AN L_1 METHOD FOR BANDLIMITING DISCRETE-TIME SIGNALS

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 bandlimiting discrete-time signals using digital filters, which are optimal in an L_1 sense. An efficient procedure for the design of such filters is proposed and it is shown that the resulting filters admit several desired properties.

1. INTRODUCTION

Bandlimiting a discrete-time signal is an important operation in several sampling procedures. Examples include the decimation and interpolation of discrete-time signals, where in the former, the signal must be bandlimited prior to downsampling in order to prevent aliasing. In the latter it is required to use a bandlimiting filter in order to reject the images created by the upsampling process [1].

The common method for bandlimiting a discrete-time signal to a desired bandwidth, is to filter it with a low-pass digital filter (or a band-pass filter in the case of bandpass sampling), having the desired bandwidth. Ideally, we would like to obtain perfect transmission in the passband, and zero transmission in the stopband. However, such ideal filters cannot be implemented in practice [2]. Instead, an approximation to the ideal filter is performed, where different approximation schemes lead to filters having different properties. In order for the approximating filter to be realizable as well, further restrictions are imposed, such as stability and causality. In addition, in bandlimiting it is often important not to distort the original signal in the desired bandwidth, in which case the choice of linear-phase filters is made, resulting in a signal which is a time-delayed version of the original signal. Linear-phase digital filters with real coefficients, which are casual and stable necessarily have a finite impulse response (FIR) [2]. Thus, a natural choice for a bandlimiting filter is a linear-phase FIR.

The design of linear-phase FIR filters has been thoroughly investigated in the literature, where the essence of all methods is to efficiently trade-off between maximum flatness in the passband and stopband, and minimal transition width; see [3] and the many references therein. On one extreme is the minimax approach, which generates filters (known as equiripple filters) with the narrowest transition band for a given number of coefficients, but results in a highly non-flat passband and stopband. On the other extreme is the maximally flat method, which yields filters possessing a high degree of flatness in certain frequencies, but have very wide transition band. Another common approach is the least-squares, which minimizes the energy of the error between the ideal filter and the approximating one. The least-squares method offers some trade-off between the above extremes, in that it results in a flatter response than the minimax, while having a transition band, which is narrower than the maximally flat filters. Nonetheless, it still admits a non-smooth behavior in the passband and stopband, and also possesses an undesirable high overshoot around the edge of the filter.

Motivated by the need to design filters having flat transmission in the pass bandwidth, yet maintaining a reasonable transition band, we consider the problem of approximating the ideal bandlimiting filter using an L_1 criterion. The L_1 optimality criterion is considered to be a good choice in robust applications, and 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. Furthermore, as in the case of the least-squares and the minimax methods, filters designed under the L_1 criterion enjoy the property of being optimal in some sense, as opposed to the window design method, which is typically ad-hoc. In addition, similar to the other two optimal methods (the least-squares and the minimax), L_1 filters can also incorporate a weighting function into the design process, which allows for a more flexible design than the window method and the maximally flat filters [2].

The design of filters under the L_1 criterion has not received much attention in the literature of filter design. We believe that the main reason for this fact is that contrary to the minimax and least-squares design methods, no efficient procedure for the design of L_1 filters exists. Indeed, we are aware of only one work, where digital differentiators are designed using an L_1 criterion [4]. In [4], the optimal solution is accomplished by first sampling the desired frequency response and then solving a discrete L_1 problem, using linear programming methods. The general method of sampling the desired frequency response and its major disadvantages has been investigated in the past [3]. Furthermore, as the sampling grid becomes more dense, the computation of the filter by linear programming becomes very slow.

In 1981 Watson suggested an efficient procedure for approximating continuous functions defined on an interval under the L_1 criterion [5]. The algorithm is a modified version of the Newton method for solving unconstrained minimization problems [6]. In his paper, Watson stated specific conditions for the L_1 problem, under which a second order rate of convergence is guaranteed, which is the same rate of convergence of the Remez exchange algorithm for the minimax case [6], [7]. In contrast to the linear pro-

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

gramming approach, this method does not rely on sampling the approximated function. However, Watson's algorithm is not directly applicable to the filter design problem, since in filter design the desired frequency response is typically approximated over a union of disjoint closed intervals (the passbands and the stopbands). In addition, it is also common to assign weights to the different intervals.

