SIMULATION OPTIMIZATION SYSTEMS
Research Laboratory

AN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION FOR DESIGN AND DIAGNOSIS OF NETWORKS

J.W. Bandler, W. Kellermann and K. Madsen

SOS-85-14-R

July 1985

McMASTER UNIVERSITY
Hamilton, Canada L8S 4L7
Department of Electrical and Computer Engineering

AN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION FOR DESIGN AND DIAGNOSIS OF NETWORKS

J.W. Bandler, W. Kellermann and K. Madsen

SOS-85-14-R

July 1985

AN ALGORITHM FOR NONLINEAR ℓ_1 OPTIMIZATION FOR DESIGN AND DIAGNOSIS OF NETWORKS

J.W. Bandler, Fellow, IEEE, W. Kellermann, Student Member, IEEE and K. Madsen

Abstract

This paper presents a new and highly efficient algorithm for nonlinear ℓ_1 optimization and its applications to circuits employing the properties of the ℓ_1 norm. The algorithm, based on the work of Hald and Madsen, is similar to a minimax algorithm originated by the same authors. It is a combination of a first-order method that approximates the solution by successive linear programming and a quasi-Newton method using approximate second-order information to solve a system of nonlinear equations resulting from the first-order necessary conditions for an optimum. The definitions of singular and regular ℓ_1 problems are given and a criterion for determining a singularity present in the ℓ_1 problem has been formulated. The versatility of the algorithm is proved by implementing it on three different computers. A single precision version on a Cyber 170/730 is compared with two double precision versions, one on a VAX 11/780 and one on a Texas Instruments Professional Computer. The new ℓ_1 algorithm is particularly useful in fault location methods using the ℓ_1 norm. A new technique for isolating the most likely faulty elements, based on an exact penalty function, is presented. Another important application of the algorithm is the design of contiguous band multiplexers consisting of multi-cavity filters distributed along a waveguide manifold which is illustrated by a 12 channel multiplexer design.

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants A7239 and G1135. This paper is based on material presented at the 1985 IEEE International Symposium on Circuits and Systems, Kyoto, Japan, June 5-7, 1985.

J.W. Bandler and W. Kellermann are with the Simulation Optimization Systems Research Laboratory and the Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada L8S 4L7.

K. Madsen is with the Institute for Numerical Analysis, Technical University of Denmark, Building 303, DK-2800 Lyngby, Denmark.

J.W. Bandler is also with Optimization Systems Associates, 163 Watson's Lane, Dundas, Ontario, Canada L9H 6L1.

I. INTRODUCTION

The optimization problem to be considered has the following mathematical formulation.

Let $f_j(\mathbf{x})=f_j(x_1,...,x_n)$, $j=1,...,m$, be a set of m nonlinear, continuously differentiable functions. The vector $\mathbf{x}\triangleq[x_1\ x_2\ ...\ x_n]^T$ is the set of n parameters to be optimized. We consider the following problem,

$$\underset{\mathbf{x}}{\text{minimize}} \quad F(\mathbf{x}) \triangleq \sum_{j=1}^{m} |f_j(\mathbf{x})|$$

subject to (1)

$$\mathbf{a}_i^T\mathbf{x} + b_i = 0 \qquad i = 1, ..., \ell_{eq},$$

$$\mathbf{a}_i^T\mathbf{x} + b_i \geq 0 \qquad i = (\ell_{eq}+1), ..., \ell,$$

where $\mathbf{a}_i$ and b_i, $i=1,...,\ell$, are constants. This is called the linearly constrained ℓ_1 problem.

The problem arises in a variety of areas. The most popular application of the ℓ_1 norm is the problem of approximating a function to data that might be contaminated with some wild points or gross errors. In this case the minimization of the ℓ_1 norm residual is superior to using other norms ℓ_p with $p > 1$ [1]. The larger the value of p, the more focus is put on the data points with largest deviation from the approximating function.

The number of applications of the ℓ_1 norm to circuit problems is increasing. The ℓ_1 norm has been successfully used to isolate the most likely faulty elements in fault isolation techniques for linear analog circuits [2]. We present a new technique for isolating the most likely faulty elements which is based on an exact penalty function.

Another important application of the ℓ_1 norm is the functional approach to post-production tuning [3], where the ℓ_1 type of objective function is used to select the number of tunable parameters needed to tune all possible outcomes of a manufactured design.

In this paper the ℓ_1 norm is employed in a general multiplexer design procedure. Therefore, a highly efficient and fast algorithm for ℓ_1 optimization is of great importance to

many circuit designers and engineers. It is the purpose of this paper to present such an algorithm.

We present an iterative algorithm for solving (1) which requires the user to supply function and gradient values of the nonlinear functions f_j. The algorithm also uses some second-order information, i.e., information about the second-order derivatives of the functions. This is approximated from the user supplied gradients.

The algorithm is similar to that of Hald and Madsen in [4]. It has been reported by Hald in [5], which describes and lists a Fortran subroutine implementing a version of the algorithm. Hald and Madsen [6] have demonstrated that the algorithm has sure convergence properties. Their results indicate that this algorithm may be the best of its class currently available.

The plan of the paper is as follows. In Section II previous work in the area of nonlinear ℓ_1 optimization is briefly reviewed. The algorithm of this paper is described in more detail in Section III, where the principles are explained by means of simple examples and the two methods, namely, the first-order method and the approximate second-order method are presented and the switching conditions between the two methods are given. The definitions of regular and singular ℓ_1 problems are also given. Section IV contains the comparison of different implementations of the present algorithm (Cyber 170/730, VAX 11/780 and TI/PC) using six test problems commonly used in the literature. A new technique for isolating the most likely faulty elements, based on an exact penalty function, is presented in Section V and illustrated by a simple mesh network example. In Section VI we describe an optimization procedure using the ℓ_1 norm for contiguous band multiplexer design. We conclude in Section VII with an assessment of the potential impact of the ℓ_1 algorithm in the area of circuit design and fault location.

II. REVIEW OF ℓ_1 ALGORITHMS

The problem (1) is, in principle, very similar to the linearly constrained minimax problem where the objective function is $F(\mathbf{x}) \triangleq \max|f_j(\mathbf{x})|$. Therefore, many of the algorithms for solving the minimax problem may be revised into algorithms for solving (1) and vice versa. For this reason most of the methods mentioned below have minimax counterparts. A survey of minimax algorithms has recently been given in Bandler, Kellermann and Madsen [7].

Most of the methods for minimizing the ℓ_1 function solve only the unconstrained problem (i.e. (1) with $\ell = 0$). For the type of methods to be described in the present paper, however, it is no complication and computationally costless to add the linear constraints.

One of the first attempts to solve the ℓ_1 problem was published in the paper of Osborne and Watson [8] in 1971. The method is iterative and at the kth iterate $\mathbf{x}_k$ the following linear approximation of the nonlinear ℓ_1 problem is used,

$$\underset{\mathbf{h}}{\text{minimize}} \quad \overline{F}(\mathbf{x}_k, \mathbf{h}) \triangleq \sum_{j=1}^{m} |f_j(\mathbf{x}_k) + \mathbf{f}_j'(\mathbf{x}_k)^T \mathbf{h}| . \tag{2}$$

This linear model problem is solved using linear programming. The direction $\mathbf{h}_k$ found is then used in a line search. This method has quadratic final convergence under special circumstances but normally the final convergence is much slower. The global convergence properties of this method are rather poor, and like the Gauss-Newton method for nonlinear least squares (which is similar) the Osborne and Watson method may provide fast convergence to a non-stationary point, i.e., a point which is not a local minimum.

The more recent papers on the ℓ_1 problem use some second-order information. Most of the methods require that the user supplies exact second (as well as first) derivatives. To the best of our knowledge the method to be described in this paper is the first which uses approximate second-order information (i.e., it is a second-order method, but the user supplies only first derivatives). The methods of the next paragraph use exact second-order information.

El-Attar, Vidyasagar and Dutta [9] use a sequence of smooth problems approximating the (nondifferentiable) ℓ_1 problem. Each of the smooth problems is solved by standard techniques and the sequence of solutions will often converge to a solution of the ℓ_1 problem. However, this kind of method may have severe ill-conditioning problems near an ℓ_1 solution because a nondifferentiable function with a kink is approximated by smooth functions. This gives curvatures in the smooth functions which tend to infinity as the ℓ_1 solution is approached. Murray and Overton [10] use a nonlinear programming formulation of the ℓ_1 problem and apply successive quadratic programming. A special line search algorithm is used to obtain a reduction in the ℓ_1 objective function. The algorithm of McLean and Watson [11] is a hybrid method like the method presented in this paper. It combines a first-order method based on (2) using trust regions with a Newton iteration. The first-order method is intended to be used initially, and close to a solution the Newton method should be used. This method often converges rapidly to a solution but the rules for switching between the two stages do not guarantee convergence. In fact the method may converge to a nonstationary point.

The algorithm of this paper is based on the work of Hald and Madsen [6]. It is a hybrid method combining a first-order method with an approximate second-order method. The first-order method is a robust trust region method which provides convergence to a neighbourhood of a solution. It is based on linear model problems of the type (2). These are solved subject to the constraints of the original problem (1) and a bound on the step length $\|\mathbf{h}\|$. The latter bound reflects the neighbourhood of the iterate $\mathbf{x}_k$ in which the kth model function (see (2)) is a good approximation to the nonlinear ℓ_1 function. If the solution approached by the first-order method is "singular" (see below) then a higher-order method must be used in order to obtain a fast ultimate rate of convergence. Therefore a switch is made to a quasi-Newton method that solves a set of nonlinear equations that necessarily hold at a solution of (1). This method has superlinear final convergence. Several switches between the first-order and the quasi-Newton method may take place. The reason for allowing this is that the latter method

works only close to a solution, so if it is started too early a switch back to the (more robust) trust region method is necessary. Notice that the user of this algorithm is required to supply function values and first-order derivatives, whereas the necessary second derivative information is generated by the algorithm.

The linearly constrained ℓ_1 problem may be formulated as a nonlinear programming problem. Then it can be solved by standard techniques from that field. When Powell's [12] method for nonlinear programming is applied to the ℓ_1 problem we obtain a method which in its final stages is very similar to our method. It can be shown that in the neighbourhood of a local solution of (1) our method generates the same points as Powell's method. However, in the latter method a quadratic program must be solved in every iteration, whereas we have to solve only a set of linear equations in the neighbourhood of a solution. Therefore, the computational effort used per iteration with our method is normally much smaller.

III. DESCRIPTION OF THE ALGORITHM

Illustrative Example

Before giving the detailed description of the algorithm we explain the principles through some simple examples.

