Short Papers

An Efficient Algorithm for Finding Zeros of a Real Function of Two Variables

MICHAŁ MROZOWSKI

Abstract—**A new fast algorithm is described for finding the roots of a real function of two variables. The procedure searches the interval in which a function changes sign and then automatically locks to a curve $f(\lambda, p) = 0$, following it inside a given rectangular region. The method ensures that at each step a new pair of points with different function signs is generated, and in effect it minimizes the number of function evaluations. Complementary algorithms opening up opportunities for the further automation of the search process are also presented in outline. An example of the application of the new procedure is included. The proposed algorithm is particularly suitable for solving the electromagnetic problems leading to transcendent equations.**

I. Introduction

Among the most frequently occurring problems in electromagnetic field theory are the so-called eigenvalue problems [1], [2] (both general or nonstandard and ordinary ones) in which the solution exists for the infinite number of a certain parameter λ. The physical meaning of this parameter depends on the problem and is usually chosen to be the propagation constant or the resonance frequency, in which case it may be called the eigenvalue. However λ may also be any other physical parameter of the structure under investigation and then the term "eigenvalue" may be inadequate [3], [4]. The parameter λ can be determined by solving an intricate nonlinear characteristic equation. In most cases one is interested in the behavior of λ as a function of a second parameter, say p, and therefore the characteristic equation is solved for subsequent discrete values of p. As a result one gets a number of points scattered on the $\lambda - p$ plane which have to be linked so as to form $\lambda(p)$ curves. Such a process may sometimes be difficult, especially if the curves approach one another and it is hard to separate them. Accordingly it may be necessary to carry out the computations several times in order to obtain a $\lambda - p$ diagram. Another drawback is that all existing procedures for nonlinear equation solving search for an interval in which the function changes sign. However the sign change may also be caused by a pole of the characteristic equation or a spurious solution (i.e., a solution having no physical meaning [5]). Consequently one has to verify each computed point, which is time consuming, especially if the characteristic equation is obtained by means of a variational method.

In this paper a new computer algorithm tracing the run of the zeros of a function of two variables inside a given rectangular region is proposed. The algorithm minimizes the number of function evaluations needed to compute the locations of roots. Also, it enables one to eliminate the curves caused by the poles of the function, as well as spurious solutions at the early stage of computations, and offers a restart facility. It is also worth mentioning that with the algorithm described herein one may obtain $\lambda - p$ diagrams, even complicated ones, in a single program run.

Manuscript received March 18, 1987; revised October 16, 1987. This work was supported in part by the Polish Academy of Sciences under Contract CPBP 02-02-VII-32 with IPPT PAN.

The author is with the Telecommunication Institute, Technical University of Gdańsk, Majakowski 11/12 Street, 80-952 Gdańsk, Poland.

IEEE Log Number 8718858.

II. The Algorithm

We consider a nonlinear equation:

$$f(\lambda, p) = 0 \tag{1}$$

subject to the constraints

$$\lambda_{\min} < \lambda < \lambda_{\max} \quad \text{and} \quad p_{\min} < p < p_{\max} \tag{2}$$

where f is a continuous function of the two variables λ and p and $\lambda_{\min}, \lambda_{\max}, p_{\min}, p_{\max}$ are constants. Note that $z = f(\lambda, p)$ is a surface in three-dimensional space (λ, p, z). Hence, (1) describes the curve resulting from the intersection of this surface with the plane $z = 0$. Let us consider a rectangular cell with vertices located inside the examined region. Assuming that at none of the nodes $f(\lambda, p) = 0$, we have exactly 16 combinations of the function signs at the vertices, and these possible situations can be divided into the four following groups:

A $(+\,+\,+\,+), (-\,-\,-\,-)$

B $(-\,+\,+\,-), (+\,+\,-\,-), (+\,-\,-\,+), (-\,-\,+\,+)$

C $(-\,+\,+\,+), (+\,+\,+\,-), (+\,+\,-\,+), (+\,-\,+\,+),$
$(+\,-\,-\,-), (-\,-\,-\,+), (-\,-\,+\,-), (-\,+\,-\,-)$

D $(-\,+\,-\,+), (+\,-\,+\,-)$.

