MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	30 Aug 1994	Final 15 may 91 - 30 Jun 94

4. TITLE AND SUBTITLE

New Methods for Nonlinear Optimization

5. FUNDING NUMBERS

DAAL03-91-G-0151

6. AUTHOR(S)

Robert B. Schnabel and Richard H. Byrd

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Colorado
Department of Computer Science
Boulder, CO 80309

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

ARO 28701.10-MA

11. 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.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Over the course of this research grant, considerable progress was made in all the areas discussed in the proposal, namely limited memory methods for problems with constraints, tensor methods for large sparse nonlinear problems and for constrained optimization, and trust regions methods for nonlinearly constrained optimization. In addition, substantial progress was made in the development of large scale global optimization methods for molecular configuration problems, a topic supported in part by other agencies but one in which ARO has expressed considerable interest as well. We summarize the work in these areas in Sections 1-4. In addition, we have worked on several other topics, including symmetric-rank one update methods for unconstrained optimization, implementations of the linear algebraic operations of the BFGS method on sequential and parallel computers, and parallel methods for solving block bordered systems of nonlinear equations. We summarize this work very briefly in Section 5. Section 6 contains a listing of publications and reports supported by this grant, and Section 7 contains a list of research personnel supported by this grant.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Nonlinear optimization, limited memory methods, tensor methods, trust region methods, global optimization	11
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

FINAL REPORT for

ARO Grant DAAL03-91-G-0151
New Methods for Nonlinear Optimization
May 15, 1991 - June 30, 1994

Principal Investigator: Robert B. Schnabel
Co-Principal Investigator: Richard H. Byrd

Over the course of this research grant, considerable progress was made in all the areas discussed in the proposal, namely limited memory methods for problems with constraints, tensor methods for large sparse nonlinear problems and for constrained optimization, and trust regions methods for nonlinearly constrained optimization. In addition, substantial progress was made in the development of large scale global optimization methods for molecular configuration problems, a topic supported in part by other agencies but one in which ARO has expressed considerable interest as well. We summarize the work in these areas in Sections 1-4. In addition, we have worked on several other topics, including symmetric-rank one update methods for unconstrained optimization, implementations of the linear algebraic operations of the BFGS method on sequential and parallel computers, and parallel methods for solving block bordered systems of nonlinear equations. We summarize this work very briefly in Section 5. Section 6 contains a listing of publications and reports supported by this grant, and Section 7 contains a list of research personnel supported by this grant.

1. Limited Memory Methods for Problems with Constraints.

Our research on limited memory quasi-Newton methods has continued and moved in some new directions. Limited memory methods work by generating a quasi-Newton approximation to the Hessian of the objective function that uses only the most recent updates, resulting in great savings in storage, independent of whether the Hessian matrix is sparse or its sparsity pattern. Thus they are an important approach to very large optimization problems where the number of variables is too large to allow a full Hessian approximation to be stored.

Our work on a new compact, closed form representation of limited memory quasi-Newton matrices has been essentially completed. A paper describing this work by the principal investigators together with Jorge Nocedal of Northwestern University was completed during this research period has appeared in *Mathematical Programming*. This representation is a important part of our extensions of the limited memory approach to constrained optimization and to trust region methods discussed below, and several researchers have shown interest in using it also.

For optimization problems with bound constraints, we have developed an algorithm using this new approach, that is extremely efficient in linear algebra cost. In numerical experiments our method is competitive with the partially separable update method of Conn, Gould and Toint, but it is applicable to a broader class of problems, and in some cases it is easier to implement for a particular application. We have written a paper based on this work, with Jorge Nocedal. In addition we have developed and tested software using this method to solve bound constrained optimization problems, which we plan to make publicly available, and which we will describe in a separate paper.

We have also developed, together with doctoral student Xuehua Lu, an efficient implementation of a limited memory symmetric rank one method, using a trust region approach. The trust region allows us to model negative curvature, and the compact representation allows the trust region computations to be preformed efficiently. Preliminary tests for this approach look promising. A talk on this work was given at the TIMS/ORSA Joint National Meeting in Chicago, May 1993.

