Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 21, No. 2, pp. 339–357, 2025

A NEW SPECTRAL CONJUGATE GRADIENT ALGORITHM

Hao Wu, Liping Wang*, and Hongchao Zhang

Abstract: This paper introduces a novel conjugate gradient method that exploits the m-th order Taylor expansion of the objective function and cubic Hermite interpolation conditions. We derive a set of modified secant equations with enhanced accuracy in approximating the Hessian matrix of the objective function. Additionally, we develop a modified Wolfe line search to address the limitations of the conventional constraint imposed on modified secant equations while ensuring the fulfillment of the curvature condition. Consequently, an improved spectral conjugate gradient algorithm is proposed based on the modified secant equation and Wolfe line search. Under standard assumptions, the algorithm is proven to be globally convergent for minimizing general nonconvex functions. Numerical results are provided to demonstrate the effectiveness of this new proposed algorithm.

Key words: *modified secant equation, spectral conjugate gradient method, Wolfe line search, nonconvex problems, global convergence*

Mathematics Subject Classification: *90C32, 90C262, 65F15*

---

1 Introduction

In this paper, we consider the following unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where the objective function $f:\mathbb{R}^n\to\mathbb{R}$ is continuously differentiable. Due to its extensive theoretical developments and efficient numerical performance, the nonlinear conjugate gradient (NCG) method is a well-known method for solving unconstrained optimization problem (1.1). Denoting the gradient of $f(x)$ at x by $g(x)$, i.e., $\nabla f(x) = g(x)$, the general iterative formula of NCG can be given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

$$d_k = \begin{cases} -g(x_0), & \text{if } k = 0,\\ -g(x_k) + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{1.3}$$

where β_k is a parameter corresponding to different NCG methods, d_k is the search direction, and the step size $\alpha_k > 0$ is obtained by some line search. In this paper, we denote $f_k = f(x_k)$, $g_k = g(x_k)$.

*Corresponding author

 DOI: https:// doi.org/10.61208/pjo-2024-013

For different choices of parameter β_k, the convergence and numerical performance of nonlinear conjugate gradient methods can be fundamentally different. The following is a list of several well-known formulas of β_k for different NCG methods such as Fletcher-Reeves (FR) [10], Polak-Ribiere-Polak (PRP) [20], Hestenes-Stiefel (HS) [13], Dai-Yuan (DY) [8] and so on [17, 19]:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{PRP} = \frac{g_k^\top y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{HS} = \frac{g_k^\top y_{k-1}}{d_{k-1}^\top y_{k-1}},$$

$$\beta_k^{LS} = \frac{g_k^\top y_{k-1}}{-d_{k-1}^\top g_{k-1}}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^\top y_{k-1}}, \qquad \beta_k^{P} = \frac{g_k^\top (y_{k-1} - s_{k-1})}{d_{k-1}^\top y_{k-1}},$$

where $y_k = g_{k+1} - g_k$, $s_k = \alpha_k d_k = x_{k+1} - x_k$ and $\|\cdot\|$ denotes the 2-norm.

In order to incorporate second-order information of the objective function, Dai and Kou [7] proposed a family of CG methods where the search direction is closest to the direction of the scaled memoryless BFGS method [18, 22]. The corresponding parameter β_k of Dai-Kou (DK) method is as follows:

$$\beta_{k+1}^{DK}(\tau_k) = \frac{y_k^\top g_{k+1}}{d_k^\top y_k} - (\tau_k + \frac{\|y_k\|^2}{s_k^\top y_k} - \frac{s_k^\top y_k}{\|s_k\|^2})\frac{s_k^\top g_{k+1}}{d_k^\top y_k}, \tag{1.4}$$

where τ_k is a hyperparameter with the default choice $\tau_k = \frac{s_k^\top y_k}{\|s_k\|^2}$, which gives

$$\beta_{k+1}^{H} = \frac{y_k^\top g_{k+1}}{d_k^\top y_k} - \frac{\|y_k\|^2 d_k^\top g_{k+1}}{(d_k^\top y_k)^2}. \tag{1.5}$$

Actually, (1.5) is also a special case of β_k^{θ} from Hager and Zhang's CG_DESCENT [12].

Inspired by the idea and good practical performance of the spectral gradient method [4, 21], Brigin and Martínez [6] made an effort to combine the CG method with the spectral gradient method and proposed a spectral CG method. The search direction is yielded by

$$\begin{aligned} d_{k+1} &= -\theta_{k+1} g_{k+1} + \beta_{k+1} d_k,\ d_0 = -g_0, \\ \theta_{k+1} &= \frac{\|s_k\|^2}{s_k^\top y_k},\ \beta_{k+1} = \frac{(\theta_k y_k - s_k)^\top g_{k+1}}{s_k^\top y_k}, \end{aligned} \tag{1.6}$$

where θ_{k+1} in (1.6) is the spectral parameter same as one case of Barzilai-Borwein (BB) [4] stepsizes, $\frac{\|s_k\|^2}{s_k^\top y_k}$ and $\frac{s_k^\top y_k}{\|y_k\|^2}$. In [6], a large amount of experiments were carried showing that the proposed spectral CG method has a better numerical performance than many traditional CG methods, such as FR and PRP methods. However, the direction d_k generated by (1.6) may not be a descent search direction. In 2010, Andrei [2] proposed a spectral CG method with sufficient descent property, where the search direction is yielded by

$$\begin{aligned} d_{k+1} &= -\theta_{k+1} g_{k+1} + \beta_{k+1}^N s_k,\ d_0 = -g_0, \\ \theta_{k+1} &= \frac{1}{y_k^\top g_{k+1}}(\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|^2 s_k^\top g_{k+1}}{y_k^\top s_k}),\ \beta_{k+1}^N = \frac{\|g_{k+1}\|^2}{y_k^\top s_k} - \frac{\|g_{k+1}\|^2 s_k^\top g_{k+1}}{(y_k^\top s_k)^2}. \end{aligned} \tag{1.7}$$

The direction d_{k+1} satisfies $g_{k+1}^\top d_{k+1} \le -(\theta_{k+1} - \frac{1}{4})\|g_{k+1}\|^2$, which possesses the sufficient descent property in case $\theta_{k+1} > \frac{1}{4}$. Therefore, to ensure sufficient descent property, Andrei suggested to reset $\theta_{k+1} = 1$ when $\theta_{k+1} \le \frac{1}{4}$. More recently, Jian [14] proposed a new spectral CG method. It combines the Dai-Kou conjugate parameter given by (1.5) with the quasi-Newton direction $d_{k+1} = -B_{k+1}^{-1} g_{k+1}$, where B_{k+1} is an approximation of the Hessian $\nabla^2 f(x_{k+1})$. The search direction given in [14] is as follows:

$$d_{k+1} = \theta_{k+1}^{JC} g_{k+1} + \beta_{k+1}^{JC} d_k,\ d_0 = -g_0, \tag{1.8}$$

where

$$\begin{aligned}
\theta_k^{JC} &= \begin{cases} \theta_k^{JC+}, & \text{if } \theta_k^{JC+} \in [\frac{1}{4} + \eta, \tau], \\ 1, & \text{otherwise}, \end{cases} \\
\theta_k^{JC+} &= 1 - \frac{1}{y_k^\mathsf{T} g_{k+1}} \left(\frac{\|y_k\|^2 d_k^\mathsf{T} g_{k+1}}{d_k^\mathsf{T} g_{k+1}} - s_k^\mathsf{T} g_{k+1} \right), \quad \text{and} \\
\beta_{k+1}^{JC} &= \frac{y_k^\mathsf{T} g_{k+1}}{d_k^\mathsf{T} y_k} - \frac{\|y_k\|^2 d_k^\mathsf{T} g_{k+1}}{(d_k^\mathsf{T} y_k)^2}.
\end{aligned}$$

Here, $\eta > 0$ and $\tau > 0$ are two positive constants such that $\frac{1}{4} + \eta < \tau$. Although global convergence of the spectral CG method is only established for uniformly convex functions [14], this method shows good numerical performance for minimizing a publicly available set of general testing functions [1].

