Optimal decentralized sigma-delta modulators for quantized feedback control

Yuki Minami\textsuperscript{1a)}, Shun-ichi Azuma\textsuperscript{2b)}, and Toshiharu Sugie\textsuperscript{2c)}

\textsuperscript{1} Department of Control Engineering, Maizuru National College of Technology
234 Shiroya, Maizuru, Kyoto 625-8511, Japan

\textsuperscript{2} Graduate School of Informatics, Kyoto University
Gokasho, Uji, Kyoto 611-0011, Japan

\textsuperscript{a)} minami@maizuru-ct.ac.jp
\textsuperscript{b)} sazuma@i.kyoto-u.ac.jp
\textsuperscript{c)} sugie@i.kyoto-u.ac.jp

Received November 14, 2011; Revised February 28, 2012; Published July 1, 2012

Abstract: This paper addresses a problem of designing decentralized sigma-delta modulators for quantized control, i.e., feedback control subject to quantized signal constraints. The sigma-delta modulators to be considered here have a limited information structure so as to be implemented in a decentralized manner, which poses a challenging design problem. We first analytically derive a solution to the problem such that the resulting quantized feedback system optimally approximates the corresponding unquantized system. Next, the performance is demonstrated by a numerical simulation and an experiment for the stabilization problem of a seesaw-cart system.

Key Words: sigma-delta modulator, feedback control, quantized control

1. Introduction

In feedback control systems, some of signals are often restricted to be \textit{discrete-valued}, due to the use of discrete-level actuators/sensors, A/D and D/A converters, and digital communication channels [1–3]. One of the promising approaches to construct control systems in such a situation is to adopt a sigma-delta modulator [4–6] for appropriately converting continuous-valued signals to discrete-valued ones. So various results on designing sigma-delta modulators or equivalents have been reported in the control community [7–15].

The authors also have considered the use of sigma-delta modulators (where the modulators are called the dynamic quantizers) [16, 17] and the following problem has been studied: when a plant $P$ and a controller $K$ are given in the system in Fig. 1 (a), find a sigma-delta modulator $M$ optimizing the control performance. So far, a closed-form solution has been derived and some experiments have been performed successively [16, 18]. On the other hand, unlike the quantized feedback systems with
a single modulator (in Fig. 1 (a)), it is often necessary to have a decentralized structure in sigma-delta modulators. For example, when two sigma-delta modulators have to be separately embedded in systems, i.e., quantization is necessary for both actuator and sensor signals as shown in Fig. 1 (b), the modulators should have a decentralized information structure. This is required, for example, in the case that the plant and the controller are remotely placed and are connected over a digital network. Therefore, the design problem of decentralized sigma-delta modulators is one of important issues, but it has been still open.

This paper thus addresses a decentralized sigma-delta modulator design problem, formulated as follows: when a plant $P$ and a controller $K$ are given in Fig. 1 (b), find sigma-delta modulators $M_1$ and $M_2$ such that the resulting quantized system in Fig. 1 (b) optimally approximates its unquantized version in Fig. 2, in the sense of the input-output relation. This is a decentralized version of the authors’ sigma-delta modulator design problem in [16, 18], and it is more challenging.

To this problem, our contributions are summarized as follows. First, we derive optimal decentralized sigma-delta modulators in a closed-form. This result is somewhat surprising because, in general, the decentralized object design in control systems, such as the decentralized controller design, is known to be very difficult (or NP-hard) [19, 20]. Second, by performing an experiment for a seesaw-cart system, it is demonstrated that the optimal decentralized sigma-delta modulators exhibit satisfactory performance in the real world. In particular, we show here that the unstable mechanical system is stabilized with the severe constraint that, at each time instant, the plant input can take one of three kinds of values and the controller input can take one of seven. Note here that most of the existing studies on quantized control [7–17] have focused on theoretical aspects, and thus few result on experimental validation can be found.

This paper is organized as follows. In Section 2, our design problem is formulated. A solution to the problem is next presented in Section 3 and is demonstrated by an experiment in Section 4. Section 5 concludes this paper.

Finally, we note that this paper is based on our preliminary version [21], which is published in a conference proceedings, and, as a journal paper, this paper contains full explanations, proofs, and simulations omitted there.
Consider the feedback system $\Sigma_M$ shown in Fig. 3, composed of the discrete-time linear system $G$ and the decentralized sigma-delta modulator $M$. The system $G$ is given by

$$G : \begin{cases} x(k + 1) = Ax(k) + B_1 r(k) + B_2 v(k), \\ z(k) = C_1 x(k) + D_1 r(k), \\ u(k) = C_2 x(k) + D_2 r(k) \end{cases}$$

where $x(k)$ is the state, $r(k) \in \mathbb{R}^p$ and $u(k) \in \mathbb{R}^m$ are the inputs, $z(k) \in \mathbb{R}^l$ and $u(k) \in \mathbb{R}^m$ are the outputs, $k \in \{0\} \cup \mathbb{N}$ is the discrete-time, and $A \in \mathbb{R}^{n \times n}$, $B_1 \in \mathbb{R}^{n \times p}$, $B_2 \in \mathbb{R}^{m \times m}$, $C_1 \in \mathbb{R}^{l \times n}$, $C_2 \in \mathbb{R}^{m \times n}$, $D_1 \in \mathbb{R}^{l \times p}$, $D_2 \in \mathbb{R}^{m \times p}$ are constant matrices. The signals $r$ and $z$ correspond to the external input and the controlled output, respectively.

On the other hand, $M$ is the decentralized sigma-delta modulator composed of the $s$ sub-modulators

$$M_i : \begin{cases} \xi_i(k + 1) = A_i \xi_i(k) + B_{1i} u_i(k) + B_{2i} v_i(k), \\ v_i(k) = Q_i [\xi_i(k) + u_i(k)] \\ \end{cases} \quad (i \in \{1, 2, \ldots, s\}).$$

Here, $\xi_i(k) \in \mathbb{R}^{N_i}$ is the state, $u_i(k) \in \mathbb{R}^{m_i}$ is the input, $v_i(k) \in \mathbb{V}_{m_i}^l$ is the output, $\mathbb{V}_{m_i}^l \subset \mathbb{R}^{m_i}$ expresses the discrete set $\{0, \pm d_i, \pm 2d_i, \ldots\}$ for the quantization interval $d_i \in \mathbb{R}_+$. Furthermore, $A_i \in \mathbb{R}^{N_i \times N_i}$, $B_{1i}, B_{2i} \in \mathbb{R}^{N_i \times m_i}$, $C_i \in \mathbb{R}^{m_i \times N_i}$ are constant matrices. The function $Q_i : \mathbb{R}^{m_i} \rightarrow \mathbb{V}_{m_i}^l$ is the typical uniform quantizer and an example (for $m_i := 1$) is illustrated in Fig. 5. Note for $Q_i$ that

$$\text{abs}(Q_i[\mu] - \mu) \leq \frac{d_i}{2} 1_{m_i}, \quad (\forall \mu \in \mathbb{R}^{m_i})$$

holds. For $M_i$, we assume that the initial state is given by $\xi_i(0) = 0$.

