Hindawi
Journal of Mathematics
Volume 2022, Article ID 8939770, 15 pages
https://doi.org/10.1155/2022/8939770

Research Article

A New Spectral Three-Term Conjugate Gradient Method with Random Parameter Based on Modified Secant Equation and Its Application to Low-Carbon Supply Chain Optimization

Guoling Zhou,[1] Yueting Yang ,[1,2] and Mingyuan Cao [1]

[1]*School of Mathematics and Statistics, Beihua University, Jilin 132013, China*
[2]*School of Science, Jiangsu University, Zhenjiang, Jiangsu 212013, China*

Correspondence should be addressed to Yueting Yang; yyt2858@163.com and Mingyuan Cao; cmy0918@beihua.edu.cn

Received 13 August 2022; Revised 7 September 2022; Accepted 17 September 2022; Published 15 October 2022

Academic Editor: Shaojian Qu

In this work, considering the advantages of spectral conjugate gradient method and quasi-Newton method, a spectral three-term conjugate gradient method with random parameter is proposed. The parameter in the search direction of the new method is determined by minimizing the Frobenius norm of difference between search direction matrix and self-scaled memoryless BFGS matrix based on modified secant equation. Then, the search direction satisfying the sufficient descent condition is obtained. The global convergence of new method is proved under appropriate assumptions. Numerical experiments show that our method has better performance by comparing with the up-to-date method. Furthermore, the new method has been successfully applied to the optimization of low-carbon supply chain.

1. Introduction

Consider the following unconstrained optimization problem

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

where $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ is continuous differentiable and bounded from below.

Spectral conjugate gradient (SCG) method is one of the most effective methods for solving (1). It has some advantages, such as simple iterative scheme, low memory requirement, and strong global convergence, as well as the traditional conjugate gradient (CG) method [1], and outperforms the traditional CG method in numerical performance. SCG method generates a sequence of solutions $\{x_k\}$ with the following formula:

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

in which $\alpha_k > 0$ is the stepsize, and the search direction d_k is defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -\theta_k g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{3}$$

where $g_k = \nabla f(x_k)$ is the gradient of $f(x)$ at iterate point x_k, θ_k is the spectral parameter, and β_k is the conjugate parameter. The choices of θ_k and β_k are crucial for the global convergence and numerical performance of the algorithm, which have been widely studied by many scholars (see [2–9]).

Deng et al. [8] proposed an improved spectral conjugate gradient (ISCG) method for nonconvex unconstrained optimization, where the parameters θ_k and β_k in (3) are determined by

$$\theta_k = \begin{cases} \dfrac{d_{k-1}^T\left(y_{k-1} - \left(g_k g_k^T s_{k-1}/\|g_k\|^2\right)\right)}{d_{k-1}^T \bar{y}_{k-1}}, & \text{if } d_{k-1}^T \bar{y}_{k-1} > \eta \|g_{k-1}\|^2, \\ \dfrac{d_{k-1}^T\left(y_{k-1} - \left(g_k g_k^T/\|g_k\|^2\right) g_{k-1}\right)}{-d_{k-1}^T g_{k-1}}, & \text{otherwise,} \end{cases}$$

$$\beta_k = \begin{cases} \dfrac{g_k^{\mathrm{T}}(y_{k-1} - s_{k-1})}{d_{k-1}^{\mathrm{T}}\bar{y}_{k-1}}, & \text{if } d_{k-1}^{\mathrm{T}}\bar{y}_{k-1} > \eta\|g_{k-1}\|^2, \\ \dfrac{g_k^{\mathrm{T}} y_{k-1}}{\|g_{k-1}\|^2}, & \text{otherwise}, \end{cases} \tag{4}$$

where $y_{k-1} = g_k - g_{k-1}, s_{k-1} = x_k - x_{k-1}$, $\bar{y}_{k-1} = (I - (g_k g_k^{\mathrm{T}}/\|g_k\|^2))y_{k-1}$, and η is a small constant. The obtained search direction d_k satisfies the sufficient descent and approaches the quasi-Newton direction. Numerical experiments showed that ISCG algorithm was effective for solving large-scale problems.

Li et al. [9] proposed a spectral three-term conjugate gradient method on three-dimensional subspace $\Omega_k = \operatorname{span}\{g_k, d_{k-1}, y_{k-1}\}$. The search direction d_k is expressed as

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -\theta_k g_k + \beta_k^{\mathrm{MPRP}} d_{k-1} + \nu_k y_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{5}$$

where $\beta_k^{MPRP} = g_k^{\mathrm{T}} y_{k-1}/(\mu|g_k^T d_{k-1}| + \|g_{k-1}\|^2)$ and $\nu_k = g_k^{\mathrm{T}} d_{k-1}/(\mu|g_k^{\mathrm{T}} d_{k-1}| + \|g_{k-1}\|^2)$ $(\mu \geq 0)$ are given by [10]. They made d_k close to general quasi-Newton direction and obtained the expression for θ_k as

$$\begin{aligned} \theta_k^{\mathrm{JCJ}} &= \begin{cases} \theta_k^{\mathrm{JCJ+}}, & \text{if } \theta_k^{\mathrm{JCJ+}} \in [\tau_1, \tau_2], \\ 1, & \text{otherwise}, \end{cases} \\ \text{or } \theta_k^{\mathrm{JCJ}} &= \begin{cases} \theta_k^{\mathrm{JCJ-}}, & \text{if } \theta_k^{\mathrm{JCJ-}} \in [\tau_1, \tau_2], \\ 1, & \text{otherwise}, \end{cases} \end{aligned} \tag{6}$$

in which $0 < \tau_1 < \tau_2$, and

$$\begin{aligned} \theta_k^{\mathrm{JCJ+}} &= \frac{1}{\mu|g_k^{\mathrm{T}} d_{k-1}| + \|g_{k-1}\|^2}\left(y_{k-1}^{\mathrm{T}} d_{k-1} - \frac{g_k^{\mathrm{T}} d_{k-1}\|y_{k-1}\|^2}{y_{k-1}^T g_k}\right) \\ &\quad + \frac{s_{k-1}^{\mathrm{T}} g_k}{y_{k-1}^{\mathrm{T}} g_k}, \\ \theta_k^{\mathrm{JCJ-}} &= \frac{1}{\mu|g_k^{\mathrm{T}} d_{k-1}| + \|g_{k-1}\|^2}\left(y_{k-1}^{\mathrm{T}} d_{k-1} - \frac{g_k^{\mathrm{T}} d_{k-1}\|y_{k-1}\|^2}{y_{k-1}^{\mathrm{T}} g_k}\right). \end{aligned} \tag{7}$$

In addition, they used modified secant equation [11]

$$B_k s_{k-1} = z_{k-1}, \tag{8}$$

where $z_{k-1} = y_{k-1} + \rho_{k-1}(\max\{\tau_{k-1}, 0\}/s_{k-1}^{\mathrm{T}} u_{k-1})u_{k-1}$, $\tau_{k-1} = 6(f_{k-1} - f_{k-1}) + 3(g_k + g_{k-1})^{\mathrm{T}} s_{k-1}$, $u_{k-1} = (1 - \lambda_k)y_{k-1} + \lambda_k s_{k-1}$, and $\rho_{k-1} \in \{0, 1\}$. If $\|s_k\| \leq 1$, then $\rho_{k-1} = 1$; otherwise, $\rho_{k-1} = 0$. Another way to choose spectral parameters was proposed, in which $\theta_k^{\mathrm{JCJ+}}$ and $\theta_k^{\mathrm{JCJ-}}$ are obtained by replacing y_{k-1} with z_{k-1} in (7). Their methods had global convergence and were superior to the three-term conjugate gradient method proposed by Sun and Liu [10].

Neculai [6] proposed a new scaled conjugate gradient (SCALCG) algorithm by using a hybridization of the memoryless Broyden-Fletcher-Goldfarb-Shanno (MBFGS) preconditioned CG method [12] and SCG method [13] for solving large-scale unconstrained optimization. The search direction d_{k+1} is defined by

$$\begin{aligned} d_{k+1} &= -D_{k+1} g_{k+1} \\ &= -\theta_{k+1} g_{k+1} + \theta_{k+1}\left(\frac{g_{k+1}^T s_k}{y_k^T s_k}\right) y_k \\ &\quad - \left[\left(1 + \theta_{k+1}\frac{y_k^T y_k}{y_k^T s_k}\right)\frac{g_{k+1}^T s_k}{y_k^T s_k} - \theta_{k+1}\frac{g_{k+1}^T y_k}{y_k^T s_k}\right] s_k, \end{aligned} \tag{9}$$

where D_{k+1} is called search direction matrix and $\theta_{k+1} = s_k^{\mathrm{T}} s_k/y_k^{\mathrm{T}} s_k$ is determined according to a two-point approximation of the standard secant equation. Numerical experiments showed that the SCALCG algorithm outperformed several well-known CG algorithms [13–15].

Babaie-Kafaki and Ghanbari [16] rewrote the search direction of Dai and Liao method [17] as

$$d_{k+1} = -D_{k+1} g_{k+1} = -\left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + t_k \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) g_{k+1}. \tag{10}$$

They obtained the following relation

$$d_{k+1}^{\mathrm{T}} g_{k+1} = -g_{k+1}^{\mathrm{T}} D_{k+1} g_{k+1} = -g_{k+1}^{\mathrm{T}} A_{k+1} g_{k+1}, \tag{11}$$

in which

$$A_{k+1} \triangleq \frac{D_{k+1}^{\mathrm{T}} + D_{k+1}}{2} = I - \frac{1}{2}\frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + t_k \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}, \tag{12}$$

and analyzed the eigenvalues of the matrix A_{k+1} to determine the parameter t_k.

