Elimination of Overflow Oscillations in Fixed-Point State-Space Digital Filters With Saturation Arithmetic: An LMI Approach

Haranath Kar and Vimal Singh

***Abstract*—A novel, linear-matrix inequality (LMI) based, criterion for the nonexistence of overflow oscillations in fixed-point state-space digital filter employing saturation arithmetic is presented. The criterion is based on a unique characterization (as prevailing in the filter under consideration) of the saturation nonlinearities, namely, an "effective" reduction of the sector.**

***Index Terms*—Asymptotic stability, digital filters, finite-wordlength effects, nonlinear systems.**

I. Introduction

THIS PAPER deals with the problem of global asymptotic stability of digital filters in state-space realization using saturation arithmetic. Specifically, the system under consideration is

$$\begin{aligned}\boldsymbol{x}(r+1) &= \boldsymbol{f}(\boldsymbol{y}(r)) \\ &= [f_1(y_1(r)) \quad f_2(y_2(r)) \quad \cdots \quad f_n(y_n(r))]^T \quad \text{(1a)}\\ \boldsymbol{y}(r) &= [y_1(r) \quad y_2(r) \quad \cdots \quad y_n(r)]^T \\ &= \boldsymbol{A}\boldsymbol{x}(r) \quad \text{(1b)}\end{aligned}$$

where $\boldsymbol{x}(r)$ is an n-vector state, $\boldsymbol{A} = [a_{ij}]$ is the $n \times n$ coefficient matrix, and T denotes the transpose. The saturation nonlinearities given by

$$f_i(y_i(r)) = \begin{cases} 1, & y_i(r) > 1 \\ y_i(r), & -1 \le y_i(r) \le 1 \\ -1, & y_i(r) < -1 \end{cases} \tag{1c}$$

where $i = 1, 2, \ldots, n$, are under consideration. Throughout the paper, it is understood that $\boldsymbol{A}$ is stable, i.e.,

$$\det(z\boldsymbol{I_n} - \boldsymbol{A}) \neq 0, \quad \text{for all } |z| \ge 1 \tag{1d}$$

where $\boldsymbol{I_n}$ denotes the $n \times n$ identity matrix.

Equation (1) may be used to describe a class of discrete-time dynamical systems with symmetric state saturation, which include digital filters using saturation overflow arithmetic, digital control systems with saturation nonlinearities, a class of neural networks, and so on.

A few criteria have been proposed for the nonexistence of overflow oscillations in system (1). In this context, [1] and [2] may be specially mentioned. According to [1], system (1) is globally asymptotically stable if

$$\begin{bmatrix} \boldsymbol{P} - \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A} & \boldsymbol{A}^T(\boldsymbol{P} - \boldsymbol{C}^T) \\ (\boldsymbol{P} - \boldsymbol{C})\boldsymbol{A} & \boldsymbol{C} + \boldsymbol{C}^T - \boldsymbol{P} \end{bmatrix} \ge \boldsymbol{0} \tag{2a}$$

where $\boldsymbol{P} = \boldsymbol{P}^T > \boldsymbol{0}$ and $\boldsymbol{C} = [c_{ij}]$ is an $n \times n$ matrix characterized by

$$c_{ii} = \sum_{j=1, j\neq i}^{n} \gamma_{ij}, \quad i = 1, 2, \ldots, n \tag{2b}$$

$$c_{ij} = -\gamma_{ij}\lambda_{ij}, \quad i, j = 1, 2, \ldots, n \quad (i \neq j) \tag{2c}$$

and $\gamma_{ij}, \lambda_{ij}$ are real numbers such that

$$\gamma_{ij} > 0, \quad i, j = 1, 2, \ldots, n \quad (i \neq j) \tag{2d}$$

$$|\lambda_{ij}| < 1, \quad i, j = 1, 2, \ldots, n \quad (i \neq j). \tag{2e}$$

On the other hand, according to [2], system (1) is globally asymptotically stable if

$$\boldsymbol{H} - \boldsymbol{A}^T\boldsymbol{H}\boldsymbol{A} \ge \boldsymbol{0} \tag{3a}$$

$$h_{ii} \ge \sum_{j=1, j\neq i}^{n} |h_{ij}|, \quad i = 1, 2, \ldots, n \tag{3b}$$

where $\boldsymbol{H} = \boldsymbol{H^T} > \boldsymbol{0}$.

