Chebyshev Approximation for Two-Dimensional Nonrecursive Digital Filters

Y. KAMP AND J. P. THIRAN, MEMBER, IEEE

Abstract—**The Remez exchange algorithm is extended for the design of two-dimensional nonrecursive digital filters approximating circularly symmetrical low-pass specifications according to a weighted Chebyshev error norm. Since the approximating function does not satisfy the Haar condition, the optimal solution is not necessarily unique and a straightforward extension of the one-dimensional exchange method may fail to converge. It is shown how the algorithm has to be complemented with a perturbation technique in order to force convergence under all circumstances. In the case of nonuniqueness the solution provided by the algorithm is a vertex of the polyhedron containing all optimal solutions, and a method is given which allows one to compute an adjacent vertex located on the same edge, thereby allowing the successive determination of all the vertices defining the polyhedron. For this case also, a procedure is described which selects, among all optimal solutions, the best one according to some additional criterion. Finally, the efficiency and accuracy aspects of the algorithm are considered and practical conclusions are drawn as an aid for the designer.**

I. INTRODUCTION

WORK IN THE DOMAIN of two-dimensional digital filters has been up to now mainly concerned with nonrecursive designs because they escape the stability problem which is inherent in recursive filters [1] and because they can achieve a strictly linear phase which is a desirable feature for the distortionless processing of images.

Two-dimensional nonrecursive digital filters can be obtained by extending some of the methods already available for the design of one-dimensional filters. This was done for the windowing technique [2], linear programming [3], and for designs according to a maximally flat criterion [4]. Recently [5], an ingenious solution was proposed to transform a one-dimensional Chebyshev design into a two-dimensional design through a change of variables. The method is limited by the fact that the optimality of the result is tied to special classes of cutoff boundaries and, in particular, optimal approximations of specifications with exact circular symmetry cannot be achieved. This demonstrates the need for a method in which the two-dimensional nature of the problem is effectively taken into account. In this line the present paper investigates the problems connected with the Chebyshev approximation by functions of two variables. As a result, an extension of the Remez exchange algorithm [6] is presented for the design of two-dimensional nonrecursive filters which approximate low-pass specifications having exactly circular pass- and stop-band limits. Moreover, most of the results obtained here also apply to the design of a much wider class of filters.

Manuscript received March 11, 1974; revised July 5, 1974.
The authors are with the MBLE Research Laboratory, Brussells, Belgium.

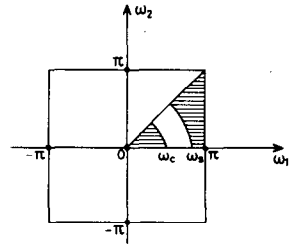

Fig. 1. Approximation domain.

II. THE APPROXIMATION PROBLEM AND ITS SOLUTION

A. Statement of the Problem

It has been shown in [4] that the requirement of circular symmetry imposes the following form for the frequency response of a nonrecursive two-dimensional digital filter

$$G(e^{j\omega_1},e^{j\omega_2}) = \sum_{I=0}^{N}\sum_{J=0}^{I} B_{IJ}(\cos I\omega_1 \cos J\omega_2 + \cos J\omega_1 \cos I\omega_2). \quad (1)$$

For the discussion of the approximation problem, it will prove more convenient to rewrite it as

$$f(x) = \sum_{i=1}^{m} a_i f_i(x) \quad (2)$$

where

$$\begin{aligned} f_i(x) &= \cos I\omega_1 \cos J\omega_2 + \cos J\omega_1 \cos I\omega_2, \\ x &= (\omega_1,\omega_2), \\ a_i &= B_{IJ}, \\ i &= I(I+1)/2 + J + 1, \\ m &= (N+1)(N+2)/2. \end{aligned} \quad (3)$$

The purpose is to design a circularly symmetric low-pass filter and the function to be approximated is thus

$$F(x) = \begin{cases} 1 & \text{if } \omega_1^2 + \omega_2^2 \le \omega_c^2 \\ 0 & \text{if } \omega_1^2 + \omega_2^2 \ge \omega_s^2 \end{cases}$$

where ω_c and ω_s are, respectively, the pass- and stopband limits. Since the frequency response (1) is already symmetrical with respect to both axes and both diagonals of the frequency plane, it will be sufficient to control its behavior in the shaded region of Fig. 1. In fact, this continuous domain will be replaced by a discrete finite point set $X = \{x_1,x_2,\cdots,x_n\}$ with $n > m$.

If V denotes the linear space of approximating functions (2) when $\{a_i\} \in R^m$, the following linear Chebyshev approx-

imation problem has to be solved: find $f^* \in V$ which minimizes the norm

$$\|f - F\| = \max_{x \in X} w(x)|f(x) - F(x)| \tag{4}$$

where the weight function $w(x)$ is strictly positive on X. The minimum of this norm is the Chebyshev deviation

$$H = \min_{f \in V} \|f - F\|. \tag{5}$$

The solution lies in a generalization of the Remez algorithm which will be summarized in the following two sections by using the approach of Stiefel [6].

B. Reference and Reference Function

The theory of the exchange algorithm relies on the following definitions. For the approximating function (2) the *characteristic vector* ϕ_j relative to a point x_j is defined by $\phi_j = [f_1(x_j), f_2(x_j), \cdots, f_m(x_j)]^T$.

A *reference* $R \subset X$ is a set of $(m + 1)$ distinct points $R = \{x_1, x_2, \cdots, x_{m+1}\}$ such that the corresponding matrix $[\phi_1, \phi_2, \cdots, \phi_{m+1}]$ has rank m. The existence of at least one reference in X is equivalent to the claim that the base functions $f_i(x)$ are linearly independent on X.

A *basis* is a set of m linearly independent characteristic vectors. By definition, a reference contains at least one basis. Let then $B = [\phi_1, \phi_2, \cdots, \phi_m]$ be a basis for R. The *characteristic relation of a reference* expresses the characteristic vector ϕ_{m+1} of the reference as a linear combination of the other vectors

$$\phi_{m+1} = \sum_{j=1}^{m} \lambda_j \phi_j \tag{6}$$

where

$$\Lambda = [\lambda_1, \lambda_2, \cdots, \lambda_m]^T = B^{-1}\phi_{m+1}. \tag{7}$$

By putting

$$\lambda_{m+1} = -1 \tag{8}$$

one derives from (6)

$$\sum_{j=1}^{m+1} \lambda_j f(x_j) = 0, \qquad \forall f \in V \tag{9}$$

and $[\hat{\lambda}_1]$ is then called the *vector of characteristic coefficients.*

The *error* $h(x)$ associated with an approximating function is defined by

$$h(x) = w(x)[f(x) - F(x)]. \tag{10}$$

A function $f(x)$ is a *reference function for* R if

$$\operatorname{sgn} h(x_j) = \eta \operatorname{sgn} \lambda_j, \qquad 1 \le j \le m + 1 \tag{11}$$

where $\eta = \pm 1$ and where sgn x is arbitrary for $x = 0$. By (8) the latter definition is equivalent to

$$h(x_j)h(x_{m+1})\lambda_j \le 0, \qquad 1 \le j \le m. \tag{12}$$

In the one-dimensional problem the characteristic coefficients have alternating signs and (11) becomes the well-known alternation property of the reference function. A *leveled reference function for* R satisfies (11) and in addition $|h(x_j)| = |h|$ for $1 \le j \le m + 1$. Owing to (9), (10), and (11), the *reference deviation* $|h|$ of a reference can be computed as

$$|h| = \left|\sum_{j=1}^{m+1} \lambda_j F(x_j)\right| \Big/ \left|\sum_{j=1}^{m+1} |\lambda_j| w^{-1}(x_j)\right| \tag{13}$$

and the *coefficient vector* $A = [a_1, a_2, \cdots, a_m]^T$ of the leveled reference function by

$$A^T = \left[F(x_1) + \frac{h(x_1)}{w(x_1)}, \cdots, F(x_m) + \frac{h(x_m)}{w(x_m)}\right] B^{-1}. \tag{14}$$

From (9), (10), (11), and (13), Theorem 1 follows.

