Reprinted from PROCEEDINGS OF THE IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS, San Jose, California, May 5-7, 1986

EFFICIENT GRADIENT APPROXIMATIONS FOR NONLINEAR OPTIMIZATION OF CIRCUITS AND SYSTEMS

J.W. Bandler*, S.H. Chen*, S. Daijavad* and K. Madsen**

*Simulation Optimization Systems Research Laboratory
and Department of Electrical and Computer Engineering
McMaster University, Hamilton, Canada L8S 4L7

**Institute for Numerical Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark

ABSTRACT

A flexible and powerful algorithm is proposed for efficient gradient approximations. It combines the techniques of perturbations, the Broyden update and special iterations. Utilizing this method, powerful gradient-based algorithms for nonlinear optimization of circuits and systems can be effectively employed without calculating exact derivatives. Examples of applications to fault location, worst-case tolerance design and design optimization are presented.

INTRODUCTION

Many powerful algorithms for nonlinear optimization have been developed and applied to circuit design problems, for example, the algorithms for linearly constrained ℓ_1 and minimax optimizations described by Bandler, Kellermann and Madsen [1], [2]. One difficulty in extending their practical applications, however, is that exact gradients of all functions with respect to all variables are usually required. For some applications, either an explicit expression of the exact gradients is not available or the computational labor for evaluating such gradients is prohibitive. Moreover, it is highly desirable to utilize many existing circuit simulation programs which provide only the values of the functions (or responses).

In this paper, we propose a flexible and powerful approach to gradient approximation for nonlinear optimization. It is a hybrid method which utilizes parameter perturbations (i.e., finite differencing), the Broyden update [3] and the special iterations of Powell [4]. Finite differencing requires one additional function evaluation to obtain the gradient with respect to each variable. It is the most reliable but also the most expensive method. The Broyden rank-one formula has been used in conjunction with the special iterations of Powell to update the approximate gradients, see, for example, Madsen [5] and Zuberek [6]. Such an update does not require additional function evaluations but its accuracy may not be satisfactory for some highly nonlinear problems or for a certain stage of the optimization. In our algorithm, parameter perturbations may be used to obtain an initial approximation and to provide regular corrections. The subsequent approximations are updated using the Broyden formula. Special iterations are introduced to improve the performance of the Broyden update. We also propose a modification of the Broyden formula which incorporates a knowledge, if available, of the structure of the Jacobian (e.g., the sparsity of the Jacobian).

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under Grants G1135 and A7239.

The hybrid method proposed is quite flexible in handling a large variety of problems, as proved through fault location of a resistive mesh network, worst-case tolerance design of a microwave amplifier and practical design of a 3-channel multiplexer involving 45 nonlinear variables. An interface is also developed for the gradient approximation module such that it is rather independent of the optimization technique and the circuit simulator. In the following sections, our algorithm is described. Its implementation in conjunction with an ℓ_1 algorithm and a minimax algorithm are illustrated.

GRADIENT APPROXIMATIONS

Method of Perturbations

A nonlinear optimization problem usually involves a set of, say, m nonlinear functions $f_j(\mathbf{x})$, $j = 1, ..., m$, where $\mathbf{x} = [x_1 \ldots x_n]^T$ is the vector of n variables.

The first-order derivative of $f_j(\mathbf{x})$ with respect to x_i can be approximated by

$$\frac{\partial f_j(\mathbf{x})}{\partial x_i} \approx \frac{f_j(\mathbf{x} + h\,\mathbf{e}_i) - f_j(\mathbf{x})}{h}, \qquad (1)$$

where $\mathbf{e}_i$ is a unit vector and h is the perturbation on x_i. An approximation of the Jacobian

$$\mathbf{G}(\mathbf{x}) \triangleq \left[\frac{\partial \mathbf{f}^T}{\partial \mathbf{x}}\right]^T$$

using perturbations requires $n+1$ evaluations of the functions $\mathbf{f}(\mathbf{x})$.

Broyden Update

Having an approximate Jacobian $\mathbf{G}_k$ at a point $\mathbf{x}_k$ and the function values at $\mathbf{x}_k$ and $\mathbf{x}_k + \mathbf{h}_k$, we can obtain $\mathbf{G}_{k+1}$ using the Broyden rank-one update [3]

$$\mathbf{G}_{k+1} = \mathbf{G}_k + \frac{\mathbf{f}(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}(\mathbf{x}_k) - \mathbf{G}_k \mathbf{h}_k}{\mathbf{h}_k^T \mathbf{h}_k} \mathbf{h}_k^T . \qquad (2)$$

The new approximation $\mathbf{G}_{k+1}$ provides a linearized model between two points $\mathbf{x}_k$ and $\mathbf{x}_k + \mathbf{h}_k$:

$$\mathbf{f}(\mathbf{x}_k + \mathbf{h}_k) - \mathbf{f}(\mathbf{x}_k) = \mathbf{G}_{k+1} \mathbf{h}_k . \qquad (3)$$

Notice that if $\mathbf{x}_k$ and $\mathbf{x}_k + \mathbf{h}_k$ are iterates of optimization the Broyden formula does not require additional function evaluations.

