Dynamic Quantizer Design Under Communication Rate Constraints

journal or publication title	IEEE Transactions on Automatic Control
volume	61
number	10
page range	3190-3196
year	2016-10
URL	http://hdl.handle.net/2298/00045878

doi: 10.1109/TAC.2015.2509438

Dynamic Quantizer Design under Communication Rate Constraints

Hiroshi Okajima , Kenji Sawada and Nobutomo Matsunaga

***Abstract*—Feedback type dynamic quantizers such as delta-sigma modulators are typically effective for encoding high-resolution data into lower resolution data. The dynamic quantizers include a filter and a static quantizer. When it is required to control under a communication rate constraint, the data rate of the quantizer output should be minimized appropriately by quantization. This technical note provides numerical methods for the complete design of a type of dynamic quantizers, including the selection of all the quantizer parameters in order to minimize a specific performance index and satisfy a communication constraint. The design method of the dynamic quantizer is proposed using a particle swarm optimization (PSO) method. A part of the initial quantizers in PSO are designed based on an invariant set analysis and an iteration algorithm. Effectiveness of the system with the proposed quantizer is assessed through numerical examples.**

***Index Terms*—Quantizer Design, Communication Rate Constraint, Networked Control Systems.**

I. Introduction

The analysis and synthesis of the networked control systems (NCS) have recently attracted significant attention [1]-[3]. Because the data rate is limited in communication channels, overcoming performance degradation is a crucial topic in NCS. To use the communication channel, the control signals should be compressed using quantizer because of the limited communication rate [4]-[8]. There exist performance degradations caused by the quantization because plants are controlled by compressed (quantized) signals. Therefore, the type of quantization used to achieve good performance in NCS needs to be considered carefully. Feedback type dynamic quantizers have proven effectiveness for overcoming performance degradation [9]-[15]. Consisting of a filter and a static quantizer, they utilize previous quantization error information to generate quantizer output. Such quantization methods are widely used in signal processing [9], [10] such as in AD/DA converters, data compressor for music audio signals, and switched-mode power supplies. In recent years, feedback type dynamic quantization methods have been exploited to a large extent in control engineering [11]-[15]. Performance degradation decreases if an appropriate filter is chosen in the dynamic quantizer. A dynamic quantizer designed based on ℓ_∞ optimization has been proposed for stable minimum-phase plants [11]. This quantizer was expressed analytically as a function of plant parameters. Its static quantizer component was assumed to be given and the quantization interval was fixed. Moreover, feedback control [12], the non minimum-phase plants [13] and non-linear systems [14] have minimized performance degradation by quantization with high efficiency.

When we want to use the dynamic quantizers under communication rate constraint, the output level number in dynamic quantizers should be explicit. However, this number has not been analyzed explicitly in the past. If the channel data rate is given as M bits per sampling, this number must be smaller than or equal to 2^M. The quantization interval in the static quantizer component is closely linked to this number. Therefore, filter and static quantizer components both need to be accounted for in dynamic quantizer design.

This study aims to design a quantizer that satisfies communication rate constraints. First, assuming that the filter parameters are given and an analysis method is proposed for the design of an optimal quantization interval that satisfies communication rate constraints in Section III. The design of the quantization interval is reduced to the ℓ_1 optimization problem. Then, the control performance can be analyzed explicitly for a given filter parameters.

In Section IV, a design method is formulated as a numerical optimization problem for filter and quantization interval design using a particle swarm optimization (PSO) algorithm with the evaluation function of Section III. Initial quantizers and its velocity terms in the PSO algorithm are designed using the design method in [16], which is composed of an invariant set analysis and an iteration algorithm. The effectiveness of the method is assessed through numerical examples in Section V.

Note that this technical note is based on our preliminary version [23], published in the conference proceedings. This technical note uses an iterative algorithm to give the appropriate initial quantizers in the PSO algorithm, contains full explanations and adds numerical simulation.

In the remainder of the manuscript, a set of $n \times m$ real matrices is denoted as $\mathcal{R}^{n\times m}$. $\mathcal{R}_+$ is the set of positive real numbers and I is the identity matrix. For a matrix H, H^T and $\rho(H)$ correspond its transpose and spectral radius, respectively. For a vector $X = \{x_1, x_2, \cdots, x_k, \cdots\}$, $\|X\|$ represents the infinity norm. Consequently, $\|X\| = \sup_k \|x_k\|$ holds.

II. Problem Formulation

A. Control systems with a communication channel

A single input single output (SISO) discrete-time plant P is defined as

$$P\colon \begin{cases} x_p(k+1) & = A_p x_p(k) + B_p u_p(k), \\ y_p(k) & = C_p x_p(k), \end{cases} \tag{1}$$

where $x_p \in \mathcal{R}^{n_p\times 1}$ is the state, $u_p \in \mathcal{R}$ is the control input, $y_p \in \mathcal{R}$ is the control output. $A_p \in \mathcal{R}^{n_p\times n_p}$, $B_p \in \mathcal{R}^{n_p\times 1}$ and $C_p \in \mathcal{R}^{1\times n_p}$ are constant matrices, and $x_p(0)$ is the initial state. Plant P is assumed stable.

Fig. 1 shows the structure of a control system equipped with a communication channel, in which u is an outer signal and y is the plant output of (1), respectively. Signal u may be regarded as an operating signal or command, such as a telesurgery operation. The quantizer Q transforms the high-resolution outer signal u into a lower resolution signal v and ENC encodes this rounded signal. The encoded signal passes through the communication channel before undergoing decoding in DEC. No delay or loss in precision is assumed to occur. Therefore, $u_p = v$ is the control input for P in this system.

The number of quantization levels N depends on the communication rate of the channel. When M [bits] of data are transmitted through the channel over a sampling period, N should satisfy the following inequality.

$$N \leq 2^M \tag{2}$$

Although, N is assumed to be even in this study, the same discussion

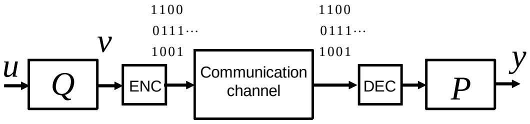


Fig. 1. Feedforward control system with quantizer Q

H. Okajima is with the Graduate School of Science and Technology, Kumamoto University, Kurokami 2-39-1, Kumamoto, Japan e-mail: okajima@cs.kumamoto-u.ac.jp, Phone & Fax: 81-96-342-3603.