Theorem 1 [6]: If $f(x)$ is a reference function for R with associated error $h(x)$, then the reference deviation of R is given by

$$|h| = \sum_{j=1}^{m+1} |\lambda_j| |h(x_j)| w^{-1}(x_j) \Big/ \sum_{j=1}^{m+1} |\lambda_j| w^{-1}(x_j).$$

C. Exchange Algorithm

Theorem 2 [6]: Let $f(x)$ be a reference function for R, $h(x)$ the corresponding error and $x_k \notin R$ an arbitrary point in X. There exists a point $x_l \in R$ such that by exchanging x_l with x_k one obtains a new reference R' for which $f(x)$ is still a reference function. The point to be exchanged is found according to the following rule. For the reference $\{x_1, x_2, \cdots, x_m, x_k\}$, let

$$\phi_k = \sum_{j=1}^{m} \mu_j \phi_j \tag{15}$$

be the characteristic relation with

$$M = [\mu_1, \mu_2, \cdots, \mu_m]^T = B^{-1}\phi_k. \tag{16}$$

Case 1: If

$$h(x_j)h(x_k)\mu_j \le 0, \qquad 1 \le j \le m \tag{17}$$

then $x_l = x_{m+1}$.

Case 2: If

$$E = \{x_j,\ 1 \le j \le m \mid h(x_j)h(x_k)\mu_j > 0\} \neq \text{—} \tag{18}$$

the index l of the point to be exchanged is such that

$$|\lambda_l/\mu_l| = \min_{j \in E} |\lambda_j/\mu_j|. \tag{19}$$

It is indeed easily shown that for the new reference R' a condition of the type (12) is satisfied. The new characteristic relation is (15) in Case 1 and results from the elimination of ϕ_l between (6) and (15) in Case 2, giving

$$\phi_{m+1} = \frac{\lambda_l}{\mu_l}\phi_k + \sum_{\substack{j=1 \\ \neq l}}^{m} \lambda_j' \phi_j \tag{20}$$

with

$$\lambda_j' = \lambda_j - \lambda_l \mu_j/\mu_l, \qquad 1 \le j \neq l \le m. \tag{21}$$

Let it be noted also that for Case 2 it is not necessary to compute the inverse of the new basis B' by matrix inversion.

Since B and B' only differ by one characteristic vector, $(B')^{-1}$ can directly be derived from B^{-1} [7]. This may, however, lead to important error accumulation, and it is therefore suggested to periodically restart the algorithm with a fresh inverse generated by matrix inversion in order to keep the reference well behaved.

Method of Ascent: Using the preceding theorems a monotonically increasing sequence of reference deviations is constructed as follows.

Let $f(x)$ be the leveled reference function for R and $h(x)$ the associated error. If

$$\max_X |h(x)| = |h| \tag{22}$$

$f(x)$ is an *optimal solution* for the Chebyshev approximation $F(x)$ on X [6]. Else let $x_k \notin R$ be a point of X such that

$$|h(x_k)| > |h|. \tag{23}$$

By the above exchange Theorem 2, a new reference R' is constructed whose characteristic relation is either (15) or (20). The new reference deviation $|h'|$ results then from Theorem 2.

Case 1:

$$|h'| = \frac{|h(x_k)|w^{-1}(x_k) + |h| \sum_{j=1}^{m} |\mu_j| w^{-1}(x_j)}{w^{-1}(x_k) + \sum_{j=1}^{m} |\mu_j| w^{-1}(x_j)} \tag{24}$$

and by (23), $|h'| > |h|$.

Case 2:

$$|h'| = \frac{|h(x_k)\lambda_l \mu_l^{-1}|w^{-1}(x_k) + |h| \left[w^{-1}(x_{m+1}) + \sum_{\substack{j=1 \\ \neq l}}^{m} |\lambda_j'| w^{-1}(x_j) \right]}{|\lambda_l \mu_l^{-1}|w^{-1}(x_k) + w^{-1}(x_{m+1}) + \sum_{\substack{j=1 \\ \neq l}}^{m} |\lambda_j'| w^{-1}(x_j)} \tag{25}$$

and (23) again implies $|h'| > |h|$ except in the case $\lambda_l = 0$ which will be considered in the next section. The fact that the reference deviation is strictly increasing is of prime importance for the convergence proof of the algorithm because in a finite number of iterations a final reference is obtained for which (22) holds.

Multiple Exchange: In order to reduce the number of iterations, it is advantageous to increase as much as possible the new reference deviation $|h'|$. To this end, it would be preferable to introduce in the new reference not only a single point but several points where the absolute value of the error exceeds the reference deviation $|h|$. For the one-dimensional case [8], this is easy to achieve because the sign alternation of the λ_j allows the simultaneous exchange of all the reference points. For two-variable approximations there is no such simple rule for the sign of the characteristic coefficients but one can try to perform a multiple exchange in the following way.

In the neighborhood of each reference point, the local maximum of the absolute value of the error will be retained if its value is larger than the reference deviation. Let then $(x_k^{(1)}, x_k^{(2)}, \cdots, x_k^{(r)})(1 \leq r \leq m+1)$ be the set of the distinct local maxima. One calculates the vector $M_1 = B^{-1}\phi_k^{(1)}$ where $\phi_k^{(1)}$ denotes the characteristic vector of $x_k^{(1)}$ and the exchange Theorem 2 gives then the point x_l of the reference R which has to be exchanged. Let R_1 be the new reference, B_1 the corresponding basis. The same sequence of operations is repeated for the second maximum $x_k^{(2)}$ by computing $M_2 = B_1^{-1}\phi_k^{(2)}$ but this second exchange is effectively performed only if it does not eliminate $x_k^{(1)}$. One proceeds in this way for all local maxima with the rule that an exchange will not be carried out if it would eliminate one of the previously introduced maxima.

D. Perturbation Technique for Degenerate Problems

An approximation problem becomes *degenerate for a reference* R if the corresponding characteristic relation contains at least one vanishing λ_j. The occurrence of degeneracies is a direct consequence of the fact that practically no two-variable functions satisfy the Haar condition [9]. According to the methods of ascent described in the preceding section, one has Theorem 3.

Theorem 3: A *static exchange* $|h'| = |h|$ occurs if and only if the problem is degenerate with respect to R and if among the points x_j of R for which $\lambda_j = 0$, one at least satisfies the inequality $h(x_j)h(x_k)\,\mu_j > 0$ where μ_j is defined by (16).

The possibility of static exchanges for degenerate problems entails that the same reference can be repeated periodically. In these cases of cycling, the above exchange algorithm does not converge and one has to resort to a special perturbation technique. This technique was originally developed in linear programming [7] to cope with similar convergence problems and was later on adapted to suit the exchange algorithm. The adaptation presented here is similar to but more efficient than that of Bittner [10] and Descloux [11]. Like in linear programming, the perturbation technique is only of theoretical importance because it ensures convergence without exception, but in practice no degenerate cases have been known to cycle for this problem.

