Discussion Paper No. 189

A Modified Conjugate Gradient Algorithm for Unconstrained Nonlinear Optimization

by Avinoam Perry

December 31, 1975

Department of Managerial Economics
and Decision Sciences
Graduate School of Management
Northwestern University

Avinoam Perry
Assistant Professor of Decision Sciences
Graduate School of Management
Northwestern University

Scope and Purpose

This paper deals with the problem of minimizing the unconstrained nonlinear continuous and differentiable function $f(x_1, x_2,...,x_n)$. Algorithms which are constructed for this purpose are decomposed into two major parts. In the first part a direction of improvement is found while in the second part a search is conducted along this direction in order to determine the optimal step size. Survey articles on the topic of unconstrained optimization are given in references [10], [12], [17] at the end of this paper.

Avinoam Perry
Assistant Professor of Decision Sciences
Graduate School of Management
Northwestern University

Abstract

In this paper we suggest a modification of the Fletcher-Reeves conjugate gradient algorithm. This modification results in an improved algorithm which is extremely powerful in the absence of an accurate line search procedure.

I. Introduction

Conjugate directions algorithms for minimizing unconstrained nonlinear programs can be divided into two major classes, in the first class we find algorithms with no memory such as Fletcher-Reeves conjugate gradient algorithm [8] and a modified version of the conjugate gradient which is presented here. In the second class we have the quasi-Newton methods which apply a matrix update approximating the hessian inverse of x. Among the most popular quasi-Newton procedures we have the Davidon [4] Fletcher & Powell [7] - DFP update, Broyden's-Rank-one-[1] BR1, update, Pearson's algorithms [16], Broyden [2], Fletcher [6], Goldfarb [9], Shanno [18] the BFGS update, and Huang [11] general family. Recent developments in the field of unconstrained optimization concentrate their efforts on algorithms with inaccurate or no line search, this is due to the fact that the line search part of an algorithm is the most time consuming part, and algorithms which exhibit superlinear convergence rate without having to go through accurate line search are preferred to the others. One of the most recent examples of an algorithm belonging to this class is the one developed by Davidon [5]. Optimal conditioning and self scaling procedures such as the ones developed by Oren and Luenberger [13,14], Oren and Spedicato [15], Shanno [18] and others, contribute greatly to the overall efficiency and local convergence properties of algorithms with inaccurate line search. However, some of these algorithms result in a departure from the pure* quadradic convergence phenomenon [13,14].

*By pure quadradic convergence we refer to algorithms which minimize a quadradic function in, at most, n+1 steps.

Computatational experience with quasi-Newton type algorithms leads to the conclusion that in order to eliminate severe accumulated rounding errors and ineffective updated matrices, the procedure should be restarted after every prespecified number of iterations. The most popular heuristic is the one which restarts the quasi-Newton procedure after every n steps. Algorithms without memory utilize information obtained in step k and k-1 only. As a result the danger of accumulating rounding errors and constructing erroneous updates due to inaccurate line search and rounding error accumulation is reduced considerably. The relatively poor performance of the Fletcher-Reeves [8] conjugate gradient algorithm is not a result of its inability to accumulate information, but rather, its strong dependence on problem structure. In the construction of the direction equation of the Fletcher-Reeves conjugate gradient the assumption $\nabla f(x_k)' \cdot \nabla f(x_{k+1}) = 0$ is made explicitly. This assumption always holds for quadradic programs with perfect line search and does not hold under more general conditions.

Relaxation of the above orthogonality assumption and an additional correction leads to a modification of the Fletcher-Reeves equation that results in a superior algorithm, the performance of which is competetive with the most successful quasi-Newton methods.

II. Derivation of the Modified C. G. Method.

Let $p_k \equiv x_{k+1} - x_k$ and $q_k \equiv g_{k+1} - g_k$ where $g_k \equiv \nabla f(x_k)$ and the unconstrained nonlinear program is min f(x). Let $d_k \equiv (1/\alpha_k) p_k$ where α_k is a scalar minimizing the one dimensional program $\min_{\alpha \geq 0} f(x_k - \alpha d_k)$.

A conjugate direction in E^n has the property

(1) $q_k' p_{k+1} = \alpha_k q_k' d_{k+1} = 0$

A conjugate gradient direction at stage k+1 is constructed by taking a linear combination of the gradient at stage k+1 and the direction vector at stage k.

