An L_2-Based Method for the Design of 1-D Zero Phase FIR Digital Filters

Emmanouil Z. Psarakis and George V. Moustakides

***Abstract*—Finite impulse response (FIR) filters obtained with the classical L_2 method have performance that is very sensitive to the form of the ideal response selected for the transition region. It is known that design requirements do not constraint in any way the ideal response inside this region. Most existing techniques utilize this flexibility. By selecting various classes of functions to describe the undefined part of the ideal response they develop methods that improve the performance of the L_2 based filters. In this paper we propose a means for selecting the unknown part of the ideal response optimally. Specifically by using a well-known property of the Fourier approximation theory we introduce a suitable quality measure. The proposed measure is a functional of the ideal response and depends on its actual form inside the transition region. Using variational techniques we succeed in minimizing the introduced criterion with respect to the ideal response and thus obtain its corresponding optimum form. The complete solution to the problem can be obtained by solving a simple system of linear equations suggesting a reduced complexity for the proposed method. An extensive number of design examples show the definite superiority of our method over most existing non min-max design techniques, while the method compares very favorably with min–max optimum methods. Finally we prove that the approximation error function of our filter has the right number of alternating extrema, required by the L_∞ criterion, in the passband and stopband. This results in a significant convergence speed up, if our optimum solution is used as an initialization scheme, of the Remez exchange algorithm.**

***Index Terms*—1-D digital filters, Fourier approximation, zero phase FIR filters.**

I. Introduction

THE DESIGN of one-dimensional (1-D) digital filters, although is an old problem with much existing literature, it has been of a growing interest over the last decade. This is because digital filters are widely used for a variety of signal processing applications such as speech and image processing, communications, seismology, radar, sonar and medical signal processing.

A very important class of 1-D filters is the class of finite impulse response (FIR) filters. This class is tractable because of certain desirable characteristics, as stability, linear phase and simplicity of design. The most common techniques used for the determination of FIR filters use as approximation criterion the minimization of the L_p measure, where the power p satisfies $1 \leq p \leq \infty$. The complexity of the corresponding optimization algorithms depends heavily on the value of p as does the corresponding performance of the resulting filters. It is well known that all values of p result in a nonlinear optimization problem except the case $p = 2$.

The L_∞ (or min-max) criterion is considered as the most desirable criterion to use in the FIR filter design problem because it exhibits equiripple behavior in the frequency bands of interest. Its basic drawback is the need of sophisticated optimization tools as the Remez exchange algorithm [23], [25], [27], [29], iterative linear programming [14], [27] or iterated weighted least squares [1], [3], [7], [17], [30], that require a large computational effort. Other values of the power p (but different from 2) are also considered in [22] but not as widely as the L_∞ case.

The L_2 (or mean square) criterion is the most tractable and the simplest criterion from a mathematical and computational point of view. It results in the well-known Fourier approximation that can be easily obtained analytically. Unfortunately this method is known for its poor performance that is more pronounced at the discontinuity points of the ideal response (Gibb's phenomenon) [22], [24], [26].

The use of windowing techniques is a classical method for reducing the undesirable ripples due to the Gibb's phenomenon. Standard windows found in the literature are the Adams, Blackman, Bartlett, Dolph–Chebyshev, Hamming, Hanning and Kaiser [2], [14], [22], [24], [26]. Although this method is very simple the resulting filters do not satisfy any optimality criterion and their performance is significantly lower than the L_∞ optimum filters.

The performance of the L_2 method can be improved if transition regions are introduced between passbands and stopbands. There are two categories of design techniques based on this idea. The first includes methods that define the ideal response inside the transition region using some arbitrary class of functions such as splines or trigonometric polynomials [24] and use it to compute the Fourier approximation. In the second category, the corresponding methods consider the transition region as a "don't care" region and simply remove it from the error measure [21], [24], [30]. All the above methods succeed in reducing the Gibb's phenomenon, but their performance is still inferior as compared to L_∞ filters. Also, the selection of the form of the ideal response inside the transition region, as we said, is arbitrary since it does not satisfy any optimality criterion.

Further improvement in the performance of the L_2 method, especially when the desired attenuation is not particularly high, can be achieved by using the modified window method

Manuscript received November 24, 1994; revised November 18, 1996. This paper was recommended by Associate Editor I. Pitas.

The authors are with the Department of Computer Engineering and Informatics, University of Patras, Patras 26500, Greece and the Computer Technology Institute (CTI), Patras 26110, Greece.

Publisher Item Identifier S 1057-7122(97)03488-0.

presented in [20]. In this approach certain parameters of the design process are optimally selected by minimizing the maximum deviation. The minimization process is computationally heavy, thus empirical formulas for the optimum values of these parameters are given, but only for the low-pass filter case and for a limited range of attenuation values. The performance of the resulting filters using the empirical formulas is very good. Unfortunately generalization of the formulas to more complicated filters (as bandpass or multiband) seems very difficult.

In this paper we present a method that belongs to the L_2 optimum design techniques. The important difference, as compared to existing methods, is the fact that the unknown part of the ideal response is obtained by minimizing an optimization criterion that is directly related to the approximation problem. Specifically, by using a well-known property of the Fourier approximation theory we define a suitable L_2 criterion. This measure, being a function of the ideal response, is minimized to define optimally the ideal response inside the transition region. Necessary continuity constraints guarantee the unicity of the solution. The complexity of the resulting method is small since it can be reduced to the solution of a symmetric Toeplitz system of linear equations. A large number of examples show the definite superiority of the proposed method as compared to most well known non min-max design techniques.

A very important feature of the proposed method is the fact that the resulting filters have error functions with the required, by the L_∞ criterion, number of alternating extrema inside the passband and stopband. This statement is proved theoretically. Thus our Fourier approximation, if used as an initialization scheme in the Reméz exchange algorithm, improves its convergence speed considerably.

The paper is organized as follows, Section I contains the Introduction. In Section II we present the basic background for the Fourier approximation and introduce our performance criterion. Section III contains the complete optimization problem along with the necessary constraints and also its general solution in the form of a linear system of equations. In Section IV by exploiting the special Toeplitz plus Hankel form of the linear system we present a special symmetric Levinson algorithm for its efficient solution. In Section V we apply the proposed method to several 1-D filter design problems and we make comparisons with existing techniques. In Section VI we investigate the problem of the initialization of the Remez exchange algorithm. We prove that our filter has the necessary number of alternating extrema inside the passband and stopband, required by the L_∞ criterion and we use it as an alternative initialization procedure for the Remez exchange algorithm. Comparisons are made with the existing initialization technique. Finally Section VII has the conclusion.

II. Properties of the Fourier Expansion

Let us consider a symmetric[1] function $D(\omega)$ that we like to approximate using a trigonometric polynomial of order $N-1$. Specifically let us define the following set of N orthonormal functions

$$\phi_0(\omega)=\frac{1}{\sqrt{\pi}},\quad \phi_n(\omega)=\sqrt{\frac{2}{\pi}}\cos(n\omega),\quad n=1,\cdots,N-1 \tag{1}$$

and the vector function

$$\boldsymbol{\phi}(\omega)=[\phi_0(\omega)\cdots\phi_{N-1}(\omega)]^t. \tag{2}$$

Let us also denote the usual inner product of two real functions $f(\omega), g(\omega)$ as

$$\langle f,g\rangle=\int_0^\pi f(\omega)g(\omega)\,d\omega. \tag{3}$$

By selecting a vector of N coefficients $\boldsymbol{h}=[h_0\cdots h_{N-1}]^t$ we can form the trigonometric polynomial $H(\omega)=\boldsymbol{\phi}^t(\omega)\boldsymbol{h}$ and use it to approximate $D(\omega)$. Thus we can define the mean square error (MSE) between $D(\omega)$ and $H(\omega)$ as $\langle D-H, D-H\rangle$. The optimum coefficients (also known as Fourier coefficients) that minimize the MSE are given by

$$\boldsymbol{h}_D=\langle\boldsymbol{\phi},D\rangle \tag{4}$$

and the corresponding optimum approximation (also known as Fourier approximation) is $H_D(\omega)=\boldsymbol{\phi}^t(\omega)\boldsymbol{h}_D$. The resulting minimum mean square error (MMSE) can be considered as a quality measure for our optimum approximation. Namely we can define the following criterion:

$$\mathcal{E}_0(D)=\langle D-H_D, D-H_D\rangle \tag{5}$$

which is a functional of $D(\omega)$.

A very remarkable property satisfied by the Fourier approximation can now be used to generalize the measure defined in (5). This property is stated without proof (a proof can be found in [6, p. 172]) in the following lemma.

Lemma 1: Let the function $D(\omega)$ be mean square optimally approximated by $H_D(\omega)$. Let also $D(\omega)$ have a piecewise continuous kth order derivative. Then this kth order derivative $D^{(k)}(\omega)$ is mean square optimally approximated by the corresponding kth-order derivative $H_D^{(k)}(\omega)$.