The above mentioned methods proposed by Dai and Kou (1.4), Andrei (1.7) and Jian (1.8) are all based on the standard secant (also called quasi-Newton) equation

$$B_{k+1} s_k = y_k, \tag{1.9}$$

where B_{k+1} is some approximation of Hessian of f at x_{k+1}. However, the standard secant equation only employs gradient information but ignores function value information of the objective function. In order to get a better approximation of the Hessian matrix, techniques using both gradients and function values have been studied by many researchers. For example, Yuan [24], Khoshgam [15], Biglari [5] and Yuan [25], respectively, proposed the following modified secant equations deriving from the Taylor expansion of the objective function with some interpolation conditions:

$$B_{k+1} s_k = z_k^{(\infty)}, z_k^{(\infty)} = y_k + \frac{\mu_k^{(\infty)}}{\|s_k\|^2} s_k, \tag{1.10}$$

$$B_{k+1} s_k = z_k^{(5)}, z_k^{(5)} = y_k + \frac{\mu_k^{(5)}}{\|s_k\|^2} s_k, \tag{1.11}$$

$$B_{k+1} s_k = z_k^{(4)}, z_k^{(4)} = y_k + \frac{\mu_k^{(4)}}{\|s_k\|^2} s_k, \tag{1.12}$$

$$B_{k+1} s_k = z_k^{(3)}, z_k^{(3)} = y_k + \frac{\mu_k^{(3)}}{\|s_k\|^2} s_k, \tag{1.13}$$

where

$$\begin{aligned}
\mu_k^{(\infty)} &= 2(f_k - f_{k+1}) + (g_k + g_{k+1})^\mathsf{T} s_k, \\
\mu_k^{(5)} &= \frac{10}{3}(f_k - f_{k+1}) + \frac{5}{3}(g_k + g_{k+1})^\mathsf{T} s_k, \\
\mu_k^{(4)} &= 4(f_k - f_{k+1}) + 2(g_k + g_{k+1})^\mathsf{T} s_k, \\
\mu_k^{(3)} &= 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^\mathsf{T} s_k.
\end{aligned}$$

In this paper, we propose a new NCG method combined with a modified line search method to ensure global convergence. Our main contributions are as follows. Firstly, based on a unified derivation framework, we propose a family of modified secant equations, containing all the previous variants (1.10)-(1.13). Our best choice among the modified secant equation family can be interpreted from the perspective of polynomial interpolation. Then, from our choice of secant equation a new improved CG search direction can be derived. Secondly, it is well-known that under the standard Wolfe line search, the positive curvature condition $s_k^\mathsf{T} y_k > 0$ always holds. However, we usually can not guarantee $s_k^\mathsf{T} z_k^{(m)} > 0$ for $z_k^{(m)}$ given by (1.10)-(1.13). To ensure the curvature condition $s_k^\mathsf{T} z_k^{(m)} > 0$ holds, a frequently-used correction is suggested in [3, 23, 15] to replace $\mu_k^{(m)}$ by $\max\{\mu_k^{(m)}, 0\}$. In fact, restricting $\mu_k^{(m)}$ to be nonnegative may destroy the negative curvature information of the Hessian inherited in $\mu_k^{(m)}$. In addition, it is also observed from numerical experiments that negative $\mu_k^{(m)}$ appears quite often in practice and simply restricting $\mu_k^{(m)}$ to be nonnegative can significantly deteriorate the numerical performance. Hence, we propose a modified Wolfe line search in this paper which can retain some negative $\mu_k^{(m)}$ without losing global convergence.

The paper is organized as follows: In Section 2, we derive a series of modified secant equations, based on which a family of modified spectral CG methods is obtained. We then propose a modified Wolfe line search and present a new spectral CG algorithm. The global convergence of our new spectral CG algorithm for minimizing general nonconvex functions is established in Section 3. Section 4 conducts numerical experiments comparing our new method with some other well-known NCG methods. We finally draw some conclusion remarks in Section 5.

$\boxed{2}$ A New Spectral CG Algorithm

$\boxed{2.1}$ A new CG search direction

Assume the function f is sufficiently smooth and consider the m-th order Taylor expansion model at x_{k+1}:

$$\begin{aligned}\varphi_{k+1}^{(m)}(d) = f_{k+1} + d^\mathsf{T} g_{k+1} + \frac{1}{2} d^\mathsf{T} \nabla^2 f(x_{k+1}) d + \frac{1}{3!} d^\mathsf{T} [\nabla^3 f(x_{k+1}) d] d + \cdots \\ + \frac{1}{m!} d^\mathsf{T} [\nabla^m f(x_{k+1}) \underbrace{d \cdots d}_{m-2}] d,\end{aligned} \tag{2.1}$$

where $\nabla^m f(x_{k+1}) \in \mathbb{R}^{\overbrace{n \times n \times n \times \cdots \times n}^{m}}$ is a m-th order tensor and

$$d^\mathsf{T} [\nabla^m f(x_{k+1}) \underbrace{d \cdots d}_{m-2}] d = \sum_{n_1, \cdots, n_m = 1}^{n} \frac{\partial^m f(x_{k+1})}{\partial x^{n_1}, \cdots, \partial x^{n_m}} d^{n_1}, \cdots, d^{n_m}.$$

Then, we have

$$\varphi_{k+1}^{(m)}(d) = f(x_{k+1} + d) + o(\|d\|^m) \tag{2.2}$$

for all $d \in \mathbb{R}^n$ with $\|d\|$ sufficiently small. In addition, it follows from the definition of $\varphi_{k+1}^{(m)}(d)$ in (2.1) that

$$\varphi_{k+1}^{(m)}(0) = f_{k+1} \quad \text{and} \quad \nabla \varphi_{k+1}^{(m)}(0) = g_{k+1}. \tag{2.3}$$

Furthermore, by (2.2), the following cubic Hermite interpolation conditions also hold

$$\varphi_{k+1}^{(m)}(x_k - x_{k+1}) = f_k + o(\|d\|^m) \quad \text{and} \quad \nabla\varphi_{k+1}^{(m)}(x_k - x_{k+1}) = g_k + o(\|d\|^{m-1}), \tag{2.4}$$

which are equivalent to

$$\begin{aligned} f_k = f_{k+1} - s_k^\mathsf{T} g_{k+1} + \frac{1}{2} s_k^\mathsf{T} \nabla^2 f(x_{k+1}) s_k + (-1)^3 \frac{1}{3!} s_k^\mathsf{T} [\nabla^3 f(x_{k+1}) s_k] s_k + \cdots \\ + (-1)^m \frac{1}{m!} s_k^\mathsf{T} [\nabla^m f(x_{k+1}) \underbrace{s_k \cdots s_k}_{m-2}] s_k + o(\|s_k\|^m) \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} g_k = g_{k+1} - \nabla^2 f(x_{k+1}) s_k + (-1)^2 \frac{1}{2} [\nabla^3 f(x_{k+1}) s_k] s_k + \cdots \\ + (-1)^{m-1} \frac{1}{(m-1)!} [\nabla^m f(x_{k+1}) \underbrace{s_k \cdots s_k}_{m-2}] s_k + o(\|s_k\|^{m-1}). \end{aligned} \tag{2.6}$$

Using the above two equations, by direct calculations of $m \times f_k + s_k^\mathsf{T} g_k$, we obtain

$$\begin{aligned} s_k^\mathsf{T} \nabla^2 f(x_{x_{k+1}}) s_k = {} & s_k^\mathsf{T} y_k + \frac{m}{m-2} (2(f_k - f_{k+1}) + (g_k + g_{k+1})^\mathsf{T} s_k) \\ & + (-1)^2 \frac{m-3}{3(m-2)} s_k^\mathsf{T} [\nabla^3 f(x_{k+1}) s_k] s_k \\ & + (-1)^3 \frac{m-4}{12(m-2)} s_k^\mathsf{T} [\nabla^4 f(x_{k+1}) s_k s_k] s_k \\ & \cdots \\ & + (-1)^{m-2} \Big(\frac{(3-m)m}{(m-1)!(m-2)} + \frac{1}{(m-2)!} \Big) s_k^\mathsf{T} [\nabla^{m-1} f(x_{k+1}) \underbrace{s_k \cdots s_k}_{m-3}] s_k \\ & + o(\|s_k\|^m), \end{aligned} \tag{2.7}$$

which suggests us to define the following series of modified secant equations

$$\begin{aligned} & B_{k+1} s_k = z_k^{(m)}, \ z_k^{(m)} = y_k + \frac{m\mu_k}{(m-2)\|s_k\|^2} s_k, \ m \ge 3, \ m \in \mathbb{Z}, \\ & \mu_k = 2(f_k - f_{k+1}) + (g_k + g_{k+1})^\mathsf{T} s_k. \end{aligned} \tag{2.8}$$

Setting $m = 3, 4, 5$ in (2.8) yields the secant equations (1.13), (1.12) and (1.11). It is also notable that (1.10) can be regarded as (2.8) with $m = +\infty$. Based on the above discussions, we easily establish the following approximation properties.