K. Sawada is with The University of Electro-Communications.

N. Matsunaga is with Kumamoto University.

also applies to odd N values. In particular, odd number is more suitable for the case that v should become zero when u equals zero.

The outer signal u is constrained by upper and lower boundaries, giving the signal range $U = [u_{\min}, u_{\max}]$. Hence, the outer signal u is assumed to satisfy the following relation.

$$u(k) \in U,\ \forall k \tag{3}$$

B. Dynamic quantizer form

The feedback-type dynamic quantizer Q is defined as:

$$Q : \begin{cases} \xi(k+1) & = \mathcal{A}\xi(k) - \mathcal{B}u(k) + \mathcal{B}v(k), \\ v(k) & = Q_{st}\left[\mathcal{C}\xi(k) + u(k)\right], \end{cases} \tag{4}$$

where Q_{st} is the mid-riser type uniform static quantizer with saturation for an even permissible number of quantization levels N. Fig. 2 shows an example of Q_{st} (Solid line, $N = 4$). $\mathcal{A} \in \mathcal{R}^{n_q \times n_q}$, $\mathcal{B} \in \mathcal{R}^{n_q \times 1}$ and $\mathcal{C} \in \mathcal{R}^{1 \times n_q}$ are constant matrices. The initial state is given as $\xi(0) = 0$. The quantizer output $v(k)$ is obtained by static quantization of $C\xi + u$. Q_{st} is defined using a quantization interval $d \in \mathcal{R}_+$ and a center point $c \in \mathcal{R}$. Its level interval is the same for input and output axes (Fig. 2).

C. Design problem of dynamic quantizer based on error system

Fig. 3 shows a quantizer performance evaluation system based on an error signal. The desired output $y_r(k)$ is an output of P using $u(k)$ as the input signal. Signal $v(k)$, which is quantized by Q, is applied to P in the control system involving the communication channel, resulting in output $y(k)$, which differs from $y_r(k)$. The error signal $e(k) = y(k) - y_r(k)$ needs to be minimized using the appropriate parameter set $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, d\}$ so that y approximate y_r. The quantizer is designed based on the following performance index $E(Q)$ defined as

$$E(Q) = \sup_{u(k) \in U} \|\mathcal{Y} - \mathcal{Y}_r\|, \tag{5}$$

where $\mathcal{Y} = \{y(1), y(2), \cdots\}$ and $\mathcal{Y}_r = \{y_r(1), y_r(2), \cdots\}$ are the output time series. Because $E(Q)$ produces the maximum value for $e(k)$, y expected to be similar to y_r if $E(Q)$ is small. In existing dynamic quantizer designs [11]-[14], $E(Q)$ is used as a performance index for these quantizers.

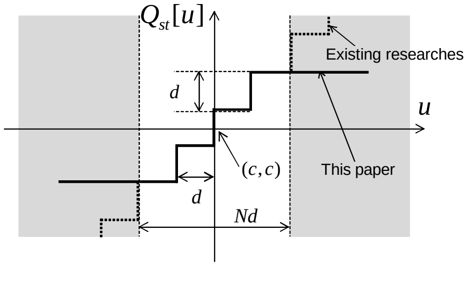


Fig. 2. Comparison of a mid-riser type uniform quantizer Q_{st} (Solid line) with existing researches (Dashed line)

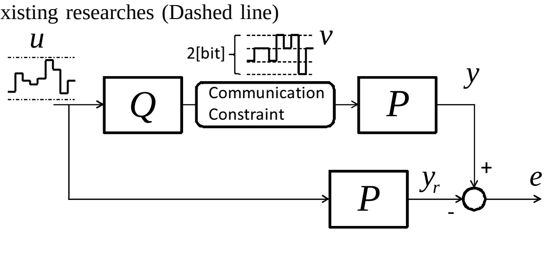


Fig. 3. Error system for quantizer performance evaluation

III. Quantizer Analysis

A. Minimum quantization interval for given dynamic quantizers

In this section, assuming that $\{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$ are known, an analytical method is developed for communication rate constraints along with a derivation of the minimum quantization interval.

The relationship between the number of quantization levels and the quantization interval is evaluated first. The quantization error of $Q_{st}(u)$ is written as

$$\epsilon = Q_{st}(u) - u. \tag{6}$$

Then the following inequality holds for $u(k)$ which satisfies $u(k) \in [c - Nd/2, c + Nd/2]$.

$$|\epsilon(k)| \leq \frac{d}{2} \tag{7}$$

On the other hand, when $u(k)$ is out of range (Fig. 2, gray area), $|\epsilon(k)| > d/2$ holds by the saturation of the quantizer. Therefore, c and d need tuning to satisfy (7) for all $u(k) \in U$. For dynamic quantizers, the input signal for Q_{st} is written as $\bar{u} = u + \mathcal{C}\xi$ to avoid confusion. The inequality for the range constraint on a the given communication rate in the network channel is defined as

$$Nd \geq \bar{u}_{\max} - \bar{u}_{\min}, \tag{8}$$

and the range of $\bar{u}(k)$ is expressed as $\bar{U}$. Because the signal $u(k)$ always takes any value in U, $\bar{U}$ can be characterized using U and a range of $\phi = \mathcal{C}\xi$. The range U_ϕ is denoted as $U_\phi = [\phi_{\min}, \phi_{\max}]$. The relationship between $\bar{U}$ and the range of ϕ is summarized as

$$\bar{U} = [u_{\min} + \phi_{\min},\ u_{\max} + \phi_{\max}]. \tag{9}$$

Therefore, U_ϕ depends on the given filter parameters $\{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$.

If we can find the range of $\bar{u}$, d can be decided using $Nd \geq \bar{u}_{\max} - \bar{u}_{\min}$. We denote a set $[\phi_{\min}^{opt},\ \phi_{\max}^{opt}]$ which minimizes $\phi_{\max} - \phi_{\min}$. Then, it is equivalent to find the set $[\phi_{\min}^{opt}, \phi_{\max}^{opt}]$ and to find d^{opt} using

$$Nd^{opt} = \bar{u}_{\max} - \bar{u}_{\min}. \tag{10}$$

Therefore, a solution leading to $[\phi_{\min}^{opt}, \phi_{\max}^{opt}]$ is proposed.

