Bisimilar Finite Abstractions of Interconnected Systems

Yuichi Tazaki and Jun-ichi Imura

Tokyo Institute of Technology,
Ōokayama 2-12-1, Meguro, Tokyo, Japan
{tazaki,imura}@cyb.mei.titech.ac.jp
http://www.cyb.mei.titech.ac.jp

Abstract. This paper addresses the design of approximately bisimilar finite abstractions of systems that are composed of the interconnection of smaller subsystems. First, it is shown that the ordinary notion of approximate bisimulation does not preserve the interconnection structure of the concrete model. Next, a new definition of approximate bisimulation that is compatible with interconnection is proposed. Based on this definition of approximate bisimulation, the design of interconnection-compatible finite abstractions of linear subsystems is discussed.

1 Introduction

Discrete abstractions simplify concrete continuous systems by cutting off the details, while preserving the essential characteristics. Moreover, it reduces the computational cost of numerical methods such as reachability analysis and controller synthesis. During this decade, there have been a variety of researches on this topic. Lafferriere et al [1] investigated a class of autonomous planar hybrid systems with finite bisimulations. Alur et al [2] presented algorithms for reachability analysis of hybrid systems by combining the notion of predicate abstraction with polytopic approximation of reachable sets. Lunze [3] considered continuous-time, continuous-state systems that can only be observed through discrete events triggered when the state hits one of the boundaries placed on the state space, and modeled the occurrence of discrete-event sequences as stochastic automata. Tsumura [4] considered systems whose state is stored in a digital memory and analyzed the relation between necessary bit-length to achieve a certain bound on input-to-output approximation error, and systems properties such as stability.

Recently, there has been several researches on finite-state abstractions using the notion of approximate bisimulation [6], which is an extension of the classical bisimulation. Girard [7] derived a procedure for constructing approximately bisimilar finite abstractions of stable discrete-time linear systems. Tabuada [8] addressed a design of approximately similar finite abstractions of continuous-time nonlinear dynamical systems under a certain stabilizability assumption. Tazaki [12] discussed the application of approximate bisimilar abstractions to optimal control problems.

To date, these approaches on discrete abstractions have been successful only to systems of a relatively small size. This is mainly due to the fact that the number of the state of discrete abstractions often grows exponentially with respect to the state dimension of the concrete system. If one knows the internal interconnection structure of a complex system, it is natural to take advantage of such knowledge to reduce the complexity of the computation of the abstraction process and that of the resultant abstraction itself. Tabuada et al [9] discussed the relation between bisimulations and compositional operators in a general setting. They showed that for a concrete system given by a composition of subsystems, there exists a bisimilar abstraction that is expressed as a composition of bisimilar abstractions of subsystems. Julius et al [11] addressed approximate syncronization and showed that approximate (bi)simulation is preserved under approximate syncronization. However, in the case of input-output interconnection, which is a special class of composition, the interconnection structure of the concrete system is not in general preserved in its abstraction. This means that, under the conventional bisimulation, one cannot simply interconnect the abstractions of subsystems to construct an abstraction of the original interconnected system.

In this research, motivated by the above background, we propose a new variant of approximate bisimulation that is compatible with interconnected systems. Furthermore, based on the proposed interconnection-compatible bisimulation, the design of finite abstractions of linear subsystems is developed.

The rest of this paper is organized as follows. In Section 2, we define the basic form of discrete-time dynamical systems treated in this paper, and the notion of approximate simulation and approximate bisimulation on this class of systems. In Section 3, after introducing the framework of input-output interconnected systems, we show with a simple example that the ordinary approximate bisimulation does not preserve the interconnection structure. To overcome this problem, we propose a new notion of approximate bisimulation that is compatible with interconnection. In Section 4 we discuss the finite abstraction problem of linear subsystems, according to the definition of approximate bisimulation introduced in Section 3. Section 5 concludes this paper with some remarks for future works.

Notation: The symbol $[\boldsymbol{v}_1; \boldsymbol{v}_2; ...; \boldsymbol{v}_N]$ denotes the vertical concatenation of vectors or that of matrices, which is equivalent to $[\boldsymbol{v}_1^{\mathrm{T}}\ \boldsymbol{v}_2^{\mathrm{T}}\ \ldots\ \boldsymbol{v}_N^{\mathrm{T}}]^{\mathrm{T}}$. Throughout the paper, the symbol $\|\cdot\|$ denotes the 2-norm unless otherwise stated. Moreover, the symbol $\|\boldsymbol{v}\|_M$ is defined as $\sqrt{\boldsymbol{v}^{\mathrm{T}} M \boldsymbol{v}}$. For matrices, $\|A\|$ denotes the largest singular value of A.

2 Approximate simulations and bisimulations of discrete-time dynamical systems

In this section, we introduce the definition of approximate (bi)simulation on a class of discrete-time dynamical systems. Let us first define the basic form of discrete-time dynamical systems.

Definition 1. *Discrete-time dynamical system*
A discrete-time dynamical system (or simply a system) is a 5-tuple $\langle X, U, Y, f, h \rangle$*, where* $X \subset \mathbb{R}^n$ *is the set of states,* $U \subset \mathbb{R}^m$ *is the set of inputs,* $Y \subset \mathbb{R}^l$ *is the set of outputs,* $f : X \times U \mapsto X$ *is the state transition function, and* $h : X \times U \mapsto Y$ *is the measurement function. The state, input, and output of the system at time* $t \in \mathcal{T} = \{0\} \cup \mathbb{N}$ *are expressed as* $\boldsymbol{x}_t \in X$*,* $\boldsymbol{u}_t \in U$*, and* $\boldsymbol{y}_t \in Y$*, respectively. The state transition and the measurement at time* t *are expressed as*

$$\boldsymbol{x}(t+1) = f(\boldsymbol{x}(t), \boldsymbol{u}(t)), \tag{1}$$

$$\boldsymbol{y}(t) = h(\boldsymbol{x}(t), \boldsymbol{u}(t)), \tag{2}$$

respectively.

Throughout this paper, we use the symbol $\Sigma\langle X, U, Y, f, h\rangle$ or simply Σ to express a system. Let us introduce the notion of approximate simulation and approximate bisimulation on the class of systems just defined.

