Admissibility of Unstable Second-order Digital Filters with Two's Complement Arithmetic

Bingo Wing-Kuen Ling, Charlotte Yuk-Fan Ho and Peter Kwong-Shun Tam

Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

SUMMARY

In this letter, we have extended the existing results on the admissible set of periodic symbolic sequences of a second-order digital filter with marginally stable system matrix to the unstable case. Based on this result, the initial conditions can be computed using the symbolic sequences. The truncation error of the representation of an initial condition due to the use of a finite number of symbols is studied.

KEY WORDS: Admissibility, two's complement arithmetic, symbolic sequences.

1. INTRODUCTION

It is well known that the autonomous response of a second-order digital filter with marginally stable system matrix implemented using two's complement arithmetic may exhibit chaotic behaviors, dependent on the initial conditions [1]-[4]. In order to analyze these complex behaviors, symbolic sequences are introduced. The symbolic sequences depend on the initial conditions. It is found that the map from the set of initial conditions to the set of symbolic sequences is neither injective nor surjective. Some researchers have worked out the admissible set of periodic sequences [2]-[4]. In this letter, we extend the results to the case with unstable system matrix and some interesting phenomenon was found.

The organization of this letter is as follows: The system is described in section 2. The admissible set of symbolic sequences of an unstable second-order digital filter with two's complement arithmetic is discussed in section 3. Finally, a conclusion is summarized in section 4.

2. SYSTEM DESCRIPTION

Assume a second-order digital filter with two's complement arithmetic is realized in direct form. The state space model of the feedback system can be represented as follows:

$$\mathbf{x}(k+1) \equiv \begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} x_2(k) \\ f(b \cdot x_1(k) + a \cdot x_2(k)) \end{bmatrix}, \text{ for } k \geq 0, \tag{1}$$

where a and b are the filter parameters, $x_1(k)$ and $x_2(k)$ are the state variables, and f is the nonlinearity due to the use of two's complement arithmetic. The nonlinearity f can be modeled as:

$$f(v) = v - 2 \cdot n \tag{2}$$

such that

$$2 \cdot n - 1 \leq v < 2 \cdot n + 1 \text{ and } n \in Z . \tag{3}$$

Hence, the state vector is confined in a square defined as follows:

$$\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} \in I^2 \equiv \left\{ \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} : -1 \leq x_1(k) < 1, -1 \leq x_2(k) < 1 \right\}, \text{ for } \forall k \geq 0. \tag{4}$$

By introducing the symbolic sequences, the state space model of the digital filter can further be represented as:

$$\mathbf{x}(k+1) = \begin{bmatrix} x_2(k) \\ b \cdot x_2(k) + a \cdot x_1(k) + 2 \cdot s(k) \end{bmatrix} \tag{5}$$

$$= \mathbf{A} \cdot \mathbf{x}(k) + \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot s(k), \text{ for } \forall k \geq 0, \tag{6}$$

where

$$\mathbf{A} \equiv \begin{bmatrix} 0 & 1 \\ b & a \end{bmatrix}, \tag{7}$$

and

$$s(k) \in \{-m, \cdots, -1, 0, 1, \cdots, m\}, \text{ for } k \geq 0, \tag{8}$$

in which m is the minimum integers satisfying:

$$-2 \cdot m - 1 \leq b \cdot x_1(k) + a \cdot x_2(k) < 2 \cdot m + 1, \text{ for } \forall k \geq 0. \tag{9}$$

The admissible set of periodic symbolic sequences with period M [2] is given by

$$\left\{ s : -1 \leq \frac{\sum_{j=0}^{M-1} s(\mathrm{mod}(i+j, M)) \cdot \cos\left(\left(\frac{M}{2} - j - 1\right) \cdot \theta\right)}{\sin\left(\frac{M \cdot \theta}{2}\right) \cdot \sin \theta} < 1 \right\}, \text{ for } i = 0, 1, \cdots, M-1, \tag{10}$$

where

$$\theta = \cos^{-1}\left(\frac{a}{2}\right), \tag{11}$$

$\mathrm{mod}(p, q)$ is the reminder of $\frac{p}{q}$,

and

$$s = (s(0), s(1), \cdots). \tag{12}$$

3. ADMISSIBILITY AND INVERTIBILITY OF A SECOND-ORDER DIGITAL FILTER WITH TWO'S COMPLEMENT ARITHMETIC

Let λ_1 and λ_2 be the eignevalues of $\mathbf{A}$. In this section, we assume that:

$$|\lambda_1| > 1 \tag{13}$$

and

$$|\lambda_2| > 1. \tag{14}$$

Define

$$\Sigma = \{s : s(k) \in \{-m,\cdots,-1,0,1,\cdots,m\}\}, \tag{15}$$

and

$$S : I^2 \to \Sigma . \tag{16}$$

Obviously, S is not surjective and the set Σ is not admissible.

