J Control Theory Appl 2010 8 (1) 27–33
DOI 10.1007/s11768-010-9186-8

Asymptotic stabilization of dynamically quantized nonlinear systems in feedforward form

Qiang LING [1], Michael D. LEMMON [2], Hai LIN [3]

(1.Department of Automation, University of Science and Technology of China, Hefei Anhui 230027, China;

2.Department of Electrical Engineering, University of Notre Dame, Notre Dame IN 46556, USA;

3.Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576)

Abstract: This paper studies the stabilizability of an n-dimensional quantized feedforward nonlinear system. The state of that system is first quantized into a finite number of bits, and then sent through a digital network to the controller. We want to minimize the number of transmitted bits subject to maintaining asymptotic stability. In the prior literature, n bits are used to stabilize the n-dimensional system by assigning one bit to each state variable (dimension). Under the stronger assumption of global Lipschitz continuity, this paper extends that result by stabilizing the system with a single bit. Its key contribution is a dynamic quantization policy which dynamically assigns the single bit to the most "important" state variable. Under this policy, the quantization error exponentially converges to 0 and the stability of the system can, therefore, be guaranteed. Because 1 is the minimum number of quantization bits (per sampling step), the proposed dynamic quantization policy achieves the minimum stabilizable bit number for that n-dimensional feedforward nonlinear system.

Keywords: Quantization; Stability; Nonlinear; Feedforward

1 Introduction

Consider an n-dimensional nonlinear system in the following feedforward form [1],

$$\dot{x} = f(x,u) = \begin{pmatrix} f_1(X_2,u) \\ f_2(X_3,u) \\ \vdots \\ f_{n-1}(X_n,u) \\ f_n(u) \end{pmatrix}, \tag{1}$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$ and the set of state variables $x_i, x_{i+1}, \cdots, x_n$ is denoted by X_i. When the above nonlinear system is controlled over a digital network, a typical configuration is shown in Fig.1.

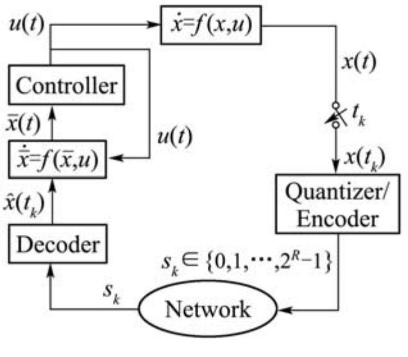

Fig. 1 Quantized nonlinear control systems.

Now we explain the signal flow in Fig.1. At sampling instants $\{t_k\}_{k=0}^{\infty}$, the state $x(t_k)$ is measured, quantized (encoded) into a symbol with R bits, $s_k(\in S = \{0, 1, \cdots, 2^R-1\})$, and transmitted over a digital network. The sampling instants are assumed to satisfy

$$0 < T_m \leqslant t_{k+1} - t_k \leqslant T_M < \infty,\ \ \forall k \geqslant 0. \tag{2}$$

It is assumed that the transmitted symbol s_k is correctly received without delay. The received symbol s_k is used to construct an estimate of the state $x(t_k)$, $\hat{x}(t_k)$. Of course, $\hat{x}(t_k)$ may be different from $x(t_k)$ due to quantization error. $\hat{x}(t_k)$ is used to generate a continuous-time state estimate $\bar{x}(t)$. The controller will make use of $\bar{x}(t)$, instead of the true state $x(t)$, to devise the control $u(t)$ [2].

This paper addresses the following questions. Does there exist an appropriate quantization policy to maintain its stability under finite R? What is the minimum quantization bit number R required to maintain stability? These two questions have generated much interest in the past few years.

In [2], nonlinear systems more general than the one in Eq.(1) are investigated. It is shown that any nonlinear control system that can be globally asymptotically stabilized by true state feedback can also be globally asymptotically stabilized by quantized state feedback, under the condition that the number of quantization bits, R, is big enough. In [3], it is shown that a finite number of quantization bits can make a class of nonlinear systems input-to-state stable (ISS) with respect to measurement errors. More quantization bits, however, mean that more network bandwidth is occupied. Therefore, it makes much sense to determine the smallest R that still asymptotically stabilize the control system. The minimality of the number of quantization bits (per sampling step) required to stabilize a nonlinear system is addressed in [4], where a notion of topological feedback en-

Received 7 September 2009.
This work was supported by the National Natural Science Foundation of China (No.60904012), the Natural Science Foundation of Anhui Province (No.090412050) and the Talent Development Program of Anhui Province (No.2008Z012).

tropy (TFE) is introduced and it is proven that a system can be stabilized if and only if R exceeds the inherent TFE of that system. When the concerned system is linear, there are many ways to compute the TFE and the required minimum bit number (see [5, 6] and references therein). When a system is nonlinear, there is no systematic approach to computing its TFE and the minimum bit number to stabilize a general nonlinear system is usually unknown. Researchers, therefore, pursue a less aggressive goal: to stabilize a nonlinear system with as few quantization bits as possible. In order to save quantization bits, the knowledge of the concerned system has to be taken into account. The nonlinear system in Eq.(1) takes an upper triangular structure, which falls into the class of the feedforward systems [7]. For this type of n-dimensional systems, $R = n$ ($R = n + 1$) can be enough to achieve semiglobal asymptotic stabilization (global stabilization) [1] under three assumptions.

Assumption 1 Functions $f_i(X_{i+1}, u)$, with $i = 1, 2, \cdots, n-1$ and $f_n(u)$ are locally Lipschitz.

Assumption 2 There exists a constant $U > 0$ for which $u(t) < U$ for all $t \geqslant t_0$.

Assumption 3 Each function $f_i(\cdot)$ ($i = 1, 2, \cdots, n$) is zero at the origin and is such that the linearization of Eq.(1) at the origin exists and is stabilizable; there exist class-$\mathcal{K}_+$ function $\phi_i(\cdot)$ for which *

$$|f_i(X_{i+1}, u+v) - f_i(X_{i+1}, u)| \leqslant \phi_i(|(X_{i+1}, u)|)|v|.$$

The results in [1] are quite significant in the sense that the provided bit number is independent of the set of initial conditions of the system, of the time-varying sampling period and can be simply assessed from the dimension of the system.

