Applied Mathematics and Computational Intelligence
Volume 14, No. 2, 2025 [149-155]

Modified Quasi-Newton Method via Linear Gradient Flow System

Chui Ying Yap[1], Wah June Leong[1]*, Keat Hee Lim[1]

[1]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

* Corresponding author: leongwj@upm.edu.my

Received: 4 November 2024
Revised: 3 February 2025
Accepted: 24 February 2025

ABSTRACT

This paper presents an efficient approach to unconstrained optimization, built upon the gradient flow derived from the objective function. By incorporating a linear approximation on the gradient flow, the proposed method introduces a modified and enhanced BFGS update that improves upon the standard BFGS method. To validate its effectiveness, the approach is implemented within a line search framework. Numerical results underscore the significance of the modified BFGS method, demonstrating superior performance compared to the standard BFGS method, thereby offering a valuable advancement in unconstrained optimization techniques.

Keywords: BFGS update, gradient flow, line search, quasi-Newton method.

1 INTRODUCTION

This paper considers the following unconstrained optimization problems:

$$\min_{x \in R^n} f(x), \tag{1}$$

where the objective function $f(x)$ is assumed to be twice continuously differentiable for all x in R^n.

Considerable progress has been made in developing a robust suite of algorithms for numerically solving (1), with most of these methods being iterative [8, 10–13]. In an iterative algorithm, an initial point x_0 is provided, and a new iterative point x_{k+1} is computed based on the information available at the current point x_k. Ideally, the sequence $\{x_k\}$ will converge to the solution of the optimization problem, x^* satisfying the first- and second-order conditions for a local minimum, $\nabla f(x^*) = 0$, and $\nabla^2 f(x^*)$ is positive definite (or at least positive semi-definite). A broad class of iterative algorithms is the so-called line search methods, which is expressible as:

$$x_{k+1} = x_k + \alpha_k d_k, \ \ \forall k \geq 0, \tag{2}$$

where α_k is a stepsize and d_k denotes a search direction.

Quasi-Newton methods [3, 5, 9, 15, 16] are widely used when the analytical expression of the second derivative of $f(x)$, known as the Hessian, is difficult or costly to compute, or store. Instead, these methods compute a search direction as follows:

$$d_k = -B_k^{-1}\nabla f(x_k),$$

where B_k is an $n \times n$ symmetric matrix approximating the Hessian through an update formula, such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula [5, 7, 9, 14, 15]:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{3}$$

where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$.

Using the Sherman-Morrison formula [6], the update formula for the inverse Hessian approximation $H_k = B_k^{-1}$ can be written as:

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}. \tag{4}$$

After determining the search direction, a step length α_k is found by performing a line search to ensure global convergence. and the process repeats until the sequence $\{x_k\}$ meets the termination criterion.

2 LINEAR GRADIENT FLOW SYSTEM AND MODIFIED QUASI-NEWTON UPDATE

Historically, partial differential equations (PDEs) and optimization have emerged from distinct branches of mathematics and have often been considered separate areas. However, their interconnections have captivated the interest of many researchers. In 1941, Courant [4] was the first to introduce the gradient method for solving variational PDEs, proposing the gradient flow system:

$$\dot{x}(t) = -\nabla f(x(t)), \tag{5}$$

with the initial condition $x(0) = x_0$, to find an equilibrium point x^* such that

$$\nabla f(x^*) = 0.$$

To solve this system using ordinary differential equations (ODEs), one can discretize time in (5) to derive a difference equation. As the focus is on the long-term behaviour of (5) rather than precise intermediate solutions, the implicit (backward) Euler method is employed due to its unconditional stability, regardless of the time step size. Applying this method to (5) yields:

$$x_{k+1} = x_k - \alpha_k \nabla f(x_{k+1}), \; k = 0, 1, 2, \ldots. \tag{6}$$

Equation (6) can be viewed as a line search method, where α_k is a step length and $d_k = -\nabla f(x_{k+1})$ is a search direction. In contrast, using the explicit (forward) Euler method in (5) would yield the steepest descent method, where $d_k = -\nabla f(x_k)$, though it requires careful time step selection for stability.

