TABLE III

Iteration	$\hat{K}$	$\hat{\delta}_1$	KM	JD
1	601.0	0.025003	2.0	0
2	200.4	0.015567	2.0	1
3	66.8	0.007426	2.0	1
4	133.6	0.012181	1.0	-1
5	89.0	0.009207	0.5	1
6	111.3	0.010786	0.25	-1
7	98.9	0.009933	0.125	1

Finally, for Case 5, δ_1 was unspecified. The initial estimate $\hat{\delta}_1$ was 0.002655 or $\hat{K} = 26.55$. It required ten iterations for $\hat{\delta}_2$ to be within 1 percent of δ_2. The run time for this case was about 1.0 s.

Summary

An optimal FIR low-pass filter can now be designed where any four of the five parameters N, F_p, F_s, δ_1, and δ_2 are specified, and the remaining parameter is chosen so as to meet or exceed specifications on all given parameters. A set of simple, approximate formulas was given for obtaining initial estimates of the unspecified parameter. Finally, simple iterative rules were given for varying the unspecified parameter from its initial estimate so as to meet input specifications to within a given tolerance.

References

[1] J. F. Kaiser, "Design methods for sampled data filters," in *Proc. 1st Allerton Conf. Circuit and System Theory*, 1963, pp. 221-236.
[2] L. R. Rabiner, "Techniques for designing finite-duration impulse-response digital filters," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 188-195, Apr. 1971.
[3] O. Herrmann, "Design of nonrecursive digital filters with linear phase," *Electron. Lett.*, vol. 6, no. 11, pp. 328-329, 1970.
[4] 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.
[5] L. R. Rabiner, "The design of finite impulse response digital filters using linear programming techniques," *Bell Syst. Tech. J.*, vol. 51, pp. 1177-1198, July-Aug. 1972.
[6] T. W. Parks and J. H. McClellan, "A program for the design of linear phase finite impulse response digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 195-199, Aug. 1972.
[7] O. Herrmann, L. R. Rabiner, and D. S. K. Chan, "Practical design rules for optimum finite impulse response low-pass digital filters," *Bell Syst. Tech. J.*, vol. 52, no. 6, pp. 769-799, July-Aug. 1973.

Least *p*th Optimization of Recursive Digital Filters

JOHN W. BANDLER and BERJ L. BARDAKJIAN

Abstract—**The application of the Bandler–Charalambous method using extremely large values of *p*, typically 10 000 to the problem of choosing the coefficients of a recursive digital filter to meet arbitrary specifications on the magnitude characteristics, is described. The Fletcher (1970) method is used in conjunction with least *p*th optimization and is compared with the well-known Fletcher–Powell method. Some relevant design ideas, such as local optimality checking by perturbation, increasing the order complexity of the filter through growing filter sections, and meeting the stability requirements by using a pole inversion technique, have been implemented. A general description of a computer program package that uses these ideas, along with some illustrative examples are given.**

Manuscript received November 10, 1972; revised May 7, 1973. This work was supported by the National Research Council of Canada under Grants A7239 and C154. This paper was presented at the 1973 International Symposium on Circuit Theory, Toronto, Canada, April 9-11, 1973.
The authors are with the Department of Electrical Engineering, McMaster University, Hamilton, Ont., Canada.

I. Introduction

Two main approaches have been taken to approximation problems in digital filter design. The first of these is an analytical approach through classical approximation theories [1]-[3]. The second is an iterative approach that is particularly appropriate for use on a digital computer [4]-[6]. Sablatash [7] discussed many contributions to both approaches.

Haykin [3] presented a unified treatment of recursive digital filtering by using the convolution integral to derive an integro-difference equation for defining the input-output relation of a linear time invariant filter. Then he used that equation to obtain various analog-to-digital filter transformations for the digitization of a continuous transfer function, with each transformation corresponding to a specific way of approximating the continuous time excitation.

An iterative method for designing recursive digital filters with arbitrary prescribed magnitude characteristics was described by Steiglitz [4]. The method uses the Fletcher-Powell algorithm [8] to minimize a square-error criterion in the frequency domain. A strategy was described whereby stability and minimum phase constraints were observed, while still using the unconstrained optimization algorithm.

Helms [5] reviewed and occasionally extended techniques for determining the coefficients of digital filters that have equiripple or minimax errors. Deczky [6] outlined a method for the design of recursive digital filters using a weighted minimum p-error criterion. The largest value of p tried was reported to have been 40, but the largest p shown in the given examples seemed to be 10.

A new and practical approach to computer-aided design optimization was presented by Bandler and Charalambous [9]. Central to the process was the application of least pth approximation using extremely large values of p, typically 1000 to 1 000 000. It was shown how suitable and reasonably well-conditioned objective functions could be formulated, giving particular emphasis to general approximation problems as, for example, in filter design. It was demonstrated how easily and efficiently extremely near minimax results could be achieved on a discrete set of sample points.

This paper describes the application of the Bandler–Charalambous method for choosing the coefficients of a recursive digital filter to meet arbitrary specifications of the magnitude characteristics. The local optimality of the least pth solution is checked by perturbation. The order complexity of the filter can be increased through growing of filter sections to meet the prescribed specifications. A pole inversion technique [4] to meet the stability requirements is also implemented. A comparison is made between the Fletcher–Powell method [8] and the more recent Fletcher method [10] in conjunction with the application of least pth optimization to a recursive digital filter design example. An example where effectively negative values of p were used is also presented.

