REPORT DOCUMENTATION PAGE

1. REPORT NUMBER
ARIO 18197.6-MA

2. GOVT ACCESSION NO.
N/A

3. RECIPIENT'S CATALOG NUMBER
N/A

4. TITLE (and Subtitle)
Trust Region Methods for Minimization

5. TYPE OF REPORT & PERIOD COVERED
Final Report
15 June 1981 - 14 June 1984

6. PERFORMING ORG. REPORT NUMBER
DAAG-81-K-0108

7. AUTHOR(s)
Robert B. Schnabel

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Dept. of Computer Science
University of Colorado, Boulder, CO

10. PROGRAM ELEMENT, PROJECT, TASK, AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

12. REPORT DATE
5 Oct. 1984

13. NUMBER OF PAGES
15

14. MONITORING AGENCY NAME & ADDRESS (IF DIFFERENT FROM CONTROLLING OFFICE)

15. SECURITY CLASS. (OF THIS REPORT)
Unclassified

15a. DECCLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (OF THIS REPORT)
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (OF ABSTRACT ENTERED IN BLOCK 20, IF DIFFERENT FROM REPORT)
NA

18. SUPPLEMENTARY NOTES
The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)
Unconstrained optimization, constrained optimization, trust region, tensor method.

20. ABSTRACT (CONTINUE ON REVERSE SIDE IF NECESSARY AND IDENTIFY BY BLOCK NUMBER)
(see reverse side)
This research project has investigated a number of topics in unconstrained optimization, constrained optimization, and solving systems of nonlinear equations. The biggest accomplishment has been the development of a new class of methods, called tensor methods, for solving systems of nonlinear equations. These methods have led to large increases in efficiency over standard methods on extensive batteries of test problems, with especially large gains on problems with singular Jacobians at the solution. The other major accomplishment has been the development of a unified theory of trust region methods for unconstrained optimization. Our theory shows how line search, dogleg, or optimal step methods can be constructed that satisfy first and second order conditions for convergence. Research has also been completed on conic methods for optimization, on secant methods that satisfy multiple secant equations, and on issues concerned with the computation of null space bases in constrained optimization. Research has been initiated on computational methods for nonlinear least squares problems with errors in the independent variables, and in parallel algorithms for optimization.
1. Introduction

Our research under this contract has mainly been concerned with developing, testing, and analyzing algorithms for numerical optimization problems. The standard continuous optimization problem is

$$\text{minimize } f : \mathbb{R}^n \rightarrow \mathbb{R}$$

subject to $c_i(z) \leq 0, \ i = 1, \ldots, p$

$$c_i(z) = 0, \ i = p + 1, \ldots, p + m$$

where each $c_i : \mathbb{R}^n \rightarrow \mathbb{R}$ and $m \leq n$. Most of our research has concerned the unconstrained optimization problem, where $p = m = 0$, and the closely related systems of nonlinear equations problem,

given $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$, find $z^* \in \mathbb{R}^n$ such that $F(z^*) = 0$.

Some of our research has been concerned with the linearly constrained optimization problem, where the constraints $c_i$ are all linear functions of $z$, or with the nonlinearly constrained optimization problem where the constraints are nonlinear.

The main topics we have addressed during this research period are:

2. Trust region methods for unconstrained and constrained optimization.
4. Multiple secant methods for nonlinear equations and unconstrained optimization.
5. Unconstrained optimization software.
7. Nonlinear least squares with errors in independent variables.
8. Parallel algorithms for numerical optimization.

Our findings on each of these topics is discussed briefly in Sections 2 - 9.
first two topics have been the most important and have consumed a large portion of our attention during this research period, however each of these topics has produced or is in the process of producing at least one technical publication. Our work on topics 3, 4, 5, and 6 is essentially complete, while our work on topics 1, 2, and 7 is continuing and forms the basis of our new ARO contract DAAC29-84-K-0140, which started on August 1, 1984. We have conducted preliminary work on topic 8 in the past year and anticipate this will become one of our major research interests in the future.

A list of publications and technical reports produced during the research period is given in section 10. Technical papers currently in progress that report research conducted under this contract are also listed at the end of section 10. Scientific personnel supported under this contract and theses produced are reported in Section 11. Invited and contributed presentations at scientific meetings that have reported research conducted under this contract are reported in Section 12.

