PROCEEDINGS
OF THE
FOURTEENTH

MIDWEST SYMPOSIUM ON CIRCUIT THEORY

MAY 6-7, 1971

UNIVERSITY OF DENVER

DENVER, COLORADO

A NEW GRADIENT ALGORITHM FOR MINIMAX OPTIMIZATION OF NETWORKS AND SYSTEMS

John W. Bandler and T.V. Srinivasan

Department of Electrical Engineering
McMaster University
Hamilton, Ontario, Canada

Abstract

A new algorithm for nonlinear minimax approximation has been developed. It is suitable for minimax optimization of network and system responses. A linear programming problem using gradient information of one or more highest ripples in the error function to produce a downhill direction followed by a linear search to find a minimum in that direction is central to the algorithm. It is compared numerically with another algorithm based on the method of Osborne and Watson on the optimization of commensurate and noncommensurate transmission-line matching networks, for which the optima are known.

INTRODUCTION

The minimax algorithm due to Osborne and Watson [1] deals with minimax formulations by following two steps - a linear programming part which provides a given step in the parameter space, followed by a linear search along the direction of the step. This algorithm is very similar to the one proposed by Ishizaki and Watanabe [2] and works very well if the objective function is not highly non-linear in the vicinity of the optimum. In cases when the linear approximation is not very good in the vicinity of the optimum, this method may fail to converge towards the optimum for successive iterations.

The razor search algorithm due to Bandler and Macdonald [3] is based on pattern search, where a few random moves are used in an effort to negotiate certain kinds of razor-sharp valleys in multi-dimensional space. This method is good if the gradient information is not available. A more recent algorithm due to Bandler and Lee-Chan [4] exploits the gradient information of the extrema of the error function to get a downhill direction by solving a set of simultaneous equations. The method works well except that in the case of

---

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.

linear dependence of the equations, some problems may arise in the convergence towards the optimum. Another method proposed by Heller [5] uses a quadratic programming approach to solve the minimax problem, but consumes a considerable amount of computer time.

A new algorithm has been developed in which gradient information of one or more of the highest ripples in the error function is used to produce a downhill direction by solving a suitable linear programming problem. A linear search follows to find the minimum in that direction, and the procedure is repeated. The algorithm is compared numerically with another based on the Osborne and Watson algorithm on the optimization of commensurate and non-commensurate transmission-line matching networks, for which the optima are known.

ALGORITHM 1

The first algorithm that was programmed is based on the Osborne and Watson algorithm. Linearizing the real and differentiable error functions $e_i(\underset{\sim}{\phi})$, $i=1,2,\ldots,n>k$, at some feasible point $\underset{\sim}{\phi}^j$, where $\underset{\sim}{\phi}$ denotes the k independent parameters, we have the following constrained minimax approximation problem:

$$\text{minimize} \quad U = x_{k+1} \tag{1}$$

subject to

$$\pm \left\{ e_i(\underset{\sim}{\phi}^j) + \underset{\sim}{\nabla} e_i^T(\underset{\sim}{\phi}^j) \begin{bmatrix} \phi_1^j x_1 - \phi_1^j + \phi_{\ell 1} \\ \phi_2^j x_2 - \phi_2^j + \phi_{\ell 2} \\ \vdots \\ \phi_k^j x_k - \phi_k^j + \phi_{\ell k} \end{bmatrix} \right\} \leq x_{k+1} \qquad i=1,2,\ldots,n>k \tag{2}$$

$$\underset{\sim}{x} \geq \underset{\sim}{0} \tag{3}$$

$$x_i \leq \frac{\phi_{ui} - \phi_{\ell i}}{\phi_i^j} \qquad i=1,2,\ldots,k \tag{4}$$

where

$$\underset{\sim}{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{k+1} \end{bmatrix}$$

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

and where

$$x_i \triangleq 1 + \frac{\Delta\phi_i^j - \phi_{\ell i}}{\phi_i^j} \qquad i=1,2,\ldots,k \tag{5}$$

The solution to this linear program produces an incremental change in parameter values given by $\Delta\underset{\sim}{\phi}^j$. Observe that the following upper and lower bounds on the parameters have been incorporated:

$$0 \leq \phi_{\ell i} \leq \phi_i \leq \phi_{ui} \qquad i=1,2,\ldots,k \tag{6}$$

Next we use a method based on golden section search suggested by Temes [6] to find γ^j corresponding to the constrained minimum value of $\max_i |e_i(\underset{\sim}{\phi}^j + \gamma^j \Delta\underset{\sim}{\phi}^j)|$ for $i=1,2,\ldots,n$. We now set $\underset{\sim}{\phi}^{j+1} = \underset{\sim}{\phi}^j + \gamma^j \Delta\underset{\sim}{\phi}^j$ and repeat the process of linearizing the error functions and solving a linear programming problem.