We define a signal $w(k)$ as follow.

$$w(k) := \frac{2}{d}(Q_{st}[\mathcal{C}\xi(k) + u(k)] - \mathcal{C}\xi(k) - u(k)) \tag{11}$$

The definition of Q_{st} indicates that $\|w\| \leq 1$. Using (11), the state equation of ϕ is written as follows:

$$\begin{aligned} \xi(k+1) &= (\mathcal{A} + \mathcal{B}\mathcal{C})\xi(k) + \frac{d}{2}\mathcal{B}w(k), & (12) \\ \phi(k) &= \mathcal{C}\xi(k),\ ||w|| \leq 1. & (13) \end{aligned}$$

Moreover, using the coordinate transformation $\xi = (d/2)\tilde{\xi}$, the state equation becomes

$$\begin{aligned} \tilde{\xi}(k+1) &= (\mathcal{A} + \mathcal{B}\mathcal{C})\tilde{\xi}(k) + \mathcal{B}w(k), & (14) \\ \tilde{\psi}(k) &= \mathcal{C}\tilde{\xi}(k),\ ||w|| \leq 1, & (15) \end{aligned}$$

where $\tilde{\psi} := 2\phi/d$. Note that (14) and (15) are independent of d. To find $U_{\tilde{\psi}} := [\tilde{\psi}_{\min}, \tilde{\psi}_{\max}]$ is equivalent to find U_ϕ. For convenience, $\tilde{\psi}_{\max}$ is denoted as ψ. Then, we obtain $\tilde{\psi}_{\min} = -\psi$ because of the solution symmetry. The problem to find $U_{\tilde{\psi}}$ is written as follow:

Problem 1: Considering (14), (15) and $\tilde{\xi}(0) = 0$, find ψ value that satisfies the inequality condition

$$-\psi \leq \mathcal{C}\tilde{\xi}(k) \leq \psi,\ \forall \tilde{\xi}(k) \in \Xi, \tag{16}$$

where Ξ is the reachable set of $\tilde{\xi}$ for w. ■

Moreover, by using the optimal solution of minimization problem of ψ under the condition given in *Problem* 1, the minimum quantization interval d^{opt} is obtained by the following theorem:

Theorem 1: Using optimal solution ψ^{opt}, the minimum quantization interval d^{opt} and c^{opt} are expressed as

$$d^{opt} = \frac{u_{\max} - u_{\min}}{N - \psi^{opt}}, \;\; c^{opt} = \frac{u_{\max} + u_{\min}}{2}. \tag{17}$$

If $N - \psi^{opt} \leq 0$, no d satisfies the condition regarding the permissible number of quantization levels. ■

Theorem 1 provides a relationship between d^{opt} and ψ^{opt} which is derived from (10). ψ^{opt}, which characterizes the signal amplitude caused by dynamic quantization, is obtained using matrices $\{\mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$. If ψ^{opt} is small, d^{opt} is also set small. On the contrary, a large d^{opt} is chosen when ψ^{opt} is large. In particular, when $N - \psi^{opt} \leq 0$, $\{\mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$ should be redesigned to satisfy the communication rate constraint. The amplitude of ψ^{opt} is regarded as an index of the usability of the dynamic quantizer for signal communication, providing valuable information for the construction of the networked control system.

B. Estimation of small ψ^{opt}

It is required to solve ψ^{opt} to satisfy the communication rate constraint. Our previous studies focus on the fact that the reachable set is covered by the invariant set from outside [17] to estimate ψ. If the invariant set clipped by $C\tilde{\xi}$ and $-C\tilde{\xi}$ is minimized, the real value of ψ is minimized indirectly. The detailed derivation sequence is shown in Appendix A. In case of Appendix A, we obtain the upper bound of ψ which leads to estimate $U_{\tilde{\psi}}$ from outside. The estimated value of ψ^{opt} is conservative in our previous researches.

In contrast, the maximum value ψ corresponds to the ℓ_1-norm of the impulse response of the linear time invariant system in (14) and (15). The value of ψ^{opt} can be estimated from outside by the following sequence.

At first, a positive integer value L is selected and calculate the following term about $\psi^{opt,L}$.

$$\psi^{opt,L} = \sum_{i=0}^{L} \left| \mathcal{C}(\mathcal{A}+\mathcal{B}\mathcal{C})^i \mathcal{B} \right| \tag{18}$$

Secondly, the value $\psi^{*,L}$ is derived by using a controllability pair $(\mathcal{A}+\mathcal{B}\mathcal{C}, (\mathcal{A}+\mathcal{B}\mathcal{C})^{L+1}\mathcal{B})$ with LMI problem in Appendix A. Then, the following inequality holds for any given L.

$$\psi^{opt,L} < \psi^{opt} \leq \psi^{opt,L} + \psi^{*,L} \tag{19}$$

When a large value L is selected, $\psi^{*,L}$ is close to zero. The value $\psi^{opt,L} + \psi^{*,L}$ can be used as an estimated value of ψ^{opt} that satisfies the communication rate constraint.

By using the result in [11] and above result, the evaluation value (5) can be given by the following remark.

Remark 1: The filter parameters $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ are given. The following equation holds with $L \to \infty$.

$$E(Q) = \left(\sum_{i=0}^{\infty} \left| \bar{C}\bar{A}^i \bar{B} \right| \right) \frac{d^{opt,L}}{2} \tag{20}$$

$$d^{opt,L} = \frac{u_{\max} - u_{\min}}{N - (\psi^{opt,L} + \psi^{*,L})} \tag{21}$$

Matrices $\bar{A}$, $\bar{B}$, $\bar{C}$ are given as follow:

$$\bar{A} = \begin{bmatrix} A_p & B_p\mathcal{C} \\ 0 & \mathcal{A}+\mathcal{B}\mathcal{C} \end{bmatrix}, \; \bar{B} = \begin{bmatrix} B_p \\ \mathcal{B} \end{bmatrix}, \bar{C} = \begin{bmatrix} C_p & 0 \end{bmatrix}$$

■
Therefore, in case the filter parameters $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ are known, the quantization interval in Theorem 1 is optimized by solving an ℓ_1 optimization problem. This remark is one of the contribution of this technical note.

