On the Numerical Optimization Design of Continuous-Time Quantizer: A Matrix Uncertainty Approach

Kenji Sawada and Seiichi Shin

Abstract—For the networked control systems, the quantized control problem is one of the challenging problems since the continuous-valued signals are compressed and quantized to the discrete-valued signals via the communication channel and such the quantization often degrades the control performance. In terms of the broadbandization and the robustness of the networked control systems, this paper considers the continuous-time quantized control. In the quantized control, it is important to design a quantizer that minimizes the output difference between before and after the quantizer implementation. This paper describes a numerical optimization method of a continuous-time quantizer considering the switching speed. Using a matrix uncertainty approach of sampled-data control, we clarify that both of the temporal and spatial resolution constraints can be considered in analysis and synthesis, simultaneously.

I. INTRODUCTION

With the rapid network technology development, a number of studies of the networked control systems (NCSs) have been done actively. One of the challenging problems of the NCSs is the quantized control problem [1]–[5]. In NCSs, the continuous-valued signals are compressed and quantized to the discrete-valued signals via the quantizer of the communication channel and such the quantization often degrades the control performance. Hence, one of the desirable quantizers is one which minimizes the performance error between before and after the quantizer insert. Motivated by this, the papers [6]–[9] have provided optimal dynamic quantizers for the following problem formulation in the discrete-time domain: For a given plant \( P \), synthesize a “dynamic” quantizer \( Q_d \) such that the system \( \Sigma_Q \) composed of \( P \) and \( Q_d \) in Fig. 1 (a) “optimally” approximates the plant \( P \) in Fig. 1 (b) in the sense of the input-output relation. The obtained quantizer allows us to design various controllers for the plant \( P \) based on the conventional control theories. Also, this framework is helpful in not only the NCS problem but also various control problems such as hybrid control, embedded system control, on-off actuator control and so on.

When we consider control of mechanical system with on-off actuator, first, the controlled object and its uncertainties are usually modeled in the continuous-time domain. Second, the model and its uncertainties are discretized to apply the above dynamic quantizer. However, the discretization sometimes results in complicated uncertainties compared with the original model and creates undesirable complexity in robust control. The continuous-time setting quantizer is suitable for the robust control of the quantized system in comparison with the discrete-time one. Then, our early works [10], [11] have considered the continuous-time setting. In the works, it is assumed that the switching process of discretizing the continuous-valued signal is sufficiently quick relative to the control frequency and only the spatial determination (quantized accuracy) is considered as the quantization effect. This is because the switching speed of the continuous-time delta-sigma modulator for wireless broadband network systems is from 1MHz to 100MHz [12], [13].

On the other hand, the above assumption is essentially weak in the case of the slow switching such as the mechanical systems with on-off actuators [14]. For the slow switching, we need to consider the quantization effect on both the switching speed and the spatial constraints in continuous-time. For example, Ishikawa et al. [15] proposed a two-step design of feedback modulator: (i) the control performance of the modulator is considered under only the spatial constraint, (ii) the modulator is tuned in terms of the switching speed constraint. However, the structure of the modulator is restricted compared with the dynamic quantizer and the obtained modulator is not always optimal.

In the paper, we propose a numerical optimization method of the continuous-time dynamic quantizer considering temporal resolution (switching speed) and spatial resolution (quantized accuracy) constraints. In addition to the invariant set analysis [16], [17] (similarly to [9]–[11]), the paper utilizes a matrix uncertainty approach [18], [19] which is proposed in sampled-data control framework [20]. Although the obtained results can be more conservative than the early works [10], [11] from the viewpoint of the class of the exogenous input and the applicable plants, both of the temporal and spatial resolution constraints can be considered in

K. Sawada is with the Department of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications, Chofu, Tokyo, 182-8585, Japan. knj.sawada@uec.ac.jp
S. Shin is with the Department of Mechanical Engineering and Intelligent Systems, The University of Electro-Communications, Chofu, Tokyo, 182-8585, Japan. shin@se.uec.ac.jp

Fig. 1. Two systems.
analysis and synthesis, simultaneously. For the fast switching case, the proposed conditions converge to the corresponding conditions of our early works. For the slow switching, we make a comparison between the proposed and the existing methods in [10], [11] through numerical examples. New insight is also offered into the existing methods.