A practical challenge with (6) is that it requires computing the gradient at the unknown future point, x_{k+1}. Approximating $\nabla f(x_{k+1})$ using first-order Taylor expansion gives

$$d_k = -\nabla f(x_{k+1}) \approx -(\nabla f(x_k) + \alpha_k \nabla^2 f(x_k) d_k),$$

where $\nabla^2 f(x_k)$ is the Hessian of f at x_k. Thus, the iterative update, after some rearrangement becomes

$$x_{k+1} = x_k - \alpha_k \left(I + \alpha_k \nabla^2 f(x_k)\right)^{-1} \nabla f(x_k). \tag{7}$$

The implicit Euler method's stability allows any arbitrarily small α_k, though convergence speed can be improved via variable step length. Obtaining $\left(I + \alpha_k \nabla^2 f(x_k)\right)^{-1}$ exactly in (7) is generally infeasible as it may require multiple linear solves for various step lengths. Therefore, we propose approximation using an updating matrix $\hat{B}_{k+1}$ that satisfies the quasi-Newton equation

$$\hat{B}_{k+1} s_k \approx \left(I + \alpha_k \nabla^2 f(x_k)\right) s_k = s_k + \alpha_k \nabla^2 f(x_k) s_k. \tag{8}$$

Note that the mean value theorem would imply that there exists a point $\tau \in (x_k, x_{k+1})$ such that

$$\nabla^2 f(\tau) s_k = y_k,$$

and hence, for a sufficiently small α_k, we have

$$\nabla^2 f(x_k) s_k \approx \nabla^2 f(\tau) s_k = y_k.$$

Therefore, we can rewrite the (modified) quasi-Newton equation (8) as

$$\hat{B}_{k+1} s_k = \hat{y}_k, \tag{9}$$

where $\hat{y}_k = s_k + \alpha_k y_k$. The modified inverse BFGS update based on (9) is then given by

$$\hat{H}_{k+1} = \left(I - \frac{s_k \hat{y}_k^T}{\hat{y}_k^T s_k}\right) \hat{H}_k \left(I - \frac{\hat{y}_k s_k^T}{\hat{y}_k^T s_k}\right) + \frac{s_k s_k^T}{\hat{y}_k^T s_k}, \tag{10}$$

where $\hat{H}_0 = I$.

To ensure convergence, the Armijo backtracking line search [2] is the most used strategy where its algorithm is given as follows:

Algorithm 1 (Armijo backtracking line search):
Step 0. Given some constants $c_1, c_2 \in (0, 1)$. Set $\alpha_k = 1$.
Step 1. Evaluate the following relation

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k \nabla f(x_k)^T d_k. \tag{11}$$

Step 3. If (11) is not satisfied, choose a new $\alpha_k \leftarrow c_2\alpha_2$ and go to Step 1. Otherwise, set $x_{k+1} = x_k + \alpha_k d_k$.

Together with the Armijo line search (**Algorithm 1**), we can now state our main algorithm:

Algorithm 2
Step 0. Choose an initial point $x_0 \in R^n$, and $\widehat{H}_0 = I$. Set $k = 0$.
Step 1. Compute $\nabla f(x_k)$. If the stopping criterion $\|\nabla f(x_k)\| \leq \epsilon$ is reached, then stop. Else go to Step 2.
Step 2. Compute $d_k = -\widehat{H}_\mathrm{k} \nabla f(x_k)$, using (4) and calculate α_k using **Algorithm 1**. Set $x_{k+1} = x_k - \alpha_k \widehat{H}_k \nabla f(x_k)$ and update $\widehat{H}_{k+1}$.
Step 3. Set $k \coloneqq k + 1$ and return to Step 1.
For global convergence to a stationary point (i.e., a point where $\nabla f(x^*) = 0$), we shall make the following assumptions:

Assumption 1.
i. The objective function $f(x)$ is twice continuously differentiable,
ii. There exist positive constants $m_1 \leq m_2$ such that
$m_1\|s_k\|^2 \leq {s_k}^T y_k \leq m_2\|s_k\|^2, \ \forall k \geq 0$.
iii. The gradient $\nabla f(x)$ is Lipschitz continuous, i.e. there exists a positive constant M such that:

$$\|\nabla f(x) - \nabla f(y)\| \leq M \|x - y\|, \ \forall x, y \in R^n.$$

Theorem 1. (see [3])
Under Assumption 1, if the Armijo condition (4) is satisfied at each iteration, the sequence $\{x_k\}$ generated by the standard BFGS algorithm (i.e. Algorithm 1 with standard BFGS update (4)) obeys $\lim_{k\to\infty}\|\nabla f(x_k)\| = 0$. □

Lemma 1.
Under Assumption 1, if the stepsize α_k is chosen by Algorithm 1, then the following holds.

$$\|s_k\|^2 \leq s_k^T \hat{y}_k \leq (1 + m_2)\|s_k\|^2, \ \forall k \geq 0. \tag{12}$$

Proof. Since $\alpha_k \in (0, 1]$, we have
$\|s_k\|^2 = s_k^T s_k \leq s_k^T \hat{y}_k = s_k^T s_k + \alpha_k s_k^T y_k \leq (1 + m_2)\|s_k\|^2$. □

Inequality (12) implies that Assumption 1 still holds for the modified $\hat{y}_k$ and thus, this ensures the global convergence of the modified BFGS algorithm under Armijo line search.

3 NUMERICAL SIMULATION

To validate the performance of our method, a set of 7 test problems with standard starting point is considered [1], i.e. Extended DENSCHNB, FH3, Generalized Quartic, HIMMELBG, Diagonal 7, Diagonal 9 and Extended BD1. We use MATLAB to code the algorithms and implement them on a PC with CPU 2.5GHz processor with 4.00GB RAM. For each test problem, 3 different dimensions are considered, which are $n = 10, 100, 1000$. The value of constants is set as: $c_1 = 0.1, c_2 = 0.5, \epsilon = 10^{-4}$ and we also restrict the number of iterations to within 1000. For all the runs which the termination criterion is not reached within the allowable number of iterations, we consider it as failure (F).

Table 1 : Standard BFGS vs Modified BFGS Method

Problem	Dimension	BFGS		Modified BFGS	
		Iteration	Function call	Iteration	Function call
Extended DENSCHNB	10	6	9	6	9
	100	6	9	6	9
	1000	6	9	6	9
FH3	10	F	F	45	218
	100	F	F	50	238
	1000	F	F	56	268
Generalized Quartic	10	25	45	29	47
	100	115	240	44	111
	1000	170	351	43	112
HIMMELBG	10	48	203	40	169
	100	48	203	41	170
	1000	49	204	41	170
Diagonal 7	10	4	5	4	5
	100	4	5	4	5
	1000	5	6	5	6
Diagonal 9	10	39	68	5	6
	100	F	F	6	7
	1000	F	F	7	8
Extended BD1	10	11	13	9	13
	100	11	13	9	13
	1000	13	18	10	14

In comparing the performance of the standard BFGS method with the proposed method across a suite of test problems, it was consistently observed that the standard BFGS method required a higher number of iterations and function evaluations to reach convergence. This implies that the proposed method demonstrates enhanced efficiency, potentially due to combining two search directions, namely the steepest descent and the standard BFGS direction. Hence, in general we can conclude that Algorithm 2 is a promising alternative to BFGS method when solving unconstrained optimization problems.

