Sequential Design of IIR Digital Filters for Low-Power DSP Applications

S. Saab, W.-S. Lu, and A. Antoniou

Department of Electrical and Computer Engineering
University of Victoria
PO Box 3055 Stn CSC, Victoria BC V8W 3P6 Canada
ssaab@ece.uvic.ca, wslu@ece.uvic.ca, andreas@ece.uvic.ca

Abstract—A method for the design of linear-phase IIR digital filters for low-power applications is proposed. In this method, the digital filter is implemented as a cascade arrangement of 2nd-order sections. The first section is designed through optimization techniques so as to satisfy as far as possible the overall required specifications. A second section is then added which is designed through optimization techniques so that the first two sections satisfy as far as possible the overall required specifications. This process is repeated until a multisection filter is obtained which satisfies the required specifications under the most critical circumstances imposed by the application at hand. The minimum number of sections required to process a particular input signal can then by switched on through the use of a simple adaptation mechanism and, in this way, the power consumption can be minimized. This design structure is achieved by formulating the design on the kth section as a weighted quadratic programming problem with stability, amplitude response, and the first k − 1 sections as constraints.

I. Introduction

Digital filters are commonly designed to assure a desirable performance for all possible signals that may be encountered. However, for many DSP applications the worst-case conditions are rarely present, which leads to wasted power and computational resources. On the other hand, many electrical systems such as portable wireless systems are dependent on power availability. In addition, many other systems such as multitasking systems are dependent on the availability of computational resources. Recent studies have shown that these dependencies can be alleviated by dynamically adjusting the complexity of the algorithms used according to predetermined criteria and the changing signal characteristics [1],[2].

In [3], a sequential design of FIR digital filters for low-power DSP applications was proposed. In this paper, a method for the design of linear-phase IIR digital filters for low-power applications is proposed. The main idea is to design a collection of 2nd-order cascaded sections that can be turned off when they are not needed so as to conserve power and/or reduce the computational complexity. Each section is designed so as to satisfy as far as possible the overall required specifications with stability, amplitude response, and the previous section designs as constraints. The solution for each section is set up as a weighted quadratic programming problem.

II. Problem Formulation

Consider an IIR digital filter represented by

$$H(z) = \frac{B(z)}{A(z)} = \prod_{i=1}^{K} \frac{b_i(z)}{a_i(z)}$$

with

$$\frac{b_i(z)}{a_i(z)} = \frac{b_{i_0} + b_{i_1} z^{-1} + b_{i_2} z^{-2}}{1 + a_{i_1} z^{-1} + a_{i_2} z^{-2}}$$

The frequency response of the filter can be examined by performing the substitution $z = e^{j\omega T}$ where ω is the frequency of the signal in rad/s, and T is the sampling period of the filter, which will be assumed to be 1 s throughout.

At the kth stage of the proposed design, it is assumed that the sections represented by $b_i(z)/a_i(z), i = 1, \ldots, k-1$ have been designed and one seeks a stable 2nd-order section with a transfer function $b_k(z)/a_k(z)$ such that the objective function

$$\Psi(\mathbf{c}_k) = \int_{-\pi}^{\pi} W_k(\omega)|e(\omega, \mathbf{c}_k)|^2 d\omega \qquad (1)$$

is minimized where $W_k(\omega)$ is a weighting function,

$$\begin{aligned} e(\omega, \mathbf{c}_k) &= H_k(e^{j\omega}) - D_k(\omega) \\ H_k(z) &= \frac{B_k(z)}{A_k(z)} = \prod_{i=1}^{k} \frac{b_i(z)}{a_i(z)} \\ \mathbf{c}_k &= \begin{bmatrix} \mathbf{b}_k^T & \mathbf{a}_k^T \end{bmatrix}^T \\ \mathbf{b}_k &= \begin{bmatrix} b_{k_0} & b_{k_1} & b_{k_2} \end{bmatrix}^T \\ \mathbf{a}_k &= \begin{bmatrix} a_{k_1} & a_{k_2} \end{bmatrix}^T \end{aligned}$$