The structure of the sigma-delta modulator $M_i$ is shown in Fig. 6. In the modulator $M_i$, the state $\xi_i(k)$ is determined by the past input sequence $\{u_i(0), u_i(1), \ldots, u_i(k - 1)\}$ and the past output $v_i(k - 1)$, and the current output $v_i(k)$ is given by the uniform quantization of the signal depending on $\xi_i(k)$ and $u_i(k)$. Thus, in this modulator, the quantization error is filtered through the feedback loop. From this, it is clear that the structure of the modulator considered in this paper is same as that of the conventional single-loop sigma-delta modulator.
Under these assumptions, the pseudo-inverse matrices of $D_v$ and that the signals $u_i(k)$ of $G$ are respectively composed of $v_i(k) \in \mathbb{R}^{m_i}$, $i = 1, 2, \ldots, s$ and $u_i(k) \in \mathbb{R}^{m_i}$, $i = 1, 2, \ldots, s$, i.e., $v(k) = [v_i^T(k) \ v_{i+1}^T(k) \ \ldots \ v_s^T(k)]^T$ and $u = [u_1^T(k) \ u_2^T(k) \ \ldots \ u_s^T(k)]^T$. Notice also that the system $\Sigma_M$ is a generalized version of the feedback system in Fig. 1 (b), and so various types of feedback systems, which include multiple modulators in the form of Eq. (2), can be treated by the framework for $\Sigma_M$.

Now, some symbols and assumptions are introduced. Let $B_{2i} \in \mathbb{R}^{n \times m_i}$, $i = 1, 2, \ldots, s$ denote the matrices such that $B_2 = [B_{21} \ B_{22} \ \ldots \ B_{2s}]$. We denote by $\tau_i$ the smallest integer $k \in \{0\} \cup \mathbb{N}$ satisfying $C_1(A + B_2 C_2)^k B_{2i} \neq 0$. For the system $G$, we assume that

(A1) $\text{rank}(D_2) = m$ (full row rank),

(A2) $\text{rank}(C_1(A + B_2 C_2)^\tau_i B_{2i}) = l$ (full row rank) for each $i \in \{1, 2, \ldots, s\}$.

Under these assumptions, the pseudo-inverse matrices of $D_2$ and $F_i := C_1(A + B_2 C_2)^\tau_i B_{2i}$ can be respectively expressed as $D_2^\dagger = D_2^\dagger (D_2 D_2^\dagger)^{-1}$ and $F_i^\dagger = F_i^\dagger (F_i F_i^\dagger)^{-1}$. The former assumption can be weakened as will be detailed in Section 3.3, while the latter is essential to derive our solution.

Now, we define the performance index. For the system $\Sigma_M$ with the initial state $x(0) = x_0 \in \mathbb{R}^n$ and the external input sequence $R := \{r_0, r_1, \ldots\} \in \ell_\infty$, let $Z_M(x_0, R)$ be the controlled output sequence for $k = 1, 2, \ldots, i.e.,$ $z(1), z(2), \ldots$ and let $z_M(k, x_0, R)$ be the output at the $k$-th time. In addition, we consider the unquantized feedback system $\Sigma$ in Fig. 4 as an ideal system to $\Sigma_M$, for which the symbols $Z(x_0, R)$ and $z(k, x_0, R)$ are similarly defined. Then the (worst-case) output difference between $\Sigma_M$ and $\Sigma$ is expressed as

$$E(M) := \sup_{(x_0, R) \in \mathbb{R}^n \times \ell_\infty} \|Z_M(x_0, R) - Z(x_0, R)\|,$$

which is used as our performance index.

**Problem 1** For the system $\Sigma_M$, assume (A1) and (A2). Then, find a decentralized sigma-delta modulator $M = (M_1, M_2, \ldots, M_s)$ (i.e., find dimensions $N_i$ and matrices $A_i, B_{1i}, B_{2i}, C_i$ for $i = 1, 2, \ldots, s$) minimizing $E(M)$.
The significance of the solution to Problem 1 is explained as follows. For example, consider the system in Fig. 1 (b). If $E(M)$ is small enough for the sigma-delta modulator $M := (M_1, M_2)$, the output behavior of the quantized feedback system in Fig. 1 (b) is similar to that of the unquantized system in Fig. 2. This implies that, if the quantized feedback system is constructed with a controller $K$ designed for the unquantized feedback in Fig. 2, which can be easily obtained by the conventional control theory, and the solution to Problem 1, the resulting quantized system would have a good performance. In other words, the solution to Problem 1 presents a practical design method of feedback control systems with quantized signal constraints.

3. Optimal decentralized sigma-delta modulators

3.1 A closed-form solution to Problem 1

The following lemma [16] is prepared.

Lemma 1 For the system $\Sigma_M$, assume (A1). If

$$\bar{C} \bar{A}^k \bar{B}_1 = 0 \quad (k = 0, 1, \ldots),$$

then

$$E(M) = \left\| \sum_{k=0}^{\infty} \text{abs}(\bar{C} \bar{A}^k \bar{B}_2) \right\| \frac{d_1}{2};$$

otherwise

$$E(M) = \infty$$

where

$$\bar{A} := \begin{bmatrix} \bar{A} & B_2 \bar{C} \\ 0 & A + B_2 \bar{C} \end{bmatrix}, \quad \bar{B}_1 := \begin{bmatrix} 0 \\ B_1 + B_2 \end{bmatrix}, \quad \bar{B}_2 := \begin{bmatrix} B_2 \\ \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \frac{d_3}{d_1} I_{m_3}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) \end{bmatrix},$$

$$\bar{C} := [ C_1 \ 0 ], \quad \bar{A} := A + B_2 C_2,$$

$$A := \begin{bmatrix} A_0 \cdots 0 \\ 0 \ A_2 \cdots \cdots \cdots 0 \\ \vdots \ \cdots \ \cdots \ \cdots 0 \\ 0 \cdots 0 \ A_s \end{bmatrix}, \quad B_1 := \begin{bmatrix} B_{11} \cdots 0 \cdots 0 \\ 0 \ B_{12} \cdots \cdots \cdots 0 \\ \vdots \ \cdots \ \cdots \ \cdots 0 \\ 0 \cdots 0 \ B_{1s} \end{bmatrix}, \quad B_2 := \begin{bmatrix} B_{21} \cdots 0 \cdots 0 \\ 0 \ B_{22} \cdots \cdots \cdots 0 \\ \vdots \ \cdots \ \cdots \ \cdots 0 \\ 0 \cdots 0 \ B_{2s} \end{bmatrix}, \quad \bar{c} := \begin{bmatrix} c_1 \cdots 0 \cdots 0 \\ 0 \ c_2 \cdots \cdots \cdots 0 \\ \vdots \ \cdots \ \cdots \ \cdots 0 \\ 0 \cdots 0 \ c_s \end{bmatrix}.$$

(8)

Lemma 1 provides an explicit expression of the performance index $E(M)$, which helps us to compute the value of $E(M)$ for given $M$. An intuitive interpretation of this lemma is as follows. Let us introduce the new variable $w(k) \in [-d_1/2, d_1/2]^m$:

$$w(k) := Q \begin{bmatrix} \text{diag} \left( I_{m_1}, \frac{d_1}{d_2} I_{m_2}, \frac{d_1}{d_2} I_{m_3}, \ldots, \frac{d_1}{d_2} I_{m_s} \right) (\bar{c} \xi(k) + u(k)) \\ - \text{diag} \left( I_{m_1}, \frac{d_1}{d_2} I_{m_2}, \frac{d_1}{d_2} I_{m_3}, \ldots, \frac{d_1}{d_2} I_{m_s} \right) (\bar{c} \xi(k) + u(k)) \end{bmatrix}$$

