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

Liron D. Grossmann and Yonina C. Eldar, *Member, IEEE*

***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 optimal 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 design techniques is made, demonstrating that the L_1 approach may be a good alternative in several applications.**

***Index Terms*—Differentiability, digital filtering, L_1 norm, Newton method, uniqueness.**

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"; clearly, various optimality criteria lead to different filter behavior, providing a convenient tool to control the inherent tradeoff of the design problem. Finally, developing a computational method to find the desired filter coefficients. The last two steps are closely related, since 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 [3]. The use of L_p norm, $2 \leq p \leq \infty$, has also been suggested as a successful alternative to the L_2 and L_∞ criteria [4].

Least-squares filters are popular and are used in many applications [5]–[11]. They gained their popularity due to two main reasons. First, the resulting optimal filters require the solution of a single linear system of equations, which can be solved efficiently. The eigenfilter method is one of the fastest ways to obtain an approximate L_2 filter [12]. Second, minimizing the least-squares error has the physical interpretation of energy minimization, which is also related to the '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]–[18]. Minimax filters result in the minimum number of coefficients for a given tolerance scheme, i.e., when the design specifications are given in terms of the transition bandwidth and 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 extremes. 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 response.

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 approach. The maximally flat filters are easy to compute and several closed form formula exist for their design [21]–[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, in this paper we consider using the weighted L_1 norm for approximating digital filters. The use

Manuscript received April 3, 2006; revised December 21, 2006. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mariane R. Petraglia. This work was supported in part by the Israel Science Foundation under Grant 536-04.

L. D. Grossmann is with the department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: lirongr@tx.technion.ac.il).

Y. C. Eldar is with the department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2007.896088

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]–[27]. In [25], the design of high-order differentiators was considered, and in [27] an arbitrary amplitude function was designed using the discrete l_1 norm. A general algorithm for the approximation under L_p was proposed in [26], but convergence is not guaranteed for $p = 1$, and when it exists, it 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 the lack of motivation and an efficient algorithm that solves the original problem. It is the goal of this paper to provide 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.

We start by formulating our L_1 design problem in Section II, after which we provide motivation for using L_1 filters. The mathematical framework of L_1-based filters is introduced 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 theorem 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 discussed. In Section V we compare between the proposed algorithm and the Remez exchange method, and show that both algorithms share several properties, especially fast running time. Some examples are given in Section VI, 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, $\|\mathbf{a}\| = \sqrt{\sum_{n=0}^{M} |\mathbf{a}_n|^2}$. 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

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

For given ω_p and ω_s, the set $[0,\omega_p] \cup [\omega_s,\pi]$ is denoted by Ω. The weighted inner product between two real functions, f and g, on Ω is written as

$$\langle f, g \rangle = \int_{\Omega} W(\omega) f(\omega) g(\omega) d\omega \tag{2}$$

where $W(\omega)$ is a positive weighting function and is at least twice differentiable on Ω.

Given a function $A(\omega, \mathbf{a})$ defined for every $\omega \in \Omega$ and every $\mathbf{a} \in \mathbb{R}^M$ for some $M > 0$, we denote its derivative (assuming it exists) with respect to ω by $A'(\omega, \mathbf{a})$. The set of zeros of $A(\omega, \mathbf{a})$ is

$$Z(\mathbf{a}) = \{\omega \in \Omega \,|\, A(\omega, \mathbf{a}) = 0\}. \tag{3}$$

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, $\mu(Z(\mathbf{a})) = 0$. The following definition will be useful in our development:

Definition 1 (Simple Zero): A zero ω_1 of a differentiable function $X(\omega)$ is called simple if $X(\omega_1) = 0$ 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

$$D(\omega) = \begin{cases} 1, & \omega \in [0,\omega_c] \\ 0, & \omega \in (\omega_c,\pi]. \end{cases} \tag{4}$$

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) of its impulseresponse $\mathbf{h}_n$:

$$H(\omega) = \sum_{n=0}^{N} \mathbf{h}_n e^{-j\omega n}. \tag{5}$$

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

$$H(\omega) = A(\omega) e^{-jM\omega} \tag{6}$$

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

$$\begin{aligned} A(\omega) &= \mathbf{h}_M + \sum_{n=1}^{M} 2\mathbf{h}_{M-n} \cos(n\omega) \\ &\triangleq \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega). \end{aligned} \tag{7}$$

As we discussed in [28], our design problem is to approximate $D(\omega)$ by a linear combination of $M+1$ functions $\{\cos(n\omega), n = 0,\ldots,M\}$, that is by the cosine polynomial $A(\omega)$.

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

$$E(\omega, \mathbf{a}) = \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega) - D(\omega). \tag{8}$$

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 as follows.

1) Weighted least-squares

$$||E(\omega, \mathbf{a})||_2 = \sqrt{\int_{\Omega} W(\omega)|E(\omega, \mathbf{a})|^2 d\omega}. \tag{9}$$

2) Weighted Chebyshev

$$||E(\omega, \mathbf{a})||_\infty = \max_{\omega \in \Omega} W(\omega)|E(\omega, \mathbf{a})|. \tag{10}$$

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

$$||E(\omega, \mathbf{a})||_1 = \int_{\Omega} W(\omega)|E(\omega, \mathbf{a})| d\omega. \tag{11}$$

The function $W(\omega)$ is a positive weighting function, used in order to weight certain frequencies. Thus, our problem is to find $\mathbf{a}$ such that $||E(\omega, \mathbf{a})||_1$ is minimized.

C. Motivation

From the pure mathematical point of view, the theory of L_1 approximation is much more challenging than that of L_2 and L_∞. As we shall see in the next section, the L_1 norm of the error may not always be differentiable, and the optimal solution is not guaranteed to be unique. In addition, there is no closed-form solution to the optimal filter as in the least-squares case, nor is there a known efficient algorithm such as Remez for calculating the resulting filter. Thus, investigating the mathematical properties of L_1 filters and finding an efficient algorithm is an interesting mathematical challenge.

From the engineering point of view, the L_1 norm is very popular in other fields such as robust estimation problems and sparse representations. Thus, it is natural to investigate it in the filter design context. In robust estimation, for example, the L_1 norm is used to attenuate large spikes contaminating a desired signal. The L_1 measure is suitable in this case since it tends to give weight to small error, ignoring rare error events. Considering the discontinuity in the ideal frequency response as a rare event (most of time the response is either one or zero), it is therefore reasonable to expect that when the L_1 norm is applied in the filter design problem, it will smear the transition band while keeping the passband and stopband as smooth as possible.

More insight into the properties of the L_1 norm can be gained by comparing the L_1, L_2, L_∞ solutions when the interval $[0, \pi]$ is not split, i.e., $\Omega = [0, \pi]$, and $W(\omega) = 1$. Fig. 1 shows the 65-length L_1, L_2, L_∞ FIR filters approximating an ideal low-pass with $\omega_c = 0.485\ \pi$. Since we are approximating a discontinuous function, the L_1 and L_∞ filters were obtained by discretization methods, which do not require continuity.