2. Tensor methods for nonlinear equations and unconstrained optimization

Our most significant accomplishment during the research period supported by this contract has been the development of a new class of methods, called tensor methods, for solving systems of nonlinear equations. Tensor methods are general purpose methods for solving systems of nonlinear equations that are especially intended for the case when the Jacobian matrix at the solution is singular or ill-conditioned. They base each iteration on a second order model of the system of equations around the current iterate $x_c$,

$$M(x_c + d) = F(x_c) + J_c d + \frac{1}{2} T_c d d$$

that includes an approximation $T_c$ to the three dimensional second derivative tensor. Their most important and novel feature is the ability to form and solve such a model at each iteration without appreciably increasing the computational
overhead of the algorithm above standard methods based on linear models. This has led to computationally efficient methods that demonstrate considerable computational advantages over standard methods. We have developed two classes of tensor methods, one for the case when the Jacobian matrix is available analytically or by finite differences, the second for the case when the Jacobian matrix is approximated by a secant approximation. We are just beginning to consider tensor methods for unconstrained optimization.

Tensor methods for nonlinear equations that use first derivatives are described by Schnabel [1982] and Schnabel and Frank [1984], as well as in the M.S. and Ph.D. theses by Frank [1982, 1984]. These methods choose the second derivative term to allow the local model at each iteration to interpolate a small number \( p \) of past function values (vectors), as well as the current function value and Jacobian. They then choose the smallest second derivative approximation, in the Frobenius norm, that satisfies the resultant interpolation conditions; this is shown to be a rank \( p \) tensor. Given this choice and the restriction \( p < \sqrt{n} \), the additional cost of forming the tensor model is shown to be \( O(n^2p) \), and the additional cost of solving the tensor model, the solution of a \( p \times p \) system of quadratic equations, is smaller still. Thus the total additional cost is unimportant in comparison to the \( O(n^2) \) cost of solving a system of linear equations at each iteration. In extensive computational tests, the new tensor method required fewer function evaluations and iterations than a comparable method based on the standard linear model on 77% of the test problems, the same on 14%, and more on 7%, with an average savings of 23% on nonsingular problems and 20-40% on singular problems. On problems where either method required 10 or more iterations, the savings were even more dramatic. In addition, the tensor method solved more problems than the standard method. Based on these results, the major libraries are considering incorporating a tensor method for nonlinear equations. Frank [1984] has shown that the tensor method is 3-step
superlinearly convergent on problems where the Jacobian at the solution has rank \( n-1 \).

The extension of tensor methods for nonlinear equations to the case when the Jacobian matrix is also approximated from function values (secant methods) is considered in the Ph.D. thesis of Frank and in an upcoming paper by Frank and Schnabel. This case is considerably more difficult because the same past function information must be used to derive first and second derivative approximations. We have successfully shown how this can be done, again without appreciably increasing the computational overhead. In extensive computational tests, the new secant tensor method again displays steady gains over a standard Broyden's method implementation such as the one in MINPACK; the gains are roughly 10% on nonsingular problems and 25-33% on singular problems.

We have just begun to consider tensor methods for unconstrained optimization. These methods will augment the standard quadratic model by low rank approximations to the third and fourth derivative tensors. (The fourth derivative is necessary to obtain improvement on singular problems.) We have devised a preliminary approach in both the derivative and secant cases. Development of these methods is a primary objective of our new ARO contract.

3. Trust region methods for unconstrained and constrained optimization

The other main topic we have pursued under this research contract is the development of trust region methods for a variety of optimization problems. Trust region methods are one of the leading approaches for solving optimization problems from far starting points, and have been shown by Sorensen [1982] and Moré and Sorensen [1981] to have strong local and global convergence properties. Our main objective was to show that the convergence properties attained by Moré and Sorensen for the "optimal-step" algorithms could be achieved by a
wide variety of strategies, including dogleg methods and line searches, when cast in a trust region framework. This objective is completely satisfied in Shultz, Schnabel, and Byrd [1984], where we exhibit a general set of step-acceptance conditions under which a wide variety of trust region algorithms have the same strong global and local convergence properties. We also propose a new dogleg algorithm that is the first to remain applicable when the model Hessian matrix is indefinite. Test results using this algorithm are reported in the Ph.D. thesis by Shultz [1983]; the new algorithm is competitive with existing optimal-step algorithms. Shultz' thesis also derives a class of trust region algorithms for linearly constrained optimization with similar convergence properties.

So far there has been little reported success in deriving elliptical trust region algorithms for nonlinearly constrained optimization, but our work on implicitly defined nonlinear least squares problems has developed a need for such methods in the case of equality constraints only. We have begun work on developing such algorithms and appear to be meeting with success; this work will be an important component of our research under our new ARO contract.

