MINIMAX DESIGN OF FIR FILTERS WITH LOW GROUP DELAY USING ENHANCED SEQUENTIAL QUADRATIC PROGRAMMING

Wu-Sheng Lu

Dept. of Elec. and Comp. Engineering
University of Victoria
Victoria, BC, Canada V8W 3P6

ABSTRACT

Sequential quadratic programming (SQP) algorithms are widely recognized to be among the most successful algorithms for general constrained optimization. This paper attempts to develop an SQP-based method for equiripple FIR filters with low group delay. We explain how the complementarity conditions in an SQP design algorithm help reduce the amount of computation required to update the Lagrange multipliers in a significant manner. Design examples are presented to illustrate the algorithm's performance that compares favorably with several existing design methods.

1. INTRODUCTION

Minimax design of linear-phase FIR filters can be carried out efficiently using the Parks-McClellan algorithm [1] and its variants [2]. These algorithms however do not apply to the class of FIR filters with low group delay (in passband), which is the subject of study in this paper. A motivation for this type of design is that an FIR filter of given length may achieve an amplitude response comparable to that of its linear-phase counterpart, but with *reduced* passband group dely, a property desirable in many applications. The coefficients of such a filter are no longer symmetrical about its midpoint, hence the entire set of filter coefficients is involved in the design process.

Several methods are available for the design of this class of filters. These include the adaptive approach [3], constrained optimization approach [4], semidefinite programming (SDP) approach [5], and least-pth approach [6]. Among others, the SDP method appears to perform very well, especially when constraints on the approximation bound are imposed over a dense grid of frequencies. However, the large number of constrained inevitably effects the design efficiency and, in the case of high-order filters, may cause numerical difficulties. In this paper, we describe an attempt to develop an sequential quadratic programming (SQP) method for the design. SQP algorithms have been widely recognized to be among the most successful for constrained nonlinear optimization problems [7], and as such it is naturally

considered to be a candidate design tool for the problem at hand. More important, as will become clear shortly, the complementarity conditions in an SQP-based design formulation lead to a very small number of nonzero Lagrange multipliers relative to the number of the constraints imposed in the design. As a result, the amount of computation involved in a single solution iteration is fairly moderate even for high-order filters. Simulations are presented to demonstrated the proposed algorithm's performance in comparison with several existing design methods.

2. PROBLEM FORMULATION

We seek to find an FIR transfer function

$$H(z) = \sum_{i=0}^{N-1} h_i z^{-i}$$

to approximate a desired frequency response $H_d(\omega)$ in accordance with a weighted minimax criterion

$$\underset{\boldsymbol{h}}{\text{minimize}}\{\max_{\omega \in \Omega} |W(\omega)[H(e^{j\omega}) - H_d(\omega)]|\} \qquad (1)$$

where $\boldsymbol{h} = [h_0 \ h_1 \ \cdots \ h_{N-1}]^T$, $W(\omega) \geq 0$ is a weighting function, and $\Omega = \{\omega : \ 0 \leq \omega \leq \pi\}$. Let η be an upper bound of $W^2(\omega)|H(e^{j\omega}) - H_d(\omega)|^2$, then (1) can be converted to

$$\text{minimize} \quad \eta \qquad (2a)$$

subject to: $W^2(\omega)|H(e^{j\omega}) - H_d(\omega)|^2 \leq \eta$ for $\omega \in \Omega$ (2b)

The desired frequency response $H_d(\omega)$ is complex-valued and assumes the form $H_d(\omega) = H_{dr}(\omega) - jH_{dj}(\omega)$, and the frequency response of the filter to be designed can be expressed as $H(e^{j\omega}) = \boldsymbol{h}^T \boldsymbol{c}(\omega) - j\boldsymbol{h}^T \boldsymbol{s}(\omega)$ where $\boldsymbol{c}(\omega) = [1 \ \cos\omega \ \cdots \ \cos(N-1)\omega]^T$ and $\boldsymbol{s}(\omega) = [0 \ \sin\omega \ \cdots \ \sin(N-1)\omega]^T$. Hence the constraint in (2b) can be written as $a(\omega, \eta, \boldsymbol{h}) \geq 0$ for $\omega \in \Omega$, where