Theorem 2.1. *Assume the function f is sufficiently smooth. When $\|s_k\|$ is sufficiently* *small, we have*

$$
\begin{aligned}
s_k^\top \nabla^2 f(x_{k+1}) s_k - s_k^\top y_k &= \frac{1}{2} s_k^\top [\nabla^3 f(x_{k+1}) s_k] s_k + o(\|s_k\|^3), \\
s_k^\top \nabla^2 f(x_{k+1}) s_k - s_k^\top z_k^{(\infty)} &= \frac{1}{3} s_k^\top [\nabla^3 f(x_{k+1}) s_k] s_k + o(\|s_k\|^3), \\
s_k^\top \nabla^2 f(x_{k+1}) s_k - s_k^\top z_k^{(5)} &= \frac{2}{9} s_k^\top [\nabla^3 f(x_{k+1}) s_k] s_k + o(\|s_k\|^3), \\
s_k^\top \nabla^2 f(x_{k+1}) s_k - s_k^\top z_k^{(4)} &= \frac{1}{6} s_k^\top [\nabla^3 f(x_{k+1}) s_k] s_k + o(\|s_k\|^3), \\
s_k^\top \nabla^2 f(x_{k+1}) s_k - s_k^\top z_k^{(3)} &= o(\|s_k\|^3), \\
s_k^\top \nabla^2 f(x_{k+1}) s_k - s_k^\top z_k^{(m)} &= \frac{m-3}{3(m-2)} s_k^\top [\nabla^3 f(x_{k+1}) s_k] s_k + o(\|s_k\|^3).
\end{aligned}
$$

Proof. The results directly follows from the equality (2.7). □

Remark 2.2. We have the following comments on the above results obtained by polynomial interpolation. For the modified secant equation (1.13), Yuan [25] supposed a cubic approximation to the objective function and used two-point cubic Hermite interpolation conditions (2.3)-(2.4) at x_k and x_{k+1}. So, $s_k^\top z_k^{(3)}$ approximates $s_k^\top \nabla^2 f(x_{k+1}) s_k$ more accurately. For (1.12) and (1.11), Biglari [5] and Khoshgam [15] considered the quartic and quintic model approximation to the function f, but still employed with the cubic Hermite interpolation conditions. It is known that at least $m+1$ interpolation conditions are required to determine a m-th order interpolation function. Hence, interpolation conditions (2.3)-(2.4) are insufficient for a higher-order ($m > 3$) model. As a result, we can see from Theorem 2.1 that $s_k^\top z_k^{(5)}$ and $s_k^\top z_k^{(4)}$ possesses low approximation quality compared to $s_k^\top z_k^{(3)}$. Furthermore, by Theorem 2.1, larger values of m will in fact lead to lower approximation accuracy. So, $s_k^\top z_k^{(\infty)}$, which corresponds to $m = \infty$, has the lowest quality to approximate $s_k^\top \nabla^2 f(x_{k+1}) s_k$ compared with other choices of $z_k^{(m)}$ with $m = 3, 4, 5$.

In this paper, we propose the following spectral conjugate search direction

$$
d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_{k+1} d_k, \ d_0 = -g_0, \tag{2.9}
$$

where the conjugate parameter β_{k+1} is given by a truncated form of (1.5) but with y_k replaced by $z_k^{(m)}$,

$$
\begin{aligned}
\beta_{k+1}^M &= \max\{\beta_{k+1}^L, \beta_{k+1}^R\}, \qquad (2.10) \\
\beta_{k+1}^L &= \frac{g_{k+1}^\top z_k^{(m)}}{d_k^\top z_k^{(m)}} - \frac{||z_k^{(m)}||^2}{d_k^\top z_k^{(m)}} \frac{g_{k+1}^\top d_k}{d_k^\top z_k^{(m)}}, \\
\beta_{k+1}^R &= \frac{g_k^\top d_k}{||d_k||^2}.
\end{aligned}
$$

For the choice of the spectral parameter θ_{k+1}, we hope d_{k+1} satisfies the quasi-Newton direction, i.e.

$$
B_{k+1} d_{k+1} = -g_{k+1}. \tag{2.11}
$$

It follows that

$$
-g_{k+1} = -\theta_{k+1} B_{k+1} g_{k+1} + \beta_{k+1} B_{k+1} d_k. \tag{2.12}
$$

Taking inner product of s_k with both sides of (2.12), we have

$$\theta_{k+1} = \frac{1}{s_k^\top B_{k+1} g_{k+1}} (s_k^\top g_{k+1} + \beta_{k+1} d_k^\top B_{k+1} s_k). \tag{2.13}$$

Applying the modified secant condition $B_{k+1}s_k = z_k^{(m)}$ to (2.13), for all $k \geq 0$, we can obtain our choice of θ_{k+1} as

$$\tilde{\theta}_{k+1} = \frac{1}{g_{k+1}^\top z_k^{(m)}} (s_k^\top g_{k+1} + \beta_{k+1} d_k^\top z_k^{(m)}). \tag{2.14}$$

The following theorem shows that the spectral conjugate direction d_{k+1} with $\theta_{k+1} = \tilde{\theta}_{k+1}$ and $\beta_{k+1} = \beta_{k+1}^L$ has sufficient descent property when $\theta_{k+1} > 1/4$.

Theorem 2.3. *If $g_{k+1}^\top z_k^{(m)} \neq 0$ and $d_k^\top z_k^{(m)} \neq 0$, for the direction d_{k+1} defined by* (2.9) *with $\theta_{k+1} = \tilde{\theta}_{k+1}$ and $\beta_{k+1} = \beta_{k+1}^L$, we have*

$$g_{k+1}^\top d_{k+1} \leq -(\theta_{k+1} - \frac{1}{4})\|g_{k+1}\|^2, \ k \geq 0. \tag{2.15}$$

Proof. Multiplying (2.10) by $g_{k+1}^\top$, we obtain

$$g_{k+1}^\top d_{k+1} = -\theta_{k+1}\|g_{k+1}\|^2 + \frac{g_{k+1}^\top z_k^{(m)} g_{k+1}^\top d_k}{d_k^\top z_k^{(m)}} - \frac{\|z_k^{(m)}\|^2}{d_k^\top z_k^{(m)}} \frac{(g_{k+1}^\top d_k)^2}{d_k^\top z_k^{(m)}}. \tag{2.16}$$

From the basic inequality $u^\top v \leq \frac{1}{2}(\|u\|^2 + \|v\|^2)$ for any u, $v \in \mathbb{R}^n$, we take $u = \frac{g_{k+1}}{\sqrt{2}}$ and $v = \frac{\sqrt{2} z_k^{(m)} g_{k+1}^\top d_k}{d_k^\top z_k^{(m)}}$. Then we obtain

$$\frac{g_{k+1}^\top z_k^{(m)} g_{k+1}^\top d_k}{d_k^\top z_k^{(m)}} \leq \frac{1}{4}\|g_{k+1}\|^2 + \frac{\|z_k^{(m)}\|^2}{d_k^\top z_k^{(m)}} \frac{(g_{k+1}^\top d_k)^2}{d_k^\top z_k^{(m)}}. \tag{2.17}$$

Combining (2.16) with (2.17) yields (2.15). □

Theorem 2.3 ensures that d_k is a sufficient descent direction for all $k \geq 1$ as long as $\theta_k > \frac{1}{4}$. So, to ensure sufficient descent search direction and boundedness of spectral parameter, we set a truncated value of $\tilde{\theta}_{k+1}$ as

$$\bar{\theta}_{k+1} = \begin{cases} \tilde{\theta}_{k+1}, & \text{if } \tilde{\theta}_{k+1} \in [\frac{1}{4} + \eta, \tau], \\ 1, & \text{otherwise,} \end{cases} \tag{2.18}$$

where $\eta \in (0, 3/4)$ is a small positive constant and $\tau > 1$ is an upper bound for the spectral parameter. Then, by (2.18), we obviously have $\bar{\theta}_{k+1} \in [\frac{1}{4} + \eta, \tau]$ and the following corollary directly follows from Theorem 2.3.

