Transactions of Illinois Academy of Sciences
(1984), Volume 77, 3 and 4, pp. 261-270

ON THE DESIGN OF DIGITAL FILTERS WITH MONOTONIC TRANSITION REGIONS USING "DON'T CARE" BANDS.

G. Cortelazzo and M.R. Lightner
Department of Electrical and Computer Engineering
University of Colorado
Boulder, CO 80309
and
Coordinated Science Lab
University of Illinois
Urbana, Illinois

ABSTRACT

This paper is concerned with the problem of imposing monotonicity over the transition regions of the transfer functions of digital filters via conditions both mathematically sufficient and computationally convenient.

Ignoring the transition regions in the design formulation, i.e. treating them as "don't care" bands is, in some cases, sufficient for obtaining monotonic behavior over these regions. In this paper we give the conditions under which the use of a single "don't care" band is sufficient for imposing monotonicity over the transition region. The L_2 and L_∞ design formulations satisfy these conditions. This non obvious fact is well-known to filter designers, but its mathematical proof, to the authors' knowledge, has not yet appeared. The proof is valid for both Finite Impulse Response (FIR) filters and Infinite Impulse Response (IIR) filters.

However, when there are two or more transition regions, as in the case of multiband filters, ignoring the transition regions will, in general, not lead to the desired monotonic behavior. Therefore the design of multiband filters requires either the use of algorithms for unconstrained approximation with additional special techniques (e.g., [1,2]) or the use of algorithms for constrained approximation (e.g., [3-5]).

This paper presents the difficulties encountered in the multiband case when attempting to obtain monotonic transition regions by using alternate constraint sets based upon the use of "don't care" bands plus boundary conditions.

This work was supported by the National Science Foundation under grant ECS - 8007075.

I. INTRODUCTION

The traditional design requirements in the frequency domain for a filter are given in terms of desired performance over the pass and stop bands. In addition, there is an implicit requirement for monotonic behavior over those bands separating the pass and stop bands, i.e. the transition bands. In formulating the filter design problem as an approximation problem the typical approach is to approximate only over the pass and stop bands. This paper considers methods for imposing monotonicity over the transition bands which do not interfere with the approximation over the pass and stop bands [8].

One obvious approach for obtaining monotonicity over the transition bands is to formulate objective or desired functions over these bands and include these extra functions in any approximation or optimization. However, formulating this monotonicity objective is not a natual task for the filter designer. Furthermore, almost any monotonicity objective will interfere with the approximation over the pass and stop bands.

A second obvious approach is to either put upper bounds on the approximant or to constrain the sign of the derivative of the approximant over the transition bands. The computational efficiency of these "brute force" techniques is related to the filter order and to the size of the transition regions. Furthermore, they cannot be proven to be mathematically sufficient. Nevertheless they have been used in practice both because they can be effective and for the lack of better alternatives.

A third method that has been extensively used in the literature is to completely ignore the transition bands in the formulation of the approximation problem. That is, the transition bands are treated as "don't care" bands. There are cases where this approach does not work ([1,6]) and cases where this approach works, i.e. produces filters with monotonic transition bands. We will show that the success of this approach can only be guaranteed for filters with a single transition band. This technique, when it works, is the computationally simplest way to achieve a monotonic transition region.

Section II thoroughly examines the case of the single "don't care" band and presents a complete monotonicity theory. The analysis is accomplished by first stating general monotonicity conditions and then by showing that these are automatically met in certain design formulations with one "don't care" band, namely the L_∞ and L_2 formulations.

A consequence of this theory is that unconstrained minimax optimizaton algorithms are sufficient for the design of low and high pass filters formulated with one "don't care" band both for FIR and IIR filters.

Section III considers the case of multiple "don't care" bands. The sufficient conditions for monotonicity for two "don't care" bands require certain boundary conditions. These boundary conditions are shown to be too complicated to be automatically met by L_∞ or L_2 approximants. As a consequence the design of filters with two or more "don't care" bands requires algorithms for constrained approximation (unless the designer wants to try techniques based on empirical trial and error procedures [1,2]).

