DESIGN OF NONLINEAR PHASE FIR DIGITAL FILTERS USING QUADRATIC PROGRAMMING

Mathias C. Lang

Department of Communications and Radio-Frequency Engineering
Vienna University of Technology
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Tel: +43 1 58801 3523; Fax: +43 1 587 05 83
E-mail: `mlang@neptun.nt.tuwien.ac.at`

ABSTRACT

This paper presents two methods for the design of FIR filters with arbitrary magnitude and phase responses according to a weighted mean squared error criterion with constraints on the resulting magnitude and phase errors. This constrained least square criterion allows for an arbitrary trade-off between pure L_2 filters and Chebyshev filters. The resulting nonlinear optimization problem is either converted into a standard quadratic programming problem (method 1) or exactly solved by a sequence of quadratic programs (method 2). The quadratic programming problems can be solved efficiently using standard software.

1. INTRODUCTION

The least squares optimality criterion for the design of digital filters is a reasonable choice for many practical applications. Especially in the stopbands of frequency selective filters it may be desirable to minimize the energy of the error. However, the maximum error can not be controlled. Especially at the bandedges large errors occur due to Gibbs' phenomenon. This problem can be overcome by imposing constraints on the resulting error. This constrained least square (CLS) criterion has first been proposed by Adams [1] for the design of linear phase FIR filters. An extension to the nonlinear phase case was given in [2], where the L_2 design with constraints on the complex error is considered. The limitation of this approach is that magnitude and phase of the frequency response cannot be constrained separately. In [3] a CLS design with constraints on the magnitude and group delay is proposed.

In this paper we propose the minimization of the weighted L_2 norm of the complex error function subject to constraints on the magnitude and phase of the frequency response. The magnitude and phase errors of the frequency response directly contribute to the linear signal distortions and hence it seems reasonable to impose constraints on these quantities. According to the author's knowledge the solution to this problem has not been considered so far.

This work has been supported by FWF grant P11133-ÖMA.

2. PROBLEM FORMULATION

The frequency response of an FIR filter with impulse response $h(n)$ of length N is given by[1]

$$H(e^{j\theta}) = \sum_{n=0}^{N-1} h(n)e^{-jn\theta} = \mathbf{h}^T\mathbf{e}(\theta), \tag{1}$$

with $\mathbf{h} = [h(0), h(1), \ldots, h(N-1)]^T$ and $\mathbf{e}(\theta) = [1, e^{-j\theta}, e^{-j2\theta}, \ldots, e^{-j(N-1)\theta}]^T$. The weighted L_2 norm of the complex frequency domain error function is given by

$$E = \sqrt{\frac{1}{\pi}\int_0^{\pi} W(\theta)|H(e^{j\theta}) - D(e^{j\theta})|^2 d\theta}, \tag{2}$$

where $W(\theta)$ is a real non-negative weighting function and $D(e^{j\theta})$ is the desired complex frequency response. Squaring (2) and ignoring terms which are independent of the filter coefficients gives the following strictly convex quadratic objective function to be minimized:

$$F(\mathbf{h}) = \mathbf{h}^T\mathbf{Q}\mathbf{h} - 2\mathbf{h}^T\mathbf{q}. \tag{3}$$

The $N \times N$ matrix $\mathbf{Q}$ and the length N column vector $\mathbf{q}$ are defined as follows:

$$\begin{aligned} \mathbf{Q} &= \frac{1}{\pi}\int_0^{\pi} W(\theta)\mathbf{e}(\theta)\mathbf{e}^H(\theta)d\theta \\ \mathbf{q} &= \frac{1}{\pi}\int_0^{\pi} W(\theta)\mathrm{Re}\{D^*(e^{j\theta})\mathbf{e}(\theta)\}d\theta. \end{aligned} \tag{4}$$

These integrals can be calculated analytically for many practical filter design specifications. The unconstrained minimization of (3) gives the coefficients

$$\mathbf{h} = \mathbf{Q}^{-1}\mathbf{q}. \tag{5}$$

This is the solution of the complex approximation problem according to the weighted L_2 norm. In order to bound the maximum value of the resulting errors, additional constraints will be considered. The quantities to be constrained

[1] The superscripts *, T and H represent conjugate, transpose, and conjugate-transpose operations, respectively.