Lemma 1:

Define $\Sigma_b \equiv \left\{ s : \sum_{j=n}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{\,n-j-1} - \lambda_1^{\,n-j-1} \\ \lambda_2^{\,n-j} - \lambda_1^{\,n-j} \end{bmatrix} \in I^2 \text{ for } n = 0,1,\cdots \right\} \subset \Sigma$, then the set Σ_b is admissible and $S_b : I^2 \to \Sigma_b$ is surjective.

Proof:

Since $\forall s \in \Sigma_b$, we have:

$$\sum_{j=n}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{\,-j-1} - \lambda_1^{\,-j-1} \\ \lambda_2^{\,-j} - \lambda_1^{\,-j} \end{bmatrix} \in I^2 . \tag{17}$$

Let

$$\mathbf{x}(0) = \sum_{j=n}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{\,-j-1} - \lambda_1^{\,-j-1} \\ \lambda_2^{\,-j} - \lambda_1^{\,-j} \end{bmatrix} . \tag{18}$$

Then

$$\mathbf{x}(k) = \mathbf{A}^k \cdot \mathbf{x}(0) + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot s'(j) \tag{19}$$

$$= \sum_{j=0}^{k-1} \frac{2 \cdot (s'(j) - s(j))}{\lambda_2 - \lambda_1} \cdot \begin{bmatrix} \lambda_2^{\,k-j-1} - \lambda_1^{\,k-j-1} \\ \lambda_2^{\,k-j} - \lambda_1^{\,k-j} \end{bmatrix} + \sum_{j=k}^{+\infty} \frac{2 \cdot s(j)}{\lambda_2 - \lambda_1} \cdot \begin{bmatrix} \lambda_2^{\,k-j-1} - \lambda_1^{\,k-j-1} \\ \lambda_2^{\,k-j} - \lambda_1^{\,k-j} \end{bmatrix}, \text{ for } \forall k \geq 0 . \tag{20}$$

Since

$$\sum_{j=k}^{+\infty} \frac{2 \cdot s(j)}{\lambda_2 - \lambda_1} \cdot \begin{bmatrix} \lambda_2^{\,k-j-1} - \lambda_1^{\,k-j-1} \\ \lambda_2^{\,k-j} - \lambda_1^{\,k-j} \end{bmatrix} \in I^2, \text{ for } \forall k \geq 0 , \tag{21}$$

$$|\lambda_1| > 1 , \tag{22}$$

$$|\lambda_2| > 1 , \tag{23}$$

and

$$\mathbf{x}(k) \in I^2 \text{, for } \forall k \geq 0 \text{,} \tag{24}$$

we have

$$s'(j) = s(j) \text{, for } j = 0,1,\cdots,k-1 \text{ and } \forall k \geq 0 \text{.} \tag{25}$$

This implies

$$s' = s \tag{26}$$

and

$$S(\mathbf{x}(0)) = s \in \Sigma_b \text{.} \tag{27}$$

Hence, the set Σ_b is admissible and $S_b : I^2 \to \Sigma_b$ is surjective. This completes the proof. ■

Lemma 2:

$S_b : I^2 \to \Sigma_b$ is injective.

Proof:

Let

$$\mathbf{x}^1(0), \mathbf{x}^2(0) \in I^2 \text{.} \tag{28}$$

Assume

$$\mathbf{x}^1(0) \neq \mathbf{x}^2(0) \tag{29}$$

and

$$S_b(\mathbf{x}^1(0)) = S_b(\mathbf{x}^2(0)) = s \in \Sigma_b \text{.} \tag{30}$$

Since

$$\mathbf{x}^1(k) = \mathbf{A}^k \cdot \mathbf{x}^1(0) + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot s(j) \text{, for } \forall k \geq 0 \text{,} \tag{31}$$

and

$$\mathbf{x}^2(k) = \mathbf{A}^k \cdot \mathbf{x}^2(0) + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot s(j) \text{, for } \forall k \geq 0 \text{,} \tag{32}$$

we have

$$\mathbf{x}^1(k)-\mathbf{x}^2(k)=\mathbf{A}^k\cdot\left(\mathbf{x}^1(0)-\mathbf{x}^2(0)\right),\ \text{for}\ \forall k\geq 0. \tag{33}$$