Under certain conditions this iterative process will converge to the nonlinear minimax optimum. For conditions of convergence of the original Osborne and Watson process see reference [1]. An important point is that the Haar condition should be satisfied by $[\underset{\sim}{\nabla} e_1 \;\; \underset{\sim}{\nabla} e_2 \; \ldots \; \underset{\sim}{\nabla} e_n]$, that is, every kxk submatrix is nonsingular.

ALGORITHM 2

The second algorithm is a generalization of the gradient razor search method [4], and is basically of the steepest descent type. In this case, suppose we have the problem:

$$\text{minimize } U = \max_{i \in I} f_i(\underset{\sim}{\phi}) \tag{7}$$

where the $f_i(\underset{\sim}{\phi})$ are real nonlinear differentiable functions generally. Linearizing $f_i(\underset{\sim}{\phi})$ and letting

$$J \triangleq \{i \mid f_i(\underset{\sim}{\phi}) = \max_i f_i(\underset{\sim}{\phi}), \;\; i \in I\} \tag{8}$$

we can write at some feasible point $\underset{\sim}{\phi}^j$

$$\Delta f_i(\underset{\sim}{\phi}^j) = \underset{\sim}{\nabla} f_i^T(\underset{\sim}{\phi}^j)\ \Delta\underset{\sim}{\phi}^j \qquad i \in J \tag{9}$$

In order for $\Delta\underset{\sim}{\phi}^j$ to define a descent direction for $\max_{i \in I} f_i(\underset{\sim}{\phi})$ we must have

$$\underset{\sim}{\nabla} f_i^T(\underset{\sim}{\phi}^j)\ \Delta\underset{\sim}{\phi}^j < 0 \qquad i \in J \tag{10}$$

Consider

$$\Delta\underset{\sim}{\phi}^j = -\sum_{i \in J} \alpha_i^j\ \underset{\sim}{\nabla} f_i(\underset{\sim}{\phi}^j) \tag{11}$$

$$\sum_{i \in J} \alpha_i^j = 1 \tag{12}$$

$$\alpha_i^j \geq 0 \qquad i \in J \tag{13}$$

which suggests the linear program:

$$\text{maximize } \alpha_{r+1}^j \geq 0 \tag{14}$$

subject to

$$-\underset{\sim}{\nabla} f_i^T(\underset{\sim}{\phi}^j) \sum_{i \in J} \alpha_i^j\ \underset{\sim}{\nabla} f_i(\underset{\sim}{\phi}^j) \leq -\alpha_{r+1}^j \qquad i \in J \tag{15}$$

plus (12) and (13).

We should have $\Delta\underset{\sim}{\phi}^j = \underset{\sim}{0}$ if $\underset{\sim}{\phi}^j$ is optimal since the necessary conditions for a minimax optimum are then satisfied. See Bandler [7]. Observe that J is nonempty, and that if J has only one element we obtain the steepest descent direction for the corresponding maximum of the $f_i(\underset{\sim}{\phi})$.

We may also impose the following parameter constraints

$$\begin{bmatrix} \phi_{\ell 1} \\ \phi_{\ell 2} \\ \vdots \\ \phi_{\ell k} \end{bmatrix} \leq \underset{\sim}{\phi}^j - \sum_{i \in J} \alpha_i^j\ \underset{\sim}{\nabla} f_i(\underset{\sim}{\phi}^j) \leq \begin{bmatrix} \phi_{u1} \\ \phi_{u2} \\ \vdots \\ \phi_{uk} \end{bmatrix} \tag{16}$$

however, a feasible downhill direction might not be obtained.

