THE DESIGN OF OPTIMAL L_1 LINEAR PHASE FIR DIGITAL FILTERS

Liron D. Grosssmann *Yonina C. Eldar*

Department of Electrical Engineering
Technion—Israel Institute of Technology
Haifa 32000, Israel.
Email: lirongr@tx.technion.ac.il, yonina@ee.technion.ac.il.

ABSTRACT

This paper addresses the problem of designing linear-phase finite impulse response digital filters using an L_1 optimality criterion. An efficient procedure for the design of such filters is proposed and it is shown that the resulting filter has several desirable properties. Specifically, the L_1 filter possesses flat passbands and stopbands while keeping the transition band comparable to that of the least-squares filters. A comparison with existing methods is made, and it is shown that our method is as efficient as the celebrated Remez exchange algorithm.

1. INTRODUCTION

The problem of designing a linear-phase finite impulse response (FIR) filter has long been recognized as an approximation problem, where an ideal frequency response, usually a discontinuous one, is approximated by a finite number of smooth functions. The design process of such filters typically consists of two contradictory requirements [1]. On the one hand, the FIR filter should preserve the discontinuous behavior of the ideal response, i.e. sharp transitions, while on the other hand, the filter should be as flat as possible in the passbands and stopbands. One way to handle this tradeoff is by changing the measure of "goodness", under which the approximation is carried out.

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 [1]. At the one extreme are the filters which are optimal in the minimax (also called the Chebyshev norm) sense, and exhibit a sharp transition from the passband to stopband, but result in an equiripple behavior between them [2]. At the other extreme are the maximally flat filters, which offer a very smooth passband and stopband at the expense of a wide transition band [3]. The least-squares, another common used criterion, may be viewed as a tradeoff between these two extremes.

It is worthwhile to mention that other criteria for designing FIR filters have also been proposed in the literature. For example, Adams suggested a combined criterion for achieving a tradeoff between the least-squares and the minimax approaches [4]. The use of L_p norm, $2 \leq p \leq \infty$ has also been suggested [5].

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 discontinuous frequency responses (multiband filters in the general case). The use of L_1 norm as a measure of goodness is very common in several engineering applications [6], but 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 [7, 8, 9]. In [7], the design of high-order differentiators was considered, and in [9] an arbitrary amplitude function was designed using the L_1 criterion. A general algorithm for the approximation under L_p was proposed in [8], but convergence is not guaranteed for $p = 1$, and when exists is often very slow. In all three papers, however, 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 should be made 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.

The paper is organized as follows. In Section 2 we mathematically formulate the low-pass filter design problem as an approximation problem using the L_1 criterion. We also motivate the use of the L_1 measure for handling the inherent tradeoffs in the problem. In Section 3 we give the necessary mathematical background needed to characterize the best L_1 filter. A simple Newton-type algorithm is proposed in Section 4. Its convergence and computational complexity are discussed in Section 5, in which we also compare its efficiency to the well known Remez exchange algorithm for the design of minimax filters. Finally, a design example is given in Section 6, with an emphasis on the properties of L_1 filters with respect to other existing design methods.

2. PROBLEM FORMULATION AND MOTIVATION

We consider the problem of designing a low-pass Nth order FIR filter with impulse response $\{\mathbf{h}_n, 0 \leq n \leq N\}$ to approximate the ideal response

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

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

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

sponse $\mathbf{h}_n$:

$$H(\omega) = \sum_{n=0}^{N} \mathbf{h}_n e^{-jwn}. \tag{2}$$

For simplicity, we consider symmetric odd length filters (known as type-1 filters), in which case $H(\omega)$ can be written as

$$H(\omega) = A(\omega)e^{-jM\omega} \tag{3}$$

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

$$\begin{aligned} A(\omega) &= \mathbf{h}_M + \sum_{n=1}^{M} 2\mathbf{h}_{M-n}\cos(n\omega) \\ &\triangleq \sum_{n=0}^{M} \mathbf{a}_n \cos(n\omega). \end{aligned} \tag{4}$$

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 it causal. Thus, we wish to approximate $D(\omega)$ by a linear combination of the $M+1$ functions $\{\cos(n\omega), n = 0, ..., M\}$.

