Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 11(1): 767–784.
DOI:10.3934/math.2026033
Received: 05 July 2025
Revised: 15 December 2025
Accepted: 06 January 2026
Published: 12 January 2026

Research article

A modified BFGS quasi-Newton method with Wolfe line search for unconstrained optimization

Wen Zhang[1,*], Tingting Guo[1], Junfeng Wu[2], Zhousheng Ruan[1] and Shufang Qiu[1,3]

[1] School of Science, East China University of Technology, Nanchang 330013, China
[2] School of Information and Artificial Intelligence, Nanchang Institute of Science and Technology, Nanchang 330108, China
[3] School of Artificial Intelligence, Guangzhou Maritime University, Guangzhou 510725, China

* **Correspondence:** Email: zhangw@ecut.edu.cn.

Abstract: In this article, we construct a modified BFGS quasi-Newton method with Wolfe line search to solve a nonlinear equations. Firstly, we propose a new quasi-Newton secant equation and the corresponding practical implementation algorithm by combining the classic BFGS with the strong Wolfe conditions. Furthermore, the local and superlinear rate of convergence of the modified quasi-Newton updates are derived theoretically. Lastly, numerical examples illustrate the effectiveness and stability of the proposed method.

Keywords: nonlinear equations; quasi-Newton method; BFGS; Wolfe line search; convergence
Mathematics Subject Classification: 49J35, 49M15, 49M37, 90C53

1. Introduction

In this paper, we focus our attention on solving nonlinear system

$$\boldsymbol{F}(\boldsymbol{x}) = 0, \quad \boldsymbol{x} = (x_1, x_2, \cdots, x_N)^T \in \mathbb{R}^N, \tag{1.1}$$

where $\boldsymbol{F} = (F_1, F_2, \cdots, F_N)^T, F_i : \mathbb{R}^N \to \mathbb{R} (i = 1, 2, \cdots, N)$ are continuously differentiable and the Jacobian is symmetric, i.e.,

$$J(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_N} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_N}{\partial x_1} & \frac{\partial F_N}{\partial x_2} & \cdots & \frac{\partial F_N}{\partial x_N} \end{bmatrix} = J^T(\boldsymbol{x}).$$

From an optimization perspective, the problem of solving nonlinear system (1.1) can be formulated as the following minimization problem: $\min_{\boldsymbol{x}\in\mathbb{R}^N} f(\boldsymbol{x})$

$$f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{F}(\boldsymbol{x})\|^2, \tag{1.2}$$

and the line search iterative method

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \alpha^k \boldsymbol{d}^k, \tag{1.3}$$

where $\alpha^k \in \mathbb{R}$, $\boldsymbol{d}^k \in \mathbb{R}^N$ are stepsize and search direction on step k, respectively.

The classic Newton's method is efficient and possesses quadratic convergence rate for solving nonlinear equations when the size is not too large. The main drawback of the Newton direction is the need for the Hessian $\nabla^2 f(\boldsymbol{x})$. Explicit computation of this matrix of second derivatives can sometimes be a cumbersome, error-prone, and expensive process. And the pure Newton iteration is not guaranteed to produce descent directions when the current iterate is not close to a solution. Quasi-Newton methods are recognized as one of the most powerful methods with locally superlinear convergence and need not compute the Hessian for solving deterministic optimization problems, these methods build quadratic models of the objective information using only gradient information. Among the various quasi-Newton update schemes, the BFGS formula (named for its discoverers Broyden, Fletcher, Goldfarb, and Shanno) stands out as the most effective [1–3]. For some unconstrained stochastic optimization problems with no available gradient information, which arise in settings from derivative-free simulation optimization to reinforcement learning, adaptive sampling quasi-Newton method was employed to estimate the gradients by using finite differences of stochastic function evaluations within a common random number framework [5]. For large-scale problems, derivative-free type methods, the limited memory L-BFGS, truncated Newton method, or Conjugate Gradient methods would be more effective. Derivative-free optimization (DFO) is vital in solving complex optimization problems where only noisy function evaluations are available through an oracle or black-box interface. For example, based on DFO methods to optimize many machine learning models and complex systems, Google Vizier has executed millions of optimizations, accelerating numerous research and production systems at Google [6, 7]. A fully derivative-free conjugate residual method is invoked to solve general large-scale nonlinear equations and obtain the global linear rate of convergence by means of some secant conditions and backtracking type line search method [4]. Zhang et al. [8] build a new quasi-Newton equation and provide the local and superlinear convergence. Zhou [13] present an inexact Modified BFGS method with line search for solving symmetric nonlinear equations and its global convergence. Wan et al. [14] study the BFGS method by modifying the Armijo line search technique and deduce a global convergence. Yuan et al. [15, 16] propose a BFGS method with modified weak Wolfe-Powell line search, and obtain a global convergence under suitable conditions for general functions. Upadhavay et al. [17] give two nonmonotone quasi-Newton algorithms with Wolfe line searches for unconstrained multiobjective optimization problems and leads to a global convergence. Cheng et al. and Zhang et al. [18, 19] consider the memoryless BFGS quasi-Newton method to solve the unconstrained optimization problem.

Comparing with the computational challenge of Hessian $\nabla^2 f(\boldsymbol{x}^k)$ and its inverse matrix $\nabla^2 f(\boldsymbol{x}^k)^{-1}$ in Newton iterative method

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \nabla^2 f(\boldsymbol{x}^k)^{-1} \nabla f(\boldsymbol{x}^k), \tag{1.4}$$

the quasi-Newton method is a class of methods which need not compute the Hessian, but generates a series of Hessian approximations, and at the same time maintains a fast rate of convergence. The classic BFGS quasi-Newton iterative method approximates the Hessian $\nabla^2 f(\boldsymbol{x}^k)$ and its inverse matrix $\nabla^2 f(\boldsymbol{x}^k)^{-1}$ by a symmetric matrix $B^k \in \mathbb{R}^{N\times N}$ and $H^k \in \mathbb{R}^{N\times N}$ respectively, from the updating Rank-two formula

$$B^{k+1} = B^k + \frac{\boldsymbol{y}^k(\boldsymbol{y}^k)^T}{(\boldsymbol{s}^k)^T\boldsymbol{y}^k} - \frac{B^k\boldsymbol{s}^k(B^k\boldsymbol{s}^k)^T}{(\boldsymbol{s}^k)^T B^k\boldsymbol{s}^k}, \tag{1.5}$$

$$H^{k+1} = (V^k)^T H^k V^k + t^k\boldsymbol{s}^k(\boldsymbol{s}^k)^T, \tag{1.6}$$

where $t^k = \frac{1}{(\boldsymbol{s}^k)^T\hat{\boldsymbol{y}}^k}$, $V^k = I - t^k\hat{\boldsymbol{y}}^k(\boldsymbol{s}^k)^T$, and

$$B^{k+1}\boldsymbol{s}^k = \boldsymbol{y}^k, \quad \boldsymbol{s}^k = \boldsymbol{x}^{k+1} - \boldsymbol{x}^k, \quad \boldsymbol{y}^k = \nabla f(\boldsymbol{x}^{k+1}) - \nabla f(\boldsymbol{x}^k).$$

In order to obtain a better approximation of B^{k+1}, we construct a new quasi-Newton secant equation

$$B^{k+1}\boldsymbol{s}^k = \tilde{\boldsymbol{y}}^k, \tag{1.7}$$

where $\tilde{\boldsymbol{y}}^k$ is to be determined by the following section, and then investigate its theoretical properties and practical implementation.

The rest of this paper is organized as follows. In section 2, we construct a new quasi-Newton secant equation. Some properties of the new quasi-Newton secant equation and the Algorithm are presented in section 3. In section 4, the local and superlinear rate of convergence is described. Several numerical tests are performed to demonstrate the accuracy and efficiency of the proposed method in section 5. The paper is concluded in section 6.

2. New quasi-Newton secant equation

Based on the symmetric system (1.1) and the first-order Taylor expansion, we have

$$\frac{\partial \boldsymbol{F}(\boldsymbol{x} + t\alpha\boldsymbol{F}(\boldsymbol{x}))}{\partial t} = J(\boldsymbol{x} + t\alpha\boldsymbol{F}(\boldsymbol{x}))\alpha\boldsymbol{F}(\boldsymbol{x}), \tag{2.1}$$

when $\|\alpha\boldsymbol{F}(\boldsymbol{x})\|$ is arbitrarily small, it yields

$$\begin{aligned}\frac{\boldsymbol{F}(\boldsymbol{x} + \alpha\boldsymbol{F}(\boldsymbol{x})) - \boldsymbol{F}(\boldsymbol{x})}{\alpha} &= \int_0^1 J(\boldsymbol{x} + t\alpha\boldsymbol{F}(\boldsymbol{x}))dt\boldsymbol{F}(\boldsymbol{x}) \\ &\approx J(\boldsymbol{x})\boldsymbol{F}(\boldsymbol{x}) = \nabla f(\boldsymbol{x}).\end{aligned} \tag{2.2}$$

Let

$$\boldsymbol{x}(\tau) = \boldsymbol{x}^k + \tau\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|}, \quad \boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x}),$$

then $\boldsymbol{x}(0) = \boldsymbol{x}^k$, $\boldsymbol{x}(\|\boldsymbol{s}^k\|) = \boldsymbol{x}^{k+1}$. To simplify the formulation, we set

$$f^k = f(\boldsymbol{x}^k), \quad \boldsymbol{g}^k = \boldsymbol{g}(\boldsymbol{x}^k) = \nabla f(\boldsymbol{x}^k), \tag{2.3}$$

notice that $\frac{dg(x(\tau))}{d\tau} = \nabla^2 f(x(\tau))\frac{s^k}{\|s^k\|}$, we have

$$\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k = \|\boldsymbol{s}^k\| \left.\frac{d\boldsymbol{g}(\boldsymbol{x}(\tau))}{d\tau}\right|_{\tau=\|\boldsymbol{s}^k\|}. \tag{2.4}$$

It reveals that the construction of $\boldsymbol{g}(\boldsymbol{x}(\tau))$ plays an important role in the approximation of Hessian $\nabla^2 f(\boldsymbol{x}^{k+1})$.