The perturbation technique consists in replacing the degenerate problem P by a perturbed problem $P(\varepsilon)$ satisfying the twofold condition that it never becomes degenerate and that it coincides with the original problem when ε tends to zero. In this way, the convergence of the perturbed problem also insures that of the degenerate problem. For convenience it will be assumed that the points of R have been ordered such that $\lambda_j = 0\ (1 \leq j \leq p)$ and $\lambda_j \neq 0\ (p+1 \leq j \leq m)$. According to Theorem 3, it is also assumed that at least one of the points $x_j\ (1 \leq j \leq p)$ is an element of the set E defined in (18). To construct the perturbed problem, the vector ϕ_{m+1} of the degenerate problem is replaced by $\phi_{m+1}(\varepsilon) = \phi_{m+1} + \sum_{j=1}^{p} \varepsilon^j \phi_j$

where ε is a small positive quantity. The characteristic relation of the perturbed problem is then

$$\phi_{m+1}(\varepsilon) = \sum_{j=1}^{m} \lambda_j(\varepsilon)\phi_j \tag{26}$$

where, by (7),

$$\begin{cases} \lambda_j(\varepsilon) = \varepsilon^j, & 1 \le j \le p \\ \lambda_j(\varepsilon) = \lambda_j \neq 0, & p+1 \le j \le m \end{cases} \tag{27}$$

thus showing that the perturbed problem is nondegenerate. Since by hypothesis some indices j $(1 \le j \le p)$ belong to E, the exchange Theorem 2 will give a point x_l with $1 \le l \le p$, whence

$$\lambda_l(\varepsilon) = \varepsilon^l. \tag{28}$$

The characteristic relation (20) of the new reference R' becomes here

$$\phi_{m+1} = \lambda_k'(\varepsilon)\phi_k + \sum_{\substack{j=1 \\ \neq l}}^{m} \lambda_j'(\varepsilon)\phi_j$$

with

$$\begin{cases} \lambda_j'(\varepsilon) = \lambda_j(\varepsilon) - \lambda_l(\varepsilon)\mu_j/\mu_l, & (1 \le j \neq l \le m) \\ \lambda_k'(\varepsilon) = \lambda_l(\varepsilon)/\mu_l. \end{cases} \tag{29}$$

The new reference deviation $|h'(\varepsilon)|$ is still given by (25) provided that the characteristic coefficients are replaced by (29). Owing to (28),

$$\lim_{\varepsilon \to 0} |h'(\varepsilon)| = |h|$$

and the exchange of the perturbed problem corresponds to a static exchange for the original problem. In addition, it is seen that the $\lambda_j'(\varepsilon)$ $(j = k,\ 1 \le j \neq l \le m)$ are polynomials without constant term and which are linearly independent because $\lambda_j'(\varepsilon)$ is the only polynomial having a term in ε^j. In particular, $\lambda_j'(\varepsilon) \not\equiv 0$ and the perturbed problem is still nondegenerate.

In order to compute the leveled reference function for R' at the next iteration, one should take

$$\operatorname{sgn} h'(x_j) = \eta \lim_{\varepsilon \to 0} [\operatorname{sgn} \lambda_j'(\varepsilon)], \qquad j = k,\ 1 \le j \neq l \le p$$

which means that the sign is determined by the coefficient of the lowest power of ε in $\lambda_j'(\varepsilon)$. For this iteration, a new set E' has to be defined analogously to (18) and the perturbed problem continues if among the $\lambda_j'(\varepsilon)$ $(j = k, 1 \le j \neq l \le p)$ some have still indices belonging to E'. During the subsequent iterations, the perturbed problem remains nondegenerate because linear combinations of the linearly independent polynomials $\lambda_j'(\varepsilon)$ $(j = k, 1 \le j \neq l \le p)$ are again linearly independent. All the exchanges of the perturbed problem correspond to static exchanges for the original one.

If among the $\lambda_j'(\varepsilon)$ $(j = k, 1 \le j \neq l \le p)$ no index belongs to E', then the perturbed problem stops. Indeed, a point x_q $(k \neq q \ge p+1)$ is then exchanged such that by (29) and (27) $\lambda_q'(0) = \lambda_q \neq 0$. A relation similar to (25) for the new reference R'' shows then that

$$\lim_{\varepsilon \to 0} |h''(\varepsilon)| > |h|$$

which corresponds to a nonstatic exchange for the original problem.

Bittner [10] makes the perturbed problem unnecessarily complicated by defining $\phi_{m+1}(\varepsilon) = \phi_{m+1} + \sum_{j=1}^{n} \varepsilon^j \phi_j$ instead of (26). The resulting computations are much more time-consuming since normally $p < m \ll n$. Descloux [11], on the other hand, oversimplifies the perturbed problem by taking $\phi_{m+1}(\varepsilon) = \phi_{m+1} + \varepsilon \sum_{j=1}^{p} \alpha_j \phi_j$ where the α_j are arbitrary constants. Indeed, (27) becomes then $\lambda_j(\varepsilon) = \varepsilon\alpha_j$ $(1 \le j \le p)$ and it is no longer possible to guarantee that the coefficients $\lambda_j'(\varepsilon)$ $(1 \le j \neq l \le p)$ in (29) are nonzero. This means that the perturbed problem can be degenerate and its convergence is thus no longer insured.

For the problem defined in Section II-A several cases of degeneracy can theoretically exist. However, experience has shown that in practice only one type of degeneracy actually occurs which is characterized by the fact that the reference contains exactly $(N+2)$ points on the frequency axis ω_1, in short, degeneracy with respect to the ω_1-axis. At these points the base functions $f_i(x)$ in (3) reduce to linear combinations of $\cos I\omega_1$ $(0 \le I \le N)$ and therefore, like in the one-variable case, the characteristic vector of an arbitrary point on ω_1 is a linear combination of the characteristic vectors corresponding to the remaining $(N+1)$ points on ω_1. By the definitions of Section II-B, a basis for such a reference contains exactly $(N+1)$ points on ω_1 while the $(N+2)$th point plays the role of x_{m+1} in R. Consequently, in the vector Λ only the characteristic coefficients corresponding to the points on ω_1 are nonzero.

III. Nonuniqueness

Since the approximating function (2) does not satisfy the Haar condition on X, the Chebyshev approximation problem stated in Section II-A has not necessarily a unique solution.

A. The Set of Best Chebyshev Approximations

In order to compute all best approximations, the approach used in [10] will now be extended. It is based on the concept of *extreme optimal solution* which is an optimal solution such that 1) the error has at least $m+1$ *extremal points*, i.e., points where it attains the Chebyshev deviation, and 2) the approximating function is a reference function for $m+1$ extremal points.

By the assumption of independent base functions on X, the set of best Chebyshev approximations are characterized by the following theorem [11].

Theorem 4: The set C of best Chebyshev approximations is bounded. In the m-dimensional space of the coefficient vector A, it is a convex polyhedron whose vertices are the extreme optimal solutions.

Since a convex polyhedron is the convex hull of its vertices, the determination of C amounts to computing all extreme optimal solutions. Starting from the extreme optimal solution obtained by the exchange algorithm, we will first determine an *adjacent extreme optimal solution* which belongs to the same *edge* of C, i.e., the set of optimal solutions corresponding to the same $m-1$ extremal points with independent characteristic vectors.

In order to express the dependence of $h(x)$ on the coefficient vector A, one rewrites the error function (10) at the point x_i as

$$h_i(A) = w_i(\phi_i{}^T A - F_i) \tag{30}$$

with $w_i = w(x_i)$ and $F_i = F(x_i)$. One can then define C as the admissible solutions A of the set of inequalities

$$\begin{cases} h_i(A) \le H \\ -h_i(A) \le H, \qquad (1 \le i \le n) \end{cases} \tag{31}$$

where H is the Chebyshev deviation.