The application of the original Broyden update is not trouble free. As has been observed by Zuberek [6], if some functions are linear in some variables and if the corresponding components of $\mathbf{h}_k$ are nonzero, then the approximation to constant derivatives are updated by nonzero values. We have developed a method where the Broyden formula is applied to each $f_j(\mathbf{x})$ as a single function. Associated with f_j, a weighting vector is defined by

$$\mathbf{w}_j \triangleq [w_{1j} \ldots w_{nj}]^T , \qquad w_{ij} \geq 0 . \tag{4}$$

The approximation to $\mathbf{f}_j'(\mathbf{x})$ is then updated by

$$(\mathbf{f}_j')_{k+1} = (\mathbf{f}_j')_k + \frac{f_j(\mathbf{x}_k + \mathbf{h}_k) - f_j(\mathbf{x}_k) - (\mathbf{f}_j')_k^T \mathbf{h}_k}{\mathbf{q}_{jk}^T \mathbf{h}_k} \mathbf{q}_{jk} , \tag{5}$$

where

$$\mathbf{q}_{jk} \triangleq [w_{1j} h_{1k} \ldots w_{nj} h_{nk}]^T . \tag{6}$$

If f_j is linear in x_i, we set $w_{ij} = 0$. In circuit design problems, it may be known that the performance function is linear in or independent of some parameters over certain frequency or time intervals. It can be verified that an approximate Jacobian given by (5) also satisfies equation (3).

Special Iterations

The Broyden formula updates the approximate gradients along the direction $\mathbf{h}_k$. If the directions of some consecutive steps of optimization are collinear, the Broyden update may not converge. To cure this problem, Powell [4] suggested the method of "strictly linearly independent directions" generated by special iterations. Unlike an ordinary iteration where a step is taken in order to reduce the objective function, a special iteration is intended to improve the gradient approximation. After every p ordinary iterations the function values are calculated at a point obtained using the formula given by Powell [4] and a Broyden update is applied. We found that $p=2$ is satisfactory, which is also suggested by other authors (see, e.g., [4], [5] and [6]).

INTERFACING TO OPTIMIZATION ROUTINES

We have implemented our algorithm in a subroutine which calls a user-written routine (e.g., a simulator) for function values and calculates the approximate gradients required by a gradient-based optimization routine. It has the following features:

1. It is independent of and transparent to the optimizer and the simulator.
2. The user controls how frequently perturbations are used to obtain approximate gradients. Between these perturbations the gradient approximation will be updated using the Broyden formula and special iterations.
3. Some sophisticated optimization methods employ distinct stages of optimization. In this case, the user may prescribe different patterns of gradient approximation for different stages. For example, when it is close to a solution, approximate gradients of better accuracy may be desired, which can be achieved by using perturbations more frequently.
4. Any linearity and sparsity present in the sensitivity matrix can be exploited by assigning appropriate weightings to the Broyden update.

Typically, a gradient-based optimizer calls a user's routine when function values and derivatives are needed. A simple interface is to re-direct these calls to a routine which implements gradient approximation. However, it can be made more effective and efficient by suitable modifications to the optimization routine.

1. Assuming that exact derivatives are available, an optimization algorithm usually uses quite restrictive rules for accepting and bounding the increment of an iteration ($\mathbf{h}_k$ in eqn. (2)). These rules should be relaxed when the gradients are only approximate.
2. The optimization algorithm updates the gradients only when an increment (a trial point) is accepted. If we start with a very poor gradient approximation this may lead to a dead cycle. Actually even if a trial point fails, the function values at that point can and should still be used to improve the gradient approximation.

These modifications will not alter the essential body of an optimization algorithm but are necessarily algorithm-dependent.

EXAMPLES OF APPLICATIONS

Our method has been applied to two general-purpose algorithms for gradient-based nonlinear optimization. These two algorithms, as described in [1] and [2], employ a 2-stage combined LP and quasi-Newton method to solve linearly constrained ℓ_1 and minimax optimization problems, respectively.

