[10] A. R. Conn and T. Pietrzykowski, "A penalty function method converging directly to a constrained optimum," Dep. Combinatorics and Optimization, Univ. Waterloo, Waterloo, Ont. Canada, Rep. 73–11, 1973.
[11] J. B. Rosen, "The gradient projection method for nonlinear programming, Part 1: linear constraints," *J. Soc. Ind. Appl. Math.*, vol. 8, 1960,
[12] C. Charalambous and A. R. Conn, "A new algorithm for discrete L_∞ linear approximation," to be published.

Minimax Optimization by Algorithms Employing Modified Lagrangians

OLOV EINARSSON

Abstract—Two general, modified Langrangian algorithms related to recent developments in nonlinear programming are presented. The methods give accurate results and are easy to program. An N-section transmission-line transformer is used as a test problem for minimax (equal ripple) optimization and the methods are compared to existing algorithms for network optimization.

I. INTRODUCTION

There exists a large class of optimization problems of engineering interest where some finite-dimensional functional is minimized (or maximized) subject to an equal ripple condition. The purpose of this short paper is to draw attention to the existence of two effective, recent algorithms which can be applied with advantage to this type of problem. While not new, the methods do not appear to have been applied to microwave problems before. The methods proposed are quite general and the choice of a transmission-line transformer problem as an example is only dictated by its use as a test problem in previous works on minimax optimization [1]–[3].

Consider the following minimax problem. Find the vector x which minimizes the real-valued function $f(x)$; i.e., find

$$\min_{x \in R^n} f(x) \tag{1}$$

where $f(x)$ is defined as

$$f(x) \triangleq \max_{\nu \in I} \tfrac{1}{2} \mid \rho(x,\nu) \mid^2 \tag{2}$$

and ρ is the reflection coefficient of the N-section lossless transmission-line transformer shown in Fig. 1. In (2) the frequency ν, normalized to some suitable frequency ν_0, is varied either over a closed interval

$$I = [\nu_1, \nu_M] \tag{3a}$$

or over a finite set

$$I = \{\nu_i\}_1^M. \tag{3b}$$

The components of the n-dimensional vector x in (1) are the (real) characteristic impedances and the lengths of the transmission-line sections. In one version the lengths of the sections are kept constant and equal to $\lambda_0/4$ where $\lambda_0 = c/\nu_0$. The corresponding x vector is

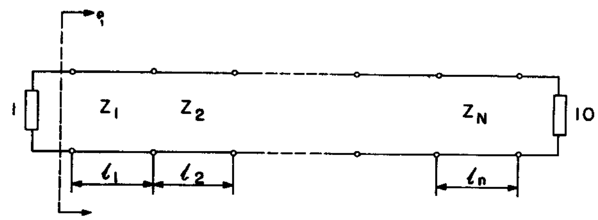

Fig. 1. 100-percent relative bandwidth 10:1 transmission-line transformer.

$$x = (Z_1, Z_2, \cdots, Z_N), \qquad n = N$$

$$l_1 = l_2 = \cdots = l_N = \lambda_0/4. \tag{4a}$$

Alternatively, the lengths of the transmission-line sections may also be varied, resulting in an x vector

$$x = (Z_1, Z_2, \cdots, Z_N, l_1, l_2, \cdots, l_N), \qquad n = 2N. \tag{4b}$$

The solution of the transmission-line transformer problem is known in terms of Chebyshev polynomials. The optimal lengths all turn out to be equal to $\lambda_0/4$ and the optimal impedances can be determined from the polynomial expressing the insertion loss function [4]. However, the methods of this investigation do not rely on this special polynomial structure of the problem.

II. DISCUSSION OF METHODS

It is readily seen that the unconstrained minimization problem given by (1), (2), and (3b) is equivalent to the following problem:

$$\min_{x \in R^n} f_1(x) \tag{5}$$

subject to the $M - 1$ nonlinear constraints

$$f_i(x) - f_1(x) \leq 0, \qquad i = 2,3,\cdots,M \tag{6}$$

where we have defined

$$f_i(x) \triangleq \tfrac{1}{2} \mid \rho(x,\nu_i) \mid^2 \tag{7}$$

and where we have used the fact that $\mid \rho(x,\nu) \mid$ takes its maximum value at the left end point ν_1 of the frequency range.

