REDUCEREA ERORII DE ROTUNJIRE ÎN ARITMETICA CU VIRGULĂ FIXĂ

conf. dr. ing. Ovidiu Radu
dr. ing. Alexandru Vasile
Universitatea Politehnica *Bucureşti*

Rezumat: Una din tehnicile utilizate pentru reducerea erorilor de rotunjire la filtrele digitale recursive o constituie reacţia de eroare (RE). Aceasta corespunde faptului că evaluarea unui polinom prin metoda Hornor este echivalentă cu un filtru recursiv, de ordinul întâi. RE este o tehnică utilizată pentru a reduce erorile de rotunjire, care apar în evaluarea unei funcţii prin aproximare polinomială.

Cuvinte cheie: reacţie de eroare, eroare de rotunjire, aproximare polinomială, filtru recursiv, funcţie exponenţială, funcţie logaritmică.

1. Introducere

Când se evaluează unele funcţii, ca cele de tip exponenţial, se utilizează în mod uzual aproximarea polinomială. Un polinom este aproximat în mod curent prin metoda Hornor, similar funcţionării unui filtru recursiv de ordinul întâi.

Pe de altă parte, aprecierea erorilor spectrului (ES) prin reacţia de eroare (RE) este o tehnică importantă, care poate reduce erorile de rotunjire, provocate de efectele lungimii finite a cuvântului asupra filtrelor digitale recursive cu virgulă fixă.

În consecinţă, se întrevede posibilitatea reducerii erorii prin tehnica RE, când se evaluează o funcţie prin aproximarea polinomială, folosind aritmetica cu virgulă fixă.

În continuare, se arată în partea a II a, relaţia dintre metoda Hornor şi un filtru recursiv, de ordinul întâi cu RE.

În partea a III a, se prezintă câteva exemple de evaluare a unor funcţii exponenţiale şi logaritmice cu RE.

2. Metoda Hornor şi filtrul recursiv cu RE

Un polinom, f(x), de ordinul N, poate fi evaluat prin metoda Hornor, astfel:

$$f(x)=(\ldots((c_N x+c_{N-1})x+c_{N-2})x+\ldots+c_1)x+c_0, \quad (1)$$

Ecuaţia (1) poate fi evaluată prin algoritm recursiv, astfel:

f[-1]:=0

pentru k:=0 la N , f[k]:=x*f[k-1]+c[N-k].

În finalul acestei iteraţii, se obţine f[N]. În acest algoritm, dacă se consideră că c[N-k] este semnalul de intrare, f[k] este semnalul de ieşire şi x este un coeficient, atunci el reprezintă ecuaţia unui filtru recursiv de ordinul întâi. Adică, ieşirea de ordinul N a acestui filtru reprezintă evaluarea unui polinom f(x), de ordinul N.

Se ştie că, dacă se implementează un filtru recursiv prin aritmetică cu virgulă fixă, pot apare erori, dacă rezultatele multiplicărilor sunt cuantizate, [1], [2].

RE reprezintă una din tehnicile utilizate pentru a reduce erorile de cuantizare.

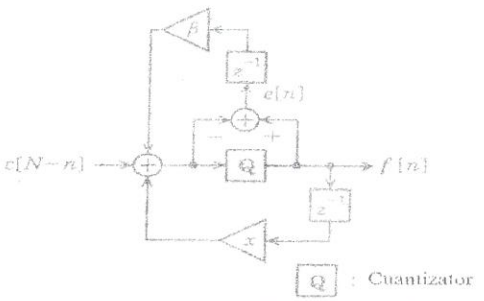

Figura 1. Filtru recursiv de ordinul întâi cu reacţie de eroare

În figura 1, se dă schema bloc a unui filtru recursiv de ordinul întâi, cu RE, unde e[n] este eroarea de cuantizare şi β reprezintă coeficientul reacţiei de eroare.

Dacă se limitează β la partea întreagă, pentru a nu creşte numărul de operaţii aritmetice, se arată în [3] că cele mai bune valori sunt :

$$\beta = \begin{cases} ...0.....|x| \le 0.5 \\ -1.....0.5 < x < 1 \\ ..1.....-1 < x \le -0.5 \end{cases} \qquad (2)$$

În consecinţă, algoritmul pentru evaluarea RE este următorul:

f[-1]: =0

e[-1]: =0

pentru k: =0 la N

begin

w: =x* f [k-1] + c [N- k] + β * e [k – 1] ;

f [k] : = Q (w) ;

e [k] : = f [k] – w ;

end

unde Q ( ) reprezintă o funcţie de cuantizare.