It may be noted that neither (2) nor (3) is a linear matrix inequality (LMI) [3], [4]. In other words, both (2) and (3) are computationally demanding. In this paper, we present an LMI based criterion for the global asymptotic stability of system (1). The presented criterion makes use of an unique characterization (as prevailing in the system under consideration) of the saturation nonlinearities, namely, an "effective" reduction of the sector. The criterion may uncover some new $\boldsymbol{A}$ [i.e., not covered by (2) and (3)] for which (1) is globally asymptotically stable.

II. Main Results

In view of (1), we have

$$|x_i(r)| \le 1, \quad i = 1, 2, \ldots, n. \tag{4}$$

Define

$$k_i = \sum_{j=1}^{n} |a_{ij}|, \quad i = 1, 2, \ldots, n. \tag{5}$$

Now, we prove the following result.

Manuscript received November 3, 2002; revised May 28, 2003. This paper was recommended by Associate Editor E. A. B. da Silva.

The authors are with the Department of Electrical Electronics Engineering, Atilim University, Incek, Ankara 06836, Turkey (e-mail: hnkar1@ rediffmail.com; vsingh11@rediffmail.com).

Digital Object Identifier 10.1109/TCSII.2003.821526

Theorem 1: The system described by (1) is globally asymptotically stable if

$$\|A\|_\infty \leq 1. \tag{6}$$

Proof: Condition (6) implies that

$$k_i \leq 1, \quad i = 1, 2, \ldots, n. \tag{7}$$

Therefore

$$\begin{aligned} |y_i(r)| &= \left| \sum_{j=1}^{n} a_{ij} x_j(r) \right| \\ &\leq \sum_{j=1}^{n} |a_{ij}||x_j(r)| \\ &\leq \sum_{j=1}^{n} |a_{ij}| \\ &= k_i \\ &\leq 1 \quad i = 1, 2, \ldots, n \end{aligned} \tag{8}$$

where use has been made of (4) and (7). Thus, no overflow occurs in (1) if (7) holds true. In other words, with the condition expressed by (6), system under consideration reduces to a linear stable system. This completes the Proof of Theorem 1. □

The following lemma is needed in the proof of our next theorem.

Lemma 1: If

$$k_i > 1, \quad i = 1, 2, \ldots, m \tag{9a}$$

$$k_i \leq 1, \quad i = m+1, m+2, \ldots, n \tag{9b}$$

then the global asymptotic stability of (1) is equivalent to that of (10), shown at the bottom of the page.

Proof: Using similar steps as in the Proof of Theorem 1, it follows, from (9b), that

$$|y_i(r)| \leq 1, \quad i = m+1, m+2, \ldots, n. \tag{11}$$

Thus

$$f_i(y_i(r)) = y_i(r), \quad i = m+1, m+2, \ldots, n. \tag{12}$$

This completes the Proof of Lemma 1. □

Next, we have the following result.

Theorem 2: The system described by (1) and (9) is globally asymptotically stable if there exists an $n \times n$ positive definite matrix $\boldsymbol{P} = \boldsymbol{P}^T$ and an $m \times m$ positive definite diagonal matrix $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \ldots, d_m)$ such that

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{P} - \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A} & \boldsymbol{A}^T(\boldsymbol{P}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{D}\boldsymbol{K}) \\ (\boldsymbol{B}^T\boldsymbol{P} - \boldsymbol{K}\boldsymbol{D}\boldsymbol{B}^T)\boldsymbol{A} & 2\boldsymbol{D} - \boldsymbol{B}^T\boldsymbol{P}\boldsymbol{B} \end{bmatrix} > \boldsymbol{0} \tag{13a}$$

where

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{I}_m \\ \cdots \\ \boldsymbol{0} \end{bmatrix} \in \mathrm{R}^{n \times m} \tag{13b}$$

$$\boldsymbol{K} = \mathrm{diag}\left(1 - \frac{1}{k_1}, 1 - \frac{1}{k_2}, \ldots, 1 - \frac{1}{k_m}\right). \tag{13c}$$

Proof: From Lemma 1, it follows that the global asymptotic stability of the system described by (1) and (9) is equivalent to that of system (10). Define

$$\begin{aligned} \boldsymbol{e}_I(\boldsymbol{y}_I(r)) &= [e_1(y_1(r)) \quad e_2(y_2(r)) \quad \cdots \quad e_m(y_m(r))]^T \\ &= \boldsymbol{y}_I(r) - \boldsymbol{f}_I(\boldsymbol{y}_I(r)). \end{aligned} \tag{14}$$