Definition 2. *Approximate simulation of dynamical systems*
Consider two systems $\Sigma\langle X, U, Y, f, h\rangle$*,* $\hat{\Sigma}\langle \hat{X}, \hat{U}, Y, \hat{f}, \hat{h}\rangle$ *and positive constant* ϵ*. A binary relation* $R \subset X \times \hat{X}$ *is called an* ϵ*-approximate simulation relation from* Σ *to* $\hat{\Sigma}$ *if and only if for every* $(\boldsymbol{x}, \hat{\boldsymbol{x}}) \in R$*, the following holds:*
for all $\boldsymbol{u} \in U$*, there exists a* $\hat{\boldsymbol{u}} \in \hat{U}$ *such that*

$$\|h(\boldsymbol{x}, \boldsymbol{u}) - \hat{h}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}})\| \leq \epsilon, \tag{3}$$

$$(f(\boldsymbol{x}, \boldsymbol{u}),\ \hat{f}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}})) \in R. \tag{4}$$

Moreover, if such an R *exists,* $\hat{\Sigma}$ *is said to be approximately similar to* Σ *with respect to* R *and the precision* ϵ*.*

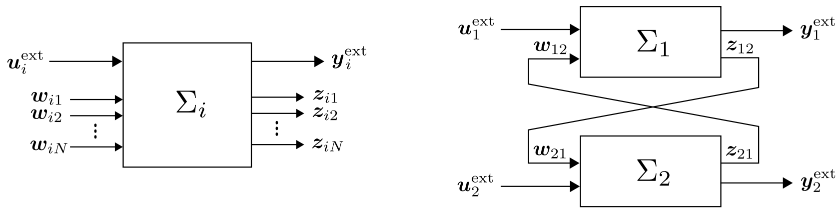
Definition 3. *Approximate bisimulation of dynamical systems*
Consider two systems $\Sigma\langle X, U, Y, f, h\rangle$*,* $\hat{\Sigma}\langle \hat{X}, \hat{U}, Y, \hat{f}, \hat{h}\rangle$ *and positive constant* ϵ*. A binary relation* $R \subset X \times \hat{X}$ *is called an* ϵ*-approximate bisimulation relation between* Σ *and* $\hat{\Sigma}$ *if and only if* R *is an* ϵ*-approximate simulation relation from* Σ *to* $\hat{\Sigma}$ *and its inverse relation* $R^{-1} = \{(\hat{\boldsymbol{x}}, \boldsymbol{x}) \,|\, (\boldsymbol{x}, \hat{\boldsymbol{x}}) \in R\}$ *is an* ϵ*-approximate simulation relation from* $\hat{\Sigma}$ *to* Σ*. Moreover, if such an* R *exists,* Σ *and* $\hat{\Sigma}$ *are said to be approximately bisimilar with respect to* R *and the precision* ϵ*, and this relation is denoted by* $\Sigma \sim_\epsilon \hat{\Sigma}$*.*

The major difference between the above definitions and those introduced in the literature (see [6], for example) is that the measurement variable of dynamical systems depend not only on states but also on control inputs, and therefore the definitions of approximate (bi)simulation are extended accordingly.

3 Approximate Bisimulation of Interconnected Systems

3.1 Expression of interconnected systems

This subsection introduces the general expression of interconnected systems treated in this paper. Consider a complex system composed of N subsystems



(a) i-th subsystem of interconnected system. (b) Interconnection of two sub-systems.

Fig. 1. Schematics of interconnected system.

interconnected with each other. The i-th subsystem is described as

$$\Sigma_i \langle X_i, U_i, Y_i, f_i, h_i \rangle.$$

Here, the input variable $\boldsymbol{u}_i \in U_i$ and the output variable $\boldsymbol{y}_i \in Y_i$ are decomposed into subvectors as shown below.

$$\boldsymbol{u}_i = \left[\boldsymbol{u}_i^{\text{ext}}; \boldsymbol{w}_i\right], \quad \boldsymbol{w}_i = \begin{cases} \left[\boldsymbol{w}_{i2}; \ldots; \boldsymbol{w}_{iN}\right] & (i = 1) \\ \left[\boldsymbol{w}_{i1}; \ldots; \boldsymbol{w}_{i,N-1}\right] & (i = N) \\ \left[\boldsymbol{w}_{i1}; \ldots; \boldsymbol{w}_{i,i-1}; \boldsymbol{w}_{i,i+1}; \ldots; \boldsymbol{w}_{iN}\right] & \text{otherwise.} \end{cases} \tag{5}$$

$$\boldsymbol{y}_i = \left[\boldsymbol{y}_i^{\text{ext}}; \boldsymbol{z}_i\right], \quad \boldsymbol{z}_i = \begin{cases} \left[\boldsymbol{z}_{i2}; \ldots; \boldsymbol{z}_{iN}\right] & (i = 1) \\ \left[\boldsymbol{z}_{i1}; \ldots; \boldsymbol{z}_{i,N-1}\right] & (i = N) \\ \left[\boldsymbol{z}_{i1}; \ldots; \boldsymbol{z}_{i,i-1}; \boldsymbol{z}_{i,i+1}; \ldots; \boldsymbol{z}_{iN}\right] & \text{otherwise.} \end{cases} \tag{6}$$

Each subsystem has two groups of input signals ($\boldsymbol{u}_i^{\text{ext}}$ and $\boldsymbol{w}_i$) and two groups of output signals ($\boldsymbol{y}_i^{\text{ext}}$ and $\boldsymbol{z}_i$). The signals $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ are *internal signals*, used to construct interconnections between other subsystems. On the other hand, $\boldsymbol{u}_i^{\text{ext}}$ and $\boldsymbol{y}_i^{\text{ext}}$ are *external signals*, which compose, together with those of other subsystems, the input/output interface of the whole interconnected system.

Here, for simplicity of discussion, we introduce the following assumption to ensure that the interconnected system is well-posed.

Assumption 1. *The internal output variables at time t, $\boldsymbol{z}_i(t)$, are independent of the internal input variables at time t, $\boldsymbol{w}_i(t)$.*

The measurement function is then decomposed as

$$\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_i^{\text{ext}} \\ \boldsymbol{z}_i \end{bmatrix} = \begin{bmatrix} h_i^y(\boldsymbol{x}_i, \boldsymbol{u}_i^{\text{ext}}, \boldsymbol{w}_i) \\ h_i^z(\boldsymbol{x}_i, \boldsymbol{u}_i^{\text{ext}}) \end{bmatrix}. \tag{7}$$

We first define the parallel composition of two subsystems.