If the distance between vertices is small, we may assume that the curve $f(\lambda, p) = 0$ does not pass through the rectangles belonging to the first group. For the remaining groups, the function changes sign and we may conclude that each such rectangular cell contains a part of one (B,C) or even two (D) curves $f(\lambda, p) = 0$. All these cases can be transformed into four situations, shown schematically in Fig. 1. Note, that group D is ambiguous because the curves can either cross one another inside the cell (Fig. 1(c)), which means that the surface $z = f(\lambda, p)$ has a saddle point, or can have no common points (Fig. 1(d)). Provided we know an interval inside which the function changes sign, we can construct on such an interval a rectangle and then, having examined the signs at the vertices, determine the side on which the exit point is situated. Obviously, this side is a base for the construction of the adjacent cell where the described process can be repeated.

This analysis leads to an algorithm which locks to the curve $f(\lambda, p) = 0$ and follows it inside the examined region. We may assume that the solutions of (1) do not form closed curves inside the rectangle given by (2). Dividing the intervals $\langle \lambda_{\min}, \lambda_{\max} \rangle$ and $\langle p_{\min}, p_{\max} \rangle$ into N and M parts, respectively, we define the uniform mesh of points (λ_i, p_j), $i = 1 \cdots N$, $j = 1 \cdots M$. Next, for nodes located on the perimeter, we compute the function values and if they do not have the same sign at two subsequent points we start the tracing process. At each step we evaluate the function at two new points located on a grid line parallel to the line determined by the two most recently found nodes with opposite function signs. These four points form a new rectangu-

0018-9480/88/0300-0601$01.00 ©1988 IEEE

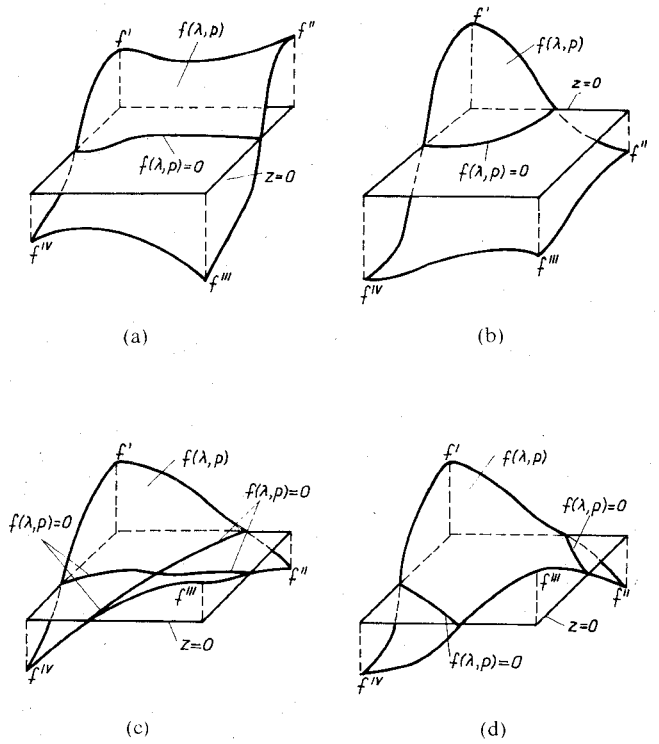

Fig. 1. Intersection of the plane $z = 0$ with the surface $z = f(\lambda, p)$.

lar cell. Depending on the combination of function signs at the vertices of the rectangle, we pass from one cell to another following the curve $f(\lambda, p) = 0$ until it reaches one of the sides of the frame. The search process on the perimeter is then resumed. However, the end point of each curve should be eliminated from further calculations so as to avoid following the curve twice, i.e., one time from the start point and a second time from the other end in the opposite direction.

III. Complementary Algorithms

Presently we will discuss several aspects of the algorithm which allow one to achieve greater efficiency and versatility in the numerical procedure. Passing from one cell to another, we compute the function values at two new points so that we know the function signs at four nodes. Note, however, that if the entry and exit points are situated on two perpendicular sides of a cell, the information about the function values in all four vertices is redundant. This is because the function signs at three of them only suffice to determine the transition of a curve through a cell. (For example, in the situation shown in Fig. 2, it is necessary to know the function values at the indicated nodes.) Hence, the procedure would be still more efficient if, instead of computing function values at the nodes of a cell in a fixed order, we decided on which side the exit point is possibly situated, and accordingly chose the proper vertex to compute the function. If the calculations confirm our assumption, we pass to the adjacent cell without checking the function sign at the fourth vertex. The slope of the curve on entry to a cell may be a criterion here for the sequence of computations.