First, consider the following one-dimensional case (with $m = 2$ and no constraints)

$$f_1(x) = x^2 + 3x \quad \text{and} \quad f_2(x) = x^2 - 2x + 1 . \tag{3}$$

The problem defined in (1) has a solution at $x = 0$. Suppose the iterate $x_k = -0.5$ is given. Then the linearized problem (2) is the following

$$\underset{h}{\text{minimize}} \{|-1.25 + 2h| + |2.25 - 3h|\} . \tag{4}$$

The objective function F and the model (4) are illustrated in Fig. 1a and Fig. 1b, respectively. Near x_k the model is a good approximation to F. It also has a kink near 0, but it is not a minimum. This is one of the motivations for using a trust region method, i.e., solving (4) subject to a bound on h so we consider the model only in the domain where it is close to F.

When x_k is closer to the minimum of F at $x = 0$ the situation is better. In this case the model minimum exists and is close to the minimum of F (see Fig. 2).

The method presented here is called Method 1 below. It has the following structure. At the iterate $\mathbf{x}_k$ the model problem (2) is solved subject to a bound on the step length, $\|\mathbf{h}\| \leq \Lambda_k$. The parameter Λ_k is updated during the iteration. If the model and the nonlinear function F are very similar at the new point $(\mathbf{x}_k + \mathbf{h}_k)$ then the bound is increased. On the other hand, if the two quantities are rather distinct then the bound is decreased.

In the example (3) above Method 1 converges very fast to the solution $x^* = 0$. When x_k is close enough to x^* the model problem has a solution with a distance from x^* which is approximately the square of the previous distance, $|x_k - x^*|^2$. This is called quadratic convergence. There are, however, many cases where Method 1 slows down when a solution is approached. This is illustrated in the next example. Consider the two dimensional case (with $m = 2$ and no constraints)

$$f_1(\mathbf{x}) = (x_1 - 1)^2 + x_2^2 ,$$
$$f_2(\mathbf{x}) = x_1^2 - x_2 . \tag{5}$$

Some level curves of F near the solution are given in Fig. 3. The solution is at $\mathbf{x}^* = [0.59\ 0.35]^T$, where $f_1(\mathbf{x}^*) > 0$ and $f_2(\mathbf{x}^*) = 0$. F has a kink at the dotted curve, otherwise it is smooth. The dotted curve represents the bottom of a valley. When $\mathbf{x}_k$ is close to $\mathbf{x}^*$ the model F of (2) has a valley the bottom of which is a line that approximates the dotted curve near $\mathbf{x}_k$ (and $\mathbf{x}^*$). Therefore, Method 1 provides rapid convergence to the dotted curve through $\mathbf{x}^*$ but along this curve (in the bottom of the valley) the convergence becomes slow because of the lack of second-order information. This motivates a shift to another method when $\mathbf{x}_k$ is close to $\mathbf{x}^*$. The method we use (called Method 2) is a quasi-Newton iteration to solve a set of nonlinear equations that hold at a solution $\mathbf{x}^*$. This set is given in Appendix A. In order to set up the equations we must know which functions are zero at the solution being approached, and therefore Method 2 can only be used close to a solution.

Regular and Singular ℓ_1 Problems

The essential difference between the two examples (3) and (5) lies in the number of functions being zero at the solution, compared to the number of variables n. In (3) the number of zero functions is n, in (5) it is less than n. Zero functions are helpful because they provide kinks which also appear in the model. Therefore n zero functions will normally mean n "linearly independent" kinks which means that as the nonlinear solution is approached the model has a solution very close to the nonlinear solution, as illustrated in Fig. 2.

Definition 1 We say that the solution $\mathbf{x}^*$ of the unconstrained ($\ell = 0$) ℓ_1 problem is regular if the set

$$\{ \mathbf{f}_j'(\mathbf{x}^*) \mid f_j(\mathbf{x}^*) = 0 \}$$

spans the space R^n. Otherwise the solution is singular.

When constraints are present, active constraints play the same role as zero functions. This motivates the next definition.

Definition 2 We say that the solution $\mathbf{x}^*$ of the linearly constrained ℓ_1 problem is regular if the set

$$\{\mathbf{f}_j'(\mathbf{x}^*) \mid f_j(\mathbf{x}^*) = 0\} \cup \{\mathbf{a}_i \mid \mathbf{a}_i^T \mathbf{x}^* + b_i = 0 \}$$

spans the space R^n. Otherwise the solution is singular.

These definitions correspond to the definitions given by Madsen and Schjaer-Jacobsen [13] for the unconstrained minimax problem.

Normally a problem is regular if the (total) number of zero functions and active constraints is at least n. For regular problems the method of this paper has been shown ([6]) to be quadratically convergent. For singular problems the convergence is superlinear.

We now give a detailed description of the method which is a combination of Method 1 and Method 2. We first describe the two basic methods and next the combined method, including switching rules.

Method 1

This is a method providing global convergence. At the kth step a feasible approximation $\mathbf{x}_k$ to a solution of (1) and a local bound Λ_k are given. In order to find a better estimate the following linearized problem is solved,

$$\underset{\mathbf{h}}{\text{minimize}} \ \bar{F}(\mathbf{x}_k, \mathbf{h}) \triangleq \sum_{j=1}^{m} |f_j(\mathbf{x}_k) + \mathbf{f}_j'(\mathbf{x}_k)^T \mathbf{h}|$$

subject to

$$\begin{aligned} &\|\mathbf{h}\|_\infty \le \Lambda_k \\ &\mathbf{a}_i^T(\mathbf{x}_k + \mathbf{h}) + b_i = 0, \qquad i = 1, \ldots, \ell_{eq}, \\ &\mathbf{a}_i^T(\mathbf{x}_k + \mathbf{h}) + b_i \ge 0, \qquad i = (\ell_{eq}+1), \ldots, \ell. \end{aligned} \tag{6}$$

The solution of (6), $\mathbf{h}_k$, may be found by a standard linear programming routine. However, we use an implementation of the algorithm of Bartels, Conn and Sinclair [14], which is more efficient. Notice that $(\mathbf{x}_k + \mathbf{h}_k)$ is feasible.

The next iterate is $(\mathbf{x}_k + \mathbf{h}_k)$ provided that this point is better than $\mathbf{x}_k$ in the sense of F, i.e., if $F(\mathbf{x}_k + \mathbf{h}_k) < F(\mathbf{x}_k)$. Otherwise $\mathbf{x}_{k+1} = \mathbf{x}_k$.

The local bound Λ_k is adjusted in every iteration based on comparison between the decrease in the nonlinear objective function and the decrease predicted by the model $\bar{F}$. If the ratio between the two is small,

$$F(\mathbf{x}_k) - F(\mathbf{x}_k + \mathbf{h}_k) \le 0.25[\bar{F}(\mathbf{x}_k, \mathbf{0}) - \bar{F}(\mathbf{x}_k, \mathbf{h}_k)], \tag{7}$$

then the bound is decreased: $\Lambda_{k+1} = \Lambda_k/4$. Otherwise, if

$$F(\mathbf{x}_k) - F(\mathbf{x}_k + \mathbf{h}_k) \ge 0.75[\bar{F}(\mathbf{x}_k, \mathbf{0}) - \bar{F}(\mathbf{x}_k, \mathbf{h}_k)], \tag{8}$$

then $\Lambda_{k+1} = 2\Lambda_k$. If neither (7) nor (8) hold then we leave the bound unchanged, $\Lambda_{k+1} = \Lambda_k$.

Experiments have shown that the method is rather insensitive to small changes in the constants used in this updating procedure of the local bound. Notice that if the new point $(\mathbf{x}_k + \mathbf{h}_k)$ is not accepted then the bound is decreased.

Method 2

This is a local method. It is assumed that a point near a solution $\mathbf{x}^*$ is known and that the set of zero functions

$$Z(\mathbf{x}^*) \triangleq \{j \mid f_j(\mathbf{x}^*) = 0\} \tag{9}$$

and the set of active constraints,

$$A(\mathbf{x}^*) \triangleq \{i \mid \mathbf{a}_i^T \mathbf{x}^* + b_i = 0\} \tag{10}$$

are known.

Method 2 is an approximate Newton method for solving the nonlinear system (A1) of Appendix A (in the variables $(\mathbf{x}, \boldsymbol{\delta}, \boldsymbol{\mu})$). Exact first derivatives are used but the matrix

$$\mathbf{g}''(\mathbf{x}^*) + \sum_{j \in Z} \delta_j \mathbf{f}_j''(\mathbf{x}^*)$$

is approximated using a modified BFGS update (see Appendix B for details). In this way an approximate Jacobian $\mathbf{J}_k$ is obtained at the estimate $(\mathbf{x}_k, \boldsymbol{\delta}^{(k)}, \boldsymbol{\mu}^{(k)})$ of the solution of (A1). The next estimate is obtained by

$$\mathbf{J}_k \begin{bmatrix} \Delta \mathbf{x}_k \\ \Delta \boldsymbol{\delta}^{(k)} \\ \Delta \boldsymbol{\mu}^{(k)} \end{bmatrix} = -\mathbf{R}\left(\mathbf{x}_k, \boldsymbol{\delta}^{(k)}, \boldsymbol{\mu}^{(k)}\right) \tag{11}$$

$$\left(\mathbf{x}_{k+1}, \boldsymbol{\delta}^{(k+1)}, \boldsymbol{\mu}^{(k+1)}\right) = \left(\mathbf{x}_k, \boldsymbol{\delta}^{(k)}, \boldsymbol{\mu}^{(k)}\right) + \left(\Delta \mathbf{x}_k, \Delta \boldsymbol{\delta}^{(k)}, \Delta \boldsymbol{\mu}^{(k)}\right)$$

where $\mathbf{R}$ is defined by (A2) of Appendix A. Notice that no line search is involved.

The Combined Method

The combined method is the algorithm which we recommend to use in this paper. Method 1 is intended to provide the global convergence and Method 2 is used to obtain fast local convergence.

Initially, Method 1 is used and the sets (9) and (10) are estimated. When a local minimum seems to be approached a switch to Method 2 is made. If the Method 2 iteration is unsuccessful then Method 1 is used again. Several switches between the two methods may take place. When Method 1 is used we say that the iteration is in Stage 1, otherwise it is in Stage 2. A detailed description of the two stages follows.

The Stage 1 Iteration

We have a point $\mathbf{x}_k$, a local bound Λ_k and a matrix $\mathbf{J}_k$ which should approximate the Jacobian of (A1).