2. Tensor Methods for Large Sparse Nonlinear Problems and for Constrained Optimization.

Over the last decade we have developed a new class of methods, called tensor methods, for solving nonlinear equations and unconstrained optimization problems. These methods appear to be considerably more efficient and robust than the best standard algorithms based upon Newton's method. During this research period, we have completed some ongoing work on this topic, and have expanded this work into several new directions.

First, we have completed the development and implementation of an efficient software package for tensor methods for nonlinear equations, and extended these methods to solve nonlinear least squares problems. Extensive testing shows that the tensor methods have considerable advantages over standard methods, in efficiency and robustness. A paper describing this software has just been completed, and submitted for publication along with the software.

Second, we have developed efficient tensor methods for solving large sparse systems of nonlinear equations and nonlinear least squares problems. These methods use efficient, state of the art sparse linear algebra techniques. We conducted substantial tests of this software, that show that the tensor methods have large advantages in robustness and computational cost over standard methods. This research is contained in the Ph.D. thesis of Ali Bouaricha and has been discussed in several talks including our plenary talk on tensor methods at the July 1992 SIAM National Meeting. A paper describing the nonlinear equations work is just now being completed.

Third, we developed parallel versions of tensor methods for small to medium size problems, and began the development of Krylov-subspace based tensor methods for very large problems that are amenable to efficient parallelization. The tests of the parallel tensor methods on an Intel hypercube showed that there is no loss in parallel efficiency between tensor methods and standard, linear model based methods. This means that tensor methods have the same advantages in computational costs on parallel as on sequential computers. The preliminary results of the Krylov subspace methods showed tensor methods can still lead to substantial advantages in computational cost over analogous Krylov subspace based linear model methods, even though there is an extra system that needs to be solved at each iteration. This work was presented at the SIAM Conference of Parallel Processing for Scientific Computation and a paper describing it appears in the proceedings of that meeting. Our students Ali Bouaricha and Dan Feng have both continued to pursue this research direction in their postdoctoral work, with excellent success. Feng has shown that the Krylov subspace tensor method can be amended to only require on iterative solve per iteration, without harming its computational properties, and has obtained excellent computational results with this method on some model aerodynamics problems.

Fourth, we have successfully completed a convergence analysis of a realistic tensor method for nonlinear equations, on nonsingular and singular problems. The analysis shows that tensor methods obtain 2 or 3 step order 1.5 convergence on problems with rank deficiency one at the solution, as well as the standard quadratic convergence on nonsingular problems. The techniques of analysis are new and useful, and have been applied in the nonlinearly constrained work described next and in Dan Feng's

recent Krylov subspace research. This work has been published in **Mathematical Programming**.

Finally, we have developed tensor methods for nonlinearly constrained optimization problems. Our new methods augment the standard linear model of the constraints by a tensor term. These methods are intended to be especially helpful on problems where the active constraints are (nearly) rank deficient at the solution, an important class of problems that are not solved efficiently by current methods. We developed two types of tensor models for nonlinear constrained problems, and a full global algorithm based upon these models. In computational tests of one of the new methods, they appear to exhibit substantial gains in efficiency over standard methods. We have also analyzed a simple method from this class and shown that it has fast convergence on problems where the constraint Jacobian at the solution has a null space of dimension one, whereas standard methods are only linearly convergent on such problems. In doing this, we have developed a generalization of the standard Kuhn-Tucker conditions for constrained optimization in the case of rank deficiency of the constraint Jacobian. This generalization is both mathematically interesting and relevant to computational optimization methods. This work formed a major part of Dan Feng's Ph.D. thesis, and is continuing. A paper on the tensor algorithms for constrained optimization has recently been submitted for publication, and papers on the generalized Kuhn-Tucker conditions and on the convergence analysis of the constrained tensor method are in draft form.

3. Trust Region Methods for Nonlinearly Constrained Optimization.

Our work on trust region methods for constrained optimization has concentrated on the development of a trust region method that uses the Symmetric Rank-One (SR1) method to approximate the Hessian of the Lagrangian. This work has been joint with Professor Humaid Khalfan of the United Arab Emirates University, and has been primarily theoretical. The use of the SR1 update is attractive in the context for two reasons. First, the SR1 has proven to be very competitive with the best known quasi-Newton methods in unconstrained and bounded constrained optimization. Second, the property of the SR1 that it does not necessarily generate positive definite matrices may be an asset in approximating the Hessian of the Lagrangian, which is not generally positive definite at the solution.

