Design Techniques for Two-Dimensional Digital Filters

JAMES V. HU and LAWRENCE R. RABINER

Abstract—The theory for designing finite-duration impulse response (FIR) digital filters can readily be extended to two or more dimensions. Using linear programming techniques, both frequency sampling and optimal (in the sense of Chebyshev approximation over closed compact sets) two-dimensional filters have been successfully designed. Computational considerations have limited the filter impulse response durations (in samples) to 25 by 25 in the frequency sampling case, and to 9 by 9 in the optimal design case. However, within these restrictions, a large number of filters have been investigated. Several of the issues involved in designing two-dimensional digital filters are discussed.

Introduction

Techniques for designing one-dimensional digital filters have been investigated for several years, and are fairly well established in the literature [1]–[7]. Very few of these techniques, however, have been extended to two or more dimensions. Since there is no lack of interest in two-dimensional signal processing, it is worth investigating those design procedures that are readily extendable to two dimensions. In this paper, the frequency sampling technique and optimal filter design procedures are discussed in the case of the design of circularly symmetric two-dimensional low-pass filters.

Before discussing the results obtained, it is worth reviewing the state of the art in designing two-dimensional digital filters. Two of the most common classes of one-dimensional digital filters are usually called recursive and nonrecursive filters. (The alternate terminology infinite impulse response (IIR) and finite impulse response (FIR) has recently been proposed [8] and will be used throughout this paper.) IIR filters can be efficiently realized by factoring their z-transforms into a cascade of real or complex pole-zero networks and realizing the filter using these simple sections. In two dimensions, however, the two-dimensional z-transforms cannot generally be factored into lower order systems [9], [10]. This leads to many design problems, such as the difficulty in determining the stability of the filter, as well as in finding an efficient realization where the coefficients may be truncated to a reasonable number of bits. Because of these difficulties, very little work has been done on designing IIR two-dimensional digital filters.

For FIR digital filters, the problems of stability do not exist in one dimension as the z-transform is a finite polynomial. Thus design techniques in one dimension are often directly extendable to two or more dimensions by appropriate modifications to the design procedure. Huang [11] has already shown how the windowing technique may be extended to two dimensions by forming two-dimensional circularly symmetric windows from one-dimensional windows. The remainder of this paper discusses the design of circularly symmetric frequency sampling filters and optimal (equiripple) filters.

Manuscript received May 2, 1972. This paper was presented at the IEEE Arden House Workshop on Digital Filtering, Harriman, N. Y., January 1972.

The authors are with Bell Telephone Laboratories, Inc., Murray Hill, N. J. 07974.

Theory of Two-Dimensional FIR Filters

Let $h(n_1, n_2)$ be the impulse response of a two-dimensional filter where $-\infty < n_1, n_2 < \infty$. The two-dimensional z transform of the filter is defined as

$$H(z_1, z_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} h(n_1, n_2) z_1^{-n_1} z_2^{-n_2}. \tag{1}$$

If the resulting filter is to be stable, the impulse response must satisfy the condition

$$\sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} |h(n_1, n_2)| < \infty. \tag{2}$$

For IIR filters (when any one of the four limits on the summation is infinite), it is generally quite difficult to determine whether or not a particular filter is stable, i.e., whether (2) is satisfied. For FIR filters, on the other hand, stability is always guaranteed since all four limits on the summation are finite and $|h(n_1, n_2)| < \infty$.

If we restrict ourselves to the case when $h(n_1, n_2)$ is a finite sequence defined from $0 \le n_1 \le N_1 - 1$, $0 \le n_2 \le N_2 - 1$, then (1) becomes

$$H(z_1, z_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} h(n_1, n_2) z_1^{-n_1} z_2^{-n_2}. \tag{3}$$

The "frequency response" of the filter is obtained by evaluating (3) for $z_1 = e^{j\omega_1}$ and $z_2 = e^{j\omega_2}$, giving

$$H(e^{j\omega_1}, e^{j\omega_2}) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} h(n_1, n_2) \exp\left[-j(\omega_1 n_1 + \omega_2 n_2)\right]. \tag{4}$$

The two-dimensional FIR filter may be realized in several ways. The simplest technique is a direct-convolution realization using the relation

$$y(n_1, n_2) = \sum_{m_1=0}^{N_1-1} \sum_{m_2=0}^{N_2-1} h(m_1, m_2) x(n_1 - m_1, n_2 - m_2) \tag{5}$$

where $x(n_1, n_2)$ is the input to the filter and $y(n_1, n_2)$ is the filter output.

The discrete Fourier transform (DFT) relations for the filter may be obtained by evaluating (4) at the discrete set of values

$$\omega_{k_1} = k_1\left(\frac{2\pi}{N_1}\right), \qquad k_1 = 0, 1, \cdots, N_1 - 1 \tag{6}$$

$$\omega_{k_2} = k_2\left(\frac{2\pi}{N_2}\right), \qquad k_2 = 0, 1, \cdots, N_2 - 1 \tag{7}$$

giving[1]

$$H(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} h(n_1, n_2) \cdot \exp\left[-j2\pi\left(\frac{k_1 n_1}{N_1} + \frac{k_2 n_2}{N_1}\right)\right]. \tag{8}$$

The inverse DFT is readily obtained as

$$h(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} H(k_1, k_2) \cdot \exp\left[j2\pi\left(\frac{k_1 n_1}{N_1} + \frac{k_2 n_2}{N_2}\right)\right]. \tag{9}$$

It is interesting to note that two-dimensional DFT's, in (8), may be evaluated using one-dimensional DFT's by partitioning (8) as

$$H(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \underbrace{\left[\sum_{n_2=0}^{N_2-1} h(n_1, n_2) e^{-j2\pi(n_2 k_2/N_2)}\right]}_{g(n_1, k_2)} e^{-j2\pi(n_1 k_1/N_1)} \tag{10}$$

and realizing that the inner summation is a one-dimensional DFT where n_1 is held fixed; once all the inner summation DFT's have been calculated (calling the results $g(n_1, k_2)$, the outer summation is a DFT where k_2 is held fixed.