Notation: The set of \( n \times m \) (positive) real matrices is denoted by \( \mathbb{R}^{n \times m} \). The set of \( n \times m \) (positive) integer matrices is denoted by \( \mathbb{N}^{n \times m} \). \( I_n \) and \( I_m \) (or for simplicity of notation, 0 and \( I \)) denote the \( n \times m \) zero matrix and the \( m \times m \) identity matrix, respectively. For a matrix \( M \), \( M^T \), \( \rho(M) \) and \( \sigma_{\text{max}}(M) \) denote its transpose, its spectrum radius and its maximum singular value, respectively. For a vector \( x \), \( x_i \) is the \( i \)-th entry of \( x \). For a symmetric matrix \( X \), \( X > 0 \) \((X \geq 0)\) means that \( X \) is positive (semi) definite. For an input \( u \) and the exogenous input \( v \), the quantizer output, respectively. The continuous-valued signal state vector, the measured output, the exogenous input, the quantizer output, respectively. The set of \( \mathbb{R}^n \times \mathbb{R}^m \) is positive (semi) definite. For a vector given by

\[
\begin{align*}
\|X\| &= \sup_{\|x\| = 1} \langle x, Mx \rangle, \\
\|X\|_2 &= \sup_{\|x\| = 1} \|Mx\|_2.
\end{align*}
\]

Notation: \( A \), \( B \), \( C \), \( D \), \( e_Q \), \( J \), \( Q \), \( \sigma_0 \), \( \sigma_\infty \), \( \sigma_r \), \( \theta \), \( \|\cdot\| \), \( \langle \cdot, \cdot \rangle \)

II. PROBLEM FORMULATION

Consider the discrete-valued input system \( \Sigma_Q \) in Fig. 1 (a), which consists of the linear time invariant (LTI) continuous-time plant \( P \) and the quantizer \( v = Q_d(u) \). The system \( P \) is given by

\[
P : \begin{bmatrix}
\dot{x} \\
z
\end{bmatrix} = \begin{bmatrix}
A & B \\
C & 0
\end{bmatrix} \begin{bmatrix}
x \\
v
\end{bmatrix},
\]

where \( x \in \mathbb{R}^n \), \( z \in \mathbb{R}^m \), \( u \in \mathbb{R}^m \), and \( v \in \mathbb{R}^m \) denote the state vector, the measured output, the exogenous input, the quantizer output, respectively. The continuous-valued signal \( u \) is quantized into the discrete-valued signal \( v \) via the quantizer \( Q_d \). We assume that the matrix \( A \) is Hurwitz, that is, the usual system in Fig. 1 (b) is stable in the continuous-time domain. The initial state is given as \( x(0) = x_0 \).

For the system \( P \), consider the continuous-time dynamic quantizer \( v = Q_d(u) \) with the state vector \( x_Q \in \mathbb{R}^n_Q \) as shown in Fig. 2. Its switching speed \( h \in \mathbb{R}_+ \) (or its temporal resolution) is determined by the operator \( HS \), which converts the continuous-time signal \( g \) into the low temporal resolution signal \( \hat{g} \) as follows:

\[
HS : g \rightarrow \hat{g} : \hat{g}(kh + \theta) = g[k],
\]

\[
g[k] = g(kh), \quad k = 0, 1, 2, 3, ..., \theta \in [0, h).
\]

That is, \( \hat{u}_Q = HSu_Q \). \( S \) is the ideal sampler with the sampling period \( h \) and \( H \) is the hold operator. The spatial resolution of the quantizer \( Q_d \) is expressed by the static quantizer \( q : \mathbb{R}^m \rightarrow d\mathbb{N}^m \) with the quantization interval \( d \in \mathbb{R}_+ \), i.e.,

\[
v = q(HSu_Q), \quad u_Q = u + v_Q
\]

and the continuous-time LTI filter \( Q \) is given by

\[
\begin{bmatrix}
\dot{x}_Q \\
v_Q
\end{bmatrix} = \begin{bmatrix}
A_Q & B_Q \\
C_Q & 0
\end{bmatrix} \begin{bmatrix}
x_Q \\
e_Q
\end{bmatrix}, \quad e_Q := v - u.
\]

Note that \( q \) is of the nearest-neighbor type toward \( -\infty \) with the quantization interval \( d \in \mathbb{R}_+ \) such as the midtread type quantizer in Fig. 3 and the initial state is given by \( x_Q(0) = 0 \) for the drift-free of \( Q_d \) [6], [7].