In this paper, we propose a method for approximating a desired bandlimiting digital filter under a weighted L_1 criterion. In order to efficiently achieve the approximated filter, we extend Watson's algorithm to include a union of closed intervals and a weighting function. Similar conditions for guaranteeing second order rate of convergence are derived. Moreover, it is stated that these conditions are satisfied for the filter design problem. Thus, we obtain an efficient algorithm whose complexity and rate of convergence is the same as the Remez exchange algorithm.

The paper is organized as follows. In Section 2, we mathematically formulate the design problem of weighted L_1 measure. In Section 3, we briefly review the mathematical background required for the development of the extended algorithm. Section 4 describes the weighted L_1 algorithm. Finally in Section 5, we demonstrate the method by simulations and discuss its advantages over existing approaches.

2. PROBLEM FORMULATION

We consider the problem of bandlimiting a discrete-time signal to a bandwidth $[-\omega_c, \omega_c]$ using an order N FIR filter with impulse response $\{h_n, 0 \leq n \leq N\}$. The filter is designed 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 response h_n:

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

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 $M = \frac{(N-1)}{2}$, and $A(\omega)$ is the real-valued function

$$\begin{aligned} A(\omega) &= h_M + \sum_{n=1}^{M} 2h_{M-n} \cos(n\omega) \\ &\triangleq \sum_{n=0}^{M} a_n \cos(n\omega). \end{aligned} \tag{4}$$

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 $M+1$ functions, $\{\cos(n\omega), n = 0, ..., M\}$.

Define the error of the approximation by $E(\omega) = D(\omega) - A(\omega)$. The approximation process of $D(\omega)$ by $A(\omega)$ is essentially a minimization problem, where we wish to find the vector of coefficients, $a = (a_0, ..., a_M)$, which minimizes some goodness criterion defined on $E(\omega)$. The two most popular criteria for measuring the approximation error, $E(\omega)$, are

1) The weighted least-squares error:

$$\|E(\omega)\|_2 = \int_\Omega W(\omega)|D(\omega) - A(\omega)|^2 d\omega. \tag{5}$$

2) The weighted Chebyshev error:

$$\|E(\omega)\|_\infty = \max_{\omega \in \Omega} W(\omega)|D(\omega) - A(\omega)|. \tag{6}$$

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 $[0, \pi]$). In the case of a bandlimiting filter, $\Omega = [0, \omega_p] \cup [\omega_s, \pi]$, where $[0, \omega_p]$ is the band in which we wish to perfectly transmit the original signal, and $[\omega_s, \pi]$ is the band we wish to perfectly reject.

Here, we consider a weighted L_1 criterion defined by

$$\|E(\omega)\|_1 = \int_\Omega W(\omega)|D(\omega) - A(\omega)|d\omega. \tag{7}$$

Note that (5), (6) and (7) define a norm on $E(\omega)$, known as the weighted L_2, L_∞ and L_1, respectively. As such all three problems represent an unconstrained convex minimization problem [6]. This viewpoint will be useful later when the algorithm for the L_1 problem will be discussed.

The solutions to the L_2 and L_∞ criteria are well understood and algorithms for obtaining them have been extensively studied [3]. The solution to the weighted least-squares (L_2) is characterized by being orthogonal to the functions $\{\cos(n\omega), n = 0, ..., M\}$, which leads to a set of $M+1$ linear equations. The solution to the minimax criterion is characterized by the alternation theorem, and is iteratively solved via the Remez algorithm. Although not usually presented in this way, both characterizations can be derived from the point of view of unconstrained convex minimization, by equating the gradient (or subgradient) of the norm with respect to a to zero [8]. In this way, the solution to (7) can also be analytically characterized, which gives rise to a set of $M+1$ nonlinear equations. The essence of the algorithm we propose is to efficiently solve this set of nonlinear equations. In the next section we describe the characterization of the solution to (7), followed by the method which achieves the optimal solution.

