that

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

with $\theta = -\pi$. This means that Λ is of the form

$$\Lambda = \operatorname{diag}\left\{1, \cdots, 1, \begin{bmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{bmatrix}, \cdots, \begin{bmatrix} \cos\theta_s & -\sin\theta_s \\ \sin\theta_s & \cos\theta_s \end{bmatrix}\right\}. \quad (5)$$

Now set

$$T(x) = V'\Lambda(x)V[xH + (1-x)I] \quad (6)$$

with

$$\Lambda(x) = \operatorname{diag}\left\{1, \cdots, 1, \begin{bmatrix} \cos x\theta_1 & -\sin x\theta_1 \\ \sin x\theta_1 & \cos x\theta_1 \end{bmatrix}, \cdots, \begin{bmatrix} \cos x\theta_s & -\sin x\theta_s \\ \sin x\theta_s & \cos x\theta_s \end{bmatrix}\right\}. \quad (7)$$

It is easily checked that $T(0) = I$ and $T(1) = UH = T_1$. Obviously, $T(x)$ is continuously dependent on x, and is nonsingular, being of the form $U(x)H(x)$ for orthogonal $U(x)$ and positive definite $H(x)$. This proves the theorem and, at the same time, we have indicated how the family $T(x)$ may be found.

Remark: It follows easily from the theorem that the set of all minimal state realizations of a transfer function matrix falls into two disjoint subsets, with all members of the one subset being continuously equivalent. It is not possible for three or more realizations to be such that no pair is continuously equivalent.

Remark: Define the matrices W_0 and W_1 by

$$W_i = [B_i \; A_iB_i \; \cdots \; A_i^{n-1}B_i], \qquad i = 0, 1 \quad (8)$$

where A_i is $n \times n$. The W_i have rank n by the minimality of $\{A_i, B_i, C_i\}$. The matrix T_1 is uniquely determined by $T_1W_0 = W_1$ or $T_1W_0W_0' = W_1W_0'$; see [8]. Since W_0W_0' evidently has positive determinant, the condition $\det T_1 > 0$ is equivalent to $\det W_1W_0' > 0$.

Remark: The result extends easily to the time-varying case. Let $\{A_i(t), B_i(t), C_i(t)\}$ be two realizations of the same weighting function, related by $A_1 = T_1(t)A_0T_1^{-1}(t) + \dot{T}_1(t)T_1^{-1}(t)$, $B_1 = T_1(t)B_0$, $C_1' = C_0'T_1^{-1}(t)$; see [8]. It is implicitly assumed that $T_1(t)$ is differentiable and nonsingular for all t. For fixed t, a family $T(x, t)$, $0 \le x \le 1$, exists taking A_0 to A_1 if and only if $\det T_1(t) > 0$. Also, if $\det T_1(t)$ is positive for one particular value of t, it is positive for all t because $T_1(\cdot)$ is continuous and always nonsingular. Accordingly, the family $T(x, t)$ exists for all t.

There is no question that there remains a gap between the statement of this theorem and its use in the design problem of varying the elements of a network so as to preserve the terminal behavior but achieve a more satisfactory internal configuration. The bridging of the gap will undoubtedly require use of the state–space approach to network synthesis; see [10], for example.

References

[1] J. D. Schoeffler, "The synthesis of minimum sensitivity networks," *IEEE Trans. Circuit Theory*, vol. CT-11, pp. 271–276, June 1964.
[2] ——, "Continuously equivalent networks and their application," *IEEE Trans. Commun. Electron.*, vol. 83, pp. 763–767, Nov. 1964.
[3] N. Howitt, "Group theory and the electric circuit," *Phys. Rev.*, vol. 37, pp. 1583–1595, June 15, 1931.
[4] D. A. Calahan, *Modern Network Synthesis*, vol. II. New York: Hayden, 1963.
[5] R. W. Newcomb, "The noncompleteness of continuously equivalent networks," *IEEE Trans. Circuit Theory* (Corresp.), vol. CT-13, pp. 207–208, June 1966.
[6] ——, *Linear Multiport Synthesis*, New York: McGraw-Hill, 1966, pp. 345–347.
[7] B. D. O. Anderson, R. W. Newcomb, R. E. Kalman, and D. C. Youla, "Equivalence of linear time-invariant dynamical systems," *J. Franklin Inst.*, vol. 279, pp. 371–378, May 1966
[8] R. W. Brockett, *Finite Dimensional Linear Systems*. New York: Wiley, 1970.
[9] R. R. Stoll, *Linear Algebra and Matrix Theory*. New York: McGraw-Hill, 1962.
[10] B. D. O. Anderson and R. W. Newcomb, "Impedance synthesis via state-space techniques," *Proc. Inst. Elec. Eng.*, vol. 115, pp. 928–936, July 1968.

