A NEW APPROACH TO FINITE WORDLENGTH COEFFICIENT FIR DIGITAL FILTER DESIGN USING THE BRANCH AND BOUND TECHNIQUE

A.N. Belbachir[1,2], B. Boulerial[1] & M.F. Belbachir[1]

[1]Signal and System Laboratory, Electronic Institute U.S.T.O.
B.P. 1505 El Mnouer, Oran –ALGERIA
[2]Vienna University of Technology, Pattern Recognition and Image Processing Group
Favoritenstr. 9/1832, A-1040 Vienna – AUSTRIA
E-mail: (belbachiran@yahoo.com) or (nabil@prip.tuwien.ac.at)

ABSTRACT

It has been shown that the branch and bound technique is effective for the design of finite wordlength optimal digital filters. This technique is however expensive in computing time. In this paper, we present a robust branch and bound branching strategy named Sequential and Progressive Search, improving the design of filters on a large wordlength processor in a reasonable computing cost. The details of the algorithm and many examples are given and compared to the other methods.

I. INTRODUCTION

For finite wordlength coefficients digital filter design, it is often desirable to use algorithms whose output quality can be adjusted depending on the availability of resources such as computing time and precision.
Remez Exchange Algorithm is usually applied for the design of infinite precision linear phase (FIR) filters [6]. When these filters are implemented on a Digital Signal Processor with a special purpose-hardware, each filter coefficient has to be represented by a finite number of bits (bwl) smaller than that used on a computer. The simplest and the most widely used approach to the problem are the rounding of the optimal infinite precision coefficients to its (bwl) bits representation. However, the filters obtained are degraded and in most case there exists another set of finite word length coefficients, which gives the best Chebyshev approximation to the desired frequency response. To find these coefficients, it is necessary to include the finite word length restriction into the filter design. In this case, the optimisation problem becomes a complex problem, where a general investigation of the optimal solution requires a prohibitive computing time. To solve this problem, many optimisation methods have been applied to discrete coefficients FIR digital filters design. Simulated annealing technique [2,3,4] has proven to be effective in many cases, but requires a large number of function evaluations and does not guarantee the optimal solution.
The linear integer programming formulation [1], [7]-[9] was applied as a discrete optimisation method on the minmax criterion. Although, it is possible to obtain an optimum result, the computing time required even with the high-speed supercomputer of today, prohibits the application of these techniques for high order filters.
Optimisation technique in the coefficients discrete space, and in particular branch and bound method, was used to solve this discrete optimum problem. This method based on implicit enumeration techniques, also requires an expensive computing cost [7,8,10,14,15].
In many local search methods based on the branch and bound technique, such as the Depth First Search (DFS) and the Breadth First Search (BFS), the solution found is better than that obtained from the direct quantizing from infinite precision filter. However, the computing cost is expensive for large wordlength processor and high order filters [10], [14], [15]. Even for (BFS) the solution could not be optimal. This is related to the estimation accuracy of the branch susceptible to contain the best solution.
In this paper, we present a different view on the branching strategy and present a new method named « Sequential and Progressive Search » (SPS), based on the branch and bound technique in the minmax sense. Compared to the Depth First Search method (DFS), the number of function evaluations is smaller and it depends on the filter length without degrading the performance of the algorithm.
In section II, we present the problem statements and the characteristics of the error criterion chosen. In section III, the description of the proposed optimisation method, the sequential and progressive search method (SPS) is given. The results reported on section IV deal with conventional minmax optimisation of FIR digitals filters and are compared to those of other methods.

II. PROBLEM STATEMENTS

Let us consider the design of N-1 order linear phase FIR digital filter with a frequency response H (f) usually written as

$$H(f) = \sum_{k=0}^{N-1} h_k e^{-j2\pi fk} \qquad (1)$$

In [5], It was shown that the frequency response amplitude of the four cases of linear phase (FIR) filters could be written in the form of:

$$P_n(f) = \sum_{k=0}^{n-1} a_k \cos 2\pi fk \qquad (2)$$

Where the number of terms, n, is:

$$n = N/2 \text{ or } (N-1)/2 \text{ or } (N+1)/2$$

and a_k is the resulting shifted sequence depending on the considered case. The function $P_n(f)$ is compared with desired frequency response amplitude D(f) using a minmax criterion, as done in the usual optimal (FIR) filter design with infinite precision [6]. The weighted approximation error e_n is given by

$$e_n = \min_{(coeff.a)} \max_{f \in F} W(f)\left|D(f) - P_n(f)\right| \qquad (3)$$

- F: the union of all the frequency bands of interest.
- W (f): a weighting function defined on F.
- D (f): the desired frequency response amplitude.