and

$$D_k(\omega) = M(\omega) e^{-j\omega\tau_k}$$

0-7803-4957-1/98/$10.00©1998 IEEE

where $M(\omega)$ is the desired amplitude response, and τ_k is the desired passband delay. The design can be accomplished by using the iterative quadratic programming method proposed in [4] as follows. We first write the objective function in (1) as

$$\Psi(\mathbf{c}_k) = \int_{-\pi}^{\pi} W_k(\omega) \left| \frac{1}{A_k(e^{j\omega})} \right|^2 \cdot \left| B_k(e^{j\omega}) - D_k(\omega) A_k(e^{j\omega}) \right|^2 d\omega \qquad (2)$$

which, in the lth iteration of the optimization procedure (see Section IV), can be approximated with

$$\Psi_d^{(l)}(\mathbf{c}_k) = \sum_{i=0}^{N} W_k^{(l)}(\omega_i) \left| \frac{1}{A_k^{(l-1)}(e^{j\omega_i})} \right|^2 \cdot \left| B_k^{(l)}(e^{j\omega_i}) - D_k(\omega_i) A_k^{(l)}(e^{j\omega_i}) \right|^2 \qquad (3)$$

where $\{\omega_i\}$ is the set of frequencies in the band(s) of interest.

III. Initial Guess

The initial guess used for $\mathbf{c}_k$ to minimize $\Psi(\mathbf{c}_k)$ in (2) is such that

$$W_k(\omega) = \left| A_k(e^{j\omega}) \right|^2$$

This is a reasonable initial weighting function as it will place more emphasis on the stopband, thus placing the zeros in a suitable initial position. As well, the objective function can then be formulated in closed quadratic form as

$$\Psi(\mathbf{c}_k) = \frac{1}{2} \mathbf{c}_k^{\mathrm{T}} \mathbf{\Gamma}_{k-1}^{\mathrm{T}} \mathbf{H}_k \mathbf{\Gamma}_{k-1} \mathbf{c}_k - \tilde{\mathbf{a}}_{k-1}^{\mathrm{T}} \mathbf{P}_k \mathbf{\Gamma}_{k-1} \mathbf{c}_k + c \qquad (4)$$

where

$$\mathbf{\Gamma}_{k-1} = \begin{bmatrix} \mathbf{B}_{k-1} & \mathbf{0} \\ 0 & \mathbf{A}_{k-1} \end{bmatrix}$$

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{Q}_k & -\mathbf{S}_{0_k} \\ -\mathbf{S}_{0_k}^{\mathrm{T}} & \mathbf{R}_{0_k} \end{bmatrix}$$

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{S}_{1_k} & -\mathbf{R}_{1_k} \end{bmatrix}$$

$\tilde{\mathbf{a}}_{k-1}$ is the vector of coefficients corresponding to the polynomial $A_{k-1}(z)$, c is a constant, and for lowpass filters with passband and stopband edges ω_p and ω_a respectively,

$$\mathbf{Q}_k[m, m \pm n] = \frac{\sin(n\pi) + \sin(n\omega_p) - \sin(n\omega_a)}{n}$$

$$\mathbf{R}_k[m, m \pm n] = \frac{\sin(n\omega_p)}{n}$$

$$\mathbf{S}_k[m, m+q] = \frac{\sin[(\tau_k + q)\omega_p]}{\tau_k + q}$$

for $1 \le m \le 2k+1$, $0 \le n \le 2k$, and $-2k \le q \le 2k$. In the above equations, $\mathbf{R}_{0_k}$ is $\mathbf{R}_k$ with the first column and row removed, $\mathbf{R}_{1_k}$ is $\mathbf{R}_k$ with the first column and last two rows removed, $\mathbf{S}_{0_k}$ is $\mathbf{S}_k$ with the first column removed, and $\mathbf{S}_{1_k}$ is $\mathbf{S}_k^{\mathrm{T}}$ with the last two rows removed. $\mathbf{B}_{k-1}$ and $\mathbf{A}_{k-1}$ are the *convolutional matrices* [5] (see Section III-A), of $\tilde{\mathbf{b}}_{k-1}$ and $\tilde{\mathbf{a}}_{k-1}$ respectively, where $\tilde{\mathbf{b}}_{k-1}$ is the vector of coefficients corresponding to the polynomial $B_{k-1}(z)$.

