CCIT Report #584 April 2006

An L_1-Method for the Design of Linear-Phase FIR Digital Filters

Liron D. Grossmann Yonina C. Eldar

Abstract

This paper considers the design of linear-phase finite impulse response digital filters using an L_1 optimality criterion. The motivation for using such filters as well as a mathematical framework for their design is introduced. It is shown that L_1 filters possess flat passbands and stopbands while keeping the transition band comparable to that of least-squares filters. The uniqueness of L_1-based filters is explored, and an alternation type theorem for the unique frequency response is derived. An efficient algorithm for calculating the optimal filter coefficients is proposed, which may be viewed as the analogue of the celebrated Remez exchange method. A comparison with other deign techniques is made, demonstrating that the L_1 approach may be a good alternative in several applications.

I. Introduction

Linear-phase finite impulse response (FIR) digital filters play an important role in many signal processing applications, for example, in multirate systems, image processing, and communication systems, to mention a few. Consequently, design methods for linear-phase filters have been intensively researched in the digital signal processing literature for over almost half a decade; see [1] and references therein.

The design of FIR filters has long been recognized as an approximation problem, where an ideal frequency response, usually a discontinuous function, is approximated by a finite number of smooth functions. Such an approximation problem usually consists of a tradeoff. On the one hand, the resulting filter should preserve the discontinuous behavior of the ideal response, i.e. sharp transitions. On the other hand, these filters are also required to be as flat as possible in the passbands and stopbands. It is widely known that these two requirements are contradictory [2].

The design process typically involves four steps [1]. First, defining a desired ideal frequency response. Second, choosing an allowed class of filters (e.g. length N FIR filters). Third, establishing a measure of "goodness" between the allowed filter and the ideal one. Clearly, different optimality criteria lead to different filter behavior, so this step provides a convenient way to handle the inherent tradeoff of the design problem. Finally, developing a computational method to find the coefficients of the best linear-phase filter. The choice of optimality criterion is often dictated by the existence of an efficient algorithm for calculating the optimal filter.

During the past forty years, numerous techniques for designing digital FIR filters have been suggested. The majority of them rely on one or a combination of the following optimality criteria: least-squares (L_2), minimax (L_∞) and maximally flat. A combined criterion for achieving a tradeoff between the least-squares and the minimax [3]. The use of L_p norm, $2 \leq p \leq \infty$ has also been suggested as successful measure for replacing the L_2 and L_∞ [4].

The authors are with the department of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. e-mails: lirongr@tx.technion.ac.il, yonina@ee.technion.ac.il.

This work was supported by the European Union's Human Potential Programme, under the contract HPRN-CT-2003-00285 (HASSIP).

Least-squares filters are popular due to two main reasons. First, the resulting optimal filters require the solution of a single linear system of equations [5], [6], [7], [8], [9], [10], [11], [12], which can be solved efficiently. The second reason for its popularity stems from the fact that minimizing the least-squares error has the physical meaning of energy minimization, which is also related to 'signal to noise ratio' associated with the signals to be filtered [3]. Another common method, related to the least-squares approach, is the windowing technique, which is also very easy to implement [13], [14].

Nevertheless, there are applications in which the overshoot resulting from the least-squares filter is not acceptable, and further reduction in the maximum error is required. In such cases, filters which are optimal in the minimax sense can be designed [15], [16], [17], [18]. The minimax filters are very common since they result in the minimum number of coefficients (and therefore minimum number of computations) for a given tolerance scheme, i.e. when the design specifications are given in terms of the maximum deviations in the passbands and stopbands. In addition, the existence of an efficient iterative design algorithm, the Remez method [15], [19], [20], makes them easy to implement. Referring to the inherent tradeoff of the filter design problem, minimax filters may be viewed as one of its extreme. For a given number of coefficients, they result in the narrowest transition band, but their passbands and stopbands exhibit the most non-flat behavior, i.e. an equiripple behavior.

Maximally flat filters are on the opposite extreme: for a given filter length, they possess a very flat passband and stopband, but the associated transition band is much wider than that of the minimax. The maximally flat filters are easy to compute and several closed form formula exist for their design [21], [22], [23]. However, the resulting transition bandwidth is usually very hard to control, which often makes them less attractive.

Attempting to further explore meaningful criteria for designing linear-phase FIR filters, we consider in this paper using the weighted L_1 norm for approximating digital filters. The use of L_1 norm as a measure of goodness is very common in several engineering applications, in particular in robust estimation problems, basis pursuit and sparse representations [24]. However, it has not received much attention and serious treatment in the filter design literature. In fact, we are aware of very few works dealing with the L_1 criterion [25], [26], [27]. In [25], the design of high-order differentiators was considered, and in [27] an arbitrary amplitude function was designed using the L_1 approach. A general algorithm for the approximation under L_p was proposed in [26], but convergence is not guaranteed for $p = 1$, and when exists is often very slow. Moreover, in all three papers the suggested algorithms are based on a discretization of the original continuous problem, which yields only an approximate solution. In order for the approximation to be accurate, the sampling grid must be dense, which becomes computationally demanding. In addition, no clear justification was given for the use of the L_1 measure in the context of filter design. We therefore believe that the two major reasons for the absence of L_1 filters are a lack of motivation and an efficient algorithm that solves the original problem. It is the goal of this paper to provide a strong motivation for the use of the L_1 criterion in the design of FIR linear-phase filters, and to propose an efficient and accurate algorithm for computing the optimal L_1 filter without the need for discretization.

After formulating our L_1 design problem in Section II, we introduce the mathematical framework of L_1-based filters in Section III. It is first shown that the error function is differentiable and sometimes even twice differentiable. A characterization of the optimal error function is also derived, which is analogous to the alternation thereom of the minimax filters. In addition, we develop a necessary and sufficient condition for the optimal filter to be unique. A simple Newton-type algorithm is proposed in Section IV, and its convergence and efficiency are further discussed in Section V. In Section VI we compare between the proposed algorithm and the Remez exchange method, and show that both methods shares several properties, especially fast running time. Some design examples are given in Section VII, with an emphasis on the properties of optimal L_1 filters with respect to other existing design approaches.

II. Problem Formulation and Motivation

A. Notations and Definitions

Matrices and vectors are denoted by bold font, with lowercase letters corresponding to vectors and uppercase letters to matrices. The nth element of a vector $\mathbf{a}$ is denoted by $\mathbf{a}_n$, and the (k,l) element of a matrix $\mathbf{H}$ by $\mathbf{H}_{k,l}$. We let $\{\mathbf{a}^k\}_{k=0}^{\infty}$ denote a sequence of vectors. Unless stated otherwise the norm of a vector $\mathbf{a}$ is taken to be the l_2 norm, and is denoted by $\|\mathbf{a}\|$, that is, $\|\mathbf{a}\| = \sqrt{\sum_{n=0}^{M} \mathbf{a}_n}$. The minimal and maximal eigenvalues of a matrix $\mathbf{H}$ are denoted by $\lambda_{min}(\mathbf{H})$ and $\lambda_{max}(\mathbf{H})$, respectively. The minimal and maximal singular values of a matrix $\mathbf{H}$ are denoted by $\sigma_{min}(\mathbf{H})$ and $\sigma_{max}(\mathbf{H})$, respectively.

The sign function of a function $X(\omega)$ is defined by

$$\text{sign}(X(\omega)) = \begin{cases} 1 & X(\omega) > 0 \\ 0 & X(\omega) = 0 \\ -1 & X(\omega) < 0. \end{cases} \tag{1}$$

The inner product between two real functions, f and g, on $[0,\pi]$ is written as

$$\langle f, g \rangle = \int_0^{\pi} W(\omega) f(\omega) g(\omega) d\omega. \tag{2}$$

For given ω_p and ω_s, the set $[0,\omega_p] \cup [\omega_s,\pi]$ is denoted by Ω.

We define the error function between a desired frequency response $D(\omega)$ and its approximation to be

$$E(\omega, \mathbf{a}) = \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega) - D(\omega). \tag{3}$$

The derivative of $E(\omega, \mathbf{a})$ with respect to ω is denoted by $E^{'}(\omega, \mathbf{a})$. The set of zeros of the error function is

$$Z(\mathbf{a}) = \{\omega \in [0,\pi] | E(\omega, \mathbf{a}) = 0 \ \text{ and } \ \mathrm{W}(\omega) > 0\}, \tag{4}$$

where $W(\omega)$ is a nonnegative weighting function, which we assume that is at least twice differentiable on Ω. The Lebesgue measure of $Z(\mathbf{a})$ is denoted by $\mu(Z(\mathbf{a}))$. When $Z(\mathbf{a})$ is an interval $\mu(Z(\mathbf{a}))$ is its length, and when $Z(\mathbf{a})$ is a finite set, $\mu(Z(\mathbf{a})) = 0$.

Definition 1 (Simple Zero): A zero of a function $X(\omega)$ is called **simple** if $X(\omega)$ changes sign at that point. Alternatively, a simple zero ω_1 satisfies $X(\omega_1) = 0$ and $X^{'}(\omega_1) \neq 0$.

B. The L_1 Design Problem

We consider the problem of approximating an ideal frequency response $D(\omega)$, $\omega \in [0,\pi]$, using an order N FIR filter with impulse response $\{\mathbf{h}_n, 0 \leq n \leq N\}$. We shall develop our design procedure by considering the basic low-pass filter

$$D(\omega) = \begin{cases} 1, & \omega \in [0,\omega_c], \\ 0, & \omega \in (\omega_c,\pi], \end{cases} \tag{5}$$

shown in Figure 1, for $\omega_c = 0.485\pi$ rad. High-pass filters are treated in the exact same manner. We briefly discuss multiband filters in Section VII.

The frequency response of the approximating filter $H(\omega)$ is given by the discrete time Fourier transform (DTFT)

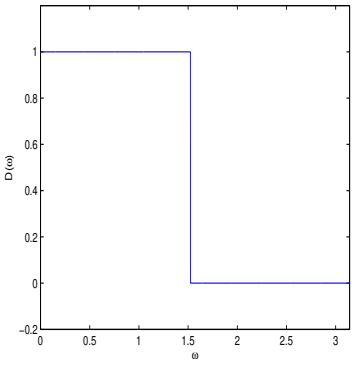


Fig. 1. An example of a desired frequency response.

of its impulse response $\mathbf{h}_n$:

$$H(\omega) = \sum_{n=0}^{N} \mathbf{h}_n e^{-jwn}. \tag{6}$$

For simplicity, we consider symmetric odd length filters (known as type-I filters), in which case $H(\omega)$ can be written as

$$H(\omega) = A(\omega)e^{-jM\omega} \tag{7}$$

where $M = \frac{N}{2}$, and $A(\omega)$ is the real-valued function

$$\begin{aligned} A(\omega) &= \mathbf{h}_M + \sum_{n=1}^{M} 2\mathbf{h}_{M-n}\cos(n\omega) \\ &\triangleq \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega). \end{aligned} \tag{8}$$

Since $D(\omega)$ is zero-phase, approximating it by $H(\omega)$ is equivalent to approximating it by $A(\omega)$, and then adding a delay of M taps to $A(\omega)$ to make $H(\omega)$ causal. Thus, our problem is to approximate $D(\omega)$ by a linear combination of $M+1$ functions $\{\cos(n\omega), n = 0, ..., M\}$.

Let $E(\omega, \mathbf{a})$ denote the error of approximation defined in 3. When when the filter to be designed must be optimal under a certain norm, the approximation problem can be mathematically written as $\min_{\mathbf{a}} \|E(\omega, \mathbf{a})\|$, where $\|\cdot\|$ denotes a weighted norm. The two most popular norms used in FIR design are:

1) Weighted least-squares

$$\|E(\omega, \mathbf{a})\|_2 = \sqrt{\int_0^{\pi} W(\omega)|A(\omega) - D(\omega)|^2 d\omega}. \tag{9}$$

2) Weighted Chebyshev

$$\|E(\omega, \mathbf{a})\|_\infty = \max_{\omega \in [0,\pi]} W(\omega)|A(\omega) - D(\omega)|. \tag{10}$$

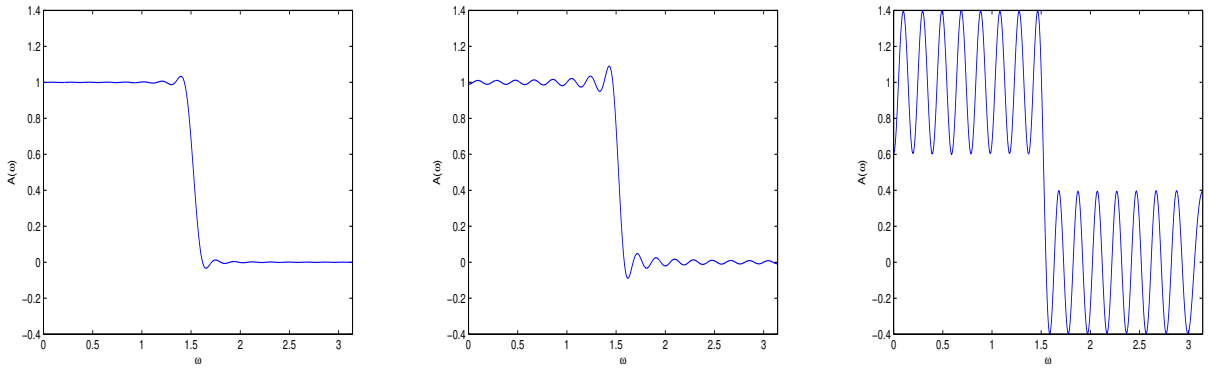
In this paper, we propose using the weighted L_1 norm given by

$$\|E(\omega, \mathbf{a})\|_1 = \int_0^{\pi} W(\omega)|A(\omega) - D(\omega)| d\omega. \tag{11}$$

Thus, our problem is to find $\mathbf{a}$ such that $\|E(\omega, \mathbf{a})\|_1$ is minimized.