(9)

with $\xi(k) := [ \xi_1^T(k) \ 
\xi_2^T(k) \ 
\cdots \ 
\xi_s^T(k) ]^T$ where $Q : \mathbb{R}^m \to \mathbb{V}^m$ is the uniform quantizer with quantization interval $d_1 \in \mathbb{R}_+$. Using this, the error system between $\Sigma_M$ and $\Sigma$, which is illustrated in Fig. 7, can be formally regarded as a linear system whose output is $z_M - z$ and inputs are $r$ and $w$ [16]. Then, the matrices $\bar{C} \bar{A}^k \bar{B}_1$ and $\bar{C} \bar{A}^k \bar{B}_2$ in Eq. (5) and Eq. (6) correspond, respectively, to the impulse response matrices from $r$ to $z_M - z$ and from $w$ to $z_M - z$. By considering these facts, it follows that

$$z_M(T, x_0, R) - z(T, x_0, R) = \sum_{k=0}^{T-1} \bar{C} \bar{A}^{(T-1)-k} \bar{B}_2 w(k)$$

(10)

390
subject to Eq. (5). This leads to Eq. (6). On the other hand, if Eq. (5) does not hold, then \( z_M - z \) depends on \( r \), which gives Eq. (7).

Now, let us derive a solution to Problem 1 by using Lemma 1. A solution to Problem 1 is a sigma-delta modulator \( M \) such that Eq. (5) holds and the right hand side of Eq. (6) is minimized. In order to obtain such a solution, we rewrite Eq. (5) and Eq. (6) as follows.

Let \( B_{11} \in \mathbb{R}^{(n+N_1+N_2+\cdots+N_s)\times m_i} \) be the matrices satisfying \( \bar{B}_1 = [\bar{B}_{11} \bar{B}_{12} \cdots \bar{B}_{1s}] \). Then, we can rewrite Eq. (5) as

\[
\bar{C} \bar{A}^k \bar{B}_{11} \bar{C} \bar{A}^k \bar{B}_{12} \cdots \bar{C} \bar{A}^k \bar{B}_{1s} = 0 \quad (k = 0, 1, \ldots).
\]

Here, from Eq. (8), the matrices \( \bar{A}, \bar{B}_1, \) and \( \bar{C} \) can be represented by

\[
\bar{A} = \begin{bmatrix} \bar{A}_{21} \bar{C}_1 & \bar{B}_{22} \bar{C}_2 & \cdots & \bar{B}_{2s} \bar{C}_s \\ 0 & A_1 + \bar{B}_{21} \bar{C}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_s + \bar{B}_{2s} \bar{C}_s \\ \end{bmatrix}, \quad \bar{B}_1 = \begin{bmatrix} \bar{B}_{11} & \bar{B}_{12} & \cdots & \bar{B}_{1s} \\ \bar{B}_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \\ \end{bmatrix}, \quad (12)
\]

Thus, we obtain

\[
\bar{C} \bar{A}^k \bar{B}_{11} = \begin{bmatrix} C_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{A} & \bar{B}_{2i} \bar{C}_i \\ 0 & A_i + \bar{B}_{2i} \bar{C}_i \end{bmatrix}^k \begin{bmatrix} 0 \\ \bar{B}_{1i} + \bar{B}_{2i} \end{bmatrix} \quad (k = 1, 2, \ldots, \ i = 1, 2, \ldots, s),
\]

which depends not on \( M_j \) (\( j \neq i \)) but \( G \) and \( M_i \). Therefore, if

\[
\begin{bmatrix} C_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{A} & \bar{B}_{2i} \bar{C}_i \\ 0 & A_i + \bar{B}_{2i} \bar{C}_i \end{bmatrix}^k \begin{bmatrix} 0 \\ \bar{B}_{1i} + \bar{B}_{2i} \end{bmatrix} = 0
\]

holds for each \( M_i \), then Eq. (5) holds.

On the other hand, if (A1) and Eq. (5) hold for \( \Sigma_M \), the value of \( E(M) \) is given by the right hand side of Eq. (6), where the matrices \( \bar{A} \) and \( \bar{C} \) are given by Eq. (12), and the matrix \( \bar{B}_2 \) can be represented by

\[
\bar{B}_2 = \begin{bmatrix} B_{21} & B_{22} & \cdots & B_{2s} \\ B_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & B_{2s} \end{bmatrix} \quad \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right)
\]

from Eq. (8). Letting \( \bar{B}_{2i} \in \mathbb{R}^{(n+N_1+N_2+\cdots+N_s)\times m_i} \) be the matrices satisfying \( \bar{B}_2 = [\bar{B}_{21} \bar{B}_{22} \cdots \bar{B}_{2s}] \), we obtain

\[
E(M) = \left\| \sum_{k=0}^{\infty} \text{abs} \left( \bar{C} \bar{A}^k | \bar{B}_{21} \bar{B}_{22} \cdots \bar{B}_{2s} \right) \right\| \frac{d_1}{2} = \left\| \sum_{k=0}^{\infty} \text{abs} \left( \bar{C} \bar{A}^k \bar{B}_{21} \right) \sum_{k=0}^{\infty} \text{abs} \left( \bar{C} \bar{A}^k \bar{B}_{22} \right) \cdots \sum_{k=0}^{\infty} \text{abs} \left( \bar{C} \bar{A}^k \bar{B}_{2s} \right) \right\| \frac{d_1}{2}.
\]

In this equation, the matrices \( \bar{C} \bar{A}^k \bar{B}_{2i} \) can be expressed as

\[
\bar{C} \bar{A}^k \bar{B}_{2i} = \begin{bmatrix} C_1 & 0 \end{bmatrix} \begin{bmatrix} \bar{A} & \bar{B}_{2i} \bar{C}_i \\ 0 & A_i + \bar{B}_{2i} \bar{C}_i \end{bmatrix}^k \begin{bmatrix} \bar{B}_{2i} \\ \bar{B}_{2i} \end{bmatrix} \frac{d_i}{d_1},
\]

(16)
for which we have \( \tilde{C} \tilde{A}^k \tilde{B}_{2i} = 0 \) for \( k \leq \tau_i - 1 \) and \( \tilde{C} \tilde{A}^k \tilde{B}_{2i} = C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i}(d_i/d_1) \) for \( k = \tau_i \). Note that \( \tau_i \in \{0\} \cup \mathbb{N} \) is the smallest integer satisfying \( C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i} \neq 0 \) as defined in Section 2. Therefore, it follows that

\[
E(M) = \left\| \begin{bmatrix} \text{abs}(C_1 \tilde{A}^{\tau_1} \tilde{B}_{21}) & \text{abs}(C_1 \tilde{A}^{\tau_2} \tilde{B}_{22}) & \cdots & \text{abs}(C_1 \tilde{A}^{\tau_s} \tilde{B}_{2s}) \end{bmatrix} \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) \right\|_2.
\]

In the norm of the right hand side, the first term depends only on the parameters of \( G \) and the second term depends on the parameters of \( G \) and \( M \). In addition, the matrices \( \text{abs}(\tilde{C} \tilde{A}^k \tilde{B}_{2i}) \) \( (i = 1, 2, \ldots, s) \) have nonnegative elements, and the matrix \( \text{abs}(\tilde{C} \tilde{A}^k \tilde{B}_{2i}) \) includes only the parameters of \( G \) and \( M_i \).

