Design of Optimal Frequency-Selective FIR Filters Using a Memetic Algorithm

Luis M. San-José-Revuelta
Dept. Signal Theory and Communications,
Universidad de Valladolid
Valladolid, Spain
lsanjose@tel.uva.es

Abstract—Finite Impulse Response (FIR) digital filters are widely used due to their capabilities of low computational load, stability and linear phase. Most traditional design approaches can not control with enough accuracy the frequency response of the designed filter. For this reason, we propose the use of a memetic algorithm along with a weighted fitness function, to estimate optimized filter coefficients that best approximate ideal specifications. Results have been compared to both traditional methods (mainly windowing and the Parks-McClellan algorithm) as well as to several bio-inspired techniques. Numerical results show that proposed method achieves better fit to filter specifications, a larger attenuation in the stop band and a narrower transition band, at the expense of slightly increasing the passband ripple (0.5 – 0.7 dB), the later in about 68% of the cases.

Index Terms—FIR design, memetic algorithm, genetic algorithm, k-opt algorithm

I. INTRODUCTION

Filters are important components in most computing and electronic devices. Filtering aims to extract information about the signal of interest, either by eliminating noise, extracting frequency components or by separating desired components from unwanted signals.

FIR (Finite Impulse Response) filters present many advantages such as stability and the possibility of obtaining generalized linear phase. Like any linear time-invariant system, they are completely defined by knowing the output signal – known as impulse response– when an impulse signal is placed as input. The length of this output is known as the filter order. The larger this order is, the more similar becomes the filter response to the ideal filter. This increases computational load and processing time. Optimal filter design according to a specified criterion can be seen as an optimization problem [1].

Many conventional methods allow the design of FIR filters. For instance, the method of windowing [2], the Parks and McClellan (PM) algorithm [3], etc. However, these methods do not provide accurate control of various parameters such as the bands cut-off frequencies and the width of the transition band. Besides, methods based on classical optimization have a great tendency to get trapped in local minima, due, in part, to a wrong selection of initial conditions. Their slow rate of convergence also limits its practical use, particularity in online applications. Many novel design strategies, including bio-inspired methods, aim to overcome these drawbacks. These methods have several interesting properties, such as simplicity and the ability to handle different data coding and representations [4]. However, conventional evolutionary algorithms (EAs) are only capable of identifying the high performance region at an affordable time and present inherent difficulties in performing local search. To overcome these problems, once the high performance regions of the search space are identified by the EAs, it is useful to apply a local search procedure to optimize the members of the final population [4]. The efficiency of the algorithm and the quality of the solutions are well improved in this way. These hybrid schemes with closer analogy to cultural evolution are often categorized as memetic algorithms (MAs) [5].

In this paper, a memetic algorithm that combines a standard genetic algorithm (GA) with a k-opt heuristic local search scheme to improve convergence speed and solution accuracy, is applied to design FIR filters. The proposed algorithm employs a two-phase strategy where it starts with a GA that searches for the optimal region with limited computing time. Afterwards, a k-opt heuristic local search is employed to do the fine-tuning so as to obtain the exact global optimum – whenever previous GA has converged properly--, with light computational expense [6], [7]. The thus obtained MA improves the characteristics of the designed filter, showing a good trade-off between frequency response characteristics (mainly, ripple, attenuation and transition bandwidth) and processing time. Besides, it allows a better control of frequency features than most conventional techniques which aim to reduce implementation complexity [8].

II. LITERATURE REVIEW

Nature-inspired methods so far used for FIR filter design belong to a vast variety of strategies. This section summarizes the state-of-the-art in FIR filter design using nature-inspired methods, focusing into two big categories: evolutionary algorithms and swarm methods.

Initial approaches made use of genetic algorithms [9] and some hybrid variants [10]. Simulations showed that GAs achieved better solutions, in terms of passband ripple and group delay, although stopband attenuation was slightly poorer.
Some hybrid methods were developed in order to reduce computational cost [11].