II. Description of the Problem

Suppose that upper and/or lower bounds on the magnitude characteristics of a recursive digital filter are prescribed at a discrete set of frequencies f_1, $f_2, \cdots, f_m$. These correspond to a discrete set of values of the variable z evaluated on the unit circle in the z domain:

$$z_i = e^{j\psi_i \pi}, \qquad i = 1, 2, \cdots, m \tag{1}$$

where

$$\psi_i = \frac{2f_i}{f_s}, \qquad i = 1, 2, \cdots, m \tag{2}$$

and f_s is the sampling frequency.

The transfer function of a recursive digital filter is chosen usually to be either of the cascade form or the parallel form [5], namely,

$$H(z) = A \prod_{k=1}^{K} \frac{1 + a_k z^{-1} + b_k z^{-2}}{1 + c_k z^{-1} + d_k z^{-2}} \tag{3}$$

or

$$H(z) = A + \sum_{k=1}^{K} \frac{a_k + b_k z^{-1}}{1 + c_k z^{-1} + d_k z^{-2}}. \tag{4}$$

All the poles of the transfer function should lie within the unit circle in the z domain in order that the filter be stable.

III. Problem Formulation

Find the n-dimensional parameter vector $\phi = [a_1 b_1 c_1 d_1 a_2 b_2 c_2 d_2 \cdots A]^T$, where $n = 4K + 1$, to minimize an appropriately chosen objective function comprising real error functions related to the upper and lower specified bounds.

Definitions [9]

$F(\phi, \psi)$	the approximating function ($F(\phi, \psi) = \lvert H(\phi, z)\rvert$);
$S_u(\psi)$	an upper specified function (desired response bound);
$S'_u(\psi, \xi)$	an artificial upper specified function;
$S_l(\psi)$	a lower specified function (desired response bound);
$S'_l(\psi, \xi)$	an artificial lower specified function;
$w_u(\psi)$	an upper positive weighting function;
$w_l(\psi)$	a lower positive weighting function; and
ξ	margin of errors with respect to the artificial and desired specifications.

$$e_u(\phi, \psi) \triangleq w_u(\psi)\,(F(\phi, \psi) - S_u(\psi)) \tag{5}$$

$$e'_u(\phi, \psi, \xi) \triangleq w_u(\psi)\,(F(\phi, \psi) - S'_u(\psi, \xi)) = e_u(\phi, \psi) - \xi \tag{6}$$

$$e_l(\phi, \psi) \triangleq w_l(\psi)\,(F(\phi, \psi) - S_l(\psi)) \tag{7}$$

$$e'_l(\phi, \psi, \xi) \triangleq w_l(\psi)\,(F(\phi, \psi) - S'_l(\psi, \xi)) = e_l(\phi, \psi) + \xi \tag{8}$$

where $S'_u(\psi, \xi)$ and $S'_l(\psi, \xi)$ are taken, respectively, as

$$S'_u(\psi, \xi) = S_u(\psi) + \frac{\xi}{w_u(\psi)} \tag{9}$$

$$S'_l(\psi, \xi) = S_l(\psi) - \frac{\xi}{w_l(\psi)}. \tag{10}$$

All the functions will be evaluated at a finite discrete set of values of ψ taken from one or more closed intervals. Therefore we define the functions

$$e'_{ui}(\phi, \xi) \triangleq e'_u(\phi, \psi_i, \xi), \qquad i \in I_u \tag{11}$$

$$e'_{li}(\phi, \xi) \triangleq e'_l(\phi, \psi_i, \xi), \qquad i \in I_l \tag{12}$$

where I_u and I_l are appropriate index sets related to all the specified discrete frequencies.

The following objective function proposed by Bandler and Charalambous [9] will be used:

$$U(\phi, \xi) = M(\phi, \xi) \left(\sum_{i \in K_u} \left[\frac{e'_{ui}(\phi, \xi)}{M(\phi, \xi)} \right]^q + \sum_{i \in K_l} \left[- \frac{e'_{li}(\phi, \xi)}{M(\phi, \xi)} \right]^q \right)^{1/q} \tag{13}$$

where

$$M(\phi, \xi) \triangleq \max_{i,j} \; [e'_{ui}(\phi, \xi), -e'_{lj}(\phi, \xi)], \qquad i \in I_u, \; j \in I_l \tag{14}$$

$$J_u(\phi, \xi) \triangleq \{ i \mid e'_{ui}(\phi, \xi) \geqslant 0, \quad i \in I_u \} \tag{15}$$

$$J_l(\phi, \xi) \triangleq \{ i \mid -e'_{li}(\phi, \xi) \geqslant 0, \quad i \in I_l \} \tag{16}$$

$$K_u \triangleq \begin{cases} J_u(\phi, \xi) & \text{if } M(\phi, \xi) > 0 \\ I_u & \text{if } M(\phi, \xi) < 0 \end{cases} \tag{17}$$

$$K_l \triangleq \begin{cases} J_l(\phi, \xi) & \text{if } M(\phi, \xi) > 0 \\ I_l & \text{if } M(\phi, \xi) < 0 \end{cases} \tag{18}$$

and

$$q \triangleq p \operatorname{sgn} M(\phi, \xi) \qquad \begin{cases} 1 < p < \infty & \text{for } M > 0 \\ 1 \leqslant p < \infty & \text{for } M < 0. \end{cases} \tag{19}$$

