A NEW SELF–INITIATED OPTIMUM WLS APPROXIMATION METHOD FOR THE DESIGN OF LINEAR PHASE FIR DIGITAL FILTERS

Chong–Yung Chi and Yung–Teng Kou

Department of Electrical Engineering
National Tsing Hua University, Hsinchu, Taiwan, Republic of China

ABSTRACT

In this paper, we propose a new self–initiated iterative optimum weighted least–squares (WLS) approximation method for the design of linear phase FIR digital filters. The designed optimum filters have nearly equiripple approximation error in every frequency band which is smaller than that of Parks–McClellan's optimum equiripple linear phase FIR filters [3,4] with the same filter length.

I. INTRODUCTION

The finite impulse response (FIR) filter has been popularly used in various signal processing areas because it is stable and linear phase can be easily attained. The windowing method [1,2] and the optimum approximation method [3,4] are the two major FIR filter design methods. Although the windowing method is simple, it needs to go through the time consuming tradeoff procedure of the sidelobe level and the width of the transition band due to windowing effects in order to meet the design specifications. On the other hand, the optimum approximation method, which is free from windowing effects, estimates the filter coefficients by minimizing a selected objective function of approximation error in the frequency domain. However, optimum filter coefficients are usually obtained iteratively because the objective function is generally nonlinear without a closed–form solution for optimum filter coefficients. The optimum approximation method outperforms the windowing method in that the required order of the filter associated with the former is smaller than that associated with the latter for the same specifications, and in that the former can accommodate different approximation errors in different frequency bands but the latter cannot.

In this paper, we propose a self–initiated iterative optimum WLS approximation method for the design of linear phase FIR digital filters. In Section II, we present the new optimum WLS approximation method. We then present two design examples to demonstrate its good performance in Section III. Finally, we draw some conclusions.

II. THE NEW OPTIMUM WLS APPROXIMATION METHOD

Let us present the new optimum WLS approximation method for the design of linear phase FIR filters through the design of lowpass filter. It is well–known that the coefficients h(n) of an Mth order FIR filter with linear phase are either symmetric or anti–symmetric with respect to its center. Thus, we further assume that M is even and

$$h(n)=h(M-n) \qquad (1)$$

without loss of generality. The frequency response $H(f)=H(z=e^{j2\pi f})$ [1] is known to be

$$H(f)=\exp\{-j2\pi fL\}\cdot\left\{h(L)+\sum_{n=1}^{L} 2\,h(L-n)\cdot\cos(2\pi fn)\right\} \qquad (2)$$

where $L=M/2$.

The frequency response of the desired lowpass filter with linear phase is given as follows:

$$H_d(f) = d(f)\cdot\exp\{-j2\pi fL\} \qquad (3)$$

where

$$d(f)=\begin{cases} 1, & \text{for } 0\leq|f|\leq f_p \\ 0, & \text{for } f_s\leq|f|\leq 1/2 \end{cases}, \qquad (4)$$

f_p and f_s denotes the cutoff frequencies in the passband and the stopband, respectively. Thus, the interval (f_p,f_s) is the transition band. The frequency response, H(f), of the optimum FIR filter to be designed must meet the following specifications:

(1) $|H(f)|\leq(1\pm\delta_p)$, $0\leq|f|\leq f_p$,

(2) $|H(f)|\leq\delta_s$, $f_s\leq|f|\leq 1/2$,

where $\delta_p>0$ as well as $\delta_s>0$ denote the tolerable approximation errors in the passband and the stopband, respectively. Next, we present the WLS estimator for the desired filter coefficients on which the new design method is based.

We define the approximation error e(f) in the frequency domain as

$$e(f)=d(f)-\left\{h(L)+\sum_{n=1}^{L} 2\,h(L-n)\cdot\cos(2\pi fn)\right\} \qquad (5)$$

so that the linear phase of H(f) is identical with that of $H_d(f)$. For notational simplicity, let d(k) and e(k) also denote $d(f=k/2N)$ and $e(f=k/2N)$, respectively, where N is the total number of samples to be used. Thus, we can express e(k) for k=0, 1, ..., N–1, in the following linear vector form:

$$\underline{e} = \underline{d} - D\,\underline{h} \qquad (6)$$

where

$$\underline{e} = [e(0), e(1), ..., e(N-1)]', \qquad (7)$$

$$\underline{d} = [d(0), d(1), ..., d(N-1)]', \qquad (8)$$

$$\underline{h} = [h(0), h(1), ..., h(L)]' \qquad (9)$$

and D is an $N\times(L+1)$ matrix with the (k,i)th element equal to

$$[D]_{ki} = \begin{cases} 2\cos(\frac{\pi k}{N}(L-i)), & i \neq L \\ 1, & i = L \end{cases}. \qquad (10)$$