To approximate $\boldsymbol{g}(\boldsymbol{x}(\tau))$, we set

$$\tilde{\boldsymbol{g}}(\tau) = \boldsymbol{a} + \boldsymbol{b}e^{\tau} + \boldsymbol{c}e^{2\tau}, \tag{2.5}$$

where $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^N$ satisfy the following conditions

$$\begin{cases} \tilde{\boldsymbol{g}}(0) = \boldsymbol{g}(\boldsymbol{x}(0)) = \boldsymbol{g}^k, \\ \tilde{\boldsymbol{g}}(\|\boldsymbol{s}^k\|) = \boldsymbol{g}(\boldsymbol{x}(\|\boldsymbol{s}^k\|)) = \boldsymbol{g}^{k+1}, \\ \displaystyle\int_0^{\|\boldsymbol{s}^k\|} \tilde{\boldsymbol{g}}^T(\tau)\frac{d\boldsymbol{x}(\tau)}{d\tau}d\tau = f^{k+1} - f^k, \end{cases}$$

it yields

$$\begin{cases} \boldsymbol{a} + \boldsymbol{b} + \boldsymbol{c} = \boldsymbol{g}^k, \\ \boldsymbol{a} + \boldsymbol{b}e^{\|\boldsymbol{s}^k\|} + \boldsymbol{c}e^{2\|\boldsymbol{s}^k\|} = \boldsymbol{g}^{k+1}, \\ \left(\boldsymbol{a}\|\boldsymbol{s}^k\| + \boldsymbol{b}(e^{\|\boldsymbol{s}^k\|} - 1) + \dfrac{\boldsymbol{c}}{2}(e^{2\|\boldsymbol{s}^k\|} - 1)\right)^T \boldsymbol{s}^k = (f^{k+1} - f^k)\|\boldsymbol{s}^k\|. \end{cases} \tag{2.6}$$

Benifits from (2.4)

$$\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k \approx \|\boldsymbol{s}^k\| \left.\frac{d\tilde{\boldsymbol{g}}(\tau)}{d\tau}\right|_{\tau=\|\boldsymbol{s}^k\|} = \|\boldsymbol{s}^k\|\left(\boldsymbol{b}e^{\|\boldsymbol{s}^k\|} + 2\boldsymbol{c}e^{2\|\boldsymbol{s}^k\|}\right),$$

and after a simple algebraic calculation from the linear system (2.6), we obtain

$$\left(\boldsymbol{g}^k + \boldsymbol{c}e^{\|\boldsymbol{s}^k\|} - \frac{\boldsymbol{c}(e^{2\|\boldsymbol{s}^k\|} - 1)}{2\|\boldsymbol{s}^k\|} - \frac{\boldsymbol{y}^k}{e^{\|\boldsymbol{s}^k\|} - 1} + \frac{\boldsymbol{y}^k}{\|\boldsymbol{s}^k\|}\right)^T \boldsymbol{s}^k = f^{k+1} - f^k, \tag{2.7}$$

$$(\boldsymbol{b}e^{\|\boldsymbol{s}^k\|})^T\boldsymbol{s}^k = \frac{e^{\|\boldsymbol{s}^k\|}}{e^{\|\boldsymbol{s}^k\|} - 1}(\boldsymbol{y}^k)^T\boldsymbol{s}^k - e^{\|\boldsymbol{s}^k\|}(e^{\|\boldsymbol{s}^k\|} + 1)\boldsymbol{c}^T\boldsymbol{s}^k, \tag{2.8}$$

$$\begin{aligned}(2\boldsymbol{c}e^{2\|\boldsymbol{s}^k\|})^T\boldsymbol{s}^k = {} & \frac{4\|\boldsymbol{s}^k\|e^{2\|\boldsymbol{s}^k\|}(f^{k+1} - f^k)}{2\|\boldsymbol{s}^k\|e^{\|\boldsymbol{s}^k\|} - e^{2\|\boldsymbol{s}^k\|} + 1} - \Big(\frac{4\|\boldsymbol{s}^k\|e^{2\|\boldsymbol{s}^k\|}}{2\|\boldsymbol{s}^k\|e^{\|\boldsymbol{s}^k\|} - e^{2\|\boldsymbol{s}^k\|} + 1}\boldsymbol{g}^k \\ & + \frac{4e^{2\|\boldsymbol{s}^k\|}(e^{\|\boldsymbol{s}^k\|} - 1 - \|\boldsymbol{s}^k\|)}{(2\|\boldsymbol{s}^k\|e^{\|\boldsymbol{s}^k\|} - e^{2\|\boldsymbol{s}^k\|} + 1)(e^{\|\boldsymbol{s}^k\|} - 1)}\boldsymbol{y}^k\Big)^T\boldsymbol{s}^k,\end{aligned} \tag{2.9}$$

and finally we have

$$\|\boldsymbol{s}^k\|\left(\boldsymbol{b}e^{\|\boldsymbol{s}^k\|} + 2\boldsymbol{c}e^{2\|\boldsymbol{s}^k\|}\right)^T \boldsymbol{s}^k = (\boldsymbol{y}^k)^T\boldsymbol{s}^k + \gamma,$$

where

$$\gamma = \mathcal{A}(\boldsymbol{g}^{k+1})^T \boldsymbol{s}^k + \mathcal{B}(\boldsymbol{g}^k)^T \boldsymbol{s}^k + C(f^{k+1} - f^k), \tag{2.10}$$

$$\mathcal{A} = \frac{(1 - 3\|\boldsymbol{s}^k\|)e^{3\|\boldsymbol{s}^k\|} + (4\|\boldsymbol{s}^k\|^2 + 2\|\boldsymbol{s}^k\| - 1)e^{2\|\boldsymbol{s}^k\|} + (-2\|\boldsymbol{s}^k\|^2 + \|\boldsymbol{s}^k\| - 1)e^{\|\boldsymbol{s}^k\|} + 1}{(e^{\|\boldsymbol{s}^k\|} - 1)(2\|\boldsymbol{s}^k\|e^{\|\boldsymbol{s}^k\|} - e^{2\|\boldsymbol{s}^k\|} + 1)}, \tag{2.11}$$

$$\mathcal{B} = \frac{(-2\|\boldsymbol{s}^k\|^2 + 3\|\boldsymbol{s}^k\| - 1)e^{3\|\boldsymbol{s}^k\|} + (1 - 2\|\boldsymbol{s}^k\|)e^{2\|\boldsymbol{s}^k\|} + (1 - \|\boldsymbol{s}^k\|)e^{\|\boldsymbol{s}^k\|} - 1}{(e^{\|\boldsymbol{s}^k\|} - 1)(2\|\boldsymbol{s}^k\|e^{\|\boldsymbol{s}^k\|} - e^{2\|\boldsymbol{s}^k\|} + 1)}, \tag{2.12}$$

$$C = \frac{2\|\boldsymbol{s}^k\|^2(e^{2\|\boldsymbol{s}^k\|} - e^{\|\boldsymbol{s}^k\|})}{2\|\boldsymbol{s}^k\|e^{\|\boldsymbol{s}^k\|} - e^{2\|\boldsymbol{s}^k\|} + 1}, \tag{2.13}$$

and $\mathcal{A} + \mathcal{B} + C = 0$.

Therefore, we obtain our new quasi-Newton secant equation as follows

$$B^{k+1}\boldsymbol{s}^k = \tilde{\boldsymbol{y}}^k, \tag{2.14}$$

$$\tilde{\boldsymbol{y}}^k = \boldsymbol{y}^k + \gamma\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}. \tag{2.15}$$

3. Properties and algorithm

In this section, we will at first show that $\tilde{\boldsymbol{y}}^k$ defined by (2.15) in the new quasi-Newton secant equation (2.14) is a better approximation to $\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k$ than $\boldsymbol{y}^k$, and then we will present the modified quasi-Newton algorithm.

Lemma 1. [8, Lemma 3.1] *If the function f is smooth enough, then*

$$(\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \boldsymbol{y}^k) = \frac{\|\boldsymbol{s}^k\|^2}{2}\sum_{i,j,l=1}^{N}\frac{\partial^3 f(\boldsymbol{x}^k + \xi\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})}{\partial x_i \partial x_j \partial x_l}\frac{s_i}{\|\boldsymbol{s}^k\|}\frac{s_j}{\|\boldsymbol{s}^k\|}\frac{s_l}{\|\boldsymbol{s}^k\|},$$

where $\xi \in (0, \|\boldsymbol{s}^k\|)$, and x_i, s_i are the ith elements of vector $\boldsymbol{x}^k$ and $\boldsymbol{s}^k$ on the step k, respectively.

