Joint 48th IEEE Conference on Decision and Control and
28th Chinese Control Conference
Shanghai, P.R. China, December 16-18, 2009

Stabilize an n-dimensional quantized nonlinear feedforward system with 1 bit

Qiang Ling, Michael D. Lemmon and Hai Lin

***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, then sent through a digital network to the controller. In order to save network bandwidth, people pursue as less quantization bits as possible to maintain stability of such a system. In DePersis' paper [1], n bits are used to stabilize the n-dimensional system by assigning one bit for each state variable (dimension). This paper extends that result by stabilizing the whole system with a single bit under the same assumptions. 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 asymptotic stability of the system can, therefore, be guaranteed. Because 1 bit is the smallest bit number in each packet, the new policy achieves the minimum stabilizable bit number for that n-dimensional feedforward nonlinear system.**

I. INTRODUCTION

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

$$\dot{x} = f(x,u) = \begin{pmatrix} f_1(X_2,u) \\ \vdots \\ f_n(u) \end{pmatrix} \tag{1}$$

where $x \in R^n$, $u \in R^m$ and $X_i(t) = [x_i(t), x_{i+1}(t), \cdots, x_n(t)]^T$. When the above nonlinear system is controlled over a digital network, a typical configuration is shown in Fig. 1.

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

$$0 < T_m \le t_{k+1} - t_k \le T_M < \infty, \forall k \ge 0 \tag{2}$$

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, $\hat{x}(t_k)$, of the state $x(t_k)$. 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

Qiang Ling is with Department of Automation, University of Science and Technology of China, Hefei, Anhui 230027, China; *Email:* qling@ustc.edu.cn; *Phone:* (86)551-360-0504. His research was supported by the National Natural Science Foundation of China (60904012), the Natural Science Foundation of Anhui Province (No. 090412050) and the Talent Development Program of Anhui Province (No. 2008Z012).

Michael D. Lemmon is with Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

Hai Lin is with Department of Electrical and Computer Engineering, National University of Singapore 4 Engineering Drive 3, Singapore 117576. His research was supported by Singapore Ministry of Education (AcRF Tier 1 funding, TDSI, and TL).

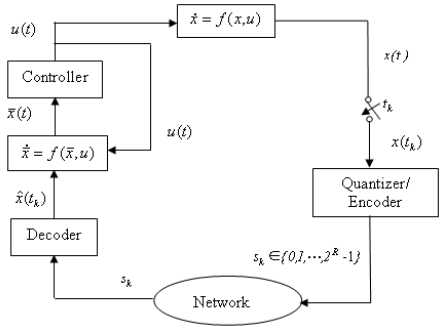

Fig. 1. Quantized nonlinear control systems

$\overline{x}(t)$. The controller will make use of $\overline{x}(t)$, instead of the true state $x(t)$, to devise the control $u(t)$[2].

This paper addresses the following questions. *Is there any quantization policy to maintain its stability under finite R? What is the minimum quanitzation bit number R (per sample) required to maintain stability?* These two questions have generated much interest in the last years.

It is shown in [2] that when the number of quantization bits, R, is big enough, the global asymptotic stabilizability of nonlinear control systems can be preserved under the feedback quantization. In [3], a finite number of quantization bits are shown to be able to stabilize a class of nonlinear systems, which are input-to-state stable (ISS) with respect to measurement errors. More quantization bits, however, require more network bandwidth. So it is of great importance to determine the smallest R that still asymptotically stabilize the system. The minimality of the number of quantization bits (per sample) required to stabilize a nonlinear system is addressed in [4], where a notion of topological feedback entropy (TFE) is introduced and it is proven that a system can be stabilized *locally* 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 compute its TFE and the minimum R to stabilize a general nonlinear system is usually unknown. Therefore a less aggressive goal is persued, *to stabilize a nonlinear system with as few quantization bits as possible*. In order to save the 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, and $R = n$ ($R = n+1$) can be enough to achieve semiglobal asymptotic stabilization (global stabilization) [1] under three assumptions:

Assumption 1: Functions $f_i(X_{i+1}, u)$ $(i = 1, 2, \cdots, n-1)$

and $f_n(u)$ are locally Lipschitz.

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

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-$\mathcal{K}_+$ (nonnegative, continuous and nondecreasing) function $\phi_i(\cdot)$,

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

Is it possible to use *fewer* than n bits to achieve the desired asymptotic stability? As R is the number of successfully transmitted bits per sample (packet), it has a lower bound

$$R \geq 1 \tag{3}$$

The present paper proposes a dynamic quantization policy that uses 1 bit to globally asymptotically stabilize the n-dimensional nonlinear system in eq. 1 under the same assumptions of [1], i.e., Assumptions 1-3. Due to the bound in eq. 3, we know the minimum bit number has been achieved. Now we remark on that policy. 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 provided single bit. This policy for the nonlinear systems is motivated by the dynamic bit assignment policy for linear systems [7].

