RICE UNIVERSITY

NUMERICAL EXPERIMENTS ON QUADRATICALLY CONVERGENT ALGORITHMS FOR FUNCTION MINIMIZATION

by

A.V. LEVY

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

Thesis Director's signature

Houston, Texas

May, 1970

Abstract

NUMERICAL EXPERIMENTS ON QUADRATICALLY CONVERGENT ALGORITHMS FOR FUNCTION MINIMIZATION

by

A.V. LEVY

The nine quadratically convergent algorithms for function minimization appearing in Ref. 1 are tested through numerical examples. Both a quadratic function and a nonquadratic function are investigated. For the quadratic function, the results show that, if high-precision arithmetic together with high accuracy in the one-dimensional search is employed, all the algorithms behave identically: they all produce the same sequence of points and they all lead to the minimal point in the same number of iterations (this number is equal at most to the number of variables). For the nonquadratic function, the results show that some of the algorithms behave identically and, therefore, any one of them can be considered to be representative of the entire class. The effect of different restarting conditions on the convergence characteristics of the algorithms is studied. Proper restarting conditions for faster convergence are given.

ACKNOWLEDGMENTS

The author is indebted to Dr. Ho-Yi Huang for suggesting the topic and helpful discussions. This research was supported by the Office of Scientific Research, Office of Aerospace Research, United States Air Force, Grant No. AF-AFOSR-828-67.

TABLE OF CONTENTS

1. Introduction

In a previous report (Ref. 1), Huang developed a unified method for constructing a generalized quadratically convergent algorithm for the minimization of a function f(x), where f is a scalar and x an n-vector. He proved that all the existing conjugate-gradient algorithms and variable-metric algorithms can be obtained as particular cases. In addition, several new pratical algorithms were generated. The following properties were required: (a) the algorithms use one-dimensional search only; (b) for a quadratic function, the algorithms are capable of converging quadratically to the minimal point, that is, the minimal point is achieved in a number of iterations at most equal to the number of variables; (c) the algorithms employ the function and its gradient g(x) only; and (d) the algorithms employ only information at the present stage and the stage immediately previous to the present.[1]

In this thesis, the algorithms presented in Ref. 1 are studied and compared through numerical examples. For conciseness, a summary of these algorithms is given in Section 2. Quadratic functions are discussed in Sections 3-5 and nonquadratic functions are discussed in Sections 6-8.

---

[1] The terminology employed here is identical with that of Ref. 1.

2. Summary of Quadratically Convergent Algorithms

A minimizing algorithm is represented by the relations

$$p_i = H_i^T g_i \quad , \quad \Delta x_i = -\alpha_i p_i \quad , \quad x_{i+1} = x_i + \Delta x_i \tag{1}$$

where the subscript i denotes the present point and the subscript i+1 denotes the next point. Here, p_i, an n-vector, is the search direction; H_i, an n x n matrix, characterizes a particular algorithm; Δx_i, an n-vector, is the displacement vector; and α_i, a scalar, is the stepsize. The sequence of steps is as follows: (a) at a given point x_i, the gradient g_i is computed; (b) the matrix H_i is either given a priori or is updated in a way to be discussed later; (c) the search direction p_i is computed according to Eq. (1-1); (d) the stepsize α_i is determined by a one-dimensional search on the function $f(x_{i+1}) = f(x_i - \alpha_i p_i)$; in this search, the function $f(x_i - \alpha_i p_i)$ is minimized with respect to the stepsize; (e) the displacement Δx_i is computed according to Eq. (1-2); and (f) the next point x_{i+1} is obtained from Eq. (1-3). Steps (a) through (f) form a complete iteration leading from the present point x_i to the next point x_{i+1}.

For a given point x_i, steps (a) through (f) are completely fixed except for the specification of the matrix H_i in step (b). The specification of the matrix H_i is given below.