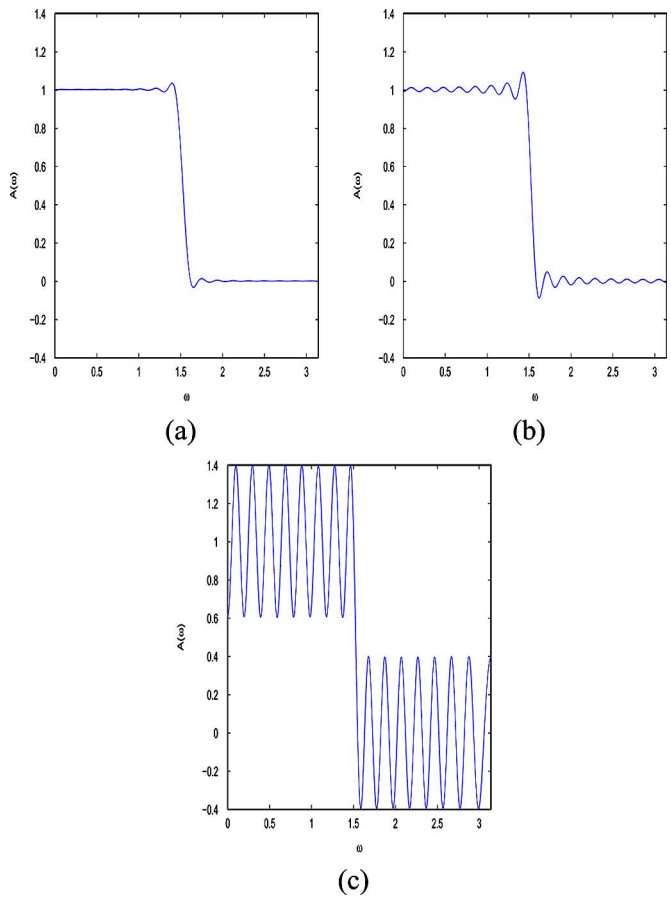

Fig. 1. Approximations of an ideal low-pass filter with $\omega_c = 0.485\ \pi$ over $[0, \pi]$: (a) L_1; (b) L_2; (c) L_∞.

From Fig. 1 it is evident that all three filters result in an overshoot around the discontinuity, and that the L_1 filter possesses the smallest one. The existence of the overshoot is well known in the least-squares case as the Gibbs phenomenon [2], and is equal to 9% of the size of the jump. The fact that Gibbs phenomenon (i.e., the overshoot behavior near the edge) is the least severe when the approximation norm is L_1 was conjectured in [29], 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.

It is important to understand 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. In practice, however, all three overshoots are usually too high and therefore the cutoff frequency must be excluded from the approximation. In this case Ω is the disjoint union of the passband and the stopband and the overshoot of the minimax filters is the smallest. In most of the paper, we consider the approximation when the cutoff frequency is excluded. The special case when $\Omega = [0, \pi]$ is discussed in Section V.

Fig. 1 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.

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 characterize L_1 filters. These results lay the ground for the development of an efficient L_1 algorithm, which is derived in Section IV.

III. The Mathematical Framework of 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], [30], [31]. Discretization techniques are also common for solving the corresponding L_2 and L_∞ measures [32]. 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 [33]. 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 [34]. Thus, solving the continuous L_1 problem is difficult to avoid. 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) [35]. 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.

The following theorem, which is a straightforward generalization of a theorem in [36], states a sufficient condition on a vector $\mathbf{a}^*$ such that $||E(\omega,\mathbf{a})||_1$ is differentiable at $\mathbf{a}^*$. Recall that μ denotes the Lebesgue measure of a given set.

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 (4), $\mu(Z(\mathbf{a})) = 0$ for every $\mathbf{a} \neq (1,0,0,\ldots,0)$ and $\mathbf{a} \neq (0,0,0,\ldots,0)$.

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 [36]. Since we assume $D(\omega)$ is piecewise constant, in each interval of Ω where $D(\omega)$ is constant, $E(\omega,\mathbf{a})$ is an M degree polynomial. Thus, each such interval 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 an interval of Ω 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 L_1 solution.

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

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

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 $\{\cos(n\omega), n = 0,\ldots,M\}$ [36]. 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 $(M \geq 1)$, corresponding to a vector $\mathbf{a}$. Then $E(\omega,\mathbf{a})$ changes sign either M or $M+1$ times in Ω.

Proof: See Appendix II. ■

Theorem 2 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.

We now address 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 [37]. Here we generalize it to a weighted norm over a disjoint union of intervals. Since the generalization and its further consequences are not straightforward, we provide a full proof of the theorem.

Theorem 3 (Second-Order Derivative): Let $Z(\mathbf{a}) = \{z_1,\dots,z_t\}$ 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 $\mathbf{D} = \mathrm{diag}\{d_1,\dots,d_t\}$ with $d_i = 2W(z_i)/|E^{'}(z_i,\mathbf{a})|$.

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

$$\mathbf{g}_i(\mathbf{a}) = \int_0^{\omega_p} W(\omega)\cos(i\omega)\mathrm{sign}(E(\omega,\mathbf{a}))d\omega + \int_{\omega_s}^{\pi} W(\omega)\cos(i\omega)\mathrm{sign}(E(\omega,\mathbf{a}))d\omega. \tag{15}$$

Suppose there are $l+1$ zeros in $[0,\omega_p]$ and $t-l-1$ zeros in $[\omega_s,\pi]$. Thus,

$$\begin{aligned}\int_0^{\omega_p} W(\omega)\cos(i\omega)\mathrm{sign}(E(\omega,\mathbf{a}))d\omega &= \int_0^{z_1} v_p W(\omega)\cos(i\omega)d\omega \\ &\quad + \sum_{k=1}^{l}\int_{z_k}^{z_{k+1}} v_k W(\omega)\cos(i\omega)d\omega \\ &\quad + \int_{z_{l+1}}^{\omega_p} v_{l+1} W(\omega)\cos(i\omega)d\omega\end{aligned}$$

where $v_p = -\mathrm{sign}(E^{'}(z_1,\mathbf{a}))$, and $v_k = \mathrm{sign}(E^{'}(z_k,\mathbf{a}))$. The last equality is justified by the assumption that the zeros are simple. Similarly with $v_s = -\mathrm{sign}(E^{'}(z_{l+2},\mathbf{a}))$

$$\begin{aligned}\int_{\omega_s}^{\pi} W(\omega)\cos(i\omega)\mathrm{sign}(E(\omega,\mathbf{a}))d\omega &= \int_{\omega_s}^{z_{l+2}} v_s W(\omega)\cos(i\omega)d\omega \\ &\quad + \sum_{k=l+2}^{t-1}\int_{z_k}^{z_{k+1}} v_k W(\omega)\cos(i\omega)d\omega \\ &\quad + \int_{z_t}^{\pi} v_t W(\omega)\cos(i\omega)d\omega.\end{aligned}$$

Using the chain rule we have that

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

Denoting $u_k = E^{'}(z_k,\mathbf{a})$, we write $v_k = \mathrm{sign}(u_k)$ and $\partial v_k/\partial\mathbf{a}_j = (\partial v_k/\partial u_k)\cdot(\partial u_k/\partial\mathbf{a}_j)$. Since z_k is assumed to be simple, $u_k \neq 0$, and thus $\partial v_k/\partial u_k = 0$, resulting in $\partial v_k/\partial\mathbf{a}_j = 0$. The same argument holds for $\partial v_p/\partial\mathbf{a}_j$ and $\partial v_s/\partial\mathbf{a}_j$. Now, for $k\neq 1, l+2$ we have

$$\begin{aligned}\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial z_k} &= W(z_k)\cos(iz_k)(v_{k-1}-v_k) \\ &= -2v_k W(z_k)\cos(iz_k),\end{aligned} \tag{16}$$

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

$$\begin{aligned}\frac{\partial\mathbf{g}_i(\mathbf{a})}{\partial z_1} &= W(z_1)\cos(iz_1)(v_p - v_1) \\ &= -2v_1 W(z_1)\cos(iz_1).\end{aligned}$$

The same is true for z_{l+2}, concluding that $\partial\mathbf{g}_i(\mathbf{a})/\partial z_k = -2v_k W(z_k)\cos(iz_k)$ for all k.