Let A be the extreme optimal solution obtained by the exchange algorithm with no more than $m + 1$ extremal points $\{x_j;\ 1 \le j \le m + 1\}$ such that one has (12) with (6), and suppose we want to compute the adjacent extreme optimal solution A^* located on the edge $h_j(A^*) = h_j(A)$ $(1 \le j \le m - 1)$. From (30), this edge is still defined by

$$\phi_j{}^T(A^* - A) = 0, \qquad (1 \le j \le m - 1). \tag{32}$$

The extreme optimal solution A^* will be different from A if $\operatorname{sgn} h_m(A)h_m(A^*) < H$ or, by (30), if

$$h_m(A)\phi_m{}^T(A^* - A) < 0. \tag{33}$$

The infinite number of A^* satisfying (32) together with (33) must now be restricted by the remaining constraints of (31). First, one has $\operatorname{sgn} h_{m+1}(A)h_{m+1}(A^*) \le H$ which, by means of (30), (6), (32), and (33), can be rewritten as $h_{m+1}(A) \cdot h_m(A)\lambda_m \ge 0$. Obviously, this inequality is compatible with (12) only if $\lambda_m = 0$ which yields

$$h_{m+1}(A^*) = h_{m+1}(A). \tag{34}$$

Therefore, the number of extreme optimal solutions which are adjacent to the solution obtained by the exchange method is equal to the number of zero characteristic coefficients corresponding to the final reference of the exchange algorithm. In particular, if every characteristic coefficient is different from zero, the Chebyshev approximation is unique. On the other hand, all best approximations have the same error at the extremal points relative to the nonzero characteristic coefficients: according to the terminology used in [12], they form a *critical point set.*

In addition to (32)–(34), the admissible values of A^* are finally restricted by the set of inequalities

$$\begin{cases} -\operatorname{sgn} h_m(A)h_m(A^*) \le H \\ \pm h_i(A^*) \qquad\qquad \le H, \qquad (m + 1 < i \le n) \end{cases}$$

which, by means of (30), (32), (33), and the relations $\phi_i = \sum_{j=1}^{m} \mu_{ij}\phi_j$ $(m + 1 < i \le n)$, are equivalent to

$$\begin{cases} -\operatorname{sgn} h_m(A)\phi_m{}^T(A^* - A) \le \dfrac{2H}{w_m} \\ -\operatorname{sgn} h_m(A)\phi_m{}^T(A^* - A) \le \dfrac{H \mp h_i(A)}{\mp w_i\mu_{im} \operatorname{sgn} h_m(A)}, \\ \qquad\qquad \text{for } \{m + 1 < i \le n \mid \mp\mu_{im}h_m(A) > 0\} \end{cases}$$

where the three upper and lower signs correspond to each other. Consequently, one has Theorem 5.

Theorem 5: The extreme optimal solution A^* adjacent to A and located on the edge (32) with $\lambda_m = 0$, is characterized by the extremal point set $\{x_1, \cdots, x_{m-1}, x_m{}^*, x_{m+1}\}$ where $x_m{}^*$ corresponds to the minimum of the following expressions

$$\begin{cases} \dfrac{2H}{w_m} \\ \dfrac{H \mp h_i(A)}{w_i|\mu_{im}|}, \qquad \text{for } \{m + 1 < i \le n \mid \mp\mu_{im}h_m(A) > 0\}. \end{cases}$$

The errors at the points $\{x_1, \cdots, x_{m-1}, x_{m+1}\}$ of the old reference are unchanged whereas the error sign at $x_m{}^*$ is given by $\operatorname{sgn}[-h_m(A)]$ if $x_m{}^* = x_m$ and by $\operatorname{sgn}[-\mu_{im}h_m(A)]$ if $x_m{}^* = x_i$ $(m + 1 < i \le n)$.

When there are t zero characteristic coefficients in (6), the above theorem allows one to compute t adjacent extreme optimal solutions. If the t points $x_m{}^*$ together with the error sign are identical, it is readily seen that the number of vertices of C is equal to $t + 1$: the set of best approximations is thus completely defined. Otherwise, the process must be repeated for each new extreme optimal solution and the number of vertices may be very high.

Finally, it should be pointed out that, if the optimal solution obtained by the exchange algorithm has more than $m + 1$ extremal points, the number of adjacent extreme optimal solutions can be smaller than t: in fact, the best approximation may be unique even if some coefficients of the characteristic relation (6) are zero.

B. Strict Approximation

In case of nonuniqueness of the best approximation, one is faced with the problem of choosing one of the infinite number of solutions belonging to C. One way to overcome this difficulty is the strict approximation as proposed by Rice [12]. Suppose that in the characteristic relation corresponding to the extremal points of the optimal solution obtained by the exchange method, one has $\lambda_i \ne 0$ $(1 \le i \le t)$ and $\lambda_i = 0$ $(t < i \le m)$. As mentioned above, all elements of C have the same error at the critical point set $\{x_1, x_2, \cdots, x_t, x_{m+1}\}$ and, consequently, at all points x_k depending on this set, i.e., such that $\phi_k = \sum_{i=1}^{t} \mu_{ki}\phi_i$. If X^* denotes this set of points where all best approximations have identical errors, the strict approximation is the element of C whose error has the minimum deviation in $X - X^*$.

Since the matrix of the characteristic vectors corresponding to the points $x^* \in X^*$ has rank t, each base function $f_i(x^*)$ can be expressed in terms of t independent functions $\{g_1(x^*), \cdots, g_t(x^*)\}$ as

$$f_i(x^*) = \sum_{j=1}^{t} b_{ij}g_j(x^*), \qquad (1 \le i \le m) \tag{35}$$

where the matrix $[b_{ij}]$ has rank t. By suitable ordering of the base functions f_i, it is always possible to make the first t rows of $[b_{ij}]$ linearly independent. Let $\{g_1, \cdots, g_m\}$ be a set of new base functions such that the first t functions are solution of the nonsingular system of equations

$$f_i(x) = \sum_{j=1}^{t} b_{ij}g_j(x), \qquad (1 \le i \le t)$$

whereas the $m - t$ last ones are given by

$$g_k(x) = f_k(x) - \sum_{j=1}^{t} b_{kj} g_j(x), \qquad (t < k \leq m). \quad (36)$$

The approximating function (2) then becomes

$$f(x) = G_1(x) + G_2(x) \quad (37)$$

in which

$$\begin{cases} G_1(x) = \sum_{i=1}^{t} \alpha_i g_i(x) \\ G_2(x) = \sum_{i=t+1}^{m} a_i g_i(x) \end{cases} \quad (38)$$

with $\alpha_i = \sum_{j=1}^{m} a_j b_{ij}$. For $x = x^*$, by (36) and (35), $G_2(x^*) = 0$ and $f(x^*) = G_1(x^*)$. Therefore, since the functions $\{g_1, \cdots, g_t\}$ are independent on X^*, all optimal approximating functions have the same coefficients α_i. The strict approximation will then be obtained by approximating $F(x) - G_1(x)$ by $G_2(x)$ on the set $X - X^*$. It may be possible that this new problem has no unique solution either in which case a further decomposition of the approximating function must be carried out.

As mentioned in Section II-D, exhaustive computations have yielded only degeneracies with respect to the ω_1-axis, in which case $t = N + 1$ and $X^* = \{x^* = (\omega_1, 0)\}$. Applying the above procedure to (1), one gets

$$G_1(x) = \sum_{I=0}^{N} \alpha_I \cos I\omega_1 \cos I\omega_2$$

with

$$\alpha_I = 2B_{II} + \sum_{J=0}^{I-1} B_{IJ} + \sum_{J=I+1}^{N} B_{JI} \quad (39)$$

and

$$G_2(x) = -\sum_{I=1}^{N} \sum_{J=0}^{I-1} B_{IJ}(\cos I\omega_1 - \cos J\omega_1) \cdot (\cos I\omega_2 - \cos J\omega_2). \quad (40)$$

In fact, $G_1(x^*)$ is the best approximation on the ω_1-axis and, as is well known [8], approximation by one-dimensional filters is very fast and accurate. In order to take advantage of this property, one can directly compute the strict approximation by first performing the approximation on the ω_1-axis which yields the optimal solution $\{\alpha_I;\ 0 \leq I \leq N\}$ with Chebyshev deviation H^*. Once the coefficients B_{IJ} $(1 \leq I \leq N;\ 0 \leq J \leq I - 1)$ have been determined by the approximation outside the ω_1-axis, it remains to deduce B_{II} $(0 \leq I \leq N)$ from relation (39). Obviously, one does not know from the outset whether the problem is degenerate with respect to the final reference of the exchange algorithm but, thanks to the following theorem, the method turns out to be advantageous even in case of nondegeneracy.

