Proceedings of the World Congress on Engineering and Computer Science 2017 Vol I
WCECS 2017, October 25-27, 2017, San Francisco, USA

A Rational Model for Curvature Quasi-Newton Methods

Issam A.R. Moghrabi

***Abstract*—Multi-step Methods of the quasi-Newton type, derived in [10,11], have shown promising numerical improvement for solving nonlinear unconstrained optimization problems over methods based on the linear Secant equation. Minimum curvature methods [6] have resulted in further performance gains as they ensure the 'smoothness' of the interpolating curves in the multi-step methods. In this work, we derive new methods of this type using a rational model. The results of the numerical tests reveal further gains over the methods developed earlier.**

***Index Terms*— Unconstrained optimization, quasi-Newton methods, multi-step methods, curvature algorithms**

I. Introduction

THIS work addresses problems of the form:

$$minmize\ f(x), x \in R^n, where\ f: R \to R^n.$$

Quasi-Newton methods require an approximation to the Hessian matrix that is updated at each iteration. Given B_i, the current approximation to the Hessian, the new Hessian approximation, B_{i+1} is chosen to satisfy the standard Secant equation:

$$B_{i+1} s_i = y_i \tag{1}$$

where

$$s_i = x_{i+1} - x_i,$$

and

$$y_i = g(x_{i+1}) - g(x_i) = g_{i+1} - g_i.$$

This matrix is used in the computation of the search direction as follows

The BFGS formula [1,2,3] is the mostly used update that satisfies the Secant equation and that seems to work well with inexact line search algorithms [12], [13] [14]. This rank-two update is given by

$$B_{i+1} = B_i - \frac{B_i s_i s_i^T B_i}{s_i^T B_i s_i} + \frac{y_i y_i^T}{s_i^T y_i}.$$

A quasi-Newton method algorithm is generally outlined as ([9]):

a. Start with any estimate x_0, of the minimum.
b. Start with a symmetric positive-definite matrix H_0 (usually $H_0=I$).
c. $i=0$
d. Find $g_0=g(x_0)$;
e. Repeat

{
- compute $p_i = - H_i g_i$;
- Determine the step length α_i using some line search technique (e.g., Cubic Interpolation [2],[8]);
- $x_{i+1} = x_i + \alpha_i p_i$;
- compute $s_i = x_{i+1} - x_i$ and $y_i = g_{i+1} - g_i$;
- $i = i + 1$

}

Until $\|g_i\|_2 < \varepsilon$, where $\varepsilon \in R$ $(\varepsilon > 0)$ is a convergence parameter.

The paper introduces some of the successful multi-step algorithms that will be used in the numerical benchmarking in this paper. Then the rational model employed utilized in the derivation of the new minimum curvature method is presented. The derivation aims at ensuring that the interpolating curve, on which the idea of the multi-step methods is based, is `smooth'. We finally present the numerical results.

II. Multi-step Quasi-Newton Methods

In the standard Secant equation, a straight line L is used to find a new iterate x_{i+1}, given the previous iterate x_i, while in the multi-step methods higher order polynomials are used.

Let $\{x(\tau)\}$ or X denote a differentiable path in R^n, where $\tau \in R$. Then upon applying the chain rule to the gradient vector $g(x(\tau))$ in order to find the derivative of the gradient g with respect to τ we get

$$\frac{dg}{d\tau} = G(x(\tau))\frac{dx}{d\tau}. \tag{2}$$

Thus, at any point on the path X, the Hessian G must satisfy (2) for any value of τ. More specifically for $\tau = \tau_c$, where $\tau_c \in R$. This will result in the following relation

$$\frac{dg}{d\tau}|_{\tau=\tau_c} = G(x(\tau))\frac{dx}{d\tau}|_{\tau=\tau_c}.$$

By analogy with the Secant equation, the aim is to derive a relation satisfied by the Hessian at the new iterate x_{i+1}, we choose a value for the parameter τ, namely τ_m, that corresponds to the most recent iterate as follows

$$g'(\tau_m) = B_{i+1} x'(\tau_m)$$

or, equivalently,

$$w_i = B_{i+1} r_i, \tag{3}$$

where the vectors r_i and w_i are defined in terms of the m most recent step vectors $\{s_k\}_{k=i-m+1}^{i}$ and the m most recent

Manuscript received June 30, 2017; revised July 30, 2017. This work was supported in part by the KUWAIT Foundation for the Advancement of Science under Grant Kfas 2346.