Differential Evolution (DE) algorithms have also been proposed [12]. Simulations show that performance was close to the GA, though computational load of DE was less than that of GA and much lower than the least squares method. In [13], Simulated Annealing (SA) was proposed for the design of both FIR and IIR filters. Results were quite good although computational load was very high. Tabu Search (TS) has also been proposed for FIR design [14].

Swarm methods such as the Ant-Colony (ACO) algorithm emerged as a need to find the global optimum in medium order filters [15]. A bee-colony algorithm was recently proposed by Dwivedi [16] and the algorithm of cloud of particles (Particle Swarm Optimization, PSO) was first proposed for low-pass filter design in [17]. Recently, the algorithm of Herd of Cats (CSO) was used in [18], and [19] developed a new bacteria food search algorithm (BFO), both of them for FIR design.

Even a few attempts using memetic algorithms for digital filter design have been proposed [20], [21].

III. THEORETICAL FOUNDATION

A. Digital FIR filter design

As linear systems, a FIR filter is completely defined by giving either its transfer function $H(z)$, or its corresponding difference equation. The latter can be written as:

$$y[n] = h[0]x[n] + h[1]x[n-1] + \ldots + h[N-1]x[n-(N-1)]$$

where $x[n]$ and $y[n]$ stand for the input and output signal, respectively, and $N$ is the number of filter coefficients (thus, $N-1$ is known as the order of the filter). The purpose of digital filter design is to obtain the coefficients of the filter impulse response $h[n]$. The values of $h[n]$ will determine the type of frequency selective filter being designed: low-pass filter (LPF), high-pass filter (HPF), band-pass filter (BPF) or band-stop filter (BSF).

Filter specifications are often given in the frequency domain as, for example, the pass-band ripple ($\delta_p$) and maximum attenuation ($A_p$), the bands cut-off frequencies, the minimum attenuation of the stop band ($A_s$), the stop-band ripple ($\delta_s$), and the transition bandwidth ($\Delta F$), among others.

This work studies the design of linear-phase symmetrical FIR filters. It is well-known that linear phase is guaranteed if the corresponding impulse response satisfies one of the following symmetry conditions [2]:

$$h[n] = h^*[−n] \text{ or } h[n] = −h^*[−n].$$

Due to this symmetry, only half of the coefficients need to be estimated. At the end of the design process, coefficients are duplicated to form the other half of the impulse response. Thus, complexity of the problem is reduced to half and the desired property of linear phase is satisfied.

The design process of an optimal filter might be seen as an optimization problem where some error measure must be minimized [1]. For instance, some error function between the ideal frequency response and the real frequency response of the estimated filter coefficients.

IV. DESIGN METHODOLOGY

A. GA encoding and fitness function

This paper proposes the use of a MA, originally developed in [4] and here extended for FIR design. The first stage of the MA is a GA where each chromosome encodes a different filter impulse response. This is represented by vector $x_i$:

$$x_i^n = [h_{1,t}(1), h_{1,t}(2), h_{1,t}(3), \ldots, h_{1,t}(\frac{N+1}{2})]$$

which represents the $(N+1)/2$ coefficients of the filter, taking advantage of the imposed symmetry; index $i$ indicates a particular chromosome from the total population of $PS$ ($1 \leq i \leq PS$), and $t$ stands for current generation. The population will consist of a total of $PS$ possible solutions.

