Acta Mathematica Scientia, 2024, **44B**(5): 1735–1746
https://doi.org/10.1007/s10473-024-0506-3

Acta Mathematica Scientia 数学物理学报
http://actams.apm.ac.cn

GLOBAL CONVERGENCE OF A CAUTIOUS PROJECTION BFGS ALGORITHM FOR NONCONVEX PROBLEMS WITHOUT GRADIENT LIPSCHITZ CONTINUITY*

Gonglin YUAN (袁功林) *Xiong ZHAO* (赵雄) *Jiajia YU* (余家佳)†
School of Mathematics and Information Science, Center for Applied Mathematics of Guangxi (Guangxi University), Guangxi University, Nanning 530004, China
E-mail: *glyuan@gxu.edu.cn; chao18874188217@163.com; 8409855@qq.com*

Abstract A cautious projection BFGS method is proposed for solving nonconvex unconstrained optimization problems. The global convergence of this method as well as a stronger general convergence result can be proven without a gradient Lipschitz continuity assumption, which is more in line with the actual problems than the existing modified BFGS methods and the traditional BFGS method. Under some additional conditions, the method presented has a superlinear convergence rate, which can be regarded as an extension and supplement of BFGS-type methods with the projection technique. Finally, the effectiveness and application prospects of the proposed method are verified by numerical experiments.

Key words cautious BFGS; nonconvex problems; Lipschitz continuity; projection technique; global convergence

2020 MR Subject Classification 90C30; 90C26; 65Y20; 65K05

1 Introduction

The problem we consider is as follows:

$$\min\{f(x) \mid x \in \mathbb{R}^n\}. \tag{1.1}$$

Here $f : \mathbb{R}^n \to \mathbb{R}$ and is continuously differentiable (possibly nonconvex). The optimization model of (1.1) has been widely used in many fields [1–5]. As one of the most effective quasi-Newton methods for solving (1.1), the popular Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [6–9] updates B_j via

$$B_{j+1} = B_j - \frac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \frac{y_j y_j^T}{s_j^T y_j}, \tag{1.2}$$

*Received February 16, 2023; revised June 16, 2024. Yuan's research was supported by the Guangxi Science and Technology base and Talent Project (AD22080047), the National Natural Science Foundation of Guangxi Province (2023GXNFSBA 026063), the Innovation Funds of Chinese University (2021BCF03001), and the special foundation for Guangxi Ba Gui Scholars.

†Corresponding author

where $s_j = x_{j+1} - x_j$ (x_j is the value of x at iteration j), $y_j = g_{j+1} - g_j$ ($g_j = g(x_j) = \nabla f(x_j)$ is the gradient of f at x_j), and B_j is the Hessian approximation at iteration j. Correspondingly, the iterative formula requires steps

$$x_{j+1} = x_j - \alpha_j H_j g_j, \tag{1.3}$$

where $H_j = B_j^{-1}$ and α_j is the step-size, which is determined by weak Wolfe-Powell (WWP) line search conditions

$$f(x_j + \alpha_j d_j) - f(x_j) \le \zeta_1 \alpha_j g(x_j)^T d_j, \tag{1.4}$$

$$g(x_j + \alpha_j d_j)^T d_j \ge \zeta_2 g(x_j)^T d_j, \tag{1.5}$$

where $0 < \zeta_1 < \zeta_2 < 1$. For convex problems, the global convergence of the BFGS method is well established under some line search conditions [10–15]. Under inexact Wolfe line search conditions Powell [16] proved that BFGS method has a superlinear convergence rate. To improve the performance of the BFGS method, some modified BFGS methods and design skills have been proposed [17–23]. However, for nonconvex problems, even under the Wolfe line search conditions, the BFGS method may not converge [24]. The main reason for that may be the following inequality cannot be deduced [10] that

$$\frac{\|y_j\|^2}{s_j^T y_j} \le M, \tag{1.6}$$

where M is a positive constant. To overcome this difficulty, many scholars have proposed some modified BFGS methods [25–38]. For example, there is the BFGS update rule modified by Li and Fukushima [27] (called cautious BFGS),

$$B_{j+1} = \begin{cases} B_j - \dfrac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \dfrac{y_j y_j^T}{y_j^T s_j}, & \text{if } \dfrac{y_j^T s_j}{\|s_j\|^2} \ge \epsilon \|g_j\|^\alpha, \\ B_j, & \text{otherwise}, \end{cases} \tag{1.7}$$

where $\epsilon > 0$ and $\alpha > 0$ are constants. In [25], for when a specific sufficient descent condition is not satisfied, Yuan et al. proposed a projection iteration approach:

$$x_{j+1} = x_j + \frac{\gamma \|\mathcal{X}_j - x_j\|^2 + (\mathcal{X}_j - x_j)^T g(\mathcal{X}_j)}{\|g(\mathcal{X}_j) - g(x_j)\|^2} [g(\mathcal{X}_j) - g(x_j)]. \tag{1.8}$$

Here $\mathcal{X}_j = x_j + \alpha_j d_j$, $d_j = -H_j g_j$ is the search direction and γ is a positive constant; otherwise, (1.3) proceeds as usual. The reason that (1.8) is called a projection is that when the special sufficient descent condition is not satisfied, the current point x_j is projected onto the paraboloid $\gamma\|\mathcal{X}_j - x\|^2 + (W_j - x)^T g(W_j)$ and then through the appropriate transformation to obtain the formula (1.8) (this treatment makes the convergence analysis possible). Combined with a modified weak-Wolfe-Powell line search [29], this method can also converge globally for nonconvex problems. All of the methods above, as well as the classical methods, require the assumption of gradient Lipschitz continuity (even stronger gradient assumptions are required for the projection iteration method [25]: $\|g(x) - g(y)\| = O(\|x - y\|)$, $x, y \in \mathbb{R}^n$). However, this assumption is not entirely reasonable for the general problems. For example, the functions $f_1(x) = \ln x$ and $f_2(x) = x^{\frac{1}{2}}$, $x \in \mathbb{R}^n$ do not satisfy the gradient Lipschitz continuity condition in the neighborhood of zero. In addition, these methods show that the sequence $\{x_j\}$ can only converge to some accumulation point rather than a unique stationary point.