The rest is organized as follows. In Section II, we present the dynamic quantization policy. 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 III, the paper is concluded with some final remarks. To improve readability, technical proofs are moved to Appendix, Section IV.

II. Main results: Dynamic quantization policy

A. Uncertainty region of the state

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

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

where $C(t_k)$ is the centroid of $P(t_k)$, $rect(L(t_k))$ represents a rectangle with the origin as the centroid and the side length vector $L(t_k)$, i.e.,

$$rect(L(t_k)) = \prod_{i=1}^{n} \left[-\frac{1}{2}L_i(t_k), \frac{1}{2}L_i(t_k) \right]$$

with $\prod$ standing for the Cartesian product. In order to minimize the worst-case estimation error, the decoder sets

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

So $\hat{x}(t_k)$ is used to represent the centroid of $P(t_k)$. The estimation error $\tilde{x}(t_k) = x(t_k) - \hat{x}(t_k)$ is bounded by

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

After receiving s_k, the decoder updates its centroid and side length vector as

$$\begin{cases} (\hat{x}(t_k), s_k) \to \hat{x}(t_{k+1}) \\ (L(t_k), s_k) \to L(t_{k+1}) \end{cases} \tag{7}$$

Of course, discretion is required to guarantee no overflow would occur, i.e., $x(t_{k+1}) \in \hat{x}(t_{k+1}) + rect(L(t_{k+1}))$. The symbol s_k in eq. 7 is sent by the encoder. So the encoder 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 $\lim_{t\to\infty} x(t) = 0$, we have to guarantee the convergence of the continuous-time estimation error $\tilde{x}(t) = x(t) - \overline{x}(t)$. Due to Assumption 2 (the boundedness of the control $u(t)$), Assumption 1 (the local Lipschitz property of $f(\cdot) = [f_1(\cdot), f_2(\cdot), \cdots, f_n(\cdot)]^T$) and eq. 2 (bounded sampling intervals), we know

Proposition 2.1: $\lim_{t\to\infty} \tilde{x}(t) = 0$ is equivalent to

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

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

Its proof is not difficult and omited here. Later we will design a quantization policy to satisfy eq. 8.

B. Dynamic quantization policy

Due to Assumptions 1 and 2, we know, for each i $(\in \{2, 3, \cdots, n\})$ and $\forall W_i > 0$, there exists a finite positive F_{i-1} such that

$$|f_{i-1}(X_i, u) - f_{i-1}(Y_i, u)| \leq F_{i-1}\|X_i - Y_i\|_\infty \tag{9}$$

for any $\|X_i\|_\infty \leq W_i$, $\|Y_i\|_\infty \leq W_i$ and $u(t) \leq U$. Here we consider a particular structure of Y_i,

$$Y_i(t_k) = X_i(t_k) + \tilde{X}_i(t_k) \tag{10}$$

where $\tilde{X}_i(t_k) = [\tilde{x}_i(t_k), \tilde{x}_{i+1}(t_k), \cdots, \tilde{x}_n(t_k)]^T$ is the quantization error vector. Correspondingly we define a vector

$$L^{(i)}(t_k) = [L_i(t_k), L_{i+1}(t_k), \cdots, L_n(t_k)]^T \tag{11}$$

By the bounds on quantization errors in eq. 6, we get

$$\|\tilde{X}_i(t_k)\|_\infty \leq \frac{1}{2}\|L^{(i)}(t_k)\|_\infty \tag{12}$$

Suppose both $\{X_i(t_k)\}_k$ and $\{L^{(i)}(t_k)\}_k$ are bounded, i.e., for any i $(= 1, 2, \cdots, n)$, there exist $Z_i > 0$ and $S_i > 0$, such that

$$\begin{cases} \max_{k\geq 0} \|X^{(i)}(t_k)\|_\infty \leq Z_i \\ \max_{k\geq 0} \|L^{(i)}(t_k)\|_\infty \leq 2S_i \end{cases} \tag{13}$$

Define $W_i = Z_i + S_i$. So $\|X_i(t_k)\|_\infty \leq Z_i < W_i$ and $\|Y_i(t_k)\|_\infty \leq W_i$. For any given Z_i and S_i, there must exist F_{i-1} so that eq. 9 holds. Then we design a quantization

policy with the knowledge of Z_i, S_i and F_{i-1}. Under that policy, the quantization error $\tilde{x}_i(t_k)$, more precisely $L_i(t_k)$, exponentially converges to 0 as k goes to ∞. Such exponential convergence guarantees that $x(t)$ ($X_i(t)$) converges to 0 as $t \to \infty$, i.e., the nonlinear system in eq. 1 is asymptotically stable. The only potential hole of the above arguement is *whether do such Z_i and S_i exist* for $i = 1, \cdots, n$? Our answer is definitely *"Yes"* and will give a constructive way to compute them.