Differentiating the objective function we get

$$\nabla U(\phi,\xi) = \left(\sum_{i \in K_u} \left[\frac{e'_{ui}(\phi,\xi)}{M(\phi,\xi)} \right]^q + \sum_{i \in K_l} \left[\frac{-e'_{li}(\phi,\xi)}{M(\phi,\xi)} \right]^q \right)^{1/q - 1} \cdot \left(\sum_{i \in K_u} \left[\frac{e'_{ui}(\phi,\xi)}{M(\phi,\xi)} \right]^{q-1} \nabla e'_{ui}(\phi,\xi) - \sum_{i \in K_l} \left[\frac{-e'_{li}(\phi,\xi)}{M(\phi,\xi)} \right]^{q-1} \nabla e'_{li}(\phi,\xi) \right), \tag{20}$$

where

$$\nabla = \begin{bmatrix} \dfrac{\partial}{\partial a_1} \\ \dfrac{\partial}{\partial b_1} \\ \dfrac{\partial}{\partial c_1} \\ \dfrac{\partial}{\partial d_1} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{\partial}{\partial A} \end{bmatrix}. \tag{21}$$

If the transfer function is chosen to be of the cascade form we get

$$\frac{\partial}{\partial a_k} e'_{ui}(\phi, \xi) = w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{z_i^{-1}}{1 + a_k z_i^{-1} + b_k z_i^{-2}} \right\} \tag{22}$$

$$\frac{\partial}{\partial b_k} e'_{ui}(\phi, \xi) = w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{z_i^{-2}}{1 + a_k z_i^{-1} + b_k z_i^{-2}} \right\} \tag{23}$$

$$\frac{\partial}{\partial c_k} e'_{ui}(\phi, \xi) = -w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{z_i^{-1}}{1 + c_k z_i^{-1} + d_k z_i^{-2}} \right\} \tag{24}$$

$$\frac{\partial}{\partial d_k} e'_{ui}(\phi, \xi) = -w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{z_i^{-2}}{1 + c_k z_i^{-1} + d_k z_i^{-2}} \right\} \tag{25}$$

$$\frac{\partial}{\partial A} e'_{ui}(\phi, \xi) = w_u(\psi_i) \frac{|H(\phi, z_i)|}{A} \tag{26}$$

$$i = 1, 2, \cdots, m \quad \text{and} \quad k = 1, \cdots, K.$$

On the other hand, if the transfer function is chosen to be of the parallel form we get

$$\frac{\partial}{\partial a_k} e'_{ui}(\phi, \xi) = w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{1}{H(\phi, z_i) \, [1 + c_k z_i^{-1} + d_k z_i^{-2}]} \right\} \tag{27}$$

$$\frac{\partial}{\partial b_k} e'_{ui}(\phi, \xi) = w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{z_i^{-1}}{H(\phi, z_i) \, [1 + c_k z_i^{-1} + d_k z_i^{-2}]} \right\} \tag{28}$$

$$\frac{\partial}{\partial c_k} e'_{ui}(\phi, \xi) = -w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{a_k z_i^{-1} + b_k z_i^{-2}}{H(\phi, z_i) \, [1 + c_k z_i^{-1} + d_k z_i^{-2}]^2} \right\} \tag{29}$$

$$\frac{\partial}{\partial d_k} e'_{ui}(\phi, \xi) = -w_u(\psi_i) \, |H(\phi, z_i)| \cdot \operatorname{Re} \left\{ \frac{a_k z_i^{-2} + b_k z_i^{-3}}{H(\phi, z_i) \, [1 + c_k z_i^{-1} + d_k z_i^{-2}]^2} \right\} \tag{30}$$

$$\frac{\partial}{\partial A} e'_{ui}(\phi, \xi) = w_u(\psi_i) \, |H(\phi, z_i)| \operatorname{Re} \left\{ \frac{1}{H(\phi, z_i)} \right\} \tag{31}$$

$$i = 1, 2, \cdots, m \quad \text{and} \quad k = 1, \cdots, K.$$

Using either one of the two forms of the transfer function, we get

$$\nabla e'_{li}(\phi, \xi) = \frac{w_l(\psi_i)}{w_u(\psi_i)} \nabla e'_{ui}(\phi, \xi), \qquad i = 1, 2, \cdots, m. \tag{32}$$

IV. Stability Constraints

Suppose that $H(z)$ has a pole at $z = z_p$ located outside the unit circle in the z domain, such that

$$H(z) = \frac{N(z)}{z - z_p}. \tag{33}$$

Consider the function

$$Q(z_p) = \frac{z - z_p}{z - \dfrac{1}{z_p^*}} \tag{34}$$

where the asterisk denotes the complex conjugate, and where $|Q(z_p)| = |z_p|$ when $|z| = 1$. Let

$$H'(z) = H(z) \times Q(z_p) = \frac{N(z)}{z - \dfrac{1}{z_p^*}}. \tag{35}$$

Hence, the inversion of real or complex conjugate pairs of poles of the transfer function with respect to the unit circle in the z domain is equivalent to multiplying the transfer function by a particular all-pass function, implying that the inversion of such poles of the transfer function with respect to the unit circle does not affect the shape of the magnitude characteristics. Thus, all such poles that do not lie within the unit circle in the z domain can be inverted with respect to the unit circle to ensure the stability of the filter. It is to be noted that the magnitude characteristics of the stable filter should be divided by the magnitude of each inverted pole, that is, if there are r poles $z_{p1}, \cdots, z_{pr}$ that do not lie within the unit circle in the z domain, then, after their inversion, we get