$R = n$ bits are used in [1] to guarantee asymptotic stability. Is that possible to use fewer bits to accomplish that task? As R is the number of transmitted bits, it has a hard lower bound

$$R \geqslant 1. \tag{3}$$

The present paper proposes a dynamic quantization policy that uses a single bit to globally asymptotically stabilize the n-dimensional nonlinear system in Eq.(1) by strengthening Assumption 1 into:

Assumption 4 Functions $f_i(X_{i+1}, u)$, with $i = 1, 2, \cdots, n-1$ are globally Lipschitz with respect to X_{i+1} for bounded u. Function $f_n(u)$ is locally Lipschitz.

Because of the hard lower bound in Eq.(3), we know the minimum bit number has already been achieved, of course, at the cost of a stronger assumption. Now we remark on that achievement. In [1], the system is n-dimensional and there are n bits. Each dimension is assigned 1 bit. In this paper, there is only 1 bit, which is assigned to the most needed dimension at every time step. Its bit assignment is dynamic, compared with the static policy in [1]. We will show that it is the dynamic bit assignment policy that makes the best use of the available single bit. This policy for the nonlinear systems is motivated by the dynamic bit assignment policy for linear systems [8].

This paper is organized as follows. In Section 2, we present the dynamic quantization policy, which is the major difference from [1]. It is shown that the quantization error exponentially converges to 0 as [1]. Based on this convergence property, we prove the asymptotic stability of the feedforward nonlinear systems. In Section 3, the paper is concluded with some final remarks. In order to improve readability, we move technical proofs into the appendix.

2 Main results: Dynamic quantization policy

2.1 Uncertainty region of the state

The quantizer/ encoder is usually connected with sensors and can know exactly the state at the sampling instants, $x(t_k)$. However, the decoder is spatially separated from the sensors, so it cannot know the exact value of $x(t_k)$. However, the decoder keeps receiving state symbols $\{s_k\}$, and can use these symbols to determine an uncertainty region $P(t_k)$ which the state $x(t_k)$ lies in, i.e.,

$$x(t_k) \in P(t_k) = C(t_k) + \mathrm{rect}(L(t_k)), \tag{4}$$

where the uncertainty region $P(t_k)$ is characterized by its centroid $C(t_k)$ and side length vector $L(t_k)$ with

$$\begin{cases} C(t_k) = [C_1(t_k), C_2(t_k), \cdots, C_n(t_k)], \\ L(t_k) = [L_1(t_k), L_2(t_k), \cdots, L_n(t_k)], \\ \mathrm{rect}(L(t_k)) = \prod_{i=1}^{n} \left[-\frac{1}{2}L_i(t_k), \frac{1}{2}L_i(t_k)\right]. \end{cases}$$

Here, $\prod$ stands for Cartesian product. Because of Eq.(4), it is reasonable to estimate $x(t_k)$ as** for the decoder to set

$$\hat{x}(t_k) = C(t_k). \tag{5}$$

So we will use $\hat{x}(t_k)$ to represent the centroid of $P(t_k)$ in the sequel. It can be seen that the estimation error $\tilde{x}(t_k) = x(t_k) - \hat{x}(t_k)$ is bounded as

$$|\tilde{x}_i(t_k)| \leqslant \frac{1}{2}L_i(t_k),\ i = 1, 2, \cdots, n. \tag{6}$$

With the received symbol s_k, the decoder updates its centroid and side length vector as

$$(\hat{x}(t_k), s_k) \to \hat{x}(t_{k+1}),\quad (L(t_k), s_k) \to L(t_{k+1}). \tag{7}$$

Of course, discretion is required to guarantee no overflow would occur, i.e., $x(t_{k+1}) \in \hat{x}(t_{k+1}) + \mathrm{rect}(L(t_{k+1}))$. The symbol s_k in Eq.(7) is sent by the encoder. Thus, the decoder surely knows s_k. As long as the encoder and the decoder agree upon the initial condition $\hat{x}(t_0)$ and $L(t_0)$, they will generate the same sequences $\{\hat{x}(t_k)\}_k$ and $\{L(t_k)\}_k$ under the same updating rule in Eq.(7).

In order to achieve asymptotic stability, i.e., $\lim\limits_{t\to\infty} x(t) = 0$, we have to guarantee the convergence of the continuous-time estimation error $\tilde{x}(t) = x(t) - \bar{x}(t)$. Because of Assumption 2 (the boundedness of the control $u(t)$), Assumption 4 (the global Lipschitz property of $f(\cdot) = [f_1(\cdot), f_2(\cdot), \cdots, f_n(\cdot)]^{\mathrm{T}}$) and Eq.(2) (bounded sampling intervals), we get

Proposition 1 The convergence of $\tilde{x}(t) = x(t) - \bar{x}(t)$, i.e., $\lim\limits_{t\to\infty} \tilde{x}(t) = 0$ is equivalent to

$$\lim_{k\to\infty} \|L(t_k)\|_\infty = 0, \tag{8}$$

* Class-$\mathcal{K}_+$ functions are nonnegative, continuous and nondecreasing functions. For $i = n$, function $\phi_n(\cdot)$ depends on $|u|$ only.

** The estimation in Eq.(5) minimizes the maximum estimation error, which is measured by the infinity norm of the state estimation error.

where $\|\cdot\|_\infty$ denotes the infinity norm.

The proof of Proposition 1 is not difficult and omitted here. Later, we will design a quantization policy so that $L(t_k)$ exponentially converges to 0.

2.2 Dynamic quantization policy

Because of Assumptions 2 and 4, we know, for each i ($\in \{1,2,\cdots,n\}$), there exists a finite positive number F_i such that

$$|f_i(X_i,u)-f_i(Y_i,u)| \leqslant F_i\|X_i-Y_i\|_\infty \tag{9}$$

for any $X_i\in\mathbb{R}^{n-i+1}$, $Y_i\in\mathbb{R}^{n-i+1}$ and $u(t)\leqslant U$.

Choose positive numbers γ and ρ by

$$\sqrt[n]{\frac{1}{2}}<\gamma<1, \tag{10}$$

$$\rho>2+\frac{(n-1)FT_M}{1-\dfrac{1}{\sqrt[n]{2}\gamma}}, \tag{11}$$

where $F=\max\limits_i F_i$. Similar to the quantization policy for a linear system in [9], we propose the following algorithm.