I. A. R. Moghrabi, an IAENG member, is a full professor at the Gulf University for Science and Technology, MIS Department, College of Business Administration, Kuwait. e-mail: moughrabi.i@gust.edu.kw.

gradient difference vectors $\{y_k\}_{k=i-m+1}^{i}$ respectively, as follows

$$r_i = \sum_{j=0}^{m-1} s_{i-j}\{\sum_{k=m-j}^{m} L_k'(\tau_m)\}$$

and

$$w_i = \sum_{j=0}^{m-1} y_{i-j}\{\sum_{k=m-j}^{m} L_k'(\tau_m)\},$$

where

$$L_k'(\tau_m) = (\tau_k - \tau_m)^{-1}[(\tau_m - \tau_j)/(\tau_k - \tau_j)], k < m,$$
$$L_m'(\tau_m) = \sum_{j=0}^{m-1}(\tau_m - \tau_j)^{-1},$$

are the standard Lagrange polynomials.

III. STRATEGIES FOR THE PARAMETERIZATIONS OF THE CURVES

Ford and Moghrabi [10],[11] examined several choices for the parameters $\{\tau_k\}_{k=0}^{m}$ where such choices influence the structure of the interpolating curve. The choices examined in [6] have proven to be far more efficient than the natural choice

$$\tau_k = k - m + 1, \text{ for } k = 0, 1, 2, \ldots, m.$$

Methods corresponding to this choice are labeled as the "unit-spaced" methods. Of the approaches considered in [6,7,11], we elect here the most numerically successful choice that is based on, what the authors termed as the Accumulative Approach.

The choices made for the parameters $\{\tau_k\}_{k=0}^{m}$ rely on some metric of the following form

$$\phi_M(z_1, z_2) = [(z_1 - z_2)^T M (z_1 - z_2)]^{1/2},$$

where M is a symmetric positive-definite matrix.

The Accumulative approach chooses one of the iterates, say x_j, as a "base-point" and set it to 0. Then, any value τ_k corresponds to the point $x_{i-m+k+1}$ for any k except for $k=j$ is computed by accumulating the distance (measured by the chosen metric Φ_M between each two consecutive pair of points in the sequence from $x_{i-m+j+1}\}$ to $x_{i-m+k+1}$. Therefore, any value τ_k, for $k=0,1,...,m$, is obtainable using

$$\tau_k = -\sum_{P=k+1}^{j}[\phi_M(x_{i-m+p+1}, x_{i-m+p})], k < j, \tag{4}$$
$$= 0 \qquad , k = j,$$
$$= \sum_{P=j+1}^{k}[\phi_M(x_{i-m+p+1}, x_{i-m+p})], k > j.$$

This approach will yield values of τ that satisfy

$$\tau_k < \tau_{k+1} \;, \quad for\; k = 0, 1, \ldots, m-1,$$

Under the assumption that no consecutive points overlap.

It is those values of the parameters $\{\tau_k\}$ that are used to compute the vectors $x'(\tau_m)$ and $g'(\tau_m)$ in (3) (or vectors r_i and w_i, respectively). The two vectors r_i and w_i are then used to compute the new Hessian approximation B_{i+1} satisfying (3).

It should be noted that different choices of the metric matrix M in Φ_M will result in different algorithms. Numerically speaking, Ford and Moghrabi [11] indicate that values of $m > 2$ do not seem to result in substantial gains in performance. This may be due to the non-smoothness of the interpolant. Thus, $m = 2$ is used in this paper. Methods using $m = 2$ are referred to as 2-step methods as they use vectors from the two most recent iterations to update the Hessian approximation at each iteration.