One well-established way of handling a nonlinear constrained optimization problem is to introduce a Lagrange multiplier for each constraint and construct a Lagrangian which will be stationary at the solution point. However, in the treatment of nonconvex problems, the usefulness of the Lagrangian is limited by the fact that a stationary point may not necessarily correspond to a minimum. Hestenes [5] and Powell [6] independently discovered that this drawback could be overcome by augmenting the Lagrangian with a quadratic penalty function term making the problem locally convex in a neighborhood of the solution point. A large number of different modified Lagrangians (also called "exact penalty functions" or "augmented Lagrangians") have been proposed both for problems with equality and inequality constraints [7], [8]. The modified Lagrangians employed in this short paper are given explicitly by (A1) and (A4) of Appendix A.

In most cases the Lagrange multipliers related to an optimization problem are unknown and have to be calculated during the minimization procedure. One way of doing this is to employ the saddle-point property of the modified Lagrangian $L(x,\mu)$ in the product space $[x,\mu]$ where μ is the multiplier vector, and solve the dual problem, i.e., to find

$$\max_{\mu \in R^m} \min_{x \in R^n} L(x,\mu) \tag{8}$$

by iterating alternately in x and μ space. The updating of the multipliers can be done in different ways. A very simple rule is used in the Hestenes–Powell method. For the routine VF01A [9] advantage is taken of the fact that if the unconstrained minimizations are

Manuscript received November 4, 1974; revised May 5, 1975.
The author was on leave from the University of Lund, at the Division of Radio and Electrical Engineering, National Research Council of Canada, Ottawa, Ont., Canada. He is now back at the College of Engineering, University of Lund, Lund, Sweden.

performed by a quasi-Newton algorithm, the approximation obtained to the second derivatives in x space can be utilized to approximate the corresponding derivatives in μ space, giving rapid (almost quadratic) convergence in both spaces.

III. NUMERICAL RESULTS

We have performed computations on the problem under consideration using two approaches: 1) an algorithm of the Hestenes–Powell type; 2) a Fortran routine VF01A developed by R. Fletcher at The Atomic Energy Research Establishment, Harwell, England [9], and available in the Harwell Subroutine Library. This routine employs a modified Lagrangian proposed by Rockafellar [8]. In both cases the unconstrained minimizations are performed by the Harwell subroutine VA09A which is a quasi-Newton algorithm employing the Broyden–Shanno–Fletcher formula to update an approximation to the Hessian, factorized in triangular form [10], [11].

The Hestenes–Powell algorithm requires that the constraints be given as equalities. If inequality constraints are present, this means that the "set of active constraints," i.e., the constraints satisfied as equalities at the solution point x^*, has to be determined. For minimax problems of the type we consider, information about the active set is usually available in advance and it is feasible to start with an approximation of the active set and adjust this set, if necessary, during the optimization procedure. We denote the active set of frequencies by

$$I_a \triangleq \{\nu_i,\ i = 2,3,\cdots,M \mid f_1(x^*) - f_i(x^*) = 0\}. \tag{9}$$

Assuming there are M_a constraints belonging to I_a we can, without loss of generality, arrange them so that

$$I_a = \{\nu_i\}_2^{M_a+1}. \tag{10}$$

Of course the formulation (3a) where the frequency is varied over a continuous interval is an optimization problem in an infinite-dimensional space and to apply finite-dimensional methods some approximations have to be made. The most obvious one is simply to sample the interval at a finite number of points. However, it should be remembered that the result obtained will be only an approximation of the continuous problem related to (3a).

Increasing the accuracy of the locations of the maxima by making the sampling interval smaller has the drawback of rapidly increasing the number of constraints which are almost active, with computational difficulties as a result. For that reason, for the five-section transformer a modification of the Hestenes–Powell algorithm was used where a simple quadratic interpolation search procedure for the ν_i was employed after every second outer iteration (i.e., change of μ).

All computations were performed on a Univac 1108 computer at the Computation Center of the University of Lund. A brief description of the algorithms can be found in Appendix A and the details of the cases treated are given in Appendix B.