A frequency interpolation formula from the DFT coefficients may be derived by inserting (9) into (4), giving

$$H(e^{j\omega_1}, e^{j\omega_2}) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left[\frac{1}{N_1 N_2} \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} H(k_1, k_2) \cdot \exp\left[j2\pi\left(\frac{k_1 n_1}{N_1} + \frac{k_2 n_2}{N_2}\right)\right]\right] \cdot \exp\left[-j(\omega_1 n_1 + \omega_2 n_2)\right]. \tag{11}$$

Interchanging orders of summation, and summing over the n_1 and n_2 indices, gives

$$H(e^{j\omega_1}, e^{j\omega_2}) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} H(k_1, k_2) A(k_1, k_2, \omega_1, \omega_2) \tag{12}$$

where

$$A(k_1, k_2, \omega_1, \omega_2) = \frac{1}{N_1 N_2}\left[\frac{1-\exp(-jN_1\omega_1)}{1-\exp\left[j\left(2\pi\frac{k_1}{N_1} - \omega_1\right)\right]}\right] \cdot \left[\frac{1-\exp(-jN_2\omega_2)}{1-\exp\left[j\left(2\pi\frac{k_2}{N_2} - \omega_2\right)\right]}\right]. \tag{13}$$

Equation (12) serves as the basis for designing frequency sampling two-dimensional filters. As seen from (12), the continuous frequency response of the filter is a linear combination of shifted interpolating functions ($A(k_1, k_2, \omega_1, \omega_2)$) weighted by the DFT coefficients $H(k_1, k_2)$. The DFT coefficients are called frequency samples as they exactly specify the value of the frequency response at uniformly spaced frequencies. For designing frequency sampling filters, the majority of the frequency samples are given specific values depending on the frequency response being approximated. The remaining unspecified frequency samples are left as free variables to be optimized according to some minimization criterion. The details of frequency sampling design of two-dimensional low-pass filters will be left for the next section.

For designing optimal (equiripple) filters whose peak error of approximation, over several closed areas in the ω_1, ω_2 plane, is minimized, (4) can be applied directly.[2] In this case, the filter coefficients are the impulse response samples that are chosen to minimize the approximation error using an optimization procedure. Since the filter frequency response of (4) is still linear in the filter coefficients ($h(n_1, n_2)$), optimization procedures similar to those used to design frequency sampling filters may be applied. The details of this method are discussed in a later section.

Finally, it can be observed from (4) that if certain symmetries are maintained in the filter impulse response coefficients (similar statements can be made about the DFT coefficients), the filter's frequency response becomes purely real, to within a linear phase shift in each of the two dimensions. The necessary symmetries are

$$h(n_1, n_2) = h(N_1 - 1 - n_1, n_2) = h(n_1, N_2 - 1 - n_2). \tag{14}$$

The application of (14) to (4) gives (assuming N_1, N_2 even)

[1] The terminology $H(k_1, k_2)$ rather than $H(e^{j\omega_{k_1}}, e^{j\omega_{k_2}})$ is used for notational convenience.

[2] Equivalently, by making all the DFT coefficients variable, (12) can be applied directly for designing optimal (equiripple) filters.

$$H(e^{j\omega_1}, e^{j\omega_2}) = e^{-j\omega_1((N_1-1)/2)}e^{-j\omega_2((N_2-1)/2)} \cdot \sum_{n_1=0}^{(N_1/2)-1} \sum_{n_2=0}^{(N_2/2)-1} h(n_1, n_2) \cdot \cos\left[\left(\frac{N_1-1}{2}\right) - n_1\right] \cdot \cos\left[\left(\frac{N_2-1}{2}\right) - n_2\right] \tag{15}$$

which is purely real, except for the linear phase terms outside the summation.

Design of Frequency Sampling Filters

For the specific case of linear phase frequency sampling filters, (12) and (13) can be modified through use of the symmetry relations on the DFT coefficients which are of the form

$$H(k_1, k_2) = |H(k_1, k_2)|\, e^{j\theta(k_1,k_2)} \tag{16}$$

where

$$|H(k_1, k_2)| = |H(k_1, N_2 - k_2)| = |H(N_1 - k_1, k_2)| \tag{17}$$

$$\theta(k_1, k_2) = \theta(k_1) + \theta(k_2) \tag{18}$$

$$\theta(k_1) = \begin{cases} -\dfrac{2\pi}{N_1} k_1\left(\dfrac{N_1-1}{2}\right), & k_1 = 0, 1, \cdots, NU \\ \dfrac{2\pi}{N_1}(N_1 - k_1)\left(\dfrac{N_1-1}{2}\right), & k_1 = NU+1, \cdots, N_1 - 1 \end{cases} \tag{19}$$

$$\theta(k_2) = \begin{cases} -\dfrac{2\pi}{N_2} k_2\left(\dfrac{N_2-1}{2}\right), & k_2 = 0, 1, \cdots, NV \\ \dfrac{2\pi}{N_2}(N_2 - k_2)\left(\dfrac{N_2-1}{2}\right), & k_2 = NV+1, \cdots, N_2 - 1 \end{cases} \tag{20}$$

and

$$NU = \begin{cases} N_1/2, & N_1 \text{ even} \\ \dfrac{N_1-1}{2}, & N_1 \text{ odd} \end{cases} \tag{21}$$

$$NV = \begin{cases} N_2/2, & N_2 \text{ even} \\ \dfrac{N_2-1}{2}, & N_2 \text{ odd.} \end{cases} \tag{22}$$

Whenever N_1 or N_2 is even, the additional constraints

$$\theta(k_1, N_2/2) = \theta(N_1/2, k_2) = 0 \tag{23}$$

$$H(k_1, N_2/2) = H(N_1/2, k_2) = 0 \tag{24}$$

must be maintained.