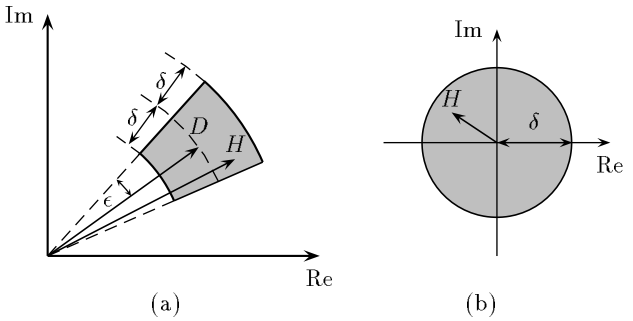

Figure 1: Constraints on magnitude and phase of the frequency response in the complex plane. (a) passbands. (b) stopbands.

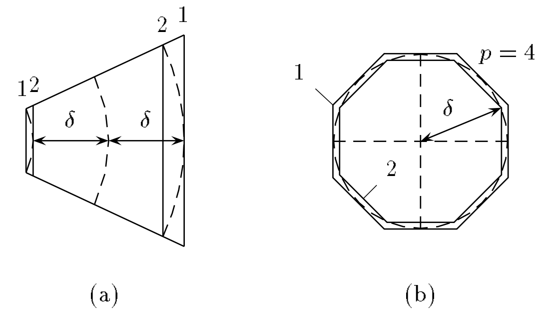

Figure 2: Linear magnitude constraints with (2) and without (1) reduction of the original feasible region. (a) passbands. (b) stopbands (with $p = 4$).

are the magnitude error $E_m(\theta) = |H(e^{j\theta})| - |D(e^{j\theta})|$ and the phase error $E_\phi(\theta) = \arg\{H(e^{j\theta})\} - \arg\{D(e^{j\theta})\}$. The constraints are formulated as

$$|E_m(\theta)| \leq \delta(\theta) \quad \text{and} \quad |E_\phi(\theta)| \leq \epsilon(\theta), \tag{6}$$

where $\delta(\theta)$ and $\epsilon(\theta)$ are real, strictly positive functions specified by the filter designer. A graphical explanation of the constraints (6) in the complex plane is shown in fig. 1. For every frequency point in the passbands and in the stopbands the frequency response $H(e^{j\theta})$ is confined to the shaded regions in fig. 1 (a) and (b). The resulting optimization problem with the objective function (3) and the constraints (6) is a nonlinearly constrained non-convex problem. This is due to the fact that the inequalities in (6) define a non-convex region in the N-dimensional filter coefficient vector space. A direct solution of this problem is difficult and standard optimization methods are slow and only yield local solutions. Hence, in this paper two alternatives are proposed to overcome these problems. In the first method the constraints (6) are linearized and the resulting convex quadratic programming problem is solved by an efficient standard method. The errors introduced by the linearization of the constraints are of second order in the small quantities $\delta(\theta)$ and $\epsilon(\theta)$ in the passbands, and can be made arbitrarily small in the stopbands by increasing the size of the problem. The second method, however, gives an exact solution of the problem by solving a sequence of quadratic programs. The linearizations used in this method are different from the ones used in method 1.

3. DESIGN METHODS

3.1. Method 1

Here, the original problem is slightly altered by linearizing the constraints (6). This yields a linearly constrained convex quadratic minimization problem, which can be solved efficiently using robust standard algorithms. The linearization of the constraints only affects the magnitude constraints. Let B_p and B_s denote the set of passbands ($|D(e^{j\theta})| > 0$) and the set of stopbands ($|D(e^{j\theta})| = 0$), respectively. The original constraints $-\delta(\theta) \leq E_m(\theta) \leq \delta(\theta)$ are replaced by the linear constraints

