A Three-Dimensional Subspace Algorithm Based on the Symmetry of the Approximation Model and WYL Conjugate Gradient Method

Guoxin Wang 1, Shengwei Yao 2, Mingyang Pei 1 and Jieqiong Xu 1,*

1 School of Mathematics and Information Science, Guangxi University, Nanning 530004, China; wgxin@st.gxu.edu.cn (G.W.); peimingyang@st.gxu.edu.cn (M.P.)
2 Guangxi (ASEAN) Financial Research Center, Guangxi University of Finance and Economics, Nanning 530007, China; ysw@gxu.edu.cn
* Correspondence: xjq@gxu.edu.cn

Abstract: In this paper, a three-dimensional subspace method is proposed, in which the search direction is generated by minimizing the approximation model of the objective function in a three-dimensional subspace. The approximation model of the objective function is not unique, and alternatives can be chosen between a symmetric quadratic model and a conic model by specific criteria. Moreover, the idea of a WLY conjugate gradient method is applied to characterize the change of gradient direction between adjacent iteration points. The strategy of initial stepsize and nonmonotone line search are adopted, and the global convergence of the presented algorithm is established under mild assumptions. In numerical experiments, we use a collection of 80 unconstrained optimization test problems to show the competitive performance of the presented method.

Keywords: unconstrained optimization; subspace; conic model; global convergence

1. Introduction

Considering the following unconstrained optimization problem:

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

with an initial point $x_0$, the following iterative formula is often used to solve (1):

$$x_{k+1} = x_k + \alpha_k d_k,$$  \hspace{1cm} (2)

where $x_k$ is the $k$th iteration point, $\alpha_k \in \mathbb{R}$ is the stepsize determined by a line search procedure, and $d_k$ is the search direction obtained by special ways.

The search direction of the conjugate gradient (CG) methods is obtained by

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \hspace{1cm} k = 0, 1, \ldots,$$  \hspace{1cm} (3)

where $d_0 = -g_0$, $g_{k+1} = \nabla f(x_{k+1})$ and $\beta_k \in \mathbb{R}$ is a scalar called the conjugate gradient parameter. Corresponding to different choices for the parameter $\beta_k$, various nonlinear conjugate gradient methods have been proposed. Some classical CG methods include HS method (Hestenes and Stiefel [1]), FR method (Fletcher and Reeves [2]), PRP method (Polak et al. [3]), CD method (Fletcher [4]), LS method (Liu and Storey [5]) and DY method (Dai and Yuan [6]).

Among these mentioned CG methods, the PRP, HS and LS methods share the common numerator $g_{k+1}^T y_k$, where $y_k = g_{k+1} - g_k$. When the step $x_{k+1} - x_k$ is small, the factor $y_k$ tends to zero, so $\beta_k$ becomes small and the next search direction is essentially the steepest descent direction which can avoid jamming. Hence, these methods have better...
numerical performance than those lacking in \( g_k^T y_k \). However, the PRP, HS and LS methods possess weak convergence properties. Based on the PRP method, Wei et al. [7] proposed a new formula
\[
\beta_k^{WYL} = \frac{g_k^T y_k^*}{g_k^T g_k},
\]
where \( y_k^* = g_{k+1} - \frac{\| g_{k+1} \|}{\| g_k \|} g_k \) and \( \| \cdot \| \) denotes the Euclidean norm. When step \( x_{k+1} - x_k \) is small, \( y_k^* \) also tends to zero so that the WYL method can also avoid jamming. Moreover, \( y_k^* \) can avoid the drawback: if the gradients of \( x_{k+1} \) and \( x_k \) are significantly different, the structure of \( y_k \) makes the smaller one between \( g_{k+1} \) and \( g_k \) nearly insignificant so that the method cannot make full use of the information. Furthermore, the WYL method not only has nice numerical performance but also has global convergence properties. More applications and properties of \( y_k^* \) can be found in [7–13].

The procedure for determining the stepsize \( \alpha_k \) used in (2) is usually called line search, which can be classified into an exact one and an inexact one. One of the most used inexact line search is the so-called standard Wolfe line search [14]:
\[
f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k,
\]
\[
\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k,
\]
where \( 0 < \delta < \sigma < 1 \). Obviously, it is a monotone procedure that seeks a suitable \( \alpha_k \), making the function value decrease to some extent. Zhang and Hager [15] proposed a nonmonotone version (ZH line search) that modifies condition (4) to
\[
f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k g_k^T d_k,
\]
where \( C_0 = f(x_0), Q_0 = 1, C_{k+1} \) and \( Q_{k+1} \) are updated by
\[
C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}}, \quad Q_{k+1} = \eta_k Q_k + 1,
\]
where \( \eta_k \in [\eta_{\min}, \eta_{\max}] \) and \( 0 \leq \eta_{\min} \leq \eta_{\max} \leq 1 \). The choice of \( \eta_k \) controls the degree of nonmonotonicity. Such a line search can not only overcome some drawbacks in the monotone line search, but is particularly efficient for unconstrained problems in numerical experiments [15]. In addition to ZH line search, there are many other efficient nonmonotone line search procedures, which can be found in [16–25].

The subspace technique is one of the effective means for solving large-scale optimization problems, which is receiving more and more attention. Yuan reviewed various subspace techniques that have been used in constructing numerical methods for solving nonlinear optimization problems in [26,27]. There are many optimization methods adopting the subspace technique, such as the limited-memory quasi-Newton method [28], the subspace trust region methods [29,30], the subspace SQP method [31], and more research can be found in [32–38]. Moreover, the combination between subspace technique and conjugate gradient method has been extensively studied. In the earliest research (see [39]), Yuan and Stoer computed the search direction \( d_{k+1} \) by minimizing the approximation quadratic model in the two dimensional subspace spanned by \( g_{k+1} \) and \( s_k \), namely \( \Omega_{k+1} = \text{Span}\{g_{k+1}, s_k\} \) where \( s_k = x_{k+1} - x_k \), and proposed the subspace minimization conjugate gradient method (SMCG), in which \( d_{k+1} \) is formed by
\[
d_{k+1} = \lambda g_{k+1} + \mu s_k,
\]
where \( t \) and \( \mu \) are undetermined parameters. Based on the above idea, Andrei [40] extended the subspace to \( \Omega_{k+1} = \text{Span}\{g_{k+1}, s_k, y_k\} \) and exploited the acceleration scheme, finally presenting a three-term conjugate gradient method (TTS), in which

\[
d_{k+1} = -g_{k+1} + \mu s_k + vy_k,
\]

and \( \mu, v \) are also scalar parameters. Inspired by Andrei, Yang et al. [41] changed the subspace into \( \Omega_{k+1} = \text{Span}\{g_{k+1}, s_k, s_{k-1}\} \), and put forward the subspace three-term conjugate gradient method (STT). For the same subspace, Li et al. [42] added more parameters to the computation of search direction so that

\[
d_{k+1} = tg_{k+1} + \mu s_k + vs_{k-1},
\]

and adopted the strategy of nonmonotone line search, eventually proposing the subspace minimization conjugate gradient method with nonmonotone line search (SMCG_NLS). Huo et al. [13] combined the idea of the WYL method with subspace method, then constructed the subspace \( \Omega_{k+1} = \text{Span}\{g_{k+1}, s_k, y_k^T\} \), in which

\[
d_{k+1} = tg_{k+1} + \mu s_k + vy_k^T,
\]

and finally proposed a new three-dimensional subspace conjugate gradient method (TSCG).

