Studies on the Synthesis of Complete Finite-Settling-Time Systems under the Minimum Integrated-Square Sampled Error and the Minimum Quadratic Control Area Criteria

By

Yoshikazu SAWARAGI* and Hiroshi FUKAWA*

(Received July 31, 1962)

Finite-settling-time systems which are non-minimal have some constants which can be arbitrarily assigned. This opens up the possibility of selecting extra system constants with a view to optimizing the system under some criterion. An approach to this problem is to minimize the integrated-square sampled error, and another is to minimize the quadratic control area. This paper is concerned with the problem of the synthesis of non-minimal finite-settling-time systems, which have no ripples between sampling instants, under these two criteria. The treatment of these optimization problems makes use of the z transform and the modified z transform methods respectively. By representing polynomials by vectors or matrices, and by making use of the linear algebra techniques, these two synthetic studies are carried out in a systematic manner. Fundamental design equations of the optimum system, and evaluating formulas of the minimum integrated-square sampled error and the minimum quadratic control area are derived.

1. Introduction

In the synthesis of finite-settling-time systems, it is usual to design a system consistent with the requirement that it be stable and that it have the shortest settling-time. When so designed, it is called a minimal system. In such a system, system parameters are determined uniquely, so the system has no parameter to be adjusted at all. In consequence, a minimal system is not always the optimum one under some other criterion. Here if some additional constants are used and the settling-time is made a little longer, the system becomes non-minimal and gets to have some constants which can be arbitrarily assigned. This opens up the possibility of selecting the extra system constants with a view to optimizing the system under some criterion. From this point of view, authors reported studies on the synthesis of this kind of problem in another paper[1]. However the system treated in it was one which contained ripples between sampling

* Department of Applied Mathematics and Physics

instants. In this paper the same kind of problem will be inquired into further. That is, this paper is concerned with the problem of synthesis of non-minimal finite-settling-time systems which contain no ripples after a suitable short transient period has elapsed. The word 'complete' given in the title means 'ripple-free' after systems once settled to the desired value.

In general, the z transform and the modified z transform of the error of complete finite-settling-time systems are both polynomials in z^{-1} so that

$$E(z) = e_0+e_1z^{-1}+\cdots+e_{N-1}z^{-N+1} \tag{1}$$

$$E(z, m) = e_1(m)z^{-1}+e_2(m)z^{-2}+\cdots+e_N(m)z^{-N} \tag{2}$$

where m is a real number and $1\geq m\geq 0$. By making use of vectors $\boldsymbol{e}=(e_0, e_1, \cdots, e_{N-1})$ and $\boldsymbol{e}(m)=(e_1(m), e_2(m), \cdots, e_N(m))$, the integrated-square sampled error and the quadratic control area are expressed as follows:

$$\sum_{n=0}^{\infty} e_n^2 = (\boldsymbol{e}, \boldsymbol{e}) \tag{3}$$

and

$$\int_0^{\infty} \{e(t)\}^2 dt = \int_0^1 (\boldsymbol{e}(m), \boldsymbol{e}(m))dm \tag{4}$$

respectively, where the round brackets denote the inner product of vectors $\boldsymbol{e}$ or $\boldsymbol{e}(m)$. The form of Eqs. (3) and (4) imply the possibility of employment of the linear algebra techniques, with which synthetic studies will be carried out readily and systematically in this report.

2. Mathematical Preliminaries

Prior to main discussion it seems to be necessary to explain new mathematical methods employed in this paper. The methods are to make operations of polynomials correspond to those of vectors or matrices.

2.1 Reprecentation of polynomials by vectors or matrices

Representing a polynomial by a vector or a matrix whose elements are coefficients of the polynomial, we can put the four rules of arithmetic for polynomials in correspondence with the operation of vectors or matrices as mentioned in the following subsections. The correspondence will be denoted hereafter by the notation '$\Leftrightarrow$'.

1) Addition and subtraction

Let

$$f(z) = a_0+a_1z^{-1}+\cdots+a_nz^{-n} \Leftrightarrow \boldsymbol{f} = (a_i) \quad (n\geq i\geq 0), \tag{5}$$

$$g(z) = b_0+b_1z^{-1}+\cdots+b_mz^{-m} \Leftrightarrow \boldsymbol{g} = (b_i) \quad (m\geq i\geq 0), \tag{6}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are both column vectors. Then the following correspondences may be readily verified

$$f(z)\pm g(z) \Leftrightarrow \boldsymbol{f}\pm\boldsymbol{g} = (a_i+b_i) \quad (\max(m, n)\geq i\geq 0)\,. \tag{7}$$

2) Multiplication

Suppose that $g(z)\times f(z)$ makes $h(z)$ which is a polynomial of the degree $m+n$ so that

$$h(z) = c_0+c_1z^{-1}+c_2z^{-2}+\cdots+c_{m+n}z^{-(m+n)} \tag{8}$$

and let $\boldsymbol{h}=(c_i)$, $(m+n\geq i\geq 0)$ denote the column vector corresponding to $h(z)$. Then we can regard the relation $h(z)=g(z)f(z)=f(z)g(z)$ as a transformation from a vector $\boldsymbol{f}$ to a vector $\boldsymbol{h}$ by G, or a transformation from a vector $\boldsymbol{g}$ to a vector $\boldsymbol{h}$ by F, namely:

$$h(z) = g(z)f(z) \Leftrightarrow \boldsymbol{h} = G\cdot\boldsymbol{f} = F\cdot\boldsymbol{g} \tag{9}$$

where the two transformations G and F represent the following matrices

$$f(z) \Leftrightarrow F = \begin{Vmatrix} a_0 & & & \\ a_1 & a_0 & & \\ \vdots & a_1 & \ddots & \\ a_n & \vdots & \ddots & a_0 \\ & a_n & & a_1 \\ & & & \vdots \\ & & & a_n \end{Vmatrix}, \quad g(z) \Leftrightarrow G = \begin{Vmatrix} b_0 & & & \\ b_1 & b_0 & & \\ \vdots & b_1 & \ddots & \\ b_m & \vdots & \ddots & b_0 \\ & b_m & & b_1 \\ & & & \vdots \\ & & & b_m \end{Vmatrix}. \tag{10}$$

And F and G, respectively, are $(m+n+1)\times(m+1)$ and $(m+n+1)\times(n+1)$ arrays, and all other elements than a's or b's are zero. Thus the multiplication of a polynomial by another is found equivalent to the linear transformation of the vector corresponding to one polynomial by the matrix corresponding to the other.

