Stability analysis of digital filters under finite word-length effects

Prof. B.-S. Chen
C.-T. Kuo

Indexing terms: Digital filters

Abstract: An approach is introduced for analysing the effects of finite word length and testing the stability in digital filters. Based on the analysis of both the quantisation errors of filter coefficients and the computational roundoff errors of multiplications and additions, a sufficient stability condition is introduced to give an insight into the hardware/cost tradeoff for word-length design choices under the effects of finite word length. The Bellman-Gronwall lemma in discrete form is employed to tackle this problem. In addition, some numerical examples are presented to illustrate the results.

1 Introduction

Digital filters are normally designed under the assumption that both the filter states and the coefficients have an arbitrary, large word length. When implemented on finite word-length hardware, the practical consequences of having to represent both the states and coefficients with a finite number of bits need careful consideration. State and coefficient quantisation both lead to a deterioration in the ideal (i.e. infinite precision) performance of the digital filter. If the effects of finite word length are not considered in digital filter design, limit cycles or unstable responses may occur even when the poles of the quantised filter lie within the unit circle.

In the design of digital filters, it has been amply demonstrated that one must consider the effects of finite word length in the digital implementation. As we know, the effects have caused many kinds of problem in digital filters; therefore, a considerable amount of research has been done recently into these problems. For example, Williamson and Sridharan [1] have dealt with the subject of an approach to coefficient word-length reduction in digital filters. The sensitivity of state-space digital filters to coefficient quantisation errors has been studied by Bhaskar Rao [2]. Kawamata and Higuchi [8] have proposed deterministic and statistical approaches to the output error variance due to the coefficient quantisation of fixed-point state-space digital filters in the time domain. Bomar [3] introduced a new technique for computationally synthesising efficient low-noise second state-space structures. The conditions of the existence of limit-cycle oscillations in floating-point digital filters are derived by Kaneko [10]. Sripad and Snyder [11] have established a necessary and adequate condition for the mantissa in a normalised floating-point number to have a reciprocal probability density. The error analysis of digital filters realised with floating-point arithmetic is discussed by Liu and Kaneko [5]. However, in the above papers, the effects of coefficient quantisation errors and computational roundoff errors in floating-point digital filters are not studied together.

In this paper, the effects of both coefficient quantisation errors and computational roundoff errors due to multiplication or addition are taken into account in digital filters. Both the errors are considered as perturbations of the filter's system using floating-point arithmetic. Based on the Bellman-Gronwall lemma, a stability criterion [9] is derived to tolerate the perturbation due to coefficient quantisation and roundoff accumulation without leading to a limit cycle or instability. According to this stability criterion, an adequate word length is proposed under the consideration of both the quantisation error and the roundoff error for a realisation of a floating-point digital filter. The signal bounds in the digital filter are also discussed.

Paper 6759G (E10), first received 17th November 1987 and in revised form 20th March 1989

Prof. Chen is with the Department of Electrical Engineering, National Tsing Hua University, National Tsing Hua University, Kuang Fu Road, Hsinchu, Taiwan, ROC

Mr. Kuo is with the Department of Electrical Engineering, Tatung Institute of Technology, 40 Chungshan North Road, 3rd. Sec. Taipei, Taiwan, ROC

2 Problem formulation

An infinite word-length digital filter can be described by the following state equations:

$$x(k+1) = Ax(k) + bu(k) \tag{1a}$$

$$y(k) = cx(k) \tag{1b}$$

where $u(k)$ is the scalar input, $y(k)$ is the scalar output, and $x(k)$ is the nth-order state vector. A, b and c are $n \times n$, $n \times 1$, $1 \times n$ real constant matrices, respectively. Different kinds of realisations of a digital filter correspond to different state-space structures of (A, b, c).

To implement a digital filter with floating-point computation, (where we consider the computational roundoff errors due to the finite word-length effects), let $fl_i\{a + b\}$ denote the result of floating-point addition, and let $FL_i[ab]$ denote the result of floating-point multiplication. An actual state-space digital filter implemented with a finite word-length machine is described by the following state equations:

$$x^*(k+1) = fl_1\{FL_1[A^*x^*(k)] + FL_2[b^*u^*(k)]\} \tag{2a}$$

$$y^*(k) = FL_3[c^*x^*(k)] \tag{2b}$$

where $x^*(k)$, $u^*(k)$ and $y^*(k)$ are the actual state, the actual input and the actual output, respectively, and matrices A^*, b^* and c^* are represented with finite word length.