Theorem 6: Let $R^* = \{x_i^*;\ 1 \leq i \leq t + 1\}$ be the final reference of the exchange algorithm on X^* with Chebyshev deviation H^*. If $R' = \{x_i;\ 1 \leq i \leq m - t + 1\}$ denotes a reference on $X - X^*$ with reference deviation $|h|$ greater than H^*, it is possible to build from R^* and R' a reference R on the whole region X, whose reference deviation is greater than H^*.

The two references R^* and R' correspond to the characteristic relations

$$G_1(x_{t+1}^*) = \sum_{i=1}^{t} \lambda_i^* G_1(x_i^*), \qquad (\lambda_i^* \neq 0)$$

and

$$G_2(x_{m-t+1}) = \sum_{i=1}^{m-t} \mu_i G_2(x_i) \quad (41)$$

with

$$h(x_{m-t+1})h(x_i)\mu_i \leq 0, \qquad (1 \leq i \leq m - t). \quad (42)$$

As $G_1(x^*) = f(x^*)$, $f(x)$ is also a reference function for $\bar{R} = \{x_1^*, \cdots, x_t^*, x_1, \cdots, x_{m-t}, x_{t+1}^*\}$ with reference deviation H^* and characteristic relation

$$f(x_{t+1}^*) = \sum_{i=1}^{t} \lambda_i^* f(x_i^*) + \sum_{i=1}^{m-t} \lambda_i f(x_i) \quad (43)$$

in which all λ_i are zero. By (37), (41) becomes

$$f(x_{m-t+1}) = \sum_{i=1}^{m-t} \mu_i f(x_i) + G_1(x_{m-t+1}) - \sum_{i=1}^{m-t} \mu_i G_1(x_i). \quad (44)$$

Since $G_1(x)$ is given in (38) in terms of t independent functions, any $G_1(x_j)$ for $1 \leq j \leq m - t + 1$ can be expressed as a linear combination of t independent values $G_1(x_i^*) = f(x_i^*)$ $(1 \leq i \leq t)$ and (44) becomes

$$f(x_{m-t+1}) = \sum_{i=1}^{t} \mu_i^* f(x_i^*) + \sum_{i=1}^{m-t} \mu_i f(x_i). \quad (45)$$

By applying the exchange theorem to the characteristic relations (43) and (45), it is now possible to introduce the point x_{m-t+1} with deviation $|h| > H^*$ in $\bar{R}$ in order to form R. This exchange cannot be a static one because condition (42) holds for all zero λ_i in (43) and the reference deviation of R will thus be larger than H^*.

In consequence of this theorem, as soon as the reference deviation $|h|$ of the exchange method for $X - X^*$ becomes greater than H^*, one can restart the exchange algorithm for the whole region X with a reference deviation greater than H^*, thus avoiding any further degeneracy with respect to the ω_1-axis.

IV. Numerical Results

In order to save computation time, it is advantageous to use a Cartesian grid for the discrete point set X which is to replace the continuous approximation domain of Fig. 1. Indeed, evaluating (1) along a vertical or horizontal line reduces to univariable computation. Nevertheless, an accurate representation of the continuous domain requires to include the intersections of the grid lines with the passband and stopband boundary circles. If one characterizes the quality of approximation by the precision factor $(h_{\max} - H)/H$ where $h_{\max}$ is the maximum of the absolute value of the error function on the continuous approximation domain, experience shows that a Cartesian grid of about $25\,N$ by $25\,N$ points is sufficient for getting a precision factor on the order of 10^{-3}, which is quite acceptable for practical purposes.

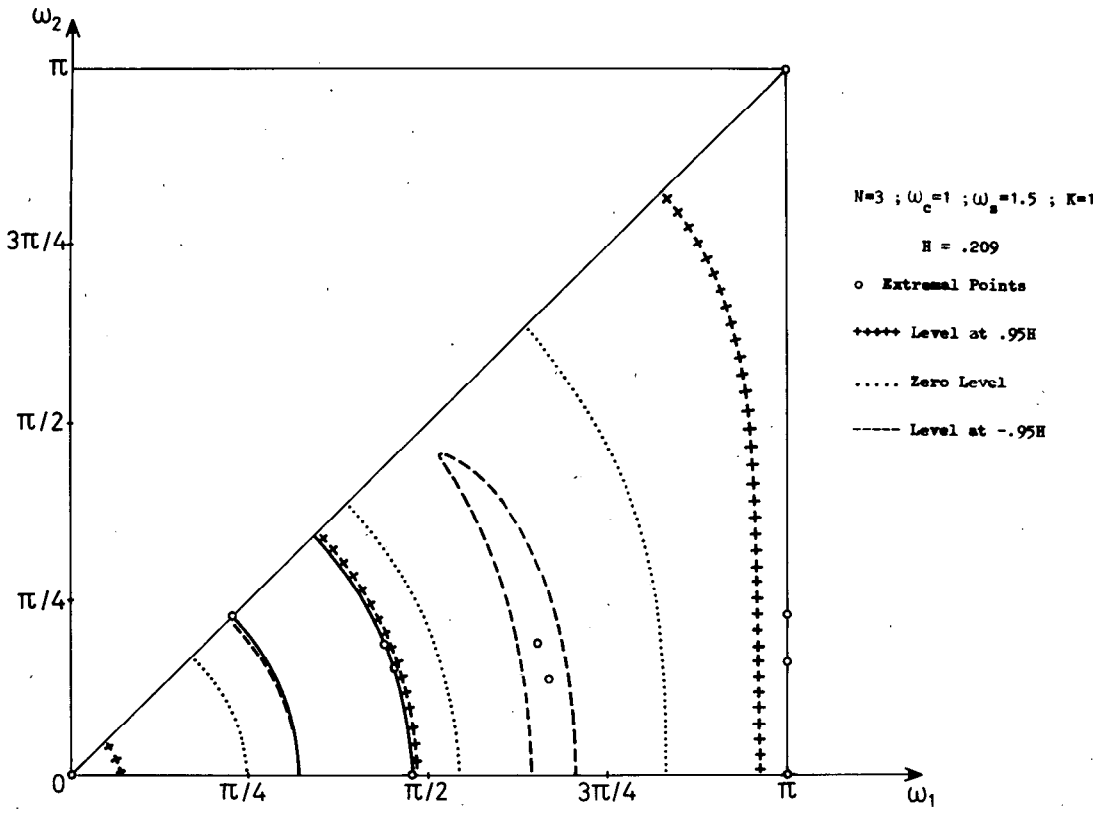

Fig. 2. Typical shape of the optimal error function.

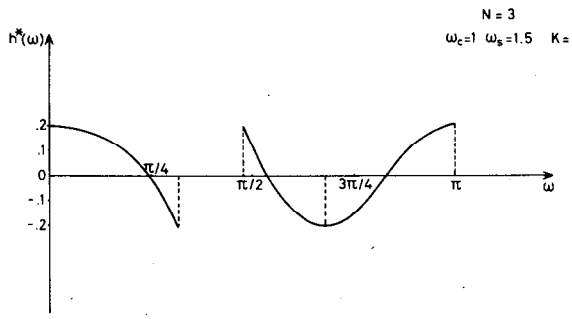

Fig. 3. Optimal one-dimensional error curve.

The weight function $w(x)$ introduced in the norm (4) is chosen equal to K in the passband and to unity in the stopband to get different tolerances in both bands.