We denote the error of the approximation by

$$E(\omega, \mathbf{a}) = A(\omega) - D(\omega). \tag{5}$$

The approximation process of $D(\omega)$ by $A(\omega)$ is essentially a minimization problem, where we wish to find the vector of coefficients, $\mathbf{a} = (\mathbf{a}_0, ..., \mathbf{a}_M)$, which minimizes some goodness criterion defined on $E(\omega, \mathbf{a})$. The two most popular criteria for measuring the approximation error are

1) The weighted least-squares error:

$$\|E(\omega, \mathbf{a})\|_2 = \int_\Omega W(\omega)|E(\omega, \mathbf{a})|^2 d\omega. \tag{6}$$

2) The weighted Chebyshev error:

$$\|E(\omega, \mathbf{a})\|_\infty = \max_{\omega\in\Omega} W(\omega)|E(\omega, \mathbf{a})|. \tag{7}$$

The function $W(\omega)$ is a positive error weighting function, and Ω stands for the union of passbands and stopbands (which are assumed to be disjoint closed intervals in $[0, \pi]$).

Here, we propose minimizing the weighted L_1 criterion

$$\|E(\omega, \mathbf{a})\|_1 = \int_\Omega W(\omega)|E(\omega, \mathbf{a})| d\omega. \tag{8}$$

The motivation for using this choice in connection with the filter design problem is twofold. First, the L_1 measure is very popular in robust estimation methods. Therefore it is intuitively reasonable to expect that when applied to the filter design problem, it would tend to smear the transition band, while at the same time yield less ripples in the passband and stopband. Indeed, as will be shown, filters which are optimal in the L_1 sense possess these properties, and result in a flatter response than the least-squares, at the expense of a small increase in the transition width. Second, one of the main drawbacks of a least-squares approximation of discontinuous functions is the Gibbs phenomenon. It turns out the approximation under L_1 results in a substantially smaller overshoot, leading to better frequency response behavior [10].

Nevertheless, two main issues regarding the L_1 norm form the theoretical and algorithmic difference from the norm L_2 and L_∞, and make its analysis more complicated and intriguing. First, like the Chebyshev norm, the L_1 measure is usually non-differentiable. This questions the use of standard optimization approaches to minimize (8), as most of them use at least the gradient of the norm. In the minimax case, this problem is bypassed via the alternation theorem, which gives rise to the efficient Remez exchange algorithm [2]. The second issue is that of uniqueness. Unlike the L_2 and L_∞ norms, the optimal solution may not be unique. Aside from its theoretical importance, the non-uniqueness may affect the performance as we discuss in the following section.

The next section addresses the problem of L_1 differentiability, where it is shown that in the case of the filter design problem, the norm may be differentiated. This allows us to use gradient methods. Moreover, it is shown that under certain circumstances, the norm is even twice differentiable, enabling the use of second order techniques such as the Newton method. The question of uniqueness will be explored after we introduce the algorithm.

3. MATHEMATICAL BACKGROUND

3.1. Notations

Let Ω denote the set $[0, \omega_p] \cup [\omega_s, \pi]$. For a vector $\mathbf{a} = (\mathbf{a}_0, ..., \mathbf{a}_M) \in \mathbb{R}^{M+1}$ we denote by $Z(\mathbf{a})$ the set

$$Z(\mathbf{a}) = \{\omega \in \Omega | E(\omega, \mathbf{a}) = 0\}. \tag{9}$$

A zero z_1 of $E(\omega, \mathbf{a})$ is called simple if $\frac{\partial E(\omega, \mathbf{a})}{\partial \omega}|_{\omega=z_1} \neq 0$. The sign function of $E(\omega, \mathbf{a})$ is defined as

$$\text{sign}(E(\omega, \mathbf{a})) = \begin{cases} 1 & E(\omega, \mathbf{a}) > 0 \\ 0 & E(\omega, \mathbf{a}) = 0 \\ -1 & E(\omega, \mathbf{a}) < 0. \end{cases} \tag{10}$$