First we build our quantizer under the conditions in eq. 13. Choose a positive number γ by

$$\sqrt[n]{\frac{1}{2}} < \gamma < 1 \tag{14}$$

Choose large enough positive numbers ρ_i so that

$$\begin{cases} 1 - \frac{(n-1)F_iT_M}{\rho_i} > 0 \\ \left(\frac{1}{1-\frac{(n-1)F_iT_M}{\rho_i}}\right)^n < 2\gamma^n \end{cases}, i = 1, \cdots, n-1. \tag{15}$$

and $\rho_n = 1$. For notational convenience, define

$$\begin{cases} \rho_{b,i} = \prod_{j=i}^{n} \rho_j \\ \rho_{f,i} = \prod_{j=1}^{i-1} \rho_j \end{cases}, i = 1, \cdots, n \tag{16}$$

where $\rho_{f,1}$ is specially defined as 1. Similar to the quantization policy of linear systems in [8], we propose the following policy.

Algorithm 1: **Dynamic quantization policy:**

Encoder/Decoder initialization:

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

Encoder Algorithm:

1) **Select** the index I_k by

$$I_k = \arg\max_i 4^i \rho_{f,i} L_{e,i}(t_k) \tag{17}$$

2) **Quantize** the state $x(t_k)$ by setting

$$s_k = \begin{cases} 1, & x_{I_k}(t_k) \geq \hat{x}_{I_k}(t_k) \\ 0, & \text{otherwise} \end{cases}$$

3) **Transmit** the quantized symbol s_k.

4) **Update** $L(t_{k+1})$ at time instant t_{k+1} as [1]

$$L_i(t_{k+1}) = \begin{cases} L_i(t_k)/2 + F_iT_M \sum_{j=I_k+1}^{n} L_j(t_{k+1}), & i = I_k \\ L_i(t_k) + F_iT_M \sum_{j=I_k+1}^{n} L_j(t_{k+1}), & i \neq I_k \end{cases} \tag{18}$$

$\hat{x}(t_{k+1})$ is updated by running the differential equation in Fig. 1

$$\frac{d}{dt}\overline{x}_{e,i}(t) = f_i(\overline{X}_{e,i+1}(t), u(t)), \tag{19}$$

$$\overline{x}_{e,i}(t_k) = \begin{cases} \hat{x}_{e,i}(t_k) + L_i(t_k)/4, & i = I_k, s_k = 1 \\ \hat{x}_{e,i}(t_k) - L_i(t_k)/4, & i = I_k, s_k = 0 \\ \hat{x}_{e,i}(t_k), & i \neq I_k \end{cases}$$

[1]The computation in eq. 18 is done reversely from $i = n$ to $i = 0$.

where $\overline{X}_{e,i}(t) = [\overline{x}_{e,i}(t), \overline{x}_{e,i+1}(t), \cdots, \overline{x}_{e,n}(t)]^T$, $t \in [t_k, t_{k+1})$ and the control $u(t)$ is generated by the controller in Fig. 1 with the estimated state $x_e(t) (= \overline{X}_{e,1}(t))$ in the place of $\overline{x}(t)$. At time $t = t_{k+1}$, $\hat{x}_i(t_{k+1})$ is updated as

$$\hat{x}_i(t_{k+1}) = \overline{x}_{e,i}(t_{k+1}^-), i = 1, 2, \cdots, n$$

5) **Update time index**, $k = k + 1$ and return to step 1.

Decoder Algorithm:

1) **Select** the index I_k by

$$I_k = \arg\max_i 4^i \rho_{f,i} L_{d,i}(t_k) \tag{20}$$

2) **Wait** for quantized data, s_k, from encoder.

3) **Update** the state estimate at t_k as

$$\hat{x}_{d,i}(t_k) := \begin{cases} \hat{x}_{d,i}(t_k) + L_i(t_k)/4, & i = I_k, s_k = 1 \\ \hat{x}_{d,i}(t_k) - L_i(t_k)/4, & i = I_k, s_k = 0 \\ \hat{x}_{d,i}(t_k), & i \neq I_k \end{cases} \tag{21}$$

4) **Generate** the continuous-time state estimate as

$$\begin{aligned} &\frac{d}{dt}\overline{x}_{d,i}(t) = f_i(\overline{X}_{d,i+1}(t), u(t)), \\ &\overline{x}_{d,i}(t_k) = \hat{x}_{d,i}(t_k) \end{aligned} \tag{22}$$

where $t \in [t_k, t_{k+1})$.