1. $\mathbf{x}_{k+1}$ and Λ_{k+1} are found using Method 1, and approximations Z_{k+1} and A_{k+1} of the sets (9) and (10) are found via the zero and active sets at the solution $\mathbf{h}_k$ of the linear model problem (6).
2. An estimate $(\boldsymbol{\delta}^{(k+1)}, \boldsymbol{\mu}^{(k+1)})$ of the multipliers is found through a least squares solution of (A1) with $(\mathbf{x}_{k+1}, Z_{k+1}, A_{k+1})$ inserted for $(\mathbf{x}, Z(\mathbf{x}), A(\mathbf{x}))$. This estimate is used for finding a new Jacobian estimate $\mathbf{J}_{k+1}$ by the BFGS method as described in Appendix B.
3. A switch to Stage 2 is made if the following two conditions hold:
   (a) The estimates Z_{k+1} and A_{k+1} have been constant over v consecutive different Stage 1 iterates ($v \geq 3$).
   (b) The multiplier estimates are in the correct ranges,

$$|\delta_j^{(k+1)}| \leq 1 ,$$

$$\mu_j^{(k+1)} \geq 0 .$$

The Stage 2 Iteration

We have an estimate $(\mathbf{x}_k, \boldsymbol{\delta}^{(k)}, \boldsymbol{\mu}^{(k)})$, estimates Z_k and A_k of (9) and (10), and a matrix $\mathbf{J}_k$ which should approximate the Jacobian of (A1).

1. Find $(\mathbf{x}_{k+1}, \boldsymbol{\delta}^{(k+1)}, \boldsymbol{\mu}^{(k+1)})$ and $\mathbf{J}_{k+1}$ using Method 2 with (Z_k, A_k) inserted for $(Z(\mathbf{x}^*), A(\mathbf{x}^*))$.

2. Let $A_{k+1} = A_k$, $Z_{k+1} = Z_k$ and $\Lambda_{k+1} = \Lambda_k$.

3. Switch to Stage 1 if one of the following conditions holds:

   (a) A function f_j with $j \notin Z_k$ has changed sign, or a constraint corresponding to an index i with $i \notin A_k$ has become violated.

   (b) A component of $\boldsymbol{\delta}^{(k+1)}$ or of $\boldsymbol{\mu}^{(k+1)}$ is outside its range:

   $$|\delta_j^{(k+1)}| > 1 ,$$

   or

   $$\mu_j^{(k+1)} < 0 .$$

   (c) $$\|\mathbf{R}(\mathbf{x}_{k+1}, \boldsymbol{\delta}^{(k+1)}, \boldsymbol{\mu}^{(k+1)}\| > 0.999\|\mathbf{R}(\mathbf{x}_k, \boldsymbol{\delta}^{(k)}, \boldsymbol{\mu}^{(k)})\|$$

   (see (A2) for the definition of $\mathbf{R}$).

This completes the description of the combined method.

It has been shown by Hald and Madsen [6], that the method has safe global convergence properties: it can only converge to stationary points. Furthermore, the final rate of convergence is at least superlinear, i.e.,

$$\|\mathbf{x}_{k+1} - \mathbf{x}^*\| \leq \varepsilon_k \|\mathbf{x}_k - \mathbf{x}^*\| , \tag{12}$$

where $\varepsilon_k \to 0$ for $k \to \infty$.

IV. COMPARISON OF DIFFERENT IMPLEMENTATIONS OF THE ALGORITHM

The algorithm described here has been reported by Hald and Madsen in [6], which contains also a comparison of two versions (for an IBM 3033 computer) of the algorithm with other ℓ_1 algorithms. The implementation of the algorithm used in our paper corresponds to

version B described in [6], where the user can choose parameters Λ_0 (initial step length of the algorithm) and v (the number of successive iterations with identical sets of active residual functions that is required before a switch to Stage 2 is made).

The numerical results presented here have been obtained using three implementations of the algorithm, each developed for a different machine. The machines used were a Cyber 170/730, a VAX 11/780 and a Texas Instruments PC. The versions on VAX and TI/PC are double precision versions while the version on Cyber is a single precision one.

To compare the performance of the present algorithm on different computers the following test problems have been used.

Test Problem 1 (El-Attar et al. [9]).

The ℓ_1 functions are:

$$f_1(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 - 1 ,$$

$$f_2(\mathbf{x}) = x_1^2 + x_2^2 + (x_3 - 2)^2 ,$$

$$f_3(\mathbf{x}) = x_1 + x_2 + x_3 - 1 ,$$

$$f_4(\mathbf{x}) = x_1 + x_2 - x_3 + 1 ,$$

$$f_5(\mathbf{x}) = 2x_1^3 + 6x_2^2 + 2(5x_3 - x_1 + 1)^2 ,$$

$$f_6(\mathbf{x}) = x_1^2 - 9x_3 .$$

Starting point: $[1 \quad 1 \quad 1]^T$

Solution: $\mathbf{x} = [0.53597 \quad 0.0 \quad 0.03192]^T$

The solution is singular and the ℓ_1 objective function value is 7.89423.

Test Problem 2 (Madsen [15])

The ℓ_1 functions are

$$f_1(\mathbf{x}) = x_1^2 + x_2^2 + x_1 x_2 ,$$

$$f_2(\mathbf{x}) = \sin x_1 ,$$

$$f_3(\mathbf{x}) = \cos x_2 .$$

Starting point: $[3\ 1]^T$

Solution: $[0.0\ \ 0.0]^T$

The solution is regular and the ℓ_1 objective function value is 1.00000.

Test Problem 3 (Kowalik and Osborne, see Watson [16])

$$f_i(\mathbf{x}) = v_i - \frac{x_1(y_i^2 + x_2 y_i)}{y_i^2 + x_3 y_i + x_4}, \qquad i = 1, \ldots, 11.$$

The constants v_i and y_i can be found in [16].

Starting point: $[0.25\ \ 0.39\ \ 0.415\ \ 0.39]^T$

Solution: $[0.19337\ \ 0.19377\ \ 0.10893\ \ 0.13973]^T$.

The solution is regular and the ℓ_1 objective function value is 3.876797×10^{-2}.

Test Problem 4 (Bard, see Watson [16]).

$$f_i(\mathbf{x}) = y_i - x_1 - \frac{i}{(16-i)x_2 + \min(i, 16-i)\, x_3}, \qquad i = 1, \ldots, 15.$$

Starting point: $[1\ \ 1\ \ 1]^T$

Solution: $[0.10094\ \ 1.52516\ \ 1.97211]^T$

The solution is regular and the ℓ_1 objective function value is 0.12434.

Test Problem 5 (Hettich, see Watson [16])

$$f_i(\mathbf{x}) = \sqrt{t_i} + ((x_1 t_i + x_2)t_i + x_3)^2 - x_4, \qquad i = 1, 2, \ldots, 5$$

$$t_i = 0.25 + (i - 1) \times 0.75/4 .$$

Starting point: $[0.0\ -\ 0.5\ \ 1.0\ \ 1.5]^T$

Solution: $[0.08273\ -\ 0.48321\ \ 1.13571\ \ 1.54057]^T$

The solution is singular and the ℓ_1 objective function value is 7.56472×10^{-3}.

Test Problem 6 (El-Attar et al. [9])

$$f_i(\mathbf{x}) = x_1 e^{-x_2 t_i} \cos(x_3 t_i + x_4) + x_5 e^{-x_6 t_i} - y_i$$

$$y_i = \frac{1}{2} e^{-t_i} - e^{-2t_i} + \frac{1}{2} e^{-3t_i} + \frac{3}{2} e^{-3t_i/2} \sin(7\, t_i) + e^{-5t_i/2} \sin(5\, t_i)$$

$$t_i = (i - 1)/10\,, \qquad i = 1, \ldots, 51.$$

Starting point: [2.0 2.0 7.0 0.0 -2.0 1.0]^T

Solution: [2.24074 1.85769 6.77005 -1.64490 0.16589 0.74228]^T

The solution is regular and the objective function value is 0.559813.

Table I shows the performance of the algorithm on three different computers. In all cases the same solutions have been obtained. In all examples we use $\Lambda_0 = 0.5$ and $v = 3$.

V. FAULT ISOLATION USING THE ℓ_1 NORM

Formulation of the Problem

In this section we deal with fault isolation in linear analog circuits under an insufficient number of independent voltage measurements. The ℓ_1 norm is used to isolate the most likely faulty elements. Practically, the faulty components are very few and the relative change in their values is significantly larger than in the nonfaulty ones [17].

The method presented here is a modification of the method utilizing multiple test vectors to obtain the measurements [2].

For k different excitations applied to the faulty network we consider the following optimization problem.

$$\underset{\mathbf{x}}{\text{Minimize}} \quad \sum_{i=1}^{n} |\Delta x_i / x_i^0| \tag{13a}$$

subject to

$$\mathbf{V}_1^c - \mathbf{V}_1^m = \mathbf{0},$$

$$\vdots \qquad (13b)$$

$$\mathbf{V}_k^c - \mathbf{V}_k^m = \mathbf{0},$$

where $\mathbf{x} \triangleq [x_1\ x_2\ \ldots\ x_n]^T$ is a vector of network parameters, $\mathbf{x}^0$ represents the nominal parameter values, $\Delta x_i \triangleq x_i - x_i^0$, $i = 1, 2, \ldots, n$, represent the deviations in network parameters from nominal values, $\mathbf{V}_k^m$ is a p-dimensional vector of voltage measurements performed at the accessible nodes for the kth excitation and $\mathbf{V}_k^c$ is a p-dimensional vector of voltages at accessible nodes calculated using the vector $\mathbf{x}$ as parameter values.

The corresponding nonlinear ℓ_1 problem can be formulated based on an exact penalty function [18] as follows.

$$\underset{\mathbf{x}}{\text{Minimize}} \quad \sum_{j=1}^{n+kxp} |f_j(\mathbf{x})| \qquad (14)$$

where

$$f_i(\mathbf{x}) \triangleq \Delta x_i / x_i^0, \qquad i = 1, 2, \ldots, n, \qquad (15)$$

$$f_{n+i}(\mathbf{x}) \triangleq \beta_i (V_i^c - V_i^m), \qquad i = 1, 2, \ldots, kxp, \qquad (16)$$

and β_i, $i = 1,2,\ldots,kxp$, are apropriate multipliers (satisfying certain conditions stated in [18]).

Mesh Network Example [2]

Consider the resistive network shown in Fig. 4 with the nominal values of elements $G_i = 1.0$ and tolerances $\varepsilon_i = \pm\ 0.05$, $i = 1, 2, \ldots, 20$. All outside nodes are assumed to be accessible with node 12 taken as the reference node. Nodes 4,5,8 and 9 are assumed internal, where no measurements can be performed.