A. Constraints

When designing the kth section, the first $k-1$ sections must be taken into account in order to find the optimal 2nd-order addition to give the optimal overall response. However, the first $k-1$ sections must not be altered. This constraint is accomplished in the formulation of equation (4) through the use of a *convolutional matrix* [5].

The convolutional matrix of a vector $\mathbf{x}$ of length n, convolved with a vector $\mathbf{y}$ of length m is defined as

$$\mathbf{X} = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ x_2 & x_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n-1} & \cdots & x_{n-m+1} \\ 0 & x_n & \cdots & x_{n-m+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix}$$

so that the convolution of $\mathbf{x}$ and $\mathbf{y}$ is $\mathbf{Xy}$. Transforming the first $k-1$ sections into its convolutional matrix $\mathbf{\Gamma}_{k-1}$ allows its absorption into $\mathbf{H}_k$ in equation (4). The resulting expression is an error of the overall response with respect to the kth section only.

To ensure that the resulting filter sections are stable, the linear inequality constraints (see p. 524 of [6])

$$\begin{bmatrix} 0 & 1 \\ 1 & -1 \\ -1 & -1 \end{bmatrix} \mathbf{a}_k \le \begin{bmatrix} \delta^2 \\ \delta^2 \\ \delta^2 \end{bmatrix} \qquad (5)$$

are imposed where δ is the maximum distance a pole can be from the origin.

Treating equation (4) as a quadratic programming problem with equation (5) as linear inequality constraints, a suitable initial guess for $\mathbf{c}_k$ is attained.

IV. Optimization Procedure

Allowing an arbitrary weighting function in equation (2) prevents the error from being expressed as a quadratic equation because $|A_k(e^{j\omega})|^2$ is an unknown divider. This problem can be alleviated by formulating an initial guess as in Section III, and using it to evaluate $|A_k(e^{j\omega})|^2$. A new $H_k(z)$ can then be found by minimizing the resulting quadratic function [4]. However, the new $\mathbf{a}_k$ will most likely be different from the initial guess, so the minimization must become an iterative procedure where successive solutions converge. The cycle is repeated until

$$\left\| \mathbf{c}_k^{(l)} - \mathbf{c}_k^{(l-1)} \right\|_2 < \varepsilon$$

where $\mathbf{c}_k^{(l)}$ is the solution to the lth iteration of the kth section, and ε is a small positive constant.

If the new $\mathbf{a}_k$ is used to reevaluate $|A_k(e^{j\omega})|^2$, this procedure becomes susceptible to unstable oscillations. This is avoided by assigning a linear combination of the old and new solutions. As will be seen in Section IV-A, the previous $\mathbf{b}_k$ is also used so that a next guess such as

$$\mathbf{c}_k^{(l)} = \frac{1}{2}\mathbf{c}_k^{(l)} + \frac{1}{2}\mathbf{c}_k^{(l-1)}$$

is required.

As well, because an arbitrary weighting function divided by a complex expression is located inside the integral of equation (2), it cannot be evaluated as a closed quadratic form and must be approximated with equation (3) which can be formulated as the open quadratic form

$$\begin{aligned}\Psi_d^{(l)}(\mathbf{c}_k) = \; & \frac{1}{2}\mathbf{c}_k^{(l)\mathrm{T}}\mathbf{\Gamma}_{k-1}^{\mathrm{T}}\mathbf{H}_{d_k}^{(l)}\mathbf{\Gamma}_{k-1}\mathbf{c}_k^{(l)} - \\ & \tilde{\mathbf{a}}_{k-1}^{\mathrm{T}}\mathbf{P}_{d_k}^{(l)}\mathbf{\Gamma}_{k-1}\mathbf{c}_k^{(l)} + c\end{aligned}$$

