

ALGORITHMS AND SOFTWARE FOR LARGE SCALE OPTIMIZATION

Progress Report

for Period Aug 1, 1989 - May 5, 1990

Jorge Nocedal

Northwestern University
Evanston, IL 60208

May 1990

Prepared for

THE U.S. DEPARTMENT OF ENERGY
AGREEMENT NO. DE-FG02-82ERO183

MASTER

db

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Continuation Proposal for the Project
"Algorithms and Software for Large Scale Optimization"

Principal Investigator: Jorge Nocedal
(Grant No. DE-FG02-87ER25047-A001)

This is a request for a continuation of this three-year grant. The research has proceeded as planned. There is no change in the scope of the project nor in the proposed budget for the coming year.

1. Report of Technical Progress.

We have developed and tested optimization algorithms capable of solving non-linear problems with thousands of variables. The main focus of our research was the limited memory BFGS method (L-BFGS), proposed by the principal investigator in 1980 and re-defined and improved in 1989. This method is considerably faster than conjugate gradient methods, and is capable of solving very large problems (possibly with as many as 10^6 variables). Its speed is due to the use of a low-storage dynamic preconditioner based on the BFGS update formula, and due to an automatic scaling technique. Our L-BFGS code has been distributed widely and is now being used, for example, to solve problems in meteorology and oceanography.

We have studied the performance of the L-BFGS method on the Alliant FX/8 at Argonne National Laboratory, using problems with up to 50,000 variables. The computationally intensive part of the method is the computation of the search direction, which is obtained by a recursive algorithm that exploits the symmetry of the BFGS update. This recursion is optimal in a sequential machine (reducing the cost by almost 50%), and it is important to study its performance on a parallel machine. The implementation described in [1] gave quite satisfactory results. Even though significant memory management problems were observed for large test problems, we obtained an execution rate of 13 megaflops, using 8 processors.

The main application area considered during the last year was the simulation of transonic flow around an airplane wing. A nonlinear least squares formulation of the problem, with 461 degrees of freedom, was solved using the L-BFGS method. The method was substantially faster than the code of Buckley-LeNir and the CONMIN code of Shanno-Phua. Two other application problems have been prepared for future testing; one is a problem in meteorology and the other in crystallography.

Another attractive method for solving large optimization problems is the discrete Newton method, in which Hessian-vector products are approximated by finite differences of gradients. We studied the relative performance of such a method with respect to that of the L-BFGS method [2]. It appears that L-BFGS is more efficient in the types of problems that arise more frequently in practice, and that the cost of additional gradient evaluations in the discrete Newton method does not pay off. We have concluded from this study that the best approaches currently available

for solving large problems are: (i) Newton methods with sparse matrix techniques; (ii) the partitioned quasi-Newton method of Griewank and Toint; (iii) the limited memory L-BFGS method.

An open question since the early 1970s is whether the popular BFGS method is in fact the best variable metric method for optimization. Many recent studies have shown that the BFGS method has some excellent theoretical properties. (which are missing from the DFP method, for example). There are also good reasons for assuming that the best variable metric method belongs to the linear class of Broyden methods. This class, which depends on a parameter ϕ , contains the BFGS and DFP (which are obtained by setting $\phi = 0$ and $\phi = 1$, respectively). An analysis of the Broyden class suggests that it may be advantageous to let the parameter ϕ become negative, at least in some iterations. We have analyzed this strategy [3], and found that using negative values of ϕ can lead to a deterioration in the rate of convergence. In particular we showed that for some apparently natural strategies for choosing $\phi < 0$, superlinear convergence is lost. Our work shows that much care is required when implementing such strategies. Nevertheless, we were able to show that it is possible to always use a negative value for the parameter ϕ and retain the good properties of BFGS, but the rule for finding ϕ is difficult to implement in practice. A numerical investigation of several new methods showed that an improvement over BFGS is possible, but is not likely to be dramatic. The analysis given in [3] suggests that a significantly better method than BFGS may not exist within the Broyden class. The analysis also provides several new general results on convergence and rate of convergence of variable metric methods.

2. Research Plans for the Coming Year.

We are currently investigating the use of full-rank dynamic scaling techniques for large scale optimization problems. The idea is based on the column scaling technique developed by Powell for linearly constrained problems. It consists of updating a factorization of the inverse Hessian approximation: $E_k^{-1} = Z_k Z_k^T$, and controlling the size of the columns of Z . The first column of Z is guaranteed to have the correct size, because an orthogonal rotation is used to pre-process the Hessian approximation, and this pre-processing ensures that the size of the first column equals the average curvature detected during the iteration. For the rest of the columns, the strategy is to increase them in size (scale them up) if they are small compared to the first column.