Substituting $\boldsymbol{f}_I(\boldsymbol{y}_I(r))$ from (14) in (10a) and noting (10b) yields

$$\boldsymbol{x}(r+1) = \boldsymbol{A}\boldsymbol{x}(r) - \boldsymbol{B}\boldsymbol{e}_I(\boldsymbol{y}_I(r)) \tag{15}$$

where $\boldsymbol{B}$ is defined in (13b). Now, consider a quadratic Lyapunov function

$$v(\boldsymbol{x}(r)) = \boldsymbol{x}^T(r)\boldsymbol{P}\boldsymbol{x}(r). \tag{16}$$

Application of (16) to (15) yields

$$\begin{aligned} \Delta v(\boldsymbol{x}(r)) &= v(\boldsymbol{x}(r+1)) - v(\boldsymbol{x}(r)) \\ &= \boldsymbol{x}^T(r)[\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A} - \boldsymbol{P}]\boldsymbol{x}(r) - \boldsymbol{x}^T(r) \\ &\quad \times \boldsymbol{A}^T\boldsymbol{P}\boldsymbol{B}\boldsymbol{e}_I(\boldsymbol{y}_I(r)) \\ &\quad - \boldsymbol{e}_I^T(\boldsymbol{y}_I(r))\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{A}\boldsymbol{x}(r) \\ &\quad + \boldsymbol{e}_I^T(\boldsymbol{y}_I(r))\boldsymbol{B}^T\boldsymbol{P}\boldsymbol{B}\boldsymbol{e}_I(\boldsymbol{y}_I(r)). \end{aligned} \tag{17}$$

$$\boldsymbol{x}(r+1) = \begin{bmatrix} \boldsymbol{f}_I(\boldsymbol{y}_I(r)) \\ \cdots\cdots\cdots \\ \boldsymbol{y}_{II}(r) \end{bmatrix} \tag{10a}$$

$$\begin{aligned} \boldsymbol{y}(r) &= \begin{bmatrix} \boldsymbol{y}_I(r) \\ \cdots\cdots\cdots \\ \boldsymbol{y}_{II}(r) \end{bmatrix} \\ &= \begin{bmatrix} y_1(r) & y_2(r) & \cdots & y_m(r) \vdots y_{m+1}(r) & y_{m+2}(r) & \cdots & y_n(r) \end{bmatrix}^T \\ &= \boldsymbol{A}\boldsymbol{x}(r) \end{aligned} \tag{10b}$$

$$\boldsymbol{f}_I(\boldsymbol{y}_I(r)) = [f_1(y_1(r)) \quad f_2(y_2(r)) \quad \cdots \quad f_m(y_m(r))]^T. \tag{10c}$$

Now consider the quantity "β" given by

$$\begin{aligned}\beta &= \sum_{i=1}^{m} 2d_i[y_i(r) - f_i(y_i(r))]\left[f_i(y_i(r)) - \frac{y_i(r)}{k_i}\right] \\ &= \boldsymbol{x}^T(r)\boldsymbol{A}^T\boldsymbol{BDK}\boldsymbol{e}_I(\boldsymbol{y_I}(r)) + \boldsymbol{e}_I^T(\boldsymbol{y_I}(r))\boldsymbol{KDB}^T\boldsymbol{Ax}(r) \\ &\quad - 2\boldsymbol{e}_I^T(\boldsymbol{y_I}(r))\boldsymbol{D}\boldsymbol{e}_I(\boldsymbol{y_I}(r)). \end{aligned} \tag{18}$$

For the nonlinearities given by (1c), β is nonnegative.

Equation (17) can be rearranged as

$$\Delta v(\boldsymbol{x}(r)) = -\left[\boldsymbol{x}^T(r) \vdots \boldsymbol{e}_I^T(\boldsymbol{y_I}(r))\right] \times \boldsymbol{M} \begin{bmatrix} \boldsymbol{x}(r) \\ \cdots\cdots\cdots \\ \boldsymbol{e}_I(\boldsymbol{y_I}(r)) \end{bmatrix} - \beta \tag{19}$$

which satisfies

$$\Delta v(\boldsymbol{x}(r)) \le 0 \tag{20}$$

if (13) holds true. This completes the Proof of Theorem 2. □

Theorem 3: If

$$k_i > 1, \quad i = 1, 2, \ldots, n \tag{21a}$$

then system (1) is globally asymptotically stable provided there exists an $n \times n$ positive definite matrix $\boldsymbol{P} = \boldsymbol{P}^T$ and an $n \times n$ positive definite diagonal matrix $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ such that