Our Contributions The contributions of this paper are as follows:

⋆ A new cautious BFGS method combined with projection technique is proposed to solve (1.1).

⋆ The presented algorithm converges globally for general problems without a gradient Lipschitz continuity condition.

⋆ It is also demonstrated that the presented algorithm with a projection technique has a superlinear convergence rate, which is an extension and supplement to the existing projection BFGS methods [25, 28].

⋆ A stronger general convergence result is demonstrated, which shows that under some assumptions, the sequence $\{x_j\}$ generated by BFGS-type methods can converge to a unique stationary point of f.

⋆ The performance of numerical experiments suggests that the algorithm has strong competitiveness and research prospects.

2 Motivation and Background of the Algorithm

The idea given by Yuan *et al.* [25] is relative to the descent condition $-d_j^T g_j > \rho\alpha_j\|d_j\|^2$, where the process (1.8) will be carried out when the condition is not satisfied; otherwise, iteration will proceed normally. This technique can ensure that the presented algorithm is globally convergent for non-convex functions if the following assumptions are satisfied: $\|g(x) - g(y)\| = O(\|x-y\|)$, $x, y \in \mathbb{R}^n$. In our opinion, nonconvex unconstrained optimization is an interesting problem, and it is valuable to study this method. However, there are two questions that deserve our attention: 1) is it possible to weaken or even remove the above assumption to achieve the global convergence? 2) how can we ensure the global convergence if the assumption is removed? Motivated by the above section, we know that the idea of the cautious BFGS method (1.7) may be used to satisfy some special conditions, so it is not difficult for us to use this sufficiently and to answer the above two questions. We fully expect the designed algorithm to satisfy the inequality (1.6), and we will further discuss the convergence rate of the projection algorithm. Thus a cautious projection BFGS method is designed. For the sake of simplicity, a special sufficient descent set is defined by

$$SD_j := \{x_j | d_j^T g_j \le -\rho\alpha_j\|d_j\|^2\|g_j\|^\alpha, j \ge 0\}, \tag{2.1}$$

and an index set by

$$\tau_j := \{j | \|\alpha_j d_j\| \|g_j\|^\alpha \le \|y_j\| \le M'\|\alpha_j d_j\|\}, \tag{2.2}$$

where $M' = O(\|g(x_j)\|^{-\alpha})$ (which can be taken as a large parameter), $\rho > 0$, and α is a positive tuning parameter. The setting of these two sets is indispensable for discussing the global convergence and convergence rate of the proposed algorithm. As for the case $x_j \notin SD_j$, x_j does not meet the sufficient descent condition. We hope to achieve the same effect by projection, so an adaptive surface needs to be introduced:

$$\{x \in \mathbb{R}^n | \mu(x)\|\mathcal{X}_j - x\|^2 + g(x)^T(\mathcal{X}_j - x) = 0\}. \tag{2.3}$$

Here $\mathcal{X}_j = x_j + \alpha_j d_j$, $\mu(x) = \mu\|g(x)\|^\alpha$ is an adaptive term and $\mu > \rho$ is a given positive constant. It is worth noting that the global convergence of the proposed algorithm can still be proven when $\alpha = 0$, according to the difference of α, the performance of the proposed algorithm will also be different, this will be shown in the numerical results section later.

The iterative process is now discussed briefly. Considering that the current point x_j and assuming the step-size α_j and the search direction d_j have been obtained, the generation of the next point x_{j+1} will take the following steps:

Case (i) $j \in SD_j$. $x_{j+1} = \mathcal{X}_j = x_j + \alpha_j d_j$ proceeds as usual.

Case (ii) $j \notin SD_j$. Project the current point x_j onto the surface (2.3) to get P_j, and the next iteration x_{j+1} is defined as

$$x_{j+1} = x_j + \frac{P_j}{\|g(\mathcal{X}_j) - g(x_j)\|^2}[g(\mathcal{X}_j) - g(x_j)], \tag{2.4}$$

where

$$P_j = \mu_j\|\mathcal{X}_j - x_j\|^2 + (\mathcal{X}_j - x_j)^T g(x_j), \text{ and } \mu_j = \mu\|g(x_j)\|^\alpha. \tag{2.5}$$

Finally we adopt an iterative approach similar to (1.7) to update B_j, which completes a brief iteration. More specifically, the complete process of a cautious projection BFGS algorithm is shown in Algorithm 1.

Algorithm 1 Cautious Projection BFGS algorithm (CPBFGS)

Initialization Choose an initial point $x_0 \in \mathbb{R}^n$, an initial symmetric and positive definite matrix $B_0 \in \mathbb{R}^{n\times n}$, the necessary parameters $\rho \in (0, +\infty)$, $\mu \in (\rho, +\infty)$, $\zeta_1 \in (0, \frac{1}{2})$, $\zeta_2 \in (\zeta_1, 1)$ and $\epsilon \in (0, \frac{1}{2})$, and a tuning parameter $\alpha \in (0, +\infty)$. Set that $j := 0$. Compute that $\|g_j\| = \nabla f(x_j)$.

