EVOLUTIONARY DESIGN OF FIR DIGITAL FILTERS WITH POWER-OF-TWO COEFFICIENTS

P. Gentili, F. Piazza, A. Uncini

Dip. di Elettronica ed Automatica, University of Ancona
Via Brecce Bianche, 60131 Ancona, Italy

Abstract

The paper presents a genetic approach to the design of Finite Impulse Response filters with coefficients constrained to be sums of power-of-two terms. The evolutionary algorithm is explained and compared experimentally with other state-of-the-art design methods. The proposed technique is able to attain good results and can be easily implemented on parallel hardware.

1. Introduction

In many Digital Signal Processing (DSP) applications, the Finite Impulse Response (FIR) discrete-time filters are widely used, mainly due to their capability to guarantee precise linear phase responses. When the filter coefficients are supposed to be unquantized, several (optimal) design procedures are available, as the well known Parks-McClelland algorithm based on the minmax criterion [1]. However, when implemented with small wordlength digital circuits, such filters cannot be designed with the same techniques of the unquantized case, they often require special design techniques in order to cope with the discretization errors [2]. Among the possible quantizations, the case when the filter coefficients are constrained to be powers-of-two or simple sums of them, receives a particular interest [3,4,5,6]. Such filters are based on the property that, in binary representation, the multiplication between two integer numbers can be substituted by a shift if one of them is a power of two. If a more accurate representation with a sum of two power-of-two terms is preferred, two shifts and an addition are involved. The advantages are a higher speed and a reduced chip area, because shift registers and adders are much faster and smaller than the equivalent binary multipliers. Also the widely used multi-layer neural networks, when implemented with digital circuits, can be designed with power-of-two synaptic weights [7].

As a matter of fact, designing optimum FIR filters with power-of-two coefficients is an optimization problem over a discrete parameter space. It is well known that rounding to the nearest power-of-two the unquantized coefficients obtained by an optimum classical design method, such as the Remez algorithm [1], does not allow to find the optimum set of quantized coefficients.

In [3] the use of integer programming techniques (MILP) is proposed to obtain optimum filters in the minmax sense. Such approach, however, requires a high computation time and a large amount of memory, so that only short filters can be designed in practice. For longer filters, the suboptimal solution of minimizing a mean-square-error criterion is often employed [2], which, however, can yield poor performance. Another kind of suboptimal approach is proposed in [4], where a minmax design method (PRP) is presented which first establishes the gain and then performs a local optimization of the coefficients of the filter. Since the coefficient domain is discrete, the optimum solution could be find by a random or guided search. In such case, moreover, the filter specifications can be a free mixture of frequency and time domain constraints, eventually non-linear in the coefficients. Unfortunately, the random search approach can only be applied to very simple filters due to the complexity of the coefficient space; also the suboptimal procedure [6] does not seem to be able to solve the complete design problem. In [5] a design method based upon the global optimization algorithm, known as simulated annealing (SA), is presented. Although it
cannot guarantee the optimality of the minmax design, the performance of the method is very good also for medium length filters. For longer filters, however, the high computational cost of the SA algorithm can be a real problem, since the sequential nature of the algorithm does not allow to implement it on powerful parallel machines.

2. The genetic approach

It is well known that genetic algorithms (GA) [8] are search algorithms based on the genetic and natural selection paradigm and that they can be successfully employed for minimizing or maximizing a cost function. The design of quantized digital FIR filters, being an optimization problem over a discrete coefficient space, can therefore be faced using a genetic approach. In [9], a classical quantization problem (without the power-of-two constraint) is solved using a simple evolutionary algorithm, where a population of several filter design realizations (there called "filomorphs") was allowed to evolve and find a (sub)optimal solution by the mutation mechanism. No crossing-over or other evolutionary paradigms were employed in that study.

In this paper, we present a design method for digital FIR filters with power-of-two coefficients, which effectively uses the genetic approach. The advantages of such method are similar to those of the SA technique of [5]:

a) can cope with a free mixture of frequency and time domain filter specifications, eventually non-linear in the coefficients;
b) is able to find solutions also for reasonably long filters;
c) provides suboptimal solutions with good performance.

Moreover, due to its implicit parallel nature, the GA approach has some additional advantages over the SA:
d) can explore many possible solutions at each generation;
e) can be easily and efficiently implemented on parallel machines.

On the other hand, the GA design method cannot guarantee an optimal solution as the MILP approach, and is usually slower than the SA when implemented on a sequential single CPU machine.

3. The proposed design method

Let us consider only the case of linear-phase symmetric real FIR filters with odd length, since the extension to other linear-phase filters is straightforward. The frequency response $H(f)$ of such a filter is given by:

$$H(f) = h(0) + \sum_{n=1}^{\frac{N-1}{2}} 2h(n) \cos(2\pi fn)$$  \hspace{1cm} (1)

where $f$ is the normalized frequency $0 \leq f \leq 0.5$, $N$ is the (odd) length and $h(n)$, $n = -(N-1), \ldots, 0, \ldots, (N-1)$, are the filter coefficients. Due to the symmetry, the following relationship holds:

$$h(-n) = h(n)$$  \hspace{1cm} (2)

which allows to reduce the free parameters of the filter to $h(n)$, $n = 0, \ldots, \frac{N-1}{2}$. Such coefficients are moreover constrained to be a simple sum of powers-of-two terms, i.e. they must belongs to the domain $D$:

$$D = \left\{ h : h = \sum_{k=1}^{\lambda} c_k 2^{-\lambda_k}, \begin{array}{l} c_k \in [-1,0,1], \ g_k \in \{1,2,\ldots,B\} \end{array} \right\}$$  \hspace{1cm} (3)

where $\lambda$ is the number of powers of two which forms the coefficients (usually $\lambda=1$ or $\lambda=2$), and $B$ is the maximum number of shifts allowed by the domain. Since it is known that the passband gain $G$ of the quantized filter assumes a particular relevance [4,5], the $G$ value is considered an additional free parameter of the problem.