Possible choices for the matrix M examined in [11], include $M = I$, $M = B_i$, $M = B_{i+1}$. The update done at each iteration generally satisfies:

$$H_{i+1}\left(y_i - \tfrac{\delta^2}{2\delta+1} y_{i-1}\right) = s_i - \tfrac{\delta^2}{2\delta+1} s_{i-1} \tag{5}$$

where

$$\delta = \tfrac{\tau_2 - \tau_1}{\tau_1 - \tau_0}.$$

For the algorithm used in the numerical comparisons here, the particular choices of the τ values are as follows

$$\tau_0 = -(\|s_i\|_2 + \|s_{i-1}\|_2), \tau_2 = 0, \tau_1 = -\|s_i\|_2.$$

The new B-version BFGS formula is given by:

$$B_{i+1}^{Multi-Step} = B_i + \tfrac{w_i w_i^T}{w_i^T r_i} - \tfrac{B_i r_i r_i^T B_i}{r_i^T B_i r_i} \quad . \tag{6}$$

IV. A NEW CURVATURE ALGORITHM (MC)

The idea here is to determine the parameters $\{\tau_k\}$, such that the curve that interpolates the iterates as well as the gradient points has a minimum curvature. The minimum curvature idea in a similar context was first proposed in [6]. We follow a different approach here as we propose a rational model for the interpolating curve. The rational model is given as follows:

$$x(\tau,\theta) = q(\tau)) / (1+\theta\tau), \tag{7}$$

where θ serves as a tuning parameter, and q(τ) is a quadratic expressed as:

$$q(\tau) = \sum_{k=0}^{2} L_k(1 + \theta\tau_k) x_{i-m+k+1} \tag{8}$$

where $L_j(\tau)$ is the Lagrange polynomial of degree 2 associated with the abscissae $\{\tau\}_{k=0}^{2}$. Equation (6) satisfies the relation

$$x(\tau_k, \theta) = x_{i+k+1} \quad , for \;\; k = 0, 1, 2.$$

We now determine an expression for the parameter θ such that the curvature of the interpolant is minimized. From (6) and (7) we obtain

$$x'(\tau,\theta) \equiv \frac{(1+\theta\tau)q' - \theta q}{(1+\theta\tau)^2} \tag{9}$$

or, equivalently,

$$x'(\tau_2,\theta) = (s_i(\tau_2 - \tau_1)((-\tau_1 - \tau_0 + 2\tau^2) + \theta(\tau_2^2 - \tau_0\tau_1)) - s_{i-1}((\tau_2 - \tau_1)^2(1 + \vartheta\tau_0))) / ((1 + \vartheta\tau_2)(\tau_2 - \tau_0)(\tau_1 - \tau_0)(\tau_2 - \tau_1)).$$

Also,

$$x''(\tau_2,\theta) = \frac{1}{(1+\theta\tau_2)^2 v}(\theta^2(s_i\alpha + s_{i-1}\beta) + \theta(s_i\gamma + s_{i-1}\lambda) + (s_i\pi + s_{i-1}\varepsilon)), \tag{10}$$

where

$$v = (\tau_1 - \tau_0)(\tau_2 - \tau_0)(\tau_2 - \tau_1),$$
$$\alpha = 2\tau_0\tau_1(\tau_1 - \tau_0), \beta = 2\tau_0\tau_1(\tau_1 - \tau_2), \gamma = 2(\tau_1^2 - \tau_0^2)$$
$$\lambda = 2(\tau_1^2 + \tau_0\tau_1 - \tau_0\tau_2 - \tau_1\tau_2), \pi = 2(\tau_1 - \tau_0) \text{ and } \varepsilon = 2(\tau_1 - \tau_2).$$