Corollary 2.4. *If $g_{k+1}^\top z_k^{(m)} \neq 0$ and $d_k^\top z_k^{(m)} \neq 0$, for the direction d_{k+1} defined by* (2.9) *with $\theta_{k+1} = \bar{\theta}_{k+1}$ and $\beta_{k+1} = \beta_{k+1}^L$, we have*

$$g_{k+1}^\top d_{k+1} \leq -\eta\|g_{k+1}\|^2, \ k \geq 0. \tag{2.19}$$

We now show that d_{k+1} defined by (2.11) with $\theta_{k+1} = \bar{\theta}_{k+1}$ and $\beta_{k+1} = \beta^M_{k+1}$ is a sufficient descent search direction.

Theorem 2.5. *If $g^{\mathsf{T}}_{k+1} z^{(m)}_k \neq 0$ and $d^{\mathsf{T}}_k z^{(m)}_k \neq 0$, for the direction d_{k+1} defined by (2.9) with $\theta_{k+1} = \bar{\theta}_{k+1}$ and $\beta_{k+1} = \beta^M_{k+1}$ given by (2.18) and (2.10), respectively, we have*

$$g^{\mathsf{T}}_{k+1} d_{k+1} \leq -\eta \|g_{k+1}\|^2, \ k \geq 0. \tag{2.20}$$

Proof. If $\beta_{k+1} = \beta^L_{k+1}$, then (2.20) follows from Corollary 2.4. Now, supposing $\beta_{k+1} = \beta^R_{k+1}$, we prove (2.20) holds by induction. First, d_0 satisfies that $g^{\mathsf{T}}_0 d_0 = -\|g_0\|^2 < -\eta \|g_0\|^2$ since $\eta \in (0, 3/4)$. Suppose d_k satisfies (2.20). Since $\beta_{k+1} = \beta^R_{k+1}$, we have from (2.18) and $d^{\mathsf{T}}_k g_k < 0$ that $\beta^L_{k+1} \leq \beta^R_{k+1} < 0$. By (2.11), we have

$$g^{\mathsf{T}}_{k+1} d_{k+1} = -\theta_{k+1} \|g_{k+1}\|^2 + \beta^R_{k+1} g^{\mathsf{T}}_{k+1} d_k.$$

If $g^{\mathsf{T}}_{k+1} d_k \geq 0$, (2.20) follows immediately from $\bar{\theta}_{k+1} \in [\frac{1}{4} + \eta, \tau]$. If $g^{\mathsf{T}}_{k+1} d_k < 0$, we have from $\beta^L_{k+1} \leq \beta^R_{k+1} < 0$ that

$$g^{\mathsf{T}}_{k+1} d_{k+1} = -\theta_{k+1} \|g_{k+1}\|^2 + \beta^R_{k+1} g^{\mathsf{T}}_{k+1} d_k \leq -\theta_{k+1} \|g_{k+1}\|^2 + \beta^L_{k+1} g^{\mathsf{T}}_{k+1} d_k.$$

So, (2.20) follows from Corollary 2.4. Hence, d_{k+1} satisfies (2.20). The proof is complete. □

2.2 A modified Wolfe line search

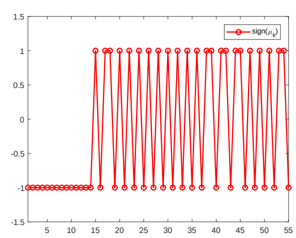

Figure 1: The sign of μ_k

iteration	1	2	3	4	5	6	7	8
μ_k	-1.107E+05	-2.634E+03	-3.307E+03	-2.750E-01	-8.784E-01	-1.102E-07	-4.256E-07	-6.419E-04
iteration	9	10	11	12	13	14	15	16
μ_k	-2.595E-09	-2.381E-07	-5.132E-14	-7.026E-13	-2.812E-13	-3.412E-13	2.983E-14	-3.223E-14
iteration	17	18	19	20	21	22	23	24
μ_k	7.043E-16	2.123E-14	-2.165E-14	1.962E-13	-1.958E-13	4.604E-13	-4.602E-13	5.752E-13

Table 1: The value of μ_k in the initial 24 iterations

For general nonlinear functions, the value of $d^{\mathsf{T}}_k z^{(m)}_k$ may be zero or negative, while $d^{\mathsf{T}}_k y_k$ is always positive when standard Wolfe line search is used in the optimization algorithms. To maintain $d^{\mathsf{T}}_k z^{(m)}_k > 0$, a common remedy procedure is to replace μ_k in (2.8) by $\mu^+_k := \max\{\mu_k, 0\}$. By this way, when μ_k is negative, we have $\mu^+_k = 0$ and the secant equation (2.8) will reduce to the standard secant equation (1.9). However, negative μ_k may inherit some negative curvature information of the Hessian when the iterates are far from a local minimizer. In our numerical experiments, we find that significant negative values of μ_k often appear in early iterations. As a typical example, the sign and values of μ_k against iteration number for minimizing the ARWHEAD function [1] are displayed in Figure 1 and Table 1. We can see from this experiment that μ_k is significantly negative in the first a few iterations while it is nearly zero with alternating sign in later iterations when approaching a local minimizer. To maintain certain negative values of μ_k in our algorithm without losing global convergence, we propose to modify $z_k^{(m)}$ in (2.8) as

$$z_k^{(m)} = y_k + t_k s_k,\ m \geq 3,\ m \in \mathbb{Z}, \tag{2.21}$$

where

$$t_k = \begin{cases} \frac{m\mu_k}{(m-2)\|s_k\|^2}, & \text{if } \mu_k > 0, \\ \kappa \frac{\mu_k}{\|s_k\|^2}, & \text{otherwise,} \end{cases} \tag{2.22}$$

with

$$\mu_k = 2(f_k - f_{k+1}) + (g_k + g_{k+1})^\mathsf{T} s_k \quad \text{and} \quad \kappa = \frac{\sigma - \rho}{1 - 2\rho + \sigma}$$

for some $0 < \rho < \sigma < 1$, and apply the following modified Wolfe line search

$$\begin{aligned} &f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^\mathsf{T} d_k, \\ &(g(x_k + \alpha_k d_k) + \min\{t_k, 0\} s_k)^\mathsf{T} d_k \geq \sigma g_k^\mathsf{T} d_k. \end{aligned} \tag{2.23}$$

It is obvious that when $t_k > 0$, the above modified Wolfe line search will be reduced to the standard Wolfe line search.

Now our new spectral conjugate gradient algorithm, denoted as NSCG, with modified Wolfe line search is described in Alg. 1.

Algorithm 1 A New Spectral Conjugate Gradient (NSCG) Algorithm

Input: An initial point $x_0 \in \mathbb{R}^n$ and tolerance $\varepsilon > 0$.
1: Set $k = 0,\ d_0 = -g_0$.
2: If $\|g_k\|_\infty \leq \varepsilon$, stop.
3: Compute α_k by the modified Wolfe conditions (2.23).
4: Update $x_{k+1} = x_k + \alpha_k d_k$ and g_{k+1}.
5: Compute β_{k+1} , θ_{k+1} and d_{k+1} by (2.10), (2.18) and (2.9), respectively, where $z_k^{(m)}$ is determined by (2.21).
6: Set $k = k + 1$ and go to step 2.

3 The Global Convergence

In this section, we establish global convergence of the NSCG algorithm proposed in Alg. 1. For this purpose, we need the following assumptions on the objective function f.