Definition 4. *Parallel composition*
Suppose two systems $\Sigma_i \langle X_i, U_i, Y_i, f_i, h_i \rangle$ $(i = 1, 2)$ are given. The parallel composition of Σ_1 and Σ_2 is the system $\langle X_1 \times X_2, U_1 \times U_2, Y_1 \times Y_2, f_1 \| f_2, h_1 \| h_2 \rangle$

whose state transition function and measurement function are defined as follows.

$$\begin{aligned}\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}(t+1) &= (f_1\|f_2)\left(\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}(t), \begin{bmatrix}\boldsymbol{u}_1\\ \boldsymbol{u}_2\end{bmatrix}(t)\right) = \begin{bmatrix}f_1(\boldsymbol{x}_1(t),\boldsymbol{u}_1(t))\\ f_2(\boldsymbol{x}_2(t),\boldsymbol{u}_2(t))\end{bmatrix},\\ \boldsymbol{y}(t) &= (h_1\|h_2)\left(\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\end{bmatrix}(t), \begin{bmatrix}\boldsymbol{u}_1\\ \boldsymbol{u}_2\end{bmatrix}(t)\right) = \begin{bmatrix}h_1(\boldsymbol{x}_1(t),\boldsymbol{u}_1(t))\\ h_2(\boldsymbol{x}_2(t),\boldsymbol{u}_2(t))\end{bmatrix}.\end{aligned} \tag{8}$$

We denote by $\Sigma_1\|\Sigma_2$ the parallel composition of Σ_1 and Σ_2. The parallel composition of more than two systems are defined recursively as follows.

$$\Sigma_1\|\Sigma_2\|\dots\|\Sigma_N := \Sigma_1\|(\Sigma_2\|\dots\|\Sigma_N). \tag{9}$$

The interconnection of subsystems are obtained by imposing restrictions representing the interconnections of the internal signals on their parallel composition.

Definition 5. *Interconnection of subsystems*
Suppose N subsystems $\Sigma_i\langle X_i, U_i, Y_i, f_i, h_i\rangle$ $(i=1,2,\dots,N)$, whose input vectors and output vectors are decomposed as in (5) and (6), respectively, are given. Moreover, suppose the size of subvectors $\boldsymbol{w}_{ij}$ and $\boldsymbol{z}_{ji}$ matches for all $i\in\{1,2,\dots,N\}, j\in\{1,2,\dots,N\}\backslash\{i\}$. The interconnection of $\Sigma_1,\Sigma_2,\dots,$ and Σ_N, denoted by $\mathcal{I}(\Sigma_1,\Sigma_2,\dots,\Sigma_N)$, is defined as the parallel composition $\Sigma_1\|\Sigma_2\|\dots\|\Sigma_N$ subject to the constraints

$$\boldsymbol{w}_{ij} = \boldsymbol{z}_{ji}\ \ (i\in\{1,2,\dots,N\},\, j\in\{1,2,\dots,N\}\backslash\{i\}) \tag{10}$$

and whose input and output variables are defined as

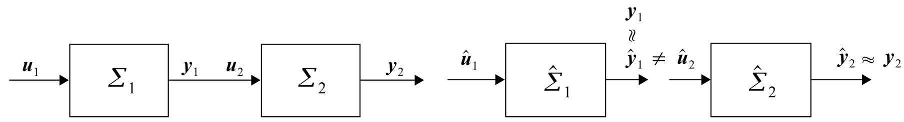
$$\bar{\boldsymbol{u}} = \left[\boldsymbol{u}_1^{\text{ext}};\boldsymbol{u}_2^{\text{ext}};\dots;\boldsymbol{u}_N^{\text{ext}}\right],\ \bar{\boldsymbol{y}} = \left[\boldsymbol{y}_1^{\text{ext}};\boldsymbol{y}_2^{\text{ext}};\dots;\boldsymbol{y}_N^{\text{ext}}\right]. \tag{11}$$

Fig. 1(a) illustrates the block diagram of the i-th subsystem, and Fig. 1(b) shows the interconnected system with two subsystems.

3.2 Composition-compatible bisimulation

There are some important aspects that the abstractions of interconnected systems should be equipped with. First, it should preserve the interconnection structure of the original system. In addition, it is preferable if we could design the abstraction separately for each subsystem. However, using the ordinary definition of bisimulation, the interconnection of the bisimilar abstractions of subsystems is *not* in general bisimilar with the original interconnected system. This can be shown in the following simple example.

Let us consider two systems Σ_1 and Σ_2 connected in a cascade (Fig. 2(a)). We denote by $\hat{\Sigma}_i$ an abstraction of Σ_i, which is ϵ_i-approximately bisimilar to Σ_i with the binary relation R_i. Our question is whether the cascade of the abstractions (shown in Fig 2(b)) is approximately bisimilar to the cascade of the original systems, under the notion of the conventional bisimulation.



(a) Cascade of subsystems. (b) Cascade of abstractions.

Fig. 2. Cascade of two subsystems.

In the cascade of Σ_1 and Σ_2, the equality

$$\boldsymbol{y}_1 = \boldsymbol{u}_2 \tag{12}$$

holds. In the cascade of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, on the other hand, assuming that the interconnection is temporarily cut off, approximate bisimilarity implies that

$$\|\boldsymbol{y}_i - \hat{\boldsymbol{y}}_i\| \le \epsilon_i \ \ (i = 1, 2)$$

hold for some $\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, \hat{\boldsymbol{u}}_1$ and $\hat{\boldsymbol{u}}_2$. The equality constraint $\hat{\boldsymbol{y}}_1 = \hat{\boldsymbol{u}}_2$ is in general not met, unless $\epsilon_1 = 0$ (meaning that $\hat{\Sigma}_1$ is strictly similar to Σ_1) and $\hat{\Sigma}_2 = \Sigma_2$. Thus, the conventional bisimulation is not compatible even with this simple interconnection.

In the following, we propose a new variation of approximate bisimulation that is compatible to interconnection operation.