The WLS estimate, $\hat{\underline{h}}$, of $\underline{h}$ is well-known [5] to be

$$\hat{\underline{h}} = [D' W D]^{-1} D' W \underline{d} \qquad (11)$$

which minimizes the following sum of weighted error squares:

$$J(\underline{h}) = \underline{e}' W \underline{e} = \sum_{k=0}^{N-1} w(k)\, e^2(k) \qquad (12)$$

where W is an N×N diagonal weighting matrix

$$W = \mathrm{diag}\,[w(0), w(1), \ldots, w(N-1)] \qquad (13)$$

with $w(k) \geq 0$ for all $0 \leq k \leq N-1$. A well-known property of WLS estimators is:

(P1) The larger the weight w(k), the smaller is the associated estimation error e(k).

Before describing the proposed algorithm, we need to define some notations for easy latter use.

(1) The error ripple: We define the ith error ripple, $e_i^p(k)$, in the passband as:

$$e_i^p(k) = \begin{cases} e(k), & f_{i-1} \leq \frac{k}{2N} \leq f_i \\ 0, & \text{otherwise} \end{cases}, \text{ for } i=1, 2,.., q$$

where q is the total number of error ripples, f_i is associated with $e(f=f_i)=0$, $i=1, 2,.., q-1$ (see (5)), and f_0 and f_q are cutoff frequencies in the passband. For the case of lowpass filter design, $f_0=0$ and $f_q=f_p$. Similarly, the jth error ripple in the stopband, denoted $e_j^s(k)$, is defined in the same manner.

(2) Amplitude of error ripple: We define the amplitude, $|e_p(k_i)|$, of the error ripple $e_i^p(k)$, as $|e_p(k_i)| = \max\{|e_i^p(k)|\}$ for $1 \leq i \leq q$. However, if $|e_p(k_1)|$ (associated with the first error ripple) is not a local maximum, we set $|e_p(k_1)| = |e_p(k_2)|$. If $|e_p(k_q)|$ (associated with the last error ripple) is not a local maximum, we also set $|e_p(k_q)| = |e_p(k_{q-1})|$. The error ripple amplitudes, $|e_s(k_j)|$ for all j, are defined in the same manner.

(3) The maximum and the minimum of error ripple amplitudes:

$$a_p = \max_i\{|e_p(k_i)|\}, \; \rho_p = \min_i\{|e_p(k_i)|\},$$
$$a_s = \max_j\{|e_s(k_j)|\}, \; \rho_s = \min_j\{|e_s(k_j)|\}.$$

The proposed iterative design method, which is shown in Figure 1, is based on the property (P1) of WLS estimators. Therefore, in the inner loop, we update the weight w(k) according to square of error ripple amplitudes (i.e., $|e_p(k_i)|$ and $|e_s(k_j)|$) in order to yield a equiripple frequency response. Contrast to the Parks-McClellan's method [3,4], the proposed design method can also accommodate different approximation errors in different frequency bands in the outer loop by adjusting w(k) according to the approximation error ratio a_s/a_p in order to direct (a_p/a_s) to the desired approximation error ratio (δ_p/δ_s). Next, we describe what the proposed design method includes in the inner loop and the outer loop, respectively.

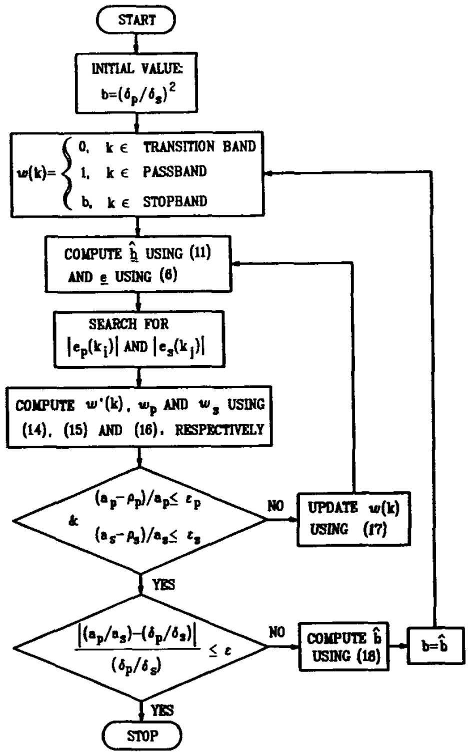


Figure 1. Optimum WLS approximation method.