Theory of Generalized Least pth Approximation

JOHN W. BANDLER AND C. CHARALAMBOUS

Abstract—A unified discussion of least pth approximation as it relates to optimal computer-aided design of networks and systems is presented. General objective functions are proposed and their properties discussed. The main result is that a wider variety of design problems and a wider range of specifications than appear to have been considered previously from the least pth point of view should now be tractable.

Introduction

It is well known to network and system designers that least pth approximation with a sufficiently large value of p can result in an optimal solution very close to the optimal minimax solution [1]–[4]. Many designers continually express their preference for least pth approximation because of its flexibility [5]. Gradient optimization methods suitable for least pth approximation, such as the Fletcher–Powell method [6], are widely available as computer subroutines and are often easier to use than minimax algorithms. The results obtained are usually almost as good for practical purposes as the minimax solution.

To the authors' knowledge, a generalization of least pth approximation to design with upper and lower response specifications, such as encountered in filter design, does not appear to have received serious attention in the literature. Usually, least pth approximation is applied to the approximation of a single specified function by a network or system response. Minimax approximation using nonlinear programming methods, on the other hand, has been applied to more general problems [7], [8]. See also [2] and [3].

This correspondence presents a unified discussion of least pth approximation. General objective functions are proposed and their properties discussed. The usefulness of least pth approximation is extended to a wider variety of network and system design problems and a wider range of specifications than appear to have been considered previously from the least pth point of view.

The Objective Functions

Definitions

Define real error functions related to the "upper" and "lower" specifications, respectively, as [2]

$$e_u(\phi, \psi) \triangleq w_u(\psi)(F(\phi, \psi) - S_u(\psi))$$
$$e_l(\phi, \psi) \triangleq w_l(\psi)(F(\phi, \psi) - S_l(\psi)) \quad (1)$$

where the symbols are

- $F(\phi, \psi)$ approximating function (actual response);
- $S_u(\psi)$ upper specified function (desired response bound);
- $S_l(\psi)$ lower specified function (desired response bound);
- $w_u(\psi)$ upper positive weighting function;
- $w_l(\psi)$ lower positive weighting function;
- ϕ vector containing the k independent parameters;
- ψ independent variable (e.g., frequency or time).

In filter design problems, for example, $F(\phi, \psi)$ will be the response' ϕ may represent the network parameters, ψ could be the frequency' $S_u(\psi)$ would refer to the passband specification, and $S_l(\psi)$ to the stopband specification. $F(\phi, \psi)$ is often continuous in ϕ and ψ, but $S_l(\psi)$, $S_u(\psi)$, $w_l(\psi)$, and $w_u(\psi)$ are most likely discontinuous in ψ, but with $S_u(\psi) \ge S_l(\psi)$.[1] See, for example, Figs. 1 and 2. Time-domain approxima-

Manuscript received August 3, 1971; revised November 15, 1971. This work was supported by the McMaster University Science and Engineering Division Research Board, and by the National Research Council of Canada under Grant A7239. This paper is based on a paper presented at the 14th Midwest Symposium on Circuit Theory, Denver, Colo., May 6–7, 1971.

The authors are with the Communications Research Laboratory, Department of Electrical Engineering, McMaster University, Hamilton, Ont., Canada.

[1] S_u may be considered as being ∞ and S_l as $-\infty$ in those bands of ψ where they are not explicitly indicated.

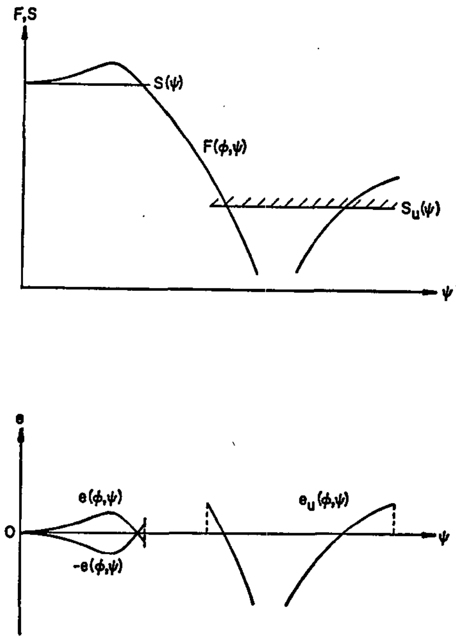

Fig. 1. Example of a design problem for which it is generally impossible for the response to exceed the specification. Case 1 is applicable.