A typical example of the optimal error function is presented in Fig. 2 for the third-degree filter with $\omega_c = 1$, $\omega_s = 1.5$, and $K = 1$. One recognizes the error shape of the optimal one-dimensional filter drawn in Fig. 3, rotated in all directions with the final reference points located approximately on circular ridges whose radii correspond to the abscissas of the extremal points of Fig. 3. In fact, on a circle ($\omega_1 = R \cos t$, $\omega_2 = R \sin t$), each base function (3) can be expanded in terms of Bessel functions of the first kind as

$$\cos I\omega_1 \cos J\omega_2 + \cos J\omega_1 \cos I\omega_2 = 2J_0[R(\sqrt{I^2 + J^2})] + 4 \sum_{k=1}^{\infty} J_{4k}[R(\sqrt{I^2 + J^2})] \cos 4k\alpha_{IJ} \cos 4kt$$

where $\alpha_{IJ} = \sin^{-1} [I/(I^2 + J^2)^{1/2}]$. Since the Bessel function $J_n(x)$ has a zero of multiplicity n at the origin, it has negligible values when its argument x is small with respect to the order n. For the above-mentioned third-degree filter, on the boundary circle $\omega_s = 1.5$, the approximating function can be represented with an error on the order of 10^{-3}, as a second-degree trigonometric polynomial in $4t$. Hence, since there are three extremal points on ω_1, the error function attains approximately the Chebyshev deviation at all points of the circle. The circular symmetry of the error function is less pronounced for circles with larger radii and, in particular, the ridge corresponding to the last extremal point of Fig. 3, i.e., the Nyquist frequency, is the straight line $\omega_1 = \pi$. As, for $\omega_1 = \pi$, the approximating function is a third-degree polynomial in ω_2, the four extremal points located on $\omega_1 = \pi$ produce a ridge of strictly constant height for all ω_2. In order to take advantage of this property during the exchange algorithm, one can split the approximating function into two parts according to (37) such that

$$G_1(x) = \sum_{I=0}^{N} \alpha_I (-1)^I \cos I\omega_1 \cos I\omega_2$$

$$G_2(x) = -\sum_{I=1}^{N} \sum_{J=0}^{I-1} B_{IJ}[\cos I\omega_1 - (-1)^{I+J} \cos J\omega_1] \cdot [(-1)^{I+J} \cos I\omega_2 - \cos J\omega_2]$$

with

$$\alpha_I = (-1)^I 2B_{II} + \sum_{J=0}^{I-1} (-1)^J B_{IJ} + \sum_{J=I+1}^{N} (-1)^J B_{JI}.$$

TABLE I

$\omega_c=1$; $\omega_s=2$; $K=1$

Precision factor = 10^{-3}

N	Single Exchange		Multiple Exchange	
	Iteration Number	Computation Time (Sec)	Iteration Number	Computation Time (Sec)
1	1	9	1	9
2	3	9	3	9
3	16	46	18	54
4	25	114	15	118
5	85	564	30	401
6	184	2276	51	1276

TABLE II

$\omega_c=1$; $\omega_s=2$; $K=1$

Precision factor = 10^{-2}

N	Single Exchange		Multiple Exchange	
	Iteration Number	Computation Time (Sec)	Iteration Number	Computation Time (Sec)
1	1	9	1	9
2	2	8	2	8
3	9	29	9	41
4	15	85	9	77
5	68	534	21	360
6	128	1714	16	463

Since $G_2(x) = 0$ for $\omega_1 = \pi$, imposing a constant value of $f(x)$ on $\omega_1 = \pi$ amounts to canceling all coefficients α_I except α_0. It turns out that the exchange algorithm modified this way yields very good results, at least within the precision factor of 10^{-3}, if π belongs to the extremal point set of the best univariable approximation. On the other hand, since nonuniqueness generally occurs only when π is not an extremal point of the optimal unidimensional error, the following rule of thumb was found to be very efficient for solving the approximation problem. It consists in performing the approximation in two steps:

1) approximation along the ω_1-axis by a univariable approximating function which yields the extremal point set $R^* = \{x_k^*; 1 \leq k \leq N+2\}$;

2) if $x_{N+2}^* = \pi$, approximation in the whole region by maintaining a constant value of the error on $\omega_1 = \pi$; and

2′) if $x_{N+2}^* \neq \pi$, approximation by means of the procedure of Section III-B which allows a direct computation of the strict approximation.

In both cases, the points of the initial reference for starting the exchange algorithm are equally spaced on circles of radius x_k^* and their number is proportional to the index k.

As regards the rate of convergence of the iterative process, a necessary condition [13] for the Remez algorithm, applied to univariable functions consisting of m base functions, to converge quadratically, is that the optimal error function has exactly $m + 1$ extremal points: on the contrary, the convergence is slower if their number exceeds $m + 1$. In particular, this property has been verified for one-dimensional filters [14]. Although such a theorem has not been proved for bivariable functions, it is reasonable to expect a similar phenomenon for two-dimensional filters. Since the optimal error exhibits ridges of extrema, the convergence is slow as evidenced by Table I which gives the iteration number and the computation time (on the Electrologica X8 computer: cycle time = 1 μs) of the single and multiple exchange algorithms in the case of $\omega_c = 1$, $\omega_s = 2$, $K = 1$ and for various degrees. If both algorithms are comparable for N ranging from 1 to 4, the multiple exchange method becomes quite advantageous when the degree is higher. In order to stress the importance of the precision factor on the computation time, Table II indicates how the figures of Table I vary when the precision factor is relaxed from 10^{-3} to 10^{-2}: from these results, it appears that the difficulty for designing the filters, especially the highest degree filter, strongly depends on the quality of approximation which is required by the practical problem.

TABLE III

$N=1$; $\omega_c=1.5$; $\omega_s=2$; $K=10$

Filter Coefficients	Extreme Optimal Solution A_1	Extreme Optimal Solution A_2	Strict Approximation
A_{00}	.793917	.790117	.790699
$A_{10} = A_{01}$	.0552223	.0571223	.0568314
A_{11}	.017049	.016099	.0162445

To serve as an illustration for nonuniqueness, Table III presents several optimal solutions relative to the parameters $N = 1$, $\omega_c = 1.5$, $\omega_s = 2$, and $K = 10$. When degeneracy with respect to the ω_1-axis occurs for a first-degree polynomial, there is only one vanishing characteristic coefficient, and the number of extreme optimal solutions is exactly equal to 2. Therefore, the two extreme optimal solutions A_1 and A_2 given in the first two columns of Table III completely define the set of best approximations and one can verify that the strict approximation of the third column is the convex combination $\delta A_1 + (1 - \delta)A_2$ with $\delta = 0.153124$.

Another example of nonuniqueness results from the data set $N = 3$, $\omega_c = 1.5$, $\omega_s = 2$, and $K = 1$. In that case, there are tens of different extreme optimal solutions and it is no longer possible to mention all of them. As regards the strict approximation, it is worth pointing out that performing the approximation outside the ω_1-axis with the function (40) yields four extremal points located on the parallel of the Cartesian grid closest to the ω_1-axis, which corresponds to a new degenerate problem. To remove this difficulty, it is thus necessary to apply the procedure of Section III-B to the reduced approximation problem. Fig. 4 shows the optimal error function of the strict approximation together with the three critical point sets.