Algorithm 1 Dynamic quantization policy

Encoder/Decoder initialization

Initialize $\hat{x}(t_0)$ and $L(t_0)$ so that $x(t_0)\in\hat{x}(t_0)+\mathrm{rect}(L(t_0))$. Set $\hat{x}_\mathrm{e}(t_0)=\hat{x}(t_0)$, $\hat{x}_\mathrm{d}(t_0)=\hat{x}(t_0)$, $L_\mathrm{e}(t_0)=L(t_0)$, $L_\mathrm{d}(t_0)=L(t_0)$, and $k=0$. Note that the subscripts e and d are used to emphasize that the variables are updated on the encoder and decoder sides, respectively.

Encoder algorithm

1) Select the index I_k by
$$I_k=\arg\max_i \rho^{2i}L_i(t_k), \tag{12}$$
2) Quantize the state $x(t_k)$ by setting
$$s_k=\begin{cases}1, & x_{I_k}(t_k)\geqslant\hat{x}_{I_k}(t_k),\\ 0, & \text{otherwise.}\end{cases}$$
3) Transmit the quantized symbol s_k.
4) Update $L(t_{k+1})$ at time instant t_{k+1} as***
$$L_i(t_{k+1})=\begin{cases}\dfrac{L_i(t_k)}{2}+FT_M\sum\limits_{j=I_k+1}^{n}L_j(t_{k+1}), & i=I_k,\\ L_i(t_k)+FT_M\sum\limits_{j=I_k+1}^{n}L_j(t_{k+1}), & i\neq I_k.\end{cases} \tag{13}$$
$\hat{x}(t_{k+1})$ is updated by running the differential equation in Fig.1 as
$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{x}_{\mathrm{e},i}(t)=f_i(\bar{X}_{\mathrm{e},i+1}(t),u(t)),$$
$$\bar{x}_{\mathrm{e},i}(t_k)=\begin{cases}\hat{x}_i(t_k)+\dfrac{L_i(t_k)}{4}, & i=I_k \text{ and } s_k=1,\\ \hat{x}_i(t_k)-\dfrac{L_i(t_k)}{4}, & i=I_k \text{ and } s_k=0,\\ \hat{x}_i(t_k), & i\neq I_k,\end{cases} \tag{14}$$
where $\bar{X}_{\mathrm{e},i}(t)=[\bar{x}_{\mathrm{e},i}(t),\bar{x}_{\mathrm{e},i+1}(t),\cdots,\bar{x}_{\mathrm{e},n}(t)]^\mathrm{T}$ for $t\in[t_k,t_{k+1})$ and the control $u(t)$ is generated by the controller in Fig.1 with the estimated state $x_\mathrm{e}(t)$ ($=\bar{X}_{\mathrm{e},1}(t)$) in the place of $\bar{x}(t)$. At time $t=t_{k+1}$, update $\hat{x}_i(t_{k+1})$ as
$$\hat{x}_i(t_{k+1})=\bar{x}_{\mathrm{e},i}(t_{k+1}^-),\ \ i=1,2,\cdots,n. \tag{15}$$
5) Update time index, $k=k+1$ and return to step 1).

Decoder algorithm

1) Select the index I_k by
$$I_k=\arg\max_i \rho^{2i}L_i(t_k). \tag{16}$$
2) Wait for quantized data, s_k, from encoder.
3) Update the state estimate at t_k as
$$\hat{x}_{\mathrm{d},i}(t_k):=\begin{cases}\hat{x}_{\mathrm{d},i}(t_k)+\dfrac{L_i(t_k)}{4}, & i=I_k \text{ and } s_k=1,\\ \hat{x}_{\mathrm{d},i}(t_k)-\dfrac{L_i(t_k)}{4}, & i=I_k \text{ and } s_k=0,\\ \hat{x}_{\mathrm{d},i}(t_k), & i\neq I_k.\end{cases}$$
4) Generate the continuous-time state estimate as
$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{x}_{\mathrm{d},i}(t)=f_i(\bar{X}_{\mathrm{d},i+1}(t),u(t)), \tag{17}$$
$$\bar{x}_{\mathrm{d},i}(t_k)=\hat{x}_{\mathrm{d},i}(t_k),$$
where $t\in[t_k,t_{k+1})$.
5) Control variable $u(t)$ is constructed from the controller in Fig.1 by replacing $\bar{x}(t)$ with $\bar{x}_\mathrm{d}(t)(=\bar{X}_{\mathrm{d},1}(t))$.
6) Update $L(t_{k+1})$ at time instant t_{k+1} as
$$L_i(t_{k+1})=\begin{cases}\dfrac{L_i(t_k)}{2}+FT_M\sum\limits_{j=I_k+1}^{n}L_j(t_{k+1}), & i=I_k,\\ L_i(t_k)+FT_M\sum\limits_{j=I_k+1}^{n}L_j(t_{k+1}), & i\neq I_k.\end{cases} \tag{18}$$
At time $t=t_{k+1}$, $\hat{x}_{\mathrm{d},i}(t_{k+1})$ is updated as
$$\hat{x}_{\mathrm{d},i}(t_{k+1})=\bar{x}_{\mathrm{d},i}(t_{k+1}^-),\ i=1,2,\cdots,n. \tag{19}$$
7) Update time index, $k=k+1$, and return to step 1).

Remark 1 Because $L_\mathrm{e}(t_0)=L_\mathrm{d}(t_0)$, $L_\mathrm{e}(t_k)$ and $L_\mathrm{d}(t_k)$ are updated by the same rule in Eqs.(13) and (18) and the transmitted symbol s_k is always received correctly, we have

$$L_\mathrm{e}(t_k)=L_\mathrm{d}(t_k),\ \forall k\geqslant 0. \tag{20}$$

Therefore, we may shorten $L_\mathrm{e}(t_k)$ and $L_\mathrm{d}(t_k)$ into the same variable $L(t_k)$ without confusion. Similarly, we can show

$$\begin{cases}\hat{x}_\mathrm{e}(t_k)=\hat{x}_\mathrm{d}(t_k), & \forall k,\\ \bar{x}_\mathrm{e}(t)=\bar{x}_\mathrm{d}(t), & \forall t\geqslant t_0.\end{cases} \tag{21}$$

Therefore, $\hat{x}_\mathrm{e}(t_k)$ and $\hat{x}_\mathrm{d}(t_k)$ are shortened into $\hat{x}(t_k)$, $\bar{x}_\mathrm{e}(t)$ and $\bar{x}_\mathrm{d}(t)$ into $\bar{x}(t)$ as well. The same $\bar{x}(t)$ is used to compute the control variable by the same rule on both encoder and decoder sides, which surely yields the $u(t)$. Our quantization policy guarantees there is no overflow, which is presented as the following proposition. See the appendix for its proof.