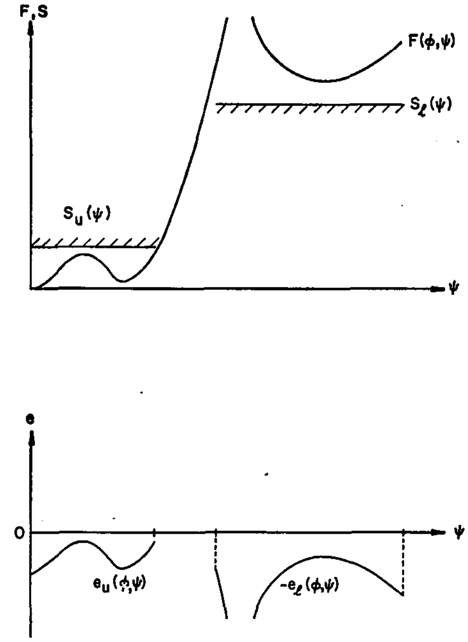

Fig. 2. Example of a design problem in which the response exceeds the specification Case 2 is applicable.

tion problems can also be readily formulated in these terms as is well known.

A special case of (1) occurs when $S_u = S_l = S$ and $w_u = w_l = w$, leading to the more common form

$$e(\phi, \psi) \triangleq w(\psi)(F(\phi, \psi) - S(\psi)). \tag{2}$$

In practice, we will evaluate all the functions at a finite discrete set of values of ψ taken from one or more closed intervals. Therefore, we will let

$$\begin{aligned} e_{ui}(\phi) &\triangleq e_u(\phi, \psi_i), \qquad i \in I_u \\ e_{li}(\phi) &\triangleq e_l(\phi, \psi_i), \qquad i \in I_l \end{aligned} \tag{3}$$

where it is assumed that a sufficient number of sample points have been chosen so that the discrete approximation problem adequately approximates the continuous problem. I_u and I_l are appropriate index sets. For the special case, $I_u = I_l = I$.

Case 1—Specification Violated

In the case when the specification is violated, some of the $e_{ui}(\phi)$ or $-e_{li}(\phi)$ are positive. For the special case mentioned earlier, we will generally always have some $\pm e_i(\phi)$ positive. In an effort to meet the specification we propose the following objective function to be minimized:

$$U(\phi) = \sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_l} [-e_{li}(\phi)]^p \tag{4}$$

where

$$\begin{aligned} J_u &\triangleq \{i \mid e_{ui}(\phi) \geq 0, \qquad i \in I_u\} \\ J_l &\triangleq \{i \mid -e_{li}(\phi) \geq 0, \qquad i \in I_l\} \end{aligned} \tag{5}$$

and $p \geq 1$. If J_u and J_l are empty then $U(\phi)$ is set to zero and optimization is terminated. If the minimum value of $U(\phi)$ is zero then we have just met or exceeded the specification. In general, of course, this will not be possible if the design problem is similar to the example depicted in Fig. 1. Note that the special case is readily accommodated since the objective function reduces to

$$U(\phi) = \sum_{i \in I} | e_i(\phi) |^p. \tag{6}$$

The larger the value of p the more nearly would we expect the maximum error to be emphasized, since

$$\max_i [e_{ui}(\phi), -e_{li}(\phi)] = \lim_{p \to \infty} \left\{ \sum_{i \in J_u} [e_{ui}(\phi)]^p + \sum_{i \in J_l} [-e_{li}(\phi)]^p \right\}^{1/p}. \tag{7}$$

In using gradient methods of minimization we would be concerned if the objective function in (4) had discontinuous derivatives. Note that $U(\phi)$ is continuous if the appropriate $e_{ui}(\phi)$ and $e_{li}(\phi)$ are continuous. Differentiating the objective function we have

$$\nabla U(\phi) = \sum_{i \in J_u} p[e_{ui}(\phi)]^{p-1} \nabla e_{ui}(\phi) - \sum_{i \in J_l} p[-e_{li}(\phi)]^{p-1} \nabla e_{li}(\phi) \tag{8}$$

where

$$\nabla \triangleq \begin{bmatrix} \frac{\partial}{\partial \phi_1} \\ \frac{\partial}{\partial \phi_2} \\ \vdots \\ \frac{\partial}{\partial \phi_k} \end{bmatrix}.$$

For $p > 1$, and with $e_{ui}(\phi)$ and $e_{li}(\phi)$ continuous with continuous derivatives for $e_{ui}(\phi) \geq 0$ and $-e_{li}(\phi) \geq 0$, $\nabla U(\phi)$ will be continuous, becoming **0** at the minimum.