Yao and Ning [18] proposed a three-term conjugate gradient method, in which the search direction was expressed as

$$\begin{aligned} d_{k+1} &= -D_{k+1} g_{k+1} \\ &= -\left(I - t_k \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + \frac{s_k s_k^T}{s_k^T y_k}\right) g_{k+1} \\ &= -g_{k+1} + \beta_k^+ d_k + \delta_k y_k, \end{aligned} \tag{13}$$

where the optimal parameter t_k was derived by minimizing the distance between D_{k+1} and the self-scaled memoryless BFGS (ML-BFGS) matrix in the Frobenius norm, that is,

$$t_k = \min\left\{\frac{1}{1 + \left(\|s_k\|^2\|y_k\|^2/(s_k^T y_k)^2\right)}, \frac{s_k^T y_k}{\|y_k\|^2}\right\}, \tag{14}$$

and the parameters $\beta_k^+ = \max\{(t_k g_{k+1}^T y_k - g_{k+1}^T s_k)/d_k^T y_k, 0\}$, $\delta_k = t_k g_{k+1}^T s_k/s_k^T y_k$. The search direction d_{k+1} was always sufficiently descent at every iteration independent of any line search strategy, and this method had global convergence for general nonconvex functions.

Based on the above work, it is shown that spectral parameter plays an important role in improving the conjugate gradient method, and modified secant equation uses more information of function value and gradient value. Therefore, in order to obtain a new algorithm with good numerical performance, especially for the objective function with sharp curvature change, we introduce the spectral parameter into (12) and construct the following search direction matrix

$$Q_{k+1} = \theta_{k+1} I - \frac{1}{2} \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + t_k \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}, \tag{15}$$

and improve ML-BFGS matrix based on the modified secant equation. The parameter t_k in (15) is determined by minimizing the Frobenius norm of difference between Q_{k+1} and ML-BFGS matrix based on modified secant equation, and we propose a spectral three-term conjugate gradient method with random parameter. The contributions of this article are listed as follows:

(i) A random parameter is introduced to simplify the format of the parameter t_k in the search direction, and the search direction satisfying the sufficient descent condition is obtained

(ii) Under appropriate assumptions, global convergence of new method for general functions is given

(iii) The new method has good numerical performance for the objective function with sharp curvature change

(iv) The new method is applied to the low-carbon supply chain optimization model, which shows that the new method is effective

The rest of this paper is organized as follows: in the next section, a new random parameter is given to present spectral three-term conjugate gradient method. In Section 3, global convergence of the new method for uniformly convex functions and general functions is proved under appropriate conditions. In Section 4, some numerical experiments are implemented. In Section 5, the application of new method in low-carbon supply chain optimization is studied. Conclusions are made in the last section.

2. A Spectral Three-Term Conjugate Gradient Method with Random Parameter

In this section, our main aim is to propose a new spectral three-term conjugate gradient method based on modified secant equation. Consider the following modified secant equation:

$$B_{k+1} s_k = z_k, \tag{16}$$

where $z_k = y_k + (\max\{\tau_k, 0\}/s_k^{\mathrm{T}} u_k) u_k$, $\tau_k = 6(f_{k+1} - f_k) + 3(g_{k+1} + g_k)^{\mathrm{T}} s_k$, and $\mu_k = y_k$; we design ML-BFGS matrix based on modified secant equation as follows:

$$B_{k+1}^{-1} = \theta_{k+1} I - \theta_{k+1} \frac{s_k y_k^{\mathrm{T}} + y_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \left(\frac{1}{\rho_k^{+}} + \theta_{k+1} \frac{\|y_k\|^2}{s_k^{\mathrm{T}} y_k} \right) \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}, \tag{17}$$

where $\rho_k^{+} = 1 + (\max\{\tau_k, 0\}/s_k^{T} y_k)$.

The parameter t_k is determined by minimizing the Frobenius norm of difference between search direction matrix and ML-BFGS matrix based on modified secant equation, that is,

$$\min \left\| Q_{k+1} - B_{k+1}^{-1} \right\|_F^2, \tag{18}$$

where $\|\cdot\|_F$ is the Frobenius norm and Q_{k+1} and B_{k+1}^{-1} are determined by (15) and (17), respectively.

From (15) and (17), we have

$$\begin{aligned} \left\| Q_{k+1} - B_{k+1}^{-1} \right\|_F^2 &= \operatorname{tr}\left(\left(Q_{k+1} - B_{k+1}^{-1} \right)^{\mathrm{T}} \left(Q_{k+1} - B_{k+1}^{-1} \right) \right) \\ &= \frac{\|s_k\|^4}{\left(s_k^{\mathrm{T}} y_k\right)^2} t^2 + 2 \left[\left(2\theta_{k+1} - 1\right) \frac{\|s_k\|^2}{s_k^{\mathrm{T}} y_k} \right. \\ &\quad \left. - \frac{1}{\rho_k^{+}} \frac{\|s_k\|^4}{\left(s_k^{\mathrm{T}} y_k\right)^2} - \theta_{k+1} \frac{\|y_k\|^2 \|s_k\|^4}{\left(s_k^{\mathrm{T}} y_k\right)^3} \right] t + \xi, \end{aligned} \tag{19}$$

where ξ is a constant independent of t. Therefore, the minimum of problem (18) is

$$\begin{aligned} t_k &= \arg\min \left\{ \operatorname{tr}\left(\left(Q_{k+1} - B_{k+1}^{-1} \right)^{\mathrm{T}} \left(Q_{k+1} - B_{k+1}^{-1} \right) \right) \right\} \\ &= \frac{1}{\rho_k^{+}} + \frac{\|y_k\|^2 \left(\theta_{k+1} - \left(2\theta_{k+1} - 1\right) m_k\right)}{s_k^{\mathrm{T}} y_k}, \end{aligned} \tag{20}$$

in which $m_k = \cos^2 \eta_k$, $\eta_k = \langle s_k, y_k \rangle$ is the angle between s_k and y_k. Instead of the mean value to $\cos^2 \eta_k = 1/2$ in [19], let m_k be a random number in the interval $[\underline{c}, \bar{c}]$, where $0 < \underline{c} < \bar{c} < 1/2$. Therefore, t_k in (20) can be regarded as a random parameter. There are many possible ways to choose θ_{k+1}; we set

$$\begin{aligned} \theta_{k+1} &= \max\left\{ 1, \frac{\|s_k\|^2}{s_k^{\mathrm{T}} y_k} \right\} \\ \text{or } \theta_{k+1} &= \max\left\{ 1, \frac{s_k^{\mathrm{T}} y_k}{\|y_k\|^2} \right\}. \end{aligned} \tag{21}$$

Substitute (20) into (15), and let $d_{k+1} = -Q_{k+1}g_{k+1}$; then,

$$\begin{aligned} d_{k+1} &= -\left(\theta_{k+1}I - \frac{1}{2}\frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + t\frac{s_k s_k^T}{s_k^T y_k}\right)g_{k+1} \\ &\triangleq -\theta_{k+1}g_{k+1} + a_k s_k + b_k y_k, \end{aligned} \tag{22}$$

where

$$\begin{aligned} a_k &= \frac{1}{2}\frac{y_k^T g_{k+1}}{s_k^T y_k} - \left[\frac{1}{\rho_k^+} + \frac{\|y_k\|^2(\theta_{k+1} - (2\theta_{k+1} - 1)m_k)}{s_k^T y_k}\right]\frac{s_k^T g_{k+1}}{s_k^T y_k}, \\ b_k &= \frac{1}{2}\frac{s_k^T g_{k+1}}{s_k^T y_k}. \end{aligned} \tag{23}$$

Based on the above analysis, a new spectral three-term conjugate gradient (STCG) algorithm can be presented as follows.

Algorithm 1. STCG algorithm.

Step 0. Given $x_0 \in \mathbb{R}^n$, $\varepsilon > 0$, $0 < \underline{c} < \overline{c} < 1/2$ and $0 < \omega < \sigma < 1$. Compute $f_0 = f(x_0)$ and $g_0 = \nabla f(x_0)$; let $d_0 := -g_0$ and $k := 0$.

Step 1. If $\|g_k\| \le \varepsilon$, stop; else, go to step 2.

Step 2. Compute a step length α_k satisfying strong Wolfe line search conditions

$$f(x_k + \alpha d_k) - f(x_k) \le \omega\alpha g_k^T d_k, \tag{24}$$

$$\left|g^T(x_k + \alpha d_k)d_k\right| \le -\sigma g_k^T d_k. \tag{25}$$

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$; calculate f_{k+1}, g_{k+1}, s_k, y_k, and ρ_k^+.

Step 4. Compute t_k by (20) and (21) and search direction d_{k+1} by (22). Set $k := k + 1$ and go to step 1.

The following lemma shows that the search direction satisfies the sufficient descent property, which plays an important role in proving the convergence of the algorithm.