In Table I the numerical results are collected and compared to the work by Bandler *et al.* [2] and by Charalambous and Bandler [3]. Their approach to the problem differs somewhat from ours. For example, the endpoint of the frequency range is not automatically included in the active set. For full details the reader is referred to the original papers. Note that one function evaluation means one evaluation of the function and its gradient.

When evaluating the results of Table I it should be borne in mind that the larger part of the number of function evaluations often is attributed to the first unconstrained minimization and consequently

TABLE I
NUMBER OF FUNCTION EVALUATIONS AND COMPUTER CPU TIME

Problem	Starting Point	Algorithm	Number of Function Evaluations	Computer CPU-time (seconds)
I N = 3 1 = $\lambda_0/4$	1, 3.1623, 10	Hestenes-Powell	26 + 10 + 6 = 42	0.71
		VF01A	33|[a] + 9 + 5 = 47 2.1-2[b] 1.5-3 3.5-6	1.09
		Bandler et al. [7]	219	
II N = 3 variable lengths	1, 3.1623, 10, 0.8, 1.2, 0.8	Hestenes-Powell	40 + 17 + 10 + 4 = 71	1.84
		VF01A	63| + 10 + 6 + 5 = 84 1.6-2 1.1-3 1.6-4 1.6-6	4.38
	1.5, 3, 6, 0.8, 1.2, 0.8	VF01A	39| + 7 + 4 = 50 3.2-3 1.2-4 7.4-6	2.76
		Charalambous and Bandler [9]	105, 95[c]	
	1, 3.1623, 10, 1, 1, 1	VF01A	70| + 7 + 4 = 81 3.2-3 2.8-4 2.4-6	4.14
		Charalambous and Bandler [9]	165, 155	
III N = 5 1 = $\lambda_0/4$	1.5, 3.5, 5.5, 7.5, 9.5	Hestenes-Powell	(23+30)[d] + (6+5) + (3+3) + + (6+4) + (5+3) = 88	1.72
		VF01A	53 + 58 + 22| + 14 = 147 4.3-3 5.4-4 2.1-4 2.4-5	16.56

Note: Starting points are given according to (4a) or (4b) with the lengths normalized to $\lambda_0/4$.
[a] Correct active set established.
[b] Gives values of K with notation $2.1\text{–}2 \triangleq 2.1 \cdot 10^{-2}$.
[c] Average number of function evaluations for Algorithm 1 and 2, respectively.
[d] Search for maxima performed.

is strongly dependent on the choice of starting point. The termination criteria chosen yield a relative error of the reflection coefficient for the three-section transformer ranging between $0.8 \cdot 10^{-5}$ and $0.5 \cdot 10^{-6}$. The essentially quadratic convergence of the methods makes an increase of this accuracy possible with little extra computational effort.

IV. CONCLUDING REMARKS

Based on the limited numerical results presented, it seems reasonable to conclude that minimax optimization can be effectively achieved through algorithms employing a modified Lagrangian, and that the efficiency of this approach is at least comparable to that of available algorithms specifically designed for minimax optimization.

The results obtained indicate that is is better to use a search procedure to locate the maxima and treat them as equality constraints than to create a large number of inequality constraints by a fine subdivision of the frequency interval.

The methods presented are quite general and can be applied to most optimization problems where equal ripples in some design quantity are desired. The modified Lagrangians employed are well-behaved functions resulting in rapid convergence of the minimizations required and making an increase of the desired accuracy of little cost in computational effort. They are insensitive to the values of the algorithmic parameters as long as the penalty factors are chosen large enough to ensure local convexity around the solution point. Furthermore, the Hestenes–Powell method is especially easy to program by incorporating any available unconstrained minimization routine into the program.

APPENDIX A

DESCRIPTION OF ALGORITHMS

Define the modified Lagrangian

$$L(x,\mu,Q) \triangleq f_1 + \sum_{i=2}^{M_a+1} \mu_i(f_i - f_1) + \sum_{i=2}^{M_a+1} Q_i(f_i - f_1)^2 \quad \text{(A1)}$$

where $\mu,Q \in R^{M_a}$ are the multiplier vector and the weight factor vector for the penalty function terms. With this notation the Hestenes–Powell-type algorithm used is as follows.