$$-\delta_l(\theta) \leq \operatorname{Re}\{H(e^{j\theta})e^{-j\phi_D(\theta)}\} - |D(e^{j\theta})| \leq \delta_u(\theta) \tag{7}$$

for $\theta \in B_p$ with $\phi_D(\theta) = \arg\{D(e^{j\theta})\}$, and

$$\operatorname{Re}\{H(e^{j\theta})e^{j\alpha_i}\} \leq \delta_s(\theta), \quad i = 0(1)2p-1 \tag{8}$$

for $\theta \in B_s$ with $\alpha_i = i\pi/p$ and any integer $p \geq 2$. The constraints (7) were first used in [4] for the design of FIR allpass filters, and the linearization used in (8) was proposed in the context of FIR filter design according to the complex Chebyshev criterion in [5]. Fig. 2 shows the linearization of the magnitude constraints according to (7) and (8) in the complex plane. The alternatives 1 and 2 in fig. 2 correspond to the following choices of the functions $\delta_l(\theta)$, $\delta_u(\theta)$ and $\delta_s(\theta)$ in (7) and (8):

Alternative 1: (9)

$$\delta_l(\theta) = \delta(\theta)\cos\epsilon(\theta) + |D(e^{j\theta})|(1-\cos\epsilon(\theta)), \theta \in B_p,$$
$$\delta_u(\theta) = \delta(\theta), \theta \in B_p, \quad \delta_s(\theta) = \delta(\theta), \theta \in B_s.$$

Alternative 2: (10)

$$\delta_u(\theta) = \delta(\theta)\cos\epsilon(\theta) - |D(e^{j\theta})|(1-\cos\epsilon(\theta)), \theta \in B_p,$$
$$\delta_l(\theta) = \delta(\theta), \theta \in B_p, \quad \delta_s(\theta) = \delta(\theta)\cos(\tfrac{\pi}{2p}), \theta \in B_s.$$

Choosing alternative 1 will give a solution with an optimal value of the objective function that is guaranteed to be less or equal to the optimal value of the original problem. However, the original constraints might be violated. The maximum violation $\Delta_p(\theta)$ of the original magnitude constraints in the passbands is given by

$$\begin{aligned}\Delta_p(\theta) &= \left(|D(e^{j\theta})| + \delta(\theta)\right)(\sec\epsilon(\theta) - 1) \\ &= \left(|D(e^{j\theta})| + \delta(\theta)\right)\left(\tfrac{1}{2}\epsilon^2(\theta) + \tfrac{5}{24}\epsilon^4(\theta) + \ldots\right),\end{aligned} \tag{11}$$

which consists only of terms of second and higher order of the specified functions $\delta(\theta)$ and $\epsilon(\theta)$. The maximum violation $\Delta_s(\theta)$ of the original magnitude constraints in the stopbands is

$$\Delta_s(\theta) = \delta(\theta)(\sec(\pi/2p) - 1), \tag{12}$$

which approaches zero quadratically with increasing p. The original constraints are guaranteed to be satisfied if the linearized magnitude constraints are chosen according to alternative 2. However, the value of the objective function might be larger than its value at the optimal solution of the original problem.

Fig. 1 shows that the constraints on the phase error are linear functions in $\text{Re}\{H(e^{j\theta})\}$ and $\text{Im}\{H(e^{j\theta})\}$. Since these functions are linear functions of the filter coefficients $\mathbf{h}$, it is obvious that constraints on the phase error function $E_\phi(\theta)$ can exactly be imposed using linear constraint functions. The constraints $E_\phi(\theta) \leq \epsilon(\theta)$ ($E_\phi(\theta) \geq -\epsilon(\theta)$) can be formulated as

$$
\begin{aligned}
&\text{Im}\{H(e^{j\theta})\} \leq (\geq) \tan[\phi_D(\theta) + (-)\epsilon(\theta)]\text{Re}\{H(e^{j\theta})\} \\
&\text{for} \quad |\phi_D(\theta) + (-)\epsilon(\theta)| < \pi/2, \\
&\text{Im}\{H(e^{j\theta})\} \geq (\leq) \tan[\phi_D(\theta) + (-)\epsilon(\theta)]\text{Re}\{H(e^{j\theta})\} \\
&\text{for} \quad |\phi_D(\theta) + (-)\epsilon(\theta)| > \pi/2, \\
&\text{sign}[\phi_D(\theta) + (-)\epsilon(\theta)]\text{Re}\{H(e^{j\theta})\} \geq (\leq) 0 \\
&\text{for} \quad |\phi_D(\theta) + (-)\epsilon(\theta)| = \pi/2, \\
&\theta \in B_p. \qquad (13)
\end{aligned}
$$

