymptotic behaviors and that only their transient behaviors can differ. A natural extension of quadratic functions consists in searching for bisimulation functions of the form

$$V_{\mathcal{B}}(x)=\max(\alpha,\sqrt{x^TMx}). \tag{10}$$

In this function, the term $\sqrt{x^TMx}$ accounts for the error of approximation between the transient behaviors of Δ_1 and Δ_2 whereas α accounts for the error of approximation between their asymptotic behaviors and is therefore independent of the initial states x.

A characterization of bisimulation functions under that form is given in the following result:

Theorem 3.4: If there exists $\lambda>0$, such that

$$M\geq C^TC \tag{11}$$
$$A^TM+MA+2\lambda M\leq 0 \tag{12}$$
$$\alpha\geq\frac{1}{\lambda}\max_{x^TMx=1}\left(\max_{u_1\in U_1}\min_{u_2\in U_2}x^TM(\overline{B}_1u_1+\overline{B}_2u_2)\right) \tag{13}$$
$$\alpha\geq\frac{1}{\lambda}\max_{x^TMx=1}\left(\max_{u_2\in U_2}\min_{u_1\in U_1}x^TM(\overline{B}_1u_1+\overline{B}_2u_2)\right) \tag{14}$$

then the function $V_{\mathcal{B}}(x)=\max(\alpha,\sqrt{x^TMx})$ is a bisimulation function between Δ_1 and Δ_2.

Proof: Let $q(x)=\max(\alpha^2,x^TMx)$, then $V_{\mathcal{B}}(x)=\sqrt{q(x)}$. Let us show that $q(x)$ satisfies the conditions of

Proposition 3.1. First, it is clear from equation (11) that equation (2) is satisfied. Let $x \in \mathbb{R}^{n_1+n_2}$ such that $x^T Mx \geq \alpha^2$, then equation (13) implies that

$$\max_{u_1 \in U_1} \min_{u_2 \in U_2} x^T M(\overline{B}_1 u_1 + \overline{B}_2 u_2) \leq \lambda\alpha\sqrt{x^T Mx}.$$

Therefore, it is straightforward that

$$\max_{u_1 \in U_1} \min_{u_2 \in U_2} \nabla q(x)^T (Ax + \overline{B}_1 u_1 + \overline{B}_2 u_2) \leq \\ x^T A^T Mx + x^T MAx + 2\lambda\alpha\sqrt{x^T Mx}.$$

Then, from equation (12),

$$\max_{u_1 \in U_1} \min_{u_2 \in U_2} \nabla q(x)^T (Ax + \overline{B}_1 u_1 + \overline{B}_2 u_2) \leq \\ -2\lambda x^T Mx + 2\lambda\alpha\sqrt{x^T Mx} \leq \\ -2\lambda\sqrt{x^T Mx}(\sqrt{x^T Mx} - \alpha) \leq \quad 0.$$

Hence, if $x^T Mx \geq \alpha^2$ then equation (3) holds. If $x^T Mx \leq \alpha^2$, then $\nabla q(x) = 0$ and therefore equation (3) holds as well. Using symmetrical arguments, it can be shown that equation (4) holds as well and therefore $V_{\mathcal{B}}(x) = \max(\alpha, \sqrt{x^T Mx})$ is a bisimulation function between Δ_1 and Δ_2. ∎

Remark 3.5: If $\ker(M) + \overline{B}_1 U_1 = \ker(M) - \overline{B}_2 U_2$, then we can obviously choose $\alpha = 0$. In that case, there exists a quadratic bisimulation function between Δ_1 and Δ_2 which implies that their asymptotic behaviors are identical.

A small example shall help to understand the proposed methodology for the construction of bisimulation functions.