It remains to compute $\partial z_k/\partial\mathbf{a}_j$, where we note that the dependence of z_k on $\mathbf{a}_j$ is implicit, through the equations $E(z_k,\mathbf{a}) = 0, k = 1,\dots,t$. Define the matrices $\mathbf{Z}_\mathbf{a} = (\partial z_k/\partial\mathbf{a}_j), \mathbf{F}_z = (\partial E(z_i,\mathbf{a})/\partial z_j)$, and $\mathbf{F}_\mathbf{a} = (\partial E(z_i,\mathbf{a})/\partial\mathbf{a}_j)$. By the implicit function theorem, $\mathbf{Z}_\mathbf{a} = -\mathbf{F}_z^{-1}\mathbf{F}_\mathbf{a}$. Now, from the definition of $E(z_k,\mathbf{a}), \mathbf{F}_z$ is seen to be diagonal with diagonal elements equal to $\partial E(z_i,\mathbf{a})/\partial z_i = E^{'}(z_i,\mathbf{a})$. The ijth element of $\mathbf{F}_a$ is given by $\partial E(z_i,\mathbf{a})/\partial\mathbf{a}_j = \cos(jz_i)$. Therefore

$$\frac{\partial z_k}{\partial\mathbf{a}_j} = -\frac{\cos(jz_k)}{E^{'}(z_k,\mathbf{a})}. \tag{17}$$

Combining (16) and (17) 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 differentiable for almost every choice of filter coefficients and that if its zeros are simple and their number exceeds $M+1$, then $||E(\omega,\mathbf{a})||_1$ 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 [38], $A_1(\omega)$ must satisfy

$$\begin{aligned}&a)\ (D(\omega)-A(\omega))(D(\omega)-A_1(\omega)) \geq 0 \quad \text{on } \Omega \\ &b)\ \int_\Omega \mathrm{sign}(D(\omega)-A(\omega))(A_1(\omega)-A(\omega))d\omega = 0.\end{aligned} \tag{18}$$

Condition (b) follows from 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 $A(\omega) = A_1(\omega)$.

The proof of the converse is complicated and is therefore relegated to Appendix III. ∎

We now characterize the zeros of the unique solution.

Corollary 1: Let $A(\omega)$ be the unique optimal weighted L_1 approximation to $D(\omega)$ on Ω. If $A(\omega)$ is not constant, 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 that

$$E^{'}(\omega, \mathbf{a}) = \sum_{k=1}^{M} k\mathbf{a}_k \sin k\omega = A^{'}(\omega) \tag{19}$$

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 (28), $A^{'}(z_1) = 0$ 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 it is identically zero, contradicting the fact that $A(\omega)$ is not constant. ∎

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

The problem of approximating a general continuous function defined over an interval by a finite number of basis functions under the (unweighted) L_1 criterion was considered by Watson in [37]. 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 develop a modified Newton method for computing the L_1 filter coefficients. By exploiting the special structure of the filter design problem, we accelerate several steps and improve the computational complexity of the resulting algorithm.

A. Algorithm Description

In essence, our algorithm generates the sequence

$$\mathbf{a}^{k+1} = \mathbf{a}^k - \gamma^k [\mathbf{H}^k]^{-1} \mathbf{g}^k, \tag{20}$$

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 [39]. We now explain in detail how to implement each stage of our method.

Initialization: We begin by choosing an initial guess of the optimal solution and set the relevant constants, $\epsilon, \sigma, \beta, \delta_1, \delta_2$, and μ. Here ϵ determines the accuracy of the stopping condition, σ and β are related to the step size selection of the fifth stage as will be explained below. The role of δ_1, δ_2, and μ is to control the positive definiteness of the matrix $\mathbf{H}^k$ in (20). Typical values are $\epsilon = 10^{-6}, \sigma = 10^{-3}, \beta = 0.5, \delta_1 = 10^{-15}, \delta_2 = 10^{15}$, and $\mu = 10^{-10}$.

For an initial guess, we would like to choose a vector that is "close" to the optimal solution. Furthermore, 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 (20)). According to Proposition 1, the only two problematic vectors are $\mathbf{a}_0 = (0, 0, \ldots, 0)$ and $\mathbf{a}_1 = (1, 0, \ldots, 0)$. 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,

$$\|E(\omega, \mathbf{a}^1)\|_1 < \min(\|E(\omega, \mathbf{a}_0)\|_1, \|E(\omega, \mathbf{a}_1)\|_1). \tag{21}$$

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

$$z_i = \frac{(2i-1)\pi}{2(M+1)}, \quad i = 1, \ldots, M+1; \tag{22}$$

see [36]. Thus, $\mathbf{a}^1$ is the solution of the linear system

$$\begin{pmatrix} \cos(0z_1) & \ldots & \cos(Mz_1) \\ \cos(0z_2) & \ldots & \cos(Mz_2) \\ \vdots & \vdots & \vdots \\ \cos(0z_{M+1}) & \ldots & \cos(Mz_{M+1}) \end{pmatrix} \mathbf{a}^1 = \begin{pmatrix} D(z_1) \\ D(z_2) \\ \vdots \\ D(z_{M+1}) \end{pmatrix}. \tag{23}$$

The matrix in this system is always invertible, since the functions $\{1, \cos(\omega), \ldots, \cos(M\omega)\}$ form a Chebyshev set on $[0, \pi]$ [36]. Simulations show that this is an excellent choice, since it is usually very close to the optimal solution and also satisfies (21). If some of the points in (22) lie in the transition band (ω_p, ω_s), then 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 $\mathbf{A}_{\mathrm{LS}}\mathbf{a}^1 = \mathbf{b}$ where $\mathbf{A}_{\mathrm{LS}}$ and $\mathbf{b}$ are described in [12]. This guess also appears to be a good choice that satisfies (21), however simulations show that the vector corresponding to (23) often performs better.

Matrix Determination: This step determines the matrix $\mathbf{H}^k$ in (20), which can take on 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 technique. 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 equal to $E(\omega, \mathbf{a}) = \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega)$, and replace each $\cos(n\omega)$ with $(z^n + z^{-n})/2$, where $z = e^{j\omega}$. The

zeros of the new algebraic polynomial in z are efficiently calculated, from which the corresponding zeros in ω are extracted.