Theorem 1. *If the function f is smooth enough, let $\varphi(\tau) = f(\boldsymbol{x}(\tau))$, then we have*

$$\begin{aligned}(\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \tilde{\boldsymbol{y}}^k) =& \frac{(2\mathcal{B} + C)\|\boldsymbol{s}^k\|}{2}\varphi''(\|\boldsymbol{s}^k\|) + \frac{(3 - 3\mathcal{B} - C)\|\boldsymbol{s}^k\|^2}{6}\varphi'''(\|\boldsymbol{s}^k\|)\\ &+ \frac{(\mathcal{B} - 1)\|\boldsymbol{s}^k\|^3}{3!}\varphi^{(4)}(\xi_1) + \frac{C\|\boldsymbol{s}^k\|^3}{4!}\varphi^{(4)}(\xi_2),\end{aligned}$$

where $\xi_1, \xi_2 \in (0, \|\boldsymbol{s}^k\|)$.

Proof. Since

$$\boldsymbol{x}(\tau) = \boldsymbol{x}^k + \tau\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|}, \quad \boldsymbol{g}(\boldsymbol{x}) = \nabla f(\boldsymbol{x}), \quad \varphi(\tau) = f(\boldsymbol{x}(\tau)), \tag{3.1}$$

and the function $\varphi(\tau)$ is smooth enough too, we obtain

$$\varphi'(\tau) = (\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T\boldsymbol{g}(\boldsymbol{x}^k + \tau\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|}) = \sum_{i=1}^{N}\frac{\partial f(\boldsymbol{x}^k + \tau\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})}{\partial x_i}\frac{s_i}{\|\boldsymbol{s}^k\|}, \tag{3.2}$$

$$\varphi''(\tau) = (\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2})^T \nabla^2 f(\boldsymbol{x}^k + \tau\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})\boldsymbol{s}^k = \sum_{i,j=1}^{N} \frac{\partial^2 f(\boldsymbol{x}^k + \tau\frac{s^k}{\|s^k\|})}{\partial x_i \partial x_j}\frac{s_i}{\|\boldsymbol{s}^k\|}\frac{s_j}{\|\boldsymbol{s}^k\|}, \tag{3.3}$$

$$\varphi'''(\tau) = \sum_{i,j,l=1}^{N} \frac{\partial^3 f(\boldsymbol{x}^k + \tau\frac{s^k}{\|s^k\|})}{\partial x_i \partial x_j \partial x_l}\frac{s_i}{\|\boldsymbol{s}^k\|}\frac{s_j}{\|\boldsymbol{s}^k\|}\frac{s_l}{\|\boldsymbol{s}^k\|}, \tag{3.4}$$

$$\varphi^{(4)}(\tau) = \sum_{i,j,l,m=1}^{N} \frac{\partial^4 f(\boldsymbol{x}^k + \tau\frac{s^k}{\|s^k\|})}{\partial x_i \partial x_j \partial x_l \partial x_m}\frac{s_i}{\|\boldsymbol{s}^k\|}\frac{s_j}{\|\boldsymbol{s}^k\|}\frac{s_l}{\|\boldsymbol{s}^k\|}\frac{s_m}{\|\boldsymbol{s}^k\|}. \tag{3.5}$$

and

$$\varphi(\|\boldsymbol{s}^k\|) - \varphi(0) = f^{k+1} - f^k, \quad \varphi'(0) = \frac{(\boldsymbol{s}^k)^T \boldsymbol{g}^k}{\|\boldsymbol{s}^k\|}, \quad \varphi'(\|\boldsymbol{s}^k\|) = \frac{(\boldsymbol{s}^k)^T \boldsymbol{g}^{k+1}}{\|\boldsymbol{s}^k\|},$$
$$\varphi''(0) = \frac{(\boldsymbol{s}^k)^T \nabla^2 f(\boldsymbol{x}^k)\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}, \quad \varphi''(\|\boldsymbol{s}^k\|) = \frac{(\boldsymbol{s}^k)^T \nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2},$$

According to the Taylor expansion, there exists $\xi_1, \xi_2 \in (0, \|\boldsymbol{s}^k\|)$, such that

$$\varphi'(0) = \varphi'(\|\boldsymbol{s}^k\|) - \|\boldsymbol{s}^k\|\varphi''(\|\boldsymbol{s}^k\|) + \frac{\|\boldsymbol{s}^k\|^2}{2}\varphi'''(\|\boldsymbol{s}^k\|) - \frac{\|\boldsymbol{s}^k\|^3}{3!}\varphi^{(4)}(\xi_1), \tag{3.6}$$

$$\varphi(0) = \varphi(\|\boldsymbol{s}^k\|) - \|\boldsymbol{s}^k\|\varphi'(\|\boldsymbol{s}^k\|) + \frac{\|\boldsymbol{s}^k\|^2}{2}\varphi''(\|\boldsymbol{s}^k\|) - \frac{\|\boldsymbol{s}^k\|^3}{3!}\varphi^{(3)}(\|\boldsymbol{s}^k\|) + \frac{\|\boldsymbol{s}^k\|^4}{4!}\varphi^{(4)}(\xi_2). \tag{3.7}$$

It yields from the new quasi-Newton secant equation (2.15), Lemma 1 and (3.6-3.7)

$$\begin{aligned}
(\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \tilde{\boldsymbol{y}}^k) &= (\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \boldsymbol{y}^k - \gamma\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}) \\
&= \|\boldsymbol{s}^k\|\varphi''(\|\boldsymbol{s}^k\|) - (\mathcal{A}+1)\varphi'(\|\boldsymbol{s}^k\|) - (\mathcal{B}-1)\varphi'(0) - \frac{C(\varphi(\|\boldsymbol{s}^k\|) - \varphi(0))}{\|\boldsymbol{s}^k\|} \\
&= \frac{C(\varphi(0) - \varphi(\|\boldsymbol{s}^k\|))}{\|\boldsymbol{s}^k\|} - (\mathcal{A}+\mathcal{B})\varphi'(\|\boldsymbol{s}^k\|) + \mathcal{B}\|\boldsymbol{s}^k\|\varphi''(\|\boldsymbol{s}^k\|) \\
&\quad + \frac{\|\boldsymbol{s}^k\|^2}{2}(1-\mathcal{B})\varphi'''(\|\boldsymbol{s}^k\|) - \frac{\|\boldsymbol{s}^k\|^3}{3!}(1-\mathcal{B})\varphi^{(4)}(\xi_1) \\
&= \frac{(2\mathcal{B}+C)\|\boldsymbol{s}^k\|}{2}\varphi''(\|\boldsymbol{s}^k\|) + \frac{(3-3\mathcal{B}-C)\|\boldsymbol{s}^k\|^2}{6}\varphi'''(\|\boldsymbol{s}^k\|) \\
&\quad + \frac{(\mathcal{B}-1)\|\boldsymbol{s}^k\|^3}{3!}\varphi^{(4)}(\xi_1) + \frac{C\|\boldsymbol{s}^k\|^3}{4!}\varphi^{(4)}(\xi_2).
\end{aligned} \tag{3.8}$$

and the proof is completed. □

Remark 1. *The advantages of new algorithm* (2.14)-(2.15) *are not sensitive obviously from Theorem 1. In the light of six-order Taylor expansion of the coefficients $\mathcal{A}, \mathcal{B}, C$, Theorem 1 becomes*

$$(\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \tilde{\boldsymbol{y}}^k) = \|\boldsymbol{s}^k\|^3\Big\{(\frac{1}{6} + \frac{\|\boldsymbol{s}^k\|}{20} + \frac{\|\boldsymbol{s}^k\|^2}{180} + o(\|\boldsymbol{s}^k\|^2))\varphi''(\|\boldsymbol{s}^k\|)$$

$$- (\frac{1}{4} + \frac{17\|\boldsymbol{s}^k\|}{120} + \frac{\|\boldsymbol{s}^k\|^2}{30} + o(\|\boldsymbol{s}^k\|^2))\varphi'''(\|\boldsymbol{s}^k\|)$$
$$+ (\frac{1}{3} + \frac{\|\boldsymbol{s}^k\|}{4} + \frac{31\|\boldsymbol{s}^k\|^2}{360} + o(\|\boldsymbol{s}^k\|^2))\varphi^{(4)}(\xi_1)$$
$$- (\frac{1}{4} + \frac{\|\boldsymbol{s}^k\|}{8} + \frac{7\|\boldsymbol{s}^k\|^2}{240} + o(\|\boldsymbol{s}^k\|^2))\varphi^{(4)}(\xi_2)\Big\}. \quad (3.9)$$

Therefore, if the function f is smooth enough, we observe that

$$(\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \boldsymbol{y}^k) = O(\|\boldsymbol{s}^k\|^2), \quad (3.10)$$
$$(\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|})^T(\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k - \tilde{\boldsymbol{y}}^k) = O(\|\boldsymbol{s}^k\|^3). \quad (3.11)$$

It reveals that the projection $\tilde{\boldsymbol{y}}^k$ on the direction $\boldsymbol{s}^k$ is of a higher-order precision approximation to $\nabla^2 f(\boldsymbol{x}^{k+1})\boldsymbol{s}^k$ than $\boldsymbol{y}^k$ does as $\|\boldsymbol{s}^k\|$ approaches to zero.

The modified quasi-Newton method to solve the nonlinear system (1.1) is presented as follows.

Algorithm 1. Given a initial point $\boldsymbol{x}^0 \in \mathbb{R}^N$, a symmetric and positive definite matrix $B^0 \in \mathbb{R}^{N\times N}$ or $H^0 \in \mathbb{R}^{N\times N}$, several constants $\alpha \in (0, \frac{1}{2}), \beta \in (\alpha, 1)$ and a sufficiently small positive constant $\epsilon > 0$, let $k = 0$.

