DIXIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS

NICE du 20 au 24 MAI 1985

ADVANTAGEOUS ESTIMATION FORMULA OF THE LEAST ROUNDOFF NOISE FOR THE CASCADE FIXED-POINT DIGITAL FILTERS

S.KAWARAI

Anritsu Electric Co., Ltd.   1800 Onna, Atsugi-shi, 243 Japan

RESUME

Lorsqu'un filtre digital à point fixe est exécuté par cascade des sections de deuxième ordre sous contraintes de plage dynamique, le bruit d'arrondissage résultant causé par la longueur de mot finie dépend beacoup du pairage de zéro de pôle et de l'ordre des sections. Deux procédures principales d'optimisation pour la réduction du bruit d'arrondissage sont disponibles, à savoir la procédure de programmation dynamique et celle heuristique. La procédure heuristique est plus pratique que celle de programmation dynamique, car sa durée de calcul est nettement plus courte et une solution presque optimale peut s'obtenir. C'est un grand avantage pour la procédure heuristique de présenter un critère pouvant juger si la solution presque optimale est acceptable. Pour la première fois, nous avons introduit un tel critère sous forme d'une formule d'estimation. Cette formule présente une limite inférieure de variation la plus importante possible du bruit d'arrondissage de sortie, et l'estimation peut être obtenue à partir de cette formule à travers un simple calcul. Nous avons étudié dans ce document la différence entre l'estimation et la solution exacte pour tous genres de filtres avec calculs d'ordinateur.

Les estimations peuvent être calculées immédiatement et sont proches des solutions exactes correspondantes avec des différences de moins de 1,5 dB. Nous sommes convaincus que cette formule est très avantagause pour n'importe quelle procédure heuristique ayant des difficultés à trouver la solution exacte.

SUMMARY

When a fixed-point digital filter is realized by cascading second-order sections under dynamic range constraints, the resulting roundoff noise due to the finite word-length is highly dependent on the pole-zero pairing and ordering of the sections. There are two main optimization procedures for the minimization of the roundoff noise, that is, dynamic programming and heuristic procedures. The heuristic procedure is much more practical than the dynamic programming procedure because of requiring a considerably smaller computing time and gives a near optimal solution. It is very beneficial for the heuristic procedure to have a criterion which judges if the near optimal solution is tolerable. We have introduced, for the first time, such a criterion in the form of an estimation formula. The formula presents the greatest possible lower bound of the variance of the output roundoff noise and the estimate can be obtained from the formula with a simple calculation. In this paper, the difference between the estimate and the exact solution is investigated for all kinds of filters with computer calculations.

The estimates are readily computable and very close to the corresponding exact solutions with the differences of less than 1.5 dB. It has been convinced that the formula is very advantageous for any heuristic procedure which can scarcely find the exact solution.

ADVANTAGEOUS ESTIMATION FORMULA OF THE LEAST ROUNDOFF NOISE FOR THE CASCADE FIXED-POINT DIGITAL FILTERS

INTRODUCTION

This paper discusses the validity and usefulness of an entirely new estimation formula in the literature [1]. The formula can evaluate a near optimal solution of minimization problem for the roundoff noise in the cascade fixed-point digital filters.

When a fixed-point digital filter is realized by cascading second-order sections under dynamic range constraints, the resulting roundoff noise due to the finite word-length is highly dependent on the pole-zero pairing and ordering of the sections. Hence, the analysis and minimization of the output roundoff noise from the cascade fixed-point digital filters have been the subject of various papers [1]–[7].

There are two main optimization procedures for the minimization problem : dynamic programming procedure [2], [3] and heuristic procedure [1], [4] – [7]. The optimization procedure using the principle of dynamic programming produces the exact optimal solution. However, the computation time required becomes prohibitive, even when a moderate number of second-order sections are involved. So, the dynamic programming procedure is impractical. On the other hand, the heuristic procedure is much more practical than the dynamic programming procedure because of requiring a considerably smaller computing time and gives a near optimal solution. It cannot, however, be known how close the near optimal solution is to the exact one. Hence, it is very beneficial for the heuristic procedure to have a criterion which judges if the near optimal solution is tolerable. We have introduced, for the first time, such a criterion in the form of an estimation formula [1]. The formula presents the greatest possible lower bound of the variance of the output roundoff noise and the estimate can be obtained from the formula with a simple calculation. In [1], we did not discuss the validity of the formula in detail.