$$|H(z)| = \frac{|H'(z)|}{\prod_{i=1}^{r} |z_{pi}|} \tag{36}$$

V. Description of the Program

A general computer program package CADRDF was developed utilizing the aforementioned ideas. The user should specify the desired form of the transfer function, the initial number of second-order filter sections, the maximum acceptable number of filter sections, the required precision, and an option for local optimality checking either by perturbation of the parameters or by increasing ξ and then restarting the optimization process. The user should also specify the upper and lower bounds on the magnitude response, the discrete set of frequency points, and the required optimization algorithm (Fletcher method [10] or Fletcher–Powell method [8]).

Starting with the initial number of filter sections, the coefficients of the digital filter are evaluated by minimizing the objective function (13) using the chosen optimization algorithm. It is to be noted that the value of p is increased successively, and the optimization is carried out for each subsequent value of p, until the absolute value of the relative change in maximum error

$$\left| \frac{\Delta M(\phi, \xi)}{M(\phi, \xi)} \right|$$

becomes less than some small quantity (taken to be 0.001). To ensure the stability of the filter a stability checking is provided whereby all the poles that do not lie within the unit circle in the z domain are inverted with respect to the unit circle, as discussed in Section IV.

Local optimality of the solution $\check{\phi}$ is checked by either perturbing $\check{\phi}$ or increasing ξ, restarting the optimization process using the highest attained value of p and comparing the solutions before and after perturbation. It is to be noted that a minimax optimum will not be affected by increasing ξ [11].

If the specifications are not satisfied and the maximum specified number of second-order filter sections has not been exceeded, a second-order filter section is grown by increasing the number of the independent parameters n by 4, assigning a starting value of zero for each of the grown filter coefficients a_k, b_k, c_k and d_k, then repeating the whole design process.

VI. Examples

Example 1

Consider the design of a low-pass digital filter of the cascade form whose ideal magnitude response is specified by

$$\text{ideal magnitude response} = \begin{cases} 1 & \psi \in W_p \\ 0 & \psi \in W_s \end{cases}$$

where W_p = [0.0, 0.09] is the passband and W_s = [0.11, 1.0] is the stopband.

Let

$$\begin{aligned} S_u(\psi) = S_l(\psi) &= 1 \qquad \psi \in W_p \\ S_u(\psi) &= 0 \qquad \psi \in W_s \\ w_u(\psi) = w_l(\psi) &= 1 \qquad \psi \in W_p \cup W_s. \end{aligned}$$

All the functions of ψ will be evaluated at a finite discrete set of values of ψ taken from the closed intervals W_p and W_s as follows:

TABLE I
Results for Example 1 Using One Section

$q = p$	Maximum Error	Objective Function	Solution $\check{\phi}$	Number of Function Evaluations: Fletcher Method	Number of Function Evaluations: Fletcher–Powell Method
2	0.446136	0.905656	-1.672551 1. -1.790717 0.850218 0.151727	106	393
10	0.278126	0.308162	-1.799881 1. -1.831101 0.889290 0.221937	33	68
100	0.248415	0.251337	-1.815636 1. -1.836826 0.895616 0.240215	27	97
1000	0.246809	0.247100	-1.816441 1. -1.837210 0.896016 0.241345	37	48
10000	0.246713	0.246741	-1.816464 0.999999 -1.837254 0.896054 0.241338	29	38

$$\psi = 0.0, 0.08\ (0.01) \qquad S_u(\psi_i) = S_l(\psi_i) = 1 \qquad i = 1, \cdots, 9$$
$$\psi = 0.0801, 0.09\ (0.00045) \qquad S_u(\psi_i) = S_l(\psi_i) = 1 \qquad i = 10, \cdots, 32$$
$$\psi = 0.11, 0.2\ (0.01) \qquad S_u(\psi_i) = 0 \qquad i = 33, \cdots, 42$$
$$\psi = 0.3, 1.\ (0.1) \qquad S_u(\psi_i) = 0 \qquad i = 43, \cdots, 50.$$

Using a second-order filter section of the cascade form $\phi = [a_1\, b_1\, c_1\, d_1\, A]^T$. A starting point $\phi^0 = [0\ 0\ 0\ -0.25\ 0.1]^T$ was taken. Test quantities for the Fletcher and Fletcher–Powell methods were 10^{-6} and $\xi = 0$.

Optimization using both the Fletcher method and the Fletcher–Powell method in accordance with the aforementioned ideas has given the results shown in Table I. Growing another filter section and restarting the optimization process has given the results shown in Table II. The magnitude response is depicted in Figs. 1 and 2, where the passband response is shown in Fig. 1 and the stopband response is shown in Fig. 2. The pole-zero configuration is shown in Fig. 3.

It is to be noted that 201 equidistant values of ψ were used for response evaluation and plotting in each frequency band. Local optimality checking was provided through perturbation of $\check{\phi}$, namely, replacing the obtained solution $\check{\phi}$ by $1.0001\check{\phi}$ and restarting the optimization process. The results are shown in Table III.