Applying these constaints into (12) and (13) yields the following equations (after considerable arithmetic manipulations):

$$H(e^{j\omega_1}, e^{j\omega_2}) = \exp\left[-j\left[\omega_1\left(\frac{N_1-1}{2}\right) + \omega_2\left(\frac{N_2-1}{2}\right)\right]\right] \cdot \frac{1}{N_1N_2} \cdot \left[H(0,0)\alpha(\omega_1, N_1)\alpha(\omega_2, N_2) + \sum_{k_1=1}^{NU} |H(k_1, 0)|\, \alpha(\omega_2, N_2)\beta(\omega_1, k_1, N_1) + \sum_{k_2=1}^{NV} |H(0, k_2)|\, \alpha(\omega_1, N_1)\beta(\omega_2, k_2, N_2) + \sum_{k_1=1}^{NU}\sum_{k_2=1}^{NV} |H(k_1, k_2)|\, \beta(\omega_1, k_1, N_1)\beta(\omega_2, k_2, N_2)\right] \tag{25}$$

where

$$\alpha(\omega, N) = \frac{\sin \frac{\omega N}{2}}{\sin \frac{\omega}{2}} \tag{26}$$

$$\beta(\omega, k, N) = \frac{\sin\left[\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)N\right]}{\sin\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)} + \frac{\sin\left[\left(\frac{\omega}{2} + \frac{\pi k}{N}\right)N\right]}{\sin\left(\frac{\omega}{2} + \frac{\pi k}{N}\right)} \tag{27}$$

$$NU = \begin{cases} N_1/2 - 1, & N_1 \text{ even} \\ (N_1 - 1)/2, & N_1 \text{ odd} \end{cases} \tag{28}$$

$$NV = \begin{cases} N_2/2 - 1, & N_2 \text{ even} \\ (N_2 - 1)/2, & N_2 \text{ odd.} \end{cases} \tag{29}$$

Although cumbersome in appearance, (25) is seen to consist basically of a sum of simple interpolating functions that are shifted (up or down) in frequency depending on k_1 and k_2. As such, (25)—without the leading linear phase terms—is the basic design equation for frequency sampling filters.

To apply (25) to the design of a circularly symmetric low-pass filter, consider Fig. 1 in which the filter response regions are plotted in the ω_1, ω_2 plane. The passband is the region where

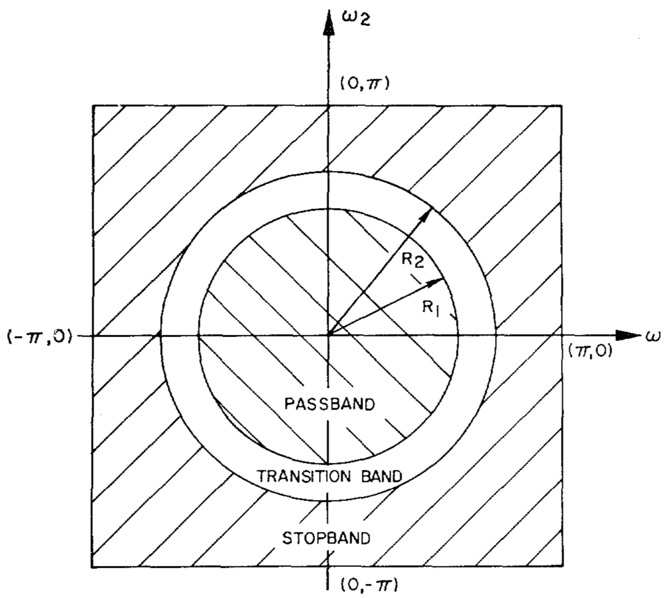

Fig. 1. Regions in the (ω_1, ω_2) plane over which the filter response is specified for a low-pass filter.

$$\rho(\omega_1, \omega_2) = (\omega_1^2 + \omega_2^2)^{1/2} \leq R_1. \tag{30}$$

The transition band is the region where

$$R_1 < \rho(\omega_1, \omega_2) < R_2 \tag{31}$$

and the stopband is the region where

$$\rho(\omega_1, \omega_2) \geq R_2. \tag{32}$$

With respect to Fig. 1, the DFT points, or frequency samples, are located on an N_1 by N_2 grid of points in the ω_1, ω_2 plane as shown in Fig. 2 for $N_1 = 9$, $N_2 = 9$. To design a frequency sampling approximation to the ideal circularly symmetric response

$$\hat{H}(e^{j\omega_1}, e^{j\omega_2}) = \begin{cases} 1, & \rho(\omega_1, \omega_2) \leq R_1 \\ 0, & \rho(\omega_1, \omega_2) \geq R_2 \end{cases} \tag{33}$$

the DFT values that occur inside the passband are given a value of 1.0; those that occur in the stopband are given a value 0.0; those that lie in the transition band are free variables whose value will be chosen by a minimization criterion.

To simplify both the equations and the computation involved in performing the desired minimizations, the following assumptions will be used in the remainder of this paper. First, we assume $N_1 = N_2 =$ odd integer. Next, we assume

$$H(k_1, k_2) = H(k_2, k_1) \tag{34}$$

i.e., there is symmetry across the 45° diagonals in the ω_1, ω_2 plane. This is a reasonable assumption for designing circularly symmetric filters as (34) is valid in the ideal case of zero approximation error. Finally, we assume the linear phase term in (25) can be disregarded in terms of the design procedure.