These constraints are linear in the unknown filter coefficients $\mathbf{h}$. The optimal filter coefficients can be computed by solving the standard quadratic programming problem with the objective function (3) and the constraints (7), (8), and (13) evaluated on a dense frequency grid. If m_p and m_s denote the number of frequency points in the passbands and stopbands, respectively, then the number of constraints is $2(2m_p + pm_s)$. The larger the integer p is chosen the more constraints have to be considered, but the smaller is the difference to the original constraints in the stopbands.

3.2. Method 2

With this method the solution of the original problem is computed by solving a sequence of quadratic programming problems. The nonlinear magnitude constraints $|E_m(\theta)| \leq \delta(\theta)$ can be written as

$$
\begin{aligned}
|H(e^{j\theta})|^2 &\geq \left(|D(e^{j\theta})| - \delta(\theta)\right)^2, \quad \theta \in B_p, \\
|H(e^{j\theta})|^2 &\leq \left(|D(e^{j\theta})| + \delta(\theta)\right)^2, \quad \theta \in B_p, \\
|H(e^{j\theta})|^2 &\leq \delta^2(\theta), \quad \theta \in B_s. \qquad (14)
\end{aligned}
$$

The squared magnitude of the frequency response is given by

$$|H(e^{j\theta})|^2 = \mathbf{h}^T \mathbf{e}(\theta)\mathbf{e}^H(\theta)\mathbf{h}. \qquad (15)$$

In iteration k we use the linearization

$$|H(e^{j\theta})|^2 \approx \mathbf{h}_{k-1}^T \mathbf{e}(\theta)\mathbf{e}^H(\theta)\mathbf{h} \qquad (16)$$

with fixed $\mathbf{h}_{k-1}$ from the previous iteration. In the first iteration ($k = 1$) $\mathbf{h}_0$ is chosen according to (5) as the solution of the unconstrained problem. In all further iterations

$$\mathbf{h}_k = \tau \mathbf{h}_k^{opt} + (1 - \tau)\mathbf{h}_{k-1}, \qquad (17)$$

where $\mathbf{h}_k^{opt}$ is the optimal solution of the k-th subproblem and $\tau \in (0, 1)$ is an update factor. The choice $\tau = 0.2..0.5$ usually results in good convergence. This simple trick for reducing the order of a nonlinear function was first proposed in [6] for the design of QMF banks using unconstrained optimization. In each iteration we solve the quadratic programming problem with the objective function (3), the linear constraints (13), and the linearized constraints

$$
\begin{aligned}
\mathbf{h}_{k-1}^T \mathbf{e}(\theta)\mathbf{e}^H(\theta)\mathbf{h} &\geq \left(|D(e^{j\theta})| - \delta(\theta)\right)^2, \quad \theta \in B_p, \\
\mathbf{h}_{k-1}^T \mathbf{e}(\theta)\mathbf{e}^H(\theta)\mathbf{h} &\leq \left(|D(e^{j\theta})| + \delta(\theta)\right)^2, \quad \theta \in B_p, \\
\mathbf{h}_{k-1}^T \mathbf{e}(\theta)\mathbf{e}^H(\theta)\mathbf{h} &\leq \delta^2(\theta), \quad \theta \in B_s. \qquad (18)
\end{aligned}
$$