Another important aspect of the algorithm is that the tracing process starts from the point located on the perimeter of the specified region. If a change of the function sign is encountered, it is possible to carry out tests verifying whether it is caused by a pole of the function or a spurious solution (i.e., a solution having

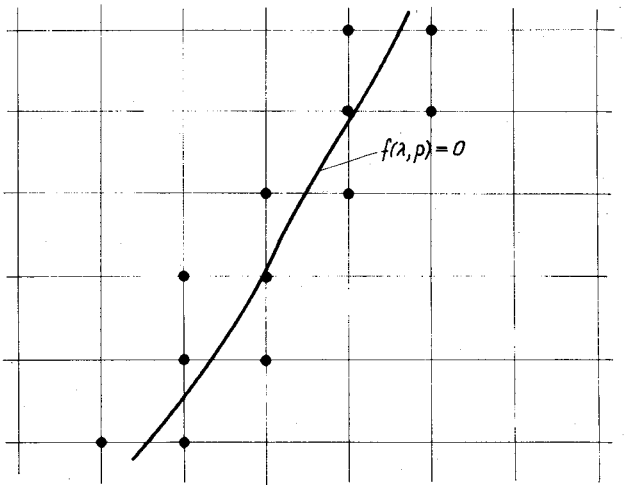

Fig. 2. Nodes at which a function has to be computed in order to determine a curve run.

no physical meaning [5]) and eliminate these improper curves from the diagram with low cost at an early stage of computations. We can also exclude fragments of the frame or even the whole perimeter from calculations providing only the known start points (for instance cutoff frequencies), in which case the curves are computed using the minimal number of function calls. The search process on the perimeter may also be used to control the computations. As was pointed out earlier, some intervals, namely those including the ends of curves, should be automatically eliminated from the frame during the computations. If we keep the record of already verified frame points, we can interrupt the computations at any moment and then resume the searching from the first node not yet examined. This property gives us an excellent tool for the construction of a restart procedure, a facility which is offered only on mainframes. The restart procedure is based on the verified frame points record and the two last zero loci (dumped periodically on a disk). These data are sufficient for determining the direction in which the curve was followed and then resuming the tracing process inside the examined region. Such a facility is extremely useful, in particular if the calculations are lengthy, because it prevents the program from accidental supply breakdowns and gives the user absolute control over the whole process.

Finally, we will highlight the pitfalls of the described algorithm. Generally speaking, the procedure suffers from the choice of mesh size. The curve is interpolated inside a cell by a segment, and accordingly we obtain a broken line instead of a smooth curve. One solution to minimize the error is to diminish the spacing between grid lines. However, a finer mesh would entail a rapid increase in the cost of computations. An alternative way is to compute the broken lines in the low-resolution process and subsequently to interpolate the curves between nodes using, for instance, spline functions [6]. Another problem arises if the curves approach one another, since, in this case, they will not be separated unless the mesh is fine enough (Fig. 3). Owing to the great distance between the grid lines, the procedure would find two lines, namely EFRQ and HGLS, losing the fragments FG and RL. Again, to cover the entire region with a denser mesh is by no means a remedy for this problem. This is because first it would increase the cost and, second, the mesh size that would be sufficient to separate the curves is not known exactly. The solution is to compute the diagram and then find all return points (these points are labeled R and L in Fig. 3). The return