With these assumptions in mind, (25) can be simplified to the form

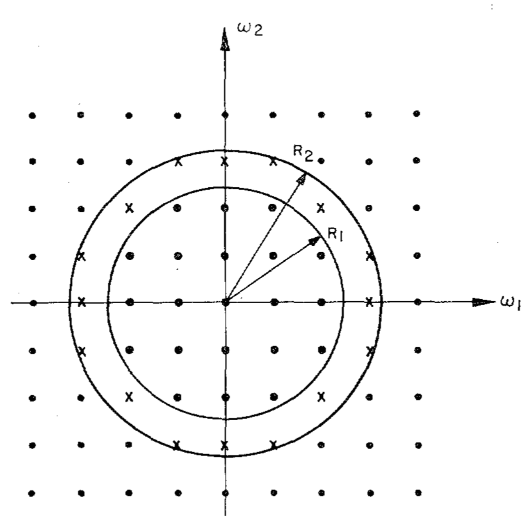

Fig. 2. Locations of frequency samples (DFT points) for a 9 by 9 grid.

$$H(e^{j\omega_1}, e^{j\omega_2}) = H_F(\omega_1, \omega_2) + \sum_{m=1}^{M} T_m H_m(\omega_1, \omega_2) \tag{35}$$

where $H_F(\omega_1, \omega_2)$ represents the contribution of the fixed DFT coefficients (the 1.0's), T_m represents the amplitude of the mth DFT coefficient in the transition band, $H_m(\omega_1, \omega_2)$ represents the interpolation function appropriate to the mth transition coefficient, and M is the total number of transition coefficients. The pertinent design equations can now be put in the form

$$1 - \alpha\delta \leq H(e^{j\omega_1}, e^{j\omega_2}) \leq 1 + \alpha\delta, \quad \text{for } \rho(\omega_1, \omega_2) \leq R_1 \tag{36}$$

$$-\delta \leq H(e^{j\omega_1}, e^{j\omega_2}) \leq \delta, \quad \text{for } \rho(\omega_1, \omega_2) \geq R_2 \tag{37}$$

where δ represents the peak approximation error in the stopband, and $\alpha\delta$ (α any constant) represents the peak approximation error in the passband. The design goal is to choose the transition samples so as to minimize δ.

The above problem is readily seen to be a linear programming [7] problem by evaluating (36) and (37) at a dense set of points in the passband and stopband and by writing the equations explicitly as linear inequalities in the form

$$\left.\begin{aligned} \sum_{m=1}^{M} T_m H_m(\omega_1, \omega_2) - \alpha\delta &\leq 1 - H_F(\omega_1, \omega_2) \\ -\sum_{m=1}^{M} T_m H_m(\omega_1, \omega_2) - \alpha\delta &\leq -1 + H_F(\omega_1, \omega_2) \end{aligned}\right\} \quad \text{for all points with } \rho(\omega_1, \omega_2) \leq R_1 \tag{38}$$

$$\left.\begin{aligned} \sum_{m=1}^{M} T_m H_m(\omega_1, \omega_2) - \delta &\leq -H_F(\omega_1, \omega_2) \\ -\sum_{m=1}^{M} T_m H_m(\omega_1, \omega_2) - \delta &\leq H_F(\omega_1, \omega_2) \end{aligned}\right\} \quad \text{for all points with } \rho(\omega_1, \omega_2) \geq R_2 \tag{39}$$

choosing $(T_1, T_2, \cdots, T_M)$ to minimize δ.

TABLE I
Stopband Attenuation for Low-Pass Filters
(Filter Attenuation in Decibels)

	$(R_2-R_1)\cdot(12.5/\pi)$				
	1	1.5	2	2.5	3
$R_1\cdot(12.5/\pi)$ 0			−42.8		−69.28
1			−37.37		−65.08
2			−42.06		
3			AF		−67.92
4			−38.56		
5			−41.83		NC
6			AF		
7			AF		
8			NC		
8.5				NC	
9			−44.68		
9.5		−31.85		NC	
10	−14.42	NC	−48.35		
10.5	−21.87	−33.72			
11	−11.25				

AF = algorithm failed for undetermined reasons.
NC = convergence was not attained within a specified number of iterations.

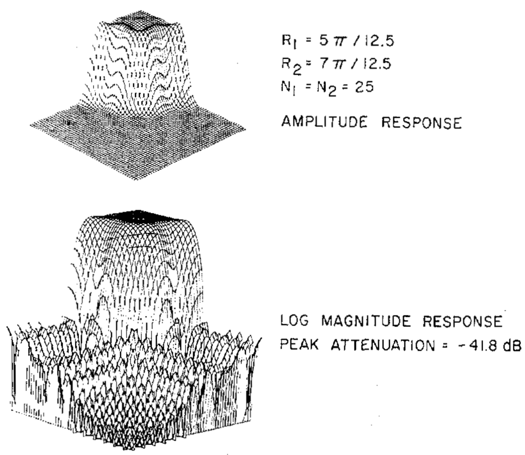

Fig. 3. Amplitude and log magnitude response of a typical low-pass filter for a transition width of $R_2-R_1=2\pi/12.5$.