While $\|g_j\| \geq \epsilon$ **do**

1. Compute the direction d_j by solving the linear equation

$$B_j d_j + g_j = 0. \tag{2.6}$$

2. Find a step-size $\alpha_j > 0$ satisfying the WWP line search conditions (1.4) and (1.5).
3. Set that $\mathcal{X}_j := x_j + \alpha_j d_j$ and $y_j := g(\mathcal{X}_j) - g(x_j)$.
4. If $x_j \in SD_j$, set that $x_{j+1} := \mathcal{X}_j$ and $s_j := x_{j+1} - x_j$, then go to step 6.
5. Otherwise $x_j \notin SD_j$. Determine the next iteration point x_{j+1} generated by the projection technique (2.4) and set that $s_j := x_{j+1} - x_j$.
6. Define τ_j by (2.2) and update B_{j+1} by the cautious BFGS formula

$$B_{j+1} = \begin{cases} B_j - \dfrac{B_j s_j s_j^T B_j}{s_j^T B_j s_j} + \dfrac{y_j y_j^T}{y_j^T s_j}, & \text{if } j \in \tau_j, \\ B_j, & \text{otherwise .} \end{cases} \tag{2.7}$$

7. Set that $j := j + 1$.

End while

3 Global Convergence of CPBFGS

In this section, we discuss the global convergence of the CPBFGS method and its convergence rate under the weakened assumptions. We assume that f satisfies the following conditions:

Assumption 3.1 The level set $S_0 = \{x \in \mathbb{R}^n | f(x) \le f(x_0)\}$ is bounded, and $f(x)$ is bounded below.

Assumption 3.2 The gradient function $g(x) = \nabla f(x)$ is continuous, and $\|g(x)\|$ is bounded above on S_0; namely, there is a positive constant $G_0 > 0$ satisfying that

$$\|g(x)\| \le G_0, \quad \forall x \in S_0. \tag{3.1}$$

Remark 1 Note that the existing modified BFGS methods [25–28] need for the gradient function to be Lipschitz continuous (projection algorithms require stronger one [25]).

Assumption 3.3 When $j \notin SD_j$, there will always be at least one step-size α_j that satisfies the condition that

$$f(x_{j+1}) - f(x_j) \le \zeta_1 \alpha_j g(x_j)^T d_j. \tag{3.2}$$

Lemma 3.4 Suppose that Assumptions 3.1–3.2 hold and that $\{x_j\}$ is generated by Algorithm 1. Then the following inequality holds that

$$s_j^T y_j \ge \lambda \|\alpha_j d_j\|^2 \|g_j\|^\alpha, \quad \forall j \ge 0, \tag{3.3}$$

where $\lambda > 0$ is a constant.

Proof The proof can be divided into two cases.

Case (i) $x_j \in SD_j$. From step 4 of Algorithm 1, we can easily obtain that

$$\begin{aligned} s_j^T y_j &= s_j^T [g(\mathcal{X}_j) - g(x_j)] \\ &= s_j^T [g(x_{j+1}) - g(x_j)] \\ &\ge -(1 - \zeta_2) g(x_j)^T s_j \\ &= \alpha_j (1 - \zeta_2)(-g_j^T d_j) \\ &\ge (1 - \zeta_2)\rho \|\alpha_j d_j\|^2 \|g_j\|^\alpha, \end{aligned}$$

where the first and the last inequality follow the second WWP line search condition (1.5) and the definition of the sufficient descent set SD_j, respectively.

Case (ii) $x_j \notin SD_j$. According to step 5 of Algorithm 1 and the definition of P_j, we can derive that

$$\begin{aligned} s_j^T y_j &= \frac{P_j}{\|g(\mathcal{X}_j) - g(x_j)\|^2} [g(\mathcal{X}_j) - g(x_j)]^T [g(\mathcal{X}_j) - g(x_j)] \\ &= g(x_j)^T (\mathcal{X}_j - x_j) + \mu_j \|\mathcal{X}_j - x_j\|^2 \\ &= \alpha_j g(x_j)^T d_j + \mu_j \|\alpha_j d_j\|^2 \\ &> \alpha_j (-\rho \alpha_j \|d_j\|^2 \|g_j\|^\alpha) + \mu \|g_j\|^\alpha \|\alpha_j d_j\|^2 \\ &= (\mu - \rho) \|\alpha_j d_j\|^2 \|g_j\|^\alpha, \end{aligned}$$

where the inequality can be deduced from the definition $x_j \notin SD_j$. Let $\lambda := \max\{1 - \zeta_2, \mu - \rho\}$, the proof is complete. □

Remark 2 By this Lemma, we have that $s_j^T y_j > 0$ which means that B_j is positive definite for all j in Algorithm 1.