3) Division

In the first place, assume that $h(z)$ is divisible by $g(z)$ and the quotient is $f(z)$, then

$$f(z) = \frac{h(z)}{g(z)}\,. \tag{11}$$

Multiplying both sides of (11) by $g(z)$, there results $h(z)=g(z)f(z)$ which is expressed by $\boldsymbol{h}=G\cdot\boldsymbol{f}$. Here if there exists a $(n+1)\times(m+n+1)$ matrix $\tilde{G}$ such that

$$\tilde{G}\cdot G = E \tag{12}$$

where E is the unit matrix of the order $n+1$, multiplying both sides of $\boldsymbol{h}=G\cdot\boldsymbol{f}$ by $\tilde{G}$ leads to the following expression

$$f(z) = \frac{1}{g(z)} \cdot h(z) \quad \Leftrightarrow \quad \boldsymbol{f} = \tilde{G} \cdot \boldsymbol{h} . \tag{13}$$

In the next place, consider the case that $h(z)$ is undivisible by $g(z)$. Assume that the expansion of $1/g(z)$ into an infinite series is

$$\frac{1}{g(z)} = d_0 + d_1 z^{-1} + d_2 z^{-2} + \cdots . \tag{14}$$

Then the right hand side of (14) may be represented by the same kind of matrix G^* as (10). Needless to say, the number of rows of G^* is infinite. The first $n+1$ rows of G^* is equal to G.

For example, let

$$a_0 + a_1 z^{-1} = \frac{c_0 + c_1 z^{-1} + c_2 z^{-2}}{1 - \beta z^{-1}} \tag{15}$$

or

$$c_0 + c_1 z^{-1} + c_2 z^{-2} = (1 - \beta z^{-1})(a_0 + a_1 z^{-1}) . \tag{16}$$

Then, according to the above discussions, (16) may be expressed by using matrix notation, namely:

$$\begin{Vmatrix} c_0 \\ c_1 \\ c_2 \end{Vmatrix} = G \cdot \begin{Vmatrix} a_0 \\ a_1 \end{Vmatrix} \tag{17}$$

$$1 - \beta z^{-1} \quad \Leftrightarrow \quad G = \begin{Vmatrix} 1 & 0 \\ -\beta & 1 \\ 0 & -\beta \end{Vmatrix} . \tag{18}$$

Hereupon

$$\begin{Vmatrix} 1 & 0 & 0 \\ \beta & 1 & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 1 & 0 \\ -\beta & 1 \\ 0 & -\beta \end{Vmatrix} = \begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} , \tag{19}$$

hence

$$\frac{1}{1 - \beta z^{-1}} \quad \Leftrightarrow \quad \tilde{G} = \begin{Vmatrix} 1 & 0 & 0 \\ \beta & 1 & 0 \end{Vmatrix} \tag{20}$$

and

$$\begin{Vmatrix} a_0 \\ a_1 \end{Vmatrix} = \tilde{G} \cdot \begin{Vmatrix} c_0 \\ c_1 \\ c_2 \end{Vmatrix} \tag{21}$$

which is the expression of (15) by means of matrix notation. If $c_0 + c_1 z^{-1} + c_2 z^{-2}$ is undivisible by $1 - \beta z^{-1}$, the expansion of $1/(1 - \beta z^{-1})$ into $1 + \beta z^{-1} + \beta^2 z^{-2} + \cdots$ gives

$$\begin{Vmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ \vdots \end{Vmatrix} = G^* \cdot \begin{Vmatrix} c_0 \\ c_1 \\ c_2 \end{Vmatrix} \tag{22}$$

where

$$G^* = \begin{Vmatrix} 1 & 0 & 0 \\ \beta & 1 & 0 \\ \cdots & \cdots & \cdots \\ \beta^2 & \beta & 1 \\ \beta^3 & \beta^2 & \beta \\ \vdots & \vdots & \vdots \end{Vmatrix}. \tag{23}$$

A glance at (23) shows that the first two rows of G^* is equal to $\tilde{G}$. In divisible cases, the condition that $c_0+c_1z^{-1}+c_2z^{-2}$ is divisible by $1-\beta z^{-1}$ is given by $c_0\beta^2+c_1\beta+c_2=0$, so we have $a_2=a_3=\cdots=0$ in (22).

2.2 Structure of transformation matrices

Studies on the structure of the transformation matrix G defined in 2.1 lead one to find $\tilde{G}$. By making use of matrices of the following form

$$R_i = \underbrace{\left\| \begin{matrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{matrix} \right\|}_{n+1} \left.\vphantom{\begin{matrix} a \\ b \\ c \\ d \end{matrix}}\right\} \begin{matrix} i+1 \\ \\ \\ m+n+1 \end{matrix} \tag{24}$$

where every element at the position $(i+k, k)$ $(n+1\geq k\geq 1)$ is unity, and zeros elsewhere, the transformation matrix G in (10) can be expressed as follows:

$$G = \begin{Vmatrix} b_0 & & & \\ b_1 & b_0 & & \\ \vdots & b_1 & \ddots & \\ b_m & \vdots & \ddots & b_0 \\ & b_m & & b_1 \\ & & \ddots & \vdots \\ & & & b_m \end{Vmatrix} = b_0R_0+b_1R_1+\cdots b_mR_m. \tag{25}$$