5) **Control** variable $u(t)$ is constructed from the controller in Fig. 1 by replacing $\overline{x}(t)$ with $\overline{x}_d(t) (= \overline{X}_{d,1}(t))$.

6) **Update** $L(t_{k+1})$ at time instant t_{k+1} as

$$L_i(t_{k+1}) = \begin{cases} L_i(t_k)/2 + F_iT_M \sum_{j=I_k+1}^{n} L_j(t_{k+1}), & i = I_k \\ L_i(t_k) + F_iT_M \sum_{j=I_k+1}^{n} L_j(t_{k+1}), & i \neq I_k \end{cases} \tag{23}$$

At time $t = t_{k+1}$, $\hat{x}_{d,i}(t_{k+1})$ is updated as

$$\hat{x}_{d,i}(t_{k+1}) = \overline{x}_{d,i}(t_{k+1}^-), i = 1, 2, \cdots, n \tag{24}$$

7) **Update time index**, $k = k + 1$, and return to step 1.

Remark: Because the transmitted symbol s_k is always received correctly, $L_e(t_0) = L_d(t_0)$ and $L_e(t_k)$ and $L_d(t_k)$ are updated by the same rule in eq. 18 and 23, we have

$$L_e(t_k) = L_d(t_k), \forall k \tag{25}$$

Therefore we may shorten $L_e(t_k)$ and $L_d(t_k)$ into the same variable $L(t_k)$ without confusion. Similarly we can show that

$$\begin{cases} \hat{x}_e(t_k) = \hat{x}_d(t_k), & \forall k \\ \overline{x}_e(t) = \overline{x}_d(t), & \forall t \geq t_0 \end{cases} \tag{26}$$

$\hat{x}_e(t_k)$ and $\hat{x}_d(t_k)$ are shortened into $\hat{x}(t_k)$, $\overline{x}_e(t)$ and $\overline{x}_d(t)$ into $\overline{x}(t)$ as well. The same $\overline{x}(t)$ is used to compute control variable by the same rule at both encoder and decoder sides. Of course, the same control variable $u(t)$ will be obtained at both sides. Our quantization policy guarantees there is no state overflow, which is wrapped up in the following Proposition. The Proposition 2.2 can be proven similarly as [1].

Proposition 2.2: Under Assumptions 1-3, we choose γ and ρ by eq. 14 and 15. The dynamic quantization policy in Algorithm 1 is implemented to the quantized nonlinear system in eq. 1. For any $k \geq 0$,

$$x(t_k) \in \hat{x}(t_k) + rect(L(t_k)) \tag{27}$$

Remark: In Algorithm 1, the side is measured by the weighted length $4^i \rho_{f,i} L_i(t_k)$ rather than the direct length $L_i(t_k)$. That policy assigns the highest priority to the n-th dimension. The motivation lies in the feedforward struction of eq. 1, i.e., the n-th dimension affects the other dimensions, but **NOT** reversely. After $L_n(t_k)$ is reduced enoughly, we get almost precise state estimate $\overline{x}_n(t)$ 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 balancing has to be made when assigning the single bit among n dimensions, which is carried out by the appropriate choice of ρ in eq. 15. It will be shown in Proposition 2.3 that $L(t_k)$ exponentially converges to 0, whose proof can be found in the appendix.

Proposition 2.3: Under Assumptions 1-3, we choose γ and ρ_i by eq. 14 and 15. The dynamic quantization policy in Algorithm 1 is implemented on the quantized nonlinear system in eq. 1. The side length vector $L(t_k)$ is bounded as

$$\|L_i(t_k)\|_\infty \leq 2^{2n+1} \rho_{b,i} \gamma^k \|L(t_0)\|_\infty, i = 1, \cdots, n \tag{28}$$