Lemma 1 suggests that, under the indicated continuity property, we can identify the Fourier coefficients not only through the approximation of $D(\omega)$ but also by approximating the derivative $D^{(k)}(\omega)$. Using this idea we can easily generalize the quality measure for the Fourier approximation $H_D(\omega)$ as follows:

$$\mathcal{E}_k(D)=\langle D^{(k)}-H_D^{(k)}, D^{(k)}-H_D^{(k)}\rangle. \tag{6}$$

Since derivatives amplify high frequencies, $\mathcal{E}_k(D)$ can be regarded as a measure that intends to amplify the difference between the two functions involved. The fact that the measure $\mathcal{E}_k(D)$ depends on $D(\omega)$ is very desirable. If $D(\omega)$ is not some specific function, but a member from a class of functions, then we can further optimize the measure and identify an optimum function $D(\omega)$. Notice that the filter design problem has exactly this characteristic. This is so because, as we said, the ideal response is not explicitly given inside the transition region and thus can be the subject of an optimum selection

[1] The results presented in Sections II–IV for the symmetric case can be easily extended to the antisymmetric case as well.

method. Also notice that requiring in Lemma 1 the derivative $D^{(k)}(\omega)$ to be a well behaved function seems necessary for a meaningful definition of the criterion $\mathcal{E}_k(D)$. If for example $D^{(k)}(\omega)$ contains Dirac functions then, when attempting to compute (6), we will be confronted with integrals of products of Dirac functions that cannot be properly handled.

In the next section we are going to present the optimization problem focused on the approximation of symmetric functions by trigonometric polynomials. The complete solution yielding the optimum symmetric function and the corresponding Fourier approximation will be presented in detail.

III. Optimization Criterion and Optimum Approximation

Let $0=\omega_0<\omega_1<\omega_2<\omega_3<\cdots<\omega_{M-2}<\omega_{M-1}=\pi$ be any M distinct points on the interval $[0\ \pi]$ and let $D(\omega)$ be a symmetric function defined on this interval as follows:

$$D(\omega)=\begin{cases} F(\omega)\ \omega\in[\omega_{2(i-1)}\ \omega_{2i-1}], & i=1,\cdots,N_u \\ G(\omega)\ \omega\in(\omega_{2i-1}\ \omega_{2i}), & i=1,\cdots,N_t \end{cases} \quad (7)$$

where $N_u=\lceil M/2\rceil, N_t=\lfloor M/2\rfloor$ and $F(\omega)$ is assumed known while $G(\omega)$ is unknown. Let us denote with $\mathcal{U}_i,\mathcal{T}_i$ the intervals $[\omega_{2(i-1)}\ \omega_{2i-1}]$ and $(\omega_{2i-1}\ \omega_{2i})$, respectively, and $\mathcal{U}=\cup_{i=1}^{N_u}\mathcal{U}_i, \mathcal{T}=\cup_{i=1}^{N_t}\mathcal{T}_i$. Notice that the region $\mathcal{U}$ is the union of the N_u closed disjoint intervals $\mathcal{U}_i$ where $D(\omega)$ is assumed known, while the region $\mathcal{T}$ is the union of the N_t open disjoint intervals $\mathcal{T}_i$ where $D(\omega)$ is assumed unknown.

Since the part of the symmetric function $D(\omega)$ inside the region $\mathcal{T}$ is not explicitly given this means that, by varying $G(\omega)$, we can have a whole class of possible functions $D(\omega)$.

A. Constrained Optimization Criterion

It is clear that our intention is to use the measure $\mathcal{E}_k(D)$ defined in (6) and apply it to $D(\omega)$ of (7). For this to be possible we need to make certain assumptions on the given function $F(\omega)$ and impose certain constraints on the symmetric function $D(\omega)$ for the validity of Lemma 1. The basic requirement of Lemma 1 is the piecewise continuity of the kth derivative $D^{(k)}(\omega)$. To ensure this property we make the following assumption $\mathcal{A}$ for $F(\omega)$:

- $\mathcal{A}$ The function $F(\omega)$ describes the behavior of the symmetric function $D(\omega)$ inside the region $\mathcal{U}$. We assume that this function is continuous and has continuous derivatives of order up to $k-1$ and a piecewise continuous derivative of order k inside each closed interval $\mathcal{U}_i$.

We also impose the following constraints on $D(\omega)$:

- $\mathcal{C}1$. The symmetric function $D(\omega)$ (and thus $G(\omega)$) is continuous and has continuous derivatives of any order and for any ω inside the open intervals $\mathcal{T}_i, i=1,\cdots,N_t$.
- $\mathcal{C}2$. The symmetric function $D(\omega)$ is continuous and has continuous derivatives of order up to $k-1$ at all end points $\omega_i, i=1,\cdots,M$.

It is easy to see that constraints $\mathcal{C}1,\mathcal{C}2$ combined with assumption $\mathcal{A}$ ensure the piecewise continuity of the kth derivative of the symmetric function $D(\omega)$ and thus the validity of Lemma 1. This in turn allows for the use of $\mathcal{E}_k(D)$ as a proper performance measure.

By taking into account the definition of the intervals $\mathcal{U}_i,\mathcal{T}_i$, we can now write $\mathcal{C}2$ more formally with the help of the following conditions:

$$D^{(m)}(\omega_{2i-1})=F^{(m)}(\omega_{2i-1}-), \qquad m=0,\cdots,k-1,\ i=1,\cdots,N_t \quad (8)$$

$$D^{(m)}(\omega_{2i})=F^{(m)}(\omega_{2i}+), \quad m=0,\cdots,k-1,\ i=1,\cdots,N_t \quad (9)$$

where $F^{(m)}(\alpha-), F^{(m)}(\alpha+)$ denote the mth order left and right derivatives of the function $F(\omega)$ evaluated at the point α. Assumption $\mathcal{A}$ combined with constraint $\mathcal{C}1$ and (8), (9) are sufficient, as we are going to see, to uniquely identify the optimum symmetric function.

Concluding, for a given function $F(\omega)$ defined over $\mathcal{U}$ and satisfying $\mathcal{A}$, we like to minimize $\mathcal{E}_k(D)$ over all symmetric functions $D(\omega)$ given by (7) and satisfying $\mathcal{C}1$, (8) and (9). It is clear that the symmetric function $D(\omega)$ needs to be defined only inside the region $\mathcal{T}$ since in the region $\mathcal{U}$ it is already known.

B. The Optimum Symmetric Function $D_o(\omega)$

Let $D_o(\omega)$ denote the optimum symmetric function solving the constrained minimization problem defined in the previous subsection. Let also $H_o(\omega)$ denote the corresponding Fourier approximation of $D_o(\omega)$. We can now prove the following theorem for $D_o(\omega), H_o(\omega)$

Theorem 1: If we like to minimize $\mathcal{E}_k(D)$ over $D(\omega)$ under the constraints $\mathcal{C}1$, (8) and (9) then the following differential equation is a necessary and sufficient condition that must be satisfied by the optimum symmetric function and its corresponding Fourier approximation

$$D_o^{(2k)}(\omega)-H_o^{(2k)}(\omega)=0, \quad \text{for } \omega\in\mathcal{T}_i, \qquad i=1,\cdots,N_t. \quad (10)$$

Proof: The proof is given in Appendix A. ■

The differential equation defined by Theorem 1, as we can see, is valid inside every open interval $\mathcal{T}_i$. Since these intervals are disjoint we have a different solution for each $\mathcal{T}_i$. Solving the differential equation we obtain the following relation for $D_o(\omega), H_o(\omega)$:

$$D_o(\omega)=H_o(\omega)+Q_i(\omega), \quad \text{for } \omega\in\mathcal{T}_i, \qquad i=1,\cdots,N_t \quad (11)$$

where $Q_i(\omega)$ is a polynomial of degree $2k-1$, being in general different for each interval $\mathcal{T}_i$. We realize from (11) that the optimum form of the symmetric function inside the region $\mathcal{T}$ is a combination of a regular and a trigonometric polynomial.

To this end, let $\boldsymbol{h}_o$ denote the Fourier coefficients corresponding to the optimum symmetric function $D_o(\omega)$, then $H_o(\omega)=\boldsymbol{\phi}^t(\omega)\boldsymbol{h}_o$. Define now the following vector function of length $2k$:

$$\boldsymbol{\psi}(\omega)=[1 \quad \omega \quad \omega^2 \quad \cdots \quad \omega^{2k-1}]^t \quad (12)$$

and for each polynomial $Q_i(\omega)$ the corresponding vector of coefficients

$$\boldsymbol{q}_i = [q_{i,0} \quad q_{i,1} \quad q_{i,2} \quad \cdots \quad q_{i,2k-1}]^t \tag{13}$$

we then have $Q_i(\omega) = \boldsymbol{\psi}^t(\omega)\boldsymbol{q}_i$. Following our derivation up to this point, we conclude that the number of unknowns has increased from N to $N + 2kN_t$ (N for the Fourier coefficients and $2k$ for the coefficients of each polynomial $Q_i(\omega)$). Specifically the vectors $\boldsymbol{h}_o, \boldsymbol{q}_i, i = 1, \cdots, N_t$ form the complete set of unknowns needed to be specified to solve the optimization problem.