Lemma 1. *Let the sequence $\{d_{k+1}\}$ be generated by STCG algorithm; then, there exists a positive constant c, such that*

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2. \tag{26}$$

Proof. From the search direction (22), we have

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\theta_{k+1}\|g_{k+1}\|^2 + \frac{y_k^T g_{k+1} g_{k+1}^T s_k s_k^T y_k}{(s_k^T y_k)^2} \\ &\quad - \left[\frac{1}{\rho_k^+} + \frac{\|y_k\|^2(\theta_{k+1} - (2\theta_{k+1} - 1)m_k)}{s_k^T y_k}\right]\frac{(s_k^T g_{k+1})^2}{s_k^T y_k} \\ &\le -\theta_{k+1}\|g_{k+1}\|^2 + \frac{1}{2}\frac{(g_{k+1}^T s_k)^2\|y_k\|^2 + (s_k^T y_k)^2\|g_{k+1}\|^2}{(s_k^T y_k)^2} \\ &\quad - \left[\frac{1}{\rho_k^+} + \frac{\|y_k\|^2(\theta_{k+1} - (2\theta_{k+1} - 1)m_k)}{s_k^T y_k}\right]\frac{(s_k^T g_{k+1})^2}{s_k^T y_k} \\ &= \left(\frac{1}{2} - \theta_{k+1}\right)\|g_{k+1}\|^2 - \frac{(s_k^T g_{k+1})^2}{s_k^T y_k} \\ &\quad \cdot \left[\frac{1}{\rho_k^+} + \frac{\|y_k\|^2}{s_k^T y_k}\left((\theta_{k+1} - (2\theta_{k+1} - 1)m_k) - \frac{1}{2}\right)\right] \\ &\le \left(\frac{1}{2} - \theta_{k+1}\right)\|g_{k+1}\|^2 - \frac{(s_k^T g_{k+1})^2}{s_k^T y_k} \\ &\quad \cdot \left[\frac{1}{\rho_k^+} + \frac{\|y_k\|^2}{s_k^T y_k}\left((\theta_{k+1} - (2\theta_{k+1} - 1)\overline{c}) - \frac{1}{2}\right)\right] \\ &\le -\frac{1}{2}\|g_{k+1}\|^2. \end{aligned} \tag{27}$$

The second of the above inequalities comes from the fact$u^T v \le 1/2(\|u\|^2 + \|v\|^2)$, in which $u = g_{k+1}^T s_k y_k$ and $v = s_k^T y_k g_{k+1}$. In the fourth of the above inequalities, $s_k^T y_k > 0$ can be ensured by the strong Wolfe line search, and $\rho_k^+ = 1 + (\max\{\tau_k, 0\}/s_k^T y_k) > 0$. Combining (21), the proof is completed. □

3. Convergence Analysis

To prove the global convergence of STCG algorithm, we give the following assumptions.

Assumption 2. The level set $\Omega = \{x \in \mathbb{R}^n : f(x) \le f(x_0)\}$ is bounded; namely, there exists a positive constant δ such that $\|x\| \le \delta, \forall x \in \Omega$.

Assumption 3. The gradient of function f is Lipschitz continuous in some neighborhood $\mathbb{N}$ of Ω; namely, there exists $L > 0$ satisfying

$$\|g(x) - g(y)\| \le L\|x - y\|, \forall x, y \in \mathbb{N}. \tag{28}$$

Based on the above assumptions, we can easily have that $g(x)$ is bounded; i.e., there exists a positive constant M such that

$$\|g(x)\| \le M, \forall x \in \Omega. \tag{29}$$

Lemma 4. *If Assumption 3 holds, then τ_k is bounded, i.e.,*

$$|\tau_k| \le 3L\|s_k\|^2. \tag{30}$$

The proof of Lemma 4 is similar to the proof of Lemma 2 in [20], so we omit it here.

According to Lemma 4, we can see $1/\rho_k^+ = s_k^T y_k/(s_k^T y_k + \max\{\tau_k, 0\}) < = 1$.

Lemma 5. *Let the sequence* $\{d_k\}$ *be generated by STCG algorithm. If Assumption 3 holds, then*

$$\alpha_k \geq \frac{(1-\sigma)\left|g_k^{\mathrm{T}} d_k\right|}{L\|d_k\|^2}. \tag{31}$$

Proof. According to (25), we have $g_{k+1}^{\mathrm{T}} d_k \geq \sigma g_k^{\mathrm{T}} d_k$, then both side to subtracte $g_k^{\mathrm{T}} d_k$, and using Lipschitz condition, we get

$$(\sigma-1) g_k^{\mathrm{T}} d_k \leq (g_{k+1}-g_k)^{\mathrm{T}} d_k = y_k^{\mathrm{T}} d_k \leq \|y_k\|\|d_k\| \leq \alpha_k L\|d_k\|^2. \tag{32}$$

Since d_k is a descent direction and $0<\sigma<1$, (31) follows immediately. □

Lemma 6. *Let the sequence* $\{d_k\}$ *be generated by STCG algorithm. If Assumption 3 holds, we have*

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty. \tag{33}$$

Proof. From the first inequality (24) of strong Wolfe conditions, Assumption 3, and Lemma 5, we have

$$f_k - f_{k+1} \geq -\omega \alpha_k g_k^{\mathrm{T}} d_k \geq -\rho \frac{(1-\sigma)\left(g_k^{\mathrm{T}} d_k\right)^2}{L\|d_k\|^2}. \tag{34}$$

Since $f(x)$ is bounded from below, the proof is completed. □

Theorem 7. *Suppose that Assumption 2 and Assumption 3 hold. The sequence* $\{x_k\}$ *is generated by STCG algorithm. If* f *is a uniformly convex function on* Ω*, namely, there exists a positive constant* μ *such that*

$$(\nabla f(x)-\nabla f(y))^T (x-y) \geq \mu\|x-y\|^2, \forall x, y \in \mathbb{N}, \tag{35}$$

then we have

$$\lim_{k\longrightarrow\infty} \|g_k\| = 0. \tag{36}$$

Proof. From the Lipschitz condition (28), we have

$$\|y_k\| = \|g_{k+1}-g_k\| \leq L\|s_k\|. \tag{37}$$

It follows (35) that

$$y_k^{\mathrm{T}} s_k \geq \mu\|s_k\|^2. \tag{38}$$

Using Cauchy inequality and (38), we obtain $\mu\|s_k\|^2 \leq y_k^{\mathrm{T}} s_k \leq \|y_k\|\|s_k\|$, i.e.,

$$\mu\|s_k\| \leq \|y_k\|. \tag{39}$$

Then, from (37), (38), and (39), we have

$$\frac{1}{L} = \frac{\|s_k\|^2}{L\|s_k\|^2} \leq \frac{\|s_k\|^2}{\|s_k\|\|y_k\|} \leq \frac{\|s_k\|^2}{s_k^{\mathrm{T}} y_k} \leq \frac{\|s_k\|^2}{\mu\|s_k\|^2} \leq \frac{1}{\mu}, \tag{40}$$

$$\frac{\mu}{L^2} \leq \frac{\mu\|s_k\|^2}{L^2\|s_k\|^2} \leq \frac{\mu\|s_k\|^2}{\|y_k\|^2} \leq \frac{s_k^{\mathrm{T}} y_k}{\|y_k\|^2} \leq \frac{\|s_k\|\|y_k\|}{\|y_k\|^2} \leq \frac{1}{\mu}. \tag{41}$$

Let $\theta_{\max} = \max\{1, 1/\mu\}$; we get $\theta_{k+1} \leq \theta_{\max}$. From (40), we obtain

$$\begin{aligned} t_k &= \frac{1}{\rho_k^+} + \frac{\|y_k\|^2(\theta_{k+1}-(2\theta_{k+1}-1)m_k)}{s_k^{\mathrm{T}} y_k} \\ &\leq 1 + \frac{L^2}{\mu}(\theta_{\max}-(2\theta_{\max}-1)\underline{c}). \end{aligned} \tag{42}$$

Therefore, from (22), (37), (38), and (42), we have

$$\begin{aligned} \|d_{k+1}\| &= \|-\theta_{k+1} g_{k+1} + a_k s_k + b_k y_k\| \\ &\leq \theta_{\max}\|g_{k+1}\| + \frac{1}{2}\left|\frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k}\right|\|y_k\| \\ &\quad + \frac{1}{2}\left|\frac{y_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k}\right|\|s_k\| + |t_k|\left|\frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k}\right|\|s_k\| \\ &\leq \theta_{\max}\|g_{k+1}\| + \frac{1}{2}\frac{\|g_{k+1}\|\|y_k\|}{\mu\|s_k\|} + \frac{1}{2}\frac{\|g_{k+1}\|\|y_k\|}{\mu\|s_k\|} \\ &\quad + \left(1 + \frac{L^2}{\mu}(\theta_{\max}-(2\theta_{\max}-1)\underline{c})\right)\frac{\|g_{k+1}\|}{\mu} \\ &\leq \left(1 + \frac{L+1}{\mu} + (\theta_{\max}-(2\theta_{\max}-1)\underline{c})\frac{L^2}{\mu^2}\right)\|g_{k+1}\| \\ &\triangleq M_1\|g_{k+1}\|. \end{aligned} \tag{43}$$

From Lemma 1 and (43), we get

$$\frac{\left(g_{k+1}^{\mathrm{T}} d_{k+1}\right)^2}{\|d_{k+1}\|^2} \geq \frac{c^2\|g_{k+1}\|^2}{M_1^2}. \tag{44}$$

Combined with Lemma 6, then

$$\sum_{k=0}^{\infty} \|g_k\|^2 < \infty. \tag{45}$$

The proof is completed. □

For general nonlinear functions, we can establish a weaker convergence result:

$$\liminf_{k\longrightarrow\infty} \|g_k\| = 0. \tag{46}$$

Lemma 8. *Suppose that Assumption 2 and Assumption 3 hold. Let the sequence* $\{x_k\}$ *be generated by STCG algorithm;*

then, we have $d_k \neq 0$ *and*

$$\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 < \infty, \quad (47)$$

whenever $\inf \{\|g_k\|: k \geq 0\} > 0$, *in which* $u_k = d_k / \|d_k\|$.