Remark: By Proposition 2.3, we can simply choose S_i in eq. 13 as

$$S_i = 2^{2n+1} \rho_{b,i} \|L(t_0)\|_\infty, i = 1, \cdots, n \tag{29}$$

Suppose there exist Z_i $(i = 1, \cdots, n)$ to satisfy eq. 13. We first choose S_n by eq. 29. The updating rule of $L_n(t_k)$ in eq. 18 and 23 guarantees $\{L_n(t_k)\}$ is non-increasing w.r.t. k. So we have made a right choice of S_n. Z_n and S_n are used together to determine F_{n-1} in eq. 9. With F_{n-1}, we can select ρ_{n-1} by $t \in [t_k, t_{k+1})$ and then determine S_{n-1} by eq. 29. Repeat the above story with S_j and Z_j for $j = n-1, n-2, \cdots, 2$. We get ρ_{n-2}, ρ_{n-3}, $\cdots$, ρ_1. Therefore we get all parameters of Algorithms 1. Under these ρ_i $(i = 1, \cdots, n-1)$ and $\rho_n = 1$, Proposition 2.3 guarantee that all choices of S_i $(i = n-1, n-2, \cdots, 1)$ in eq. 29 are valid. So the existence of S_i is no longer a problem. We only need to justify the existence of Z_i $(i = 1, \cdots, n)$.

Remark: Algorithm 1 and Proposition 2.3 assume both the encoder and the decoder know the initial uncertainty region $P(t_0)(= \hat{x}(t_0) + rect(L(t_0)))$, 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 uses the i-th bit of the n-bit packet to signal overflowing in the i-th dimension and expands the uncertaint set accordingly. Furthermore, that 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. So the synchronization assumption can be relaxed.

C. Asymptotic stabilization by quantized feedback

As shown in eq. 28, 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 state

Proposition 2.4: Under Assumptions 3 and 3, there exist positive numbers Z_i for $i = n, n-1, \cdots, 2$, positive numbers and vectors λ_i^* and, respectively, k_i, for $i = 1, 2, \cdots, n$, which can be used to construct the following controller

$$\begin{aligned} u &= \lambda_n \sigma\left(\frac{k_n \overline{X}_{d,n} + v_{n-1}}{\lambda_n}\right) \\ v_{n-i} &= \lambda_{n-i} \sigma\left(\frac{k_{n-i} \overline{X}_{d,n-i} + v_{n-i-1}}{\lambda_{n-i}}\right) \\ v_1 &= \lambda_1 \sigma\left(\frac{k_1 \overline{X}_{d,1}}{\lambda_1}\right) \end{aligned} \tag{30}$$

where, for $i = 1, 2, \cdots, n$, $\lambda_i \in (0, \lambda_i^*]$ and $\overline{X}_{d,i}(t)$ $(\overline{x}_d(t))$ is generated by the decoder in eq. 22 [2] and the function $\sigma(\cdot)$ denotes a saturation function.

The quantization policy in Algorithm 1 and the controller in eq. 30 guarantees 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 $\|x(t_0)\|_\infty \leq \|L(t_0)\|_\infty / 2 \leq \delta(\epsilon)$ implies

$$\|x(t)\|_\infty \leq \epsilon, \forall t \geq t_0 \tag{31}$$

- The state converges to 0, i.e.,

$$\lim_{t \to \infty} \|x(t)\|_\infty = 0 \tag{32}$$

Remark: We do not pursue strict proof here. What we do is to show the key ideas in proving Proposition 2.4. By eq. 30, we get

$$\begin{aligned} u(t) &= \lambda_n \sigma\left(\frac{k_n \overline{x}_n(t) + v_{n-1}(t)}{\lambda_n}\right) \\ &= \lambda_n \sigma\left(\frac{k_n x_n(t) + \phi_{n-1}(t)}{\lambda_n}\right) \end{aligned} \tag{33}$$

where $\phi_{n-1}(t) = k_n(x_n(t) - \overline{x}_n(t)) + v_{n-1}(t)$. $|x_n(t) - \overline{x}_n(t)|$ is bounded by $L_n(t_{k+1})$ with $t_k \leq t \leq t_{k+1}$. Because S_n is an upper bound on $\{L_n(t_k)\}$, it is also an upper bound on $|x_n(t) - \overline{x}_n(t)|$. $v_{n-1}(t)$ is bounded by λ_{n-1}. So we have a well-defined bound on $\phi_{n-1}(t)$ for all t. Assumption 3 (stabilizability assumption) guarantees, under the control $u(t)$ in eq. 33, the following equation

$$\dot{x}_n(t) = f_n(u) \tag{34}$$

has a bounded solution $x_n(t)$. Of course we can get its bound, which is chosen as Z_n. $\overline{X}_n(t)(\overline{x}_n(t))$ is, therefore, bounded by $(S_n + Z_n)$.