The main problems considered in this paper are as follows:

(i) Under what conditions is the digital filter, affected by the finite word length, still stable?

(ii) By what values are the state signal and output signal bounded in (i)?

3 Stability analysis of digital filters

In this Section, we present a method to analyse the stability of digital filters influenced by finite word length. Before further analysis, some mathematical tools for the problems are introduced. Let the norm of real vector $x \in R^n$, denoted by $\|x\|$, be defined by References 6 and 7

$$\|x\|\sqrt{E[x^T x]} \tag{3}$$

Then

$$\begin{aligned}\|Ax\|^2 &= E[x^T A^T A x] \\ &= tr(E[x^T A^T A x]) \\ &= tr(E[A^T A x x^T])\end{aligned} \tag{4}$$

where $tr(\cdot)$ denotes the trace operator. Corresponding to the vector norm, the following inequality is satisfied:

$$\begin{aligned}&\|Ax\|^2 \leqslant tr(E[\lambda_{max}(A^T A) x x^T]) \\ &= \lambda_{max}(E[A^T A]) tr(E[x x^T]) \\ &= \begin{cases}\lambda_{max}(E[A^T A])\|x\|^2, & \text{while } A \text{ is stochastic} \\ \lambda_{max}(A^T A)\|x\|^2, & \text{while } A \text{ is deterministic}\end{cases}\end{aligned} \tag{5}$$

where $\lambda_{max}(\cdot)$ denotes the maximum eigenvalue of $\cdot$, and so we obtain

$$\|Ax\|^2 \leqslant \|A\|^2 \|x\|^2$$

where $\|A\|$ denotes the induced norm defined as follows:

$$\|A\|\begin{cases}\sqrt{(\lambda_{max}(A^T A))}, & \text{for } A \text{ is deterministic} \\ \sqrt{(\lambda_{max}(E[A^T A]))}, & \text{for } A \text{ is stochastic}\end{cases} \tag{6}$$

The rounded floating-point sum of two M-vectors can be expressed by References 4 and 12

$$fl\{a + b\} = (I + \Delta R)(a + b) \tag{7}$$

where $\Delta = 2^{-L}$, L being the mantissa length and

$$R = \text{diag}\,[r_1, r_2, \ldots, r_M]$$

where each r_i is uniformly distributed between -1 and 1, and so

$$\begin{aligned}&E[r_i] = 0, \quad i = 1, 2, \ldots, M \\ &E[r_i r_j] = \tfrac{1}{3}\delta_{ij}\end{aligned} \tag{8}$$

The rounded floating-point product of a $M \times M$ matrix A and a M-vector x [4, 12] is given as follows:

$$FL[Ax] = A(I + \Delta H)x \tag{9}$$

with

$$H = \text{diag}\,[h_1, h_2, \ldots, h_M]$$

where h_i has zero mean, and the variances are given approximately as

$$\begin{aligned}&E[h_i^2] = \frac{(i+1)}{3}, \quad \text{for } i = 1, 2, \ldots, M-1 \\ &E[h_M^2] = \frac{M}{3} \\ &E[h_i h_j] = \frac{i}{3}, \quad \text{for } j > i.\end{aligned} \tag{10}$$

Using the representations eqns. 7 and 9, eqn. 2 becomes

$$\begin{aligned}x^*(k+1) &= (I + \Delta R_1)[A^*(I + \Delta H_1)x^*(k) \\ &\quad + b^*(I + \Delta H_2)u^*(k)] \\ &= Ax^*(k) + (A^* - A)x^*(k) + (\Delta R_1 A^* \\ &\quad + A^*\Delta H_1)x^*(k) + bu^*(k) \\ &\quad + (b^* - b)u^*(k) + (\Delta R_1 b^* \\ &\quad + b^*\Delta H_2)u^*(k) \\ &\quad + \text{terms of higher order } \Delta^2 \\ &\simeq Ax^*(k) + (A^* - A)x^*(k) \\ &\quad + (\Delta R_1 A^* + A^*\Delta H_1)x^*(k) \\ &\quad + bu^*(k) + (b^* - b)u^*(k) \\ &\quad + (\Delta R_1 b^* + b^*\Delta H_2)u^*(k)\end{aligned} \tag{11a}$$

$$\begin{aligned}y^*(k) &= c^*(I + \Delta H_3)x^*(k) \\ &= cx^*(k) + (c^* - c)x^*(k) \\ &\quad + c^*\Delta H_3 x^*(k)\end{aligned} \tag{11b}$$

In eqn. 11a, because the terms of higher order Δ^2 are small, they are ignored.