Definition 6. *Interconnection-compatible approximate simulation*
Suppose subsystems $\Sigma_i\langle X_i, U_i, Y_i, f_i, h_i\rangle$, $\hat{\Sigma}_i\langle \hat{X}_i, \hat{U}_i, \hat{Y}_i, \hat{f}_i, \hat{h}_i\rangle$, and a set of positive constants

$$\boldsymbol{\epsilon}_i = \{\epsilon_i^y,\ \{\epsilon_{ij}^w\}_{j\in\{1,\ldots,N\}\backslash\{i\}},\ \{\epsilon_{ij}^z\}_{j\in\{1,\ldots,N\}\backslash\{i\}}\} \tag{13}$$

are given. A binary relation $R_i \in X_i \times \hat{X}_i$ is an interconnection-compatible (IC in short) $\boldsymbol{\epsilon}_i$-approximate simulation relation from Σ_i to $\hat{\Sigma}_i$ if and only if for every $(\boldsymbol{x}_i, \hat{\boldsymbol{x}}_i) \in R_i$, the following holds:
for all $\boldsymbol{u}_i = [\boldsymbol{u}_i^{\text{ext}}; \boldsymbol{w}_i]$ $(\boldsymbol{w}_i = [\boldsymbol{w}_{i1}; \ldots; \boldsymbol{w}_{iN}])$, there exists a $\hat{\boldsymbol{u}}_i^{\text{ext}}$ that satisfies the following two conditions.
1. For $[\boldsymbol{z}_{i1}; \ldots; \boldsymbol{z}_{iN}] = h_i^z(\boldsymbol{x}_i, \boldsymbol{u}_i^{\text{ext}})$ and $[\hat{\boldsymbol{z}}_{i1}; \ldots; \hat{\boldsymbol{z}}_{iN}] = \hat{h}_i^z(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{u}}_i^{\text{ext}})$,

$$\|\boldsymbol{z}_{ij} - \hat{\boldsymbol{z}}_{ij}\| \le \epsilon_{ij}^z \ \ (j \in \{1, \ldots, N\}\backslash\{i\}). \tag{14}$$

2. For all $\hat{\boldsymbol{w}}_i = [\hat{\boldsymbol{w}}_{i1}; \ldots; \hat{\boldsymbol{w}}_{iN}]$ within the range $\|\boldsymbol{w}_{ij} - \hat{\boldsymbol{w}}_{ij}\| \le \epsilon_{ij}^w$,

$$\|h_i^y(\boldsymbol{x}_i, \boldsymbol{u}_i^{\text{ext}}, \boldsymbol{w}_i) - \hat{h}_i^y(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{u}}_i^{\text{ext}}, \hat{\boldsymbol{w}}_i)\| \le \epsilon_i^y \ \ (j \in \{1, \ldots, N\}\backslash\{i\}) \tag{15}$$

and

$$(f_i(\boldsymbol{x}_i, \boldsymbol{u}_i^{\text{ext}}, \boldsymbol{w}_i),\ \hat{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{u}}_i^{\text{ext}}, \hat{\boldsymbol{w}}_i)) \in R_i \tag{16}$$

hold.

Definition 7. *Interconnection-compatible approximate bisimulation*
Suppose subsystems $\Sigma_i\langle X_i, U_i, Y_i, f_i, h_i\rangle$, $\hat{\Sigma}_i\langle \hat{X}_i, \hat{U}_i, \hat{Y}_i, \hat{f}_i, \hat{h}_i\rangle$ *and a set of positive constants* $\boldsymbol{\epsilon}_i$ *defined as in (13) are given. A binary relation* $R_i \subset X_i \times \hat{X}_i$ *is called an IC* $\boldsymbol{\epsilon}_i$*-approximate bisimulation relation between* Σ_i *and* $\hat{\Sigma}_i$ *if and only if* R_i *is an IC* $\boldsymbol{\epsilon}_i$*-approximate simulation relation from* Σ_i *to* $\hat{\Sigma}_i$ *and its inverse relation* R_i^{-1} *is an IC* $\boldsymbol{\epsilon}_i$*-approximate simulation relation from* $\hat{\Sigma}_i$ *to* Σ_i*. Moreover, if such an* R_i *exists,* Σ_i *and* $\hat{\Sigma}_i$ *are said to be IC-approximately bisimilar with respect to* R_i *and the precision* $\boldsymbol{\epsilon}_i$*, and this relation is denoted by* $\Sigma_i \sim^{\mathcal{I}}_{\boldsymbol{\epsilon}_i} \hat{\Sigma}_i$.

The major difference between the above definition and the ordinary approximate bisimulation is that, the internal input signals $\hat{\boldsymbol{w}}_{ij}$ are regarded as disturbances rather than control inputs.

Remark 1. For systems without internal input signals, Definition 7 reduces to the definition of ordinary approximate bisimulation with the output error bound separately specified to $\boldsymbol{y}_i^{\text{ext}}$ and $\boldsymbol{z}_{ij}$s. For systems without external input signals, Definition 7 becomes a bounded output error condition under bounded disturbances.

The following theorem states that IC-approximately bisimilar abstractions are actually compatible with interconnection.

Theorem 1. *Suppose* N *subsystems* Σ_i $(i = 1, \ldots, N)$ *are given, and for each of them,* $\hat{\Sigma}_i$ *is an IC-approximately bisimilar abstraction with respect to the binary relation* R_i *and the precision set* $\boldsymbol{\epsilon}_i$*. If the condition*

$$\epsilon_{ij}^{w} \geq \epsilon_{ji}^{z} \ (i \in \{1, 2, \ldots, N\},\ j \in \{1, 2, \ldots, N\}\backslash\{i\}) \tag{17}$$

is satisfied, the two interconnected systems $\mathcal{I}(\Sigma_1, \ldots, \Sigma_N)$ *and* $\mathcal{I}(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_N)$ *are approximately bisimilar (in the sense of Definition 3) with respect to the relation*

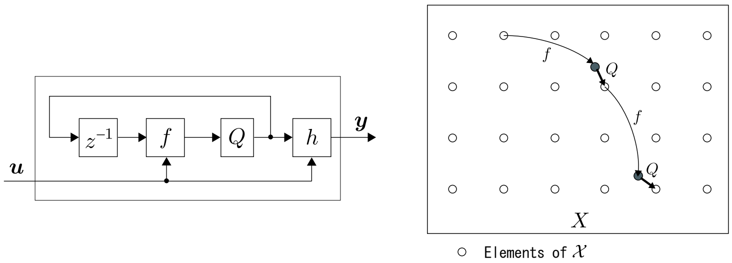
$$R = \{((\boldsymbol{x}_1; \boldsymbol{x}_2; \ldots; \boldsymbol{x}_N), (\hat{\boldsymbol{x}}_1; \hat{\boldsymbol{x}}_2; \ldots; \hat{\boldsymbol{x}}_N)) \,|\, (\boldsymbol{x}_i, \hat{\boldsymbol{x}}_i) \in R_i \ \ (i = 1, 2, \ldots, N)\} \tag{18}$$

and the precision

$$\epsilon = \sum_i \epsilon_i^{y}. \tag{19}$$

The proof is given in Appendix A.