Initially set $x = x^0$, $\mu = 0$, $Q = Q^0$, $ep = ep^0$, then do the following.

1) Find $\min_x L(x,\mu,Q)$. Assume $x = x^*$ is this minimum with the final step towards the minimum

$$|x_i^* - x_i^k| \leq ep, \qquad i = 1,\cdots,n. \quad \text{(A2)}$$

2) Set $\mu_i^* = \mu_i + 2Q_i[f_i(x^*) - f_1(x^*)]$, $i = 2,3,\cdots,M_a + 1$ and $ep^* = 0.05ep$.

3) Finish if $|f_i(x^*) - f_1(x^*)| \leq ep1$, $i = 2,3,\cdots,M_a + 1$. If $f_i(x^*) - f_1(x^*) \geq \frac{1}{4}(f_i(x) - f_1(x))$ set $Q_i^* = 10Q_i$, $i = 2,3,\cdots, M_a + 1$.

4) Set $x = x^*$, $\mu = \mu^*$, $Q = Q^*$, $ep = ep^*$, go to 1).

The Harwell subroutine VF01A [20] solves the problem

$$\min_{x \in R^n} \{F(x) \mid c_i(x) \geq 0, \qquad i = 1,2,\cdots,M\} \quad \text{(A3)}$$

employing the modified Lagrangian

$$L_1(x,\theta,\sigma) \triangleq F(x) + \tfrac{1}{2} \sum_{i=1}^{M} \sigma_i\{[(c_i(x) - \theta_i)_-]^2 - \theta_i^2\} \quad \text{(A4)}$$

where θ, $\sigma \in R^M$, and the notation

$$a_- = \begin{cases} a & \text{if } a \leq 0 \\ 0 & \text{if } a > 0 \end{cases} \quad \text{(A5)}$$

is used. One outer iteration (in θ space) consists of minimizing L_1 over x for fixed θ and σ with the termination criterion (A2). In contrast to the earlier algorithm the same value of ep is used in all iterations. Due to the rapid convergence of the unconstrained minimization algorithm this should make little difference in the number of function evaluations. The components of σ are specified initially and then adjusted by the algorithm so that in two successive iterations $K^{(k+1)} < K^{(k)}$ where

$$K \triangleq \max_i |\min(c_i,\theta_i)|. \quad \text{(A6)}$$

The outer iterations are terminated when K is less than a specified number (*akmin*).

APPENDIX B

NUMERICAL DATA

The following cases have been treated where, because of symmetry, only $\nu_i \leq 1$ have been considered for the fixed length case. The parameters Q^0, ep, $ep1$, σ^0, and *akmin* relate to the initialization of the algorithms and the termination criteria. They are defined in connection with the description of the algorithms in Appendix A.

Problem I: Three-section transformer, fixed lengths, $l_i = \lambda_0/4$, $i = 1,2,3$; $n = 3$.

Hestenes–Powell: $M_a = 1$; $\{\nu_i\}_1^2 = 0.50,0.77$
$Q_2^0 = 10$, $ep = 10^{-2}$, $ep1 = 0.5 \cdot 10^{-6}$.

VF01A: $M = 5$; $\{\nu_i\}_1^6 = 0.50,0.60,0.70,0.77,0.90,1.00$
$\{\sigma^0\}_1^5 = 10$ (initial values), $akmin = 10^{-4}$, $ep = 10^{-5}$.

Problem II: Three-section transformer, variable lengths; $n = 6$.

Hestenes–Powell: $M_a = 3$, $\{\nu_i\}_1^4 = 0.50,0.77,1.23,1.50$
$\{Q_i^0\}_1^3 = 10$, $ep = 10^{-2}$, $ep1 = 0.5 \cdot 10^{-6}$.

VF01A: $M = 10$; $\{\nu_i\}_1^{11} = 0.50,0.60,0.70,0.77,0.90,1.00,1.10,$
$1.23,1.30,1.40,1.50$
$\{\sigma_i^0\}_1^{10} = 10,100,10,$[1] $akmin = 10^{-4}$,
$ep = 10^{-5}$.

Problem III: Five-section transformer, fixed lengths, $l_i = \lambda/4$, $i = 1,\cdots,5$; $n = 5$.