(2) $d_{k+1} = -g_{k+1} + \beta d_k$

equation (1) implies

(3) $$\beta = \frac{q_k' \, g_{k+1}}{q_k' \, d_k}$$

and (2) becomes

(4) $$d_{k+1} = -g_{k+1} + \frac{q_k' g_{k+1}}{q_k' d_k} \cdot d_k$$

If f(x) is quadradic and α is computed with perfect accuracy we have

(5) $q_k' g_{k+1} = (g_{k+1} - g_k)' g_{k+1} = g_{k+1}' g_{k+1}$

and

(6) $q_k' d_k = (g_{k+1} - g_k)' d_k = (g_{k+1} - g_k)'(-g_k + \beta_k d_{k-1}) = g_k' g_k$

and (4) becomes

(7) $$d_{k+1} = -g_{k+1} + \frac{g_{k+1}' \, g_{k+1}}{g_k' g_k} \, d_k .$$

(7) is the well known Fletcher-Reeves conjugate gradient direction [8].

Relaxing the assumption regarding the orthogonality of consecutive gradient vectors we can rewrite (4) as follows

(8) $$d_{k+1} = -g_{k+1} + \frac{q_k' g_{k+1}}{q_k' \, p_k} \cdot p_k = -g_{k+1} + \frac{p_k q_k'}{p_k' q_k} \cdot g_{k+1} = -\left[I - \frac{p_k q_k'}{p_k' q_k} \right] g_{k+1}$$

Denote the matrix $I - \dfrac{p_k q_k'}{p_k' q_k}$ as D_{k+1} and (8) becomes

(9) $d_{k+1} = -D_{k+1} g_{k+1}$

Note that D_{k+1} does not depend on D_k and hence it is a conjugate direction algorithm with no memory.

<u>Proposition 1</u>: The matrix D_{k+1} is of rank n-1.

<u>proof</u>: The matrix D_{k+1} has a right eigen vector $\equiv p_k$, and a left eigen vector $\equiv q_k$, the eigen values of which are equal to zero.

<u>Proposition 2</u>: The right orthogonal complement of D_{k+1} is the rank 1 matrix $\frac{p_k p_k'}{p_k' q_k}$ and the left orthogonal complement of D_{k+1} is the rank 1 matrix $\frac{q_k q_k'}{q_k' p_k}$

<u>proof</u>: The product $D_{k+1} \cdot \frac{p_k p_k'}{p_k' q_k}$ and the product $\frac{q_k q_k'}{q_k' p_k} \cdot D_{k+1}$ yield the nxn null matrix.

<u>Proposition 3</u>: The matrix $S_{k+1} = D_{k+1} + \frac{p_k p_k'}{p_k' q_k}$ is of full rank.

<u>proof</u>: D_{k+1} is of rank (n-1) and $\frac{p_k p_k'}{p_k' q_k}$ is its right orthogonal complement. By adding the two complementary matrices we obtain full rank.

<u>Proposition 4</u>: A direction vector $d_{k+1} = -S_{k+1} g_{k+1}$ is equal to the Fletcher-Reeves conjugate gradient direction if f(x) is quadradic and α_k is computed with perfect accuracy.

proof: Under the above conditions $p_k' g_{k+1} = 0$, and hence, $-S_{k+1} q_{k+1} = -D_{k+1} g_{k+1} = d_{k+1}$ of Fletcher-Reeves [8].

The product $S_{k+1} g_{k+1}$ can be expressed as

$$(10) \quad d_{k+1} = -S_{k+1} g_{k+1} = -\left[I - \frac{p_k q_k'}{p_k' q_k} + \frac{p_k p_k'}{p_k' q_k} \right] g_k = -g_k + \frac{(q_k - p_k)' g_{k+1}}{p_k' q_k} \cdot p_k$$

Denoting

$$(11) \quad \gamma_k = \frac{(q_k - p_k)' \, g_{k+1}}{p_k' q_k}$$

we obtain a modified conjugate gradient equation

$$(12) \quad d_{k+1} = -g_{k+1} + \gamma_k p_k$$

Based on proposition 4 we conclude that the modified conjugate gradient method, which is constructed by applying the direction vector in (12) to $f(x_k + \alpha_k d_k)$, is a method possessing a quadradic convergence rate.

Computational experience with this method indicates, that the method of the modified conjugate gradient is competitive with the most efficient quasi-Newton methods and more efficient than the Fletcher-Reeves conjugate gradient method [8].

III Further Observations and Concluding Remarks

The matrix $S_{k+1} = I - \frac{p_k q_k'}{p_k' q_k} + \frac{p_k p_k'}{p_k' q_k}$ is not symmetric and

$S_{k+1} q_k \neq p_k$ but, nevertheless, we have $q_k' S_{k+1} q_k = p_k' q_k$. $p_k' q_k$ is assumed to be positive, a fact which is always true for convex programs. The matrix S_{k+1} is not symmetric due to the fact that D_{k+1} is not symmetric. Symmetry can be forced into the equation by replacing S_{k+1} with H_{k+1} which is defined as

(13) $H_{k+1} = D_{k+1} \cdot D'_{k+1} + \frac{p_k p'_k}{p'_k q_k} = (I - \frac{p_k q'_k}{p'_k q_k})(I - \frac{q_k p'_k}{p'_k q_k}) + \frac{p_k p'_k}{p'_k q_k}$

Note that $H_{k+1} q_k = p_k$ and that H_{k+1} is the BFGS update with no memory. Symmetry can be forced through the transpose operation of (13) and its complement.