In this paper, the difference between the estimate and the exact solution is investigated for all kinds of filters with computer calculations.

ESTIMATION FORMULA OF THE LEAST ROUNDOFF NOISE

Let the given transfer function be

$$H^*(z) = a_0 \prod_{i=1}^{L} \frac{\alpha_i^*(z)}{\beta_i^*(z)} = a_0 \prod_{i=1}^{L} \frac{\alpha_{0i} + \alpha_{1i} z^{-1} + \alpha_{2i} z^{-2}}{1 + \beta_{1i} z^{-1} + \beta_{2i} z^{-2}} \tag{1}$$

which is realized in the cascade form with the proper scalings [4]. The two most commonly employed configurations for the cascade form are of the 1D and 2D forms shown in Figs. 1(a) and 1(b), respectively. The noise flow graphs of the cascade 1D and 2D forms are obtained from Figs. 1(a) and 1(b), as shown in Figs. 2(a) and 2(b), respectively. $e_i(n)(i=1\sim L+1)$ in the Figure are the noises generated due to product quantizations in the second-order sections, and the small circles and the solid dots represent summation nodes and branch nodes, respectively. The scaling factors s_i $(i=1\sim L)$ are readily determined by imposing the dynamic range constraints at branch nodes in the cascade 1D form as follows :

$$s_i = 1 \Big/ \left\| \frac{1}{\beta_i} \prod_{k=1}^{i-1} \frac{\alpha_k}{\beta_k} \right\|_p \quad (i=1 \sim L ; p \geqq 1) \tag{2a}$$

while for the cascade 2D form

$$s_i = 1 \Big/ \left\| \prod_{k=1}^{i} \frac{\alpha_k}{\beta_k} \right\|_p \quad (i=1 \sim L ; p \geqq 1) \tag{2b}$$

where $\| \cdot \|_p$ denotes the L_p norm, defined for an arbitrary periodic function $X(\cdot)$ with period ω_s by

$$\| X \|_p = \left[\frac{1}{\omega_s} \int_0^{\omega_s} |X(\omega)|^p \, d\omega \right]^{\frac{1}{p}} \tag{3}$$

for each real p $\geqq 1$. $\alpha_k(\omega) \triangleq \alpha_k^*(e^{j\omega Ts})$, $\beta_k(\omega) \triangleq \beta_k^*(e^{j\omega Ts})$ and T_s $(=2\pi/\omega_s)$ is the sampling period. The values of p are fixed to be 2 and ∞ for random and deterministic inputs, respectively.

We shall present the estimation formula of the least output roundoff niose from a digital filter in the cascade form. Referring to Fig. 2, the variance of the output roundoff noise $e(n)$ from a digital filter in the cascade 1D form is given by

$$E[\{e(n)\}^2] = \sigma_0^2 \left\{ k_{L+1} + \sum_{i=1}^{L} k_i A_i^p (\alpha_1, \ldots, \alpha_L, \beta_1, \ldots, \beta_L) \right\} \tag{4a}$$

while for the cascade 2D form

$$E[\{e(n)\}^2] = \sigma_0^2 \left\{ 1 + \sum_{i=1}^{L} k_i B_i^p (\alpha_1, \ldots, \alpha_L, \beta_1, \ldots, \beta_L) \right\} \tag{4b}$$

where

$$A_i^p(\alpha_1, \ldots, \beta_1, \ldots) \triangleq a_0^2 \left\| \frac{1}{\beta_i} \prod_{j=1}^{i-1} \frac{\alpha_j}{\beta_j} \right\|_p^2 \left\| \prod_{j=i}^{L} \frac{\alpha_j}{\beta_j} \right\|_2^2 , \tag{5a}$$

$$B_i^p(\alpha_1, \ldots, \beta_1, \ldots) \triangleq a_0^2 \left\| \prod_{j=1}^{i} \frac{\alpha_j}{\beta_j} \right\|_p^2 \left\| \frac{1}{\beta_i} \prod_{j=i+1}^{L} \frac{\alpha_j}{\beta_j} \right\|_2^2 , \tag{5b}$$

σ_0^2 is the variance of the noise from each rounding operation and k_i is the number of noise sources inputting to the i-th summation node.

A discussion of the estimation of the least roundoff noise will be presented on the basis of (4) and (5).