$a(\omega, \eta, \boldsymbol{h}) =$

$\eta - W^2(\omega)\{[\boldsymbol{h}^T \boldsymbol{c}(\omega) - H_{dr}(\omega)]^2 + [\boldsymbol{h}^T \boldsymbol{s}(\omega) - H_{dj}(\omega)]^2\}$

$$(3)$$

For feasible exercise of optimization algorithms, the constraint in (2b) is imposed on a dense grid of frequencies, $\Omega_d = \{\omega_1, \omega_2, \ldots, \omega_K\} \subset \Omega$, in the frequency bands of interest. Consequently, the problem in (2) becomes

$$\text{minimize} \quad \eta \tag{4a}$$
$$\text{subject to:} \quad a(\omega_i, \eta, \boldsymbol{h}) \geq 0 \quad 1 \leq i \leq K \tag{4b}$$

If we define $\boldsymbol{x} = [\eta \ \ \boldsymbol{h}^T]^T$, $a_i(\boldsymbol{x}) = a(\omega_i, \eta, \boldsymbol{h})$, and $\boldsymbol{e} = [1 \ 0 \ \cdots \ 0]^T$, then problem (4) can be written as

$$\text{minimize} \quad \boldsymbol{e}^T \boldsymbol{x} \tag{5a}$$
$$\text{subject to:} \quad a_i(\boldsymbol{x}) \geq 0 \qquad 1 \leq i \leq K \tag{5b}$$

3. DESIGN METHOD

3.1. An SQP-based algorithm for problem (5)

The Lagrangian of (5) is defined by

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = \boldsymbol{e}^T \boldsymbol{x} - \sum_{i=1}^{K} \mu_i a_i(\boldsymbol{x})$$

where $\boldsymbol{\mu} = [\mu_1 \ \cdots \ \mu_K]^T$ collects K Lagrange multipliers. The Karush-Kuhn-Tucker (KKT) conditions of problem (5) are the first-order necessary conditions for a solution of (5) which are given as follows [8]:

$$\nabla L(\boldsymbol{x}, \boldsymbol{\mu}) = \boldsymbol{e} - \sum_{i=1}^{K} \mu_i \nabla a_i(\boldsymbol{x}) = \boldsymbol{0} \tag{6a}$$
$$a_i(\boldsymbol{x}) \geq 0 \qquad 1 \leq i \leq K \tag{6b}$$
$$\mu_i \geq 0 \qquad 1 \leq i \leq K \tag{6c}$$
$$\mu_i a_i(\boldsymbol{x}) = 0 \qquad 1 \leq i \leq K \tag{6d}$$

The conditions in (6d) are called *complementarity conditions* which imply that if $a_i(\boldsymbol{x}) \neq 0$ (which in conjunction with (6b) means $a_i(\boldsymbol{x}) > 0$), then the ith Lagrange multiplier μ_i must be zero. In the context of filter design, this simply means that if at a certain frequency ω_i the approximation error is smaller than necessary (i.e., smaller than bound η), then the associated constraint $a_i(\boldsymbol{x}) \geq 0$ does not have to be involved in the next iteration (this degree of freedom is ensured because of $\mu_i = 0$). In all computer simulations we have conducted, the number of nonzero Lagrange multipliers, say $\hat{K}$, is only a small fraction of the total number of constraints $K (\hat{K} < 0.1K)$. As a result, the amount of computation requested for updating the Lagrange multiplier $\boldsymbol{\mu}_k$ is not significant even for high-order filters. We shall come back to this point when implementation details of the proposed algorithm are described.