On the other hand, Dai and Kou [43] also focused on the analysis of Yuan and Stoer, but they paid more attention to the estimate of the parameter \( \rho_{k+1} = g_{k+1}^T B_{k+1} s_{k+1} \) during the calculation of \( d_{k+1} \). They combined the Barzilai–Borwein [44] idea and provided some efficient Barzilai–Borwein conjugate gradient methods (BBCG). It is remarkable that the idea of BCG to estimate \( \rho_{k+1} \) is employed in this paper.

It is noteworthy that all of the above mentioned subspace minimization conjugate gradient methods obtain the search direction by minimizing the approximate quadratic model of objective function in the presented subspace. However, Sun and Yuan [45,46] have pointed out that, when the current iterative point is not close to the minimizer, the quadratic model may lead to a poor prediction of the minimizer if the objective function possesses strong non-quadratic behavior. Furthermore, a quadratic model does not take into account more information instead of the gradient value in the current iteration, which means that it does not have enough degrees freedom for incorporating all of the information in the iterative procedure.

Thus, the research for approximate nonquadratic model is of the essence. Up to now, many nonquadratic models have been applied to optimization problems, such as conic model, tensor model and regularization model. The conic model can incorporate more function information than the quadratic model, and its application in unconstrained optimization was first studied by Davidon [47]. A typical conic model for unconstrained optimization is

\[
\phi_{k+1}(s) = \frac{g_{k+1}^T s}{1 + B_{k+1} s} + \frac{1}{2} \frac{s^T B_{k+1} s}{(1 + b_{k+1}^T s)^2},
\]

which is an approximation to \( f(x_k + s) - f(x_k) \), and \( B_{k+1} \) is a symmetric positive definite matrix approximating to the Hessian of \( f(x) \) at \( x_{k+1} \) which satisfies the secant equation \( B_{k+1} s_k = y_k \). The vector \( b_{k+1} \) is normally called the horizontal vector satisfying

\[
1 + b_{k+1}^T s > 0.
\]

Such a conic model has been investigated by many scholars. Sorensen [48] discussed a class of conic methods called “optimization by collinear scaling” for unconstrained optimization and shown that a particular member of this algorithm class has a Q-superlinear convergence. Ariyawansa [49] modified the procedure of Sorensen [48] and established the duality between the collinear scaling DFP and BFGS methods. Sheng [50] further discussed the interpolation properties of conic model method. Di and Sun [51] proposed a trust region method for conic models to solve unconstrained optimization problems. The trust region methods based on conic model have brought about a great number of research.
Li et al. [52] paid attention to the combination of subspace method and conic model. They considered the following conic approximation model:
\[
\phi_{k+1}(d) = \frac{g_{k+1}^T d}{1 + b_{k+1}^T d} + \frac{1}{2} \frac{d^T B_{k+1} d}{(1 + b_{k+1}^T d)^2}
\]
where
\[
b_{k+1} = -\frac{1 - \gamma_{k+1}}{\gamma_{k+1} s_k} g_{k+1},
\]
\[
\gamma_{k+1} = \frac{-s_k^T s_k}{\Delta_{k+1} + f_k - f_{k-1}},
\]
\[
\Delta_{k+1} = (f_k - f_{k+1})^2 - (g_{k+1}^T s_k)(s_k^T s_k).
\]

In the two-dimensional subspace \( \Omega_{k+1} = \text{Span}\{s_{k+1}, s_k\} \), they minimized the above conic model to compute the search direction, and finally developed a subspace minimization conjugate method based on the conic model (SMCG\_Conic). Sun et al. [53] extended the subspace to \( \Omega_{k+1} = \text{Span}\{s_{k+1}, s_k, s_{k-1}\} \) and presented a three-dimensional subspace minimization conjugate gradient method based on a conic model (CONIC\_CG3).

The adoption of a nonquadratic model does not mean abandoning the quadratic model; here we refer to the analysis of Yuan [54], in which a quantity \( u_k \) is defined by
\[
u_k = \left| \frac{2(f_k - f_{k+1} + g_{k+1}^T s_k)}{s_k^T y_k} - 1 \right|
\]
which shows the extent of how the objective function \( f(x) \) is close to a quadratic on the line segment between \( x_k \) and \( x_{k+1} \). Dai et al. [55] indicate that if the following condition
\[
u_k \leq c_1 \text{ or } \max\{u_k, u_{k-1}\} \leq c_2
\]
holds, where \( 0 < c_1 < c_2 \) are two small constants, then they think that \( f(x) \) is very close to a quadratic on the line segment between \( x_k \) and \( x_{k+1} \). If the condition (10) is satisfied, then the choice of the quadratic approximation model is preferable; otherwise, the conic model is more suitable. The utilization of such a quantity can be found in [56,57].

Since the combination of subspace technique and conic model has been investigated, we come up with the idea of whether we can make further studies. SMCG\_Conic and CONIC\_CG3 are efficient methods that possess good theoretical and numerical results, but they only considered the two-dimensional subspace \( \text{Span}\{s_{k+1}, s_k\} \) and three-dimensional subspace \( \text{Span}\{s_{k+1}, s_k, s_{k-1}\} \), respectively. Hence, we expect to extend the application of conic model in the subspace conjugate gradient method. In addition, the WYL method with the ingredient \( y_k^p \) has good performance both in theory and numerical experiment, and the subspace method TSCG based on quadratic model and subspace \( \text{Span}\{s_{k+1}, s_k, y_k^p\} \) also is an efficient method. In this paper, we investigate a new application of WYL method and subspace method based on the conic model by extending SMCG\_Conic to subspace \( \text{Span}\{s_{k+1}, s_k, y_k^p\} \). Furthermore, we adopt the initial stepsize strategy and nonmonotone line search, under which we establish the sufficient descent property and global convergence property. Since our method is based on three-dimensional subspace and conic model, it has efficient performance in numerical experiment and can solve more problems than subspace method based on two-dimensional subspace and quadratic model.

This paper is organized as follows: in Section 2, the search directions on the subspace \( \Omega_{k+1} \) based on two different models are derived, and the criteria for how to choose the approximate model and search direction are presented. In Section 3, we obtain the stepsize by the strategies of initial stepsize and nonmonotone line search, and elaborate the generated algorithm. In Section 4, we give the proofs for some important lemmas of the
search direction, then state the convergence performance of the generated algorithm under suitable assumptions. In Section 5, we compare the numerical results of our algorithm with another two methods.

2. The Search Direction

For an optimization method, the search direction attaches great importance to the iterative formula, and the computation of search direction is always the first job. Hence, the main content of this section is to construct the formula for the search direction. Since the approximation model is selected between conic model and quadratic model, the computation will be divided into two parts.

2.1. Conic Model

In this subsection, we will give the computation of the search direction in the case that conic model is used.

When \( (10) \) is violated, the conic model is more suitable to approximate the objective function, so we compute the search direction by minimizing the conic model in subspace \( \Omega_{k+1} = \text{Span}\{g_{k+1}, s_k, y_k^s\} \). We consider the subproblem

\[
\min_{d \in \Omega_{k+1}} \phi_{k+1}(d),
\]

where \( \phi_{k+1}(d) \) is the same as \( (9) \).