Since $\boldsymbol{h}_o$ are the Fourier coefficients corresponding to $D_o(\omega)$ they must satisfy (4), this means

$$\boldsymbol{h}_o = \langle \boldsymbol{\phi}, D_o \rangle = \langle \boldsymbol{\phi}, 1_{\mathcal{U}} D_o \rangle + \langle \boldsymbol{\phi}, 1_{\mathcal{T}} D_o \rangle \tag{14}$$

where $1_{\mathcal{X}}(\omega)$ denotes the index function of the set $\mathcal{X}$. Using (7), (11), and the parametrization of the polynomials $Q_i(\omega)$, we have that (14) can be written

$$\boldsymbol{h}_o = \langle \boldsymbol{\phi}, 1_{\mathcal{U}} F \rangle + \left(\sum_{i=1}^{N_t} \langle \boldsymbol{\phi}, 1_{\mathcal{T}_i} \boldsymbol{\phi}^t \rangle \right) \boldsymbol{h}_o + \sum_{i=1}^{N_t} \langle \boldsymbol{\phi}, 1_{\mathcal{T}_i} \boldsymbol{\psi}^t \rangle \boldsymbol{q}_i. \tag{15}$$

Let us now call

$$\begin{aligned} \boldsymbol{h}_{\mathcal{U}} &= \langle \boldsymbol{\phi}, 1_{\mathcal{U}} F \rangle, \quad A = \sum_{i=1}^{N_t} \langle \boldsymbol{\phi}, 1_{\mathcal{T}_i} \boldsymbol{\phi}^t \rangle, \\ B_i &= \langle \boldsymbol{\phi}, 1_{\mathcal{T}_i} \boldsymbol{\psi}^t \rangle \end{aligned} \tag{16}$$

then $\boldsymbol{h}_{\mathcal{U}}, A, B_i$ are quantities that depend only on known functions integrated over known sets. Thus they can be considered given. Notice that $\boldsymbol{h}_{\mathcal{U}}$ is a vector of length N, A is a matrix of dimensions $N \times N$ and B_i are matrices of dimensions $N \times 2k$. Equation (15) can now be written as

$$(I - A)\boldsymbol{h}_o - \sum_{i=1}^{N_t} B_i \boldsymbol{q}_i = \boldsymbol{h}_{\mathcal{U}} \tag{17}$$

which constitutes the first set of N equations involving the unknowns of our problem.

The remaining $2kN_t$ linear equations can be obtained by requiring the optimum solution to satisfy relations (8) and (9). Let us first define the following two matrix functions using the vector functions $\boldsymbol{\phi}(\omega), \boldsymbol{\psi}(\omega)$, and their derivatives:

$$M(\omega) = [\boldsymbol{\phi}(\omega) \boldsymbol{\phi}^{(1)}(\omega) \cdots \boldsymbol{\phi}^{(k-1)}(\omega)]^t \tag{18}$$

$$L(\omega) = [\boldsymbol{\psi}(\omega) \boldsymbol{\psi}^{(1)}(\omega) \cdots \boldsymbol{\psi}^{(k-1)}(\omega)]^t \tag{19}$$

where $M(\omega)$ has dimensions $k \times N$ and $L(\omega)$ has dimensions $k \times 2k$. Conditions (8) and (9) can now be, respectively, written

$$M(\omega_{2i-1})\boldsymbol{h}_o + L(\omega_{2i-1})\boldsymbol{q}_i = \boldsymbol{v}_{2i-1}, \qquad i = 1, \cdots, N_t \tag{20}$$

$$M(\omega_{2i})\boldsymbol{h}_o + L(\omega_{2i})\boldsymbol{q}_i = \boldsymbol{v}_{2i}, \qquad i = 1, \cdots, N_t \tag{21}$$

where the vectors $\boldsymbol{v}_{2i-1}, \boldsymbol{v}_{2i}$ are defined as follows:

$$\boldsymbol{v}_{2i-1} = [F(\omega_{2i-1}-) F^{(1)}(\omega_{2i-1}-) \cdots F^{(k-1)}(\omega_{2i-1}-)]^t \tag{22}$$

$$\boldsymbol{v}_{2i} = [F(\omega_{2i}+) F^{(1)}(\omega_{2i}+) \cdots F^{(k-1)}(\omega_{2i}+)]^t. \tag{23}$$

Since the matrices $M(\omega), L(\omega)$ and the end points of the intervals $\mathcal{T}_i$ are known we can see that (20), (21) constitute the remaining $2kN_t$ equations needed to solve our problem. Concluding we obtain the complete solution to the constrained optimization problem by solving the set of linear equations defined by (17), (20), and (21).

One point that needs to be stressed is the fact that the optimum solution depends on the integer k we select to use. In Section V we are going to propose a method for solving this parameter selection problem. Before we concentrate ourselves on the computation of the optimum solution, let us consider the special case which arises for $k = 0$ and study its correspondence to existing results.

C. Optimum Solution for $k = 0$

For this case there are no continuity constraints and the relation corresponding to the differential equation defined in (10) takes the form

$$D_o(\omega) - H_o(\omega) = 0 \quad \text{for } \omega \in \mathcal{T}_i, \qquad i = 1, \cdots, N_t. \tag{24}$$

From (24) we conclude that, inside the region $\mathcal{T}$, the optimum symmetric function $D_o(\omega)$ coincides with its Fourier approximation. This means that, when we form the MSE, the region $\mathcal{T}$ has no contribution to the corresponding integral. Equivalently, the region $\mathcal{T}$ becomes a "don't care" region. If we apply this idea to the filter design problem then the resulting method (for $k = 0$) coincides with the method proposed in [24, p. 70] and presented in detail in [4] and [5].

IV. Computation of the Optimal Solution, Complexity Issues

Let us now concentrate on the solution of the linear system defined by (17), (20), and (25). Solving (17) for $\boldsymbol{h}_o$ we obtain

$$\boldsymbol{h}_o = (I - A)^{-1}\boldsymbol{h}_{\mathcal{U}} + \sum_{i=1}^{N_t} (I - A)^{-1} B_i \boldsymbol{q}_i. \tag{25}$$

Substituting (25) into (20) and (21) we have

$$\begin{aligned} &M(\omega_{2i-1}) \sum_{j=1}^{N_t} (I - A)^{-1} B_j \boldsymbol{q}_j + L(\omega_{2i-1})\boldsymbol{q}_i \\ &\quad = \boldsymbol{v}_{2i-1} - M(\omega_{2i-1})(I - A)^{-1}\boldsymbol{h}_{\mathcal{U}} \end{aligned} \tag{26}$$

$$\begin{aligned} &M(\omega_{2i}) \sum_{j=1}^{N_t} (I - A)^{-1} B_j \boldsymbol{q}_j + L(\omega_{2i})\boldsymbol{q}_i \\ &\quad = \boldsymbol{v}_{2i} - M(\omega_{2i})(I - A)^{-1}\boldsymbol{h}_{\mathcal{U}}. \end{aligned} \tag{27}$$

Notice that (26) and (27) constitute a linear system of $2kN_t$ equations with $2kN_t$ unknowns (the vectors $\boldsymbol{q}_i, i = 1, \cdots, N_t$). In order to specify the above system the main computational cost is devoted in obtaining the matrices $(I - A)^{-1}B_j, j = 1, \cdots, N_t$ and the vector $(I - A)^{-1}\boldsymbol{h}_{\mathcal{U}}$. This is so because A is a matrix of size $N \times N$ and, as we will see in the examples, we have $N \gg 2kN_t$. It is clear that all desired quantities can be computed by solving a number of linear systems of the form

$$(I - A)\boldsymbol{x} = \boldsymbol{y} \tag{28}$$

where the matrix $C = I - A$ is common to all systems.