If A is a stable transition matrix for digital filters in infinite precision, then

$$\|A^k\| \leqslant mr^k, \quad k > 0 \tag{12}$$

for some constant $m > 0$ and $0 \leqslant r < 1$. Simply choose $r = \max_i |\lambda_i(A)|$, where $\lambda_i(A)$, for $i = 1, 2, \ldots, n$, denotes the eigenvalues of A. That is, r is the absolute value of the eigenvalue of A (or the pole of digital filter) nearest to the unit circle. m can be estimated from $\|A^k\|/r^k \leqslant m$ for all k. To obtain m is sometimes very difficult, but fortunately it can be obtained with the aid of a computer.

The Bellman-Gronwall lemma is introduced for the following theorem:

Lemma: [6]
Let $(u(k))_0^\infty$, $(f(k))_0^\infty$, $(h(k))_0^\infty$ be real-valued sequences on the set of the positive integer Z_+. Let

$$h(k) \geqslant 0, \quad \forall k \in Z_+ \tag{13}$$

Under these conditions, if

$$u(k) \leqslant f(k) + \sum_{i=0}^{k-1} h(i)u(i), \quad k = 0, 1, 2, \ldots \tag{14}$$

then

$$u(k) \leqslant f(k) + \sum_{i=0}^{k-1}\left\{\prod_{j=i+1}^{k-1}[1 + h(j)]h(i)f(i)\right\}, \quad k = 0, 1, 2, \ldots \tag{15}$$

where $\prod_{j=i+1}^{k-1}[1 + h(j)]$ is set equal to 1 when $i = k - 1$.

Remark:

(i) If for some constant values of h, $h(i) \leqslant h$, $\forall i$, then eqn. 15 becomes

$$u(k) \leqslant f(k) + h \sum_{i=0}^{k-1} (1+h)^{k-1-i} f(i) \quad (16)$$

(ii) If for some constant values of f, $f(i) \leqslant f$, $\forall i$, then eqn. 15 becomes

$$u(k) \leqslant f \prod_{i=1}^{k-1} [1 + h(i)] \quad (17)$$

Based on the lemma, we have the following theorem: for the digital filter with the state eqns. 11*a* and 11*b*, suppose that the induced norms of the coefficient quantisation errors are bounded, respectively, by the following inequalities:

$$\|A^* - A\| \leqslant \alpha_1; \quad \|b^* - b\| \leqslant \beta_1; \quad \|c^* - c\| \leqslant \gamma_1 \quad (18)$$

Assume that

$$\|\Delta R_1 A^* + A^* \Delta H_1\| \cong \alpha_2; \quad \|\Delta R_1 b^* + b^* \Delta H_2\| \cong \beta_2;$$
$$\|c^* \Delta H_3\| \cong \gamma_2; \quad \|b\| = \beta_3; \quad \|c\| = \gamma_3 \quad (19)$$

Under these conditions, if the stability inequality

$$r + m(\alpha_1 + \alpha_2) < 1 \quad (20)$$

is satisfied, then the deteriorated digital filter is still stable, the state signal $\|x^*(k)\|$ of the digital filter is bounded by

$$\frac{m(\beta_1 + \beta_2 + \beta_3)}{1 - [r + m(\alpha_1 + \alpha_2)]} \sup_{0 \leqslant i < \infty} \|u^*(i)\|, \quad \text{as } k \to \infty \quad (21)$$

and the output signal $\|y^*(k)\|$ is bounded by

$$\frac{m(\gamma_1 + \gamma_2 + \gamma_3)(\beta_1 + \beta_2 + \beta_3)}{1 - [r + m(\alpha_1 + \alpha_2)]} \sup_{0 \leqslant i < \infty} \|u^*(i)\|, \quad \text{as } k \to \infty \quad (22)$$

Proof: See Appendix 7.