If we define

$$\sigma_j \stackrel{def}{=} \|s_j\|_M^2 \ge 0, \sigma_{j-1} \stackrel{def}{=} \|s_{j-1}\|_M^2 \ge 0 \text{ and } \mu_j \stackrel{def}{=} s_{j-1}^T M s_j,$$

then the curvature function is given by

$$\phi(\tau_2,\theta) = \|x''(\tau_2,\theta)\|_M^2$$

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

or, more explicitly, the aim to is solve for θ the following cubic equation

$$\phi(\tau_2,\theta) = \frac{1}{(1+\theta\tau_2)^5 \nu^2}[\theta^3(4k_1 - k_2\tau_2) + \theta^2(3k_2 - 2k_3\tau_2) + \theta(2k_3 - 3k_4\tau_2) + (k_4 - 4k_5\tau_2)] = 0,$$

where

$$k_1 = \alpha^2\sigma_i + \beta^2\sigma_{i-1} + 2\alpha\beta\mu_i,$$
$$k_2 = 2\alpha\gamma\sigma_i + 2\beta\lambda\sigma_{i-1} + 2(\alpha\lambda + \beta\gamma)\mu_i,$$
$$k_3 = (\gamma^2 + 2\alpha\pi)\sigma_i + (\lambda^2 + 2\beta\varepsilon)\sigma_{i-1} + 2(\gamma\lambda + \alpha\varepsilon + \beta\pi)\mu_i,$$
$$k_4 = 2\gamma\pi\sigma_i + 2\lambda\varepsilon\sigma_{i-1} + 2(\gamma\varepsilon + \lambda\pi)\mu_i,$$
$$k_5 = \pi^2\sigma_i + \varepsilon^2\sigma_{i-1} + 2\pi\varepsilon\mu_i.$$

Given the values for the parameters τ_0, τ_1 and prescribed in (5), we proceed to determine the minimum curvature at τ_2, the parameter corresponding to the most recent iterate. From (10), we have

$$\phi(\tau_2,\theta) = \frac{\left((\tau_1-\tau_0)^2\sigma_i + \left(\tau_1^2\right)\sigma_{i-1} + 2\tau_1(\tau_1-\tau_0)\mu_i\right)}{\tau_0\tau_1(\tau_1-\tau_0)} \begin{bmatrix} \theta^3(16\tau_0^2\tau_1^2) + \theta^2(24\tau_0\tau_1(\tau_1+\tau_0)) + \\ \theta(8(\tau_1+\tau_0)^2 + 16\tau_0\tau_1) + 8(\tau_1+\tau_0) \end{bmatrix} = 0,$$

whose three real roots are given by

$$\theta_1 = (\|s_{i-1}\|_2 + \|s_i\|_2)^{-1}, \theta_2 = (\|s_i\|_2)^{-1}$$

And

$$\theta_3 = -1/2((\|s_{i-1}\|_2 + \|s_i\|_2)^{-1} + (\|s_i\|_2)^{-1}) .$$

The roots θ_1 and θ_2 are points of singularity for the update (5). As for θ_3, the curvature expression is given by

$\Phi(\tau_2,\theta_3) = -28(-\|s_i\|_2 - (\|s_i\|+)^2 + 16/(\tau_0\tau_1)$.

For θ_3 to correspond to a minimum the values of τ_0 and τ_1 should both be less than one.

V. NUMERICAL RESULTS

Our numerical experiments were conducted on sixty functions classified into "low" ($2 \le n \le 15$), "medium" ($16 \le n \le 45$), and "high" ($46 \le n \le 80$) dimension as in [7], [13] and [14]. Each problem was tested with four different starting-points each. The total number of problems tested is a total of 876. The overall numerical results are presented in Table 1. Tables 2 to 5 show the total results for each dimension. The results report the total number of function/gradient evaluations, the total iterations, the scores and the total execution time. Those are reported in Table 1 and further illustrated in Figures 1 and 2. For each problem, the method with the least number of gradient/function evaluations is awarded one point adding to the "Scores". In our tests, we have employed a safeguarded cubic interpolation line search technique where a new estimate to the minimum, x_{i+1}, is accepted if it satisfies the following two standard stability conditions (see [2,4,5], for example):