The following useful conclusion needs to be cited (Theorem 2.1 in [12]):

Lemma 3.5 Suppose that Assumptions 3.1–3.2 hold and that $\{x_j\}$ is generated by Algorithm 1. If there are positive constants $\mathfrak{m}_1 \le \mathfrak{m}_2$ such that

$$\frac{y_j^T s_j}{\|s_j\|^2} \ge \mathfrak{m}_1 \quad \text{and} \quad \frac{\|y_j\|^2}{y_j^T s_j} \le \mathfrak{m}_2, \quad \forall j \ge 0 \tag{3.4}$$

hold, then there exist constants $\mathfrak{M}_1 > \mathfrak{M}_2 > 0, \mathfrak{M}_3 > 0$ such that, for any integer $t > 0$,

$$\|B_j s_j\| \le \mathfrak{M}_1 \|s_j\| \tag{3.5}$$

and

$$\mathfrak{M}_2 \|s_j\|^2 \le s_j^T B_j s_j \le \mathfrak{M}_3 \|s_j\|^2 \tag{3.6}$$

hold for at least $\lceil t/2 \rceil$ values of $j \in \{1, \cdots, t\}$.

Now we analyze the global convergence of the CPBFGS method.

Theorem 3.6 Suppose that Assumptions 3.1–3.2 hold and that $\{x_j\}$ is generated by Algorithm 1. Then the CPBFGS converges globally, i.e., we have that

$$\liminf_{j \to +\infty} \|g_j\| = 0. \tag{3.7}$$

Proof We use the method of contradiction. Without loss of generality, there must be a positive constant ϵ such that

$$\|g_j\| \ge \epsilon, \ \forall j. \tag{3.8}$$

According to the definition of M in (2.2) and the inequality (3.1), there are positive constants, a_1 and a_2, satisfying that $a_1 \le M \le a_2$.

First, when τ_j is infinite, for all $j \in \tau_j$, we consider the following two cases:

Case (i) $x_j \in SD_j$. We can easily conclude the following inequalities from Lemma 3.4 and the definition of τ_j:

$$\frac{s_j^T y_j}{\|s_j\|^2} = \frac{s_j^T y_j}{\|\alpha_j d_j\|^2} \ge \lambda \|g_j\|^\alpha \ge \lambda \epsilon^\alpha \tag{3.9}$$

and

$$\frac{\|y_j\|^2}{y_j^T s_j} \le \frac{M'^2 \|\alpha_j d_j\|^2}{\lambda \|\alpha_j d_j\|^2 \|g_j\|^\alpha} \le \frac{M'^2}{\lambda \epsilon^\alpha}. \tag{3.10}$$

Case (ii) $x_j \notin SD_j$, then $0 \le -g_j^T d_j < \rho \alpha_j \|d_j\|^2 \|g_j\|^\alpha$. By the definition of s_j and τ_j, we have that

$$\begin{aligned}
\|s_j\| &= \|\frac{P_j}{\|g(\mathcal{X}_j) - g(x_j)\|^2} [g(\mathcal{X}_j) - g(x_j)]\| \\
&= \frac{\|P_j\|}{\|g(x_j + \alpha_j d_j) - g(x_j)\|} \\
&\le \frac{\|P_j\|}{\|\alpha_j d_j\| \|g_j\|^\alpha} \\
&\le \frac{\mu_j \|\alpha_j d_j\|^2 + \alpha_j |g_j^T d_j|}{\|\alpha_j d_j\| \epsilon^\alpha} \\
&\le \frac{\mu \|\alpha_j d_j\|^2 \|g_j\|^\alpha + \rho \|\alpha_j d_j\|^2 \|g_j\|^\alpha}{\|\alpha_j d_j\| \epsilon^\alpha} \\
&\le \frac{(\mu + \rho) G_0^\alpha}{\epsilon^\alpha} \|\alpha_j d_j\|,
\end{aligned}$$

where the second inequality is derived from the Cauchy inequality, and the last inequality can be obtained by (3.1). From Lemma 3.4 and the definition of τ_j, we have that

$$\frac{s_j^T y_j}{\|s_j\|^2} \geq \frac{\lambda\|\alpha_j d_j\|^2\|g_j\|^\alpha}{(\frac{(\mu+\rho)G_0^\alpha}{\epsilon^\alpha})^2\|\alpha_j d_j\|^2} \geq \frac{\lambda\epsilon^{3\alpha}}{((\mu+\rho)G_0^\alpha)^2} \tag{3.11}$$

and

$$\frac{\|y_j\|^2}{y_j^T s_j} \leq \frac{M'^2\|\alpha_j d_j\|^2}{\lambda\|\alpha_j d_j\|^2\|g_j\|^\alpha} \leq \frac{M'^2}{\lambda\epsilon^\alpha}. \tag{3.12}$$

Overall, for all $j \in \tau_j$ and both cases, there are positive constants λ_1 and λ_2 such that

$$\frac{y_j^T s_j}{\|s_j\|^2} \geq \lambda_1 \text{ and } \frac{\|y_j\|^2}{y_j^T s_j} \leq \lambda_2, \tag{3.13}$$

where $\lambda_1 = \max\{\lambda\epsilon^\alpha, \frac{\lambda\epsilon^{3\alpha}}{((\mu+\rho)G_0^\alpha)^2}\}$ and $\lambda_2 = \frac{M'^2}{\lambda\epsilon^\alpha}$. Without loss of generality, having $\mathfrak{m}_1 = \lambda_1$ and $\mathfrak{m}_2 = \lambda_2$, according to Lemma 3.5, we immediately get (3.5) and (3.6). Under Assumption 3.1 and 3.3, summing both sides of (1.4) from $j = 0$ to $+\infty$ yields that