Now, consider a nilpotent matrix N of the order $m+n+1$ such that

$$N = \begin{Vmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{Vmatrix}, \quad N^{m+n} \neq 0, \quad N^{m+n+1} = 0 \tag{26}$$

where every underdiagonal element is unity, and all other elements are zero. Then

$$R_k = N^k \cdot R_0 ,$$

hence

$$G = (b_0E + b_1N + \cdots + b_mN^m)R_0 \tag{27}$$

where E is the unit matrix of the order $m+n+1$. Now let

$$G' = b_0E + b_1N + \cdots + b_mN^m . \tag{28}$$

If there exists G'^{-1}, the inverse of G', it is easy to show that

$$\tilde{G} = {}^tR_0G'^{-1} \tag{29}$$

G'^{-1} may be found without any difficulty by the use of the fact that N is a nilpotent matrix as shown in (26).

To illustrate with a practical example, consider the case $g(z)=1-\beta z^{-1}$. Then

$$g(z) = 1-\beta z^{-1} \Leftrightarrow G' = E-\beta N$$

$$\frac{1}{g(z)} = 1+\beta z^{-1}+\beta^2 z^{-1}+\cdots \Leftrightarrow G'^{-1} = E+\beta N+\beta^2N^2+\cdots+\beta^{m+n}N^{m+n},$$

and evidently

$$G'^{-1}G' = E-\alpha^{m+n+1}N^{m+n+1} = E$$

the last equality follows (26). Thus, by replacement of z^{-1} by N we can obtain G', and from the expansion of $1/g(z)$ G'^{-1} can also be obtained with the same replacement.

3. Synthesis under the Minimum Integrated-Square Sampled Error Criterion

3.1 Error vector

Consider the sampled-data system with a series compensator $C(z)$ shown in Fig. 1. For simplicity, the synthetic study will be carried out on the assumption

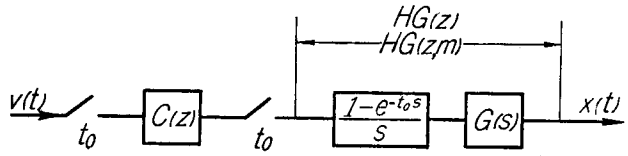

Fig. 1. Sampled-data system with series compensator $C(z)$.

that an input is a step function. To the synthesis for other inputs the procedure mentioned below can be applied in the same manner.

The z transform of the error of the system subjected to the unit step input

is given by

$$E(z) = \frac{1-C(z)HG(z)}{1-z^{-1}} \tag{30}$$

where $HG(z)$ is the pulse transfer function of a controlled system with a data hold. In order to make the output settle completely to the desired value after N sampling periods have elapsed, $C(z)$ must be a polynomial in z^{-1} of the degree N and have the denominator of $HG(z)$ as a factor. Hence let

$$HG(z) = \frac{G_1(z)}{G_2(z)} \quad \text{and} \quad C(z) = D(z)G_2(z) \tag{31}$$

where $G_1(z)$ and $G_2(z)$ are both polynomials in z^{-1} of the degree n, and $D(z)$ is a polynomial in z^{-1} of the degree $N-n$. Substitution of (31) in (30) gives

$$E(z) = \frac{1-G_1(z)D(z)}{1-z^{-1}} \tag{32}$$

where

$$\begin{aligned} D(z) &= a_0+a_1z^{-1}+\cdots+a_{N-n}z^{-N+n} \\ G_1(z) &= g_0+g_1z^{-1}+\cdots+g_nz^{-n}\,. \end{aligned} \tag{33}$$

In the present instance, the right-hand side of (32) is divisible. Then $E(z)$ becomes a polynomial of the degree $N-1$ because the degree of $G_1(z)D(z)$ is N. According to the discussion of the previous section, make $E(z)$, 1 and $D(z)$ correspond to the following vectors

$$\begin{aligned} E(z) &\Leftrightarrow \boldsymbol{e}: \ N \text{ dimensional vector} \\ 1 &\Leftrightarrow \boldsymbol{u}: \ N+1 \text{ dimentional vector} \\ D(z) &\Leftrightarrow \boldsymbol{a}: \ N-n+1 \text{ dimentional vector} \end{aligned} \tag{34}$$

and make $G_1(z)$, $1-z^{-1}$ and $1/(1-z^{-1})$ correspond to matrices as follows:

$$\begin{aligned} G_1(z) &\Leftrightarrow G: \ (N+1)\times(N-n+1) \text{ matrix} \\ 1-z^{-1} &\Leftrightarrow V: \ (N+1)\times N \text{ matrix} \\ 1/(1-z^{-1}) &\Leftrightarrow \tilde{V}: \ N\times(N+1) \text{ matrix}. \end{aligned} \tag{35}$$

Then it follows that (32) can be represented in the following vector form

$$\boldsymbol{e} = \tilde{V}(\boldsymbol{u}-G\boldsymbol{a}) = \tilde{V}\left(\begin{Vmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{Vmatrix} - G \begin{Vmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-n} \end{Vmatrix}\right) = \boldsymbol{v}-B\cdot\boldsymbol{a} \tag{36}$$

where

$$G = \overbrace{\begin{Vmatrix} g_0 & & & \\ g_1 & g_0 & & \\ \vdots & g_1 & \ddots & \\ g_n & \vdots & \ddots & g_0 \\ & g_n & & g_1 \\ & & \ddots & \vdots \\ & & & g_n \end{Vmatrix}}^{N-n+1} \Bigg\} N+1, \quad \tilde{V} = \overbrace{\begin{Vmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & \cdots & 1 & 0 \end{Vmatrix}}^{N+1} \Bigg\} N \tag{37}$$

$$\boldsymbol{v} = \tilde{V}\boldsymbol{u} \quad \text{and} \quad B = \tilde{V}G .$$

Expression (36) is the desired error vector every component of which is equal to the error at the sampling instant.