Proof. Define $\gamma = \inf \{\|g_k\|: k \geq 0\}$, then $\|g_k\| \geq \gamma > 0$. From the sufficient descent condition (26), we know $d_k \neq 0$ for each k, so u_k is well defined. To prove global convergence, we define $a_k^+ = \max \{a_k', 0\}$, where $a_k' = (1/2)(y_k^T g_{k+1}/d_k^T y_k) - ((1/\rho_k) + ((\|y_k\|^2 (\theta_{k+1} - (2\theta_{k+1} - 1)m_k))/s_k^T y_k))(s_k^T g_{k+1}/d_k^T y_k)$. By (22), we have

$$\begin{aligned} \frac{d_{k+1}}{\|d_{k+1}\|} &= \frac{-\theta_{k+1} g_{k+1}}{\|d_{k+1}\|} + a_k^+ \frac{d_k}{\|d_{k+1}\|} + b_k \frac{y_k}{\|d_{k+1}\|} \\ &= \frac{-\theta_{k+1} g_{k+1} + b_k y_k}{\|d_{k+1}\|} + a_k^+ \frac{\|d_k\|}{\|d_{k+1}\|} \frac{d_k}{\|d_k\|}, \end{aligned} \quad (48)$$

namely,

$$u_{k+1} = \omega_k + \delta_k u_k, \quad (49)$$

where

$$\begin{aligned} \omega_k &= \frac{-\theta_{k+1} g_{k+1} + b_k y_k}{\|d_{k+1}\|}, \\ \delta_k &= a_k^+ \frac{\|d_k\|}{\|d_{k+1}\|} \geq 0. \end{aligned} \quad (50)$$

Using the identity $\|u_{k+1}\| = \|u_k\| = 1$, we have

$$\|\omega_k\| = \|u_{k+1} - \delta_k u_k\| = \|\delta_k u_{k+1} - u_k\|. \quad (51)$$

Since $\delta_k \geq 0$, then

$$\begin{aligned} \|u_{k+1} - u_k\| &\leq \|(1+\delta_k) u_{k+1} - (1+\delta_k) u_k\| \\ &\leq \|u_{k+1} - \delta_k u_k\| + \|\delta_k u_{k+1} - u_k\| \\ &= 2\|\omega_k\|. \end{aligned} \quad (52)$$

From (25), we have

$$\begin{aligned} \sigma g_k^T d_k - \sigma g_{k+1}^T d_k &\leq g_{k+1}^T d_k - \sigma g_{k+1}^T d_k - \sigma y_k^T d_k \\ &\leq (1-\sigma) g_{k+1}^T d_k \frac{-\sigma}{1-\sigma} \\ &\leq \frac{g_{k+1}^T d_k}{y_k^T d_k}, \\ g_{k+1}^T d_k + \sigma g_{k+1}^T d_k &\leq -\sigma g_k^T d_k + \sigma g_{k+1}^T d_k g_{k+1}^T d_k + \sigma g_{k+1}^T d_k \\ &\leq \sigma y_k^T d_k \frac{g_{k+1}^T d_k}{y_k^T d_k} \\ &\leq \frac{\sigma}{1+\sigma}. \end{aligned} \quad (53)$$

Thus,

$$\left| \frac{s_k^T g_{k+1}}{s_k^T y_k} \right| = \left| \frac{d_k^T g_{k+1}}{d_k^T y_k} \right| \leq \frac{\sigma}{1-\sigma}, \quad (54)$$

$$\|y_k\| \leq \|g_{k+1}\| + \frac{\|g_k\|}{\|g_{k+1}\|} \|g_{k+1}\| \leq 1 + \frac{M}{\gamma} \|g_{k+1}\|. \quad (55)$$

By the definition of ω_k, b_k, (54) and (55), we get

$$\begin{aligned} \|\omega_k\| &= \frac{\|-\theta_{k+1} g_{k+1} + b_k y_k\|}{\|d_{k+1}\|} \\ &\leq \frac{\theta_{k+1} \|g_{k+1}\| + 1/2 |s_k^T g_{k+1}/s_k^T y_k| \cdot \|y_k\|}{\|d_{k+1}\|} \\ &\leq \left[\theta_{\max} + \frac{\sigma}{2(1-\sigma)} \left(1 + \frac{M}{\gamma}\right)\right] \frac{\|g_{k+1}\|}{\|d_{k+1}\|}. \end{aligned} \quad (56)$$

If $\|g_{k+1}\| > \gamma$, from Lemma 1 and Lemma 6, we have

$$\sum_{k=0}^{\infty} \frac{c^2 \gamma^2 \|g_{k+1}\|^2}{\|d_{k+1}\|^2} \leq \sum_{k=0}^{\infty} \frac{c^2 \|g_{k+1}\|^4}{\|d_{k+1}\|^2} \leq \sum_{k=0}^{\infty} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < +\infty. \quad (57)$$

Thus, (47) holds. □

Property(∗). Consider a method of form (2) and (22), and suppose

$$0 < \gamma \leq \|g_k\| \leq \bar{\gamma}, k \geq 0. \quad (58)$$

We call that a method has Property(∗) if there exist constants $b > 1$ and $\lambda > 0$ such that $|a_k'| < b$ and $\|s_k\| \leq \lambda \Rightarrow |a_k'| \leq 1/2b$.

Lemma 9. *Suppose that Assumption 2 and Assumption 3 hold. Let the sequence* $\{d_k\}$ *be generated by STCG algorithm; then, STCG algorithm has Property(∗).*

Proof. By (25) and (26), we obtain

$$d_k^T y_k \geq (\sigma - 1) g_k^T d_k \geq c(1-\sigma) \|g_k\|^2. \quad (59)$$

□

Using (29), (58), Assumption 2, and (59), we obtain

$$\begin{aligned}
\left|a_k'\right| &= \left|\frac{1}{2}\frac{y_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k} - \left(\frac{1}{\rho_k^+} + \frac{\|y_k\|^2(\theta_{k+1} - (2\theta_{k+1} - 1)m_k)}{s_k^{\mathrm{T}} y_k}\right)\frac{s_k^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k}\right| \\
&\le \frac{1}{2}\frac{\|y_k\|\|g_{k+1}\|}{c(1-\sigma)\|g_k\|^2} \\
&\quad + \left(\frac{1}{\rho_k^+} + \frac{\|y_k\|^2(\theta_{k+1} - (2\theta_{k+1} - 1)m_k)}{s_k^{\mathrm{T}} y_k}\right)\frac{\|s_k\|\|g_{k+1}\|}{c(1-\sigma)\|g_k\|^2} \\
&\le \frac{1}{2}\frac{\|g_{k+1} - g_k\|\|g_{k+1}\|}{c(1-\sigma)\|g_k\|^2} \\
&\quad + \left(1 + \frac{\|g_{k+1} - g_k\|^2\theta_{\max}}{c(1-\sigma)\|g_k\|^2}\right)\frac{\|s_k\|\|g_{k+1}\|}{c(1-\sigma)\|g_k\|^2} \\
&\le \frac{\bar{\gamma}^2}{c(1-\sigma)\gamma^2} + \left(1 + \frac{4\bar{\gamma}^2\theta_{\max}}{c(1-\sigma)\gamma^2}\right)\frac{2\delta\bar{\gamma}}{c(1-\sigma)\gamma^2} \\
&\coloneqq b.
\end{aligned} \tag{60}$$

Let

$$\lambda \coloneqq \frac{c^2(1-\sigma)^2\gamma^4}{2\bar{\gamma}^2[\bar{\gamma} + (1 + (4\bar{\gamma}^2\theta_{\max}/c(1-\sigma)\gamma^2))2\delta][(L/2) + (1 + (4\bar{\gamma}^2\theta_{\max}/c(1-\sigma)\gamma^2))]}. \tag{61}$$

If $\|s_k\| \le \lambda$, from (60) and (61), we obtain

$$\begin{aligned}
\left|a_k'\right| &\le \frac{1}{2}\frac{L\|s_k\|\|g_{k+1}\|}{c(1-\sigma)\gamma^2} + \left(\frac{1}{\rho_k^+} + \frac{L^2\|s_k\|^2\theta_{\max}}{c(1-\sigma)\gamma^2}\right)\frac{\|s_k\|\|g_{k+1}\|}{c(1-\sigma)\gamma^2} \\
&\le \left[\frac{1}{2}\frac{L\bar{\gamma}}{c(1-\sigma)\gamma^2} + \left(1 + \frac{4\bar{\gamma}^2\theta_{\max}}{c(1-\sigma)\gamma^2}\right)\frac{\bar{\gamma}}{c(1-\sigma)\gamma^2}\right]\|s_k\| \\
&\le \left[\frac{1}{2}\frac{L\bar{\gamma}}{c(1-\sigma)\gamma^2} + \left(1 + \frac{4\bar{\gamma}^2\theta_{\max}}{c(1-\sigma)\gamma^2}\right)\frac{\bar{\gamma}}{c(1-\sigma)\gamma^2}\right]\lambda \\
&= \frac{1}{2b}.
\end{aligned} \tag{62}$$

In the next lemma, we show that if gradient sequence is bounded away form zero, then a fraction of the steps cannot be too small.. Let $\mathbb{N}$ be the set of positive integers, $K^\lambda \coloneqq \{i \in \mathbb{N} : i \ge 1, \|s_i\| > \lambda\}$, for $\lambda > 0$, namely, the set of integers corresponding to steps greater than λ. Now, we need to discuss groups of Δ consecutive iterates. Let $K_{k,\Delta}^\lambda \coloneqq \{i \in \mathbb{N} : k \le i \le k + \Delta - 1, \|s_i\| > \lambda\}$, and $|K_{k,\Delta}^\lambda|$ denote the number of elements of $K_{k,\Delta}^\lambda$.