$$\sum_{j=0}^{+\infty} -\zeta_1\alpha_j g(x_j)^T d_j < +\infty, \tag{3.14}$$

which implies that

$$-\alpha_j g_j^T d_j \to 0, \quad \text{as}\, j \to +\infty. \tag{3.15}$$

Then, define that $A = \limsup_{j\to+\infty} \alpha_j$. When $A > 0$, there exist a subsequence $\{\alpha_{j_l}\}$ of sequence $\{\alpha_j\}$ and a positive constant a such that $\alpha_{j_l} > a$, as $l \to +\infty$. From (3.15), we can infer that $-g_{j_l}^T d_{j_l} \to 0$, as $l \to +\infty$. By (3.6), we have that

$$\mathfrak{M}_2\|d_{j_l}\|^2 \leq d_{j_l}^T B_{j_l} d_{j_l} = -g_{j_l}^T d_{j_l} \to 0, \quad \text{as} \quad l \to +\infty, \tag{3.16}$$

so $\|d_{j_l}\| \to 0$, which further yields that

$$\|g_{j_l}\| = \|B_{j_l} d_{j_l}\| \leq \mathfrak{M}_1\|d_{j_l}\| \to 0, \quad \text{as} \quad l \to +\infty, \tag{3.17}$$

where the inequality follows (3.5), which contradicts (3.8). When $A = 0$, $\lim_{j\to+\infty} \alpha_j = 0$. Note that, from (3.6), we obtain that

$$\mathfrak{M}_2\|d_j\|^2 \leq d_j^T B_j d_j = -g_j^T d_j \leq \|g_j\|\|d_j\| \leq G_0\|d_j\|, \tag{3.18}$$

which means that $\|d_j\|$ is bounded, and further yields that $\alpha_j d_j \to 0$, as $j \to +\infty$. Then $\rho\alpha_j\|d_j\|^2\|g_j\|^\alpha \leq \mathfrak{M}_2\|d_j\|^2 \leq -g_j^T d_j$ is always true when $j \to +\infty$, which also means that the sufficient descent condition (2.1) is automatically satisfied, i.e., iteration $x_{j+1} = x_j + \alpha_j d_j$ always holds. Due to the fact that level set is bounded, there must be a subsequence $\{x_{j_k}\}$ of $\{x_j\}$ that converges to some accumulation point x^*, as $k \to +\infty$. By the continuity of $g(x)$, we have that

$$\|g(x_{j_k}) - g(x^*)\| \to 0, \quad \text{as} \quad k \to +\infty. \tag{3.19}$$

From (3.6) and the WWP condition (1.5), it can be deduced that

$$\begin{aligned}(1-\zeta_2)\mathfrak{M}_2\|d_{j_k}\|^2 &\leq (1-\zeta_2)d_{j_k}^T B_{j_k} d_{j_k}\\ &= -(1-\zeta_2)g(x_{j_k})^T d_{j_k}\\ &\leq (g(x_{j_k+1}) - g(x_{j_k}))^T d_{j_k}\end{aligned}$$

$$\leq \|g(x_{j_k+1}) - g(x_{j_k})\| \|d_{j_k}\|$$
$$\leq (\|g(x_{j_k}) - g(x^*)\| + \|g(x_{j_k+1}) - g(x^*)\|) \|d_{j_k}\|,$$

which combined with (3.19), gives that $\|d_{j_k}\| \to 0$ as $k \to +\infty$. Using (3.6), we can derive the same contradiction as (3.17). The proof of the first part is complete.

Second, when τ_j is finite, after a finite number of iterations, B_j is not updated (i.e. B_j is uniformly positive definite for all $j > 0$). Therefore, for infinitely many $j > 0$, there exist constants $\mathfrak{M}_1, \mathfrak{M}_2$ and $\mathfrak{M}_3 > 0$ such that (3.5) and (3.6) hold. Thus we can easily complete the proof of this part with the same steps. □

In the next theorem we analyze the convergence rate of the CPBFGS method under additional assumptions.

Assumption 3.7 1) $f(x)$ is twice continuously differentiable.

2) $\{x_j\}$ converges to an isolated accumulation point x^* at which $g(x^*) = 0$.

3) The Hessian matrix $G(x)$ of $f(x)$ is Hölder continuous and is positive definite at x^*.

Theorem 3.8 Supposing that Assumptions 3.1–3.3 and 3.7 hold, $\{\alpha_j\}$ satisfies the strong Wolfe condition, and $\{x_j\}$ is generated by Algorithm 1. When j is large enough, the convergence rate of the CPBFGS method is superlinear.

Proof The strong Wolfe condition means that (1.5) is replaced by

$$|g(x_j + \alpha_j d_j)^T d_j| \leq -\zeta_2 g(x_j)^T d_j. \tag{3.20}$$

Due to the mean-value theorem and Assumption 3.7, we can deduce that when j is large enough, $y_j = \int_0^1 G(x_j + r\alpha_j d_j)\alpha_j d_j dr$ holds, which further yields that there is a positive constant m such that

$$y_j^T(\alpha_j d_j) \geq m\|\alpha_j d_j\|^2, \tag{3.21}$$

i.e., $(g(x_j + \alpha_j d_j) - g_j)^T d_j \geq m\alpha_j \|d_j\|^2$. By combining with (3.20) we get that

$$-(\zeta_2 + 1)d_j^T g_j \geq d_j^T(g(x_j + \alpha_j d_j) - g_j) \geq m\alpha_j\|d_j\|^2 \geq (\zeta_2 + 1)\rho\alpha_j\|d_j\|^2\|g_j\|^\alpha, \tag{3.22}$$