Assumption 1. (a) f is bounded from below, i.e. $f(x) > -\infty,\ \forall x \in \mathbb{R}^n$.
(b) The level set $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$is bounded, namely, there exists a constant M such that

$$||x|| \leq M,\ \forall x \in S. \tag{3.1}$$

(c) In some neighborhood Ω of S, f is continuously differentiable and its gradient is Lipschitz continuous, i.e. there exists a constant $L > 0$ such that

$$||g(x) - g(y)|| \leq L||x - y||,\ \forall x, y \in \Omega, \tag{3.2}$$

which implies that there exists a constant $\gamma > 0$ such that

$$||g(x)|| \leq \gamma, \ \forall x \in S. \tag{3.3}$$

Under Assumption 1, we first show that the modified line search (2.23) is well-defined.

Lemma 3.1. *Suppose that Assumption 1 holds and d_k is a descent direction. There exists a suitable stepsize α_k satisfying modified Wolfe conditons* (2.23).

Proof. Since f is bounded from below by Assumption 1 (a), it follows from smoothness of f and $0 < \rho < 1$ that there exists a stepsize $\alpha_k' > 0$ such that

$$f(x_k + \alpha_k' d_k) = f(x_k) + \alpha_k' \rho g_k^\top d_k, \tag{3.4}$$

and $f(x_k + \alpha_k d_k) \leq f(x_k) + \rho \alpha_k g_k^\top d_k$ for all $0 < \alpha_k < \alpha_k'$. By the mean value theorem, there exists $\alpha_k'' \in (0, \alpha_k')$ such that

$$f(x_k + \alpha_k' d_k) - f(x_k) = \alpha_k' g(x_k + \alpha_k'' d_k)^\top d_k. \tag{3.5}$$

Combining (3.4) with (3.5), we obtain

$$g(x_k + \alpha_k'' d_k)^\top d_k = \rho g_k^\top d_k > \sigma g_k^\top d_k. \tag{3.6}$$

Let $s_k = \alpha_k'' d_k$. If $\mu_k > 0$, the second condition in (2.23) holds obviously by (3.6). If $\mu_k \leq 0$, we have from the definition of t_k and $\kappa = (\sigma - \rho)/(1 - 2\rho + \sigma)$ in (2.22) that

$$\begin{aligned} g(x_k + \alpha_k'' d_k)^\top s_k + t_k ||s_k||^2 &= \rho g_k^\top s_k + \kappa \mu_k = \rho g_k^\top s_k + \kappa (2(f_k - f_{k+1}) + (g_k + g_{k+1})^\top s_k) \\ &\geq \rho g_k^\top s_k + \kappa(-2\rho g_k^\top s_k + g_k^\top s_k + \sigma g_k^\top s_k) \\ &= \rho g_k^\top s_k + (\sigma - \rho) g_k^\top s_k \\ &= \sigma g_k^\top s_k. \end{aligned} \tag{3.7}$$

So, the second condition in (2.23) also holds. The proof is complete. □

We now establish the Zoutendijk condition [26] for our new modified Wolfe line search, which often plays an important role for showing global convergence of line search methods.

Lemma 3.2. *Suppose that Assumption 1 holds. Consider the iterative scheme* (1.2)*, where d_k is a descent direction and α_k is determined by the modified Wolfe conditions* (2.23)*. Then, we have*

$$\sum_{k=0}^{\infty} \frac{\left(g_k^\top d_k\right)^2}{||d_k||^2} < \infty. \tag{3.8}$$

Proof. First, note that

$$\begin{aligned} \mu_k &= 2(f_k - f_{k+1}) + (g_k + g_{k+1})^\top s_k = 2g(\bar{x}_k)^\top (x_k - x_{k+1}) + (g_k + g_{k+1})^\top s_k \\ &= -2g(\bar{x}_k)^\top s_k + (g_k + g_{k+1})^\top s_k = (g_k - g(\bar{x}_k) + g_{k+1} - g(\bar{x}_k))^\top s_k, \end{aligned}$$

where $\bar{x}_k = v x_k + (1 - v) x_{k+1}$ for some $v \in [0, 1]$. Then, by (3.2), we have

$$\begin{aligned} |\mu_k| &\leq (||g_k - g(\bar{x}_k)|| + ||g_{k+1} - g(\bar{x}_k)||) ||s_k|| \\ &\leq L\left(||x_k - \bar{x}_k|| + ||\bar{x}_k - x_{k+1}||\right) ||s_k|| = L||s_k||^2. \end{aligned} \tag{3.9}$$

By the definition of t_k in (2.22), if $\mu_k > 0$, we have $0 < t_k \le \frac{mL}{m-2}$. On the other hand, if $\mu_k \le 0$, we have from (3.9) and (2.22) that $-\kappa L \le t_k \le 0$, where $0 < \kappa = (\sigma - \rho)/(1 - 2\rho + \sigma) < 1$. Thus we have

$$-\kappa L \le t_k \le \frac{mL}{m-2}. \tag{3.10}$$

Furthermore, we have from (3.2) and the second condition in (2.23) that

$$L\alpha_k ||d_k||^2 \ge (g_{k+1} - g_k)^\top d_k \ge (\sigma - 1) g_k^\top d_k - t_k \alpha_k ||d_k||^2,$$

which with (3.10) gives

$$\alpha_k \ge \frac{\sigma - 1}{L + t_k} \frac{g_k^\top d_k}{||d_k||^2} \ge \frac{\sigma - 1}{L(1 + \frac{m}{m-2})} \frac{g_k^\top d_k}{||d_k||^2}. \tag{3.11}$$

Then, it follows from (3.11) and the first condition in (2.23) that

$$f_k - f_{k+1} \ge q \frac{(g_k^\top d_k)^2}{||d_k||^2}, \tag{3.12}$$

where $q = \frac{\rho(1-\sigma)}{L(1+\frac{m}{m-2})}$. Summing (3.12) over k and noting that f is bounded from below, we have (3.8) holds. The proof is complete. □

In the following, we show global convergence of NSCG in the sense that

$$\liminf_{k \to \infty} \ \|g_k\| = 0. \tag{3.13}$$

Assuming that (3.13) does not hold, i.e. there exists a constant $\xi > 0$ such that

$$\|g_k\| \ge \xi \tag{3.14}$$

for all $k \ge 0$, we will show (3.13) holds by way of contradiction. We first establish two necessary lemmas.

Lemma 3.3. *Suppose Assumption 1 holds. Consider the iterative form* (1.2)*, where α_k is determined by the modified Wolfe conditions* (2.23) *and d_k is given by* (2.9) *with β_{k+1} and θ_{k+1} being generated by* (2.10) *and* (2.18)*, respectively. If* (3.14) *holds and $d_k \ne 0$, we have*

$$\sum_{k \ge 1} ||u_k - u_{k-1}||^2 < \infty, \tag{3.15}$$

where $u_k = \frac{d_k}{\|d_k\|}$.

Proof. Divide β_{k+1} given by (2.10) into two parts as follows:

$$\beta_{k+1}^1 = \max\{\beta_{k+1}, 0\} \tag{3.16}$$

and

$$\beta_{k+1}^2 = \min\{\beta_{k+1}, 0\}. \tag{3.17}$$

Define

$$\omega_k = \frac{-\theta_k g_k + \beta_k^2 d_{k-1}}{||d_k||} \tag{3.18}$$

and

$$\delta_k = \frac{\beta_k^1 ||d_{k-1}||}{||d_k||}. \tag{3.19}$$

From $d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_{k+1} d_k$, we have

$$u_k = \omega_k + \delta_k u_{k-1},\ \ k \geq 1. \tag{3.20}$$

Since $\{u_k, k \geq 0\}$ are unit vectors, we have

$$||\omega_k|| = ||u_k - \delta_k u_{k-1}|| = ||\delta_k u_k - u_{k-1}||. \tag{3.21}$$

Using (3.21) and the condition $\delta_k \geq 0$, we derivate

$$\begin{aligned} ||u_k - u_{k-1}|| &\leq ||(1+\delta_k)u_k - (1+\delta_k)u_{k-1}|| \\ &\leq ||u_k - \delta_k u_{k-1}|| + ||\delta_k u_k - u_{k-1}|| = 2||\omega_k||. \end{aligned} \tag{3.22}$$