Now suppose we start with a reasonable initial point $\boldsymbol{x}_0$ and an initial $\boldsymbol{\mu}_0 = \boldsymbol{0}$. In the kth iteration, $\{\boldsymbol{x}_k, \boldsymbol{\mu}_k\}$ is updated to $\{\boldsymbol{x}_{k+1}, \boldsymbol{\mu}_{k+1}\} = \{\boldsymbol{x}_k + \boldsymbol{\delta}_x, \boldsymbol{\mu}_k + \boldsymbol{\delta}_\mu\}$ such that

the KKT conditions in (6) are approximately satisfied up to the first order. The first-order approximation of (6) leads to

$$\boldsymbol{Y}_k \boldsymbol{\delta}_x + \boldsymbol{e} - \boldsymbol{A}_k^T \boldsymbol{\mu}_{k+1} = \boldsymbol{0} \tag{7a}$$
$$\boldsymbol{A}_k \boldsymbol{\delta}_x \geq -\boldsymbol{a}_k \tag{7b}$$
$$\boldsymbol{\mu}_{k+1} \geq \boldsymbol{0} \tag{7c}$$
$$(\boldsymbol{\mu}_{k+i})_i (\boldsymbol{A}_k \boldsymbol{\delta}_k + \boldsymbol{a}_k)_i = 0 \quad 1 \leq i \leq K \tag{7d}$$

where $\boldsymbol{Y}_k = \nabla^2 L(\boldsymbol{x}_k, \boldsymbol{\mu}_k)$, $\boldsymbol{a}_k = [a_1(\boldsymbol{x}_k) \ \cdots \ a_K(\boldsymbol{x}_k)]^T$, and

$$\boldsymbol{A}_k = \begin{bmatrix} \nabla^T a_1(\boldsymbol{x}_k) \\ \vdots \\ \nabla^T a_K(\boldsymbol{x}_k) \end{bmatrix}_{K \times (N+1)} \tag{7e}$$

The ith row of $\boldsymbol{A}_k$ assumes the form $[1 \ \ W^2(H_{dr}\boldsymbol{c}^T + H_{dj}\boldsymbol{s}^T - (\boldsymbol{c}\boldsymbol{c}^T + \boldsymbol{s}\boldsymbol{s}^T)\boldsymbol{h}_k^T)]$ where the frequency-dependence of W, $\boldsymbol{c}$, $\boldsymbol{x}$, H_{dr}, and H_{dj} at $\omega = \omega_i$ has been omitted. Eq. (7a)–(7d) are the *exact* KKT conditions of the quadratic programming (QP) problem

$$\text{minimize} \quad \frac{1}{2} \boldsymbol{\delta}^T \boldsymbol{Y}_k \boldsymbol{\delta} + \boldsymbol{\delta}^T \boldsymbol{e} \tag{8a}$$
$$\text{subject to:} \quad \boldsymbol{A}_k \boldsymbol{\delta} \geq -\boldsymbol{a}_k \tag{8b}$$

Now if $\boldsymbol{\delta}_x$ is the solution of (8), then $\boldsymbol{x}_k$ is updated to $\boldsymbol{x}_{k+1} = \boldsymbol{x} - k + \boldsymbol{\delta}_x$. Based on the analysis earlier, the nonzero Lagrange multipliers can be updated using (7a) as

$$\hat{\boldsymbol{\mu}}_{k+1} = (\boldsymbol{A}_{ak} \boldsymbol{A}_{ak}^T)^{-1} \boldsymbol{A}_{ak} (\boldsymbol{Y}_k \boldsymbol{\delta}_x + \boldsymbol{e}) \tag{9}$$

where the rows of $\boldsymbol{A}_{ak}$ are those rows of $\boldsymbol{A}_k$ satisfying $(\boldsymbol{A}_k \boldsymbol{\delta}_x + \boldsymbol{a}_k)_i = 0$ and $\boldsymbol{\mu}_{k+1}$ is obtained by inserting zeros wherever necessary into $\hat{\boldsymbol{\mu}}_{k+1}$. The fact that only a small fraction of constraints in (8b) is active also implies that the QP problem in (8) can be solved efficiently using for example an active-set method [8].

The iteration continues until a convergence criterion in terms of $\|\boldsymbol{\delta}_x\|_2$ or the total number of iterations is met. The optimal impulse response of the length-N FIR filter can be obtained from the solution vector $\boldsymbol{x}^*$ as $\boldsymbol{h}^* = \boldsymbol{x}^*(2:end)$.