with $\mathbf{h}_{k-1}$ computed in the previous iteration. This iterative algorithm stops if the violation of the original magnitude constraints (14) is smaller than some prescribed tolerance or, equivalently, if $\|\mathbf{h}_k^{opt} - \mathbf{h}_{k-1}^{opt}\|$ is smaller than a prescribed bound. The constraints (13) and (18) are evaluated on a dense frequency grid. The total number of constraints per iteration is $4m_p + m_s$. In practice, the number of constraints can be reduced considerably after a few iterations because the set of active constraints remains essentially unchanged. As an alternative to (16) we could use a Taylor series for $|H(e^{j\theta})|$ about $\mathbf{h}_{k-1}$ as in [3]. However, experiments showed that using the linearization (16) results in better convergence. It should be noted that convergence of this method cannot be guaranteed. However, many filter designs using this method assured us that it does in fact converge fast for most practical specifications. Convergence problems may arise if the constraints in the stopbands are very restrictive. In this case the optimal solution exhibits an equiripple stopband behavior. However, for all examples we tried it was always possible to ensure convergence by decreasing the update factor τ in (17), albeit at a slower rate.

The advantage of method 1 is that convergence can not become a problem. Only one quadratic programming problem has to be solved and there exist reliable algorithms with guaranteed convergence to the optimal solution. However, linearization of the magnitude constraints is necessary, and hence only an approximate solution of the original problem can be computed. Method 2 directly solves the original problem. Optimality of the solution is guaranteed because the original non-convex problem is iteratively approximated by convex problems with unique optimal solutions. In the case of convergence, the original problem is, up to a specified tolerance, exactly represented by the convex problem of the final iteration. The number of constraints of each quadratic programming problem is smaller than in method 1. However, a sequence of these problems has to be solved and convergence can not be guaranteed in general. The disadvantages of both methods are not severe. For practical specifications the linearization errors of method 1 are negligible, and convergence of the algorithm in method 2 is usually fast.

4. DESIGN EXAMPLE

A length 51 low-delay bandpass filter with a constant desired passband magnitude response of 1 was designed using the proposed methods. The phase response was required to be approximately linear in the passband, but the desired delay was chosen to be 15 samples instead of 25 samples as in the exact linear phase case. The stopband edges are $f_{s1} = 0.1$, $f_{s2} = 0.35$, and the passband edges are

$f_{p1} = 0.15$, $f_{p2} = 0.3$ ($f = \theta/2\pi$). The weighting function $W(\theta)$ in (2) was chosen to be 10^3 in the first stopband, 1 in the passband, 10^4 in the second stopband, and 0 in the transition bands. The maximum passband errors were specified to be 0.04 for the magnitude error and 0.03 radians for the phase error. The minimum stopband attenuation was chosen to be 50 dB in the first stopband and 60 dB in the second stopband. Two solutions were computed using method 1. The first with linearized magnitude constraints according to (9) (alternative 1) and the second according to (10) (alternative 2). A third solution was computed using method 2. For the designs with method 1 $p = 4$ was chosen. The number of equidistant frequency points was $m_p = 107$ and $m_s = 178$ (71 points in stopband 1 and 107 points in stopband 2) for all 3 designs. Fig. 3 shows the magnitudes of the frequency responses of the designed filters. The filters are almost identical. However, method 1/alternative 1 (dotted curve) gives a result which slightly violates the original magnitude constraints. The maximum violation of the magnitude constraints in the passband is $2.94 \cdot 10^{-4}$ which is smaller than the bound (11). The maximum stopband violation is 0.69 dB which equals the bound (12). This maximum violation in the stopbands occurs only once in each stopband at the first relative maximum next to the bandedges. The stopband violations can be made arbitrarily small by increasing p. The passband and stopband details are shown in fig. 4. It took 8 iterations for the algorithm of method 2 to converge to a solution with a maximum constraint violation of $7.5 \cdot 10^{-7}$ with an update factor $\tau = 0.5$. The smallest value of the objective function is achieved by the solution computed with method 1/alternative 1. This is due to the fact that this problem formulation has the least restrictive constraints. The most restrictive constraints occur for method 1/alternative 2 which yields the largest value of the objective function. The different decays of the stopband magnitude in fig. 4 (c) and (d) show this fact.

