**Preprints of the 21st IFAC World Congress (Virtual)**
**Berlin, Germany, July 12-17, 2020**

# Mitigation of Quantization Effect in Observer-based State-feedback Control with Quantized Observation

**Shuichi Ohno** * **Kuga Atarashi** ** **Tohru Katayama** ***

* *Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, 739-8527, JAPAN (e-mail: ohno@hiroshima-u.ac.jp).*
** *Hiroshima University, (e-mail: kuga_atarashi@hiroshima-u.ac.jp)*
*** *Professor Emeritus Kyoto University, (e-mail: tohru.katayama.65z@st.kyoto-u.ac.jp)*

**Abstract:** This paper studies observer-based state-feedback control of a linear system, where the observation signal is corrupted by the quantization error of an error feedback quantizer that consists of a uniform quantizer and a feedback filter. To mitigate the effect of the quantization error on the system output, we minimize the $H_2$ norm of the transfer function from the round-off error of the uniform quantizer to the system output with respect to the observer gain and the feedback filter alternatively. The minimization with respect to the observer gain is formulated into an optimization under a linear matrix inequality and a bilinear matrix inequality, whose minimum is numerically found by an XY-centering algorithm. The minimization with respect to the feedback filter is a convex optimization whose solution can be obtained numerically. The two minimization problems are iteratively solved, until the $H_2$ norm converges. A numerical example is provided to see the performance of our method.

*Keywords:* networked control system, observer, quantization, error feedback quantizer

## 1. INTRODUCTION

In a networked or a distributed control system, information between the controller and the plant is transmitted through rate limited communication channels. To transmit signals over rate-limited digital communication channels, continuous-valued or even discrete-valued signals have to be quantized into low-resolution signals. When only a small number of bits is assigned to represent the signals, the quantization error may cause serious stability/performance degradation. This motivates the research on control under limited data rates.

In Wong and Brockett (1999), keeping the states of a closed-loop system in a bounded region with state-feedback control is called *containability* and the minimum data rate for the containability is provided. Stabilizability and observability under a communication constraint have been studied in Tatikonda and Mitter (2004) for discrete-time linear time-invariant (LTI) systems. A necessary and sufficient condition on the information rate for asymptotic stabilizability and observability has been presented. The same condition has been shown in Nair and Evans (2003), together with a necessary and sufficient condition for exponential stabilizability of discrete-time LTI systems with random initial states.

The discussions on the minimum data rate based on the information theory give valuable insights into control under limited data rates. However, even if the rate is assured, the closed-loop system may not always be stabilized, since an ideal quantization is necessitated.

Indeed, Sontag (1984) shows that any unstable LTI systems cannot be globally exponentially stabilized with bounded control signals. Thus, to guarantee the global stability, an infinite-level quantization is required. If one can utilize a dynamic quantizer that changes its quantization law dynamically depending on its input, a finite-level quantization enables the global stability. For example, Fu and Xie (2009) prove a finite-level logarithmic quantizer can globally stabilize unstable feedback control systems. In Brockett and Liberzon (2000); Liberzon (2003), adaptive logarithmic quantizers that change their ranges and resolutions have been used to attain global asymptotic stabilization with quantized signals. Using the adaptive logarithmic quantizers, Chang et al. (2018) design a static output feedback control for discrete-time systems. However, since a fine quantization is necessary around the convergent point of the signal to be quantized, the global stability is possible only if there exists only one convergent point which is known *a priori* to the quantizer.

On the other hand, in Wakaiki et al. (2019), bounds for the ranges of the quantizers have been provided by using an observer to obtain sufficient conditions for local practical stability under bounded disturbances and noises when there is no reference signal. However, since the bounds depend on the reference signal, the bounds cannot be determined *a priori* without the knowledge of the reference signal. When a fixed known reference signal is not available, we can develop a control that minimizes the region of the state-vector in a certain sense. Ferrante et al. (2014) designs an observer to minimize the ellipsoid that contains the state-vector for a feedback control system with quantized output.