Thus, if there exists sub-modulators \( M_1, M_2, \ldots, M_s \) satisfying Eq. (14) and

\[
\text{abs}(\tilde{C} \tilde{A}^k \tilde{B}_{2i}) = \text{abs}\left( C_1 \begin{bmatrix} \tilde{A} & 0 \\ 0 & \tilde{A}_i + \tilde{B}_{2i} e_i \end{bmatrix}^k \begin{bmatrix} \tilde{B}_{2i} \\ \tilde{B}_{2i} \end{bmatrix} \right) \frac{d_i}{d_1} = 0 \quad (k = \tau_i + 1, \tau_i + 2, \ldots, i = 1, 2, \ldots, s),
\]

such a decentralized sigma-delta modulator \( M = (M_1, M_2, \ldots, M_s) \) is a solution to Problem 1. So we have the following result.

**Theorem 1** For the system \( \Sigma_M \), assume (A1) and (A2). Then an optimal decentralized sigma-delta modulator is \( M^* = (M_1^*, M_2^*, \ldots, M_s^*) \) with

\[
M_i^* : \begin{cases} 
\xi_i(k + 1) = \tilde{A} \xi_i(k) - \tilde{B}_{2i} u_i(k) + B_{2i} v_i(k) \\
u_i(k) = Q_i[-(C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i})^\top C_1 \tilde{A}^{\tau_i+1} \xi_i(k) + u_i(k)]
\end{cases}
\]

where \( N_i := n \ (i = 1, 2, \ldots, s) \), and the minimum value of the performance is given by

\[
E(M^*) = \left\| \text{abs}\left( C_1 [\tilde{A}^{\tau_1} \tilde{B}_{21} \cdots \tilde{A}^{\tau_s} \tilde{B}_{2s}] \right) \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) \right\|_2.
\]

**Proof 1** As mentioned above, a decentralized sigma-delta modulator \( M \) satisfying Eq. (5) and Eq. (18) is optimal under (A1). Meanwhile, Eq. (5) holds for \( M^* \) because \( \tilde{B}_2 = -\tilde{B}_2 \) in \( M^* \). In addition, since \( (C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i})^\top = (C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i})^\top ((C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i}) (C_1 \tilde{A}^{\tau_i} \tilde{B}_{2i})^\top)^{-1} \) under (A2), Eq. (18) also holds for \( M^* \). These complete the proof. \( \blacksquare \)

Theorem 1 gives a closed-form solution to Problem 1, and the minimum value of the performance index, which shows the performance limitation of the decentralized sigma-delta modulators in the form of Eq. (2). The difference between the proposed optimal sigma-delta modulator and the conventional modulator used in the field of signal processing is that the proposed modulator is custom-made for a given feedback system, while the conventional one is general-purpose. The proposed modulator in Eq. (19) is characterized by the parameters of the system \( G \), i.e., the parameters of plant and controller. On the other hand, the conventional sigma-delta modulator has been studied in the absent of feedback loop, and thus, it has been designed without the information of feedback system.

![Fig. 7. Error system between \( \Sigma_M \) and \( \Sigma \).](image)
3.2 Example

Consider the system in Fig. 1 (b). The plant $P$ and the controller $K$ are given by

$$P : \begin{align*}
x_p(k+1) &= \begin{bmatrix} 0.952 & 0.09358 \\ -0.9358 & 0.8584 \end{bmatrix} x_p(k) + \begin{bmatrix} 0.004798 \\ 0.09358 \end{bmatrix} v_1(k), \\
z(k) &= \begin{bmatrix} 6.5 & 1 \end{bmatrix} x_p(k), \\
y(k) &= \begin{bmatrix} 6.5 & 1 \end{bmatrix} x_p(k),
\end{align*}$$

$$K : \begin{align*}
x_k(k+1) &= \begin{bmatrix} 0.533 & 0.002 \\ -2.631 & 0.405 \end{bmatrix} x_k(k) + \begin{bmatrix} 0 \\ 0 \end{bmatrix} r_R(k) + \begin{bmatrix} 0.0712 \\ 0.1538 \end{bmatrix} v_2(k), \\
u_1(k) &= [-14.25 -3.0] x_k(k) + r_R(k).
\end{align*}$$

This controller has desirable performance in the unquantized system in Fig. 2. For the system $G$ defined by the above $P$ and $K$, (A1) and (A2) hold. We assume here that the inputs of $P$ and $K$, i.e., $v_1$ and $v_2$, take values on $\{0, \pm 1, \pm 2, \ldots\}$, which is supposed to be a severe constraint. Then the optimal decentralized sigma-delta modulator $M^*$ in Theorem 1 is used with $d_1 = d_2 := 1$. From the parameters of the plant $P$ and the controller $K$, the optimal modulator $M^* = (M^*_1, M^*_2)$ is given by

$$M^*_1 : \begin{align*}
\xi_1(k+1) &= \begin{bmatrix} 0.952 & 0.09358 & -0.06836 & -0.01439 \\ -0.9358 & 0.8584 & -1.334 & -0.2808 \\ 0.4628 & 0.0712 & 0.533 & 0.002 \\ 1.029 & 0.1583 & -2.631 & 0.405 \end{bmatrix} \xi_1(k) - \begin{bmatrix} 0.004798 \\ 0.09358 \\ 0 \\ 0 \end{bmatrix} u_1(k) + \begin{bmatrix} 0.004798 \\ 0.09358 \\ 0 \\ 0 \end{bmatrix} v_1(k), \\
v_1(k) &= Q_1 \begin{bmatrix} -42.1 & -11.76 & 14.25 & 3 \end{bmatrix} \xi_1(k) + u_1(k),
\end{align*}$$

$$M^*_2 : \begin{align*}
\xi_2(k+1) &= \begin{bmatrix} 0.952 & 0.09358 & -0.06836 & -0.01439 \\ -0.9358 & 0.8584 & -1.334 & -0.2808 \\ 0.4628 & 0.0712 & 0.533 & 0.002 \\ 1.029 & 0.1583 & -2.631 & 0.405 \end{bmatrix} \xi_2(k) - \begin{bmatrix} 0 \\ 0 \\ 0.0712 \\ 0.1583 \end{bmatrix} u_2(k) + \begin{bmatrix} 0 \\ 0 \\ 0.0712 \\ 0.1583 \end{bmatrix} v_2(k), \\
v_2(k) &= Q_2 \begin{bmatrix} 13.02 & 8.42 & -12.26 & -3.457 \end{bmatrix} \xi_2(k) + u_2(k).
\end{align*}$$

Figure 8 shows the time responses of $u_1, v_1, u_2, v_2,$ and $z$ for the conditions $x_p(0) := [-0.10 -0.24]^\top$, $x_k(0) := [0 0]^\top$, $r_R(k) \equiv 0$, and $r_N(k) \equiv 0$. The thick and thin lines correspond to the responses of the quantized system in Fig. 1 (b) and the unquantized system in Fig. 2, respectively. It turns out that the output response of the quantized system with the optimal decentralized sigma-delta modulator is quite similar to that of the unquantized (ideal) system.

For comparison, we also show the responses of the quantized system in Fig. 1 (b) with the uniform quantizer $M := (Q_1, Q_2)$ in the fifth figure by the dotted line. We can see that the output response is completely different from that of the unquantized system.

These results illustrate that the proposed decentralized sigma-delta modulator has much better performance than the uniform quantizer and will be a useful tool for control subject to discrete-valued signal constraints.
3.3 Extensions to more general cases
This section extends Theorem 1 to more general cases.