We are now going to investigate the special structure of the matrix C and see how we can use it to efficiently solve a system of the form of (28). From the definition of A in (16) we have that the elements of C satisfy

$$C_{n,m} = \begin{cases} 1 - \sum_{i=1}^{N_t} \int_{\mathcal{T}_i} \phi_n^2(\omega)\, d\omega, & \text{for } n = m \\ -\sum_{i=1}^{N_t} \int_{\mathcal{T}_i} \phi_n(\omega)\phi_m(\omega)\, d\omega, & \text{for } n \neq m. \end{cases} \tag{29}$$

After computing the integrals in (29), C can take the following form:

$$C = \begin{bmatrix} c_0 & \sqrt{2}\boldsymbol{c}^t \\ \sqrt{2}\boldsymbol{c} & T + H \end{bmatrix} \tag{30}$$

where T is a $(N-1)\times(N-1)$ Toeplitz symmetric matrix, H is a $(N-1)\times(N-1)$ Hankel matrix and $\boldsymbol{c}$ is a vector of length $(N-1)$ defined as follows:

$$T = \begin{bmatrix} c_0 & c_1 & \cdots & c_{N-2} \\ c_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_1 \\ c_{N-2} & \cdots & c_1 & c_0 \end{bmatrix} \tag{31}$$

$$H = \begin{bmatrix} c_2 & c_3 & \cdots & c_N \\ c_3 & c_4 & \cdots & c_{N+1} \\ \vdots & \vdots & \cdots & \vdots \\ c_N & c_{N+1} & \cdots & c_{2(N-1)} \end{bmatrix} \tag{32}$$

$$\boldsymbol{c} = [c_1 \quad c_2 \quad \cdots \quad c_{N-1}]^t \tag{33}$$

where

$$c_n = \begin{cases} 1 - \sum_{i=1}^{N_t} (\omega_{2i} - \omega_{2i-1}) & n = 0 \\ \frac{1}{n} \sum_{i=1}^{N_t} (\sin(n\omega_{2i-1}) & \\ \quad - \sin(n\omega_{2i})) & n = 1, \cdots, 2(N-1). \end{cases} \tag{34}$$

By properly applying Schur's matrix inversion formula [16, p. 658] we can reduce the linear system in (28) to one of size $N-1$ involving the part $T+H$ which is Toeplitz plus Hankel. Using methods as in [19], the Toeplitz plus Hankel structure can be reduced into a block Toeplitz structure and using a block Levinson algorithm we can efficiently obtain the solution. In our case a simpler reduction, leading to a more efficient solution, is possible. Specifically our system can be reduced to a pure symmetric Toeplitz system with the additional characteristic that the desired solution is also symmetric. Let us partition the length N vectors $\boldsymbol{x}_N, \boldsymbol{y}_N$ as follows: $\boldsymbol{x}_N = [x_0 \boldsymbol{x}_{N-1}], \boldsymbol{y}_N = [y_0 \boldsymbol{y}_{N-1}]$ then, the system $C\boldsymbol{x}_N = \boldsymbol{y}_N$ is equivalent to the following symmetric system:

$$\begin{bmatrix} T & J\boldsymbol{c} & JH \\ \boldsymbol{c}^t J & c_0 & \boldsymbol{c}^t \\ JH & \boldsymbol{c} & T \end{bmatrix} \begin{bmatrix} \frac{J\boldsymbol{x}_{N-1}}{\sqrt{2}} \\ x_0 \\ \frac{\boldsymbol{x}_{N-1}}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{J\boldsymbol{y}_{N-1}}{\sqrt{2}} \\ y_0 \\ \frac{\boldsymbol{y}_{N-1}}{\sqrt{2}} \end{bmatrix} \tag{35}$$

where the $(N-1)\times(N-1)$ matrix J denotes the exchange matrix with unities on the antidiagonal, and zeros elsewhere. It is easy to verify that the coefficient matrix in (35) is Toeplitz and symmetric. Although we have increased the size of the problem by a factor of two, as we can see, the desired solution is symmetric. Thus instead of applying the normal Levinson algorithm [13] to obtain the solution, we can use a symmetric version that is more efficient. Another possibility is to apply the symmetric Split Levinson algorithm [9] that uses only symmetric quantities which is more proper for our problem and has lower complexity. Both algorithms are presented in Table I. It is well known [9], [13] that the Levinson and the Split Levinson algorithms have complexity of the order of $O(N^2)$ as compared to the $O(N^3)$ required by a general solution algorithm. There also exist other methods based on FFT that can reduce the complexity to $O(N\log(N))$ [12], [15] but these methods are more complicated and more susceptible to numerical errors. We must stress one more time that since the systems we like to solve have all the same matrix C this means that the prediction part of both algorithms in Table I (which is the most computationally expensive) needs to be computed only once for all problems of the form of (28).

V. Design of Zero Phase FIR Digital Filters

In this section we are going to adapt our results to the design of the odd length[2] zero phase FIR digital filters. To this end let us assume that the region $\mathcal{U}$ is the union of the passbands and stopbands of the desired filter while the region $\mathcal{T}$ coincides with the union of the transition bands. Then, the function $F(\omega)$ constitutes the ideal response of the desired filter inside the passbands and stopbands while the function $G(\omega)$ describes the behavior of the ideal response inside the transition regions of the filter and is not explicitly given. Under these assumptions it is easy to see that the filter design problem can be considered as a special case of the general approximation problem defined in Section III.

We have seen that the performance measure $\mathcal{E}_k(D)$ introduced in the previous sections is tractable because it yields optimum ideal responses and corresponding Fourier approximations that are easily computable. Since in this section our intention is to make comparisons, it seems unfair to use our performance measure for this purpose. The reason is that the proposed measure clearly favors our design method. The correct thing to do is to select a measure that is consistent with the idea of optimality existing in practice.

It is well known that min–max filters are considered as the most desirable filters. Since they result from the minimization of the L_∞ criterion it is reasonable to ask how any other filter compares with this criterion. Practically this means the need to identify the maximum approximation error ripple inside the passband and stopband. Thus when comparing any number of filters, the filter with the smallest maximum ripple can be characterized as the best. A consequence of this idea is of course the fact that no filter is better than the min–max equiripple filter.

[2]The correspondence between the Fourier coefficients defined in (4) and the filter coefficients $a_i, i = 0, \pm 1, \cdots, \pm(N-1)$ is $a_0 = h_0/\sqrt{\pi}$ and $a_i = a_{-i} = h_i/(\sqrt{2\pi}), i = 1, \cdots, (N-1)$.

TABLE I
THE SYMMETRIC LEVINSON AND SPLIT LEVINSON ALGORITHMS FOR THE SOLUTION OF SYMMETRIC TOEPLITZ LINEAR SYSTEMS

We like to solve the systems of equations:
$R_{2m-1}\boldsymbol{x}_{2m-1} = \boldsymbol{y}_{2m-1}$, $m = 1, \ldots, N$, where
R_{2m-1} is a symmetric Toeplitz matrix with first column $[r_0 \ \cdots \ r_{2m-2}]^t$
$\boldsymbol{x}_{2m-1}$ is the symmetric vector $[x_{m-1} \ \cdots \ x_0 \ \cdots \ x_{m-1}]^t$
$\boldsymbol{y}_{2m-1}$ is the symmetric vector $[y_{m-1} \ \cdots \ y_0 \ \cdots \ y_{m-1}]^t$

Symmetric Levinson Algorithm

Prediction Part: From order $2m-3$ we have available
the predictor $\boldsymbol{a}_{2m-3}$ and the power α_{2m-3}

For $n = 2m-2, \ 2m-1$ apply the formulas

$$k_n = \frac{[r_{n-1} \ \cdots \ r_1]\boldsymbol{a}_{n-1}}{\alpha_{n-1}}$$
$$\alpha_n = \alpha_{n-1}(1-k_n^2)$$
$$\boldsymbol{a}_n = \begin{bmatrix} \boldsymbol{a}_{n-1} \\ 0 \end{bmatrix} - k_n \begin{bmatrix} 0 \\ J_{n-1}\boldsymbol{a}_{n-1} \end{bmatrix}$$

Filtering Part: From order $2m-3$ we have available the solution $\boldsymbol{x}_{2m-3}$

$$\epsilon_{2m-1} = y_{m-1} - [r_1 \ \cdots \ r_{2m-3}]\boldsymbol{x}_{2m-3}$$
$$\boldsymbol{x}_{2m-1} = \begin{bmatrix} 0 \\ \boldsymbol{x}_{2m-3} \\ 0 \end{bmatrix} + \left(\frac{\epsilon_{2m-1}}{\alpha_{2m-1}}\right)\left(\boldsymbol{a}_{2m-1} + J_{2m-1}\boldsymbol{a}_{2m-1}\right)$$

Initial Conditions: $\boldsymbol{a}_1 = 1$, $\alpha_1 = r_0$, $\boldsymbol{x}_1 = y_0/r_0$

Symmetric Split Levinson Algorithm

Prediction Part: From order $2m-3$ we have available
the predictors $\boldsymbol{a}_{2m-3}$, $\boldsymbol{a}_{2m-4}$ and the powers α_{2m-3}, α_{2m-4}, β_{2m-3}

For $n = 2m-2, \ 2m-1$ apply the formulas

$$\beta_n = [r_1 \ \cdots \ r_{n-1}]\boldsymbol{a}_{n-1}$$
$$k_n = \frac{\alpha_{n-1}}{\alpha_{n-2}}$$
$$\boldsymbol{a}_n = \begin{bmatrix} \boldsymbol{a}_{n-1} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \boldsymbol{a}_{n-1} \end{bmatrix} - k_n \begin{bmatrix} 0 \\ \boldsymbol{a}_{n-2} \\ 0 \end{bmatrix}$$
$$\alpha_n = \alpha_{n-1} + \beta_n - k_n\beta_{n-1}$$