2.1. Matrix H_i. For any point x_i other than the initial point x_o, the matrix H_i is updated according to one of the following formulas:

Algorithm I (Fletcher-Powell-Davidon algorithm)

$$H_i = H_{i-1} + \frac{\Delta x_{i-1}\Delta x_{i-1}^T}{\Delta x_{i-1}^T \Delta g_{i-1}} - \frac{H_{i-1}\Delta g_{i-1}\Delta g_{i-1}^T H_{i-1}}{\Delta g_{i-1}^T H_{i-1}\Delta g_{i-1}} \tag{2}$$

Algorithm II (McCormick algorithm)

$$H_i = H_{i-1} + \frac{(\Delta x_{i-1} - H_{i-1}\Delta g_{i-1})\Delta x_{i-1}^T}{\Delta x_{i-1}^T \Delta g_{i-1}} \tag{3}$$

Algorithm III (Pearson algorithm)

$$H_i = H_{i-1} + \frac{(\Delta x_{i-1} - H_{i-1}\Delta g_{i-1})\Delta g_{i-1}^T H_{i-1}}{\Delta g_{i-1}^T H_{i-1}\Delta g_{i-1}} \tag{4}$$

Algorithm IV

$$H_i = H_{i-1} + \frac{(\Delta x_{i-1} - H_{i-1}\Delta g_{i-1})(\Delta x_{i-1} - H_{i-1}^T \Delta g_{i-1})^T}{(\Delta x_{i-1} - H_{i-1}^T \Delta g_{i-1})^T \Delta g_{i-1}} \tag{5}$$

Algorithm V

$$H_i = H_{i-1} - \frac{H_{i-1}\Delta g_{i-1}\Delta g_{i-1}^T H_{i-1}}{\Delta g_{i-1}^T H_{i-1}\Delta g_{i-1}} \tag{6}$$

Algorithm VI

$$H_i = H_{i-1} - \frac{H_{i-1}\Delta g_{i-1}\Delta x_{i-1}^T}{\Delta x_{i-1}^T \Delta g_{i-1}} \tag{7}$$

Algorithm VII

$$H_i = H_{i-1} - \frac{H_{i-1}\Delta g_{i-1}(\Delta x_{i-1} - H_{i-1}^T \Delta g_{i-1})^T}{(\Delta x_{i-1} - H_{i-1}^T \Delta g_{i-1})^T \Delta g_{i-1}} \tag{8}$$

Algorithm VIII

$$H_i = H_{i-1} - \frac{H_o \Delta g_{i-1} \Delta x_{i-1}^T}{\Delta x_{i-1}^T \Delta g_{i-1}} \tag{9}$$

Algorithm IX (Generalized Fletcher-Reeves algorithm)

$$H_i = H_o + \frac{H_o g_i p_{i-1}^T}{p_{i-1}^T g_{i-1}} \tag{10}$$

In Eqs. (2)-(10), the gradient difference Δg_{i-1} is defined by

$$\Delta g_{i-1} = g_i - g_{i-1} \tag{11}$$

and H_o denotes the matrix used at the starting point x_o.

2.2. Initial Matrix H_o. At the starting point x_o, the matrix H_o has to be assigned. For Algorithms I-VIII, any matrix H_o, such that the matrix $(H_o + H_o^T)/2$ is either positive definite or negative definite, can be used. For Algorithm IX, the matrix H_o must be symmetric.

2.3. Remark. At convergence, Algorithms I-IV yield a final matrix H_n equal to the inverse of the second derivative matrix. Also at convergence, Algorithms V-VII yield a final matrix H_n equal to the null matrix. Algorithms VIII and IX are simplified algorithms obtained from the general algorithm.

3. Quadratic Functions: General Considerations

To apply Algorithms I-IX to the minimization of a quadratic function, the following considerations are in order.

3.1. Starting Condition. For Algorithms I-VIII, any initial matrix H_o, such that the matrix $(H_o + H_o^T)/2$ is either positive definite or negative definite, can be used. For Algorithm IX, the matrix H_o must be symmetric.

3.2. Stopping Condition. Since the minimum condition

$$g(x) = 0 \tag{12}$$

can be achieved at any iteration, the algorithms are stopped (a) when the inequality

$$g_i^T g_i \leq \epsilon_1 \tag{13}$$

is satisfied or (b) when $i = n$. In Ineq. (13), ϵ_1 is a prescribed small number and defines the accuracy required in the optimum condition (12).

3.3. Precision Requirement. In practice, the realization of quadratic convergence on a computer requires that high-precision arithmetic be used together with high accuracy in the one-dimensional search. In this connection, the search is stopped when the inequality

$$|g_{i+1}^T p_i| \leq \epsilon_2 \tag{14}$$

is satisfied. Here, ϵ_2 is a prescribed small number.

4. Quadratic Functions: Numerical Example

In this section, a particular quadratic function is studied using a Burroughs B-5500 computer and double-precision arithmetic. The experimental conditions are now indicated.