3.2. Differentiability of the L_1 Norm

In general the L_1 norm is non-differentiable. However, the next theorem states a condition under which the norm can be differentiated at a point $\mathbf{a}$:

Theorem 1 (First Derivative). *If $Z(\mathbf{a})$ has zero measure, then the derivative of $\|E(\omega, \mathbf{a})\|_1$ exists at $\mathbf{a}$, and the components of its gradient are given by*

$$\mathbf{g}_n(\mathbf{a}) \triangleq \int_0^\pi W(\omega)\cos(n\omega)\text{sign}(E(\omega, \mathbf{a}))d\omega. \tag{11}$$

As a consequence of this theorem, it can be shown that the L_1 norm in (8) is differentiable for all $\mathbf{a} \in \mathbb{R}^{M+1}$, except for two particular points, $\mathbf{a} = (1, 0, ..., 0)$, and $\mathbf{a} = (0, ..., 0)$; see [11]. However, these points refer to the degenerate case of a constant filter, and therefore can be ignored. The theorem also shows that the gradient at $\mathbf{a}$ depends on the sign function of the error at that point, which depends on the set $Z(\mathbf{a})$. Indeed, the set of zeros at $\mathbf{a}$ plays an analogue role to the set of exterma in the Chebyshev norm case.

The next theorem addresses the question of second differentiability.

Theorem 2 (Second-Order Derivative). *Let $Z(\mathbf{a}) = \{z_1, ..., z_t\}$ be the set of zeros of $E(\omega, \mathbf{a})$ and assume that each zero is simple. Then the Hessian matrix $\|E(\omega, \mathbf{a})\|_1$ is given by*

$$\mathbf{H}(\mathbf{a}) = \mathbf{A}^T\mathbf{D}^{-1}\mathbf{A} \tag{12}$$

where $\mathbf{A}$ *is a* $t \times (M+1)$ *matrix whose ijth element is* $\sqrt{W(z_i)}\cos((j-1)z_i)$*, and* $\mathbf{D} = diag\{d_1, \ldots, d_t\}$ *with* $d_i = \frac{1}{2}|\frac{\partial E(\omega,\mathbf{a})}{\partial \omega}|_{\omega=z_i}|$.

Using these two theorems, a Newton-type algorithm for minimizing (8) is proposed in the following section.

4. THE WEIGHTED L_1 ALGORITHM

The above discussion showed that differentiability (and sometimes twice differentiability) of the L_1 norm is guaranteed in the filter design problem. As a result the minimizing filter may be obtained by applying a modified Newton method [12]. A modified Newton generates a sequence $\mathbf{a}^k$ given by

$$\mathbf{a}^k = \mathbf{a}^{k-1} - \gamma^{k-1}[\mathbf{F}^{k-1}]^{-1}\mathbf{g}^{k-1}, \tag{13}$$

where $\mathbf{g}^{k-1}$ is the gradient at $\mathbf{a}^{k-1}$ given by (11), γ^{k-1} is the step size, and $\mathbf{F}^{k-1}$ equals one of three matrices. It is equal to the Hessian matrix, if it exists and is positive definite, while if it is not positive definite, then the identity matrix multiplied by a positive scalar is added to it to form $\mathbf{F}^{k-1}$. In case the Hessian does not exist it equals the identity matrix to ensure a descent direction.

The following steps describe the algorithm for computing the best weighted L_1 approximation to $D(\omega)$ on Ω.

Step 1 - Initialization. Determine an initial vector $\mathbf{a}^1 \in \mathbb{R}^{M+1}$, $\epsilon > 0$, $0 < \sigma < 1/2$, $0 < \beta < 1$. Set $k = 1$. The initialization of $\mathbf{a}^1$ is described in [11].