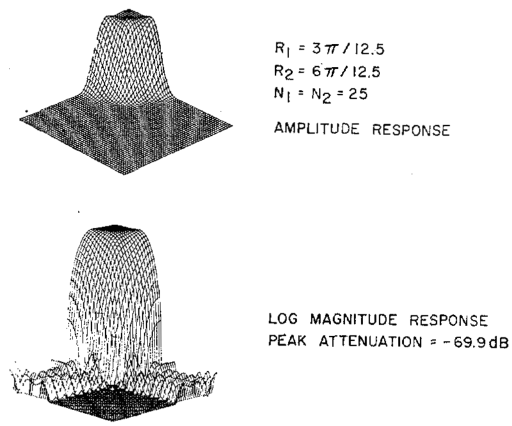

Fig. 4. Amplitude and log magnitude response of a typical low-pass filter for a transition width of $R_2-R_1=3\pi/12.5$.

Results of Low-Pass Filter Design

The linear programming problem of (38) and (39) was solved on a large central computer (GE-635) using the scientific subroutine APMM [12]. A 64–1 interpolation was used throughout the stopband and passband to evaluate $H(e^{j\omega_1}, e^{j\omega_2})$ over a fine grid. Table I presents a summary of results[3] obtained for the case where $N_1=N_2=N=25$ and $\alpha \gg 1$, i.e., in-band ripple was not constrained. The entries in this table give values for 20 $\log_{10}\delta$ as a function of R_1, the in-band radius, and R_2-R_1, the width of the transition band. For many of the entries in the table, the linear program subroutine did not converge within a specified number of iterations (on the order of 200) and the run was terminated. Each filter that was designed required on the order of 6-min computation time to determine the transition coefficients. This was due to the large number of constraint equations (on the order of several thousand), even though the number of variable coefficients was small (on the order of 10).

If the passband ripple is unconstrained, the results of Table I indicate that a transition width of $2(\pi/12.5)$ gives about 40-dB stopband attenuation, whereas a width of $3(\pi/12.5)$ gives about 60-dB attenuation. Figs. 3 and 4 show two-dimensional perspectives [14] of the frequency response of two of the filters in Table I. Fig. 3 shows linear and log magnitude responses for a filter with $R_1=5\pi/12.5$, $R_2-R_1=2\pi/12.5$. Fig. 4 shows linear and log magnitude responses for a filter with $R_1=3\pi/12.5$, $R_2-R_1=3\pi/12.5$. These figures clearly show that the peak in-band ripple occurs at the edges of the passband, and the peak out-of-band ripple occurs at the edges of the stopband. (The filter response only in the region $0\leq\omega_1$, $\omega_2\leq\pi$ is shown in these figures as the response in the rest of the ω_1, ω_2 plane is determined from symmetry considerations.)

[3] The exact filter coefficients corresponding to the entries in Table I are available in [13].

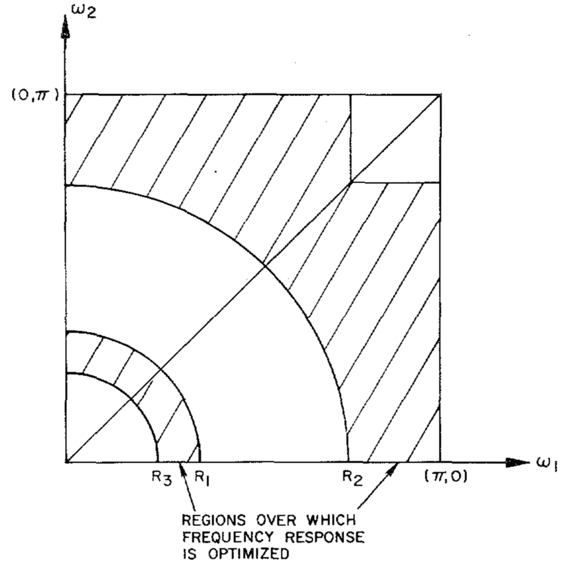

Fig. 5. Modified regions in the (ω_1, ω_2) plane over which the filter response is optimized for a circularly symmetric low-pass filter.

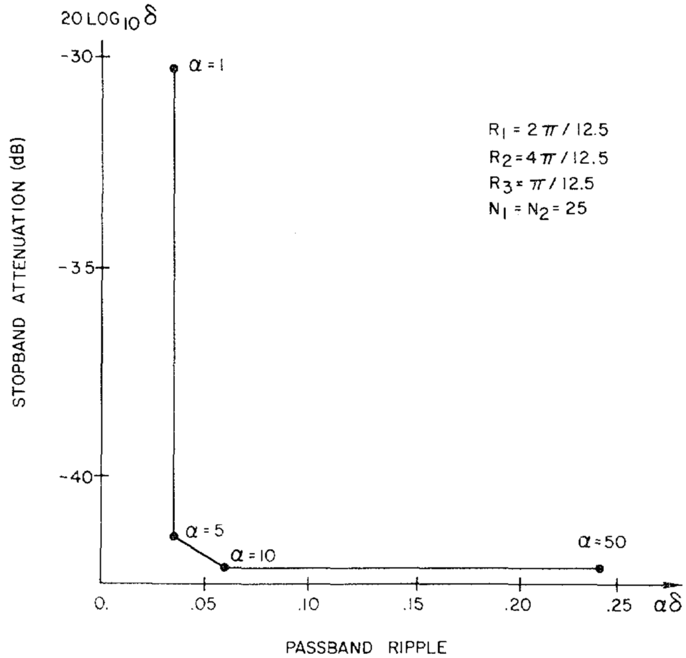

Fig. 6. Plot of tradeoffs obtainable between passband ripple and peak stopband attenuation for a fixed transition width.