The discussion will be divided into three parts depending on the three different dimensions of the subspace \( \Omega_{k+1} \).

**Situation 1:** \( \text{dim}(\Omega_{k+1}) = 3 \).

In this situation, the search direction is

\[
d_{k+1} = t g_{k+1} + \mu s_k + v y_k^s.
\]

By substituting \( (12) \) into \( (11) \) and using the secant equation, the problem \( (11) \) turns into

\[
\min_{(t, \mu, v)} \phi_{k+1}(u) = \frac{a^T u}{1 + c^T u} + \frac{1}{2} \frac{u^T A_{k+1} u}{(1 + c^T u)^2},
\]

where

\[
A_{k+1} = \begin{pmatrix}
\rho_{k+1} & g_{k+1}^T & \omega_k \\
g_{k+1} & s_k^T & y_k^s \\
\omega_k & y_k^s & y_k^s
\end{pmatrix},
\]

\[
a = \begin{pmatrix}
\|g_{k+1}\|^2 \\
\|g_{k+1} - s_k\|^2 \\
\|s_k^T y_k^s\|^2
\end{pmatrix},
\]

\[
c = \begin{pmatrix}
b_{k+1}^T s_k \\
b_{k+1}^T y_k^s \\
b_{k+1}^T y_k^s
\end{pmatrix},
\]

\[
u = \begin{pmatrix}
t \\
\mu \\
v
\end{pmatrix},
\]

and \( \rho_{k+1} = g_{k+1}^T b_{k+1} s_k, \omega_k = g_{k+1}^T b_{k+1} y_k^s, \tau_k = y_k^s b_{k+1} y_k^s \).

To solve \( (13) \), we first figure out the zero solution of the function

\[
\nabla \phi_{k+1}(u) = \frac{1}{1 + c^T u} \left( I - \frac{c u^T}{1 + c^T u} \right) \left( a + A_{k+1} u \right),
\]

and obviously \( I - \frac{c u^T}{1 + c^T u} \) is invertible, so the problem is reduced to \( a + A_{k+1} u = 0 \). After calculation, we can deduce that, when \( A_{k+1} \) is positive definite and \( 1 + c^T A_{k+1} a \neq 0 \), the solution of \( (13) \) is

\[
u_{k+1} = -\frac{A_{k+1}^{-1} a}{1 + c^T A_{k+1}^{-1} a}.
\]
By using the relationship \( A_{k+1}^{-1} = \frac{A_{k+1}^{-1}}{|A_{k+1}|} \), where

\[
A_{k+1}^* = \begin{pmatrix} X & \theta_1 & \theta_2 \\ \theta_1 & \theta & \theta_3 \\ \theta_2 & \theta_3 & Y \end{pmatrix}
\] (15)

is the adjugate matrix of \( A_{k+1} \), and \( X, \theta_1, \theta_2, \theta_3 \), \( Y \) denote the cofactor of \( A_{k+1}^{11}, A_{k+1}^{21}, A_{k+1}^{31}, A_{k+1}^{22}, A_{k+1}^{32}, A_{k+1}^{33} \), respectively, in which \( A_{k+1}^{ij} \) denotes the \( ij \)-th element of \( A_{k+1} \) and

\[
X = (s_{kj}^T y_k) \tau_k - (y_k^T y_k')^2,
\]

\[
\theta_1 = (y_k^T y_k') \omega_k - (g_{k+1}^T y_k) \tau_k,
\]

\[
\theta_2 = (s_{k+1,j}^T y_k') (y_k^T y_k') - (s_{k}^T y_k) \omega_k,
\]

\[
\theta = \rho_{k+1} \tau_k - \omega_k^2,
\]

\[
\theta_3 = (g_{k+1}^T y_k) \omega_k - \rho_{k+1} (y_k^T y_k'),
\]

\[
Y = \rho_{k+1} (s_k^T y_k') - (s_k^T y_k)^2,
\]

then we finally obtain the solution of (13),

\[
u_{k+1} = \begin{pmatrix} \nu_{k+1} \\ h_{k+1} \\ v_{k+1} \end{pmatrix} = -\frac{1}{D_{k+1}} \begin{pmatrix} X & \theta_1 & \theta_2 \\ \theta_1 & \theta & \theta_3 \\ \theta_2 & \theta_3 & Y \end{pmatrix} \begin{pmatrix} ||g_{k+1}||^2 \\ s_{k+1}^T s_k \\ s_{k+1}^T y_k \end{pmatrix} = \frac{1}{D_{k+1}} \begin{pmatrix} -q_1 \\ -q_2 \\ -q_3 \end{pmatrix},
\] (16)

where

\[
|A_{k+1}| = \rho_{k+1} X + \theta_1 s_{k+1}^T y_k + \theta_2 \omega_k,
\]

\[
D_{k+1} = |A_{k+1}| + q_1 b_{k+1}^T g_{k+1} + q_2 b_{k+1}^T s_k + q_3 b_{k+1}^T y_k,
\]

\[
q_1 = X ||g_{k+1}||^2 + \theta_1 s_{k+1}^T s_k + \theta_2 s_{k+1}^T y_k,
\]

\[
q_2 = \theta_1 ||s_{k+1}||^2 + \theta s_{k+1}^T s_k + \theta_3 s_{k+1}^T y_k,
\]

\[
q_3 = \theta_2 ||s_{k+1}||^2 + \theta_3 s_{k+1}^T s_k + Y s_{k+1}^T y_k.
\]

So the solution of (11) in the three-dimensional subspace \( \Omega_{k+1} = \text{Span} \{ s_{k+1}, s_k, y_k^* \} \) is

\[
d_{k+1} = (s_{k+1} \quad s_k \quad y_k^*) \nu_{k+1}.
\]

During evaluation, it is important to find appropriate ways to estimate \( \rho_{k+1}, \omega_k \) and \( \tau_k \) in order to avoid matrix–vector multiplication and improve efficiency. Yuan and Stoer [39] proposed two ways to calculate such quantities containing \( B_{k+1} \), one of which is to obtain \( B_{k+1} \) by using the scaled memoryless BFGS formula. The approach of Dai and Kou [43] is to combine the Barzilai–Borwein [44] idea by approximating the Hessian by \( (1/\alpha_{k+1}^{BB_1}) I \) or \( (1/\alpha_{k+1}^{BB_2}) I \), where

\[
\alpha_{k+1}^{BB_1} = ||s_k||^2, \quad \alpha_{k+1}^{BB_2} = ||y_k||^2.
\]