This paper considers an observer-based state-feedback control for linear systems, whose feedback control is synthesized with an existing technique, e.g., the conventional linear quadratic regulator technique, and the state-vector is estimated by an observer based on the quantized observation. To reduce the quantization error, we adopt the error feedback quantizer, which consists of a uniform quantizer and an error feedback filter. The error feedback quantizer is equivalent to a linearized discrete $\Delta\Sigma$ modulator which is widely used for analog to digital (A/D) conversion. For $\Delta\Sigma$ modulators, see, e.g., Schreier and Temes (2004) and references therein.

The quantization induces an error signal at the system output. We evaluate the mean squared error (MSE) of the error signal, assuming that the quantization range is large enough to prevent overflows, and the quantization error is an i.i.d. white noise with zero mean. The MSE is expressed as the $H_2$ norm of the transfer function from the round-off error of the uniform quantizer in the error feedback quantizer to the system output, which is a function of the observer gain and the noise transfer function (NTF) of the error feedback quantizer.

Our objective is to find the optimal observer gain and NTF that minimize the MSE. However, it is difficult to simultaneously minimize the MSE with respect to the observer gain and the NTF. To obtain reasonable observer gain and NTF, we utilize an alternate minimization. For a given error feedback filter, the observer is obtained by numerically solving an optimization problem under a linear matrix inequality (LMI) and a bilinear matrix inequality (BMI) by an XY-centering algorithm originally developed by Iwasaki and Skelton (1995). On the other hand, for a given observer gain, the optimal NTF that minimizes the MSE is given by solving a convex optimization problem as shown in Ohno and Tariq (2017). We iteratively minimize the MSE with respect to the observer gain and the error feedback filter, until the MSE converges. A numerical example is provided to see that our method exhibits reasonable performance.

**Notation:** $\mathbb{R}$ is the set of real numbers and $\mathbb{R}_+$ is the set of non-negative numbers. $I_n$ is the identity matrix of size $n \times n$. The z transform of a sequence (or a vector) $h = \{h_k\}_{k=0}^{\infty}$ is denoted as $H[\mathrm{z}] = \sum_{k=0}^{\infty} h_k \mathrm{z}^{-k}$. The output sequence $y$ of a discrete-time LTI system $H[\mathrm{z}]$ with the input sequence $x$ (i.e. $y = h * x$ where $*$ denotes the convolution) is expressed as $y = H[\mathrm{z}]x$. The $H_2$ norm of $H[\mathrm{z}]$ is defined as

$$\|H[\mathrm{z}]\|_2 = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{trace}\left( H^*[e^{j\omega}] H[e^{j\omega}] \right) d\omega \right\}^{\frac{1}{2}}$$

where $(\cdot)^*$ is the complex conjugate transpose operator, whereas the $H_\infty$ norm of $H[\mathrm{z}]$ is

$$\|H[\mathrm{z}]\|_\infty = \max_{\theta \in [-\pi,\pi)} \sigma_{\max}(H[e^{j\omega}])$$

where $\sigma_{\max}(H[e^{j\omega}])$ is the maximum singular value of $H[e^{j\omega}]$ for $\omega \in [-\pi, \pi)$.

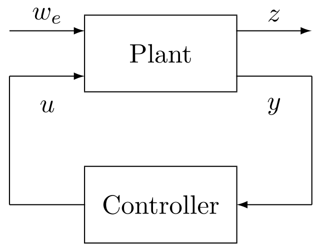


Fig. 1. Generalized plant

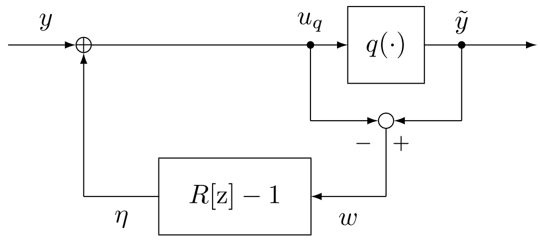


Fig. 2. Error feedback quantizer

## 2. FEEDBACK SYSTEM AND QUANTIZATION

Fig. 1 depicts a generalized plant, in which the plant is assumed to be LTI and the signals $w_e$, $z$, $y$, and $u$ are functions of time. The inputs to the plant are the exogenous input $w_e$ and the output $u$ of the controller. The controller generates the input $u$ to the plant, using the observation signal $y$ of the plant.