We have developed a practical trust region method for nonlinearly constrained optimization using the SR1 update, and have been able to establish superlinear convergence results for the this method similar to the results we previously derived for the SR1 in the unconstrained case (see Section 5). In particular, with a trust region, we have been able to dispense with the assumption that the SR1 approximation is positive definite. A paper describing this research is in progress.

4. Large Scale Global Optimization Methods for Molecular Configuration Problems.

We have been developing global optimization methods for finding the lowest energy configurations of molecular structures. To do this, one must find the lowest (global) minimum of energy functions that generally have very many parameters and huge numbers of local minimizers. Therefore, these are very difficult global optimization problems. Our approach is to develop fairly general purpose methods that do not utilize any knowledge of the solution structure, and are applicable to a broad class of partially separable large scale global optimization problems. The methods combine efficient stochastic global optimization techniques with several new, more deterministic perturbation techniques. So far we have applied our methods to Lennard-Jones problems with up to 76 atoms, to water clusters with up to 32 molecules whose energy is given by the Coker/Watts potential, and to polymers with up to 58 amino acids with potential energy given by the CHARMM package. The results appear to be the best so far by general purpose optimization methods, and appear to be producing some interesting chemistry issues.

Our methods combine an initial phase that locates some initial low local minimizers and is derived from previous stochastic methods, and a second, more deterministic phase for progressing from low to even lower local minimizers that is new and accounts for most of the computational effort, and the success, of the methods. Both phases make critical use of new portions that vary only a small subset of variables (an atom for Lennard-Jones, a molecule for water, or a small set of torsion angles for polymers) at once. In the initial phase this is used to improve the sample points. In the second phase, it is used to move one atom or molecule, or a set of torsion angles, in an existing configuration to better positions by solving a global optimization problem in only these variables. These steps are relatively inexpensive due to the small number of variables involved and the separability of the energy function. An expansion of the cluster before the one molecule global optimization was added for the water problem, and is crucial to the success of the method because it permits the method to move to significantly different structures. It has also proven useful for larger Lennard-Jones problems. Heuristics for deciding which configuration and molecule to improve next have also been important to the success of the water method. These methods could be applied to any partially separable function, although the determination of the unit to vary at once would be problem dependent. For the polymer problems, there are some significant differences in a number of these stages due to the chain structure of these problems.

For the Lennard-Jones problems, we first ran a simplified version of our algorithm, with no expansion phase, on problems with up to 55 atoms. For up to 30 atoms we found the best known solutions, including a solution for 22 atoms that was unknown prior to Northby's special purpose Lennard-Jones method. For over 30 atoms we usually did not find quite as good a solution as the Northby method. More recently, we have added expansion, and also conducted considerably longer runs on parallel computers, which have permitted us to consider configurations derived from a far larger set of initial configurations. In these experiments, we have run problems with up to 76 atoms, and have found the best known solutions for all of them, and a new improved solution for 75 atoms. These solutions include improved solutions for several of the larger problems that had been found only recently by other researchers, some using specialized methods. Our solutions are apparently by the far the best that have been produced for these problems by a general purpose global optimization method.

For water, we have mainly run our algorithm on clusters of 20 and 21 water molecules, because results of minimizing these same clusters and energy function, using a dynamic simulated annealing procedure, have been obtained by X. Long at University of California, San Diego. We have obtained many configurations with significantly lower energies than Long's. At present, the best solutions obtained by running our algorithm have energies of -0.3482 and -0.3690 atomic units (a.u.) for 20 and 21 molecules, respectively. These are approximately 0.005 and 0.011 a.u. lower than the best structures found by Long, respectively, whereas at room temperature, only vibrational states with energies about 0.001 a.u. above the ground state are likely. These values have been obtained through fairly long runs on 64 processors of the Intel Delta computer at Caltech, and it is beginning to seem likely that we are near the global minimum for these problems. We still do not know, however, whether the structures we have found are global minima. Of the best structures, some have the expected dodecahedral (for 21) or collapsed dodecahedral (for 20) shapes, but some that are very close to the current minimum have more irregular shapes. If these are indeed possible vibrational states, this raises interesting questions about either the possible structure of water clusters or the validity of the Coker/Watts energy function.