Next, we construct the matrix $\mathbf{H}^k$ according to the following rule which guarantees that it is positive definite. If there are no zeros 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 exists. If it is positive definite then $\mathbf{H}^k$ is equal to the expression given by (14). Otherwise $\mathbf{H}^k$ is the result of the modified Cholesky method applied to (14), which is guaranteed to be positive definite [39].

The positive definiteness of the Hessian matrix in (14) is related to the number of zeros (all simple). If their number is less than $M+1$ then the Hessian is positive semidefinite, whereas if it is $M+1$ (it cannot be greater) then the Hessian is positive definite. This can be shown by noting that $\mathbf{V}^k$ in (14) is full rank if and only if the number of zeros is $M+1$.

In order to ensure that $\mathbf{H}^k$ is well conditioned in case it is equal to (14), we require that the zeros will be far apart from each other. This is obtained by checking that their minimal distance is larger than a specified threshold μ. Furthermore, in order to guarantee global convergence we require that the elements of the diagonal matrix in (14) are bounded below by δ_1 and above by δ_2. This is explained in detail in Appendix IV.

Descent Direction: The third step is the most time consuming step, since it involves solving the linear system of equations

$$\mathbf{H}^k\mathbf{d}^k = -\mathbf{g}^k \tag{24}$$

where $\mathbf{g}^k$ is the gradient at $\mathbf{a}^k$ and $\mathbf{d}^k$ is the descent direction. Usually, solving a linear system 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 (24) 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, (24) can be solved in $O(M^2)$ steps. Specifically, in this case (24) becomes ${\mathbf{V}^k}^T\mathbf{D}^k\mathbf{V}^k\mathbf{d}^k = -\mathbf{g}^k$. Denoting $\mathbf{f}^k = \mathbf{D}^k\mathbf{V}^k\mathbf{d}^k$ we first solve the equation ${\mathbf{V}^k}^T\mathbf{f}^k = -\mathbf{g}^k$, and then solve $\mathbf{V}^k\mathbf{d}^k = \mathbf{D}^{k-1}\mathbf{f}^k$. Due to the fact that $\mathbf{V}^k$ is a Vandermonde-like matrix, 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 (24) 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 (24) in $O(M^2)$ operations.

Stopping Criterion: The algorithm stops when $|(\mathbf{d}^k)^T\mathbf{g}^k|$ is less than a predetermined threshold, ϵ.

Step Size Selection: The step size, γ^k, is determined according to the Armijo rule given by

$$\frac{||E(\omega,\mathbf{a}^k+\gamma^k\mathbf{d}^k)||_1 - ||E(\omega,\mathbf{a}^k)||_1}{\gamma^k(\mathbf{d}^k)^T\mathbf{g}^k} \geq \sigma \tag{25}$$

such that a sufficient decrease of $||E(\omega,\mathbf{a})||_1$ is guaranteed.

The algorithm is summarized in Table I.

TABLE I
THE L_1 ALGORITHM

Given ω_p, ω_s, N and $W(\omega)$:
Step 1 - Initialization. Set $M = \frac{N}{2}$ and determine an initial vector $\mathbf{a}^1 \in \mathbb{R}^{M+1}$, $\epsilon > 0$, $0 < \sigma < 1/2$, $0 < \beta < 1$, $\delta_1 > 0$, $\delta_2 > 0$, and $\mu > 0$. Set $k = 1$.
Step 2 - Positive-definite matrix determination. Calculate the simple zeros of the error function $\{z_1, \ldots, z_t\}$. Determine a positive definite $(M+1)\times(M+1)$ matrix $\mathbf{H}^k$ according to one of the following cases:
I. If $t = 0$, then set $\mathbf{H}^k = \mathbf{I}$.
Otherwise, construct the matrix $\mathbf{D}^k$ as given in Theorem 3.
II. If $d_{min} < \delta_1$ or $\delta_2 < d_{max}$, then set $\mathbf{H}^k = \mathbf{I}$.
Otherwise, form the matrix $\mathbf{V}^k$ as defined in Theorem 3.
III. 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 $\mathbf{H}^k = {\mathbf{V}^k}^T\mathbf{D}^k\mathbf{V}^k$.
IV. 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 ${\mathbf{V}^k}^T\mathbf{D}^k\mathbf{V}^k$ [39]. Set $\mathbf{H}^k = {\mathbf{V}^k}^T\mathbf{D}^k\mathbf{V}^k + \mathbf{C}^k$, where $\mathbf{C}^k$ is the matrix added to the Hessian by the Cholesky decomposition algorithm.
Step 3 - Descent Direction. Compute the $(M+1)$-dimensional vector $\mathbf{g}^k$ whose nth component is given by (12). The current descent direction $\mathbf{d}^k$ is the unique solution of (24).
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 $\max\{1, \beta, \beta^2, \ldots\}$ such that (25) holds.
Step 6 - Updating. Set $\mathbf{a}^{k+1} = \mathbf{a}^k + \gamma^k\mathbf{d}^k$, $k = k+1$, and go to Step 2.

B. 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 IV. ■

Theorem 6: Let $\mathbf{a}^*$ be a minimizer of (11). If $E(\omega,\mathbf{a}^*)$ changes sign $M+1$ times in Ω, then the algorithm in Table I converges at a second order rate.

Note the condition in Theorem 6 is equivalent to the uniqueness of $\mathbf{a}^*$.

Proof: In order to prove that the rate of convergence is second order, we need to show that the Hessian is Lipschitz continuous in a neighborhood of an optimum point $\mathbf{a}^*$, and that there exists a $k_0 \geq 1$ such that in (25) $\gamma^k = 1$ for all $k \geq k_0$ [42]. Showing that $\gamma^k = 1$ for $k \geq k_0$ requires that $\nabla^2||E(\omega,\mathbf{a}^*)||_1$ is positive definite [42], so let us first prove that this is the case when the solution is unique.

If $E(\omega,\mathbf{a}^*)$ changes sign $M+1$ times in Ω, i.e., it has $M+1$ simple zeros in Ω, then 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.

The proof that $\nabla^2||E(\omega,\mathbf{a})||_1$ is Lipschitz continuous in the vicinity of $\mathbf{a}^*$ is given in 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 next compare our algorithm with the Remez exchange method for designing equiripple filters.

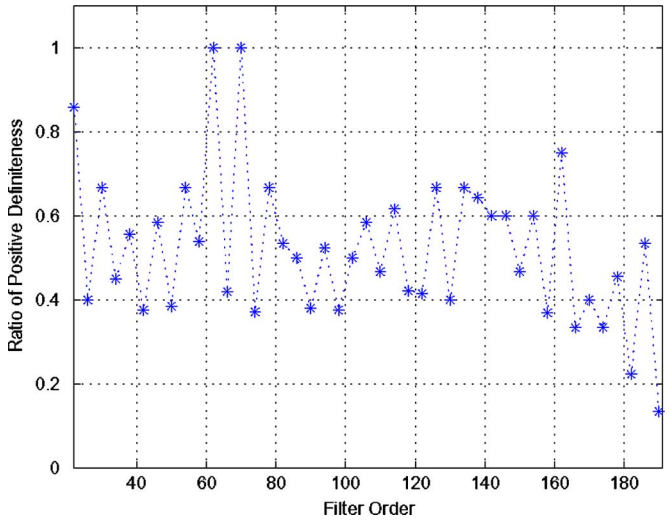

Fig. 2. Ratio of the time an $O(M^2)$ algorithm was used.