3.2. Convex relaxation of problem (8)

A strictly convex relaxation of problem (8) can be made by replacing $\boldsymbol{Y}_k$ in (8a) with a positive definite matrix, still denoted by $\boldsymbol{Y}_k$, recursively as follows: with $\boldsymbol{Y}_0 = \boldsymbol{I}$ the Broyden-Fletcher-Goldfarb-Shanno (BFGS) recursion [8] updates $\boldsymbol{Y}_k$ to

$$\boldsymbol{Y}_{k+1} = \boldsymbol{Y}_k + \frac{\boldsymbol{\eta}_k \boldsymbol{\eta}_k^T}{\boldsymbol{\delta}_x^T \boldsymbol{\eta}_k} - \frac{\boldsymbol{v}_k \boldsymbol{v}_k^T}{\boldsymbol{\delta}_x^T \boldsymbol{v}_k} \tag{10a}$$

where

$$\boldsymbol{v}_k = \boldsymbol{Y}_k \boldsymbol{\delta}_x \tag{10b}$$

$$\boldsymbol{\eta}_k = \theta \boldsymbol{\gamma}_k + (1 - \theta) \boldsymbol{v}_k \tag{10c}$$

$$\boldsymbol{\gamma}_k = -(\boldsymbol{A}_{k+1} - \boldsymbol{A}_k)^T \boldsymbol{\mu}_{k+1} \tag{10d}$$

$$\theta = \begin{cases} 1 & \text{if } \boldsymbol{\delta}_x^T (\boldsymbol{\gamma}_k - 0.2 \boldsymbol{v}_k) \geq 0 \\ \dfrac{0.8 \boldsymbol{\delta}_x^T \boldsymbol{v}_k}{\boldsymbol{\delta}_x^T (\boldsymbol{v}_k - \boldsymbol{\gamma}_k)} & \text{otherwise} \end{cases} \tag{10e}$$

A desirable feature of the BFGS update is that if $\boldsymbol{Y}_k$ is positive definite, then $\boldsymbol{Y}_{k+1}$ is also positive definite. With $\boldsymbol{Y}_0 = \boldsymbol{I}$, therefore, every QP subproblems in the entire design process is guaranteed to be a convex QP which can be solved efficiently using an interior-point or active set algorithm.

3.3. A closed-form line search

A further enhancement of the SQP algorithm described above is made by including a line search step in the algorithms that yields a positive scalar α_k that minimizes a potential function $\psi(\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x)$ over $\alpha \in [0, 1]$ where $\boldsymbol{\delta}_x$ is the solution of (8) with $\boldsymbol{Y}_k$ produced using (10). The potential function adopted here assumes the form

$$\psi(\boldsymbol{x}) = \boldsymbol{e}^T \boldsymbol{x} - \sum_{i=1}^{K} (\boldsymbol{\mu}_k)_i a_i(\boldsymbol{x}) \tag{11}$$

With $\boldsymbol{x} = \boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x$, (11) becomes

$$\psi(\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x) = \boldsymbol{e}^T (\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x) - \sum_{i=1}^{K} (\boldsymbol{\mu}_k)_i a_i(\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x) \tag{12}$$

Since $\boldsymbol{\mu}_k \geq \boldsymbol{0}$, minimizing $\psi(\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x)$ helps reducing the objective function in (5a) *and*, in case some $a_i(\boldsymbol{x}_k + \boldsymbol{\delta}_x)$ fail to hold the constraints in (5b), reducing the degree of violation of these constraints. In this way, the inclusion of a line search step in an SQP algorithm turns out to be of great benefit, although it adds an additional amount of computation to the algorithm. It follows from (3) and (12) (and the fact that $\boldsymbol{\mu}_k \geq 0$) that the potential function $\psi(\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x)$ in (12) is a second order polynomial of α with a positive coefficient for α^2, i.e., $\psi(\boldsymbol{x}_k + \alpha \boldsymbol{\delta}_x) = a_2^{(k)} \alpha^2 + a_1^{(k)} \alpha + a_0^{(k)}$ with $a_2^{(k)} > 0$, which achieves its minimum at $\alpha^* = -a_1^{(k)}/2a_2^{(k)}$. Thus we may choose $\alpha_k = \min\{\alpha^*, 1\}$.