If the exchange algorithm is used to compute an extreme optimal solution, since the problem is degenerate with respect to the final reference, the last iterations consist of static exchanges. In order to ensure unconditional convergence, one has to resort to the perturbation technique of Section I-D. For the example at hand, the number of iterations of the unperturbed and perturbed exchange algorithms is equal to 36 and 41, respectively, while the corresponding computation times are 92 s and 137 s. Therefore, although both iteration numbers are comparable,

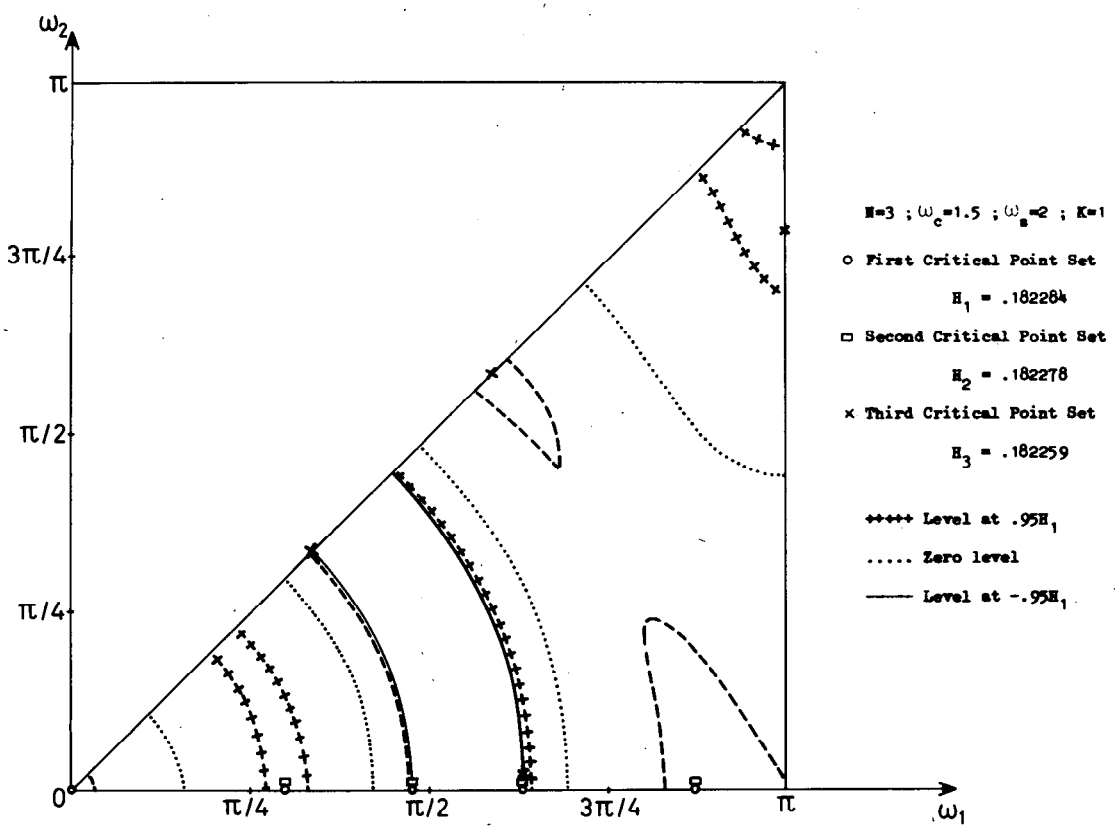

Fig. 4. Optimal error function and critical point sets of a strict approximation.

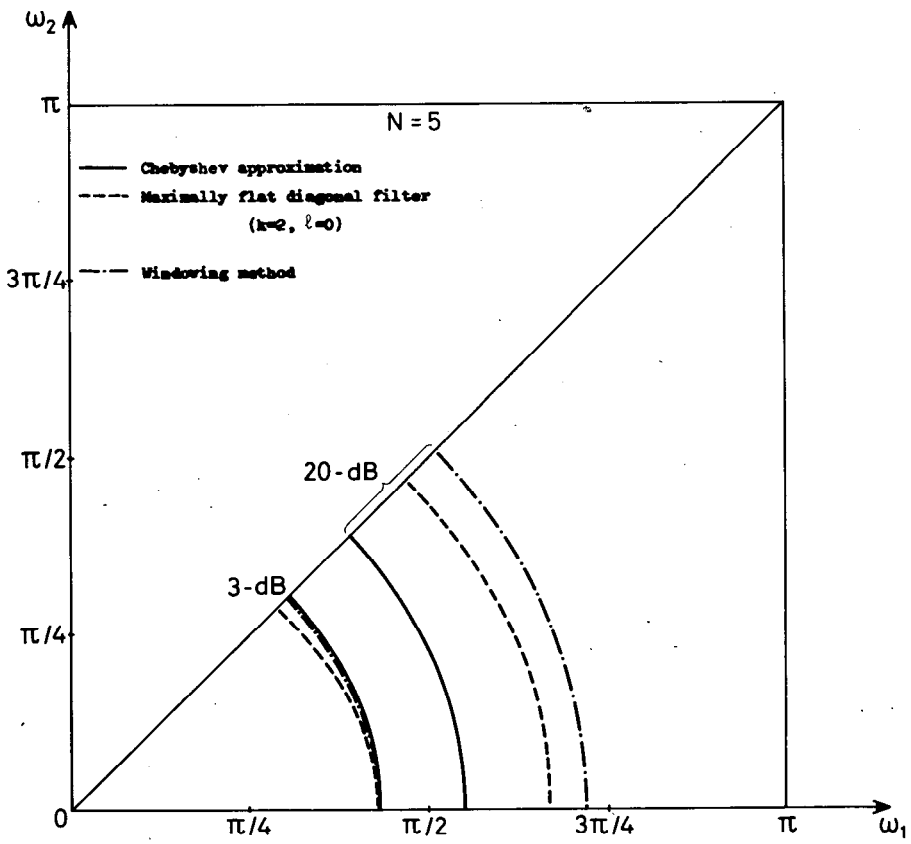

Fig. 5. Comparison with other design methods.

the computation time of the perturbed algorithm is higher owing to the more involved computations at each iteration step. On the other hand, experience has proven that the unperturbed exchange algorithm never failed to converge. Hence, as pointed out in Section I-D, the perturbation method is more important from the theoretical standpoint than from the practical one.

Hu and Rabiner [3] have used linear programming to design frequency sampling and optimal (in the sense of Chebyshev approximation) two-dimensional filters. Results for the latter design have been reported only for a few filters of the fourth degree and were found to be comparable with those obtained by the exchange algorithm of this paper. By lack of a detailed description of their methods, it seems likely that the small discrepancies are due to different discretization schemes of the continuous approximation domain. As regards the frequency sampling design, the comparison was made on a few samples of the filters of degree 12 published in [3]. For these examples, design by Chebyshev approximation showed that a degree 7 was already sufficient to achieve the same filtering performance.

To visualize the superiority of the Chebyshev approximation on other design methods, a fifth-degree filter was computed having a 3-dB passband and a 20-dB stopband (Fig. 5). For comparison, the 3- and 20-dB contour lines have been plotted for a maximally flat diagonal filter [4] and for a filter designed with a windowing technique [2], both having the same degree and the same passband limit along the ω_1-axis. The circular symmetry of these designs is comparable but the selectivity given by the Chebyshev approximation is significantly better, i.e., smaller transition bandwidth for substantially the same passband boundary.

V. Conclusion

The preceding analysis has clearly shown that the Chebyshev approximation of two-dimensional nonrecursive filters is fundamentally more difficult and inherently slower than in the one-dimensional case. The reasons behind this deserve some further comments.

First, two-variable approximating functions do generally not satisfy the Haar condition and this is reflected by the existence of degeneracies. If degeneracy occurs before the end of the algorithm it may fail to converge because the same reference is repeated periodically. To force convergence even in this case, it was shown how to complement the algorithm with a suitable perturbation technique. This technique is only of theoretical interest, however, because it is computationally very costly and because in practice cycling of the algorithm has never been observed.

A second difference is that the optimal solution is not characterized by a simple alternation of the error like in the one-dimensional problem and, as a consequence, the rule for exchanging one point of the reference becomes more complicated. This fact makes the one-by-one exchange more time-consuming and even prevents a simultaneous exchange, a procedure which precisely makes the algorithm so efficient in the single variable case. An improvement in this direction was made, however, by proposing a method which allows to exchange several points simultaneously, and the experiments have demonstrated that it leads to important reductions of the run time when the degree is high.