The function $W(\omega)$ is a nonnegative weighting function, used in order to control or exclude certain frequency bands.

In order to motivate the use of L_1 filters, let us compare the L_1, L_2, L_∞ solutions with $W(\omega) = 1$ for all $\omega \in [0, \pi]$. Figure 2 shows the 65-length L_1, L_2, L_∞ FIR filters approximating the ideal low-pass filter shown in Fig. 1. Since we are approximating a discontinuous function, the L_1 and L_∞ filters were obtained by discretization methods, which do not require continuity.



(a) (b) (c)

Fig. 2. Approximations to the ideal low-pass filter in Fig. 1 over $[0, \pi]$ (a) L_1 (b) L_2 (c) L_∞.

For the least-squares problem in (9) with $W(\omega) = 1$ over $[0, \pi]$ it is well known that the resulting filter suffers from the Gibbs phenomenon [2]. This effect gives rise to an overshoot of 9% of the size of the jump. As can be inferred from the figure, the situation is even worse when trying to approximate $D(\omega)$ under the Chebyshev norm with $W(\omega) = 1$ in (10). In this case an overshoot of around 50% of the size of the jump is observed. The L_1 filter exhibits the smallest overshoot around the discontinuity. The fact that the Gibbs phenomenon (i.e. the overshoot behavior near the edge) is the least severe when the approximation norm is L_1 was conjectured in [28], where it was also stated that the overshoot becomes larger as p of the L_p norm increases. Thus, as long as the approximation of a discontinuous function over the entire frequency band $[0, \pi]$ is considered, the L_1 solution seems to be a better alternative to the least-squares (and, of course, to the minimax) method. Note that the fact that the L_1 solution results in the smallest overshoot does not contradict the fact that the minimax solution leads to the smallest maximal error. Indeed, the largest deviation is obtained at the cutoff frequency where the minimax error is the minimal among the three filters. In the rest of the interval the other perform better.

Figure 2 suggests another interesting property of the L_1 filter. In particular, the L_1 measure can provide a suitable solution to the inherent tradeoff of the filter design problem, that is flat passband and stopband versus sharp transition. Indeed, the L_1 filter possesses a flatter response in the passband and stopband than those of the L_2 and L_∞ filters. At the same time, the transition between the passband and stopband is only slightly wider than that of the L_2 approximation.

The overshoot of the L_1, L_2 and L_∞ filters can be further reduced by splitting the approximation of the ideal filter into a union of disjoint intervals of $[0, \pi]$, making the approximated function continuous. The separation into passband and stopband is accomplished by setting $W(\omega)$ to zero in (11). The approximation of a continuous function greatly simplifies the mathematical and algorithmic treatment of the underlying problem. For example, both the alternation theorem and the Remez exchange method for designing minimax filters rely on the fact that

$D(\omega)$ is continuous in Ω [15]. In addition, as opposed to the discontinuous case, the overshoot of the minimax filter will now be the smallest since the cutoff frequency is excluded. Nevertheless, the largest error is spread over the entire passband and stopband. The behavior of the overshoot of the L_1 filters when we approximate $D(\omega)$ over Ω will be explored in Section VII. In what follows, we therefore assume that the designer defines the passband and stopband edges, ω_p and ω_s, and that $W(\omega) = 0$ for $\omega \in (\omega_p, \omega_s)$.

Despite the apparent advantages of the L_1 approach, there are still two main questions with respect to the best L_1 approximation. While approximating $D(\omega)$ over Ω ensures a unique solution for both the L_2 and the L_∞ problems, uniqueness is not always guaranteed in the L_1 case. Uniqueness is important from both the theoretical and the algorithmic aspects. For instance, it is the uniqueness of the minimax filters that allows the use of the Remez method for computing it. The second question concerns the existence of an efficient algorithm for computing the coefficients of the optimal L_1 filter.

In the next section, we address the issue of uniqueness and provide a characterization of optimal L_1 filters, which lay the ground for the development of an efficient algorithm for calculating them in Section IV.

III. The Mathematical Framework of Optimal L_1 Filters

We begin by considering the differentiability of the L_1 norm with respect to the vector of coefficients.

A. Characterization of the Optimal Filter

Typically L_1 problems are solved by replacing the original continuous problem by a discrete one, and then using linear programming techniques [27], [29], [30]. Discretization techniques are also common for solving the corresponding L_2 and L_∞ measures [31]. While discretization algorithms yield very accurate solutions when applied to L_2 and L_∞ problems, they are usually impractical for tasks involving the L_1 criterion. Indeed, achieving the same degree of accuracy in an L_1 problem requires a much denser discretization grid than that of an L_2 or L_∞ one, which increases the computational complexity [32]. In addition, unlike the L_2 and L_∞ solutions, the L_1 norm is very sensitive to small perturbations, that is, a slight change in the coefficients may lead to a very different approximating function [33]. Thus, solving the continuous L_1 problem is rather inevitable. At the same time, however, we would like it to be as efficient as possible. There are many efficient optimization algorithms for smooth functions (such as the L_2 norm) and several methods for non-differentiable functions (such as the Chebyshev norm) [34]. Unfortunately, none is known to be efficient for solving the general L_1 approximation problem. Nevertheless, under a mild assumption, which holds true in our FIR design case, the L_1 norm can be differentiated and sometimes even twice differentiated. This allows the use of smooth optimization methods to solve our L_1 filter design problem.

The following theorem, which is a straightforward generalization of a theorem in [35], states a sufficient condition on a vector $\mathbf{a}^*$ such that $\|E(\omega, \mathbf{a})\|_1$ is differentiable at $\mathbf{a}^*$.

Theorem 1 (First Derivative): If $\mu(Z(\mathbf{a}^*)) = 0$, then $\|E(\omega, \mathbf{a})\|_1$ is differentiable at $\mathbf{a}^*$, and the nth component of the gradient at $\mathbf{a}^*$ is given by

$$\mathbf{g}_n(\mathbf{a}^*) = \langle \cos(n\omega), \operatorname{sign}(E(\omega, \mathbf{a}^*)) \rangle \,. \tag{12}$$

Fortunately, as shown in the next proposition, essentially all choices of filter coefficients satisfy the condition of Theorem 1.

Proposition 1: For the function $D(\omega)$ in (5), $\mu(Z(\mathbf{a})) = 0$ for every $\mathbf{a} \neq (1, 0, 0, ..., 0)$ and $\mathbf{a} \neq (0, 0, 0, ..., 0)$.

Proof: Since $A(\omega)$ is a degree M cosine polynomial on $[0,\pi]$, it cannot have more than M zeros on $[0,\pi]$ unless it is identically zero [35]. Since we assume $D(\omega)$ is piecewise constant, then in each subinterval where $D(\omega)$ is constant, $E(\omega,\mathbf{a})$ is an M degree polynomial. Thus, each such subinterval contains at most M zeros, resulting in an overall finite number of zeros in $[0,\pi]$. The case where $E(\omega,\mathbf{a})$ is identically zero on a subinterval may occur only if $A(\omega)$ equals one of the constant values $D(\omega)$ takes on, i.e. 0 or 1. This in turn can happen if and only if $\mathbf{a}$ has this constant as its first component, keeping the rest of the components zero. ■

Thus, except for the trivial case, where $A(\omega)$ is constant (corresponding to a length-one filter), we are guaranteed that the L_1 norm of the error is differentiable. As a result, we shall assume from now on that $N \geq 2$ and therefore $M \geq 1$. The case where $N = 0$ is treated in Appendix I.

Since $\|E(\omega,\mathbf{a})\|_1$ is convex in $\mathbf{a}$, we obtain the following characterization of the the L_1 solution.

Proposition 2: A vector $\mathbf{a} \in \mathbb{R}^{M+1}$ ($M \geq 1$) minimizes (11) if and only if

$$\langle \cos(n\omega), \operatorname{sign}(E(\omega,\mathbf{a}))\rangle = 0, \quad n = 0,...,M. \tag{13}$$

As the inner product is linear in the first term, it follows from this proposition that every M degree cosine polynomial has to be orthogonal to the sign of the optimal error function. It is also interesting to note the close resemblance of the weighted L_1 characterization to the weighted least-squares one, where for the latter the solution is characterized by the property that $E(\omega,\mathbf{a})$ (instead of $\operatorname{sign}(E(\omega,\mathbf{a}))$) is orthogonal to the functions $\{\cos(n\omega), n = 0,...,M\}$ [35]. However, whereas the orthogonality condition in the least-squares case leads to a set of linear equations, (13) describes a set of $M+1$ nonlinear equations in $M+1$ unknowns, the components of $\mathbf{a}$.

The next theorem allows us to further characterize the optimal L_1 solution.

Theorem 2 (Characterization of the Optimal Filter): Let $A(\omega)$ be an optimal weighted L_1 approximation to $D(\omega)$ on Ω of degree M, corresponding to a vector $\mathbf{a}$. Then $E(\omega,\mathbf{a})$ changes sign either M or $M+1$ times in Ω.

It is interesting to note that this theorem resembles the famous alternation theorem of the corresponding minimax filter problem [15], where the role of the zeros of the error function in the former is played by the extrema in the latter. Nevertheless, while the alternation theorem states that the optimal solution is also unique, the L_1 filter is not guaranteed to be so. The question of uniqueness is explored in the next subsection. The proof of Theorem 2 and its proof is deferred to Appendix II. From this theorem, we come to an important corollary.

Corollary 1: Let $\mathbf{a}^*$ be the coefficients of the optimal Nth order filter, $N \geq 2$. Then $\|E(\omega,\mathbf{a}^*)\|_1$ is differentiable.

Proof: By Theorem 1 we know that if $\mu(Z(\mathbf{a}^*)) = 0$, then $\|E(\omega,\mathbf{a}^*)\|_1$ is differentiable. We also know from Proposition 1 that $\mu(Z(\mathbf{a}^*)) = 0$ for every $\mathbf{a}^* \neq (1,0,0,...,0)$ and $\mathbf{a}^* \neq (0,0,0,...,0)$. We now show that $\mathbf{a}^*$ cannot be any one of these vectors. If $\mathbf{a}^* = (1,0,0,...,0)$ then $E(\omega,\mathbf{a}^*)$ does not change sign in Ω. However, by Theorem 2, $E(\omega,\mathbf{a}^*)$ must have at least $M \geq 1$ sign changes in Ω. A similar argument holds for the case when $\mathbf{a}^*$ is assumed to be the zero vector. ■

We now address the question of second order differentiability of $\|E(\omega,\mathbf{a})\|_1$. The existence of the Hessian matrix will allow us to develop an efficient algorithm for minimizing $\|E(\omega,\mathbf{a})\|_1$. The next theorem was originally stated for the unweighted L_1 norm defined over an interval, and given without proof in [36]. Here we generalize it to the weighted, disjoint union of intervals case. Since the generalization and its further consequences are not straightforward, we provide a full proof of the theorem.

Theorem 3 (Second-Order Derivative): Let $Z(\mathbf{a}) = \{z_1,...,z_t\}$ be the set of zeros of $E(\omega,\mathbf{a})$ in Ω and assume

that each zero is **simple**. Then the Hessian matrix of $\|E(\omega,\mathbf{a})\|_1$ is given by

$$\mathbf{H}(\mathbf{a}) = \mathbf{V}^T\mathbf{D}\mathbf{V} \tag{14}$$

where $\mathbf{V}$ is an $(M+1)\times t$ matrix with $\mathbf{V}_{ij}=\cos((j-1)z_i)$, and $\mathbf{D}=diag\{d_1,\ldots,d_t\}$ with $d_i=\frac{2W(z_i)}{|E^{'}(z_i,\mathbf{a})|}$.

Proof: We wish to compute $\frac{\partial^2\|E(\omega,\mathbf{a})\|_1}{\partial\mathbf{a}_i\partial\mathbf{a}_j}=\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial\mathbf{a}_j}$. By (12),

$$\mathbf{g}_i(\mathbf{a})=\int_0^{\pi}W(\omega)\cos(i\omega)\mathrm{sign}(E(\omega,\mathbf{a}))d\omega=\int_0^{z_1}v_0W(\omega)\cos(i\omega)d\omega+$$
$$\sum_{k=1}^{t-1}v_k\int_{z_k}^{z_{k+1}}W(\omega)\cos(i\omega)d\omega+\int_{z_t}^{\pi}v_tW(\omega)\cos(i\omega)d\omega,$$

where $v_0=\mathrm{sign}(E^{'}(0,\mathbf{a}))$, and $v_k=\mathrm{sign}(E^{'}(z_k,\mathbf{a}))$. The last equality is justified by the assumption that the zeros are simple. Using the chain rule we get,

$$\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial\mathbf{a}_j}=\sum_{k=1}^{t}\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial z_k}\frac{\partial z_k}{\partial\mathbf{a}_j}+\sum_{k=1}^{t}\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial v_k}\frac{\partial v_k}{\partial\mathbf{a}_j}.$$

Denoting $u_k=E^{'}(z_k,\mathbf{a})$, we write $v_k=\mathrm{sign}(u_k)$ and

$$\frac{\partial v_k}{\partial\mathbf{a}_j}=\frac{\partial v_k}{\partial u_k}\frac{\partial u_k}{\partial\mathbf{a}_j}.$$

Since z_k is assumed to be simple, $u_k\neq 0$, and thus $\frac{\partial v_k}{\partial u_k}=0$, resulting in $\frac{\partial v_k}{\partial\mathbf{a}_j}=0$. Now,

$$\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial z_k}=W(z_k)\cos(iz_k)(v_{k-1}-v_k)=-2v_kW(z_k)\cos(iz_k), \tag{15}$$

where we used the fact the zeros are assumed to be simple, and therefore v_{k-1} and v_k have opposite signs.