Example 3.6: Let us consider the following systems:

$$\Delta_1 : \dot{x}_1(t) = -x_1(t) + u_1(t),\ u_1(t) \in [0, 1],\ y_1(t) = x_1(t)$$
$$\Delta_2 : \dot{x}_2(t) = -x_2(t),\ y_2(t) = x_2(t)$$

Let us define

$$M = C^T C = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Equation (11) holds. We can check that $A^T M + MA = -2M$. Hence, equation (12) holds for $\lambda = 1$. Equation (13) becomes

$$\alpha \geq \max_{(x_1 - x_2)^2 = 1} \left(\max_{u_1 \in [0,1]} (x_1 - x_2) u_1 \right) = 1.$$

Equation (14) becomes

$$\alpha \geq \max_{(x_1 - x_2)^2 = 1} \left(\min_{u_1 \in [0,1]} (x_1 - x_2) u_1 \right) = 0.$$

From Theorem 3.4, $V_{\mathcal{B}}(x) = \max(|x_1 - x_2|, 1)$ is a bisimulation function between Δ_1 and Δ_2.

The characterization given by Theorem 3.4 is quite effective from a computational point of view. Indeed, the matrix M can be computed by solving a set of linear matrix inequalities. Then, α is chosen by computing the optimal value (or an over-approximation) of the optimization problems given by equations (13) and (14).

Similar to Proposition 3.3, we can show that bisimulation functions of the form (10) are universal for stable linear systems with constrained inputs.

Proposition 3.7: Let Δ_1 and Δ_2 be asymptotically stable linear systems with constrained inputs. Then, there exists a bisimulation function of the form (10) between Δ_1 and Δ_2.

Proof: Similar to the proof of Proposition 3.3, we can show that there exists a symmetric positive semidefinite matrix M satisfying equations (11) and (12). Then,

$$\max_{x^T Mx = 1} \left(\max_{u_1 \in U_1} \min_{u_2 \in U_2} x^T M(\overline{B}_1 u_1 + \overline{B}_2 u_2) \right) \leq \\ \max_{x^T Mx = 1} \left(\max_{u_1 \in U_1} \max_{u_2 \in U_2} x^T M(\overline{B}_1 u_1 + \overline{B}_2 u_2) \right) \leq \\ \max_{u_1 \in U_1} \max_{u_2 \in U_2} \left(\max_{x^T Mx = 1} x^T M(\overline{B}_1 u_1 + \overline{B}_2 u_2) \right) \leq \\ \max_{u_1 \in U_1} \max_{u_2 \in U_2} \sqrt{(\overline{B}_1 u_1 + \overline{B}_2 u_2)^T M(\overline{B}_1 u_1 + \overline{B}_2 u_2)}.$$

Since the set of inputs U_1 and U_2 are compact sets there exists α such that equation (13) and by symmetry equation (14) hold. ∎

We assumed that the initial sets of Δ_1 and Δ_2 are compact and thus bounded. Hence, the value of the game (1) is necessarily finite. Then, we have the following result:

Corollary 3.8: Let Δ_1 and Δ_2 be asymptotically stable constrained linear systems. Then, Δ_1 and Δ_2 are approximately bisimilar and the precision of the approximate bisimulation can be evaluated by solving game (1).

B. Bisimulation Functions for Non-Stable Systems

When Δ_1 and Δ_2 are not stable, the previous technique cannot be used since Proposition 3.1 implicitly assumes that there exists a bisimulation function with finite values on $\mathbb{R}^{n_1+n_2}$. This implies that for any $(x_1, x_2) \in \mathbb{R}^{n_1+n_2}$, for any trajectory of Δ_1 starting in x_1, there exists a trajectory of Δ_2 starting in x_2 and such that the distance between the observations of these trajectories remains bounded (and conversely). When dealing with unstable dynamics, it is not hard to see that this is generally not the case and that bisimulation functions with finite values on $\mathbb{R}^{n_1+n_2}$ cannot exist. In the following, we search for simulation functions whose values are finite on a subspace of $\mathbb{R}^{n_1+n_2}$.