Remark:

(*a*) The relationship between the location of the pole nearest to the unit circle, the coefficient quantisation error and the computational roundoff error is revealed. From the stability inequality in eqn. 20, it is seen that the smaller r is, the stronger the stability will be

(*b*) The values of $\|b\|$ and $\|c\|$ are obtained directly from the design of ideal filters. In other words, when getting the filter in infinite precision, we can compute the values of the induced norms $\|b\|$ and $\|c\|$

(*c*) $\|A^* - A\|$, $\|b^* - b\|$ and $\|c^* - c\|$ are the coefficient quantisation errors. The following result is the key tool used to evaluate the bounds of this kind of error. Define a new matrix $A^{\#}$ and decide the element of $A^{\#}$ as follows:

(i) If $|A_{ij}| > 1$ and $|A_{ij}| \neq 2^k$, where A_{ij} is the ijth element of A and $k = \cdots -2, -1, 0, 1, 2, \ldots$, then

$$A_{ij}^{\#} = \text{sgn}(A_{ij}) \times [|A_{ij}| + 2^{-L-1} \times 2^{int(\log_2|A_{ij}|)+1}] \quad (23)$$

where A_{ij} is an element of $A^{\#}$, int$(\cdot)$ the integer function and sgn$(\cdot)$ the sign function

(ii) If $|A_{ij}| < 1$, $|A_{ij}| \neq 0$ and $|A_{ij}| \neq 2^k$, where $k = \cdots -2, -1, 0, 1, 2, \ldots$, then

$$A_{ij}^{\#} = \text{sgn}(A_{ij})[|A_{ij}| + 2^{-L-1} \times 2^{int(\log_2|A_{ij}|)}] \quad (24)$$

(iii) If $|A_{ij}| = 2^k$, where $k = \cdots -2, -1, 0, 1, 2, \ldots$, then

$$A_{ij}^{\#} = \text{sgn}(A_{ij}) \cdot 2^k \quad (25a)$$

If $|A_{ij}| = 0$, then

$$A_{ij}^{\#} = 0 \quad (25b)$$

Because $\|A^* - A\| \leqslant \|A^{\#} - A\|$, let $\alpha_1 = \|A^{\#} - A\|$. In the same way, define $b^{\#}$ and $c^{\#}$, then β_1 and γ_1 can be obtained

(*d*) $\|\Delta R_1 A^* + A^* \Delta H_1\|$, $\|\Delta R_1 b^* + b^* \Delta H_2\|$ and $\|c^* \Delta H_3\|$ are the rounded errors of the results of the multiplications and additions. To simplify the analysis, we substitute A for A^*, because the difference between ΔA and ΔA^* can be ignored. Thus, we have

$$\alpha_2 = \|\Delta R_1 A + A \Delta H_1\| \quad (26)$$

Similarly,

$$\beta_2 = \|\Delta R_1 b + b \Delta H_2\| \quad (27)$$

$$\gamma_2 = \|c \Delta H_3\| \quad (28)$$

(*e*) From the above analysis, it is seen that the stability of the finite word-length filter strongly depends on its realisations (i.e. structures of A, b and c). Different kinds of realisations of a digital filter have different values of α_1 and α_2 and lead to different degrees of stability, so the inequality of eqn. 20 refers to a stability criterion for word-length consideration under finite word-length effects.

We do not claim that this stability criterion given here is the ultimate tool for solving the problem of the minimal word length when implementing digital filters. However, it can be used as a reference for conservative designs, and it permits an approximate analysis of the performance/cost tradeoff for word-length design choices.