By (3.14), the relation $\theta_k \leq \tau$, the definition of β_{k+1}^2 and the fact $\frac{g_k^\top d_k}{||d_k||^2} \leq \beta_{k+1} \leq 0$, we have

$$||-\theta_k g_k + \beta_k^2 d_{k-1}|| \leq \tau ||g_k|| + \frac{|g_{k-1}^\top d_{k-1}|}{||d_{k-1}||^2} ||d_{k-1}|| \leq (1+\tau)\gamma, \tag{3.23}$$

where $\gamma > 0$ is the constant defined in Assumption (1)(c). From (3.18), (3.22) and (3.23), it follows that

$$||u_k - u_{k-1}|| \leq \frac{2(1+\tau)\gamma}{||d_k||}. \tag{3.24}$$

Then, the descent property of d_k (2.20) with the relation (3.14) yields

$$\frac{1}{||d_k||^2} \leq \frac{1}{\xi^4} \frac{||g_k||^4}{||d_k||^2} \leq \frac{1}{\xi^4 \eta^2} \frac{(g_k^\top d_k)^2}{||d_k||^2}. \tag{3.25}$$

Thus (3.15) holds by Lemma 3.2. The proof is complete. □

Lemma 3.4. *Suppose Assumption 1 holds. Consider the iterative form* (1.2)*, where α_k is determined by the modified Wolfe conditions* (2.23) *and d_k is given by* (2.9) *with β_{k+1} and θ_{k+1} being generated by* (2.10) *and* (2.18)*, respectively. Then, β_{k+1} has the Property (*) given in [11]:*
(1) There exists a constant $b > 1$, such that $|\beta_{k+1}| \leq b,\ \forall k \geq 0$,
(2) There exists a constant $c > 0$, such that if $||s_k|| \leq c$, then $|\beta_{k+1}| \leq \frac{1}{b},\ \forall k \geq 0$.

Proof. By the definition of β_{k+1} in (2.10), we have

$$\begin{aligned} \beta_{k+1} = \beta_{k+1}^L,&\ \text{ if } \beta_{k+1}^L \geq 0, \\ 0 > \beta_{k+1} \geq \beta_{k+1}^L,&\ \text{ if } \beta_{k+1}^L < 0. \end{aligned}$$

Hence, $|\beta_{k+1}| \leq |\beta_{k+1}^L|$. From (2.20), (3.14) and the second condition in (2.23), we have

$$d_k^\top z_k^{(m)} \geq -(1-\sigma) g_k^\top d_k \geq \eta(1-\sigma)||g_k||^2 \geq \eta(1-\sigma)\xi^2. \tag{3.26}$$

If $\mu_k > 0$, it is easy to obtain

$$\left| \frac{d_k^\top g_{k+1}}{d_k^\top z_k^{(m)}} \right| \leq \max\left\{ \frac{\sigma}{1-\sigma}, 1 \right\}. \tag{3.27}$$

If $\mu_k \le 0$, we have

$$g_{k+1}^{\mathsf{T}} d_k = d_k^{\mathsf{T}} z_k^{(m)} - (-g_k^{\mathsf{T}} d_k + t_k s_k^{\mathsf{T}} d_k) < d_k^{\mathsf{T}} z_k^{(m)}, \tag{3.28}$$

where the inequality above is due to that

$$\begin{aligned}
-g_k^{\mathsf{T}} d_k + t_k s_k^{\mathsf{T}} d_k &= \frac{1}{\alpha_k}(-g_k^{\mathsf{T}} s_k + \kappa\mu_k) \\
&= \frac{1}{\alpha_k}(-g_k^{\mathsf{T}} s_k + \kappa(2(f_k - f_{k+1}) + (g_k + g_{k+1})^{\mathsf{T}} s_k)) \\
&\ge \frac{1}{\alpha_k}(-g_k^{\mathsf{T}} s_k + \kappa(-2\rho g_k^{\mathsf{T}} s_k + g_k^{\mathsf{T}} s_k + \sigma g_k^{\mathsf{T}} s_k)) \\
&= \frac{1}{\alpha_k}(-g_k^{\mathsf{T}} s_k + (\sigma - \rho) g_k^{\mathsf{T}} s_k) > 0,
\end{aligned}$$

which by $t_k = \kappa\mu_k/\|s_k\|^2$ also implies $\kappa\mu_k/g_k^{\mathsf{T}} s_k \ge (\sigma - \rho)$. This inequality together with the second condition in (2.23) gives

$$\frac{d_k^{\mathsf{T}} g_{k+1}}{d_k^{\mathsf{T}} z_k^{(m)}} \ge \frac{\sigma - (\kappa\mu_k)/(g_k^{\mathsf{T}} s_k)}{\sigma - 1} \ge \frac{\rho}{\sigma - 1}. \tag{3.29}$$

By (3.28) and (3.29), we have

$$\left| \frac{d_k^{\mathsf{T}} g_{k+1}}{d_k^{\mathsf{T}} z_k^{(m)}} \right| \le \max\left\{ \frac{\rho}{1-\sigma}, 1 \right\}, \tag{3.30}$$

which with (3.27) implies

$$\left| \frac{d_k^{\mathsf{T}} g_{k+1}}{d_k^{\mathsf{T}} z_k^{(m)}} \right| \le \max\left\{ \frac{\sigma}{1-\sigma}, 1 \right\}, \tag{3.31}$$

Using (3.2), (3.9) and $\kappa \in (0,1)$, we obtain

$$||z_k^{(m)}|| \le ||y_k|| + \max\left\{ \kappa, \frac{m}{m-2} \right\} \frac{|\mu_k|}{||s_k||^2} ||s_k|| \le \frac{mL}{m-2} ||s_k||. \tag{3.32}$$

It can be shown that (3.32) together with (3.1), (3.3), (3.26), (3.31) and (3.32) implies

$$|\beta_{k+1}| \le \bar{c}||s_k||, \forall k \ge 0,$$

where $\bar{c} = \frac{m(m-2)\gamma L + 2\max\{\frac{\sigma}{1-\sigma}, 1\} M m^2 L^2}{(m-2)^2 \eta (1-\sigma)\xi^2}$. Let $b = \max\{1, 2\bar{c}M\},\ c = \frac{1}{b\bar{c}}$, then for all $k \ge 0$, we have

$$|\beta_{k+1}| \le b \tag{3.33}$$

and

$$||s_k|| \le c \Rightarrow |\beta_{k+1}| \le \frac{1}{b}. \tag{3.34}$$

The relations (3.33) and (3.34) indicate that β_{k+1} has the Property (*) in [11]. The proof is complete. □

Based on the above lemmas, the global convergence of NSCG can be established as follows.

Theorem 3.5. *Suppose Assumption 1 holds. Consider the iterative form* (1.2)*, where* α_k *is determined by the modified Wolfe conditions* (2.23) *and* d_k *is given by* (2.9) *with* β_{k+1} *and* θ_{k+1} *being generated by* (2.10) *and* (2.18)*, respectively. We have*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.35}$$

Proof. The proof of this theorem essentially follows the same outline of proofs of Lemma 4.2 and Theorem 4.3 in [11], we only outline its proof here.