The matrix of second partial derivatives is given by

$$\begin{aligned} H = p(p-1) &\left[\sum_{i \in J_u} e_{ui}^{p-2} \nabla e_{ui} (\nabla e_{ui})^T + \sum_{i \in J_l} (-e_{li})^{p-2} \nabla e_{li} (\nabla e_{li})^T \right] \\ &+ p \left[\sum_{i \in J_u} e_{ui}^{p-1} \nabla (\nabla e_{ui})^T - \sum_{i \in J_l} (-e_{li})^{p-1} \nabla (\nabla e_{li})^T \right]. \end{aligned} \tag{9}$$

It is easily shown that $\nabla e_{ui}(\nabla e_{ui})^T$ and $\nabla e_{li}(\nabla e_{li})^T$ are positive-semidefinite. If $e_{ui}(\phi)$ and $-e_{li}(\phi)$ are convex, then $\nabla(\nabla e_{ui})^T$ and $-\nabla(\nabla e_{li})^T$ are also positive-semidefinite. For large enough values of p, however, the first two terms are likely to be much greater than the last two, so H is usually likely to be positive-semidefinite anyway.

Case 2—Specification Satisfied

For the case when the specification is satisfied all the $e_{ui}(\phi)$ and $-e_{li}(\phi)$ will be negative. It is usually impossible to make $\pm e_i(\phi)$ negative, however, so this case need not be considered here. This time, in an effort to exceed the specification by as much as possible, we propose the following objective function to be minimized:

$$U(\phi) = \sum_{i\in I_u} [-e_{ui}(\phi)]^{-p} + \sum_{i\in I_l} [e_{li}(\phi)]^{-p} \quad (10)$$

where we assume

$$\begin{aligned} -e_{ui}(\phi) > 0, &\quad i \in I_u \\ e_{li}(\phi) > 0, &\quad i \in I_l \end{aligned} \quad (11)$$

and $p \geq 1$.

The larger the value of p the more nearly would we expect the minimum "error" to be emphasized, since

$$\min_i [-e_{ui}(\phi), e_{li}(\phi)] = \lim_{p\to\infty} \left\{ \sum_{i\in I_u} [-e_{ui}(\phi)]^{-p} + \sum_{i\in I_l} [e_{li}(\phi)]^{-p} \right\}^{-1/p} \quad (12)$$

so that minimizing the objective function in (10) will tend to maximize the minimum amount by which the specification is exceeded.

Differentiating (10) we have

$$\nabla U(\phi) = \sum_{i\in I_u} p[-e_{ui}(\phi)]^{-p-1} \nabla e_{ui}(\phi) - \sum_{i\in I_l} p[e_{li}(\phi)]^{-p-1} \nabla e_{li}(\phi). \quad (13)$$

In this case, with $1/e_{ui}(\phi)$ and $1/e_{li}(\phi)$ continuous with continuous derivatives for $-e_{ui}(\phi)>0$ and $e_{li}(\phi)>0$, $\nabla U(\phi)$ will be continuous, becoming **0** at the minimum.

The matrix of second partial derivatives is given by

$$H = p(p+1)\left[\sum_{i\in I_u} (-e_{ui})^{-p-2} \nabla e_{ui} (\nabla e_{ui})^T + \sum_{i\in I_l} e_{li}^{-p-2} \nabla e_{li} (\nabla e_{li})^T\right] + p\left[\sum_{i\in I_u} (-e_{ui})^{-p-1} \nabla (\nabla e_{ui})^T - \sum_{i\in I_l} e_{li}^{-p-1} \nabla (\nabla e_{li})^T\right]. \quad (14)$$

H will be positive-semidefinitive under conditions rather similar to those for case 1.

Discussion

Much of the foregoing analysis is intuitively obvious. Fig. 3 shows sketches which can be used as an aid to understanding the procedure.

Probably the most useful parallel to the objective function of case 1 is the simple penalty function approach for dealing with nonfeasible points in constrained optimization [2]. A suitable penalty term including only the violated constraints is minimized, commonly with $p=2$. If the penalty term is zero, a feasible solution is indicated. If the minimum is nonzero, the constraints remain violated. In the case of generalized least pth approximation such a situation may indicate the impossibility of satisfying the specification.