4 Numerical examples

Example 1:
To illustrate the stability criterion of the digital filter proposed in this paper, we consider a second-order Butterworth highpass filter with the transfer function

$$H(z) = \frac{-0.576617z + 0.434501}{z^2 - 1.38762z + 0.492423} \quad (29)$$

The filter can be one section in the parallel realisation of a fourth-order or sixth-order Butterworth highpass filter. The direct form 2 of the filter is shown in Fig. 1, and its state-space representation is

$$x(k+1) = \begin{bmatrix} 1.38762 & -0.492423 \\ 1 & 0 \end{bmatrix} x(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(k) \quad (30a)$$

$$y(k) = [-0.576617 \quad 0.434501] x(k) \quad (30b)$$

Hence

$$A = \begin{bmatrix} 1.38762 & -0.492423 \\ 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

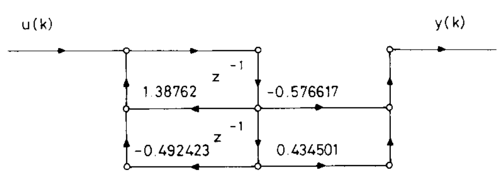

Fig. 1 *Direct form 2 of second-order filter in Example 1*

and

$c = [-0.576617 \quad 0.434501]$.

From eqns. 23, 24 and 25, we have

$$A^{\#} = \begin{bmatrix} 1.38762 + 2^{-L-1} \cdot 2 & -(0.492423 + 2^{-L-1}) \\ 1 & 0 \end{bmatrix},$$

$$b^{\#} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

and

$$c^{\#} = [-(0.576617 + 2^{-L-1}) \quad 0.434501 + 2^{-L-1} \cdot 2^{-1}]$$

Hence

$$\alpha_1 = 2^{-L-1} \cdot 2.236068, \quad \beta_1 = 0$$

and

$$\gamma_1 = 2^{-L-1} \cdot 1.118034$$

From eqns. 26, 27 and 28, we obtain

$$\alpha_2 = 2^{-L} \cdot 1.732245, \quad \beta_2 = 2^{-L} \cdot 1.154700$$

and

$$\gamma_2 = 2^{-L} \cdot 0.519645$$

The values of $\|A\|$, $\|b\|$ and $\|c\|$ are

$$\|A\| = 1.757693, \quad \|b\| = \beta_3 = 1$$

and

$$\|c\| = \gamma_3 = 0.721996$$

From eqn. 12, it is found that

$$r = 0.701729$$

When $k = 1$, $m = \|A\|/r = 2.504803$ satisfies the requirement in eqn. 12. Using the stability criterion (eqn. 20), we obtain

$$L > 4.581105$$

From the above analysis, we suggest that the required word length must be greater than or equal to 5 bits; otherwise it may lead to unstable responses. Substituting these values of α_1, α_2, β_1, β_2, β_3, γ_1, γ_2 and γ_3 into eqns. 21 and 22, we can estimate the bounds of the state signal $\|x^*(k)\|$ and the output signal $\|y^*(k)\|$.

The digital filter was simulated on the IBM-AT computer system using a signal input obtained from a psuedorandom Gaussian number generator. Simulations were carried out truncating the word lengths to 1 up to 10 bits, with the results shown in Fig. 2. It is seen from Fig. 2 that the digital filter is not well behaved when the word length is small.

Example 2:

Consider a second-order lowpass Butterworth filter of which the transfer function is

$$H(z) = \frac{(z+1)^2}{z^2 - 1.96413z + 0.96802} \tag{31}$$

Similarly, it can be a section of a fourth-order or sixth-order lowpass Butterworth filter.

It is easy to obtain

$$A = \begin{bmatrix} 0.9821 & -0.0598 \\ 0.0597 & 0.9821 \end{bmatrix}, \quad b = \begin{bmatrix} 0.04993 \\ 0.00810 \end{bmatrix}$$

and

$$c = [-0.12409 \quad 1.2288]$$

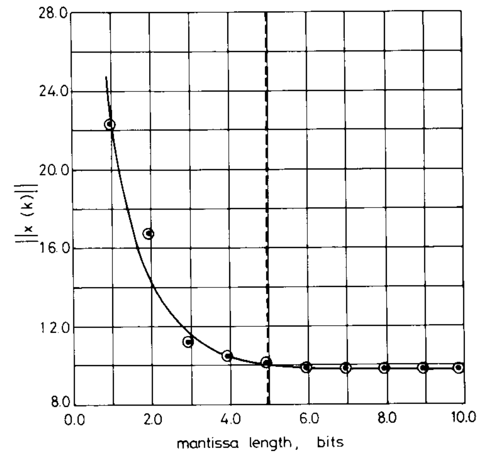

Fig. 2 *State signal $\|x(k)\|$ for Example 1*
- - - - Word length bound
-·-·— State signal