Two faults are assumed in the network in elements G_2 and G_{18}. For Case 1 we applied the new ℓ_1 algorithm to optimization problem (14) with a single excitation at node 1. For Case 2 we considered two excitations applied at nodes 3 and 6 sequentially. The results of both

optimization problems are summarized in Table II. The nominal component values have been used as a starting point since just a few elements change significantly from nominal.

In both cases the actual faulty elements have been identified, but in Case 2, the estimated changes in the faulty elements are closer to their true values. Also some of the changes in the nonfaulty components approach better their true values in Case 2.

The estimated changes in the faulty elements are much closer to the actual changes as compared to the results reported in [2].

VI. CONTIGUOUS-BAND MULTIPLEXER DESIGN USING THE ℓ_1 NORM

Introductory Remarks

Practical design and manufacture of contiguous and non-contiguous band microwave multiplexers consisting of multi-cavity filters distributed along a waveguide manifold has been a problem of significant interest [19-22]. Recently, a general multiplexer optimal design procedure using a powerful gradient-based minimax algorithm has been described [7]. The simulation and sensitivity analysis aspect of the problem together with a number of examples of multiplexer optimization have been presented in [23]. A typical structure under consideration is shown in Fig. 5. All design parameters of interest, e.g., waveguide spacings, input-output and filter coupling parameters, can be directly optimized. A wide range of possible multiplexer optimization problems can be formulated and solved by appropriately defining specifications on common port return loss and individual channel insertion loss functions.

A major task in designing a multiplexer is to determine the location of the channel filters along the waveguide manifold [21]. This is very important for designs using the common port return loss as the only optimization criterion. A typical value of lower specification on return loss over the passbands of all multiplexer channels is 20 dB.

The error functions $f_j(\mathbf{x})$, $j \in J$, are of the form

$$- w_L(\omega_i)\,(F(\mathbf{x}, \omega_i) - S_L(\omega_i))\,, \tag{17}$$

where $F(\mathbf{x}, \omega_i)$ is the return loss at the common port at the ith frequency, $S_L(\omega_i)$ is the lower specification on return loss at the ith frequency, $\mathbf{x} \triangleq [x_1 \;\; x_2 \;\ldots\; x_n]^T$ is the vector of design parameters and w_L is an arbitrary user chosen non-negative weighting factor.

If we perform a minimax optimization based on these error functions and at the solution the minimax objective function value is negative then the goal has been achieved. In many cases, however, using the filter spacings as the only optimization variables may not be sufficient to satisfy all specifications and minimax optimization gives results corresponding to the situation where the specification violations are distributed over all multiplexer channels. In that case the use of the one-sided ℓ_1 optimization of the same error functions may lead to more desirable results where the violations occur only over a few multiplexer channels. This process of identifying "bad channels" has two very important consequences. First, the results indicate in which channels the additional variables have to be released to improve locally (in the frequency domain) the performance of the multiplexer and second, it gives very good starting values of the waveguide spacings to be used in the subsequent minimax optimization. The idea presented is illustrated by designing a 12 GHz, 12 channel multiplexer without dummy channels. The 12-channel contiguous band multiplexer has a channel frequency separation of 40 MHz and a usable bandwidth of 39 MHz with the center frequency of channel no. 1 12 180.0 MHz.

12-Channel 12 GHz Multiplexer Design

Suppose we want to design this multiplexer such that a lower specification of 20 dB on the common port return loss over the passbands of all 12 channels should be satisfied.

We start the design process with twelve identical 6th order filters with the coupling coefficients given in the following matrix [24]

$$\mathbf{M} = \begin{bmatrix} 0 & 0.594 & 0 & 0 & 0 & 0 \\ 0.594 & 0 & 0.535 & 0 & 0 & 0 \\ 0 & 0.535 & 0 & 0.425 & 0 & -0.400 \\ 0 & 0 & 0.425 & 0 & 0.834 & 0 \\ 0 & 0 & 0 & 0.834 & 0 & 0.763 \\ 0 & 0 & -0.400 & 0 & 0.763 & 0 \end{bmatrix}$$

Initially we select the spacing lengths along the waveguide manifold as the only optimization variables with starting values set equal to $\lambda_{gk}/2$ (half the wavelength corresponding to the kth center frequency). For the kth channel the waveguide spacing is measured along the manifold from the adjacent (k – 1)th channel. For the first channel the spacing is the distance from the short circuit. The filters are assumed lossy and dispersive. Waveguide junctions are assumed nonideal.

Fig. 6 shows the return loss response of the multiplexer at the start of the optimization process. The specification on the common port return loss is seriously violated, especially in the lower frequencies range (corresponding to channels 8-12).

The filter spacings are the dominant variables of the problem. This is based on the initial sensitivity analysis of the common port return loss function w.r.t. all variables at selected frequency points.

We perform the one-sided ℓ_1 optimization which is defined in the following way

$$\underset{\mathbf{x}}{\text{minimize}} \quad \sum_{i=1}^{m} |f_i^*(\mathbf{x})|, \tag{18}$$

where

$$f_i^* \triangleq \begin{cases} f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \geq 0 \\ 0 & \text{if } f_i(\mathbf{x}) < 0 \end{cases} \tag{19}$$

The functions $f_i(\mathbf{x})$ are the original error functions defined in (1).

We define also the gradients of the functions $f_i^*(x)$ in the following way

$$\frac{\partial f_i^*(\mathbf{x})}{\partial \mathbf{x}} \triangleq \begin{cases} \dfrac{\partial f_i(\mathbf{x})}{\partial \mathbf{x}} & \text{if } f_i(\mathbf{x}) \geq 0 \\ 0 & \text{if } f_i(\mathbf{x}) < 0 \end{cases} \tag{20}$$

The results of the ℓ_1 optimization defined above are shown in Fig. 7. The violations of the 20 dB specification are most serious in the frequency range corresponding to channels 1-2 and 8-12. This motivates us to release additional optimization variables in the filters corresponding to these channels. As additional optimization variables we release the input-output transformer ratios, cavity resonant frequencies as well as intercavity couplings. From that point minimax optimization is employed using the ℓ_1 optimized spacings as the starting values for the spacings. The final optimized return loss of the 12 channel multiplexer is shown in Fig. 8. The problem involves 60 nonlinear design variables.

VII. CONCLUSIONS

We have described a highly efficient algorithm for nonlinear ℓ_1 optimization problems. The algorithm combines linear programming methods with quasi-Newton methods and the convergence is at least superlinear.

Singular and regular ℓ_1 problems have been defined and a criterion for determining a singularity present in the ℓ_1 problem has been formulated.

The importance of the algorithm stems from the fact that the number of applications of the ℓ_1 norm to circuit and system problems has been increasing in recent years. The necessary conditions for optimality of the nonlinear ℓ_1 problem (see, e.g., [18]) indicate that zeros of the nonlinear functions $f_j(\mathbf{x})$ play an important role in the characteristics of the ℓ_1 problem. This fact has been used in fault isolation techniques for linear analog circuits and we have demonstrated that the new ℓ_1 algorithm is very successful in methods for fault isolation in linear analog circuits under an insufficient number of independent voltage measurements.

We have also shown that the algorithm is very reliable and implementations on different machines are possible, including microcomputers.

A formulation using the ℓ_1 norm for the initial stage of multiplexer design has been presented and illustrated by a 12 channel 12 GHz multiplexer. The one-sided ℓ_1 optimization sets to zero as many error functions as possible and this results in identifying channels of the multiplexer where the specification violations are most serious.

We feel that the properties of the ℓ_1 norm will be used more and more frequently in solving circuit and system problems, including diagnosis of networks, selection of tunable parameters in post-production tuning and model parameter identification from measurements.

ACKNOWLEDGEMENT

The authors are grateful to Mr. S. Pan of McMaster University for developing the TI/PC version of the ℓ_1 algorithm and for running the test problems. They also acknowledge the contribution of Mr. S. Daijavad of McMaster University to this research project. The assistance of Ms. M.L. Renault of McMaster University in preparing the numerical examples is also gratefully acknowledged.

REFERENCES

[1] R.H. Bartels and A.R. Conn, "An approach to nonlinear ℓ_1 data fitting", University of Waterloo, Computer Science Department, Report CS-81-17, 1981.

[2] J.W. Bandler, R.M. Biernacki, A.E. Salama and J.A. Starzyk, "Fault isolation in linear analog circuits using the ℓ_1 norm", Proc. IEEE Int. Symp. Circuits and Systems (Rome, Italy, 1982), pp. 1140-1143.

[3] J.W. Bandler and A.E. Salama, "Functional approach to microwave postproduction tuning", IEEE Trans. Microwave Theory Tech., vol. MTT-33, 1985, pp. 302-310.

[4] J. Hald and K. Madsen, "Combined LP and quasi-Newton methods for minimax optimization", Mathematical Programming, vol. 20, 1981, pp. 49-62.

[5] J. Hald, "A 2-stage algorithm for nonlinear ℓ_1 optimization", Report No. NI-81-03, Inst. for Num. Analysis, Tech. University of Denmark, 1981.

[6] J. Hald and K. Madsen, "Combined LP and quasi-Newton methods for nonlinear ℓ_1 optimization", SIAM J. on Numerical Analysis, vol. 22, 1985, pp. 68-80.

[7] J.W. Bandler, W. Kellermann and K. Madsen, "A superlinearly convergent minimax algorithm for microwave circuit design", to appear in the IEEE Trans. Microwave Theory Tech., 1985.

[8] M.R. Osborne and G.A. Watson, "On an algorithm for discrete nonlinear ℓ_1 approximation", The Computer Journal, vol. 14, 1971, pp. 184-188.

[9] R.A. El-Attar, M. Vidyasagar and S.R.K. Dutta, "An algorithm for ℓ_1-norm minimization with application to nonlinear ℓ_1-approximation", SIAM J. Numer. Anal., vol. 16, 1979, pp. 70-86.

[10] W. Murray and M.L. Overton, "A projected Lagrangian algorithm for nonlinear ℓ_1 optimization", SIAM J. Scient. Stat. Comp., vol. 2, 1981, pp. 207-224.

[11] R.A. McLean and G.A. Watson, "Numerical methods for nonlinear discrete ℓ_1 approximation problems", in Numerical Methods of Approximation Theory, Birkhauser Verlag, vol. 5, 1980, pp. 169-183.

[12] M.J.D. Powell, "The convergence of variable metric methods for non-linearly constrained optimization calculations", in Nonlinear Programming 3, eds. Mangasarian, Meyer, Robinson, Academic Press, 1978.