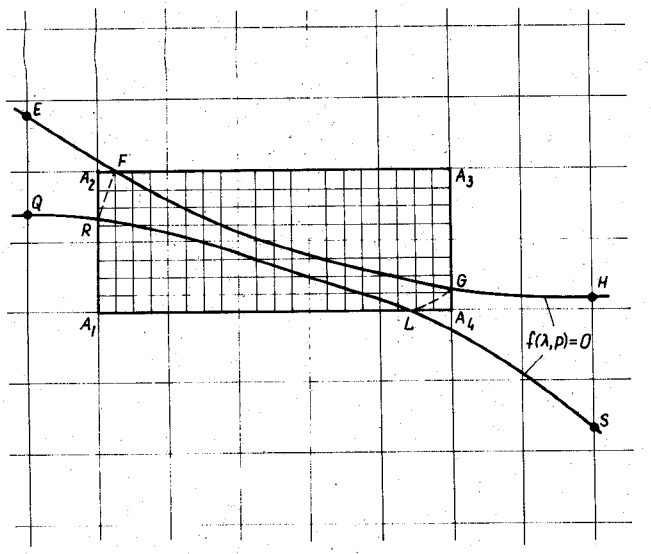

Fig. 3. Illustration of the correction process.

points can be classified into two categories, namely left and right ones. The next step of the correction procedure would be to match each left return point with its right counterpart and for each such pair to construct a new local frame (the rectangle A_1–A_4 in Fig. 3) covering its interior with a finer mesh. Obviously such a procedure can be iterated so that the proper diagram can be obtained in a single program run.

The method described in this paper is similar to the contour following scheme in three dimensions. However, function contouring is slightly different from zero searching. First of all, the contour following schemes allow contours to form closed curves. Consequently, they search the whole examined region for start points, assuming that function values are provided for each node. One could modify one of the existing library procedures so that it would follow the contour $f(\lambda, p) = 0$ searching the start points on the perimeter only. Such a modification would immediately bring about the problem of curve separation. Second, standard contouring procedures perform three-dimensional interpolation of a function inside a cell, which means that they require function values at additional points. Consequently, these procedures would be less efficient than the one proposed herein. For instance, in procedure OB14 from the Harwell Subroutine Library [9] the surface $z = f(\lambda, p)$ is approximated inside a cell by quadratics spanned over triangles. In effect this procedure requires as many as 16 values of the function in order to plot a part of a line which is inside a cell (compared to four or even three in the proposed algorithm). Finally, none of the standard procedures uses any of the complementary algorithms described above enhancing the efficiency and versatility of the new method.

IV. Numerical Example

The algorithm presented in this paper was successfully applied to solve characteristic equations in various nonstandard eigenvalue problems. All complementary algorithms outlined in the previous section, including the restart facility, were also tested and found very useful and reliable during the two years of their exploitation. As an example of the application, a sample dispersion diagram for a layered parallel-plate line containing a perpendicularly magnetized ferrite is presented in Fig. 4. Such a structure has been previously investigated by the author [7], and it was shown that the dispersion equation can be obtained using

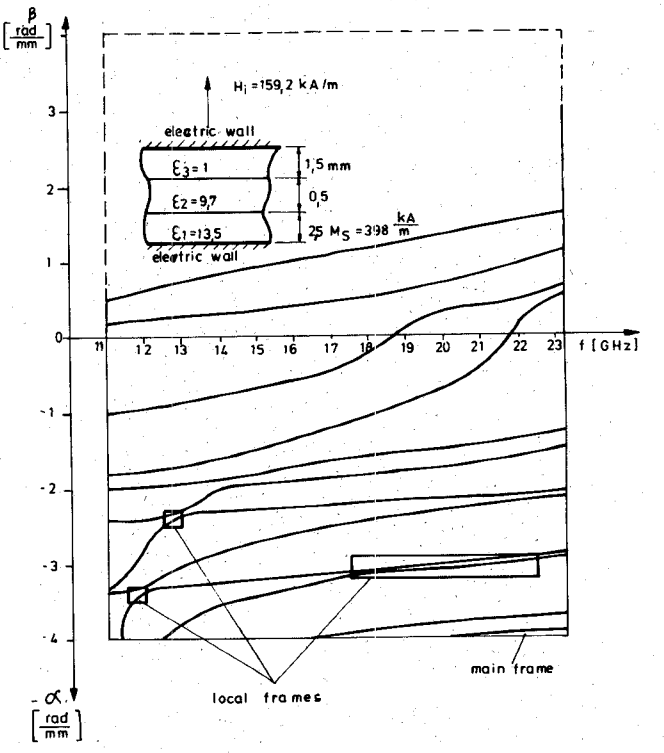

Fig. 4. A sample dispersion diagram obtained with the proposed procedure (a parallel-plate line with a perpendicularly magnetized layer of a gyromagnetic medium).