We can also obtain the following values:

$$\alpha_1 = 2^{-L-1} \cdot 1.007782; \quad \alpha_2 = 2^{-L} \cdot 0.983913;$$
$$\|A\| = 0.983919; \quad \beta_1 = 2^{-L-1} \cdot 0.064424;$$
$$\beta_2 = 2^{-L} \cdot 0.050583; \quad \beta_3 = 0.050583;$$
$$\gamma_1 = 2^{-L-1} \cdot 2.003902; \quad \gamma_2 = 2^{-L} \cdot 1.009219;$$
$$\gamma_3 = 1.235050; \quad r = 0.982126;$$
$$m = 1.001826$$

Applying the stability criterion (eqn. 20), we get

$$L > 6.802618$$

Hence, the suitable word length is equal to or greater than 7.

5 Conclusion

This paper has proposed an approach to the analysis of the errors due to coefficient quantisations of state-space digital filters and rounded quantisations of the results of multiplications and additions. A stability criterion is introduced to choose an adequte word length for the realisation of a digital filter. This method of analysis of the finite word-length effects can be extended to the design of 2-D digital filters, discrete-time Kalman filters and digital controllers etc.

As we know, although several state-space structures may be equivalent (with regard to their input–output characteristics) for infinite-precision representations of the coefficients and variables, they may have greatly different characteristics when the precision is limited. So this stability criterion will be used as a basis for us to choose an adequate realisation among different kinds of realisations in digical filter design. One disadvantage for the use of this method from the state-space approach to synthesise digital filters is that the bounds obtained by this method are only a sufficient condition and may be conservative in some cases.

6 References

1 WILLIAMSON, D., and SRIDHARAN, S.: 'An approach to coefficient word length reduction in digital filters', *IEEE Trans.*, 1985, **CAS-32,** pp. 893–903

2 BHASKAR RAO, D.V.: 'Analysis of coefficient quantization errors in state-space digital filters', *IEEE Trans.*, 1986, **ASSP-34,** pp. 131–139

3 BOMAR, B.W.: 'Computationally efficient low roundoff noise second-order state-space structure', *IEEE Trans.*, 1986, **CAS-33,** pp. 35–41

4 WILKINSON, J.H.: 'Rounding errors in algebraic processes' (Prentice-Hall, Englewood Cliffs, NJ, 1963)

5 LIU, B., and KANEKO, T.: 'Error analysis of digital filters realized with floating point arithmetic', *Proc. IEEE*, 1969, **57,** pp. 1735–1748

6 DESOER, C.A., and VIDYASAGAR, M.: 'Feedback systems: input-output properties' (Academic Press, New York, 1975)

7 GOODWIN, G.C., and SIN, K.S.: 'Adaptive filtering prediction and control' (Prentice-Hall, Englewood Cliffs, NJ, 1984)

8 KAWAMATA, M., and HIGUCHI, T.: 'A unified approach to the optimal synthesis of fixed-point state-space digital filters', *IEEE Trans.*, 1985, **ASSP-33,** pp. 911–920

9 KUO, C.-T., and CHEN, B.-S.: 'Robust stability analysis of digital filters under finite word length effects'. International Computer Symposium, Tainan, Taiwan, 1986

10 KANEKO, T.: 'Limit cycle oscillations in floating-point digital filters', *IEEE Trans.*, 1973, **AU-21,** pp. 100–106

11 SRIPAD, A.B., and SNYDER, D.L.: 'Quantisation errors in floating-point arithmetic', *IEEE Trans.*, 1978, **ASSP-26,** pp. 456–463

12 RINK, R.E., and CHONG, H.Y.: 'Performance of state regulator systems with floating-point computation', *IEEE Trans.*, 1979, **AC-24,** pp. 411–421

7 Appendix

Proof of theorem:
The proof is divided into two parts:

(i) From eqn. 11*a*, it is seen that

$$\begin{aligned} x^*(k+1) \cong\ & Ax^*(k) + (A^* - A)x^*(k) \\ & + (\Delta R_1 A^* + A^*\Delta H_1)x^*(k) \\ & + bu^*(k) + (b^* - b)u^*(k) \\ & + (\Delta R_1 b^* + b^*\Delta H_2)u^*(k) \end{aligned}$$