4. DESIGN EXAMPLES

The method described above was applied to three FIR filters with low group delay.

Example 1 A minimax design of lowpass FIR filter of length 91 with group delay $d = 40$, passband edge $\omega_p = 0.475\pi$

and stopband edge $\omega_a = 0.525\pi$ was designed using the proposed method with $K = 1100$. The weighting function is piecewise constant with $W(\omega) \equiv 0.2$ in the passband, $W(\omega) \equiv 2$ in the stopband and 0 elsewhere. It took the algorithm 150 iterations to converge to a solution FIR filter with maximum passband ripple $e_p = 0.0189$, minimum stopband attenuation $e_a = 54.41$ dB, and maximum relative group delay deviation in passband $e_{gd} = 0.026$. The average number of active constraints was found to be $\hat{K} = 22$. The amplitude amplitude response and passband ripple of the filter obtained are depicted in Fig. 1a and b, respectively. For comparison, an FIR filter with identical design specifications was designed in [5] using the SDP method, and the minimax design gave $e_p = 0.0191$, $e_a = 54.26$ dB, and $e_{gd} = 0.027$.
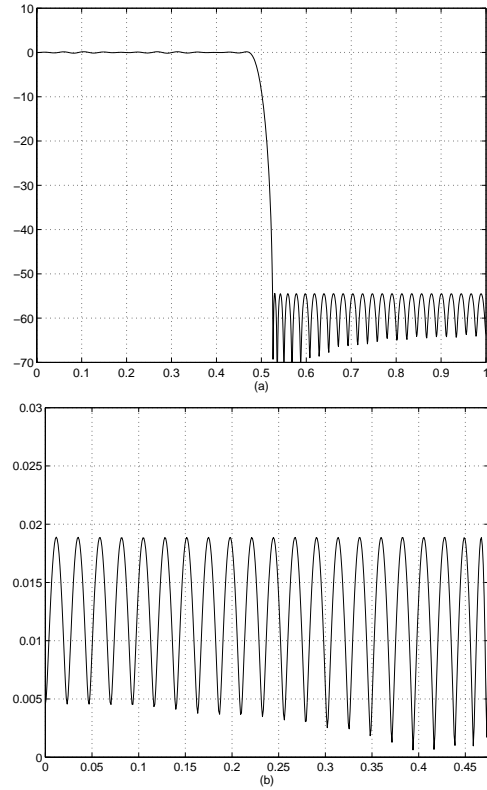
Fig. 1. (a) The amplitude response and (b) passband ripple of the filter in Example 1.

Example 2 A bandpass filter of length 161 with group delay $d = 65$, $\omega_{a1} = 0.375\pi$, $\omega_{p1} = 0.4\pi$, $\omega_{p2} = 0.6\pi$, $\omega_{a2} = 0.625\pi$ was designed by the proposed method with $K = 1200$, and $W(\omega) \equiv 1$ in the passband and stopband and 0 elsewhere. It took the algorithm 100 iterations to converge to a solution filter with $e_p = 0.0127$, $e_a = 38.04$ dB, and $e_{gd} = 0.041$. The average number of active constraints was $\hat{K} = 40$. The amplitude response and passband ripple of the filter obtained are shown in Fig. 2(a) and (b), respec-

