A Robust Initialization Scheme for the Remez Exchange Algorithm

Emmanouil Z. Psarakis and George V. Moustakides, *Senior Member, IEEE*

***Abstract*—A well-known least squares optimum approximation method is proposed as an efficient initialization scheme for the Remez exchange algorithm. More specifically, we theoretically demonstrate that the "don't care" least squares optimum solution guarantees, inside the bands of interest, the correct number of alternating in-sign extrema of the error function, thus satisfying one of the two basic conditions that are sufficient for obtaining the L_∞ optimum solution. Although convergence of Remez is theoretically assured, its practical implementations may fail to converge in "difficult" design problems when classical initialization is used. In particular, Matlab's realization of Remez, when initialized with the proposed scheme, exhibits a significantly better overall performance that translates into faster convergence and more robust behavior, especially in difficult designs problems.**

***Index Terms*—Alternation theorem, don't care filter design method, FIR filter design theory, min–max approximations, Remez exchange algorithm.**

I. Introduction

In FINITE-IMPULSE response (FIR) filter design theory, L_∞ is considered as the most popular approximation measure. Approximations obtained with this criterion, also known as Chebyshev or min–max, exhibit equiripple behavior in the frequency bands of interest (passband and stopband) resulting in the complete elimination of the annoying Gibb's phenomenon, observed in window- and certain L_2-based techniques.

Unfortunately, L_∞ optimum solutions are obtained through iterative schemes that tend to be computationally demanding. Most well-known techniques are the Remez exchange algorithm (REA), the iterative weighted least squares, and algorithms based on constrained linear optimization. In this letter, we will focus on REA [7], [10], which tends to be the most popular method for designing classical filters as well as digital differentiators and Hilbert transformers.

In most iterative techniques, convergence speed and therefore convergence time strongly depend on the initial "guess" of the solution. The first initialization method for REA was presented in [7] and additional methods aiming in speeding up the algorithm in [1], [2], [6], [11], and [12].

Manuscript received September 21, 2001; revised July 24, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Steven Kay.

E. Z. Psarakis is with the Department of Business Planning and Information Systems, TEI of Patras, 26334 Patras, Greece, and also with the Computer Technology Institute (CTI), 26110 Patras, Greece.

G. V. Moustakides is with the Institut National de Recherche en Informatique et en Automatique (INRIA), 35042 Rennes, France and also with the Department of Computer Engineering and Communications, University of Thessaly, 38221 Volos, Greece.

Digital Object Identifier 10.1109/LSP.2002.806701

In this letter, we propose the use of the simple "don't care" least squares optimum filter [9, p. 70] as an alternative initialization scheme for REA. This recommendation, however, is supported by an important theoretical result stating that the don't care optimum solution has the correct number of alternating in-sign local extrema, required by the alternation theorem, inside the bands of interest. This suggests that one of the two conditions (in fact, the most difficult one) that define the optimum L_∞ solution is readily satisfied. Furthermore, since the don't care method is known to produce efficient filters that, in the L_∞ sense, are close to the L_∞ optimum solution, the second condition of the alternation theorem is also well approximated. These two properties clearly suggest that the don't care optimum filter can be an appropriate initializer for REA.

It should be noted that similar results were impossible to demonstrate for any other filter design method as window- and spline-based [9] or eigenfilters [13]. In other words, filters obtained by these alternative methods cannot be considered as safe starting points for REA.

II. L_∞ Approximations and the Alternation Theorem

Let us first introduce the approximation problem of interest. Consider a collection of a finite number of closed nonoverlapping intervals I_i, $i = 1, \ldots, K$ that are subsets of $[0\ \ \pi]$. Consider also a real function $D(\omega)$, continuous and known on each interval I_i, which we like to approximate in the L_∞ sense. Let $\phi_n(\omega) = \cos(n\omega)$, or $\phi_n(\omega) = \sin((n+1)\omega)$, $n = 0, \ldots, N-1$ be two sets of base functions we like to use to approximate the function $D(\omega)$. Finally, let $W(\omega)$ be a weighting function that is known, continuous, and nonnegative on each interval I_i.

The correspondence with the filter design problem is apparent. The intervals I_i are either passbands or stopbands; $D(\omega)$ is the desired response; and $W(\omega) > 0$ is the weight in each band. The first set of base functions can be used to approximate even symmetric responses (defined on $[-\pi\ \ \pi]$), while the second odd symmetric. It is also clear that the open intervals between consecutive I_i constitute the transition regions between the bands of interest.