Lemma 10. *Suppose that Assumption 2 and Assumption 3 hold. Let the sequences $\{x_k\}$ and $\{d_k\}$ be generated by STCG algorithm. When (58) holds, there exists $\lambda > 0$ such that*

$$\left|K_{k,\Delta}^\lambda\right| > \frac{\Delta}{2}, \textit{for } \Delta \in \mathbb{N}, \tag{63}$$

where $k \ge k_0$, in which k_0 is any index.

Proof. Suppose on the contrary that there exists $\lambda > 0$, such that $|K_{k,\Delta}^\lambda| \le \Delta/2$ for $\Delta \in \mathbb{N}$ and for any $k \ge k_0$.

By (54) and (55), we have

$$\begin{aligned}
\|b_k y_k\| &= \frac{1}{2}\left|\frac{s_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k}\right|\|y_k\| \\
&\le \frac{\sigma}{2(1-\sigma)}\left(1 + \frac{\bar{\gamma}}{\gamma}\right)\|g_{k+1}\| \\
&\triangleq M_2\|g_{k+1}\|.
\end{aligned} \tag{64}$$

According to (22), we have

$$\begin{aligned}
\|d_{k+1}\|^2 &\le \left(a_k'\|d_k\| + \|-\theta_{k+1}g_{k+1} + b_k y_k\|\right)^2 \\
&\le 2a_k'2\|d_k\|^2 + 2\|-\theta_{k+1}g_{k+1} + b_k y_k\|^2 \\
&\le 2a_k'2\|d_k\|^2 + 2\left(2\|\theta_{\max}g_{k+1}\|^2 + 2\|b_k y_k\|^2\right) \\
&\le 2a_k'2\|d_k\|^2 + 4\left(\theta_{\max}^2 + M_2^2\right)\|g_{k+1}\|^2.
\end{aligned} \tag{65}$$

By induction, we have

$$\|d_l\|^2 \le c_1\left(1 + 2a_{l-1}'2 + 2a_{l-1}'22a_{l-2}'2 + \cdots + 2a_{l-1}'22a_{l-2}'2 \cdots 2a_{k_0}'2\right), \tag{66}$$

for any given index $l \ge k_0 + 1$, where c_1 depends on $\|d_{k_0-1}\|$, not on l. Next, we consider $2a_{l-1}'22a_{l-2}'2 \cdots 2a_k'2$, where $k_0 \le k \le l-1$. Now, we divide $2(l-k)$ factors of (66) into groups of each 2Δ elements; namely, if $\Lambda \coloneqq (l-k)/\Delta$, then (66) can be divided into Λ or $\Lambda + 1$ groups

$$\left(2a_{l_1}'2 \cdots 2a_{k_1}'2\right), \cdots, \left(2a_{l_\Lambda}'2 \cdots 2a_{k_\Lambda}'2\right), \tag{67}$$

and a possible group

$$\left(2a_{l_{\Lambda+1}}'2 \cdots 2a_k'2\right), \tag{68}$$

where $l_i = l - 1 - (i-1)\Delta$ for $i = 1, 2, \cdots, \Lambda + 1$, and $k_i = l_{i+1} + 1$ for $i = 1, 2, \cdots, \Lambda$. It is clear that $k_i \ge k_0$ for $i = 1, 2, \cdots, \Lambda$; from assumption condition, we get $p_i \coloneqq |K_{k_i,\Delta}^\lambda| \le \Delta/2$. Thus, there are p_i indices j such that $\|s_j\| > \lambda$ and $(\Delta - p_i)$ indices j such that $\|s_j\| \le \lambda$ on $[k_i, k_i + \Delta - 1]$.

From (60), we have $b > (\bar{\gamma}^2/(c(1-\sigma)\gamma^2)) > 1$, i.e., $2b^2 > 1$. In conjunction with $2p_i - \Delta \le 0$, we have $2a_{l_i}'2 \cdots 2a_{k_i}'2 \le 2^\Delta b^{2p_i}(1/2b)^{2(\Delta - p_i)} = \left(2b^2\right)^{2p_i - \Delta} \le 1$. So every item in (67) is less than or equal to 1, and so is their product. In (68), we have $2a_{l_{\Lambda+1}}'2 \cdots 2a_k'2 \le \left(2b^2\right)^\Delta$. Then, we get

$$\|d_l\|^2 \le c_2(l - k_0 + 2), \tag{69}$$

Table 1: The test functions and dimensions.

Problems	Functions	Dimensions	Problems	Functions	Dimensions
1	The penalty function I	1000	31	Variable dimension	1000
2	The penalty function I	5000	32	Variable dimension	10000
3	The penalty function I	10000	33	Variable dimension	50000
4	The penalty function I	50000	34	Watson	1000
5	The penalty function I	100000	35	Watson	10000
6	Boundary value function	1000	36	Watson	50000
7	Boundary value function	5000	37	Chebyquad	5000
8	Boundary value function	10000	38	Chebyquad	10000
9	Boundary value function	50000	39	Broyden banded	1000
10	Boundary value function	100000	40	Broyden banded	5000
11	Broyden tridiagonal function	1000	41	Broyden banded	10000
12	Broyden tridiagonal function	5000	42	Generalized Rosebrock	1000
13	Broyden tridiagonal function	10000	43	Generalized Rosebrock	5000
14	Broyden tridiagonal function	50000	44	Generalized Rosebrock	10000
15	Separable cubic function	1000	45	Boundary value	1000
16	Separable cubic function	5000	46	Boundary value	5000
17	Separable cubic function	10000	47	Boundary value	10000
18	Separable cubic function	50000	48	Integral equation	1000
19	Separable cubic function	100000	49	Integral equation	5000
20	The variable dimension function	1000	50	Integral equation	10000
21	The Chebyquad function	1000	51	Yang tridiagonal	1000
22	Nearly separable function	1000	52	Yang tridiagonal	5000
23	Nearly separable function	5000	53	Yang tridiagonal	10000
24	Schittkowski function 302	1000	54	Allgower	1000
25	Extended Rosenbrock	1000	55	Freudenstein and Roth	2
26	Extended Rosenbrock	10000	56	Powell badly scaled	2
27	Extended Rosenbrock	50000	57	Brown badly scaled	2
28	Extended Powell singular	1000	58	Beale	2
29	Extended Powell singular	10000	59	Helical valley	3
30	Extended Powell singular	50000	60	Wood	4

where $c_2 > 0$ and independent of l. Furthermore, $\sum_{k\geq 0}(1/\|d_k\|^2) = \infty$. But from (26), (33), and (58), we have

$$c^2\gamma^4\sum_{k\geq 0}\frac{1}{\|d_k\|^2} \leq c^2\sum_{k\geq 0}\frac{\|g_k\|^4}{\|d_k\|^2} \leq \sum_{k\geq 0}\frac{\left(g_k^{\mathrm{T}}d_k\right)^2}{\|d_k\|^2} < \infty. \quad (70)$$

It leads to a contradiction. The proof is completed. □

Theorem 11. *Suppose that Assumption 2 and Assumption 3 hold. Let the sequence $\{x_k\}$ be generated by STCG algorithm; then, (46) holds.*

Proof. The proof by contradiction is adopted. We can obtain the proof similarly to Theorem 4.3 in [21]. □

4. Numerical Results

In this section, we show the numerical performance of STCG algorithm. The test problems are unconstrained problems from CUTEr library [22] and Andrei [23], in which the dimensions vary from 2 to 100000. All codes are written on MATLAB R2015b and run on PC with 1.19 GHz CPU processor, 8.00 GB RAM memory. We list these test problems and their dimensions in Table 1.

We compare STCG algorithm against the descent Dai-Liao (DDL) method [16] and the modified Polak-Ribière-Polyak (PRP^+) method [21], which have better numerical performance. When $\theta_{k+1} = \max\{1, \|s_k\|^2/s_k^{\mathrm{T}}y_k\}$ and $\theta_{k+1} = \max\{1, s_k^{\mathrm{T}}y_k/\|y_k\|^2\}$ are chosen, STCG algorithm is denoted by "New1" and "New2," respectively.

All test methods are terminated when satisfying the following condition:

$$\|g_k\| \leq \varepsilon \text{ or } |f(x_{k+1}) - f(x_k)| \leq \varepsilon \max\{1.0, |f(x_k)|\}, \quad (71)$$

where $\varepsilon = 10^{-6}$ and $\omega = 0.1$ and $\sigma = 0.6$ in Wolfe conditions (24) and (25).

The result of computational experiments from partial problems in Table 1 are listed in Table 2. In Table 2, k, nf,

Table 2: Partial numerical results of several methods.