It remains to compute $\frac{\partial z_k}{\partial\mathbf{a}_j}$, where we note that the dependence of z_k on $\mathbf{a}_j$ is implicit, through the equations

$$E(z_k,\mathbf{a})=0,\qquad k=1,...,t.$$

Define the matrices $\mathbf{Z_a}=\left(\frac{\partial z_k}{\partial\mathbf{a}_j}\right)$, $\mathbf{F}_z=\left(\frac{\partial E(z_i,\mathbf{a})}{\partial z_j}\right)$, and $\mathbf{F_a}=\left(\frac{\partial E(z_i,\mathbf{a})}{\partial\mathbf{a}_j}\right)$. By the implicit function theorem,

$$\mathbf{Z_a}=-\mathbf{F}_z^{-1}\mathbf{F_a}.$$

Now, from the definition of $E(z_k,\mathbf{a})$, $\mathbf{F}_z$ is seen to be diagonal with diagonal elements equal to $\frac{\partial E(z_i,\mathbf{a})}{\partial z_i}=E^{'}(z_i,\mathbf{a})$. The ijth element of $\mathbf{F}_a$ is given by $\frac{\partial E(z_i,\mathbf{a})}{\partial\mathbf{a}_j}=\cos(jz_i)$. Therefore,

$$\frac{\partial z_k}{\partial\mathbf{a}_j}=-\frac{\cos(jz_k)}{E^{'}(z_k,\mathbf{a})}. \tag{16}$$

Combining (15) and (16) proves the theorem. ■

Theorem 3 is useful only when the zeros of the error function at $\mathbf{a}$ are simple. Otherwise, $\mathbf{D}$ is not defined, and we are not even guaranteed that $\|E(\omega,\mathbf{a})\|_1$ is twice differentiable.

In summary, we have shown that the L_1 norm of the error function is practically differentiable for every choice of filter coefficients and that if its zeros are simple and their number exceeds $M+1$, then $\|E(\omega,\mathbf{a})\|$ is also twice differentiable. We now address the uniqueness of the optimal solution.

B. The Problem of Uniqueness

While the L_2 and the L_∞ minimizers are always unique, uniqueness is not guaranteed in the L_1 case. The uniqueness of the optimal solution is important for two main reasons. The first one is purely theoretical, while the second is of practical nature. Specifically, as we shall see in the next section, uniqueness has paramount influence on the fast convergence of our algorithm. Therefore, our goal in this section is to state conditions under which uniqueness is ensured.

Theorem 4 (Uniqueness): Let $A(\omega)$ be an L_1 optimal frequency response. Then $A(\omega)$ is unique if and only if the corresponding error function changes sign $M+1$ times in Ω.

Proof: Assume first that $E(\omega,\mathbf{a})$ has $M+1$ sign changes, and let us show that $A(\omega)$ is unique. Suppose to the contrary that there exists another optimal solution, $A_1(\omega)$. By [Thm. 2.4, 37], $A_1(\omega)$ must satisfy

$$\begin{aligned} &a)(D(\omega)-A(\omega))(D(\omega)-A_1(\omega))\geq 0 \quad \text{on } \Omega \\ &b)\int_B \operatorname{sign}(D(\omega)-A(\omega))(A_1(\omega)-A(\omega))d\omega = 0. \end{aligned} \tag{17}$$

Condition (b) follows from the the fact that $A_1(\omega)-A(\omega)$ is a degree M cosine polynomial, and that $A(\omega)$ is optimal. From condition (a), we conclude that $(A(\omega)-D(\omega))$ and $(A_1(\omega)-D(\omega))$ must change signs at the same points. So, $A_1(\omega)$ must interpolate $D(\omega)$ at the same points of $A(\omega)$, which means that $A(\omega)-A_1(\omega)$ has $M+1$ zeros, and therefore $A(\omega)=A_1(\omega)$.

Proving that if the optimal solution is unique then $\operatorname{sign}(E(\omega,\mathbf{a}))$ has exactly $M+1$ sign changes in Ω is complicated and long. The proof is found in Appendix III. ■

We now characterize the zeros of the unique solution.

Corollary 2: Let $A(\omega)$ be the unique best weighted L_1 approximation to $D(\omega)$ on Ω. Then each zero of the error function is simple, i.e. it is of multiplicity one.

Proof: From Theorem 4, $E(\omega,\mathbf{a})$ has $M+1$ zeros in Ω. By the Rolle theorem, between each two zeros $E^{'}(\omega,\mathbf{a})$ must cross zero. Excluding the transition band, and noting

$$E^{'}(\omega,\mathbf{a}) = \sum_{k=1}^{M} k\mathbf{a}_k \sin k\omega = A^{'}(\omega), \tag{18}$$

we conclude that $A^{'}(\omega)$ must have at least $M-1$ zeros in Ω.

Now, suppose that one of the zeros of $E(\omega,\mathbf{a})$, say z_1, is not simple, i.e. $E^{'}(z_1,\mathbf{a})=0$. By (35), $A^{'}(z_1)=0$ as well. Thus, $A^{'}(\omega)$ has M zeros ($M-1$ from the extrema of $E(\omega,\mathbf{a})$ and z_1). However, since $A^{'}(\omega)$ is the derivative of a degree M polynomial ($A(\omega)$) it cannot have more than $M-1$ zeros, unless identically zero, a contradiction. ■

Using the differentiability and uniqueness results, in the next section we propose an algorithm for computing the best Nth order L_1 filter for a given Ω.

IV. The Weighted L_1 Algorithm

In 1981 Watson considered the problem of approximating a general continuous function defined over an interval by a finite number of basis functions under the L_1 criterion (unweighted) [36]. Assuming that the L_1 error is differentiable (and sometimes even twice differentiable), Watson suggested the use of the modified Newton method in order to obtain the optimal coefficients of the basis functions. Using the fact that the error function in our case is indeed differentiable and may be also twice differentiable, we generalize Watson's idea to a weighted L_1 error and apply the modified Newton method to our problem. We further exploit the special structure of the problem, and accelerate several steps to improve the computational complexity of the algorithm.

In the following subsection, we first overview several general facts about the modified Newton method, which we shall use in subsequent sections. We then describe our version of the algorithm, and discuss the implementation of each step.

A. The Modified Newton Method

Given a twice continuously differentiable cost function $f(\mathbf{a}) : \mathbb{R}^{M+1} \rightarrow \mathbb{R}$, the modified Newton algorithm generates the sequence

$$\mathbf{a}^{k+1} = \mathbf{a}^k - \gamma^k [\mathbf{H}^k]^{-1} \mathbf{g}^k, \tag{19}$$

where $\mathbf{g}^k$ is the gradient of $f(\mathbf{a})$ at $\mathbf{a}^k$, γ^k is the step size, and $\mathbf{H}^k$ equals either the Hessian matrix when it is positive definite or a modified version of the Hessian, which is positive definite [37]. The fact that $\mathbf{H}^k$ must be positive definite ensures that the value of the cost function decreases at each iteration if γ^k is chosen properly. The step size γ^k should also guarantee a sufficient decrease in the cost function. This requirement can be met if γ^k satisfies the Armijo rule given by

$$f(\mathbf{a}^k - \gamma^k [\mathbf{H}^k]^{-1} \mathbf{g}^k) \leq f(\mathbf{a}^k) - \sigma \gamma^k \mathbf{g}^{k^T} [\mathbf{H}^k]^{-1} \mathbf{g}^k, \tag{20}$$

for some predetermined small σ.

Assuming the condition numbers of $\mathbf{H}^k$ are bounded from above, it is possible to guarantee that the sequence generated by (19) is globally convergent [38]. That is, it converges to a local minimum of $f(\mathbf{a})$ (a point where the gradient is zero), independent of the starting point. If $f(\mathbf{a})$ is convex, then a local minimum is also a global one.

When the function $f(\mathbf{a})$ is not twice differentiable at every point (our filter design problem is an example of such a case), the Hessian matrix may not exist at each iteration. In this case it is common to replace $\mathbf{H}^k$ with the identity matrix [38]. The following proposition states a condition on the matrices $\mathbf{H}^k$ such that the generated sequence is globally convergent [36].

Proposition 3 (Global Convergence): Let there exist positive numbers λ_1 and λ_2 such that

$$\lambda_1 \leq \lambda_{min} \left(\mathbf{H}^k \right) \quad \text{and} \quad \lambda_{\max} \left(\mathbf{H}^{\mathrm{k}} \right) \leq \lambda_2 \tag{21}$$

for all k. Then $\|\mathbf{g}^k\| \rightarrow 0 \;\; as \;\; k \rightarrow \infty$.

The most attractive property of the Newton method is that when the sequence is close to the optimal solution, it converges with a local second order rate of convergence. A sequence $\mathbf{a}^k$ converging to $\mathbf{a}^*$ is said to converge at a second order rate if there exists a constant $r > 0$, such that

$$\|\mathbf{a}^{k+1} - \mathbf{a}^*\| \leq r \|\mathbf{a}^k - \mathbf{a}^*\|^2 \tag{22}$$

in a neighborhood of $\mathbf{a}^*$. Assuming the sequence in (19) converges to an optimum (it will, if global convergence conditions hold), the next proposition specifies the conditions under which a second order rate of convergence is obtained [39].

Proposition 4 (Local Convergence): Suppose that $f(\mathbf{a})$ is twice differentiable and that the Hessian $\nabla^2 f(\mathbf{a})$ is Lipschitz continuous in a neighborhood of an optimum point $\mathbf{a}^*$. Suppose that there exists a $k_0 \geq 1$ such that in (19) $\gamma^k = 1$ for all $k \geq k_0$. Then the rate of convergence of $\{\mathbf{a}^k\}$ is of second order.

Note that Proposition 4 assumes the step size is equal to 1 in a neighborhood of the solution. The following proposition shows that when σ in (20) is chosen properly there exists a neighborhood of the optimal solution such that $\gamma^k = 1$ [36].

Proposition 5: If $\sigma < 0.5$ in (20), and $\nabla^2 f(\mathbf{a}^*)$ is positive definite, then there exists a $k_0 \geq 1$ such that for all $k \geq k_0$, $\gamma^k = 1$ in (19).

We are now ready to introduce our method for calculating the L_1 optimal filter coefficients. In Section V we make use of Propositions 3 and 4 to prove that the proposed method is globally convergent and that when the solution is unique it has a second order rate of convergence.

B. The L_1 Algorithm

Table I summarizes the six steps required for computing the coefficients of an N-order (N is even) linear phase FIR filter, which is closest to $D(\omega)$ in the weighted L_1 sense with a positive weighting function, $W(\omega)$. In what follows, we give a detailed description of each of the steps.

Given $\omega_p,\ \omega_s,\ N$ and $W(\omega)$.
Step 1 - Initialization. Set $M = \frac{N}{2}$and determine an initial vector $\mathbf{a}^1 \in \mathbb{R}^{M+1}$,$\epsilon > 0$, $0 < \sigma < 1/2$, $0 < \beta < 1$, $0 < \delta_1$, $0 < \delta_2$, and $0 < \mu$. Set $k = 1$.
Step 2 - Positive-definite matrix determination. Calculate the simple zeros of the error function at the kth $\{z_1, \ldots, z_t\}$. If $t > 0$ then form the matrix $\mathbf{D}^k$, as defined in Theorem 3. Set d_{min} and d_{max} to be the minimal and maximal elements of $\mathbf{D}^k$. Determine a positive definite $(M+1) \times (M+1)$ matrix $\mathbf{H}^k$ according to one of the following cases:
I. If $t = 0$ or $d_{min} < \delta_1$ or $\delta_2 < d_{max}$, then set $\mathbf{H}^k = \mathbf{I}$.
Otherwise, form the matrix $\mathbf{V}^k$ as defined in Theorem 3.
II. If $t = M+1$, $\delta_1 \leq d_{min}$, $d_{max} \leq \delta_2$ and $\mu < \min_{i,j,i\neq j} |\cos(z_i) - \cos(z_j)|$ then set $\mathbf{H}^k = {\mathbf{V}^k}^T \mathbf{D}^k \mathbf{V}^k$.
III. If $0 < t < M+1$, $\delta_1 \leq d_{min}$, $d_{max} \leq \delta_2$, or $\min_{i,j,i\neq j} |\cos(z_i) - \cos(z_j)| \leq \mu$, then apply the modified Cholesky decomposition to ${\mathbf{V}^k}^T \mathbf{D}^k \mathbf{V}^k$ [37]. Set $\mathbf{H}^k = {\mathbf{V}^k}^T \mathbf{D}^k \mathbf{V}^k + \mathbf{C}^k$.
Step 3 - Descent Direction. Compute the $(M+1)$-dimensional vector $\mathbf{g}^k$ whose nth component is given by (12). Determine $\mathbf{d}^k$, the current descent direction, as the unique solution of

$$\mathbf{H}^k \mathbf{d}^k = -\mathbf{g}^k. \tag{23}$$

Step 4 - Stopping Criterion. If $|(\mathbf{d}^k)^T \mathbf{g}^k| < \epsilon$, then stop.
Step 5 - Step Size. Determine the step size γ^k to be $\max\{1, \beta, \beta^2, ...\}$ such that

$$\frac{\|E(\omega, \mathbf{a}^k + \gamma^k \mathbf{d}^k)\|_1 - \|E(\omega, \mathbf{a}^k)\|_1}{\gamma^k (\mathbf{d}^k)^T \mathbf{g}^k} \geq \sigma \tag{24}$$

Step 6 - Updating. Set $\mathbf{a}^{k+1} = \mathbf{a}^k + \gamma^k \mathbf{d}^k$, $k = k+1$, and go to Step 2.