Using Eq. (2) in Eq. (3) gives

$$e_n = \min_{(coeff.a)} \max_{f \in F} W(f)\left|D(f) - \sum_{k=0}^{n-1} a_k \cos 2\pi fk\right| \qquad (4)$$

The filter coefficients are restricted to the discrete values allowed by (bwl) bit binary word length.

III. OPTIMIZATION METHOD 'SEQUENTIAL AND PROGRESSIVE SEARCH' (SPS)

We consider the problem of filter design with an extra constraint imposing a limit on the word length of the coefficients a_k, k=0,1,...., N/2. Using the fixed point representation, we can express the discrete coefficient a_k as a linear combination:

$$|a_k| = \sum_{j=1}^{bwl-1} y_{j,k} 2^{-j} \qquad k = 0, 1,...., N/2. \qquad (5)$$

Where (bwl) is the binary bit allowed for the filter discrete design and 'j' is the binary bit indication. $y_{j,k}$ is a bivalent variable only fixed to only take the values '0' or '1'. Hence, the frequency response amplitude $P_n(f)$ could be expressed as

$$P_n(f) = \sum_{=0}^{-1} s\left(\sum_{j=1}^{bwl-1} y_{j,k} 2^{-j}\right) \cos 2\pi fk. \qquad (6)$$

Where s is the sign of 'a_k' , s (= -1 or +1).
It is not possible to find the optimal filter coefficients at (bwl) bits wordlength processor using the DFS method, owing to the long computing time required. But we can calculate these filter coefficients filter in a lower wordlength, (bbN) binary bit (bbN ≤ bwl) with such a method. a_k could be expressed as

$$|a_{k(opt)}| = \sum_{j=1}^{bb-1} y_{j,k(opt)} 2^{-j} \qquad k = 0, 1,...., N/2. \qquad (7)$$

$a_{k(opt)}$: the coefficients related to the optimal digital filter at (bbN) wordlength.
$y_{j,k(opt)}$: 'i' binary bit value of the $a_{k(opt)}$ coefficient.

The proposed method, named Sequential and Progressive Search (SPS) takes this optimal solution as a starting point (starting solution) for its branching strategy in order to design filters with a higher wordlength discrete coefficients. Using the optimal solution in the minmax sense found by the (DFS) method in the (bbN) binary bits wordlength, we calculate with the SPS algorithm the solution at (bwl) bits (bbN ≤ bwl).
First, the coefficients are found at the (bbN+1) binary bits on the minimax sense using a local investigation in a reduced discrete search space. This reduced space is defined using the previous solution. Then we increase gradually the wordlength and calculate the new solution till reaching the (bwl) (the required word length). For each step 'i', we define a lower bounding function $e_{n\,i}$, which can be written as

$$e_{n\,i}(a) = \min_{(coeff.a)} \max_{f \in F} W(f)\left|D(f) - \sum_{k=0}^{n-1} a_k \cos 2\pi fk\right| \qquad (8)$$

$e_{n\ i}(a)$ is defined as the lowest value of the error function for a_k solving the following program which satisfy (8) under the conditions:

$$a_k^{bbN+1} \le a_k^{bbN} + \mu \quad k=1,..., N/2 \qquad (9a)$$

$$a_k^{bbN+1} \ge a_k^{bbN} - \mu \quad k=1,..., N/2 \qquad (9b)$$

a_k^{bbN} : is the coefficients a_k represented by bbN bits
μ interval chosen to contain the solution.

The implementation of a continuous value in two different wordlengths of the fixed point representation does not offer two equal discrete values. The maximum difference between these discrete values is the quantization error '±LSB' (least sided bit) due to both truncation (±LSB) and rounding (±1/2.LSB).
Therefore, we have overestimate the μ value between the (bbN) and (bwl) wordlengths to

$$\mu = LSB = 2^{-(bbN-1)}. \qquad (10)$$

Substituting Eq. (10) and Eq. (7) in Eq. (9)

$$\sum_{j=1}^{bbN} y_{j,k} 2^{-j} \le \sum_{j=1}^{bbN-1} y_{j,k(opt)} 2^{-j} + 2^{-(bbN-1)} \qquad (11a)$$

$$\sum_{j=1}^{bbN} y_{j,k} 2^{-j} \ge \sum_{j=1}^{bbN-1} y_{j,k(opt)} 2^{-j} - 2^{-(bbN-1)} \qquad (11b)$$