Unfortunately, a useful generalization of the "don't care" band approach, i.e. the determination of the boundary constraints on the derivatives that are necessary and sufficient to impose monotonicity over two to more "don't care" bands is a very difficult task. The specific difficulty is finding the most general such constraints that are still practically interesting in an approximation setting. A way to pursue such a generalization together with the associated difficulties is discussed in Section III. As an examples, sufficient conditions for the "don't care" bands, are presented.

Section IV presents conclusions and indicates directions for future research.

II. ONE TRANSITION BAND

We will consider filters whose transfer functions are members of the class:

$$S[0,1] = \left\{ R_{l,m} = \frac{\sum_{i=0}^{l} a_i x^i}{\sum_{i=0}^{m} b_i x^i} \;\middle|\; b_0 = 1;\quad x \in [0,1]; \sum_{i=0}^{m} b_i x^i \neq 0 \quad x \in [0,1] \right\} \tag{1}$$

Each $R_{l,m}(x)$ is assumed to have no pole-zero cancellations. Furthermore, notice that the derivative with respect to x of any $R_{l,m}(x)$, which will be denoted as $R'_{l,m}(x)$ has no poles on $[0,1]$. Let $M = m + l$. The numerator of $R'_{l,m}(x)$ is a polynomial with degree at most $M-1$, therefore $R'_{l,m}(x)$ can have at most $M-1$ zeros on $[0,1]$.

Let $K = [0,b] \cup [c,1]$ with $0 < b < c < 1$, and $\mathring{K} = (0,b) \cup (c,1)$.

Finally define

$$d(x) = \begin{cases} 1 & x \in [0,b] \\ 0 & x \in [c,1] \end{cases} \tag{2}$$

$$e(x) = R_{l,m}(x) - d(x) \tag{3}$$

$e'(x)$ denotes the derivative of $e(x)$ with respect to x, using appropriate right and left hand derivatives at the boundary points.

The first proposition we introduce states conditions over $[0,b] \cup [c,1]$ and at the boundary points $\{b,c\}$ which are sufficient to guarantee that $R_{l,m}(x)$ is strictly monotonic in (b,c). This corresponds to the idea of characterizing certain sufficient conditions for monotonicity over the transition regions.

Proposition 1: If $R_{l,m}(x) \in S[0,1]$ approximates $d(x)$ and satisfies:

(i) $R_{l,m}(x)$ interpolates $d(x)$ at, at least M points in $\mathring{K}$

(ii) $e'(b) < 0$, $e'(c) < 0$

then $R_{l,m}(x)$ is strictly monotonic on (b,c), i.e.:

$$R'_{l,m}(x) \neq 0 \quad x \in (b,c) \tag{4}$$

Proof: Assume that $R_{l,m}(x)$ interpolates $d(x)$ at M_1 points in $(0,b)$ and at M_2 points in $(c,1)$, where $M = M_1 + M_2$. By the mean value theorem, $R'_{l,m}(x)$ has $M_1 - 1$ zeroes on $(0,b)$ and $M_2 - 1$ zeros on $(c,1)$, i.e. $R'_{l,m}(x)$ has a total of $M-2$ on $\mathring{K}$. Since $R'_{l,m}(x)$ has a numerator polynomial of degree $M-1$ and a denominator polynomial of degree 2m, these observations allow $R'_{l,m}(x)$ to be written as

$$R'_{l,m}(x) = (x - \alpha)Q(x) \qquad Q(x) \neq 0 \quad x \in [b,c]. \tag{5}$$

$Q(x)$ represents the rational function having as numerator the $M-2$ zeros of $R'(x)$ laying in $\mathring{K}$ and denominator equal of that of $R'_{l,m}(x)$.

Assumption (i) and decomposition (5) emphasize that $R'_{l,m}(x)$ has only one zero free to occur on (b,c).