Hölder's inequality has the following relation for $p, q \geqq 1$:

$$\|f\|_p \cdot \|g\|_q \geqq \|fg\|_1 \quad \left(\frac{1}{p} + \frac{1}{q} = 1 \right) \tag{6}$$

where $f(\cdot) \in L_p(0, \omega_s)$ and $g(\cdot) \in L_p(0, \omega_s)$.
Applying the Hölder's inequality (6) to (5), we have the following relaitons for random inputs ($p = 2$):

$$A_i^2(\alpha_1, \ldots, \beta_1, \ldots) \geqq \left\| \frac{H}{\beta_i} \right\|_1^2 \triangleq R_1(\beta_i) \tag{7a}$$

$$B_i^2(\alpha_1, \ldots, \beta_1, \ldots) \geqq \left\| \frac{H}{\beta_i} \right\|_1^2 = R_1(\beta_i) \tag{7b}$$

When $p = \infty$, on the other hand, the left hand side of (6) becomes

$$\begin{aligned} \|f\|_\infty \cdot \|g\|_q &= \max_{0 \leqq \omega \leqq \omega_s} |f(\omega)| \cdot \|g\|_q = \| \max |f(\omega)| \cdot g \|_q \\ &\geqq \|fg\|_q \quad (q \geqq 1) \end{aligned} \tag{8}$$

Using the inequality (8), the following relations are obtained from (5) for deterministic inputs ($p = \infty$):

$$A_i^\infty(\alpha_1, \ldots, \beta_1, \ldots) \geqq \left\| \frac{H}{\beta_i} \right\|_2^2 \triangleq R_2(\beta_i) \tag{9a}$$

$$B_i^\infty(\alpha_1, \ldots, \beta_1, \ldots) \geqq \left\| \frac{H}{\beta_i} \right\|_2^2 = R_2(\beta_i) \tag{9b}$$

From (4), (7) and (9), one can obtain the following inequality:

$$E[\{e(n)\}^2] \geqq \sigma_0^2 \Gamma_j \quad (j = 1, 2) \tag{10}$$

where

$$\Gamma_j \triangleq \min_{\beta's} \left\{ k_{L+1} + \sum_{i=1}^{L} k_i R_j(\beta_i) \right\} (j = 1, 2) \tag{11}$$

and $k_{L+1} = 1$ for the 2D form. In the above, „$\underset{\beta's}{\min}$" means the minimum value of the argument for all possible orderings of β's and then Γ_j is uniquely determined.

It follows from Eq. (10) that the value of Γ_j presents the greatest possible lower bound of the variance of the output roundoff noise normalized by σ_0. In other words, the exact solution equals to Γ_j or lies between the near optimal solution and Γ_j. So we call Eq. (11) an estimation formula, in which the estimate Γ_j shows how far the near optimal solution is away from the exact one at most. According to Eq. (11), it is also known that Γ_j can be readily obtained with a simple calculation.

A similar estimation formula can be introduced for FIR digital filters.

COMPUTER IMPLEMENTATIONS

Computer calculations have been made to investigate the validity and usefulness of the estimate Γ_j. In this paper, we illustrate, for simplicity, the difference between the estimate Γ_1 and the exact solution for the cascade 1D form with L_2 scaling. Sample filters in use are all kinds of elliptic filters of 6th to 12th orders. The estimate Γ_1 and the exact solution are tabulated in Table I. The result for highpass filter is equal to that for lowpass filter and is ignored from the Table.

It is known from the Table that all of the estimates are so close to the corresponding exact solutions that the differences between them are less than 1.5 dB. Similar results will be obtained for the cascade 1D form with L_∞ scaling and the cascade 2D form.

CONCLUSION

We have investigated the validity and usefulness of the estimation formula of the least roundoff noise for the cascade fixed-point digital filters. The estimates are readily computable and very close to the corresponding exact solutions with the differences of less than 1.5 dB.

It has been convinced that the formula is very advantageous for any heuristic procedure which can scarcely find the exact solution.

ACKNOWLEDGMENTS

The author would like to thank Dr. S. Hidaka for his encouragement of the research and T. Sumi for many beneficial discussions.

REFERENCES

[1] S. Kawarai, "A new approach to the optimization of cascade fixed-point digital filters", IEEE 1984 ISCAS Proceedings, pp.246-249, May 1984.