Since

$$|\lambda_1|>1, \tag{34}$$

$$|\lambda_2|>1, \tag{35}$$

and

$$\mathbf{x}^1(k),\mathbf{x}^2(k)\in I^2,\ \text{for}\ \forall k\geq 0, \tag{36}$$

we have

$$\mathbf{x}^1(0)=\mathbf{x}^2(0). \tag{37}$$

This contradicts equation (29). Hence, S_b is injective, and completing the proof. ■

Remark 1:

According to Lemma 1 and 2, S_b is bijective.

Lemma 3:

Define $T_b:\Sigma_b\rightarrow I^2$, then T_b is bijective and

$$T_b(s)=\sum_{j=0}^{+\infty}\frac{2\cdot s(j)}{\lambda_1-\lambda_2}\cdot\begin{bmatrix}\lambda_2^{-j-1}-\lambda_1^{-j-1}\\ \lambda_2^{-j}-\lambda_1^{-j}\end{bmatrix}.$$

Proof:

By applying similar methods in Lemma 1 and 2, we can easily prove that T_b is bijective. To show

$$T_b(s)=\sum_{j=0}^{+\infty}\frac{2\cdot s(j)}{\lambda_1-\lambda_2}\cdot\begin{bmatrix}\lambda_2^{-j-1}-\lambda_1^{-j-1}\\ \lambda_2^{-j}-\lambda_1^{-j}\end{bmatrix}, \tag{38}$$

since $\forall\mathbf{x}(0)\in I^2$, we have

$$\mathbf{x}(k)=\mathbf{A}^k\cdot\mathbf{x}(0)+\sum_{j=0}^{k-1}\mathbf{A}^{k-1-j}\cdot\begin{bmatrix}0\\2\end{bmatrix}\cdot s(j),\ \text{for}\ \forall k\geq 0. \tag{39}$$

This implies that

$$\mathbf{x}(0)=\left(\mathbf{A}^{-1}\right)^k\cdot\mathbf{x}(k)-\sum_{j=0}^{k-1}\mathbf{A}^{-1-j}\cdot\begin{bmatrix}0\\2\end{bmatrix}\cdot s(j),\ \text{for}\ \forall k\geq 0. \tag{40}$$

$$\Rightarrow \mathbf{x}(0) = \lim_{k\to+\infty} \left(\mathbf{A}^{-1}\right)^k \cdot \mathbf{x}(k) - \lim_{k\to+\infty} \sum_{j=0}^{k-1} \mathbf{A}^{-1-j} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot s(j) \tag{41}$$

$$= \sum_{j=0}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{-j-1} - \lambda_1^{-j-1} \\ \lambda_2^{-j} - \lambda_1^{-j} \end{bmatrix}. \tag{42}$$

This completes the proof. ■

Remark 2:

$$\mathbf{x}(n) = \sum_{j=n}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{n-j-1} - \lambda_1^{n-j-1} \\ \lambda_2^{n-j} - \lambda_1^{n-j} \end{bmatrix}, \text{ for } \forall n \geq 0.$$

Proof:

Since

$$\mathbf{x}(k) = \mathbf{A}^k \cdot \mathbf{x}(0) + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j} \cdot \begin{bmatrix} 0 \\ 2 \end{bmatrix} \cdot s(j), \text{ for } \forall k \geq 0, \tag{43}$$

and

$$T_b(s) = \mathbf{x}(0) = \sum_{j=0}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{-j-1} - \lambda_1^{-j-1} \\ \lambda_2^{-j} - \lambda_1^{-j} \end{bmatrix}, \tag{44}$$

we have

$$\mathbf{x}(n) = \sum_{j=n}^{+\infty} \frac{2 \cdot s(j)}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{n-j-1} - \lambda_1^{n-j-1} \\ \lambda_2^{n-j} - \lambda_1^{n-j} \end{bmatrix}, \text{ for } \forall n \geq 0. \tag{45}$$

■

Remark 3:

For a one-dimensional case, any number $x \in [-1,1)$ can be represented as an M-ary number with each bit $b(j) \in \{1-M, \cdots, -1, 0, 1, \cdots, M-1\}$, that is:

$$x = \sum_{j=0}^{+\infty} b(j) \cdot M^{-(1+j)}. \tag{46}$$

Define

$$\bar{b} = \{b(0), b(1), \cdots\}, \tag{47}$$

and

$$\Sigma_{one} = \{\overline{b}\}. \tag{48}$$

Since

$$\sum_{j=0}^{+\infty} \frac{M-1}{M^{j+1}} = 1, \tag{49}$$

$$\Rightarrow -1 \le \sum_{j=0}^{+\infty} \frac{b(j)}{M^{j+1}} \le 1. \tag{50}$$

Hence, the mapping

$$S_{one} : [-1,1) \to \Sigma_{one} \tag{51}$$

is surjective.