For the system \( \Sigma_Q \) in Fig. 1 (a) with the initial state \( x_0 \) and the exogenous input \( u \in L_p^\infty \), \( z(t, x_0, Q_d(u)) \) denotes the output of \( z \) at the time \( t \). Also, for the system in Fig. 1 (b) without \( Q_d \), \( z^*(t, x_0, u) \) denotes its output at the time \( t \). Consider the following cost function:

\[
J(Q_d) := \sup_{(x_0, u) \in \mathbb{R}^n \times L_p^\infty} z_p(x_0, u),
\]

\[
z_p(x_0, u) := \sup_t \|z(t, x_0, Q_d(u)) - z^*(t, x_0, u)\|.
\]

If the quantizer minimizes \( J(Q_d) \), the system \( \Sigma_Q \) “optimally” approximates the usual system \( P \) in the sense of the input-output relation.

Our early works [10], [11] have proposed an optimal dynamic quantizer for the cost function \( J(Q_d) \) for the fast switching case \( h = 0 \). That is, only the spatial deterioration has been concerned. To consider the temporal resolution constraint caused by the operator \( HS \), this paper modifies the cost function as follow:

\[
J_h(Q_d) := \sup_{(x_0, u) \in \mathbb{R}^n \times L_p^\infty} \sup_{k \in \mathbb{N}_+} \sup_{\theta \in [0, h)} \|z(kh + \theta, x_0, Q_d(u)) - z^*(kh + \theta, x_0, u)\|.
\]

Such the optimal quantizer minimizes the output error between the systems in Figs. 1 (a) and (b) in terms of the input-output relation under the temporal and spatial resolution constraints.

Motivated by the above, our objective is to solve the following continuous-time dynamic quantizer synthesis problem (E): For the system \( \Sigma_Q \) composed of \( P \) and \( Q_d \) with the initial state \( x_0 \in \mathbb{R}^n \) and the exogenous input \( u \in L_p^\infty \).
suppose that the quantization interval \( d \in \mathbb{R}_+ \), the switching speed \( h \in \mathbb{R}_+ \) and the performance level \( \gamma \in \mathbb{R}_+ \) are given. Characterize a continuous-time dynamic quantizer \( Q_d \) (i.e., find parameters \( (n_Q, A_Q, B_Q, C_Q) \)) achieving \( J_h(Q_d) \leq \gamma \).

III. MAIN RESULT

A. SYSTEM EXPRESSION

In this subsection, we consider the system expression for the quantizer analysis. Define the quantization error \( e \) as

\[
e := q(\hat{u}_Q) - \hat{u}_Q = v - \hat{u}_Q.
\]  

(3)

From the properties of the quantizer \( q \) and the operator \( HS \),

\[
e(kh + \theta) = e[k] \in \left[-\frac{d}{2}, \frac{d}{2}\right], \quad k = 0, 1, 2, \ldots, \theta \in [0, h)
\]

holds where \( e[k] = e(kh) \). Then, one gets

\[
v(kh + \theta) = v_Q[k] + u[k] + e[k]
\]  

(4)

where \( v_Q[k] = v_Q(kh) \) and \( u[k] = u(kh) \) for \( k = 0, 1, 2, \ldots, \theta \in [0, h) \). In this case, by using the sampled-data control technique, the following lemma holds.

Lemma 1: For the cost function \( J_h(Q_d) \), the difference between \( z(kh + \theta, x_0, Q_d(u)) \) and \( z*(kh + \theta, x_0, u) \) for \( k = 0, 1, 2, \ldots, \theta \in [0, h) \) is given by the following system:

\[
\begin{align*}
\Sigma : & \quad \xi[k + 1] = A_{\xi}[k] + B_{\xi}[k] \\
& \quad + \int_{0}^{h} e^{A_{\xi}(h-\tau)} B_{\xi}(kh + \tau)d\tau \\
& \quad + C_{\xi} \int_{0}^{h} e^{A_{\xi}(h-\tau)} B_{\xi}(kh + \tau)d\tau \\
& \quad e_{\xi}(\theta) = \xi[k] + \mathcal{D}(\theta) e[k]
\end{align*}
\]  

(5)

where \( \xi[0] = 0 \), \( u(kh + \tau) := u[k] - u(kh + \tau) \), the matrices \( A, B, C(\theta) \) and \( D(\theta) \) are defined as follows:

\[
A := \begin{bmatrix}
e^{Ah} & 0 \\
e^{A_\theta h} & e^{A_\theta h}
\end{bmatrix}, \quad B := \begin{bmatrix}
e^{A_\theta k} & 0 \\
e^{A_\theta k} & e^{A_\theta k}
\end{bmatrix}, \quad C := \begin{bmatrix}
C e^{A_\theta} \\
0
\end{bmatrix}, \quad D(\theta) := C^T \int_{0}^{\theta} e^{A_\theta \tau} d\tau,
\]