Step 2 - Positive-definite matrix determination. Form the matrices $\mathbf{A}^k$ and $\mathbf{D}^k$, as defined in theorem 2, and determine a positive definite $(M+1)\times(M+1)$ matrix $\mathbf{H}_k$ according to one of the following cases. If $t = 0$ or $\mathbf{D}^k$ is singular, then set $\mathbf{H}^k = I$. If $t \geq M+1$, $\mathbf{D}^k$ is non-singular, and $\text{rank}(\mathbf{A}^k) = M+1$, then set $\mathbf{H}^k = (\mathbf{A}^k)^T(\mathbf{D}^k)^{-1}\mathbf{A}^k$. If $t > 0$, $\mathbf{D}^k$ is non-singular and $\text{rank}(\mathbf{A}^k) < M+1$, then set $\mathbf{H}^k = (\mathbf{A}^k)^T(\mathbf{D}^k)^{-1}\mathbf{A}^k + \lambda^k I$, where $\lambda^k > 0$ is given.

Step 3 - Descent Direction. Compute the $(M+1)$-dimensional vector $\mathbf{g}^k$ whose nth component is given by (11). Determine $\mathbf{d}^k$, the current descent direction, which is the unique solution of

$$\mathbf{H}^k\mathbf{d}^k = -\mathbf{g}^k. \tag{14}$$

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

$$T(\mathbf{a}^k, \gamma^k) \geq \sigma, \tag{15}$$

where

$$T(\mathbf{a}^k, \gamma^k) = \frac{\|E(\omega, \mathbf{a}^k + \gamma^k\mathbf{d}^k)\|_1 - \|E(\omega, \mathbf{a}^k)\|_1}{\gamma^k(\mathbf{d}^k)^T\mathbf{g}^k}. \tag{16}$$

Note that the step size is selected in such a way that the weighted L_1 norm of the error is decreased.

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

In the next section, we discuss convergence issues, and the computational complexity of the above algorithm. In addition, a characterization of the unique optimal solution is given, which is reminiscent of the alternation theorem for the minimax design problem.

5. CONVERGENCE, UNIQUENESS AND COMPUTATIONAL COMPLEXITY

We start by summarizing the convergence properties of the our algorithm. The proof of the following theorem is given in [11].

Theorem 3 (Global and Local Convergence). *The proposed algorithm is globally convergent. Furthermore, if the optimal solution is unique, then the algorithm has a second order rate of convergence.*

The Remez algorithm for minimax filters also is also globally convergent and admits a second order rate of convergence [13]. Thus, when uniqueness holds in our problem our algorithm and the Remez mehtod are comparable. The next theorem (whose proof may be found in [11]), which may be viewed as the L_1 analogue of the alternation theorem, states a necessary and sufficient condition on the optimal solution to be unique.

Theorem 4 (Uniqueness). *Let* $A(\omega)$ *be a best (but not necessarily unique) weighted* L_1 *approximation to* $D(\omega)$ *on* Ω *of degree* M*. Then* $E(\omega, \mathbf{a})$ *changes sign either* M *or* $M+1$ *times in* Ω*. The number of sign changes is* $M+1$ *if and only if* $A(\omega)$ *is the unique best approximation.*

Indeed, in our problem the number of the sign changes of the error function plays the role of extrema in the minimax case.

Finally, we note that the computational complexity of the Remez algorithm is dominated by the solution of a linear system of equations. It was shown in [2] that for equiripple filters, the solution of the linear system may be efficiently found in $O(M^2)$ operations (instead of $O(M^3)$). The complexity of our method is also determined by the solution of a linear system of equations (Step 3). We have developed a method, described in [11], for solving our linear system in an $O(M^2)$ when the Hessian matrix is positive definite. Simulation results show that this is often the case. A more thorough complexity analysis is provided in [11].

6. SIMULATIONS RESULTS

In this section, we compare our approach for L_1 filters with existing methods. The L_1 algorithm was coded in Matlab, and simulations show that its running time may be compared with the Remez exchange algorithm. We consider the design of a low-pass filter

$$D(\omega) = \begin{cases} 1, & \omega \in [0, 0.63\pi] \\ 0, & \omega \in (0.63\pi, \pi]. \end{cases} \tag{17}$$

by approximating it with a type 1 linear-phase FIR filter of order $N = 42$. We define the passband region to be $[0, 0.6\pi]$ and the stopband region as $[0.66\pi, \pi]$, with the weighting function equal one. The approximating filters using various methods are shown in Figs. 1-2, in a logarithmic scale.