Using Property (*) and (3.14), we can show that $||d_k||^2$ grows at most linearly. Then, we can show similarly to Lemma 4.2 in [11] that there exists $\lambda > 0$ such that, for any $\Delta \in \mathbb{N}^*$ which is the set of positive integers and any index k_0, there is a index $k \geq k_0$ such that $|\mathcal{K}_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}$, where $\mathcal{K}_{k,\Delta}^{\lambda} := \{i \in \mathbb{N}^* : k \leq i \leq k+\Delta-1, ||s_{i-1}|| > \lambda\}$ and $|\mathcal{K}_{k,\Delta}^{\lambda}|$ denotes the cardinality of $\mathcal{K}_{k,\Delta}^{\lambda}$. We proceed by contradiction. Suppose that for any $\lambda > 0$, there exsits $\Delta \in \mathbb{N}^*$ and k_0 such that for any $k \leq k_0$, we have that $|\mathcal{K}_{k,\Delta}^{\lambda}| \leq \frac{\Delta}{2}$. By induction, we obtain for any index $l \geq k_0 + 1$

$$\|d_l\|^2 \leq \zeta \left(1 + 2\beta_l^2 + 2\beta_l^2 2\beta_{l-1}^2 + \cdots + 2\beta_l^2 2\beta_{l-1}^2 \cdots 2\beta_{k_0}^2\right), \tag{3.36}$$

where ζ depends on $\|d_{k_0+1}\|^2$. Using the assumptions of contradiction, we can prove that each term on the right hand of (3.36) is bounded by 1. As a result, we have

$$\|d_l\|^2 \leq \zeta \left(l - k_0 + 2\right). \tag{3.37}$$

By (2.20), (3.8) and (3.14), we obtain

$$\xi^4 \eta^2 \sum_{k\geq 1} \frac{1}{\|d_k\|^2} \leq \eta^2 \sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty, \tag{3.38}$$

which contradicts (3.37).

Then, by Lemma 3.3 and the boundedness of $\{x_k\}$ in (3.1), we can obtain a contradiction similarly to the proof of Theorem 4.3 in [11] that $2M \geq \frac{1}{2}\sum_{i=k}^{k+\Delta-1} ||s_{i-1}|| > \frac{\Delta}{2}|\mathcal{K}_{k,\Delta}^{\lambda}| > \frac{\lambda\Delta}{4} \geq 2M$, where Δ is chosen as $\lceil \frac{8M}{\lambda} \rceil$. $\square$

4 Numerical Experiments

In this section, we show numerical comparison of the following nonlinear CG algorithms:
NSCG: Our new spectral conjugate gradient method Alg. 1;
DKCG: The algorithm proposed by Dai and Kou [7];
JSCG: The algorithm proposed by Jian et al. [14];
SCG_+: The spectral CG algorithm with standard Wolfe line search and the search direction

d_{k+1} is given by

$$
\begin{aligned}
&d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_{k+1} d_k,\ d_0 = -g_0,\\
&\beta_{k+1} = \max\left\{ \frac{g_{k+1}^{\mathsf{T}} v_k^{(m)}}{d_k^{\mathsf{T}} v_k^{(m)}} - \frac{||v_k^{(m)}||^2}{d_k^{\mathsf{T}} v_k^{(m)}} \frac{g_{k+1}^{\mathsf{T}} d_k}{d_k^{\mathsf{T}} v_k^{(m)}}, \frac{g_k^{\mathsf{T}} d_k}{||d_k||^2} \right\},\\
&\theta_{k+1} = \begin{cases} \tilde{\theta}_{k+1}, & \text{if } \tilde{\theta}_{k+1} \in [\frac{1}{4} + \eta, \tau],\\ 1, & \text{otherwise,} \end{cases}\\
&\tilde{\theta}_{k+1} = \frac{1}{g_{k+1}^{\mathsf{T}} v_k^{(m)}} (s_k^{\mathsf{T}} g_{k+1} + \beta_{k+1} d_k^{\mathsf{T}} v_k^{(m)}),\\
&v_k^{(m)} = y_k + \frac{m}{m-2} \frac{\max\{\mu_k, 0\}}{\|s_k\|^2} s_k,\ m \geq 3,\ m \in \mathbb{Z},\\
&\mu_k = 2(f_k - f_{k+1}) + (g_k + g_{k+1})^{\mathsf{T}} s_k.
\end{aligned}
$$

All experiments are coded in VC++6.0 and run on a laptop with Inter Core i5-9300H CPU, 16GB RAM memory and the Windows 10 operating system. Our testing problems are extracted from the collection of unconstrained optimization test functions [1] with problem dimensions varying from 100 to 10000. We only select the problems for which all the comparison solvers converge to the same local minimizer. Finally, a total of 85 testing problems are used in our testing problem set. The parameters $\eta = 0.001$ and $\tau = 10$ are used in NSCG and SCG$_+$. For better performance of each comparison algorithm, the parameters (ρ, σ) in the modified Wolfe line search (2.23) and in the standard Wolfe line search used in the rest algorithms are set as

$$
(\rho, \sigma) = \begin{cases} (0.18, 0.2), & \text{for NSCG,}\\ (0.1, 0.9), & \text{otherwise.} \end{cases}
$$

All the comparison algorithms are stopped when either $||g_k||_\infty \leq 10^{-8}$ or the number of iterations exceeds 10000.

We use the performance profiles of Dolan and Moré [9] to evaluate the performance of algorithms. Define $\mathcal{P}$ and $\mathcal{S}$ to be the set consisting of n_p test problems and the set of compared solvers, respectively. Define $NI_{p,s}, NF_{p,s}$ and $NG_{p,s}$, respectively, to be the number of iterations, the number of function evaluations and the number of gradient evaluations for the solver s to solve problem p to the required accuracy. Define the performance ratio as

$$
r_{p,s}^I = \frac{NI_{p,s}}{NI_p^*}, \qquad r_{p,s}^F = \frac{NF_{p,s}}{NF_p^*} \qquad \text{and} \qquad r_{p,s}^G = \frac{NG_{p,s}}{NG_p^*},
$$

where

$$
NI_p^* = \min\{NI_{p,s}, s \in \mathcal{S}\}, \quad NF_p^* = \min\{NF_{p,s}, s \in \mathcal{S}\} \quad \text{and} \quad NG_p^* = \min\{NG_{p,s}, s \in \mathcal{S}\}.
$$

It is obvious that $r_{p,s}^I, r_{p,s}^F, r_{p,s}^G \geq 1$ for all p, s. If a solver fails to solve a problem, then the ratio $r_{p,s}^I, r_{p,s}^F, r_{p,s}^G$ will set to positive infinity in our experiment. For any $\tau \geq 1$, the following cumulative distribution function, related to the performance ratio on the number of iterations, is defined as

$$
\rho_s^I(\tau) := \frac{\left|\{p \in \mathcal{P}, r_{p,s}^I \leq \tau\}\right|}{n_p}
$$

to measure the performance of each solver s with respect to the number of iterations. So, $\rho_s^I(1)$ gives the percentage of problems for which solver s takes the least number of iterations. For any $\tau \geq 1$, the cumulative distribution functions measuring the performance of solver s with respect to the number of function and gradient evaluations are similarly defined as $\rho_s^F(\tau)$ and $\rho_s^G(\tau)$, respectively.

Figure 2 presents the performance profiles of NSCG with $m = \infty, 5, 4, 3$. We observe that NSCG($m = 3$) performs best among the NSCGs with different choices of parameter m. And as the parameter m increases its performance gets worse which matches our theoretical analysis. It is also worth noticing that NSCG($m = 5$) and NSCG($m = \infty$) have nearly identical performance. Hence, we take NSCG($m = 3$) as default of NSCG in the following numerical comparisons with other algorithms.

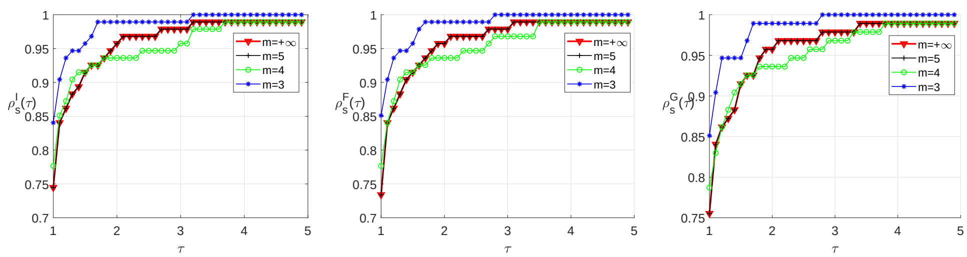

Figure 2: Performance profiles of NSCG with $m = \infty, 5, 4, 3$ on the number of iterations, function and gradient evaluations

SCG$_+$ is actually same as NSCG but with the t_k in (2.22) replaced by $t_k = \frac{m}{m-2} \frac{\max\{\mu_k,0\}}{\|s_k\|^2}$ and the modified Wolfe line search replaced by the standard Wolfe line search. In Figure 3, we can see that NSCG wins about 80%, 70% and 63%, while SCG$_+$ wins about 35%, 45% and 52%, on solving the test problems in terms of iterations, function and gradient evaluations, respectively. Hence, allowing negative μ_k combined with the modified line search (2.23) indeed improves the numerical performance of NSCG. We think the reason of improving numerical experiments might be that more negative curvature information of the Hessian can be inherited to the search direction by allowing negative μ_k.