We focus on \( u(kh + \tau) \) of \( \Sigma \). For the operator \( HS \),

\[
\lim_{h \to 0} \|[(I - HS)u](t)| \neq 0
\]

holds. This implies that we cannot ignore the temporal resolution constraint on the cost function \( J_h(Q_d) \) even if \( h \to 0 \). On the other hand, low-pass prefiltering rectifies this situation [20]. In fact, for the stable LTI system \( F \),

\[
\lim_{h \to 0} \|[(I - HS)Fu](t)| = 0, \quad F := \begin{bmatrix}
A_{F} & B_{F} \\
C_{F} & 0
\end{bmatrix}
\]

holds. For the evaluation of the cost function \( J_h(Q_d) \), then this paper utilizes

\[
u = HSFr, \quad r \in L_0^p
\]  

(6)

as the exogenous input. Note that \( u \in L_0^p \) if stable \( F \) is strictly proper and \( r \in L_0^p \). Along with this, \( \Sigma \) in (5) is rewritten as

\[
\begin{align*}
\Sigma' : & \quad \xi[k + 1] = A_{\xi}[k] + B_{\xi}[k] \\
& \quad + \int_{0}^{h} e^{A_{\xi}(h-\tau)} B_{\xi}(kh + \tau)d\tau \\
& \quad + C_{\xi} \int_{0}^{h} e^{A_{\xi}(h-\tau)} B_{\xi}(kh + \tau)d\tau \\
& \quad e_{\xi}(\theta) = \xi[k] + \mathcal{D}(\theta) e[k]
\end{align*}
\]  

(7)

Also, this paper solves the following synthesis problem (\( \mathcal{E}' \)):

For the system \( \Sigma_Q \) composed of \( P \) and \( Q_d \) with the initial state \( x_0 \in \mathbb{R}^m \) and the exogenous input \( u \) in (6), suppose that the quantization interval \( d \in \mathbb{R}_+ \), the switching speed \( h \in \mathbb{R}_+ \) and the performance level \( \gamma \in \mathbb{R}_+ \) are given. Characterize a continuous-time dynamic quantizer \( Q_d \) (i.e., find parameters \( (n_Q, A_Q, B_Q, C_Q) \)) achieving \( J_h(Q_d) \leq \gamma \).

B. QUANTIZER ANALYSIS

The quantization error \( e \) of (7) is bounded as mentioned earlier. The reachable set and the invariant set characterize such a system with a bounded input. Consider the LTI discrete-time system given by

\[
\hat{\xi}[k + 1] = A_{\hat{\xi}}[k] + B_{\hat{\xi}}w[k]
\]  

(8)

where \( \hat{\xi} \in \mathbb{R}^{m_{\xi}} \) and \( w \in \mathbb{R}^{m_w} \) denote the state vector and disturbance input, respectively. We define the reachable set and the invariant set.

Definition 1: Define the reachable set of the system (8) to be a set \( \mathbb{R}_\infty \) which satisfies

\[
\mathbb{R}_\infty := \left\{ \hat{\xi} \in \mathbb{R}^{m_{\xi}} : \exists k \in \mathbb{N}_+: \exists [w]\in \mathbb{W}, \right. \left. \hat{\xi}[k] = \sum_{i=0}^{k-1} A_{\hat{\xi}}^{k-1-i} B_{\hat{\xi}}w[i] \right\}, \quad \mathbb{W} := \{ w \in \mathbb{R}^{m_w} : w^T w \leq 1 \}.
\]

Definition 2: Define the invariant set of the system (8) to be a set \( \mathcal{X} \) which satisfies

\[
\hat{\xi} \in \mathcal{X}, \quad \forall [w] \in \mathbb{W} \Rightarrow A_{\hat{\xi}} + B_{\hat{\xi}} w \in \mathcal{X}.
\]

The analysis condition can be expressed in terms of matrix inequalities as summarized in the following proposition [17].