[13] K. Madsen and H. Schjaer-Jacobsen, "Singularities in minimax optimization of networks", IEEE Trans. Circuits and Systems, vol. CAS-23, 1976, pp. 456-460.

[14] R.H. Bartels, A.R. Conn and J.W. Sinclair, "Minimization techniques for piecewise differentiable functions: the ℓ_1 solution to an overdetermined linear system", SIAM J. Numer. Anal., vol. 15, 1978, pp. 224-241.

[15] K. Madsen, "An algorithm for minimax solution of overdetermined systems of nonlinear equations", J. Inst. Maths. Appl., vol. 16, 1975, pp. 321-328.

[16] G.A. Watson, "The minimax solution of an overdetermined system of nonlinear equations", J. Inst. Maths. Appl., vol. 23, 1979, pp. 167-180.

[17] H.M. Merrill, "Failure diagnosis using quadratic programming", IEEE Trans. Reliability, vol. R-22, 1973, pp. 207-213.

[18] C. Charalambous, "On conditions for optimality of the nonlinear ℓ_1 problem", Mathematical Programming, vol. 17, 1979, pp. 123-135.

[19] A.E. Atia, "Computer-aided design of waveguide multiplexers", IEEE Trans. Microwave Theory Tech., vol. MTT-22, 1974, pp. 332-336.

[20] M.H. Chen, F. Assal and C. Mahle, "A contiguous band multiplexer", COMSAT Technical Review, vol. 6, 1976, pp. 285-306.

[21] M.H. Chen, "A 12-channel contiguous band multiplexer at KU-band", 1983 IEEE Int. Microwave Symp. Digest (Boston, 1983), pp. 77-79.

[22] E.R. Egri, A.E. Williams and A.E. Atia, "A contiguous-band multiplexer design", 1983 IEEE Int. Microwave Symp. Digest (Boston, 1983) pp. 86-88.

[23] J.W. Bandler, S.H. Chen, S. Daijavad and W. Kellermann, "Optimal design of multi-cavity filters and contiguous-band multiplexers", Proc. 14th European Microwave Conference (Liege, Belgium), 1984, pp. 863-868.

[24] R. Tong and D. Smith, "A 12-channel contiguous band multiplexer for satellite application", 1984 IEEE Int. Microwave Symp. Digest (San Francisco, 1984), pp. 297-298.

APPENDIX A

NECESSARY CONDITIONS FOR A SOLUTION

At a solution $\mathbf{x}^*$ of the linearly constrained ℓ_1 problem (1) the functions which are zero play a special role since they contribute to the kinks of F. The functions which are non-zero at $\mathbf{x}^*$ give smooth contributions to F since $|f_j(\mathbf{x})|$ is smooth near $\mathbf{x}^*$ when $f_j(\mathbf{x}^*) \neq 0$. Therefore we partition F into a smooth and a non-smooth part,

$$F(\mathbf{x}) = \sum_{j \notin Z} |f_j(\mathbf{x})| + \sum_{j \in Z} |f_j(\mathbf{x})|$$

$$= g(\mathbf{x}) + \sum_{j \in Z} |f_j(\mathbf{x})|$$

where $Z = Z(\mathbf{x}^*)$ is defined by (9) and $g = g_{\mathbf{x}^*}$ is smooth in a neighbourhood of $\mathbf{x}^*$.

It is easily shown (see for instance Charalambous, [18]) that the following set of equations hold at the local minimum $\mathbf{x} = \mathbf{x}^*$

$$\begin{aligned} &\mathbf{g}'(\mathbf{x}) + \sum_{j \in Z} \delta_j \mathbf{f}_j'(\mathbf{x}) - \sum_{i \in A} \mu_i \mathbf{a}_i = \mathbf{0}\,, \\ &f_j(\mathbf{x}) = 0\,, \qquad\qquad j \in Z\,, \\ &\mathbf{a}^T\mathbf{x} + b_i = 0\,, \qquad\qquad i \in A\,, \end{aligned} \tag{A1}$$

where $|\delta_j| \leq 1$, $\mu_i \geq 0$, $Z = Z(\mathbf{x}^*)$ and $A = A(\mathbf{x}^*)$ are defined by (9) and (10), and

$$g(\mathbf{x}) = \sum_{j \notin Z} |f_j(\mathbf{x})|\,.$$

This set of equations corresponds to the Kuhn-Tucker conditions for the nonlinear programming problem which is equivalent to (1). The unknowns are $\mathbf{x}$, δ_j and μ_i, and it is seen that the number of unknowns equals the number of equations. If we use a vector notation (A1) can be expressed as follows,

$$\mathbf{R}(\mathbf{x}, \boldsymbol{\delta}, \boldsymbol{\mu}) = \mathbf{0}\,. \tag{A2}$$

APPENDIX B

UPDATING THE MATRICES J_k

The Jacobian of the nonlinear system (A1) is

$$\mathbf{R}'(\mathbf{x}, \boldsymbol{\delta}, \boldsymbol{\mu}) = \begin{bmatrix} \mathbf{g}''(\mathbf{x}) + \sum_{j \in Z} \delta_j \mathbf{f}_j''(\mathbf{x}) & \mathbf{E} & \mathbf{F} \\ \mathbf{E}^T & \mathbf{0} & \mathbf{0} \\ -\mathbf{F}^T & \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{B1}$$

where $\mathbf{E}$ and $\mathbf{F}$ are matrices with columns $\mathbf{f}_j'(\mathbf{x})$, $j \in Z$, and $-\mathbf{a}_i$, $i \in A$, respectively.

In Method 2 we need to find an estimate $\mathbf{J}_{k+1}$ to $\mathbf{R}'(\mathbf{x}^{k+1}, \boldsymbol{\delta}^{(k+1)}, \boldsymbol{\mu}^{(k+1)})$. This is done as follows. At the iterate $\mathbf{x}_k$ estimates Z_k and A_k replace Z and A. The submatrices $\mathbf{E}$ and $\mathbf{F}$ are calculated exactly using $\mathbf{f}_j'(\mathbf{x}_k)$, $j \in Z_k$, and $-\mathbf{a}_i$, $i \in A_k$.

Only the upper left hand side part of $\mathbf{R}'$ needs to be approximated. This is done via a modification of the BFGS method, due to Powell [12]. The modification is necessary because the upper left hand side of $\mathbf{R}'$ is not necessarily positive semidefinite at a solution of (1). However, it is more stable and not less efficient to keep the approximation positive definite. The updating procedure becomes

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k / [\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k] + \mathbf{z}_k \mathbf{z}_k^T / [\mathbf{s}_k^T \mathbf{z}_k]$$

with

$$\begin{aligned} \mathbf{s}_k &= \mathbf{h}_k , \\ \mathbf{z}_k &= \theta \mathbf{y}_k + (1 - \theta)\mathbf{B}_k \mathbf{s}_k , \qquad 0 < \theta \le 1 , \\ \mathbf{y}_k &= G(\mathbf{x}_k + \mathbf{h}_k , \boldsymbol{\delta}^{(k)} , \boldsymbol{\mu}^{(k)}) - G(\mathbf{x}_k , \boldsymbol{\delta}^{(k)} , \boldsymbol{\mu}^{(k)}) , \\ G(\mathbf{x} , \boldsymbol{\delta} , \boldsymbol{\mu}) &= \mathbf{g}'(\mathbf{x}) + \sum_{j \in Z_k} \delta_j \mathbf{f}_j'(\mathbf{x}) , \end{aligned} \tag{B2}$$

where θ is defined such that $\mathbf{s}_k^T \mathbf{z}_k > 0$ which implies that positive definiteness is maintained. Notice that when θ is close enough to 0 this inequality will hold provided $\mathbf{B}_k$ is positive

definite. Normally, however, θ can be chosen to be 1. Powell's formula for calculating θ is

$$\theta = \begin{cases} 1 & \text{if } \mathbf{s}_k^T \mathbf{y}_k \geq 0.2\, t_k \\ 0.8\, t_k / [\, t_k - \mathbf{s}_k^T \mathbf{y}_k] & \text{otherwise} \end{cases} \tag{B3}$$

with $t_k = \mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k$. We have found, however, that Powell's updating procedure becomes unstable when θ is too close to 0 and therefore we have modified (B3). If θ found by (B3) becomes less than 0.5 then we use $\theta = 0$. This implies that when (B3) gives $\theta < 0.5$ then $\mathbf{B}_{k+1} = \mathbf{B}_k$.

This completes the description of the Jacobian approximation procedure.

TABLE I

COMPARISON OF THE THREE IMPLEMENTATIONS OF THE ℓ_1 ALGORITHM

	CYBER		VAX		TI/PC	
Test Problem	NF(NS)	CPU	NF(NS)	CPU	NF(NS)	CPU
1	13(1)	0.14	11(1)	0.09	11(1)	2.2
2	61*(1)	0.29	57*(1)	0.23	57*(1)	5.4
3	8(1)	0.16	8(0)	0.15	8(0)	4.2
4	6(0)	0.16	6(0)	0.13	6(0)	3.5
5	22(1)	0.33	25(1)	0.34	25(1)	9.2
6	11(0)	1.37	11(0)	1.31	11(0)	43.1

NF - number of function evaluations

NS - number of shifts to Stage 2

CPU - execution time in seconds

* machine accuracy reached

TABLE II

RESULTS FOR THE MESH NETWORK EXAMPLE

Element	Nominal Value	Actual Value	Percentage Deviation		
			Actual	Case 1	Case 2
G_1	1.0	0.98	-2.0	0.00	0.13
G_2	1.0	0.50	-50.0*	-48.78	-49.44
G_3	1.0	1.04	4.0	0.00	3.60
G_4	1.0	0.97	-3.0	0.00	0.00
G_5	1.0	0.95	-5.0	-2.26	-1.71
G_6	1.0	0.99	-1.0	0.00	0.00
G_7	1.0	1.02	2.0	0.00	0.00
G_8	1.0	1.05	5.0	0.00	0.00
G_9	1.0	1.02	2.0	2.80	0.97
G_{10}	1.0	0.98	-2.0	0.00	0.00
G_{11}	1.0	1.04	4.0	0.00	0.00
G_{12}	1.0	1.01	1.0	3.45	2.08
G_{13}	1.0	0.99	-1.0	0.00	-0.44
G_{14}	1.0	0.98	-2.0	0.00	0.00
G_{15}	1.0	1.02	2.0	0.00	1.55
G_{16}	1.0	0.96	-4.0	-2.42	-5.71
G_{17}	1.0	1.02	2.0	0.00	2.67
G_{18}	1.0	0.50	-50.0*	-52.16	-48.94
G_{19}	1.0	0.98	-2.0	0.00	-1.95
G_{20}	1.0	0.96	-4.0	-3.67	-4.88
Number of Function Evaluations				8	8
Execution Time (secs) on Cyber 170/815				3.0	3.9