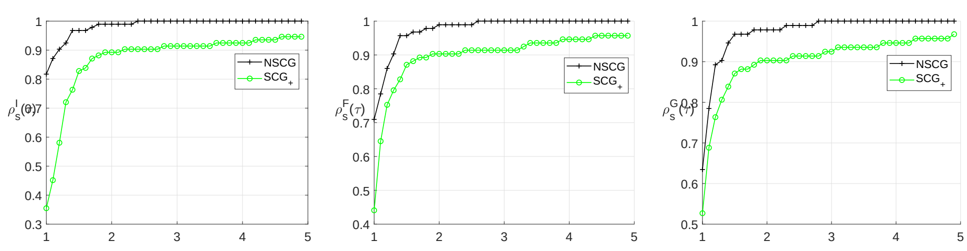

Figure 3: Performance profiles of NSCG and SCG$_+$ based on the number of iterations, function evaluations and gradient evaluations

Figure 4 shows the comparison performance profile of NSCG with the nonlinear CG solvers DKCG and JSCG. We can see that NSCG performs relatively better among these three solvers for solving this set of test problems. From Figure 4, we see that NSCG is fastest for about 75%, 60% and 55% on solving the test problems in terms of iterations, function and gradient evaluations. We believe the superior performance of NSCG is due to both the modified secant equation used by NSCG, which explores a more accurate approximation of the Hessian matrix on the direction s_k as indicated in Theorem 2.1, and the modified Wolfe line search (2.23), which could incorporate certain negative curvature information of the Hessian.

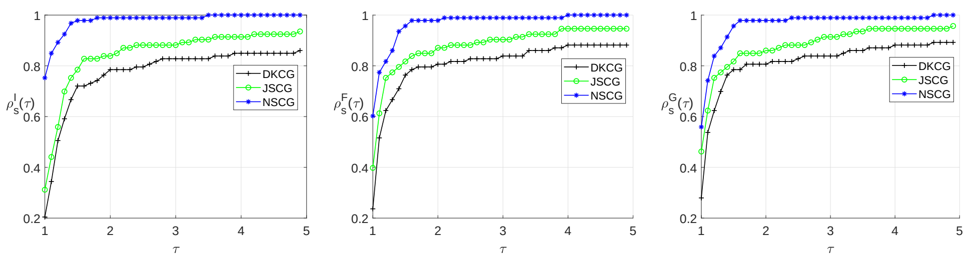

Figure 4: Performance profiles of DKCG, JSCG and NSCG based on the number of iterations, function evaluations and gradient evaluations

5 Conclusions

In this paper, based on a new family of modified secant equations and a modified Wolfe line search, we propose a new spectral conjugate gradient (NSCG) algorithm for solving unconstrained smoothing optimization problems. According to the m-th order Taylor expansion of the objective function and cubic Hermite interpolation conditions, we derive a family of modified secant equations with higher accuracy in approximation of the Hessian of the objective function. It contains a series of variants of secant conditions in the literature. To keep the negative curvature information of the Hessian, a negative value μ_k resulted from cubic Hermite interpolation is used in our new formula for generating search directions. Combined with a new modified Wolfe line search, global convergence of NSCG is established for minimizing smooth unconstrained nonlinear optimization. Our numerical experiments show NSCG($m = 3$) has the best performance in the family of the proposed algorithms NSCG($m \geq 3$). Our preliminary numerical experiments also show NSCG($m = 3$) performs better than the well-known nonlinear CG methods DKCG, SCG$_+$ and JSCG for our set of testing problems.

References

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

[2] N. Andrei, New accelerated conjugate gradient algorithms as a modification of dai–yuan's computational scheme for unconstrained optimization, *J. Comput. Appl. Math.* 234 (2010) 3397–3410.

[3] S. Babaie-Kafaki, R. Ghanbari, and N. Mahdavi-Amiri, Two new conjugate gradient methods based on modified secant equations, *J. Comput. Appl. Math.* 234 (2010) 1374–1386.

[4] J. Barzilai and J.M. Borwein, Two-point step size gradient methods, *IMA J. Numer. Anal.* 8 (1988) 141–148.

[5] F. Biglari, M.A. Hassan, and W.J. Leong, New quasi-newton methods via higher order tensor models, *J. Comput. Appl. Math.* 235 (2011) 2412–2422.

[6] E.G. Birgin and J.M. Martínez, A spectral conjugate gradient method for unconstrained optimization, *Appl. Math. Optim.* 43 (2001) 117–128.

[7] Y.H. Dai and C.X. Kou, A nonlinear conjugate gradient algorithm with an optimal property and an improved wolfe line search, *SIAM J. Optim.* 23 (2013) 296–320.

[8] Y.H. Dai and Y.X. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* 10 (1999) 177–182.

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

[10] R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964) 149–154.

[11] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* 2 (1992) 21–42.

[12] W.W. Hager and H.C. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.* 2 (2006) 35–58.

[13] M.R. Hestenes, E.Stiefel, Methods of conjugate gradients for solving linear systems, *Journal of research of the National Bureau of Standards.* 49 (1952) 409–436.

[14] J.B. Jian, Q. Chen, X.Z. Jiang, Y.F. Zeng, and J.H. Yin, A new spectral conjugate gradient method for large-scale unconstrained optimization, *Optim. Methods Softw.* 32 (2017) 503–515.

[15] Z. Khoshgam and A. Ashrafi, A new modified scaled conjugate gradient method for large-scale unconstrained optimization with non-convex objective function, *Optim. Methods Softw.* 34 (2019) 783–796.

[16] C.X. Kou, An improved nonlinear conjugate gradient method with an optimal property, *Sci. China Math.* 57 (2014) 635–648.

[17] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, part 1: theory, *J. Optim. Theory Appl.* 69 (1991) 129–137.

[18] A. Perry, A class of conjugate gradient algorithms with a two-step variable metric memory, *Technical report, Discussion paper.* 1977.

[19] A. Perry, A modified conjugate gradient algorithm, *Oper. Res.* 26 (1978) 1073–1078.

[20] B.T. Polyak, The conjugate gradient method in extremal problems, *Comput. Math. Math. Phys.* 9 (1969) 94–112.

[21] M. Raydan, The barzilai and borwein gradient method for the large scale unconstrained minimization problem, *SIAM J. Optim.* 7 (1997) 26–33.

[22] D.F. Shanno, On the convergence of a new conjugate gradient algorithm, *SIAM J. Numer. Anal.* 15 (1978) 1247–1257.

[23] Z.B. Sun, H.Y. Li, J. Wang, and Y.T. Tian, Two modified spectral conjugate gradient methods and their global convergence for unconstrained optimization, *Int. J. Comput. Math.* 95 (2018) 2082–2099.

[24] Y.X. Yuan, A modified bfgs algorithm for unconstrained optimization, *IMA J. Numer. Anal.* 11 (1991) 325–332.

[25] Y.X. Yuan and R.H. Byrd, Non-quasi-newton updates for unconstrained optimization, *J. Comput. Math.* (1995) 95–107.

[26] G. Zoutendijk, Nonlinear programming, computational methods, *Integer and nonlinear programming.* (1970) 37–86.

Manuscript received 23 August 2023
revised 12 December 2023
accepted for publication 4 February 2024

Hao Wu
School of Mathematics
Nanjing University of Aeronautics and Astronautics
Nanjing, Jiangsu, 211106, China
E-mail address: wuhoo104@nuaa.edu.cn

Liping Wang
School of Mathematics
Nanjing University of Aeronautics and Astronautics
Nanjing, Jiangsu, 211106, China
E-mail address: wlpmath@nuaa.edu.cn

Hongchao Zhang
Department of Mathematics
Louisiana State University
Baton Rouge, LA, 70803-4918, USA
E-mail address: hozhang@math.lsu.edu