Now we work on boundedness of $X_{n-1}(t)$. $u(t)$ is composed of $k_{n-1}\overline{X}_{d,n-1}$, $k_n \overline{X}_n(t)$ and $v_{n-2}(t)$. The latter two items, $k_n \overline{X}_n(t)$ and $v_{n-2}(t)$, are bounded. And $X_{n-1}(t) - \overline{X}_{n-1}(t)$ is also bounded. By the stabilizabiity assumption,

[2] Because $\overline{x}_e(t) = \overline{x}_d(t)$ for any t, $\overline{x}_d(t)$ or $\overline{X}_{d,i}(t)$ is known by the encoder.

we get an upper bound on $X_{n-1}(t)$, the solution of the following equaiton

$$\dot{X}_{n-1} = \begin{cases} f_{n-1}(X_n, u) \\ f_n(u) \end{cases} \tag{35}$$

We choose Z_{n-1} as the upper bound on $X_{n-1}(t)$, which is determined only by Z_n. We can keep working on $X_{n-2}(t)$ and get Z_{n-2} that is a function of Z_n and Z_{n-1}. Following the similar procedure, we get all Z_i $(i = n-3, \cdots, 2)$.

III. Conclusion

In summary, the present paper proposes a dynamic quantization policy to stabilize with only 1 bit(per sample) a class of n-dimensional quantized feedforward nonlinear systems. Because 1 bit per sample is the lowest constant bit rate, the proposed quantization policy achieves the minimum bit rate for the given nonlinear systems, which is rarely reported in the current literature. These results on minimum constant bit rate are, however, achieved under the perfect newtwork transmission assumption(without either dropout or delay). For linear systems with dropouts and network transmission delay, there are already some results on the minimum stabilizing bit rate [7]. For certain nonlinear systems, it is shown that bounded network transmission delay may not increase the stabilizing (average) bit rate [9]. Built upon these achievements, we will try to relax our assumptions in future.

References

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

[2] C. D. Persis and A. Isidori, "Stabilizability by state feedback implies stabilizability by encoded state feedback," *Systems and Control Letters*, vol. 53, pp. 249–258, 2004.

[3] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information," *IEEE Transactions on Automatic Control*, vol. 50(6), pp. 910–916, 2005.

[4] G. Nair, R. Evans, I. Mareels, and B. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Transactions on Automatic Control, special issue on Networked Control Systems*, vol. 49(9), pp. 1585 – 1597, 2004.

[5] S. Tatikonda and S. Mitter, "Control under communication constraints," *IEEE Transactions on Automatic Control*, vol. 49(7), pp. 1056–1068, 2004.

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

[7] Q. Ling and M. Lemmon, "Stability of quantized control systems under dynamicc bit assignment," *IEEE Trans. on Automatic Control*, vol. 50(5), 2005.

[8] ——, "Stability of quantized linear systems with bounded noise under dynamic bit assignment," in *IEEE Conference on Decision and Control*, Atlantis, Paradise Island, Bahamas, 2004.

[9] C. D. Persis, "Minimal data rate stabilization of nonlinear systems over networks with large delays," *arXiv.org*, 2007.

IV. Appendix: Technical proofs

Define generalized side lengths as

$$\begin{cases} \overline{L}_n(t_k) = \max\left(L_n(t_k), \rho_n \gamma^k \|L(t_0)\|_\infty\right) \\ \overline{L}_i(t_k) = \max\left(L_i(t_k), \rho_i \overline{L}_{i+1}(t_k)\right) \end{cases} \tag{36}$$

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

Based on the above definition, we can easily get

Lemma 4.1:

$$\overline{L}_i(t_k) \geq \rho_{b,i} \gamma^k \|L(t_0)\|_\infty \tag{37}$$

$L(t_k)$ is updated by eq. 18 (23). Based on the definitions of γ and ρ_i (in eq. 14 and 15) and the definition in eq. 36, we get the following results.

Lemma 4.2: Let $\beta = \sqrt[n]{2}\gamma$. For any k and any $i = 1, \cdots, n$,

$$\frac{\overline{L}_i(t_{k+1})}{\overline{L}_i(t_k)} \leq \beta \tag{38}$$

For the "longest" side chosen by eq. 17(20), if $\overline{L}_{I_k}(t_k) \geq 4\rho_{b,I_k}\gamma^k\|L(t_0)\|_\infty$, then

$$\frac{\overline{L}_{I_k}(t_{k+1})}{\overline{L}_{I_k}(t_k)} \leq \frac{1}{2}\beta \tag{39}$$

Proof: We first prove eq. 38.