Discussion

The results of the low-pass filter reported by Steiglitz [4] were reproduced using the package CADRDF (taking the transition region specification into account). However, using the set of sample points given by Steiglitz in the optimization process and then evaluating the magnitude response of the obtained solution at 201 equidistant values of ψ in the passband and, similarly, for the stopband, an error peak was detected in the interval [0.08, 0.09] that was not taken into account by the previous scheme of specified sample points. The passband response is depicted in Fig. 4 and the stopband response is depicted in Fig. 5.

From the results of Table I, it can be seen that for this example, the Fletcher method was more efficient than the Fletcher–Powell method in the sense that it required a smaller number of function evaluations to produce the solution. When the number of independent variables was increased to 9 through growing the second-filter section, the Fletcher–Powell method was remarkably slower than the Fletcher method, as it ran into time limit without yielding even the least squares solution, while for the same specified time, the Fletcher method gave the least 10 000th solution.

As the magnitude response was specified on a closed continuous interval, it was plotted along the specified closed continuous interval in Figs. 1 and 2.

TABLE II
Results for Example 1 Using Two Sections

$q = p$	Maximum Error	Objective Function	Solution $\check{\phi}$	Number of Function Evaluations Using Fletcher Method
2	0.094612	0.183426	-1.858902 0.999999 -1.864459 0.945603 -1.304232 0.999999 -1.739599 0.775075 0.028496	130
10	0.046434	0.055707	-1.870517 0.999999 -1.872787 0.952596 -1.480946 0.999999 -1.750330 0.785957 0.040564	80
100	0.043999	0.044486	-1.870794 1. -1.873968 0.953359 -1.517114 1. -1.752070 0.787578 0.043228	58
1000	0.043639	0.043688	-1.870747 1. -1.874085 0.953433 -1.520148 0.999999 -1.752504 0.787952 0.043373	49
10000	0.043610	0.043614	-1.870741 0.999999 -1.874095 0.953439 -1.520276 1. -1.752557 0.787996 0.043369	52

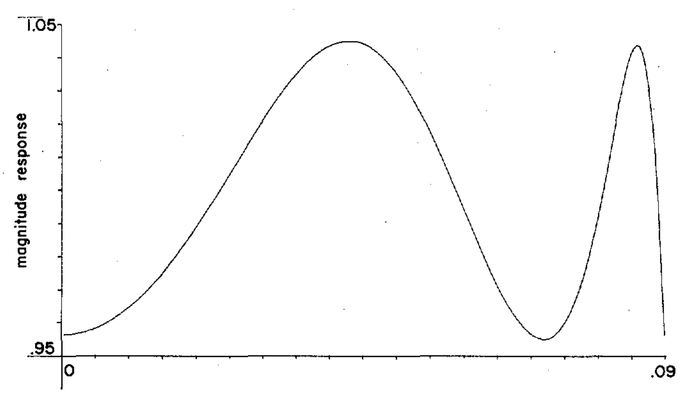

Fig. 1. Magnitude characteristic of the passband of the lowpass filter of Example 1.

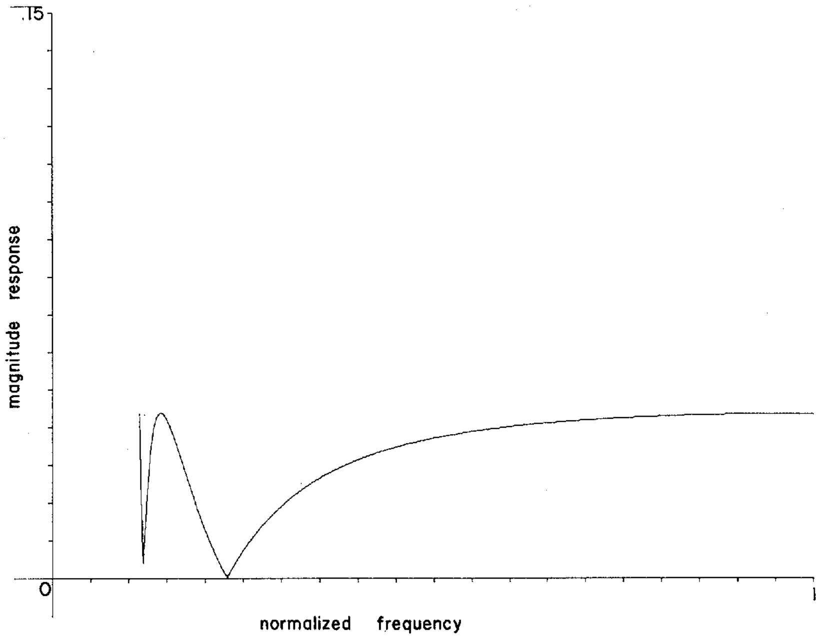

Fig. 2. Magnitude characteristic of the stopband of the low-pass filter of Example 1.

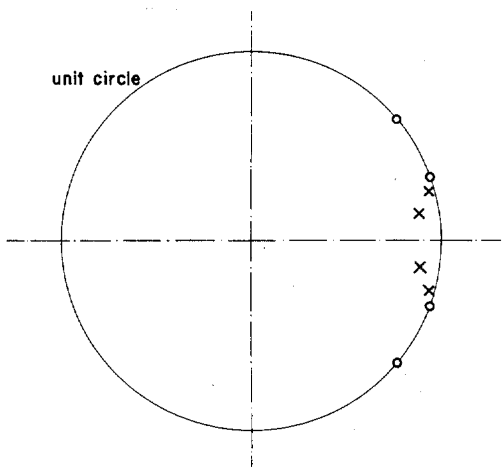

Fig. 3. Pole-zero configuration for the low-pass filter of Example 1.