[2] S.Y.Hwang, "On optimization of cascade fixed-point digital filters", IEEE Trans. Circuits Syst.,vol. CAS-21, pp. 163-166, Jan. 1974.

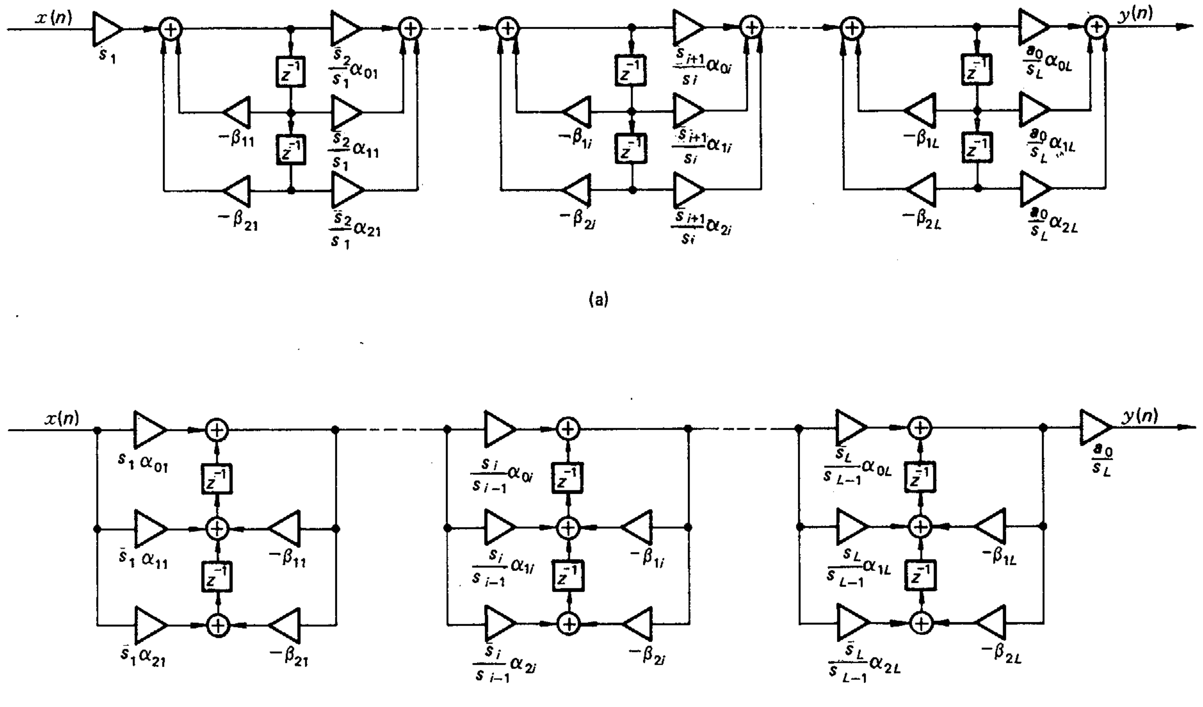
[3] E.Lueder, "Minimizing the roundoff noise in digital filters by dynamic programming", Frequenz, 29, pp.211-214, 1975.

[4] L.B.Jackson, "Roundoff-noise analysis for fixed-point digital filters realized in cascade or parallel form", IEEE Trans. Audio Electroacoust., vol. AU-18, pp.107-122, June 1970.

[5] W.S.Lee, "Optimization of digital filters for low roundoff noise", IEEE Trans. Circuits Syst., vol. CAS-21, pp.424-431, May 1974.

[6] G.Dehner, "A contribution to the optimization of roundoff-noise in recursive digital filters", Arch. Elektron & Uebertragungstech., 29, 12, pp.505-510, Dec. 1975.