which means that when j is large enough, $x_j \in SD_j$ is always satisfied. Therefore, we only need to consider $x_j \in SD_j$, where $s_j = x_{j+1} - x_j = \alpha_j d_j$. Since the matrices $\{\int_0^1 G(x_j + r\alpha_j d_j)\mathrm{d}r\}$ are uniformly positive definite, we can know that there are obviously positive constants m_1 and m_2 such that

$$\|y_j\| \leq m_1\|s_j\| \text{ and } y_j^T s_j \geq m_2\|s_j\|^2, \tag{3.23}$$

which further yields that

$$\|y_j\| \geq m_2\|s_j\|, \tag{3.24}$$

by the Cauchy inequality. Thus, the inequality $m_2\|s_j\| \leq \|y_j\| \leq m_1\|s_j\|$ implies that, when j is large enough, $\|\alpha_j d_j\|\|g_j\|^\alpha \leq \|y_j\| \leq M'\|\alpha_j d_j\|$ always holds (parameter $M' = O(\|g_j\|^{-\alpha})$), which means that $j \in \tau_j$ is always satisfied. From the above derivation, the CPBFGS method reduces to the ordinary BFGS method, and the superlinear convergence of Algorithm 1 has been established (see, [12, 26, 27]). □

So far, we have completed the convergence analysis. Next, under some assumptions, we will prove a general but stronger convergence result.

Theorem 3.9 Assuming that the same assumptions as in Theorem 3.6 hold, $g(x)$ is uniformly continuous and $\mathcal{M}_1 \prec H_j \prec \mathcal{M}_2$ for all $j > 0$ (this means that $H_j - \mathcal{M}_1 I$ and $\mathcal{M}_2 I - H_j$ are positive definite matrixes), where $H_j = B_j^{-1}$, $\mathcal{M}_1$ and $\mathcal{M}_2 > 0$. Then

$$\lim_{j \to +\infty} \|g_j\| = 0. \tag{3.25}$$

Proof For any given $\epsilon > 0$, according to (3.7), there exist infinitely many iterates x_j such that $\|g(x_j)\| < \epsilon$. We suppose that (3.25) is not true. Then there must be two infinite subsequences, $\{x_{m_k}\}$ and $\{x_{n_k}\}$, with $m_k < n_k$, for $k = 1, 2, \cdots$, which satisfies the following inequalities:

$$\|g(x_{m_k})\| \geq 2\epsilon, \quad \|g(x_{n_k})\| < \epsilon, \quad \|g(x_j)\| \geq \epsilon, \quad j = m_k + 1, \cdots, n_k - 1. \tag{3.26}$$

Then, from (1.4), the lower boundedness of f and $\mathcal{M}_1 \prec H_j \prec \mathcal{M}_2$ for all j, it follows that

$$\begin{aligned} +\infty > -\sum_{j=1}^{+\infty} \alpha_j g_j^T d_j = \sum_{j=1}^{+\infty} \alpha_j g_j^T H_j g_j \geq \sum_{k=1}^{+\infty} \sum_{j=m_k}^{n_k-1} \alpha_j g_j^T H_j g_j \\ \geq \sum_{k=1}^{+\infty} \sum_{j=m_k}^{n_k-1} \mathcal{M}_1 \alpha_j \|g_j\|^2 \geq \mathcal{M}_1 \epsilon^2 \sum_{k=1}^{+\infty} \sum_{j=m_k}^{n_k-1} \alpha_j, \end{aligned} \tag{3.27}$$

which, together with $\alpha_j \geq 0$, yields that

$$\lim_{k \to +\infty} \sum_{j=m_k}^{n_k-1} \alpha_j = 0. \tag{3.28}$$

According to (1.3), we have that

$$\|x_{j+1} - x_j\| = \|a_j H_j g_j\| \leq \alpha_j \|H_j\| \|g_j\| \leq \alpha_j M^*, \tag{3.29}$$

where $M^* = \mathcal{M}_2 G_0$. Then it follows from (3.29) that

$$\|x_{n_k} - x_{m_k}\| \leq \sum_{j=m_k}^{n_k-1} \|x_{j+1} - x_j\| \leq M^* \sum_{j=m_k}^{n_k-1} \alpha_j, \tag{3.30}$$

which, together with (3.28), implies that $\lim\limits_{k \to +\infty} \|x_{n_k} - x_{m_k}\| = 0$. Therefore, combined with the uniform continuity of $g(x)$, it follows that $\lim\limits_{k \to +\infty} \|g(x_{n_k}) - g(x_{m_k})\| = 0$, which contradicts (3.26). □

Using Theorem 3.9 and considering the second part of the proof of Theorem 3.6, when τ_j is finite, we will get a better conclusion: $\lim\limits_{j \to +\infty} \|g_j\| = 0$.

4 Numerical Results

4.1 Test of Unconstrained Optimization Problems

In this subsection, some numerical experiments of Algorithm 1 are reported. We show the total test results for 63 unconstrained optimization problems [30–32] on MATLAB R2019a.

Dimension The dimensions of the tested problems we consider are 300,1200, and 2100.

PC Requirements Windows 10 operating system with an Intel (R) Core (TM) i7-6700HQ CPU at 2.60GHz (8 CPUs), 2.6GHz, 8066MB of RAM.

The tested algorithms CBFGS [27] (with $\alpha = 1$), PBFGS1 [25], PBFGS2 [28], CPBFGS1 (with $\rho = 5$, $\mu = 6$, $M = \|g_j\|^\alpha$ and $\alpha = 0.1$ (if $\|g_j\| > 1$), 1 (else $\|g_j\| \le 1$)) and CPBFGS2 (with $\rho = 13$, $\mu = 15$, $M = 1000$ and $\alpha = 0$).