The solution to the linear program provides $\Delta\underset{\sim}{\phi}^j$. The same golden section search method mentioned earlier is then used to find γ^j corresponding to the constrained minimum value of $\max_{i \in I} f_i(\underset{\sim}{\phi}^j + \gamma^j \Delta\underset{\sim}{\phi}^j)$. $\underset{\sim}{\phi}^{j+1}$ is set to

$\underset{\sim}{\phi}^j+\gamma^j\Delta\underset{\sim}{\phi}^j$ and the process is repeated.

To summarize this algorithm as it is used in practice: we try to generate a descent direction based on the gradient vectors of those functions (selected as follows) whose current values fall within a specified tolerance, proceed to the minimum of $\max_{i\epsilon I} f_i(\underset{\sim}{\phi})$ in that direction and repeat the process. It is important to note that since this algorithm is intended for discrete minimax approximation it is assumed that there is a sufficient number of functions $f_i(\underset{\sim}{\phi})$, say n, and, furthermore, only $f_j(\underset{\sim}{\phi})$ such that

$$\begin{aligned} f_j(\underset{\sim}{\phi}) &> f_{j+1}(\underset{\sim}{\phi}) \qquad j=1,2,\ldots,n-1 \\ f_j(\underset{\sim}{\phi}) &\geq f_{j-1}(\underset{\sim}{\phi}) \qquad j=2,3,\ldots,n \end{aligned} \tag{17}$$

are chosen as possible candidates, where $f_i(\underset{\sim}{\phi})$, $i=1,2,\ldots,n$ indicates, for example, sequential sampling of a continuous function $f(\underset{\sim}{\phi},\psi)$ having, in general, a number of local maxima for increasing values of ψ on a closed interval*.

If the linear program using these functions chosen does not yield a direction of decreasing $\max_{i\epsilon I} f_i(\underset{\sim}{\phi})$, the procedure is repeated after including the function corresponding to the next largest of the possible candidates if it exists. When all possible candidates have been included and $\max_{i\epsilon I} f_i(\underset{\sim}{\phi})$ can still not be reduced, we repeat the procedure with r functions corresponding to the first r largest of the candidates, beginning with r=1, in another series of attempts to reduce $\max_{i\epsilon I} f_i(\underset{\sim}{\phi})$. Only when there are no more suitable functions left does the algorithm terminate. If, at any time, $\max_{i\epsilon I} f_i(\underset{\sim}{\phi})$ is reduced we resume with the original procedure.

EXAMPLES

Both algorithms have been compared numerically on a CDC 6400 computer on the problem of minimizing $\max|\rho|$, where ρ is the reflection coefficient, on eleven frequencies ω_i in the band 0.5 to 1.5 GHz for the network shown in Figure 1. This network has already received attention from the optimization point of view [3], [8].

---

*We can easily extend this algorithm, however, to deal with problems involving discontinuous upper and lower specifications such as are encountered in filter design.

For Algorithm 1 we took $e_i(\underset{\sim}{\phi}) = \frac{1}{2}|\rho(\underset{\sim}{\phi}, j\omega_i)|^2$ and for Algorithm 2 we took $f_i(\underset{\sim}{\phi}) = \frac{1}{2}|\rho(\underset{\sim}{\phi}, j\omega_i)|^2$. Appropriate gradient vectors were evaluated using the adjoint network method [9]. In the 2-section examples the 11 frequencies were uniformly spaced. In the 3-section examples the frequencies were 0.5, 0.6, 0.7, 0.77, 0.9, 1.0, 1.1, 1.23, 1.30, 1.40, 1.50 GHz.

The progress of both algorithms from identical starting points with respect to number of function evaluations (one function evaluation corresponds to 11 evaluations of ρ) is recorded in Figures 2 and 3.

Unless otherwise noted, the linear searches for a minimum terminated when the interval of uncertainty fell below 10^{-7}. The points shown mark the end of a linear search or the beginning of a linear programming problem. Constraints on the parameters as indicated in the figures were imposed.

DISCUSSION

The examples tested are considered to be good ones for observing the behaviour of the two algorithms since, depending on which parameters are chosen as variable, linear dependence of the gradient vectors of $\frac{1}{2}|\rho|^2$ at different frequencies can occur and the Haar condition may also not be satisfied. The convergence of Algorithm 1 may not be guaranteed.