Let $E_{u,i}$ (respectively $E_{s,i}$) be the subspace of $\mathbb{R}^{n_i}$ spanned by the generalized eigenvectors of A_i associated to eigenvalues whose real part is positive (respectively strictly negative). Note that we have $E_{u,i} \oplus E_{s,i} = \mathbb{R}^{n_i}$. Let $P_{u,i}$ and $P_{s,i}$ denote the associated projections. $E_{u,i}$ and $E_{s,i}$ are invariant under A_i and are called the unstable and the stable subspaces of the system Δ_i. Using a change of coordinates, the matrices of system Δ_i can be transformed into the following form

$$A_i = \begin{bmatrix} A_{u,i} & 0 \\ 0 & A_{s,i} \end{bmatrix}, B_i = \begin{bmatrix} B_{u,i} \\ B_{s,i} \end{bmatrix}, C_i = [C_{u,i}\ C_{s,i}], \tag{15}$$

where all the eigenvalues of $A_{u,i}$ have a positive real part and all the eigenvalues of $A_{s,i}$ have a strictly negative real part. Let us define the unstable subsystems of Δ_1 and Δ_2

$$\Delta_{u,i} : \begin{cases} \dot{x}_{u,i}(t) &= A_{u,i} x_{u,i}(t) + B_{u,i} u_i(t), \\ y_{u,i}(t) &= C_{u,i} x_{u,i}(t) \end{cases} \tag{16}$$

where $y_{u,i}(t) \in \mathbb{R}^p$, $x_{u,i}(t) \in E_{u,i}$, $x_{u,i}(0) \in P_{u,i}I_i$ and $u_i(t) \in U_i$. For $j \in \{u, s\}$, we define the matrices

$$A_j = \begin{bmatrix} A_{j,1} & 0 \\ 0 & A_{j,2} \end{bmatrix},\ C_j = \begin{bmatrix} C_{j,1} \mid -C_{j,2} \end{bmatrix}$$
$$\overline{B}_{j,1} = \begin{bmatrix} B_{j,1} \\ 0 \end{bmatrix},\ \overline{B}_{j,2} = \begin{bmatrix} 0 \\ B_{j,2} \end{bmatrix}.$$

and the projection defined by

$$P_j x = \begin{bmatrix} P_{j,1}x_1 \\ P_{j,2}x_2 \end{bmatrix}.$$

The following theorem generalizes the result of Proposition 3.1 to systems with unstable modes.

Theorem 3.9: Let $\mathcal{R}_u \subseteq E_{u,1} \times E_{u,2}$ be a subspace satisfying:

$$\mathcal{R}_u \subseteq \ker(C_u), \tag{17}$$
$$A_u \mathcal{R}_u \subseteq \mathcal{R}_u, \tag{18}$$
$$\mathcal{R}_u + \overline{B}_{u,1}U_1 = \mathcal{R}_u - \overline{B}_{u,2}U_2. \tag{19}$$

Let $q_s : E_{s,1} \times E_{s,2} \to \mathbb{R}^+$ be differentiable and let ∇q_s denote its gradient. If for all $x_s \in E_{s,1} \times E_{s,2}$,

$$q_s(x_s) \geq x_s^T C_s^T C_s x_s \tag{20}$$

$$\max_{u_1 \in U_1} \min_{\substack{u_2 \in U_2 \\ \overline{B}_{u,1}u_1 + \overline{B}_{u,2}u_2 \in \mathcal{R}_u}} \nabla q_s^T(x_s)\left(A_s x_s + \overline{B}_{s,1}u_1 + \overline{B}_{s,2}u_2\right) \leq 0 \tag{21}$$

$$\max_{u_2 \in U_2} \min_{\substack{u_1 \in U_1 \\ \overline{B}_{u,1}u_1 + \overline{B}_{u,2}u_2 \in \mathcal{R}_u}} \nabla q_s^T(x_s)\left(A_s x_s + \overline{B}_{s,1}u_1 + \overline{B}_{s,2}u_2\right) \leq 0 \tag{22}$$

then the function $V_{\mathcal{B}} : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^+ \cup \{+\infty\}$ defined by $V_{\mathcal{B}}(x) = \sqrt{q_s(P_s x)}$ if $P_u x \in \mathcal{B}_u$ and $V_{\mathcal{B}}(x) = +\infty$ otherwise, is a bisimulation function between Δ_1 and Δ_2.