Now let us return to the previous cascade system example and make sure that the cascade of abstractions based on Definition 7 is actually approximately bisimilar to the original system. Suppose $\Sigma_1 \sim^{\mathcal{I}}_{\boldsymbol{\epsilon}_1} \hat{\Sigma}_1$ with respect to the relation R_1 and $\Sigma_2 \sim^{\mathcal{I}}_{\boldsymbol{\epsilon}_2} \hat{\Sigma}_2$ with respect to the relation R_2, where $\boldsymbol{\epsilon}_1 = \{\epsilon_{12}^{z}\}$, $\boldsymbol{\epsilon}_2 = \{\epsilon_2^{y}, \epsilon_{21}^{w}\}$. For any $\boldsymbol{u}_1^{\text{ext}}$, there exists $\hat{\boldsymbol{u}}_1^{\text{ext}}$ that satisfies $\|\boldsymbol{z}_{12} - \hat{\boldsymbol{z}}_{12}\| \leq \epsilon_{12}^{z}$ (and vice versa). Moreover, the condition $\|\boldsymbol{y}_2^{\text{ext}} - \hat{\boldsymbol{y}}_2^{\text{ext}}\| \leq \epsilon_2^{y}$ holds as long as the condition $\|\boldsymbol{w}_{21} - \hat{\boldsymbol{w}}_{21}\| \leq \epsilon_{21}^{w}$ is satisfied. Therefore, if the condition $\epsilon_{21}^{w} \geq \epsilon_{12}^{z}$



(a) Block diagram of $Q(\Sigma)$ (b) State transition of $Q(\Sigma)$

Fig. 3. System with state quantizer.

holds, the cascade of $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ is approximately bisimilar to the cascade of Σ_1 and Σ_2 with the precision $\epsilon = \epsilon_2^y$.

In order to design the abstractions of subsystems that preserve approximate bismilarity under interconnection, one should first divide the input-output signals into groups of those used for external interface and those used for interconnection, and then design the abstractions satisfying the conditions stated in Theorem 1. Moreover, the precision parameter of each abstraction should be chosen according to the condition (17). The proposed method enables us to design the abstraction of each subsystem in a separate way by regarding errors on internal signals as disturbances. One should keep in mind, however, that this could produce a conservative result compared to designing the abstraction by viewing the original interconnected system as a whole, if such a method is available.

4 Finite abstractions of linear subsystems

In the previous section, we have introduced the notion of interconnection-compatible approximate bisimulation. As the next step, in this section we address the design of approximately bisimilar finite abstractions of subsystems of interconnected systems. The term *finite abstraction* refers to a finite state system that approximates a continuous-state system.

4.1 Expression of finite abstractions via state quantization

One of the most important issues of the finite abstraction problem is about the expression of finite automata. In the following, we propose a way of expressing finite automata via *state-quantization* of continuous-state systems.

First of all, a quantization function is defined as follows.

$$Q : X \mapsto \mathcal{X}. \tag{20}$$

Here, the set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ is a finite subset of X.

The following definition introduces the notion of finite automata induced by the state-quantization of continuous-state systems.

Definition 8. *Finite automata induced by state quantization*
Consider a dynamical system $\Sigma\langle X, U, Y, f, h\rangle$ and a quantization function Q : $X \mapsto \mathcal{X}$. The following system is called a finite automaton induced by the state quantization of Σ, denoted by $Q(\Sigma)$.

$$Q(\Sigma) \,:\, \begin{cases} \boldsymbol{x}(t+1) = Q(f(\boldsymbol{x}(t), \boldsymbol{u}(t))) \\ \boldsymbol{y}(t) = h(\boldsymbol{x}(t), \boldsymbol{u}(t)) \end{cases} \quad (\boldsymbol{x}(0) \in \mathcal{X}). \tag{21}$$

The block diagram and the state transition of a $Q(\Sigma)$ is illustrated in Fig. 3. Notice that, in this state equation, the state transition is closed in $\mathcal{X}$ as long as the initial state is chosen from $\mathcal{X}$. To clarify this property, we define the subset of the input set as

$$\mathcal{U}_{ij} = \{\boldsymbol{u} \in U \,|\, Q(f(\boldsymbol{x}_i, \boldsymbol{u})) = \boldsymbol{x}_j\}, \tag{22}$$

which refers to the set of control inputs that drives the state $\boldsymbol{x}_i$ to $\boldsymbol{x}_j$. Note that from the property of the quantization function Q, $\{\mathcal{U}_{ij}\}_j$ forms a partition of U for each i. Using this notation, the state transition of $Q(\Sigma)$ is rewritten as

$$\boldsymbol{x}(t) = \boldsymbol{x}_i \;\wedge\; \boldsymbol{u}(t) \in \mathcal{U}_{ij} \;\Rightarrow\; \boldsymbol{x}(t+1) = \boldsymbol{x}_j. \tag{23}$$

In this way, the input set is also discretized as a class induced from the state transition over finite states. Therefore, the partition of the input set is dependent on the current state. This implicit fashion of the input discretization differs from the other researches (like [7],[8]), where explicit input quantization or originally discrete input systems are considered. Moreover, in the case that the measurement function is a function of states only (written as $h(\boldsymbol{x})$), the state quantization results in indirect quantization of the output set; i.e., $Y \mapsto \mathcal{Y} = \{h(\boldsymbol{x}) \,|\, \boldsymbol{x} \in \mathcal{X}\}$.

4.2 Approximate bisimulation condition of finite abstraction

In this subsection, we address the design of finite abstractions of linear subsystems. Linear subsystems are expressed as follows.

$$\Sigma_i \,:\, \begin{cases} \boldsymbol{x}_i(t+1) = A_i\boldsymbol{x}_i(t) + B_i^u\boldsymbol{u}_i^{\mathrm{ext}}(t) + \displaystyle\sum_{j\in\{1,2,\dots,N\}\setminus\{i\}} B_{ij}^w\boldsymbol{w}_{ij}, \\ \boldsymbol{y}_i^{\mathrm{ext}}(t) = C_i^y\boldsymbol{x}_i(t) + D_i^{yu}\boldsymbol{u}_i^{\mathrm{ext}}(t) + \displaystyle\sum_{j\in\{1,2,\dots,N\}\setminus\{i\}} D_{ij}^{yw}\boldsymbol{w}_{ij}, \\ \boldsymbol{z}_{ij}(t) = C_{ij}^z\boldsymbol{x}_i(t) \qquad (j \in \{1,2,\dots,N\}\setminus\{i\}). \end{cases} \tag{24}$$

We assume that the state set X_i of Σ_i is bounded. For systems of this form, we express their abstractions as state-quantized systems $Q_i(\Sigma_i)$, defined in the previous subsection. Then the problem of concern reduces to deriving a quantization function Q_i whose resultant $Q_i(\Sigma_i)$ is IC-approximately bisimilar to Σ_i with respect a binary relation R_i satisfying the condition.

$$\text{For any } \boldsymbol{x} \in X_i, \text{ there exists a } \hat{\boldsymbol{x}} \in \mathcal{X}_i \text{ such that } (\boldsymbol{x}, \hat{\boldsymbol{x}}) \in R_i, \tag{25}$$

where $\mathcal{X}_i$ denotes the state set of $Q_i(\Sigma_i)$.