Space precludes an extensive discussion of these problems. We will simply note that Algorithm 1 progressed very slowly in the examples of Figures 2(a) and 2(d), and very rapidly in the examples of Figures 2(c) and 3(a). Algorithm 2 in all cases ultimately produced very good results and progressed very rapidly in the examples of Figures 2(a) and 2(b). Appropriate contours for the 2-section examples may be found in reference [8].

CONCLUSIONS

Our results indicate that Algorithm 2 is generally more reliable in reaching an optimal minimax solution than Algorithm 1. Typically one or two minutes are sufficient to optimize a six-parameter design, depending on how far from the optimum one starts and how close one wishes to get. Both algorithms are currently being tested on filter problems.

It is felt that suitably integrating the two algorithms into one package should result in a program which combines the efficiency of the first one on certain problems with the reliability of the second on others.

REFERENCES

[1] M.R. Osborne and G.A. Watson, "An algorithm for minimax approximation in the non-linear case", Computer J., vol. 12, pp. 63-68, February 1969.

[2] Y. Ishizaki and H. Watanabe, "An iterative Chebyshev approximation method for network design", IEEE Trans. Circuit Theory, vol. CT-15, pp. 326-336, December 1968.

[3] J.W. Bandler and P.A. Macdonald, "Optimization of microwave networks by razor search", IEEE Trans. Microwave Theory and Techniques, vol. MTT-17, pp. 552-562, August 1969.

[4] J.W. Bandler and A.G. Lee-Chan, "Gradient razor search method for optimization", International Microwave Symposium, Washington, D.C., May 1971.

[5] J.E. Heller, "A gradient algorithm for minimax design", Report R-406, Coordinated Science Laboratory, University of Illinois, Urbana, Illinois, January 1969.

[6] 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.

[7] J.W. Bandler, "Conditions for a minimax optimum", IEEE Trans. Circuit Theory, vol. CT-18, July 1971.

[8] J.W. Bandler and P.A. Macdonald, "Cascaded noncommensurate transmission-line networks as optimization problems", IEEE Trans. Circuit Theory, vol. CT-16, pp. 391-394, August 1969.

[9] J.W. Bandler and R.E. Seviora, "Current trends in network optimization", IEEE Trans. Microwave Theory and Techniques, vol. MTT-18, pp. 1159-1170, December 1970.