Step 1. Calculate $\boldsymbol{g}^k$ by (2.2). If $\|\boldsymbol{g}^k\| = 0$, then stop. Otherwise, go to **Step 2**.

Step 2. Solve the search direction $\boldsymbol{s}^k$ by $B^k\boldsymbol{s}^k = -\boldsymbol{g}^k$ or $\boldsymbol{s}^k = -H^k\boldsymbol{g}^k$.

Step 3. Compute $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \lambda^k\boldsymbol{s}^k$ for $\lambda^k > 0$ with initial number $\lambda = 1$, such that the strong Wolfe conditions

$$f^{k+1} \le f^k + \alpha\lambda^k(\boldsymbol{g}^k)^T\boldsymbol{s}^k, \quad (3.12)$$
$$|(\boldsymbol{g}^{k+1})^T\boldsymbol{s}^k| \le \beta|(\boldsymbol{g}^k)^T\boldsymbol{s}^k|. \quad (3.13)$$

Step 4. Update B^k or H^k by the following Rank-two formula:

$$B^{k+1} = B^k + \frac{(\hat{\boldsymbol{y}}^k - B^k\boldsymbol{s}^k)(\boldsymbol{u}^k)^T + \boldsymbol{u}^k(\hat{\boldsymbol{y}}^k - B^k\boldsymbol{s}^k)^T}{(\boldsymbol{u}^k)^T\boldsymbol{s}^k} - \frac{(\boldsymbol{s}^k)^T(\hat{\boldsymbol{y}}^k - B^k\boldsymbol{s}^k)\boldsymbol{u}^k(\boldsymbol{u}^k)^T}{((\boldsymbol{u}^k)^T\boldsymbol{s}^k)^2}, \quad (3.14)$$
$$H^{k+1} = H^k + \frac{(\boldsymbol{s}^k - H^k\hat{\boldsymbol{y}}^k)(\boldsymbol{v}^k)^T + \boldsymbol{v}^k(\boldsymbol{s}^k - H^k\hat{\boldsymbol{y}}^k)^T}{(\boldsymbol{v}^k)^T\hat{\boldsymbol{y}}^k} - \frac{(\hat{\boldsymbol{y}}^k)^T(\boldsymbol{s}^k - H^k\hat{\boldsymbol{y}}^k)\boldsymbol{v}^k(\boldsymbol{v}^k)^T}{((\boldsymbol{v}^k)^T\hat{\boldsymbol{y}}^k)^2}, \quad (3.15)$$

where $\hat{\boldsymbol{y}}^k = \boldsymbol{y}^k + \hat{\gamma}\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}$, and

$$\hat{\gamma} = \begin{cases} \gamma, & \text{if } (\boldsymbol{y}^k)^T\boldsymbol{s}^k + \gamma \ge \epsilon\|\boldsymbol{s}^k\|^2, \\ 0, & \text{otherwise.} \end{cases} \quad (3.16)$$

and γ is defined by (2.10), parameter vectors $\boldsymbol{u}^k$ and $\boldsymbol{v}^k$ are not orthogonal to $\boldsymbol{s}^k$ and $\hat{\boldsymbol{y}}^k$, respectively.

Step 5. Let $k = k + 1$ and go to **Step 1**.

Remark 2. *It is interesting to note that the* ***curvature condition*** $(\boldsymbol{y}^k)^T\boldsymbol{s}^k > 0$ *is an important property to ensure* B^{k+1} *is positive definite in the quasi-Newton method. The setting* $\hat{\gamma}$ *in* (3.16) *will guarantee the curvature condition* $(\hat{\boldsymbol{y}}^k)^T\boldsymbol{s}^k > 0$ *when* $(\boldsymbol{y}^k)^T\boldsymbol{s}^k > 0$*. Moreover, if* $\hat{\boldsymbol{y}}^k = \boldsymbol{y}^k$*, then B-update formula* (3.14) *is called* ***General PSB update*** *and H-update formula* (3.15) *is* ***Greenstadt update****. In particular, B-update* (3.14) *becomes Rank-one SR1 update as* $\boldsymbol{u}^k = \hat{\boldsymbol{y}}^k - B^k\boldsymbol{s}^k$*, and* (3.14) *becomes DFP update as* $\boldsymbol{u}^k = \hat{\boldsymbol{y}}^k$*, and* (3.14) *becomes PSB update as* $\boldsymbol{u}^k = \boldsymbol{s}^k$*, and* (3.14) *becomes BFGS update* (1.5) *as* $\boldsymbol{u}^k = \frac{1}{1+w^k}\hat{\boldsymbol{y}}^k + \frac{w^k}{1+w^k}B^k\boldsymbol{s}^k$*, where* $w^k = \sqrt{\frac{(\hat{\boldsymbol{y}}^k)^T\boldsymbol{s}^k}{(\boldsymbol{s}^k)^TB^k\boldsymbol{s}^k}}$*. About the details of General PSB update* (3.14)*, we refer the reader to textbooks on optimization, such as [1, Chapter 5], [3, Chapter 6].*

4. Local convergence analysis

Without loss of generality, we will discuss the local and superlinear convergence of the modified quasi-Newton method with $\lambda^k = 1$.

Lemma 2. [9, Lemma 3.1] *Assume that function* $f(\boldsymbol{x})$ *is twice differentiable over a convex open set D, and* $\boldsymbol{x}^* \in D$ *be the stationary point of function* $f(\boldsymbol{x})$*, namely* $\nabla f(\boldsymbol{x}^*) = \boldsymbol{g}(\boldsymbol{x}^*) = 0$*, and* $\nabla^2 f(\boldsymbol{x}^*)$ *is nonsingular and* $\nabla^2 f(\boldsymbol{x})$ *is continuous at the stationary point* $\boldsymbol{x}^*$*, and there exist positive constants* ρ *and K, for all* $\boldsymbol{x} \in D$ *such that*

$$\|\nabla^2 f(\boldsymbol{x}) - \nabla^2 f(\boldsymbol{x}^*)\| \le K\|\boldsymbol{x} - \boldsymbol{x}^*\|^\rho, \tag{4.1}$$

then, it leads to

$$\|\boldsymbol{g}(\boldsymbol{p}) - \boldsymbol{g}(\boldsymbol{q}) - \nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{p} - \boldsymbol{q})\| \le K\max\{\|\boldsymbol{p} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{q} - \boldsymbol{x}^*\|^\rho\}\|\boldsymbol{p} - \boldsymbol{q}\|, \quad \forall \boldsymbol{p}, \boldsymbol{q} \in D, \tag{4.2}$$

and there exist positive constants $\bar{\epsilon}, L$*, and a closed neighborhood* $\bar{X}(\boldsymbol{x}^*, \bar{\epsilon}) \subseteq D$*, such that*

$$\frac{1}{L}\|\boldsymbol{p} - \boldsymbol{q}\| \le \|\boldsymbol{g}(\boldsymbol{p}) - \boldsymbol{g}(\boldsymbol{q})\| \le L\|\boldsymbol{p} - \boldsymbol{q}\|, \quad \forall \boldsymbol{p}, \boldsymbol{q} \in \bar{X}(\boldsymbol{x}^*, \bar{\epsilon}). \tag{4.3}$$

Lemma 3. [9, Theorem 3.2 – Corollary 3.5] *For any B-update formula in the quasi-Newton algorithm, for all* $(\boldsymbol{x}^k, B^k)$ *in a neighborhood of* $(\boldsymbol{x}^*, \nabla^2 f(\boldsymbol{x}^*))$*, if*

$$\begin{aligned}\|B^{k+1} - \nabla^2 f(\boldsymbol{x}^*)\|_M \le & \Big(1 + \alpha_1\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\}\Big)\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M \\ & + \alpha_2\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\},\end{aligned} \tag{4.4}$$

where nonnegative constants $\alpha_1, \alpha_2 > 0$*, and M is a nonsingular symmetric matrix, define the matrix norm* $\|\cdot\|_M$ *by* $\|Q\|_M = \|MQM\|$ *and the initial* $(\boldsymbol{x}^0, B^0)$ *is close enough to* $(\boldsymbol{x}^*, \nabla^2 f(\boldsymbol{x}^*))$*. Then the sequence* $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - (B^k)^{-1}\nabla f(\boldsymbol{x})$ *from quasi-Newton method is locally convergent at* $\boldsymbol{x}^*$*. And the iterative sequence* $\{B^k\}$ *and* $\{(B^k)^{-1}\}$ *are uniformly bounded and there exists* $\sigma \in (0, 1)$*,* $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\| \le \sigma\|\boldsymbol{x}^k - \boldsymbol{x}^*\|$ *for all* $k \ge 0$*. Furthermore, if some subsequence of* $\{\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M\}$ *converges to zero, then* $\{\boldsymbol{x}^k\}$ *converges Q-superlinearly at* $\boldsymbol{x}^*$*. Namely,* $\lim\limits_{k\to+\infty}\frac{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}^k-\boldsymbol{x}^*\|} = 0$*.*

There is a similar result regarding H-update formula. More precisely, if

$$\|H^{k+1} - \nabla^2 f(\boldsymbol{x}^*)^{-1}\|_M \le\Big(1 + \alpha_1\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\}\Big)\|H^k - \nabla^2 f(\boldsymbol{x}^*)^{-1}\|_M$$

$$+\alpha_2 \max\{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|^{\rho}, \|\boldsymbol{x}^k-\boldsymbol{x}^*\|^{\rho}\}, \tag{4.5}$$

then $\{\boldsymbol{x}^k\}$ *is locally convergent at* $\boldsymbol{x}^*$*, the iterative sequence* $\{H^k\}$ *and* $\{(H^k)^{-1}\}$ *are uniformly bounded and* $\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\| \le \sigma\|\boldsymbol{x}^k-\boldsymbol{x}^*\|$*. And moreover if some subsequence of* $\{\|H^k-\nabla^2 f(\boldsymbol{x}^*)^{-1}\|_M\}$ *converges to zero, then* $\{\boldsymbol{x}^k\}$ *converges Q-superlinearly to* $\boldsymbol{x}^*$*.*

Hereafter, we suppose that function f satisfies all of the assumptions of Lemma 2, and then depict some properties of $\hat{\boldsymbol{y}}^k$ in Algorithm 1 by the following propositions.