Starting Condition. Three different initial matrices are tested. They are given by

$$H_0 = I \ , \ H_0 = -I \ , \ H_0 = I + S \tag{15}$$

where I is the identity matrix and S is a skew-symmetric matrix whose elements are defined by

$$s_{\ell k} = -s_{k\ell} = \ell - k \tag{16}$$

The initial matrices (15-1) and (15-2) are symmetric and, therefore, are used in Algorithms I-IX. The initial matrix (15-3) is not symmetric and, therefore, is used in Algorithms I- VIII only.

Stopping Condition. An algorithm is stopped when Ineq. (13) is satisfied. In Ineq. (13), the small number ε_1 is given by

$$\varepsilon_1 = 10^{-12} \tag{17}$$

One-Dimensional Search. The one-dimensional search is stopped when Ineq. (14) is satisfied. In Ineq. (14), the small number ε_2 is given by

$$\varepsilon_2 = 10^{-16} \tag{18}$$

Example 4.1. We consider the problem of minimizing the function[2]

$$f = (x + y + 0.5u)^2 + (x + 2y + z + u)^2 + (y + z + 1.5u)^2$$

$$+ (0.5x + y + 1.5z - 0.5)^2 \qquad (19)$$

This function admits the minimum $f = 0$ at the point defined by

$$x = 0.5 \ , \quad y = -0.5 \ , \quad z = 0.5 \ , \quad u = 0 \qquad (20)$$

The nominal point chosen for starting an algorithm is the point of coordinates

$$x = 4 \ , \quad y = 4 \ , \quad z = 4 \ , \quad u = 4 \qquad (21)$$

For a given initial matrix, all the algorithms generate the same sequence of points and, therefore, converge to the solution in the same number of iterations. The results pertaining to the initial matrices (15-1) and (15-2) are summarized in Table 1; those pertaining to the initial matrix (15-3) are given in Table 2.[3]

---

[2] In this example, as well as the subsequent example, all the symbols are scalar.

[3] The asterisk in Tables 1-2 marks those numbers which should be zero in an exact solution. Because of computer limitations, the value zero cannot be achieved. Therefore, only the highest error in the solution is indicated.

Table 1

Example 4.1, $H_o = I$ or $H_o = -I$, Algorithms I-IX.

i	x	y	z	u	f	$g^T g$
i	4.0000	4.0000	4.0000	4.0000	0.828×10^{3}	0.4×10^{5}
1	1.4755	-1.3315	0.3809	0.7517	0.577×10^{0}	0.2×10^{1}
2	1.3252	-1.3823	0.8605	0.4065	0.638×10^{-1}	0.1×10^{-1}
3	1.3017	-1.2926	0.8163	0.3265	0.565×10^{-1}	0.2×10^{-1}
4	0.5000	-0.5000	0.5000	0.0000	$^*0.653 \times 10^{-34}$	$^*0.3 \times 10^{-31}$

Table 2

Example 4.1, $H_o = I + S$, Algorithms I-VIII.

i	x	y	z	u	f	$g^T g$
0	4.0000	4.0000	4.0000	4.0000	0.828×10^{3}	0.4×10^{5}
1	-4.6710	-0.5111	5.2264	10.496	0.539×10^{3}	0.2×10^{5}
2	0.1399	0.0073	-0.0056	0.0155	0.237×10^{0}	0.1×10^{1}
3	0.0685	-0.0497	0.3189	-0.2015	0.166×10^{-1}	0.2×10^{-2}
4	0.5000	-0.5000	0.5000	0.0000	$^*0.856 \times 10^{-40}$	$^*0.2 \times 10^{-37}$

5. Quadratic Functions: Discussion and Conclusions

The numerical example of the previous section shows that, if high-precision arithmetic and high accuracy in the search are used, all the algorithms behave identically for a given initial matrix: they all produce the same sequence of points and lead to the minimal point in no more than n iterations, where n is the number of variables. In view of these results, Algorithm IX with the initial matrix given by (15-1) should be favored for the minimization of a quadratic function. If Eqs. (1), (10), (15-1) are combined, the search direction p_i simplifies to

$$p_i = g_i + \frac{g_i^T g_i}{g_{i-1}^T g_{i-1}} p_{i-1} \tag{22}$$

This simplified version characterizes the Fletcher-Reeves algorithm and involves the least amount of computational work per iteration.

6. Nonquadratic Functions: General Considerations

To apply Algorithms I-IX to the minimization of a nonquadratic function, the following considerations are in order.