Assumption (ii) is equivalent to

$$R'_{l,m}(b) < 0 \quad R'_{l,m}(c) < 0 \tag{6}$$

If $\alpha \in (b,c)$ then $b - \alpha < 0$ and $c - \alpha > 0$, therefore (6) implies $Q(b) > 0$ and $Q(c) < 0$. This is a contradiction since it implies $Q(x) = 0, x \in (b,c)$, contrary to the assumption (i) which is embodied in (5). Thus the (4) must hold.

It should be clear from the structure of $R'_{l,m}(x)$ in (5) that $R'_{l,m}(b)$ and $R'_{l,m}(c)$ cannot be simultaneously zero. $R'_{l,m}(b) = 0$ (and hence $R'_{l,m}(c) < 0$) implies $\alpha = b$, which still satisfies (4).

Similarly for $R'_{l,m}(c) = 0$.

□

Remark: The proof has been given for $R_{l,m}(x)$ interpolating $d(x)$ at M points only. The case of $M + 1$ interpolating points is trivially proven (assumption (i) being sufficient to prove (4), since $R'_{l,m}(x)$ has all its $M - 1$ zeros on $\mathring{K}$).

Given a sequence of points $x_1, x_2, \ldots$ in $[0,1]$ if

$$|e(x_i)| = \max_{x \in [0,1]} |e(x)| \tag{7}$$

and

$$e(x_i) = -e(x_{i-1}) \quad i = 1,2,\ldots.$$

$e(x)$ is said to alternate on the x_i's.

Proposition 2 strongly characterizes the behavior on $[0,b] \cup [c,1]$ of the rational approximants to $d(x)$ exhibiting the alternating property (7). Figure 1 illustrates the behavior by Proposition 2.

Proposition 2: A rational function $R_{l,m}(x) \in S[0,1]$ approximating $d(x)$ in such a way that $e(x)$ has, at least, $M + 2$ alternations on K has the following properties:

(a) b and c are extremal points of $e(x)$ on K

(b) $e(b) < 0 \quad e'(b) < 0 \quad e(c) > 0 \quad e'(c) < 0$

Proof: Since all the alternations of $e(x)$ correspond to local extrema of $R_{l,m}(x)$ (i.e. to zeros of $R'_{l,m}$) unless they occur at the boundary points $\{0,b,c,1\}$ and since $R'_{l,m}(x)$ cannot have more than $M - 1$ zeros on $[0,1]$, the assumption of, at least, $M + 2$ alternations for $e(x)$ implies one of the following cases, namely $e(x)$ with:

(i) $M - 1$ extrema in $\mathring{K}$ and 4 extrema at $\{0,b,c,1\}$

(ii) $M - 1$ extrema in $\mathring{K}$ and 3 extrema at $\{0,b,1\}$

(iii) $M - 1$ extrema in $\mathring{K}$ and 3 extrema at $\{0,c,1\}$

(iv) $M - 1$ extrema in $\mathring{K}$ and 3 extrema at $\{0,b,c\}$

(v) $M - 1$ extrema in $\mathring{K}$ and 3 extrema at $\{b,c,1\}$

(vi) $M - 2$ extrema in $\mathring{K}$ and 4 extrema at $\{0,b,c,1\}$

Notice first that case (ii) or (iii) are contradictory, therefore they cannot occur. Case (ii), in fact can occur only by contradicting alternation as shown in Figure 2, or by calling for $R'_{l,m}(b) = 0$ as in Figure 3 which implies $R'_{l,m}(x)$ with one zero more that it can have. Similar argument proves the contradictory nature of case (iii).

Property a is satisfied in all the cases that can occur.

Property b is easily seen to be satisfied in all cases except (vi). If not, a further extremum of $R_{l,m}(x)$ would exist in $[b,c]$, but $R_{l,m}(x)$ is already assumed to have $M-1$ extrema in K, and therefore cannot have another one.

One way to prove property b for case (vi) is exhaust all the possible situations for $e'(b)$ and $e'(c)$ to see that only those corresponding to property b are immune from contradiction. Let us distinguish the following 4 subcases:

Subcase I: $e'(b) = 0$, $e'(c) = 0$. Not admissible since it implies that $R'_{l,m}(x)$ has M zeros.

