A Three-Term Gradient Descent Method with Subspace Techniques

Shengwei Yao,1,2,3 Yuping Wu,1 Jielan Yang,1 and Jieqiong Xu1

1College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, China
2Guangxi (ASEAN) Financial Research Center, Guangxi University of Finance and Economics, Nanning, Guangxi, China
3Guangxi Key Laboratory Cultivation Base of Cross-Border E-commerce Intelligent Information Processing, Guangxi University of Finance and Economics, Nanning, Guangxi, China

Correspondence should be addressed to Jieqiong Xu; xjq@gxu.edu.cn

Received 1 September 2020; Revised 21 September 2020; Accepted 18 December 2020; Published 7 January 2021

Copyright © 2021 Shengwei Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We proposed a three-term gradient descent method that can be well applied to address the optimization problems in this article. The search direction of the obtained method is generated in a specific subspace. Specifically, a quadratic approximation model is applied in the process of generating the search direction. In order to reduce the amount of calculation and make the best use of the existing information, the subspace was made up of the gradient of the current and prior iteration point and the previous search direction. By using the subspace-based optimization technology, the global convergence result is established under Wolfe line search. The results of numerical experiments show that the new method is effective and robust.

1. Introduction

Gradient descent and conjugate gradient (CG) methods have profound significance for dealing with unconstrained optimization problems:

\[ \min_{x \in \mathbb{R}^n} f(x), \]  

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is a continuous-differential function. They are widely used because of the low storage requirements and the strong convergence property. Starting from an initial iteration point \( x_0 \in \mathbb{R}^n \), in each iteration, the method produces an approximate solution sequence \( \{x_k\} \) of (1):

\[ x_{k+1} = x_k + \alpha_k d_k, \quad k \geq 0, \]  

in which \( x_k \) is the current iteration point, \( \alpha_k \geq 0 \) is the step length, and \( d_k \) is the search direction that usually has the following form (for typical CG methods):

\[ d_k = \begin{cases} 
- g_k, & \text{if } k = 0, \\
- g_k + \beta_k d_{k-1}, & \text{if } k \geq 1,
\end{cases} \]  

where \( g_k = \nabla f(x_k) \) is the gradient of \( f(x) \) at \( x_k \) and \( \beta_k \) is the CG parameter. The step-size \( \alpha_k \) is usually determined by performing a specific line search, among which Wolfe line search \([1, 2]\) is one of the most commonly used:

\[ f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k g_k^T d_k, \]  

and

\[ g_k^T d_k \geq \sigma g_k^T d_k, \]  

where \( 0 < \rho \leq \sigma < 1 \). Sometimes, the strong Wolfe line search given in (4) and

\[ |g_k^T d_k| \leq \sigma |g_k^T d_k|, \]  

is also widely used in the CG method for the establishment of convergence results. Different choices of \( \beta_k \) form different versions of the CG methods. The research results of conjugate gradient algorithm are very rich, including PRP, HS, LS, and DY \([3–10]\). For a detailed survey of conjugate gradient methods, refer to \([11]\).
Recently, subspace technique has attracted more and more researchers’ attention. Various subspace techniques are applied to construct different approaches to deal with various optimization problems. For detailed description of subspace technique, refer to [12–14]. In [15], Yuan and Stoer came up with the SMCG method by using subspace technique in the conjugate gradient method. Specifically, a two-dimensional subspace \( \Omega_{k+1} = \text{Span}\{g_{k+1}, s_k\} \) is being used in [15], to determine the search direction, that is,
\[
d_{k+1} = \xi_k g_{k+1} + \psi_k s_k,\tag{7}
\]
where \( \xi_k \) and \( \psi_k \) are parameters and \( s_k = x_{k+1} - x_k \). In [16, 17], Li et al. obtained the values of \( \xi_k \) and \( \psi_k \) by minimizing the dynamically selected conic approximate model and tensor model, respectively. In order to solve nonlinear optimization problems, Conn et al. [18] brought up a class of iterated-subspace minimization methods. Similar to Yuan’s idea [15], Yang et al. [19] raised a new CG method in the same subspace \( \Omega_{k+1} \).