For simplicity, we consider discrete-time equivalent systems and express signals and systems in discrete-time and assume that the plant has a strictly proper single-input and single-output LTI system of order $n_p$. Let us denote its minimal state-space matrices of the plant as $(A_p, B_p, C_p, 0)$. Then, the plant is expressed as

$$x_{k+1} = A_p x_k + B_p u_k \quad (1)$$
$$y_k = C_p x_k \quad (2)$$

where $x_k$ is the state-vector at time $k$.

Suppose that the observation signal $y$ is transmitted over a digital communication channel. We have to convert the observation signal $y$ into a discrete-valued signal by using a quantizer. We adopt an error feedback quantizer that exhibits better performance than the conventional uniform quantizer.

Fig. 2 shows an error feedback quantizer that consists of a uniform quantizer $q(\cdot)$ and an error feedback filter. The error feedback quantizer is equivalent to a $\Delta\Sigma$ modulator which is widely used for analog to digital (A/D) conversion (see Schreier and Temes (2004) and references therein).

Let $d \in \mathbb{R}_+$ be the quantization step-size, and $N+1$ be the number of bits to represent a binary number, where $N$ be a positive integer. For a scalar input $\xi$, the output of the mid-tread uniform quantizer is given by

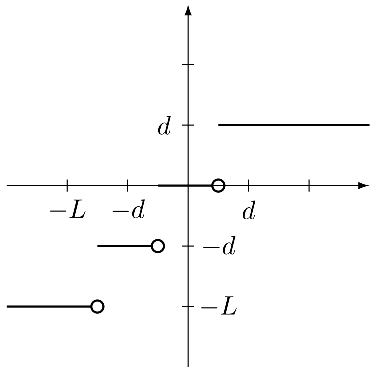


Fig. 3. Input/output characteristics of a mid-tread uniform 2-bit quantizer

$$q(\xi) = \begin{cases} L-d, & \xi > L - \frac{d}{2} \\ ld, & \xi \in [(l-\frac{1}{2})d, (l+\frac{1}{2})d) \\ & \text{for an integer } l \text{ and } -L-\frac{d}{2} \le \xi \le L - \frac{d}{2} \\ -L, & \xi < -L - \frac{d}{2} \end{cases} \tag{3}$$

where $L = 2^N d$. Fig. 3 illustrates the input/output characteristics of a mid-tread uniform quantizer when $N = 1$.

We assume that the dynamic range of the quantizer is large enough to prevent overflows. Under this assumption, we model the quantization error as a uniform random variable with zero mean and variance $\sigma_w^2 = d^2/12$.

For scalar inputs, the validity of this model has been well demonstrated (see, e.g., Widrow and Kollár (2008)). The so-called white noise assumption further assumes that the errors at different time are statistically independent of each other. Bennett (1948) demonstrates that the whiteness assumption on the quantization error approximately holds true when there is no overflow, the quantization interval is sufficiently small, and a sufficiently large number of bits is assigned to each value. On the other hand, Gray (1990) shows that the whiteness assumption for the quantization error of the uniform quantizer in the error feedback quantizer is not always true. However, even if the whiteness assumption does not strictly hold, our design method below is still valid, since it minimizes the $H_2$ norm of the transfer function from the error to the output.

The round-off error of the uniform quantizer $q(\cdot)$ is given by

$$w = \tilde{y} - u_q \tag{4}$$

where $u_q$ is the input to the uniform quantizer; see Fig. 2. It is filtered by the error feedback filter $R[\mathrm{z}] - 1$ and then it is fed back to the input to the uniform quantizer. (Here $-1$ in $R[\mathrm{z}] - 1$ is just for simplicity of presentation. ) The error feedback filter $R[\mathrm{z}] - 1$ is assumed to strictly proper, that is, $R[\infty] = 1$. It should be noted that if $R[\mathrm{z}] = 1$, there is no feedback and the quantizer boils down into a conventional uniform quantizer.