Let $h = [h_0\ \ h_1 \cdots h_{N-1}]^t$ be a vector of coefficients, and denote by $H(h, \omega) = \sum_{n=0}^{N-1} h_n \phi_n(\omega)$ a linear combination of the base functions. We are interested in obtaining the optimum vector h_o, in the L_∞ sense, that satisfies

$$h_o = \arg \left\{ \inf_h \sup_{\omega \in \bigcup_{i=1}^{K} I_i} W(\omega) |D(\omega) - H(h, \omega)| \right\}. \quad (1)$$

We have now the following theorem that provides necessary and sufficient conditions for h_o.

Theorem 1: The vector h_o is optimum in the L_∞ sense if and only if the following two conditions are satisfied.

1) The weighted error function $W(\omega)[D(\omega) - H(h_o, \omega)]$ has at least $N+1$ local extrema with alternating sign at points $\omega_1 < \omega_2 < \cdots < \omega_{N+1}$ that **belong to the bands of interest** (i.e., $\omega_n \in \bigcup_{i=1}^{K} I_i$).
2) The local extrema of condition 1) are all equal, in absolute value, to the maximum weighted absolute deviation $\sup_{\omega \in \bigcup_{i=1}^{K} I_i} W(\omega)|D(\omega) - H(h_o, \omega)|$.

Proof: The proof can be found in [5]. ∎

The above theorem, known as the *alternation theorem*, completely characterizes the optimum solution h_o. Conditions 1) and 2) also suggest that a vector h can be regarded as a suitable initial guess if it satisfies condition 1) exactly and condition 2) approximately. In fact, the main difficulty in all existing initialization schemes is to ensure existence of at least $N+1$ local extrema, with alternating sign, *inside* the set $\bigcup_{i=1}^{K} I_i$.

III. Don't Care Optimum Least Squares Approximations

Let $D(\omega)$, $W(\omega)$, $\phi_n(\omega)$ be as in Section II. Recall that $W(\omega)$ was defined only on the set $\bigcup_{i=1}^{K} I_i$. If we extend $W(\omega)$ to the whole interval $[0\ \pi]$, by setting $W(\omega) = 0$ for $\omega \in [0\ \pi] - \bigcup_{i=1}^{K} I_i$, we can then define a vector of optimum coefficients $h_{\rm dc}$ by solving the following least squares problem

$$h_{\rm dc} = \arg\left\{ \inf_{h} \int_0^\pi W^2(\omega)[D(\omega) - H(h, \omega)]^2\, d\omega \right\}. \quad (2)$$

Since the weighting function $W(\omega)$ is zero outside the set of interest $\bigcup_{i=1}^{K} I_i$, the values of $D(\omega)$ outside $\bigcup_{i=1}^{K} I_i$ play absolutely no role (this is why the term "don't care" is applied). The optimum vector $h_{\rm dc}$ is the solution to the linear system defined by the equations

$$\int_0^\pi W^2(\omega)[D(\omega) - H(h_{\rm dc}, \omega)]\phi_i(\omega)\, d\omega = 0, \qquad i = 0, \ldots, N-1. \quad (3)$$

It is easy to show that the linear system in (3) has a Toeplitz plus Hankel structure. This allows for the employment of specialized algorithms with reduced $O(N^2)$ complexity [8], as compared to the $O(N^3)$ complexity required for the solution of a general linear system.

In Theorem 2, we will show that $h_{\rm dc}$ generates the correct number of local extrema with alternating sign inside the bands of interest. In other words, it satisfies exactly condition 1) of Theorem 1.

Theorem 2: The don't care least squares optimum solution has a weighted error with at least $N+1$ local extrema with alternating sign inside the set $\bigcup_{i=1}^{K} I_i$.

Proof: Extending the proof presented in [5] for vectors to the case of functions of ω, let us consider the weighted optimum error function $E(\omega) = W(\omega)[D(\omega) - H(h_{\rm dc}, \omega)]$. This function is piecewise, continuous, and well defined on the whole interval $[0\ \pi]$. If $E(\omega)$ is not identically zero, then we can divide $[0\ \pi]$ into $M+1$ intervals $U_i = [\alpha_i\ \alpha_{i+1}]$, $i = 0, \ldots, M$, with $\alpha_0 = 0, \alpha_{M+1} = \pi$, satisfying the following conditions.