TABLE I
THE L_1 ALGORITHM.

C. Algorithm Description

Exploiting the special structure of the our problem we shall now explain how to implement each stage of the above algorithm.

Initialization. In this first step we choose an initial guess of the optimal solution and set the relevant constants. Specifically, ϵ determines the accuracy of the stopping condition, while σ and β are related to the step size selection

of the fifth stage as will be explained below. Typical values are $\epsilon = 10^{-6}$, $\sigma = 10^{-3}$, $\beta = 0.5$, $\delta_1 = 10^{-15}$, $\delta_2 = 10^{15}$ and $\mu = 10^{-10}$.

A good starting vector can save a large number of iterations, and therefore time. In our problem, we impose another constraint on the initial guess: we would like it to be such that our algorithm passes only through vectors at which the derivative exists (allowing the use of the gradient in (19)). According to Proposition 1, the only two problematic vectors are $(0, 0, ..., 0)$ and $(1, 0, ..., 0)$. Since the value of $\|E(\omega, \mathbf{a})\|_1$ is decreased in each iteration, avoiding non-differentiable points can be accomplished if the initial vector $\mathbf{a}^1$ satisfies,

$$\|E(\omega, \mathbf{a}^1)\|_1 < \min\left(\|E(\omega, (0, 0, ..., 0))\|_1,\ \|E(\omega, (1, 0, ..., 0))\|_1\right). \tag{25}$$

As a good initial guess, we choose $\mathbf{a}^1$ such that the corresponding $A(\omega)$ interpolates the desired response $D(\omega)$ at the points

$$z_i = \frac{(2i-1)\pi}{2(M+1)}, \quad i = 1, ..., M+1; \tag{26}$$

see [35]. Thus, $\mathbf{a}^1$ is the solution of the linear system

$$\begin{pmatrix} 1 & \cos(z_1) & \dots & \cos(Mz_1) \\ 1 & \cos(z_2) & \dots & \cos(Mz_2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cos(z_{M+1}) & \dots & \cos(Mz_{M+1}) \end{pmatrix} \mathbf{a}^1 = \begin{pmatrix} D(z_1) \\ D(z_2) \\ \vdots \\ D(z_{M+1}) \end{pmatrix}. \tag{27}$$

The matrix in this system is always invertible, since the functions $\{1, \cos(\omega), ..., \cos(M\omega)\}$ form a Chebyshev set on $[0, \pi]$ [35]. Simulations show that this is an excellent choice, since it is usually very close to the optimal solution and also satisfies (25). If some of the points in (26) happen to lie in the transition band (ω_p, ω_s), some intermediate values between zero and one are chosen as the value of the desired response $D(\omega)$.

Another possible initial vector is the solution of the weighted least-squares approximation given by $\mathbf{A}_{LS}\mathbf{a}^1 = \mathbf{b}$ where $\mathbf{A}_{LS}$ and $\mathbf{b}$ are described in [11]. This guess has also proven to be a good choice that satisfies (25), as observed in simulations, however the vector corresponding to (27) often performs better.

Matrix Determination This step determines the matrix $\mathbf{H}^k$ in (19), which can assume one of three forms, depending on the number of zeros of $E(\omega, \mathbf{a}^k)$. Thus, this stage has to start with the calculation of the zeros of the error function.

The brute force computation is essentially an exhaustive search over a grid, similar to the way the extrema are computed in the Remez exchange algorithm [15]. However, since the error function is a cosine polynomial in each band, $[0, \omega_p]$ and $[\omega_s, \pi]$, both efficiency and accuracy may be improved by using a polynomial root finding algorithm. There are several efficient methods for computing the zeros of an algebraic polynomial, e.g. [40]. Fortunately, we can convert the problem of finding the zeros of a trigonometric polynomial in each band into an algebraic one. To see this consider, for example, $E(\omega, \mathbf{a})$ on $[\omega_s, \pi]$, which is given by

$$E(\omega, \mathbf{a}) = \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega).$$

Replacing each $\cos(n\omega)$ with $\frac{z^n + z^{-n}}{2}$, where $z = e^{j\omega}$, we obtain the complex valued polynomial in z

$$E(z) = z^{-M} \sum_{n=0}^{2M} \mathbf{c}_n z^n,$$

where

$$\mathbf{c}_n = \begin{cases} \frac{\mathbf{a}_{M-n}}{2}, & n = 0, ..., M-1, \\ \mathbf{a}_0, & n = M, \\ \frac{\mathbf{a}_{n-M}}{2} & n = M+1, ..., 2M. \end{cases} \tag{28}$$

We now apply any of the efficient algorithms to find the zeros of $E(z)$ [40]. To extract the corresponding zeros of $E(\omega, \mathbf{a})$ from those of $E(z)$, we compute the angles of the zeros of $E(z)$, which are located on the unit circle. The total number of zeros in the passband and stopband is denoted by t. A zero of $E(\omega, \mathbf{a}^k)$ z_i is simple if $E^{'}(z_i, \mathbf{a}^k) \neq 0$. This is easily checked since $E^{'}(\omega, \mathbf{a}^k)$ is given by (35).

Next, we construct the matrix $\mathbf{H}^k$ according to the following rule. If there are no zeros ($t = 0$) or some of them are not simple, then $\mathbf{H}^k$ equals the identity matrix. If all the zeros are simple, then according to Theorem 3 the Hessian matrix is defined and given by (14). This theorem also provides us with a very convenient and efficient way of checking the positive definiteness of the Hessian without computing any eigenvalues, for example. Specifically, we show in Appendix II, that the positive definiteness of $\mathbf{H}^k$ is directly related to that of $\mathbf{D}^k$ (which is easily checked since it is diagonal) and to the non-singularity of $\mathbf{V}^k$. Keeping $\mathbf{D}^k$ positive definite is achieved by bounding its elements from below, i.e. $d_m in$ has to be larger than a threshold (δ_1). Checking the singularity of $\mathbf{V}^k$ (in case it is indeed square) is done by checking the minimal distance (μ) of the zeros of the error function at the kth iteration as further explained in Appendix II, where the numbers δ_1 and δ_2 are also seen to guarantee global convergence.

If, however the number of zeros (which are all assumed to be simple) is less than $M + 1$, then the Hessian matrix is only positive semidefinite, in which case we add to it a diagonal matrix to form $\mathbf{H}^k$. The choice of the diagonal matrix should be such that the new matrix will be positive definite and well conditioned. A good choice was suggested in [37], using the Cholesky decomposition.

Descent Direction The third step is the most time consuming step, since it involves solving a linear system of equations defined in (23) for obtaining the descent direction. Usually, solving a linear system of equations with $M + 1$ unknowns (the length of $\mathbf{d}^k$) requires $O(M^3)$ operations. In our case, however, the special structure of the matrix $\mathbf{H}^k$ in (23) can be exploited to reduce the computational complexity.

If $\mathbf{H}^k = \mathbf{I}$ then the solution is straightforward and requires no computation. When $\mathbf{H}^k$ equals the Hessian matrix, (23) can be solved in $O(M^2)$ steps. Specifically, in this case (23) becomes

$$\mathbf{V}^{k^T}\mathbf{D}^k\mathbf{V}^k\mathbf{d}^k = -\mathbf{g}^k. \tag{29}$$

Denoting $\mathbf{f}^k = \mathbf{D}^k\mathbf{V}^k\mathbf{d}^k$ we first solve the equation,

$$\mathbf{V}^{k^T}\mathbf{f}^k = -\mathbf{g}^k, \tag{30}$$

and then solve

$$\mathbf{V}^k\mathbf{d}^k = \mathbf{D}^{k^{-1}}\mathbf{f}^k. \tag{31}$$

Each system can be solved in $O(M^2)$ operations, as explained in detail in [41].

Finally, if the Hessian matrix is only positive semidefinite, then using its Cholesky decomposition from the previous step, we can solve (23) in $O(M^2)$ operations. Of course, there is no real gain in terms of complexity in this case, since in order to perform the Choleksy decomposition, an $O(M^3)$ number of operations are required.

If the case where the Hessian matrix is positive definite is frequent enough, then we can gain a lot in terms of

complexity using the above method for solving (23) in $O(M^2)$ operations. In the next section, we shall see how frequent this case it.

Stopping Criterion The algorithm is stopped if $|(\mathbf{d}^k)^T\mathbf{g}^k|$ which equals to the directional derivative in the direction $\mathbf{d}^k$, is less than a predetermined threshold, ϵ. In this case, the gradient is close enough to zero.

Step Size Selection In the fifth step the size of the progress along $\mathbf{d}^k$ is determined. This step size, γ^k, is determined according to the Armijo rule in (20), such that a sufficient decrease of $\|E(\omega,\mathbf{a})\|_1$ is guaranteed.

V. Convergence Issues

We now show that the algorithm of Table I is globally convergent, and state conditions which guarantee a second order rate of local convergence.

Theorem 5: The algorithm in Table I is globally convergent.

Proof: See Appendix II. ■

We now address the local convergence of the proposed method.

Theorem 6: If $\mathbf{a}^*$ is the unique minimizer of (11), then the algorithm in Table I converges at a second order rate.

Proof: In order to prove that the rate of convergence is second order, we shall show that the conditions of Proposition 4 hold when the solution is unique. By Proposition 5 showing that $\gamma^k = 1$ for $k \geq k_0$ requires that $\nabla^2\|E(\omega,\mathbf{a}^*)\|_1$ is positive definite, so let us first prove that this is the case when the solution is unique.

From Theorem 4, if $\mathbf{a}^*$ is the unique minimizer then the error changes sign $M+1$ times in Ω, i.e. it has $M+1$ simple zeros in Ω. Thus, according to Theorem 3 the Hessian matrix at $\mathbf{a}^*$, $\nabla^2\|E(\omega,\mathbf{a}^*)\|_1$, exists, its rank is $M+1$, and is therefore positive definite. Since the Hessian is continuous in $\mathbf{a}$ (it is seen to be continuous in z_i which are continuous in $\mathbf{a}$), there is a neighborhood of $\mathbf{a}^*$, $B(\mathbf{a}^*,\epsilon)$, for which the Hessian matrix exists and is positive definite. Thus, we conclude that Proposition 5 is valid and that the first condition of Proposition 4 is satisfied. Proving that $\nabla^2\|E(\omega,\mathbf{a})\|_1$ is Lipschitz continuous in the vicinity of $\mathbf{a}^*$ is mathematically challenging. We defer the proof to Appendix V. ■

In summary, our discussion suggests that for the filter design problem, the algorithm will converge to the optimal solution from any starting point. Furthermore, when the optimal solution is unique convergence will be very fast if the initial starting iterate is close enough to the optimal solution.

We now compare our algorithm with the Remez exchange method for designing equiripple filters.

VI. Comparison with the Remez Algorithm

It is interesting to note that our method shares several features with the celebrated Remez exchange algorithm for designing minimax filters, and therefore may be viewed as its dual in the following sense. First, the Remez procedure is an ascent algorithm, while ours is a descent one. Second, the role of the extrema in Remez is played by the zeros in the L_1 case, that is in each iteration the set of zeros is replaced by a "better" set in such a way that the error is decreased, whereas in the minimax case the set of extrema is exchanged by a different set to increase the error. In terms of convergence, both methods are globally convergent, and under the assumption that the L_1 solution is unique (which is always the case in the minimax problem), both methods enjoy second order rate of convergence [42]. Simulation results show that in most cases, the L_1 solution is indeed unique.

The following figures compare the convergence properties of our algorithm and those of Remez. Figure 3 shows a graph of the number of iterations as a function of the filter order for our algorithm. For each filter order ten different filters (five low-pass and five high pass with different transition bands and different weighting functions)

were implemented and the number of iterations on the graph is taken to be the average of these ten. From this graph it can be deduced that on the average 16 iterations are required to compute the coefficients of order 20 to 200 filters. Figure 4 shows a similar graph for the Remez exchange method. The Remez algorithm was implemented as

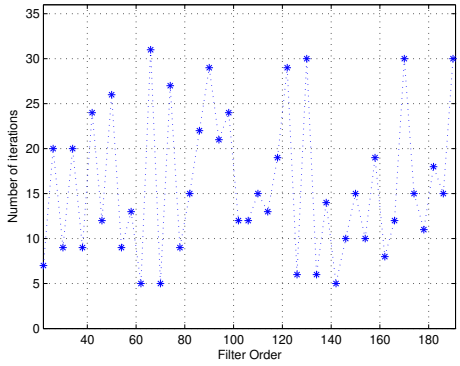


Fig. 3. The average number of iterations versus the filter order using the proposed algorithm.