V. 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 [43]. Simulation results show that in most cases, the L_1 solution is indeed unique.

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 $O(M^2)$ flops when the Hessian matrix is positive definite. Fig. 2 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. 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 ratios on the graph are taken to be the average of these ten. The average ratio is 0.53, which implies that half of the time the solution is achieved in $O(M^2)$. Nevertheless, the overall complexity of our algorithm is $0.5O(M^2)+0.5O(M^3)=O(M^3)$, while the complexity of Remez algorithm is always $O(M^2)$. The number of iterations required to achieve a certain degree of accuracy in both algorithms was the same.

The complexity of our method can be reduced to $O(M^2)$, if we take $\mathbf{H}^k$ to be the identity matrix when the Hessian is not positive definite. However, in this case the number of iterations is usually increased, and the running time is not always improved.

It is interesting to note that our algorithm is also applicable in the case where $\Omega=[0,\pi]$, that is it can be applied to the weighted approximation of a discontinuous function. This is in contrast with the Remez method, which relies on the continuity of the approximated function. Since our algorithm is based on the calculation of the first and second derivatives of the L_1 norm of the error, all we need to check is their existence in the discontinuous case.

Since the first derivative does not depend on the continuity of $D(\omega)$, its existence is not affected by the inclusion of the cutoff frequency. Thus, as in the disjoint case the L_1 norm of the error is differentiable for all vectors in $\mathbb{R}^{M+1}$ except for $\mathbf{a}=(0,0,\ldots,0)$ and $\mathbf{a}=(1,0,\ldots,0)$.

As we saw in Theorem 3, the existence of the Hessian and its expression are dependent on the first derivative of $E(\omega,\mathbf{a})$ with respect to ω. Specifically, the diagonal elements of $\mathbf{D}$ in (14) are equal to

$$d_i=(2W(z_i))/(|E^{'}(z_i,\mathbf{a})|)$$

where $E^{'}(z_i,\mathbf{a})=A^{'}(z_i)-D^{'}(z_i)$. Thus, as long as z_i is not equal to the cutoff frequency ω_c, d_i is well defined (when z_i is also a simple zero, of course) and therefore the Hessian matrix is defined. If, however, $z_i=\omega_c$ for some i, then we can no longer use (14) and substitute it with the identity matrix. Thus, the case when $\Omega=[0,\pi]$ imposes no restrictions and our algorithm includes this setting as well.

VI. Design Examples

We now illustrate the properties of L_1 filters through two examples.

Example 1—Low-Pass Filter: We first consider the low-pass filter described in (4) with $\omega_c=0.485\,\pi$ rad. We choose $\omega_p=0.474\pi,\omega_s=0.493\pi,N=64$, and $W(\omega)=1$. Fig. 3 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 Fig. 4. 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. Fig. 5(a) 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. Fig. 5(b) 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.

Thus, we see that L_1 filters have flatter response than the least-squares (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 applications 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 a good choice.

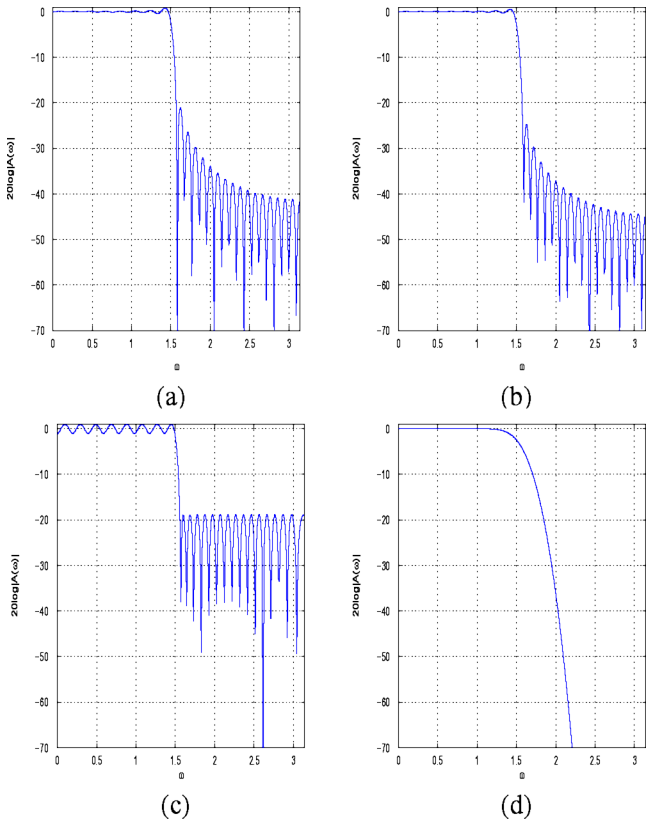

Fig. 3. Low-pass filter of length 65: (a) Kaiser; (b) least-squares; (c) minimax; (d) maximally-flat.

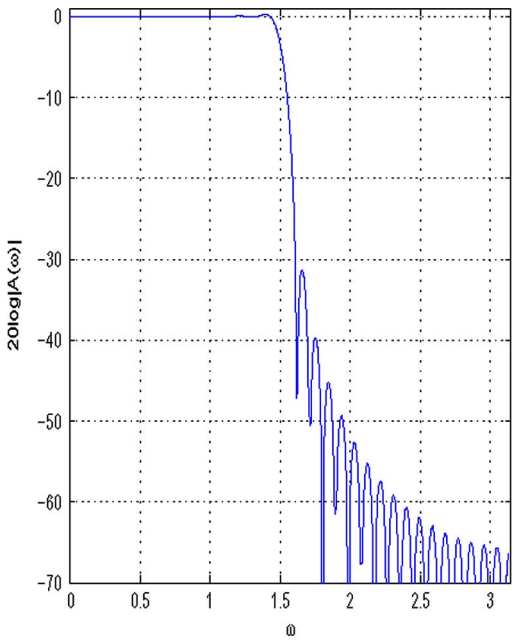

Fig. 4. The 65 length L_1 filter.

Example 2—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. It is easy to see that the conditions on the gradient and Hessian

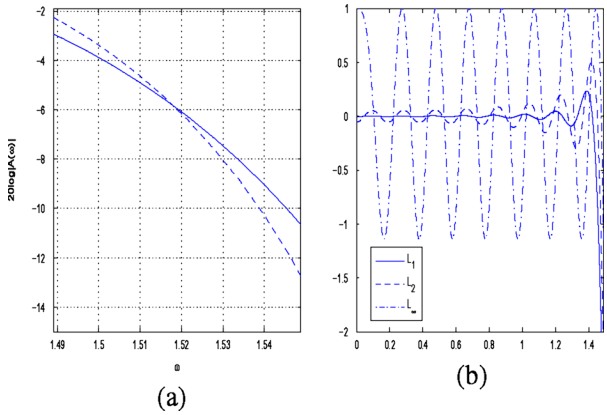

Fig. 5. (a) Transition bands of the L_1 (solid line) and L_2 (dashed line). (b) Enlarged part of the passband.