$$f(x_{i+1}) \le f(x_i) + 10^{-4} s_i^T g(x_i)$$

and

$$s_i^T g(x_{i+1}) > 0.9 s_i^T g(x_i).$$

Algorithm M1 corresponds to the standard BFGS method while method A1 corresponds to following choices of the parameters in (6):

$$\tau_0 = -(\|s_i\|_2 + \|s_{i-1}\|_2), \tau_2 = 0, \tau_1 = -\|s_i\|_2.$$

The results presented here show clearly that the new algorithm MC method exhibits a superior numerical performance, by comparison with the other algorithms against which it is compared. In general, the curvature methods show also numerical improvement in the low dimension.

Table 1 : Overall Results for 648 problems

Method	Evaluations	Iterations	Time (sec.)	Scores
M1	86401 (100.00%)	73090 (100.00%)	39171.18 (100.00%)	89
A1	76164 (88.15%)	61335 (83.92%)	31474.71 (80.35%)	241
MC	75975 (87.93%)	60490 (82.76%)	31247.42 (79.77%)	546

Table 2 : Results for dimensions from 2 to 15 (440 problems)

Method	Evaluations	Iterations	Time (sec.)	Scores
M1	25589 (100.00%)	21648 (100.00%)	319.83 (100.00%)	73
A1	24494 (95.72%)	19525 (90.19%)	267.25 (83.56%)	168
MC	24343 (95.13%)	18991 (87.72%)	265.31 (82.95%)	199

Table 3 : Results for dimensions from 16 to 45 (240 problems)

Method	Evaluations	Iterations	Time (sec.)	Scores
M1	27058 (100.00%)	23844 (100.00%)	3429.78 (100.00%)	44
A1	22995 (84.98%)	19578 (82.11%)	2745.05 (80.04%)	85
MC	22001 (81.31%)	18650 (78.22%)	2593.11 (75.60%)	111

Table 4 : Results for dimensions from 46 to 80 (134 problems)

Method	Evaluations	Iterations	Time (sec.)	Scores
M1	21146 (100.00%)	17431 (100.00%)	13426.87 (100.00%)	25
A1	18009 (85.17%)	14122 (81.02%)	10835.33 (80.70%)	46
MC1	17930 (84.79%)	13819 (79.27%)	10315.66 (76.82%)	63

Table 5 : Results for dimensions from 81 to 100 (62 problems)

Method	Evaluations	Iterations	Time (sec.)	Scores
M1	12608 (100.00%)	10167 (100.00%)	21994.7 (100.00%)	12
A1	10666 (84.60%)	8110 (79.77%)	17627.08 (80.14%)	19
MC1	10091 (80.04%)	8011 (78.79%)	17453.34 (79.35%)	31

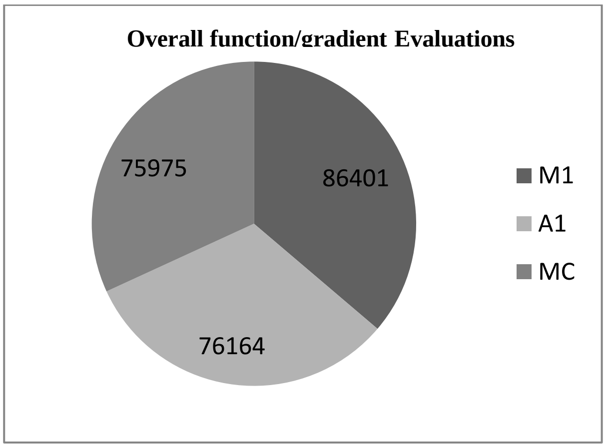

Fig. 1. Overall function and gradient evaluations for each method.