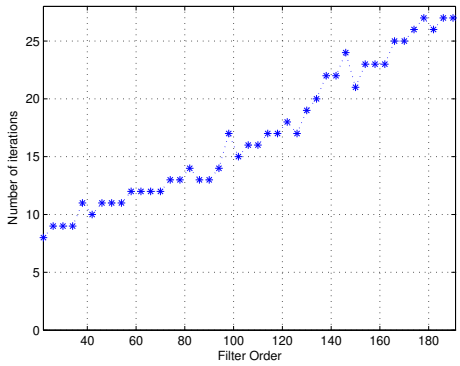


Fig. 4. The average number of iterations versus the filter order using the Remez algorithm.

it was given in [15], i.e. without any improvements. The degree of accuracy in both methods was taken to be the same. It is seen that the average number of iterations for filters designed using the Remez algorithm is 22. Thus, on the average our method operates slightly better than the Remez one although they are comparable.

The computational complexity of both methods is also comparable. The most time consuming step in the Remez as well as in the proposed algorithm is the solution of a system of linear equations, which usually has a complexity $O(M^3)$. Parks and McClellan showed that for the filter design problem this step requires only $O(M^2)$ flops, using the Barycentric Lagrange interpolation [15]. As shown earlier, the linear system in our case may also be computed in an $O(M^2)$ flops when the Hessian matrix is positive definite. Figure 5 shows the ratio of the number of times the Hessian matrix was positive definite to the number of iterations as a function of M for filters with different cutoff frequencies. The average ratio is 0.53, which implies that we have reduced the complexity by half.

VII. Design Examples

In this section, the properties of L_1 filters are demonstrated and discussed. Specifically, we are interested in understanding how L_1 filters handle the tradeoff between flatness in the passbands and stopbands and the width of the transition band. We present three different types of filters: low-pass, high-pass and band-pass. We begin with the low-pass filter introduced in Section II.

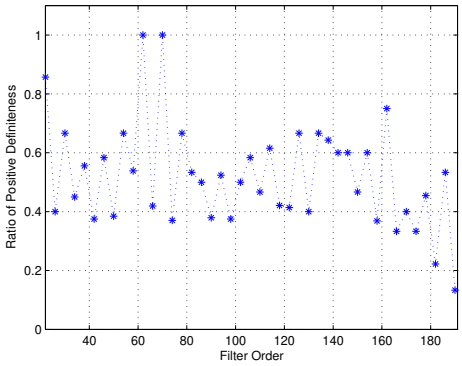


Fig. 5. The ratio of the time an $O(M^2)$ algorithm was used.

Example 1 - Low-pass filter. We consider the low-pass filter described in (5) with $\omega_c = 0.485\pi$ rad. The first filter we design has the intervals $[0, 0.474\pi]$ and $[0.493\pi, \pi]$ as its passband and stopband, respectively, using $W(\omega) = 1$ in both bands. The length of the filter is designed to be 65. Figure 6 shows the frequency responses of several methods used to design this low-pass: the Kaiser windowing method, least-squares, minimax and the maximally-flat approach. The corresponding L_1 filter is shown in Figure 7. The figures suggest that the L_1 filter results in the flattest response among the L_2, minimax and windowing methods. At the same time it has a much narrower transition band than the maximally flat filter. Figure 8 shows the transition band of the L_1 filter (the solid line) and that of the least-squares, from which it is seen that the L_2 filter performs slightly better than the L_1 method, however, the improvement is minor. Figure 9 shows the passband response of the L_1, the least-squares, the minimax filters. It is obvious that the L_1 admits the flattest response.

Example 2 - Wideband high-pass filter with weights. In this example, we design a wideband high-pass with cutoff frequency equal to $w_c = 0.11\pi$ rad. The stopband is taken to be $[0, 0.1\pi]$ and the passband is $[0.12\pi, \pi]$. In order to ensure a good attenuation in the stopband, we consider the following weighting function

$$W(\omega) = \begin{cases} 20, & \omega \in [0, 0.1\pi], \\ 1, & \omega \in [0.12\pi, \pi]. \end{cases} \tag{32}$$

Since a weighting function can be incorporated only in the L_1, L_2 and L_∞ design methods, only these are considered in this example. Figure 10 shows the resulting frequency responses. The figure shows again the flatness property of the L_1 filters, and this time it is even more visible in the passband. Furthermore, we see that the incorporation of the weighting function in the stopband results in the highest attenuation in the L_1 filter.

Example 3 - Bandpass Filter Although it was mentioned in Section II that we consider only low-pass and high-pass filters, in this example we apply our method to design a bandpass filter. In fact, as we discuss in Section VIII, the results regarding the characterization and uniqueness of the filter may no longer be valid for the multiband case. Nevertheless, using the same algorithm of Table I to compute the optimal coefficients, we can gain more insight into the meaning of L_1 filters. The bandpass filter is designed according to the following specifications: its order is $N = 50$, the passband is $[0.35\pi, 0.45\pi]$ and the stopband is $[0, 0.3\pi] \cup [0.5\pi, \pi]$. The weighting function equals one over the passband and stopband. Figure 11 shows the L_1, L_2 and L_∞ responses.

The figure demonstrates that the transition bands of all filters is comparable, and that the L_1 bandpass offers the strongest attenuation in the stopband. In addition, the running time of the bandpass design was no longer than that of the other examples (specifically, it required 9 iterations).

In summary, from the above examples it can be deduced that L_1 filters have flatter response than the least-squares

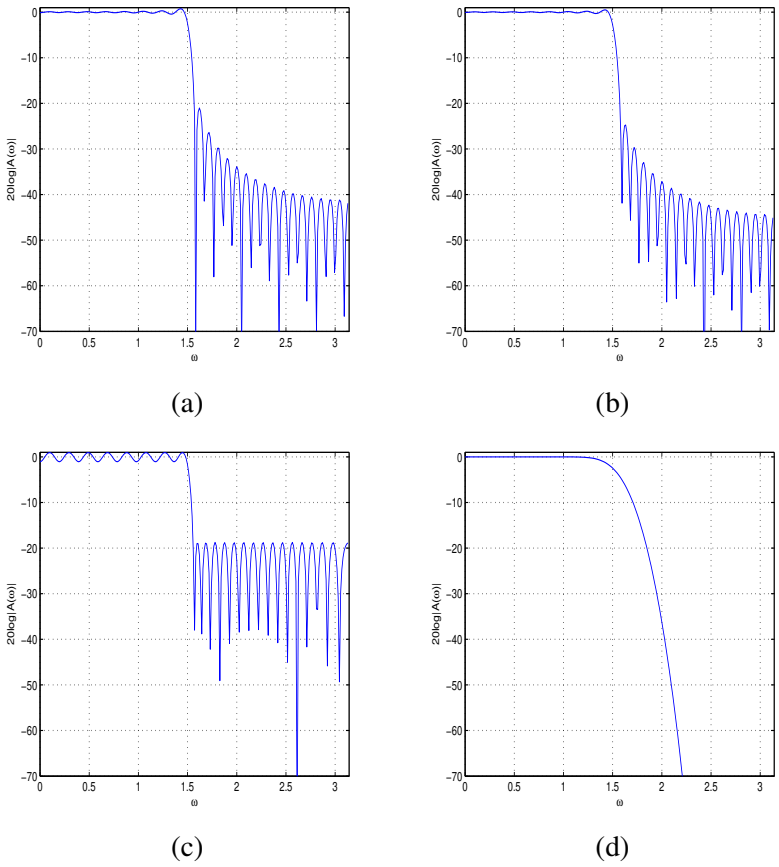


Fig. 6. Low-pass filter of length 65 (a) Kaiser (b) Least-squares (c) Minimax (d) Maximally-flat.

(and than the minimax, of course), but their transition band remains comparable. The maximally flat filters possess the flattest response, but their transition width is unacceptably wide. Thus, L_1 filters provide a suitable tradeoff between flatness and the transition bandwidth. In application where flatness is of utmost importance together with a reasonable transition bandwidth, for example in antialiasing filters for multirate systems, the L_1 solution may be the best choice.

VIII. Discussion

This paper addressed the problem of designing FIR linear-phase digital filters, which are optimal in the L_1 sense. Most common filter design methods rely on either the minimax or the least-squares criteria, while the L_1 measure was ignored. In this respect, this paper fills in the gap by describing both the theory and the algorithmic aspects of L_1 filter design.

Optimal L_1 filters suggest a very flat response in the passband and stopband while retaining a transition band, which is comparable to the least-squares filters. Thus, the L_1 filters constitute a better tradeoff between the minimax and the maximally flat filters than the least-squares approach.

The mathematical theory of L_1 filters was developed in this paper, where it was shown that the error function is differentiable and a formula for the Hessian matrix was derived. An explicit condition for uniqueness was stated, which is reminiscent of the famous alternation theorem for the minimax case. Following the mathematical results, a modified Newton algorithm was proposed to calculate the optimal L_1 filter. The special structure of the problem enabled us to further accelerate the running time of computationally demanding steps. It was also shown that the

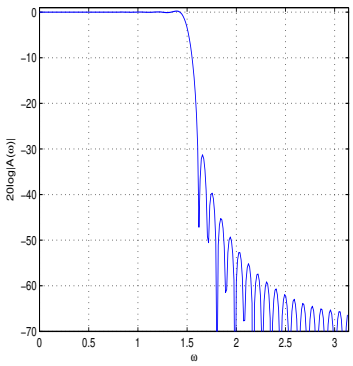


Fig. 7. The 65 length L_1 filter.

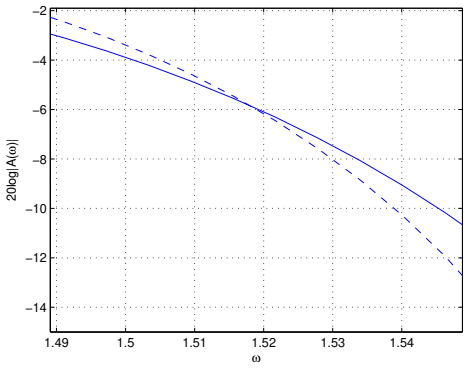


Fig. 8. The transition bands of the L_1 (solid) and L_2 (dashed).

algorithm is globally convergent and that under the uniqueness assumption it has a second order rate of convergence. In practice, uniqueness usually holds, and even when it does not, fast convergence was observed. We also compared our method with the celebrated Remez exchange algorithm, and it was shown that on the average our technique requires a smaller number of iterations. The computational complexity of the corresponding steps in each algorithm was shown to be comparable.

In this paper we considered only the approximation of low-pass filters. High-pass filters are treated in the exact same manner, that is all the theorems proved for the low-pass case are valid for the high-pass filter as well. However, several results do not hold in the multiband case. In particular, the characterization Theorem 2 and the uniqueness condition change when considering multiband filters. Nevertheless, as we showed, the algorithm of Table I can still be applied to compute the coefficient of a multiband filter, but different conditions on its convergence rate will be imposed.

Characterization, differentiability and uniqueness of type-II, III and IV filters is also an interesting problem.

Finally, nonlinear phase filters are also important in many signal processing applications. The approximation problem in this case become a complex one. Many papers dealing with nonlinear phase minimax filters have been published, and it would be very interesting to conisder their L_1 analogues.

We hope that this paper will encourage further research of the L_1 criterion in the field of filter design.

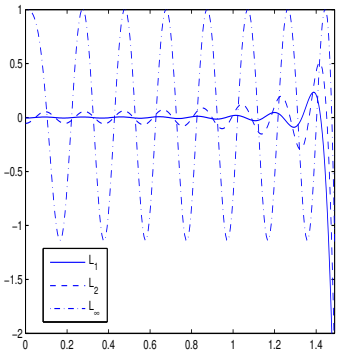
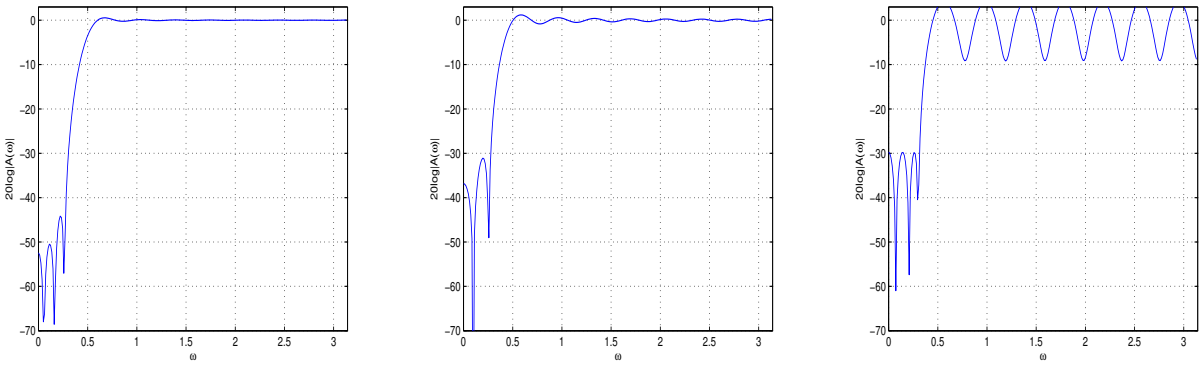


Fig. 9. An enlarged part of the passband of the 65 length L_1, L_2 and L_∞ filters.



(a) (b) (c)

Fig. 10. High-pass filter of length 35 (a) L_1 (b) Least-squares (c) Minimax.