Subcase II: $e'(b) = 0$, $e'(c) \neq 0$. Not admissible since if $e'(c) < 0$ then alternation is violated (see Fig. 4) and if $e'(c) > 0$ a further extremum exists in (b,c) (see Fig. 5). Therefore $R_{l,m}(x)$ ends up with M extrema in $[0,1]$ (namely $M-1$ in K, one at b and one in (b,c)).

Subcase III: $e'(b) \neq 0$, $e'(c) = 0$. Not admissible by the same argument used in Subcase II.

Subcase IV: $e'(b) \neq 0$, $e'(c) \neq 0$.

This subcase specializes into 16 situations according to the specific signs of $e'(b)$, $e'(c)$ and $e(b)$, $e(c)$. Each one of these situations except the one $e(b) > 0$, $e(c) < 0$, $e'(b) < 0$, $e'(b) > 0$ (i.e. that satisfies property b) is contradictory since it violates at least one of the following points:

(A) $e(b)$ and $e(c)$ don't alternate

(B) the situation implies the existence of two further zeros of $R'_{l,m}(x)$ in $[b,c]$ (this happens in the four subcases corresponding to $e'(b) > 0$ and $e'(c) > 0$).

(C) the sign of the derivative $R'_{l,m}(x)$ at b or c contradicts the assumption that b or c is an extremal point. (An example of violation of point C is the situation $e(b) < 0$, $e(c) < 0$, $e'(b) < 0$, $e'(c) < 0$ shown in Fig. 6. Clearly b cannot be an extremal point.) Thus both (a) and (b) are satisfied.

□

A result embedded in Proposition 2 is that the rational approximants to d(x) exhibiting the alternating property "naturally" satisfy the assumptions of Proposition 1. The following Corollary makes this observation explicit.

Corollary: A rational function $R_{l,m}(x) \in S[0,1]$ approximating $d(x)$ in such a way that $e(x)$ has at least $M+2$ alternations on K is strictly monotonic in (b,c).

Proof: The occurrence of at least $M+2$ alternations of $e(x)$ in K implies that $R_{l,m}(x)$ interpolates $d(x)$ has at least M points in K. Therefore assumption i) of Proposition 1 is satisfied. Remember that by Proposition 2, property b $e'(b) < 0$ and $e'(c) < 0$. Therefore assumption ii) of Proposition 1 is also satisfied. Proposition 1 then guarantees that $R_{l,m}(x)$ is strictly monotonic on $[b,c]$.

□

Thus the behavior of a rational approximant to $d(x)$ over the whole $[0,1]$ is strongly characterized by Proposition 2 and the Corollary. Fig. 7 illustrates this behavior, the key point being the values of $e(b)$, $e(c)$ and monotonicity over the transition region (b,c).

GENERALIZATIONS

1. In the above Propositions and Corollary, with obvious modifications, d(x) can be replaced by any f(x) corresponding to an e(x) = f(x) − R(x), with

$$e'(x) = 0 \; x \in K => R'_{l,m}(x) = 0 \; x \in K \tag{13}$$

   For example high-pass filters satisfy (13). In general, all practical spectrum-shaping filters satisfy (13).

2. In the above Propositions and Corollary the error $e(x)$ can be defined as

$$e(x) = W(x)(d(x) - R_{l,m}(x))$$

   with $W(x) > 0$ and such that (13) is still fulfilled.

APPLICATIONS

The main advantage of Proposition 1 is that the sufficient conditions for monotonicity are in harmony with good approximating properties. As a confirmation that the monotonicity constraints over the transition band don't spoil the approximation over the pass and stop bands we note that the required conditions are spontaneously exhibited in several classes of approximants. For example, L_2 − optimal approximants satisfy the hypotheses of Proposition 1 [7,pp.111] as well as the L_∞ − optimal approximants, both polynomial and rational, as pointed out in the Corollary. Proposition 1 may also be useful in the construction of new filter design algorithms.