This work proposes to use a fitness function (FF) defined as a combination of two previously proposed FFs. A first one, that we will refer as $J_1$, obtained in [18], [19], and a second one, $J_2$, taken from [17]. This way, the FF proposed for use in the MA is,

$$J = w_1 J_1 + w_2 J_2$$

where coefficients $w_1$ and $w_2$ are adjusted in order to get a filter with better frequency characteristics. Error function $E(\omega)$ is defined in [3] as

$$E(\omega) = G(\omega)[H_d(\omega) - H_i(\omega)]$$

where $G(\omega)$ is a weighting function and $\omega_p$ and $\omega_s$ represent the pass- and stop-band limits, respectively. Notice that frequencies are normalized to sampling frequency. $H_d(\omega)$ represents the frequency response of the designed filter, and $H_i(\omega)$ represents the frequency response of the ideal desired filter. This way, error function $E(\omega)$ is higher in that band where tolerances are narrower. Notice that $w_1 J_1$ weights the maximum errors obtained in pass- and stop-bands, measured with respect to maximum allowed ripples, $\delta_p$ and $\delta_s$, respectively. This expression allows to differently weight errors in pass- and stop-bands, since $E(\omega)$ depends on $G(\omega)$, which is given by Eq. (3), where $k$ allows to weight asymmetrically each filter band. On the other hand, term $w_2 J_2$ gathers together errors at every frequency within both pass- and stop-bands (and not only the maximum error values) and references them with respect to maximum deviations, $\delta_p$ and $\delta_s$, respectively.

Finally, notice that error within transition bandwidth is taken into account in second term of $J_1$.

1This expression assumes an ideal amplitude value equal to 1 within passband, and equal to 0 within stopband.
B. Steps of the MA

Proposed MA combines a standard GA and a heuristic local search method. First, the GA estimates a potential quasi-optimum solution, then, the \(k\)-opt local search algorithm is applied to the GA fittest individual so as to quickly estimate a precise solution (always assuming that the GA has reported a solution estimate near the global optimum).

1) Genetic algorithm (Step 1): The metric given in Eq. (1) is used as the fitness function in the GA. In contrast to [4], the here proposed GA makes use of two genetic operators: mutation and crossover, though the later is applied with low probability. Mutation is implemented using the scheme proposed in [4], i.e., offspring vector \(o_i\) is calculated as

\[
o_{ik} = \text{sign}(p_{ik} + N_k(0, \sigma)), \quad k = 1, 2, \ldots, K
\]  

(4)

where \(o_{ik}\) and \(p_{ik}\) denote the \(k\)-th element of the \(i\)-th offspring and parent chromosomes, respectively, \(N_k(\mu, \sigma)\) represents a Gaussian random variable with mean \(\mu\) and standard deviation \(\sigma\). A new random number is generated for each value of \(k\). Standard deviation controls the proximity of the offspring around the parents.

Next, crossover is performed with probability \(P_c = 0.01\). The fittest \(PS\) individuals from the parents and offspring are selected to obtain next generation. GA stops when the algorithm starts converging. The heuristic local search of Step-2 is then called to fine-tune the best individual found so far.

2) \(k\)-opt Local Search (Step-2): Local search methods are improvement heuristics that search in the neighborhood of the current solution for a better one until no better solution is found. Neighborhood of a binary encoded vector is defined as the set of solutions which can be reached by changing the bits value within the vector.

The smallest neighborhood can be achieved by flipping a single bit of the solution vector. The so-called 1-opt neighborhood, searches for a flip with the highest associated gain in fitness \((g = f(b') - f(b))\) in each iteration. An expression for the gain of flipping bit \(k\) in current solution is given in [4]. The \(k\)-opt neighborhood is obtained by flipping one up to \(k\) elements in the solution vector simultaneously:

\[
N_{k-\text{opt}}(b) = b' \in S/d_H(b, b') \leq k
\]  

(5)

where \(S\) is the solution space and \(d_H\) denotes the hamming distance between bit vectors. The \(k\)-opt neighborhood size, \(|N_{k-\text{opt}}(b)| = \sum_{i=0}^{k} \binom{k}{i}\), grows exponentially with \(k\).

To efficiently search a subset of the \(k\)-opt neighborhood, the main ideas of the Lin-Kernighan methods can be used [6], [7]. The main aim is to find a better solution by flipping a variable number of bits in the current solution in every iteration. A sequence of \(k\) solutions is produced in each iteration by flipping the bit with the highest associated gain. Every bit of the solution is only flipped once. The best solution from \(PS\) solutions is then selected as input for next iteration. Thus, a variable number of bits are flipped in each iteration to find a better solution in the neighborhood of current solution.