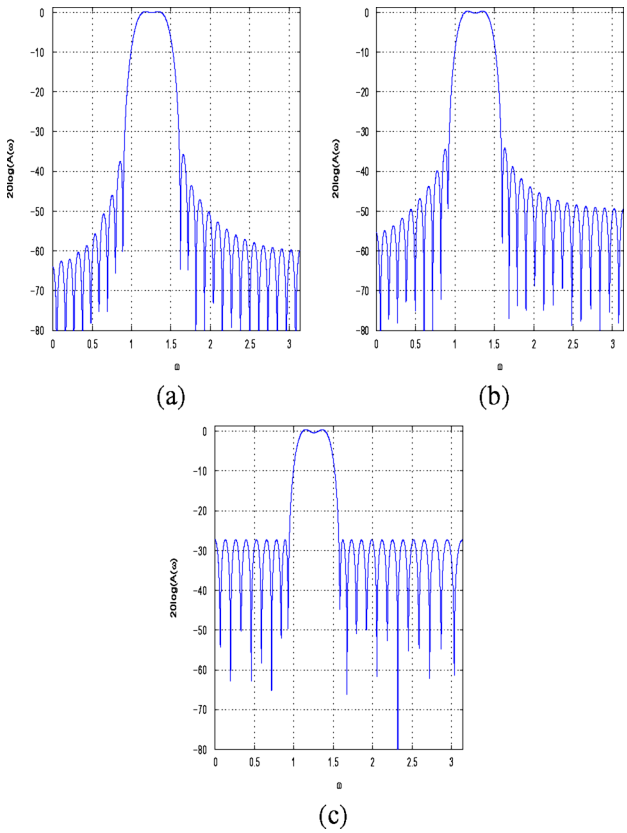

Fig. 6. Bandpass filter of length 35: (a) L_1; (b) least-squares; (c) minimax.

remain the same in this case as well. In addition, global convergence is also guaranteed. However, the condition of the local rate of convergence may change.

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. Fig. 6 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 approach offers the strongest attenuation in the stopband. The running time of our method was the same in both examples.

VII. 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 the gap by describing both the theory and the algorithmic aspects of L_1 filter design.

Optimal L_1 filters result in a very flat response in the passband and stopband while retaining a transition band which is comparable to that of the least-squares. Thus, L_1 filters may 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. In particular, 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 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.

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 may not hold in the multiband case. In particular, the characterization in Theorem 2 and the uniqueness condition change when considering multiband filters. Nevertheless, the algorithm of Table I can still be applied to compute the coefficients of a multiband filter, but different conditions on its convergence rate will be imposed.

The design approach we presented assumes that the specifications are given in terms of the filter order and the passband and stopband. Sometimes the specifications are given in terms of the passband and stopband and the maximal deviations in these bands. We believe that a formula relating the latter specification to the filter order can be developed for L_1 filters in a similar manner to that of minimax filters. This is a topic for future research.

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

Appendix I
The Optimal $N = 0$ Filter

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

$$\begin{aligned}\|E(\omega,a)\|_1 &= \int_\Omega W(\omega)|a - D(\omega)\,|\,d\omega \\ &= \int_0^{\omega_p} W(\omega)|a-1|d\omega \\ &\quad + \int_{\omega_s}^{\pi} W(\omega)|a|d\omega \end{aligned} \tag{26}$$

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

$$\|E(\omega,a)\|_1 = \begin{cases} (A+B)a - A, & a \geq 1 \\ (B-A)a + A, & 0 < a < 1 \\ -(A+B)a + A & a \leq 0. \end{cases} \tag{27}$$

Clearly, the optimal solution a^* depends on A and B. If $A > B$ then $a^* = 1$, whereas if $A < B$, then $a^* = 0$. 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.

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 to the contrary that $E(\omega,\mathbf{a})$ has more than $M+1$ sign changes in Ω. Since each sign change is also a zero of $E(\omega,\mathbf{a}), E(\omega,\mathbf{a})$ 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

$$E^{'}(\omega,\mathbf{a}) = \sum_{k=1}^{M} k\mathbf{a}_k \sin k\omega = A^{'}(\omega) \tag{28}$$

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. 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 ω_p, ω_s and $M = 4$ are given and that the number of sign changes is $L = 3$. Without loss of generality, let us assume that the error changes sign one time in $[0,\omega_p]$, and two times in $[\omega_s,\pi]$ at the points $\{z_1,z_2,z_3\}$, and that $E(0,\mathbf{a}) > 0$. Now, we construct the cosine polynomial of degree 4

$$\begin{aligned} T(\omega) &= (\cos(\omega) - \cos(z_1))(\cos(\omega) - \cos(z_2)) \\ &\quad \times (\cos(\omega) - \cos(z_t))(\cos(\omega) - \cos(z_3)) \end{aligned} \tag{29}$$

where z_t is any number larger than ω_p and smaller than ω_s. 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

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

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 $0\le\omega_1<\omega_p$, $\omega_s<\omega_2\le\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ 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.

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 $0\le\omega_1<\omega_p,\omega_s<\omega_2\le\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in(\omega_s,\omega_2]$. In this case, 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 $<T(\omega),\mathrm{sign}(E(\omega,\mathbf{a}))>\neq 0$.
2) Suppose that there exist ω_1,ω_2, such that $0\le\omega_1<\omega_p,\omega_s<\omega_2\le\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in(\omega_s,\omega_2]$. We treat this case in the exact same manner as the first one.
3) Suppose that there exist ω_1,ω_2, such that $0\le\omega_1<\omega_p,\omega_s<\omega_2\le\pi$, $\mathrm{sign}(E(\omega,\mathbf{a}))=1$ for $\omega\in[\omega_1,\omega_p)$ and $\mathrm{sign}(E(\omega,\mathbf{a}))=-1$ for $\omega\in(\omega_s,\omega_2]$. 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 $A(\omega)\ge 1$ for $\omega\in[\omega_1,\omega_p)$ and $A(\omega)\le 0$ for $\omega\in(\omega_s,\omega_2]$, then $A(\omega)$ must cross zero in $[\omega_p,\omega_s]$ adding one more zero to $A'(\omega)$. In total, $A'(\omega)$ has M zeros, which means that $A(\omega)$ is identically zero.

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 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 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.

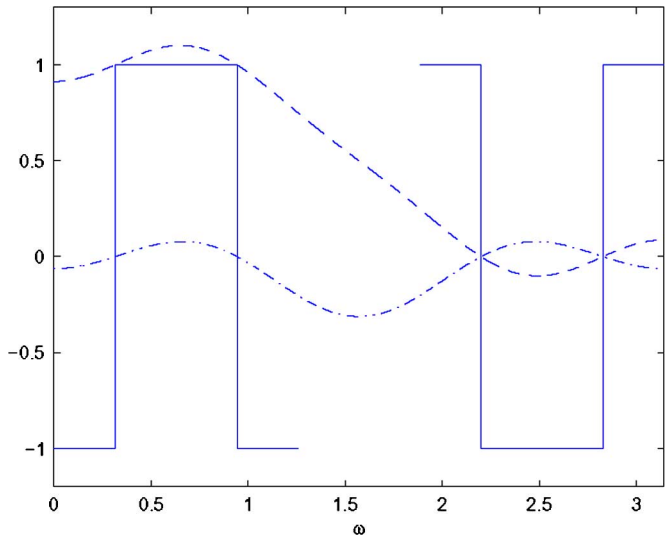

Fig. 7. $A(\omega)$ (dashed line), $\mathrm{sign}(E(\omega,\mathbf{a}))$ (solid line), and $T(\omega)$ (dashed–dotted line), $M=4$.