For better convergence and robust numerical performance, some authors pay attention to study conjugate gradient methods with different forms. Zhang et al. [20] presented a three-term conjugate gradient method, drawing on the idea of multistep quasi-Newton method. In [21], Narushima et al. proposed a specific three-term conjugate gradient method, in which the direction is formed by \( g_{k+1} \), \( y_k = g_{k+1} - g_k \), and \( d_k \). For some other three-term conjugate gradient methods, refer to [23–27].

The outline of the article is shown below. In Section 2, we use the technique of subspace minimization to derive the search direction on the subspace constituted by \( \{−g_{k+1}, s_k, g_k\} \). We elaborate the proposed algorithm in Section 3. Section 4 provides the convergence analysis of the given algorithm under some suitable conditions. In Section 5, the numerical experiments compared with other methods are presented.

2. Derivation of the Search Direction

Different from the abovementioned subspace \( \Omega_{k+1} = \text{Span}\{−g_{k+1}, s_k\} \), in this paper, the following subspace \( \Omega_{k+1} = \text{Span}\{−g_{k+1}, s_k, g_k\} \), is used to construct the search direction. In order to simplify the calculation process, for each iteration point \( x_{k+1} \), the following quadratic model is used to approximate the function \( f(x) \)
\[
\min_{d \in \Omega_{k+1}} \phi_{k+1}(d) = g_{k+1}^T d + \frac{1}{2} d^T B_{k+1} d, \tag{9}
\]
where \( B_{k+1} \) can be viewed as an approximation of Hessian matrix. Moreover, \( B_{k+1} \) is supposed to be positive definite and meet quasi-Newton equation \( B_{k+1} s_k = y_k \). Model \( \phi_{k+1}(d) \) has two advantages, which is not only a good approximation of \( f(x) \) [14] but also convenient to minimize in the subspace \( \Omega_{k+1} \). Since \( g_{k+1} \) and \( g_k \) are usually not collinear, we discuss the case of dimensions 3 and 2.

Case I. \( \text{dim}(\Omega_{k+1}) = 3 \).

In this case, the direction has the form
\[
d_{k+1} = −g_{k+1} + a_k s_k + b_k g_k, \tag{10}
\]
where \( a_k \) and \( b_k \) are parameters to be calculated.

Adding \( d_{k+1} \) to the minimizing problem (9), we get the following minimizing problem:
\[
\min_{a_k, b_k \in \mathbb{R}} \phi_{k+1}(−g_{k+1} + a_k s_k + b_k g_k), \tag{11}
\]
then \( a_k \) and \( b_k \) are the solutions of the following linear system:
\[
a_k (s_k^T y_k) + b_k (g_k^T s_k) = g_k^T y_k - g_k^T s_k, \tag{12}
a_k (g_k^T y_k) + b_k (g_k^T B_{k+1} g_k) = g_k^T B_{k+1} g_k - g_k^T g_k.
\]

To solve system (12), we need to estimate two quantities \( \eta_k = g_k^T B_{k+1} g_k \) and \( \omega_k = g_k^T + \omega_k \). Inspired by BBCG [28], we adopt
\[
\eta_k = g_k^T B_{k+1} g_k = \frac{3}{2} \frac{\|y_k\|^2 \|g_k\|^2}{s_k^T y_k}. \tag{13}
\]

Next, we choose the BFGS [29–32] update initialized with the identity matrix to compute \( \omega_k \):
\[
\omega_k = g_k^T B_{k+1} g_k = g_k^T \left( I + \frac{s_k y_k^T}{s_k^T y_k} - \frac{s_k s_k^T}{s_k^T s_k} \right) g_k
\]
\[
= g_k^T g_k + \frac{(g_k^T y_k)(g_k^T s_k) - (g_k^T s_k)(g_k^T s_k)}{s_k^T y_k} s_k^T s_k \tag{14}
\]

This choice not only retains useful information but also gives numerical results that perform better than the scaling matrix \( s_k^T y_k/s_k^T s_k \). Substituting (13) and (14) into the linear system (12), we get
\[
a_k (s_k^T y_k) + b_k (g_k^T y_k) = g_k^T y_k - g_k^T s_k, \tag{15}
a_k (g_k^T y_k) + b_k \eta_k = \omega_k - g_k^T g_k.
\]