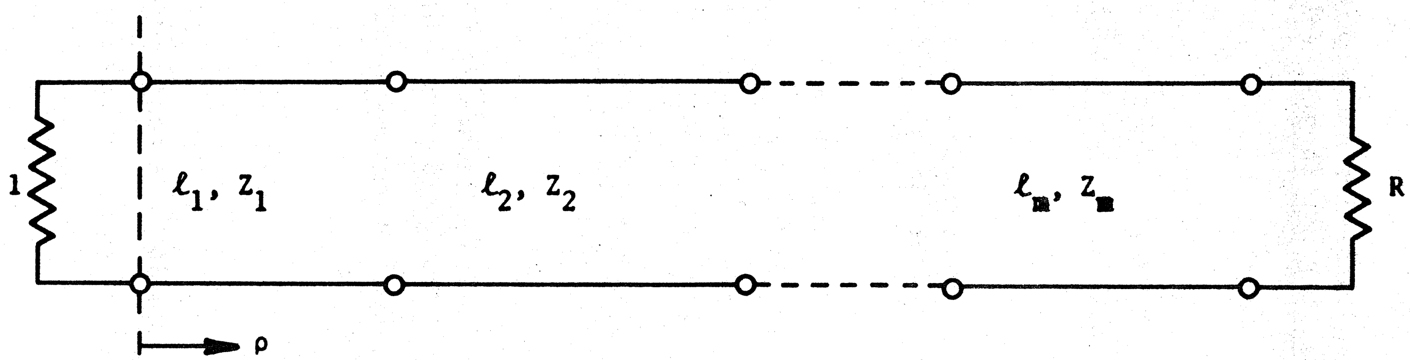


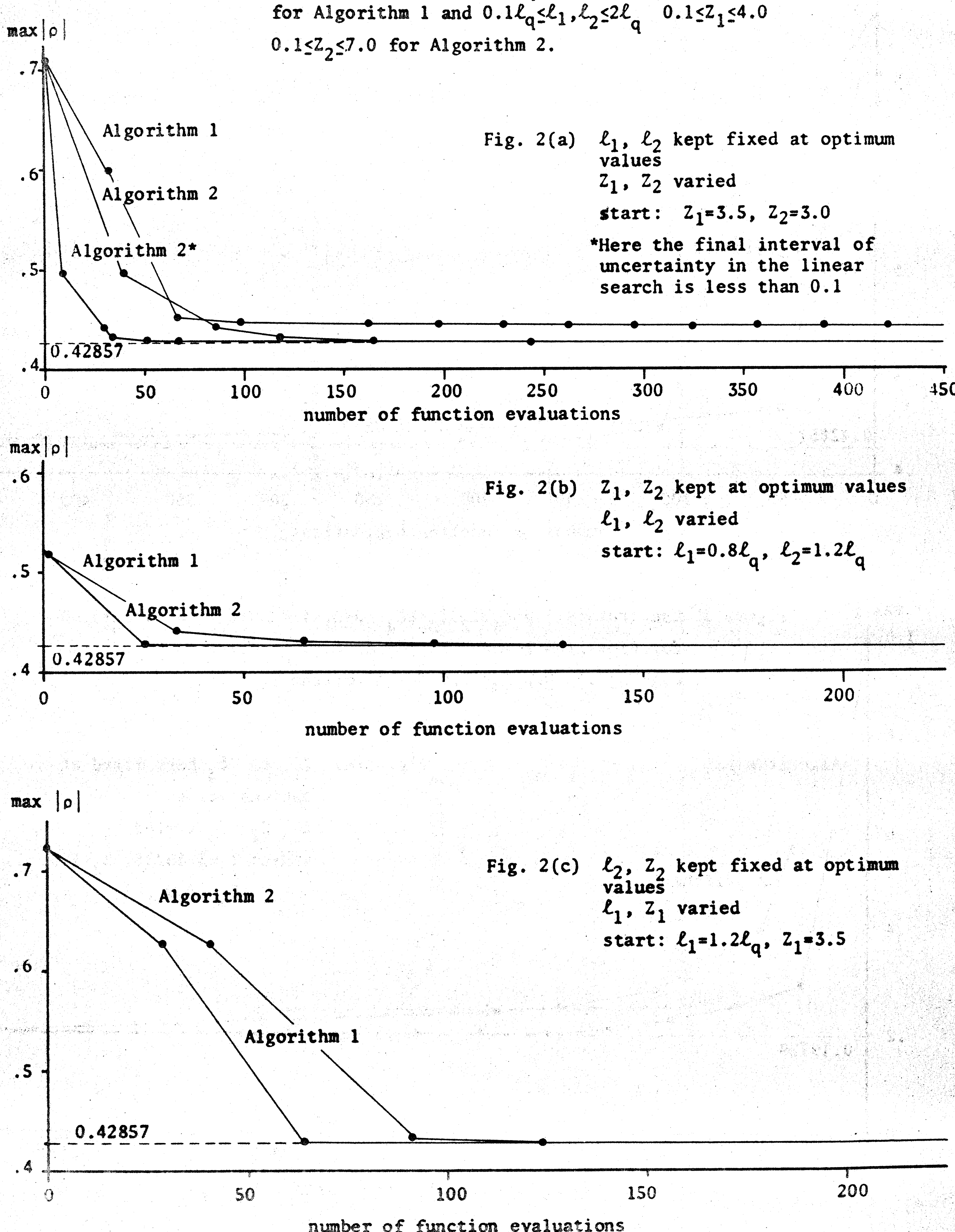
Figure 1 m-section resistively terminated cascade of transmission lines. Optimum matching over 100% band centered at 1GHz for R=10 occurs for the following parameter values.

2-section: $\ell_1=\ell_2=\ell_q$, $Z_1=2.23605$, $Z_2=4.4721$

3-section: $\ell_1=\ell_2=\ell_3=\ell_q$, $Z_1=1.63471$, $Z_2=3.16228$, $Z_3=6.11729$.

$\ell_q=7.49481$ cm is the quarter-wavelength at center frequency.

Figure 2 Constraints: $0 \le \ell_1, \ell_2 \le 2\ell_q$ $0 \le Z_1 \le 4.0$ $0 \le Z_2 \le 7.0$ for Algorithm 1 and $0.1\ell_q \le \ell_1, \ell_2 \le 2\ell_q$ $0.1 \le Z_1 \le 4.0$ $0.1 \le Z_2 \le 7.0$ for Algorithm 2.