The quantization error of the error feedback quantizer is defined as

$$e = \tilde{y} - y \tag{5}$$

which is found from (4) and $u_q = y + (R[\mathrm{z}] - 1)w$ to be

$$e = R[\mathrm{z}]w. \tag{6}$$

Then, the output signal of the error feedback quantizer can be expressed as

$$\tilde{y} = y + R[\mathrm{z}]w. \tag{7}$$

The filter $R[\mathrm{z}]$ is called a *noise transfer function* (NTF) or *noise shaping filter* (NSF).

Azuma and Sugie (2012); Okajima et al. (2016) have utilized a type of $\Delta\Sigma$ modulators different from ours, named as a *dynamic quantizer*, to reduce the effect of quantization in which the output of the uniform quantizer rather than the round-off error is filtered and fed back to the input of the uniform quantizer. As pointed out by Ohno et al. (2018), our error feedback quantizer subsumes the *dynamic quantizer*; if the NTF $R[\mathrm{z}]$ has minimum phase, the two quantizers are equivalent but the *dynamic quantizer* cannot be optimal if the optimal $R[\mathrm{z}]$ has non-minimum phase.

## 3. EFFECT OF QUANTIZATION IN OBSERVER-BASED STATE-FEEDBACK CONTROL

We adopt the conventional state-feedback control. Let $w_{e,k}$ be the reference signal for the state-vector at time $k$. If the state-vector is available at the controller, the input to the plant is generated by

$$u_k = K(w_{e,k} - x_k) \tag{8}$$

where $K$ is the feedback gain, which is designed, e.g., by a control law for the optimal linear quadratic regulator.

Since the state-vector is not available at the controller, we estimate the state-vector based on quantized output $\tilde{y}_k$. The linear minimum variance filter is known as the optimal linear filter that minimizes the estimation variance of the state-vector; however, it is not always optimal for other purposes. This paper utlizes a Luenberger observer to minimize the effect of the round-off error on the output of the plant.

The observer generates the estimate $\hat{x}_{k+1}$ of the state-vector at time $k+1$ by

$$\hat{x}_{k+1} = A_p\hat{x}_k + B_p u_k + F(\tilde{y}_k - C_p\hat{x}_k) \tag{9}$$

where $F$ is the observer gain. Let us define an augmented state as $[x_k^T, \xi_k^T]^T$ with the state-estimation error

$$\xi_k = x_k - \hat{x}_k. \tag{10}$$

The augmented state-space equations can be expressed as

$$\begin{bmatrix} x_{k+1} \\ \xi_{k+1} \end{bmatrix} = A_h \begin{bmatrix} x_k \\ \xi_k \end{bmatrix} + \begin{bmatrix} B_p K \\ \mathbf{0} \end{bmatrix} r_k + B_h e_k \tag{11}$$

$$y_k = C_h \begin{bmatrix} x_k \\ \xi_k \end{bmatrix} \tag{12}$$

where

$$A_h = \begin{bmatrix} A_p - B_p K & B_p K \\ \mathbf{0} & A_p - FC_p \end{bmatrix} \tag{13}$$

$$B_h = \begin{bmatrix} \mathbf{0} \\ F \end{bmatrix} \tag{14}$$

$$C_h = [\, C_p \;\; \mathbf{0} \,]. \tag{15}$$

Suppose that $A_h$ is of Shur. Since the round-off error is bounded if there is no overflow, the quantization error $e$ is bounded for a stable NTF $R[\mathrm{z}]$; then each entry of the state-vector is bounded. Thus, there exists an ellipsoid described by a positive definite matrix $\tilde{P}$ that satisfies

$$\begin{bmatrix} x_{k+1} \\ \xi_{k+1} \end{bmatrix}^T \tilde{P} \begin{bmatrix} x_{k+1} \\ \xi_{k+1} \end{bmatrix} \le 1 \tag{16}$$

for all $k \ge \bar{N}$ for a given integer $\bar{N} \ge 0$. Ferrante et al. (2014) construct an observer to find a minimum ellipsoid that contains the state-vector at any time, while keeping the global ultimate boundedness stabilization property, defined in Khalil (2002), of the observer-based control for continuous-time linear systems with quantized observation.