Combining (13), the determinant of the system (15) is calculated as
\[
\Delta_k = \eta_k (s_k^T y_k) - (g_k^T y_k)^2 = \frac{3}{2} \|y_k\|^2 \|g_k\|^2 - (g_k^T y_k)^2 > 0. \tag{16}
\]
Then, $a_k$ and $b_k$ are obtained as
\[
 a_k = \frac{1}{\Delta_k} \left[ \eta_k (g_{k+1}^T y_k - g_{k+1}^T s_k) - (g_k^T y_k)(w_k - g_{k+1}^T g_k) \right],
\]
\[
 b_k = \frac{1}{\Delta_k} \left[ (s_k^T y_k)(w_k - g_{k+1}^T g_k) - (g_k^T y_k)(g_{k+1}^T y_k - g_{k+1}^T s_k) \right].
\]

Case II. $\dim(\Omega_{k+1}) = 2$.

By observing numerical experiments, we find that $\text{span}\{-g_{k+1}^T s_k\}$ produces better numerical results. In this case, we chose $\Omega_{k+1} = \text{span}\{-g_{k+1}^T s_k\}$ as the subspace. The direction is expressed as
\[
 d_{k+1} = -g_{k+1}^T + a_k s_k.
\]

Substituting (18) into (9), we get
\[
 \min_{i \in \mathbb{R}} \phi_{k+1} (-g_{k+1} + a_k s_k).
\]

Then, the solution of (19) can be expressed as
\[
 a_k = \frac{g_{k+1}^T B_{k+1} s_k - g_{k+1}^T s_k}{s_k^T B_{k+1} s_k} = \frac{g_{k+1}^T y_k - g_{k+1}^T s_k}{s_k^T y_k}.
\]

Therefore,
\[
 d_{k+1} = -g_{k+1}^T + a_k s_k = -g_{k+1}^T + \frac{g_{k+1}^T y_k - g_{k+1}^T s_k}{s_k^T y_k} s_k.
\]

If it is an exact line search, namely, $g_{k+1}^T s_k = 0$, then, from (17) and (21), we get
\[
 d_{k+1} = -g_{k+1}^T + \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k,
\]
which means $a_k = g_{k+1}^T y_k / s_k^T y_k$, $b_k = 0$, and the obtained method is degraded to the HS method (Hestenes and Stiefel) [4].

3. The Proposed TCGS Algorithm

This section is mainly to elaborate the three-term CG method with subspace techniques (TCGS), in which the acceleration scheme [33] is used (Algorithm 1).

4. Convergence Analysis

The main content of this section is to study the convergence properties of the TCGS algorithm. Some necessary assumptions for the objective function are given as follows.

Assumption 1. The level set $D = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded.

Assumption 2. Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously-differentiable, and its gradient $g(x)$ is Lipshitz continuous with the constant $L > 0$, i.e.,
\[
 \|g(x) - g(y)\| \leq L\|x - y\|, \quad x, y \in \mathbb{R}^n.
\]

Based on the above assumptions, it can be showed that there exists a constant $\Gamma > 0$ such that
\[
 \|g(x)\| \leq \Gamma, \quad x \in D.
\]

It is well known that, for optimization algorithm, the relevant properties of the search direction are very important for the convergence of the algorithm. Firstly, we will study some properties of the search direction generated by the TCGS algorithm. The following lemmas show that the direction is decreasing, and Dai–Liao conjugacy condition is satisfied.

Lemma 1. Suppose that $B_{k+1} > 0$. Then, $d_{k+1}$ generated by the TCGS algorithm is a descent direction.

Proof. According to (9), we can get $\phi_{k+1}(0) = 0$. Since $B_{k+1} > 0$ and $d_{k+1}$ is generated by the TCGS algorithm, $\phi_{k+1}(d_{k+1}) \leq 0$. Furthermore,
\[
 g_{k+1}^T d_{k+1} = -\frac{1}{2}d_{k+1}^T B_{k+1} d_{k+1} < 0.
\]

Lemma 2. The search direction $d_{k+1}$ satisfies the Dai–Liao conjugacy condition with $t = 1$, namely, $y_k^T d_{k+1} = -g_{k+1}^T s_k$.