6.1. Starting Condition. For Algorithms I-VIII, any initial matrix H_o, such that the matrix $(H_o + H_o^T)/2$ is either positive definite or negative definite, can be used. For Algorithm IX, the matrix H_o must be symmetric.

6.2. Stopping Condition. An algorithm is stopped when Ineq. (13) is satisfied. For a nonquadratic function, this may correspond to either a minimal point or a nonminimal stationary point. The probability of occurrence of a nonminimal point can be reduced materially if the constant ϵ_1 in Ineq. (13) is very small.

6.3. Precision Requirement. The precision requirement of Section 3 also holds for nonquadratic functions. This ensures fast convergence in the neighborhood of the minimal point.

Sometimes, strict satisfaction of Ineq. (14) may require considerable computational work in the one-dimensional search. In order to avoid this situation, the one-dimensional search is also stopped when the inequality

$$|\Delta\alpha_i| \leq \epsilon_3 \, |\alpha_i| \tag{23}$$

is satisfied. Here, α_i is the nominal stepsize, $\Delta\alpha_i$ is the correction to the nominal value, and ϵ_3 is a prescribed small number.

6.4. Restarting Condition. In general, the minimal point of a nonquadratic function cannot be reached in n iterations, and further iterations may be needed. This being the case, an algorithm can be restarted by setting

$$H_i = H_o \tag{24}$$

In this connection, a necessary restarting condition and an additional restarting condition are discussed below.

Necessary Restarting Condition. Restart of an algorithm is necessary when the inequality

$$|g_i^T p_i| \leq \epsilon_2 \tag{25}$$

is satisfied, while the stopping condition (13) is violated. In Ineq. (25), ϵ_2 dènotes a prescribed small number, which is identical with that in Ineq. (14). When Ineq. (25) is satisfied, the starting point for the one-dimensional search satisfies automatically the search stopping condition (14), so that $\alpha_i = 0$.

For the simplified algorithms VIII and IX, restart at the nth or (n+1)th point, counted from the previous starting or restarting point, is also necessary. An intuitive explanation is given below. In Eqs. (9) and (10), the initial matrix H_o appears explicitly on the right-hand side and, therefore, the basic cycle is composed of n iterations counted from the point where H_o is used. On the other hand, this restarting condition is not necessary for Algorithms I-VII. Since the initial matrix H_o does not appear explicitly in the updating formulas (2)-(8), the matrix H_i can be regarded as the initial matrix from the ith iteration on.

Additional Restarting Condition. The realization of the quadratic convergence for Algorithms I-IX depends not only on the previous precision requirement but also on how well the function can be approximated by a quadratic. If the behavior of the function deviates considerably from that of a quadratic, Algorithms I-IX may exhibit slow convergence. Therefore, these algorithms may be restarted as soon as some severe violation of the quadratic behavior is detected.

For a quadratic function, the following relation can be shown to hold:

$$f_i - f_{i-1} + (\alpha_{i-1}/2)(g_{i-1}^T p_{i-1} + g_i^T p_{i-1}) = 0 \tag{26}$$

For a nonquadratic function, the left-hand side of Eq. (26) is other than zero, and its absolute value can be used as a measure of the violation of the quadratic behavior. This being the case, an algorithm may be restarted when the inequality

$$|f_i - f_{i-1} + (\alpha_{i-1}/2)(g_{i-1}^T p_{i-1} + g_i^T p_{i-1})| \geq \varepsilon_4 \tag{27}$$

is satisfied, where ε_4 is a prescribed small number.

7. Nonquadratic Functions: Numerical Example

In this section, a particular nonquadratic function is studied using a Burroughs B-5500 computer and double-precision arithmetic. The experimental conditions are now indicated.

Starting Condition. As noted before, an algorithm may be restarted several times before convergence is achieved. If one assumes that the second-derivative matrix is not available, the most reliable matrix is the identity matrix, that is,

$$H_o = I \tag{28}$$

This initial matrix is employed in the subsequent example.