1) Case where (A1) does not hold: If (A1) does not hold, the performance index $E(M)$ cannot be expressed as Eq. (6) and Eq. (7). In this case, if Eq. (5) holds, then we have

$$E(M) \leq \left\| \sum_{k=0}^{\infty} \text{abs}(\tilde{C} \tilde{A}^k \tilde{B}_2) \right\| d_1 \frac{d_1}{2};$$

otherwise Eq. (7). Therefore we cannot directly derive an optimal sigma-delta modulator in a similar way to Section 3.1. However, we can prove that $M^*$ given in Theorem 1 is optimal under (A2) and...
an assumption which is a weaker condition than \((A1)\).

Let \(D_2 \in \mathbb{R}^{m \times p} (i = 1, 2, \ldots, s)\) denote the matrices such that \(D_2 = [D_{21}^\top, D_{22}^\top, \ldots, D_{2s}^\top]^\top\), and we define the set \(I := \{i \in \{1, 2, \ldots, s\} : \tau_i \neq \tau_{\text{max}}\}\) for \(\tau_{\text{max}} := \max_{i \in \{1, 2, \ldots, s\}} \tau_i\) (note that \(\tau_i \in (0) \cup \mathbb{N}\) is the minimum integer satisfying \(C_1 \tilde{A}_\tau B_2i \neq 0\)). In addition, if \(I\) is not empty, we define the matrix \(\hat{D}_2\) composed of the matrices \(D_{2i} (i \in I)\). Then, the optimality of \(M^*\) is proven under \((A2)\) and

\[(A1') \quad \text{rank}([C_2 \hat{D}_2]) = m \quad (\text{full row rank}),\]

and the set \(I\) is empty or the matrix \(\hat{D}_2\) is full row rank.

The above fact is proven as follows. Letting \(Z_{M'}(x_0, R)\) denote the output sequence \(Z_M(x_0, R)\) for \(M := M^*\), the relation

\[
\|\begin{pmatrix} C_1[\tilde{A}^{\tau_1} B_{21} & \tilde{A}^{\tau_2} B_{22} & \cdots & \tilde{A}^{\tau_s} B_{2s}\end{pmatrix} \begin{pmatrix} I_{m_1} & 0 & 0 & 0 \\ d_2 & I_{m_2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ d_s & 0 & \cdots & I_{m_s}\end{pmatrix} \begin{pmatrix} \xi_1(k+1) \\ \xi_2(k+1) \\ \vdots \\ \xi_s(k+1) \end{pmatrix}\| \quad \frac{d_1}{2}
\]

holds under \((A1')\) and \((A2)\) (see Appendix A). This shows that \(M^*\) in Theorem 1 is optimal.

On the other hand, if \((A1')\) is not satisfied, the sigma-delta modulator \(M^*\) in Theorem 1 is not always optimal, while the modulator minimizes the upper bound of \(E(M)\), i.e., the right hand side of Eq. (21). In this sense, \(M^*\) is a practical sigma-delta modulator.

2) Case where the extra term \(D_3v\) exists in \(G\): Some systems cannot be expressed in the form of Eq. (1). For example, when \(K\) is a state feedback controller in Fig. 1 (b), the system \(\Sigma_M\) in Fig. 3 is composed of the linear system with the term \(D_3v\):

\[
\hat{G} : \begin{cases} x(k+1) = Ax(k) + B_1r(k) + B_2v(k), \\ z(k) = C_1x(k) + D_1r(k), \\ u(k) = C_2x(k) + D_2r(k) + D_3v(k), \end{cases}
\]

and the sigma-delta modulator \(M\). In this case, Theorem 1 cannot be directly applied. However, if

\[(A3) \quad D_3^k = 0 \quad (k \geq m)\]

holds, then we can obtain an optimal decentralized sigma-delta modulator under \((A1)\) and \((A2)\). More concretely, it is given by \(M^* := (M_1^*, M_2^*, \ldots, M_s^*)\) composed of

\[
M_i^* : \begin{cases} \xi_i(k+1) = \tilde{A}_i \xi_i(k) - \tilde{B}_i u_i(k) + \tilde{B}_i v_i(k) \\ v_i(k) = Q_i(-C_i \tilde{A}_i \tilde{B}_2i)^r C_i \tilde{A}_i^{\tau_i+1} \xi_i(k) + u_i(k) \end{cases}
\]

where \(\tilde{A} := A + B_2(I - D_3)^{-1} C_2, \tilde{B}_2i \in \mathbb{R}^{n \times m_i} (i = 1, 2, \ldots, s)\) are the matrices such that \(B_2(I - D_3)^{-1} = [\tilde{B}_{21}, \tilde{B}_{22}, \ldots, \tilde{B}_{2s}]\), and \(\tau_i\) is the minimum integer satisfying \(C_i \tilde{A}_i \tilde{B}_2i \neq 0\) (the proof is given in Appendix B). Note under \((A3)\) that the matrix \(I - D_3\) is nonsingular.

4. Experimental evaluation with seesaw-cart system

In this section, we evaluate the real-world performance of the proposed sigma-delta modulators by the experiment using a seesaw-cart system.

4.1 Description of experimental setup

We consider the seesaw-cart system shown in Fig. 9. The system is composed of the cart and the seesaw, and it is \(0.8(L) \times 0.3(W) \times 0.3(H)\) meters long. The mass of the cart and the seesaw are

\[\text{For example, when } I := \{1, 2, \ldots, s - 1\}, \hat{D}_2 \text{ is defined as } \hat{D}_2 := [D_{21}^\top, D_{22}^\top, \ldots, D_{2(s-1)}^\top]^\top.\]
0.57 [kg] and 2.29 [kg], respectively. The cart moves along the seesaw rail by applying voltage onto the DC motor (0.023 [Nm/A]), and the seesaw rotates only in the vertical plane. The position of the cart and the angle of the seesaw are measured by the potentiometers. The computer (using MATLAB Real-time Workshop and Simulink) is connected to the seesaw-cart system for controlling the system. The control objective is to stabilize the system under the constraint that the plant input takes a value on \{-8, 0, +8\} and the controller input does a value on \{0, \pm 0.02, \pm 0.04, \ldots, \pm 0.06\} at each time.

To this end, we construct the quantized feedback system in Fig. 1 (b). By using the Lagrange equations and the linear approximation, we have the following continuous-time model of the seesaw-cart system:

\[
P_c: \begin{cases}
    \dot{x}_p(t) = \\
    z(t) = \\
    y(t) = 
\end{cases} \begin{bmatrix}
    0 & 1.00 & 0 & 0 \\
    -1.92 & -4.15 & 5.38 & 0 \\
    25.63 & 1.27 & 9.91 & 0 \\
    0 & 0 & 0 & 1.00
\end{bmatrix} x_p(t) + \begin{bmatrix}
    0 & 0.51 & 0 \\
    0 & 0 & 0 \\
    0 & -0.15 & 0
\end{bmatrix} v_1(t),
\]

where \( x_p := [\alpha \ \dot{\alpha} \ \theta \ \dot{\theta}]^\top \in \mathbb{R}^4 \) is the state variable, \( \alpha \) is the position of the cart, \( \theta \) is the angle of seesaw, and \( v_1 \) is the voltage applied to the motor as a control input. The controlled output \( z \) is chosen so as to reflect the relative importance of \( \theta \) and \( \dot{\theta} \) compared to \( \alpha \) and \( \dot{\alpha} \) and to satisfy (A2). The measured output \( y \) is composed of \( \alpha \) and \( \theta \). For the discrete-time model of \( P_c \) with the sampling period \( h := 0.01 \) [s], \( K \) is given by the observer based integral servo controller:
be too complex for powerless hardware devices. As a future work, it is expected to develop a design
system in Fig. 1 (b) where the initial states of \( P \) and \( K \) are given by \( x_r(0) := [0 \ 0 \ -0.013 \ 0]^T \) and \( x_k(0) := [0 \ 0 \ 0 \ 0]^T \) and \( r_R(k) \equiv 0 \). In addition, the simulation result of the unquantized feedback
system in Fig. 2 is illustrated by the thin lines in the figures, although the thin lines are invisible by the overlap with the thick lines in the forth and fifth figures. For comparison, the result for the uniform quantizer case is shown in the forth and fifth figures by the dotted lines. From Fig. 10, we have a similar observation to that in Section 3.2.