Proposition 1. *There exists* $\bar{K}>0$*, for all* $\boldsymbol{x}^{k+1}, \boldsymbol{x}^k \in D$*, it comes*

$$\|\hat{\boldsymbol{y}}^k-\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\| \le \bar{K}\max\{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|^{\rho}, \|\boldsymbol{x}^k-\boldsymbol{x}^*\|^{\rho}\}\|\boldsymbol{s}^k\|. \tag{4.6}$$

Proof. In view of the definition $\hat{\boldsymbol{y}}^k=\boldsymbol{y}^k+\hat{\gamma}\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}$ and $\hat{\gamma}$ (3.16), we know $\hat{\boldsymbol{y}}^k=\boldsymbol{y}^k$ when $(\boldsymbol{y}^k)^T\boldsymbol{s}^k+\gamma<\epsilon\|\boldsymbol{s}^k\|^2$. It yields further (4.6) holds by combining (4.2).

In other case, if $(\boldsymbol{y}^k)^T\boldsymbol{s}^k+\gamma \ge \epsilon\|\boldsymbol{s}^k\|^2$, then we have $\hat{\boldsymbol{y}}^k=\tilde{\boldsymbol{y}}^k$, it gives

$$\begin{aligned}\|\hat{\boldsymbol{y}}^k-\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\| &\le \|\boldsymbol{y}^k-\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\|+\|\gamma\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}\|\\ &\le \|\boldsymbol{y}^k-\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\|+\frac{\|\mathcal{A}(\boldsymbol{g}^{k+1})^T\boldsymbol{s}^k+\mathcal{B}(\boldsymbol{g}^k)^T\boldsymbol{s}^k+C(f^{k+1}-f^k)\|}{\|\boldsymbol{s}^k\|}.\end{aligned} \tag{4.7}$$

By virtue of the Taylor expansion, there exist $\xi,\eta\in(0,1)$, such that

$$f^{k+1}-f^k=\frac{1}{2}(\boldsymbol{g}^{k+1}+\boldsymbol{g}^k)^T-\frac{1}{4}(\boldsymbol{s}^k)^T(\nabla^2 f(\boldsymbol{x}^{\xi})-\nabla^2 f(\boldsymbol{x}^{\eta}))\boldsymbol{s}^k, \tag{4.8}$$

where $\boldsymbol{x}^{\xi}=\xi\boldsymbol{x}^k+(1-\xi)\boldsymbol{x}^{k+1}, \boldsymbol{x}^{\eta}=\eta\boldsymbol{x}^k+(1-\eta)\boldsymbol{x}^{k+1}$.

Since $\nabla^2 f(\boldsymbol{x})$ is continuous at $\boldsymbol{x}^*$, then there exists $K_1>0$, $\|\nabla^2 f(\boldsymbol{x}^*)\|\le K_1$. It yields by invoking (4.2)

$$\|\nabla^2 f(\boldsymbol{x}^{\xi})-\nabla^2 f(\boldsymbol{x}^{\eta})\| \le 2K\max\{\|\boldsymbol{x}^{\xi}-\boldsymbol{x}^*\|^{\rho}, \|\boldsymbol{x}^{\eta}-\boldsymbol{x}^*\|^{\rho}\}, \tag{4.9}$$

and there exist $K_2>0$, such that

$$\begin{aligned}(\boldsymbol{y}^k)^T\boldsymbol{s}^k &\le (\|\nabla^2 f(\boldsymbol{x}^*)\|+K\max\{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|^{\rho}, \|\boldsymbol{x}^k-\boldsymbol{x}^*\|^{\rho}\})\|\boldsymbol{s}^k\|^2\\ &\le K_2\max\{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|^{\rho}, \|\boldsymbol{x}^k-\boldsymbol{x}^*\|^{\rho}\}\|\boldsymbol{s}^k\|^2,\end{aligned} \tag{4.10}$$

and therefore we obtain (4.6) by combining (4.7)-(4.10) and taking $\bar{K}=\frac{K(1+|C|)}{2}+K_2|\mathcal{A}+\frac{C}{2}|$. □

Proposition 2. *There exists* $\tilde{\epsilon}, \tilde{L}>0$*, for all* $\boldsymbol{x}^{k+1}, \boldsymbol{x}^k \in \bar{X}(\boldsymbol{x}^*, \tilde{\epsilon})$*, we have*

$$\frac{1}{L}\|\boldsymbol{s}^k\| \le \|\hat{\boldsymbol{y}}^k\| \le \tilde{L}\|\boldsymbol{s}^k\|. \tag{4.11}$$

Proof. If $(\boldsymbol{y}^k)^T\boldsymbol{s}^k+\gamma<\epsilon\|\boldsymbol{s}^k\|^2$, then (4.11) holds obviously from (4.3).

Otherwise, $\hat{\boldsymbol{y}}^k=\tilde{\boldsymbol{y}}^k$, and then

$$\|\tilde{\boldsymbol{y}}^k\| \le \|\boldsymbol{y}^k\|+\|\gamma\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}\|$$

$$\leq \Big(L + (\frac{K|C|}{2} + K_2|\mathcal{A} + \frac{C}{2}|) \max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\}\Big)\|\boldsymbol{s}^k\|, \tag{4.12}$$

it derives the upper boundary of (4.11) by taking $\epsilon_1 = \max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|\}$, $\tilde{L} = L + (\frac{K|C|}{2} + K_2|\mathcal{A} + \frac{C}{2}|)\epsilon_1^\rho$.

About the lower boundary of (4.11), by using the continuity of $\nabla^2 f(\boldsymbol{x})$ at $\boldsymbol{x}^*$, there exist $\epsilon_2 > 0$, for all $\boldsymbol{x} \in \bar{X}(\boldsymbol{x}^*, \epsilon_2)$,

$$\|\nabla^2 f(\boldsymbol{x}) - \nabla^2 f(\boldsymbol{x}^*)\| \leq \frac{1}{9\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|}, \tag{4.13}$$

$$K\|\boldsymbol{x} - \boldsymbol{x}^*\|^\rho \leq \frac{1}{3\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|}, \tag{4.14}$$

and then for all $\boldsymbol{x}^{k+1}, \boldsymbol{x}^k \in \bar{X}(\boldsymbol{x}^*, \epsilon_2)$, we have

$$\|\nabla^2 f(\boldsymbol{x}^\xi) - \nabla^2 f(\boldsymbol{x}^\eta)\| \leq \frac{2}{9\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|}, \tag{4.15}$$

$$K_2 \max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\} \leq \frac{2K_2}{3K\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|}. \tag{4.16}$$

At the same time,

$$\begin{aligned}\|\gamma\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}\| &= \frac{1}{\|\boldsymbol{s}^k\|}\|(\mathcal{A} + \frac{C}{2})(\boldsymbol{y}^k)^T\boldsymbol{s}^k - \frac{C}{4}(\boldsymbol{s}^k)^T(\nabla^2 f(\boldsymbol{x}^\xi) - \nabla^2 f(\boldsymbol{x}^\eta))\boldsymbol{s}^k\| \\ &\leq (\frac{2K_2|\mathcal{A} + \frac{C}{2}|}{3K} + \frac{C}{18})\frac{\|\boldsymbol{s}^k\|}{\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|},\end{aligned} \tag{4.17}$$

and

$$\|\boldsymbol{y}^k\| \geq \|\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\| - \|\boldsymbol{g}^{k+1} - \boldsymbol{g}^k - \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\| \geq \frac{2\|\boldsymbol{s}^k\|}{3\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|},$$

then

$$\|\hat{\boldsymbol{y}}^k\| \geq \Big|\frac{2}{3} - \frac{2K_2|\mathcal{A} + \frac{C}{2}|}{3K} - \frac{C}{18}\Big|\frac{\|\boldsymbol{s}^k\|}{\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|},$$

and (4.11) holds with $\tilde{\epsilon} = \min\{\epsilon_1, \epsilon_2\}$ and $\tilde{L} = \max\Big\{\frac{\|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|}{\left|\frac{2}{3} - \frac{2K_2|\mathcal{A} + \frac{C}{2}|}{3K} - \frac{C}{18}\right|}, L + (\frac{K|C|}{2} + K_2|\mathcal{A} + \frac{C}{2}|)\tilde{\epsilon}^\rho\Big\}$. □

Proposition 3. *For $k \geq 1, \sigma \in (0, 1)$, suppose $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\| \leq \sigma\|\boldsymbol{x}^k - \boldsymbol{x}^*\|$ and both $\{H^k\}$ and $\{(H^k)^{-1}\}$ are bounded, then*

$$\lim_{k\to+\infty} \frac{\|(H^k - \nabla^2 f(\boldsymbol{x}^*)^{-1})\hat{\boldsymbol{y}}^k\|}{\|\hat{\boldsymbol{y}}^k\|} = 0,$$

implies that $\{\boldsymbol{x}^k\}$ superlinearly converges to $\boldsymbol{x}^*$.