TABLE III
Local Optimality Table for Example 1

	Solution $\check{\phi}$ Before Perturbation	Starting Point ϕ^0 After Perturbation	Solution ϕ^* After Perturbation	Absolute Differences in Components of $\check{\phi}$ and ϕ^*
	-1.870741	-1.870928	-1.870741	0.0
	0.999999	1.000100	1.	10^{-6}
	-1.874095	-1.874283	-1.874095	0.0
	0.953439	0.953534	0.953439	0.0
	-1.520276	-1.520428	-1.520280	4×10^{-6}
	1.	1.000100	1.000004	4×10^{-6}
	-1.752557	-1.752732	-1.752557	0.0
	0.787996	0.788075	0.787996	0.0
	0.043369	0.043373	0.043368	10^{-6}
Maximum Error	0.043610	0.046658	0.043610	0.0

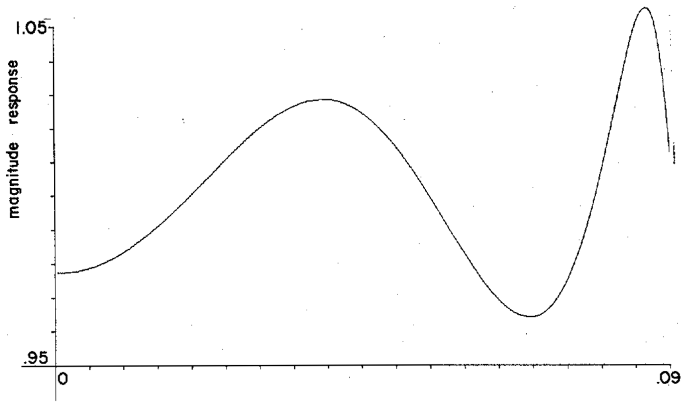

Fig. 4. Magnitude characteristic of the more densely sampled passband of the low-pass filter given by Steiglitz [4].

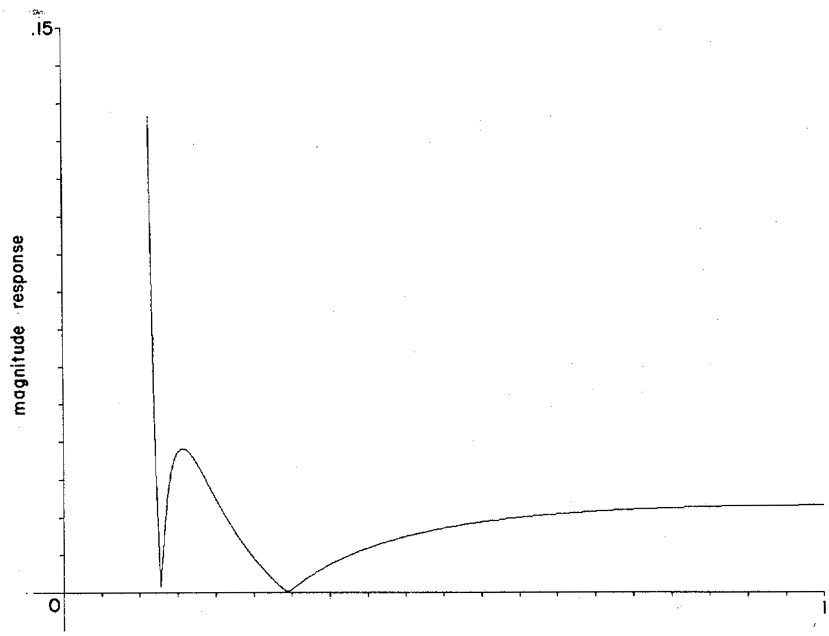

Fig. 5. Magnitude characteristic of the more densely sampled stopband of the low-pass filter given by Steiglitz [4].

Example 2

Consider the design of a low-pass recursive digital filter of the cascade form, for a 10-kHz sampling rate, whose upper and lower magnitude response bounds are specified by

$$f = 0, 900\ (100) \qquad S_u(f) = 1.1 \quad S_l(f) = 0.9.$$

$$f = 1200 \qquad S_u(f) = 0.1$$

$$f = 1500, 5000\ (500) \qquad S_u(f) = 0.1$$

The specifications can be prescribed in terms of ψ as follows ($\psi = 2f/f_s$):

$$\psi = 0.0, 0.18\ (0.02) \quad S_u(\psi_i) = 1.1 \quad S_l(\psi_i) = 0.9 \quad i = 1, \cdots, 10$$

$$\psi = 0.24 \quad S_u(\psi_i) = 0.1 \quad i = 11$$

$$\psi = 0.3, 1\ \ (0.1) \quad S_u(\psi_i) = 0.1 \quad i = 12. \cdots, 19.$$

Using a second-order recursive digital filter section, the same starting point as that used by Suk and Mitra [12], namely, $\phi^0 = [0\ 1\ -1\ 0.5\ 0.1]^T$ was taken. Weighting factors, test quantity for the Fletcher method and ξ were as in Example 1.

Optimization using the Fletcher method gave the results shown in Table IV. Growing another filter section from the one section locally optimum solution and restarting the optimization process has given the results shown in Table V.