Developing Eq. (11) we obtain

$$\sum_{j=1}^{bbN-2} y_{j,k(opt)}2^{-j} + \sum_{j=bbN-1}^{bbN} y_{j,k}2^{-j} \le \sum_{j=1}^{bbN-2} y_{j,k(opt)}2^{-j} + y_{bbN-1,k(opt)}2^{-bbN} + 2^{-(bbN-1)}. \quad (12a)$$

$$\sum_{j=1}^{bbN-2} y_{j,k(opt)}2^{-j} + \sum_{j=bbN-1}^{bbN} y_{j,k}2^{-j} \ge \sum_{j=1}^{bbN-2} y_{j,k(opt)}2^{-j} + y_{bbN-1,k(opt)}2^{-bbN} - 2^{-(bbN-1)}. \quad (12b)$$

After simplifications, we have

$$y_{bbN-1,k} \cdot 2^{-bbN+1} + y_{bbN,k} \cdot 2^{-bbN} \le y_{bbN-1,k(opt)} \cdot 2^{-bbN} + 2^{-(bbN-1)} \quad (13a)$$

$$y_{bbN-1,k} \cdot 2^{-bbN+1} + y_{bbN,k} \cdot 2^{-bbN} \ge y_{bbN-1,k(opt)} \cdot 2^{-bbN} - 2^{-(bbN-1)} \quad (13b)$$

hence,

$$y_{bbN-1,k} + y_{bbN,k}2^{-1} \le y_{bbN-1,k(opt)} + 1 \quad (14a)$$

$$y_{bbN-1,k} + y_{bbN,k}2^{-1} \ge y_{bbN-1,k(opt)} - 1 \quad (14b)$$

The problem is restricted as resolving two equations with two variables (x1,x2) on the form of

$$a1x1 + a2\ x2 \le b1 \quad (15a)$$

$$a1x1 + a2\ x2 \ge b2 \quad (15b)$$

(a1, a2, b1, b2) are constants.

$a1=1$, $a2= 2^{-1}$, $b1=y_p+1$, $b2=y_p-1$,

and where y_p is the previous calculated optimal solution. (in this case $y_p= y_{bbN-1,k(opt)}$)

Hence, we obtain a small grid containing admissible values, from which we choose the solution sequence (x1, x2) which gives the smallest value of maximal weighted error.

This procedure is iterated for each coefficient increasing the word length, until reaching the desired word length, in which the design of discrete coefficients digital FIR filter was required.

Denoting that this method does not affect the bivalent variable from 1 to (bbN-2) bits obtained by the optimal method 'DFS'. Therefore, it improves the coefficient precision, adding required bits from (bbN-1) to (bwl), related to the optimal sequence, in order to well define the coefficient value.

In this paper, we have chosen the fixed point transformation as shown in (Eq5). We can show that an extension to the other binary representation such that floating point or power of two could be easily done.

VI. RESULTS

The algorithm has been tested using cases reported in literature. The software algorithm was developed in Matlab and tested on 300 MHZ Pentium machine. The results obtained are presented and compared to algorithms in [3,15]. The reference numbers indicate where the filters are taken. A filter with length 24 with 9 bits in quantization, excluded the sign bit is denoted by '24/9'.

The starting points are calculated by (DFS) in discrete coefficients wordlength of bbN=3 bits. The two first filters in the Table 1. have the passband edges (0,0.1) and the stopband edges (0.1125,0.5), the third filter has the passband edges (0,0.08) and the stopband edges (0.16,0.5), the fourth filter has the passband edges (0,0.159) and the stopband edges (0.295,0.5) and the last filter has the passband edges (0,0.307) and the stopband edges (0.35,0.5). All filters have equal weights in passbands and stopbands. In all examples the results are better than those obtained in the indicated references. The reference algorithms are Simulated Annealing (SA) [4], the Breath First Search (BFS) [15] and the Depth First Search (DFS) [15]. In table 1, the results are given both in the approximation error and the design times. A comparison measures the number of function evaluations between (DFS) and (SPS) algorithms of all filters of Table 2. The reduction in the number of function evaluations is at least in the order of 4500.

N/bwl	-I-	-II-
15/5	$8.52891033 .10^{11}$	1171875
21/6	$6.20506086 .10^{19}$	195312500
15/7	$6.76752340 .10^{16}$	1953125
20/7	$1.18059162 .10^{21}$	48828125
16/19	$1.46149048 .10^{48}$	12405426
24/9	$3.16993821 .10^{32}$	1708984375

Table 2. Number of the Function of Evaluation of the SPS Method Compared to DFS method [15].

I- Number of function evaluations by `DFS` [15].
II- Number of function evaluations by 'SPS'.