4 CONCLUSION

In this paper, a novel quasi-Newton-like approximation has been proposed based on a Euler-based method for solving unconstrained optimization problems. The key contribution of this work lies in developing a modified approach that utilizes Euler's framework to enhance the efficiency of the quasi-Newton method. From a numerical perspective, the proposed method demonstrates significant advantages, as it often requires fewer iterations to reach an acceptable solution compared to standard BFGS method. This efficiency can lead to reduced computational costs and improved performance in solving complex optimization problems. Furthermore, we have established the convergence of the modified method under suitable assumptions, ensuring its theoretical reliability when Armijo line search is applied. Numerical experiments showed the superiority of the proposed method over the standard BFGS method in terms of iteration count and function call. The significance of this work lies in its potential to provide a valuable alternative to existing quasi-Newton methods.

REFERENCES

[1] N. Andrei, An unconstrained optimization test functions collection, Adv. Model. Optim., vol. 10: pp. 147-161, 2008.

[2] L. Armijo, Minimization of Functions Having Lipschitz-Continuous First Partial Derivatives, Pacific Journal of Mathematics, vol. 16: pp. 1–3, 1966.

[3] R. H. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, SIAM J. Numer. Anal., vol. 26: pp. 727-739, 1989.

[4] R. Courant and D. Hilbert, Methods of Mathematical Physics (Vol. II), Wiley-Interscience, New York, 1962.

[5] J. E. Dennis and J. J. More, Quasi-Newton Methods, Motivation and Theory, SIAM Review, vol. 19, no. 1, pp. 46–89, 1977.

[6] R. Fletcher, Practical Methods of Optimization (2nd ed.), John Wiley & Sons, New York, 1987.

[7] D. Li and M. Fukushima, A Globally and Superlinearly Convergent BFGS Method for Nonsmooth Convex Optimization, SIAM Journal on Optimization, vol. 11, no. 4, pp. 1051–1064, 2001.

[8] J. Nocedal and S. J. Wright, Numerical Optimization (2nd ed.), Springer, New York, 2006.

[9] M. Powell, Convergence Properties of a Class of Minimization Algorithms, Nonlinear Programming, vol. 2, pp. 1–27, 1975.

[10] H. S. Sim, W. J. Leong, C. Y. Chen, and S. N. I. Ibrahim, Multi-Step Spectral Gradient Methods with Modified Weak Secant Relation for Large-Scale Unconstrained Optimization, Numer. Algebra Control Optim., vol. 8, no. 3, pp. 377–387, 2018.

[11] H. S. Sim, W. J. Leong, and C. Y. Chen, Gradient Method with Multiple Damping for Large-Scale Unconstrained Optimization, Optim. Lett., vol. 13, no. 1, pp. 617–632, 2019.

[12] H. S. Sim, C. Y. Chen, W. J. Leong, and J. Li, Nonmonotone Spectral Gradient Method based on Memoryless Symmetric Rank-One Update for Large-Scale Unconstrained Optimization, J. Ind. Manag. Optim., vol. 18, no. 6, pp. 3975-3988, 2022.

[13] H. S. Sim, W. S. Y. Ling, W. J. Leong, and C. Y. Chen, Proximal Linearized Method for Sparse Equity Portfolio Optimization with Minimum Transaction Cost, J. Inequal. Appl., vol. 2023, no. 1, paper ID. 152, 2023.

[14] G. Y. H. Woo, H. S. Sim, Y. K. Goh, and W. J. Leong, Proximal Variable Metric Method with Spectral Diagonal Update for Large Scale Sparse Optimization, Journal of the Franklin Institute, vol. 360, pp. 4640–4660, 2023.

[15] L. Xiao and M. J. D. Powell, On the Convergence of the BFGS Method for Unconstrained Optimization, IMA Journal of Numerical Analysis, vol. 17, pp. 455–472, 1997.

[16] J. Zhang and D. Zhu, A Projective Quasi-Newton Method for Nonlinear Optimization, Journal of Computational and Applied Mathematics, vol. 53, pp. 291–307, 1994.