Proposition 1: Consider the system (8). For a matrix

\[
0 \prec P = P^T \in \mathbb{R}^{m_{\xi} \times m_{\xi}}, \quad \text{the ellipsoid } \mathcal{E}P := \{ \hat{\xi} \in \mathbb{R}^{m_{\xi}} : \hat{\xi}^T P \hat{\xi} \leq 1 \}
\]

is an invariant set if and only if there exists a scalar \( \alpha \in [0, 1 - \rho(A)^2] \) satisfying

\[
A_{\hat{\xi}}^T P A_{\hat{\xi}} - \left(1 - \alpha \right) P^T B_{\hat{\xi}}^T B_{\hat{\xi}} \leq 0.
\]  

(9)

Note that the ellipsoidal set \( \mathcal{E}P \) covers the reachable set \( \mathbb{R}_\infty \) from outside. Define the set \( \mathcal{E} := \{ \hat{w} \in \mathbb{R}^{m_w} : e = \frac{\sqrt{m_w}}{2} \hat{w} \text{satisfies (3)} \} \) and rewrite the system (7) as

\[
\hat{\Sigma}' : \left\{ \begin{array}{l}
\hat{\xi}[k + 1] = A_{\hat{\xi}}[k] + B_{\hat{\xi}}\hat{w}[k] \\
\hat{z}_p(kh + \theta) = \mathcal{C}(\hat{\theta})\hat{\xi}[k] + \mathcal{D}(\hat{\theta})\hat{w}[k]
\end{array} \right.
\]

where \( \hat{\xi} = \frac{\sqrt{m_w}}{2} \hat{\xi} \) and \( \hat{z}_p = \frac{\sqrt{m_w}}{2} \hat{z}_p \). The relation \( \mathcal{E} \subseteq \mathbb{W} \) clearly holds since \( e^T e \leq \frac{m_w}{4} \) and the set \( \mathbb{W} \) is an independent bounded disturbance without the relation (3).

That is, the reachable set of \( \hat{\Sigma}' \) with \( \hat{w} \in \mathcal{E} \) is no larger than that of \( \Sigma' \) with the disturbance \( \hat{w} \in \mathbb{W} \).

Then, this paper utilizes the reachable set to estimate the influences of the quantization error and the invariant set to characterize the cost function \( J_h(Q_d) \) by substituting \( A = A_{\xi} \) and \( B = B_{\xi} \) into (9). Move on to the matrix exponential \( e^{A_\theta} \) of \( \mathcal{C}(\theta) \) and \( \mathcal{D}(\theta) \) in (10), which is rewritten as

\[
e^{A_\theta} = I + \int_{0}^{\theta} e^{A_\theta \tau} d\tau.
\]  

(11)
Along with this, \( \tilde{z}_p \) of \((10)\) is also rewritten as
\[
\tilde{z}_p(kh + \theta) = (C + C\Omega(\theta)D)\tilde{x}[k] + \Omega(\theta)\tilde{w}[k],
\]
\[
C := \begin{bmatrix} C & 0 \end{bmatrix}, \quad D := \begin{bmatrix} A & BC_Q \end{bmatrix}.
\]

In addition, from the properties of \(R_\infty\) and \(E(P)\),
\[
J_h(Q_d) \leq \sup_{\xi \in \mathcal{P}} \sup_{\theta \in [0,h]} \frac{\| (C + C\Omega(\theta)D)\xi \| \sqrt{md}}{2} + \sup_{\omega \in \mathcal{W}} \sup_{\theta \in [0,h]} \frac{\| C\Omega(\theta)\omega \| \sqrt{md}}{2} \quad (12)
\]
holds. Similarly to the papers [9]-[11], by using the \(L_1\) control technique in [16], we provide the sufficient conditions of computing \(\gamma_1\) and \(\gamma_2\) of \((12)\) as follows:
\[
\left\{ \begin{array}{l}
C + C\Omega(\theta)D + D^T\Omega(\theta)^T C^T = 0, \\
\Omega(\theta)^T C D \Omega(\theta) \leq \gamma_2^2 I_n,
\end{array} \right. \quad (13)
\]
The inequalities \((13)\) are difficult to test since we need to find \(\mathcal{P}\), \(\gamma_1\) and \(\gamma_2\) satisfy \((9)\) and \((13)\) for infinitely many values of \(\theta \in [0, h]\). Then, we consider their sufficient conditions using the matrix uncertainty technique [18], [19], which are easy to compute. Considering \(\Omega(\theta)\) in \((11)\) as a matrix uncertainty, we introduce the following lemmas regarding the matrix exponential [21], [22].

**Lemma 2:** For the matrix \(\Omega(\theta)\) in \((11)\),
\[
\sigma_{\text{max}}(\Omega(\theta)) \leq \delta(\theta) \leq \delta(h), \quad \forall \theta \in [0, h),
\]
\[
\delta(\theta) := \begin{cases} e^{\mu(A)\theta} - 1 & \mu(A) \neq 0, \\
|\theta|, & \mu(A) = 0,
\end{cases}
\]
\[
\mu(A) := \max \{ \lambda : \lambda \in \text{eig}(A + A^*)/2 \}
\]
holds.