Lemma 3. Suppose that $d_k$ is generated by the proposed TCGS algorithm and the step size $\alpha_k$ possesses the conditions (4) and (5), then
\[
 \alpha_k \geq \frac{(1 - \sigma)g_{k+1}^T d_k}{L\|d_k\|^2}.
\]

Proof. Based on condition (5) and assumption (23), we get
\[
 (\sigma - 1)g_{k+1}^T d_k \leq y_k^T d_k \leq \|y_k\|\|d_k\| \leq \alpha_k L\|d_k\|^2.
\]

Since $g_{k+1}^T d_k < 0$, $\sigma < 1$, we prove it.

Lemma 4. Assume that Assumptions 1 and 2 hold. Consider the algorithm TCGS in which Wolfe line search is used to compute the step size $\alpha_k$. Then, the Zoutendijk condition [34] holds:
\[
 \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty.
\]

Proof. Combining equation (4) and Lemma 3, we have
\[
 f(x_k) - f(x_{k+1}) \geq -\rho \alpha_k g_{k+1}^T d_k \geq \rho \frac{(1 - \sigma)g_{k+1}^T d_k^2}{L\|d_k\|^2}.
\]

Condition (29) and Assumption 1 deduce (28) directly.
Lemma 5. Suppose that the objective function \( f(x) \) satisfies Assumptions 1 and 2, consider the algorithm TCGS in which the strong Wolfe line search (4) and (6) is used to compute the step size \( \alpha_k \). If
\[
\sum_{k=1}^{\infty} \frac{1}{d_k^2} = +\infty,
\]
holds, then
\[
\lim_{k \to \infty} \inf \| g_k \| = 0. \tag{31}
\]

Theorem 1. Under the Assumptions 1 and 2, consider the sequence \( \{x_k\} \) generated by algorithm TSGS, \( \alpha_k \) satisfies the conditions (4) and (6). For uniformly convex function \( f(x) \), i.e., there exists a constant \( \mu > 0 \) such that
\[
(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu\|x - y\|^2, \quad x, y \in \mathbb{N}. \tag{32}
\]

Then, (31) holds.

Proof. From (23) and (32), we get, respectively,
\[
\| y_k \| \leq L \| s_k \| + M \| s_k \|^2. \tag{33}
\]

On the basis of (33), combined with Cauchy inequality, it follows that
\[
\mu \| s_k \|^2 \leq y_k^T s_k \leq \| y_k \| \| s_k \|, \quad \text{i.e.,} \quad \mu \| s_k \| \leq \| y_k \|. \tag{34}
\]

Furthermore, add the triangle inequality and (24), and we have
\[
\| g_{k+1} y_k - g_{k+1} s_k \| \leq \| g_{k+1} y_k \| + \| g_{k+1} s_k \|
\leq \Gamma \| s_k \| + \| s_k \|
\leq \Gamma (L + 1) \| s_k \|,
\]
\[
\| \omega_k - g_{k+1} s_k \| \leq \frac{2}{\| s_k \|^2} \| y_k^T s_k \| + \frac{2}{\| s_k \|^2} \| g_{k+1} s_k \|
\leq \frac{\| y_k^T g_{k+1} s_k \|}{\| y_k \| \| s_k \|} + \frac{\| g_{k+1} s_k \|}{\| s_k \|}
\leq \frac{\| y_k \| \| g_{k+1} \|}{\| s_k \|} + \frac{\| g_{k+1} \|}{\| s_k \|}
\leq \frac{\| y_k \| \| g_{k+1} \|}{\| s_k \|} + \| g_{k+1} \|
\leq \Gamma (L + 1) \| g_{k+1} \|. \tag{35}
\]

Next, we analyze the convergence in two cases:

Case I. \( \dim(\Omega_{k+1}) = 3 \).