Since the worst ripples occurred near the edges of the transition band, a new set of filters was designed using the constraint on passband ripple. In an effort to save computation, the filter response was optimized over the region shown in Fig. 5. The regions $0 \le \rho(\omega_1, \omega_2) \le R_3$ and $\omega_1 > \omega_2 > R_2$ were unconstrained because the ripple in the passband and stopband becomes quite small far from the band edges. A filter was designed with $R_3 = \pi/12.5$, $R_1 = 2\pi/12.5$, $R_2 = 4\pi/12.5$, and with $N_1 = N_2 = 25$. Values of α of 50, 10, 5, and 1 were used. Fig. 6 gives a plot of the measured tradeoffs between the stopband ripple 20 $\log_{10} \delta$ and the passband ripple $\alpha\delta$. This curve shows that considerable improvement can be made in passband ripple (a reduction from 0.24 to 0.04) with hardly any change in stopband attenuation (0.8 dB). On the basis of these results, it would seem worth the computational effort to constrain the in-band ripple at the edges of the passband. Figs. 7–10 show the frequency responses for the cases $\alpha = 50$ and $\alpha = 1$, as well as magnitude contour plots (magnitude$\ge$0.75) for these two cases. In the contour plots, a dotted contour is drawn each 0.05. The passband and the transition region for the $\alpha = 1$ case are significantly more circularly symmetric than for the $\alpha = 50$ case, because the $\alpha = 50$

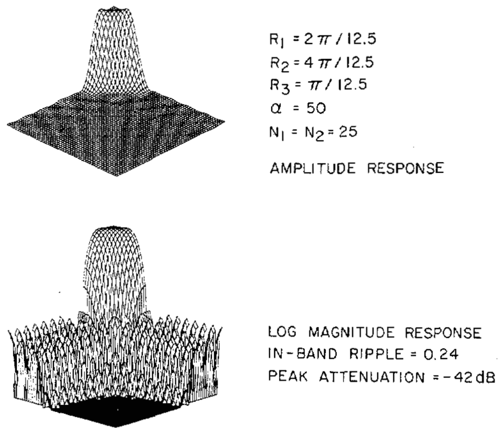

Fig. 7. Amplitude and log magnitude response of a low-pass filter with $\alpha = 50$.

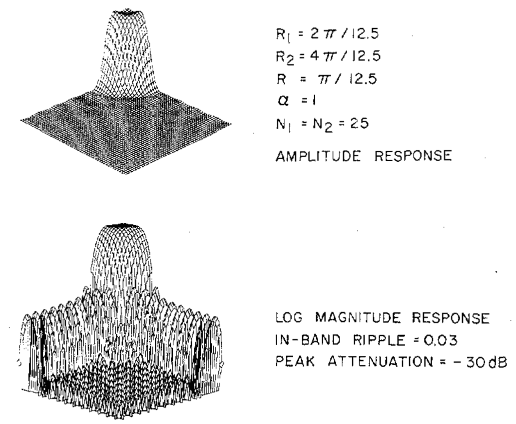

Fig. 8. Amplitude and log magnitude response of a low-pass filter with $\alpha = 1$.

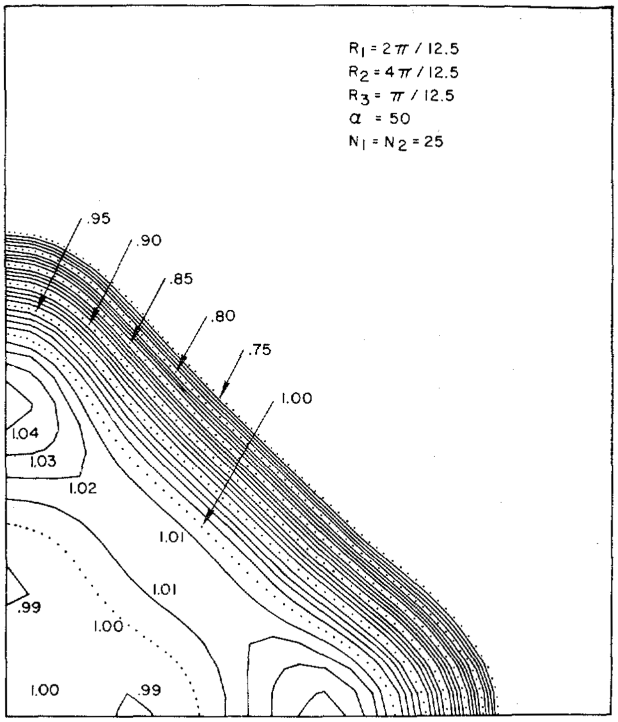

Fig. 9. Contour plot of passband and transition band response for the filter of Fig. 7.

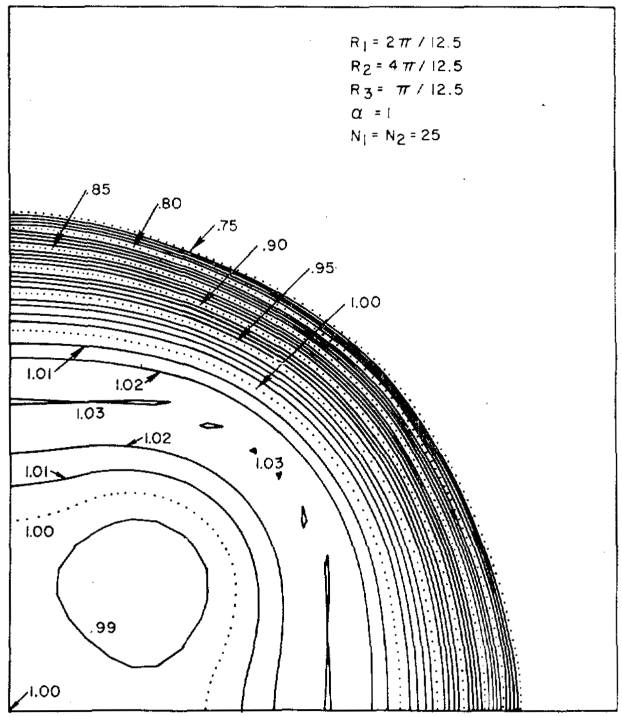

Fig. 10. Contour plot of passband and transition band response for the filters of Fig. 8.

case essentially has an unconstrained passband and thus the requirements for maintaining a good approximation to circular symmetry in the passband are not met.