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{P} - \boldsymbol{A}^T\boldsymbol{PA} & \boldsymbol{A}^T(\boldsymbol{P} - \boldsymbol{D}\tilde{\boldsymbol{K}}) \\ (\boldsymbol{P} - \tilde{\boldsymbol{K}}\boldsymbol{D})\boldsymbol{A} & 2\boldsymbol{D} - \boldsymbol{P} \end{bmatrix} > \boldsymbol{0} \tag{21b}$$

where

$$\tilde{\boldsymbol{K}} = \mathrm{diag}\left(1 - \frac{1}{k_1}, 1 - \frac{1}{k_2}, \ldots, 1 - \frac{1}{k_n}\right). \tag{21c}$$

Proof of Theorem 3 is similar to that of Theorem 2 and is, therefore, omitted.

Remark 1: The conditions expressed by (13) and (21) are LMI [3], [4].

Remark 2: The presented criteria (Theorem 2 and Theorem 3) make use of the fact that, pertaining to the system under consideration, the saturation nonlinearities $f_i(y_i(r))$ can be viewed as confined to the sector $[1/k_i, 1], 0 < 1/k_i \le 1$, rather than being viewed [5] as confined to the sector $[0, 1]$.

III. Comparative Evaluation

As shown in [6], pertaining to second-order digital filter, (2) leads to

$$|a_{11} - a_{22}| < 2\min(|a_{12}|, |a_{21}|) + 2 - |\mathrm{tr}(\boldsymbol{A})| \tag{22a}$$

if

$$|\mathrm{tr}(\boldsymbol{A})| \ge 2\det(\boldsymbol{A}). \tag{22b}$$

On the other hand, (3) yields [7]

$$|a_{11} - a_{22}| \le 2\min(|a_{12}|, |a_{21}|) + 1 - \det(\boldsymbol{A}). \tag{23}$$

Now consider a specific example of second-order digital filter with

$$\boldsymbol{A} = \begin{bmatrix} 1.2 & -2.8 \\ 0.1 & 0 \end{bmatrix}. \tag{24}$$

Both (22) and (23) are violated for this example.

In the example under consideration, $k_1 = 4 > 1, k_2 = 0.1 < 1$ and, consequently, Theorem 2 can be applied to test the global asymptotic stability. Choosing $\boldsymbol{D} = [4/3]$ and

$$\boldsymbol{P} = \begin{bmatrix} 1 & -3 \\ -3 & 10 \end{bmatrix} \tag{25}$$

which is positive definite, it is found that the matrix in (13a)

$$\boldsymbol{M} = \begin{bmatrix} 0.18 & -0.48 & -0.3 \\ -0.48 & 2.16 & 0 \\ -0.3 & 0 & 1.6666 \end{bmatrix} \tag{26}$$

is positive definite. Thus, according to Theorem 2, the filter under consideration is globally asymptotically stable.

The above example, therefore, illustrates that, for some $\boldsymbol{A}$, the new criterion may lead to results not covered by (2) and (3).

IV. Conclusion

A criterion for the nonexistence of overflow oscillations in fixed-point state-space digital filters employing saturation arithmetic is established. The criterion takes the form of LMI and, hence, is computationally tractable.

References

[1] V. Singh, "Elimination of overflow oscillations in fixed-point state-space digital filters using saturation arithmetic," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 814–818, June 1990.

[2] D. Liu and A. N. Michel, "Asymptotic stability of discrete-time systems with saturation nonlinearities with applications to digital filters," *IEEE Trans. Circuits Syst. I*, vol. 39, pp. 798–807, Oct. 1992.

[3] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.

[4] *LMI Control Toolbox—For Use with Matlab*, The MATH Works Inc., 1995.

[5] V. Singh, "A new realizability condition for limit cycle free state-space digital filters employing saturation arithmetic," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 1070–1071, Oct. 1985.

[6] H. Kar and V. Singh, "A new criterion for the overflow stability of second-order state-space digital filters using saturation arithmetic," *IEEE Trans. Circuits Syst. I*, vol. 45, pp. 311–313, Mar. 1998.

[7] J. H. F. Ritzerfeld, "A condition for the overflow stability of second-order digital filters that is satisfied by all scaled state-space structures using saturation," *IEEE Trans. Circuits Syst.*, vol. 36, pp. 1049–1057, Aug. 1989.