Parameters Setting $B_0 = I$, $\zeta_1 = 0.2$, $\zeta_2 = 0.85$ and $\epsilon = 10^{-6}$.

Stopping Rule The Himmelblau stop rule is used: if $|f(x_j)| > \epsilon_1$ holds, let $H = \frac{|f(x_j)-f(x_{j+1})|}{|f(x_j)|}$; otherwise, let $H = |f(x_j) - f(x_{j+1})|$. If $\|g(x)\| < \epsilon$ (or $H < \epsilon_2$) is true, or the total number of iteration loops exceeds one thousand, the program terminates, where $\epsilon_1 = \epsilon_2 = 10^{-5}$.

Other Settings Only when the number of WWP line search is guaranteed to be greater than ten will the step size α_j be selected. The notations used are ni: the total number of algorithm iterations; nfg: the total number of function and gradient computations; cpu time: the CPU time of the iterative process in seconds.

The performance profiles of tested algorithms are analyzed by the tool provided by Dolan and Morè [33], and the Figures 1–3 show the performance results regarding ni, nfg and cpu time respectively. The performance profile was first used to compare the efficiency in [39]. According to Figures 1–3, there is no difficulty in concluding that CPBFGS1 exhibits the best performance among the five algorithms for nfg and cpu time, while CPBFGS2 performs best for ni. In other words, the algorithms with a projection technique have better performance and robustness for ni and nfg but get lost for cpu time when compared with those without a projection technique. It can be inferred from all of the above that CPBFGS overcomes the shortcomings of the projection algorithms and CBFGS, and that it is competitive and promising.

In addition, to show the rationality of the proposed algorithm, we also report in Table 1 the percentage of projection iterations (2.4) in the total iterations and the number of B_j updates as a percentage of the total iterations.

Table 1 Projection iteration percentage and B_j update iteration percentage

Algorithm	Projection iteration percentage	B_j update iteration percentage
CBFGS1	9.8%	2.3%
CBFGS2	19.3%	69.1%

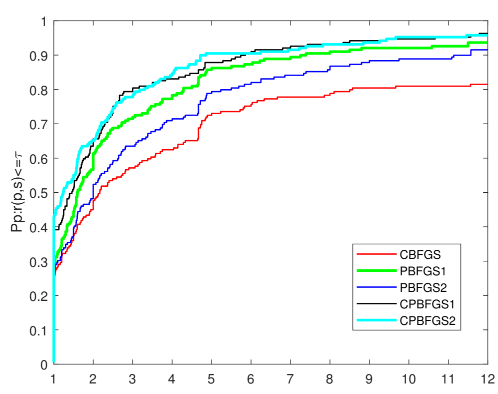

Figure 1 Performance profiles of the algorithms for ni

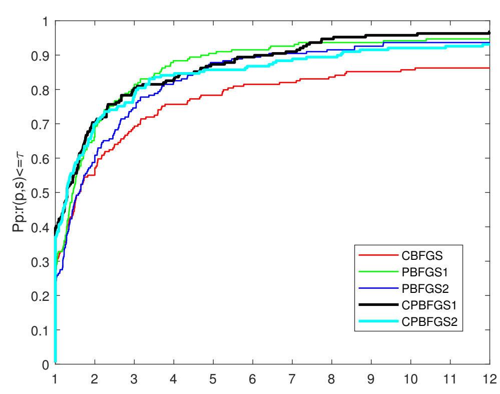

Figure 2 Performance profiles of the algorithms for nfg

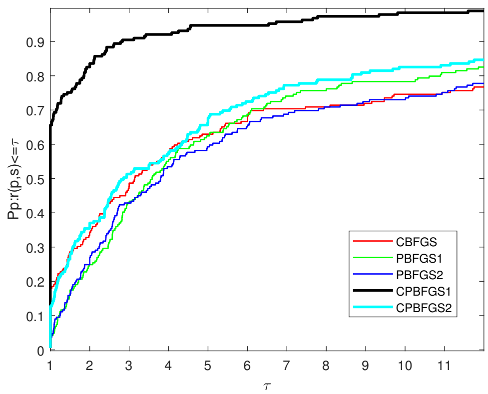

Figure 3 Performance profiles of the algorithms for cpu time

5 Conclusions

In this paper, a cautious projection BFGS method has been proposed for solving nonconvex unconstrained problems. We showed that CPBFGS with a WWP line search can converge globally without a gradient Lipschitz continuity assumption, and that it also has a superlinear convergence rate; this is a significant extension of projection-class methods and modified BFGS methods. In addition, a stronger convergence result has been given. This method not only has superior theoretical properties, but also shows good competitiveness and promising behavior in terms of numerical experiments.

Conflict of Interest The authors declare no conflict of interest.