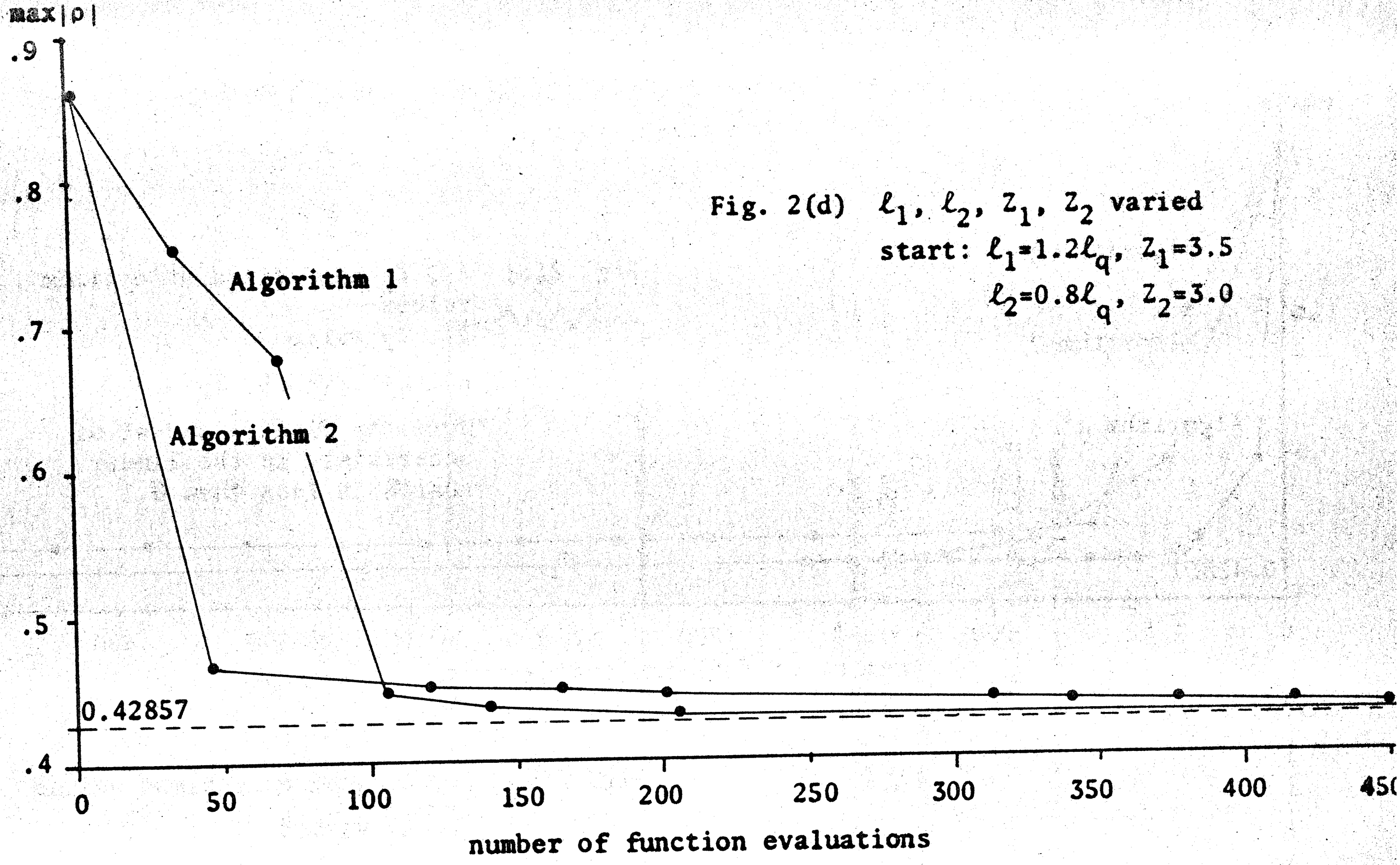


Fig. 2(d) ℓ_1, ℓ_2, Z_1, Z_2 varied
start: $\ell_1=1.2\ell_q$, $Z_1=3.5$
$\ell_2=0.8\ell_q$, $Z_2=3.0$

Figure 3 Constraints: $0\le\ell_1,\ell_2,\ell_3\le 2\ell_q$ $0\le Z_1\le 4.0$ $0\le Z_2\le 7.0$ $0\le Z_3\le 11.0$ for Algorithm 1 and $0.1\ell_q\le\ell_1,\ell_2,\ell_3\le 2\ell_q$ $0.1\le Z_1\le 4.0$ $0.1\le Z_2\le 7.0$ $0.1\le Z_3\le 11.0$ for Algorithm 2.