* Faults

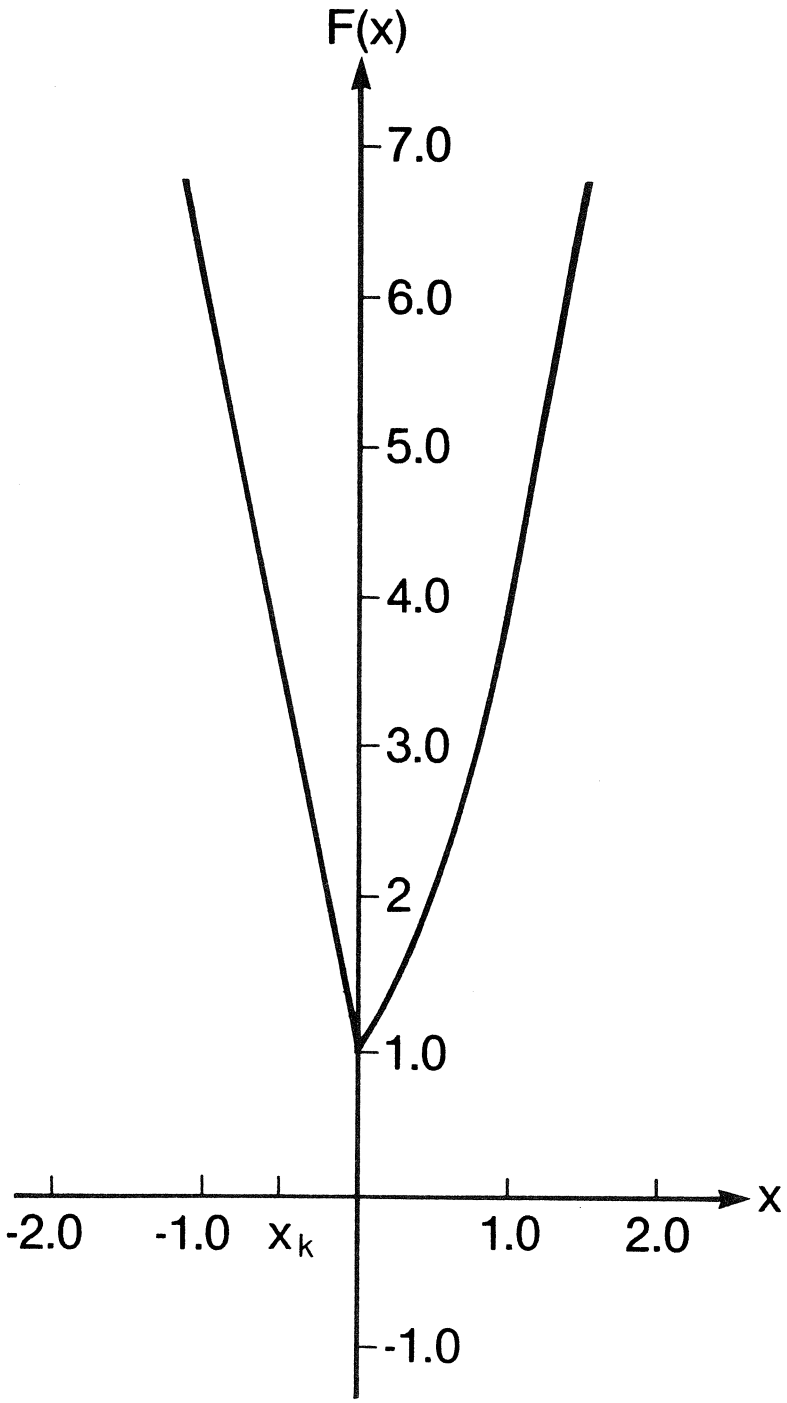

Fig. 1a The ℓ_1 objective function corresponding to one-dimensional problem with $f_1(x) = x^2 + 3x$ and $f_2(x) = x^2 - 2x + 1$.

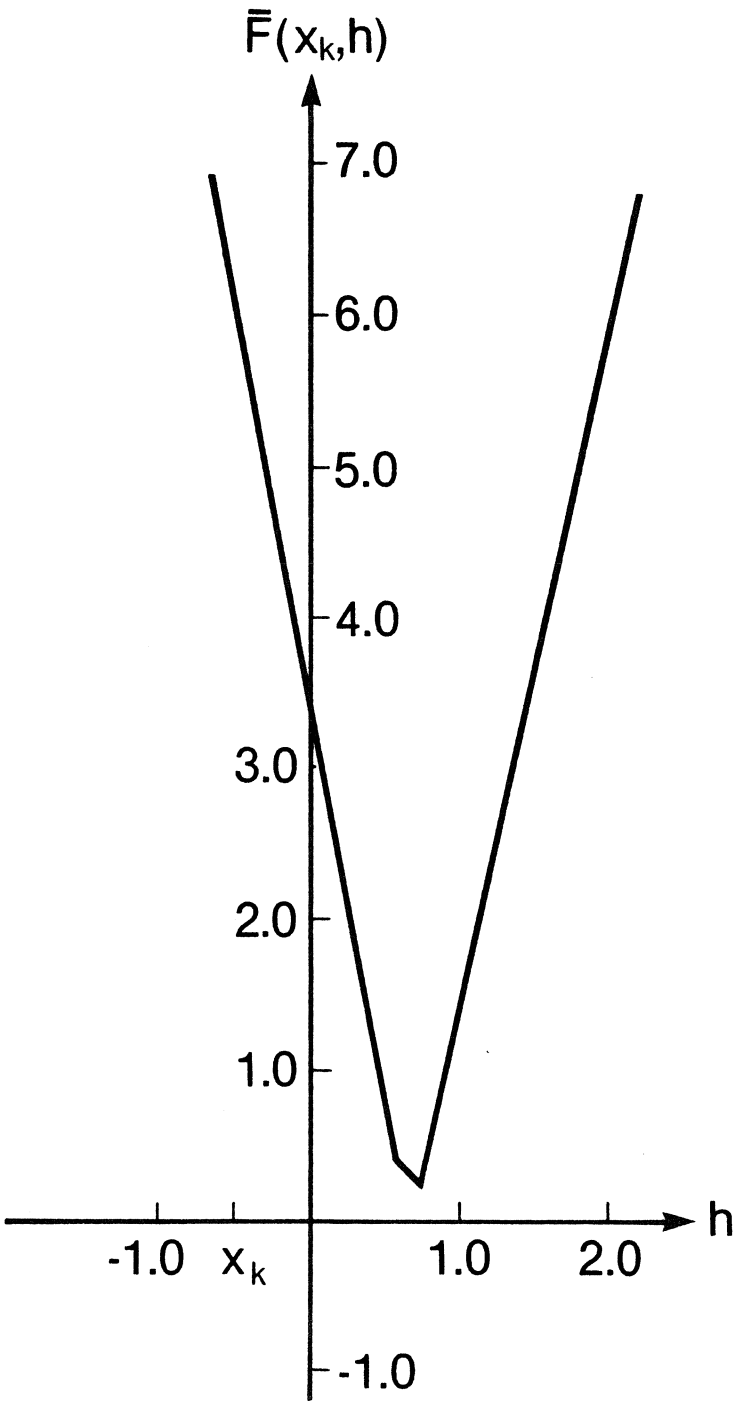

Fig. 1b The linearized model function for example (3) at $x_k = -0.5$, $\bar{F}(x_k, h) = |-1.25 + 2h| + |2.25 - 3h|$.

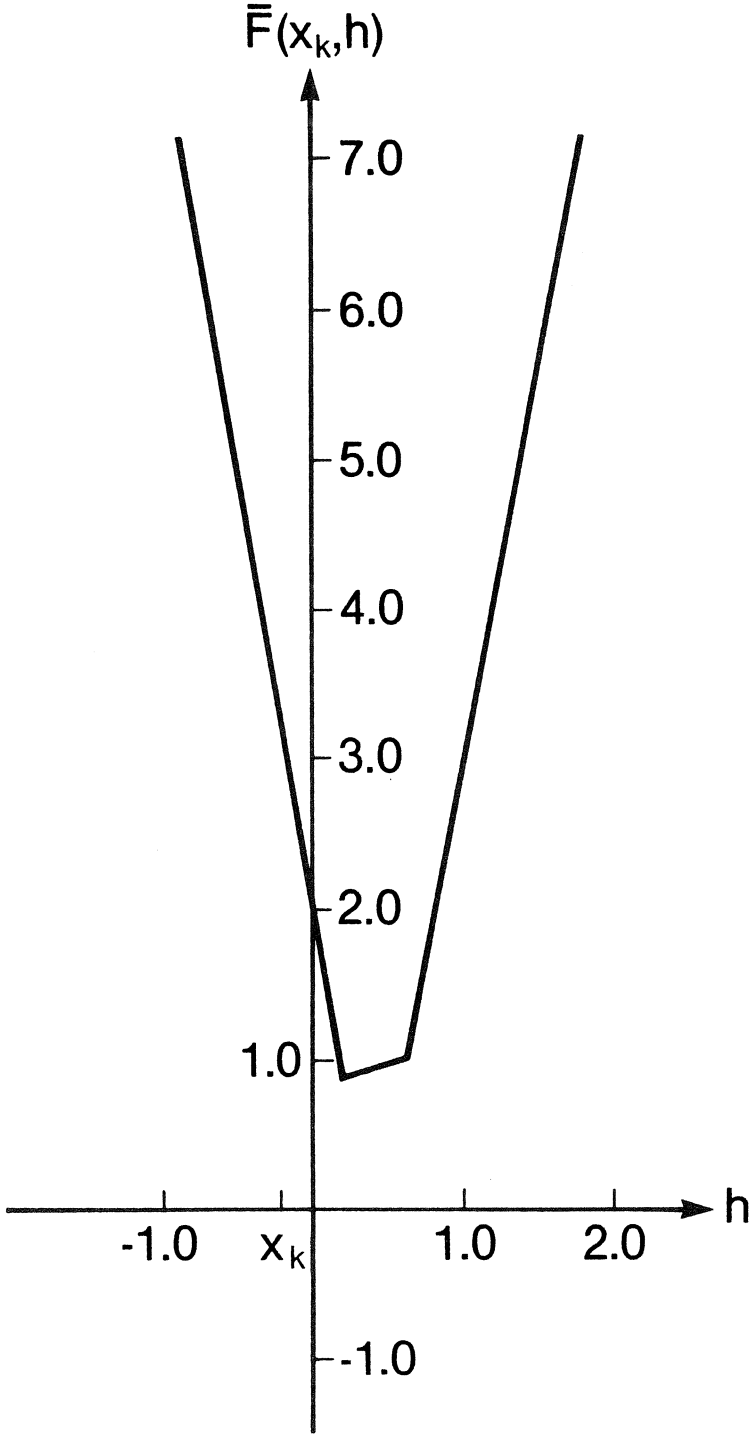

Fig. 2 The linearized model function for example (3) at $x_k = -0.2$, $\bar{F}(x_k, h) = |-0.56 + 2.6h| + |1.44 - 2.4h|$.