Proof: The sketch of the proof is the following. Let $x = (x_1, x_2) \in \mathbb{R}^{n_1+n_2}$, if $P_u x \notin \mathcal{B}_u$ then $V_{\mathcal{B}}(x) = +\infty$ and it is clear that the conditions of Definition 2.5 hold. Hence, let us assume $P_u x \in \mathcal{B}_u$, then $V_{\mathcal{B}}(x) = \sqrt{q_s(P_s x)}$. From equations (17) and (20),

$$V_{\mathcal{B}}(x) \geq \|C_s P_s x\| = \|C_s P_s x + C_u P_u x\| = \|C_x\|.$$

Then, the first condition of Definition 2.5 holds. Let $x_1 \xrightarrow{t}_1 x_1'$, let $u_1(.)$ be an input which leads Δ_1 from x_1 to x_1' in time t. Equation (19) and (21) imply that there exists an input $u_2(.)$ such that $\overline{B}_{u,1}u_1(.) + \overline{B}_{u,2}u_2(.) \in \mathcal{R}_u$ and the function q_s is decreasing under the evolution of the systems. $u_2(.)$ leads Δ_2 from x_2 to x_2' in time t, then

$$q_s(P_s x') \leq q_s(P_s x) \text{ where } x' = (x_1', x_2').$$

Moreover since $E_{u,1}$ and $E_{u,2}$ are invariant under A_1 and A_2, we have that

$$P_u x' = e^{A_u t}P_u x + \int_0^t e^{A_u(t-s)}\left(\overline{B}_{u,1}u_1(s) + \overline{B}_{u,2}u_2(s)\right) ds$$

From equation (18), it is straightforward that $P_u x' \in \mathcal{R}_u$. Hence, $x_2 \xrightarrow{t}_2 x_2'$ and $V_{\mathcal{B}}(x') \leq V_{\mathcal{B}}(x)$. Therefore, the second and by symmetry the third conditions of Definition 2.5 hold. ■

Remark 3.10: We can check (see [14], [19]), that the subspace $\mathcal{R}_u \subseteq E_{u,1} \times E_{u,2}$ satisfying equations (17), (18) and (19) is actually an exact bisimulation relation between the unstable subsystems $\Delta_{u,1}$ and $\Delta_{u,2}$.

The function q_s can be computed using a technique similar to the one we described for the computation of bisimulation functions for stable systems. Actually, the only difference is that now the inputs u_1 and u_2 are not independent anymore but related by $\overline{B}_{u,1}u_1 + \overline{B}_{u,2}u_2 \in \mathcal{R}_u$. Similar to Proposition 3.7, we can show that there always exists a function q_s of the form (10) and satisfying equations (20), (21) and (22). As a consequence, we have:

Corollary 3.11: If there exists a subspace $\mathcal{R}_u$ satisfying equations (17), (18) and (19), and such that for all $x_{u,1} \in P_{u,1}I_1$ there exists $x_{u,2} \in P_{u,2}I_2$ satisfying $(x_{u,1}, x_{u,2}) \in \mathcal{R}_u$ (*i.e.* the unstable subsystems $\Delta_{u,1}$ and $\Delta_{u,2}$ are exactly bisimilar), then Δ_1 and Δ_2 are approximately bisimilar.

Proof: For all $x_1 \in I_1$, there exists $x_2 \in I_2$ such that $P_u x \in \mathcal{R}_u$ then,

$$\max_{x_1 \in I_1} \min_{x_2 \in I_2} V(x_1, x_2) = \max_{x_1 \in I_1}\left(\min_{x_2 \in I_2,\ P_u x \in \mathcal{R}_u} \sqrt{q_s(P_s x)}\right). \tag{23}$$

Since I_1 and I_2 are compact sets, this game has a finite value and thus Δ_1 approximately simulates Δ_2. ■

IV. Safety Verification

We now show how our results can be used for the approximation of a system by a system of lower dimension in the context of safety verification.