IV. Design of Dynamic Quantizer under Communication Rate Constraint

In this section, dynamic quantizers are designed using two-step design method. Iterative design method based on an invariant set analysis [22], [16] and the particle swarm optimization method (PSO) [18] are used together to obtain quantizer which minimize (20). The PSO is a kind of the optimization method based on the swarm behavior. It requires many particles which represent the candidate of the quantizer parameters. We denote design parameter positions in the PSO as $p_i = \{\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i\}$ and parameter velocities, which are used in the PSO algorithm, as $\Delta p_i = \{\Delta\mathcal{A}_i, \Delta\mathcal{B}_i, \Delta\mathcal{C}_i\}$, respectively. The number of particles in the PSO algorithm is determined as m. In standard PSO design, initial particles and velocities are given randomly. In contrast, the quantizer parameters and its velocities would be designed using iterative design method at first step in the proposed design method. The obtained quantizers by the iterative method are used as a part of the initial quantizers in the PSO algorithm. It is expected that the dynamic quantizers, which achieve good performance, are obtained by the two-step design method.

A. An iterative algorithm of dynamic quantizer design based on invariant set analysis

As the first step of the design algorithm, iterative method using an invariant set analysis [17] is presented in this section. *Problems 3* and *4* in appendices provide inequality conditions related to performance index and communication rate constraint, respectively.

By combining *Problems 3* and *4*, the design problem using inequality conditions is addressed as follows.

Problem 2: Find the following Γ^*.

$$\Gamma^* = \min_{\mathcal{A},\mathcal{B},\mathcal{C},Z_p>0,Z_d>0,\alpha,\beta,\gamma^2,\psi^2} \Gamma(\gamma,\psi) \tag{22}$$

$$\Gamma(\gamma,\psi) := \gamma \frac{u_{\max} - u_{\min}}{2(N-\psi)} \tag{23}$$

subject to

$$\begin{bmatrix} Z_d & \mathcal{C}^T \\ \mathcal{C} & \psi^2 \end{bmatrix} \geq 0, \; \begin{bmatrix} Z_p & \bar{C}^T \\ \bar{C} & \gamma^2 \end{bmatrix} \geq 0,$$

$$\begin{bmatrix} (1-\beta)Z_d & 0 & (\mathcal{A}+\mathcal{B}\mathcal{C})^T Z_d \\ 0 & \beta I & \mathcal{B}^T Z_d \\ Z_d(\mathcal{A}+\mathcal{B}\mathcal{C}) & Z_d\mathcal{B} & Z_d \end{bmatrix} \geq 0,$$

$$\begin{bmatrix} (1-\alpha)Z_p & 0 & \bar{A}^T Z_p \\ 0 & \alpha I & \bar{B}^T Z_p \\ Z_p\bar{A} & Z_p\bar{B} & Z_p \end{bmatrix} \geq 0$$

$$\alpha \in \left[0, 1-\rho(\bar{A})^2\right], \; \beta \in \left[0, 1-\rho(\mathcal{A}+\mathcal{B}\mathcal{C})^2\right]$$

Γ is minimized using the inequality constraints of *Problems 3* and *4*. If a set $\{\mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$ are obtained by solving *Problem* 2, $E(Q) \leq \Gamma^*$ holds. Therefore, the obtained quantizer is expected to exhibit good control performance under communication rate constraints if we can obtain small Γ.

The multiple variables in inequality conditions and the nonlinear evaluation function (23) make *Problem 2* difficult to solve numerically. A design algorithm using *Problems 2, 3* and *4* has been developed to obtain appropriate numerical solutions [16].

In *Problem 2*, what is now needed is to find variables $\{\gamma,\, \psi,\, \alpha,\, \beta\}$ and matrix variables $\{Z_p,\, Z_d,\, \mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$. This leads to solve a non-convex problem in the sense that the constraints are bilinear matrix inequalities of $\{Z_p,\, Z_d,\, \mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$ and evaluation function Γ is a nonlinear function of $\{\gamma, \psi\}$. Note the fact that $\{Z_p,\, Z_d\}$ or $\{\alpha,\, \beta,\, \mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$ are fixed, the constraints becomes convex. Then, if Γ can be substituted by another linear function J, *Problem 2* can be solved by an iteration design algorithm which successively minimizes

J over variables while fixing the other variables in terms of the LMI optimization. Of course, the selection of the substituted function J is important. Consider the following substituted function form

$$J = a\gamma^2 + b\psi^2 + g, \tag{24}$$

where a, b and g are coefficients. J is a linear function of $\{\gamma^2,\, \psi^2\}$ and useful for solving the standard inner point method because $\{\gamma^2,\, \psi^2\}$ appear linearly in the constraints. This paper selects coefficient values a and b appropriately and then reduces J to a substituted function of Γ. The key idea is introduced by the iteration design algorithm proposed in [16]. In [16], appropriate $\{a,\, b,\, g\}$ are provided such that $\Gamma(\gamma, \psi) = J(\gamma, \psi)$ holds in the neighborhood of a certain set of $\{\gamma,\, \psi\}$. In this paper, only a and b are given because g does not affect the optimization problems. Denote $\{\gamma, \psi\}$ obtained from the k-th step of the iteration design algorithm by $\{\gamma_k, \psi_k\}$. Consider the linear approximation of $J(\gamma, \psi)$ and $\Gamma(\gamma, \psi)$ in the neighborhood of $\{\gamma_k,\, \psi_k\}$. The former is given by

$$\tilde{J} = 2a\gamma_k(\gamma - \gamma_k) + 2b\psi_k(\psi - \psi_k) + J(\gamma_k, \psi_k), \tag{25}$$

and the latter is given by

$$\begin{aligned}\widetilde{\Gamma} \;=\; & \frac{u_{\max} - u_{\min}}{2(N - \psi_k)}(\gamma - \gamma_k) + \gamma_k \frac{u_{\max} - u_{\min}}{2(N - \psi_k)^2}(\psi - \psi_k) \\ & + \Gamma(\gamma_k, \psi_k). \end{aligned} \tag{26}$$

A comparison between (25) and (26) gives coefficients a and b as follows:

$$a = \frac{u_{\max} - u_{\min}}{4\gamma_k(N - \psi_k)},\; b = \gamma_k \frac{u_{\max} - u_{\min}}{4\psi_k(N - \psi_k)^2}. \tag{27}$$

Therefore, for the k-step of the iteration design algorithm, (27) provides $\Gamma(\gamma_k, \psi_k) = J(\gamma_k, \psi_k)$. In other words, by updating a and b of $J(\gamma, \psi)$ based on (27), $J(\gamma, \psi)$ is applicable to the substituted function of $\Gamma(\gamma, \psi)$. Next, we consider the update steps of the iteration design algorithm. In the algorithm, $\{Z_p,\, Z_d\}$ are fixed to obtain appropriate $\{\mathcal{A},\, \mathcal{B},\, \mathcal{C},\, d\}$ in *Problem 2*. On the other hand, $\{Z_p,\, Z_d\}$ can be updated by *Problems* 4 and 3, respectively.