The figures suggest that the L_1 filters have the following attractive properties. In most of the passband and stopband region it admits a higher degree of flatness than the least-squares method, and much higher than the minimax approach, see Fig. 3 for the magnified passband. The maximum deviations in both the stopband and passband regions are slightly higher than that of the least-squares and the minimax, and are much smaller than that of the maximally flat filter, whose transition band is very wide. In addition, the maximum deviation occurs very close to the discontinuity without resulting in large ripples as in the least-squares filter. Thus, the L_1 filter may provide a suitable tradeoff between the minimax and maximally flat filters, in applications where flat passbands and stopbands are required (corresponding to maximally flat filters), and still a reasonable transition region should be kept.

7. CONCLUSIONS

An efficient algorithm for the design of linear-phase FIR filters, which are optimal in the L_1 sense, has been proposed. The method, which is a modified version of the Newton algorithm, often enjoys a fast convergence rate, which is the same rate of convergence as the Remez exchange algorithm for the design of equiripple filters. In contrast to the pure Newton method, however, it does not require the computation of second derivatives, and can be made very efficient in the case of filter design. Simulation results have been demonstrated, showing that the L_1 filters may be a good choice when, for example, bandlimiting low-pass signals. The method was described for designing low-pass filters, however, extensions to other type of filters is possible.

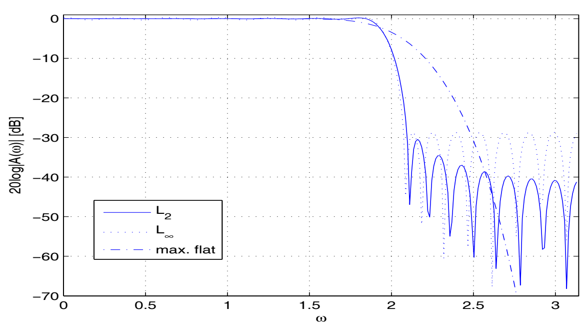

Fig. 1. Low-pass filters (L_2, L_∞, and maximally flat) with $N = 42$.

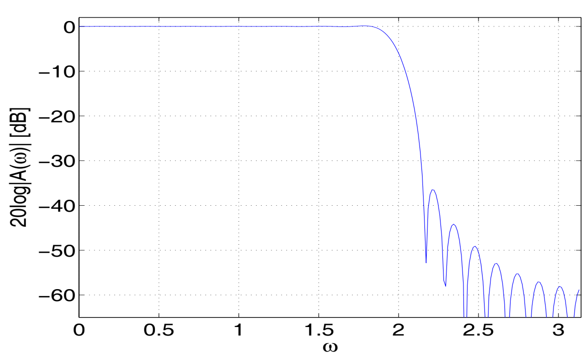

Fig. 2. L_1 low-pass filter, $N = 42$.

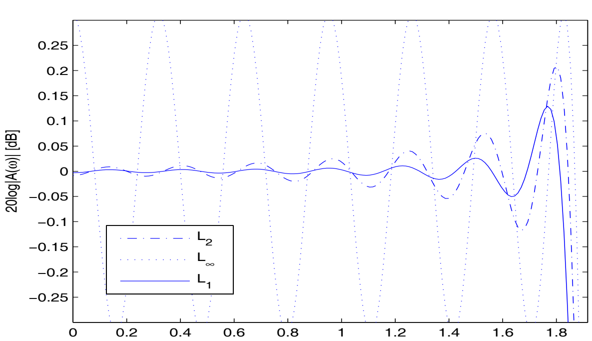

Fig. 3. Enlarged passbands for $N = 42$.

8. REFERENCES

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

[2] 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.

[3] O. Hermann, "On the approximation problem in nonrecursive digital filter design," *IEEE Trans. Circuits Theory*, vol. 18, pp. 411–413, 1971.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

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

[10] A. J. Jerri, *The Gibbs Phenomenon in Fourier Analysis, Splines, and Wavelets Approximations*, Dordrecht: Kluwer, 1998.

[11] L. D. Grossmann and Y. C. Eldar, "An L_1-method for the design of linear-phase FIR digital filters," CCIT Report EE Dept., Technion - Israel Institute of Technology.

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

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