- In each U_i, there exists a subinterval where $E(\omega)$ is nonzero.
- In each U_i, the nonzero values of $E(\omega)$ have constant sign.
- The constant sign of each U_i alternates between consecutive intervals U_i.

As a consequence of the above conditions, we have the following properties.

- In each interval U_i, the function $E(\omega)$ has a local extremum different than zero.
- All such local extrema are of alternating sign.
- Each extremum can appear only inside the set $\bigcup_{i=1}^{K} I_i$, since outside this set $W(\omega)$ was defined to be zero, and consequently $E(\omega)$ is also zero.

What is now left to show, to complete the proof, is that $M \geq N$. We will first prove the statement for the case $\phi_n(\omega) = \cos(n\omega)$. Assume $M \leq N-1$; then we can define the following function

$$P(\omega) = \mathrm{A} \prod_{i=1}^{M} (\cos(\omega) - \cos(\alpha_i)) \quad (4)$$

where A is a constant. Since $\cos(\omega)$ is monotone, we can select A so as the sign of $P(\omega)$ matches exactly the sign of $E(\omega)$. Since $E(\omega)$ is not identically zero, and $P(\omega)$ is zero for a finite number of points, this means that

$$\int_0^\pi W^2(\omega)[D(\omega) - H(h_{\rm dc}, \omega)]P(\omega)\, d\omega > 0. \quad (5)$$

Notice now that any polynomial in $\cos(\omega)$ of degree at most $N-1$ (as is the case of $P(\omega)$) can be written as a linear combination of the base functions $\cos(n\omega)$. This means that, because of (3), the integral in (5) is zero, thus contradicting (5). So we must have $M \geq N$. In exactly similar way, we can prove the theorem for the base functions $\phi_n(\omega) = \sin((n+1)\omega)$, except that only now $P(\omega)$ must be defined as

$$P(\omega) = \mathrm{A}\sin(\omega) \prod_{i=1}^{M} (\cos(\omega) - \cos(\alpha_i)). \quad (6)$$

This concludes the proof. ∎

Another important property, also stated in the Introduction, is the fact that the don't care method is known to provide solutions that, in the L_∞ sense, are only a few decibels inferior to the optimum L_∞ solution [3], [4]. This, of course, suggests that condition 2) is approximately satisfied.

Summarizing the desirable features of the don't care least squares optimum solution.

1) There exist computationally efficient schemes that take advantage of the special Toeplitz plus Hankel structure.
2) It satisfies exactly the first (and most difficult) condition of the alternation theorem.
3) From practice it is known that it efficiently approximates the second condition of the alternation theorem.

We conclude, therefore, that this filter constitutes a powerful candidate for the initialization of REA.

IV. Discussion and Simulations

The conventional initialization scheme of REA simply consists in selecting the initial local extrema with a uniform distribution inside each band, with the number of points inside each band being proportional to the length of the band [7]. It is clear that the computational cost of this initialization is insignificant. We should also mention that this is the scheme employed in Matlab's function `remez.m`.

For our initialization, we used Matlab's function `firls.m`, which computes don't care optimum filters. It should be noted, however, that this function does not employ any efficient solution of the linear system defined in (3). We also had to modify the function `remez.m` to accept the don't care solution from `firls.m` as a starting point. We designed several types of filters as lowpass, bandpass, bandstop, and multiband with varying lengths and weights. In particular, we considered filter lengths ranging from 51 up to 201, and in all cases we observed a significant speed up in REA as compared to the conventional initialization scheme. Unfortunately, this significant reduction in number of iterations did not translate into an equivalent reduction in the total computational complexity due to the overhead required by the solution of the linear system (3). Nevertheless, in all cases, the computational gain was nonnegligible and ranged from 20% to 50%.

To our opinion, the strong point of our scheme does not lie so much in its computational gain as in its robustness. In particular, when we applied Matlab's implementation of REA to "difficult" design problems, in numerous cases it failed to converge with conventional initialization, while convergence was attained with a few iterations with the proposed one. Although convergence of REA is assured theoretically, we should stress that its practical implementations may fail. As reported in [2], the source of this undesirable behavior is not necessarily the usage of a coarse-frequency grid as one might initially believe but also the uniform initialization itself. Indeed, it is possible for the classical initialization to give rise to error functions that are not unimodal between adjacent zeros violating the sign alternation requirement of the alternation theorem. Of course, we do not claim that our scheme can assure convergence; however, since our initialization is close to the optimum solution and satisfies the sign alternation requirement, one can expect that it will have a more robust behavior, especially in difficult design problems. This is, in fact, what we observed when we used it in practice.