Figure 11, on the other hand, shows by thick lines the corresponding experimental result in the same fashion. In spite of the very coarse signals as shown in the first, second, and third figures, it can be seen that the position \( \alpha \) and the angle \( \theta \) of the quantized feedback system in Fig. 1 (b) are close to those of the unquantized feedback system in Fig. 2. Moreover, the experimental result agrees with the simulation result in Fig. 10 from a qualitative point of view. It can be concluded that the proposed decentralized sigma-delta modulator would exhibit satisfactory control performance in the real-world system.

However, the remark is that the plant input \( v_1 \) and the controller input \( v_2 \) are oscillating even though the plant output \( y \) is almost converged in the steady state. This is due to the unstable plant. Thus, when the plant is stable, the signals \( v_1 \) and \( v_2 \) do not oscillate. In fact, from Fig. 8 in Section 3.2, we cannot see that the oscillations of the plant input \( v_1 \) and the controller input \( v_2 \) in the steady state, since the plant is stable.

5. Conclusion

This paper has discussed an optimization problem of a class of decentralized sigma-delta modulators for feedback control. We have derived optimal sigma-delta modulators in an analytical way, and also clarified the performance limitation of the decentralized sigma-delta modulators in feedback systems. The result is based on a generalized feedback system representation; thus our framework will be a fundamental tool in decentralized sigma-delta modulator design for various types of quantized feedback systems. Finally, an experimental evaluation has been performed by using a see-saw-cart system. It has been demonstrated that an unstable mechanical system can be satisfactorily controlled under severe quantized signal constraints by the proposed sigma-delta modulator.

The order of the proposed decentralized modulator is equal to \( ns \), i.e., the product of the order of the system \( G \) and the number of the sub-modulators. So the proposed sigma-delta modulator might be too complex for powerless hardware devices. As a future work, it is expected to develop a design
method of fixed-order decentralized sigma-delta modulators.

Since the structure of the sigma-delta modulator considered here is single-loop, the modulator can be extended to a cascade version. The sigma-delta modulators with cascade structure, i.e., the multi-stage noise shaping (MASH) sigma-delta modulators, may have superior performance over the single-loop sigma-delta modulators. Therefore, it is an interesting future work to consider a performance analysis and design problem of MASH sigma-delta modulators. In addition, it is necessary to design a high order sigma-delta modulator for noise robustness of real-world systems. In order to obtain such a modulator, we have to solve a design problem subject to the specification for robustness against noise. This is an open problem.
Fig. 11. Experimental result of the quantized system in Fig. 1 (b) with $M_i := M_i^t$ (thick lines), the unquantized system in Fig. 2 (thin line) and the quantized system in Fig. 1 (b) with the uniform quantizer $M_i := Q_i$ (dotted line).

Appendix

A. Proof for case where (A1’) holds instead of (A1)

Equation (22) is proven as follows.

First equality: Without loss of generality, it is proven by considering $s$ cases:

1) $\tau_1 \leq \tau_2 \leq \cdots \leq \tau_{s-1} < \tau_s$,
2) $\tau_1 \leq \tau_2 \leq \cdots < \tau_{s-1} = \tau_s$,
3) $\tau_1 \leq \tau_2 \leq \cdots = \tau_{s-1} = \tau_s$,

$\vdots$
8) $\tau_1 = \tau_2 = \cdots = \tau_{s-1} = \tau_s$. 

399
We first prove the case that 1) \( \tau_1 \leq \tau_2 \leq \cdots \leq \tau_{s-1} < \tau_s \), where \( \tau_{\text{max}} = \tau_s \) and the matrix \( D_{2s} \in \mathbb{R}^{m_s \times p} \) is not full row rank. We partition \( w(k) \) in Eq. (9) as \( [w_1^\top(k) \ w_2^\top(k) \ \cdots \ w_s^\top(k)]^\top \) where \( w_i(k) \in \mathbb{R}^{m_i} \). Then, using Eq. (10) and Eq. (16), we have the following relation under Eq. (5).

\[
\|z_M(\tau_s+1, x_0, R) - z(\tau_s+1, x_0, R)\| = \left\| \sum_{k=0}^{\tau_s} \tilde{C} \tilde{A}^\tau s-k [ \tilde{B}_21 \ \tilde{B}_{22} \ \cdots \ \tilde{B}_{2s} ] w(k) \right\|
\]

\[
= \left\| \sum_{k=0}^{\tau_s} \tilde{C} \tilde{A}^\tau s-k \tilde{B}_21 \ \sum_{k=0}^{\tau_s} \tilde{C} \tilde{A}^\tau s-k \tilde{B}_{22} \ \cdots \ \sum_{k=0}^{\tau_s} \tilde{C} \tilde{A}^\tau s-k \tilde{B}_{2s} \right\| w(k)
\]

\[
= \left\| C_1 \tilde{A}^\tau \tilde{B}_{21} w_1(\tau_s - \tau_1) + C_1 \tilde{A}^\tau \tilde{B}_{22} \frac{d_2}{d_1} w_2(\tau_s - \tau_2) + \cdots + C_1 \tilde{A}^\tau \tilde{B}_{2s} \frac{d_s}{d_1} w_s(0)
\]

\[
+ \sum_{k=0}^{\tau_s-\tau_1-1} \tilde{C} \tilde{A}^\tau s-k \tilde{B}_21 w_1(k) + \sum_{k=0}^{\tau_s-\tau_2-1} \tilde{C} \tilde{A}^\tau s-k \tilde{B}_{22} w_2(k) + \cdots + \sum_{k=0}^{\tau_s-\tau_{s-1}-1} \tilde{C} \tilde{A}^\tau s-k \tilde{B}_{2(s-1)} w_{s-1}(k) \right\|
\]

\[ (A-1) \]

Here, let \( H := [C_1 \tilde{A}^\tau \tilde{B}_{21} \ C_1 \tilde{A}^\tau \tilde{B}_{22} (d_2/d_1) \cdots \ C_1 \tilde{A}^\tau \tilde{B}_{2s} (d_s/d_1)] \) and let \( H_{ij} \) and \( \langle H \rangle_i \) denote the \((i, j)\)-th element and the \(i\)-th row vector of \( H \). In addition, we define \( \Psi \in \mathbb{R}^{(m-m_s) \times m} \) and \( \Phi_k \in \mathbb{R}^{m \times m} \) as

\[
\Psi := \begin{bmatrix} I & 0 & \cdots & 0 \\ 0 & I & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I \end{bmatrix}, \quad \Phi_k := \begin{bmatrix} \phi_{1k} & 0 & \cdots & 0 \\ 0 & \phi_{2k} & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \phi_{sk} \end{bmatrix}
\] \[ (A-2) \]

where

\[
\phi_{ik} := \begin{cases} I & \text{if} \ k = \tau_s - \tau_i, \\ 0 & \text{otherwise}. \end{cases}
\] \[ (A-3) \]