3.2 Derivation of design equation

As described in section 1, the integrated-square sampled error is given by the inner product of (36). Then

$$\begin{aligned} I[\boldsymbol{a}] = \sum_{n=0}^{\infty} e_n^2 &= (\boldsymbol{e}, \boldsymbol{e}) = (\boldsymbol{v}-B\boldsymbol{a}, \boldsymbol{v}-B\boldsymbol{a}) \\ &= |\boldsymbol{v}|^2 - 2(B\boldsymbol{a}, \boldsymbol{v}) + ({}^tBB\boldsymbol{a}, \boldsymbol{a}) \end{aligned} \tag{38}$$

where ${}^tB$ is the transpose of B.

On the other hand, conditions for obtaining complete finite-settling-time response can be formulated as follows:[2]

$$\sum_{i=0}^{N-n+1} f_{ij}a_i - d_j , \qquad (j = 1, 2, \cdots, l) \tag{39}$$

or in terms of vector notation

$$F\boldsymbol{a} - \boldsymbol{d} = \boldsymbol{0} . \tag{40}$$

Since conditions that $C(z)$ contains the denominator of $HG(z)$ are excluded from (39) or (40), the number of conditions for step response synthesis is unity. But in expectation of the extention to the other higher order inputs synthesis, we assume that the number of them is l. Hereupon our problem can be described mathematically as follows: "To obtain such a vector that satisfies the condition (40) and gives the minimum value of (38)", because if such a vector $\bar{\boldsymbol{a}}$ is found the optimum pulse transfer function of a controller can be determined by the use of (31).

Multiplying (40) by a Lagrange's multiplier $2\boldsymbol{\lambda} = (2\lambda_i)$ $(l \geq i \geq 1)$ and adding to (38), there results

$$J[\boldsymbol{a}, \boldsymbol{\lambda}] = |\boldsymbol{v}|^2 - 2(B\boldsymbol{a}, \boldsymbol{v}) + ({}^tBB\boldsymbol{a}, \boldsymbol{a}) + (F\boldsymbol{a}-\boldsymbol{d}, 2\boldsymbol{\lambda}) . \tag{41}$$

The condition that (41) have an extremum is then expressed in the form

$$\delta J[\boldsymbol{a}, \boldsymbol{\lambda}] = 2({}^tBB\boldsymbol{a} - {}^tB\boldsymbol{v} + {}^tF\boldsymbol{\lambda}, \delta\boldsymbol{a}) + 2(F\boldsymbol{a} - \boldsymbol{d}, \delta\boldsymbol{\lambda}) = 0\,. \tag{42}$$

It follows, therefore, that the coefficient of each variation $\delta\boldsymbol{a}$, $\delta\boldsymbol{\lambda}$ within the bracket must vanish, so that we have

$${}^tBB\boldsymbol{a} + {}^tF\boldsymbol{\lambda} = {}^tB\boldsymbol{v} \tag{43}$$

$$F\boldsymbol{a} = \boldsymbol{d} \tag{44}$$

or in a single form

$$\begin{Vmatrix} {}^tBB & {}^tF \\ F & 0 \end{Vmatrix} \cdot \begin{Vmatrix} \boldsymbol{a} \\ \boldsymbol{\lambda} \end{Vmatrix} = \begin{Vmatrix} {}^tB\boldsymbol{v} \\ \boldsymbol{d} \end{Vmatrix} \tag{45}$$

which is the fundamental design equation of the optimum system as proved below.

3.3 Uniqueness of solution of design equation

Equation (15) has the unique solution, if and only if

$$det\begin{pmatrix} {}^tBB & {}^tF \\ F & 0 \end{pmatrix} \neq 0\,. \tag{46}$$

This may be easily varified as follows.

Using Laplace's expansion theorem twice leads to

$$det\begin{pmatrix} {}^tBB & {}^tF \\ F & 0 \end{pmatrix} = \sum_{I,K} \beta_{IK} F_I \cdot F_K \tag{47}$$

where β_{IK} is a determinant of the order $N-n+1-l$ made by taking off all the elements at the intersections of l rows, which have numbers I: $i_1 < i_2 < \cdots < i_l$, and l columns, which have numbers K: $k_1 < k_2 < \cdots < k_l$, out of the $(N-n+1) \times (N-n+1)$ matrix ${}^tBB$. And F_I is a determinant of the order l formed from l columns numbered I: $i_1 < i_2 < \cdots < i_l$ in the $l \times (N-n+1)$ matrix F. The summation ranges all the combinations of $\binom{N-n+1}{l}$. If ${}^tBB$ represents a positive definite quadratic form and the rank of F is l, (47) does not vanish. Because $||\beta_{IK}||$ becomes also a matrix representing a positive definite quadratic form[3], and there exists a combination $I = (i_1 < i_2 < \cdots < i_l)$ such that $F_I \neq 0$. The second assumption is naturally satisfied, because condition (40) is composed of l linearly independent equations. And it is easy to show that ${}^tBB$ represents a positive definite quadratic form.

Note that

$$({}^tBB\boldsymbol{a}, \boldsymbol{a}) = (B\boldsymbol{a}, B\boldsymbol{a}) \geq 0 \tag{48}$$

namely, ${}^tBB$ represents a non-negative quadratic form. Hence, it is enough to show that $det({}^tBB) \neq 0$. Clearly,

$$det\,({}^tBB) = \sum \{D(i_1, i_2, \cdots, i_{N-n+1})\}^2 \tag{49}$$

where $D(i_1, i_2, \cdots, i_{N-n+1})$ is a determinant of the order $N-n+1$ formed with $N-n+1$ rows numbered $i_1<i_2\cdots<i_{N-n+1}$ in B which is a $N\times(N-n+1)$ matrix. The summation ranges all the combinations of $\binom{N}{N-n+1}$. Now from (37), if $g_0=g_1=\cdots g_{\lambda-1}=0$, $g_\lambda\neq 0$, then $D(\lambda+1, \lambda+2, \cdots, N-n+1+\lambda)=g_\lambda^{N-n+1}\neq 0$. Therefore we have $det\,({}^tBB)\neq 0$, unless $G_1(z)$, that is, $HG(z)$ is identically equal to zero. This accomplishes the proof.