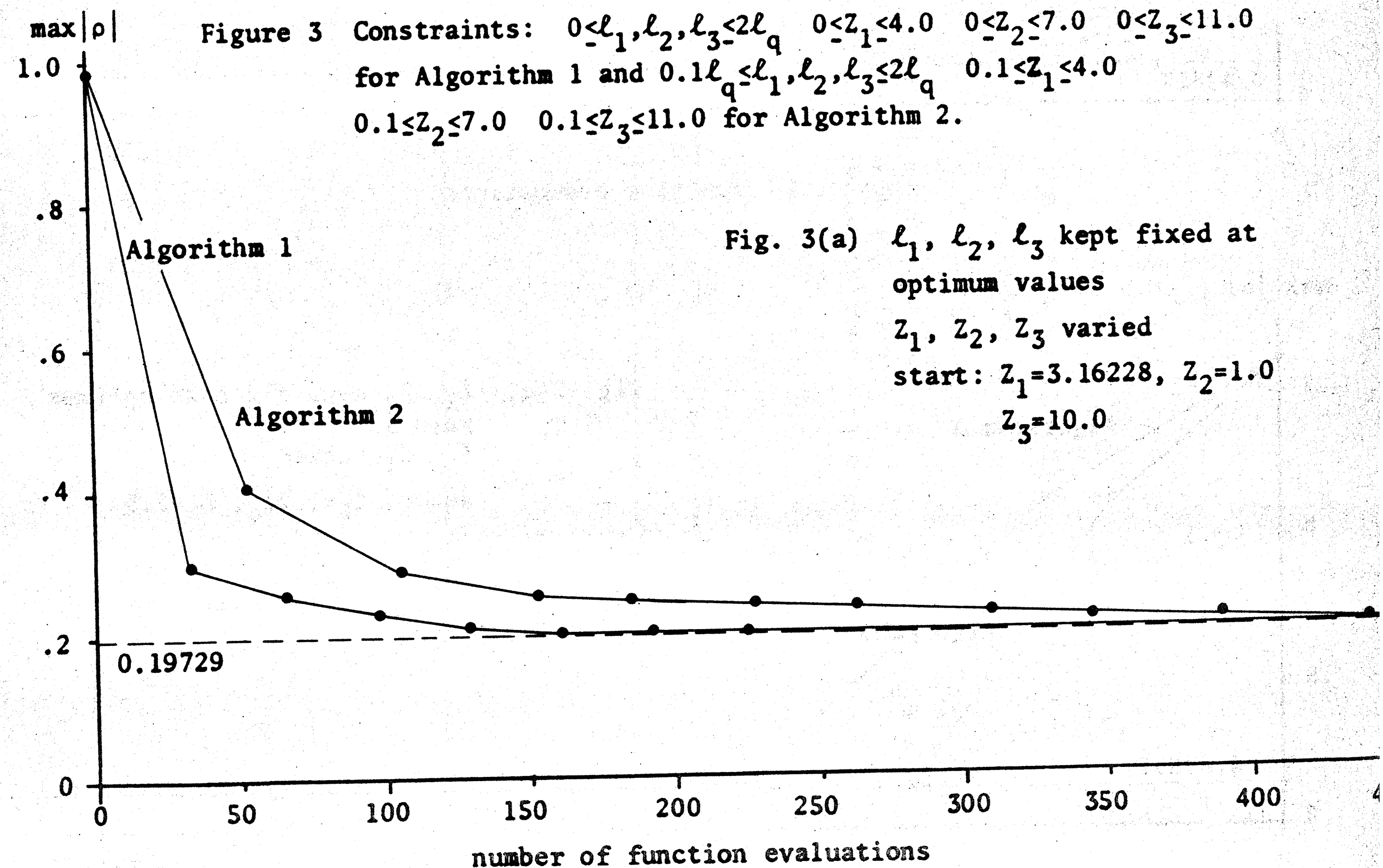


Fig. 3(a) ℓ_1, ℓ_2, ℓ_3 kept fixed at optimum values
Z_1, Z_2, Z_3 varied
start: $Z_1=3.16228$, $Z_2=1.0$
$Z_3=10.0$