Further, the maximum detection over a two-dimensional region is much longer than the corresponding work over an interval and is made even slower by the ridges in the error function whose existence was already mentioned in [3]. Finally, the computation time is also increased by the fact that in two dimensions there exists no counterpart of the Lagrange interpolation formula for the leveled reference function and by the fact that the number of coefficients is proportional to the square of the degree.

These considerations show that the nonfulfilled Haar condition is not the only reason for the slowness of the two-dimensional approximation. Several measures were taken to improve this situation. Besides the multiple exchange already mentioned, a heuristic procedure was proposed where the approximation is first performed on the ω_1-axis or where the value of the approximating function is kept constant over $\omega_1 = \pi$. Its advantage lies in a significant saving of computation time and although it does not always give the optimal solution, the accuracy of the result is in most cases satisfactory. It is clear that future efforts should be aimed at further reducing the run time.

Comparison with other design methods (frequency sampling, maximally flat criterion, and windowing technique) has made evident the superiority of the Chebyshev design for what regards selectivity. For this norm, the one-by-one exchange is known to be more economical than linear programming [15] and this speed advantage is increased by the possibility of performing a multiple exchange.

This paper has solved the typical problem of designing a two-dimensional nonrecursive filter which approximates circularly symmetrical low-pass specifications. Most of the results, however, are not restricted to circular symmetry or low-pass specifications but apply to the design of any filter whose frequency response can be put under the form (2) within a linear phase factor.

Acknowledgment

The authors wish to thank J. Praet for his programming of the exchange algorithm.

References

[1] T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image processing," *Proc. IEEE*, vol. 59, pp. 1586–1609, Nov. 1971.
[2] T. S. Huang, "Two-dimensional windows," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 88–89, Mar. 1972.
[3] J. V. Hu and L. R. Rabiner, "Design techniques for two-dimensional digital filters," *IEEE Trans. Audio Electroacoust.*, vol. AU-20, pp. 249–257, Oct. 1972.
[4] Y. Kamp and J. P. Thiran, "Maximally flat nonrecursive two-dimensional digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-21, pp. 437–449, May 1974.
[5] J. H. McClellan, "On the design of one-dimensional and two-dimensional FIR digital filters," Ph.D. dissertation, Rice University, Houston, Tex., Apr. 1973.
[6] E. Stiefel, "Über diskrete und lineare Tschebyscheff-Approx-

imationen," *Numer. Math.*, vol. 1, pp. 1–28, 1959.
[7] S. I. Gass, *Linear Programming*. New York: McGraw-Hill, 1958.
[8] 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.
[9] L. Collatz, "Approximation von Functionen bei einer und bei mehreren unabhängigen Veränderlichen," *Z. Angew. Math. Mech.*, vol. 36, pp. 198–211, May/June 1956.
[10] L. Bittner, "Das Austauschverfahren der linearen Tschebyscheff-Approximation bei nicht erfüllter Haarscher Bedingung," *Z. Angew. Math. Mech.*, vol. 41, pp. 238–256, June 1961.
[11] J. Descloux, "Dégénérescence dans les approximations de Tschebyscheff linéaires et discrètes," *Numer. Math.*, vol. 3, pp. 180–187, 1961.
[12] J. R. Rice, *The Approximation of Functions*, vol. 2, *Nonlinear and Multivariate Theory*. Reading, Mass.: Addison-Wesley, 1969.
[13] G. Meinardus, *Approximation von Funktionen und ihre numerische Behandlung*. Berlin: Springer, 1964.
[14] A. N. Godard, "Design of optimum FIR bandpass filters through frequency transformations," *Philips Res. Rep.*, vol. 29, pp. 243–252, June 1974.
[15] E. Stiefel, "Note on Jordan elimination, linear programming and Tchebycheff approximation," *Numer. Math.*, vol. 2, pp. 1–17, 1960.

Timing Jitter in Digital Filtering of Analog Signals

BEDE LIU, FELLOW, IEEE

***Abstract*—Effects of timing jitter in digital filtering of analog signals are analyzed. Jitter may be introduced into the system when the incoming analog signal is being sampled and when the samples of the output of the digital filter are being read out for reconstruction back into an analog signal. These two jitters are called readin and readout jitters, respectively. General expressions are derived for the spectrum of the analog output in the presence of jitter and for the mean-square error introduced by the jitter. Four specific cases are then considered. They are readin jitter only, readout jitter only, independent jitter, and locked jitter. In each case, the general expression for the spectrum of the analog output is reduced to a more explicit form, and a simple upper bound on the mean-square error is derived. An example is included to illustrate the effect of high-frequency jitter and the effect of jitter due to 60-cycle pickup.**

I. Introduction

TO PROCESS an analog signal by a digital filter, the analog signal is first converted into a digital signal by sampling. The digital signal is then processed by the digital filter, and the filter output is converted to an analog signal by means of a reconstruction filter. Although the input analog signal is supposed to be sampled periodically, say every T second, the actual sampling instants may be different from the nominal sampling instants by a small amount due to noise, imperfections of the clock and other factors. This difference is known as *timing jitter*. Similarly, the digital filter produces an output sample nominally every T second for converting into an analog signal. But there may also be a timing jitter present in the read out.

This paper treats the effect of these jitter on the processing of analog signals by digital filters. The effect of finite word length [1], [2] such as input quantization and roundoff accumulation is not discussed here.

The problem to be analyzed is illustrated in Fig. 1.

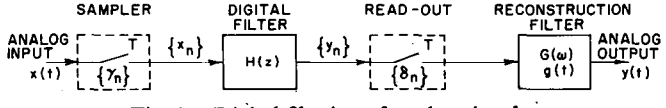

Fig. 1. Digital filtering of analog signals.

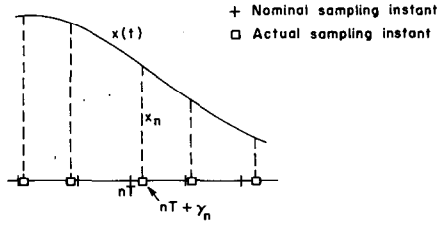

Fig. 2. Readin jitter.

$x(t)$ and $y(t)$ are, respectively, the analog input and output signals. The digital filter has a unit sample response $\{h_n\}$ and its transfer function is[1] $H(z) = \sum_n h_n z^{-n}$. The input and output of the digital filter are denoted by $\{x_n\}$ and $\{y_n\}$. The analog reconstruction filter has a transfer function $G(\omega)$ and an impulse response $g(t)$. Although the input $x(t)$ is supposed to be sampled at nT, $n = \cdots, -1, 0, 1, \cdots$, the actual sampling instant is, however, $nT + \gamma_n$ as illustrated in Fig. 2. The sequence $\{\gamma_n\}$ is called the *readin jitter* and is assumed to be small compared with T. The output of the digital filter is read out at time $nT + \delta_n$, where $\{\delta_n\}$ is called the *readout jitter* and is also assumed to be small. $x(t)$ is taken to be a wide sense stationary (WSS) random signal with autocorrelation function $R_{xx}(\tau) = E\{x(t)x(t+\tau)\}$ and power spectral density $\Phi_{xx}(\omega) = \int_{-\infty}^{\infty} R_{xx}(\tau)\, e^{-j\tau\omega}\, d\tau$. It is assumed to be bandlimited to π/T rad/s, i.e., $\Phi_{xx}(\omega) = 0$ for $|\omega| > \pi/T$. The jitter sequences $\{\gamma_n\}$ and $\{\delta_n\}$ are assumed to be jointly stationary to order 2 with zero mean

Manuscript received July 19, 1973; revised November 15, 1973. This research was supported by the Air Force Office of Scientific Research under Grant AFOSR 71-2101.

The author is with the Department of Electrical Engineering, Princeton University, Princeton, N.J. 08540.

[1] Unless otherwise specified, summation are from $-\infty$ to $+\infty$.