Proposition 2 Under Assumptions 2∼4, we choose γ and ρ by Eqs.(10) and (11). The dynamic quantization policy in Algorithm 1 is implemented to the quantized nonlinear system in Eq.(1). For any $k\geqslant 0$,

$$x(t_k)\in\hat{x}(t_k)+\mathrm{rect}(L(t_k)). \tag{22}$$

Remark 2 In Algorithm 1, the side is measured by the weighted length $\rho^{2i}L_i(t_k)$ rather than the direct length $L_i(t_k)$. That policy assigns the highest priority to the n-th

***To successfully do the computation in Eq.(13), we start from $i=n$, and proceed in the decreasing order of i.

dimension. The motivation lies in the feedforward structure of Eq.(1), i.e., the n-th dimension affects the other dimensions, but NOT reversely. After $L_n(t_k)$ is reduced enough, we get almost precise state estimate $\bar{x}_n(t)$ and the order of the state estimation problem could be reduced by 1, i.e., from n to $n-1$. That rationale keeps working for the remaining dimensions. Of course, some subtle balance has to be made when assigning the single bit among n dimensions, which is carried out by the appropriate choice of ρ in Eq.(11). It will be shown in Proposition 3 that $L(t_k)$ exponentially converges to 0. The proof can be found in the appendix.

Proposition 3 Under Assumptions 2∼4, we choose γ and ρ by Eqs.(10) and (11). The dynamic quantization policy in Algorithm 1 is implemented to the quantized nonlinear system in Eq.(1). The side length vector $L(t_k)$ is bounded as

$$\|L(t_k)\|_\infty \leqslant C\gamma^k \|L(t_0)\|_\infty, \tag{23}$$

where the constant

$$C = 2^{n+1}\rho^{n^2+n-1}.$$

Remark 3 Algorithm 1 and Proposition 3 assume both the encoder and the decoder know the initial uncertainty region $P(t_0)(= \hat{x}(t_0) + \mathrm{rect}(L(t_0)))$, which the initial state $x(t_0)$ lies within. That assumption might not hold, e.g., the decoder does not know the true initial uncertainty region. A "zooming-out" algorithm in [1] is introduced to tackle this issue, which works as follows:

1) First, the encoder and the decoder agree upon an initial compact set.
2) If the initial state $x_n(t_0)$ lies outside of that compact set, the encoder sends a packet with its n-th bit as "1" to notify the decoder of overflow. Then both the encoder and the decoder synchronously expand the n-th side length of the initial compact set, $L_n(t_0)$, into $L_n(t_1) = \lambda L_n(t_0)$ with a certain ratio λ. When the expanding ratio λ is big enough, after a finite number of steps, $L_n(t_k)$ is long enough so that $x_n(t_k)$ will not overflow. $L_n(t_k)$ will be chosen as the new "initial" n-th side length and the encoder and the decoder have been synchronized regarding the n-th dimension.
3) After the n-th dimension synchronization is achieved, the encoder and the decoder work for the $(n-1)$-th dimension by setting the $(n-1)$-th bit of a packet into 1 to signal the overflow of the $(n-1)$-th dimension of the state. Similar expanding strategy is implemented to achieve synchronization over the $(n-1)$-th dimension in finite steps.
4) The above procedure repeats until synchronization between the encoder and the decoder has been achieved for all dimensions of the state. Such synchronization again takes only finite steps.

In the above "zooming-out" algorithm, only one bit of the packet with n bits is used to signal an overflow. Furthermore, the above algorithm works consecutively from the n-th dimension to the 1-st dimension. We can, therefore, replace the n-bit packet with a single bit and also pursue synchronization consecutively from the n-th dimension to the 1-st dimension. This synchronization is done before implementing Algorithm 1. Thus, the synchronization assumption can be relaxed.

2.3 Asymptotic stabilization by quantized feedback

As shown in Eq.(23), the quantization error exponentially converges to 0, which satisfies the requirements in proving asymptotic stability in [1] (Propositions 2 and 3). Here we directly borrow these results to get

Proposition 4 Let Assumptions 3 and 4 hold. There exist positive numbers λ_i and vectors k_i, for $i = 1, 2, \cdots, n$, which can be used to construct the following controller

$$u = \lambda_n \sigma\left(\frac{k_n \bar{X}_{\mathrm{d},n} + v_{n-1}}{\lambda_n}\right), \tag{24}$$

where

$$v_{n-i} = \lambda_{n-i}\sigma\left(\frac{k_{n-i}\bar{X}_{\mathrm{d},n-i} + v_{n-i-1}}{\lambda_{n-i}}\right),\ i=1,\cdots,n-1$$

with $v_1 = \lambda_1\sigma\left(\dfrac{k_1\bar{X}_{\mathrm{d},1}}{\lambda_1}\right)$, $\bar{X}_{\mathrm{d},i}(t)$ $(\bar{x}_{\mathrm{d}}(t))$ is generated by the decoder in Eq.(17) and $\sigma(\cdot)$ is the saturation function.

The quantization policy in Algorithm 1 and the controller in Eq.(24) guarantee the response of the closed-loop system in Eq.(1) to satisfy the following properties:

- For each $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that $\|x(t_0)\|_\infty \leqslant \dfrac{\|L(t_0)\|_\infty}{2} \leqslant \delta(\epsilon)$ implies
$$\|x(t)\|_\infty \leqslant \epsilon,\ \ \forall t \geqslant t_0. \tag{25}$$
- The state converges to 0, i.e., $\lim\limits_{t\to\infty} \|x(t)\|_\infty = 0$.

We verify Propositions 2, 3, and 4 through the following example:

$$\begin{cases} \dot{x}_1 = x_2 + (\cos x_2)u, \\ \dot{x}_2 = u \end{cases} \tag{26}$$

with $T_M = 0.1\,\mathrm{s}$, $T_m = 0.05\,\mathrm{s}$, and $|u| \leqslant 1$. This system is a global Lipschitz system with $F = 2$. By the rules in Eqs.(10) and (11), we choose $\gamma = \dfrac{1+\sqrt[2]{0.5}}{2}$, $\rho = 2.2$, $\lambda_1 = \lambda_2 = 1$, $k_1 = [-1, -1]$ and $k_2 = -1$. As shown in Fig.2, the quantization errors $e_i(t_k) = x_i(t_k) - \hat{x}_i(t_k)$ $(i = 1, 2)$ are always bounded by $0.5L_i(t_k)$ (Proposition 2), the envelopes of $L_i(t_k)$ exponentially converges to 0 (Proposition 3) and the states of quantized systems, $x_i(t_k)$ converge to 0 (Proposition 4).