P.	DDL (k/nf/ng/CPU)	PRP + (k/nf/ng/CPU)	New1 (k/nf/ng/CPU)	New2 (k/nf/ng/CPU)
1	11/30/12/0.0313	13/39/15/0.0625	11/32/12/0.0313	12/34/15/0.0313
2	13/35/14/0.1094	18/50/23/0.1406	9/31/10/0.0313	12/40/16/0.1250
3	14/41/15/0.1406	12/44/17/0.1406	13/41/14/0.1250	12/39/14/0.1563
4	37/85/51/0.9531	16/49/19/0.3594	19/60/27/0.5000	11/44/15/0.2344
5	37/96/50/1.4844	18/55/19/0.8438	12/45/13/0.5469	14/49/15/0.6719
6	24/33/26/0.0469	9/19/11/0.0313	19/27/21/0.0156	17/27/20/0.0781
7	24/33/26/0.1094	9/19/11/0.0313	19/27/21/0.0938	18/27/21/0.1094
8	24/33/26/0.2656	9/19/11/0.0625	19/28/22/0.1250	21/31/24/0.1875
9	24/33/26/0.4219	9/19/11/0.2656	18/26/20/0.4531	20/30/23/0.3281
10	24/33/26/0.8438	9/19/11/0.2594	19/28/22/0.7813	19/26/21/0.7031
11	44/61/49/0.0625	37/67/47/0.0469	31/46/32/0.0313	31/46/32/0.0313
12	31/46/32/0.0938	44/76/53/0.1719	31/46/33/0.0625	31/46/32/0.1875
13	36/53/40/0.2969	18/42/25/0.1250	35/49/37/0.2813	34/46/36/0.2344
14	34/51/36/0.5625	41/75/50/1.0313	35/56/38/0.9219	36/59/39/0.7344
15	36/56/39/1.5781	49/81/58/2.1563	40/60/43/1.3125	40/62/43/1.7031
16	12/15/14/0.1719	7/16/11/0.1250	7/12/9/0.0938	7/12/9/0.0938
17	12/15/14/5.0625	13/22/18/6.2188	8/13/10/3.3125	8/13/10/3.2969
18	12/15/14/14.8906	12/21/16/17.1875	9/14/11/11.3906	9/14/11/11.4531
19	14/17/16/389.5625	14/22/18/438.8125	10/15/12/285.5156	8/13/10/234.6094
20	13/54/13/0.0156	13/54/13/0.0156	13/54/13/0.0313	13/54/13/0.0313
21	28/48/34/2.6409	10/21/13/0.9688	18/30/20/1.4375	15/25/17/1.1406
22	75/144/101/1.1406	15/45/23/0.3594	33/67/44/0.5313	35/71/44/0.5156
23	32/76/47/9.0000	59/145/98/18.7031	36/84/53/10.0156	41/88/54/10.3750
24	21/58/26/0.0156	18/56/25/0.0781	19/55/22/0.0156	25/62/29/0.0469

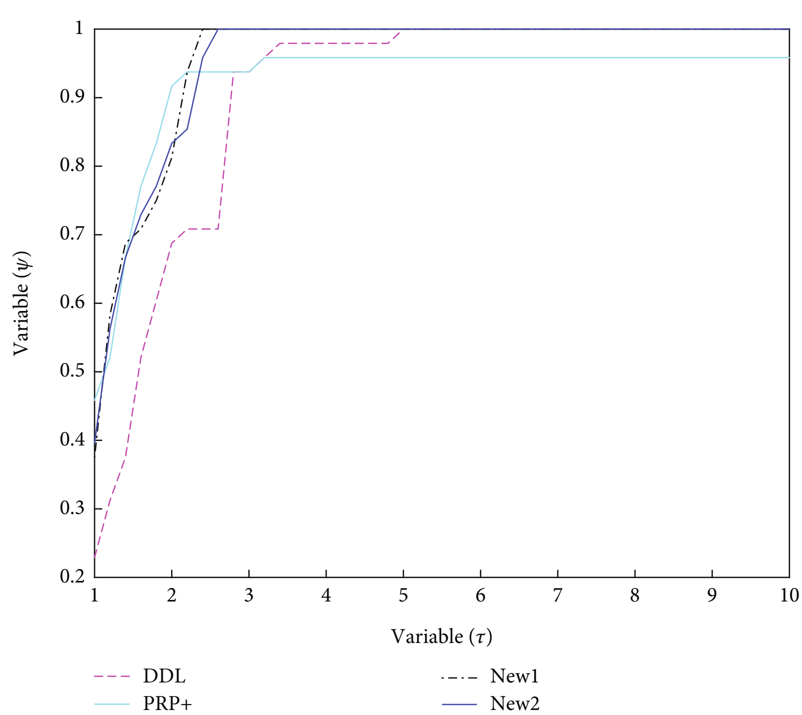

Figure 1: The number of iterations.

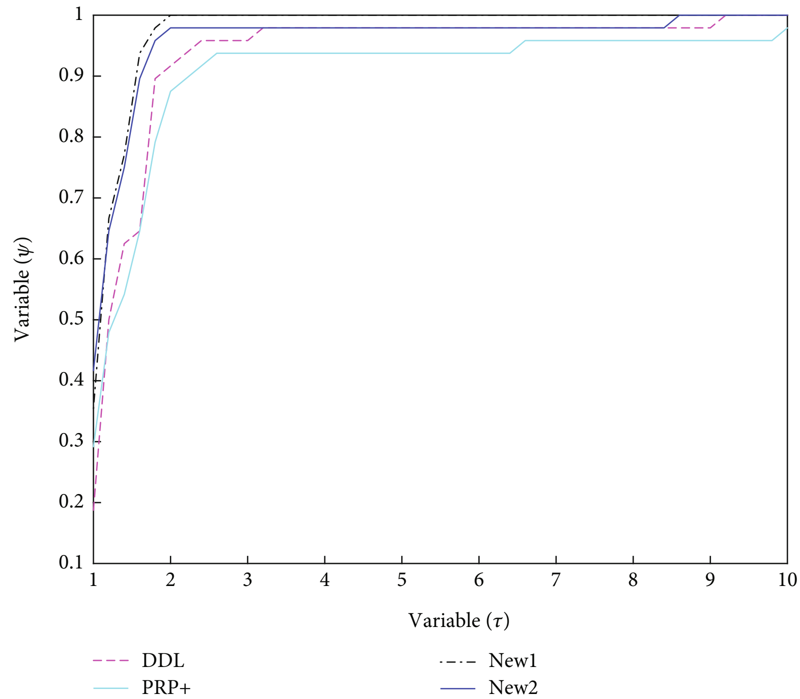

Figure 2: The number of function evaluations.

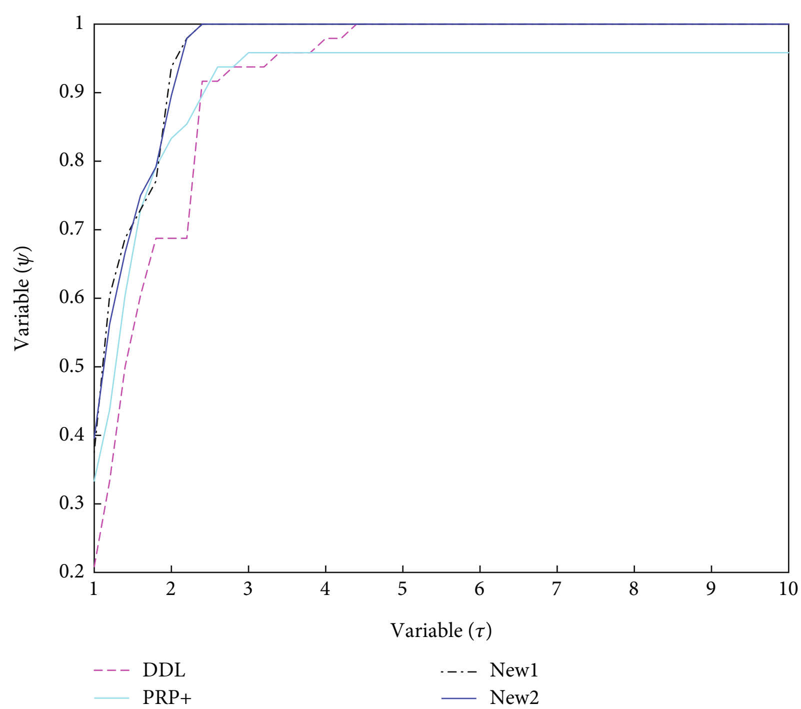

Figure 3: The number of gradient evaluations.

ng, and CPU stand for number of iterations, function evaluations, gradient evaluations, and CPU time, respectively. And based on the numerical results of all the test problems, we present the performance profile (including number of iterations, function evaluations, gradient evaluations, and CPU time) introduced by Dolan and Moré [24] to show the difference in numerical effects among the four algorithms. In a performance profile plot, the horizontal axis gives the percentage (τ) of the test problems for which a method is the fastest (efficiency), while the vertical side gives the percentage (ψ) of the test problems that are successfully solved by each of the methods.

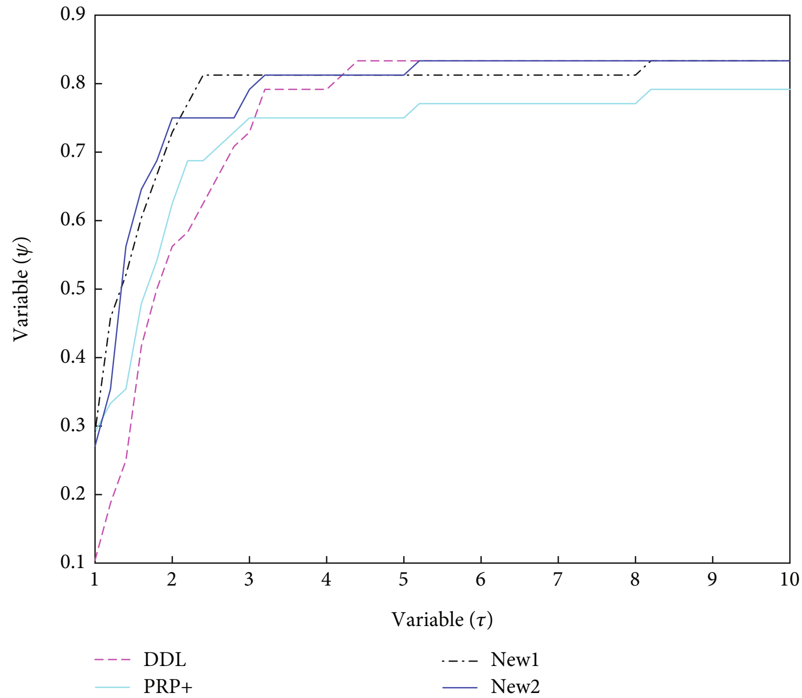

Figure 4: CPU time.