Remark 2. The condition (25) is necessary for applying bisimilar abstractions to actual analysis and control problems, assuming that the initial state is arbitrarily chosen from X_i. In [11], this condition is imposed in the definition of approximate (bi)simulation.

Theorem 2. *Interconnection-compatible approximately bisimilar finite abstractions of linear subsystems*
Let Σ_i be an (A_i, B_i^u)-stabilizable discrete-time linear system defined by (24) and let $\boldsymbol{\epsilon}_i$ be a set of positive constants defined by (13). There exist a matrix F_i, a positive definite matrix M_i and a constant $\lambda_i \in (0,1)$ satisfying the conditions

$$(A_i + B_i^u F_i)^{\mathrm{T}} M_i (A_i + B_i^u F_i) \leq \lambda_i^2 M_i, \tag{26}$$

$$\begin{aligned} M_i &\geq \frac{1}{(1-\lambda_i)^2 \alpha_i^2} (C_i^y + D_i^{yu} F_i)^{\mathrm{T}} (C_i^y + D_i^{yu} F_i), \\ M_i &\geq \frac{1}{(1-\lambda_i)^2 {\epsilon_{ij}^z}^2} (C_{ij}^z + D_{ij}^{zu} F_i)^{\mathrm{T}} (C_{ij}^z + D_{ij}^{zu} F_i) \ \ (j \in \{1,2,\ldots,N\} \backslash \{i\}) \end{aligned} \tag{27}$$

where the constant α_i is given by

$$\alpha_i := \epsilon_i^y - \sum_{j \in \{1,2,\ldots,N\} \backslash \{i\}} ||D_{ij}^{yw}|| \epsilon_{ij}^w. \tag{28}$$

Furthermore, if α_i and the constant defined as

$$\beta_i := 1 - \sum_{j \in \{1,2,\ldots,N\} \backslash \{i\}} ||{B_{ij}^w}^{\mathrm{T}} M_i B_{ij}^w|| \epsilon_{ij}^w \tag{29}$$

are both positive, then for a quantization function Q_i satisfying the condition

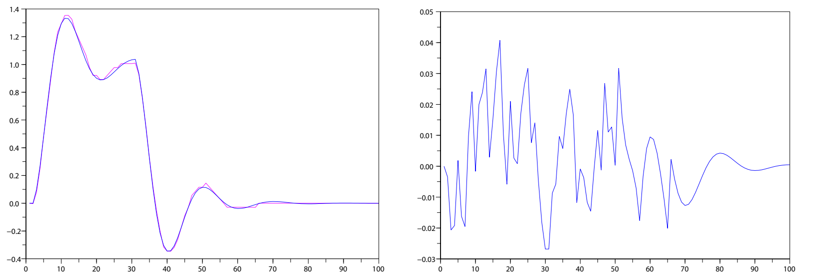
$$||\boldsymbol{x}_i - Q_i(\boldsymbol{x}_i)||_{M_i} \leq \beta_i \ \ \forall \boldsymbol{x}_i \in X_i, \tag{30}$$

the systems Σ_i and $Q_i(\Sigma_i)$ are IC-approximately bisimilar with respect to the precision $\boldsymbol{\epsilon}_i$ and the relation

$$R_i = \{(\boldsymbol{x}, \hat{\boldsymbol{x}}) \,|\, ||\boldsymbol{x} - \hat{\boldsymbol{x}}||_{M_i} \leq 1/(1-\lambda_i)\}, \tag{31}$$

which satisfies (25).

A rough explanation of Theorem 2 is as follows: Bisimulation can be captured as a tracking problem of two systems. If one system can track the other system's output trajectory with a constant error bound (say ϵ), then this system is similar to the other with the precision ϵ. Moreover, if both systems can track their opponent's trajectory, then they are bisimilar to each other. Therefore, in those cases when both systems are linear, the problem can be viewed as the stabilization problem of the error system, and in such cases, the bisimulation relation is related to the invariant set of the error system. A detailed proof is given in Appendix B.



(a) Output response.

(b) Difference of the output trajectories of $\mathcal{I}(\hat{\Sigma}_1, \Sigma_2)$ and $\mathcal{I}(\Sigma_1, \Sigma_2)$.

Fig. 4. Output response of $\mathcal{I}(\hat{\Sigma}_1, \Sigma_2)$ and $\mathcal{I}(\Sigma_1, \Sigma_2)$.

Finally, an explicit expression of the quantizer Q_i satisfying the condition (30) is given as follows.

$$Q_i(\boldsymbol{x}) = \left(\frac{\sqrt{n}}{2}U_i\right)^{-1}\left[\left(\frac{\sqrt{n}}{2}U_i\right)\boldsymbol{x}\right] \tag{32}$$

Here, the matrix U_i is given by $U_i^{\mathrm{T}}U_i = M_i/\beta_i^2$, n is the size of $\boldsymbol{x}$ and $[\boldsymbol{x}]$ is the rounding function, which maps each element of $\boldsymbol{x}$ to its nearest integer.

4.3 Example

This section shows a simple example. Consider an interconnected system composed of two subsystems. The parameters of the subsystems are given as follows.

$$\Sigma_1 : \begin{cases} \boldsymbol{x}_1(t+1) = \begin{bmatrix} 1.0 & 0.1 \\ -1.0 & 0.7 \end{bmatrix} \boldsymbol{x}_1(t) + \begin{bmatrix} 0.0 & 0.0 \\ 1.0 & 0.1 \end{bmatrix} \begin{bmatrix} u_1^{\mathrm{ext}}(t) \\ w_{12} \end{bmatrix} \\ \begin{bmatrix} y_1^{\mathrm{ext}} \\ z_{12} \end{bmatrix}(t) = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 0.1 \end{bmatrix} \boldsymbol{x}_1(t) \end{cases}$$

$$\Sigma_2 : \begin{cases} \boldsymbol{x}_2(t+1) = \begin{bmatrix} 0.7 & 0.1 \\ -1.0 & 0.5 \end{bmatrix} \boldsymbol{x}_2(t) + \begin{bmatrix} 0.0 \\ 0.2 \end{bmatrix} w_{21}(t) \\ z_{21}(t) = \begin{bmatrix} 0.0 & 0.2 \end{bmatrix} \boldsymbol{x}_2(t) \end{cases}$$

Let us derive an abstraction of this system by first designing a finite abstraction of Σ_1 only, and next making its interconnection with Σ_2.