This section shows the results of the simulations carried out with Matlab(R) and the effectiveness of the proposed filter design method. An odd number of samples is used for all filters designed in this work and the condition of symmetry of coefficients is assumed. Therefore, any type of filter (LPF, HPF, BPF and BSF) can be designed without restrictions. \(N = 20\) unless otherwise specified. Results are mean values after 50 independent runs of the algorithm.

After several initial simulations, values selected for the different parameters were: \(PS = 30\), maximum number of generations in GA \(N_{\text{axGen}} = 250\), crossover probability \(P_c = 0.2\), mutation probability \(P_m = 0.01\), \(w_1 = 0.25\) and \(w_2 = 0.75\).

V. Results

A. Low Pass Filter Design

1) Fixed Filter Order \(N\): The parameters of the desired LPF to be designed are: \(F_s = 2\) Hz, \(L = 512\), \(N - 1 = 20\), \(\delta_p = 0.1\), \(\delta_s = 0.01\), normalized pass-band cut-off frequency \(\omega_p = 0.45\), stop-band normalized cut-off frequency \(\omega_s = 0.55\), and \(\Delta F = 0.1\).

This LPF design will be performed using the following methods: windowing [2], PM [3], and MA. In Fig. 1 frequency responses are depicted. Notice that MA gets the filter with the narrowest \(\Delta F\), clearly outperforming the other two methods. In contrast, MA slightly increases \(\delta_p\).

The filter designed with MA presents higher \(A_s\) than those obtained with PM and windowing. Notice the linear phase of the three methods in the pass-band.

Table I shows the results in terms of maximum ripples \(\delta_p\) and \(\delta_s\), minimum \(A_s\), and \(\Delta F\). It can be observed that, for the case of windowing, if a reduced \(\Delta F\) is required, attenuation \(A_s\) must be reduced (lower \(\beta\)). This fact involves that \(\delta_p\) increases, as well as the maximum attenuation in this band. Also, MA implemented with proposed FF provides a better frequency response than PM, in terms of \(\Delta F\) and \(A_s\). However, it slightly increases \(\delta_p\) and, therefore, the maximum attenuation produced in this band.
2) Variable order: Let us see what happens if both filter order and transition bandwidth are modified.

![Fig. 2. Population size (PS) and number of iterations (MaxGen) required for MA convergence for different orders of a low pass filter.](image)

Fig. 2. Population size (PS) and number of iterations (MaxGen) required for MA convergence for different orders of a low pass filter.

Fig. 2 shows the relationship between order $N$ and both the population size and the number of iterations required for the design. This figure is valid as a guide for parameter tuning for HPFs and BPFs, as well. BSFs require an increase of ~33% in both parameters.

3) Comparison with other nature-inspired algorithms:

Once the comparison with traditional methods has been made, comparisons to other nature-inspired algorithms will be performed.

Fig. 3 shows the frequency responses of MA as well as BFO, PSO and RGA [19]. MA shows the best $A_s$ without an increase of $\Delta F$. In the meanwhile, $\delta_p$ is very similar to those obtained with other design methods. If we look at the phase of the designed filters, all of them present a linear phase in the pass-band, in accordance to the properties of symmetrical FIR filters described in Section III.

For the shake of brevity, results concerning the high-pass filter have not been included here. However, similar conclusions to those of the low-pass filter are valid.

B. Band-pass filter (BPF) and band-stop filter (BSF)

This section gathers the results related to both BPF and BSF filters. Common specifications are: $f_s = 2$ Hz, $L = 512$, $(N-1) = 20$, $\delta_p = 0.1$, $\delta_s = 0.01$. Cut-off frequencies for BPF: inferior stop-band cut-off frequency $\omega_{p1} = 0.25$, inferior pass-band cut-off frequency $\omega_{p2} = 0.65$, upper stop-band cutoff frequency $\omega_{s2} = 0.75$.