the normal field components approach. The ferrite medium causes coupling between the modes of the line. Accordingly, traditional methods would require great effort to determine the actual curve run. The new algorithm ensures efficiency of numerical computations. In particular, a part of the frame, indicated by the dashed line, was excluded from calculations by virtue of the theoretical analysis of the boundary value problem for such a structure [8]. Moreover, the curves were automatically separated. After the completion of the first step (low-resolution mesh: $\Delta\beta = 0.1$ rad/mm, $\Delta f = 250$ MHz), the procedure found three pairs of return points. Then the local frames, also shown in the figure, were constructed and grids with $\Delta\beta = 0.02$ rad/mm and $\Delta f = 50$ MHz were used in the correction process. One iteration was sufficient to separate the curves.

In the example discussed above, the proposed method was applied to solve a dispersion equation for a complicated waveguiding structure. The use of the new method, however, is not restricted to finding zero loci of equations resulting from the analysis of eigenvalue problems. This procedure would find application in any other situation in which a nonlinear equation of two variables occurs.

V. Conclusions

An efficient algorithm was proposed to compute the curves of zero loci of a real function of two variables. Starting from a point located on the perimeter of the prescribed region, the procedure follows each curve, minimizing the number of function evaluations. Different aspects of the algorithm were highlighted and ways were given to obtain extreme utility. The method was successfully applied in various electromagnetic problems and proved its efficiency, versatility, and reliability.

ACKNOWLEDGMENT

The author is indebted to D. Bowen of the British Council for his assistance during the preparation of the revised version of this paper.

REFERENCES

[1] R. H. Jansen, "The spectral-domain approach for microwave integrated circuits," *IEEE Trans. Microwave Theory Tech.*, vol MTT-34, pp 1043–1056, Oct. 1985.
[2] I V. Lindell, "Variational methods for nonstandard eigenvalue problems in waveguide and resonator analysis," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-30, pp. 1194–1204, Aug. 1982
[3] G. J. Gabriel and I. V. Lindell, "Comments on variational methods for nonstandard eigenvalue problems in waveguide and resonator analysis," *IEEE Trans. Microwave Theory Tech.*, vol MTT-31, pp. 786–789, Sept. 1983.
[4] G. J. Gabriel and I. V. Lindell, "Comments on 'Variational methods for nonstandard eigenvalue problems'," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-32, pp. 474–476, Apr 1984.
[5] A. Farrar and A. T. Adams, "Computation of propagation constants for the fundamental and higher order modes in microstrip," *IEEE Trans. Microwave Theory Tech.*, vol MTT-24, pp. 456–460, July 1976.
[6] C de Boor, *A Practical Guide to Splines*. New York: Springer, 1978.
[7] M. Mrozowski and J. Mazur, "General analysis of a parallel-plate waveguide inhomogeneously filled with gyromagnetic media," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-34, pp. 388–395, Apr. 1986
[8] M. Mrozowski and J Mazur, "The lower bound for the eigenvalues of the characteristic equation for an arbitrary multilayered structure with perpendicular magnetization," submitted to *IEEE Trans. Microwave Theory Tech.*
[9] Harwell Subroutine Library, Computer Science and System Division, Atomic Energy Research Establishment Report No-R9185, Sept 1980, Harwell, Oxfordshire, England