Filtering Part: From order $2m-3$ we have available the solution $\boldsymbol{x}_{2m-3}$

$$\epsilon_{2m-1} = y_{m-1} - [r_1 \ \cdots \ r_{2m-3}]\boldsymbol{x}_{2m-3}$$
$$\boldsymbol{x}_{2m-1} = \begin{bmatrix} 0 \\ \boldsymbol{x}_{2m-3} \\ 0 \end{bmatrix} + \left(\frac{\epsilon_{2m-1}}{\alpha_{2m-1}}\right)\boldsymbol{a}_{2m-1}$$

Initial Conditions: $\boldsymbol{a}_1 = 1$, $\boldsymbol{a}_2 = [1 \ 1]^t$, $\alpha_1 = r_0$, $\alpha_2 = r_0 + r_1$, $\beta_2 = r_1$, $\boldsymbol{x}_1 = y_0/r_0$

Based on what we said above we can now propose a means for selecting the parameter k of our method. We can apply the method for values of k ranging from 0 up to some prescribed value k_{max} and select the solution that yields the smallest maximum ripple. The whole process seems to require a large number of operations. Fortunately in practice we only need to consider a limited number of k values. More precisely when we are interested in filters with length larger than 50 ($N > 25$) it seems that $k = 1$ has always the best performance. This was observed in all design examples we considered, noting that case $k = 0$ never occurred as the optimum.

An additional property that must be pointed out is the form of the optimum ideal response inside the transition region $\mathcal{T}$. Notice that it is a combination of a regular and a trigonometric polynomial (the frequency response of the filter) as compared to splines or trigonometric polynomials used by other L_2 based existing methods.

Let us now apply our method to two different filter design problems and compare it to other existing techniques. Specifically we are going to compare our method against the min-max equiripple [23], the eigenvalue [30], the don't care region [4], [24, p. 70] and the first order spline [4], [24, p. 69] filters.

A. Design of Low-Pass Filters

Let $D_{LP}(\omega)$ be the ideal response of a low-pass filter defined as follows:

$$D_{LP}(\omega) = \begin{cases} 1, & \omega \in [0 \ \omega_p] \\ 0, & \omega \in [\omega_s \ \pi] \\ G(\omega), & \omega \in (\omega_p \ \omega_s) \end{cases} \tag{36}$$

TABLE II
MAXIMUM APPROXIMATION ERROR RESULTED FROM THE DESIGN OF A LOW PASS FILTER BY DIFFERENT METHODS AND FOR DIFFERENT VALUES OF N

N	k	Proposed	Eigen	Splines	Don't Care	Min-max
11	2	7.08×10^{-2}	1.23×10^{-1}	9.46×10^{-2}	1.22×10^{-1}	5.56×10^{-2}
21	1	1.68×10^{-2}	3.02×10^{-2}	4.86×10^{-2}	3.04×10^{-2}	1.05×10^{-2}
31	1	2.77×10^{-3}	5.54×10^{-3}	3.22×10^{-2}	5.53×10^{-3}	1.53×10^{-3}
41	1	5.61×10^{-4}	1.31×10^{-3}	2.50×10^{-2}	1.31×10^{-3}	3.29×10^{-4}
51	1	8.97×10^{-5}	2.41×10^{-4}	1.96×10^{-2}	2.40×10^{-4}	5.35×10^{-5}
61	1	1.94×10^{-5}	5.62×10^{-5}	1.67×10^{-2}	5.63×10^{-5}	1.19×10^{-5}
71	1	3.22×10^{-6}	1.04×10^{-5}	1.41×10^{-2}	1.04×10^{-5}	1.94×10^{-6}
81	1	7.07×10^{-7}	2.39×10^{-6}	1.26×10^{-2}	2.75×10^{-6}	4.23×10^{-7}
91	1	1.23×10^{-7}	4.51×10^{-7}	1.10×10^{-2}	4.46×10^{-7}	7.74×10^{-8}
101	1	2.66×10^{-8}	1.01×10^{-7}	1.01×10^{-2}	1.03×10^{-7}	1.69×10^{-8}

where ω_p, ω_s are the desired cutoff frequencies of the filter. For this case we have $N_u = 2, N_t = 1$. Notice also that the function $F(\omega)$ of (7), is defined through the first two branches of the ideal response $D_{LP}(\omega)$.

Constraints (8) and (9) take the following form:

$$D_{LP}(\omega_p) = 1, D_{LP}^{(m)}(\omega_p) = 0, \qquad m = 1, \cdots, k-1 \quad (37)$$

$$D_{LP}^{(m)}(\omega_s) = 0, \qquad m = 0, \cdots, k-1. \quad (38)$$

Consider the special case $\omega_p = 0.3$ and $\omega_s = 0.4$ where ω is normalized in [0 1]. In Table II we present the maximum ripple for all methods under comparison, for different values of N. Notice that the second column contains the value for the parameter k of our method that yields the smallest maximum ripple. We can conclude from Table II that our method has at least 50% smaller maximum ripple as compared to any other non min–max method. At the same time its performance is close to the optimum equiripple method (last column). We obtained similar results in all other design examples we considered with different values of the cutoff frequencies ω_p, ω_s. We must also note that our results were very close to the results obtained by using the modified window method of [20] (whenever it was possible to apply this method).

In Fig. 1 we plot the form of the optimum ideal response inside the transition region for the case $N = 21$. From Table II we have for this case that the optimum k equals unity. This means that we have continuity only in $D_o(\omega)$. In Fig. 2 we plot the approximation errors for our method (solid), min-max (half-tone), eigenvalue (dash), don't care region (dash-dot), and first-order spline (dot).

B. Design of Bandpass Filters

The ideal response $D_{BP}(\omega)$ of such a filter, is given by

$$D_{BP}(\omega) = \begin{cases} 1, & \omega \in [\omega_{p_1}\ \omega_{p_2}] \\ 0, & \omega \in [0\ \omega_{s_1}] \cup [\omega_{s_2}\ \pi] \\ G(\omega), & \omega \in (\omega_{s_1}\ \omega_{p_1}) \cup (\omega_{p_2}\ \omega_{s_2}) \end{cases} \quad (39)$$

where $\omega_{p_1}, \omega_{p_2}, \omega_{s_1}, \omega_{s_2}$ are the desired cutoff frequencies of the filter. For this case we can easily see that $N_u = 3$ and

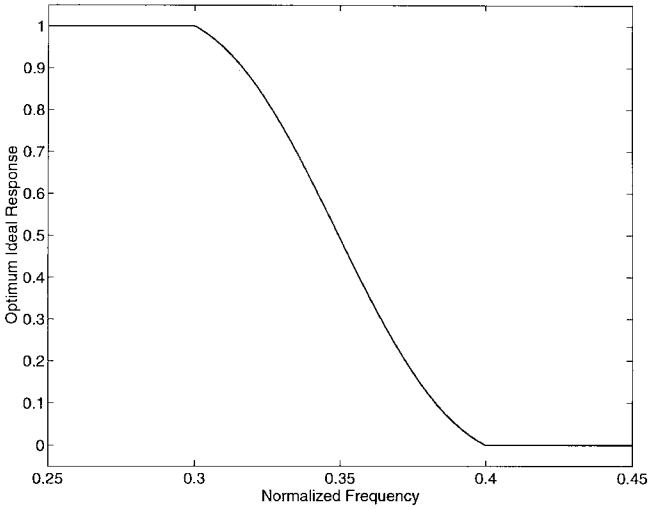

Fig. 1. Optimum ideal response for the design of a low-pass filter with $N = 21$, $\omega_p = 0.3$, and $\omega_s = 0.4$.

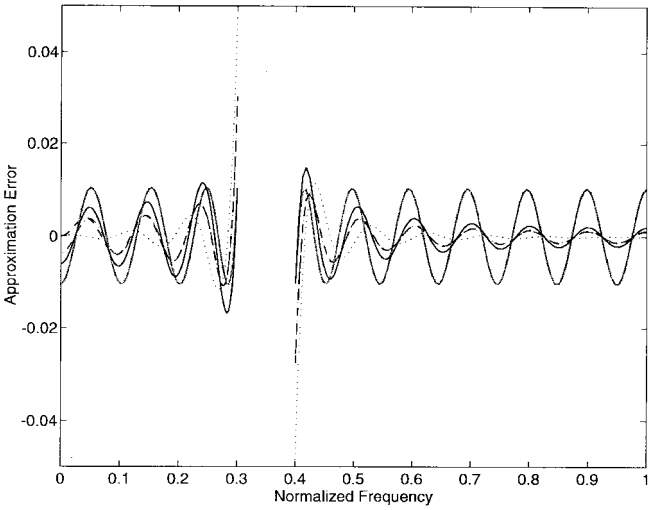

Fig. 2. Approximation errors for the design of a low pass filter with $N = 21$, $\omega_p = 0.3$, and $\omega_s = 0.4$. Proposed method (solid), min-max (half-tone), eigenfilter (dash), don't care region (dash–dot), and first-order spline (dot).