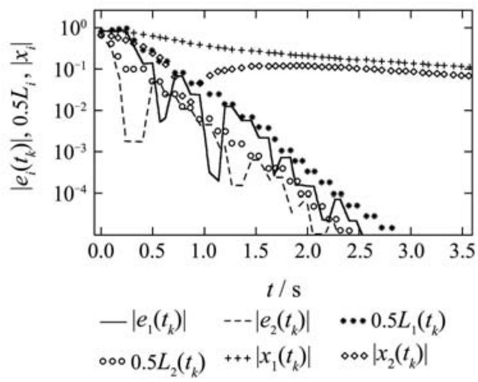

Fig. 2 Simulation results of the system in Eq.(26).

3 Conclusions

In summary, the present paper proposes a dynamic quantization policy to stabilize with only one bit (per sample) a class of n-dimensional quantized feedforward nonlinear systems. Because 1 bit (per sample) is the smallest number of quantization bits, the proposed quantization policy achieves the minimum bit number for the given nonlinear systems, which is rarely reported in the current literature. These results on the minimum bit number are, however, achieved at the cost of some strong assumptions, such as perfect network transmission (without either dropout or delay) and global Lipschitz system dynamics. For linear systems with dropouts and network transmission delay, there are already some results on the minimum stabilizing bit rate [8]. For certain nonlinear systems, it is shown that bounded network transmission delay may not increase the stabilizing (average) bit rate [10]. Built upon these achievements, we will try to relax our assumptions in the future. In the quantization literature, some papers pursue the minimum quantization density, instead of the minimum bit rate, required to stabilize a linear system [11, 12]. We may also extend them to the feedforward nonlinear here.

References

[1] C. De Persis. n-bit stabilization of n-dimensional nonlinear systems in feedforward form[J]. *IEEE Transactions on Automatic Control*, 2005, 50(3): 299 – 311.

[2] C. De Persis, A. Isidori. Stabilizability by state feedback implies stabilizability by encoded state feedback[J]. *Systems & Control Letters*, 2004, 53(3/4): 249 – 258.

[3] D. Liberzon, J. P. Hespanha. Stabilization of nonlinear systems with limited information feedback[J]. *IEEE Transactions on Automatic Control*, 2005, 50(6): 910 – 915.

[4] G. N. Nair, R. J. Evans, I. Mareels, et al. Topological feedback entropy and nonlinear stabilization[J]. *IEEE Transactions on Automatic Control*, 2004, 49(9): 1585 – 1597.

[5] S. Tatikonda, S. K. Mitter. Control under communication constraints[J]. *IEEE Transactions on Automatic Control*, 2004, 49(7): 1056 – 68.

[6] G. N. Nair, R. J. Evans. Stabilizability of stochastic linear systems with finite feedback data rates[J]. *SIAM Journal of Control and Optimization*, 2004, 43(2): 413 – 436.

[7] A. Teel. A nonlinear small gain theorem for the analysis of control system with saturations[J]. *IEEE Transactions on Automatic Control*, 1996, 41(9): 1256 – 1270.

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

[9] Q. Ling, M. D. Lemmon. Stability of quantized linear systems with bounded noise under dynamic bit assignment[C]//*Proceedings of the 43rd IEEE Conference on Decision and Control*. Piscataway: IEEE, 2004: 1454 – 1459.

[10] C. De Persis. Minimal data rate stabilization of nonlinear systems over networks with large delays[C]//*The 5th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshop*. New York: IEEE, 2007: 600 – 608.

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

[12] M. Fu, L. Xie. The sector bound approach to quantized feedback control[J]. *IEEE Transactions on Automatic Control*, 2005, 50(11): 1698 – 1711.

Appendix

A1 Proof of Proposition 2

We prove it by mathematical induction. Eq.(22) works for 0. Suppose it holds for k. Now we prove it for $k+1$.

By the remark after Algorithm 1, $\bar{x}_{\rm e}(t)$ and $\bar{x}_{\rm d}(t)$ are equal, and named $\bar{x}(t)$. Define

$$e(t)=x(t)-\bar{x}(t). \tag{a1}$$

By the definitions of s_k (Eqs.(12) and (16)) and $\dfrac{\bar{x}_{{\rm e},i}(t_k)}{\bar{x}_{{\rm d},i}(t_k)}$ (Eqs.(14) and (17)), we get

$$|e_i(t_k)|\leqslant\begin{cases}\dfrac{L_{I_k}(t_k)}{2}, & i=I_k,\\ L_i(t_k), & \text{otherwise}.\end{cases} \tag{a2}$$

By Eqs.(14) and (17), $\bar{x}(t)$ is updated as

$$\dot{\bar{x}}_i(t)=f_i(\bar{x}_{i+1}(t),\cdots,\bar{x}_n(t),u(t)),\ \ t\in[t_k,t_{k+1}),$$

where $f_i(\cdot)$ are the functions in Eq.(1). Assumption 4 yields

$$|\dot{e}_i(t)|=\left|\dot{x}_i(t)-\dot{\bar{x}}_i(t)\right|\leqslant\sum_{j=i+1}^{n}F|e_j(t)|. \tag{a3}$$

It is straightforward that

$$\begin{aligned}|e_i(t_{k+1})|&\leqslant|e_i(t_k)|+\int_{t_k}^{t_{k+1}}|\dot{e}_i(\tau)|{\rm d}\tau\\&\leqslant|e_i(t_k)|+T_M\max_{t_k\leqslant t<t_{k+1}}|\dot{e}_i(t)|\\&\leqslant|e_i(t_k)|+FT_M\sum_{j=i+1}^{n}\max_{t_k\leqslant t<t_{k+1}}|e_j(t)|,\end{aligned} \tag{a4}$$

where the last inequality comes from Eq.(a3). We can place the following lemma.