For \( \omega_k \), we utilize the memoryless BFGS formula to obtain \( B_{k+1} \) so that

\[
\omega_k = s_k^T Y_k \left( I + \frac{y_k y_k^T}{s_k y_k} - \frac{s_k s_k^T}{s_k^T s_k} \right) y_k^*
\]

\[
= s_k^T Y_k^* + \frac{s_{k+1} s_k y_k^* y_k^*}{s_k^T y_k} - \frac{s_k s_k^T y_k^* y_k^*}{s_k^T s_k},
\] (17)
Then for $\tau_k$, we combine the idea of [42,43], and estimate $\tau_k$ by

$$
\tau_k = \zeta_k \frac{\|y_k\|^2}{s_k^2 y_k} \|y_k^*\|^2,
$$

where

$$
\zeta_k = \begin{cases} 
\max\{0.9\zeta_{k-1}, 1.2\}, & \text{if } \alpha_k > 1.0, \\
\min\{1.1\zeta_{k-1}, 1.75\}, & \text{otherwise},
\end{cases}
$$

and $\zeta_0 = 1.5$. It is obvious that $\zeta_k \in [1.2, 1.75]$. For the estimate of $\rho_k+1$, we adopt the idea of Li et al. [42], because it can guarantee some good properties. Firstly, according to the analysis regarding (14), the positive definitiveness of $A_{k+1}$ is essential, which requires $\|A_{k+1}\| > 0$. It follows that

$$
\rho_{k+1} > \frac{-\theta_1 g_{k+1}^T y_k - \theta_2 y_k}{X}.
$$

By setting $X = m_k s_k^T y_k \tau_k$ with $m_k \triangleq 1 - \frac{(y_k^T y_k)^2}{s_k^T y_k}$, and combining (15), we have

$$
\rho_{k+1} > \left[ \frac{(s_{k+1}^T y_k)^2}{s_k^T y_k} + \frac{\omega_k^2}{\tau_k} - 2 \frac{g_{k+1}^T y_k y_k^T y_k^*}{s_k^T y_k \tau_k} \right] / m_k \triangleq n_k,
$$

if and only if $m_k$ is positive, which can be guaranteed by (18). In addition, the positive definitiveness of $A_{k+1}$ also requires that its first and second order leading principal minors are positive, and it means

$$
\rho_{k+1} > \frac{(s_{k+1}^T y_k)^2}{s_k^T y_k}.
$$

Secondly, $D_{k+1} > 0$ is also a necessary condition to keep the sufficient descent property, from which follows

$$
\rho_{k+1} > \frac{S_k}{s_k^T y_k \tau_k} / M_k,
$$

if $M_k > 0$, where

$$
S_k = -\theta_1 s_{k+1}^T y_k - \theta_2 y_k - b_{k+1}^T s_{k+1} \left( X \|s_{k+1}\|^2 + \theta_1 s_{k+1}^T s_k + \theta_2 s_{k+1}^T y_k^* \right) - b_{k+1}^T \left( \theta_s \|s_{k+1}\|^2 - \omega_k s_{k+1}^T y_k s_{k+1}^T y_k^* \right) - b_{k+1}^T y_k^* \left( \theta_s \|s_{k+1}\|^2 + \omega_k s_{k+1}^T y_k s_{k+1}^T y_k^* - (s_{k+1}^T y_k)^2 s_{k+1}^T y_k^* \right),
$$

and

$$
M_k = 1 - \frac{(y_k^T y_k^*)^2}{s_k^T y_k \tau_k} + \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \left[ 2 \frac{s_{k+1}^T y_k y_k^*}{s_k^T y_k \tau_k} + \frac{g_{k+1}^T s_k}{s_k^T y_k \tau_k} - \frac{(s_{k+1}^T y_k)^2}{s_k^T y_k \tau_k} \right].
$$

Likewise, we define the right-hand side of (22) as $N_k$, i.e. $N_k \triangleq \frac{S_k}{s_k^T y_k} / M_k$. After the above discussion, we can estimate $\rho_{k+1}$ as follows,

$$
\rho_{k+1} = \zeta_k \max\{K, N_k, n_k\},
$$

where $K = K_1 \|s_{k+1}\|^2$,

$$
K_1 = \max\left\{ \frac{\|y_k\|^2}{s_k^T y_k}, \left| \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \frac{\|s_{k+1}\|^2}{\|s_{k+1}\|^2} \right| \right\},
$$

and $\zeta_k$ is the same as that in (18). One of the purposes of $K$ is to ensure (21).
Before computing $d_{k+1}$ by (12) and (16), we should verify the following conditions:

\[ \Delta_{k+1} \geq 0, \]
\[ M_k \geq \rho_0, \]  
\[ \xi_1 \leq \frac{s_k^T y_k}{\|s_k\|^2} \leq \frac{\|y_k\|^2}{s_k^T y_k} \leq \xi_2, \]  
\[ \left| \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \right| \frac{\|s_{k+1}\|^2}{\|s_{k+1}^T s_k\|^2} \leq \xi_3, \]  
\[ \frac{|A_{k+1}|}{s_k^T y_k \bar{a}_{k+1}} \geq \xi_5, \]  

where $\rho_0 \in (0, 1)$ and $\xi_1, \xi_2, \xi_3, \xi_5$ are positive constants. The expressions (24) and (25) are fundamental premises of the conic model (5) and relation (22), respectively. On the basis of the Barzilai–Borwein [44] idea, (26) might indicate the suitable condition numbers of the approximation Hessian matrix. Moreover, (27) is vital to guarantee the descent property of the search direction. As for (28), obviously it makes $A_{k+1}$ more positive definite, and is also important for establishing the sufficient descent property of the search direction.

For convenience, we call (24)–(28) the first conic model conditions. Hence, if the first conic model conditions hold, then we compute the search direction by (12) and (16).

**Situation 2:** $\dim(\Omega_{k+1}) = 2$ or 1.

Li et al. [52] have made a deep study of the subspace conjugate gradient method based on the conic model in this case. Here we refer to their works. When $\dim(\Omega_{k+1}) = 2$, the search direction is formed by

\[ d_{k+1} = t g_{k+1} + \mu s_k, \]

which is the same as (8). The solution of $t$ and $\mu$ is

\[ \begin{pmatrix} t_{k+1} \\ \mu_{k+1} \end{pmatrix} = \frac{1}{D_{k+1}} \begin{pmatrix} -c_1 \\ -c_2 \end{pmatrix}, \]

which is the same as (13) in [52].

In this case, conditions (24) and (26) are still essential, but before obtaining $d_{k+1}$ by (8) and (29), the following conditions are also necessary:

\[ m_k \geq \bar{\rho}_0, \]
\[ \left| \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \right| \frac{\|s_{k+1}\|^2}{\|s_{k+1}^T s_k\|^2} \leq \xi_3, \text{ if } \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} S_{k+1} s_k < 0, \]
\[ \frac{\|s_{k+1}\|^2}{\|s_k\|^2} \leq \xi_4, \text{ if } \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} S_{k+1} s_k \geq 0, \]

where $\bar{\rho}_0 \in (0, 1)$, $\xi_4$ is a positive constant and $\xi_3$ is identical to that in (27). We call (24), (26) and (30)–(32) the second conic model conditions. If the first conic model conditions do not all hold but the second conic model conditions hold, we compute the search direction by (8) and (29).

If the first and second conic model conditions are violated but (33) and (34) hold,

\[ \bar{\theta}_1 \leq \frac{s_k^T y_k}{\|s_k\|^2}, \]
\[ \frac{|s_{k+1}^T y_k s_{k+1}^T d_k|}{d_k^T y_k \|s_{k+1}\|^2} \leq \theta_4, \]
where $\vartheta_4 \in [0, 1)$ and $\vartheta_1$ are positive constants, the HS direction is considered, i.e.,
\[ d_{k+1} = -s_{k+1} + \beta_{k}^{HS} d_k. \] (35)

We call (33) and (34) the HS conditions.