The magnitude response is depicted in Figs. 6 and 7, where the passband response is shown in Fig. 6,

TABLE IV
Results for Example 2 Using One Section

$q = p$	Maximum Error	Objective Function	Solution $\check{\phi}$	Number of Function Evaluations Using Fletcher Method
2	0.192456	0.244387	−0.858152 1. −1.500265 0.709050 0.157547	44
10	0.109918	0.121980	−1.147287 0.999999 −1.540582 0.759644 0.206694	60
100	0.101718	0.102881	−1.165964 1. −1.544701 0.764404 0.210503	27
1000	0.101302	0.101417	−1.166416 0.999999 −1.545169 0.764763 0.210425	43
10000	0.101271	0.101282	−1.166418 1.000001 −1.545223 0.764801 0.210399	24

and the stopband response is shown in Fig. 7. The pole-zero configuration is shown in Fig. 8.

Discussion

The sign of q was positive so long as the magnitude response did not lie within the specified bounds, but a change of sign of q to negative, occurred, when the specifications were met, but that did not stop the optimization process as it went on to produce a locally optimum solution for the case when the specifications were met. As the magnitude response was initially specified at a discrete set of frequency points, the magnitude response of the final solution was considered only at that discrete set of frequency points as shown in Figs. 6 and 7. Increasing the value of ξ did not affect the least 10 000th solution.

The results shown in Tables I–V are computed on the basis that the least squares solution is used as a starting point to obtain the least 10th solution, then the least 10th solution is used as a starting point to obtain the least 100th solution, and so on.

VII. Conclusions

The application of the Bandler–Charalambous method using extremely large values of p, typically 10 000, to recursive digital-filter design problems yields reasonably well-conditioned objective functions. Effectively negative values of p can be used to obtain the coefficients of a recursive digital filter that meets or exceeds the prescribed specifications. The use of the Fletcher method in conjunction with least pth optimization seems to be more efficient than that of the Fletcher–Powell method.

Local optimality of the least pth solution may be checked by perturbing the obtained solution or increasing ξ and then restarting the optimization process. Stability requirements of the filter transfer function can be met by using a pole inversion technique. The order complexity of the filter can be increased by growing of filter sections.

The results of Examples 1 and 2 indicate that the zeros of the transfer function tend to lie on the unit circle in the z domain. Thus, it seems that the efficiency of the computation may be improved by using a starting or fixed value of 1 for each of the filter coefficients b_k, where $k = 1, \cdots, K$.

Acknowledgment

The authors wish to thank Mrs. J. R. Popović, whose computer program on least pth approximation FMLPO [13] and [14] was modified and incorporated into CADRDF. Dr. C. Charalambous is also thanked for helpful suggestions. Dr. A. G. Deczky is thanked for pointing out an error in the Fletcher–Powell routine previously used by the authors.

References

[1] J. F. Kaiser, "Digital Filters," in *Systems Analysis by Digital Computer*, F. 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] S. S. Haykin, "A unified treatment of recursive digital filtering," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 113–116, Feb. 1972.
[4] K. Steiglitz, "Computer-aided design of recursive digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-18, pp. 123–129, June 1970.
[5] H. D. Helms, "Digital filters with equiripple or minimax responses," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, pp. 87–93, Mar. 1971.

TABLE V
Results for Example 2 Using Two Sections

$q = -p$	Maximum Error	Objective Function	Solution $\check{\phi}$	Number of Function Evaluations Using Fletcher Method
-2	-0.062826	-0.016925	-1.412375 1. -1.575303 0.882915 0.618033 0.999999 -1.430463 0.562704 0.026034	182
-10	-0.068650	-0.059397	-1.407856 1. -1.578990 0.885035 0.082835 0.999999 -1.438495 0.570015 0.032027	85
-100	-0.073797	-0.072980	-1.404770 0.999999 -1.584958 0.888081 -0.207829 0.999999 -1.453960 0.582477 0.035615	94
-1000	-0.074321	-0.074240	-1.404186 1. -1.585726 0.888448 -0.226099 0.999999 -1.456176 0.584152 0.035719	58
-10000	-0.074369	-0.074361	-1.404118 0.999999 -1.585808 0.888486 -0.227133 0.999999 -1.456438 0.584344 0.035707	71

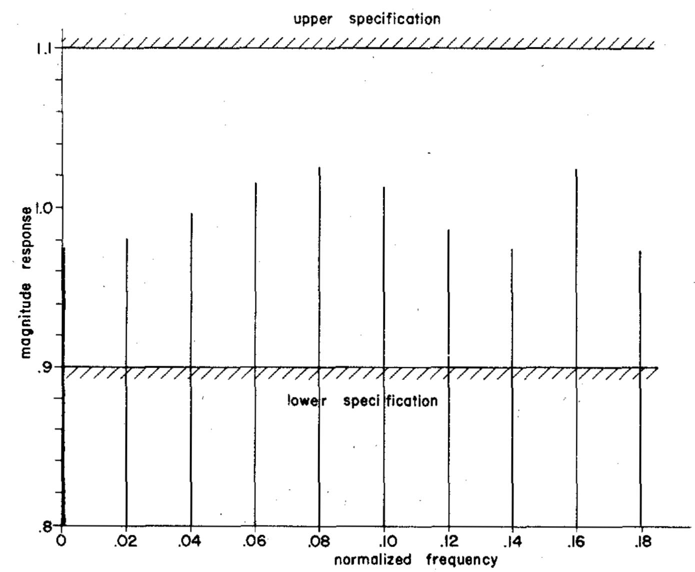

Fig. 6. Magnitude characteristic of the passband of the low-pass filter of Example 2.