It is expected to obtain small Γ and good quantizer parameters by iteration design algorithm. Then, the extended design algorithm based on [16] is expressed as follows.

Iteration design algorithm

- **Step1-0**: Initial quantizer parameters $\mathcal{A}_0, \mathcal{B}_0$ and $\mathcal{C}_0$ are given and d_0 is obtained by solving *Problem* 3. Moreover, $Z_{p,0}$, $Z_{d,0}$, γ_0 and ψ_0 are determined through *Problems* 4 and 3.
- **Step1-1**: For fixed $Z_{d,k}$ and $Z_{p,k}$, $\mathcal{A}_{k+1}$, $\mathcal{B}_{k+1}$ and $\mathcal{C}_{k+1}$ are obtained by solving *Problem* 2. The coefficients a and b of function J are defined in (27).
- **Step1-2**: For fixed $\mathcal{A}_{k+1}, \mathcal{B}_{k+1}$ and $\mathcal{C}_{k+1}$, $Z_{p,k+1}$ and $Z_{d,k+1}$ are obtained by solving *Problems* 4 and 3, respectively. γ and ψ of the solution are set as γ_{k+1} and ψ_{k+1}, respectively.
- **Step1-3**: Γ_{k+1} and Γ_k are compared. If a ratio Γ_k/Γ_{k+1} exceeds $1+\Delta$, $\Delta > 0$, this algorithm is repeated from Step1-1. Otherwise, the process ends and $\mathcal{A}_{k+1}, \mathcal{B}_{k+1}$ and $\mathcal{C}_{k+1}$ describe the obtained dynamic quantizer.
- **Step1-4**: Obtain $\psi^{*,L} + \psi^{opt,L}$ by calculating with $\mathcal{A}_{k+1}, \mathcal{B}_{k+1}, \mathcal{C}_{k+1}$. Then, the quantization interval $d^{opt,L}$, which satisfy the communication rate constraint, is derived using $\psi^{*,L} + \psi^{opt,L}$. Parameter velocity Δp is given by $\Delta p = \{\mathcal{A}_{k+1} - \mathcal{A}_k,\, \mathcal{B}_{k+1} - \mathcal{B}_k,\, \mathcal{C}_{k+1} - \mathcal{C}_k\}$.

The initial quantizer parameters are chosen to satisfy stability condition of the dynamic quantizer [11]. For example, using the existing results such as [11] is one method to give an initial quantizer for the iteration design algorithm. Then, d_0 is obtained by solving *Problem* 3 with these parameters. This design algorithm performs adequately when *Problem* 3 is solvable for the initial quantizer.

When $N - \psi_0 < 0$ that is, *Problem 3* is unsolvable for the initial quantizer $\{\mathcal{A}_0,\, \mathcal{B}_0,\, \mathcal{C}_0\}$, this initial quantizer is modified. For example, $\{\mathcal{A}_0 + (h-1)/h\mathcal{B}_0\mathcal{C}_0,\, \mathcal{B}_0,\, \mathcal{C}_0/h\}$ where $h > N/\psi_0$ is a dynamic quantizer. For appropriate h, inequality $N - \psi_0' > 0$ because $\psi_0' = \psi_0/h$. Therefore, the initial quantizer $\{\mathcal{A}_0 + (h-1)/h\mathcal{B}_0\mathcal{C}_0,\, \mathcal{B}_0,\, \mathcal{C}_0/h\}$ is solvable.

In Step1-1, $\mathcal{A}_{k+1}$, $\mathcal{B}_{k+1}$ and $\mathcal{C}_{k+1}$ are updated and the resulting J value is smaller than that obtained for $\mathcal{A}_k$, $\mathcal{B}_k$, $\mathcal{C}_k$. Therefore, Γ is also expected to be small in this case. When the quantizer parameters $\mathcal{A}_{k+1}, \mathcal{B}_{k+1}, \mathcal{C}_{k+1}$ are fixed in Step1-2, *Problems 3*and *4* are regarded as LMI optimization problems. γ and ψ are minimized by solving *Problem 3* and *4*, respectively. Γ does not increase in Step1-2 because Γ is a monotonically increasing function for γ and ψ.

In Step1-3, the performance of the designed quantizers is compared. This assessment is conducted using a parameter Δ, $0 < \Delta \ll 1$. The number of updates increases with decreasing Δ values.

This design algorithm is an iterative process that is expected to generate a quantizer $\{\mathcal{A},\, \mathcal{B},\, \mathcal{C}\}$, satisfying communication rate constraints and exhibiting good performance. At least the obtained quantizer by this iterative algorithm is better than that in [16] because $d^{opt,L}(< d^*)$ is used in the proposed quantizer.

Not only p but also Δp is obtained by using the proposed iterative design algorithm.

B. Design of dynamic quantizer using particle swarm optimization

In this section, a concrete design procedure to determine the design parameters p, which minimize (20), is presented. As a part of initial candidate solution, the iterative design algorithm in section IV-A is used.