Proposition 2 represents a strong characterization over the pass and stop bands of the approximants with the alternation property (7). The main application of Proposition 2 is in connection with the Corollary. Together they guarantee that the L_∞ filter formulation with one don't care band is appropriate when monotonic behavior is desired over the transition band. Therefore any algorithm for unconstrained L_∞ approximation, e.g. the Remez algorithms for polynomials and for rational functions, and the Maehly algorithm, and the differential-corrector algorithm, deliver approximants with monotonic behavior over the "don't care" band. The Corollary of Proposition 2 is a generalization of a theorem given in [8] for the polynomial case only.

Comment

One can find many other types of pass-stop band and boundary conditions sufficient for strict monotonicity over one "don't care" band. For instance conditions involving less interpolating points and, of course, different boundary derivative requirements. These conditions have not been investigated because they do not correspond to interesting approximating properties.

III. SOME CONSIDERATIONS FOR THE CASE OF TWO OR MORE "DON'T CARE" BANDS

The practical utility of the propositions in the previous section is that they fully justify the use of unconstrained minimax approximation algorithms in problems with a "don't care" band over which a monotonic approximant is desired. Proposition 1 captures conditions sufficient for monotonicity over the transition region that are spontaneously satified by the best L_∞ approximant, both rational and polynominal. This fact is made explicit by Proposition 2 and the Corollary.

For approximation problems (minimax or other) with several transition regions the use of "don't care" bands alone is no longer sufficient to guarantee monotonic approximants over the transition regions. An efficient family of constraints, sufficient to guarantee monotonicity of the approximants on the transition regions, is based on the use of "don't care" bands together with constraints on the sign of the derivative of the approximant on the boundary of the transition regions. Clearly these constraints can only be exploited with constrained (minimax or other) approximation algorithms [4,5].

It is worthwhile devoting special attention to minimax approximation because it is particulary suitable to filter design formulations and it can be very computationally efficient.

Since monotonicity conditions for multiple "don't care" bands necessarily involve derivatives of more than first order. It is useful to distinguish between polynomial and rational functions. The latter turn out to be much harder to characterize.

For the special case of the two "don't care" bands and polynomial approximants (FIR filters) the following proposition gives an example of sufficient monotonicity conditions.

Proposition 3:

Notation:

Let $0 < b_1 < c_1 < b_2 < c_2 < 1$.

and

$K = [0,b_1] \bigcup [c_1, b_2] \bigcup [c_2, 1]$

$K = (0,b_1) \bigcup (c_1 b_2) \bigcup (c_2, 1)$

with

$$d_2(x) = \begin{cases} 1 & x \in [0,b_1] \\ 0 & x \in [c_1,b_2] \\ 1 & x \in [c_2,1] \end{cases}$$

$P(x)$ denotes a M-order polynominal with real coefficients

$P_m^{(i)}$ denotes the i-th derivative of $P_m(x)$ with respect to x.

Statement: If $P_m(x)$ approximates $d(x)$ satisfying the following conditions:

(i) $P_m(x)$ interpolates $d(x)$ at, at least, $M-1$ points in $\mathring{K}_2$

(ii) $P_m^{(1)}(b_1) < 0 \quad P_m^{(1)}(b_2) > 0$

$P_m^{(1)}(c_1) < 0 \quad P_m^{(1)}(c_2) > 0$

(iii) $P_m^{(2)}(b_1) < 0 \quad P_m^{(2)}(b_2) > 0$

$P_m^{(2)}(c_1) > 0 \quad P_m^{(2)}(c_2) < 0$

(iv) $P_m^{(4)}(b_1) > 0 \quad P_m^{(4)}(b_2) < 0$

$P_m^{(4)}(c_1) < 0 \quad P_m^{(4)}(c_2) > 0$

then $P_m(x)$ is strictly montonic on $\bigcup_{i=1}^{2} (b_i, c_i)$, i.e., $P_m(x) \neq 0 \quad x \in \bigcup_{i=1}^{2} (b_i, c_i)$.

Proposition 3 is presented here without proof since it is primarily meant to be an illustration of the type of monotonicity constraints that could be used for polynomial approximants. It is worth noting that the constraints of Proposition 3 are linear in the derivatives. The number of the constraints is independent of the size of the transition bands and they result in approximants strictly monotonic on the transition regions. Unfortunately the number of constraints grows with the number of "don't care" bands, and even the case of three "don't care" bands turns out to be rather cumbersome.