Obviously it holds for $i = n$. Now we assume it works for $i = i_0 + 1$ and prove it also holds for $i = i_0$. By eq. 18(23), we know

$$\begin{aligned} L_{i_0}(t_{k+1}) &\leq L_{i_0}(t_k) + F_{i_0} T_M \sum_{j=i_0+1}^{n} L_j(t_{k+1}) \\ &\leq \overline{L}_{i_0}(t_k) + \frac{F_{i_0} T_M}{\rho_{i_0}} \sum_{j=i_0+1}^{n} \overline{L}_{i_0}(t_{k+1}) \\ &\leq \overline{L}_{i_0}(t_k) + (n-1)\frac{F_{i_0} T_M}{\rho_{i_0}} \overline{L}_{i_0}(t_{k+1}) \end{aligned} \tag{40}$$

Note that the above second inequality comes from the defintion of $\overline{L}_i(t_k)$ in eq. 36. If $\overline{L}_{i_0}(t_{k+1}) = L_i(t_{k+1})$, eq. 40 produces

$$\overline{L}_{i_0}(t_{k+1}) \leq \overline{L}_{i_0}(t_k) + (n-1)\frac{F_{i_0} T_M}{\rho_{i_0}} \overline{L}_{i_0}(t_{k+1})$$

Solving the above inequality w.r.t. $\overline{L}_{i_0}(t_{k+1})$, we get eq. 38.

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

$$\frac{\overline{L}_{i_0}(t_{k+1})}{\overline{L}_{i_0}(t_k)} \leq \frac{\rho_{i_0}\overline{L}_{i_0+1}(t_{k+1})}{\rho_{i_0}\overline{L}_{i_0+1}(t_k)} \leq \beta$$

By mathematical induction, we know eq. 38 works for $\forall i$.

From now on, we prove eq. 39. By the definition of I_k in eq. 17(20), we know, for any $j = I_k + 1, \cdots, n$,

$$4^{I_k}\rho_{f,I_k} L_{I_k}(t_k) \geq 4^j \rho_{f,j} L_j(t_k), \tag{41}$$

$$L_{I_k}(t_k) \geq 4 \prod_{m=I_k+1}^{j} \rho_m L_j(t_k) \tag{42}$$

When $\overline{L}_{I_k}(t_k) \geq 4\rho_{b,I_k}\gamma^k\|L(t_0)\|_\infty$, the defintion in eq. 36, together with eq. 42, yields

$$\overline{L}_{I_k}(t_k) = L_{I_k}(t_k) \geq 4\rho_{b,I_k}\gamma^k\|L(t_0)\|_\infty \tag{43}$$

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

$$L_{I_k}(t_{k+1}) \geq L_{I_k}(t_k)/2 \tag{44}$$

Combining eq. 43 and 44 yields

$$L_{I_k}(t_{k+1}) \geq 2\rho_{b,I_k}\gamma^k\|L(t_0)\|_\infty \tag{45}$$

Combining eq. 42 and 44 produces

$$L_{I_k}(t_{k+1}) \geq 2\prod_{m=I_k+1}^{j}\rho_m L_j(t_k) \quad (46)$$

Based on the defintion of $\overline{L}_j(t_k)$ and eq. 45, the above equation gives us

$$L_{I_k}(t_{k+1}) \geq 2\prod_{m=I_k+1}^{j}\rho_m \overline{L}_j(t_k) \quad (47)$$

Substituting eq. 38 into the above equation generates

$$L_{I_k}(t_{k+1}) \geq \frac{2}{\beta}\prod_{m=I_k+1}^{j}\rho_m\overline{L}_j(t_{k+1}) > \prod_{m=I_k+1}^{j}\rho_m\overline{L}_j(t_{k+1})$$

Particularlly, $L_{I_k}(t_{k+1}) > \rho_{I_k}\overline{L}_{I_{k+1}}(t_{k+1})$. So

$$\begin{aligned}\overline{L}_{I_k}(t_{k+1}) &= L_{I_k}(t_{k+1})\\ &\leq \frac{1}{2}\overline{L}_{I_k}(t_k) + \frac{F_{I_k}T_M}{\rho_{I_k}}\sum_{j=I_k+1}^{n}\overline{L}_{I_k}(t_{k+1})\\ &= \frac{1}{2}\overline{L}_{I_k}(t_k) + \frac{(n-1)F_{I_k}T_M}{\rho_{I_k}}\overline{L}_{I_k}(t_{k+1})\end{aligned}$$

Solving the above last inequality w.r.t. $\overline{L}_{I_k}(t_{k+1})$ yields eq. 39. ♢

Define

$$p(t_k) = \prod_{i=1}^{n}\overline{L}_i(t_k) \quad (48)$$

Lemma 4.3: If

$$p(t_k) \geq \prod_{i=1}^{n}\left(4\rho_{b,i}\gamma^k\|L(t_0)\|_\infty\right), \quad (49)$$

then

$$\overline{L}_{I_k}(t_k) \geq 4\rho_{b,I_k}\gamma^k\|L(t_0)\|_\infty \quad (50)$$

Proof: Under the condition of eq. 49, we first prove the following claim by contradition.