tively. For comparison, an FIR filter with the same design specifications was designed in [6] using the least pth methods. The results were $e_p = 0.0176$, $e_a = 37.49$ dB, and $e_{gd} = 0.043$.
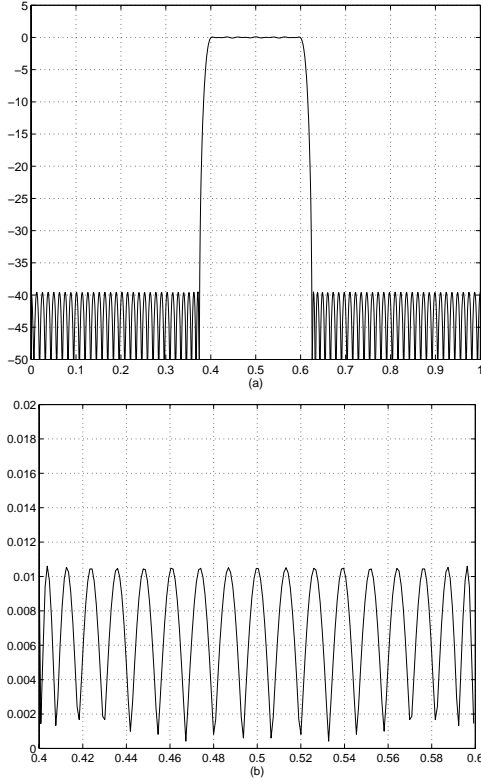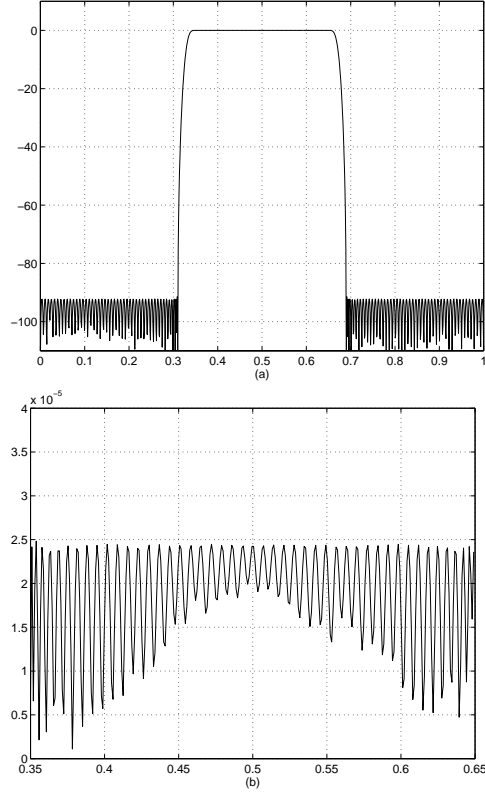

Fig. 2. (a) The amplitude response and (b) passband ripple of the filter in Example 2.

Example 3 The proposed algorithm was found reliable for the design of a variety of equiripple FIR filters with low group delay. This example presents a bandpass FIR filter of relatively high order. The design parameters were $N = 281$, $d = 120$, $f_{a1} = 0.31\pi$, $f_{p1} = 0.35\pi$, $f_{p2} = 0.65\pi$, $f_{a2} = 0.69\pi$, $W(\omega) \equiv 1$ in the passbands and stopband, and $K = 1400$. It took the algorithm 110 iterations to converge to a solution FIR filter with $e_p = 2.4833 \times 10^{-5}$, $e_a = 91.43$dB, and $e_{gd} = 2.043 \times 10^{-4}$. The amplitude response and passbnad ripple of the filter are shown in Fig. 3a and b, respectively.

5. REFERENCES

[1] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, vol. 19, pp. 189-194, 1974.

[2] A. Antoniou, *Digital Filters: Analysis, Design, and Applications*, 2nd ed., McGraw Hill, 1993.

Fig. 3. (a) The amplitude response and (b) passband ripple of the filter in Example 3.

[3] Y. C. Lim, H.-H. Lee, C. I. Chen, and R.-H. Yang, "A weighted least-squares approximation for quasi-equiripple FIR and IIR digital filter design," *IEEE Trans. Signal Processing*, vol. 40, pp. 551-558, 1992.

[4] M. Lang, "Chebyshv design of FIR filters with arbitrary magnitude magnitude and phase responses," *EU-SIPCO'1996*, vol. 3, pp. 1757-1760, 1996.

[5] W.-S. Lu, "Design of nonlinear-phase FIR digital filters: A semidefinite programming approach," *Proc. ISCAS'1999*, vol. 3, pp. 263-266, Orlando, FL, May 1999.

[6] W.-S. Lu, "Minimax design of nonlinear-phase FIR filters: A least$p$th approach," *Proc. ISCAS'2003*, vol. 1, pp. 409-412, Phoenix, AZ, May 2002.

[7] P. T. Boggs and J. W. Tolle, "Sequential quadratic programming," *Acta Numerica*, vol. 4, pp. 1-51, 1995.

[8] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., Wiley, 1987.