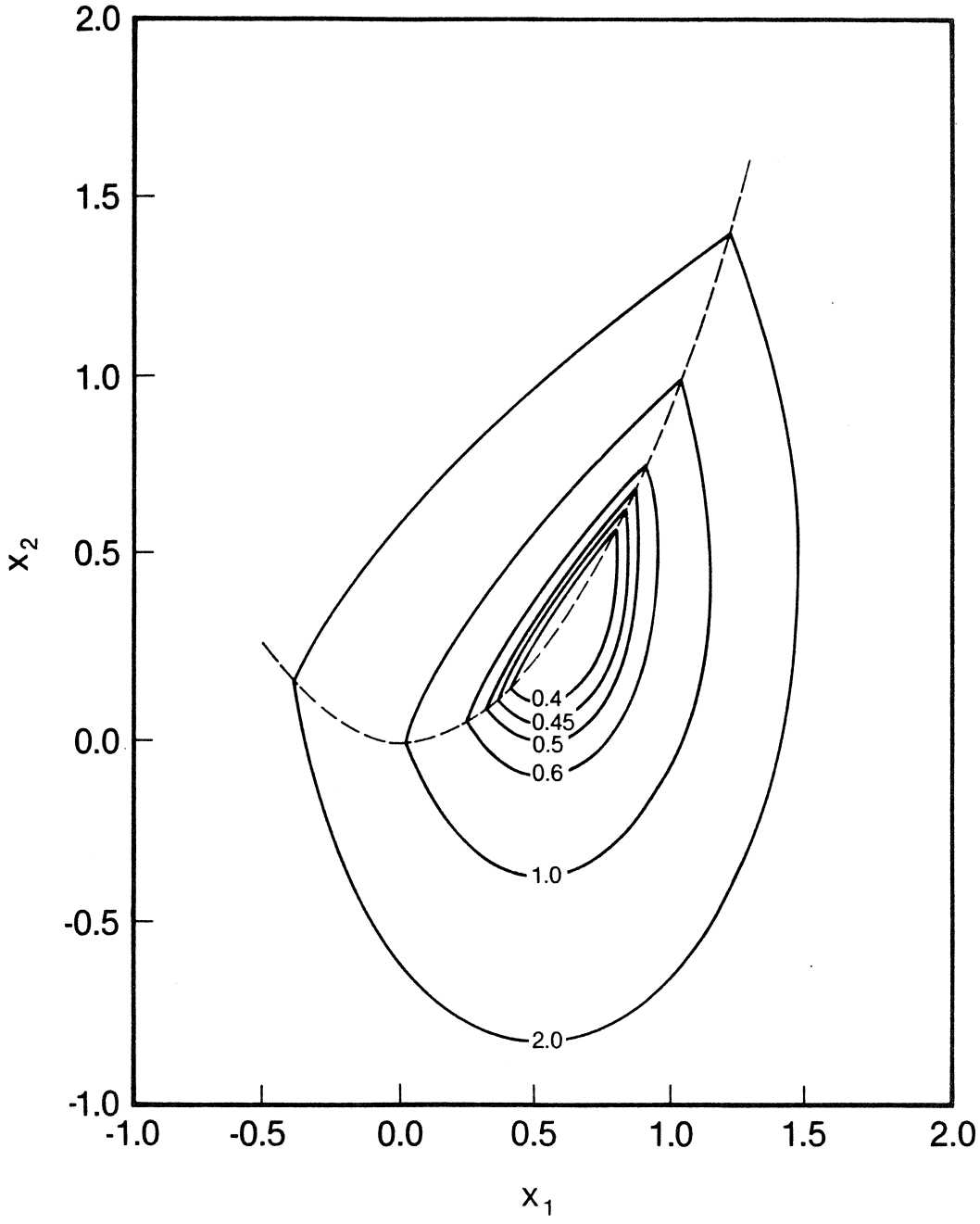

Fig. 3 Contours for the two-dimensional problem (5), $f_1(\mathbf{x}) = (x_1 - 1)^2 + x^2_2$, $f_2(x) = x^2_1 - x_2$.

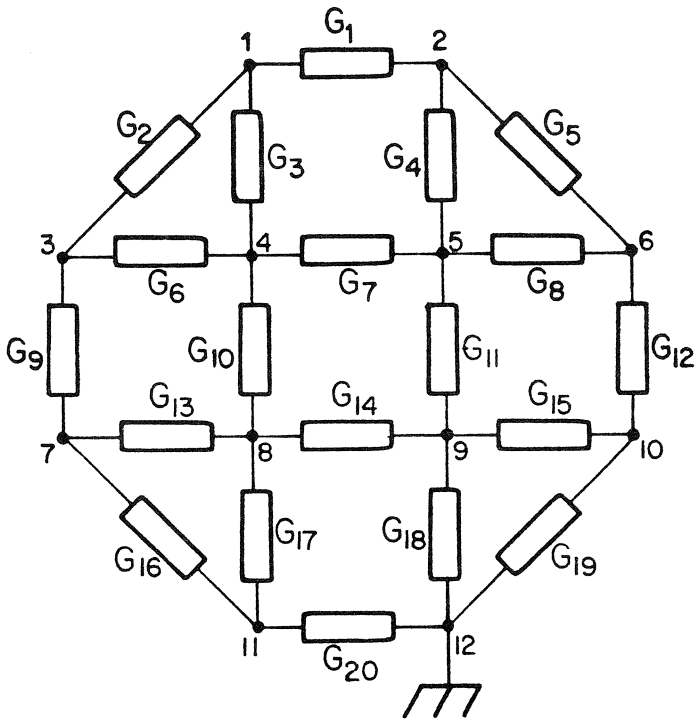

Fig. 4 The resistive mesh network.

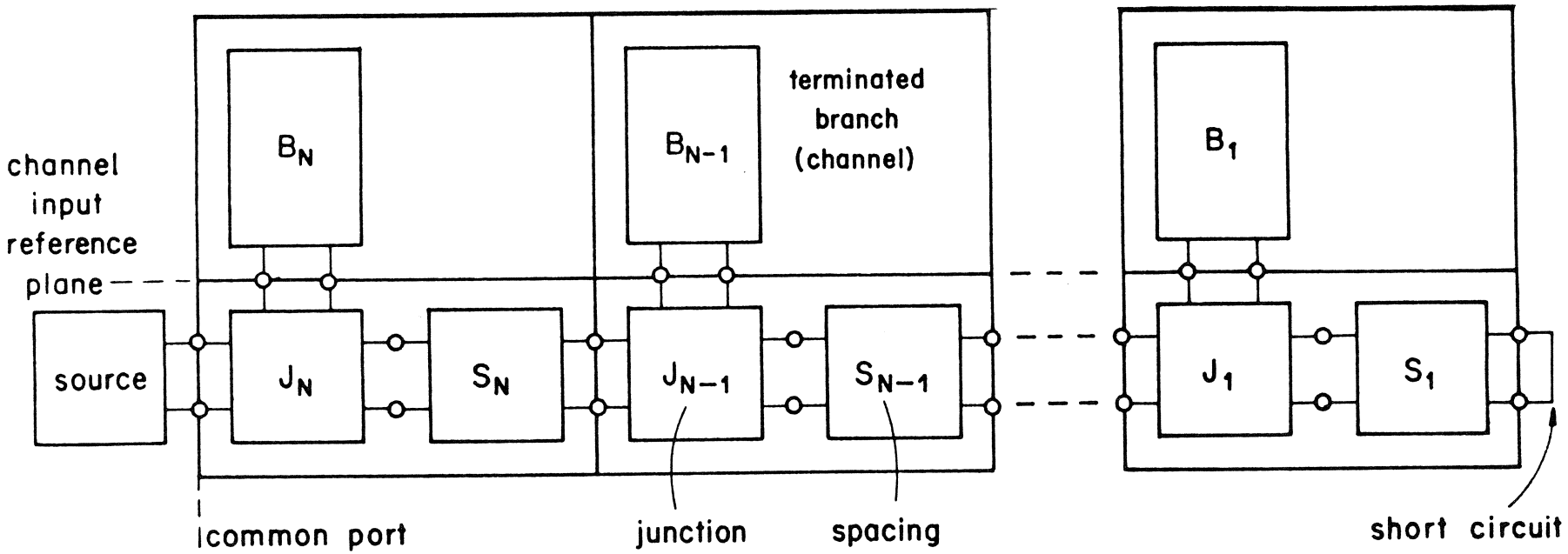

Fig. 5 The multiplexer configuration under consideration. $J_1, J_2, \ldots, J_N$ are arbitrarily defined 3-port junctions, $B_1, B_2, \ldots, B_N$ are terminated branches or channels which may each be represented in reduced cascade form and $S_1, S_2, \ldots, S_N$ are usually waveguide spacing elements.