If it is desired that the monotonicity conditions be oriented toward minimax approximation then it is reasonable to tailor them to describe the largest class of L_∞ approximants monotonic over the "don't care" bands. This is equivalent to requiring that the monotonic approximants interpolate the target function at least $M + 2 - (B \times 2 + 2) + B + 1 = M - B + 1$ points, where B represents the number of "don't care" bands. (This is the reason for the assumption (i) in Proposition 3.) Imposing a number of interpolating points greater than this can greatly simplify the boundary conditions on the derivatives. But this amounts to considering only a subclass of all the possible monotonic approximants. Similarly, it is restrictive to search among the class of monotonic approximants with zero derivatives of fixed order at the boundary points.

Conditions (ii)-(iv) of Proposition 3 are certainly sufficient to characterize monotonic behavior over the "don't care" bands among the class of polynomials satisfying condition (i) of Proposition 3. However they don't describe the entire class of polynomials having monotonic behavior over two "don't care" bands. The conditions describing this class are not yet clear. Thus we view Proposition 3 as an example of the possible conditions but do not suggest that it presents the best, or even the most practical, way of achieving monotonicity over two transition regions.

We note that there could exist nonlinear conditions on the derivatives of the approximant at the boundary point or at other specified points, capable of guaranteeing monotonicity in a more synthetic manner than conditions of the type used in Proposition 3. The search for the most synthetic or convenient way to impose monotonicity over the "don't care" bands for classes of approximants interpolating the objective functions at, at least, M-$B + 1$ points is an open and relatively unexplored research question.

The considerations in this section are meant to indicate what kind of results are presently missing in order to use multiple "don't care" bands in filter design while keeping the same generality that Section 2 illustrates for filter design with a single "don't care" band.

IV. CONCLUSION

This paper has investigated the conditons under which a "don't care" band approach to transition regions leads to monotonic behavior over these regions. Such a technique is analyzed because it seems the computationally most convenient. Section II clearly shows the cases when the use of a single "don't care" band is sufficient to impose monotonicity on the "don't care" band. L_∞ and L_2 filter design formulations belong to these cases. The results are valid both for FIR as well as IIR filters.

Section III shows why the simple "don't care" bands are not adequate to impose monotonicity if the number of transition regions is greater or equal to two. The next best approach seems to be the use of "don't care" bands together with derivative conditions at certain specified points. Secion III gives a set of sufficient conditions for the two transition region problem. However, a general approach and solution to the multiple transition region design problem remains an open research question.

REFERENCES

1. L. Rabiner, J.F. Kaiser and R.W. Schafer, "Some Considerations in the Design of Multiband Finite-Impulse-Response Digital Filters" *Trans. on Acoustics, Speech, and Sig. Proc.*, no. 6, vol. ASSP-22, Dec. 1974.
2. G. Cortelazzo, M.R. Lightner, and W.K. Jenkins, "Frequency Domain Design of Multiband Finite Impulse Response Digital Filters Based Upon the Minimax Criterion, " *Circuits, Systems, Signal Processing*, Vol. 2, No. 3, 1983, pp. 285-309.
3. M.T. McCallig and B.J. Leon, "Constrained Ripple Design of FIR Digital Filters, *IEEE Trans. on Circuits and Systems*. Vol. CAS-25, November 1978, pp. 893-902.
4. F. Grenez, "Design of Linear or Minimum Phase FIR Filters by Constrained Chebychev approximation, *Signal Proc.* No. 5, 1983, pp. 325-332.
5. G.D. Taylor and M.J. Winter, "Calculation of Best Restricted Approximations, *SIAM J. Numer. Anal.*, Vol. 7, No. 2, June 1970, pp. 248-255.
6. F. Brophy and A.C. Salazar, "Synthesis of Spectrum Shaping Digital Filters of Recursive Design," *IEEE Trans. Cir. Syst.*, Vol. CAS-22, March 1975.
7. E.W. Cheney, *Introduction to Approximation Theory*. McGraw-Hill, New York 1966.
8. G. Cortelazzo. *Frequency Domain Design of Multiband Finite Impulse Response Digital Filters Based on the Minmax Criterion*. M.S. Thesis, University of Illinois, September 1980, (Coordinated Science Lab Report R-906, December 1980).