$N_t = 2$. As in the previous example the function $F(\omega)$ of (7) is defined through the first two branches of the ideal response $D_{BP}(\omega)$. Constraints (8) and (9) take the form

$$D_{BP}(\omega_{p_i}) = 1$$

$$D_{BP}^{(m)}(\omega_{p_i}) = 0, \qquad m = 1, \cdots, k-1, \quad i = 1, 2 \quad (40)$$

$$D_{BP}^{(m)}(\omega_{s_i}) = 0, \qquad m = 0, \cdots, k-1, \quad i = 1, 2. \quad (41)$$

Let us consider the special case $\omega_{s_1} = 0.2$, $\omega_{p_1} = 0.25$, $\omega_{p_2} = 0.65$, $\omega_{s_2} = 0.7$. Table III has, as in the case of the low pass filter, the performance of all methods under comparison for different values of N. Again for this type of design our method has at least 50% smaller maximum error than any other non min–max method. As before this result is valid for other combinations of cutoff frequencies as well.

Fig. 3 depicts the optimum ideal response, for the two intervals of the transition region and for $N = 21$. From the corresponding table we can see that the optimum value of k

TABLE III
MAXIMUM APPROXIMATION ERROR RESULTED FROM THE DESIGN OF A BANDPASS FILTER BY DIFFERENT METHODS AND FOR DIFFERENT VALUES OF N

N	k	Proposed	Eigen	Splines	Don't Care	Min-max
11	4	2.88×10^{-1}	4.28×10^{-1}	2.91×10^{-1}	2.99×10^{-1}	2.09×10^{-1}
21	2	7.04×10^{-2}	1.37×10^{-1}	9.44×10^{-2}	1.20×10^{-1}	5.56×10^{-2}
31	1	3.25×10^{-2}	6.87×10^{-2}	5.36×10^{-2}	5.89×10^{-2}	2.36×10^{-2}
41	1	1.62×10^{-2}	3.66×10^{-2}	4.85×10^{-2}	2.96×10^{-2}	1.04×10^{-2}
51	1	7.98×10^{-3}	1.99×10^{-2}	4.63×10^{-2}	1.46×10^{-2}	4.69×10^{-3}
61	1	2.72×10^{-3}	6.25×10^{-3}	3.22×10^{-2}	5.43×10^{-3}	1.54×10^{-3}
71	1	1.19×10^{-3}	3.02×10^{-3}	2.60×10^{-2}	2.59×10^{-3}	7.02×10^{-4}
81	1	5.44×10^{-4}	1.57×10^{-3}	2.50×10^{-2}	1.29×10^{-3}	3.27×10^{-4}
91	1	2.48×10^{-4}	8.59×10^{-4}	2.43×10^{-2}	6.34×10^{-4}	1.52×10^{-4}
101	1	8.78×10^{-5}	2.70×10^{-4}	1.96×10^{-2}	2.35×10^{-4}	5.35×10^{-5}

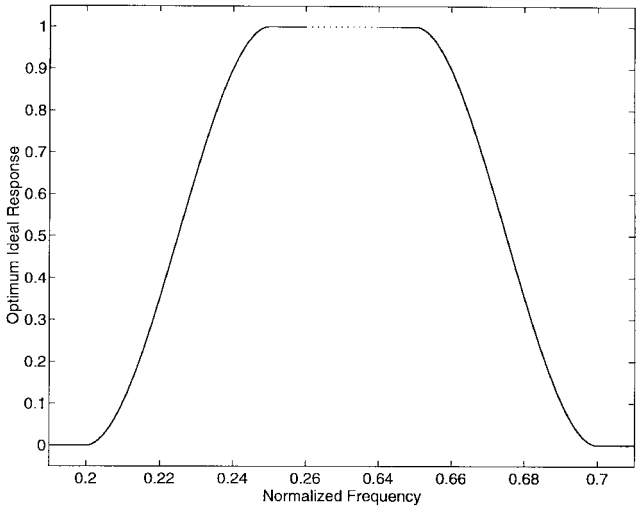

Fig. 3. Optimum ideal response for the design of a bandpass filter with $N = 21$, $\omega_{s1} = 0.2$, $\omega_{p1} = 0.25$, $\omega_{p2} = 0.65$, $\omega_{s2} = 0.7$.

is 2 meaning that $D_o(\omega)$ is continuous and has continuous derivatives. Finally, in Fig. 4 we plot the approximation error curves of the methods under comparison for $N = 21$.

C. Numerical Issues

Numerical problems can arise when designing filters with the proposed method. These problems mainly come from the solution of the Toeplitz systems of the form of (35). Regarding the two algorithms in Table I, that can be used for the solution of such systems, we observed that the normal symmetric Levinson algorithm seemed to have a slightly better numerical stability. On the other hand, the symmetric Split Levinson algorithm has the advantage of requiring less operations.

Numerical problems were also observed in other design methods as eigenfilters, "don't care" and Weighted Least Squares [4], [5]. These problems arise whenever we like to design filters with large N and large Δ (where Δ denotes the length of the largest transition band). For our method we

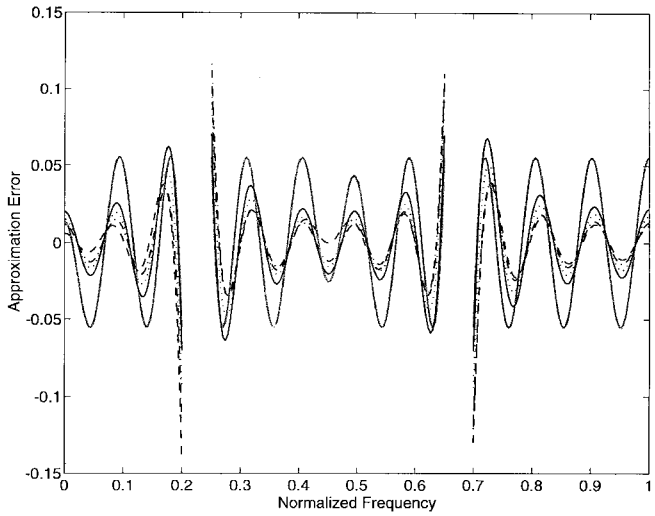

Fig. 4. Approximation errors for the design of a bandpass filter with $N = 21$, $\omega_{s1} = 0.2$, $\omega_{p1} = 0.25$, $\omega_{p2} = 0.65$, $\omega_{s2} = 0.7$. Proposed method (solid), min–max (half-tone), eigenfilter (dash), don't care region (dash–dot), and first-order spline (dot).

observed that when the symmetric Levinson is used and when $(2N-1)\Delta \leq 24$ (with frequencies normalized in [0 1]), the resulting solution is safely without numerical problems. On the other hand this upper bound is for most design problems pessimistic but, in any case, it is twice as large as the corresponding bound in [4]. In practice this limitation is not a serious problem because the filters we can design can achieve very high attenuation which is satisfactory in most cases.

VI. A NEW INITIALIZATION SCHEME FOR THE REMÉZ EXCHANGE ALGORITHM

The Reméz exchange algorithm is the most well-known method in the approximation theory for the minimization of the L_∞ criterion. The Parks–McClellan algorithm [23] constitutes an implementation of this iterative scheme dedicated to the design of min–max optimum FIR equiripple filters. Although this iterative scheme is very efficient in designing filters, it is known to require an increased number of iterations to converge. This need is more pronounced for large filters.

For most iterative algorithms, convergence speed depends on the initial solution estimate. Usually the better the initial estimate, the faster the method converges to the final solution. The fact that the convergence speed of the Parks–McClellan algorithm is very sensitive to the initial guess of the locations where the error function attains its extrema, was first pointed out in [10]. Specifically, a nonuniform distribution of the extrema, was suggested leading to an improvement in the convergence speed as compared to the uniform distribution suggested in the original Parks–McClellan algorithm. Our goal in this section is to prove that the optimum trigonometric polynomial obtained by our method for $k = 1$ can be regarded as a very desirable initial estimate for the Reméz exchange algorithm.

If $D(\omega)$ is the ideal symmetric function defined in (7) then, with the L_∞ criterion we are interested in defining a trigonometric polynomial $H(\omega)$ that approximates $D(\omega)$ in

the region $\mathcal{U}$ where $D(\omega)$ is exactly known. From [8, p. 175], [28, p. 54], we have that the optimum L_∞ approximation has an error function $E(\omega) = D(\omega) - H(\omega)$ that satisfies the alternation property. That is, there exist at least $N+1$ points in the region $\mathcal{U}$ where the error function has local minima with the same amplitude and alternating signs. We can thus conclude that a trigonometric polynomial can be regarded as an appropriate initial estimate for the Reméz exchange algorithm when its corresponding error function has at least $N+1$ local alternating extrema inside $\mathcal{U}$. Notice that we do not require the extrema to have the same amplitude, because this property is satisfied only by the L_∞ optimum solution [8, p. 175]. With the next theorem we show that the optimum trigonometric polynomial obtained by our method for $k=1$ satisfies exactly this requirement and thus can be used for initializing the Reméz exchange algorithm. We must also note that this is the main property that distinguishes our solution from other filter design techniques, namely, that with other methods we cannot guarantee the right number of alternations inside the passband and stopband.