Therefore, there are two choices of search direction when $\dim(\Omega_{k+1}) = 2$: one is to compute $d_{k+1}$ by (8) and (29) if the second conic model conditions hold; the other is to compute $d_{k+1}$ by (35) if the HS conditions hold. Otherwise, we use the negative gradient direction $-s_{k+1}$ as our search direction.

2.2. Quadratic Model

The content of this subsection is to calculate the search direction in the case that the objective function is approximated by a quadratic model.

When the condition (10) holds, the quadratic approximation model seems to perform well in approximating the objective function, so we consider the subproblem based on the quadratic approximation model as follows:
\[ \min_{d \in \Omega_{k+1}} \psi_{k+1}(d) = g_{k+1}^T d + \frac{1}{2} d^T B_{k+1} d, \] (36)
and related work can be found in [13]. The corresponding results can be applied to this situation. We state the concrete results in the following.

Firstly, there are some necessary conditions:
\[ \vartheta_1 \leq s_k^T y_k \|s_k\|^2 \leq \|y_k\|^2 \leq \vartheta_2, \] (37)
\[ \vartheta_1 \leq \frac{\tau_k}{\|y_k\|^2}, \quad 4 \|y_k\|^4 \|y_k^*\|^2 \leq \vartheta_2, \] (38)
\[ \|s_k\|^2 \geq \vartheta_3, \] (39)
we call (37)–(39) the quadratic model conditions.

When the quadratic model conditions hold,
\[ d_{k+1} = t_{k+1}^1 s_{k+1} + \mu_{k+1}^1 s_k + \nu_{k+1}^1 y_k^*, \] (40)
where
\[
\begin{pmatrix}
  t_{k+1}^1 \\
  \mu_{k+1}^1 \\
  \nu_{k+1}^1
\end{pmatrix}
= -\frac{1}{|A_{k+1}|} \begin{pmatrix}
  X & \vartheta_1 & \vartheta_2 \\
  \vartheta_1 & \vartheta & \vartheta_3 \\
  \vartheta_2 & \vartheta_3 & \vartheta
\end{pmatrix} \begin{pmatrix}
  \|s_{k+1}\|^2 \\
  \|s_{k+1}^T y_k\| \\
  \|y_k\|
\end{pmatrix}.
\]

When the quadratic model conditions do not all hold but (37) holds,
\[ d_{k+1} = t_{k+1}^2 s_{k+1} + \mu_{k+1}^2 s_k, \] (41)
where
\[
\begin{pmatrix}
  t_{k+1}^2 \\
  \mu_{k+1}^2
\end{pmatrix}
= \frac{1}{|A_{k+1}|} \begin{pmatrix}
  g_{k+1}^T y_k g_{k+1}^T y_k - \|s_{k+1}\|^2 s_k^T y_k \\
  \|s_{k+1}\|^2 s_k^T y_k - \rho_{k+1} s_{k+1}^T y_k
\end{pmatrix}.
\]

$A_{k+1}$ is
\[
A_{k+1} = \begin{pmatrix}
  \rho_{k+1} & s_{k+1}^T y_k \\
  s_{k+1}^T y_k & s_k^T y_k
\end{pmatrix}.
\]
When (37) does not holds but the HS conditions hold,

\[ d_{k+1} = -s_{k+1} + \beta_{HS}^k d_k. \]

Otherwise, \( d_{k+1} = -s_{k+1}. \)

To sum up, the generated directions possess several forms as follows: when the approximation quadratic model is selected, i.e. (10) is violated,

- \( d_{k+1} \) is calculated by (12) and (16), if the first conic model conditions hold,
- \( d_{k+1} \) is calculated by (8) and (29), if the second conic model conditions hold,
- \( d_{k+1} \) is calculated by (35), if the HS conditions hold;
when the approximation quadratic model is selected, i.e. (10) holds,

- \( d_{k+1} \) is calculated by (12) and (40), if the quadratic model conditions hold,
- \( d_{k+1} \) is calculated by (8) and (41), if only (37) holds,
- \( d_{k+1} \) is calculated by (35), if the HS conditions hold;
otherwise, \( d_{k+1} = -s_{k+1}. \)

3. The Stepsize and Algorithm

In this section, we will present the obtainment of another essential ingredient in the iterative formula of optimization method, i.e. stepsize. It requires two steps: the first is the choice of initial stepsize, and the second is the line search procedure. After these discussion, then we can propose our whole algorithm.

3.1. Strategy for the Initial Stepsize

Firstly, we show how to choose the initial stepsize.

It is generally acknowledged that the initial stepsize is of great significance for an optimization method, especially for the conjugate gradient method. For Newton and quasi-Newton methods, the choice of the initial trial stepsize may always be unit step \( \alpha^0 = 1. \) However, it is different for methods that do not produce well-scaled search directions, such as the steepest descent or the conjugate gradient methods. Thus, it is significant to make a reasonable initial guess of the stepsize by considering the current information about the objective function and algorithm for such methods [58, 59]. Many strategies of the initial stepsize have been proposed which can be found in the achievements of Hager and Zhang [60], Nocedal and Wright [58], Dai and Kou [19], Liu and Liu [57].

In the strategy of [19], Dai and Kou presented a condition,

\[
\left| \varphi_{k+1}(a_{k+1}^0) - \varphi_{k+1}(0) \right| \leq \varepsilon_2,
\]

where \( \varepsilon_1, \varepsilon_2 \) are small positive constants, \( a_{k+1}^0 \) denotes the initial trial stepsize, and \( \varphi_{k+1}(a) = f(x_{k+1} + ad_{k+1}). \) If condition (42) holds, it implies that the points \( x_{k+1} + a_{k+1}^0 d_{k+1} \) and \( x_{k+1} \) are not far away from each other, so it is reasonable to use the minimizer of the quadratic interpolation function for \( \varphi_{k+1}(0), \varphi'_{k+1}(0) \) and \( \varphi_{k+1}(a_{k+1}^0) \) as the new initial stepsize, where \( \varphi'_{k+1}(0) \) denotes the first derivative of \( \varphi_{k+1}(0). \)

In this paper, the selection of the initial stepsize has two parts, depending on whether the negative gradient direction is adopted or not, and is presented via modification of that in Liu and Liu [56].

When the search direction is negative gradient direction, according to the analysis of Liu and Liu [57], it is desirable to take the initial trial stepsize by

\[
\bar{s}_{k+1} = \begin{cases} 
\max \{ \min \{ \frac{s_k y_k}{\|y_k\|^2}, \lambda_{\max} \}, \lambda_{\min} \} & \text{if } s_{k+1}^T s_k > 0, \\
\max \{ \min \{ \|s_k\|^2 / s_k^T y_k, \lambda_{\max} \}, \lambda_{\min} \} & \text{if } s_{k+1}^T s_k \leq 0,
\end{cases}
\]
where \( \lambda_{\text{min}} \) and \( \lambda_{\text{max}} \) are two positive constants controlling the initial stepsize within the interval \([\lambda_{\text{min}}, \lambda_{\text{max}}]\) which is preferable in numerical experiments.