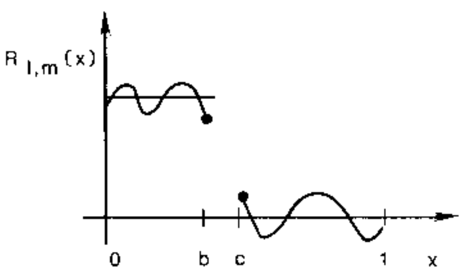

Figure 1. Filter behavior imposed by Proposition 2.

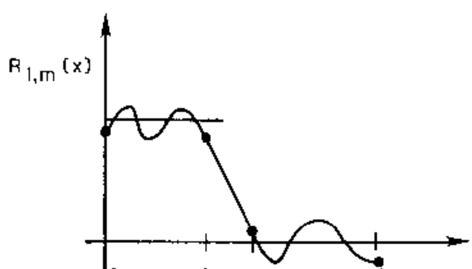

Figure 2. Illustration of contradicting alternation in Proposition 2.

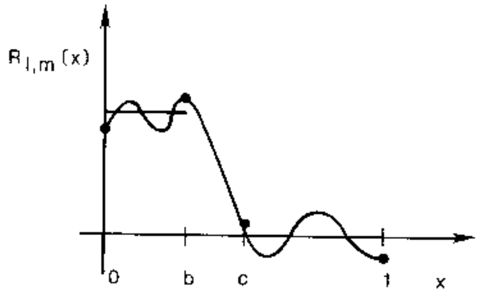

Figure 3. Illustration requiring too many zero's in Proposition 2.

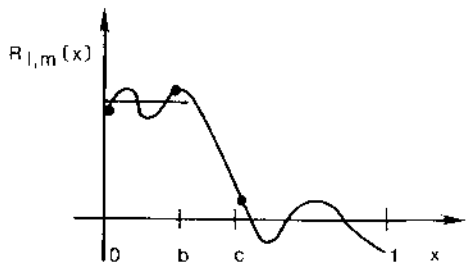

Figure 4. Illustration of Subcase II, Proposition 2.

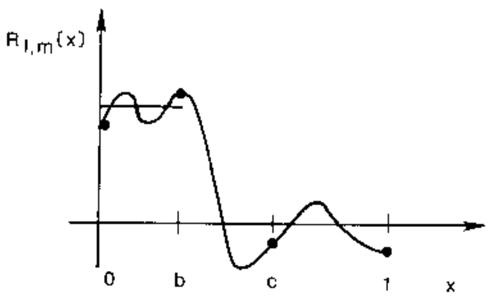

Figure 5. Illustration of Subcase II, Proposition 2.

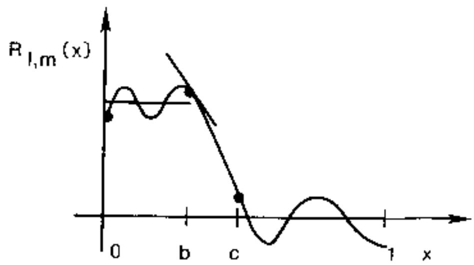

Figure 6. Illustration of Subcase IV, Proposition 2.

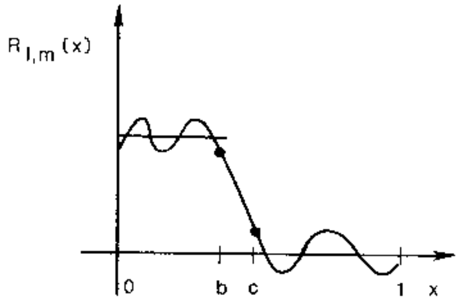

Figure 7. Behavior of rational approximant as characterized by Proposition 2 and the Corollary.