We have also begin working on configuration problems associated with polymers. The polymer problem we have begun working with is the protein polyalanine, using the CHARMM energy function to compute the potential energy. As is common in this area we have been treating the bond lengths and

bond angles as fixed, and have been trying to find the optimal values of the dihedral angles. This parameterization is natural because the dihedral angles are the crucial parameters to be varied in the optimization, and more efficient because the number of variables is greatly reduced, but the internal parameterization leads to some interesting algorithmic challenges. We have adapted our general approach to this framework in two ways. In the initial phase, the sampling is done by generating dihedral angles sequentially along the chain, and the angle is resampled if it gives a poor value for the potential up to that point. In the second phase we try to improve local minimizers by selecting a small subset of dihedral angles, and doing global minimization on a the resulting small dimensional problem, followed by full dimensional local minimization as above. So far we have been able to find what appears to be the global solution for problems with 20, 30, and 40 amino acids (40, 60, and 80 dihedral angles). We have just begun working on a problem with 58 amino acids, and the algorithm seems to be working well, although this work is still very much in progress. We plan to continue developing methods for finding the lowest energy configuration of polymers, and will soon switch to other polymers to help develop the new methods.

This work has been presented at a variety of conferences during this research period, including the SIAM Conference on Optimization, the Conference on Large-Scale Optimization in Gainesville, Florida, at NATO Advanced Study Institute, and the ORSA-TIMS national meeting, as well as at workshops at the University of California, San Diego and at Iowa State University. Papers describing this work have appeared in a conference proceedings and in two books resulting from conferences. In addition, a paper on the algorithm and its application to Lennard-Jones problems has been completed and submitted for publication, and several other papers are in preparation.

5. Other Topics in Nonlinear Optimization.

In addition to the topics described above, during this research period we have performed research on several optimization topics that are closely related to one or more of the above topics. In this section we mention these topics very briefly.

We have performed theoretical and computational research on the use of the symmetric rank-one update in unconstrained optimization. This update appears very competitive with the widely used BFGS update, and depending on the results of our research and the research of others, we may want to use it in our software such as our global optimization methods. We completed a computational study of the SR1 method and an analysis of its convergence properties using a line search; this paper has appeared in the **SIAM Journal on Optimization**. More recently we showed that an SR1 method with a trust region has even stronger convergence properties; this paper has been submitted for publication. This research forms the foundation for the research on using the SR1 update in trust region methods for constrained optimization that is discussed in Section 6.

Recently, we have also completed a computational study of various implementations of the BFGS method. There are several ways to implement the linear algebraic operations in a BFGS method that are equivalent mathematically, but that have differing costs in operations, and differing abilities to adapt to parallel computation. The form that sequences the inverse Hessian approximation seems best as far as its cost and parallelizability, but there have been concerns about its numerical stability. Our tests indicate that this form does have significant advantages over the Cholesky factor form in terms of cost and parallelizability, and no discernible disadvantages in terms of stability or robustness. Thus it is the form that we would recommend. This research was just completed and will be written up shortly.

Finally, we continued some research in developing and analyzing algorithms for solving block bordered systems of nonlinear equations on parallel computers. This is a problem of significant practical importance, because many large scale problems are naturally expressed as block bordered problems. Our research in this period showed how to construct methods with good global (and local) convergence properties in a way that is consistent with parallelization and infrequent communication. This research is described in two journal publications.

6. Publications and Technical Reports Associated with Grant DAAL03-91-G-0151

(1) A. Bouaricha and R.B. Schnabel, "Parallel tensor methods for nonlinear equations and nonlinear least squares", in *Proceedings of Sixth SIAM Conference of Parallel Processing for Scientific Computation,* SIAM, Philadelphia, 1993, pp. 639-643.

(2) R.H. Byrd, T. Derby, E. Eskow, K. P. B. Oldenkamp, and R. B. Schnabel, "A new stochastic/perturbation method for large-scale global optimization and its application to water cluster problems", in *Large Scale Optimization: State of the Art,* W.W. Hager, D.W. Hearn, and P.M. Pardalos eds., Kluwer Academic Publishers, Dordrecht, 1994.