4. Conic methods for unconstrained and constrained optimization

Conic methods for optimization are an approach introduced by Davidon [1980] and further studied by Sorensen [1980] where each iteration is based upon a conic model of the objective function around the current iterate \( x_c \).

\[
m(x_c + d) = f(x_c) + \frac{\nabla f(x_c)^T d}{1 + b_c^T d} + \frac{\nabla d^T A_c d}{1 + b_c^T d}
\]

Davidon showed how to use such a model to interpolate the function value and gradient at the previous as well as the current iteration, and Sorensen tested such a method and showed that it retained the local superlinear convergence of standard secant methods. We have been interested in the more fundamental question of how such a model can be used when analytic (or finite difference)
second derivatives are available. This question has been the subject of M.S. theses by Stordahl [1980] and Back [1984], and is also reported in Schnabel [1982] and in an upcoming paper by Schnabel and Back. We have studied several ways in which the conic model can interpolate the current function value, gradient, and Hessian, as well as information from previous iterates. Our recent work with Back uses the conic model to interpolate the gradient at the previous iterate as well as the above mentioned information. In extensive tests, the resultant method shows a small gain in efficiency over standard methods, but probably not enough to justify incorporating the conic model into software libraries. After several years of study, our opinion is that conic models will not lead to large improvements in unconstrained optimization algorithms. We have also considered using a conic function in the line search of a barrier penalty method; here the conic led to substantial savings and this may be a good use for conics. This work is reported in Back [1984] and will be part of the paper by Schnabel and Back.

5. Multiple secant methods for nonlinear equations and unconstrained optimization

We have long been interested in the subject of secant methods for nonlinear equations and unconstrained optimization that satisfy multiple secant equations; see e.g. Gay and Schnabel [1978]. Our interest in this problem was rekindled by our work on secant tensor methods that required Jacobian updates that satisfy multiple secant equations, i.e., approximations $J_c$ that satisfy several equations of the form $J_c s_1 = y_1$. These equations arise whenever the secant method attempts to interpolate multiple past function vectors. In Schnabel [1983], we give a new treatment of this question for both nonlinear equations and unconstrained optimization. For nonlinear equations, we give a much cleaner presentation of the algorithms and convergence results in Gay and
Schnabel [1978]; in particular, the new algorithm was important in the implementation of secant tensor methods in Frank's Ph.D. thesis. We then give the first clear and concise statement of when multiple secant equations for optimization are consistent with the requirement that the Hessian approximation be symmetric. When these conditions are satisfied, we show that there are natural generalizations of the standard update formulas that satisfy multiple secant equations. Finally, we propose several new algorithms that use multiple secant updates to come as close as possible to interpolating multiple past gradients, and show that these algorithms retain local superlinear convergence. These new algorithms have not yet been tested.

6. Unconstrained optimization software

In conjunction with the preparation of the book by Dennis and Schnabel [1983], we have developed a modular system of algorithms and computer software for unconstrained optimization. This software is novel in that it allows the user to select from any one of three global strategies (line search, dogleg, and optimal step) as well as a wide variety of options for calculating or approximating derivatives. The software was written as part of the M.S. thesis by Weiss [1980]; in the interim, we have tested it extensively and added a reverse communication capability. The entire package, and comparative test results obtained using it, are described in Schnabel, Weiss, and Koontz [1982]. We have used this package as the basis of most of our test codes described above, for example the tensor, conic, and new trust region methods, and it continues to be an important resource for the quick development of high quality test software.

7. Null space bases for constrained optimization

Many recent constrained optimization algorithms use a basis for the null space of the matrix of constraint gradients. The convergence proofs for these
methods often assume that this basis changes smoothly with the variable \( z \), assuming that the constraints are continuous functions. A recent paper by Coleman and Sorensen [1984] has pointed out, however, that standard methods for computing the null space basis from the QR factorization do not have this continuity property. This paper interested us and our colleague Richard Byrd to explore the question of whether such a continuous null space basis is, in fact, possible. In a very recent paper, Byrd and Schnabel [1984], we report that results from topology show that, in general, there is no continuous function that generates the null space basis for all full rank rectangular matrices of a fixed size. Thus constrained optimization algorithms cannot assume an everywhere continuous null space basis. We also propose an alternative implementation of a class of constrained optimization algorithms that uses approximations to the reduced Hessian but is independent of the choice of null space basis. This approach obviates the need for a continuously varying null space basis. We plan to test these new methods as part of the research under our new ARO contract.