From Table 2, we can see that STCG is significantly superior to DDL for 88 percent of the problems; STCG is superior to PRP^+ for 58 percent of the problems. Figures 1–4 plot the performance profiles for the number of iterations, the number of function evaluations, the number of gradient evaluations, and the CPU time, respectively. They show that the performance of New1 and New2 is superior to DDL and PRP^+ in all aspects. In the overall trend, the performance of New1 is slightly better than New2. We deem that New1 is more competitive than New2. In conclusion, STCG method is competitive.

5. Application of STCG Algorithm in Low-Carbon Supply Chain Optimization

In recent years, global warming has become increasingly serious due to the dramatic increase in carbon emissions caused by human activities. As an important means to achieve sustainable development, energy conservation and emission reduction are highly valued by the government, enterprises, and consumers. Therefore, we use STCG algorithm to study the optimal pricing, warranty decision, and carbon emission level strategy of the two low-carbon supply chain (LSCS) models under the centralized game structure in [25].

As shown in Figure 5 [25], manufacturers sell products through retailers and provide consumers with free after-sales warranty services. Manufacturers produce greenhouse gases when they produce products and provide warranty services. The government will set a certain carbon emission quota for each enterprise. When the enterprise carbon emission quota is insufficient or excessive, the enterprise can trade in the carbon emission market.

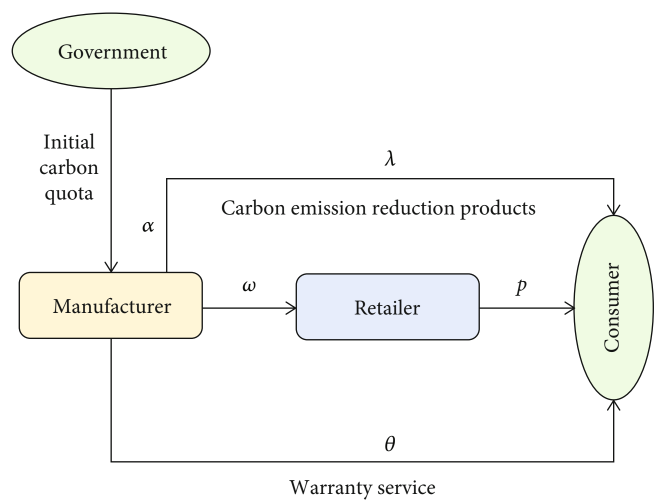

Figure 5: Supply chain operation flow chart.

For a better description of the model, the symbols are shown in Table 3.

According to the assumptions given in [25], the object problem can be transformed into profit maximization. The warranty cost function is introduced as $b\theta^2$ $(b > 0,\ \theta > 0)$ (see [26–28]). Market demand only depends on the warranty period θ, the demand function is $D = d - \alpha p + n\theta$, and the function expression for the number of products repaired by the manufacturer during the warranty period is $R = \varphi + \tau\theta$. In centralized decision-making, we regard manufacturers and retailers as subjects with identical interests, and both

Table 3: Symbolic explanation table.

Symbolic variables	c_m	Unit production cost of new products
	d	Potential market demand
	k	Unit carbon emissions trading price
	a	The initial amount of carbon emissions allocated by the government
	τ	Warranty period sensitivity coefficient for quantity of repaired products
	φ	Potential product repair quantity
	μ	Carbon emission reduction investment cost coefficient
	e	Carbon emissions from the production of one unit of new product
	b	Warranty period investment cost coefficient
	β	Revenue sharing ratio
	α	Consumer sensitivity coefficient
	D	Market demand
	R	Total number of products returned for repair
	Π_T^i	Profit function for the entire LSCS in the mode i
Decision variables	p	Unit retail price
	θ	Warranty period
	λ	Carbon emission reduction level of product

Table 4: Optimal decisions under different consumer price sensitivity coefficients.

Sensitive coefficient α	Initial point	p^{C*}	θ^{C*}	Π_T^{C*}
$\alpha = 0.1$	(0; 0)	2052.4	61.2	220090
	(50; 10)	2052.6	61.2	220090
	(100; 20)	2052.5	61.3	220090
	(1000; 40)	2052.7	61.7	220090
	(5000; 100)	2052.8	61.2	220090
$\alpha = 0.3$	(0; 0)	807.2950	8.9785	57435
	(50; 10)	807.2954	8.9828	57435
	(100; 20)	807.3985	8.8695	57435
	(1000; 40)	808.8676	8.9162	57435
	(5000; 100)	808.0076	8.9505	57435
$\alpha = 0.5$	(0; 0)	595.3961	0.0524	44145
	(50; 10)	595.3171	-0.0129	44145
	(100; 20)	595.2176	0.0495	44145

Table 5: Optimal decisions under different consumer price sensitivity coefficients.

Sensitive coefficient α	Initial point	p^{C*}	θ^{C*}	Π_T^{C*}
$\alpha = 0.5$	(1000; 40)	595.6001	0.0460	44145
	(5000; 100)	595.7273	0.0388	44145
$\alpha = 0.8$	(0; 0)	480.6638	-4.7658	53692
	(50; 10)	480.2855	-4.7935	53692
	(100; 20)	480.5973	-4.7755	53692
	(1000; 40)	480.6685	-4.7647	53692
	(5000; 100)	480.2986	-4.8443	53692
$\alpha = 1$	(0; 0)	442.6351	-6.3651	65706
	(50; 10)	442.6246	-6.3611	65706
	(100; 20)	442.6080	-6.4392	65706
	(1000; 40)	443.0113	-6.4115	65706
	(5000; 100)	442.9677	-6.4097	65706

sides cooperate to maximize LSCS profits. Therefore, the total profit function of LSCS is

$$\max_{p,\theta} \Pi_T = (p - c_m)D + k(a - eD - \varepsilon eR) - b\theta^2. \quad (72)$$

We transform (72) into the following optimization problem:

$$\min_{p,\theta} \Pi_T = -(p - c_m)D - k(a - eD - \varepsilon eR) + b\theta^2. \quad (73)$$

Based on (72), the carbon emission reduction level of the product is considered. The demand function of product $D = d - \alpha p + n\theta + \delta\lambda$ is linear, where n and δ are positive, inversely proportional to the retail price p, proportional to the warranty period θ, and the carbon emission reduction level λ. In the production process, manufacturer needs to develop carbon emission reduction technologies to increase the carbon emission reduction level of products; the investment cost function of carbon emission reduction level is $(1/2)b\lambda^2$ [29].

Table 6: Under the carbon emission reduction level of products, the optimal decision under different consumer price sensitivity coefficients.

Sensitive coefficient α	Initial point	p^{C_0*}	θ^{C_0*}	λ^{C_0*}	$\Pi_T^{C_0*}$
	(0; 0; 0)	2562.2	82.6	105.1	281600
	(50; 10; 10)	2562.9	82.6	105.2	281600
$\alpha = 0.1$	(100; 20; 20)	2561.1	82.6	105.0	281600
	(1000; 40; 40)	2559.7	82.5	105.0	281600
	(5000; 100; 100)	2564.8	82.7	105.3	281600
	(0; 0; 0)	825.3441	9.6797	12.4979	58522
	(50; 10; 10)	824.7923	9.6693	12.4181	58522
$\alpha = 0.3$	(100; 20; 20)	825.3579	9.7176	12.4898	58522
	(1000; 40; 40)	825.9481	9.7082	12.5336	58522
	(5000; 100; 100)	825.7093	9.6906	12.5188	58522
	(0; 0; 0)	595.1347	-0.0185	0.1655	44145
	(50; 10; 10)	595.2992	-0.0387	0.2185	44145
$\alpha = 0.5$	(100; 20; 20)	596.2103	0.0146	0.2719	44145
	(1000; 40; 40)	596.4003	0.0473	0.2774	44145
	(5000; 100; 100)	596.2731	0.0579	0.2949	44145
	(0; 0; 0)	477.1441	-4.9476	-6.0976	53960
	(50; 10; 10)	477.2710	-4.9305	-6.0391	53960
$\alpha = 0.8$	(100; 20; 20)	477.2881	-4.9724	-6.0170	53960
	(1000; 40; 40)	477.6123	-4.9022	-6.0851	53960
	(5000; 100; 100)	477.8681	-4.8857	-5.9646	53960
	(0; 0; 0)	439.3399	-6.5715	-8.1051	66185
	(50; 10; 10)	439.2198	-6.4765	-8.1148	66185
$\alpha = 1$	(100; 20; 20)	439.6549	-6.5136	-8.0531	66185
	(1000; 40; 40)	439.8000	-6.5487	-8.0275	66185
	(5000; 100; 100)	439.6809	-6.5573	-8.0730	66185

Considering the carbon reduction efficiency of a product under centralized decision, the overall profit function of LSCS is

$$\max_{p,\lambda,\theta} \Pi_T = (p - c_m)D + k(a - eD - \varepsilon eR) - \frac{1}{2}\mu\lambda^2 - b\theta^2. \quad (74)$$

Similarly, (74) is transformed into the following optimization problem:

$$\min_{p,\lambda,\theta} \Pi_T = -(p - c_m)D - k(a - eD - \varepsilon eR) + \frac{1}{2}\mu\lambda^2 + b\theta^2. \quad (75)$$

Then, the STCG algorithm is used to calculate the optimization problems (73) and (75), and choose different initial points (p^C, θ^C) and $(p^{C_0}, \theta^{C_0}, \lambda^{C_0})$, respectively. Based on the average unit carbon emissions trading price of April 2020 in Fujian Province of China, let $k = 9.1$. With reference to the setting of the remaining parameters in [30–32], we set $c_m = 500$, $\tau = 0.12$, $n = 0.84$, $\delta = 0.8$, $a = 5000$, $e = 10$, $d = 300$, $\varepsilon = 0.3$, $\varphi = 50$, $r = 0.4$, $\mu = 15$, $b = 10$, and $\beta = 0.2$; these values ensure that the optimal value is meaningful, and the results are shown in Tables 4–6.