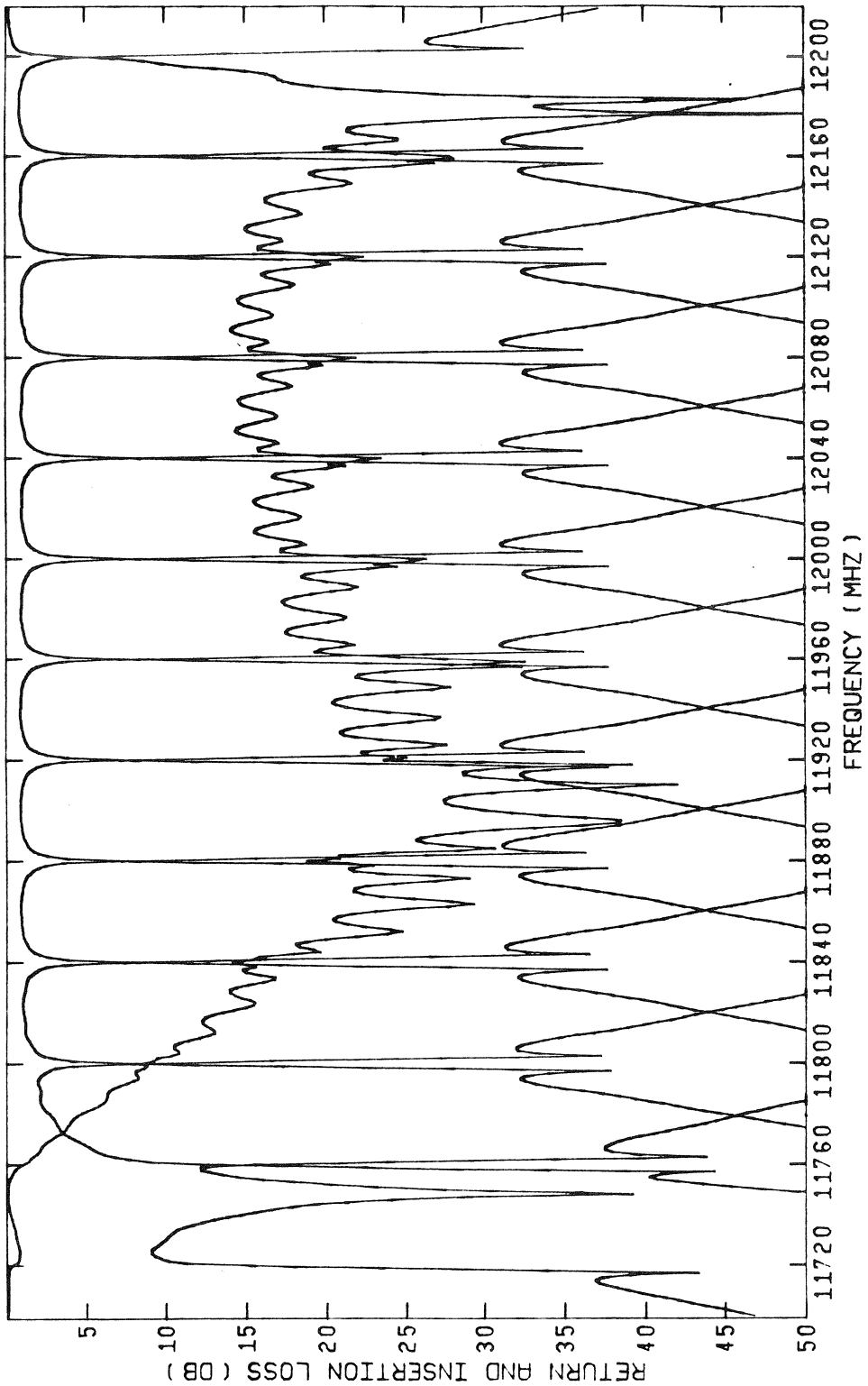

Fig. 6 Responses of the 12-channel multiplexer at the start of the optimization process.

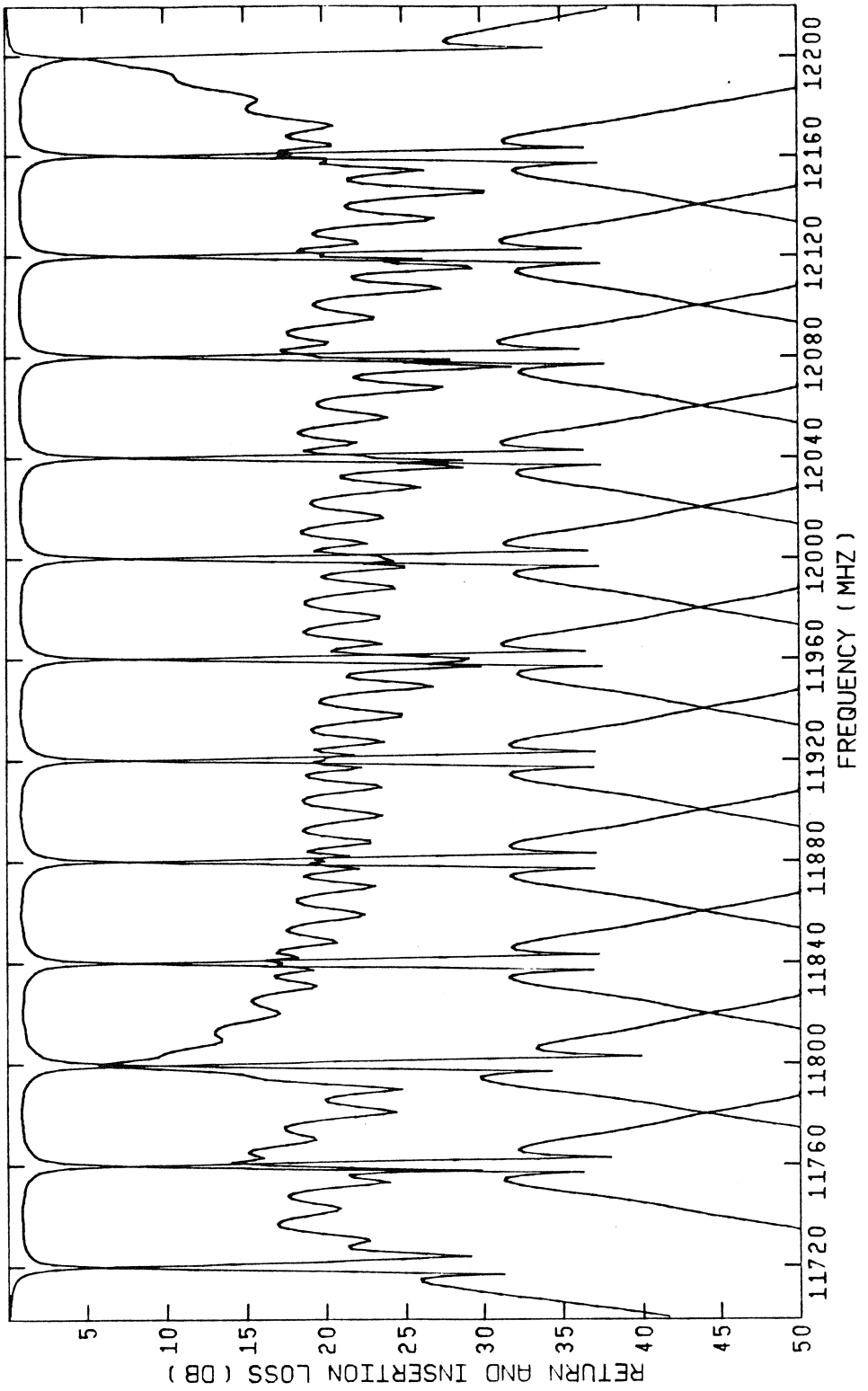

Fig. 7 Responses of the 12-channel multiplexer with optimized spacings only using ℓ_1 optimization.

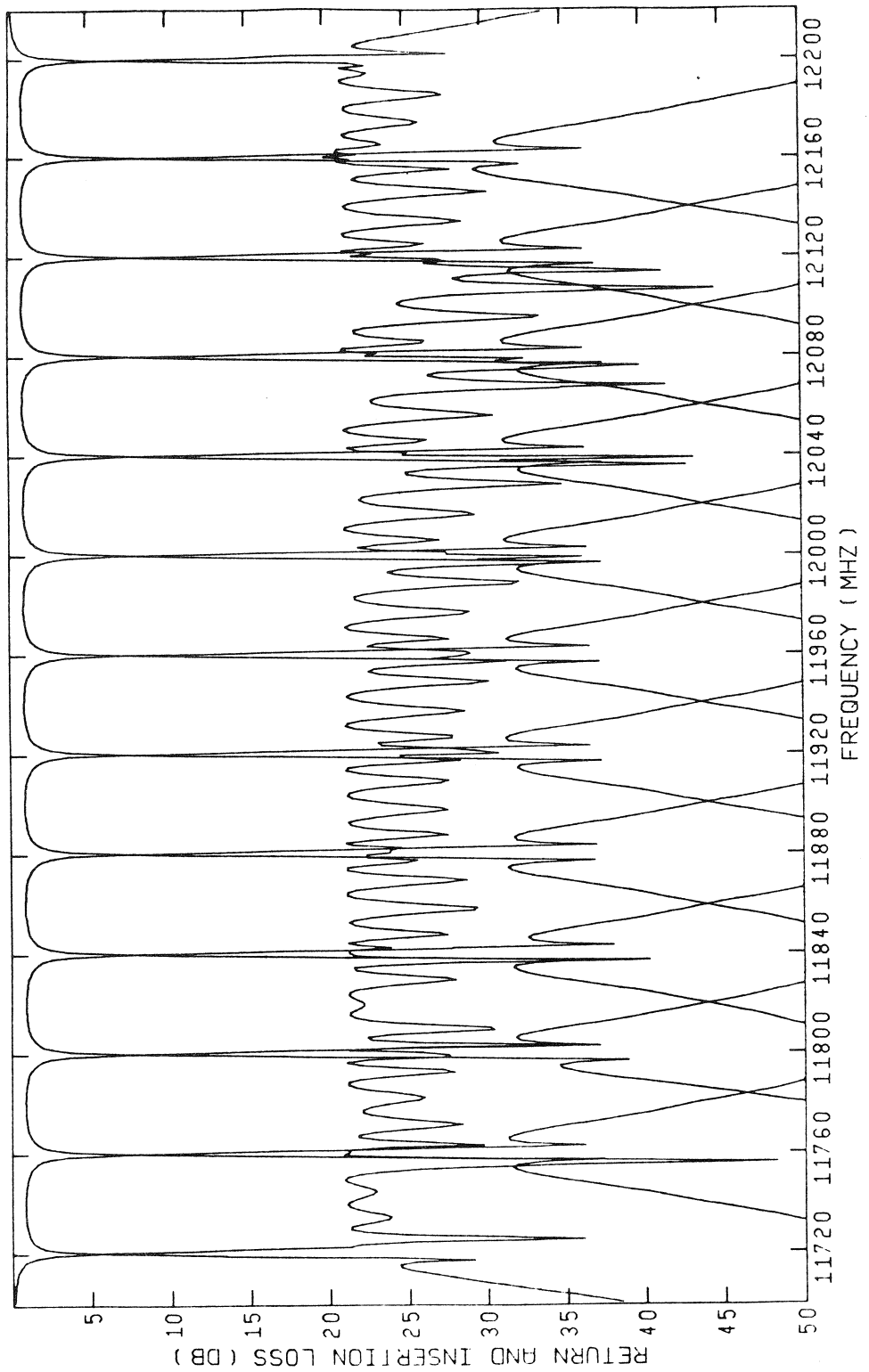

Fig. 8 Responses of the 12-channel multiplexer with optimized spacings, input-output transformer ratios, cavity resonances and coupling parameters using minimax optimization.