This paper focuses on the system output rather than the state-vector. We would like to find the optimal observer and the NTF that minimize the effect of the quantization error on the output of the plant. For simplicity, let the system output $z$ be identical with the observation signal $y$; see Fig. 1. We denote the transfer function from the quantization error $e$ to the output of the plant $y$ as $H[\mathrm{z}]$, which is given by

$$\begin{aligned} H[\mathrm{z}] &= C_h(\mathrm{z}I_{n_h} - A_h)^{-1}B_h \\ &= C_p(\mathrm{z}I_{n_p} - A_p + B_pK)^{-1}B_pK(\mathrm{z}I_{n_p} - A_p + FC_p)^{-1}F \end{aligned} \tag{17}$$

where $n_h = 2n_p$.

The quantization error $e$ goes through $H[\mathrm{z}]$. At the output, the error due to quantization is given by

$$\epsilon = H[\mathrm{z}]e = H[\mathrm{z}]R[\mathrm{z}]w. \tag{18}$$

If the round-off error $w$ of the uniform quantizer is a statistically white random signal with zero mean and variance $\sigma_w^2$, then the mean squared error (MSE) is given by

$$E\{\|\epsilon_k\|_2^2\} = \|H[\mathrm{z}]R[\mathrm{z}]\|_2^2\sigma_w^2 \tag{19}$$

where $E\{\cdot\}$ denotes the expectation operator. Similarly the feedback signal $\eta$ has zero mean and variance given by

$$E\{\|\eta_k\|_2^2\} = \|R[\mathrm{z}] - 1\|_2^2\sigma_w^2. \tag{20}$$

The variance of the feedback signal has to be suppressed to avoid overflows.

Now, we would like to minimize the MSE with respect to the observer gain $F$ and the NTF $R[\mathrm{z}]$, under the constraint on the variance of the feedback signal $\eta$. More specifically, we consider the following minimization problem:

$$\min_{F\in\mathbb{R}^{n_p\times 1}, R[\mathrm{z}]\in RH_\infty} \mu_\epsilon \tag{21}$$

subject to

$$\|H[\mathrm{z}]R[\mathrm{z}]\|_2^2 < \mu_\epsilon \tag{22}$$

$$\|R[\mathrm{z}] - 1\|_2^2 < \mu_\eta \tag{23}$$

where $RH_\infty$ is the set of stable proper rational functions with real coefficients.

## 4. SYNTHESIS OF OBSERVER AND QUANTIZER

Since $H[\mathrm{z}]R[\mathrm{z}]$ has multiplications of design variables, it is not that easy to obtain the global optimum for $\mu_\epsilon$ by simultaneously minimizing it with respect to the observer gain $F$ and the NTF $R[\mathrm{z}]$. Here, we utilize a simple iterative procedure to obtain a sub-optimal solution, which can be obtained with a reasonable cost.

The transfer function $H[\mathrm{z}]$ depends on the observer gain $F$, while the NTF $R[\mathrm{z}]$ does not. Thus, we can apply an alternate optimization which minimizes the objective with respect to $F$ and $R[\mathrm{z}]$ alternatively .

Suppose that an NTF $R[\mathrm{z}]$ is given, we can optimize our objective with respect to $F$ as follows:

$$\min_{F\in\mathbb{R}^{n_p\times 1}} \mu_\epsilon \tag{24}$$

subject to

$$\|H[\mathrm{z}]R[\mathrm{z}]\|_2^2 < \mu_\epsilon \tag{25}$$

which can be formulated as an optimization.

Let the order of $R[\mathrm{z}]$ be $n_r$. We denote the state-space matrices for $R[\mathrm{z}]$ as $(A_r, B_r, C_r, 1)$. Then, the state-space equations from $w_k$ to $\epsilon_k$ is expressed as

$$\tilde{x}_{k+1} = \mathcal{A}\tilde{x}_k + \mathcal{B}w_k \tag{26}$$

$$\epsilon_k = \mathcal{C}\tilde{x}_k \tag{27}$$

where

$$\mathcal{A} = \begin{bmatrix} A_h & B_hC_r \\ \mathbf{0} & A_r \end{bmatrix} \tag{28}$$