Hestenes–Powell with Maximum Search:

$M_a = 2$
initial $\{\nu_i\}_1^3 = 0.500,0.625,0.875$
final $\{\nu_i\}_1^3 = 0.50000,0.61231,0.85971$
(exact $\{\nu_i\}_1^3 = 0.50000,0.61229,0.85976$)
$Q_2^0 = Q_3^0 = 100$, $ep = 0.0156$.

VF01A: $M = 25$, $\{\nu_i\}_1^{26} = 0.50,0.52,\cdots,1.00$
$\{\sigma_i^0\}_1^{25} = 10$, $akmin = 10^{-4}$, $ep = 10^{-5}$.

ACKNOWLEDGMENT

The author wishes to thank Dr. R. A. Hurd who read the manuscript and made valuable suggestions for its improvement. He also wishes to thank Dr. I. Tagesson for his help with the computer programming.

REFERENCES

[1] J. W. Bandler and P. A. MacDonald, "Optimization of microwave networks by razor search," *IEEE Trans. Microwave Theory Tech.* (*Special Issue on Computer-Oriented Microwave Practices*), vol. MTT-17, pp. 552–562, Aug. 1969.
[2] J. W. Bandler, T. V. Srinivasan, and C. Charalambous, "Minimax optimization of networks by grazor search," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-20, pp. 596–604, Sept. 1972.
[3] C. Charalambous and J. W. Bandler, "New algorithms for network optimization," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-21, pp. 815–818, Dec. 1973.
[4] H. J. Riblet, "General synthesis of quarter-wave impedance transformers," *IRE Trans. Microwave Theory Tech.*, vol. MTT-5, pp. 36–43, Jan. 1957.
[5] M. R. Hestenes, "Multiplier and gradient methods," *J. Optimiz. Theory Appl.*, vol. 4, pp. 303–320, 1969.

[1] For the three starting points given in Table I.

[6] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed. New York: Academic, 1969, ch. 19, pp. 283–298.
[7] K. J. Arrow, F. J. Gould, and S. M. Howe, "A general saddle point result for constrained optimization," *Math. Program*, vol. 5, pp. 225–234, Oct. 1973.
[8] R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in non-convex programming," *SIAM J. Contr.*, vol. 12, pp. 268–285, May 1974.
[9] R. Fletcher, "An ideal penalty function for constrained optimization," Comput. Sci. and Syst. Div., Atomic Energy Research Establishment, Harwell, England, Rep. C. S. S. 2, pp. 1–27, Dec. 1973.
[10] ——, "A new approach to variable metric algorithms," *Comput. J.*, vol. 13, pp. 317–322, 1970.
[11] P. E. Gill and W. Murray, "Quasi-Newton methods for unconstrained optimization," *J. Inst. Math. Its Appl.*, vol. 9, pp. 91–108, 1972.

Integrated TRAPATT Diode Arrays

A. ROSEN, H. KAWAMOTO, MEMBER, IEEE, J. KLATSKIN, AND E. L. ALLEN, JR.

***Abstract*—This short paper is a description of the technique used to monolithically interconnect TRAPATT diodes in an array—resulting in a diode having low inductance interconnection and integrated heat capacitance which is necessary for long pulsewidths. For given power dissipation density and pulse length, the transient temperature rise in the diode decreases with the diameter. The reduction in diode diameter, however, leads to reduced power output. To take advantage of the reduction in temperature rise of small-size diodes while maintaining a large power output, a multiple-diode structure, monolithically interconnected, was fabricated.**

Pulsewidth operation of 50 μs has been achieved at a dissipation power density as high as 200 kW/cm², whereas the dissipation density must be reduced to 100 kW/cm² for the same total-area single-disk diode to operate reliably at 50 μs.

DIODE CONSIDERATIONS

The temperature rise of the TRAPATT diodes is a critical factor in limiting the device performance because it can ultimately lead to device failure. The dissipated energy per pulse, as well as the average power dissipation, are among the highest values required of any solid-state device. Therefore, it is necessary to give careful consideration to the dynamic thermal conditions in the diodes, and to the development of thermal design criteria that will lead to satisfactory performance.