\( a_k \) and \( b_k \) are computed by (16), and using (24), (33), (35), and (36), we get
\[
\| a_k \| \leq \frac{2}{\| s_k \|^2} \| y_k^T \| + \frac{2}{\| s_k \|^2} \| g_{k+1} s_k \|
\leq \frac{2}{\| s_k \|^2} \| y_k^T \| + \frac{2}{\| s_k \|^2} \| g_{k+1} s_k \|
\leq 2 \| s_k \|^2 \| y_k^T \| + \frac{2}{\| s_k \|^2} \| g_{k+1} s_k \|
\leq 2 \| s_k \|^2 \| y_k^T \| + \frac{2}{\| s_k \|^2} \| g_{k+1} s_k \|
\leq 2 \| s_k \|^2 \| y_k^T \| + \frac{2}{\| s_k \|^2} \| g_{k+1} s_k \|
\leq \frac{2}{\mu (L + 1)} \| y_k \| + \frac{2}{\mu (L + 1)} \| s_k \| \leq M \frac{1}{\| g_k \|}. \tag{36}
\]
Step 1: given $x_0 \in \mathbb{R}^n$ and $\epsilon > 0$, set $d_0 := -g_0$ and $k := 0$.
Step 2: if $\|g_k\| \leq \epsilon$ holds, then stop; else, go to step 3.
Step 3: determine the step-size $\alpha_k$ by Wolfe line search, which means that conditions (4) and (5) hold.
Step 4: generate $x_{k+1}$ by using the acceleration scheme [33], and compute $g_{k+1}$, $f(x_{k+1})$, and $s_k = x_{k+1} - x_k$.
Step 5: if $\dim(\Omega_{k+1}) = 2$, compute $d_{k+1}$ by (21), and go to step 6. If $\dim(\Omega_{k+1}) = 3$, compute $d_{k+1}$ by (10) and (17), and go to step 6.
Step 6: set $k := k + 1$, and go to step 2.

Therefore,
\[
\|d_{k+1}\| \leq \|g_{k+1}\| + |a_k| \|s_k\| + \|b_k\| \|g_k\| \\
\leq \Gamma + M_1 + M_2. 
\] (39)

Case II. $\dim(\Omega_{k+1}) = 2$.
Using (20), (24), and (33), we obtain
\[
|a_k| \leq \frac{\|g_{k+1}^T y_k\| + \|g_{k+1}^T s_k\|}{\|y_k^T s_k\|} \\
\leq \frac{\Gamma L \|s_k\| + \Gamma \|s_k\|^2}{\mu \|s_k\|} \\
= \frac{\Gamma (L + 1)}{\mu} \frac{1}{\|s_k\|} \equiv M_3 \frac{1}{\|s_k\|} 
\] (40)

Therefore,
\[
\|d_{k+1}\| \leq \|g_{k+1}\| + |a_k| \|s_k\| \\
\leq \Gamma + M_3. 
\] (41)

From Lemma 3, (31) holds.

5. Numerical Results

This section aims to observe the performance of the TCGS algorithm through numerical experiments and verify the effectiveness of the algorithm in dealing with unconstrained problems. For experimental purposes, we compared the numerical performance of TCGS with the PRP and MTHREECG [22] methods, in which MTHREECG has a structure similar to CGS, and PRP is a classic and effective CG method.

In [22], Deng and Wan presented a similar three-term conjugate gradient method (MTHREECG) with the following form:
\[
d_{k+1} = -g_{k+1} - \delta_{+k} s_k - \zeta_{+k} y_k, \\
\delta_{+k} = \left(1 - \min \left[1, \frac{\|y_k\|^2}{\|s_k\|^2} \right] \right) \frac{g_{k+1}^T s_k}{\|s_k\|^2} - \frac{g_{k+1}^T y_k}{\|s_k\|^2} \frac{g_{k+1}^T s_k}{\|s_k\|^2} - \zeta_{+k}^T y_k \\
\zeta_{+k} = \frac{g_{k+1}^T s_k}{\|s_k\|^2} - \frac{g_{k+1}^T y_k}{\|s_k\|^2} 
\] (42)

We chose a total of 75 test functions, and all the test functions come from [35]. The dimension of each function is from 1000, 2000, . . . to 10000 for the numerical experiments. The code is written by Fortran and available at https://camo.ici.ro/neculai/THREECG/threecg.for [36]. The default parameter values of the algorithm are consistent with [36].

The performance profile introduced by Dolan and Moré [37] is one of the most used tools for evaluating the performance of different methods. In this paper, we use it to investigate the numerical performance of the proposed TCGS algorithm.