An analysis of this scaling technique revealed that it is incomplete, and we propose an algorithm that sometimes scales the columns up, and sometimes scales them down. We have shown that this algorithm is globally convergent on convex problems, and we are currently investigating whether it enjoys a superlinear rate of convergence. Our first numerical tests with this strategy were remarkable: for

ill-conditioned problems, and for problems with a poor estimate of the Hessian approximation, the new method accelerates convergence dramatically, typically cutting the amount of work by 50%. The ideas therefore hold much promise and will be useful both for medium size problems as well as for large scale optimization. We will develop a BFGS code with this strategy, aimed at small and medium-size problems. We believe that this will be the most efficient BFGS code in existence. We also intend to study how to use this automatic column scaling technique in the context of a limited memory method for large problems. The extension is not simple, but we believe that it will result in a substantial improvement of performance. Graduate student Marucha Lalee has done a significant part of this research, and will participate on all aspects of this work.

The main application areas we will consider in the coming year are meteorology and oceanography. The four-dimensional assimilation approach in meteorology requires effective nonlinear optimization codes. To determine the best approach we will study the formulation of the problem and the characteristics of the function and gradient. It is possible that new optimization techniques will be required to effectively exploit the structure of the problem. We will also solve problems in oceanography, involving hundreds of thousands of variables, which are currently being solved by simple conjugate gradient techniques. For these problems it is important to be able to handle upper and lower bounds in the variables - a subject that we plan to address during the next year.

The theory of conjugate gradient methods for linear systems of equations is elegant and complete. It is therefore surprising that not much is known about the theoretical properties of nonlinear conjugate gradient methods. This lack of knowledge is well illustrated by the Polak-Ribiere method, which is considered to be the most effective nonlinear conjugate gradient method in practice, but which is known to be capable of failing on non-convex problems. On the other hand, the Fletcher-Reeves method, which does not perform very well in practice, is the only conjugate gradient method with satisfactory global convergence results. We intend to advance the theory of nonlinear conjugate gradient methods, and in so doing we expect to find an implementation that is superior to the ones known to date. It is true that, in their present form, conjugate gradient methods are not competitive with the L-BFGS method when solving large problems. However we believe that conjugate gradient methods have not been properly implemented so far, and that their full potential needs to be assessed. This topic is important: conjugate gradient methods require very low storage, and vectorize and parallelize well. If their convergence rate and robustness could be improved, they would be very attractive when solving extremely large problems. We have begun by showing that several modifications of the Polak-Ribiere method, which in some sense can be interpreted as simple restarting techniques along directions different from the

gradient, give rise to superlinear convergence. The analysis has highlighted an issue that appears to have been neglected in the literature, and which is crucial, namely the conditions that need to be imposed in the line search to ensure that descent directions are always generated.

We have found this line search condition, and it fits perfectly with the convergence theory. It is also easily implemented by modifying a line search algorithm, such as the one developed by Moré, which enforces the Wolfe conditions. We have found that nonlinear conjugate gradient methods fall into two categories, which we will characterize precisely in a forthcoming paper [4]; one them contains the Fletcher-Reeves method, and the other the Polak-Ribiere method. It is clear to us that a subset of the second class contains the most valuable methods. Preliminary numerical tests show that all the new methods outperform the existing conjugate gradient methods. The search for a method with all the desirable theoretical and numerical properties is under way.

References:

- [1] *The Performance of Several Algorithms for Large Scale Unconstrained Optimization*, (1989), Northwestern University, EECS Dept. Rep. NAM 01, to appear in the Proceedings of the Cornell Conference on Large Scale Optimization.
- [2] *A Numerical Study of the Limited Memory BFGS Method and the Truncated-Newton Method for Large Scale Optimization*, (1989), Northwestern University, EECS Dept. Rep. NAM 02, submitted to SIAM Journal on Optimization, with S. Nash
- [3] *On the Behaviour of Broyden's Class of Quasi-Newton Methods*, 1990 Northwestern University, EECS Dept. Rep. NAM 01, submitted to SIAM Journal on Optimization, with R. Byrd and D.C. Liu.
- [4] *Global Convergence Properties of Conjugate Gradient Methods for Optimization*, 1990, to appear, with J.C. Gilbert.

END

**DATE
FILMED**

3 / 23 / 92