Lemma 1 For $t\in[t_k,t_{k+1})$,

$$|e_i(t)|\leqslant L_i(t_{k+1}). \tag{a5}$$

Proof We again prove this lemma by mathematical induction. We can see that Eq.(a5) holds for $i=n$. Now suppose that Eq.(a5) holds for $i\geqslant i_0+1$. We want to prove it also works for $i=i_0$. For $t\in[t_k,t_{k+1})$,

$$\begin{aligned}|e_{i_0}(t)|&\leqslant|e_{i_0}(t_k)|+\int_{t_k}^{t}|\dot{e}_{i_0}(\tau)|{\rm d}\tau\\&\leqslant|e_{i_0}(t_k)|+T_M\max_{t_k\leqslant\tau<t_{k+1}}|\dot{e}_{i_0}(\tau)|\\&\leqslant|e_{i_0}(t_k)|+FT_M\sum_{j=i_0+1}^{n}\max_{t_k\leqslant t<t_{k+1}}|e_j(t)|\\&\leqslant|e_{i_0}(t_k)|+FT_M\sum_{j=i_0+1}^{n}L_j(t_{k+1})\\&=L_{i_0}(t_{k+1}),\end{aligned}$$

where the third inequality comes from Eq.(a3), the fourth one from the assumption that Eq.(a5) holds for $i\geqslant i_0+1$. We, therefore, complete the proof.

Because $\hat{x}(t_{k+1})=\bar{x}(t_{k+1}^-)$,

$$\begin{aligned}|x_i(t_{k+1})-\hat{x}_i(t_k+1)|&=|x_i(t_{k+1})-\bar{x}_i(t_{k+1}^-)|\\&\leqslant L_i(t_{k+1}).\end{aligned} \tag{a6}$$

So $x(t_{k+1})\in\hat{x}(t_{k+1})+{\rm rect}(L(t_{k+1}))$.

A2 Proof of Proposition 3

Define generalized side lengths as

$$\begin{cases}\bar{L}_n(t_k)=\max\left(\rho^nL_n(t_k),\rho^n\gamma^k\|L(t_0)\|_\infty\right),\\ \bar{L}_i(t_k)=\max\left(\rho^iL_i(t_k),\bar{L}_{i+1}(t_k)\right),\end{cases} \tag{a7}$$

where $i=1,2,\cdots,n-1$.

Based on the above definition, we can easily get a lower bound on $\bar{L}_i(t_k)$.

Lemma 2

$$\bar{L}_i(t_k)\geqslant\rho^n\gamma^k\|L(t_0)\|_\infty. \tag{a8}$$

$L(t_k)$ is updated by Eqs.(13) and (18). Based on the definitions of γ and ρ (in Eqs.(10) and (11)) and the definition in Eq.(a7), we get the following proposition.

Lemma 3 For any k and any $i = 1, \cdots, n$,

$$\frac{\bar{L}_i(t_{k+1})}{\bar{L}_i(t_k)} \leqslant \frac{1}{1 - \dfrac{(n-1)FT_M}{\rho}}. \quad \text{(a9)}$$

For the "longest" side chosen by Eq.(12) or Eq.(16), if $\bar{L}_{I_k}(t_k) \geqslant 2\rho^n \gamma^k \|L(t_0)\|_\infty$, then

$$\frac{\bar{L}_{I_k}(t_{k+1})}{\bar{L}_{I_k}(t_k)} \leqslant \frac{1}{2} \times \frac{1}{1 - \dfrac{(n-1)FT_M}{\rho}}. \quad \text{(a10)}$$

Proof We first prove Eq.(a9). Obviously it holds for $i = n$. Now we assume it works for $i = i_0 + 1$ and prove it for $i = i_0$. By Eq.(13), we know

$$L_{i_0}(t_{k+1}) \leqslant L_{i_0}(t_k) + FT_M \sum_{j=i_0+1}^{n} L_j(t_{k+1}).$$

Multiplying both sides of the above equation by ρ^{i_0}, we get

$$\begin{aligned}\rho^i L_{i_0}(t_{k+1}) &\leqslant \rho^{i_0} L_{i_0}(t_k) + \frac{FT_M}{\rho} \sum_{j=i_0+1}^{n} \rho^{i_0+1} L_j(t_{k+1}) \\ &\leqslant \bar{L}_{i_0}(t_k) + \frac{FT_M}{\rho} \sum_{j=i_0+1}^{n} \bar{L}_{i_0}(t_{k+1}) \\ &\leqslant \bar{L}_{i_0}(t_k) + (n-1)\frac{FT_M}{\rho} \bar{L}_{i_0}(t_{k+1}).\end{aligned}$$

If $\bar{L}_{i_0}(t_{k+1}) = \rho^{i_0} L_{i_0}(t_{k+1})$, solving the above inequality w.r.t. $\bar{L}_{i_0}(t_{k+1})$ yields Eq.(a9).

When $\bar{L}_{i_0}(t_{k+1}) \neq \rho^{i_0} L_i(t_{k+1})$, $\bar{L}_{i_0}(t_{k+1}) = \bar{L}_{i_0+1}(t_{k+1})$ and we get

$$\begin{aligned}\frac{\bar{L}_{i_0}(t_{k+1})}{\bar{L}_{i_0}(t_k)} = \frac{\bar{L}_{i_0+1}(t_{k+1})}{\bar{L}_{i_0}(t_k)} &\leqslant \frac{\bar{L}_{i_0+1}(t_{k+1})}{\bar{L}_{i_0+1}(t_k)} \\ &\leqslant \frac{1}{1 - \dfrac{(n-1)FT_M}{\rho}},\end{aligned}$$

where the last inequality comes from the assumption that Eq.(a9) holds for $i = i_0 + 1$. By mathematical induction, we know Eq.(a9) works for any i.