If a feasible solution in constrained optimization is available a penalty term formed by all the constraint functions may be defined so that an optimal solution close to the boundary of the feasible region is discouraged. Indeed, by minimizing this penalty function an attempt to move as far as possible from the boundary is made. Thus, a parallel to case 2 is the penalty function approach developed by Fiacco and McCormick [9], [10]. This is seen by letting p in (10) be 1. Unlike the Fiacco–McCormick technique, however, our objective function is in the form of a penalty term, so our aim is simply to move away from the boundary. Similar precautions to avoid nonfeasible solutions may have to be taken [see Fig. 3(d)].

It should be remarked that the role of the weighting functions is the usual one. In case 1, deviations from the specification are emphasized by relatively large weighting numbers, and a greater effort will be devoted to forcing the corresponding response closer to the specification than the rest of the response. In case 2, relatively large weighting numbers have the effect of allowing the corresponding response to remain much closer to the specification while improving other parts.

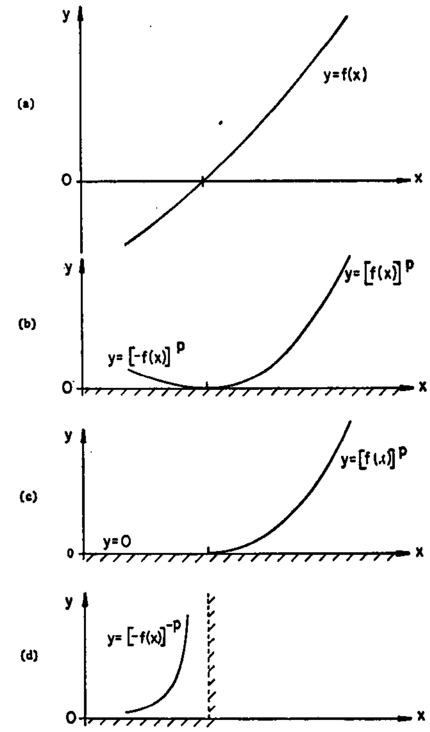

Fig. 3. Sketches to illustrate the behavior of components of possible generalized least pth objectives. $f(x)$ is convex continuous with continuous derivatives. $p>1$ in (b) and (c). $p\geq 1$ in (d).

Conclusions

Implementation of the generalized least pth objectives proposed in this correspondence should be very straightforward. The usually difficult problem of choosing suitable weighting functions to force more nearly uniform approximation is alleviated by using an appropriate value of p. Difficulties such as are encountered in attempting to use the conventional least pth objective function to force responses above or below desired levels are virtually eliminated. Poles, for example, in the stopband of filters if they are deemed desirable, which they usually are, are readily accommodated. Numerical experiments employing the objective functions in the optimal design of networks and systems are currently under way.

References

[1] G. C. Temes and D. Y. F. Zai, "Least pth approximation," *IEEE Trans. Circuit Theory* (Corresp.), vol. CT-16, pp. 235–237, May 1969.
[2] J. W. Bandler, "Optimization methods for computer-aided design," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-17, pp. 533–552, Aug. 1969.
[3] G. C. Temes, "Optimization methods in circuit design," in *Computer Oriented Circuit Design*, F. F. Kuo and W. G. Magnuson, Jr., Eds. Englewood Cliffs, N. J.: Prentice-Hall, 1969.
[4] R. Seviora, M. Sablatash, and J. W. Bandler, "Least pth and minimax objectives for automated network design," *Electron. Lett.*, vol. 6, pp. 14–15, Jan. 1970.
[5] S. W. Director, "Survey of circuit-oriented optimization techniques," *IEEE Trans. Circuit Theory*, vol. CT-18, pp. 3–10, Jan. 1971.
[6] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, vol. 6, pp. 163–168, June 1963.
[7] L. S. Lasdon and A. D. Waren, "Optimal design of filters with bounded, lossy elements," *IEEE Trans. Circuit Theory*, vol. CT-13, pp. 175–187, June 1966.
[8] A. D. Waren, L. S. Lasdon, and D. F. Suchman, "Optimization in engineering design," *Proc. IEEE*, vol. 55, pp. 1885–1897, Nov. 1967.
[9] A. V. Fiacco and G. P. McCormick, "The sequential unconstrained minimization technique for nonlinear programming, a primal-dual method," *Management Sci.*, vol. 10, pp. 360–366, Jan. 1964.
[10] ——, "Computational algorithm for the sequential unconstrained minimization technique for nonlinear programming," *Management Sci.*, vol. 10, pp. 601–617, July 1964.