where

$$\begin{aligned}
\mathbf{H}_{d_k}^{(l)} &= \begin{bmatrix} \sum_{i=0}^{N} \widehat{W}_k^{(l)}(\omega_i)\mathbf{\Omega}_k(\omega_i) & -\sum_{i=0}^{N_p} \widehat{W}_k^{(l)}(\omega_i)\mathbf{\Theta}_{0_k}(\omega_i) \\ -\sum_{i=0}^{N_p} \widehat{W}_k^{(l)}(\omega_i)\mathbf{\Theta}_{0_k}^{\mathrm{T}}(\omega_i) & \sum_{i=0}^{N_p} \widehat{W}_k^{(l)}(\omega_i)\mathbf{\Omega}_{0_k}(\omega_i) \end{bmatrix} \\
\mathbf{P}_{d_k}^{(l)} &= \begin{bmatrix} \sum_{i=0}^{N_p} \widehat{W}_k^{(l)}(\omega_i)\mathbf{\Theta}_{1_k}(\omega_i) & -\sum_{i=0}^{N_p} \widehat{W}_k^{(l)}(\omega_i)\mathbf{\Omega}_{1_k}(\omega_i) \end{bmatrix} \\
\widehat{W}_k^{(l)}(\omega_i) &= W_k^{(l)}(\omega_i)\left|\frac{1}{A_k^{(l-1)}(e^{j\omega_i})}\right|^2 \\
\mathbf{\Omega}_k(\omega_i)[m, m \pm n] &= \cos(n\omega_i) \\
\mathbf{\Theta}_k(\omega_i)[m, m+q] &= \cos[(\tau_k + q)\omega_i]
\end{aligned}$$

for $1 \le m \le 2k+1$, $0 \le n \le 2k$, and $-2k \le q \le 2k$, where $\omega_i \in \{0, \ldots, \omega_p\}$ for $i \in \{0, \ldots, N_p\}$. In the above equations, $\mathbf{\Omega}_{0_k}(\omega_i)$ is $\mathbf{\Omega}_k(\omega_i)$ with the first column and row removed, $\mathbf{\Omega}_{1_k}(\omega_i)$ is $\mathbf{\Omega}_k(\omega_i)$ with the first column and last two rows removed, $\mathbf{\Theta}_{0_k}(\omega_i)$ is $\mathbf{\Theta}_k(\omega_i)$ with the first column removed, and $\mathbf{\Theta}_{1_k}(\omega_i)$ is $\mathbf{\Theta}_k^{\mathrm{T}}(\omega_i)$ with the last two rows removed.

A. Constraints

The constraints in Section III-A obviously must be used in the optimization procedure as well. In addition, the linear inequality constraints

$$\frac{\tilde{\mathbf{b}}_k^{(l-1)\mathrm{T}}\mathbf{\Omega}_k(\omega_{i_t})\mathbf{B}_{k-1}}{\left|A_k^{(l-1)}(e^{j\omega_{i_t}})\right|^2}\mathbf{b}_k^{(l)} \le (1+\delta_p)^2 \tag{6}$$

$$-\frac{\tilde{\mathbf{b}}_k^{(l-1)\mathrm{T}}\mathbf{\Omega}_k(\omega_{i_b})\mathbf{B}_{k-1}}{\left|A_k^{(l-1)}(e^{j\omega_{i_b}})\right|^2}\mathbf{b}_k^{(l)} \le -(1-\delta_p)^2 \tag{7}$$

are used to ensure that the amplitude response stays within predefined limits in the passband and transition band, where $\omega_{i_t} \in \{0, \ldots, \omega_a\}$, $\omega_{i_b} \in \{0, \ldots, \omega_p\}$, and δ_p is the predefined maximum deviation from unity gain in the passband and maximum additional gain in the transition band. The constraints in the transition band are required because optimal stable linear-phase IIR filters naturally exhibit artifacts in the transition band(s).

B. Weights

Using the additional constraints in Section IV-A allows the control of the overall passband and transition band amplitude response during each section design. The stopband can then be minimized as much as possible subject to these constraints. However, in order to distribute the attenuation more evenly throughout the stopband, the weighting function is adjusted using a modified version of the approach proposed in [3] as

$$\begin{aligned}
W_k^{(l)}(\omega_i) &= \frac{M[W_k^{(l-1)}(\omega_i) + \epsilon]|H_k^{(l-1)}(e^{j\omega_i})|^2}{V_k^{(l-1)}} \\
V_k^{(l-1)} &= \frac{\sum_{m=N_a}^{N}[W_k^{(l-1)}(\omega_m) + \epsilon]|H_k^{(l-1)}(e^{j\omega_m})|^2}{N - N_a + 1}
\end{aligned}$$

where $\omega_i \in \{\omega_a, \ldots, \pi\}$ for $m \in \{N_a, \ldots, N\}$, ϵ is a small positive constant used to prevent the propagation of zeros in the weighting function, and M is a multiplier used to emphasize attenuation in the stopband. $W_k^{(l)}(\omega_i)$ is set to 1 in the passband.