By using Lemma 2 and the S-procedure [18], [23], [24], the sufficient condition analyzing the cost function \(J_h(Q_d)\) of the system \(\Sigma_Q\) can be expressed in terms of matrix inequality as summarized in the following theorem.

**Theorem 1:** Consider the system \(\Sigma_Q\) composed of \(P\) and \(Q_d\) with the initial state \(x_0 \in \mathbb{R}^n\) and the exogenous input \(u\) in \((6)\). For the quantization interval \(d \in \mathcal{W}\) and the switching speed \(h \in \mathbb{R}_+\), the upper bound of the cost function \(J_h(Q_d)\) is given by
\[
J_h(Q_d) \leq (\gamma + \sigma_{\text{max}}(C)\delta(h)) \frac{\sqrt{md}}{2} \quad (14)
\]
if there exist \(0 < Q = Q^T \in \mathbb{R}^{(n_q + n_A) \times (n_q + n_A)}\), \(0 < S = S^T \in \mathbb{R}^{n \times n}\), \(0 < S = S^T \in \mathbb{R}^{n \times n}\), \(\alpha_h \in [0,1/h - \rho(A)^2/h]\) and \(\gamma \in \mathbb{R}_+\) satisfying
\[
\begin{bmatrix}
\Phi_h Q + Q \Phi_h^T + \alpha_h Q & \Gamma_h & \sqrt{h} \Phi_h^T \\
\Gamma_h^T & -\alpha_h I & \sqrt{h} \Gamma_h \\
\sqrt{h} \Phi_h Q & \sqrt{h} \Gamma_h & -Q
\end{bmatrix} \leq 0,
\]
\[
\begin{bmatrix}
Q & Q C Q^T & \gamma^2 I_n - \delta(h) C S C^T & 0 \\
C Q & \gamma Q & 0 & S \\
\sqrt{\delta(h)} D Q & 0 & S
\end{bmatrix} \geq 0,
\]
\[
\Phi_h := \begin{bmatrix} \frac{1}{h} \int_0^h e^{A^T} d\tau A & \frac{1}{h} \int_0^h e^{A^T} d\tau BC_Q \\
0 & \frac{1}{h} \int_0^h e^{A^T} d\tau (A_Q + B_Q C_Q) \end{bmatrix},
\]
\[
\Gamma_h := \begin{bmatrix} \frac{1}{h} \int_0^h e^{A^T} d\tau B \\
0 \frac{1}{h} \int_0^h e^{A^T} d\tau B_Q \end{bmatrix}.
\]

An advantage of the condition \((15)\) over the conditions \((9)\) is that it can be used for a small \(h\) without numerical difficulty. This idea is due to [18], [19]. In the limit of \(h \to 0\),
\[
\Phi_h \to \begin{bmatrix} A & BC_Q \\
0 & A_Q + B_Q C_Q \end{bmatrix}, \quad \Gamma_h \to \begin{bmatrix} B \\
0 \end{bmatrix}
\]
holds. These matrices equal the matrices \(A\) and \(B\) of the system \(\Sigma\) without the operator \(HS\). In the same limit, from \(\delta(h) \to 0\), the conditions \((15)\) and \((16)\) converge to the analysis conditions of the continuous-time dynamic quantizer without the operator \(HS\) in [10], [11]. On the other hand, for a small \(h\), \(A\) and \(B\) are close to identity and zero, respectively, and the left side of \((9)\) is close to zero.

In numerical computation, it is appropriate to fix the structure of \(S\) such that \(S\Omega(\theta) = \Omega(\theta)S\) holds. For example, we can set \(S = s_0 I_n\), \(s_0 \in \mathbb{R}_+\) and this setting leads to the following optimization problem \((\text{aop})\):
\[
\min_{Q, S = s_0 I_n, \alpha, \gamma} \gamma^2 \quad \text{s.t.} \quad (15) \text{ and } (16).
\]
When scalar \(\alpha_h\) is fixed, the conditions in Theorem 1 are linear matrix inequalities (LMIs) in terms of the other variables. Using standard LMI software and the line search of \(\alpha_h\), we can obtain an upper bound of \(J_h(Q_d)\).