The proposed design method begins with the initial value $b=(\delta_p/\delta_s)^2$ and the initial

$$w(k) = \begin{cases} 0, & k \in \text{transition band} \\ 1, & k \in \text{passband} \\ b, & k \in \text{stopband} \end{cases}$$

The WLS estimate $\hat{\underline{h}}$ and the associated $\underline{e}$ are then computed by (11) and (6), respectively. We then search for all $|e_p(k_i)|$ and $|e_s(k_j)|$ with which we compute the unnormalized weight w'(k) defined as

$$w'(k) = \begin{cases} w(k) \cdot |e_p(k_i)|^2, & k \in \text{ith ripple in passband} \\ w(k) \cdot |e_s(k_j)|^2, & k \in \text{jth ripple in stopband} \end{cases} \qquad (14)$$

for all i and j, and maxima of w'(k), denoted w_p and w_s, as follows:

$$w_p = \max\{w'(k), k \in \text{passband}\}, \qquad (15)$$
$$w_s = \max\{w'(k), k \in \text{stopband}\}. \qquad (16)$$

Then we check whether the approximate frequency response has nearly equiripple by $(a_p-\rho_p)/a_p \leq \varepsilon_p$ and $(a_s-\rho_s)/a_s \leq \varepsilon_s$, where ε_p and ε_s are preassigned small positive constants. Finally, we update w(k) by

$$w(k)=\begin{cases} 0, & k \in \text{transition band} \\ w'(k)/w_p, & k \in \text{passband} \\ (w'(k)/w_s)\cdot b, & k \in \text{stopband} \end{cases} \quad (17)$$

In the outer loop, we check whether (a_p/a_s) is nearly the same as (δ_p/δ_s) by $|(a_p/a_s)-(\delta_p/\delta_s)|/(\delta_p/\delta_s) \le \varepsilon$ where ε is also a preassigned small positive constant. Then, we update the value of b by $b=\hat{b}$ where

$$\hat{b}=(a_s/a_p)^2\cdot(\delta_p/\delta_s)^2\cdot b \quad (18)$$

which will direct the ratio (a_p/a_s) to (δ_p/δ_s) in the ensuing iterations.

Several noteworthy remarks with regard to the proposed method are as follows:

(R1) Unlike other approximation methods, the initial guess for $\hat{\underline{h}}$ is not needed by the proposed design method.

(R2) Associated with the optimum filter, $a_p/a_s \approx \delta_p/\delta_s$ which is required by the proposed method. However, the maximum amplitudes of error ripples (a_p and a_s) may be greater than the specifications (i.e., δ_p and δ_s) for the selected filter order M. If this case happens, we need to increase M and then go through the previous design procedure.

(R3) Associated with the optimum filter, the maximum approximation error in the passband may be greater than a_p and that in the stopband may also be greater than a_s. In other words, either $\max\{|e_1^p(k)|\}$ or $\max\{|e_q^p(k)|\}$ or both can be greater than a_p, and either $\max\{|e_1^s(k)|\}$ or $\max\{|e_{q'}^s(k)|\}$ or both can also be greater than a_s, where q' is the number of error ripples in the stopband.

(R4) One can observe, from (17) and (18), that for each iteration, $0 \le w(k) \le 1$ for those k belonging to the passband and $0 \le w(k) \le b$ for those k belonging to the stopband, and that the parameter b is only used in updating w(k) for k belonging to the stopband and is updated according to a_s/a_p in the outer loop. In other words, a_p plays the role of reference for adjusting the approximation error level associated with the stopband through $\hat{b}$.

(R5) Because the proposed method is an approximation method for linear phase FIR filter design, unlike any consistent WLS estimators, approximation errors never tend to zero as the parameter N approaches infinity. Therefore, the optimum filters for N large enough are the same.

Although, the proposed design method shown in Figure 1 was described via the case of lowpass FIR filter with linear phase, it can also be applied for general cases of linear phase FIR filter design with some minor modifications. As a final remark, the desired d(f) can be any real function as needed by the designer.

III. DESIGN EXAMPLES

We now present two design examples of FIR filter with linear phase. One example is for a lowpass filter case and the other is for a bandpass filter case. For these two examples, the number of frequency samples used was N=2000 and the parameters ε_p, ε_s and ε used were $\varepsilon_p=\varepsilon_s=0.01$ and $\varepsilon=0.01$.