The performance profile for the number of iterations can be seen in Figure 1. Looking at Figure 1, it can be found that...
the TCGS algorithm can solve about 78% of the test problems with the least iterations. With the increasing factor \( \tau \), TCGS method outperforms both MTHREECG and PRP methods. Figure 2 represents the performance profile with respect to the number of function evaluations. The result shows the similar performance with the result in Figure 1, in which TCGS also outperforms both MTHREECG and PRP methods. From the numerical experiments, it can be seen that the proposed TCGS method is efficient and robust in dealing with a set of unconstrained test problems.

6. Conclusion

By using the idea of subspace minimization, we come up with a new type of subspace gradient descent method in this paper. In our method, the subspace is spanned by \(-g_{k+1}, s_k, g_k\), and the quadratic approximation of the objective function is minimized to obtain the search direction. Therefore, the direction has the form \( d_{k+1} = -g_{k+1} + a_k s_k + b_k g_k \), where the values of \( a_k \) and \( b_k \) are computed by discussing in two cases. In addition, we prove the descent property of the direction \( d_k \) and show that Dai–Liao conjugacy condition is satisfied. Under the Wolfe line search, the convergence of the proposed TCGS algorithm is established. Observing the numerical results, the performance of the TCGS algorithm in a set of unconstrained problems is competitive.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This research was supported by the Guangxi Natural Science Foundation (nos. 2018GXNSFAA281340 and 2020GXNSFAA159014) and Program for Innovative Team of Guangxi University of Finance and Economics.

References

[1] P. Wolfe, “Convergence conditions for ascent methods,” SIAM Review, vol. 11, no. 2, pp. 226–235, 1969.
[2] P. Wolfe, “Convergence conditions for ascent methods. II: some corrections,” SIAM Review, vol. 13, no. 2, pp. 185–188, 1971.
[3] B. T. Polyak, “The conjugate gradient method in extremal problems,” USSR Computational Mathematics and Mathematical Physics, vol. 9, no. 4, pp. 94–112, 1969.
[4] M. R. Hestenes and E. Stiefel, Methods of Conjugate Gradients for Solving Linear systems, NBS, Washington, DC, USA, 1952.
[5] Y. Liu and C. Storey, “Efficient generalized conjugate gradient algorithms, part 1: Theory,” Journal of Optimization Theory and Applications, vol. 69, no. 1, pp. 129–137, 1991.
[6] Y. H. Dai and Y. Yuan, “A nonlinear conjugate gradient method with a strong global convergence property,” SIAM Journal on Optimization, vol. 10, no. 1, pp. 177–182, 1999.
[7] Z. Wei, S. Yao, and L. Liu, “The convergence properties of some new conjugate gradient methods,” Applied Mathematics and Computation, vol. 183, no. 2, pp. 1341–1350, 2006.
[8] G. Yuan, Z. Meng, and Y. Li, “A modified Hestenes and stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations,” Journal of Optimization Theory and Applications, vol. 168, no. 1, pp. 129–152, 2016.
[9] G. Yuan, Z. Wei, and X. Lu, “Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search,” Applied Mathematical Modelling, vol. 47, pp. 811–825, 2017.
[10] G. Yuan, Z. Wei, and Y. Yang, “The global convergence of the Polak-Ribiére-Polyak conjugate gradient algorithm under inexact line search for nonconvex functions,” Journal of Computational and Applied Mathematics, vol. 362, pp. 262–275, 2019.
[11] W. W. Hager and H. Zhang, “A survey of nonlinear conjugate gradient methods,” Pacific Journal of Optimization, vol. 2, no. 1, pp. 35–58, 2006.
[12] Y.-X. Yuan, “Subspace techniques for nonlinear optimization,” Series in Contemporary Applied Mathematics, Chinese Academy of Sciences, Beijing, China, pp. 206–218, 2007.
[13] Y.-X. Yuan, “Subspace methods for large scale nonlinear equations and nonlinear least squares,” Optimization and Engineering, vol. 10, no. 2, pp. 207–218, 2009.
[14] Y. Yuan, “A review on subspace methods for nonlinear optimization,” in Proceedings of the International Congress of Mathematics, pp. 807–827, Seoul, South Korea, August 2014.
[15] Y.-X. Yuan and J. Stoer, “A subspace study on conjugate gradient algorithms,” ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik, vol. 75, no. 1, pp. 69–77, 1995.
[16] Y. Li, Z. Liu, and H. Liu, “A subspace minimization conjugate gradient method based on conic model for unconstrained optimization,” Computational and Applied Mathematics, vol. 38, no. 1, p. 16, 2019.
[17] T. Wang, Z. Liu, and H. Liu, “A new subspace minimization conjugate gradient method based on tensor model for unconstrained optimization,” International Journal of Computer Mathematics, vol. 96, no. 10, pp. 1924–1942, 2019.
[18] A. Conn, N. Gould, A. Sartenaer, and P. Toint, “On iterated-subspace minimization methods for nonlinear optimization,” in Proceedings of the Linear and Nonlinear Conjugate Gradient-Related Methods, Seattle, WA, USA, January 1996.
[19] Y. Yang, Y. Chen, and Y. Lu, “A subspace conjugate gradient algorithm for large-scale unconstrained optimization,” Numerical Algorithms, vol. 76, no. 3, pp. 813–828, 2017.
[20] L. Zhang, W. Zhou, and D. Li, “Some descent three-term conjugate gradient methods and their global convergence,” Optimization Methods and Software, vol. 22, no. 4, pp. 697–711, 2007.
[21] Y. Narushima, H. Yabe, and J. A. Ford, “A three-term conjugate gradient method with sufficient descent property for unconstrained optimization,” SIAM Journal on Optimization, vol. 21, no. 1, pp. 212–230, 2011.
[22] S. Deng and Z. Wan, “A three-term conjugate gradient algorithm for large-scale unconstrained optimization problems,” Applied Numerical Mathematics, vol. 92, pp. 70–81, 2015.
[23] N. Andrei, “A new three-term conjugate gradient algorithm for unconstrained optimization,” *Numerical Algorithms*, vol. 68, no. 2, pp. 305–321, 2015.