At first, we describe the conventional particle swarm optimization algorithm which is a kind of the optimization method based on the swarm behavior [18], [19]. The following minimization problem is considered in this section.

$$\min_{p \in \mathcal{R}^n} E(p) \tag{28}$$

$$\textbf{subject to}\, \psi^{opt,L}(p) < N \tag{29}$$

where, $E : \mathcal{R}^n \to \mathcal{R}$ is the objective function and $p = \{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$ is the design variable vector. $\psi^{opt,L}(p) < N$ denote the communication rate constraint. $\psi^{opt,L}(p)$ can be calculated by (18). The optimal solution p^{opt} for (28), (29) is required to obtain from an optimization algorithm. Patricle swarm optimization (PSO) is a computation method for optimizing a problem by iteratively trying to improve a solution. Multiple particles $p_1, \cdots, p_m$ are used in the PSO algorithm where m denotes the number of particles. To solve (28) by the PSO algorithm, the following objective function $E_f(p)$ is assumed to be given.

$$E_f = \begin{cases} E(p) & (\psi^{opt,L}(p) < N) \\ E_{pen} + E(p) & (\textbf{otherwise}) \end{cases} \tag{30}$$

The penalty E_{pen} is the larger positive value compared to a value of $E(x)$ which is an acceptable solution. The position and the velocity of i-th particle are denoted as $p_i = \{\mathcal{A}_i, \mathcal{B}_i, \mathcal{C}_i\}$ and $\Delta p_i = \{\Delta\mathcal{A}_i, \Delta\mathcal{B}_i, \Delta\mathcal{C}_i\}$, respectively. p_i is updated based on the following update laws.

$$p_i^{t+1} \;=\; p_i^t + \Delta p_i^{t+1} \tag{31}$$

$$\begin{aligned}\Delta p_i^{t+1} \;=\; & \omega_0 \Delta p_i^t + \omega_1 \mathrm{rand}_{1,i}^t (p_{pbest,i}^t - p_i^t) \\ & + \omega_2 \mathrm{rand}_{2,i}^t (p_{gbest}^t - p_i^t) \end{aligned} \tag{32}$$

t denote the iteration number and its initial value is t = 0. ω_0, ω_1 and ω_2 are the weighting coefficients which are given as positive values by the designer. The random numbers $\mathrm{rand}_{1,i}^t$ and $\mathrm{rand}_{2,i}^t$ are selected in the range $[0, 1]$. In (32), $p_{pbest,i}^t$ means the personal best solution which is determined by the following statements.

$$p_{pbest,i}^t \quad := \quad \arg \min_{x \in \{p_i^j | j=1,2\ldots,t\}} E_f(p) \tag{33}$$

p_{gbest}^t means the global best solution which is determined by the following statements.

$$p_{gbest}^t \quad := \quad \arg \min_{x \in \{p_{pbest,i}^t | i=1,2\ldots,n\}} E_f(p) \tag{34}$$

The PSO algorithm for the dynamic quantizer design is given as following steps:

PSO algorithm

- **Step2-1**: Set t = 0. For $i = 1, \cdots, m_r$, initial quantizer positions p_i^0 and its velocities Δp_i^0 are selected by using iterative design algorithm in Section IV-A. For $i = m_r + 1, \cdots, m$, initial position p_i^0 and velocity Δp_i^0 are selected randomly and evaluate the corresponding objective function at each position.
- **Step2-2**: Update $p_{pbest,i}^t$ and p_{gbest}^t by (33) and (34), respectively. Then, apply update laws (31), (32) for all particles, and go to Step2-3.
- **Step2-3**: Evaluate all position p_i^t by (30). Set $t = t+1$ and go to Step2-2 if $t < t_{\max}$. Else, update $p_{pbest,i}^{t_{\max}}$ and $p_{gbest}^{t_{\max}}$. $p_{gbest}^{t_{\max}}$ is the designed parameters. For the given $p_{gbest}^{t_{\max}}$, $\psi^{*,L} + \psi^{opt,L}$ is calculated and the quantization interval $d^{opt,L}$ is derived using $\psi^{*,L} + \psi^{opt,L}$.

m_r is a positive integer to give good initial quantizers for the PSO algorithm. In particular, p_{gbest}^0 is better initial quantizer if $m_r \geq 1$. The PSO algorithm is simple and it makes no assumption about the quantizer design problem. It might be obtained a good solution for the evaluation function $E(p)$ because we use the quantizers, which is designed by the iterative algorithm, as the initial particles.

V. Numerical Examples

The effectiveness of the proposed method is evaluated through numerical examples. The range of u is assumed to be $U = [-1, 1]$. Plant parameters are defined as

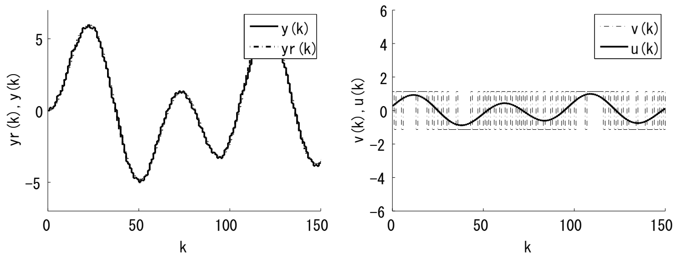
$$P_1 = \left(\begin{array}{c|c} A_p & B_p \\ \hline C_p & D_p \end{array} \right) = \left(\begin{array}{cc|c} 1.7326 & -0.7408 & 0.5 \\ 1 & 0 & 0 \\ \hline 0.3533 & -0.0083 & 0 \end{array} \right).$$

P_1 is derived by using $P_1(s) = (s+20)/(s^2+3s+2)$ with sampling time $\Delta t = 0.1$. Discrete systems P_2 and P_3 are derived by using $P_2(s) = 1/(s^2 + 3s + 2)$ and $P_3(s) = (s - 5)/(s^2 + 3s + 2)$, respectively.

For the proposed quantizer design algorithm, the parameters are selected as $m_r = 1$, $m = 1000$, $t_{\max} = 300$, $E_{pen} = 10^6$, $\omega_0 = 0.9$, $\omega_1 = \omega_2 = 1$ and $L = 100$. For $i = 2, \cdots, 1000$, initial positions p_i^0 and velocities Δp_i^0 are selected randomly so that their entries lie in the range $[-1, 1]$. In case with P_1 and $N = 2$, the following quantizer parameters are obtained by the proposed algorithm.

$$Q \quad = \quad \left(\begin{array}{c|c} \mathcal{A} & \mathcal{B} \\ \hline \mathcal{C} & d \end{array} \right) = \left(\begin{array}{cc|c} 0 & 1 & 0 \\ 0.728 & 0.160 & 1 \\ \hline -0.728 & -0.937 & 1.882 \end{array} \right)$$

$\psi^{opt} = 0.937$ is obtained for the filter parameters $\{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$, which satisfy the condition $N - \psi^{opt} > 0$. Moreover, $E(Q) = 0.757$ is obtained and it is small.