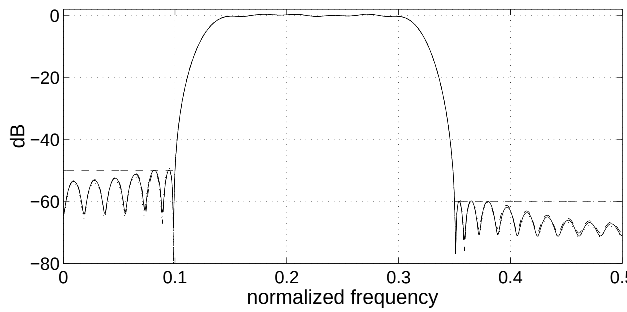

Figure 3: Design example: $|H(e^{j2\pi f})|$. Dotted curve: method 1 (alternative 1). Dashed curve: method 1 (alternative 2). Solid curve: method 2.

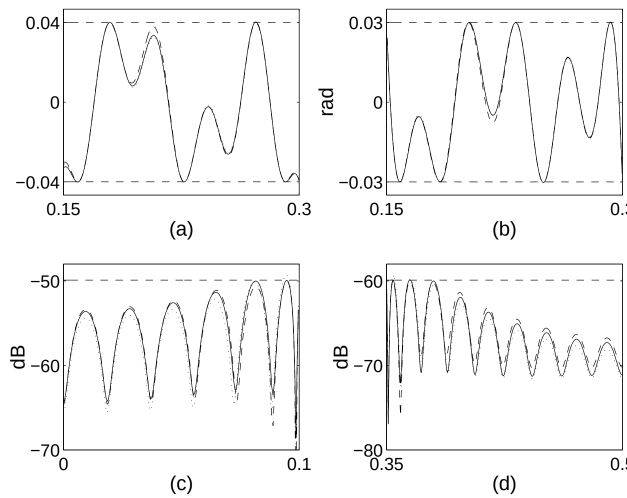

Figure 4: Design example: (a) $E_m(2\pi f)$ in the passband, (b) $E_\phi(2\pi f)$, (c) $|H(e^{j2\pi f})|$ in stopband 1, (d) $|H(e^{j2\pi f})|$ in stopband 2. Dotted curves: method 1 (alternative 1). Dashed curves: method 1 (alternative 2). Solid curves: method 2.

5. CONCLUSION

Two different methods for the design of nonlinear phase FIR filters with arbitrary magnitude and phase responses have been proposed. The solutions are computed according to a least squares criterion with additional constraints on the resulting magnitude and phase responses. In method 1 only one quadratic programming problem has to be solved and the solutions are close to the optimum. Method 2 requires the solution of a sequence of quadratic programming problems. The solutions computed with method 2 are optimal, but convergence cannot be guaranteed. However, the method converged for all practical design examples. Current work concentrates on the development of an efficient multiple exchange algorithm to avoid frequency discretization. First results are encouraging.

6. REFERENCES

[1] J. W. Adams, "FIR Digital Filters with Least-Squares Stopbands Subject to Peak-Gain Constraints," *IEEE Trans. Circuits and Systems*, vol. CAS-39, pp. 376-388, April 1991.

[2] M. Lang and J. Bamberger, "Nonlinear Phase FIR Filter Design According to the L_2 Norm with Constraints for the Complex Error," *Signal Processing*, vol. 36, pp. 31-40, March 1994.

[3] J. L. Sullivan and J. W. Adams, "A New Nonlinear Optimization Algorithm for Asymmetric FIR Digital Filters," *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 2, pp. 541-544, May-June 1994.

[4] K. Steiglitz, "Design of FIR Digital Phase Networks," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-29, pp. 171-176, April 1981.

[5] X. Chen and T. W. Parks, "Design of FIR Filters in the Complex Domain," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-35, pp. 144-153, Feb. 1987.

[6] C.-K. Chen and J.-H. Lee, "Design of Quadrature Mirror Filters with Linear Phase in the Frequency Domain," *IEEE Trans. Circuits and Systems-II*, vol. 39, pp. 593-605, September 1992.