3. Exemple pentru evaluarea funcţiilor cu RE

Se vor prezenta două exemple de evaluare a funcţiilor, prin polinoame cu RE:

- exponenţială, 2^{x-1}, cu $0 \le x < 1$;
- logaritmică, $\log_2 (x+1)$, cu $0 \le x < 1$;

A. Aproximări polinomiale

Se presupune că evaluarea este efectuată de un procesor de 16 biţi cu virgulă fixă, că partea fracţionară a datelor este de 15 biţi şi că lungimea rezultatului operaţiilor de tipul Ax + B→C din multiplicator este 16 b × 16 b + 32 b → 32 b.

Deoarece partea fracţionară a datelor este de 15 biţi, se presupune că mărimea maximă a erorii absolute tolerate este valoarea cifrei 1 a bitului cel mai puţin semnificativ (LSB), adică $2^{-15} = 3{,}0518 \cdot 10^{-5}$

Setul de coeficienţi pentru aproximări polinomiale se obţine astfel:

1. Se calculează setul de coeficienţi prin minimizarea aproximării maxime, utilizând calculul cu virgulă mobilă şi precizie dublă; maximul erorii amplitudinii poate fi inferior valorii $3{,}0518 \cdot 10^{-5}$.
2. Se cuantizează setul de coeficienţi la 15 biţi. Această cuantizare se efectuează prin rotunjire sus sau jos, astfel încât să minimizeze eroarea maximă a amplitudinii, când se evaluează polinoame, utilizând calculul cu virgulă mobilă şi precizie dublă.
3. Dacă eroarea maximă a amplitudinii pentru un set de coeficienţi cuantizaţi este mai mare decât $3{,}0518 \cdot 10^{-5}$, se măreşte ordinul polinoamelor de aproximare şi se reia punctul 1.

În continuare, sunt prezentate polinoame de aproximare şi coeficienţii lor; aceşti coeficienţi sunt daţi prin fracţii pentru a putea fi mai precis memoraţi în procesor.

1. Exponenţial: 2^{x-1}, $0 \le x < 1$

$$2^{x-1} = \sum_{k=0}^{4} c_k x^k$$

$c_1 = 11354/32768$
$c_2 = 3959/32768$
$c_3 = 847/32768$ (3)
$c_4 = 224/32768$

2. Logaritmic: $\log_2(x+1)$, $o \le x < 1$

$$\log_2(x+1) = \sum_{k=0}^{6} c_k x^k$$

$c_0 = 0/32768$
$c_1 = 47268/32768$
$c_2 = 23520/32768$
$c_3 = 14959/32768$
$c_4 = -9062/32768$
$c_5 = 3965/32768$
$c_6 = -842/32768$ (4)

Eroarea absolută maximă a amplitudinii pentru (3) şi (4) este $0{,}609 \cdot 10^{-5}$, respectiv $0{,}869 \cdot 10^{-5}$, când evaluarea se face prin calcul cu virgulă mobilă şi precizie dublă.

B. Exemple de evaluare a funcţiilor

Coeficientul reacţiei de eroare este ales –1 sau 0, conform cu (2). În consecinţă, se va utiliza rotunjirea, pentru cuantizare în figura 1.

În figura 2, se arată eroarea lui 2^{x-1} funcţie de x, fără RE, (2a) şi cu RE, (2b). Erorile sunt calculate pentru $2^{13} = 8192$ valori ale lui x, distribuite uniform, în intervalul $0 \le x < 1$.

Din figura 2, rezultă că erorile amplitudinii depăşesc valoarea 2^{-15}, crescând o dată cu apropierea lui x de valoarea 1.

Din figura 2b, reiese că erorile amplitudinii nu depăşesc valoarea 2^{-15} şi chiar se reduc, când $x=1$. Rezultă eficienţa utilizării RE.

În figura #, este reprezentată distribuţia erorii lui 2^{x-1} pentru orice x, care este reprezentat prin cuvinte de 16 biţi lungime, adică pentru $2^{15}=32768$ valori uniform distribuite în intervalul $0 \le x < 1$.

Figura 3a, (fără RE), arată că există cuvinte ale căror erori în amplitudine sunt mai mari decât 2^{-15}.

Figura 3b, (cu RE), arată că nu există cuvinte a căror erori în amplitudine să depăşească 2^{-15}.