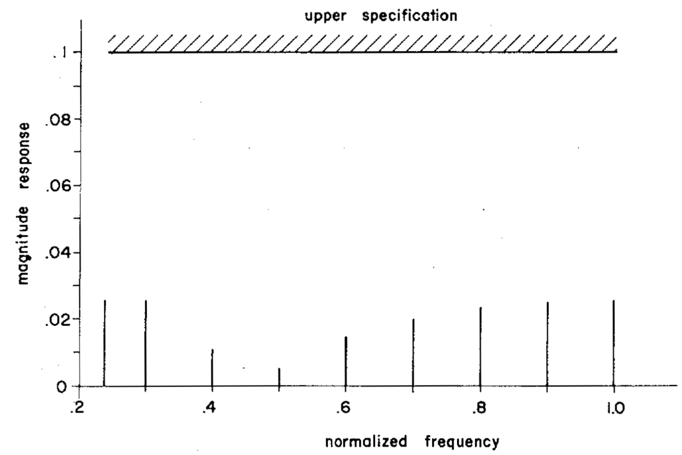

Fig. 7. Magnitude characteristic of the stopband of the low-pass filter of Example 2.

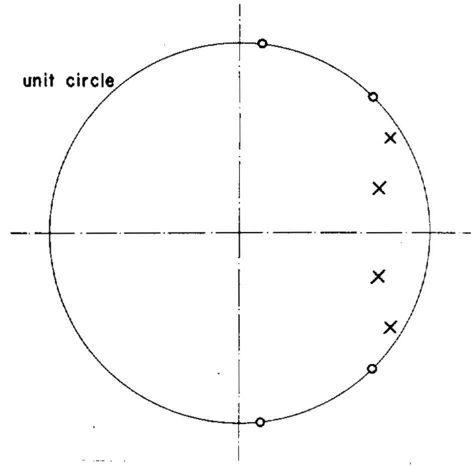

Fig. 8. Pole-zero configuration for the low-pass filter of Example 2.

[6] A. G. Deczky, "Synthesis of recursive digital filters using the minimum p-error criterion," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 257-263, Oct. 1972.
[7] M. Sablatash, "The state of the art in approximation techniques for digital filter design," presented at the 1971 Symp. Digital Filtering, Imperial College of Science and Technology, London, Aug. 31-Sept. 2, 1971.
[8] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, vol. 6, pp. 163-168, June 1963.
[9] J. W. Bandler and C. Charalambous, "Practical least pth optimization of networks," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, pp. 834-840, Dec. 1972.
[10] R. Fletcher, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, pp. 317-322, Aug. 1970.
[11] J. W. Bandler and C. Charalambous, "On conditions for optimality in least pth approximation with $p \to \infty$," *J. Optimiz. Theory and Appl.*, to be published.
[12] M. Suk and S. K. Mitra, "Computer-aided design of digital filters with finite word length," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 356-363, Dec. 1972.
[13] J. R. Popović and J. W. Bandler, "A general program for discrete least pth approximation," *IEEE Trans. Microwave Theory Tech.*, to be published.
[14] J. R. Popović, J. W. Bandler, and C. Charalambous, "General programs for least pth and near minimax approximation," *Int. J. Syst. Sci.*, to be published.

Correspondence

On the Stability of Two-Dimensional Digital Filters

G. A. MARIA and M. M. FAHMY

Abstract—This correspondence proposes a method to check the stability of two-dimensional recursive filters. In this method the Jury table is modified and used to check the first condition of Huang's theorem. Some examples are solved to illustrate the method.

I. Introduction

The stability of two-dimensional recursive filters was first studied by Shanks [1]. He introduced a theorem which calls for mapping all the points $|Z_1| \leqslant 1$ into the Z_2-plane under the function $B(Z_1, Z_2) = 0$, where $B(Z_1, Z_2)$ is the denominator of the transfer function of the recursive filter. Huang [2] simplified the above theorem by showing that it is sufficient to examine the mapping on the boundary of the unit circle $|Z_1| = 1$. Using a theorem due to Ansell [3], Huang suggested a procedure that reduces the amount of computation needed to check the stability of two-dimensional filters.

In this correspondence a new method is proposed to check the stability of two-dimensional filters. In this method the Jury table is modified to examine the roots of polynomials with complex coefficients, and then is used to check the first condition of Huang's theorem. The amount of computation needed for this method is comparable to that needed when using the procedure suggested by Huang. However, the proposed method has the advantage that all determinants used in computation are of the dimension two, while Huang's method uses determinants of an order up to the order of the filter. For the sake of comparison, the same examples solved by Huang are solved here by the proposed method.

Manuscript received December 10, 1972; revised April 16, 1973.
The authors are with the Department of Electrical Engineering, Queen's University, Kingston, Ont., Canada.

II. The Main Result

Huang's criterion for studying the stability of two-dimensional digital filters can be stated as follows.

Huang's Theorem: A causal filter with a Z-transform $H(Z_1, Z_2) = A(Z_1, Z_2)/B(Z_1, Z_2)$ where A, B are polynomials, is stable if and only if: 1) the map of $\partial d_1 \triangleq (Z_1; |Z_1| = 1)$ in the Z_2 plane, according to $B(Z_1, Z_2)$, lies inside $d_2 \triangleq (Z_2;$