8. Nonlinear least squares with errors in independent variables

The nonlinear least squares problem is a special case of unconstrained optimization where the objective function is the sum of squares of nonlinear functions. It arises most commonly in fitting the data \((t_i, y_i), i = 1, \ldots, m\) to the nonlinear model \(m(x, t)\); the parameters \(x\) that "best" fit the data are calculated by solving the nonlinear least squares problem

\[
\minimize_{x \in \mathbb{R}^n} \sum_{i=1}^{m} (m(x, t_i) - y_i)^2.
\]

This problem arises from the assumption that the data would fit the model exactly for some value of the parameters \(x\) were it not for measurement errors in the dependent variables \(y_i\). It is implicitly assumed that there are no measurement errors in the independent variables \(t_i\), and usually this is a reasonable
assumption in practice (or at least, the errors in \( t \) are very small in comparison to the errors in \( y \)). Sometimes, however, the measurement errors in the independent variables are significant in comparison to the errors in the dependent variables, and in this case it is helpful to calculate the parameters \( x \) by solving the expanded problem

\[
\minimize_{x \in \mathbb{R}^n, \delta \in \mathbb{R}^m} \sum_{i=1}^m \left( m(x, t_i + \delta_i) - y_i \right)^2 + \sigma^2 (\delta_i)^2
\]

where \( \sigma \) is an estimate of the variance in measuring \( y \) to the variance in measuring \( t \). This is another nonlinear least squares problem but the number of variables is far larger because usually \( m \gg n \). Together with Paul Boggs of the National Bureau of Standards, we have been developing an algorithm and software that solves the above problem with hardly more expense than solving the standard problem. The first report on this work currently is in preparation.

We are also considering the generalizations of both problems where the model function is an implicitly defined function of \( y \) as well as \( t \); in this case both problems become nonlinearly equality constrained optimization problems. The techniques discussed in Section 3 will be applied to this problem as part of our new ARO contract.

9. Parallel algorithms for numerical optimization

The emerging availability of a variety of parallel computer architectures opens up many new possibilities for numerical computation. We have spent some of our research time under this contract considering the opportunities for parallelism in optimization; our thoughts are contained in the recent paper Schnabel [1984]. We are particularly interested in using a loosely coupled network of computing devices, such as a network of workstations, to solve global optimization and related problems; our preliminary ideas are related in the above-mentioned paper. We have spent some time in the last year demonstrat-
ing the feasibility of using a network of Sun workstations running Berkeley Unix
4.2 and connected by an Ethernet to develop and run distributed programs;
these efforts have been quite successful. We anticipate devoting an increasing
amount of our research effort to this area in the future.

10. Publications and technical reports produced

Publications:

(1) "Determining feasibility of a set of nonlinear inequality constraints.,"

(2) "Comments on evaluating algorithms and codes for mathematical program-
    ming," in Evaluating Mathematical Programming Techniques, J.M. Mulvey,

(3) "Unconstrained optimization in 1981," in Nonlinear Optimization 1981,

(4) "Forcing sparsity by projecting with respect to a nondiagonally weighted
    L. Toint).

(5) Numerical Methods for Unconstrained Optimization and Nonlinear Equa-

(6) "Conic methods for unconstrained minimization and tensor methods for
    nonlinear equations," in Mathematical Programming, The State of the Art,
    417-438.

(7) "A family of trust region based algorithms for unconstrained minimization
    with strong global convergence properties," SIAM Journal on Numerical


Submitted for publication:


Additional forthcoming papers concerning research performed under this contract:

(13) "Secant tensor methods for nonlinear equations" (with P.D. Frank).

(14) "Conic methods using second derivatives for optimization" (with R. Back).

(15) "Nonlinear least squares with errors in independent variables (with P.T. Boggs).

11. List of scientific personnel

P.D. Frank, research assistant, June 1981 - June 1984

M.S., 1982; Thesis: "A second-order local model for solution of systems
of nonlinear equations."

Ph.D., 1984; Thesis: "Tensor methods for solving systems of nonlinear equations."

Additional theses supervised in conjunction with research for this contract:

G.A. Shultz, Ph.D., 1983; Thesis: "Computationally practical globally convergent algorithms for unconstrained and linearly constrained optimization."


J.R. Donaldson, M.S., 1984 (expected); Thesis: "Confidence regions and intervals for nonlinear least squares."

12. Presentations at scientific meetings of research conducted under this contract


(2) "Tensor models in nonlinear equations," SIAM national meeting, Stanford, California, July 1982.


13. References


P. D. Frank [1982], "A second-order local model for solution of systems of nonlinear equations", M.S. Thesis, Department of Computer Science, University of Colorado at Boulder.

P. D. Frank [1984], "Tensor methods for solving systems of nonlinear equations", Ph.D. Thesis, Department of Computer Science, University of Colorado at