References

[1] Fu Z J, Ren K, Shu J G, et al. Enabling personalized search over encrypted outsourced data with efficiency improvement. J IEEE Trans Parall Distr Syst, 2016, **27**(9): 2546–2559

[2] Fu Z J, Wu X L, Guan C W, et al. Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement. J IEEE Trans Inf Foren Sec, 2017, **11**(12): 2706–2716

[3] Gu B, Sun X M, Sheng V S. Structural minimax probability machine. J IEEE Trans Neur Net Lear, 2017, **28**(7): 1646–1656

[4] Pan Z Q, Lei J J, Zhang Y, et al. Fast motion estimation based on content property for low-complexity H.265/HEVC encoder. J IEEE Trans Broadcast, 2016, **62**(3): 675–684

[5] Pan Z Q, Zhang Y, Kwong S. Efficient motion and disparity estimation optimization for low complexity multiview video coding. J IEEE Trans Broadcast, 2015, **61**(2): 166–176

[6] Broyden C G. The convergence of a class of double-rank minimization algorithms 1. general considerations. SIMA J Appl Math, 1970, **6**(1): 76–90

[7] Fletcher R. A new approach to variable metric algorithms. Comput J, 1970, **13**(3): 317–322

[8] Goldfarb D. A family of variable-metric methods derived by variational means. Math Comp, 1970, **24**(109): 23–26

[9] Shanno D F. Conditioning of quasi-Newton methods for function minimization. Math Comp, 1970, **24**(111): 647–656

[10] Powell M J D. Some global convergence properties of a variable metric algorithm for minimization without exact line searches. Nonlinear Programming, SIAM-AMS Proceedings. 1976, **9**(1): 53–72

[11] Byrd R H, Nocedal J, Yuan Y X. Global convergence of a cass of quasi-Newton methods on convex problems. SIAM J Numer Anal, 1987, **24**(5): 1171–1190

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

[13] Dixon L C W. Variable metric algorithms: Necessary and sufficient conditions for identical behavior of nonquadratic functions. J Optim Theory Appl, 1972, **10**(1): 34–40

[14] Griewank A. The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients. Math Program, 1991, **50**(1): 141–175

[15] Powell M J D. On the convergence of the variable metric algorithm. IMA J Appl Math, 1971, **7**(1): 21–36

[16] Powell M J D. Updating conjugate directions by the BFGS formula. Math Program, 1987, **38**(1): 29–46

[17] Biggs M C. Minimization algorithms making use of non-quadratic properties of the objective function. IMA J Numer Anal, 1971, **8**(3): 123–123

[18] Nocedal J, Yuan Y X. Analysis of a self-scaling quasi-Newton method. Math Program, 1993, **61**: 19–37

[19] Oren S S, Luenberger D G. Self-scaling variable metric (ssvm) algorithms: Part i: Criteria and sufficient conditions for scaling a class of algorithms. Manag Sci, 1974, **20**(5): 845–862

[20] Yuan Y X. A modified bfgs algorithm for unconstrained optimization. IMA J Numer Anal, 1991, **3**: 325–332

[21] Cheng W Y, Li D H. Spectral scaling BFGS method. J Optim Theory Appl, 2010, **146**(2): 305–319

[22] Sheng Z, Yuan G L, Cui Z R. A new adaptive trust region algorithm for optimization problems. Acta Math Sci, 2018, **38B**(2): 479–496

[23] Yuan G L, Wei Z X. The superlinear convergence analysis of a nonmonotone BFGS algorithm on convex objective functions. Acta Math Sci, 2008, **24B**: 35–42

[24] Dai Y H. Convergence properties of the BFGS algoritm. SIAM J Optim, 2002, **13**(3): 693–701

[25] Yuan G L, Sheng Z, Wang B P, et al. The global convergence of a modified BFGS method for nonconvex functions. J Comput Appl Math, 2017, **327**: 274–294

[26] Li D H, Fukushima M. A modified BFGS method and its global convergence in nonconvex minimization. J Comput Appl Math, 2001, **129**(1/2): 15–35

[27] Li D H, Fukushima M. On the global convergence of BFGS method for nonconvex unconstrained optimization problems. SIAM J Optim, 2001, **11**(4): 1054–1064

[28] Yuan G L, Wang X L, Sheng Z. The projection technique for two open problems of unconstrained optimization problems. J Optim Theory Appl, 2020, **186**(2): 590–619

[29] Yuan G L, Wei Z X, Lu X W. Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search. Appl Math Model, 2017, **47**: 811–825

[30] Bongartz I , Conn A R, Gould N, et al. CUTE: Constrained and unconstrained testing environment. ACM Trans Math Software, 1995, **21**(1): 123–160

[31] Moré J J, Garbow B S, Hillstrom K E. Testing unconstrained optimization software. ACM Trans Math Software, 1981, **7**(1): 17–41

[32] Andrei N. An unconstrained optimization test functions collection. Adv Model Optim, 2008, **10**(1): 147–161

[33] Dolan E D, Moré, Jorge J. Benchmarking optimization software with performance profiles. Math Program, 2002, **91**(2): 201–213

[34] Zhou W J, Zhang L. A nonlinear conjugate gradient method based on the MBFGS secant condition. Optim Methods Softw, 2006, **21**(5): 707–714

[35] Yang Y G. A robust BFGS algorithm for unconstrained nonlinear optimization problems. Optim, 2022, **73**(3): 851–873

[36] Yuan G L, Zhao X, Liu K Jun, et al. An adaptive projection BFGS method for nonconvex unconstrained optimization problems. Numer Algorithms, 2024, **95**(4): 1747–1767

[37] Yuan G L, Li P Y, Lu J Y. The global convergence of the BFGS method with a modified WWP line search for nonconvex functions. Numer Algorithms, 2022, **91**(1): 353–365

[38] Yuan G L, Zhang M X, Zhou Y J. Adaptive scaling damped BFGS method without gradient Lipschitz continuity. Appl Math Lett, 2022, **124**: Art 107634

[39] Tits A L, Yang Y. Globally convergent algorithms for robust pole assignment by state feedback. IEEE Trans Automat Contr, 1996, **41**(10): 1432–1452