Then, we consider the initial state \( x_0 \in \mathbb{R}^n \) and the external input \( r_k \in \mathbb{R}^p \) \((k = 0, 1, \ldots)\)

\[
\begin{bmatrix} x_0 \\ r_0 \end{bmatrix} := -[C_2 \ D_2]^\top \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \cdots, \frac{d_s}{d_1} I_{m_s} \right) \Phi_0 \text{sign}(\langle H \rangle_i)^\top \left( \frac{d_1}{2} - \delta \right),
\] \[ (A-4) \]

\[
r_k := \dot{D}_s^k \Psi \left( - \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \cdots, \frac{d_s}{d_1} I_{m_s} \right) \Phi_k \text{sign}(\langle H \rangle_i)^\top \left( \frac{d_1}{2} - \delta \right) \\
+ Q[\xi \xi(k) + C_2 x(k)] - (\xi \xi(k) + C_2 x(k)) \right)
\] \[ (A-5) \]

for \( i^* := \arg \max_{i \in \{1, 2, \ldots, m\}} \sum_{j=1}^m |H_{ij}| \) and an arbitrarily given small number \( \delta \in (0, d_1/2) \). For Eq. (A-4) and Eq. (A-5), under \((\mathbf{A1})^*\), we have the following relations from Eq. (1) and Eq. (9).

\[
\begin{bmatrix} w_1(\tau_s - \tau_1) \\ w_2(\tau_s - \tau_2) \\ \vdots \\ w_s(0) \end{bmatrix} = \text{sign}(\langle H \rangle_i)^\top \left( \frac{d_1}{2} - \delta \right)
\] \[ (A-6) \]

and

\[
w_i(k) = 0 \quad (k \neq \tau_s - \tau_i, \ i = 1, 2, \ldots, s-1),
\] \[ (A-7) \]

which are derived by using the relations \([C_2 \ D_2]^\top := [C_2 \ D_2]^\top ([C_2 \ D_2][C_2 \ D_2]^\top)^{-1} \) and \( \dot{D}_s^k := \dot{D}_s^k (\dot{D}_s^k \dot{D}_s^k)^{-1} \). Therefore, by using Lemma 1 in [17], we can obtain
\[\|z_M(\tau_s + 1, x_0, R) - z(\tau_s + 1, x_0, R)\| \]
\[= \left\| \text{abs} \left( C_1 \tilde{A}^{\tau_s} B_2 \tilde{A}^{\tau_2} B_{22} \cdots \tilde{A}^{\tau_s} B_{2s} \right) \right\| \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) \left( \frac{d_1}{2} - \delta \right) \]

(A-8)

from Eq. (A-1), Eq. (A-6), and Eq. (A-7), which leads to the first equality of Eq. (22). In a similar way to the above proof, the other \(s - 1\) cases are proven.

**Second and third inequalities:** Obvious.

**Fourth inequality:** In a similar way to the derivation of Eq. (A-1), the right hand side of Eq. (21) is rewritten as

\[
\left\| \sum_{k=0}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_2 \right) \right\| \frac{d_1}{2} = \left\| \sum_{k=0}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{21} \tilde{B}_{22} \cdots \tilde{B}_{2s} \right) \right\| \frac{d_1}{2}
\]

\[
= \left\| \left[ \sum_{k=0}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{21} \right) \right] \sum_{k=0}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{22} \right) \cdots \sum_{k=0}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{2s} \right) \right\| \frac{d_1}{2}
\]

\[
= \left\| \left[ \text{abs} \left( C_1 \tilde{A}^{\tau_s} B_{21} \right) \right] \text{abs} \left( C_1 \tilde{A}^{\tau_s} B_{22} \right) \cdots \text{abs} \left( C_1 \tilde{A}^{\tau_s} B_{2s} \right) \right\| \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right)
\]

\[
+ \left[ \sum_{k=\tau_s+1}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{21} \right) \right] \sum_{k=\tau_s+1}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{22} \right) \cdots \sum_{k=\tau_s+1}^{\infty} \text{abs} \left( \tilde{C} \tilde{A}^k \tilde{B}_{2s} \right) \right\| \frac{d_1}{2}.
\]

Then, under (A2), the second term in the norm is equal to zero since Eq. (18) holds for \(M^*\). This leads to the fourth inequality.

**B. Proof for case where the term \(D_3 v(k)\) exists**

Letting \(\xi := [\xi_1^T \xi_2^T \cdots \xi_s^T]^T, u := [u_1^T u_2^T \cdots u_s^T]^T, v := [v_1^T v_2^T \cdots v_s^T]^T\), and \(v := [v_1^T v_2^T \cdots v_s^T]^T\), we have

\[v(k) = \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) Q \left[ \text{diag} \left( I_{m_1}, \frac{d_1}{d_2} I_{m_2}, \ldots, \frac{d_1}{d_s} I_{m_s} \right) (\xi_k + u(k)) \right].\]

(B-1)

From this and Eq. (9), the following relation can be obtained.

\[v(k) = \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) \left\{ w(k) + \text{diag} \left( I_{m_1}, \frac{d_1}{d_2} I_{m_2}, \ldots, \frac{d_1}{d_s} I_{m_s} \right) (\xi_k + u(k)) \right\}
\]

\[= \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) w(k) + \xi_k + u(k)\]

(B-2)

We substitute Eq. (B-2) into the third equation of (23), and then we have

\[u(k) = (I - D_3)^{-1} \left( C_2 x(k) + D_2 r(k) + D_3 \xi(k) + D_3 \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \ldots, \frac{d_s}{d_1} I_{m_s} \right) w(k) \right).\]

(B-3)

Note under (A3) that the matrix \((I - D_3)\) is nonsingular and \(I + (I - D_3)^{-1} D_3 = (I - D_3)^{-1}\). Here, we define

\[
\tilde{A} := \begin{bmatrix} \tilde{A} & B_2 \frac{(I - D_3)^{-1} C}{} \\ 0 & \tilde{A} + B_2 C \end{bmatrix}, \quad \tilde{B}_1 := \begin{bmatrix} 0 \\ \tilde{B}_1 + B_2 \end{bmatrix},
\]

\[
\tilde{B}_2 := B_2 \frac{(I - D_3)^{-1}}{}, \quad \text{diag} \left( I_{m_1}, \frac{d_2}{d_1} I_{m_2}, \frac{d_3}{d_1} I_{m_3}, \ldots, \frac{d_s}{d_1} I_{m_s} \right),
\]

(B-4)

Then, if the relation Eq. (5) with Eq. (B-4) holds, the output difference between the systems \(\Sigma_M\) and \(\Sigma\) is given by Eq. (10) with Eq. (B-4), which leads to a similar result to Lemma 1. Indeed, the result
can be derived by replacing $\mathcal{B}_2$ in Lemma 1 with $\mathcal{B}_2(I - D_3)^{-1}$. Thus, in a similar way to Section 3.1, we can obtain an optimal decentralized sigma-delta modulator in form of Eq. (24).