(a) (b)

Fig. 4. (a) Actual ($y(k)$) and desired outputs ($y_r(k)$) of the control system (b) Input $u(k)$ and output signals $v(k)$ of quantizer $Q_{proposed}$

TABLE I
Comparison for design algorithm in [16]

	P_1	P_2	P_3
$E(Q)$ of proposed method with $N = 2$	0.757	0.0378	0.309
$E(Q_{it})$ of Ref. [16] with $N = 2$	3.27	0.246	0.747
$E(Q)$ of proposed method with $N = 8$	0.316	0.0016	0.0303
$E(Q_{it})$ of Ref. [16] with $N = 8$	0.329	0.0552	0.0530

TABLE II
Comparison for optimal quantizer [11] with P_1 and some quantization levels N

N	2	4	8	16
$E(Q)$ of proposed method	0.757	0.1088	0.0316	0.0130
$E(Q_{oq})$ of Ref. [11]	–	0.1116	0.0316	0.0130

Simulation results for Q with

$$u(k) = 0.3\cos(0.06k) + 0.7\sin(0.13k) \in [-1, 1] \tag{35}$$

are shown in Fig 4 [1]. In Fig. 4 (a), $y_r(k)$ is the desired output (thin line) and $y(k)$ is the actual output using Q (thick line). Both outputs are similar, suggesting that Q displays good performance. Fig. 4 (b) shows the quantizer input $u(k)$ and output $v(k)$ using Q for two quantization levels. Q satisfies the communication rate constraints.

To show the effectiveness of the proposed algorithm, the quantizer Q_{it} by iteration algorithm [16] is used as a target for comparison. The quantizers are designed for each N and P_i and the obtained $E(Q)$ values are presented in Table. I. We can find that $E(Q)$ is smaller than $E(Q_{it})$ for all sets P_i and N. In particular, the difference is larger if N takes smaller value.

Then, the optimal quantizer $Q_{oq} = \{\mathcal{A}_{oq}, \mathcal{B}_{oq}, \mathcal{C}_{oq}\}$ in [11] acts as a target for comparison. P_1, which is stable minimum-phase system, is selected as the plant to design Q_{oq}. For obtained Q_{oq}, the resulting ψ^{oq} value amounts to 2.4176, which does not satisfy the condition for the communication rate (Theorem 1). Therefore, if a quantization interval is set in Q_{st}, some signals $u(k)$ cannot meet the range constraints in this quantizer. In Table. II, Q_{oq} and Q are compared by the value $E(Q)$ for different numbers of quantization levels ($N = 2, 4, 8, 16$). For $N = 2$, Q_{oq} does not satisfy the communication rate constraints. When N becomes larger, the value $E(Q)$ close to $E(Q_{oq})$. For large $N(= 8, 16)$ almost same quantizer parameters for Q_{oq} are obtained by the proposed design algorithm.

In this study, a method is developed for the design of SISO dynamic quantizers. This method may easily be extended to MIMO systems by combining the result in [16] and this paper.

[1] To confirm whether the signal range constraint is satisfied, we use $Q_{st}[\cdot]$ which have no saturation. If range constraint is not satisfied, the number of output level in Fig. 4 becomes greater than N.

Appendix A
Estimation of ψ using an invariant set analysis [20]

The controllability pair in (14) is given by $((\mathcal{A}+\mathcal{B}\mathcal{C}),\,\mathcal{B})$. The minimization problem of ψ is formulated as follow.

Problem 3: Assume that $\{\mathcal{A},\,\mathcal{B},\,\mathcal{C}\}$ are given. Find the following $\psi^*\,(>0)$.

$$\psi^* = \min_{Z_d>0,\beta} \psi \tag{36}$$

$$\textbf{subject to} \begin{bmatrix} Z_d & \mathcal{C}^T \\ \mathcal{C} & \psi^2 \end{bmatrix} \geq \mathbf{0},$$

$$\begin{bmatrix} (1-\beta)Z_d & 0 & (\mathcal{A}+\mathcal{B}\mathcal{C})^T Z_d \\ 0 & \beta I & \mathcal{B}^T Z_d \\ Z_d(\mathcal{A}+\mathcal{B}\mathcal{C}) & Z_d\mathcal{B} & Z_d \end{bmatrix} \geq 0,$$

$$\beta \in \left[0, 1-\rho(\mathcal{A}+\mathcal{B}\mathcal{C})^2\right].$$

■

Note that Z_d is the positive definite matrix that characterize the invariant sets for $(\mathcal{A}+\mathcal{B}\mathcal{C}, \mathcal{B})$. For fixed $\mathcal{A}, \mathcal{B}, \mathcal{C}$ and β, *Problem* 3 is regarded as an LMI optimization problem. Therefore, the solution of *Problem* 3 is accessible by the inner point method of ψ and Z_d with the line search about β. As a result, we can get ψ^* by solving *Problem* 3. For fixed Z_d and β, *Problem* 3 is regarded as another LMI optimization problem.

The quantization interval d^* is derived using

$$d^* = \frac{u_{\max} - u_{\min}}{N - \psi^*}. \tag{37}$$

When we set $d = d^*$ in Q_{st}, the communication rate constraint is satisfied for any $u(k) \in U$. Therefore, d^* is a solution that satisfies the condition related to permissible number of quantization levels. In addition, the corresponding center point c^* is written as $c^* = (u_{\min} + u_{\max})/2$.

The conservativeness of d^* exists but it might be small. This conservativeness may stem from several reasons. First, the obtained invariant set is larger than the reachable set. Second, $\omega(k)$ is not guaranteed to take an arbitrary value satisfying $\|\omega\| \leq 1$. On the other hand, a smaller estimated reachable set generate a smaller d.

Appendix B
Design of dynamic quantizer parameters using an invariant set analysis [21]

A design method has been proposed for $\{\mathcal{A},\,\mathcal{B},\,\mathcal{C}\}$ via invariant set analysis [17]. A minimization problem with inequality conditions is defined as follow.