3. MATHEMATICAL BACKGROUND

3.1. Notations

Let Ω denote the set $[0, \omega_p] \cup [\omega_s, \pi]$. The weighted L_1 norm of a function $f(\omega)$ on Ω is defined by

$$\|f\|_1 = \int_\Omega W(\omega)|f(\omega)|d\omega. \tag{8}$$

For a vector $a = (a_0, ..., a_M) \in \mathbb{R}^{M+1}$ we denote the approximation error by

$$E(\omega, a) = \sum_{n=0}^{M} a_n \cos(n\omega) - D(\omega), \tag{9}$$

and the set of its zeros by $Z(a)$:

$$Z(a) = \{\omega \in \Omega | E(\omega, a) = 0\}. \tag{10}$$

The sign function of $E(\omega, a)$ is defined as,

$$\text{sign}(E(\omega,a)) = \begin{cases} 1 & E(\omega,a) > 0 \\ 0 & E(\omega,a) = 0 \\ -1 & E(\omega,a) < 0. \end{cases} \quad (11)$$

We denote the inner product between two continuous functions on Ω, f and g, by

$$\langle f, g \rangle = \int_{\Omega} W(\omega) f(\omega) g(\omega) d\omega. \quad (12)$$

3.2. Characterization of the L_1 Optimal Solution

We shall now show that a minimum of (7) is characterized by the property that $\text{sign}(E(\omega, a))$, defined in (11), is orthogonal to the functions $\{\cos(n\omega), n = 0, ..., M\}$. In order for that characterization to be valid, one has to make the assumption that the set of zeros of $E(\omega, a)$, $Z(a)$ in (10), has zero measure. Such is the case, when for example, the number of zeros of $E(\omega, a)$ is finite. Fortunately, this is also the case in our problem too, since we approximate a piecewise constant function, $D(\omega)$, by a trigonometric polynomial of degree M, $A(\omega)$, and therefore the number of zeros in each interval of Ω cannot exceed M, unless $A(\omega)$ is a constant. We can therefore assume that $Z(a)$ is of zero measure.

Theorem 1. *Suppose that $Z(a)$ has zero measure. Then, a vector $a \in \mathbb{R}^{M+1}$ minimizes (7) if and only if*

$$g_n(a) \triangleq \langle \cos(n\omega), \text{sign}(E(\omega, a)) \rangle = 0, \; n = 0, ..., M. \quad (13)$$

The unweighted version of the theorem may be found in [9].

Proof. Since $Z(a)$ is of zero measure for every a, then $\|E(\omega, a)\|_1$ is differentiable for every a. Therefore, as $\|E(\omega, a)\|_1$ is convex with respect to a, we can compute the gradient of $\|E(\omega, a)\|_1$ and equate it to zero to obtain a necessary and sufficient condition for minimum. Computation of the gradient of $\|E(\omega, a)\|_1$ shows that,

$$\nabla \|E(\omega, a)\|_1 = g(a), \quad (14)$$

where $g(a)$ is an $M + 1$-dimensional vector, whose components, $g_n(a)$, are defined in (13). □

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 $E(\omega, a)$ (instead of $\text{sign}(E(\omega, a))$) is orthogonal to the functions $\{\cos(n\omega), n = 0, ..., M\}$. However, while the orthogonality condition in the least-squares case leads to a set of linear equations, the orthogonality condition in (13) describes a set of $M + 1$ nonlinear equations in $M + 1$ unknowns, the components of a.

The next section describes a descent algorithm for solving (7). A general unconstrained minimization algorithm of a convex function aims to solve the set of usually nonlinear equations defined by the optimality condition that the gradient equals zero. In our case, this set of equations is given by (13). It starts with an initial vector a_1, determines according to a fixed rule a descent direction and a suitable step size, and moves along that direction to a new point, determined by the step size. At the new point a new direction and step size are chosen and the process is repeated [6].

4. THE WEIGHTED L_1 ALGORITHM

A simple algorithm for solving an unconstrained minimization problem is the Newton method, for which second order rate of convergence is guaranteed if the Hessian matrix of the function to be minimized is positive definite in the neighborhood of the solution. Newton's method for unconstrained minimization can be regarded as a version of the Newton method for solving nonlinear equations (which describe the condition that the gradient of the function equals zero), with the requirement that the value of the function decreases with each iteration [6]. Therefore, we can apply Newton method (for unconstrained minimization) to our problem as well, in which case the set of nonlinear equations are given by (13).