Tables 4–6 show the optimal decision of LSCS under different initial points and different consumer price sensitivity coefficients α. As far as the consumer price sensitivity coefficient α is concerned, each optimal decision variable is negatively correlated with the consumer price sensitivity coefficient α. When α is small, indicate that the price has a relatively small impact on the market demand. On the premise of ensuring after-sales service and carbon emission reduction level of products, the retail and wholesale prices of products can be appropriately raised. When α is large, the price has become an important factor affecting market demand, and each decision-making variable is reduced. With the increase of α, the optimal profit shows a trend of rising first and then falling.

In terms of different initial points, on the premise of the same consumer sensitivity coefficient α, the STCG algorithm is used to solve the supply chain optimization problem. Numerical experiment results show that the solution obtained by STCG algorithm is very close to the real solution.

6. Conclusion

A new spectral three-term conjugate gradient method with random parameter is proposed. By minimizing the

Frobenius norm distance between search direction Q_{k+1} and ML-BFGS matrix based on the modified secant equation, the parameter t_k in Q_{k+1} is determined, and the random parameter is introduced to simplify the scheme of t_k. The search direction is sufficiently close to the quasi-Newton direction and satisfies the sufficient descent property. Global convergence of the new algorithm is proved under appropriate assumptions. Some classical test problems are selected for numerical experiments and compared with DDL and PRP$^+$ to verify the effectiveness of the proposed algorithm. Using STCG algorithm to solve supply chain optimization has a certain practical prospect, which reflects the effectiveness of the new algorithm. There are still many deficiencies in our research; for example, the more efficient and widely used spectral parameter θ_k is not taken into account. Therefore, the idea of alternating direction method will be used in the future work to explore the spectral parameter θ_k and parameter t_k in the search direction. The use of alternating direction method can better correct the approximate degree of search direction and quasi-Newton direction and obtain a more effective and robust method.

Data Availability

All data generated or analyzed during this study are included in this manuscript.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the Key Project of Natural Science Foundation Joint Fund of Jilin Province (YDZJ202101-ZYTS167 and YDZJ202201ZYTS303); Innovation Talent Training Program of Science and Technology of Jilin Province (20180519011JH); Science and Technology Development Project Program of Jilin Province (20190303132SF); Project of Education Department of Jilin Province (JJKH20210030KJ); Graduate Innovation Project of Beihua University (2021002).

References

[1] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin, and Y. Yuan, "Convergence properties of nonlinear conjugate gradient methods," *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 345–358, 2000.

[2] S. Aji, P. Kumam, A. M. Awwal, M. M. Yahaya, and K. Sitthithakerngkiet, "An efficient DY-type spectral conjugate gradient method for system of nonlinear monotone equations with application in signal recovery," *AIMS Mathematics*, vol. 6, no. 8, pp. 8078–8106, 2021.

[3] J. Li and S. Jing, "An improved HS type spectral conjugate gradient method," *Learning Education*, vol. 10, no. 3, pp. 142–156, 2021.

[4] G. M. Al-Naemi and A. H. Sheekoo, "Global convergence condition for a new spectral conjugate gradient method for large-scale optimization," *Journal of Physics: Conference Series*, vol. 1879, no. 3, article 032001, 2021.

[5] S. Tao, Y. Li, I. Mugume, and S. Shen, "The spectral conjugate gradient method in variational adjoint assimilation for model terrain correction I: theoretical frame," *Conference Series: Materials Science and Engineering*, vol. 768, no. 7, pp. 072044–072058, 2020.

[6] A. Neculai, "Scaled conjugate gradient algorithms for unconstrained optimization," *Computational Optimization and Applications*, vol. 38, no. 3, pp. 401–416, 2007.

[7] P. Faramarzi and K. Amini, "A modified spectral conjugate gradient method with global convergence," *Journal of Optimization Theory and Applications*, vol. 182, no. 2, pp. 667–690, 2019.

[8] S. Deng, Z. Wan, and X. Chen, "An improved spectral conjugate gradient algorithm for nonconvex unconstrained optimization problems," *Journal of Optimization Theory and Applications*, vol. 157, no. 3, pp. 820–842, 2013.

[9] X. Li, J. Shi, X. Dong, and J. Yu, "A new conjugate gradient method based on quasi-Newton equation for unconstrained optimization," *Journal of Computational and Applied Mathematics*, vol. 350, no. 18, pp. 372–379, 2019.

[10] M. Sun and J. Liu, "Three modified Polak-Ribire-Polyak conjugate gradient methods with sufficient descent property," *Journal of Inequalities and Applications*, vol. 2015, no. 1, Article ID 125, 2015.

[11] I. Livieris and P. E. Pintelas, "A new class of spectral conjugate gradient methods based on a modified secant equation for unconstrained optimization," *Journal of Computational and Applied Mathematics*, vol. 239, no. 1, pp. 396–405, 2013.

[12] D. F. Shanno, "On the convergence of a new conjugate gradient algorithm," *SIAM Journal on Numerical Analysis*, vol. 15, no. 6, pp. 1247–1257, 1978.

[13] E. G. Birgin and J. M. Martínez, "A spectral conjugate gradient method for unconstrained optimization," *Applied Mathematics and Optimization*, vol. 43, no. 2, pp. 117–128, 2001.

[14] W. Sun and Y. Yuan, "Optimization theory and methods," in *Nonlinear Programming*, vol. 1, pp. 1–700, Springer Optimization and Its Applications, 2006.

[15] W. Hager and H. Zhang, "Algorithm 851: CG-descent, a conjugate gradient method with guaranteed descent," *ACM Transactions on Mathematical Software*, vol. 32, no. 1, pp. 113–137, 2006.

[16] S. Babaie-Kafaki and R. Ghanbari, "A descent family of Dai–Liao conjugate gradient methods," *Optimization Methods and Software*, vol. 29, no. 3, pp. 583–591, 2014.

[17] Y. Dai and L. Liao, "New conjugacy conditions and related nonlinear conjugate gradient methods," *Applied Mathematics and Optimization*, vol. 43, no. 1, pp. 87–101, 2001.

[18] S. Yao and L. Ning, "An adaptive three-term conjugate gradient method based on self-scaling memoryless BFGS matrix," *Journal of Computational and Applied Mathematics*, vol. 332, no. 17, pp. 72–85, 2018.

[19] J. Stoer and Y. Yuan, "A subspace study on conjugate gradient algorithms," *Journal of Applied Mathematics and Mechanics*, vol. 75, no. 1, pp. 69–77, 1995.

[20] H. Yabe and M. Takano, "Global convergence properties of nonlinear conjugate gradient methods with modified secant condition," *Computational Optimization and Applications*, vol. 28, no. 2, pp. 203–225, 2004.

[21] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on Optimization*, vol. 2, no. 1, pp. 21–42, 1992.

[22] N. I. Gould, D. Orban, and P. L. Toint, "UTEr and SifDec: a constrained and unconstrained testing environment, revisited," *ACM Transactions on Mathematical Software*, vol. 29, no. 4, pp. 373–394, 2003.

[23] N. Andrei and N. Andrei, "Introduction: overview of unconstrained optimization," in *Nonlinear Conjugate Gradient Methods for Unconstrained Optimization*, vol. 158 of Springer Optimization and Its Applications, , pp. 1–66, Springer, 2020.

[24] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.

[25] S. Qu, H. Yang, and Y. Ji, "Low-carbon supply chain optimization considering warranty period and carbon emission reduction level under cap-and-trade regulation," *Environment, Development and Sustainability*, vol. 23, no. 12, pp. 18040–18067, 2021.

[26] S. M. Gilbert and V. Cvsa, "Strategic commitment to price to stimulate downstream innovation in a supply chain," *European Journal of Operational Research*, vol. 150, no. 3, pp. 617–639, 2003.

[27] H. Pun, "Supplier selection of a critical component when the production process can be improved," *International Journal of Production Economics*, vol. 154, pp. 127–135, 2014.

[28] X. Wu and Y. Zhou, "The optimal reverse channel choice under supply chain competition," *European Journal of Operational Research*, vol. 259, no. 1, pp. 63–66, 2017.

[29] E. Zerang, A. Taleizadeh, and J. Razmi, "Analytical comparisons in a three-echelon closed-loop supply chain with price and marketing effort-dependent demand: game theory approaches," *Environment, Development and Sustainability*, vol. 20, no. 1, pp. 451–478, 2018.

[30] X. Qian, F. Chan, J. Zhang, M. Yin, and Q. Zhang, "Channel coordination of a two-echelon sustainable supply chain with a fair- minded retailer under cap-and-trade regulation," *Journal of Cleaner Production*, vol. 244, no. 3, pp. 118715–118726, 2020.

[31] L. Yang, J. Ji, M. Wang, and Z. Wang, "The manufacturer's joint decisions of channel selections and carbon emission reductions under the cap-and-trade regulation," *Journal of Cleaner Production*, vol. 193, no. 29, pp. 506–523, 2018.

[32] M. Yu and T. Li, "Information sharing in a supply chain under cap-and-trade regulation," *Mathematical Problems in Engineering*, vol. 2018, Article ID 4573919, 18 pages, 2018.