Problem 4: Find parameters $\{\mathcal{A},\,\mathcal{B},\,\mathcal{C}\}$, which minimize γ.

$$\gamma^* = \min_{\mathcal{A},\mathcal{B},\mathcal{C},Z_p>0,\alpha,\gamma^2} \gamma \tag{38}$$

$$\textbf{subject to}$$

$$\begin{bmatrix} Z_p & \bar{C}^T \\ \bar{C} & \gamma^2 \end{bmatrix} \geq 0, \begin{bmatrix} (1-\alpha)Z_p & 0 & \bar{A}^T Z_p \\ 0 & \alpha I & \bar{B}^T Z_p \\ Z_p\bar{A} & Z_p\bar{B} & Z_p \end{bmatrix} \geq 0$$

$$\alpha \in \left[0, 1-\rho(\bar{A})^2\right]$$

■

In *Problem* 4, $\mathcal{A}, \mathcal{B}, \mathcal{C}, Z_p, \alpha$ and γ are the design variables, and Z_p is the positive definite matrix. When *Problem* 4 is solved, the obtained quantizer $\{\mathcal{A},\,\mathcal{B},\,\mathcal{C}\}$ obeys the condition

$$E(Q) \leq \gamma\frac{d}{2}. \tag{39}$$

Because the quantization interval d is fixed in [21], the minimization of $\gamma d/2$ is equivalent to minimization of γ. When α and $\{\mathcal{A},\,\mathcal{B},\,\mathcal{C}\}$ are fixed, *Problem* 4 is regarded as an LMI optimization problem with a variable matrix Z_p. Moreover, for fixed α and Z_p values, it is considered as another LMI optimization problem.

References

[1] P. Antsaklis and J. Baillieul (eds.), "Guest editional special issue on networked control systems", *IEEE Trans. Autom. Control*, Vol. 49, No. 9, pp. 1421–1423, Sep. 2004.

[2] G. Nair, F. Fagnani, S. Zampieri, and R. Evans, "Feedback control under data rate constraints: An overview", *Proc. IEEE*, Vol. 95, No. 1, pp. 108–137, Jan. 2007.

[3] K. Tsumura, H. Ishii, H. Hoshina, "Tradeoffs between quantization and packet loss in networked control of linear systems", *Automatica*, Vol. 45, No. 12, pp. 2963–2970, Dec. 2009.

[4] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints-Part I: State estimation problems", *IEEE Trans. Autom. Control*, Vol. 42, No. 9, 1294–1299, Sep. 1997.

[5] W. S. Wong and R. W. Brockett, "Systems with finite communication bandwidth constraints-Part II: Stabilization with limited information feedback", *IEEE Trans. Autom. Control*, Vol. 44, No. 5, pp. 1049–1053, May 1999.

[6] S. Tatikonda and S. Mitter, "Control under communication constraints", *IEEE Trans. Autom. Control*, Vol. 49, No. 7, pp. 1056–1068, July 2004.

[7] R. W. Brockett and D. Liberzon, "Quantized feedback stabilization of linear systems", *IEEE Trans. Autom. Control*, Vol. 45, No. 7, pp. 1279–1289, July 2000.

[8] B. Picasso and A. Bicchi, "On the stabilization of linear systems under assigned I/O quantization", *IEEE Trans. Autom. Control*, vol. 52, no. 10, pp. 1994–2000, Oct. 2007.

[9] H. Inose, Y. Yasuda and J. Murakami, "A telemetering system by code modulation-$\Delta \cdot \Sigma$ modulation", *IRE Trans. on SET*, Vol. 8, No. 3, pp. 205–209, March 1962.

[10] S. R. Norsworthy, R. Schreier and G. C. Temes, "Delta-sigma data converters", *IEEE Press*, 1996.

[11] S. Azuma and T. Sugie, "Optimal dynamic quantizers for discrete-valued input control", *Automatica*, Vol. 44, No. 2, pp. 396–406, Feb. 2008.

[12] S. Azuma and T. Sugie, "Synthesis of optimal dynamic quantizers for discrete-valued input control", *IEEE Trans. Autom. Control*, Vol. 53, No. 9, pp. 2064–2075, Sep. 2008.

[13] Y. Minami, S. Azuma and T. Sugie, "An optimal dynamic quantizer for feedback control with discrete-valued signal constraints", *46th IEEE Conference on Decision and Control*, pp. 2259–2264, Dec. 2007.

[14] S. Azuma and T. Sugie, "Dynamic quantization of nonlinear control systems",*IEEE Trans. Autom. Control*, Vol. 57, No. 4, pp. 875–888, April 2012.

[15] D. E. Quevedo, G. C. Goodwin, and J. A. De Dona, "Finite constraint set receding horizon quadratic control", *International Journal of Robust and Nonlinear Control*, Vol. 14, No. 4, pp. 355–377 April 2004.

[16] H.Okajima, K. Sawada, N. Matsunaga and Y. Minami, "Dynamic quantizer design for MIMO systems based on communication rate constraint", *Electron. Comm. Jpn.*, Vol. 96, No. 5, pp. 28-36, May. 2013.

[17] H. Shingin and Y. Ohta, "Optimal invariant sets for discrete-time systems approximation of reachable sets for bounded inputs", *10th IFAC/IFORS/IMACS/IFIP Symposium on Large Scale Systems*, pp. 401–406, July 2004.

[18] J. Kennedy and R. Eberhart, "Particle swarm optimization", *Proc. of IEEE international conference on neural networks*, Vol. 4, pp. 1942–1948, 1995.

[19] I. Maruta, T. Kim and T. Sugie, "Fixed-structure H_∞ controller synthesis: A meta-heuristic approach using simple constrained particle swarm optimization", *Automatica*, Vol. 45, No. 2, pp. 553–559, Feb. 2009.

[20] H. Okajima, N. Matsunaga and K. Sawada, "Optimal quantization interval design of dynamic quantizers which satisfy the communication rate constraints", *49th IEEE CDC*, pp. 4733–4739, Dec. 2010.

[21] K. Sawada and S. Shin, "Dynamic quantizer synthesis based on invariant set analysis for SISO systems with discrete-valued Input", *The 19th International Symposium on Mathematical Theory of Networks and Systems*, pp. 1385–1390, July 2010.

[22] H. Okajima, K. Sawada, N. Matsunaga, "Integrated design of filter and interval in dynamic quantizer under communication rate constraint", *The 18th IFAC World Congress*, pp. 8785–8791, Aug. 2011.

[23] R. Yoshino, H. Okajima, N. Matsunaga and Y. Minami, "Dynamic quantizers design under data constraints by using PSO method", *SICE Annual Conference 2014*, pp. 1041–1046, Sep. 2014.