APPENDIX I
THE OPTIMAL $N = 0$ FILTER

In this appendix we show how to find the best constant filter, corresponding to a length-one filter. Specifically, we wish to find a such that

$$\|E(\omega, a)\|_1 = \int_0^\pi W(\omega)|a - D(\omega)|d\omega = \int_0^{\omega_p} W(\omega)|a-1|d\omega + \int_{\omega_s}^\pi W(\omega)|a|d\omega. \tag{33}$$

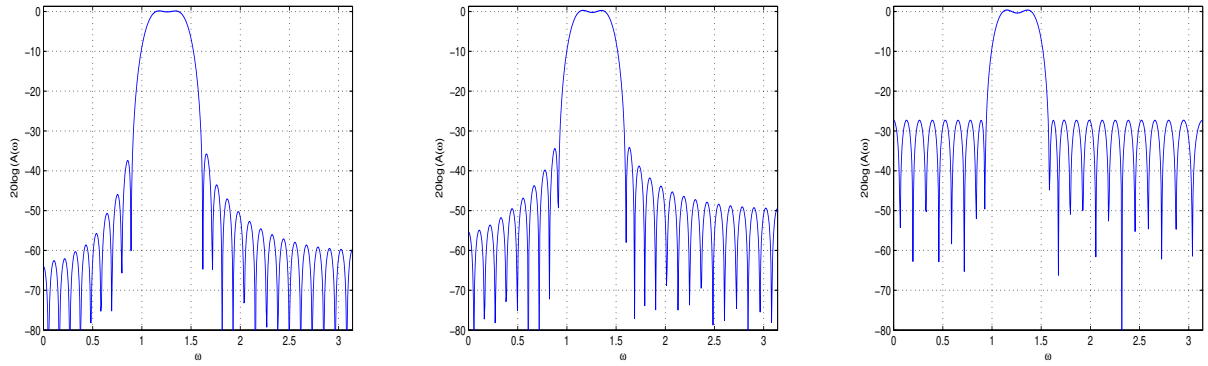
is minimized. Defining $A = \int_0^{\omega_p} W(\omega)d\omega$ and $B = \int_{\omega_s}^\pi W(\omega)d\omega$, $A > 0, B > 0$, we have

$$\|E(\omega, a)\|_1 = \begin{cases} (A+B)a - A, & a \geq 1, \\ (B-A)a + A, & 0 < a < 1, \\ -(A+B)a + A & a \leq 0. \end{cases} \tag{34}$$

Clearly, the optimal solution a^* depends on A and B. If $A > B$ then $a^* = 1$, whereas if $A < B$, $a^* = 0$. When $A = B$ every $0 < a^* < 1$ is an optimal solution. In this case, we also see that the solution is not unique.

APPENDIX II
PROOF OF THEOREM 2

The proof of the theorem relies on the following lemma.



(a) (b) (c)

Fig. 11. Bandpass filter of length 35 (a) L_1 (b) Least-squares (c) Minimax.

Lemma 1: For every $\mathbf{a} \in \mathbb{R}^{M+1}$, $(M \geq 1)$, $E(\omega, \mathbf{a})$ has no more than $M+1$ sign changes in Ω.

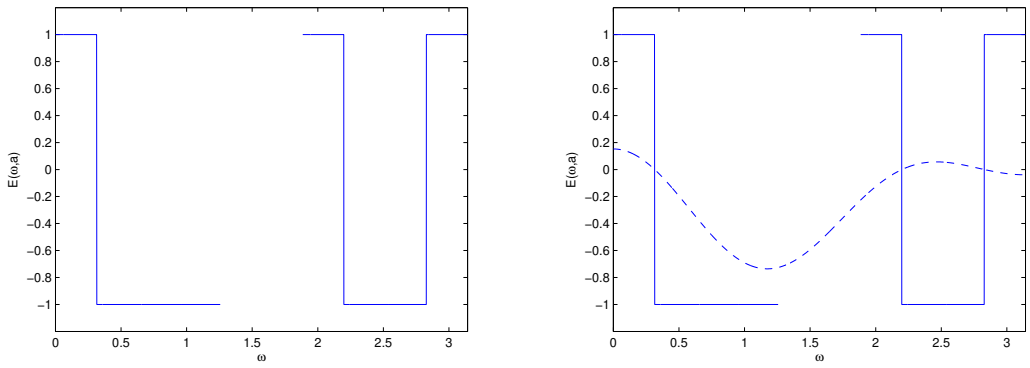
Proof: Suppose that $E(\omega, \mathbf{a})$ has more than $M+1$ sign changes in Ω. We know that each sign change is also a zero of $E(\omega, \mathbf{a})$, and therefore $E(\omega, \mathbf{a})$ has at least $M+1$ zeros in Ω. By the Rolle theorem the derivative of $E(\omega, \mathbf{a})$, $E^{'}(\omega, \mathbf{a})$, has at least $M-1$ zeros in Ω. Now, note that

$$E^{'}(\omega, \mathbf{a}) = \sum_{k=1}^{M} k\mathbf{a}_k \sin k\omega = A^{'}(\omega), \tag{35}$$

from which we conclude that $A^{'}(\omega)$ has at least $M-1$ zeros in Ω. However, since $A^{'}(\omega)$ is the derivative of an M degree polynomial it cannot have more than $M-1$ zeros unless it is identically zero, which proves that $E(\omega, \mathbf{a})$ cannot have more than $M+1$ sign changes. ■

We now prove the theorem.

Proof: Suppose first that the error function has $L < M$ sign changes. We claim that the resulting filter cannot be optimal. We prove our claim by constructing a cosine polynomial of degree M (at most), which violates the optimality condition of Proposition 2. To illustrate the idea, suppose $\omega_p = 0.4\pi$, $\omega_s = 0.6\pi$, $M = 4$ and that the number of sign changes is $L = 3$. A typical sign function is depicted in Fig. 12(a) (the transition band is not depicted), from which we see that the error changes sign three times in Ω. Specifically, it changes sign one time



(a) (b)

Fig. 12. (a) Sign error for $M = 4$ with three sign changes in Ω. (b) Sign error for $M = 4$ with three sign changes in Ω (in solid), and $T(\omega)$ (dashed).

in $[0,\omega_p]$, and two times in $[\omega_s,\pi]$. Denote the points where it changes sign by $\{z_1,z_2,z_3\}$. Now, we construct the cosine polynomial of degree 4,

$$T(\omega)=(\cos(\omega)-\cos(z_1))(\cos(\omega)-\cos(z_2))(\cos(\omega)-\cos(z_t))(\cos(\omega)-\cos(z_3)), \tag{36}$$

where z_t is any number larger than ω_p and smaller than ω_s. The polynomial $T(\omega)$ is depicted in dashed line together with $\mathrm{sign}(E(\omega,\mathbf{a}))$ for $z_t=0.5\pi$ in Fig. 12(b). We see that whenever $\mathrm{sign}(E(\omega,\mathbf{a}))$ is positive, so is $T(\omega)$, and whenever $\mathrm{sign}(E(\omega,\mathbf{a}))$ is negative, so is $T(\omega)$. As a result,

$$\langle T(\omega),\mathrm{sign}(E(\omega,\mathbf{a}))\rangle=\int_0^\pi W(\omega)\mathrm{sign}(E(\omega,\mathbf{a}))T(\omega)\neq 0.$$

Therefore $\mathrm{sign}(E(\omega,\mathbf{a}))$ violates the optimality condition for $A(\omega)$ in Proposition 2.

For general values of M and $L<M$, the proof proceeds in the same manner. Note that the construction of $T(\omega)$ was possible by the assumption that $L<M$, otherwise $T(\omega)$ would be of degree greater than M. Thus, we have established that for $A(\omega)$ to be optimal, the corresponding error function, $E(\omega,\mathbf{a})$, must change sign at least M times in Ω. The fact that $E(\omega,\mathbf{a})$ cannot have more than $M+1$ sign changes follows directly from Lemma 1. ■

APPENDIX III
PROOF OF THEOREM 4

To complete the proof of Theorem 4, we first prove a useful lemma.

Lemma 2: If the optimal error function has M sign changes, then there exist ω_1,ω_2, such that $0\leq\omega_1<\omega_p$, $\omega_s<\omega_2\leq\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in(\omega_s,\omega_2]$.

This lemma shows that if the optimal solution has M sign changes then the sign of the error function cannot be arbitrary. Specifically, it is -1 at the end of the passband and 1 at the beginning of the stopband. Figure 13 shows an example of the four possible shapes of $\mathrm{sign}(E(\omega,\mathbf{a}))$ can take on for $M=4$. The empty part in the middle of each figure corresponds to the the transition band, which is excluded. The lemma states that Fig. 13(d) is the only possible choice of $\mathrm{sign}(E(\omega,\mathbf{a}))$ for the corresponding $A(\omega)$ to be optimal.

Proof: In order to prove the lemma we shall eliminate the other three possibilities.

1. Suppose that there exist ω_1,ω_2, such that $0\leq\omega_1<\omega_p$, $\omega_s<\omega_2\leq\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in(\omega_s,\omega_2]$. This situation is depicted in Fig. 13 (a). We construct a cosine polynomial of degree M, $T(\omega)$, that contradicts the fact the $A(\omega)$ is optimal. The polynomial has zeros at the points where $E(\omega,\mathbf{a})$ changes sign, and therefore $\langle T(\omega),\mathrm{sign}(E(\omega,\mathbf{a}))\rangle\neq 0$. Figure 14(a) shows $T(\omega)$ in dashed together with possibility (a) for $M=4$.

2. Suppose that there exist ω_1,ω_2, such that $0\leq\omega_1<\omega_p$, $\omega_s<\omega_2\leq\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in(\omega_s,\omega_2]$. This situation is depicted in Fig. 13(b). We treat this case in the exact same manner as the first one. The corresponding $T(\omega)$ and the sign function are depicted in Fig. 14(b).

3. Suppose that there exist ω_1,ω_2, such that $0\leq\omega_1<\omega_p$, $\omega_s<\omega_2\leq\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in(\omega_s,\omega_2]$. This situation is depicted in Fig. 13 (c). We claim that this possibility contradicts the fact that the optimal solution, $A(\omega)$, is of degree M. In particular, since $\mathrm{sign}(E(\omega,\mathbf{a}))$ has M sign changes in Ω, $A^{'}(\omega)$ must have $M-1$ zeros there. However, since $A(\omega)\geq 1$ for $\omega\in[\omega_1,\omega_p)$ and $A(\omega)\leq 0$ for $\omega\in(\omega_s,\omega_2]$, then $A(\omega)$ must cross zero in $[\omega_p,\omega_s]$ adding one more zero to $A^{'}(\omega)$. Totally $A^{'}(\omega)$ has M zeros, which means that $A(\omega)$ is identically zero. A 4th degree polynomial $A(\omega)$ having a sign function as in Fig. 13(c) is shown in dashed in Fig. 15, from which it is clear that the degree of $A(\omega)$ must be higher than 4.

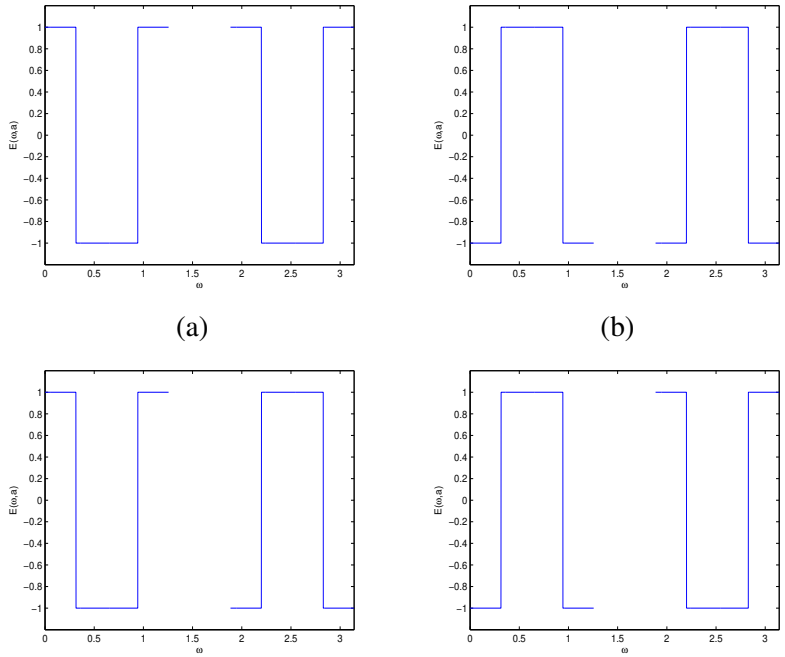


Fig. 13. Four possible shapes for the sign function of the optimal error which changes sign $M = 4$ times.

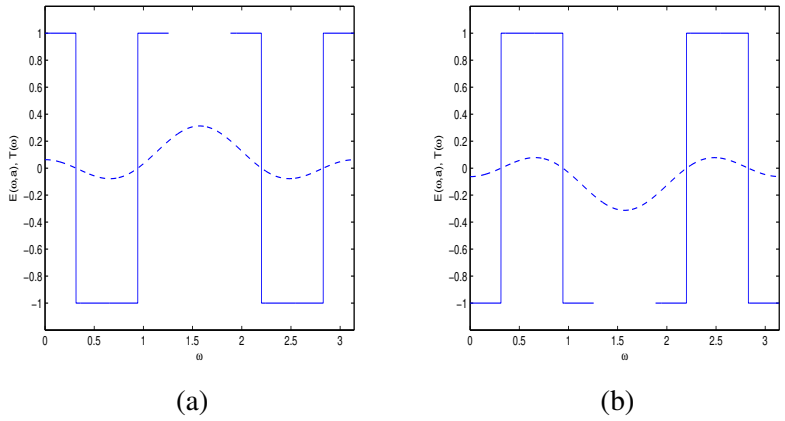


Fig. 14. $T(\omega)$ (dashed) and $\text{sign}(E(\omega, \mathbf{a}))$ for $M = 4$ corresponding to (a) Fig. 13(a), (b) Fig. 13(b).