Proof. If $(\boldsymbol{y}^k)^T\boldsymbol{s}^k + \gamma < \epsilon\|\boldsymbol{s}^k\|^2$, then $\hat{\boldsymbol{y}}^k = \boldsymbol{y}^k$ and Proposition 3 holds obviously from [9]. Otherwise, $\hat{\boldsymbol{y}}^k = \tilde{\boldsymbol{y}}^k$. From Algorithm 1, $H^k\boldsymbol{g}^k = -\boldsymbol{s}^k$ then we have $H^k\boldsymbol{y}^k = H^k\boldsymbol{g}^{k+1} + \boldsymbol{s}^k$ and

$$(H^k - \nabla^2 f(\boldsymbol{x}^*)^{-1})\hat{\boldsymbol{y}}^k = H^k\boldsymbol{g}^{k+1} + \boldsymbol{s}^k - \nabla^2 f(\boldsymbol{x}^*)^{-1}\boldsymbol{y}^k + \gamma\frac{(H^k - \nabla^2 f(\boldsymbol{x}^*)^{-1})\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}. \tag{4.18}$$

Let $\|H^k\| \le r$, $\|(H^k)^{-1}\| \le \tilde{r}$, then

$$\begin{aligned}\|\boldsymbol{g}^{k+1}\| &\le \|(H^k)^{-1}\|\|(H^k - \nabla^2 f(\boldsymbol{x}^*)^{-1})\tilde{\boldsymbol{y}}^k + \nabla^2 f(\boldsymbol{x}^*)^{-1}(\tilde{\boldsymbol{y}}^k - \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k) - \gamma\frac{H^k\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}\|\\ &\le \tilde{r}(\|(H^k - \nabla^2 f(\boldsymbol{x}^*)^{-1})\tilde{\boldsymbol{y}}^k\| + \|\nabla^2 f(\boldsymbol{x}^*)^{-1}\|\|\tilde{\boldsymbol{y}}^k - \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\| + \frac{\gamma r}{\|\boldsymbol{s}^k\|}),\end{aligned}$$

and it reveals after relating to Proposition 1-2,

$$\lim_{k\to+\infty}\frac{\|\boldsymbol{g}^{k+1}\|}{\|\boldsymbol{s}^k\|} = 0,$$

and

$$\lim_{k\to+\infty}\frac{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}^k - \boldsymbol{x}^*\|} = 0,$$

this completes the proof. □

Proposition 4. *If $\nabla^2 f(\boldsymbol{x}^*)$ is positive definite and the sequence $\{\boldsymbol{x}^k\}$ generated by Algorithm 1 converges to x^*, then there exists $k_0 > 0$, $(\boldsymbol{s}^k)^T\hat{\boldsymbol{y}}^k > 0$ holds for all $k \ge k_0$.*

Proof. In view of the positive definiteness of $\nabla^2 f(\boldsymbol{x}^*)$, let

$$\delta = \min_{w\in\mathbb{R}, w\neq 0}\frac{\boldsymbol{w}^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{w}} > 0$$

then link with Proposition 1, we obtain

$$\begin{aligned}\frac{(\boldsymbol{s}^k)^T\hat{\boldsymbol{y}}^k}{(\boldsymbol{s}^k)^T\boldsymbol{s}^k} &= \frac{(\boldsymbol{s}^k)^T(\hat{\boldsymbol{y}}^k - \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k)}{(\boldsymbol{s}^k)^T\boldsymbol{s}^k} + \frac{(\boldsymbol{s}^k)^T\nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k}{(\boldsymbol{s}^k)^T\boldsymbol{s}^k}\\ &\ge -\bar{K}\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\} + \delta.\end{aligned}$$

By using $\lim_{k\to+\infty}\boldsymbol{x}^k = \boldsymbol{x}^*$, there exists $k_0 > 0$, for all $k \ge k_0$, such that $\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\} < \frac{\delta}{\bar{K}}$. This concludes the proof. □

By combining Lemma 2-3 and Proposition 1-4, we could end up with the following Theorems to establish the local and superlinear convergence for our modified quasi-Newton algorithm.

Theorem 2. *Suppose that function f possesses the hypotheses of Lemma 3, if there exist a constant $\mu_1 \ge 0$, a nonsingular symmetric matrix M, for all $(\boldsymbol{x}^k, B^k)$ in a neighborhood of $(\boldsymbol{x}^*, \nabla^2 f(\boldsymbol{x}^*))$ such that*

$$\frac{\|M\boldsymbol{u}^k - M^{-1}\boldsymbol{s}^k\|}{\|M^{-1}\boldsymbol{s}^k\|} \le \mu_1\|\boldsymbol{s}^k\|^\rho, \tag{4.19}$$

then sequence $\{\boldsymbol{x}^k\}$ generated by the B-update formula (3.14) *of the modified quasi-Newton algorithm is locally and Q-superlinearly convergent at $\boldsymbol{x}^*$.*

Proof. If $(\boldsymbol{y}^k)^T\boldsymbol{s}^k + \gamma < \epsilon\|\boldsymbol{s}^k\|^2$, then $\hat{\boldsymbol{y}}^k = \boldsymbol{y}^k$ and Theorem 2 holds directly from [9, Theorem 5.3]. Now let's consider only the case $\hat{\boldsymbol{y}}^k = \tilde{\boldsymbol{y}}^k = \boldsymbol{y}^k + \gamma\frac{\boldsymbol{s}^k}{\|\boldsymbol{s}^k\|^2}$ for the B-update formula (3.14).

Assume that $0 < \mu_1\|\boldsymbol{s}^k\|^\rho \le \frac{1}{3}$ and set $A = \nabla^2 f(\boldsymbol{x}^*)$, after invoking [9, Lemma 5.2], (4.6) and (4.19), we have

$$\begin{aligned}\|B^{k+1} - \nabla^2 f(\boldsymbol{x}^*)\|_M \le& \Big(\sqrt{1-\alpha\theta^2} + \frac{5}{2(1-\beta)}\frac{\|M\boldsymbol{u}^k - M^{-1}\boldsymbol{s}^k\|}{\|M^{-1}\boldsymbol{s}^k\|}\Big)\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M \\ &+ 2(1+2\sqrt{N})\|M\|\frac{\|\tilde{\boldsymbol{y}}^k - \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{s}^k\|}{\|M^{-1}\boldsymbol{s}^k\|} \\ \le& \Big(\sqrt{1-\alpha\theta^2} + \alpha_1\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\}\Big)\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M \\ &+ \alpha_2\max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\},\end{aligned} \tag{4.20}$$

where $\beta \in [0, \frac{1}{3}]$, $\alpha = \frac{1-2\beta}{1-\beta^2} \in [\frac{3}{8}, 1]$, $\theta^k = \frac{\|M(B^k - \nabla^2 f(\boldsymbol{x}^*))\boldsymbol{s}^k\|}{\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M\|M^{-1}\boldsymbol{s}^k\|} \in [0, 1]$, $\alpha_1 = \frac{5\mu_1 2^\rho}{2(1-\beta)}$, $\alpha_2 = 2\bar{K}(1+2\sqrt{N}\|M\|^2)$.

It implies that the B-update formula (3.14) with iteration $\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - (B^k)^{-1}\nabla f(\boldsymbol{x})$ satisfies the hypotheses of Lemma 3 and therefore, the sequence $\{\boldsymbol{x}^k\}$ is locally convergent at $\boldsymbol{x}^*$, $\{B^k\}$ and $\{(B^k)^{-1}\}$ are uniformly bounded and there exists $\sigma \in (0, 1)$, $\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\| \le \sigma\|\boldsymbol{x}^k - \boldsymbol{x}^*\|$ for all $k \ge 0$.

To prove the Q-superlinear convergence, we can rewrite (4.20) as

$$\begin{aligned}\|B^{k+1} - \nabla^2 f(\boldsymbol{x}^*)\|_M \le& \sqrt{1 - \frac{3}{8}(\theta^k)^2}\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M \\ &+ \max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\}(\alpha_1\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M + \alpha_2).\end{aligned} \tag{4.21}$$

If there is a subsequence of $\{B^k\}$ which converges to $\nabla^2 f(\boldsymbol{x}^*)$, Lemma 3 will give the desired conclusion.