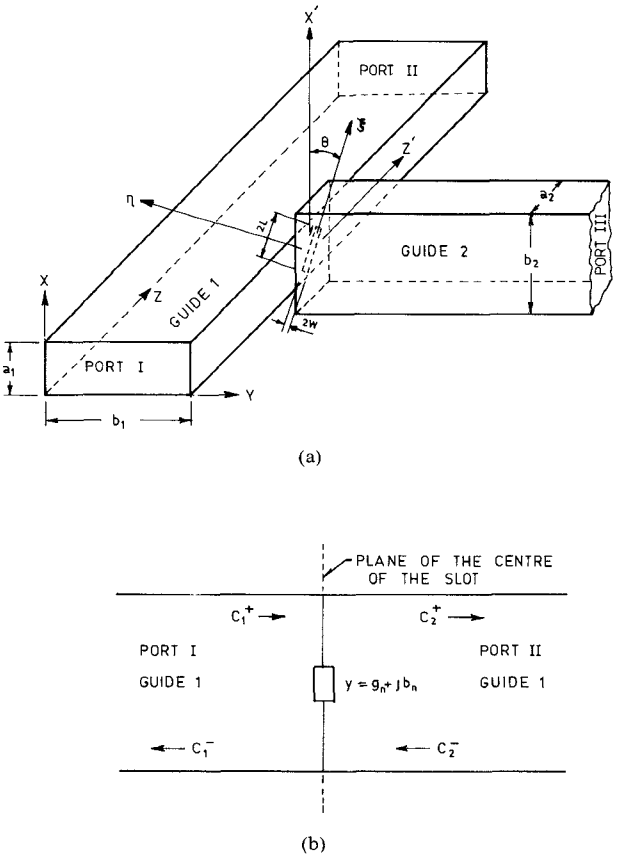

Fig. 1 (a) Coplanar $E-H$ plane T-junction. (b) Equivalent network representing impedance loading in guide 1.

Analysis of Coplanar $E-H$ Plane T-Junction Using Dissimilar Rectangular Waveguides

B. N. DAS, G S. N. RAJU, AND AJOY CHAKRABORTY

Abstract—**An analysis of a T-junction which differs from conventional H-plane T-junctions in that the T arm is rotated by 90° and coupling takes place through an inclined slot is presented. Since use of standard X-band waveguides result in such a T-junction operating above 11.7 GHz, nonstandard waveguide dimensions have been considered to bring down the operating frequency to 9.375 GHz. The effect of a change of the broad dimension of the primary feed waveguide on the resonant conductance is evaluated. The variations of resonant length with the angle of inclination of the slot, and coupling with frequency, are presented.**

I. INTRODUCTION

Investigations of H-plane T-junctions have already been reported [1]. In the present paper, thorough analytical investigations have been carried out on a T-junction (Fig. 1(a)) in which the coupling slot is in the narrow wall of the primary guide and the narrow dimension of the coupled guide is oriented along the axis of the primary guide. As a result, the E field of the coupled guide and the H field of the primary guide are coplanar and hence this type of T-junction is designated as a coplanar $E-H$ plane T-junction.[1] The power is coupled to the T-arm through an inclined slot in the narrow wall of the primary guide. It can be noted that in this type of T-junction no power can be coupled using either a longitudinal or a vertical slot. When such a coupler is made using standard X-band waveguides the maximum slot length which can be obtained is 14 mm with an inclination of 45° and the resonant frequency is around 12 GHz.

In the present work, investigations have been carried out to find the waveguide dimensions necessary for slots which resonate around 9.4 GHz. Further investigations have also been carried out to keep the normalized slot conductance loading on the primary guide as low as 0.01. This impedance is expressed in terms of self-reaction and discontinuity in modal current. Evaluation of self-reaction in the coupled guide employs TE and TM mode fields instead of the hybrid mode field used earlier [1]. To obtain the self-reaction in the primary guide, the magnetic current in the inclined slot is resolved into transverse and longitudinal components.

Computations have been carried out to obtain the waveguide dimensions for which a low value of normalized slot conductance is obtained at a resonant frequency around 9.375 GHz. Computed results of the various parameters of engineering importance, e.g., resonant slot length, slot conductance, and coupling, are presented.

Manuscript received May 13, 1987; revised September 28, 1987.

B N. Das and A. Chakraborty are with the Department of E and ECE, Indian Institute of Technology, Kharagpur 721302, India

G S. N. Raju was with the Department of ECE, A.U. College of Engineering, Visakhapatnam 530003, India He is now with the Department of E and ECE (under QIP), Indian Institute of Technology, Kharagpur 721302, India

IEEE Log Number 8718857.

[1]The Editor has kindly brought to the notice of the authors that this type of structure was reported by W. H. Watson in Fig. 34 of his paper "Resonant slots," *JIEE* (London), vol. 93, pt 3A, pp 747–777, 1946

0018-9480/88/0300-0604$01.00 ©1988 IEEE