3.4 Proof of the fact that solution of (45) gives the minimum value of (38)

The design equation (45) is a necessary condition for an extremum. In this case it is easy to show that the solution of (45), which is denoted by $\bar{a}$, makes the integrated-square sampled error (38) minimum under the condition (40).

To prove this, assume that $a=\bar{a}+\delta a$ is an arbitrary vector satisfying (40), then

$$F\delta a = 0\,. \tag{50}$$

Now from (38) we have

$$(\sum_{n=0}^{\infty} e_n^2)[a]-(\sum_{n=0}^{\infty} e_n^2)[\bar{a}] = 2({}^tBB\bar{a}-{}^tB\bar{a}, \delta a)+({}^tBB\delta a, \delta a)\,. \tag{51}$$

Since it is shown by the use of (43) and (50) that the first term of the right-hand side of (51) vanishes:

$$({}^tBB\bar{a}-{}^tB\bar{a}, \delta a) = (-{}^tF\bar{\lambda}, \delta a) = -(F\delta a, \bar{\lambda}) = 0\,, \tag{52}$$

then we obtain

$$(\sum_{n=0}^{\infty} e_n^2)[a]-(\sum_{n=0}^{\infty} e_n^2)[\bar{a}] = ({}^tBB\delta a, \delta a)\geqq 0\,, \tag{53}$$

which is the proof. The last inequality in (53) follows the fact that ${}^tBB$ represents a positive definite quadratic form.

3.5 Evaluation of $(\sum_{n=0}^{\infty} e_n^2)_{\min}$

The minimum value of the integrated-square sampled error is evaluated from the following formula.

$$(\sum_{n=0}^{\infty} e_n^2)_{\min} = N+\begin{vmatrix} {}^tBB & {}^tF & {}^tBv \\ F & 0 & d \\ {}^tvB & {}^td & 0 \end{vmatrix} \Big/ \begin{vmatrix} {}^tBB & {}^tF \\ F & 0 \end{vmatrix}\,. \tag{54}$$

Proof:

$$\begin{aligned} I[\bar{a}] = (\sum_{n=0}^{\infty} e_n^2)_{\min} &= |v|^2-2(B\bar{a}, v)+({}^tBB\bar{a}, \bar{a}) \\ &= |v|^2-({}^tBv, \bar{a})+({}^tBB\bar{a}-{}^tBv, \bar{a})\,. \end{aligned}$$

Substitution of (43): ${}^tBB\bar{a}-{}^tBv={}^tF\bar{\lambda}$ in the above expression yields

$$I[\bar{a}] = |v|^2 - ({}^tBv, \bar{a}) - ({}^tF\bar{\lambda}, \bar{a})$$
$$= |v|^2 - ({}^tBv, \bar{a}) - (F\bar{a}, \bar{\lambda}) .$$

This can be written, by the use of (44) : $F\bar{a}=d$, as

$$I[\bar{a}] = |v|^2 - ({}^tBv, \bar{a}) - (d, \bar{\lambda})$$
$$= |v|^2 - \|{}^tvB, {}^td\| \cdot \left\| \begin{matrix} \bar{a} \\ \bar{\lambda} \end{matrix} \right\|$$
$$= |v|^2 - \|{}^tvB, {}^td\| \cdot \left\| \begin{matrix} {}^tBB & {}^tF \\ F & 0 \end{matrix} \right\|^{-1} \cdot \left\| \begin{matrix} {}^tBv \\ d \end{matrix} \right\| .$$

Expression (37) shows that $|v|^2=N$. Hence we have the formula (54). This accomplishes the proof.

3.6 Example

To illustrate a practical example, consider the system whose transfer function is given by

$$G(s) = \frac{1}{s(Ts+1)} . \tag{55}$$

The pulse transfer function of the system including a zero-order hold element is

$$HG(z) = \frac{\alpha z^{-1}(z^{-1}-q)}{(1-z^{-1})(1-dz^{-1})} \tag{56}$$

where

$$\alpha = T(1-d-\mu d) , \quad q = \frac{1-d-\mu}{1-d-\mu d}$$
$$d = e^{-\mu} , \quad \mu = \frac{t_0}{T} .$$

In order that the output of the system shown in Fig. 1 settles completely to the unit step input after a short time period Nt_0 has elapsed, the form of $C(z)$ must be a finite polynomial in z^{-1}:

$$C(z) = b_0 + b_1 z^{-1} + \cdots + b_N z^{-N} \tag{57}$$

whose coefficients satisfy the following linear conditions[2)]

$$b_0 + b_1 + \cdots + b_N = 0 \tag{58}$$
$$b_0 d^N + b_1 d^{N-1} + \cdots + b_N = 0 \tag{59}$$
$$b_1 + 2b_2 + \cdots + Nb_N = -\frac{1}{t_0} . \tag{60}$$

Equations (58) and (59) show that $C(z)$ is divisible by $(1-z^{-1})(1-dz^{-1})$ which is the denominator of $HG(z)$. Let

$$C(z) = (1-z^{-1})(1-dz^{-1})(a_0 + a_1 z^{-1} + \cdots + a_{N-2} z^{-N+2}) , \tag{61}$$

then condition (60) can be rewritten as

$$a_0+a_1+\cdots+a_{N-2}=\frac{1}{t_0(1-d)}=\frac{1}{\alpha(1-q)}. \tag{62}$$

In designing a controller, we can assume that $\alpha=1$ in (56), because the optimum pulse transfer function of a controller for $\alpha\neq 1$ is obtained from the optimum one for $\alpha=1$ by multiplying it by $1/\alpha$. According to the general discussion, the numerator of (56): $G_1(z)=qz^{-1}+z^{-2}$ is represented by the following $(N+1)\times(N-1)$ matrix G

$$G=\begin{Vmatrix} 0 & & \\ -q & \ddots & \\ 1 & \ddots & 0 \\ & \ddots & -q \\ & & 1 \end{Vmatrix}.$$