ℓ_1 Optimization

Six problems of ℓ_1 optimization have been tested. The first one, due to Madsen [5], is a data-fitting problem involving 5 variables and 21 functions. The second one is a nonlinear ℓ_1 modelling problem, due to El-Attar et al. [7], of finding a third-order model for a seventh-order system involving 6 variables and 51 functions. The other four examples have been described by Bandler et al. (Example 1, 2, 5 and the mesh network example in [1]). The last example considers fault location of a resistive mesh network consisting of 20 elements. The two faulty elements deviate from their nominal values by 50 percent. Tolerances of 5 percent are associated with the other elements. Using measurements on the circuit with a single excitation, the actual faults have been identified.

The results are summarized in Table 1. In Case 1, parameter perturbations are conducted at every optimization iteration to approximate the gradients. This represents quite a traditional approach to gradient approximation. In Case 2, perturbations are used only for initialization. The subsequent approximations are updated by the Broyden formula and special iterations. Its advantage over Case 1 is clearly shown. The optimization programs that we used employ a quasi-Newton method to secure fast final convergence when a smooth valley is detected near a solution (namely, Stage 2 of [1]). The accuracy of the gradient approximation becomes crucial for such iterations. For Case 3 in Table 1, different updating schemes are used for two distinct stages. The approximate gradients are updated by the Broyden formula and special iterations for Stage 1, and perturbations are employed for every Stage 2 iteration where better accuracy is desired. The results show a similar

TABLE 1 ℓ_1 OPTIMIZATION WITH GRADIENT APPROXIMATIONS

Test Problem	Case 1	Case 2	Case 3
1	54(9)	32(19)	–
2	105(15)	63(40)	–
3	71(17)	65(48)	54(24)
4	98(32)	54(43)	58(26)
5	89(17)	51(38)	54(26)
6	147(7)	34(10)	–

Comments: The entries under each case are the number of function evaluations. The entries in parentheses are the corresponding numbers of optimization iterations.

number of function evaluations as Case 2 but the number of iterations is smaller. Such a variant can be achieved very conveniently with the flexibility of our algorithm.

Minimax Optimization

Two examples of practical circuit optimization are given here.

Worst-case tolerance design [8] is considered for a microwave amplifier consisting of an NEC70000 FET and five transmission lines (Fig. 1). The FET is characterized by tabulated scattering parameters. The design variables are the characteristic impedance Z and the lengths ℓ_i of the transmission lines. Assuming a 5 percent tolerance for each length ℓ_i, we seek an optimally centered design to best satisfy the specification given by 20 log $|S_{21}|$ = 7.625 dB between 6 GHz and 18 GHz. The approximate gradient at the starting point is used to predict the worst-case vertices. Working with these vertices a solution is obtained. If at the solution we detect any new worst-case vertices then they are added to the initial set. This procedure is repeated until the set of selected vertices is complete. The final solution is found to be $[\ell_1\ \ell_2\ \ell_3\ \ell_4\ \ell_5\ Z]$ = [69.01 152.01 18.48 5.095 36.49 126.39], which required 280 function evaluations. Fig. 2 shows the worst-case envelope at the solution. We have also solved the same problem with derivatives being calculated entirely by numerical differentiations, which required 585 function evaluations.

A large scale problem, namely a 12 GHz, 3-channel multiplexer involving 45 nonlinear variables [9], is also solved. A general multiplexer simulation program developed by Bandler et al. [9] is efficiently utilized through our interface to provide approximate gradients for the minimax optimizer. The minimax error functions are created using specifications on common port return loss and individual channel insertion losses, simulated multiplexer responses and weighting factors. The network parameters optimized include spacings, transformer ratios, cavity resonances and coupling coefficients. The network responses at the starting point and at the solution are shown in Figs. 3 and 4, respectively. To reach the solution 476 response evaluations are performed. The solution reported in [9] was obtained in more than 70 iterations, which would need more than 3000 response evaluations to provide the required exact gradient by numerical differentiations. The computational efficiency of the new algorithm is most pronounced in solving such complex problems.

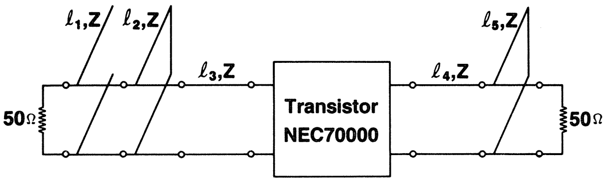

Fig. 1 A microwave amplifier.