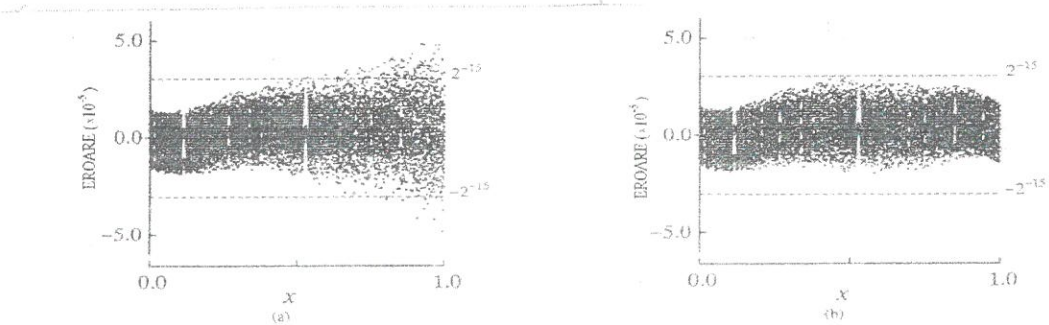

Figura 2. Eroarea funcţiei 2^{x-1}: (a) fără RE; (b) cu RE

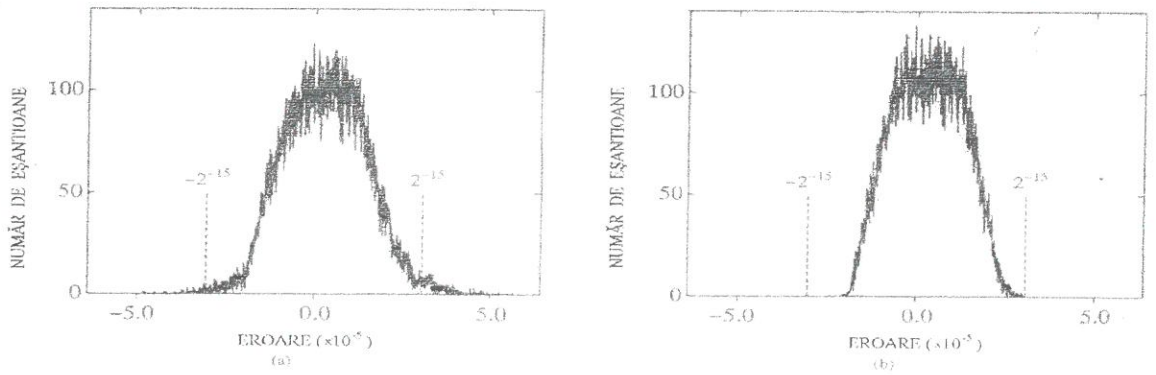

Figura 3. Distribuţia erorii funcţiei 2^{x-1}: (a) fără RE; (b) cu RE

Tabelul I

	2^{x-1}		$\log_2(x+1)$	
	Cu RE	Fără RE	Cu RE	Fără RE
Eroarea maximă a amplitudinii($\times 10^{-5}$)	2,978	5,655	3,324	6,363
Abaterea relativă a datelor(%)	0,00	1,46	0,05	2,59

Tabelul I cuprinde eroarea maximă a amplitudinii şi abaterea relativă a datelor faţă de valoarea 2^{-15}.

Rezultatele prezentate mai sus arată că, atunci când se evaluează o funcţie prin aproximare polinomială utilizând aritmetică cu virgulă fixă, RE poate fi o bună metodă pentru anumite funcţii de reducere a erorilor care apar în operaţiile aritmetice.

Concluzii

1. Algoritmul de evaluare a unui polinom folosind metoda Hornor este echivalent cu realizarea unui filtru recursiv de ordinul întâi.
2. Reacţia de eroare, care este una din metodele utilizate pentru reducerea erorilor care apar în operaţiile aritmetice ale filtrelor recursive, poate fi aplicată cu succes pentru reducerea erorilor de rotunjire, în evaluarea funcţiilor prin aproximarea polinomială.

Bibliografie

1. **RADU, O.** : Filtre numerice . Aplicaţii, Editura Tehnică, Bucureşti, 1979.
2. **CAVICCHI, T. J.**: Digital Signal Processing, John Wiley, New York, 2000.
3. **ELLIOT, D. F.** Handbook of Digital Signal Processing, New York Academic, 1987.