V. Design Example

A 7-section low-power IIR digital lowpass filter was designed with $\omega_p = 1$ rad/s, $\omega_a = 1.3$ rad/s and $\omega_s = 2\pi$ rad/s. The optimization parameters were set to $N = 100$, $\delta = 0$, $\varepsilon = 0.1$, and $\epsilon = 0.1$. Equation (6) was used with $\{\omega_{i_t}\} = \{0 \quad 0.5 \quad 1 \quad 1.15 \quad 1.3\}$, and equation (7) was used with $\{\omega_{i_b}\} = \{0 \quad 0.5 \quad 1\}$. The assignments of τ_k, δ_p, and M are discussed in Section V-A. The results obtained are displayed in Figs. 1-7.

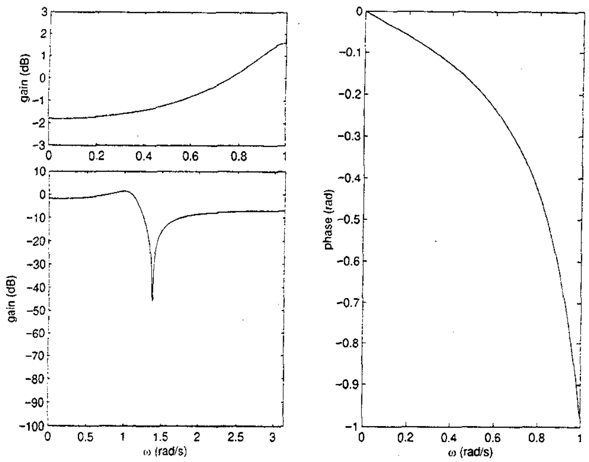

Fig. 1. First section of a low-power IIR digital filter.

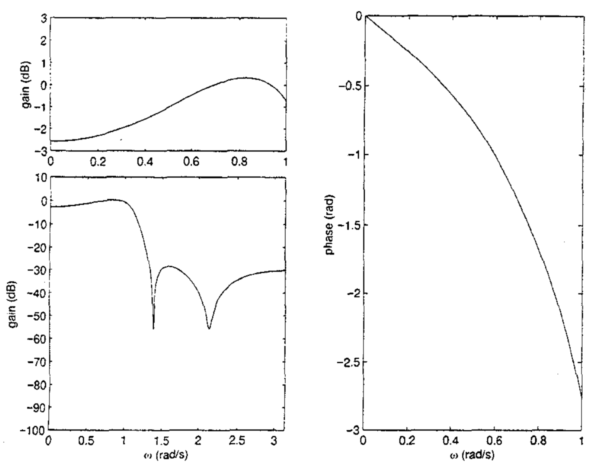

Fig. 2. First 2 sections of a low-power IIR digital filter.

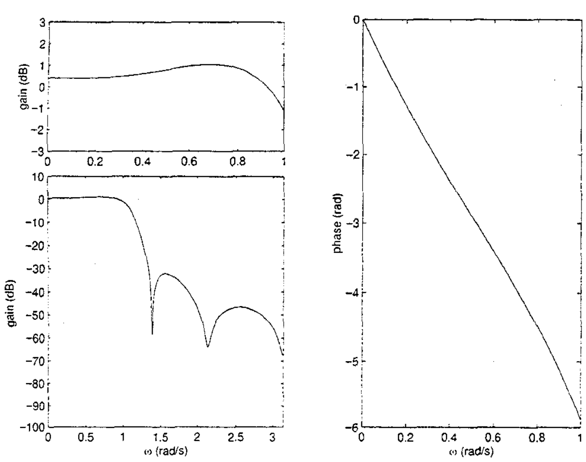

Fig. 3. First 3 sections of a low-power IIR digital filter.