For BSF: $\omega_{p1} = 0.25$, $\omega_{s1} = 0.35$, $\omega_{s2} = 0.75$, $\omega_{p2} = 0.85$. BSF is implemented with $PS=50$ and $MaxGen=300$, while BSF requires $PS=60$ and $MaxGen=500$.

Like in previous sections, MA is first compared to traditional approaches. Fig. 4 shows frequency responses of the three methods for stopband filters. It can be seen that the method with highest $A_s$ is the proposed MA, with a $\Delta F$ very similar to the rest of the methods, being, on the contrary, worse in terms of $\delta_p$, which is slightly higher than the ones retrieved using windowing and PM.

Numerical simulations show that MA with $J$ gets a larger $A_s$ as well as a narrower $\Delta F$. This does not occur with traditional techniques where, despite having a lower attenuation, $\Delta F$ is much higher. However, $\delta_p$ is much lower. Similar conclusions are valid for BSFs.

![Fig. 3. Frequency response, magnitude (dB) and phase (rad), for a 20-order low pass filters designed with different nature-inspired algorithms: MA (proposed), BFO, PSO and RGA [19].](image)

![Fig. 4. Frequency response, magnitude (in dB) and phase (in rad), of the BSFs obtained using MA, PM [3] and windowing (Kaiser with $\beta = 1.9$) [2].](image)
Table II compares frequency parameters with those obtained with other nature-inspired algorithms (BFO and CSO [19]). MA clearly outperforms the other approaches since it achieves the largest $A_s$, the smallest $\delta_p$ and $\delta_s$, while keeping transition bandwidth within specs tolerance.

<table>
<thead>
<tr>
<th>Design Method</th>
<th>Max. $A_s$ (dB)</th>
<th>Max. $\delta_p$</th>
<th>Max. $\delta_s$</th>
<th>$\Delta F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFO [19]</td>
<td>33.84</td>
<td>0.161</td>
<td>0.02033</td>
<td>0.1080</td>
</tr>
<tr>
<td>CSO [18]</td>
<td>34.47</td>
<td>0.153</td>
<td>0.01891</td>
<td>0.1086</td>
</tr>
<tr>
<td>MA</td>
<td>37.91</td>
<td>0.1499</td>
<td>0.01381</td>
<td>0.094</td>
</tr>
</tbody>
</table>

This way, proposed MA with FF $J$ given in Eq. (1), consolidates as a good alternative design method to other strategies for any type of frequency-selective digital filter.

VI. Conclusions

This paper explores the possibility of using a memetic algorithm for designing FIR filters. The proposed memetic algorithm consists in two stages. First, a GA is applied for determining the region of the solutions space where the optimal solution is, and, afterwards, a local search technique based on the $k$-opt algorithm is applied.

MA has been compared to both traditional methods for filter design (windowing [2], the minimum squares method and the algorithm of Parks-McClellan [3]), and several nature-inspired methods, mainly CSO [19], [18], RGA [19], BFO [19] and PSO [22].

Numerical results show that MA improves the frequency characteristics of filters obtained by conventional techniques. The frequency response complies with enough accuracy specifications of transition bandwidth and cut-off frequencies of stop and pass bands, giving priority to the attenuation of the bands over the width of transition bands. MA obtained a greater attenuation of the stopband (4-6 dB higher) in comparison to traditional methods, reducing transition band width, as well. However, a slight increase in the passband ripple is observed in about 68% of cases, and, therefore, an increment in the attenuation of that band around 0.5-0.7 dB.

On the other hand, when MA is compared to other nature-inspired techniques it achieves satisfactory results, as well. Simulations carried out indicate that the proposed method obtains a better performance than other techniques in terms of frequency parameters: a greater attenuation of the stop-bands, a suitable transition bandwidth and a pass-band ripple very close to those obtained with other similar techniques.

REFERENCES