$$\mathcal{B} = \begin{bmatrix} B_h \\ B_r \end{bmatrix} \tag{29}$$

$$\mathcal{C} = \begin{bmatrix} C_h & \mathbf{0} \end{bmatrix}. \tag{30}$$

It is known that $||H[\mathrm{z}]R[\mathrm{z}]||_2^2 < \mu_\epsilon$ if and only if there exists a positive definite matrix $\mathcal{P}$ such that (see e.g., Boyd et al. (1997))

$$\begin{bmatrix} \mathcal{P} & \mathcal{P}\mathcal{A} & \mathcal{P}\mathcal{B} \\ \mathcal{A}^T\mathcal{P} & \mathcal{P} & \mathbf{0} \\ \mathcal{B}^T\mathcal{P} & \mathbf{0} & 1 \end{bmatrix} \succ \mathbf{0} \tag{31}$$

$$\begin{bmatrix} \mu_\epsilon & \mathcal{C} \\ \mathcal{C}^T & \mathcal{P} \end{bmatrix} \succ \mathbf{0}. \tag{32}$$

Then, the problem is equivalent to the minimization of $\mu_\epsilon$ with respect to $F \in \mathbb{R}^{n_p\times 1}$ and a positive definite matrix $\mathcal{P}$ subject to (31) and (32).

We can express $\mathcal{A}$ as

$$\mathcal{A} = \begin{bmatrix} A_p - B_pK & -B_pK & B_pC_r \\ \mathbf{0} & A_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_r \end{bmatrix} - \begin{bmatrix} \mathbf{0} \\ F \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} & C_p & -C_r \end{bmatrix} \tag{33}$$

Although (32) is an LMI of $\mathcal{P}$, (31) is a non-convex BMI, since $\mathcal{P}A$ contains a multiplication of $\mathcal{P}$ and $F$. This also holds true for $\mathcal{P}B$. Here, we minimize $\mu_\epsilon$ with respect to $\mathcal{P}$ and $F$ alternately; for a fixed $\mathcal{P}$, $\mu_\epsilon$ is minimized with respect to $F$ and then for the obtained $F$, $\mu_\epsilon$ is minimized with respect to $\mathcal{P}$. This alternate optimization is repeated until $\mu_\epsilon$ converges. To obtain a smaller $\mu_\epsilon$, we utilize the XY-centering algorithm originally developed by Iwasaki and Skelton (1995).

On the other hand, for a given $H[\mathrm{z}]$, we can minimize our objective with respect to $R[\mathrm{z}]$, that is, we can solve the following minimization problem:

$$\min_{R[z]\in RH_\infty} \mu_\epsilon \tag{34}$$

subject to

$$\|H[z]R[z]\|_2^2 < \mu_\epsilon \tag{35}$$

$$\|R[z]-1\|_2^2 < \mu_\eta \tag{36}$$

As shown in Ohno and Tariq (2017), by putting the order of $R[z]$ to be $n_h$, which equals the order of $H[z]$, the problem above can be converted into a convex optimization; the latter can be solved numerically and efficiently by using a numerical solver.

## 5. A NUMERICAL EXAMPLE

We compare our observer-based state-feedback control with that of Ferrante et al. (2014) which uses a uniform quantizer and an observer designed based on a minimum ellipsoid that contains the state-vector. In our simulations, we utilize zero-order hold equivalent discrete-time models for continuous-time systems with sampling period $T_s = 0.01$.

The quantization interval for every quantizer is set to be $d = 1/2^4$. Each entry of initial sate-vectors is unity. The feedback gain $K$ of Ferrante et al. (2014) is used for state-feedback control, which is given by

$$K = [\,1.7047\ 1.0566\,]. \tag{37}$$

For $N = 1000$, we evaluate the empirical MSE defined as

$$\mathrm{MSE} = \frac{1}{N}\sum_{k=1}^{N} \epsilon_k^T \epsilon_k \tag{38}$$

where $\epsilon_k$ is the error signal at the output given by (18). It should be noted that the outputs for different observers are not equal even if there is no quantization error since the system depends on the observer gain.