The following parameters are chosen to meet the conditions in Theorem 2.

$$\epsilon_1^y = 0.1,\ \epsilon_{12}^z = \epsilon_{21}^w = 0.1,\ \epsilon_{21}^z = \epsilon_{12}^w = 0.05$$

$$M_1 = \begin{bmatrix} 2.3\mathrm{e}3 & 2.2\mathrm{e}2 \\ 2.2\mathrm{e}2 & 4.3\mathrm{e}1 \end{bmatrix},\ F_1 = \begin{bmatrix} -4.1 & -1.2 \end{bmatrix},\ \lambda_1 = 0.7$$

$$M_2 = \begin{bmatrix} 1.6\mathrm{e}3 & 1.6\mathrm{e}2 \\ 1.6\mathrm{e}2 & 1.9\mathrm{e}2 \end{bmatrix},\ \lambda_2 = 0.7$$

Substituting the above values of M_1 and $\beta_1 = 0.97$ to (32) we obtain the quantizer Q_1 and hence $\hat{\Sigma}_1 = Q_1(\Sigma_1)$.

Fig. 4 shows the output response of the two systems $\mathcal{I}(\hat{\Sigma}_1, \Sigma_2)$ and $\mathcal{I}(\Sigma_1, \Sigma_2)$. The input signal

$$u_1^{\mathrm{ext}}(t) = \begin{cases} 1.0 & (t < 30) \\ 0.0 & (t \geq 30) \end{cases}$$

is applied to $\mathcal{I}(\Sigma_1, \Sigma_2)$, and the input signal for $\mathcal{I}(\hat{\Sigma}_1, \Sigma_2)$ is given by $\hat{u}_1^{\mathrm{ext}} = u_1^{\mathrm{ext}}(t) + F_1(\hat{\boldsymbol{x}}_1(t) - \boldsymbol{x}_1(t))$. It is observed from the figure that the specified output error bound is achieved.

5 Conclusion

In this paper we discussed the design of finite abstractions of interconnected systems. In a general setting of interconnected systems, we introduced an extended notion of approximate bisimulation, which is compatible with interconnection. This means that the abstractions of subsystems that are based on the presented approximate bisimulation can be connected with each other to form an abstractions of the whole system. We have also presented a design procedure for the finite abstraction of linear subsystems under this new notion of approximate bisimulation. In future works, we should extend the class of subsystems whose finite abstractions are computable.

References

1. G. Lafferriere, G.J. Pappas and S. Sastry : Hybrid Systems with Finite Bisimulations, P. Antsaklis et al. (Eds.) : Hybrid Systems V, LNCS 1567, pp 186-203, 1999.
2. R. Alur, T.D. Verimag and F.Ivančić : Predicate Abstractions for Reachability Analysis of Hybrid Systems, ACM Trans. on Embedded Computing Systems, Vol.5, No.1, pp 152-199, 2006.
3. J. Lunze : A timed discrete-event abstraction of continuous-variable systems, International Journal of Control, 72:13, pp 1147-1164, 1999.
4. K. Tsumura : Stabilization of Linear Systems by Bit-Memory Controllers under Constraints of Bit-Length, 45th IEEE Conference on Decision and Control, pp5507-5512 , San Diego, CA, USA, December, 2006.
5. R. Milner : Communication and Concurrency, Prentice-Hall, 1989.
6. A. Girard and G.J. Pappas : Approximation Metrics for Discrete and Continuous Systems, IEEE Transactions on Automatic Control, 52(5), pp782-798, 2007.
7. A. Girard : Approximately Bisimilar Finite Abstractions of Stable Linear Systems, Hybrid Systems: Computation and Control, vol 4416 in LNCS, pp 231-244, Springer, 2007.
8. P. Tabuada : Approximate Simulation Relations and Finite Abstractions of Quantized Control Systems, Hybrid Systems: Computation and Control, vol 4416 in LNCS, pp 529-542, Springer, 2007.

9. P. Tabuada, G.J. Pappas and P. Lima : Composing Abstractions of Hybrid Systems, Hybrid Systems: Computation and Control, vol 2289 in LNCS, pp 436-450, Springer, 2002.
10. P. Tabuada, G.J. Pappas and P. Lima : Compositional Abstractions of Hybrid Control Systems, Discrete Event Dynamic Systems : Theory and Applications, 14, pp 203-238, 2004.
11. A.A. Julius and G.J. Pappas: Approximate Equivalence and Approximate Synchronization of Metric Transition Systems, in the Proc. 45th IEEE Conf. Decision and Control 2006, San Diego, USA.
12. Y. Tazaki and J. Imura : Finite Abstractions of Discrete-time Linear Systems and Its Application to Optimal Control, *to appear in* 17th IFAC World Congress, Seoul, Korea, 2008.

A Proof of Theorem 1

We only prove the approximate similarity from $\mathcal{I}(\Sigma_1, \ldots, \Sigma_N)$ to $\mathcal{I}(\hat{\Sigma}_1, \ldots, \hat{\Sigma}_N)$. The opposite case is treated in the same manner. Suppose, for each i, the state $\boldsymbol{x}_i$ of the subsystem Σ_i and the state $\hat{\boldsymbol{x}}_i$ of its abstraction $\hat{\Sigma}_i$ are in the relation R_i. Moreover, suppose that the input $\bar{\boldsymbol{u}} = [\boldsymbol{u}_1^{\rm ext}; \ldots; \boldsymbol{u}_N^{\rm ext}]$ of $\mathcal{I}(\Sigma_1, \ldots, \Sigma_N)$ is arbitrarily chosen. Then, for each i, the internal output, internal input, external output and state transition are subsequently determined as $\boldsymbol{z}_i = h_i^z(\boldsymbol{x}_i, \boldsymbol{u}_i^{\rm ext})$, $\boldsymbol{w}_{ij} = \boldsymbol{z}_{ji}$, $\boldsymbol{y}_i^{\rm ext} = h_i^y(\boldsymbol{x}_i, \boldsymbol{u}_i^{\rm ext}, \boldsymbol{w}_i)$, and $\boldsymbol{x}_i' = f_i(\boldsymbol{x}_i, \boldsymbol{u}_i^{\rm ext}, \boldsymbol{w}_i)$. From the approximate bisimilarity of Σ_i and $\hat{\Sigma}_i$, there exists $\hat{\bar{\boldsymbol{u}}} = [\hat{\boldsymbol{u}}_1^{\rm ext}; \ldots; \hat{\boldsymbol{u}}_N^{\rm ext}]$ satisfying (14) and also, under the assumption