By Lemma 2, we know the shape of $\mathrm{sign}(E(\omega,\mathbf{a}))$. We assume to the contrary that $E(\omega,\mathbf{a})$ has $M=4$ sign changes in Ω and denote the points of sign 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. 7 in dashed and solid respectively.

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

$$T(\omega)=(\cos(\omega)-\cos(z_1))(\cos(\omega)-\cos(z_2))\times(\cos(\omega)-\cos(z_3))(\cos(\omega)-\cos(z_4)) \quad (30)$$

which is shown in dashed–dotted lines in Fig. 7. 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 $A(\omega)$ in 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 $A(\omega)$ then 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

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

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

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

where $P(\omega) > 0$ and $Q(\omega) < 0$ 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 $\epsilon_1 = p/2Q$. Now, $\epsilon_1 > 0$, and for all $\omega \in [0.6\pi, z_3]$ we have

$$A(\omega) + \epsilon_1 T(\omega) \geq (\cos(\omega) - \cos(z_3))(P(\omega) - p/2 > 0.$$

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

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

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 $\epsilon_2 = (r)/(2S)$. Now, $\epsilon_2 > 0$, and for all $\omega \in [z_3, z_4]$ we have

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

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 $A_1(\omega) = A(\omega) + \epsilon T(\omega)$ has 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 [44]: If $\mathbf{A} \in \mathbb{R}^{n\times n}$ is positive definite, and $\mathbf{X} \in \mathbb{R}^{n\times n}$ is nonsingular, then

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

Lemma 4 [45]: If $|||\cdot|||$ is any matrix norm and $\mathbf{A}$ is an $n\times n$ matrix, then $\lambda_{\max}(\mathbf{A}) \leq |||\mathbf{A}|||$.

Lemma 5: Let $\{z_0, z_1, \ldots, z_n\}$ 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

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

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

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

Now

$$\begin{aligned}|(\mathbf{B}\mathbf{B}^T)_{ij}| &= \left|\sum_{k=1}^{n+1}\cos((i-1)z_{k-1})\cos((j-1)z_{k-1})\right| \\ &\leq \sum_{k=1}^{n+1}|\cos((i-1)z_{k-1})\cos((j-1)z_{k-1})| \\ &\leq n+1 \end{aligned} \tag{31}$$

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

$$\begin{aligned}|\det(\mathbf{B})| = \prod_{k=1}^{n+1}\sigma_k(\mathbf{B}) &\leq \sigma_{\min}(\mathbf{B})\sigma^n_{\max}(\mathbf{B}) \\ &\leq \sigma_{\min}(\mathbf{B})(n+1)^n \end{aligned} \tag{32}$$

where in the last inequality we used (31). In [46] it is shown that

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

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

$$\begin{aligned}|det(\mathbf{B})| &\geq \left|\prod_{i=0}^{n-1}\prod_{i>j}(\cos(z_i) - \cos(z_j))\right| \\ &\geq \left(\min_{i,j}|(\cos(z_i) - \cos(z_j))|\right)^{\frac{n(n-1)}{2}}. \end{aligned} \tag{33}$$

Combining (32) and (33) proves the second inequality. ■

We now prove Theorem 5. According to [37], we need to find two positive numbers λ_1 and λ_2, such that

$$\lambda_1 \leq \lambda_{\min}(\mathbf{H}^k) \quad \text{and} \quad \lambda_{\max}(\mathbf{H}^k) \leq \lambda_2. \tag{34}$$

We claim that the constants δ_1, δ_2 and μ, which are determined in the first step of the algorithm satisfy (34). 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 $\lambda_{\min}(\mathbf{H}^k) = \lambda_{\max}(\mathbf{H}^k) = 1$. Thus, any choice of constants such that $\lambda_1 < 1$ and $1 < \lambda_2$ will satisfy (34).

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

$$\begin{aligned}\delta_1 \leq \lambda_{\min}(\mathbf{D}^k), \quad \lambda_{\max}\left(\mathbf{D}^k\right) \\ \leq \delta_2, \mu \leq \min_{i,j}|\cos(z_i) - \cos(z_j)|. \end{aligned} \tag{35}$$

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 that 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 [36]. Applying Lemma 3 with $\mathbf{A}=\mathbf{D}^k$ and $\mathbf{X}=\mathbf{V}^k$, we obtain

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

From Lemma 3 with $\mathbf{B}=\mathbf{V}^k$ and (33) we therefore have

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

Thus, the choice of $\lambda_1=\delta_1(\mu^{M(M-1)})/((M+1)^{2M})$ satisfies the lower bound in (34). To obtain λ_2 note that by Lemma 3, we have

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

However, from Lemma 5 and (35) we get that

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

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 results in a bounded condition number of $\mathbf{H}^k$ [47].

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 x and $\mathbf{y}$ in $B(\mathbf{a}^*,\epsilon)$

$$|||\mathbf{H}(\mathbf{y})-\mathbf{H}(\mathrm{x})|||\leq L||\mathbf{y}-\mathrm{x}|| \tag{36}$$

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

$$|||\mathbf{A}|||=(M+1)\max_{i,j}|\mathbf{A}_{ij}|. \tag{37}$$

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

$$\begin{aligned}&|||\mathbf{H}(\mathbf{y})-\mathbf{H}(\mathbf{x})|||\\&\quad=(M+1)\max_{i,j}\left|\sum_{m=1}^{t}g_{ij}(z_m^x,\mathbf{x})-g_{ij}(z_m^y,\mathbf{y})\right|\\&\quad\leq(M+1)t\left|g_{i^*j^*}(z_l^x,\mathbf{x})-g_{i^*j^*}(z_l^y,\mathbf{y})\right|.\end{aligned} \tag{38}$$

The first equality is a substitution of the definition of $\mathbf{H}$ in (37) using the function $g_{ij}(z,\mathbf{x})$. In the second one we denote the indexes 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 (38) 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, $|E^{'}(z,\mathbf{x})|=E^{'}(z,\mathbf{x})$ 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 $\partial z/\partial\mathbf{x}_j$ equals $-\cos(jz)/E^{'}(z,\mathbf{x})$, 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 (38) we conclude that

$$|||\mathbf{H}(\mathbf{y})-\mathbf{H}(\mathbf{x})|||\leq L||\mathbf{y}-\mathbf{x}|| \tag{39}$$

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

Acknowledgment

The authors would like to thank the anonymous reviewers for their helpful comments.