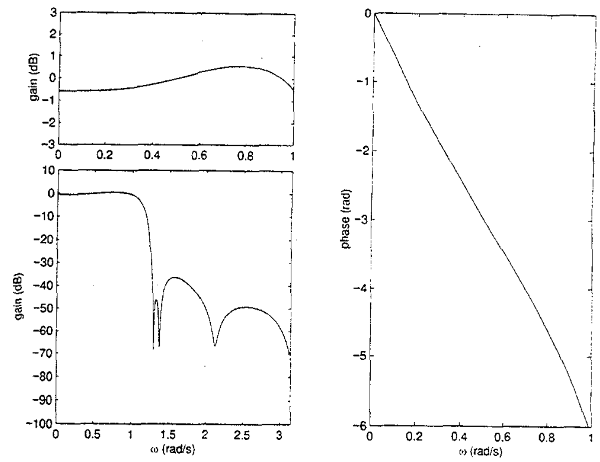

Fig. 4. First 4 sections of a low-power IIR digital filter.

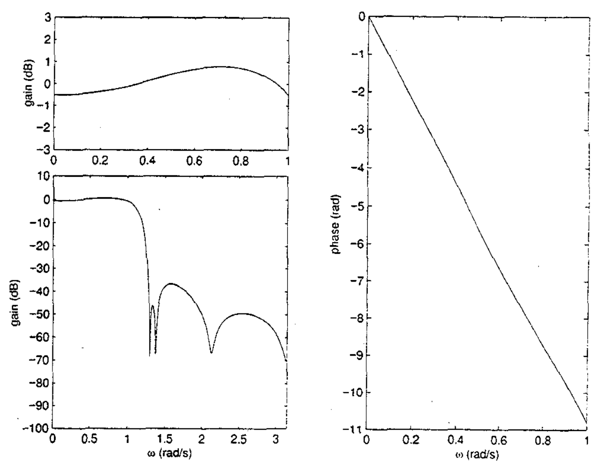

Fig. 5. First 5 sections of a low-power IIR digital filter.

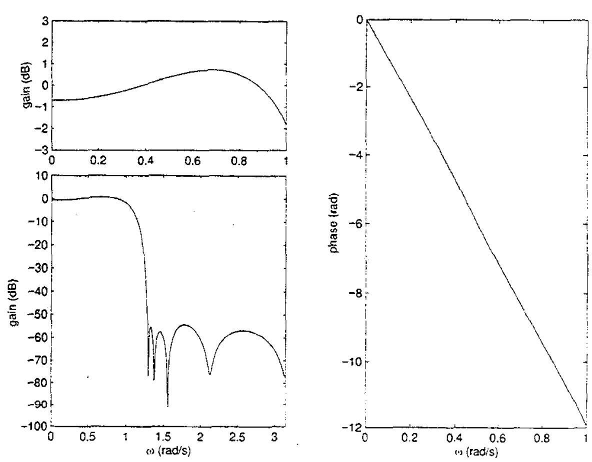

Fig. 6. First 6 sections of a low-power IIR digital filter.

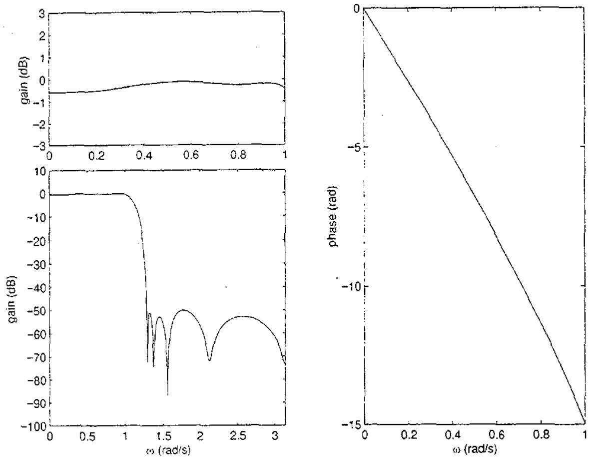

Fig. 7. First 7 sections of a low-power IIR digital filter.