$$\|\boldsymbol{w}_{ij} - \hat{\boldsymbol{w}}_{ij}\| \leq \epsilon_{ij}^w, \tag{33}$$

satisfying (15) and (16). From the condition (17), the assumption (33) is actually fulfilled. Finally, it is straightforward from (15) that $\|\bar{\boldsymbol{y}} - \hat{\bar{\boldsymbol{y}}}\| \leq \epsilon$ holds with ϵ defined by (19).

□

B Proof of Theorem 2

From the stabilizability assumption, there exists a matrix F_i making the eigenvalues of $(A_i + B_i^u F_i)$ strictly inside the unit circle and hence one can show that M_i and λ_i satisfying (26),(27) exist by following the line of Proposition 3 in [6]. Let us denote the state, the input and the output of $Q_i(\Sigma_i)$ by $\hat{\boldsymbol{x}}_i$, $\hat{\boldsymbol{u}}_i$ and $\hat{\boldsymbol{y}}_i$, respectively. We first derive the condition for $Q_i(\Sigma_i)$ to be approximately similar to Σ_i. Taking the difference of the state equations of $Q_i(\Sigma_i)$ and Σ_i, we obtain the error system

$$\boldsymbol{e}_i(t+1) = A_i \boldsymbol{e}_i(t) + B_i^u \delta \boldsymbol{u}_i^{\rm ext}(t) + \sum_{j \in \{1,2,\ldots,N\} \setminus \{i\}} B_{ij}^w \delta \boldsymbol{w}_{ij} + \boldsymbol{d}_i(t) \tag{34}$$

where $\boldsymbol{e}_i = \hat{\boldsymbol{x}}_i - \boldsymbol{x}_i$, $\delta\boldsymbol{u}_i^{\text{ext}} = \hat{\boldsymbol{u}}_i^{\text{ext}} - \boldsymbol{u}_i^{\text{ext}}$, $\delta\boldsymbol{w}_{ij} = \hat{\boldsymbol{w}}_{ij} - \boldsymbol{w}_{ij}$ and $\boldsymbol{d}_i$ is the quantization error defined by $\boldsymbol{d}_i(t) = Q_i(A_i\hat{\boldsymbol{x}}_i(t)+B_i\hat{\boldsymbol{u}}_i(t))-(A_i\hat{\boldsymbol{x}}_i(t)+B_i\hat{\boldsymbol{u}}_i(t))$. Here, we specify the control input $\hat{\boldsymbol{u}}_i^{\text{ext}}(t)$ as a function of $\boldsymbol{x}_i(t), \hat{\boldsymbol{x}}_i(t)$ and $\boldsymbol{u}_i^{\text{ext}}(t)$ defined by $\hat{\boldsymbol{u}}_i^{\text{ext}}(t) = \boldsymbol{u}_i^{\text{ext}}(t) + F_i(\hat{\boldsymbol{x}}_i(t) - \boldsymbol{x}_i(t))$ where F_i is a matrix making the matrix $(A_i + B_i^u F_i)$ asymptotically stable. Then the error dynamics is written as

$$\boldsymbol{e}_i(t+1) = (A_i + B_i^u F_i)\boldsymbol{e}_i(t) + \sum_{j\in\{1,2,\ldots,N\}\setminus\{i\}} B_{ij}^w \delta\boldsymbol{w}_{ij}(t) + \boldsymbol{d}_i(t) \tag{35}$$

and moreover, the following inequality holds.

$$\|\boldsymbol{e}_i(t+1)\|_{M_i} \le \lambda_i \|\boldsymbol{e}_i(t)\|_{M_i} + 1 - \beta_i + \|\boldsymbol{d}_i(t)\|_{M_i}.$$

Here, M_i is the positive definite matrix satisfying (26),(27) and the constant β_i is defined by (29). From the condition $\|\boldsymbol{d}_i(t)\|_{M_i} \le \beta_i$, the set E_i defined as

$$E_i = \{\boldsymbol{e} \,|\, \|\boldsymbol{e}\|_{M_i} \le 1/(1-\lambda_i)\} \tag{36}$$

is an invariant set of the error system. Moreover, from (27) and (28), every element $\boldsymbol{e} \in E_i$ satisfies the conditions

$$\|\hat{\boldsymbol{y}}_i^{\text{ext}} - \boldsymbol{y}_i^{\text{ext}}\| \le \|(C_i^y + D_i^{yu} F_i)\boldsymbol{e}\| + \sum_{j\in\{1,2,\ldots,N\}\setminus\{i\}} \|D_{ij}^{yw}\| \epsilon_{ij}^w \le \epsilon_i^y,$$
$$\|\hat{\boldsymbol{z}}_{ij} - \boldsymbol{z}_{ij}\| = \|(C_{ij}^z + D_{ij}^z F_i)\boldsymbol{e}\| \le \epsilon_{ij}^z \qquad (j = 1, \ldots, N,\ j \ne i).$$

Therefore it follows that the binary relation R_i defined as $R_i = \{(\boldsymbol{x}, \hat{\boldsymbol{x}}) \,|\, (\hat{\boldsymbol{x}} - \boldsymbol{x}) \in E_i\}$, which is written equivalently as (31), is an IC-approximate simulation relation from Σ_i to $Q_i(\Sigma_i)$ with the precision $\boldsymbol{\epsilon}_i$.

In the opposite case, choosing the control input $\boldsymbol{u}_i^{\text{ext}}(t)$ as $\boldsymbol{u}_i^{\text{ext}}(t) = \hat{\boldsymbol{u}}_i^{\text{ext}}(t) + F_i(\boldsymbol{x}_i(t) - \hat{\boldsymbol{x}}_i(t))$ yields the same error system as (35). Therefore, the relation (31) is an IC-approximate bisimulation relation between Σ_i and $Q_i(\Sigma_i)$.

□