Design of Optimal (Equiripple) Filters

By allowing all the impulse response coefficients to vary, or equivalently all the DFT coefficients, filters can be designed which are optimal in the sense of Chebyshev approximation over closed regions. For the case when all the DFT coefficients are varied, as was done here, the filter frequency response is evaluated using (25) and treating all the filter coefficients $H(k_1, k_2)$ as unknowns. Optimal approximations to circularly symmetric filters may be designed (using linear programming techniques) by evaluating (25) at a dense set of points in both the passband and the stopband and minimizing both the stopband ripple δ and the passband ripple $\alpha\delta$.

Because of the increased amount of computation, necessitated both by the need to evaluate $H(e^{j\omega_1}, e^{j\omega_2})$ over the entire ω_1, ω_2 plane (except, of course, the transition band) and by the increased number of variables, the size of the impulse responses considered was reduced to $N_1 = N_2 = 9$. Within these limitations, four optimal (equiripple) filters were designed. A summary of the results for these four cases is given in Table II. Figs. 11 and 12 show the frequency response of the

TABLE II
Optimal Filters

Filter 1	$R_1=2\pi/4.5$, $R_2=3\pi/4.5$, $\alpha=1$ in-band ripple = 0.0867 out-of-band attenuation = −21.24 dB
Filter 2	$R_1=\pi/4.5$, $R_2=2\pi/4.5$, $\alpha=10$ in-band ripple = 0.287 out-of-band attenuation = −30.84 dB
Filter 3	$R_1=1.5\pi/4.5$, $R_2=3\pi/4.5$, $\alpha=10$ in-band ripple = 0.079 out-of-band attenuation = −41.97 dB
Filter 4	$R_1=2\pi/4.5$, $R_2=3\pi/4.5$, $\alpha=10$ in-band ripple = 0.235 out-of-band attenuation = −32.51 dB

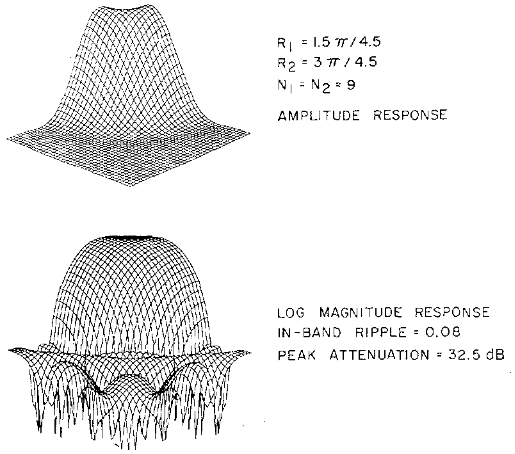

Fig. 11. Amplitude and log magnitude response of an optimal low-pass filter.

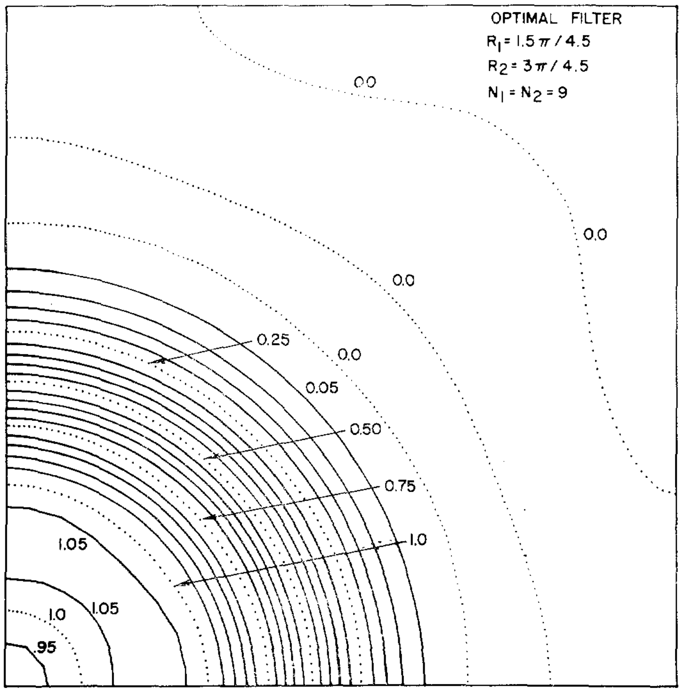

Fig. 12. Contour plot of the response of the filter of Fig. 11.

third filter in Table II and a contour plot of the log magnitude response, respectively. As the contour lines in the passband and transition region show, the passband response is a good approximation to a circularly symmetric response. Fig. 11 also shows the interesting phenomenon that there are equiripple ridges, as well as peaks. Thus the ripple of the frequency response may attain a maximum or minimum in a continuous region of the ω_1, ω_2 plane (corresponding to the equiripple ridge), as well as at isolated unconnected values. This is an added complexity to the one-dimensional design case where, because of the well-ordered frequency scale, the ripple peaks alternate between maxima and minima. This phenomenon suggests that simply counting points at which $H(e^{j\omega_1}, e^{j\omega_2})$ attains a maximum value can lead to significant errors in designing equiripple filters. For this reason several of the one-dimensional design algorithms [3], [5], [6] are not readily extended to two dimensions.

Conclusions

This paper represents preliminary work on techniques for designing two-dimensional digital filters. It shows that the frequency sampling and optimal (equiripple) design methods can be carried over to the two-dimensional design problem directly, although at considerable computational cost. Future efforts along these lines should be aimed at lowering these costs.