Andrei [61] thinks that, the more accurate the step length is, the faster convergence a conjugate gradient algorithm has. It therefore makes sense to verify if the initial trial stepsize \( \hat{\alpha}_{k+1} \) satisfies (42) or not. If so and \( d_k \neq -g_k, \|g_{k+1}\|^2 \leq 1 \), we update the initial stepsize by

\[
\hat{\alpha}_{k+1} = \max\{\min\{\hat{\alpha}_{k+1}, \lambda_{\text{max}}\}, \lambda_{\text{min}}\},
\]

where

\[
\hat{\alpha}_{k+1} = \min q(q_{k+1}(0), q'_{k+1}(0), q_{k+1}(\hat{\alpha}_{k+1})),
\]

where \( q(q_{k+1}(0), q'_{k+1}(0), q_{k+1}(\hat{\alpha}_{k+1})) \) denotes the quadratic interpolation function for \( q_{k+1}(0), q_{k+1}(0) \) and \( q_{k+1}(\hat{\alpha}_{k+1}) \).

Therefore, the initial stepsize for the negative gradient direction is

\[
\alpha^0_{k+1} = \begin{cases} 
\hat{\alpha}_{k+1}, & \text{if (42) holds, } d_k \neq -g_k, \|g_{k+1}\|^2 < 1 \text{ and } \hat{\alpha}_{k+1} > 0, \\
\hat{\alpha}_{k+1}, & \text{otherwise.}
\end{cases}
\] (44)

When it comes to the search direction that is not negative gradient direction, the similarity of the calculation between our direction and that in quasi-Newton methods implies that the unit stepsize \( \alpha^0_{k+1} = 1 \) might be a reasonable initial trial stepsize. Again we figure out the minimizer of the quadratic interpolation function

\[
\hat{\alpha}_{k+1} = \min q(q_{k+1}(0), q'_{k+1}(0), q_{k+1}(1)).
\]

If \( \alpha^0_{k+1} = 1 \) satisfies (42) and \( \hat{\alpha}_{k+1} > 0 \), we update the initial stepsize by

\[
\hat{\alpha}_{k+1} = \max\{\min\{\hat{\alpha}_{k+1}, \lambda_{\text{max}}\}, \lambda_{\text{min}}\}.
\]

Therefore, the initial stepsize for the search direction except negative gradient direction is

\[
\alpha^0_{k+1} = \begin{cases} 
\hat{\alpha}_{k+1}, & \text{if (42) holds and } \hat{\alpha}_{k+1} > 0, \\
1, & \text{otherwise.}
\end{cases}
\] (45)

3.2. The Nonmonotone Line Search

Secondly, we introduce the line search used in our algorithm.

For the variable \( \eta_k \) in ZH line search, Zhang and Hager [15] proved that if \( \eta_{\text{max}} = 1 \), then the generated sequence \( \{x_k\} \) only has the property that

\[
\liminf_{k \to \infty} \|g_k\| = 0.
\]

Liu and Liu [57] presented a formula of \( \eta_k \),

\[
\eta_k = \begin{cases} 
c, & \text{mod} (k, n) = n - 1, \\
1, & \text{mod} (k, n) \neq n - 1,
\end{cases}
\]

where \( 0 < c < 1 \) and \( \text{mod}(k, n) \) denotes the residue for \( k \) modulo \( n \), and it resulted in a better convergence.

Referring to the above study, Li et al. [52] and Sun et al. [53] made some modifications so that the improved line search is more appropriate for their algorithms.
In order to gain the decent convergence result and performance, this paper modifies the improved ZH line search used in [53]. To be specific, we set

\[
C_{k+1} = \begin{cases} 
    f_k + \min \{1, \ 0.9(C_k - f_k)\}, & k < 5, \\
    \eta_k Q_k C_k + f(x_{k+1})/Q_{k+1}, & k \geq 5,
\end{cases} 
\]

where

\[
Q_{k+1} = \begin{cases} 
    6, & k = 4, \\
    \eta_k Q_k + 1, & k \geq 5,
\end{cases} 
\]

The details of the three-dimensional subspace conjugate gradient method based on conic model are given in the next algorithm, which is Algorithm 1.

**Algorithm 1 WYL_TSCO**

**Require**: initial point \( x_0 \), initial steps as \( a_0 \), positive constants \( \epsilon, \epsilon_1, \epsilon_2, 0 < \delta < \sigma < 1 \), \( \xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \theta_1, \theta_2, \theta_3, \theta_4, \theta_6, \rho_0 \in (0, 1), \lambda_1, \lambda_2, \lambda_{\text{min}}, \lambda_{\text{max}} \\

**Ensure**: optimal \( x^* \)

1. Set MaxRestart: = 4n, IterRestart: = 0, IterQuad: = 0, NumNongrad: = 0, \( C_0 = f_0 \), \( d_0 = -g_0 \) and \( k = 0 \).
2. If \( \|g_0\|_{\infty} \leq \epsilon \), stop.
3. Calculate the steps size \( a_k \) by (5) and (6) with \( a_0 \).
4. Update \( x_{k+1} = x_k + a_k d_k \). If \( \|g_{k+1}\|_{\infty} \leq \epsilon \), stop; otherwise, let IterRestart: = IterRestart + 1. If \( |r_k - 1| \leq 10^{-8} \) or \( |f_{k+1} - f_k| - 0.5|g_{k+1}^T s_k + g_k^T s_k| \leq 6.0 \times 10^{-8} \), IterQuad: = IterQuad + 1; else, IterQuad: = 0.
5. Calculate the search direction \( d_k \). If NumNongrad = MaxRestart, or (IterQuad = MinQuad and IterRestart \( \neq \) IterQuad), go to 5.6; else if the condition (10) holds, go to 5.1; else, go to 5.3.

5.1. If the quadratic model conditions hold, compute \( d_{k+1} \) by (12) and (40), set NumNongrad: = NumNongrad + 1; else, go to 5.2.
5.2. If condition (37) holds, compute \( d_{k+1} \) by (8) and (41), set NumNongrad: = NumNongrad + 1; else, go to 5.5.
5.3. If the first conic model conditions hold, compute \( d_{k+1} \) by (12) and (16), set NumNongrad: = NumNongrad + 1; else, go to 5.4.
5.4. If the second conic model conditions hold, compute \( d_{k+1} \) by (8) and (29), set NumNongrad: = NumNongrad + 1; else, go to 5.5.
5.5. If the HS conditions hold, compute \( d_{k+1} \) by (35), set NumNongrad: = NumNongrad + 1; else, go to 5.6.
5.6. Compute \( d_{k+1} = -g_{k+1} \), set NumNongrad: = 0 and IterRestart: = 0.
6. If \( d_{k+1} = -g_{k+1} \), calculate \( a_{k+1}^* \) by (44); otherwise, calculate \( a_{k+1}^* \) by (45).
7. Update \( Q_{k+1} \) and \( C_{k+1} \) by (46)–(48).
8. Set \( k := k + 1 \), go to Step 3.
4. Properties of the Proposed Algorithm

The theoretical nature is an important criterion to determine the performance of an algorithm. In this section, we will give some properties of WYL_TSCO, including the sufficient descent property and the global convergence.

4.1. Some Lemmas

This subsection will make a deep discussion on our algorithm and prove some properties of the presented directions before establishing the global convergence. At first, we propose two assumptions as follows:

Assumption 1. The objective function \( f(x) \) is continuously differentiable and bounded from below on \( \mathbb{R}^n \).