**C. QUANTIZER SYNTHESIS**

The problem \((\text{aop})\) suggests that the quantizer synthesis problem \((\text{E})\) reduces to the following non-convex optimization problem \((\text{OP})\):
\[
\min_{Q, S = s_0 I_n, A_Q, B_Q, C_Q, \alpha, \gamma} \gamma^2 \quad \text{s.t.} \quad (15) \text{ and } (16).
\]
That is, if \((\text{OP})\) is feasible, \((\text{E})\) is feasible.

From the matrix product such as \(\int_0^h e^{A^T} d\tau\) and \(A_Q + B_Q C_Q\) in \((15)\), it is difficult to derive the synthesis condition from Theorem 1 unlike the continuous-time case without the operator \(HS\) in [10], [11]. Then, we fixed the parameters as follows:
\[
n_Q = n, \quad A_Q = A, \quad B_Q = B.
\]
\((17)\) does not impose a severe limitation on the synthesis because \(A_Q\) and \(B_Q\) of the continuous-time dynamic quantizer without the operator \(HS\) in [10], [11] are also \((17)\). In other words, \(\tilde{x}_Q = A x_Q + B e_Q\) estimates the quantization influence on the system \(P\). Along with this, we fix \(Q\) of \((15)\) as follows:
\[
Q = \begin{bmatrix} Y & V \\
Y & V \end{bmatrix}, \quad Y = Y^T > 0, \quad V = V^T > 0.
\]
\((18)\) also does not impose a severe limitation on the synthesis because an appropriate choice of the quantizer state coordinates allows us to assume that \(Q\) has the special structure for the full order case \(n_Q = n\) [25].
Under some circumstances (17) and (18), we obtain the following synthesis condition.

Theorem 2: Consider the system $\Sigma_Q$ composed of $P$ and $Q_d$ with the initial state $x_0 \in \mathbb{R}^n$ and the exogenous input $u$ in (6). Suppose that the quantization interval $d \in \mathbb{R}_+$, the switching speed $h \in \mathbb{R}_+$ and the performance level $\gamma \in \mathbb{R}_+$ are given. For a scalar $\alpha_h \in [0, 1/h]$, there exist a continuous-time dynamic quantizer $Q_d$ achieving (14) if one of the following equivalent statements holds.

(i) There exist matrices $0 < Q = Q^T \in \mathbb{R}^{(n+n_Q)(n+n_Q)}$, $0 < S = S^T \in \mathbb{R}^{n \times n}$ and a dynamic quantizer $Q_d$ satisfying (15) and (16).

(ii) There exist matrices $0 < Y = Y^T \in \mathbb{R}^{n \times n}$, $0 < V = V^T \in \mathbb{R}^{n \times n}$, $W \in \mathbb{R}^{m \times n}$, $0 < S = S^T \in \mathbb{R}^{n \times n}$ satisfying

\[
\begin{bmatrix}
\Theta_{Ah} + \Theta_{Ah}^T + \alpha_h \Theta_P & \Theta_{Bh} & \sqrt{h} \Theta_{Ah}^T \\
\sqrt{h} \Theta_{Ah} & -\alpha_h I & \sqrt{h} \Theta_{Bh} & -\Theta_P \\
0 & \Theta_{Dh} & \gamma^2 I - \delta(h) C S C^T \\
0 & S
\end{bmatrix} \leq 0,
\]

where

\[
\Theta_P := \begin{bmatrix} Y & V \\ V & V \end{bmatrix}, \quad \Psi_h := \frac{1}{h} \int_0^h e^{A \tau} d\tau,
\]

\[
\Theta_{Ah} := \begin{bmatrix} \Psi_h (AY + BW) & \Psi_h (AV + BW) \\ \Psi_h (AV + BW) & \Psi_h (AV + BW) \end{bmatrix},
\]

\[
\Theta_{Bh} := \begin{bmatrix} \Psi_h B \\ \Psi_h B \end{bmatrix}, \quad \Theta_C := \begin{bmatrix} CY & CV \end{bmatrix},
\]

\[
\Theta_{Dh} := \begin{bmatrix} \sqrt{\delta(h)} (AY + BW) & \sqrt{\delta(h)} (AV + BW) \end{bmatrix}.
\]

In this case, such a quantizer parameter is given by

\[
n_Q = n, \quad A_Q = A, \quad B_Q = B, \quad C_Q = WY^{-1}. \tag{21}
\]

In the limit of $h \to 0$, $\Psi_h$ converges to $I$, then conditions (19) and (20) also converge to the synthesis condition of the continuous-time dynamic quantizer without the operator $HS$. Also, by setting $S = s_a I_a$, for Theorem 2, the quantizer synthesis problem $(\text{Sop})'$ reduces to the following optimization problem $(\text{Sop})$:

\[
\min_{Y,V,S,s_a I_a, W, \alpha_h, \gamma} \gamma^2 \quad \text{s.t.} \ (19) \text{ and } (20).
\]

If $(\text{Sop})$ is feasible, $(\text{Sop})'$ is feasible.