And referring to (37) leads to

$$B=\hat{V}G=\left.\begin{Vmatrix} 0 & \cdots & 0 \\ -q & \ddots & \vdots \\ 1-q & \ddots & 0 \\ \vdots & \ddots & \\ 1-q & \cdots\ 1-q & -q \end{Vmatrix}\right\} N \quad \overbrace{\qquad}^{N-1}.$$

Therefore, the design equation of this case is given by

$$\begin{Vmatrix} & & 1 \\ & {}^tBB & 1 \\ & & \vdots \\ & & 1 \\ 1 & 1\cdots 1 & 0 \end{Vmatrix} \begin{Vmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-2} \\ \lambda \end{Vmatrix} = \begin{Vmatrix} -q+(N-2)(1-q) \\ -q+(N-3)(1-q) \\ \vdots \\ -q \\ \dfrac{1}{1-q} \end{Vmatrix} \tag{63}$$

where

$${}^tBB=\begin{Vmatrix} q^2+(N-2)(1-q)^2 & -q(1-q)+(N-3)(1-q)^2 & \cdots & -q(1-q) \\ -q(1-q)+(N-3)(1-q)^2 & q^2+(N-3)(1-q)^2 & \cdots & -q(1-q) \\ \vdots & \vdots & & \vdots \\ -q(1-q) & -q(1-q) & \cdots & q^2 \end{Vmatrix}.$$

The solutions of (63) are

$$\left.\begin{aligned} a_n &= \frac{q^{2N-n-1}+q^{n+2}}{q^2-q^{2N}}, \quad (n=0,1,\cdots,N-3) \\ a_{N-2} &= \frac{(1+q)q^N}{q^2-q^{2N}} \end{aligned}\right\} \tag{64}$$

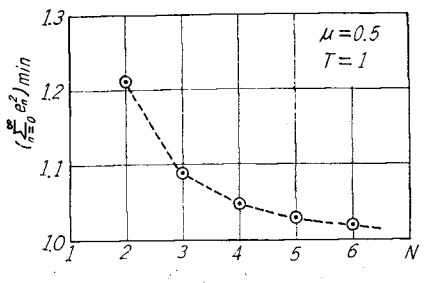

Fig. 2. Minimum integrated-square sampled error of system discussed in subsection 3.6.

And by making use of (45), we get the minimum integrated-square sampled error

$$\left(\sum_{n=0}^{\infty} e_n^2\right)_{\min} = 1+\frac{1+q}{1-q}\cdot\frac{1}{1-q^{2N-2}}. \quad (65)$$

Fig. 2 shows (65) where $\mu=t_0/T=0.5$. It is natural that the minimum integrated-square error decreases monotonically with increasing N. The transient responses for $N=2$, 3, 4, and 5 are shown in Fig. 3. Owing to the criterion adopted here, which concerns errors merely at sampling instants, it is observed in Fig. 3 that the system response has fairly large ripples. To improve this respect, some other criterion which concerns response between sampling instants must be introduced. This will be considered in the next section.

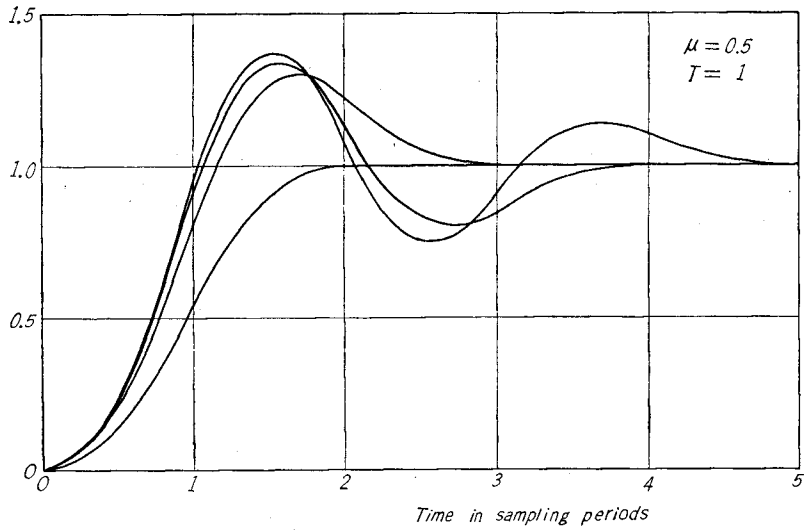

Fig. 3. Transient responses of system designed under minimum integrated-square sampled error criterion.