Solving the above difference eqn., we obtain the solution

$$\begin{aligned} x^*(k) \cong\ & A^k x^*(0) + \sum_{i=0}^{k-1} A^{k-1-i}(A^* - A)x^*(i) \\ & + \sum_{i=0}^{k-1} A^{k-1-i} bu^*(i) \\ & + \sum_{i=0}^{k-1} A^{k-1-i}(b^* - b)u^*(i) \\ & + \sum_{i=0}^{k-1} A^{k-1-i}(\Delta R_1 A^* + A^*\Delta H_1)x^*(i) \\ & + \sum_{i=0}^{k-1} A^{k-1-i}(\Delta R_1 b^* + b^*\Delta H_2)u^*(i) \end{aligned} \quad (32)$$

Applying the properties of the norm, we get

$$\begin{aligned} \|x^*(k)\| \leqslant\ & \|A^k\|\,\|x^*(0)\| + \sum_{i=0}^{k-1} \|A^{k-1-i}\| \\ & \times \|A^* - A\|\,\|x^*(i)\| + \sum_{i=0}^{k-1} \|A^{k-1-i}\| \\ & \times \|b\|\,\|u^*(i)\| + \sum_{i=0}^{k-1} \|A^{k-1-i}\| \\ & \times \|b^* - b\|\,\|u^*(i)\| + \sum_{i=0}^{k-1} \|A^{k-1-i}\| \\ & \times \|\Delta R_1 A^* + A^*\Delta H_1\|\,\|x^*(i)\| \\ & + \sum_{i=0}^{k-1} \|A^{k-1-i}\|\,\|\Delta R_1 b^* + b^*\Delta H_2\| \\ & \times \|u^*(i)\| \end{aligned} \quad (33)$$

From eqns. 12, 18 and 19, we have

$$\|A^k\| \leqslant mr^k; \quad \|A^* - A\| \leqslant \alpha_1;$$
$$\|\Delta R_1 A^* + A^*\Delta H_1\| \cong \alpha_2; \quad \|b^* - b\| \leqslant \beta_1;$$
$$\|\Delta R_1 b^* + b^*\Delta H_2\| \cong \beta_2; \quad \|b\| = \beta_3$$

Therefore,

$$\begin{aligned} \|x^*(k)\| \leqslant\ & mr^k\|x^*(0)\| + \sum_{i=0}^{k-1} mr^{k-1-i}(\alpha_1 + \alpha_2)\|x^*(i)\| \\ & + \sum_{i=0}^{k-1} mr^{k-1-i}(\beta_1 + \beta_2 + \beta_3)\|u^*(i)\| \\ \leqslant\ & mr^k\|x^*(0)\| + \sum_{i=0}^{k-1} mr^{k-1-i}(\alpha_1 + \alpha_2)\|x^*(i)\| \\ & + \sup_{0 \leqslant i < k} \|u^*(i)\| \sum_{i=0}^{k-1} mr^{k-1-i}(\beta_1 + \beta_2 + \beta_3) \end{aligned} \quad (34)$$

Dividing both sides of the inequality of eqn. 34 by r^k, we obtain

$$\begin{aligned} r^{-k}\|x^*(k)\| \leqslant\ & m\|x^*(0)\| + m(\beta_1 + \beta_2 + \beta_3) \\ & \times \sup_{0 \leqslant i < k} \|u^*(i)\| \frac{r^k - 1}{r^k(r-1)} \\ & + mr^{-1}(\alpha_1 + \alpha_2) \sum_{i=0}^{k-1} r^{-i}\|x^*(i)\| \end{aligned} \quad (35)$$

Applying the Lemma, we obtain

$$\begin{aligned} r^{-k}\|x^*(k)\| \leqslant\ & \Bigg[m\|x^*(0)\| + m(\beta_1 + \beta_2 + \beta_3) \\ & \times \sup_{0 \leqslant i < k} \|u^*(i)\| \frac{r^k - 1}{r^k(r-1)} \Bigg] \\ & + mr^{-1}(\alpha_1 + \alpha_2) \\ & \times \sum_{i=0}^{k-1} [1 + mr^{-1}(\alpha_1 + \alpha_2)]^{k-1-i} \\ & \times \Bigg[m\|x^*(0)\| + m(\beta_1 + \beta_2 + \beta_3) \\ & \times \sup_{0 \leqslant j < i} \|u^*(j)\| \frac{r^i - 1}{r^i(r-1)} \Bigg] \end{aligned} \quad (36)$$