Filter length/ Word length	N	bbN	bwl	Infinite Precision	Rounded	[ref]/ Time `s` seconds	SPS/ Time seconds
8/15[15]	8	3	15	0.05766	0.05765	0.05762/ 50	0.05759/19
21/6[4]	21	3	6	0.02099	0.07223	0.07115/ 5	0.04687/ 79256
8/7[15]	8	3	7	0.05760	0.06356	0.06355/ 93600	0.06355/ 0.06
20/7 [15]	20	3	7	0.00344	0.02191	0.02005/ 10	0.016321/20256
16/19[15]	16	3	19	0.13590	0.13590	0.13961/ 1180	0.13580/11830

Table 1. Results & Comparison for Filter Design Cases Use the SPS Method.

V. CONCLUSION

In this paper, a new approach to finite wordlength coefficient (FIR) digital filter design using the branch and bound technique is presented. The main feature of this approach is its applicability to the design of filter in a processor with a large wordlength. The computing time in such processor wordlength would be prohibitive using the Depth First Search (DFS). The obtained results when compared to the other algorithms and local search methods [4], [15] and [16] are better in all cases. In the examples, the limitation of the search domain does not seem to degrade the performance of the algorithm. As a future work, the improvement of the algorithm for long filter order will be studied.

Acknowledgement

The authors acknowledge Professor Michele MARCHESI of the Electrical and Electronical Engineering Department, University of Cagliari-Italy, for his valuable suggestions to improve this paper quality.

REFERENCES

[1] D.M. Kodek, 'Design of Optimal Finite Word length FIR Digital Filters Using Integer Programming Techniques' IEEE Tr.ASSP, pp.304-308, June 1980

[2] I. PITAS, 'Optimisation and Adaptation of Discrete-Valued Digital Filter Parameters by Simulated Annealing' IEEE Trans. SP, pp. 860-866, April 1994.

[3] N. Benvenuto, M. Marchesi, ''Digital Filters Design by Simulated Annealing,'' IEEE Trans. Circuits Syst., vol. 36, pp. 459-460, March 1989.

[4] T. Ciloglu and Z. Unver, 'A New Approach to Discrete Coefficient FIR Digital Filter Design by Simulated Annealing,' IEEE of Int. Conf. on ASSP Minnisota 93.

[5] L. R. Rabiner, B. Gold, 'Theory and Application of Digital Signal Processing,' Prentice-Hall, INC. 1975.

[6] J. H. Mc Clellan, T. W. Parks, and L. R. Rabiner, '' A Computer Program for Designing Optimum FIR Linear Phase Digital Filters,'' IEEE Trans. Audio Electroacoust., vol. AU-21, pp. 506-526, Dec. 1973.

[7] M. Minoux, ''Programmation mathématique, théorie et algorithmes,'' tomes 1 et 2, Dunod, 1983.

[8] B. Jaumard, M. Minoux, and P. Siohan, 'Finite Precision Design of FIR Digital Filters Using a Convexity Property,' IEEE Trans. ASSP, pp. 407-411, Mar. 1988.

[9] Yong C. Lim, S. R. Parker, and A. G. Constantinides, '' Finite Word Length FIR Filter Design Using Integer Programming Over a Discrete Coefficient Space,'' IEEE Trans. ASSP, vol. -30, pp. 661-664, Aug. 1982.

[10] Y. C. Lim, S. R. Parker, ''FIR Filter Design Over a Discrete Powers-of-Two Coefficient Space,'' IEEE Trans. ASSP, vol. ASSP-31, pp. 583-591, June 1983.

[11] Y. C. Lim S. R. Parker,'Discrete Coefficient FIR Digital Filter Design Based Upon an LMS Criteria,'' IEEE Trans. Circ. Syst., vol. CAS-30, pp. 723-739, Oct 1983.

[12] Y. C. Lim,'Design of D-C-Value Linear Phase FIR Filters with Optimum Normalised Peak Ripple Magnitude,' IEEE Tran. CAS pp. 1480-1486, Dec 1990

[13] Li Lee & A.V. Oppenheim, 'Properties of Approximate Parks-Mc Clellan Filters,'' Proc. ICASSP München, April 1997.

[14] B. Boulerial, M. F. Belbachir, ''Filtres RIF : Synthese Directe dans l'Espace Discret des Coefficients,'' Pro. NWSIP'98, Sidi Bel Abbes, Algeria, December 1998.

[15] B. Boulerial, ''Filtres RIF : Synthèse Directe dans l'Espace Discret des Coefficients,''thesis, Technology University of Oran, Algeria, November 1998.

[16] A. N. Belbachir, M. F. Belbachir, "Information Processing," Internal Report, Dipartimento di Ingegneria Elettrica ed Elettronica, Universita' degli Studi di Cagliari, Italy, October 1999.

[17] A. N. Belbachir, ''Conception des Filtres Numériques RIF à Phase Linéaire dans l'Espace Discret des Coefficients,'' thesis, University of Oran, Algeria, 2000.