From now on, we prove Eq.(a10). By the definition of I_k in Eqs.(12) and (16), we know

$$\rho^{2I_k} L_{I_k}(t_k) \geqslant \rho^{2j} L_j(t_k), \ \ j = I_k + 1, \cdots, n. \quad \text{(a11)}$$

So, for any $j = I_k + 1, \cdots, n$,

$$\rho^{I_k} L_{I_k}(t_k) \geqslant \rho\rho^j L_j(t_k) \quad \text{(a12)}$$
$$\geqslant \rho^j L_j(t_k). \quad \text{(a13)}$$

When $\bar{L}_{I_k}(t_k) \geqslant 2\rho^n\gamma^k\|L(t_0)\|_\infty$, the defintion in Eq.(a7), together with Eq.(a13), yields

$$\bar{L}_{I_k}(t_k) = \rho^{I_k} L_{I_k}(t_k) \geqslant 2\rho^n\gamma^k\|L(t_0)\|_\infty. \quad \text{(a14)}$$

By the updating rule of $L_{I_k}(t_k)$,

$$\begin{aligned}\rho^{I_k} L_{I_k}(t_{k+1}) &= \rho^{I_k} \frac{L_{I_k}(t_k)}{2} + \rho^{I_k} \sum_{j=I_k+1}^{n} L_j(t_{k+1}) \\ &\geqslant \rho^{I_k} \frac{L_{I_k}(t_k)}{2} \quad \text{(a15)} \\ &\geqslant 2\rho^n\gamma^k \frac{\|L(t_0)\|_\infty}{2} \\ &\geqslant \rho^n\gamma^{k+1}\|L(t_0)\|_\infty. \quad \text{(a16)}\end{aligned}$$

Let $j^* = \arg \max_{j=I_k+1,\cdots,n} \rho^j L_j(t_k)$. For any $j \geqslant I_k + 1$,

$$\begin{aligned}&\rho^j L_j(t_{k+1}) \\ &= \rho^j L_j(t_k) + \frac{FT_M}{\rho} \sum_{m=j+1}^{n} \rho^{j+1} L_m(t_{k+1}) \\ &\leqslant \rho^{j^*} L_{j^*}(t_k) + \frac{FT_M}{\rho} \sum_{m=j+1}^{n} \rho^{j^*} L_{j^*}(t_k) \max_{l=I_k+1,\cdots,n} \rho^l L_l(t_{k+1}) \\ &\leqslant \rho^{j^*} L_{j^*}(t_k) + (n-1)\frac{FT_M}{\rho} \max_{l=I_k+1,\cdots,n} \rho^l L_l(t_{k+1}).\end{aligned}$$

Because the above inequality holds for any $j \geqslant I_k + 1$, we know

$$\begin{aligned}&\max_{l=I_k+1,\cdots,n} \rho^l L_l(t_{k+1}) \\ &\leqslant \rho^{j^*} L_{j^*}(t_k) + (n-1)\frac{FT_M}{\rho} \max_{l=I_k+1,\cdots,n} \rho^l L_l(t_{k+1}).\end{aligned}$$

Solving the above inequality w.r.t. $\max_{l=I_k+1,\cdots,n} \rho^l L_l(t_{k+1})$ yields

$$\max_{l=I_k+1,\cdots,n} \rho^l L_l(t_{k+1}) \leqslant \rho^{j^*} L_{j^*}(t_k) \frac{1}{1 - \dfrac{(n-1)FT_M}{\rho}}. \quad \text{(a17)}$$

By Eq.(a12), we get

$$\rho^{I_k} L_{I_k}(t_k) \geqslant \rho\rho^{j^*} L_{j^*}(t_k). \quad \text{(a18)}$$

Combining the above equation with Eqs.(a15) and (a17) yields, for any $j = I_k + 1, \cdots, n$,

$$\begin{aligned}\rho^{I_k} L_{I_k}(t_{k+1}) &\geqslant \frac{1}{2}\rho \left(1 - (n-1)\frac{FT_M}{\rho}\right) \rho^j L_j(t_{k+1}) \\ &\geqslant \rho^j L_j(t_{k+1}), \quad \text{(a19)}\end{aligned}$$

where the last inequality comes from the definition of ρ in Eq.(11). Considering Eqs.(a7), (a16) and (a19), we get

$$\begin{aligned}\bar{L}_{I_k}(t_{k+1}) &= \frac{1}{2}\rho^{I_k} L_{I_k}(t_k) + FT_M \sum_{j=I_k+1}^{n} \rho^{I_k} L_j(t_{k+1}) \\ &\leqslant \frac{1}{2}\rho^{I_k} L_{I_k}(t_k) + \frac{FT_M}{\rho} \sum_{j=I_k+1}^{n} \rho^j L_j(t_{k+1}) \\ &\leqslant \frac{1}{2}\bar{L}_{I_k}(t_k) + \frac{FT_M}{\rho} \sum_{j=I_k+1}^{n} \bar{L}_{I_k}(t_{k+1}) \\ &\leqslant \frac{1}{2}\bar{L}_{I_k}(t_k) + \frac{(n-1)FT_M}{\rho} \bar{L}_{I_k}(t_{k+1}).\end{aligned}$$

Solving the above inequality w.r.t. $\bar{L}_{I_k}(t_{k+1})$ yields Eq.(a10). This completes the proof.

Define $p(t_k) = \prod_{i=1}^{n} \bar{L}_i(t_k)$. We get

Lemma 4 When

$$p(t_k) \geqslant \left(2\rho^{2n}\gamma^k\|L(t_0)\|_\infty\right)^n, \quad \text{(a20)}$$
$$\bar{L}_{I_k}(t_k) \geqslant 2\rho^n\gamma^k\|L(t_0)\|_\infty. \quad \text{(a21)}$$

Proof We prove Eq.(a21) by contradiction. Suppose it does NOT hold, i.e., $\bar{L}_{I_k}(t_k) < 2\rho^n\gamma^k\|L(t_0)\|_\infty$. So

$$\rho^{I_k} L_{I_k}(t_k) \leqslant \bar{L}_{I_k}(t_k) < 2\rho^n\gamma^k\|L(t_0)\|_\infty. \quad \text{(a22)}$$

By the selection rule of I_k (in Eq.(12) or Eq.(16)), we get, for any $j = 1, \cdots, n$,

$$\begin{aligned}\rho^j L_j(t_k) &\leqslant \frac{\rho^{I_k}}{\rho^j} \rho^{I_k} L_{I_k}(t_k) \\ &< \rho^n \rho^{I_k} L_{I_k}(t_k) \\ &\leqslant \rho^n 2\rho^n\gamma^k\|L(t_0)\|_\infty. \quad \text{(a23)}\end{aligned}$$

By the definition of $\bar{L}_j(t_k)$ in Eq.(a7) and the above equation, we get

$$\bar{L}_j(t_k) < 2\rho^{2n}\gamma^k\|L(t_0)\|_\infty, \ \forall j. \quad \text{(a24)}$$

Multiplying the above for $j = 1, \cdots, n$ yields

$$p(t_k) = \prod_{i=1}^{n} \bar{L}_i(t_k) < \left(2\rho^{2n}\gamma^k\|L(t_0)\|_\infty\right)^n, \quad \text{(a25)}$$

which contradicts with the condition in Eq.(a20). So Eq.(a21) holds. This completes the proof.