Now, the last option remaining is the one described in the lemma. If there are no sign changes in $[0, \omega_p)$ then $\omega_1 = 0$, whereas if there are no sign changes in $(\omega_s, \pi]$ then $\omega_2 = \pi$. Otherwise ω_1 is taken to be the last sign change in $[0, \omega_p)$ and ω_2 to be the first sign change in $(\omega_s, \pi]$.

■

We now prove that if $A(\omega)$ is the unique optimal weighted L_1 approximation to $D(\omega)$ then the number of sign changes of the corresponding error function, $E(\omega, \mathbf{a})$, is $M + 1$. To this end, we shall assume that the number of sign changes is M (it was already proved that for $A(\omega)$ to be optimal this is the minimum number of sign changes possible), and construct another optimal solution, denoted by $A_1(\omega)$, with coefficients $\mathbf{a}_1$. Specifically, we derive $A_1(\omega)$ from $A(\omega)$ in the following way. We construct an M degree cosine polynomial $T(\omega)$ with zeros at the points where $E(\omega, \mathbf{a})$ changes sign. The function $A_1(\omega)$ will be equal to the sum of $A(\omega)$ and a positive

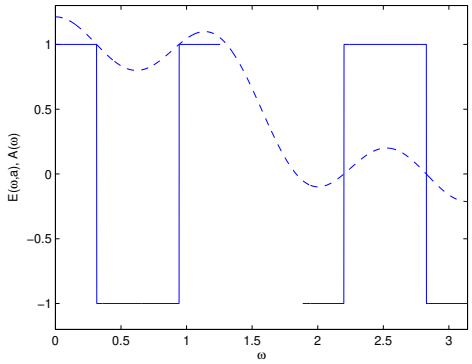


Fig. 15. A polynomial $A(\omega)$ with error sign function as in Fig. 13(c).

multiple of $T(\omega)$, and is therefore an M degree polynomial. If the coefficient multiplying $T(\omega)$ is chosen to be small enough, then the sign function of $E(\omega, \mathbf{a}_1)$ and $E(\omega, \mathbf{a})$ will be the same, implying that both $A_1(\omega)$ and $A(\omega)$ are optimal, contradicting the uniqueness of $A(\omega)$. We now demonstrate how to construct $A_1(\omega)$ using figures for $M = 4$, and the interval $[0.4\pi, 0.6\pi]$ as the transition band.

By Lemma 2, we know that the corresponding sign function of the error of $A(\omega)$, $\mathrm{sign}(E(\omega, \mathbf{a}))$, has the shape shown in Fig. 13(d). We assume on the contrary that $E(\omega, \mathbf{a})$ has $M = 4$ sign changes in Ω and denote the point of change by $\{z_1, z_2, z_3, z_4\}$. The corresponding functions, $A(\omega)$ and $\mathrm{sign}(E(\omega, \mathbf{a}))$ are shown in Fig. 16 in dashed and solid respectively.

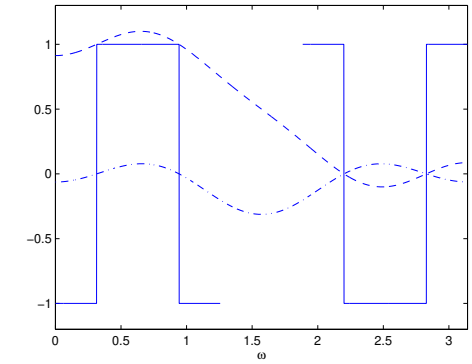


Fig. 16. $A(\omega)$ (dashed), $\mathrm{sign}(E(\omega, \mathbf{a}))$ (solid) and $T(\omega)$ (dashdot), $M = 4$.

Now, construct the cosine $M = 4$ degree cosine polynomial $T(\omega)$ given by

$$T(\omega) = (\cos(\omega) - \cos(z_1))(\cos(\omega) - \cos(z_2))(\cos(\omega) - \cos(z_3))(\cos(\omega) - \cos(z_4)), \tag{37}$$

which is shown in black in Fig.16. Note that in the passband, $T(\omega)$ and $A(\omega)$ have the same sign, and therefore any positive multiple of $T(\omega)$ added to $A(\omega)$ preserves the sign of $A(\omega)$ the passband. In the stopband, however, $T(\omega)$ and $A(\omega)$ have opposite signs and their sum may no longer coincide with the sign of $A(\omega)$. If, however, a small positive multiple of $T(\omega)$ is added to $A(\omega)$ it is possible to preserve the sign of the sum in the stopband equal to that of $A(\omega)$. Thus, our goal is to find $\epsilon > 0$, such that

$$A_1(\omega) = A(\omega) + \epsilon T(\omega), \quad \text{and} \quad \mathrm{sign}(\mathrm{E}(\omega, \mathbf{a})) = \mathrm{sign}(\mathrm{E}(\omega, \mathbf{a}_1)).$$

We now show how to choose ϵ. As mentioned in the last paragraph, the problem is in the stopband, so we concentrate on the interval $[0.6\pi, \pi]$. Consider first the interval $[0.6\pi, z_3]$. In this interval $A(\omega)$ is positive and $T(\omega)$ is negative. Since z_3 is a zero of both $A(\omega)$ and $T(\omega)$ we can write the two functions as

$$A(\omega) = (\cos(z_3) - \cos(\omega))P(\omega), \qquad T(\omega) = (\cos(z_3) - \cos(\omega))Q(\omega),$$

where $P(\omega) > 0$ and for $Q(\omega) < 0$ for all $\omega \in [0.6\pi, z_3]$. Denote $p = \min_{[0.6\pi, z_3]} P(\omega)$, $Q = \max_{[0.6\pi, z_3]} |Q(\omega)|$, and define

$$\epsilon_1 = \frac{p}{2Q}.$$

Now, $\epsilon_1 > 0$, and for all $\omega \in [0.6\pi, z_3]$ we have

$$\begin{aligned} &A(\omega) + \epsilon_1 T(\omega) = (\cos(z_3) - \cos(\omega))P(\omega) + \epsilon_1(\cos(z_3) - \cos(\omega))Q(\omega) \\ &\geq (\cos(z_3) - \cos(\omega))(P(\omega) - \frac{p}{2}) > 0. \end{aligned}$$

Therefore, $A(\omega) + \epsilon_1 T(\omega) > 0$ on $[0.6\pi, z_3]$, having the same sign as $A(\omega)$.

We now consider the interval $[z_3, z_4]$. In this interval $A(\omega)$ is negative and $T(\omega)$ is positive. Since both z_3 and z_4 are zeros of $A(\omega)$ and $T(\omega)$ we can write the two functions as

$$A(\omega) = (\cos(\omega) - \cos(z_3))(\cos(z_4) - \cos(\omega))R(\omega), \quad T(\omega) = (\cos(\omega) - \cos(z_3))(\cos(z_4) - \cos(\omega))S(\omega),$$

where $R(\omega) < 0$ and $S(\omega) > 0$ for all $\omega \in [z_3, z_4]$. Denote $r = \min_{[z_3, z_4]} |R(\omega)|$, $S = \max_{[z_3, z_4]} S(\omega)$, and define

$$\epsilon_2 = \frac{r}{2S}.$$

Now, $\epsilon_2 > 0$, and for all $\omega \in [z_3, z_4]$ we have

$$\begin{aligned} A(\omega) + \epsilon_2 T(\omega) = (\cos(\omega) - \cos(z_3))(\cos(z_4) - \cos(\omega))(R(\omega) + \epsilon_2 S(\omega)) = \\ (\cos(\omega) - \cos(z_3))(\cos(z_4) - \cos(\omega))(R(\omega) + \frac{r}{2S} S(\omega)) \leq \\ (\cos(\omega) - \cos(z_3))(\cos(z_4) - \cos(\omega))(R(\omega) + \frac{r}{2}) < 0. \end{aligned}$$

Therefore, we have $A(\omega) + \epsilon_2 T(\omega) < 0$ on $[z_3, z_4]$, having the same sign as $A(\omega)$.

Continuing in a similar manner in the interval $[z_4, \pi]$ we obtain $\epsilon_3 > 0$, satisfying $A(\omega) + \epsilon_3 T(\omega) > 0$ on $[z_4, \pi]$, and having the same sign as $A(\omega)$.

Finally, by taking

$$\epsilon = \min\{\epsilon_1, \epsilon_2, \epsilon_3\},$$

we get that $A_1(\omega) = A(\omega) + \epsilon T(\omega)$, have the same sign as $A(\omega)$ over the entire stopband. Therefore, the errors of $A_1(\omega)$ and $A(\omega)$ have the same sign functions, implying that they are both optimal. Since $A_1(\omega) \neq A(\omega)$ however, we arrive at a contradiction to the fact that $A(\omega)$ is unique.

APPENDIX IV
PROOF OF THEOREM 5

Before proving Theorem 5, we shall state three useful lemmas.

Lemma 3 ([43]): If $\mathbf{A} \in \mathbb{R}^{n\times n}$ is positive definite, and $\mathbf{X} \in \mathbb{R}^{n\times n}$ is nonsingular, then

$$\lambda_{min}\left(\mathbf{X}^T\mathbf{A}\mathbf{X}\right) \geq \lambda_{min}\left(\mathbf{A}\right)\sigma^2_{min}\left(\mathbf{X}\right) \quad \text{and} \quad \lambda_{max}\left(\mathbf{A}\right)\sigma^2_{max}\left(\mathbf{X}\right) \geq \lambda_{max}\left(\mathbf{X}^T\mathbf{A}\mathbf{X}\right),$$

Lemma 4 ([44]): If $|||\cdot|||$ is any matrix norm and $\mathbf{A}$ is an $n \times n$ matrix, then $\lambda_{max}\left(\mathbf{A}\right) \leq |||\mathbf{A}|||$.

Lemma 5: Let $\{z_0, z_1, ..., z_n\}$ be $n+1$ distinct points in $[0, \pi]$ and let $\mathbf{B}$ be an $(n+1)\times(n+1)$ matrix, whose ijth element is $\cos((i-1)z_{j-1})$. Then,

$$\sigma_{max}\left(\mathbf{B}\right) \leq n+1 \quad \text{and} \quad \frac{\left(\min_{i,j,i\neq j}|\cos(z_i)-\cos(z_j)|\right)^{\frac{n(n-1)}{2}}}{(n+1)^n} \leq \sigma_{min}\left(\mathbf{B}\right).$$

Proof: We begin with the first inequality. Using Lemma 4 and the matrix norm $|||\mathbf{A}||| = (n+1)\max_{i,j}|\mathbf{A}_{ij}|$, we have that

$$\sigma^2_{max}\left(\mathbf{B}\right) = \lambda_{max}\left(\mathbf{B}\mathbf{B}^T\right) \leq (n+1)\max_{i,j}\left|\left(\mathbf{B}\mathbf{B}^T\right)_{ij}\right|.$$

Now,

$$\left|\left(\mathbf{B}\mathbf{B}^T\right)_{ij}\right| = \left|\sum_{k=1}^{n+1}\cos((i-1)z_{k-1})\cos((j-1)z_{k-1})\right| \leq \sum_{k=1}^{n+1}|\cos((i-1)z_{k-1})\cos((j-1)z_{k-1})| \leq n+1, \quad (38)$$

proving the first inequality. To prove the second one we first note that by the singular value decomposition of $\mathbf{B}$, we have

$$|det(\mathbf{B})| = \prod_{k=1}^{n+1}\sigma_k(\mathbf{B}) \leq \sigma_{min}(\mathbf{B})\sigma^n_{max}(\mathbf{B}) \leq \sigma_{min}(\mathbf{B})(n+1)^n, \quad (39)$$

where in the last inequality we used (38). In [45] it is show that

$$|det(\mathbf{B})| = \left|\prod_{i=0}^{n-1}\theta_i^{n-i}\prod_{i>j}\left(\cos(z_i)-\cos(z_j)\right)\right|,$$

where $\theta_0 = 1$ and $\theta_i = 2, \quad i \geq 1$. Thus,

$$|det(\mathbf{B})| \geq \left|\prod_{i=0}^{n-1}\prod_{i>j}\left(\cos(z_i)-\cos(z_j)\right)\right| \geq \left(\min_{i,j}|(\cos(z_i)-\cos(z_j))|\right)^{\frac{n(n-1)}{2}}. \quad (40)$$

Combining (39) and (40) the second inequality. ■

We now prove Theorem 5.

Proof: According to Proposition 3, we need to find two positive numbers λ_1 and λ_2, such that (21) holds. We claim that the constants δ_1, δ_2 and μ, which are determined in the first step of the algorithm satisfy (21). To see this let us examine the three possibilities $\mathbf{H}^k$ can assume at the kth iteration.

If $\mathbf{H}^k = \mathbf{I}$, then $\lambda_{min}\left(\mathbf{H}^k\right) = \lambda_{max}\left(\mathbf{H}^k\right) = 1$. Any choice of constants such that $\lambda_1 < 1$ and $1 < \lambda_2$ will satisfy (21) in this case.