4. Synthesis under the Minimum Quadration Control Area Criterion

4.1 Error vector

In the case of synthesis under the minimum quadratic control area criterion, it is convinient to make use of the modified z transform method as mentioned in the introduction. The modified z transform of the error of the system shown in Fig. 1 is of the form

$$E(z, m) = \frac{z^{-1} - C(z) HG(z, m)}{1 - z^{-1}} \tag{66}$$

where $HG(z, m)$ is the modified pulse transfer function of the controlled system including a hold element. In general, $HG(z, m)$ is a rational function in z^{-1} and contains the parameter m in the numerater only. Consequently we may put

$$HG(z, m) = \frac{G_1(z, m)}{G_2(z)} \tag{67}$$

where $G_2(z)$ is a polynomial of the degree n and $G_1(z, m)$ is, in general, a polynomial of the degree $n+1$ without a constant term. So $G_1(z, m)$ may be written as

$$G_1(z, m) = \sum_{k=1}^{n+1} g_k(m) z^{-k}. \tag{68}$$

Similarly to the discussion in section 3, let

$$C(z) = G_2(z) D(z) \tag{69}$$

where

$$D(z) = a_0 + a_1 z^{-1} + \cdots + a_{N-n} z^{-N+n}.$$

Substitution of (67) and (69) in (66) gives

$$E(z, m) = \frac{z^{-1} - G_1(z, m) D(z)}{1 - z^{-1}}. \tag{70}$$

Here the right hand side of (70) is divisible and $E(z, m)$ is a polynomial of the degree N without a constant term as shown in (2). Then (70) may be represented by the vector notation

$$\begin{Vmatrix} e_1(m) \\ e_2(m) \\ \vdots \\ e_N(m) \end{Vmatrix} = \tilde{V}\left(\begin{Vmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{Vmatrix} - G_m \begin{Vmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-n} \end{Vmatrix} \right) \tag{71}$$

or more simply

$$\boldsymbol{e}(m) = \tilde{V}(\boldsymbol{u} - G_m \boldsymbol{a}) = \boldsymbol{v} - B_m \boldsymbol{a} \tag{72}$$

where $\tilde{V}$ is the same matrix as shown in (37) and G_m, which is the representation of (68), is of the following form

$$G_m = \begin{Vmatrix} g_1(m) & & & \\ g_2(m) & g_1(m) & & \\ \vdots & g_2(m) & \ddots & \\ g_{n+1}(m) & \vdots & \ddots & g_1(m) \\ & g_{n+1}(m) & g_2(m) & \\ & & \ddots & \vdots \\ & & & g_{n+1}(m) \end{Vmatrix}, \quad (N+1) \times (N-n+1) \tag{73}$$

4.2 Derivatiun of design equation

The quadratic control area can be obtained by integrating the inner product of $e(m)$ from 0 to 1 with respect to m

$$\int_0^\infty \{e(t)\}^2 dt = \int_0^1 (e(m), e(m))\, dm = \int_0^1 (v - B_m a, v - B_m a)\, dm$$
$$= \int_0^1 \{|v|^2 - 2(B_m a, v) + ({}^t B_m B_m a, a)\}\, dm . \tag{74}$$

Since only the elements of B_m and ${}^t B_m B_m$ are functions of a parameter m, (74) may be written as

$$\int_0^\infty \{e(t)\}^2 dt = |v|^2 - 2\left(\int_0^1 B_m dm\, a, v\right) + \left(\int_0^1 {}^t B_m B_m dm\, a, a\right)$$
$$= N - 2(\bar{B}a, v) + ({}^t\bar{B}\bar{B}a, a) \tag{75}$$

where $\bar{B}$ and ${}^t\bar{B}\bar{B}$ are matrices whose elements are obtained by integrating each of the corresponding elements of B_m and ${}^t B_m B_m$ with respect to m.

On the other hand, the condition imposed on a for obtaining the complete finite-settling-time response is the same with (40): $Fa - d = 0$.

Similarly to the discussion described in subsection 3.2, we can obtain the desired vector $\bar{a}$, which satisfies (40) and minimizes (75), as a solution of the following equation

$$\begin{Vmatrix} {}^t\bar{B}\bar{B} & {}^tF \\ F & 0 \end{Vmatrix} \begin{Vmatrix} a \\ \lambda \end{Vmatrix} = \begin{Vmatrix} {}^t\bar{B}v \\ d \end{Vmatrix} . \tag{76}$$

4.3 On the solution of (76)

It is easy to show that equation (76) has the unique solution. In order for this to be the case, it is enough to show that ${}^t\bar{B}\bar{B}$ represents a positive definite quadratic form.

Proof:

$$({}^t\bar{B}\bar{B}a, a) = \int_0^1 ({}^t B_m B_m a, a)\, dm = \int_0^1 |B_m a|^2 dm \geq 0 . \tag{78}$$

The left hand side of (78) does not vanish unless $|B_m a|^2$ is identically equal to zero. Now, in the expression of the error vector $e = v - B_m a$, v stands for the unit step input and $B_m a$ stands for the response to the unit step input of the system shown in Fig. 1. Consequently, if $B_m a$ is a zero vector, the output of the system can never be observed. This means that the system is cut off, that is, $C(z) \equiv 0$. Hence, we get $({}^t\bar{B}\bar{B}a, a) > 0$ whenever $a \neq 0$.

4.4 The minimum value of quadratic control area

By the procedure exactly similar to subsection 3.4, it is easy to show that the solution of (76) makes the quadratic control area (75) minimum. The minimum

value is evaluated by employing the following formula

$$\left[\int_0^\infty \{e(t)\}^2 dt\right]_{\min} = N + \frac{\begin{vmatrix} {}^t\bar{B}\bar{B} & {}^tF & {}^t\bar{B}\boldsymbol{v} \\ F & 0 & \boldsymbol{d} \\ {}^t\boldsymbol{v}\bar{B} & {}^t\boldsymbol{d} & 0 \end{vmatrix}}{\begin{vmatrix} {}^t\bar{B}\bar{B} & {}^tF \\ F & 0 \end{vmatrix}} . \tag{79}$$

The procedure of derivation of (79) is all the same as subsection 3.5.