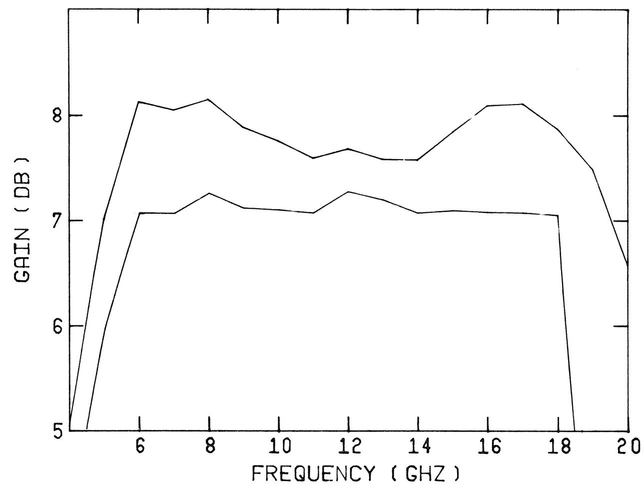

Fig. 2 The worst-case envelope for the amplifier response.

CONCLUSIONS

The computational efficiency and structural flexibility of a new algorithm for gradient approximation have been demonstrated. Implementation of this algorithm in integration with ℓ_1 and minimax optimization have been illustrated. Circuit examples of practical significance have been described. The utilization of this algorithm in conjunction with a vast variety of existing circuit simulators makes it effective and practical to take advantage of the powerful tool of gradient-based optimization in modern computer-aided design.

REFERENCES

[1] J.W. Bandler, W. Kellermann and K. Madsen, "A superlinearly convergent algorithm for nonlinear ℓ_1 optimization with circuit applications", Proc. IEEE Int. Symp. Circuits and Systems (Kyoto, Japan, 1985), pp. 977-980.

[2] J.W. Bandler, W. Kellermann and K. Madsen, "A superlinearly convergent minimax algorithm for microwave circuit design", IEEE Trans. Microwave Theory Tech., vol. MTT-33, 1985, pp. 1519-1530.

[3] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations", Math. of Computation, vol. 19, 1965, pp. 577-593.

[4] M.J.D. Powell, "A Fortran subroutine for unconstrained minimization, requiring first derivatives of the objective

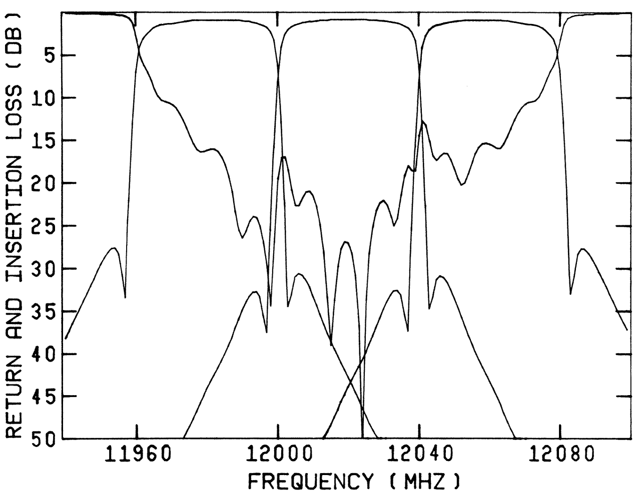

Fig. 3 Responses of the multiplexer at the starting point.

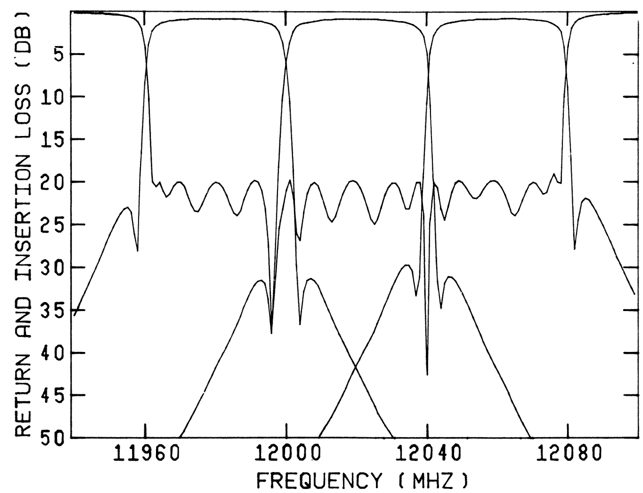

Fig. 4 Responses of the multiplexer at the solution.

functions", AERE Harwell, Oxfordshire, England, Report AERE-R.6469, 1970, pp. 20-27.

[5] K. Madsen, "Minimax solution of nonlinear equations without calculating derivatives", Mathematical Programming Study 3, 1975, pp. 110-126.

[6] W.M. Zuberek, "Numerical approximation of gradients for circuit optimization", Proc. 27th Midwest Symp. Circuits and Systems (Morgantown, WV, 1984), pp. 200-203.

[7] 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.

[8] J.W. Bandler, P.C. Liu and H. Tromp, "A nonlinear programming approach to optimal design centering, tolerancing and tuning", IEEE Trans. Circuits and Systems, vol. CAS-23, 1976, pp. 155-165.

[9] 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.