Let Δ_1 be a constrained linear system. Then $\mathrm{Reach}(\Delta_1)$ denotes the reachable set of Δ_1 and is defined as the subset of $\mathbb{R}^{p_1}$ of points reachable by the external trajectories of Δ_1. We consider the problem of checking wether the intersection of $\mathrm{Reach}(\Delta_1)$ with a set Π_F of unsafe sets is empty or not. Thus, we must verify that for any inputs the external trajectories of Δ_1 does not reach Π_F. In that case, the inputs must be seen as disturbances or uncertainties. Though recent progress has been made in the reachability analysis of high dimensional systems [7], [11], [12], [20], it remains one of the most challenging issues of the verification of continuous and hybrid systems. Our method consists in constructing a smaller system Δ_2 such that its reachable set is close enough to the one of Δ_1 in order to process the safety verification by solving a reachability problem for Δ_2. Particularly, if Δ_2 is approximately bisimilar to Δ_1 (with some precision δ) and if the distance of $\mathrm{Reach}(\Delta_2)$ to Π_F is greater than δ then Theorem 2.4 allows to conclude that Δ_1 is safe.

Without loss of generality, let us assume that the matrices of Δ_1 are of the form (15). Let $\Delta_{u,1}$ be the unstable subsystem of Δ_1. From Corollary 3.11, we know that Δ_1 and $\Delta_{u,1}$ are approximately bisimilar. We use the following methodology to compute a bisimulation function. First, we solve the linear matrix inequalities (11) and (12). The second step consists in solving the two optimization problems (13) and (14). Afterwards, the precision of the approximate bisimulation can be evaluated by solving the game (1).

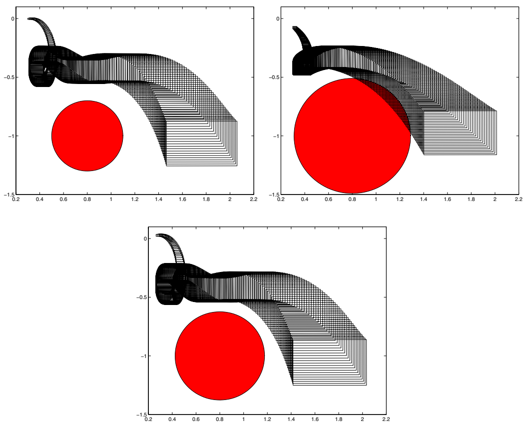

Fig. 1. Reachable sets of the original ten dimensional system (top left) and of its four dimensional and six dimensional approximations (top right and bottom). The disk on the left figure represents the unsafe set Π_F. The disks on the right and bottom figures consist of the set of points whose distance to Π_F is smaller than the precision of the approximate bisimulation between Δ_1 and its approximations.

We used this method with a ten dimensional system with ten inputs and two outputs. The associated unstable subsystem is a four dimensional system with four inputs and two outputs. We computed the reachable sets of both systems using zonotope techniques for reachability analysis of linear systems with inputs [7]. In Figure 1, we represented the reachable sets of the ten dimensional system and of its four dimensional approximation. We can see that the approximation does not allow to conclude though Δ_1 is actually safe.

Therefore, we need to refine the approximation. Our approach consists in defining the approximation Δ_2 as a combination of the unstable subsystem $\Delta_{u,1}$ with a stable subsystem. Then, from Corollary 3.11, we know that Δ_1 and Δ_2 are approximately bisimilar. The better the stable subsystems approximates the stable part of Δ_1, the better the system Δ_2 approximates system Δ_1. For our example, we chose the stable subsystem as the projection of the stable part of Δ_1 on the two dimensional space spanned by the eigenvectors associated to the two largest eigenvalues of the matrix $A_{s,1}$. We can see on Figure 1 that the approximation of Δ_1 by the six dimensional system Δ_2 allows to check the safety of Δ_1.

The example also illustrates the important point that robustness simplifies verification. Indeed, if the distance between $\mathrm{Reach}(\Delta_1)$ and Π_F would have been larger then the approximation of Δ_1 by its unstable subsystem might have been sufficient to check the safety of Δ_1. Generally, the more robustly safe a system is, the larger the distance from the unsafe safe, resulting in larger model compression and easier safety verification.

