Digital Signal Processing 28 (2014) 136–143

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

An improved criterion for the global asymptotic stability of fixed-point state-space digital filters with combinations of quantization and overflow

Neha Agarwal *, Haranath Kar

Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, India

ARTICLE INFO

Article history:
Available online 18 February 2014

Keywords:
Asymptotic stability
Digital filter
Finite wordlength effect
Lyapunov method
Nonlinear system

ABSTRACT

This paper deals with the problem of global asymptotic stability of fixed-point state-space digital filters under various combinations of quantization and overflow nonlinearities and for the situation where quantization occurs after summation only. Utilizing the structural properties of the nonlinearities in greater detail, a new global asymptotic stability criterion is proposed. A unique feature of the presented approach is that it exploits the information about the maximum normalized quantization error of the quantizer and the maximum representable number for a given wordlength. The approach leads to an enhanced stability region in the parameter-space, as compared to several previously reported criteria.

1. Introduction

Due to a broad range of applications of digital filters in many areas such as telecommunications, speech processing, image processing, medical electronics, geophysics, control systems, spectrum and vibration analyses, adaptive and tracking systems, etc., the design and analysis of various types of filters have received considerable attention [1–10] over the past few decades.

In the implementation of a stable and linear recursive digital filter using finite wordlength processors with fixed-point arithmetic, nonlinearities are introduced owing to the quantization and overflow [11,12]. Quantization nonlinearities in digital filters usually take one of the following three forms: magnitude truncation, round off and value truncation. The usual types of overflow nonlinearities employed in practice are zeroing, triangular, two's complement and saturation overflow characteristics. Due to the presence of such nonlinearities, the filter may exhibit unstable behavior (e.g., granular limit cycles, overflow oscillations etc.). When dealing with the design and implementation of fixed-point state-space digital filters, it is essential to have a criterion for choosing the values of the filter coefficients so that the designed filter is free of limit cycles. If the digital filter is implemented on a digital hardware with sufficiently long wordlength, the effects of quantization and overflow may be treated as decoupled or mutually independent [11,12]. Under this decoupling approximation, a number of researchers have studied quantization effects in digital filters without considering overflow effects [13–18], while others have investigated overflow effects in the absence of quantization [19–32]. On the other hand, several researchers have also queried the validity of decoupling approximation [33,34].

The study of the stability properties of digital filters under the combined effects of quantization and overflow nonlinearities appears to be more realistic due to the fact that the operations of practical digital filters take place in presence of both types of nonlinearities [35–41]. The problem of global asymptotic stability of fixed-point state-space digital filters with combinations of quantization and overflow has attracted the attention of a few researchers [35–41]. A few computationally simple criteria have been reported in [38] under which the global asymptotic stability of a system involving quantization and overflow nonlinearities is shown to be equivalent to the global asymptotic stability of that with only quantization nonlinearities. The key idea of [38] is based on the determination of stability regions where overflow cannot occur. In practice, the results in [38] are helpful to determine the stability of digital filters with combinations of quantization and overflow nonlinearities using available stability information for digital filters with only quantization nonlinearities. Ref. [40] presents improved versions of some of the stability results given in [38]. An alternative approach for stability analysis of digital filters with various combinations of quantization (except value truncation) and overflow nonlinearities has also been established in [40]. The state-space criterion in [40] is based on the sector information of the combined quantization and overflow nonlinearities where overflow nonlinearities are assumed for every state in the system. However,

* Corresponding author. Fax: +91 532 2545341.
E-mail addresses: nehaagarwal.mnnit@gmail.com (N. Agarwal), hnkar1@rediffmail.com (H. Kar).

http://dx.doi.org/10.1016/j.dsp.2014.02.006
1051-2004/

when dealing with the dynamic behavior of digital filters in presence of finite wordlength nonlinearities, it is worth noticing that all the states in the system may not reach to the overflow level. Depending on the values of system parameters, [41] makes distinction between the nonlinearities operating as composite nonlinearities and those operating as only quantization nonlinearities. The criterion in [41] has been derived using a more precise characterization of the nonlinearities as compared to [40] and, therefore, [41] turns out to be an improvement over [40]. It may be observed that [41] utilizes only the values of system parameters in order to identify the nonlinearities operating in the quantization region but ignores the information about the maximum normalized quantization error of the quantizer and the maximum representable number for a given wordlength. Consequently, there still remains scope for characterizing the composite nonlinearities properly with a view to achieve enhanced stability region in the parameter-space.

Motivated by the preceding discussion, in this paper, we revisit the problem of global asymptotic stability of fixed-point state-space digital filters under the influence of various combinations of quantization and overflow. Larger the global asymptotic stability region in the parameter-space of a digital filter, greater would be the flexibility to choose the filter coefficients. It remains a challenge to reduce the conservatism of the existing criteria for testing the stability of digital filters, particularly in the presence of both quantization and overflow. Therefore, we aim to develop an approach which may lead to less conservative stability conditions than those in [38,40,41]. Inspired by the works of [30,31], an improved method for extracting the composite nonlinearities which effectively operate only in quantization region is developed. This method not only captures the information of the dynamical behavior of the system more efficiently as compared to [41] but also makes use of the information about the maximum normalized quantization error of the quantizer and the maximum representable number for a given wordlength. This paper is organized as follows. Section 2 provides the notations used in this paper and introduces the system under consideration along with the existing criteria. A new sufficient condition for the nonexistence of limit cycles in fixed-point state-space digital filters under various combinations of quantization and overflow is established in Section 3. In Section 4, a comparison of the proposed criterion with previously reported criteria [38,40,41] is made in order to establish the significance of the present approach. Finally, the paper is concluded in Section 5.