Acknowledgment

The authors wish to thank Mrs. N. Graham for her work in devising a novel algorithm for making the perspective plots shown.

References

[1] J. F. Kaiser, "Digital filters," in *Systems Analysis by Digital Computer*, F. Kuo and J. F. Kaiser, Eds. New York: Wiley, 1966, ch. 7.

[2] B. Gold and C. M. Rader, *Digital Processing of Signals*. New York: McGraw-Hill, 1969.

[3] H. W. Schuessler, "On the approximation problem in the design of digital filters," in *Proc. 5th Annual Princeton Conf. Information Sciences and Systems*, Mar. 1971, pp. 54–63.
[4] L. R. Rabiner, B. Gold, and C. A. McGonegal, "An approach to the approximation problem for nonrecursive digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 83–106, June 1970.
[5] E. Hofstetter, A. V. Oppenheim, and J. Siegel, "A new technique for the design of nonrecursive digital filters," in *Proc. 5th Annual Princeton Conf. Information Sciences and Systems*, Mar. 1971, pp. 64–72.
[6] T. W. Parks and J. H. McClellan, "Chebyshev approximation for nonrecursive digital filters with linear phase," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 189–194, Mar. 1972.
[7] L. R. Rabiner, "The design of finite impulse response digital filters using linear programming techniques," *Bell Syst. Tech. J.*, Aug. 1972.
[8] ——, "Techniques for designing finite-duration impulse-response digital filters," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 188–195, Apr. 1971.
[9] J. L. Shanks, "The design of stable two-dimensional recursive filters," in *Proc. M. J. Kelly Communications Conf.*, Univ. of Missouri, Columbia, 1970.
[10] ——, "Two-dimensional recursive filters," in *SWIEECO Rec.*, 1969, pp. 19E1–19E8.
[11] T. S. Huang, "Two-dimensional windows," *IEEE Trans. Audio Electroacoust.* (Corresp.), vol. AU-20, pp. 88–89, Mar. 1972.
[12] *IBM System/360 Programming Manual*, Scientific Subroutine Package (360 A-CM-03 X), Fortran Program APMM, pp. 283–288.
[13] J. Hu, "Frequency sampling design of two-dimensional finite impulse response digital filters," M.S. thesis, Mass. Inst. Tech., Cambridge, June 1972.
[14] N. Graham, "Perspective drawing of surfaces with hidden line elimination," *Bell Syst. Tech. J.*, vol. 51, pp. 843–861, Apr. 1972.

Synthesis of Recursive Digital Filters Using the Minimum p-Error Criterion

ANDREW G. DECZKY

Abstract—The problem of designing a stable recursive digital filter to have an arbitrarily prescribed frequency response may be considered as an approximation problem. Using the minimum p-error criterion, a new problem of minimizing a function of n variables results, which is successfully solved using the Fletcher–Powell algorithm. An important theorem guaranteeing the existence of a stable optimum for a large class of synthesis problems is stated, and necessary modifications to the Fletcher–Powell algorithm to assure stability are considered. Finally a number of results of the application of this method are given.

I. The Approximation Problem

The problem of designing a recursive digital filter to have an arbitrarily prescribed frequency response may be regarded as a classical approximation problem. The advantage of this approach is that many methods of solution for such problems exist.

The classical approximation problem may be stated as follows [14]. Let $f(x)$ be a given real-valued function defined on a set X, and let $F(A, x)$ be a real-valued approximating function depending continuously on $x \in X$ and on n parameters A. Given a distance function ρ, determine the n parameters $A^* \in P$ such that

$$\rho[F(A^*, x), f(x)] \leq \rho[F(A, x), f(x)], \quad \text{for all } A \in P. \quad (1)$$

The solution to this problem is said to be a best approximation with respect to the distance function ρ chosen.

For recursive digital filter synthesis we have the following identification. Let the transfer function $H(z)$ be a function of n parameters (e.g., the filter coefficients) and order these in a vector A. The independent variable x is now the digital frequency $\phi = \omega T$, and the set X is $(\phi: 0 \leq \phi \leq \pi)$. Then the frequency response of the filter (such as magnitude or group delay) may be expressed as a real-valued function of A and ϕ, i.e., $F(A, \phi)$, while the desired frequency response becomes $f(\phi)$.

Having defined the problem in these general terms, we must now make specific choices for the distance function ρ, the form of the filter transfer function $H(z)$, the parameter vector A, and the method of solution (this depending on the distance function ρ chosen).

II. The Distance Function

The distance function we have chosen is one of the most often used, namely the weighted L_p norm

$$\|L(A)\|_p = \left\{ \int_0^\pi w(\phi) \,|\, F(A, \phi) - f(\phi) \,|^p d\phi \right\}^{1/p} \quad (2)$$

where $w(\phi)$ is a positive weighting function. The reason for this choice is that L_p approximations have been extensively studied, and their properties are well known. Furthermore, the problem of finding the best L_p approximation numerically is transformed into the minimization of a function of n variables, for which very powerful algorithms exist. Finally the two most often used error criteria, namely the minimum square and the minimax are included as special cases of this distance

Manuscript received March 7, 1972. This work was supported by the Swiss National Science Foundation. This paper is part of a dissertation submitted to the Institute of Applied Physics, Swiss Federal Institute of Technology, Zurich, in partial fulfillment of the requirements for the Ph.D. degree. This paper was presented at the IEEE Arden House Workshop on Digital Filtering, Harriman, N. Y., January 1972.

The author is with the Institute of Applied Physics, Swiss Federal Institute of Technology, Zurich 8049, Switzerland.