IV. NUMERICAL EXAMPLES

For the slow switching, we make a comparison between the proposed method and the existing continuous-time quantizer in [10], [11]. Consider the system $\Sigma_Q$. The plant $P$ is the stable minimum phase LTI system:

\[
\begin{bmatrix}
\dot{x}(t) \\
\dot{z}(t)
\end{bmatrix} = \begin{bmatrix}
-3 & 3 & 0 \\
0 & -2 & 2 \\
1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
x(t) \\
v(t)
\end{bmatrix}.
\]

In the case without the operator $HS$, an optimal form of the continuous-time quantizer in [10], [11] is given by

\[
Q_d^{op} : \begin{cases}
\dot{x}_Q = A x_Q + B(v - u) \\
v = q(-CB)^{-1}C(A + f I)x_Q + u
\end{cases} \tag{22}
\]

where its achievable performance of $J(Q_d^{op})$ is

\[
\inf \gamma_c = \frac{d \sqrt{m}}{4 \sqrt{\rho(f - \rho)}} \sigma_{\max}(CB),
\]

\[
\rho = \max \{\mu(A), \mu(A - B(CB)^{-1}C(A + f I))\}.
\]

The continuous-time quantizer $Q_d^{op}$ and its performance are parameterized by the free parameter $f \in \mathbb{R}_+$.

For the comparison, we set the switching speed $h = 0.01$ [s] and the quantization interval $d = 2$. The initial state $x_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ and the input $u(t) = \sin(\pi t) + \cos(0.7\pi t)$ are given. First, we set $f = 50$ and obtain $Q_d^{op}$ with $\gamma_c = 0.065$. Second, we solve the problem $(\text{Sop})$ and obtain $\gamma = 0.707$ and the matrix $C_Q = \begin{bmatrix} -7.43 & -28.32 \end{bmatrix}$. In this case, both of the quantizers can achieve good approximation performance. Fig. 4 illustrates the time responses of $\Sigma_Q$ with the two quantizers $Q_d^{op}$ in (22). In Fig. 4, the thin line and the thick line are for the the usual system in Fig. 1 (b) and the system $\Sigma_Q$ in Fig. 1 (a), respectively. We see that the controlled output $z(t)$ of Fig. 1 (a) approximates that of Fig. 1 (b) (rather, the two outputs are exactly similar) even if the quantizer output $v \in \{-2, 0, 2\}$ is applied.

![Fig. 4. Time responses of $\Sigma_Q$ with (22) for $h = 0.01$.](image)

Next, we consider the case $h = 0.1$. From $(\text{Sop})$, we get $\gamma = 2.790$ and the matrix $C_Q = \begin{bmatrix} -0.4143 & -2.5730 \end{bmatrix}$. Figs. 5 and 6 illustrate the time responses of $\Sigma_Q$ with (22) and the proposed quantizer, respectively. We see that $z(t)$ of the usual plant $P$ is approximated by $z(t)$ of the system $\Sigma_Q$ with the proposed quantizer, while $z(t)$ of the system $\Sigma_Q$ with (22) diverges. From this example, we see that the proposed method can address the spatial resolution and the temporal resolution issues, simultaneously. On the other hand, by setting $f = 5$, the corresponding (22) achieves approximation performance equivalent to the proposed method. For the slow switching case, the existing continuous-quantizer in [10], [11] may be suitable for a two step design such that the parameter $f$ is tuned in terms of the switching speed constraint.
V. CONCLUSION

Focusing on the broadbandization and the robustness of the networked control systems, this paper has dealt with the continuous-time quantized control. We have proposed numerical optimization methods analyzing and synthesizing the continuous-time dynamic quantizer based on the invariant set analysis and the sampled-data control technique. As a result, both of the temporal and spatial resolution constraints can be considered, simultaneously. Finally, it has been pointed out that the proposed method is helpful through numerical examples. Considering the quantized feedback control system with unstable plants and generalizing the exogenous signal for the evaluation of the cost function are future topics.

REFERENCES


http://www.control.utoronto.ca/~francis/