Assumption 2. The gradient function \( g(x) \) is Lipschitz continuous on the level set \( D = \{ x \in \mathbb{R}^n : f(x) \leq f(x_0) \} \), which means that there exists a positive constant \( L > 0 \) satisfying

\[
\| g(x) - g(y) \| \leq L \| x - y \|, \quad \forall x, y \in D,
\]

which implies that \( \| y_k \| \leq L \| s_k \| \).

With these assumptions, we can prove some properties of the new directions in the following.

**Lemma 1.** For the search direction \( d_{k+1} \) calculated by WYL_TSCO, there exists a constant \( \varsigma_1 > 0 \) such that

\[
\mathcal{G}^T_{k+1}d_{k+1} \leq -\varsigma_1 \| \mathcal{G}_{k+1} \|^2. \tag{49}
\]

**Proof.** We will discuss in several parts based on different situations and different approximation models.

**Case I:** if the negative direction is adopted, i.e. \( d_{k+1} = -\mathcal{G}_{k+1} \), then

\[
\mathcal{G}^T_{k+1}d_{k+1} = -\| \mathcal{G}_{k+1} \|^2 \leq -\frac{1}{2} \| \mathcal{G}_{k+1} \|^2.
\]

**Case II:** if \( d_{k+1} \) is determined by (35), i.e., the HS direction, combining (34), then we have

\[
\mathcal{G}^T_{k+1}d_{k+1} = -\| \mathcal{G}_{k+1} \|^2 + \beta_{k+1}^H \frac{d_{k+1}^T y_k}{d_k} \frac{\mathcal{G}_{k+1}^T y_k \mathcal{G}_{k+1} d_k}{d_{k+1}^T y_k}
\]

\[
\leq -\| \mathcal{G}_{k+1} \|^2 + \theta_4 \| \mathcal{G}_{k+1} \|^2
\]

\[
= -\theta_4 \| \mathcal{G}_{k+1} \|^2.
\]

**Case III (conic):** if \( d_{k+1} \) is calculated by (8) and (29), Li, Liu and Liu [52] have proved that

\[
\mathcal{G}^T_{k+1}d_{k+1} \leq -\zeta \| \mathcal{G}_{k+1} \|^2,
\]

where

\[
\zeta = \min \left\{ \frac{\rho_0}{(8 - 6\rho_0 \max\{\xi_2, \xi_4\})}, \frac{1}{6\varsigma} \right\},
\]

and \( \varsigma \) is a positive constant.
Case IV (conic): when $d_{k+1}$ is obtained by (12) and (16), we have

$$
S_{k+1}^T d_{k+1} = \begin{pmatrix}
\|S_{k+1}\|^2 \\
\|S_{k+1}^T 5k \| \\
\|S_{k+1}^T 6k \|
\end{pmatrix}
\begin{pmatrix}
l_{k+1} \\
h_{k+1} \\
v_{k+1}
\end{pmatrix}
+ \begin{pmatrix}
X \\
\theta_1 \\
\theta_2 \\
\theta_3 \\
y
\end{pmatrix}
\begin{pmatrix}
\|S_{k+1}\|^2 \\
\|S_{k+1}^T 5k \| \\
\|S_{k+1}^T 6k \|
\end{pmatrix}

= - \frac{1}{D_{k+1}} \begin{pmatrix}
\|S_{k+1}\|^2 \\
\|S_{k+1}^T 5k \| \\
\|S_{k+1}^T 6k \|
\end{pmatrix}
\begin{pmatrix}
l_{k+1} \\
h_{k+1} \\
v_{k+1}
\end{pmatrix}
+ \begin{pmatrix}
X \\
\theta_1 \\
\theta_2 \\
\theta_3 \\
y
\end{pmatrix}
\begin{pmatrix}
\|S_{k+1}\|^2 \\
\|S_{k+1}^T 5k \| \\
\|S_{k+1}^T 6k \|
\end{pmatrix}

= - \frac{\|S_{k+1}\|^4}{D_{k+1}} h(x, y),
$$

where $x \triangleq \frac{S_{k+1}^T 5k}{\|S_{k+1}\|^2}$, $y \triangleq \frac{S_{k+1}^T 6k}{\|S_{k+1}\|^2}$. Moreover, $h(x, y)$ is a binary quadratic function of $x$ and $y$ which can be expressed as

$$h(x, y) = Yx^2 + 2\theta_3 xy + 2\theta_2 y^2 + 2\theta_1 x + X.$$

It is easy to acquire the Hessian of $h(x, y)$

$$H_h = \begin{pmatrix}
2Y & 2\theta_3 \\
2\theta_3 & 2\theta_1
\end{pmatrix},$$

we have $Y > 0$ because $A_{k+1}$ is positive definite, and the determinant of $H_h$

$$4Y\theta - 4\theta_3^2 = 4\rho_{k+1}|A_{k+1}|,$$

is also positive, so $h(x, y)$ has a minimizer, that is

$$h(x, y)_{\text{min}} = \frac{|A_{k+1}|}{\rho_{k+1}}.$$

Therefore, we can derive

$$S_{k+1}^T d_{k+1} \leq - \frac{\|S_{k+1}\|^4}{D_{k+1}} h(x, y)_{\text{min}} \leq - \frac{|A_{k+1}|}{D_{k+1}\rho_{k+1}} \|S_{k+1}\|^4. \quad (50)$$

Because $\rho_{k+1}, D_{k+1}$ and $|A_{k+1}|$ are all positive, we just need to prove that $\frac{|A_{k+1}|}{D_{k+1}\rho_{k+1}} \|S_{k+1}\|^4$ has a lower bound. Since $D_{k+1}$ contains $\rho_{k+1}$, we first prove that $\rho_{k+1}$ has an upper bound; that is the upper bound of $N_k, n_k$ and $K$.

For $N_k$, we use the Cauchy inequality and can obtain

$$|N_k| = \frac{S_k}{s_k^T y_k} / M_k \leq \sqrt{1 - \gamma_{k+1} \gamma_{k+1}^{-1}} \left( \frac{\|S_k\|^2}{s_k^T \gamma_k} + 4 \frac{\|S_k\|^2 \|y_k\| \|y_k^*\|^2}{s_k^T \gamma_k s_k^* \gamma_k^*} + 4 \frac{\|S_k\|^2 \|y_k\| \|y_k^*\| \|\omega_k\|}{s_k^T y_k \gamma_k} + 2 \frac{\|S_k\|^3 \|y_k\|^2 \|\omega_k\|}{s_k^T \gamma_k s_k^* \gamma_k^*} + 2 \frac{\|S_k\|^3 \|y_k^*\|^2 \|\omega_k^*\|}{s_k^T y_k^* \gamma_k^*} \right) / M_k.$$
combining (17), (18) and using the Cauchy inequality again, we have

\[ |N_k| \leq \frac{1}{M_k} \|s_{k+1}\|^2 \left[ \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \left( \frac{\|s_{k+1}\|^2}{\|s_{k+1}^TS_k\|} + 4 \frac{\|s_{k+1}\|^2}{\|s_{k+1}^TS_k\|} + \frac{2 + \zeta_2}{\xi_k} \|s_{k+1}\| \|y_k\| \right) 
+ 2 \frac{\|s_{k+1}\| \|y_k\|}{s_k^Ty_k} + \frac{2 + \zeta_2}{\xi_k} \sum_{i=1}^2 \frac{y_k T s_i y_k}{\|s_{k+1}^TS_i\|} \right]. \]