Directly applying the Newton method to our problem would yield a sequence a_k given by,

$$a_k = a_{k-1} - [\nabla g(a_{k-1})]^{-1} g(a_{k-1}), \quad (15)$$

where, $g(a_{k-1})$ is given by (13), and $\nabla g(a_{k-1})$ is the Jacobian matrix of $g(a_{k-1})$. However, in our case it is not guaranteed that the Jacobian matrix of the system of equations will be nonsingular for every a_k.

Several modifications of the Newton method were suggested to include the case where the Jacobian matrix is singular. However, in order to guarantee a second order rate of convergence, the requirement for the positive definiteness of the matrix in the neighborhood of the solution still remains. Watson suggested to use one of the modified versions of Newton's method to solve the L_1 approximation of functions defined on an interval by a finite number of linearly independent functions. Watson stated a theorem, which gives explicit conditions on the optimal solution of the L_1 problem, which guarantees the positive definiteness of the Jacobian matrix, and therefore a second order rate of convergence.

The theorem of Watson, however, is not suitable for our problem, as we consider a union of closed intervals instead of a single one, and a weighted L_1 norm. A close inspection on the proof of Watson's theorem suggests that the conditions on the optimal solution can be modified to our case as well in such a way that the second order rate of convergence is preserved. The main result concerns the computation of the Jacobian matrix of $g(a)$ in (13). The components of $g(a)$, $g_n(a)$, are explicit functions of the zeros of $E(\omega, a)$, $Z(a)$, which are themselves functions of a. Thus, by the chain rule, the computation of the Jacobian matrix involves knowing the derivatives of $g_n(a)$ with respect to the zeros in $Z(a)$, in addition to the derivatives of the zeros with respect to a. The result of this computation, i.e. the differentiation of $g_n(a)$ with respect to a is given by the following lemma.

Lemma 1. *Let $a \in \mathbb{R}^{M+1}$, and $Z(a) = \{z_1, ..., z_t\}$ be the set of zeros of $E(\omega, a)$, and assume that each zero is simple. The Jacobian matrix of $g(a)$ in (13) is given by,*

$$\nabla g(a) = A^T D^{-1} A \quad (16)$$

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

Note that D is invertible because we assume that the zeros are simple. The lemma is of great importance, as it provides us with an analytic expression of the Jacobian matrix, and circumvents the

numerical computation of the second derivatives of $\|E(\omega, a)\|_1$, which is required in the general case of Newton's algorithm.

The following steps describe the algorithm for computing the best weighted L_1 approximation to $D(\omega)$ on Ω. The difference from Watson's algorithm is in the choice of the matrix A_k in step 2. Note that these steps follow the general structure of descent algorithms for unconstrained minimization problems [6].

Step 1.- Initialization. Determine an initial vector $a_1 \in \mathbb{R}^{M+1}$,$\epsilon > 0$, $0 < \sigma < 1/2$, $0 < \beta < 1$. Set $k = 1$.

Step 2.- Positive-definite matrix determination. Form the matrices A_k and D_k, as defined in lemma 1, and determine a positive definite $(M+1) \times (M+1)$ matrix H_k according to one of the following cases.

If $t = 0$ or D_k is singular, then set $H_k = I$. In this case, the corresponding direction d_k will be the steepest descent direction. If $t \geq M+1$, D_k is non-singular, and $\mathrm{rank}(A_k) = M+1$, then set $H_k = A_k^T D_k^{-1} A_k$. If $t > 0$, D_k is non-singular and $\mathrm{rank}(A_k) < M+1$, then set $H_k = A_k^T D_k^{-1} A_k + \lambda_k I$, where $\lambda_k > 0$ is given. The determination rule of λ_k is given in [5].

Step 3 - Descent Direction. Compute the $(M+1)$-dimensional vector $g_k = g_k(a_k)$ whose n component is given by (13). Determine d_k, the current descent direction, which is the unique solution of

$$H_k d_k = -g_k. \tag{17}$$

The direction d_k corresponds to the Newton direction when the Jacobian matrix ∇g_k is non-singular.

Step 4 - A Stopping Criterion. If $|d_k^T g_k| < \epsilon$ then stop.