Claim: There must exist i such that

$$\overline{L}_i(t_k) \geq 4\rho_{b,i}\gamma^k\|L(t_0)\|_\infty \quad (51)$$

Suppose the above claim is false, i.e., for any $i = 1, 2, \cdots, n$,

$$\overline{L}_i(t_k) < 4\rho_{b,i}\gamma^k\|L(t_0)\|_\infty \quad (52)$$

Then we get $p(t_k) < \prod_{i=1}^{n} 4\rho_{b,i}\gamma^k\|L(t_0)\|_\infty$, which contradicts with eq. 49. So the claim in eq. 51 must be true.

There are 3 cases for I_k.

Case (1): $I_k = i$. Eq. 50 obviously holds.

Case (2): $I_k < i$. By the selection rule of I_k, we get

$$\begin{aligned}L_{I_k}(t_k) &\geq \frac{\rho_{f,i}}{\rho_{f,I_k}}L_i(t_k)\\ &\geq 4\rho_{b,I_k}\gamma^k\|L(t_0)\|_\infty\end{aligned}$$

So eq. 50 holds.

Case (3): $I_k > i$. Similar to Case (2). ♢

By Lemmas 4.2 and 4.3 and the definitions of $p(t_k)$, ρ and γ, we get

Corollary 4.4:

$$\frac{p(t_{k+1})}{p(t_k)} \leq 2\gamma^n, \forall k \quad (53)$$

When eq. 49 holds,

$$\frac{p(t_{k+1})}{p(t_k)} \leq \gamma^n \quad (54)$$

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

Proposition 4.5:

$$p(t_k) < 2\prod_{i=1}^{n}\left(4\rho_{b,i}\gamma^k\|L(t_0)\|_\infty\right), \forall k \quad (55)$$

Proof: For $k = 0$, eq. 55 holds. Suppose it holds when $k = k_0$. Now we prove it also works for $k = k_0 + 1$. There are 2 cases.

(1) When eq. 49 holds, we know, by eq. 54,

$$\begin{aligned}p(t_{k_0+1}) &\leq \gamma^n p(t_{k_0})\\ &< 2\prod_{i=1}^{n}\left(4\rho_{b,i}\gamma^{k_0+1}\|L(t_0)\|_\infty\right)\end{aligned}$$

i.e., eq. 55 holds for $k = k_0 + 1$.

(2) When eq. 49 does NOT hold, we know, by eq. 53,

$$\begin{aligned}p(t_{k_0+1}) &\leq 2\gamma^n p(t_{k_0})\\ &< 2\gamma^n\prod_{i=1}^{n}\left(4\rho_{b,i}\gamma^{k_0}\|L(t_0)\|_\infty\right)\\ &= 2\prod_{i=1}^{n}\left(4\rho_{b,i}\gamma^{k_0+1}\|L(t_0)\|_\infty\right)\end{aligned}$$

i.e., eq. 55 holds for $k = k_0 + 1$.

In summary, eq. 55 holds for both cases. ♢

Now we are ready to prove Proposition 2.3.

Proof: We want to get an upper bound of $L_i(t_k)$ for a given i. First we try to get an upper bound for $\overline{L}_j(t_k)$ with $j \neq i$.

If $j < i$, then we know

$$\overline{L}_j(t_k) \geq \rho_j\rho_{j+1}\cdots\rho_{i+1}\overline{L}_i(t_k) \quad (56)$$

If $j > i$, we get

$$\overline{L}_j(t_k) \geq \rho_{b,j}\gamma^k\|L(t_0)\|_\infty \quad (57)$$

Multiplying eq. 56 and 57 for all j, we get a lower bound on $p(t_k)$ as

$$\begin{aligned}p(t_k) &\geq (L_i(t_k))^i\prod_{m=1}^{i-1}(\rho_m\rho_{m+1}\cdots\rho_{i-1})\\ &\quad\times\prod_{m=i+1}^{n}\left(\rho_{b,m}\gamma^k\|L(t_0)\|_\infty\right)\end{aligned}$$

Combining the above equation with the upper bound of $p(t_k)$ in eq. 55 yields

$$(L_i(t_k))^i \leq 2\times 4^n\left(\rho_{b,i}\gamma^k\|L(t_0)\|_\infty\right)^i \quad (58)$$

Taking the i-th root on both sides of the above inequality produces

$$\begin{aligned}L_i(t_k) &\leq \sqrt[i]{2\times 4^n}\rho_{b,i}\gamma^k\|L(t_0)\|_\infty\\ &\leq 2\times 4^n\rho_{b,i}\gamma^k\|L(t_0)\|_\infty\diamondsuit\end{aligned}$$