Theorem 2: Let $D_o(\omega), H_o(\omega)$ be the optimum ideal response and its corresponding Fourier approximation resulting by the proposed method for $k=1$. Then, there exist at least $N+1$ points in the region $\mathcal{U}$ where the error function $E_o(\omega) = D_o(\omega) - H_o(\omega)$ exhibits alternating extrema.

Proof: The proof of Theorem 2 is given in Appendix B. ■

Based on Theorem 2 and the fact that our method yields very good approximations under the L_∞ criterion we expect that it will speed up the convergence of the Reméz exchange algorithm considerably. We applied our idea to the MATLAB program REMEZ.M by modifying its initialization part. In Table IV we present the number of iterations needed by the Reméz exchange algorithm to converge under the two initialization schemes and for different values of N. We again considered the problem of approximating the low-pass filter of Section V-A. From the table we can easily conclude that we have a significant gain in complexity and that this gain is increasing with increasing order.

TABLE IV
NUMBER OF ITERATIONS REQUIRED BY THE REMÉZ EXCHANGE ALGORITHM TO CONVERGE UNDER THE PROPOSED AND THE EXISTING INITIALIZATION SCHEME

N	Proposed	Existing
11	3	6
21	3	6
31	4	7
41	3	8
51	4	13
61	4	9
71	4	10
81	4	10
91	4	10
101	4	14

VII. CONCLUSION

We have presented a new L_2 based method for the design of FIR digital filters. By minimizing a suitable L_2 measure we were able to optimally define the part of the ideal response that was not explicitly specified in the design requirements. Since the proposed measure was also related to the Fourier approximation this led to corresponding optimum approximations that had improved performance. In a large number of design examples the method outperformed most existing non min–max methods while at the same time compared well with the optimum min–max equiripple approximation. The complexity of the proposed method was low because it required the solution of a symmetric Toeplitz linear system. Special symmetric versions of the Levinson algorithm were presented for efficiently solving the corresponding linear systems. Finally by proving that our optimum filter guarantees the necessary number of alternating extrema inside the passband and stopband required by the L_∞ criterion, an alternative initialization scheme for the well-known Reméz exchange algorithm was presented. Under this new scheme the convergency speed of the algorithm was significantly improved.

APPPENDIX A

Proof of Theorem 1: In order to prove the theorem, we are going to follow a classical variational techniques [11]. Let us consider a variation of the symmetric function around the optimum function $D_o(\omega)$. Specifically let us consider $D(\omega)$ of the form

$$D(\omega) = D_o(\omega) + \epsilon V(\omega) \tag{42}$$

where $D_o(\omega)$ is the optimum symmetric function, ϵ is a scalar parameter and $V(\omega)$ is a variation function. Notice first that any symmetric function $D(\omega)$ can be written under the form of (42). This is true since we can always select $\epsilon = 1$ and $V(\omega) = D(\omega) - D_o(\omega)$. From (42) we have that the Fourier approximation $H_D(\omega)$ of $D(\omega)$ can be written as

$$H_D(\omega) = H_o(\omega) + \epsilon H_V(\omega) \tag{43}$$

where $H_o(\omega)$ and $H_V(\omega)$ are the Fourier approximations for $D_o(\omega)$ and $V(\omega)$, respectively. Also we can conclude from (42) that since both $D(\omega)$ and $D_o(\omega)$ are symmetric functions satisfying (7) this means that the variation function $V(\omega)$ must satisfy the following constraints:

$$V(\omega) = 0 \quad \text{for } \omega \in \mathcal{U} \tag{44}$$

$$V^{(m)}(\omega_{2i-1}) = V^{(m)}(\omega_{2i}) = 0, \quad m = 0, \cdots, k-1, \quad i = 1, \cdots, N_t. \tag{45}$$

Applying (42) and (43) to the measure $\mathcal{E}_k(D)$ and using well known properties of the inner product we have

$$\begin{aligned}\mathcal{E}_k(D) = {} & \langle D_o^{(k)} - H_o^{(k)}, D_o^{(k)} - H_o^{(k)} \rangle \\ & + 2\epsilon \langle D_o^{(k)} - H_o^{(k)}, V^{(k)} - H_V^{(k)} \rangle \\ & + \epsilon^2 \langle V^{(k)} - H_V^{(k)}, V^{(k)} - H_V^{(k)} \rangle. \end{aligned} \tag{46}$$

This equation will be the base for proving our theorem.

Let us first prove the necessity of (10) of Theorem1. Consider a fixed variation function $V(\omega)$ and let ϵ be a variable. The criterion $\mathcal{E}_k(D)$ thus becomes a function $\mathcal{E}_k(\epsilon)$ of ϵ. We can see that this function has a minimum for $\epsilon = 0$ (remember that $D_o(\omega)$ is assumed to minimize $\mathcal{E}_k(D)$). This means that the first derivative of $\mathcal{E}_k(\epsilon)$ at $\epsilon = 0$ must be zero. From this we have

$$\langle D_o^{(k)} - H_o^{(k)}, V^{(k)} - H_V^{(k)} \rangle = 0. \tag{47}$$

Notice now that since $H_o^{(k)}(\omega)$ is the Fourier approximation for $D_o^{(k)}(\omega)$, because of Lemma 1, we will have that the difference $D_o^{(k)}(\omega) - H_o^{(k)}(\omega)$ is orthogonal to any trigonometric polynomial of order $N-1$ and consequently to $H_V^{(k)}(\omega)$. Thus (47) is equivalent to

$$\langle D_o^{(k)} - H_o^{(k)}, V^{(k)} \rangle = 0. \tag{48}$$

Since $V(\omega)$ is an arbitrary variation function, (48) is true for any variation function $V(\omega)$ satisfying (44), (45). Using the definition of the inner product and the fact that $V(\omega)$ is nonzero only inside the region $\mathcal{T}$, we have that (48) can be written as

$$\sum_{i=1}^{N_t} \int_{\omega_{2i-1}}^{\omega_{2i}} (D_o^{(k)}(\omega) - H_o^{(k)}(\omega)) V^{(k)}(\omega)\, d\omega = 0. \tag{49}$$

Integrating by parts and using (45) yields

$$(-1)^k \sum_{i=1}^{N_t} \int_{\omega_{2i-1}}^{\omega_{2i}} (D_o^{(2k)}(\omega) - H_o^{(2k)}(\omega)) V(\omega)\, d\omega = 0. \tag{50}$$

As we said $V(\omega)$ is an arbitrary variation function, thus we can select $V(\omega)$ to be identically zero in all intervals $\mathcal{T}_i$ except one. This means that

$$\int_{\omega_{2i-1}}^{\omega_{2i}} (D_o^{(2k)}(\omega) - H_o^{(2k)}(\omega)) V(\omega)\, d\omega = 0, \quad i = 1, \cdots, N_t \tag{51}$$

where again $V(\omega)$ is arbitrary inside $\mathcal{T}_i$ (provided it is zero at the end points). We can now conclude [11, p. 9] that in each interval $\mathcal{T}_i$ we have

$$D_o^{(2k)}(\omega) - H_o^{(2k)}(\omega) = 0 \tag{52}$$

which is the desired relation.

To prove the sufficiency, let $D_o(\omega), H_o(\omega)$ be a symmetric function and its corresponding Fourier approximation and assume that they satisfy (52). For any other symmetric function $D(\omega)$ [satisfying $\mathcal{A}, \mathcal{C}2$, (8), (9)] define $V(\omega) = D(\omega) - D_o(\omega)$, then $V(\omega)$ satisfies (44) and (45). From this we have that (50) is true and consequently (49), (48), and (47) are true as well. Substituting (47) in (46) (with $\epsilon = 1$) yields

$$\begin{aligned} \mathcal{E}_k(D) &= \langle D_o^{(k)} - H_o^{(k)}, D_o^{(k)} - H_o^{(k)} \rangle \\ &\quad + \langle V^{(k)} - H_V^{(k)}, V^{(k)} - H_V^{(k)} \rangle \\ &\geq \langle D_o^{(k)} - H_o^{(k)}, D_o^{(k)} - H_o^{(k)} \rangle = \mathcal{E}_k(D_o) \end{aligned} \tag{53}$$

and this completes the proof. ■

APPENDIX B

Before going to the proof of Theorem 2, let us first prove the following lemma.