V. Conclusion

In this paper, we applied the framework of approximate bisimulations to the approximation of constrained linear systems. We presented a class of functions which provide universal bisimulation functions for such systems. An important consequence, is that any two systems with exactly bisimilar unstable subsystems are approximately bisimilar. A computationally tractable characterization for this class of bisimulation functions has been given. Finally, we showed how the approximate bisimulation framework could be used in the context of safety verification of constrained linear systems. Future research should deal with the development of methods for computing bisimulation functions for nonlinear, stochastic, and hybrid systems.

References

[1] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, "A survey of model reduction methods for large-scale systems," *Contemporary Mathematics*, vol. 280, pp. 193–219, 2000.

[2] C. L. Beck, J. Doyle, and K. Glover, "Model reduction of multi-dimensional and uncertain systems," *IEEE Transactions on Automatic Control*, vol. 41, no. 10, pp. 1466–1477, October 1996.

[3] E. M. Clarke, O. Grumberg, and D. A. Peled, *Model Checking*. MIT Press, 2000.

[4] L. de Alfaro, M. Faella, and M. Stoelinga, "Linear and branching metrics for quantitative transition systems," in *ICALP'04*, ser. LNCS. Springer, 2004, vol. 3142, pp. 1150–1162.

[5] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden, "Metrics for labelled markov processes," *Theoretical Computer Science*, vol. 318, no. 3, pp. 323–354, June 2004.

[6] R. A. Freeman and P. V. Kokotovic, "Inverse optimality in robust stabilization," *SIAM J. Control and Optimization*, vol. 34, no. 4, pp. 1365–1391, July 1996.

[7] A. Girard, "Reachability of uncertain linear systems using zonotopes," in *Hybrid Systems: Computation and Control*, ser. LNCS. Springer, 2005, vol. 3414, pp. 291–305.

[8] A. Girard and G. J. Pappas, "Approximate bisimulations for nonlinear systems," in *Proc. 44th IEEE Conference on Decision and Control and European Control Conference*, December 2005.

[9] ——, "Approximation metrics for discrete and continuous systems," May 2005, technical Report MS-CIS-05-10, Dept. of CIS, University of Pennsylvania.

[10] E. Haghverdi, P. Tabuada, and G. J. Pappas, "Bisimulation relations for dynamical, control, and hybrid systems," *Theoretical Computer Science*, vol. 342, pp. 229–261, 2005.

[11] Z. Han and B. H. Krogh, "Reachability of hybrid control systems using reduced-order models," in *Proc. American Control Conference*, Boston, MA, July 2004.

[12] B. H. Krogh and O. Stursberg, "On efficient representation and computation of reachable sets for hybrid systems," in *Hybrid Systems: Computation and Control*, ser. LNCS. Springer, 2003, vol. 2623, pp. 482–497.

[13] D. Liberzon, E. D. Sontag, and Y. Wang, "Universal construction of feedback laws achieving ISS and integral-ISS disturbance attenuation," *Systems and Control Letters*, vol. 46, pp. 111–127, 2002.

[14] G. J. Pappas, "Bisimilar linear systems," *Automatica*, vol. 39, no. 12, pp. 2035–2047, December 2003.

[15] J. F. Sturm, "Using SEDUMI 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Softwares*, vol. 11-12, pp. 625–653, 1999.

[16] H. Tanner and G. J. Pappas, "Abstractions of constrained linear systems," in *Proc. American Control Conference*, Denver, CO, June 2003.

[17] K. C. Toh, R. H. Tutuncu, and M. J. Todd. SDPT3 - a MATLAB software package for semidefinite quadratic linear programming. [Online]. Available: http://www.math.nus.edu.sg/~mattohkc/sdpt3.html

[18] F. van Breugel, M. Mislove, J. Ouaknine, and J. Worrell, "An intrinsic characterization of approximate probabilistic bisimilarity," in *Foundations of Software Science and Computation Structures*, ser. LNCS. Springer, 2003, vol. 2620, pp. 200–215.

[19] A. van der Schaft, "Equivalence of dynamical systems by bisimulation," *IEEE Trans. Automatic Control*, vol. 49, pp. 2160–2172, 2004.

[20] H. Yazarel and G. J. Pappas, "Geometric programming relaxations for linear system reachability," in *Proc. American Control Conference*, Boston, MA, June 2004.