Example 1. Lowpass filter

The parameters f_p, f_s, δ_p and δ_s for specifying the linear phase lowpass FIR filter to be designed were as follows:

(1) $f_p=0.2$ and $f_s=0.3$;

(2) $\delta_p=0.01$ and $\delta_s=0.001$.

The parameter M used was M=27. The frequency response of the optimum lowpass filter is shown in Figure 2a. Approximation error (unweighted) is shown in Figure 2b, from which one can see that the frequency response of the optimum filter has nearly equiripple in each frequency band. The maximum of error ripples in the passband is $a_p \approx 0.0090 < \delta_p$ and that in the stopband is $a_s \approx 0.0090 < \delta_s$. The ratio $a_p/a_s \approx \delta_p/\delta_s = 1$. However, the maximum approximation errors (not a local extremum) in the passband occurs at $f=f_p$ and its value is $0.0095 < \delta_p$ (see (R3)). On the other hand, the maximum approximation error in the stopband is the same as $a_s \approx 0.00090$.

The optimum equiripple linear phase FIR filter with the same order using the well-known iterative Parks–McClellan's method [1] was reported to have corresponding $\hat{a}_p = \hat{a}_s \approx 0.00092$. Remark that $\hat{a}_s/a_s \approx \hat{a}_p/a_p \approx 1.022$ (or 0.2 dB improvement by our method) whereas the maximum approximation error (≈ 0.0095) of the designed FIR filter by our method is slightly larger than $\hat{a}_p$.

Example 2. Bandpass filter

The desired d(f) (see (4)) for this case was given by

$$d(f)=\begin{cases} 0, & \text{for } 0 \le |f| \le f_{s1} \\ 1, & \text{for } f_{p1} \le |f| \le f_{p2} \\ 0, & \text{for } f_{s2} \le |f| \le 1/2 \end{cases}$$

where $f_{s1}=0.15$, $f_{p1}=0.175$, $f_{p2}=0.3$ and $f_{s2}=0.35$ are cutoff frequencies. The specifications of the the filter to be designed were given as follows:

(1) $|H(f)| \le \delta_{s1}$, $0 \le |f| \le f_{s1}$,

(2) $|H(f)| \le (1 \pm \delta_p)$, $f_{p1} \le |f| \le f_{p2}$,

(3) $|H(f)| \le \delta_{s2}$, $f_{s2} \le |f| \le 1/2$,

where $\delta_{s1}=0.01$, $\delta_p=0.01$ and $\delta_{s2}=0.05$ denote the tolerable approximation errors.

The parameter M used was M=74. The frequency response of the optimum bandpass filter is shown in Figures 3a. Approximation error (unweighted) is shown in Figure 3b. Again, one can also see, from Figure 3b, that the frequency response of the optimum filter has nearly equiripple in each frequency band. The maximum of error ripples in the first stopband is $a_{s1} \approx 0.0099 < \delta_{s1}$, that in the passband is $a_p \approx 0.0099 < \delta_p$ and that in the second stopband is $a_{s2} \approx 0.0499 < \delta_{s2}$. The ratio $a_p/a_{s1} \approx \delta_p/\delta_{s1} = 1$ and the ratio $a_{s2}/a_{s1} = 5.04 \approx \delta_{s2}/\delta_{s1} = 5$. Note that, for the first stopband and the passband, the maximum approximation errors are almost the same as a_{s1} and a_p, respectively, whereas that in the second stopband is the same as a_{s2}.

The optimum equiripple linear phase FIR filter with the same filter length using the Parks–McClellan's algorithm [1] was reported to have corresponding $\hat{a}_{s1} = \hat{a}_p \approx 0.011 > a_{s1} \approx a_p$ and $\hat{a}_{s2} \approx 0.055 > a_{s2}$ (or 0.9 dB and 0.85 dB improvements, respectively, by our method.)

IV. CONCLUSIONS

We have presented a new self–initiated iterative optimum WLS approximation method for the design of linear phase FIR digital filters based on the property (P1) of WLS estimators. The key weights w(k) (see (14) through (17)) used in each iteration are determined by error ripple amplitudes while those associated with the transition band are set to be zero. The proposed method tries to search for the filter with equiripple in every frequency band and then adjusts different approximation errors in different frequency bands according to the design specifications. The designed optimum filters have nearly equiripple amplitude in every frequency band which is slightly smaller than that of Parks–McClellan's optimum equiripple linear phase FIR filter with the same filter length. This fact was supported by the presented two design examples The proposed design method is also applicable for the design of any arbitrary linear phase FIR filters with some minor modifications.