Stopping Condition. An algorithm is stopped when Ineq. (13) is satisfied with

$$\epsilon_1 = 10^{-12} \tag{29}$$

One-Dimensional Search. The one-dimensional search is stopped when Ineq. (14) or Ineq. (23) is satisfied with

$$\epsilon_2 = 10^{-16} \ , \quad \epsilon_3 = 10^{-6} \tag{30}$$

Restarting Condition. Four restarting conditions are studied: (A) restart when Ineq. (25) is satisfied, with ϵ_2 given by Eq. (30-1); (B) restart either when Ineq. (25) is satisfied or at the nth point, counted from the previous starting or restarting point; (C) restart either when Ineq. (25) is satisfied or at the (n+1)th point, counted from the previous starting or restarting point; and (D) restart when either Ineq. (25) or Ineq. (27) is satisfied. Concerning Ineq. (27), three values of ϵ_4 are considered, namely,

$$\epsilon_4 = 0.01 \quad , \quad \epsilon_4 = 0.1 \quad , \quad \epsilon_4 = 1 \tag{31}$$

Example 7.1. We consider the problem of minimizing the function (Ref. 2)

$$f = 100(x^2 - y)^2 + (x - 1)^2 + (z - 1)^2 + 90(z^2 - u)^2$$

$$+ 10.1[(y - 1)^2 + (u - 1)^2] + 19.8(y - 1)(u - 1) \tag{32}$$

This function admits the minimum f = 0 at the point defined by

$$x = 1 \quad , \quad y = 1 \quad , \quad z = 1 \quad , \quad u = 1 \tag{33}$$

and the nonminimal stationary value f = 7.876 at the point defined by

$$x = -0.9679 \quad , \quad y = 0.9471 \quad , \quad z = -0.9695 \quad , \quad u = 0.9512 \tag{34}$$

The nominal point chosen for starting an algorithm is the point of coordinates

$$x = -3 \quad , \quad y = -1 \quad , \quad z = -3 \quad , \quad u = -1 \tag{35}$$

The number of iterations for convergence under different restarting conditions is given in Table 3.

Table 3

Number of iterations for convergence (Example 7.1).

Algorithm	A	B	C	D		
				$\epsilon_4=0.01$	$\epsilon_4=0.1$	$\epsilon_4=1$
I	40	60	45	27	24	21
II	40	60	45	27	24	21
III	40	60	45	27	24	21
IV	40	60	45	27	24	21
V	64	64	64	32	31	30
VI	64	64	64	32	31	30
VII	64	64	64	32	31	30
VIII	>100	74	93	39	41	39
IX	>100	38	28	74	89	57

8. Nonquadratic Functions: Discussion and Conclusions

From the numerical results of the previous section, the following comments are in order:

(a) Algorithms I-IV behave identically for a given restarting condition: they all produce the same sequence of points and lead to the minimal point in the same number of iterations. These algorithms belong to a class which, for a quadratic function, yields a final matrix H_n equal to the inverse of the second-derivative matrix. As far as the sequence of points is concerned, any one of the four algorithms can be considered to be representative of the entire class.

(b) Algorithms V-VII behave identically for a given restarting condition: they all produce the same sequence of points and lead to the minimal point in the same number of iterations. These algorithms belong to a class which, for a quadratic function, yields a final matrix H_n equal to the null matrix. As far as the sequence of points is concerned, any one of the three algorithms can be considered to be representative of the entire class.

(c) In general, Algorithms I-IV exhibit faster convergence with restarting condition A than with restarting conditions B and C. In the example, Ineq. (25) was never satisfied. This suggests that operation without restart should be preferred to that with restart at the nth or (n+1)th point, counted from the previous starting or restarting point.

(d) For Algorithms V-VII, restarting conditions A, B, C, yield the same results: the elements of the H-matrix at the nth point, counted from the previous starting or restarting point, become very small (in the absence of computational errors, the H-matrix becomes a null matrix). Therefore, restart every n points is necessary.

(e) In general, for Algorithms VIII and IX, restarting conditions B and C yield faster convergence than restarting condition A. This suggests that restart at the nth or (n+1)th point, counted from the previous starting or restarting point, is to be favored.

(f) In general, restarting condition D improves convergence of Algorithms I-IV. It also improves convergence of Algorithms V-VIII. Although the number of iterations for convergence depends on the choice of ϵ_4, there exists a wide range of values of ϵ_4 for which convergence under restarting condition D is faster than convergence under restarting conditions A, B, C. Note that, if $\epsilon_4 \to \infty$, restarting conditions D and A become identical; also, if $\epsilon_4 \to 0$, restarting condition D yields the ordinary gradient algorithm.

References

1. HUANG, H.Y., Unified Approach to Quadratically Convergent Algorithms for Function Minimization, Rice University, Aero-Astronautics Report No. 64, 1969.

2. PEARSON, J.D., On Variable Metric Methods of Minimization, Research Analysis Corporation, Technical Paper No. RAC-TP-302, 1968.