(3) R.H. Byrd, T. Derby, E. Eskow, K. Oldenkamp, R.B. Schnabel and C. Triantafillou, "Parallel global optimization methods for molecular configuration problems", in *Proceedings of Sixth SIAM Conference of Parallel Processing for Scientific Computation,* SIAM, Philadelphia, 1993, pp. 165-169.

(4) R. H. Byrd, E. Eskow, and R. B. Schnabel, "A new large-scale global optimization method and its application to Lennard-Jones problems", University of Colorado Technical Report CU-CS-630-92, submitted *SIAM Journal on Optimization.*

(5) R.H. Byrd, E. Eskow, R. Schnabel and S. Smith, "Parallel global optimization: numerical methods, dynamic scheduling methods, and application to molecular configuration", in *Parallel Computation*, A.E. Fincham and B. Ford, eds., Clarendon Press, Oxford, 1993

(6) R. H. Byrd, D. Liu, D., and J. Nocedal, "On the behavior of Broyden's class of quasi-Newton methods," *SIAM Journal on Optimization* 2, 1992, pp. 533-557.

(7) R. H. Byrd, P. Lu, J. Nocedal and C. Zhou, "A limited memory algorithm for bound constrained optimization", Technical Report NAM-07, Dept. of Electrical Engineering and Computer Science, Northwestern University, 1993, submitted to *SIAM Journal on Scientific Computing.*

(8) R. H. Byrd, H. Khalld have fan, and R. B. Schnabel, "Analysis of a symmetric rank-one trust region method", University of Colorado Technical Report CU-CS-657-93, submitted to *SIAM Journal on Optimization.*

(9) R. H. Byrd, J. Nocedal, and R. B. Schnabel, "Representations of quasi-Newton matrices and their use in limited memory methods", *Mathematical Programming 63,* 1994, pp. 129-156.

(10) R. H. Byrd, R. Tapia, and Y. Zhang, "An SQP augmented Lagrangian BFGS algorithm for constrained optimization," *SIAM Journal on Optimization* 2, 1992, pp. 210-241.

(11) D. Feng, P. Frank, and R. B. Schnabel, "Local convergence analysis of tensor methods for nonlinear equations", *Mathematical Programming 62,* 1993, pp. 427-459.

(12) E. Eskow and R. B. Schnabel, "Software for a new modified Cholesky factorization," *ACM Transactions on Mathematical Software,* 17, 1991, pp. 306-312.

(13) D. Feng and R. B. Schnabel, "Globally convergent parallel algorithms for solving block bordered systems of nonlinear equations", *Optimization Methods and Software 2,* 1993, pp. 269-295.

(14) D. Feng and R. B. Schnabel, "Tensor methods for equality constrained optimization", University of Colorado Technical Report CU-CS-729-94, submitted to *SIAM Journal on Optimization*

(15) H. F. Khalfan, R. H. Byrd and R. B. Schnabel, "A Theoretical and Experimental Study of the Symmetric Rank-One Update", *SIAM Journal on Optimization 3,* 1993, pp 1-24.

(16) R. B. Schnabel, "Parallel nonlinear optimization: limitations, opportunities, and challenges", *Algorithms for Continuous Optimization: The State of the Art,* E. Spedicato, ed., Kluwer Academic Publishers, 1994, pp. 531-559.

(17) S. Smith and R. B. Schnabel, "Centralized and distributed dynamic scheduling for adaptive parallel algorithms," in *Unstructured Scientific Computation on Scalable Multiprocessors,* P. Mehrotra, J. Saltz, and R. Voigt, eds., MIT Press, Cambridge, Mass., 1992, pp. 301-321.

(18) S. Smith and R. B. Schnabel, "Dynamic scheduling strategies for an adaptive asynchronous parallel global optimization algorithm, University of Colorado Technical Report CU-CS-625-92, submitted to *IEEE Parallel and Distributed Systems.*

(19) X. Zhang, R. H. Byrd, and R. B. Schnabel, "Parallel methods for solving nonlinear block bordered systems of equations," *SIAM Journal on Scientific and Statistical Computing* 13, 1992, pp. 841-859.

7. Students Supported (in part) by Grant DAAL03-91-G-0151

Ali Bouaricha, Ph.D. granted 1992

Dan Feng, Ph.D. granted 1993

Xuehua Lu, Ph. D. in progress (expected 1995)

Christos Triantafillou, Ph.D. student