In the rest of this section, we prove that the exact expression of $E(M)$ is given by Eq. (6) with Eq. (B-4). First, let $H_k := C\tilde{A}^{(T-1)-k}\tilde{B}_2$ for $T \in \mathbb{N}$, and we define

$$
\Gamma := \text{diag} \left( I_{m_1}, \frac{d_1}{d_2} I_{m_2}, \frac{d_1}{d_3} I_{m_3}, \ldots, \frac{d_1}{d_n} I_{m_n} \right). \quad (B-5)
$$

Under (A1), we consider the external input

$$
r(k) = r_k := D_2^\dagger \left\{ -(I - D_3)\Gamma^{-1} \text{sign}(H_k,i^*) \left( \frac{d_1}{2} - \delta \right) - D_3\Gamma^{-1} \text{sign}(H_k,i^*)^T \left( \frac{d_1}{2} - \delta \right) \right. \\
+ (I - D_3)\Gamma^{-1} \left( Q \left[ \Gamma(I - D_3)^{-1}(\xi(k) + C_2x(k)) \right] - \Gamma(I - D_3)^{-1}(\xi(k) + C_2x(k)) \right) \right\} 
$$

for $i^* := \arg\max_{i \in \{1,2,\ldots,m\}} \sum_{j=1}^{m} \sum_{k=0}^{T-1} |H_{ij}^{(k)}|$ and an arbitrarily given small number $\delta \in (0, d_1/2)$. Note that $R \in \ell_\infty^n$ is defined by $r_k$ in Eq. (B-6). For this input, we obtain

$$
w(k) = Q[\Gamma(\xi(k) + u(k))] - \Gamma(\xi(k) + u(k)) \\
= Q \left[ \Gamma(I - D_3)^{-1}(\xi(k) + C_2x(k) + D_2r(k) + D_3\Gamma^{-1}w(k)) \right] \\
- \Gamma(I - D_3)^{-1}\left( \xi(k) + C_2x(k) + D_2r(k) + D_3\Gamma^{-1}w(k) \right) \\
= \text{sign}(H_k,i^*)^T \left( \frac{d_1}{2} - \delta \right) \quad (B-7)
$$

from Eq. (9) and Eq. (B-3). Thus it follows from Eq. (10) and Lemma 1 in [17] that

$$
\|z_M(T,x_0,R) - z(T,x_0,R)\| = \left\| \sum_{k=0}^{T-1} \text{abs}(\tilde{C}\tilde{A}^{(T-1)-k}\tilde{B}_2) \left( \frac{d_1}{2} - \delta \right) \right\| 
$$

for $R \in \ell_\infty^n$. On the other hand, according to Eq. (9) and Eq. (10),

$$
\|z_M(T,x_0,R) - z(T,x_0,R)\| \leq \left\| \sum_{k=0}^{T-1} \text{abs}(\tilde{C}\tilde{A}^{(T-1)-k}\tilde{B}_2) \frac{d_1}{2} \right\| 
$$

holds for every $(x_0,R) \in \mathbb{R}^n \times \ell_\infty^n$ and $T \in \mathbb{N}$. From this and Eq. (B-8), we obtain the following relation.

$$
\sup_{(x_0,R) \in \mathbb{R}^n \times \ell_\infty^n} \|z_M(T,x_0,R) - z(T,x_0,R)\| = \left\| \sum_{k=0}^{T-1} \text{abs}(\tilde{C}\tilde{A}^{(T-1)-k}\tilde{B}_2) \left( \frac{d_1}{2} \right) \right\|. \quad (B-10)
$$

Therefore, we have Eq. (6) with Eq. (B-4) for the system $\tilde{G}$ in Eq. (23).

**C. The optimal decentralized sigma-delta modulator in Section 4.2**

For the simulation and experiment in Section 4.2, the following optimal decentralized sigma-delta modulator $M^* = (M_1^*, M_2^*)$ is used.
\[
A^*_1 := \begin{bmatrix}
0.9999 & 0.0098 & 0.0002 & 0 & -0.0083 & -0.0011 & -0.0062 & -0.0016 & 0.0014 \\
-0.0187 & 0.9592 & 0.0527 & 0.0002 & -1.659 & -0.2213 & -1.228 & -0.3121 & 0.2889 \\
0.00128 & 0 & 1 & 0.01 & 0.0025 & 0.0003 & 0.0019 & 0.0005 & -0.0004 \\
0.2563 & 0.0137 & 0.0995 & 1 & 0.5076 & 0.0677 & 0.3758 & 0.0955 & -0.0884 \\
0.764 & 0 & -0.451 & 0 & -2.371 & 0.741 & -0.664 & -0.296 & 0 \\
0.021 & 0 & 0.264 & 0 & -0.017 & 0 & 0.737 & 0.009 & 0 \\
0.821 & 0 & 2.416 & 0 & -0.079 & 0.074 & -1.959 & 1.079 & 0 \\
0 & 0 & -0.01 & 0 & 0 & 0 & 0 & 0.99 & \\
\end{bmatrix},
\]

\[
M^*_1 := \begin{bmatrix}
0 & 0.005 & 0 & -0.0015 & 0 & 0 & 0 & 0 & 0 \\
0.005 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0.0015 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix},
\]

\[
B^*_{11} := -\begin{bmatrix}
0 \\
0.005 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\end{bmatrix}, \quad B^*_{21} := \begin{bmatrix}
0.005 \\
0 \\
-0.0015 \\
0 \\
0 \\
0 \\
0 \\
0 \\
0 \\
\end{bmatrix},
\]

\[
C^*_1 := \begin{bmatrix}
-1030 \\
-619.2 \\
-1545 \\
-1390 \\
330.6 \\
44.1 \\
244.8 \\
62.2 \\
-57.59 \\
\end{bmatrix},
\]

\[
A^*_2 := \begin{bmatrix}
0.9999 & 0.0098 & 0.0002 & 0 & -0.0083 & -0.0011 & -0.0062 & -0.0016 & 0.0014 \\
-0.0187 & 0.9592 & 0.0527 & 0.0002 & -1.659 & -0.2213 & -1.228 & -0.3121 & 0.2889 \\
0.00128 & 0 & 1 & 0.01 & 0.0025 & 0.0003 & 0.0019 & 0.0005 & -0.0004 \\
0.2563 & 0.0137 & 0.0995 & 1 & 0.5076 & 0.0677 & 0.3758 & 0.0955 & -0.0884 \\
0.764 & 0 & -0.451 & 0 & -2.371 & 0.741 & -0.664 & -0.296 & 0 \\
0.021 & 0 & 0.264 & 0 & -0.017 & 0 & 0.737 & 0.009 & 0 \\
0.821 & 0 & 2.416 & 0 & -0.079 & 0.074 & -1.959 & 1.079 & 0 \\
0 & 0 & -0.01 & 0 & 0 & 0 & 0 & 0 & 0.99 \\
\end{bmatrix},
\]

\[
M^*_2 := \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix},
\]

\[
B^*_{12} := \begin{bmatrix}
0.222 & 0.003 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0.764 & -0.451 \\
0.021 & 0.264 \\
0.821 & 2.416 \\
0 & -0.01 \\
\end{bmatrix}, \quad B^*_{22} := \begin{bmatrix}
0.222 & 0.003 \\
0.764 & -0.451 \\
0.021 & 0.264 \\
0.821 & 2.416 \\
0 & -0.01 \\
\end{bmatrix},
\]

\[
C^*_2 := \begin{bmatrix}
3.033 & 1.558 & 3.802 & 3.515 & -1.168 & -0.2079 & -0.6709 & -0.295 & 0.2847 \\
3.651 & 1.876 & 4.577 & 4.232 & -1.406 & -0.2502 & -0.8077 & -0.3552 & 0.3428 \\
\end{bmatrix}.
\]

References