Next, multiplying both sides of eqn. 36 by r^k, it follows that

$$\begin{aligned} \|x^*(k)\| \leqslant\ & mr^k\|x^*(0)\| + m(\beta_1 + \beta_2 + \beta_3) \\ & \times \sup_{0 \leqslant i < k} \|u^*(i)\| \frac{r^k - 1}{r - 1} + mr^{k-1}(\alpha_1 + \alpha_2) \\ & \times \sum_{i=0}^{k-1} [1 + mr^{-1}(\alpha_1 + \alpha_2)]^{k-1-i} \\ & \times \Bigg[m\|x^*(0)\| + m(\beta_1 + \beta_2 + \beta_3) \\ & \times \sup_{0 \leqslant j < i} \|u^*(j)\| \frac{r^i - 1}{r^i(r-1)} \Bigg] \end{aligned}$$

$$= mr^k\|x^*(0)\| + m(\beta_1 + \beta_2 + \beta_3)$$

$$\times \sup_{0 \leq i \leq k} \|u^*(i)\| \frac{r^k - 1}{(r - 1)}$$

$$+ m^2(\alpha_1 + \alpha_2)\|x^*(0)\|$$

$$\times \frac{[r + m(\alpha_1 + \alpha_2)]^{k-1} - r^k[r + m(\alpha_1 + \alpha_2)]^{-1}}{1 - [1 + mr^{-1}(\alpha_1 + \alpha_2)]^{-1}}$$

$$+ m^2(\alpha_1 + \alpha_2)(\beta_1 + \beta_2 + \beta_3)$$

$$\times \sup_{0 \leq i < k} \|u^*(i)\| \left\{ \frac{1}{r - 1} \right.$$

$$\times \frac{[r + m(\alpha_1 + \alpha_2)]^{k-1} - r^k[r + m(\alpha_1 + \alpha_2)]^{-1}}{1 - [1 + mr^{-1}(\alpha_1 + \alpha_2)]^{-1}}$$

$$+ \frac{-1}{r - 1}$$

$$\left. \times \frac{[r + m(\alpha_1 + \alpha_2)]^{k-1} - [r + m(\alpha_1 + \alpha_2)]^{-1}}{1 - [r + m(\alpha_1 + \alpha_2)]^{-1}} \right\}$$

When $k \to \infty$, if $r + m(\alpha_1 + \alpha_2) < 1$, then

$$\|x^*(k)\| \leq \frac{m(\beta_1 + \beta_2 + \beta_3)}{1 - [r + m(\alpha_1 + \alpha_2)]} \sup_{0 \leq i < \infty} \|u^*(i)\| \quad (37)$$

Hence, the bound of the state signal $\|x^*(k)\|$ is

$$\frac{m(\beta_1 + \beta_2 + \beta_3)}{1 - [r + m(\alpha_1 + \alpha_2)]} \sup_{0 \leq i < \infty} \|u^*(i)\|$$

(ii) From eqns. 18 and 19,

$$\|c^* - c\| \leq \gamma_1, \quad \|c^*\Delta H_3\| \cong \gamma_2, \quad \text{and} \quad \|c\| = \gamma_3$$

From eqn. 11*b*, we have

$$y^*(k) = cx^*(k) + (c^* - c)x^*(k) + c^*\Delta H_3\, x^*(k)$$

Taking norms, we get

$$\|y^*(k)\| \leq \|c\|\,\|x^*(k)\| + \|c^* - c\|\,\|x^*(k)\| + \|c^*\Delta H_3\|\,\|x^*(k)\| = (\gamma_1 + \gamma_2 + \gamma_3)\|x^*(k)\|$$

Similarly, when $k \to \infty$, if $r + m(\alpha_1 + \alpha_2) < 1$, the output signal $\|y^*(k)\|$ will be bounded by

$$\frac{m(\gamma_1 + \gamma_2 + \gamma_3)(\beta_1 + \beta_2 + \beta_3)}{1 - [r + m(\alpha_1 + \alpha_2)]} \sup_{0 \leq i < \infty} \|u^*(i)\|$$

Now, we complete the proof.