Otherwise, the sequence $\|B^{k+1} - \nabla^2 f(\boldsymbol{x}^*)\|_M$ is bounded away from zero. After by using $\sqrt{1-\alpha} \le 1 - \frac{\alpha}{2}$, (4.21) is equivalent to

$$\begin{aligned}\frac{3(\theta^k)^2}{16}\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M \le& \|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M - \|B^{k+1} - \nabla^2 f(\boldsymbol{x}^*)\|_M \\ &+ \max\{\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\|^\rho, \|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\}(\alpha_1\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M + \alpha_2).\end{aligned} \tag{4.22}$$

Therefore

$$\sum_{k=1}^{+\infty}(\theta^k)^2\|B^k - \nabla^2 f(\boldsymbol{x}^*)\|_M < +\infty,$$

and

$$\lim_{k\to+\infty}\frac{\|(B^k - \nabla^2 f(\boldsymbol{x}^*))\boldsymbol{s}^k\|}{\|\boldsymbol{s}^k\|} = 0. \tag{4.23}$$

By taking advantage of Proposition 1-2, we observe

$$\begin{aligned}\|(B^k - \nabla^2 f(\boldsymbol{x}^*))\boldsymbol{s}^k\| \ge& \|\boldsymbol{g}^{k+1}\| - \|\boldsymbol{g}^{k+1} - \boldsymbol{g}^k - \nabla^2 f(\boldsymbol{x}^*)\| \\ \ge& \frac{1}{L}\|\boldsymbol{x}^{k+1} - \boldsymbol{x}^*\| - \bar{K}\|\boldsymbol{x}^k - \boldsymbol{x}^*\|^\rho\|\boldsymbol{s}^k\|,\end{aligned} \tag{4.24}$$

and thus

$$\frac{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}^k-\boldsymbol{x}^*\|} \le \frac{2L\|(B^k-\nabla^2 f(\boldsymbol{x}^*))\boldsymbol{s}^k\|}{\|\boldsymbol{s}^k\|} + 2L\bar{K}\|\boldsymbol{x}^k-\boldsymbol{x}^*\|^\rho$$

together with the convergence of $\{\boldsymbol{x}^k\}$ and (4.23), it yields

$$\lim_{k\to+\infty} \frac{\|\boldsymbol{x}^{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}^k-\boldsymbol{x}^*\|} = 0.$$

This completes the proof. □

We just report the local superlinear convergence theorem about the H-update formula as follows, since the proof is very similar to that of Theorem 2.

Theorem 3. *Suppose that function f possesses the assumptions of Lemma 2, if there exist a constant $\mu_2 \ge 0$, a nonsingular symmetric matrix M, for all $(\boldsymbol{x}^k, H^k)$ in a neighborhood of $(\boldsymbol{x}^*, \nabla^2 f(\boldsymbol{x}^*)^{-1})$ such that*

$$\frac{\|M\boldsymbol{v}^k - M^{-1}\hat{\boldsymbol{y}}^k\|}{\|M^{-1}\hat{\boldsymbol{y}}^k\|} \le \mu_1\|\hat{\boldsymbol{y}}^k\|^\rho, \quad \forall \hat{\boldsymbol{y}}^k \ne 0,$$

then sequence $\{\boldsymbol{x}^k\}$ generated by the H-update formula (3.15) *of the modified quasi-Newton algorithm is locally and Q-superlinearly convergent at $\boldsymbol{x}^*$.*

5. Numerical experiments

To compare the computational performance between the classic BFGS updates (1.5) and our Algorithm as $\boldsymbol{u}^k = \frac{1}{1+w^k}\hat{\boldsymbol{y}}^k + \frac{w^k}{1+w^k}B^k\boldsymbol{s}^k$ in (3.14), where $w^k = \sqrt{\frac{(\hat{\boldsymbol{y}}^k)^T\boldsymbol{s}^k}{(\boldsymbol{s}^k)^T B^k \boldsymbol{s}^k}}$, we take the following four test functions for examples with initial points $\boldsymbol{x}^0 \in \mathbb{R}^N$ (e.g., the bold number $\mathbf{1} = (1, 1, \cdots, 1)^T \in \mathbb{R}^N$), and parameters $B^0 = \boldsymbol{I}, \alpha = 10^{-4}, \beta = 0.9, \epsilon = 10^{-10}$, and terminate the routine once iterations exceed its limit $k = 1000$ or $\|\boldsymbol{F}(\boldsymbol{x})\| \le 10^{-6}$. The numerical results are listed in Table 1–4, where K_{iter} denotes iteration times, K_F, $\|\boldsymbol{F}(\boldsymbol{x})\|$ are the number of function evaluations and the norm of $\boldsymbol{F}(\boldsymbol{x})$ at the stopping point, respectively. CPU time is the total amount of time the CPU spends in seconds. We examine empirical characteristics of our proposed algorithm in both symmetric Hessian(Test 1-3) and nonsymmetric (Test 4) settings.

Test 1. The discretized Chandrasekhar's H-equation [10]: $(i = 1, 2, \cdots, N)$

$$F_i(\boldsymbol{x}) = x_i - (1 - \frac{\sigma}{2N}\sum_{j=1}^{N}\frac{t_i x_j}{t_i + t_j})^{-1},$$

where $\sigma \in [0, 1]$, $t_i = \frac{i-0.5}{N}$, and take $\sigma = 0.9$.

Test 2. The gradient function of the Engval function [11]: $(i = 2, 3, \cdots, N-1)$

$$F_1(\boldsymbol{x}) = x_1(x_1^2 + x_2^2) - 1,$$

$$F_i(\boldsymbol{x}) = x_i(x_{i-1}^2 + 2x_i^2 + x_{i+1}^2) - 1,$$
$$F_N(\boldsymbol{x}) = x_N(x_{N-1}^2 + x_N^2).$$

Test 3. The function $\boldsymbol{F}$ is given by [12]:

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} e^{x_1} - 1 \\ e^{x_2} - 1 \\ \vdots \\ e^{x_N} - 1 \end{bmatrix}.$$

Test 4. The function $\boldsymbol{F}$ is given by [13]: $(i = 1, 2, \cdots, N-1)$

$$F_i(\boldsymbol{x}) = 2x_i - x_{i+1} + \sin(x_i) - 1,$$
$$F_N(\boldsymbol{x}) = 2x_N + \sin(x_N) - 1.$$

Table 1. Numerical results for Test 1.

$\boldsymbol{x}^0$	N	BFGS				New Algorithm 1			
		K_{iter}	N_F	$\\|\boldsymbol{F}(\boldsymbol{x})\\|$	CPU time	K_{iter}	K_F	$\\|\boldsymbol{F}(\boldsymbol{x})\\|$	CPU time
1	10	1000	100	1.002818e-01	0.691437	61	7	1.491093e-07	0.066236
	50	737	74	4.103267e-08	1.450943	57	6	9.678603e-07	0.094702
	100	1000	102	3.167766e-01	5.941363	81	9	4.859871e-07	0.259126
	500	1000	100	6.734134e-03	146.157723	22	3	8.709973e-07	1.505249
−10	10	1000	101	9.236783e-02	0.307169	47	5	4.953654e-07	0.027917
	50	284	29	8.074406e-08	0.457444	130	13	1.288553e-08	0.140885
	100	1000	100	1.325485e+02	6.129278	49	5	3.985216e-07	0.189592
	500	851	88	8.125466e-07	114.946035	85	9	5.585781e-07	5.961022
−100	10	1000	100	3.634384e+01	0.342612	128	18	3.190402e-07	0.086090
	50	1000	101	2.220825e-01	1.752874	111	12	5.135561e-07	0.140591
	100	1000	101	4.458655e+00	5.450797	222	23	7.428437e-07	0.804041
	500	167	17	4.246196e-07	21.457026	238	24	7.531092e-07	16.692411
10	10	1000	101	1.319797e+00	0.242842	37	4	7.478728e-07	0.027177
	50	1000	100	2.766506e+00	1.728747	80	8	5.345034e-07	0.243167
	100	128	16	2.408735e-10	0.695889	435	44	5.562306e-07	4.770006
$-\frac{\mathbf{10}}{N}$	10	1000	100	1.235471e+00	0.261892	25	3	4.983473e-07	0.015856
	50	1000	101	9.458086e+01	2.213521	15	2	4.268824e-07	0.031304
	100	1000	101	3.549297e-01	7.079332	202	21	3.319002e-07	0.770231

Table 2. Numerical results for Test 2.

x^0	N	BFGS				New Algorithm 1			
		K_{iter}	N_F	$\|\|F(x)\|\|$	CPU time	K_{iter}	K_F	$\|\|F(x)\|\|$	CPU time
1	10	80	2	NaN	/	116	12	9.229853e-07	0.043482
	50	113	12	9.081388e-07	0.135634	208	21	9.202787e-07	0.252628
	100	111	12	8.892927e-07	0.260747	156	16	9.734373e-07	0.357513
	500	202	21	9.879407e-07	18.138592	170	17	8.923980e-07	9.839334
−1	10	30	3	4.149480e-07	0.011782	77	8	5.385229e-07	0.018419
	50	61	7	2.599932e-07	0.063876	152	16	9.436380e-07	0.145193
	100	66	7	5.261260e-07	0.176674	286	29	7.199529e-07	0.652808
	500	65	7	8.279229e-07	5.219468	199	20	6.498353e-07	11.225447
$\frac{\mathbf{1}}{N}$	10	30	457	8.707093e-07	0.178206	96	464	9.160651e-07	0.187356
	50	65	5067	8.659048e-07	6.462482	114	5072	9.029144e-07	5.012703
	500	63	7	7.057961e-07	5.208877	120	12	9.670329e-07	6.602308
$-\frac{\mathbf{1}}{N}$	10	34	4	5.573255e-07	0.493004	52	6	7.849099e-07	0.639780
	50	138	20	7.614608e-07	23.180269	86	9	8.245071e-07	0.098294
	100	658	66	9.766928e-07	84.362569	149	15	9.919319e-07	72.960420
$\frac{\mathbf{10}}{N}$	10	80	2	NaN	/	116	12	9.229853e-07	0.023389
	50	51	6	8.319896e-07	0.058299	144	15	8.920610e-07	0.125570
	100	70	450	8.706823e-07	1.474716	98	453	9.478277e-07	1.889281
	500	150	5072	9.476521e-07	632.372397	186	5076	9.603845e-07	467.089217