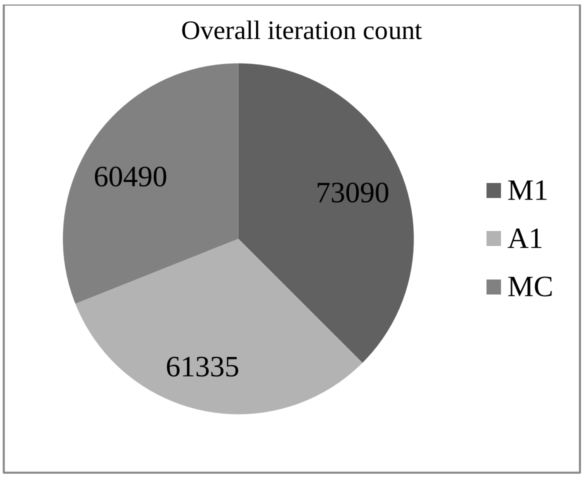

Fig. 2. Overall iteration count for each method.

VI. CONCLUSIONS AND SUGGESTIONS FOR FURTHER WORK

A new method for determining the parameterization of the interpolating curves in the two-step quasi-Newton methods was derived here. The parameters that influence the structure of the interpolating curve are obtained by minimizing a cubic equation at each iteration. It was revealed that the cubic curvature function to be minimized has cheaply computable roots. The numerical results showed that such approach yields a substantial improvement in numerical performance over the standard BFGS method. The new method has slightly improved over our best known multi-step accumulative method *A1*.

Future research might focus on issues like:

- Is there an optimal choice for the curve parameters τ?
- Can these methods improve the numerical performance if they are applied to solving systems of non-linear equations?
- Further Study of the convergence properties of the methods.

REFERENCES

[1] M. Abidi, A. Gribok and J. Paik, Unconstrained Optimization, *Optimization Techniques in Computer Vision*, pp.69-92 (2016).

[2] C.G. Broyden, The convergence of a class of double-rank minimization algorithms - Part 2: The new algorithm, *J. Inst. Math. Applic*., 6, 222-231 (1970).

[3] J.E. Dennis and R.B. Schnabel, Minimum change variable metric update formulae, SIAM Review ,21, 443-459 (1979).

[4] R. Fletcher, A new approach to variable metric algorithms, *Comput. J.*, 13, 317-322 (1970).

[5] R. Fletcher, *Practical Methods of Optimization*, Second Edition, Wiley, New York, (1987).

[6] J.A. Ford and I.A. Moghrabi, Minimum curvature quasi-Newton methods, *Computers Math. Applic.*, 31,179-186 (1996).

[7] J.A. Ford and I.A. Moghrabi, Function-value multi-step quasi-Newton methods, Department of Computer Science (University of Essex) Technical Report CSM-270 (1996).

[8] J.A. Ford and I.A. Moghrabi, On the use function value in multi-step methods, Department of Computer Science (University of Essex) Technical Report CSM-271 (1996).

[9] J.A. Ford and I.A. Moghrabi, Further investigation of multi-step quasi- Newton methods, *Scientia Iranica*, 1, 327-334 (1995).

[10] J.A. Ford and I.A. Moghrabi, Multi-step quasi-Newton methods for optimization, *J. Comput. Appl. Math*., 50, 305-323 (1994).

[11] J.A. Ford and I.A. Moghrabi, alternative parameter choices for multi-step quasi-Newton methods, *Optim. Meth. Software*, 2, 357-370 (1993).

[12] P. Gill, W. Murray and M. Wright, *Numerical Linear Algebra on Optimization*, volume I, Addison-Wesley, U.S.A., (1991).

[13] D. Goldfarb, A family of variable metric methods derived by variational means, *Maths. Comp*., 24, 23-26 (1970).

[14] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization, *Maths. Comp*., 24, 647-656 (1970).

ISBN: 978-988-14047-5-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)