2. System description and existing criteria

The following notations are used throughout in this paper:

$R^{p\times q}$	set of $p\times q$ real matrices
R^p	set of $p\times 1$ real vectors
x_i	i-th entry of a vector $\boldsymbol{x}$
J_k	list of coordinate indexes, which we regard as a subset of $\{1,2,\ldots,n\}$; the subscript k refers to the iteration number
n_k	number of indexes in J_k
J_k^c	complement set of J_k such that $J_k\cup J_k^c=\{1,2,\ldots,n\}$; the superscript c stands for the complement
$\boldsymbol{x}_{J_k}$	the vector which consists of x_i's with $i\in J_k$
$\boldsymbol{M}_{J_k,J_k}$	submatrix of $\boldsymbol{M}$ composed of $\boldsymbol{M}_{i,j}$'s with (i,j) belonging to $J_k\times J_k$
$\boldsymbol{M}_{J_k,J_k^c}$	submatrix of $\boldsymbol{M}$ composed of $\boldsymbol{M}_{i,j}$'s with (i,j) belonging to $J_k\times J_k^c$
ϕ	null set
$\boldsymbol{I}_n$	identity matrix of order n
$\boldsymbol{1}_n$	$n\times 1$ vector with all elements being equal to 1
$\boldsymbol{0}$	null matrix or null vector of appropriate dimension
δ	maximum normalized quantization error
$N_{\max}$	maximum representable number for a given wordlength
$\rho(\boldsymbol{A})$	spectral radius of a matrix $\boldsymbol{A}$
Z_+	set of positive integers
T	superscript 'T' refers to the transpose
$diag(d_{11},d_{22},\ldots,d_{nn})$	diagonal matrix with diagonal elements $d_{11},d_{22},\ldots,d_{nn}$
$\boldsymbol{B}>\boldsymbol{0}$	$\boldsymbol{B}$ is positive definite symmetric matrix

The system under consideration is represented by

$$\begin{aligned}\boldsymbol{x}(r+1)&=\boldsymbol{O}\{\boldsymbol{Q}(\boldsymbol{y}(r))\}\\&=\big[O_1\{Q_1(y_1(r))\}\quad O_2\{Q_2(y_2(r))\}\quad\ldots\\&\qquad O_n\{Q_n(y_n(r))\}\big]^T\\&=\big[f_1(y_1(r))\quad f_2(y_2(r))\quad\ldots\quad f_n(y_n(r))\big]^T\\&=\boldsymbol{f}(\boldsymbol{y}(r)),\end{aligned}\tag{1a}$$

$$\boldsymbol{y}(r)=\boldsymbol{A}\boldsymbol{x}(r),\tag{1b}$$

where $\boldsymbol{x}(r)\in R^n$ denotes the state vector, $\boldsymbol{A}=[a_{ij}]\in R^{n\times n}$ the coefficient matrix, $\boldsymbol{Q}(.)$ the quantization nonlinearities, $\boldsymbol{O}(.)$ the overflow nonlinearities and $\boldsymbol{f}(.)$ is the composite nonlinear function. In the event of $\boldsymbol{Q}(.)$ being either magnitude truncation or roundoff, the composite nonlinearities $f_i(y_i(r))$, $i=1,2,\ldots,n$, involved in (1) may be viewed as confined to the sector $[K_o,K_q]$, i.e.,

$$f_i(0)=0,\qquad K_o y_i^2(r)\leqslant f_i(y_i(r))y_i(r)\leqslant K_q y_i^2(r),\quad i=1,2,\ldots,n,\tag{2a}$$

where

$$K_o=\begin{cases}0, & \text{for zeroing or saturation}\\ -1/3, & \text{for triangular}\\ -1, & \text{for two's complement}\end{cases}\tag{2b}$$

$$K_q=\begin{cases}1, & \text{for magnitude truncation}\\ 2, & \text{for round off.}\end{cases}\tag{2c}$$

It is assumed that the infinite precision counterpart of the system is stable, i.e.,

$$\det(z\boldsymbol{I}_n-\boldsymbol{A})\neq 0,\quad \forall |z|\geqslant 1.\tag{3}$$

Note that quantization nonlinearities satisfy

$$|e_i(r)|=|Q_i(y_i(r))-y_i(r)|\leqslant\delta,\quad i=1,2,\ldots,n,\tag{4}$$

where $\boldsymbol{e}(r)=\boldsymbol{Q}(\boldsymbol{y}(r))-\boldsymbol{y}(r)$ is the quantization error vector and δ represents the maximum normalized quantization error. Due to overflow bound in (1a), it follows that

$$|x_i(r)|\leqslant N_{\max},\quad i=1,2,\ldots,n,\tag{5}$$

where $N_{\max}$ is the maximum representable number for a given wordlength.

Eqs. (1)–(3) represent a class of digital filters with different combinations of quantization and overflow nonlinearities where quantization occurs after summation only. The problem is to choose the multipliers a_{ij}, $i,j=1,2,\ldots,n$, such that the filter is free of limit cycles. For this, some existing criteria [40,41], which ensure the global asymptotic stability (thereby, imply the absence of any kind of instability including limit cycles), may be utilized. Pertaining to the system described by (1) and (2), the following result is reported in [41].