By Lemmas 3 and 4 and the definitions of $p(t_k)$, ρ and γ, we get

Corollary 1

$$\frac{p(t_{k+1})}{p(t_k)} \leqslant \left(\frac{1}{1-\dfrac{(n-1)FT_M}{\rho}}\right)^n < 2\gamma^n,\ \forall k \geqslant 0. \quad \text{(a26)}$$

When Eq.(a20) holds,

$$\frac{p(t_{k+1})}{p(t_k)} \leqslant \gamma^n. \quad \text{(a27)}$$

For $p(t_k)$, we can place the following upper bound.

Proposition 5

$$p(t_k) < 2\left(2\rho^{2n}\gamma^k\|L(t_0)\|_\infty\right)^n,\ \forall k. \quad \text{(a28)}$$

Proof For $k=0$, Eq.(a28) holds. Suppose it also holds when $k=k_0$. Now we prove it also works for $k=k_0+1$. There are 2 cases.

1) When Eq.(a20) holds, we know, by Eq.(a27),

$$\begin{aligned} p(t_{k_0+1}) &\leqslant \frac{1}{2}\frac{1}{1-\dfrac{(n-1)FT_M}{\rho}}\left(\frac{1}{1-\dfrac{(n-1)FT_M}{\rho}}\right)^{n-1} p(t_{k_0}) \\ &\leqslant \gamma^n p(t_{k_0}) \\ &< \gamma^n \times 2\left(2\rho^{2n}\gamma^{k_0}\|L(t_0)\|_\infty\right)^n \\ &= 2\left(2\rho^{2n}\gamma^{k_0+1}\|L(t_0)\|_\infty\right)^n, \end{aligned}$$

i.e., Eq.(a28) holds for $k=k_0+1$.

2) When Eq.(a20) does NOT hold, we know, by Eq.(a26),

$$\begin{aligned} p(t_{k_0+1}) &< \left(\frac{1}{1-\dfrac{(n-1)FT_M}{\rho}}\right)^n p(t_{k_0}) \\ &\leqslant 2\gamma^n p(t_{k_0}) \\ &< 2\gamma^n \times \left(2\rho^{2n}\gamma^{k_0}\|L(t_0)\|_\infty\right)^n \\ &= 2\left(2\rho^{2n}\gamma^{k_0+1}\|L(t_0)\|_\infty\right)^n, \end{aligned}$$

i.e., Eq.(a28) holds for $k=k_0+1$.

In summary, Eq.(a28) holds for both cases. This completes the proof.

Now we are ready to prove Proposition 3.

Proof Because

$$p(t_k) = \prod_{i=1}^{n} \bar{L}_i(t_k)$$

and $\bar{L}_j(t_k) \geqslant \rho^n\gamma^k\|L(t_0)\|_\infty$ for any $j=1,2,\cdots,n$, we get

$$\bar{L}_i(t_k) \leqslant \frac{p(t_k)}{(\rho^n\gamma^k\|L(t_0)\|_\infty)^{n-1}},\ \forall i. \quad \text{(a29)}$$

Furthermore, $L_i(t_k) \leqslant \dfrac{1}{\rho^i}\bar{L}_i(t_k)$. So

$$L_i(t_k) \leqslant \frac{1}{\rho^i}\frac{p(t_k)}{(\rho^n\gamma^k\|L(t_0)\|_\infty)^{n-1}},\ \forall i. \quad \text{(a30)}$$

Substituting the bound in Eq.(a28) into the above equation yields

$$\begin{aligned} L_i(t_k) &\leqslant \frac{1}{\rho^i}\frac{2\left(2\rho^{2n}\gamma^k\|L(t_0)\|_\infty\right)^n}{(\rho^n\gamma^k\|L(t_0)\|_\infty)^{n-1}} \\ &< 2^{n+1}\rho^{n^2+n-1}\gamma^k\|L(t_0)\|_\infty, \end{aligned} \quad \text{(a31)}$$

i.e., Eq.(23) holds. This completes the proof.

Qiang LING received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1997, the M.E. degree from Tsinghua University, Beijing, China, in 2000, and the Ph.D. degree from the University of Notre Dame, Notre Dame, IN, in 2005. He worked as a research staff member at Seagate Technology from 2005 to 2008. He joined University of Science and Technology of China as an associate professor in 2008. His research interests include networked control systems, signal processing and wireless sensor works. E-mail: qling@ustc.edu.cn.

Michael D. LEMMON received the B.S. degree in Electrical Engineering from Stanford University, Stanford, CA, in 1979 and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, PA, in 1990. He is a professor of Electrical Engineering at the University of Notre Dame, Notre Dame, IN. Most recently, he helped forge a consortium of academic, private and public sector partners to build one of the first metropolitan scale sensor-actuator networks (CSOnet) used in monitoring and controlling combined-sewer overflow events. His research deals with real-time networked control systems with an emphasis on understanding the impact that reduced feedback information has on overall system performance. This work has been funded by a variety of state and federal agencies that include the National Science Foundation, Army Research Office, Defense Advanced Research Project Agency, and Indiana's 21st Century Technology Fund.

He served as an associate editor for the IEEE Transactions on Neural Networks and the IEEE Transactions on Control Systems Technology. He chaired the first IEEE working group on hybrid dynamical systems and was the program chair for a hybrid systems workshop in 1997. E-mail: lemmon@nd.edu.

Hai LIN received the B.S. degree from the University of Science and Technology, Beijing, China, in 1997, the M.E. degree from the Chinese Academy of Sciences, Beijing, China, in 2000, and the Ph.D. degree from the University of Notre Dame, Notre Dame, IN, in 2005. He is currently an assistant professor in the Electrical and Computer Engineering Department, National University of Singapore. His research interests are in the multidisciplinary study of the problems at the intersection of control, communication, computation and life sciences. He is particularly interested in hybrid systems, networked embedded systems, and systems biology.

He received the 2006 Alumni Association Graduate Research Award from the University of Notre Dame. E-mail: elelh@nus.edu.sg.