(14) $T_{k+1} = D'_{k+1} D_{k+1} + \frac{q_k q'_k}{q'_k p_k} = (I - \frac{q_k p'_k}{p'_k q_k})(I - \frac{p_k q'_k}{q'_k p_k}) + \frac{q_k q'_k}{p'_k q_k}$

Note that $T_{k+1} p_k = q_k$. It can be shown that $(T_{k+1})^{-1}$ is the DFP update with no memory.

Quasi-Newton methods such as the ones described above are especially effective when dealing with nonlinear nonquadradic programs with inaccurate line search. Under these circumstances, the directions constructing the DFP and the BFGS updates are far from being conjugate to each other with respect to a hessian matrix which keeps changing from one step to another.

Experiments with the modified conjugate gradient algorithm involved eight test problems which are described in detail in Himmelblau [10, pp. 195]. We compared the performance of the DFP and BFGS methods over a wide range of line search accuracy measures. We found out that as the line search accuracy went down the performance of the M.C.G. improved relative to the DFP and the BFGS methods. This conclusion was consistent with all the problems we tested.

Our computational experience leads to the conclusion that algorithms with inaccurate line search perform better if information from previous steps is not accumulated into an update matrix and the identity matrix replaces S_k in the computation of S_{k+1}. Under these circumstances the modified conjugate gradient algorithm performs better than either the DFP or the BFGS procedures.

REFERENCES

[1] Broyden, C.G., "Quasi-Newton Methods and Their Application to Function Minimization", Math. Comp., 21, 1967 pp. 368-381.

[2] Broyden, C.G. "The Convergence of a Class of Double Rank Algorithms, Parts I and II," J. Inst. Math. Appl., 7, 1971, pp. 76-90, 222-236.

[3] Broyden, C.G., J.E. Dennis, and J.J. Moré, "On the Local and Superlinear Convergence of Quasi-Newton Methods", J. Inst. Math Appl. 12, 1973 pp. 223-246.

[4] Davidon, W.C. "Variable Metric Method for Minimization", Research and Development Report ANL-5990, U.S. Atomic Energy Commission, Argonne National Laboratories, 1959.

[5] Davidon, W.C. "Optimally Conditioned Algorithms without Line Searches", Math. Prog. Vol. 9, No. 1, 1975, pp. 1-30.

[6] Fletcher, R. "A New Approach to Variable Metric Algorithms", Comp. J., 13, 1970, pp. 317-322.

[7] Fletcher R., and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization", Comp. J., 6, 1963, pp. 163-168.

[8] Fletcher, R. and C.M. Reeves, "Function Minimization by Conjugate Gradients", Comp. J., 7, 1964, pp. 149-154.

[9] Goldfarb, D., "A Family of Variable Metric Methods Derived by Variational Means", Math. Comp. 24, 1970, pp. 23-26.

[10] Himmelblau, D.M. "Applied Nonlinear Programming", McGraw-Hill Book Co. 1972.

[11] Huang, H.Y., "Unified Approach to Quadradically Convergent Algorithms for Function Minimization J.O.T.A. 5, 1970, pp. 405-423.

[12] Luenberger, D.G. "Introduction to Linear and Nonlinear Programming", Addison-Wesley Publishing Co. 1973.

[13] Oren, S.S. "Self Scaling Variable Metric Algorithms, Part II", Management Science, Vol 20, No. 5, 1974, pp. 863-874.

[14] Oren, S.S. and D.G. Luenberger, "Self Scaling Variable Metric Algorithms, Part I", Management Science, Vol. 20, No. 5, 1974, pp. 845-862.

[15] Oren, S.S. and E. Spedicato, "Optimal Conditioning of Self Scaling Variable Metric Algorithm", Presented in ORSA-TIMS meeting in San Juan De Poerto Rico, Oct. 1974.

[16] Pearson, J.D., "Variable Metric Methods of Minimization", Comp. J., 13, 1969, pp. 171-178.

[17] Powell, M.J.D. "Recent Advances in Unconstrained Optimization", Math. Prog. 1, 1971, pp. 26-57.

[18] Shanno, D.F. "Conditioning of Quasi-Newton Methods for Function Minimization", Math. Comp. 24, 1970, pp. 647-656.