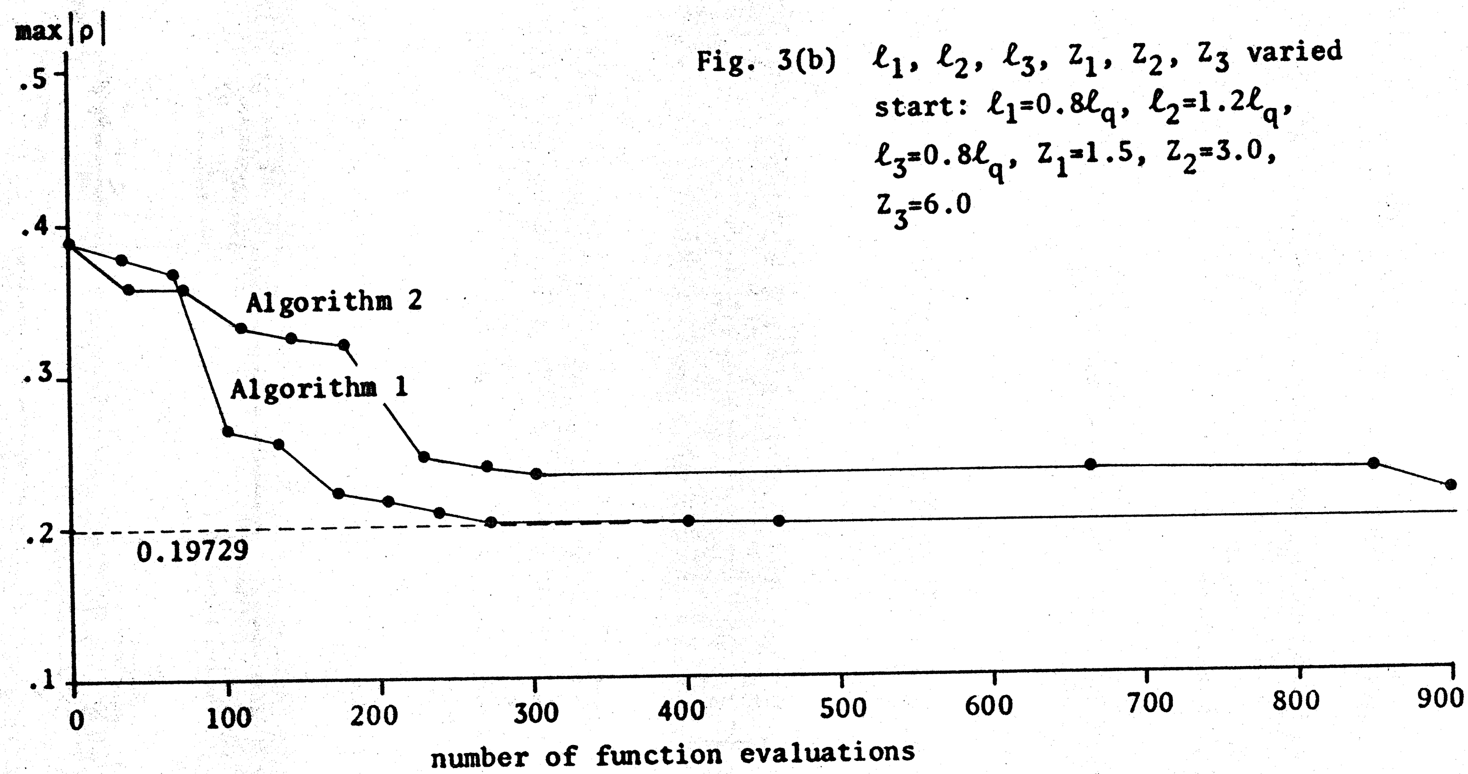


Fig. 3(b) ℓ_1, ℓ_2, ℓ_3, Z_1, Z_2, Z_3 varied
start: $\ell_1=0.8\ell_q$, $\ell_2=1.2\ell_q$, $\ell_3=0.8\ell_q$, $Z_1=1.5$, $Z_2=3.0$, $Z_3=6.0$