Lemma 2: Let $D(\omega)$ be a continuous symmetric function on the interval $[0\pi]$, and let $H(\omega)$ be its Nth order Fourier approximation (optimum trigonometric polynomial of order $N-1$). Then the error function $E(\omega) = D(\omega) - H(\omega)$ either 1) vanishes identically or 2) changes sign at least N times inside $[0\pi]$.

Proof: In order to prove Lemma 2, we are going to use similar steps as in Theorem 5 [8, p. 110]. If $E(\omega)$ is identically zero, there is nothing to prove (case i)). Let us now assume that $E(\omega)$ is not identically zero, we will then prove that it satisfies ii). Since $D(\omega)$ is continuous by assumption and $H(\omega)$ is continuous as being a trigonometric polynomial, the error function $E(\omega)$ is continuous as well. Assume that the function $E(\omega)$ changes sign K times with $K < N$. This means that inside $[0\pi]$ there exist K points $0 < \eta_1 < \eta_2 < \cdots < \eta_K < \pi$ that define the $K+1$ intervals $[0, \eta_1], [\eta_1, \eta_2], \cdots [\eta_{K-1}, \eta_K], [\eta_K, \pi]$, inside which the function $E(\omega)$ has the following properties.

1) The function $E(\omega)$ is not identically zero inside each interval.
2) The nonzero values of $E(\omega)$ inside each interval have constant sign.
3) The nonzero values of $E(\omega)$ alternate sign between consecutive intervals.

Consider now the following symmetric function $P(\omega)$:

$$P(\omega) = \alpha \prod_{i=1}^{K} (\cos(\omega) - \cos(\eta_i)). \tag{54}$$

We can see that $P(\omega)$ also satisfies the above three properties. Actually regarding the first property it can be zero only at the K points η_i. Selecting properly the constant $\alpha \neq 0$ we can completely match the signs of $P(\omega)$ and $E(\omega)$. This means that $E(\omega)P(\omega) \geq 0$, for every $\omega \in [0\pi]$. By assumption $E(\omega)$ is not identically zero, while $P(\omega)$ can be zero only at a finite number of points. Since both functions are continuous we conclude

$$\langle E(\omega), P(\omega) \rangle > 0. \tag{55}$$

Relation (55) is a contradiction because $P(\omega)$ can be written as a linear combination of the orthonormal functions $\{\phi_n(\omega)\}_{n=1}^{K}$ and is therefore orthogonal to $E(\omega)$. Thus, the error function $E(\omega)$ changes sign at least N times inside the interval $[0\pi]$ and this concludes the proof of the lemma. ■

Proof of Theorem 2: First let us see how we can use the results of Lemma 2 to identify the number of alternating extrema. From the existence of the $N+1$ intervals where the function $E(\omega)$ is of alternating sign and the continuity of $E(\omega)$ we conclude that $E(\eta_i) = 0, i = 1, \cdots, N$. Thus in each of the $N+1$ intervals, $E(\omega)$ has an extremum. In other words we have proved the existence of $N+1$ alternating extrema. What is left to show is that all extrema can occur at points outside the transition region $\mathcal{T}$. From (11) we can conclude that if $E_o(\omega) = D_o(\omega) - H_o(\omega)$ is the error function of our optimum filter, then in every interval $\mathcal{T}_i$ the function $E_o(\omega)$

is a first-order polynomial of the form $q_{i,0} + q_{i,1}\omega$. Let us consider the following cases.

Case $q_{i,1} \neq 0$*:* The error function for this case is monotone thus it cannot have an extremum in the open interval $\mathcal{T}_i$. Extrema can at most occur at the edges of $\mathcal{T}_i$ but the edges are points in $\mathcal{U}$.

Case $q_{i,1} = 0$*:* For this case the error function, inside $\mathcal{T}_i$, is a constant equal to $q_{i,0}$. If $q_{i,0} = 0$ then, because of property 1 of Lemma 2, this value cannot be an extremal value. If $q_{i,0} \neq 0$ then the whole interval $\mathcal{T}_i$ must be a subset of some interval (say) $[\eta_j \eta_{j+1}]$. If the value $q_{i,0}$ is extremal for the interval $[\eta_j \eta_{j+1}]$ then, because of continuity, the error function will also assume this extremal value at both edges of $\mathcal{T}_i$. We can thus select either edge of $\mathcal{T}_i$ as a candidate for the occurrence of the extremum, thus avoiding $\mathcal{T}_i$.

Concluding, there always exist $N+1$ points outside the transition region $\mathcal{T}$ where the error function assumes alternating extremal values. This concludes the proof. ■

REFERENCES

[1] J. W. Adams, "FIR digital filters with least squares stopbands subject to peak-gain constraints," *IEEE Trans. Circuits Syst.*, vol. 39, pp. 376–388, Apr. 1991.

[2] ———, "A new optimal window," *IEEE Trans. Signal Processing*, vol. 39, pp. 1753–1769, Aug. 1991.

[3] R. Algazi *et al.*, "Design of almost minimax FIR filters in one and two dimensions by WLS techniques," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 590–596, June 1986.

[4] 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.

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

[6] P. L. Butzer and R. J. Nessel, *Fourier Analysis and Approximations, Vol. I.* Brussels, Belgium: Birkhäuser, 1971.

[7] C. Chi and S. Chiou, "A new WLS Chebyshev approximation method for the design of FIR digital filters with arbitrary complex frequency domain," *Signal Processing*, vol. 29, pp. 335–347, Dec. 1992.

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

[9] P. Delsarte and Y. V. Genin, "The split Levinson algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 470–478, June 1986.

[10] S. Ebert and U. Heute, "Accelerated design of linear or minimum phase FIR filters with a Chebyshev magnitude response," *Proc. Inst. Elect. Eng.*, vol. 130, pp. 267–270, Dec. 1983.

[11] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Englewood Cliffs, NJ: Prentice Hall, 1963.

[12] F. G. Gustavson and D. Y. Y. Yun, "Fast algorithms for rational Hermite approximation and solution of Toeplitz systems," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 750–754, Sept. 1979.

[13] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall,1991.

[14] H. D. Helms, "Digital filters with equiripple or minimax responses," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 87–94, Mar. 1971.

[15] J. R. Jain, "An efficient algorithm for a large Toeplitz set of linear equations," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-27, pp. 612–616, Dec. 1979.

[16] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice Hall,1980.

[17] Y. C. Lim *et al.*, "A weighted least squares algorithm for quasiequiripple FIR and IIR digital filter design," *IEEE Trans. Signal Processing*, vol. 40, pp. 551–558, Mar. 1992.

[18] M. T. McCalling and B. J. Leon, "Constrained ripple design of FIR digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 893–901, Nov. 1978.

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

[20] R. M. Mersereau *et al.*, "Selection of ideal filters for the window design method," *Signal Processing IV: Theory and Applications*, pp. 927–930, EURASIP, 1988.

[21] T. Q. Nguyen, "The design of arbitrary FIR digital filters using the eigenvalue method", *IEEE Trans. Signal Processing*, vol. 41, pp. 1128–1139, Mar. 1993.

[22] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1975.

[23] 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, Mar. 1972.

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

[25] S. Pei and J. Shyu, "Design of real FIR filters with arbitrary complex frequency responses by two real Chebyshev approximations," *Signal Processing*, vol. 26, pp. 119–129, May 1992.

[26] L. R. Rabiner, *Theory and Applications of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1975.

[27] L.R. Rabiner *et al.*, "FIR digital filter design techniques using weighted Chebyshev approximation," *Proc. IEEE*, vol. 63, pp. 595–609, Apr. 1975.

[28] J. R. Rice, *The Approximation of Function, vol. I.* Reading, MA: Addison-Wesley, 1964.

[29] 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.

[30] 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.

Emmanouil Z. Psarakis was born in Crete, Greece, in 1963. He received the Diploma degree in physics and the Ph.D. degree in computer engineering and informatics from the University of Patras, Greece, in 1985 and 1990, respectively.

Since 1993 he has been with the Computer Technology Institute (CTI) of Patras, Greece. His research interests include design and implementation of one dimensional and multidimensional digital filters, voice and image compression, and biomedical signal processing.

George V. Moustakides was born in Drama, Greece, in 1955. He received the diploma in electrical engineering from the National Technical University of Athens, Greece, in 1979, the M.Sc. degree in systems engineering from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1980, and the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ, in 1983.

From 1983 to 1986 he held a research position at the Institut de Resherche en Informatique et Systemes Aleatoires (IRISA-INRIA), Rennes, France; and from 1987 to 1990, a research position at the Computer Technology Institute (CTI) of Patras, Patras, Greece. From 1991 to 1996 he was an Associate Professor with the Department of Computer Engineering and Informatics, University of Patras, Patras, Greece, and, since 1996, he has been a Professor with the same department. His interests include adaptive estimation algorithms, design of classical filters, quickest detection procedures, and biomedical signal processing.