The diodes in the array were interconnected both by utilizing monolithically connected metallized bridges and by soldering a piece of copper on top of the diodes. The additional copper mass acts as a heat capacitor, temporarily absorbing heat transients and thus extending the permissible operational pulsewidth. The diameter of each diode is sufficiently small to provide thermal spreading during the required pulse length, and, thus, the temperature rise during the pulse is reduced. The spacing between the diodes is sufficiently large to prevent thermal overlap between adjacent elements.

INTEGRATED DIODE ARRAY AND METALLIZED AIR BRIDGE

The diode array is fabricated by applying a dot pattern mark on a photoresist covered wafer, followed by a mesa etching. The process results in an array of diodes on a single integrated heat sink. This integrated heat sink (having diode arrays) is mounted on a microstrip line package, and then a gold-plated copper disk is soldered on the diodes' tops for interconnection. Later, a metallized air bridging technique [1], [2] is used to interconnect the "individual diode" mesas. The metallized air bridge provides, in addition to low inductance interconnection, an integrated heat capacitance which is necessary for long pulsewidth (50–100 μs) applications, such as in pulsed amplifiers for phased-array radar systems.

The processing steps applied to TRAPATT arrays are similar to the technique described by Basseches and Pfahnl [1] for interconnections on passive substrates. The major difference is that active semiconductor silicon mesas of small sizes, rather than large metal circuit patterns, are formed and interconnected. This semiconductor material must be protected amidst the large contour topography of the mesas, leading to a new process.

Detailed steps of this process are shown in Fig. 1 and are described as follows.

1) Batch process diodes using the standard mesa techniques.
2) Test all devices.
3) Apply positive photoresist, expose, and develop using a mask of array dots that are slightly smaller than diodes. This step serves two purposes: a) to prevent copper plating on top of the diode; and b) to protect the diode junction while plating.
4) Copper plate between the diodes and between arrays to a height slightly above mesa height.
5) Remove the photoresist above each diode.
6) Reapply the photoresist, expose, and develop using a connection pattern mask.
7) Gold plate to a thickness of 2 mils.
8) Remove the photoresist.
9) Remove the copper.
10) Remove the photoresist around mesa body.
11) Test.

Scanning electron micrographs (SEM's) of interconnected seven-diode arrays are shown in Fig. 2, and their I-V characteristics are shown in Fig. 3. The interconnection of the arrays was checked by comparing the junction capacity of the array with the junction capacity of single diodes within the array.

DIODE PERFORMANCE

The multiple diodes have been tested in S-band TRAPATT amplifier circuits. The results of the tests are summarized in Table I.

Seven-diode arrays made from a p-type wafer operated at an efficiency of 26.5 percent, 70-W output power, and 5.5-dB gain at S band in a coupled-bar circuit [4]. The 85-W output power at 50-μs pulsewidth has been achieved with a 19-diode array made from a double-diffused wafer [5] in a stagger-tuned microstrip line amplifier circuit [6]. The dissipation power density of the ordinary disk diode has to be controlled to be less than 100 kW/cm² to operate at 50-μs pulsewidth, but the array diode could dissipate at as high as 200 kW/cm² while operating at 50-μs pulsewidth. This indicates that the array structure is superior to the ordinary disk structure in power-handling capabilities.

Two 19-diode arrays made from a p-type wafer were mounted in series on a microstrip line package. Array diodes in series were operated in a stagger-tuned microstrip line circuit and demonstrated a 360-MHz bandwidth with an output power of 130 W (Fig. 4). This result has been obtained by an input level profiling technique in which the input RF power is adjusted as the frequency is varied to give a detected output waveform with no observable noise. The maximum gain was 6.5 dB and the maximum efficiency was 14.2 percent.

CONCLUSIONS

TRAPATT devices show great potential in the area of pulsed amplifiers for phased-array radar systems.

To achieve simultaneous high peak powers at high-duty cycle and long pulsewidth, special attention must be paid to the design

Manuscript received October 4, 1975; revised April 11, 1975. This work was supported in part by the U. S. Army Electronics Command under Contract DAAB07-74-C-0180.

The authors are with David Sarnoff Research Center, RCA Laboratories, Princeton, N. J. 08540.