Let us now present two indicative design examples. Consider first a lowpass filter with $\omega_p = 0.3\pi, \omega_s = 0.4\pi$. If L denotes the filter length, then REA performed as follows:

- $L = 51$, iterations: 8*(3)*, flops: 336 882*(256 023)*,
- $L = 101$, iterations: 13*(4)*, flops: 1 575 160*(879 396)*,
- $L = 201$, iterations: 15*(4)*, flops: 6 855 927*(3 697 605)*,

where normal letters correspond to the conventional scheme and italic to the proposed initialization. The number of operations was counted using Matlab's function `flops.m`.

As a second example consider a bandstop filter with specifications $\omega_{p_1} = 0.2\pi, \omega_{s_1} = 0.3\pi, \omega_{s_2} = 0.5\pi$, and $\omega_{p_2} = 0.6\pi$. The corresponding performance of REA was

$L = 51$, iterations: 19*(3)*, flops: 647 531*(449 675)*,
$L = 101$, iterations: 15*(4)*, flops: 1 500 139*(898 002)*,
$L = 201$, iterations: NC*(5)*, flops: NC*(5 264 586)*,

where "NC" denotes no convergence. Regarding the last case $L = 201$, we must say that Matlab's algorithm did not converge even when we applied a much denser frequency grid, which is in accordance with the observation reported in [2].

V. Conclusion

We have theoretically demonstrated that the don't care optimum solution constitutes a suitable initialization scheme for the Remez exchange algorithm. In numerous simulations, we observed a significant reduction in the number of iterations required by the algorithm to converge. Furthermore, REA exhibited an increased robustness, especially in difficult design cases, where practical implementations of the algorithm failed convergence under classical initialization. Taking into account that the proposed initialization induces a significant speedup in REA, accompanied by a nonnegligible computational gain and an overall increased robustness, we conclude that our scheme can be clearly considered as a possible alternative to the uniform initialization.

References

[1] A. Antoniou, "Accelerated procedure for the design of equiripple nonrecursive digital filters," *Proc. Inst. Elect. Eng. G, Electron. Circuits Syst.*, vol. 126, no. 1, pp. 1–10, Jan. 1982.

[2] ——, "New improved method for the design of weighted Chebyshev, nonrecursive, digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 740–750, Oct. 1983.

[3] C. S. Burrus *et al.*, "Least squared error FIR filter design with transition bands," *IEEE Trans. Signal Processing*, vol. 40, pp. 1327–1340, June 1992.

[4] C. S. Burrus, "Multiband least squares FIR filter design," *IEEE Trans. Signal Processing*, vol. 43, pp. 412–420, Feb. 1995.

[5] E. W. Chiney, *Introduction to Approximation Theory*. New York: McGraw-Hill, 1966.

[6] S. Ebert and U. Heute, "Accelerated design of linear or minimum phase FIR filters with a Chebyshev magnitude response," *Proc. Inst. Elect. Eng. G, Electron. Circuits Syst.*, vol. 130, no. 6, pp. 267–270, Dec. 1983.

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

[8] G. A. Merchant and T. W. Parks, "Efficient solution of a Toeplitz plus Hankel coefficient matrix system of equations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 40–44, Feb. 1982.

[9] T. W. Parks and C. S. Burrus, *Digital Filter Design*. New York: Wiley, 1987.

[10] L. R. Rabiner, J. H. McClellan, and T. W. Parks, "FIR digital filter design techniques using weighted Chebyshev approximations," *Proc. IEEE*, vol. 63, pp. 595–610, Apr. 1975.

[11] D. J. Shpak and A. Antoniou, "A generalized Remez method for the design of FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 161–174, Feb. 1990.

[12] R. H. Yang and Y. C. Lim, "Grid density for design of one and two-dimensional FIR filters," *Elec. Lett.*, vol. 27, no. 22, pp. 2053–2055, Oct. 1991.

[13] P. P. Vaidyanathan and T. Q. Nguyen, "Eigenfilters: A new approach to least squares FIR filter design and applications including Nyquist filters," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 11–23, Jan. 1987.