Step 5- Step Size. Determine the step size γ_k to be $\max\{1, \beta, \beta^2, ...\}$ such that

$$T(a_k, \gamma_k) \geq \sigma, \tag{18}$$

where

$$T(a_k, \gamma_k) = \frac{\|E(\omega, a_k + \gamma_k d_k)\|_1 - \|E(\omega, a_k)\|_1}{\gamma_k d_k^T g_k}. \tag{19}$$

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 $a_{k+1} = a_k + \gamma_k d_k$, $k = k+1$, go to Step 2.

The following theorem, which is a generalization of the convergence theorem proved in [5], states conditions under which the algorithm has a second order rate of convergence.

Theorem 2. *Let a^* be the unique minimizer of (7) , and let $Z(a^*)$ be the set of zeros corresponding to a^*. Define the $t \times (M+1)$ matrix A^* and the $t \times t$ diagonal matrix D^* as in Step 2 above. Assume that $t \geq M+1$, $\mathrm{rank}(A^*) = M+1$, and each zero is simple. Then, eventually (18) is satisfied with $\gamma_k = 1$, and the rate of convergence is second order.*

It can be shown that for the design of our bandlimiting filter, all the conditions of the theorem are satisfied, and thus the algorithm will have a second order rate of convergence. In fact, at the optimal solution $t = M+1$, i.e. the approximation error changes sign at exactly $M+1$ points.

The theorem above guarantees a good local convergence rate, but says nothing about the computational load associated with each step of the algorithm. Indeed, in its most general form, certain steps of the algorithm may be computationally demanding. However, in our case, it is possible to facilitate the computational complexity of several stages of the algorithm. Specifically, the inner products which define $g_n(a)$ in (13) of step 2, may be computed analytically, without resorting to numerical integration. In addition, the elements of the diagonal matrix D_k are the absolute value of the derivative of $E(\omega, a_k)$ with respect to ω. In general, this derivative is computed using numerical methods, however, in our case the (i,i) element of D_k can be analytically computed to yield $D_k(i,i) = \frac{1}{2}|\sum_{l=1}^{M} l a_k(l) \sin(l z_i)|$. Step 3 requires solving a system of linear equations and one of the many efficient methods for factorizing a positive definite matrix may be used for that purpose [10]. Thus, it is the solution of the linear system in step 3 and the step size selection in step 5 that comprise the computational complexity of one iteration of the algorithm. Note that the computational complexity of the Remez exchange algorithm is also determined by the solution of a linear system, and therefore each step of the proposed algorithm is comparable with that of Remez.

Finally, a good guess of the initial vector a_1 may accelerate the convergence of the algorithm. For our initial guess, it is common to choose a_1, such that the corresponding $A(\omega)$ interpolates the desired response, $D(\omega)$, at the points

$$z_i = \frac{(2i-1)\pi}{2(M+1)}, \quad i = 1, ..., M+1 \tag{20}$$

see [11]. If some of the points happen to lie in the transition band $[\omega_p, \omega_s]$, then some intermediate values between zero and one are chosen as the value of the desired response, $D(\omega)$.

5. SIMULATIONS RESULTS

In this section, we compare our method 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. The traditional design procedure requires specifying the passband and the stopband regions along with the maximum deviations in these regions. The equiripple filters (optimal in the minimax sense) leads to the smallest filter order. Nevertheless, the equiripple filters, by their definition, spread the maximal error over the entire passbands and stopbands. This property may not be desirable in several applications. For example, it may be acceptable to have a larger deviation in certain regions of the passband at the expense of a better transmission in the rest of the passband. This might be the case when bandlimiting low-pass signals, such as speech signals.

We consider the design of a low-pass filter

$$D(\omega) = \begin{cases} 1, & \omega \in [0, 0.63\pi] \\ 0, & \omega \in (0.63\pi, \pi], \end{cases} \tag{21}$$

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 ideal response together with its approximation using various methods are shown in Figs. 1-3. The approximating filters are shown 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. 4 for the

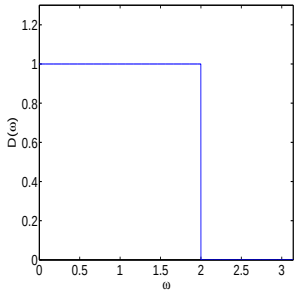

Fig. 1. An example of a desired frequency response.