[24] 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, pp. 72–85, 2018.

[25] H. Kobayashi, Y. Narushima, and H. Yabe, “Descent three-term conjugate gradient methods based on secant conditions for unconstrained optimization,” *Optimization Methods and Software*, vol. 32, no. 6, pp. 1313–1329, 2017.

[26] X. Y. Wang, S. J. Li, and X. P. Kou, “A self-adaptive three-term conjugate gradient method for monotone nonlinear equations with convex constraints,” *Calcolo*, vol. 53, no. 2, pp. 133–145, 2016.

[27] P. Gao, C. He, and Y. Liu, “An adaptive family of projection methods for constrained monotone nonlinear equations with applications,” *Applied Mathematics and Computation*, vol. 359, pp. 1–16, 2019.

[28] Y. Dai and C. Kou, “A Barzilai-Borwein conjugate gradient method,” *Science China Mathematics*, vol. 59, no. 8, pp. 1511–1524, 2016.

[29] C. G. Broyden, “The convergence of a class of double-rank minimization algorithms I. General considerations,” *IMA Journal of Applied Mathematics*, vol. 6, no. 1, pp. 76–90, 1970.

[30] R. Fletcher, “A new approach to variable metric algorithms,” *The Computer Journal*, vol. 13, no. 3, pp. 317–322, 1970.

[31] D. Goldfarb, “A family of variable-metric methods derived by variational means,” *Mathematics of Computation*, vol. 24, no. 109, p. 23, 1970.

[32] D. F. Shanno, “Conditioning of quasi-Newton methods for function minimization,” *Mathematics of Computation*, vol. 24, no. 111, p. 647, 1970.

[33] N. Andrei, “Acceleration of conjugate gradient algorithms for unconstrained optimization,” *Applied Mathematics and Computation*, vol. 213, no. 2, pp. 361–369, 2009.

[34] G. Zoutendijk, “Nonlinear programming, computational methods,” *Integer and Nonlinear Programming*, pp. 37–86, 1970.

[35] N. Andrei, “An unconstrained optimization test functions collection,” *Advanced Modeling and Optimization*, vol. 10, no. 1, pp. 147–161, 2008.

[36] N. Andrei, “A simple three-term conjugate gradient algorithm for unconstrained optimization,” *Journal of Computational and Applied Mathematics*, vol. 241, pp. 19–29, 2013.

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