Suppose now that $\mathbf{H}^k$ equals the Hessian matrix at the kth iteration, and can be therefore expressed as $\mathbf{H}^k = \mathbf{V}^{k^T}\mathbf{D}^k\mathbf{V}^k$. In this case, we also have

$$\delta_1 \leq \lambda_{min}\left(\mathbf{D}^k\right), \quad \lambda_{max}\left(\mathbf{D}^k\right) \leq \delta_2, \quad \text{and} \quad \mu \leq \min_{i,j}|\cos(z_i)-\cos(z_j)|. \quad (41)$$

Therefore, since $\mathbf{H}^k$ is positive definite, $\mathbf{V}^k$ is full-rank. We also claim that $\mathbf{V}^k$ is $(M+1)\times(M+1)$ and is invertible. To see this, note that since $\mathbf{V}^k$ is full rank, the number of its rows has to be at least $M+1$ (the number columns is by definition $M+1$). This means that there are at least $M+1$ zeros to the error function at the kth iteration. However, from Theorem 2, we know the number of zeros cannot exceed $M+1$, and therefore $\mathbf{V}^k$ must be an $(M+1)\times(M+1)$ matrix. In addition, by the property of the set of functions $\{1,\cos(\omega),\ldots,\cos(M\omega)\}$, $\mathbf{V}^k$ is invertible [35]. Applying Lemma 3 with $\mathbf{A}=\mathbf{D}^k$ and $\mathbf{X}=\mathbf{V}^k$, we obtain

$$\lambda_{min}\left(\mathbf{H}^k\right)\geq\lambda_{min}\left(\mathbf{D}^k\right)\sigma^2_{min}\left(\mathbf{V}^k\right).$$

From Lemma 5 with $\mathbf{B}=\mathbf{V}^k$ and (41) , we therefore have

$$\lambda_{min}\left(\mathbf{H}^k\right)\geq\delta_1\frac{\mu^{M(M-1)}}{(M+1)^{2M}}.$$

Thus, the choice of $\lambda_1=\delta_1\frac{\mu^{M(M-1)}}{(M+1)^{2M}}$ satisfies the lower bound in (21). To obtain λ_2 note that by Lemma 3

$$\lambda_{max}\left(\mathbf{H}^k\right)\leq\lambda_{max}\left(\mathbf{D}^k\right)\sigma^2_{max}\left(\mathbf{V}^k\right).$$

However, from Lemma 5 and (41) we get that

$$\lambda_{max}\left(\mathbf{H}^k\right)\leq\delta_2(M+1)^2,$$

establishing $\lambda_2=\delta_2(M+1)^2$.

Finally we have to consider the case when $\mathbf{H}^k$ is equal to the modified Hessian. In this case, it can be shown that the Cholesky decomposition described in Appendix C results in a bounded condition number of $\mathbf{H}^k$ [46]. ■

APPENDIX V
PROOF OF THEOREM 6

We complete the proof of Theorem 6 by showing that $\nabla^2\|E(\omega,\mathbf{a})\|_1$ is Lipschitz continuous in a neighborhood of $\mathbf{a}^*$, that is for $\mathbf{x}$ and $\mathbf{y}$ in $B(\mathbf{a}^*,\epsilon)$,

$$|||\mathbf{H}(\mathbf{y})-\mathbf{H}(\mathbf{x})|||\leq L\|\mathbf{y}-\mathbf{x}\| \tag{42}$$

for some $L>0$, where $|||\cdot|||$ is a matrix norm. To see this we shall use the matrix norm given by,

$$|||\mathbf{A}|||=(M+1)\max_{i,j}|\mathbf{A}_{ij}|. \tag{43}$$

Let z_i^x, $i=1,...,M+1$ be the set of zeros corresponding to $\mathbf{x}$, and similarly z_i^y for $\mathbf{y}$. We set $g_{ij}(z,\mathbf{x})=W(z)\cos(iz)\cos(jz)\frac{2}{|E'(z,\mathbf{x})|}$. Now,

$$|||\mathbf{H}(\mathbf{y})-\mathbf{H}(\mathbf{x})|||=(M+1)\max_{i,j}\left|\sum_{m=1}^{t}g_{ij}(z_m^x,\mathbf{x})-g_{ij}(z_m^y,\mathbf{y})\right|\leq(M+1)t\left|g_{i^*j^*}(z_l^x,\mathbf{x})-g_{i^*j^*}(z_l^y,\mathbf{y})\right| \tag{44}$$

The first equality is a substitution of the definition of $\mathbf{H}$ in (43) using the function $g_{ij}(z,\mathbf{x})$. In the second one we denote the indices of the maximal element by (i^*,j^*), whereas in the third inequality we use the triangle inequality and bound each element by the maximal element, whose index we denote by l.

We wish to show that the expression in (44) is not greater than $L\|\mathbf{y}-\mathbf{x}\|$ for some $L>0$. By the assumption that at $\mathbf{a}^*$ all the zeros are simple, and by continuity, all the zeros in the neighborhood $B(\mathbf{a}^*,\epsilon)$ are simple as well. Thus, we can assume without loss of generality that $E'(z_l^x,\mathbf{x})>0$ for all $\mathbf{x}\in B(\mathbf{a}^*,\epsilon)$. If $W(z)$ is Lipschitz continuous

in z (it is, for example, when it is constant, which is the common case), $g_{ij}(z,\mathbf{x})$ is Lipschitz continuous in z (as a function from $\mathbb{R} \to \mathbb{R}$). To see this, note that $g_{ij}(z,\mathbf{x})$ is a product of three Lipschitz continuous functions of z. In particular, $W(z)$ is Lipschitz by assumption, whereas $\cos(iz)$ and $\cos(jz)$ are also of that type since they have bounded derivatives. Finally, $\frac{1}{2}|E^{'}(z,\mathbf{x})| = \frac{1}{2}E^{'}(z,\mathbf{x})$ in $B(\mathbf{a}^*,\epsilon)$, where it also has a bounded derivative, and therefore Lipschitz continuous in z as well.

Using the proof of Theorem 3, we know that $\frac{\partial z}{\partial \mathbf{x}_j}$ equals $-\frac{\cos(jz)}{E^{'}(z,\mathbf{x})}$, which is bounded, and thus z is Lipschitz continuous in $\mathbf{x}$. As a result, $g_{ij}(z,\mathbf{x})$ is Lipschitz continuous in $\mathbf{x}$ (as a function from $\mathbb{R}^{M+1} \to \mathbb{R}$), i.e. there exists an $\bar{L} > 0$, such that

$$|g_{ij}(z_l^x,\mathbf{x}) - g_{ij}(z_l^y,\mathbf{y})| \leq \bar{L}\|\mathbf{y}-\mathbf{x}\|.$$

Combining this last inequality with (44) we conclude that

$$|||\mathbf{H}(\mathbf{y}) - \mathbf{H}(\mathbf{x})||| \leq L\|\mathbf{y}-\mathbf{x}\|, \tag{45}$$

for $L = (M+1)t\bar{L}$, establishing a second order rate of convergence.

REFERENCES

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

[2] B. Porat, *A Course in Digital Signal Processing*, New York, NY: John-Wiley & Sons, 1997.

[3] J. W. Adams, "FIR digital filters with least-squares stopbands subject to peak-gain constrains," *IEEE Trans. on Circuits and System Theory*, vol. 39, pp. 376–388, 1991.

[4] C. S. Burrus and J. A. Barreto, "Least p-power error design of filters," *Proc. IEEE Int. Symp. Circuits. Syst. ISCAS-92*, pp. 545–548, 1992.

[5] D. W. Tufts and J. T. Francis, "Designing digital low-pass filters - comparison of some methods and criteria," *IEEE Trans. Audio Electroacoustics*, vol. 18, pp. 487–494, 1970.

[6] V. R. Algazi and M. Suk, "On the frequency weighted least-squares design of finite duration filters," *IEEE Trans. on Circuits and Systems*, vol. 12, pp. 943–953, 1975.

[7] C. S. Burrus , A. W. Soewito and R. A. Gopinath, "Least squared error FIR filter design with transition bands," *IEEE Trans. on Signal Processing*, vol. 40, pp. 1327–1339, 1992.

[8] W. C. Kellogg, "Time domain design of nonrecursive least mean-square digital filters," *IEEE Trans. Audio Electroacoust.*, vol. 20, pp. 155–158, 1972.

[9] Y. C. Lim and S. R. Parker, "Discrete coefficient FIR digital filter design based upon an LMS criteria," *IEEE Trans. Circuits Syst.*, vol. 30, pp. 723–739, 1983.

[10] M. Okuda, M. Ikehara and S. Takahashi, "Fast and stable least-squares approach for the design of linear phase FIR filters," *IEEE Trans. Signal Processing*, vol. 46, pp. 1485–1493, 1998.

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

[12] E. Z. Psarakis, "A weighted L_2-based method for the design of arbitrary one-dimensional FIR digital filters," *Signal Processing*, vol. 86, pp. 937–950, 2006.

[13] J. F. Kaiser, "Design methods for sampled data filters," *Proc. 1st Allerton Conf. Circuit and System Theory*, pp. 211–236, 1963.

[14] H. D. Helms, "Nonrecursive digital filters: Design methods for achieving specifications on frequency response," *IEEE Trans. on Audio Electroacoust*, vol. 16, pp. 336–342, 1968.

[15] 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, April 1975.

[16] J.H. McClellan and T. W. Parks, "A unified approach to the design of optimum FIR linear-phase digital filter," *IEEE Trans. Circuits Systems*, vol. 20, pp. 697–701, 1973.

[17] H. D. Helms, "Digital filters with equiripple or minimax responses," *IEEE Trans. Audio Electroacoustics*, vol. 19, pp. 87–94, 1971.

[18] O. Hermann, "Design of nonrecursive digital filters with linear phase," *Electron. Lett.*, vol. 6, pp. 328–329, 1970.

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

[20] E. Z. Psarakis and G. V. Moustakides, "A robust initialization scheme for the remez exchange algorithm," *IEEE signal processing letters*, vol. 10, pp. 1–3, January 2003.

[21] P. P. Vaidyanathan, "Efficient and mutiplierless design of FIR filters with very sharp cutoff via maximally flat building blocks," *IEEE Trans. on Circuits and Systems*, vol. 3, pp. 236–244, 1985.

[22] L. R. Rajagopal and S. C. Dutta Roy, "Design of maximally flat FIR filters using the Bernstein polynomial," *IEEE Trans. on Circuits and Systems*, vol. 34, pp. 1587–1590, 1987.

[23] T. Cooklev and A. Nishihara, "Maximally flat FIR filters," *IEEE ISCAS*, pp. 96–99, 1993.

[24] P. Bloomfield, *Least Absolute Deviations: Theory, Applications, and Algorithms*, Boston, Mass: Birkhauser, 1983.

[25] C. K. Chen and J. H. Lee, "Design of high-order digital differentiatiors using L_1 error criteria," *IEEE Trans. Circ. Sys.: Analog and Digital Signal Processing*, vol. 42, pp. 287–291, April 1995.

[26] C. S. Burrus , J. A. Barreto and I. Selesnick, "Iterative reweighted least-squares design of filters," *IEEE Trans. Signal Processing*, vol. 42, pp. 2926–2936, 1994.

[27] W. S. Yu , I. K. Fong , and K. C. Chang, "An l_1 approximation based method for synthesizing fir details," *IEEE Trans. on Circuits and Systems II*, vol. 39, pp. 578–561, 1992.

[28] E. Moskona and E. Saff, "Gibbs phenonmenon for best L_1 trigonometric polynomial approximation," *Const. Approx.*, vol. 11, pp. 391–416, 1995.

[29] I. Barrodale and F. D. K. Roberts, "An improved algorithm for discete l_1 linear approximation," *SIAM J. on Num. Anal.*, vol. 10, pp. 839–848, 1973.

[30] C. A. Zala , I. Barrodale and J. S. Kennedy, "High-resolution signal and noise field estimation using the L_1 (least absolute value) norm," *IEEE J. of Oceanic Engineering*, vol. 12, pp. 253–263, 1987.

[31] L. R. Rabiner, "The design of finite impulse response digital filters using linear programming techniques," *Bell Syst. Tech. J.*, vol. 51, pp. 1177–1198, 1972.

[32] K. H. Usow, "On L_1 approximation II: Computation for discrete function and discretization effects," *SIAM J. Numer. Anal.*, vol. 4, pp. 233–244, 1967.

[33] K. H. Usow, "On L_1 approximation I: Computation for continuous function and continuous dependence," *SIAM J. Numer. Anal.*, vol. 4, pp. 70–88, 1967.

[34] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, New York, NY: McGraw-Hill Internetional Edition, 1996.

[35] J. R. Rice, *The Approximation of Functions*, vol. 1, Reading, MA: Addison-Wesley, 1964.

[36] G. A. Watson, "An algorithm for linear L_1 approximation of continuous functions," *IMA J. Num. Anal.*, vol. 1, pp. 157–167, 1981.

[37] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, London, UK: Academic Press, 1981.

[38] D. P. Bertsekas, *Nonlinear Programming*, Belmont, Mass: Athena Scientific, 1999.

[39] J. Nocedal and S. J. Wright, *Numerical Optimization*, New York, NY: Springer, 1999.

[40] M. Lang and B. C. Frenzel, "Polynomial root finding," *IEEE Signal Processing Letters*, vol. 1, pp. 141–143, 1994.

[41] N. J. Higham, "Fast solution of Vandermonde-like systems involving orthogonal polynomials," *IMA J. Num Anal.*, vol. 8, pp. 473–486, 1988.

[42] M. J. D Powell, *Approximation Theory and Methods*, Cambridge, UK: Cambridge University Press, 1981.

[43] G. H. Golub and C. F. van Loan, *Matrix Computations*, Baltimore, MD: Johns Hopkins University Press, 1997.

[44] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge University Press, 1985.

[45] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, Philadelphia, PA: SIAM, 1998.

[46] J. J. More and D. C. Sorensen, "Newton's method," *Studies in Numerical Analysis*, vol. 24, pp. 29–82, 1984.