The minmax criterion minimizes the maximum error between the filter frequency response $H(f)$ and the ideal response $T(f)$ on a dense grid of frequency points $f_m$, $m = 1, \ldots, F$, equally spaced between 0 and 0.5, i.e. minimizes:

$$\delta(G, \ h(n) \ n = 0, \ldots, \frac{N-1}{2}) = \max_{m=1, \ldots, F} \left| G^{-1}H(f_m) - T(f_m) \right|$$  \hspace{1cm} (4)
with $N$, $F$, $\lambda$ and $B$ a-priori selected.
In the proposed method a GA approach is employed for minimizing $\delta$. For this purpose, instead of coding the $h(n)$ and $G$ values, we code the deviations, inside specified ranges, from some leading values. As leading values for the coefficients $h(n)$, the corresponding unquantized optima coefficients $h_0(n)$ are chosen, computed by the Parks-McClelland algorithm [1]. As leading value of $G$, the value $G_0$ nearest to 1 is selected, which minimizes the following function [5]:

$$
\max_{n=-\frac{N-1}{2},..\frac{N-1}{2}} \left| h_0(n) - \frac{Q_D[G_0 h_0(n)]}{G_0} \right|
$$

(5)

where $Q_D[.]$ represents the rounding to the nearest power-of-two in the domain $D$.

The parameters $h(n)$, $n = 0,..,\frac{N-1}{2}$, and $G$ are therefore coded so that:

$$
\hat{h}(n) = Q_D[hcode(n) \cdot \sigma + h_0(n)]
$$

(6)

$$
G = gcode \cdot \sigma + G_0
$$

where $hcode(n)$ and $gcode$ are integer values and $\sigma$ is a fixed small constant. Assigning a given number of bits to each code and setting $\sigma$ to a suitable value, a range of deviation from the leading values is established.

Using such integer codes as genetic strings, a population of quantized filters is created. The evolution then takes place with given probabilities of mutation and crossing-over. The fitness of each individual is computed by expr. (4) and then normalized so that the lowest fitness in the population is set to 0 and the highest to 10000. To avoid the genetic drift in small populations, we have chosen to use a multiple crossing-over [8] on the genetic strings of two individuals. Figure 1 shows this operation using substrings of 16 bits. An elitist mechanism, similar to the De Jong's model [8], has been also implemented, increasing the fitness of the best individuals in the population by a fixed amount roughly proportional to the fitness itself.

After a given number of generations, the performance has been measured by the maximum weighted error $e_{dB}$ expressed in decibel [4,5]:

$$
e_{dB} = \left[ \frac{\delta}{(H_{\max}(f) + H_{\min}(f))/2} \right]_{dB}
$$

(7)

where $H_{\max}(f)$ and $H_{\min}(f)$ are respectively the maximum and minimum values of the frequency response in the passband.
The GA algorithm can also be iterated using as leading values the results of the first run.

4. Experimental results

A computer simulation study was carried out in order to evaluate the performance of the proposed design method.

The GA technique, the SA algorithm in [5], the computation of the optimum minmax filter obtained by the Parks-McClelland algorithm [1] ("quantized" label) and the rounding to the nearest power-of-two of this filter ("rounding" label), have been implemented on a computer system. Many linear-phase FIR filter design experiments have been carried out with the three implemented methods varying the filter lengths and the initial conditions for the GA and SA algorithms. All the experiments were performed with the same filter specifications used in [3,4,5].

Figures 2 and 3 report the maximum weighted error in decibel $e_{dB}$ versus the number of taps of the filter. The filter lengths and the other parameters ($\lambda=2, B=9, F=512$) have been selected in order to compare the results with the data on the MILP and PRP designs presented in [3] and [4]. The probabilities of mutation and crossing-over were set to 0.0015 and 0.9 respectively, with populations from 100 to 1000 individuals.

In Fig. 2, it can be noted that the proposed GA approach can obtain the same results of the SA method [5]. However, the SA algorithm required several runs to provide the best results, while the GA method found them in a single run. Although the GA method resulted 5 to 6 times slower than the SA, it is able to easily overcome the Simulated
Annealing when implemented on a parallel machine (as for example a Transputer network). From Fig. 3 it can be seen that the GA approach can attain better results than the PRP [4] when long filters are involved. No data are available for the MILP [3] and SA methods with these longer filter lengths. Figures 4 and 5 show the amplitude responses of two filters designed with the proposed method with 33 and 59 taps, compared with the unquantized or rounded responses.

5. References


Fig. 1 Multiple crossing-over with 16 bits substrings, used by the algorithm.

Fig. 2 Performance comparison for filter length from 15 to 35 taps.
Fig. 3 Performance comparison for filter length from 51 to 59 taps. (FIR = Unquantized)

Fig. 4 Amplitude response of the designed filter with 33 taps. (FIR = Unquantized).

Fig. 5 Amplitude response of the designed filter with 59 taps. (FIR = Unquantized).