References

[1] T. W. Parks and C. S. Burrus, *Digital Filter Design*. New York: Wiley-Interscience, 1987.
[2] B. Porat, *A Course in Digital Signal Processing*. New York: Wiley, 1997.
[3] J. W. Adams, "FIR digital filters with least-squares stopbands subject to peak-gain constraints," *IEEE Trans. Circuits Syst. Theory*, vol. 38, no. 4, pp. 376–388, Apr. 1991.
[4] C. S. Burrus and A. Barreto, "Least p-power error design of filters," in *Proc. IEEE Int. Symp. Circuits. Syst. (ISCAS)*, , 1992, pp. 545–548.
[5] D. W. Tufts and J. T. Francis, "Designing digital low-pass filters—comparison of some methods and criteria," *IEEE Trans. Audio Electroacoust.*, vol. AE-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. Circuits Syst.*, vol. CAS-22, no. 12, pp. 943–953, Dec. 1975.
[7] C. S. Burrus, A. W. Soewito, and R. A. Gopinath, "Least squared error FIR filter design with transition bands," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1327–1340, Jun. 1992.
[8] W. C. Kellogg, "Time domain design of nonrecursive least mean-square digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AE-20, no. 2, pp. 155–158, Jun. 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. CAS-30, no. 10, pp. 723–739, Oct. 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 Process.*, vol. 46, no. 6, pp. 1485–1493, Jun. 1998.
[11] E. Z. Psarakis, "A weighted LJ-based method for the design of arbitrary one-dimensional FIR digital filters," *Signal Process.*, vol. 86, pp. 937–950, 2006.

[12] 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, no. 1, pp. 11–23, Jan. 1987.
[13] J. F. Kaiser, "Design methods for sampled data filters," in *Proc. Ist Allerton Conf. Circuit Syst. Theory*, 1963, pp. 211–236.
[14] H. D. Helms, "Nonrecursive digital filters: Design methods for achieving specifications on frequency response," *IEEE Trans. Audio Electroacoust.*, vol. AE-16, no. 3, pp. 336–342, Sep. 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, Apr. 1975.
[16] J. H. McClellan and T. W. Parks, "A unified approach to the design of optimum FIR linear-phase digital filter," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 697–701, Nov. 1973.
[17] H. D. Helms, "Digital filters with equiripple or minimax responses," *IEEE Trans. Audio Electroacoust.*, vol. AE-19, no. 1, pp. 87–93, Mar. 1971.
[18] O. Hermann, "Design of nonrecursive digital niters 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, no. 10, pp. 740–750, Oct. 1983.
[20] E. Z. Psarakis and G. V. Moustakides, "A robust initialization scheme for the Remez exchange algorithm," *IEEE Signal Process. Lett.*, vol. 10, no. 1, pp. 1–3, Jan. 2003.
[21] P. P. Vaidyanathan, "Efficient and mutiplierless design of FIR filters with very sharp cutoff via maximally flat building blocks," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 236–244, Mar. 1985.
[22] L. R. Rajayopal and S. C. D. Roy, "Design of maximally flat FIR filters using the Bernstein polynomial," *IEEE Trans. Circuits Syst.*, vol. CS-34, no. 12, pp. 1587–1590, Dec. 1987.
[23] T. Cooklev and A. Nishihara, "Maximally flat FIR filters," in *Proc. IEEE Int. Symp. Circuits Systems (ISCAS)*, 1993, vol. 1, pp. 96–99.
[24] P. Bloomfield, *Least Absolute Deviations: Theory, Applications, and Algorithms*. Boston, MA: Birkhäuser, 1983.
[25] C. K. Chen and J. H. Lee, "Design of high-order digital differentiatiors using LI error criteria," *IEEE Trans. Circuits Syst., Analog Digit. Signal Process.*, vol. 42, pp. 287–291, Apr. 1995.
[26] C. S. Burrus, J. A. Barreto, and I. Selesnick, "Iterative reweighted least-squares design of filters," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2926–2936, Nov. 1994.
[27] W. S. Yu, L. K. Fong, and K. C. Chang, "An H approximation based method for synthesizing FIR details," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 39, pp. 578–561, Aug. 1992.
[28] L. D. Grossmann and Y. C. Eldar, "The design of optimal LI linear phase digital FIR filters," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, vol. 3, pp. 884–887, 2006.
[29] E. Moskona and E. Saff, "Gibbs phenonmenori for best LI trigonometric polynomial approximation," *Const. Approx.*, vol. 11, pp. 391–416, 1995.
[30] I. Barrodale and F. D. K. Roberts, "An improved algorithm for discrete l1 linear approximation," *S1AM J. Numer. Anal.*, vol. 10, pp. 839–848, 1973.
[31] C. A. Zala, I. Barrodale, and J. S. Kennedy, "High-resolution signal and noise field estimation using the L1 (least absolute value) norm," *IEEE J. Ocean. Eng.*, vol. 12, no. 1, pp. 253–263, Jan. 1987.
[32] 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.
[33] K. H. Usow, "On L1 approximation H: Computation for discrete function and discretization effects," *SIAM J. Numer. Anal.*, vol. 4, pp. 233–244, 1967.
[34] K. H. Usow, "On L1 approximation I: Computation for continuous function and continuous dependence," *SIAM J. Numer. Anal.*, vol. 4, pp. 70–88, 1967.
[35] S. G. Nash and A. Sofer, *Linear and Nonlinear Programming*, International ed. New York: McGraw-Hill, 1996.
[36] J. R. Rice, *The Approximation of Functions*. Reading, MA: Addison-Wesley, 1964, vol. I.
[37] G. A. Watson, "An algorithm for linear LI approximation of continuous functions," *IMA J. Numer. Anal.*, vol. 1, pp. 157–167, 1981.
[38] A. Pinkus, *On LI Approximation*. New York: McGraw-Hill, 1966.
[39] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. London, U.K.: Academic, 1981.
[40] M. Lang and B. C. Frenzel, "Polynomial root finding," *IEEE Signal Process. Lett.*, vol. 1, no. 10, pp. 141–143, Oct. 1994.
[41] N. J. Higham, "Fast solution of Vandermonde-like systems involving orthogonal polynomials," *IMA J. Numer. Anal.*, vol. 8, pp. 473–486, 1988.
[42] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1999.
[43] M. J. D. Powell, *Approximation Theory and Methods*. Cambridge, U.K.: Cambridge Univ. Press, 1981.
[44] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins Univ. Press, 1997.
[45] ——, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[46] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*. Philadelphia, PA: SIAM, 1998.
[47] J. J. More and D. C. Sorensen, "Newton's method," *Studies Numer. Anal.*, vol. 24, pp. 29–82, 1984.

Liron D. Grossmann received the B.Sc. degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel in 2002. He is currently working towards the M.Sc. degree in electrical engineering at the Technion and the B.Sc. degree in the Department of Mathematics at the Technion.

His research area include filter design methods, functional analysis, and approximation theory and its applications to image processing and communication.

Yonina C. Eldar (S'98–M'02) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2001.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently an Associate Professor in the Department of Electrical Engineering at the Technion—Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT. Her research interests are in the general areas of signal processing, statistical signal processing, and computational biology.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, and in 2007 the Henry Taub Prize for Excellence in Research. She is a member of the IEEE Signal Processing Theory and Methods technical committee, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the *EURASIP Journal of Signal Processing*, and on the Editorial Board of *Foundations and Trends in Signal Processing*.