We first design the observer gain for a uniform quantizer. Then, we design the observer gain and the NTF alternatively. For a fixed NTF, the MSE is numerically minimized with respect to the observer gain $L$ by the XY-centering algorithm by Iwasaki and Skelton (1995), whereas, for a fixed observer gain, the MSE is numerically minimized with respect to the NTF $R[z]$ by using a numerical solver CVX by Grant and Boyd (2014).

The plant is a continuous-time system whose state-space matrices are given by

$$A = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad C = [\,1\ \ 0\,]. \tag{39}$$

The state-space matrices of the discrete-time equivalent are

$$A_p = \begin{bmatrix} 1 & 0.0100 \\ 0.0100 & 0.9901 \end{bmatrix}, \quad B_p = \begin{bmatrix} 0.0101 \\ 0.0100 \end{bmatrix}, \quad C_p = [\,1\ \ 0\,]. \tag{40}$$

Each entry of the initial sate-estimates is unity to see only the effect of the round-off error, while the reference signal $r_k$ is set to be a zero vector.

Table 1 displays theoretical and empirical MSEs for our methods with a uniform quantizer and with our designed feedback quantizer, and Ferrante et al. (2014). For each observer, the empirical MSE is larger than its corresponding theoretical MSE. This is because our whiteness assumption on the quantization error does not hold true after the system output converges.

Table 1. Theoretical and empirical MSEs of our methods with a uniform quantizer and with our designed error feedback quantizer, and Ferrante et al. (2014).

| | theoretical MSE | empirical MSE |
|---|---|---|
| Uniform quantizer | $4.6656 \cdot 10^{-6}$ | $7.2097 \cdot 10^{-4}$ |
| Feedback quantizer | $2.3384 \cdot 10^{-10}$ | $7.6826 \cdot 10^{-10}$ |
| Ferrante et al. (2014) | $5.0689 \cdot 10^{-6}$ | $6.5724 \cdot 10^{-4}$ |

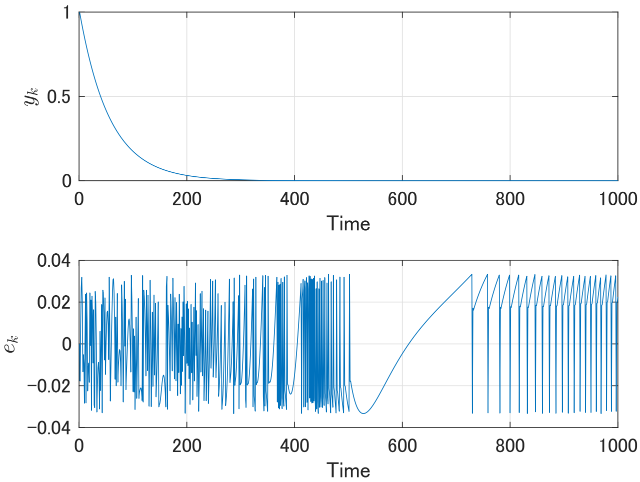


Fig. 4. System output $y$ and round-off error $e$ with the designed observer and NTF

With the uniform quantizer, our method has a smaller theoretical MSE than the method of Ferrante et al. (2014) but has a larger empirical MSE. This is also due to the loss of randomness of the round-off error after the convergence. With the designed feedback quantizer, our method enjoys the least theoretical and empirical MSEs.

Fig. 4 shows the system output $y$ and the quantization error $e$ of the system with the designed observer and NTF. It can be seen that as the system output converges, the round-off error of the uniform quantizer in our error feedback quantizer becomes deterministic. However, even if the whiteness assumption is not valid, our method based on the $H_2$ norm works well.

## 6. CONCLUSIONS

We have studied observer-based state-feedback control of a linear system whose observation signal is corrupted by the quantization error. To mitigate the effect of the quantization error, an alternate minimization of the MSE at the system output has been proposed to design the observer gain and the NTF of the error feedback quantizer. The MSE is minimized with respect to the observer gain and the NTF alternatively until it converges. A numerical example demonstrates the effectiveness of our method.

## REFERENCES


Azuma, S. and Sugie, T. (2012). Dynamic quantization of nonlinear control systems. *IEEE Transactions on Automatic Control*, 57(4), 875–888. doi:10.1109/TAC.2011.2167824.