It is well known that S_{one} is injective, so S_{one} is bijective. However, this is not true for the two-dimensional case. Since:

$$\sum_{j=0}^{+\infty} \frac{2 \cdot m}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{-j-1} - \lambda_1^{-j-1} \\ \lambda_2^{-j} - \lambda_1^{-j} \end{bmatrix} = \frac{2 \cdot m}{(\lambda_1 - 1) \cdot (\lambda_2 - 1)} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{52}$$

and

$$|a + b| < 2 \cdot m + 1, \tag{53}$$

we have

$$\frac{2 \cdot m}{(\lambda_1 - 1) \cdot (\lambda_2 - 1)} > 1. \tag{54}$$

Hence, S is not surjective and the set Σ is not admissible. However, if we confine the set Σ by its subset Σ_b, then we guarantee that there exists $\mathbf{x}(0) \in I^2$. Hence, S_b is surjective and the set Σ_b is admissible.

Although an infinite number of bits is required to represent x with infinite precision, we may truncate the representation by a finite number of bits and the quantization error is bounded by the magnitude represented by the last bit. That is

$$\sum_{j=k}^{+\infty} \frac{M-1}{M^{j}} = \frac{1}{M^{k-1}}. \tag{55}$$

However, for the two-dimensional case, the truncation error is

$$\begin{bmatrix} e_1(k) \\ e_2(k) \end{bmatrix} = \sum_{j=k}^{+\infty} \frac{2 \cdot m}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{-j-1} - \lambda_1^{-j-1} \\ \lambda_2^{-j} - \lambda_1^{-j} \end{bmatrix} - \frac{2}{\lambda_1 - \lambda_2} \cdot \begin{bmatrix} \lambda_2^{-k} - \lambda_1^{-k} \\ \lambda_2^{1-k} - \lambda_1^{1-k} \end{bmatrix}. \tag{56}$$

Since

$$\lim_{k \to +\infty} e_i(k) = 0 \tag{57}$$

and

$$e_i(k) > 0 \tag{58}$$

for $i = 1,2$, $\exists k_0 \in Z^+$ such that $e_i(k)$ for $i = 1,2$, are monotonically decreasing with respect to k for $k \geq k_0$. Hence, we still can truncate the representation of $\mathbf{x}(0)$ using a finite number of symbols.

This property suggests that an information can be coded using the successive approximation technique. Compared to the existing successive approximation coding technique, the traditional one is to code the information directly, while this coding technique is to code the symbolic sequences. The security is improved.

4. CONCLUSION

In this letter, we have extended the results on the admissible set of symbolic sequences of a marginally stable second-order system to an unstable system. Based on this result, the initial conditions can be computed by the symbolic sequences directly. Moreover, the truncation error of the representation of an initial condition due to the use of finite number of symbols is studied.

ACKNOWLEDGEMENT

The work described in this letter was substantially supported by The Hong Kong Polytechnic University, partly form a studentship and partly by a research grant with account number G-YD26.

REFERENCES

[1] L. O. Chua and T. Lin, "Chaos in digital filters," *IEEE Transactions on Circuits and Systems*, vol. 35, no. 6, pp. 648-658, June 1988.

[2] L. O. Chua and T. Lin, "Fractal pattern of second-order non-linear digital filters: a new symbolic analysis," *International Journal of Circuit Theory and Applications*, vol. 18, pp. 541-550, 1990.

[3] Z. Galias and M. J. Ogorzalek, "On symbolic dynamics of a chaotic second-order digital filter," *International Journal of Circuit Theory and Applications*, vol. 20, pp. 401-409, 1992.

[4] C. W. Wu and L. O. Chua, "Properties of admissible symbolic sequences in a second-order digital filter with overflow non-linearity," *International Journal of Circuit Theory and Applications*, vol. 21, pp. 299-307, 1993.