4.5 Example

As an example, consider the same system as discussed in subsection 3.6. The modified z transform of $1/s(Ts+1)$ including a zero-order hold element is given by

$$HG(z, m) = \frac{g_1(m)z^{-1} + g_2(m)z^{-2} + g_3(m)z^{-3}}{(1-z^{-1})(1-dz^{-1})} \cdot T \tag{80}$$

where

$$\begin{aligned} g_1(m) &= m\mu - 1 + d^m \\ g_2(m) &= -m\mu(1+d) + (\mu + d + 1) - 2d^m \\ g_3(m) &= m\mu d - d(\mu+1) + d^m \end{aligned}$$

and μ, d are the same symbols as before. In this case, the controller must satisfy the same conditions as (58), (59) and (60). By applying the general procedure discussed in preceding subsections to this case, we can design the optinum controller under the minimum quadratic control area criterion. Fig. 4 shows the minimum quadratic control area as a function of settling-time Nt_0. It is naturally observed that the minimum value decreases monotonically with increasing N. The difference of the values between $N=2$ and 3 is remarkable. In this example, the system can not be adjusted by the above method till N becomes 3.

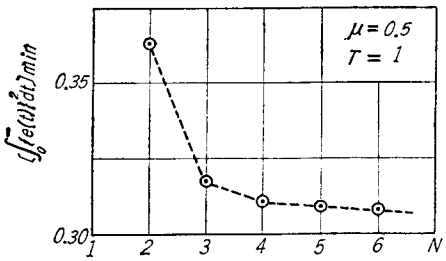

Fig. 4. Minimum quadratic control area of system discussed in subsection 4.5.

Fig. 5 shows transient responses to the unit step input of the optimum system designed by the method mentioned in this section. In Fig. 6 and Fig. 7, to discuss the comparative merits of the minimum integrated-square sampled error criterion and the minimum quadratic control area criterion, differences are shown between the unit step responses of the system designed under the former criterion and under the latter one. Dotted lines represent the former and solid lines the latter. A glance at these figures shows that the latter is superior to the former.

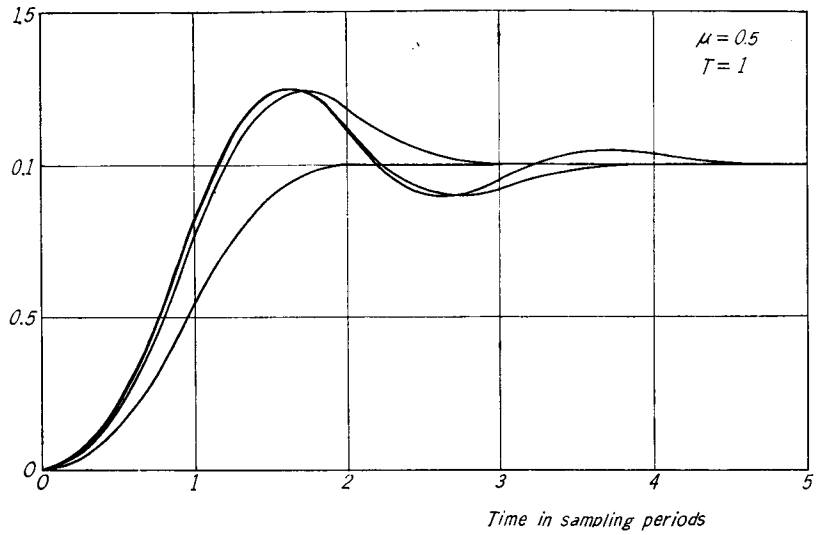

Fig. 5. Transient responses of system designed under minimum quadratic control area criterion.

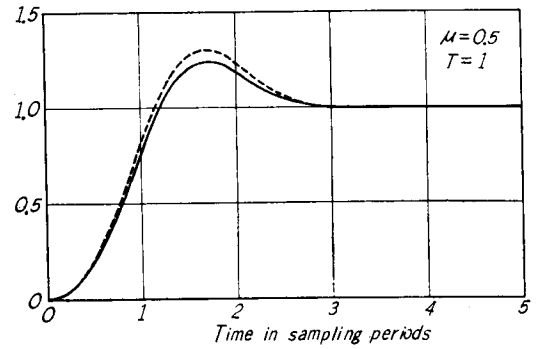

Fig. 6. Comparison between transient responses for $N=3$ under two kinds of criteria.

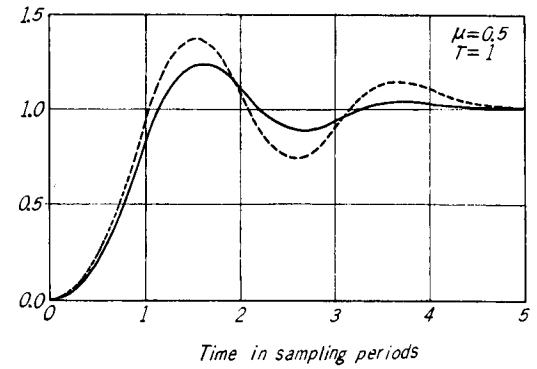

Fig. 7. Comparison between transient responses for $N=5$ under two kinds of criteria.

5. Conclusion

With a view to optimizing systems under the minimum integrated-square sampled error or the minimum quadratic control area criterion, the synthetic study is carried out in a systematic manner by representing polynomials by vectors or matrices and by making use of the linear algebra techniques. Two types of design equations and evaluating formulas of both the minimum integrated-square sampled error and the minimum quadratic control area are derived. In the case treated in this paper, the minimum quadratic control area criterion is found superior to the minimum integrated-square sampled error criterion. The linear algebra techniques used for the step response synthesis may be also applied to the other input response synthesis.

Acknowledgment

The authors would like to express their thanks to Mr. J. Tsuda who performed the calculations for this paper.

References

1) Y. Sawaragi and H. Fukawa: Synthesis of Finite-Settling-Time Systems under the Integrated-Square Sampled Error Criterion; Tech. Reps. of the Engg. Res. Inste. Kyoto Univ. Vol. XI, No. 9, Report No. 86, Oct. 1961.
2) Y. Sawaragi and H. Fukawa: Unified Method of Designing Finite-Settling-Time Systems; Proc. of the 10th JNCAM, 1960.
3) K. Asano: Determinants and Matrices, Kyoritsu Book Co., 1955. (in Japanese)