Bennett, W.R. (1948). Spectra of quantized signals. *The Bell System Technical Journal*, 27(3), 446–472. doi:10.1002/j.1538-7305.1948.tb01340.x.

Boyd, S., Ghaoul, L.E., Feron, E., and Balakrishnan, V. (1997). *Linear Matrix Inequalities in System and Control Theory.* Society for Industrial and Applied Mathematics.

Brockett, R.W. and Liberzon, D. (2000). Quantized feedback stabilization of linear systems. *IEEE Transactions on Automatic Control*, 45(7), 1279–1289. doi:10.1109/9.867021.

Chang, X., Xiong, J., Li, Z., and Park, J.H. (2018). Quantized static output feedback control for discrete-time systems. *IEEE Transactions on Industrial Informatics*, 14(8), 3426–3435. doi:10.1109/TII.2017.2774446.

Ferrante, F., Gouaisbaut, F., and Tarbouriech, S. (2014). Observer-based control for linear systems with quantized output. In *2014 European Control Conference (ECC)*, 964–969. doi:10.1109/ECC.2014.6862246.

Fu, M. and Xie, L. (2009). Finite-level quantized feedback control for linear systems. *IEEE Transactions on Automatic Control*, 54(5), 1165–1170. doi:10.1109/TAC.2009.2017815.

Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. URL `http://cvxr.com/cvx`.

Gray, R.M. (1990). Quantization noise spectra. *IEEE Transactions on Information Theory*, 36(6), 1220–1244. doi:10.1109/18.59924.

Iwasaki, T. and Skelton, R.E. (1995). The XY-centring algorithm for the dual LMI problem: a new approach to fixed-order control design. *International Journal of Control*, 62(6), 1257–1272. doi:10.1080/00207179508921598. URL `https://doi.org/10.1080/00207179508921598`.

Khalil, H. (2002). *Nonlinear Systems.* Pearson Education. Prentice Hall. URL `https://books.google.co.jp/books?id=t_d1QgAACAAJ`.

Liberzon, D. (2003). Hybrid feedback stabilization of systems with quantized signals. *Automatica*, 39(9), 1543 – 1554. doi:https://doi.org/10.1016/S0005-1098(03)00151-1.

Nair, G.N. and Evans, R.J. (2003). Exponential stabilisability of finite-dimensional linear systems with limited data rates. *Automatica*, 39(4), 585–593. doi:10.1016/S0005-1098(02)00285-6.

Ohno, S., Ishihara, Y., and Nagahara, M. (2018). Minmax design of error feedback quantizers without overloading. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 65(4), 1395–1405. doi:10.1109/TCSI.2017.2758801.

Ohno, S. and Tariq, M.R. (2017). Optimization of noise shaping filter for quantizer with error feedback. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 64(4), 918–930. doi:10.1109/TCSI.2016.2627031.

Okajima, H., Sawada, K., and Matsunaga, N. (2016). Dynamic quantizer design under communication rate constraints. *IEEE Transactions on Automatic Control*, 61(10), 3190–3196. doi:10.1109/TAC.2015.2509438.

Schreier, R. and Temes, G.C. (2004). *Understanding Delta-Sigma Data Converters.* Wiley-IEEE Press.

Sontag, E.D. (1984). An algebraic approach to bounded controllability of linear systems. *International Journal of Control*, 39(1), 181–188.

Tatikonda, S. and Mitter, S. (2004). Control Under Communication Constraints. *IEEE Transactions on Automatic Control*, 49(7), 1056–1068. doi:10.1109/TAC.2004.831187.

Wakaiki, M., Zanma, T., and Liu, K. (2019). Observer-based stabilization of systems with quantized inputs and outputs. *IEEE Transactions on Automatic Control*, 64(7), 2929–2936. doi:10.1109/TAC.2018.2873355.

Widrow, B. and Kollár, I. (2008). *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications.* Cambridge University Press New York.

Wong, W.S. and Brockett, R.W. (1999). Systems with finite communication bandwidth constraints. II. stabilization with limited information feedback. *IEEE Transactions on Automatic Control*, 44(5), 1049–1053.