REFERENCES

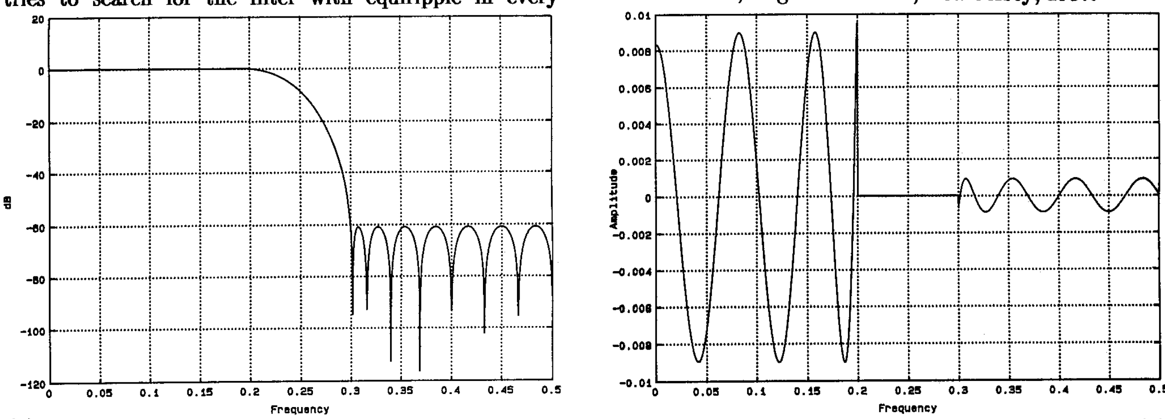
[1] A. Oppenheim and R. Schafer, Discrete–time signal processing, Prentice–Hall, Englewood Cliffs, New Jersey, 1989.

[2] R. A. Roberts and C. T. Mullis, Digital signal processing, Addison–Wesley Publishing Company, 1987.

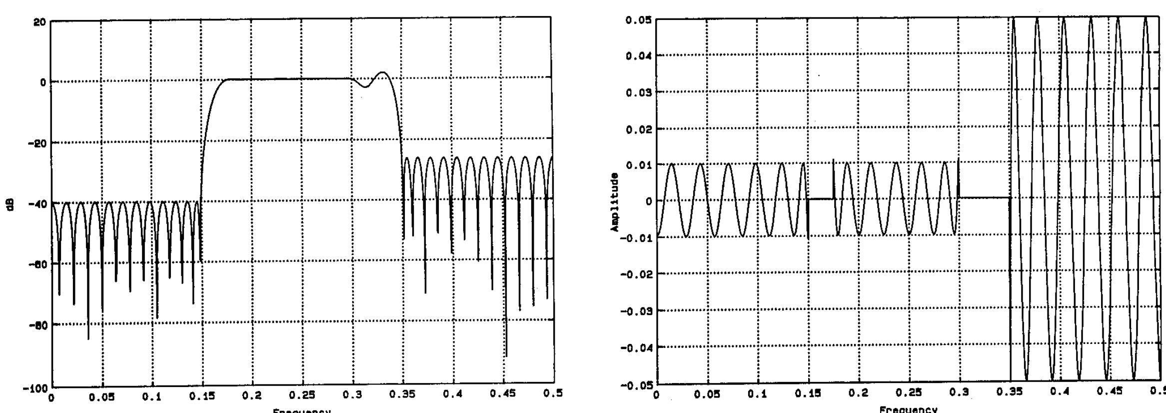
[3] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," IEEE Trans. Circuit Theory, vol. CT–19, pp. 189–194, March 1972.

[4] J. H. McClellan and T. W. Parks, "A unified approach to the design of optimum FIR linear phase filters," IEEE Trans. Circuit Theory, vol. CT–20, pp. 697–701, Nov. 1973.

[5] J. M. Mendel, Lessons in digital estimation theory, Prentice Hall, Englewood Cliffs, New Jersey, 1987.



(a) Figure 2. Optimum lowpass FIR filter with linear phase for M=27 and N=2000. (a) Log magnitude and (b) approximation error (unweighted). (b)



(a) Figure 3. Optimum bandpass FIR filter with linear phase for M=74 and N=2000. (a) Log magnitude and (b) approximation error (unweighted). (b)