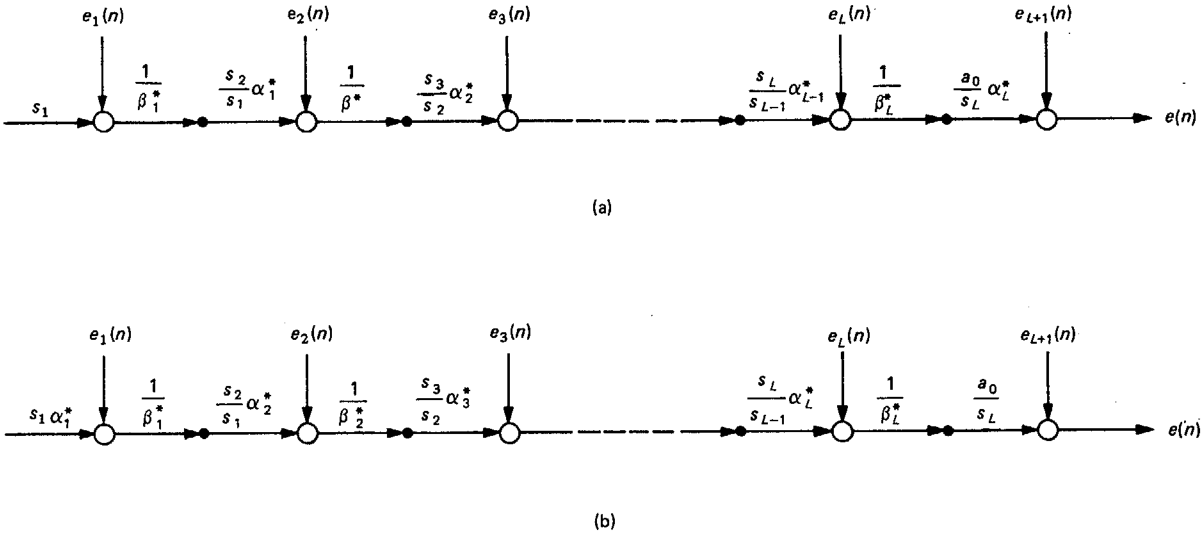
[7] B.Liu and A.Peled, "Heuristic optimization of the cascade realization of fixed-point digital filters", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23, pp.464-473, Oct. 1975.


x(n)
s_1
z^{-1}
$\frac{s_2}{s_1}\alpha_{01}$
$-\beta_{11}$
$\frac{s_2}{s_1}\alpha_{11}$
$-\beta_{21}$
$\frac{s_2}{s_1}\alpha_{21}$
$\frac{s_{i+1}}{s_i}\alpha_{0i}$
$-\beta_{1i}$
$\frac{s_{i+1}}{s_i}\alpha_{1i}$
$-\beta_{2i}$
$\frac{s_{i+1}}{s_i}\alpha_{2i}$
$\frac{a_0}{s_L}\alpha_{0L}$
$-\beta_{1L}$
$\frac{a_0}{s_L}\alpha_{1L}$
$-\beta_{2L}$
$\frac{a_0}{s_L}\alpha_{2L}$
y(n)
$s_1\alpha_{01}$
$s_1\alpha_{11}$
$-\beta_{11}$
$s_1\alpha_{21}$
$-\beta_{21}$
$\frac{s_i}{s_{i-1}}\alpha_{0i}$
$\frac{s_i}{s_{i-1}}\alpha_{1i}$
$-\beta_{1i}$
$\frac{s_i}{s_{i-1}}\alpha_{2i}$
$-\beta_{2i}$
$\frac{s_L}{s_{L-1}}\alpha_{0L}$
$\frac{s_L}{s_{L-1}}\alpha_{1L}$
$-\beta_{1L}$
$\frac{s_L}{s_{L-1}}\alpha_{2L}$
$-\beta_{2L}$
$\frac{a_0}{s_L}$

(a)

(b)

Fig. 1 Configurations for the cascade form.
(a) 1D form. (b) 2D form.



(a)

(b)

Fig. 2 Noise flow graphs of the cascade form.
(a) 1D form. (b) 2D form.

TABLE I
Results for the 1D Form with L_2 Scaling

(Unit in dB)

Filter's Order	Lowpass Filter			Bandpass Filter			Bandstop Filter		
	Estimate Γ_1	Exact Solution	Difference	Estimate Γ_1	Exact Solution	Difference	Estimate Γ_1	Exact Solution	Difference
6	11.12	11.55	0.43	10.23	11.29	1.06	12.48	13.96	1.48
8	13.08	13.75	0.67	11.75	13.06	1.31	14.31	15.63	1.32
9	13.14	14.02	0.88	—	—	—	—	—	—
10	17.44	18.06	0.62	10.78	11.66	0.88	13.84	14.52	0.68
11	17.29	18.19	0.90	—	—	—	—	—	—
12	18.98	19.67	0.69	11.76	12.74	0.98	14.84	15.56	0.72