Table 3. Numerical results for Test 3.

x^0	N	BFGS				New Algorithm 1			
		K_{iter}	N_F	$\|\|F(x)\|\|$	CPU time	K_{iter}	K_F	$\|\|F(x)\|\|$	CPU time
−50	10	141	15	1.216419e-06	0.280529	50	5	6.712844e-07	0.066971
	20	289	29	3.348452e-07	0.374504	109	11	8.229254e-07	0.156077
	30	594	60	2.921977e-07	0.992670	198	20	5.910921e-07	0.337214
	100	1000	100	1.022180e+01	6.126605	694	70	7.985341e-07	5.177154
−100	10	120	49	NaN	/	75	8	9.849854e-07	0.078148
	20	449	45	5.128064e-07	0.627262	138	15	5.450494e-07	0.191452
	30	567	57	6.160162e-07	1.026790	276	28	8.909035e-07	0.473574
	50	1000	100	8.627459e+00	2.514872	717	72	7.986347e-07	2.462645
−10	20	309	31	6.116347e-07	0.347280	114	13	7.089407e-07	0.184445
	30	298	30	3.312465e-07	0.538713	208	21	8.466287e-07	0.415286
	50	519	52	1.813603e-07	1.341700	253	26	9.311691e-07	0.774297
	100	1000	100	7.617593e+00	6.162504	355	36	7.219017e-07	2.840285
5	10	216	22	7.037623e-07	0.127426	167	17	6.258216e-07	0.156327
	20	421	43	8.108269e-07	0.522406	184	19	5.406795e-07	0.243489
	39	1000	100	1.800895e+01	1.887971	596	60	8.097367e-07	1.439646
	49	1000	100	1.558516e+01	2.578864	826	83	6.960628e-07	2.449176

Table 4. Numerical results for Test 4.

x^0	N	BFGS				New Algorithm 1			
		K_{iter}	N_F	$\|\|F(x)\|\|$	CPU time	K_{iter}	K_F	$\|\|F(x)\|\|$	CPU time
10	59	1000	107	5.550772e+05	1.855848	1000	100	9.865474e-03	1.044976
	69	1000	110	1.072615e+08	2.191505	659	66	8.710266e-07	0.733924
	99	1000	111	8.862006e+04	3.701440	675	68	9.735961e-07	1.058931
−10	30	1000	105	5.569634e+05	0.687643	740	74	8.757998e-07	0.469053
	50	1000	113	2.212649e+04	1.461766	1000	101	8.082762e-03	0.835354
	79	1000	109	3.080223e+04	2.498293	692	70	9.384950e-07	0.861929
	99	1000	113	8.054627e+05	3.881890	1000	100	9.210557e-03	1.539297
	100	1000	103	3.603290e+05	3.694832	659	66	9.089866e-07	1.047995
50	20	1000	110	5.216235e+01	0.490500	1000	100	5.732373e-05	0.490532
	40	1000	106	7.676800e+06	0.949262	1000	101	7.887017e-04	0.657374
−50	39	1000	109	1.827236e+05	1.011190	1000	100	2.082736e-03	0.680466
	59	1000	109	2.952795e+06	1.769789	680	68	6.906835e-07	0.646690
−1	10	1000	100	3.512951e+01	0.265638	261	27	9.980125e-07	0.125660
	29	1000	100	8.197235e+00	0.683372	766	77	9.100291e-07	0.426080
	39	1000	100	9.067728e+00	0.855453	591	60	9.402869e-07	0.455313
	59	1000	100	1.033973e+01	1.393942	877	88	8.623849e-07	0.883676

As reported in Table 4 for the nonsymmetric problem, the new Algorithm performs better than the classic BFGS updates. In fact, there is a little numerically instable for the latter method and without any instability for the new Algorithm. Numerical experiments from Tables 1–4 illustrate that the new Algorithm is in good agreement with the theoretical results, and performs well even for high-dimensional cases. Further more, the new Algorithm behaves more efficiently than the classic BFGS method since it requires less function calculations for most initial points.

6. Conclusions

We propose a modified quasi-Newton method based on the classic BFGS updates and Wolfe line search technique to solve a unconstrained optimization problem, and attain the local and superlinear rate of convergence of the modified quasi-Newton updates. Some numerical experiments well verified the theoretical results, and confirmed the efficiency to solve nonlinear equations not only for small-medium dimensions but also even for large-scale or nonsymmetric problems.

Author contributions

Wen Zhang: Conceptualization, investigation, methodology, resources, writing-original draft, writing-review and editing, project administration; Tingting Guo: Data curation, software, writing-original draft; Junfeng Wu: Validation, visualization, software, writing-review; Zhousheng Ruan: Methodology, supervision, writing-review; Shufang Qiu: Formal analysis, methodology, writing-review. All authors read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgements

The authors would like to thank editors and reviewers for their valuable suggestions and comments. This work is partially supported by National Natural Science Foundation of China (11861007, 12061008, 12261004), Natural Science Foundation of Jiangxi Province of China (20232BAB201019), Guangdong Basic and Applied Basic Research Foundation (2025A1515012248), Innovation Team Project of Regular Universities in Guangdong Province (2025KCXTD037).

Conflict of interest

The authors declare no conflict of interest.

References

1. W. Y. Sun, Y. X. Yuan, *Optimization theory and methods: Nonlinear programming*, New York: Springer Science+Business Media, LLC, 2006.
2. H. Y. Liu, J. Hu, Y. F. Li, Z. W. Wen, *Optimization: Modeling, algorithm and theory*, Beijing: Higher Education Press, 2020.
3. J. Nocedal, S. J. Wright, *Numerical optimization*, 2 Eds., New York: Springer, 2006.
4. L. Zhang, A derivative-free conjugate residual method using secant condition for general large-scale nonlinear equations, *Numerical Algorithms*, **83** (2020), 1277–1293. https://doi.org/10.1007/s11075-019-00725-7
5. R. Bollapragada, S. M. Wild, Adaptive sampling quasi-Newton methods for zeroth-order stochastic optimization, *Math. Program. Comput.*, **15** (2023), 327–364. https://doi.org/10.1007/s12532-023-00233-9
6. X. Y. Song, Q. Y. Zhang, C. Lee, E. Fertig, T. K. Huang, L. Belenki, et al., The vizier gaussian process bandit algorithm, *arXiv preprint arXiv:2408.11527*, 2024.
7. D. Y. Wang, G. Liang, G. W. Liu, K. Zhang, Derivative-free optimization via finite difference approximation: An experimental study, *Asia Pac. J. Oper. Res. (APJOR)*, **42** (2025), 1–23. https://doi.org/10.1142/S0217595925400056
8. J. Z. Zhang, N. Y. Deng, L. H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, *J. Optimiz. Theory Appl.*, **102** (1999), 147–167. https://doi.org/10.1023/A:1021898630001
9. C. G. Broyden, J. E. Dennis, J. J. More, On the local and superlinear convergence of quasi-Newton methods, *IMA J. Appl. Math.*, **12** (1973), 223–245. https://doi.org/10.1093/imamat/12.3.223
10. C. T. Kelley, *Iterative methods for linear and nonlinear equations*, Philadelphia: SIAM, 1995.
11. W. J. Zhou, F. Wang, A PRP-based residual method for large-scale monotone nonlinear equations, *Appl. Math. Comput.*, **261** (2015), 1–7. https://doi.org/10.1016/j.amc.2015.03.069
12. W. J. Zhou, D. M. Shen, Convergence properties of an iterative method for solving symmetric nonlinear equations, *J. Optimiz. Theory Appl.*, **164** (2015), 277–289. https://doi.org/10.1007/s10957-014-0547-1

13. W. J. Zhou, A modified BFGS type quasi-Newton method with line search for symmetric nonlinear equations problems, *J. Comput. Appl. Math.*, **367** (2020), 112454. https://doi.org/10.1016/j.cam.2019.112454
14. Z. Wan, K. L. Teo, X. L. Shen, C. M. Hu, New BFGS method for unconstrained optimization problem based on modified Armijo line search, *Optimization*, **63** (2014), 285–304. https://doi.org/10.1080/02331934.2011.644284
15. G. L. Yuan, Z. Sheng, B. P. Wang, W. J. Hu, C. N. Li, The global convergence of a modified BFGS method for nonconvex functions, *J. Comput. Appl. Math.*, **327** (2018), 274–294. https://doi.org/10.1016/j.cam.2017.05.030
16. G. L. Yuan, P. Y. Li, J. Y. Lu, The global convergence of the BFGS method with a modified WWP line search for nonconvex functions, *Numerical Algorithms*, **91** (2022), 353–365. https://doi.org/10.1007/s11075-022-01265-3
17. A. Upadhayay, D. Ghosh, K. Kumar, Nonmonotone Wolfe-type quasi-Newton methods for multiobjective optimization problems, *Optimization*, **Preprint** (2025), 1–33. https://doi.org/10.1080/02331934.2025.2475203
18. Y. L. Cheng, J. Gao, An efficient augmented memoryless quasi-Newton method for solving large-scale unconstrained optimization problems, *AIMS Mathematics*, **9** (2024), 25232–25252. https://doi.org/10.3934/math.20241231
19. X. Y. Zhang, Y. T. Yang, A new hybrid conjugate gradient method close to the memoryless BFGS quasi-Newton method and its application in image restoration and machine learning, *AIMS Mathematics*, **9** (2024), 27535–27556. https://doi.org/10.3934/math.20241337