A. Delay Assignments

When designing each section, careful attention must be paid to the assignment of τ_k and, as a consequence, δ_p and M. Aside from setting the phase, τ_k affects the location of the poles and zeros. Provided that the constraints do not prevent their placements, low values of τ_k will put zeros on the unit circle at the stopband edge. As τ_k increases, the zeros will move around the unit circle until one is at π, and the other will move up the real axis until it is sitting outside the unit circle to the right. Then, as τ_k increases further, the left zero will move left, away from the unit circle until it reappears on the far right-hand side. It will then move towards the unit circle, meet with the other zero, at which point they separate up and down outside of the unit circle.

The poles will start near the unit circle at the passband edge, then move towards the real axis within the right-hand side of the unit circle. The poles will rest on the real axis for some of the time when the zeros are also on the real axis. When both zeros are on the right-hand side, the poles will mirror their movement within the unit circle.

The constraint set by requiring linear phase in the passband will tend to pull the zeros slightly off of the unit circle as they move towards π, and place poles closer to their corresponding zeros. This can be alleviated by increasing the values of δ_p and/or M. The optimal solution is to find the lowest values of δ_p and M which will give improved response results with each section. Sometimes the compromise is too much and a section must be designed in order to prepare for another stopband amplitude response improving section. These sections will have high delays with the zeros sitting on the right-hand side.

An overall optimization procedure using the direct search method was used for finding the low delays which placed the zeros on the unit circle, and the high delays which straightened the phase and passband amplitude.

The design procedure for the example was to find the lowest delay sections possible until the constraints required a high delay section. The parameter assignments used in the design are

$$\begin{aligned}\{\tau_k\} &= \{\,0.4932 \quad 1.1800 \quad 4.2000 \\ &\qquad 0.175\times 10^{-2} \quad 5.1274 \quad 0.8639 \quad 2.2618\,\} \\ \{A_{p_k}\} &= \{\,3.0 \quad 2.5 \quad 2.0 \quad 1.5 \quad 1.0 \quad 3.0 \quad 0.5\,\} \\ \{M_k\} &= \{\,10^3 \quad 10^3 \quad 10^3 \quad 10^3 \quad 10^3 \quad 10^3 \quad 10^6\,\}\end{aligned}$$

where $A_{p_k} = 20\log(\delta_{p_k})$.

VI. Conclusion

A method for the design of linear-phase IIR digital filters for low-power applications has been proposed. The problem is set up as a quadratic function with constraints on previous section designs, stability, and amplitude response. This allows for quick design time using a quadratic programming algorithm. Heuristics are used in assigning the phase of each section. The resulting designs exhibit the desirable steep transition response inherent in IIR digital filters.

Acknowledgement

The authors are grateful to Micronet, Networks of Centres of Excellence Program, and the Natural Sciences and Engineering Research Council of Canada for supporting this work.

References

[1] J. T. Ludwig, S. H. Nawab, and A. P. Chandrakasan, "Low-power digital filtering using approximate processing," *IEEE J. Solid State Circuits*, vol. 31, no. 3, pp. 395–400, Mar. 1996.

[2] S. H. Nawab, A. V. Oppenheim, A. P. Chandrakasan, J. M. Winograd, and J. T. Ludwig, "Approximate signal processing," *J. VLSI Signal Processing*, vol. 15, no. 2, pp. 177–200, Mar. 1997.

[3] W.-S. Lu, A. Antoniou, S. Saab, "Sequential design of FIR digital filters for low-power DSP applications," *in Proc. Asilomar Conf.*, Monterey, CA., November 1997.

[4] W.-S. Lu, "Design of stable IIR digital filters with equiripple passbands and peak-constrained least squares stopbands," *in Proc. IEEE Int. Symp. Circuits Syst.*, vol. 4, pp. 2192–2195, June 1997.

[5] S. Saab, W.-S. Lu, and A. Antoniou, "Design of quadrature mirror-image filter banks for low-power applications," *in Proc. IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, vol. 1, pp. 178–181, August 1997.

[6] A. Antoniou, *Digital Filters: Analysis, Design, and Applications*, McGraw-Hill, New York, 2nd edition, 1993.