magnified passband. The maximum deviations in both the stop-

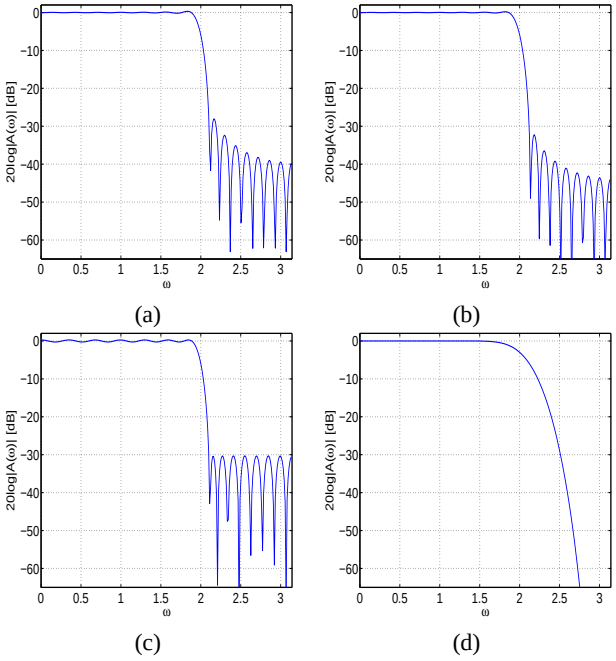

Fig. 2. Low-pass filters with $N = 42$ (a) Kaiser window (b) least-squares (c) minimax (d) maximally-flat.

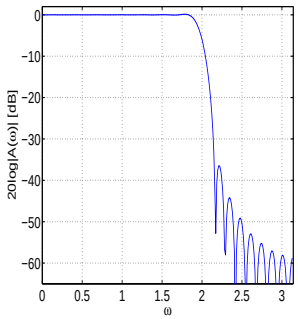

Fig. 3. L_1 low-pass filter, $N = 42$.

band and passband regions is slightly higher than that of the least-squares and the minimax, and is 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 be 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 (the minimax result in the narrowest transition band).

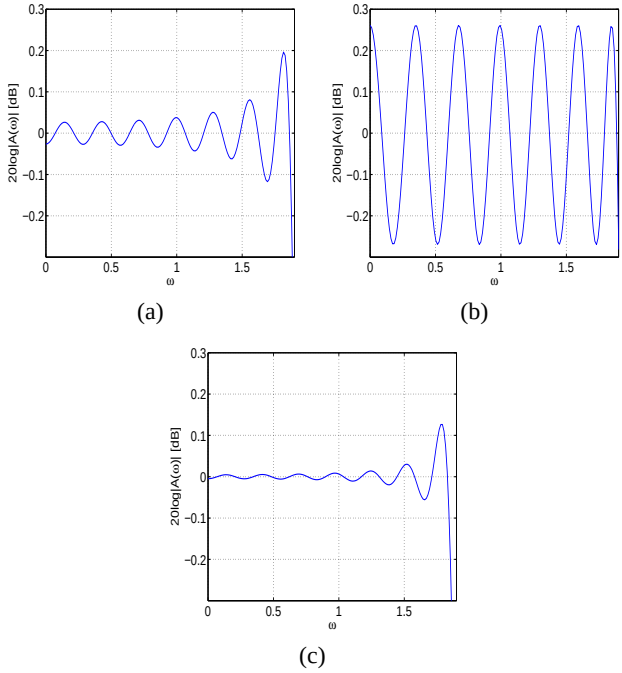

Fig. 4. Enlarged passbands for $N = 42$ (a) least-squares (b) minimax (c) L_1.

6. CONCLUSIONS

An efficient algorithm for the design of linear-phase FIR bandlimiting filters, which are optimal in the L_1 sense, has been proposed in this paper. The method is simple, and based on Watson's algorithm for approximation of functions defined on an interval, which is a Newton type algorithm. It enjoys the fast convergence of the Newton method, i.e. a second order rate, which is the same rate of convergence of the Remez exchange algorithm for the design of equiripple filters. It contrast to the 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.

7. REFERENCES

[1] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*, Englewood-Cliffs, NJ: Prentice-Hall, 1996.

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

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

[4] 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.

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

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

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

[8] D. G. Luenberger, *Optimization by Vector Space Methods*, New York, NY: John Wiley, 1969.

[9] G. A. Watson, *Approximation Theory and Numerical Methods*, London, UK: John Wileys & Sons, 1980.

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

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