Under condition (26), we have

\[ \frac{\|s_{k+1}\| \|y_k\|}{s_k^Ty_k} \leq \frac{\|s_k\| \|s_{k+1}\| \|y_k\|}{s_k^Ty_k} \leq \frac{1}{\xi_1}, \]
and in the same way,

\[ \frac{\|s_{k+1}\| \|y_k\|}{\|s_k\|} \leq \frac{\|s_{k+1}\|^2}{\|y_k\|}. \]

Note that \( \zeta_k \geq 1 \), the above inequality of \( |N_k| \) can be simplified to

\[ |N_k| \leq \frac{1}{M_k} \|s_{k+1}\|^2 \left[ \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \left( 5 + 4(2 + \zeta_2) \frac{\|s_k\|}{\|y_k\|} + \frac{2 + \zeta_2}{\xi_k} \|y_k\| \right) \right. 
+ \left. (2 + \zeta_2) \frac{\|s_k\|}{\|y_k\|} \right] \left( \frac{\|s_{k+1}\|^2}{\|s_{k+1}^TS_k\|} + \frac{1 + (2 + \zeta_2)}{s_k^Ty_k} \right)^2 \frac{\|y_k\|^2}{s_k^Ty_k}. \]

Utilizing conditions (25)–(27) and the expression of \( K \), we can obtain the upper bound of \( N_k \),

\[ |N_k| \leq \frac{\|s_{k+1}\|^2}{M_k} \left[ \left( \frac{(2 + \zeta_2)^2}{\xi_1^2} + 6 \frac{(2 + \zeta_2)}{\xi_1} + 2 \sqrt{\frac{\zeta_2}{\xi_1}} + 5 \right) \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \|s_{k+1}\|^2 
+ \left( \frac{2 + \zeta_2}{\xi_1} + 1 \right)^2 \frac{\|y_k\|^2}{s_k^Ty_k} \right] 
\leq \frac{1}{\rho_0} \left( \frac{(2 + \zeta_2)^2}{\xi_1^2} + 8 \frac{(2 + \zeta_2)}{\xi_1} + 2 \sqrt{\frac{\zeta_2}{\xi_1}} + 6 \right) K_1 \|s_{k+1}\|^2 
= \left( \frac{(2 + \zeta_2)^2}{\xi_1^2} + 8 \frac{(2 + \zeta_2)}{\xi_1} + 2 \sqrt{\frac{\zeta_2}{\xi_1}} + 6 \right) K \rho_0. \]
Next, the upper bound of $n_k$ is acquired by the same procedure,
\[
|n_k| \leq \frac{1}{m_k} \|g_{k+1}\|^2 \left( \frac{\|y_k\|^2}{s_k^T y_k} + \frac{(2 + \zeta_2)^2}{\zeta_k} \frac{\|y_k\|^2}{s_k^T y_k} + 2 \frac{(2 + \zeta_2)}{\zeta_k} \right)
\]
\[
\leq \frac{1}{m_k} \|g_{k+1}\|^2 \left( 1 + (2 + \zeta_2) \frac{s_k^T y_k}{\|y_k\|^2} \right)^2 \frac{\|y_k\|^2}{s_k^T y_k}
\]
\[
\leq \frac{1}{1 - \frac{\zeta_2}{\zeta_1}} \frac{(\zeta_1 + \zeta_2 + 2)^2}{\zeta_1^2} K_1 \|g_{k+1}\|^2
\]
\[
\leq 6 \frac{(\zeta_1 + \zeta_2 + 2)^2}{\zeta_1^2} K.
\]

For $K_1$, it is easy to know that
\[
K_1 \leq \max \{\zeta_2, \zeta_3\}.
\]

Through the above discussion, now we can give the upper bound of $\rho_{k+1}$,
\[
\rho_{k+1} = \zeta_k \max \{K, N_k, n_k\}
\]
\[
\leq 2 \max \left\{ \left( \frac{2 + \zeta_2}{\zeta_1^2} + 8 \frac{(2 + \zeta_2)}{\zeta_1} + 2 \sqrt{\frac{\zeta_2}{\zeta_1}} + 6 \right) \rho_0, 6 \frac{(\zeta_1 + \zeta_2 + 2)^2}{\zeta_1^2} \right\} K_1 \|g_{k+1}\|^2
\]
\[
\leq L_0 \|g_{k+1}\|^2,
\]
where
\[
L_0 \triangleq 2 \max \left\{ \left( \frac{2 + \zeta_2}{\zeta_1^2} + 8 \frac{(2 + \zeta_2)}{\zeta_1} + 2 \sqrt{\frac{\zeta_2}{\zeta_1}} + 6 \right) \rho_0, 6 \frac{(\zeta_1 + \zeta_2 + 2)^2}{\zeta_1^2} \right\} \max \{\zeta_2, \zeta_3\}.
\]

Since we have found the upper bound of $\rho_{k+1}$, then we now consider $D_{k+1}$. According to (22), $D_k$ can be expressed as
\[
D_{k+1} = s_k^T y_k \tau_k \rho_{k+1} M_k - S_k = s_k^T y_k \tau_k (M_k \rho_{k+1} - M_k N_k).
\] (51)

Using the formula of $M_k$ and the upper bound of $\rho_{k+1}$, we have
\[
|D_{k+1}| = D_{k+1} \leq \|g_{k+1}\|^2 s_k^T y_k \tau_k \left[ L_0 + L_0 \frac{\|y_k\|^2}{s_k^T y_k \tau_k} + L_0 \left| \frac{1}{\gamma_{k+1}} \frac{\|g_{k+1}\|^2}{\|y_k\|^2} \right| \right.
\]
\[
+ \frac{\|g_{k+1}\|^2}{s_k^T y_k} \left( 2 \frac{\|g_{k+1}\|^2}{s_k^T y_k} \|y_k\|^2 \frac{|\|g_{k+1}\|^2|}{s_k^T y_k} \right) + \frac{M_k \|N_k\|}{\|g_{k+1}\|^2} \right].
\]

Note that
\[
\frac{\|g_{k+1}\|^2}{s_k^T y_k \tau_k} = \frac{\|g_{k+1}\|^2}{s_k^T y_k \tau_k} \frac{|\|g_{k+1}\|^2|}{|\|g_{k+1}\|^2|} \leq \frac{\|s_k\|^2}{\|y_k\|^2} \frac{|\|g_{k+1}\|^2|}{\|g_{k+1}\|^2},
\]
\[
\frac{\|g_{k+1}\|^2}{s_k^T y_k} = \frac{\|g_{k+1}\|^2}{s_k^T y_k} \frac{|\|g_{k+1}\|^2|}{|\|g_{k+1}\|^2|} \leq \frac{\|s_k\|^2}{\|g_{k+1}\|^2} \frac{|\|g_{k+1}\|^2|}{\|g_{k+1}\|^2},
\]
so the above inequality of $D_{k+1}$ can be simplified to

$$D_{k+1} \leq \|g_{k+1}\|^2 s_k^T y_k \tau_k \left[ 2L_0 + L_0 \left( 2 \frac{\|s_k\|}{\|y_k\|} + \frac{s_k^T y_k}{\|y_k\|^2} \right) \frac{1 - \gamma_{k+1}}{\gamma_{k+1}} \|g_{k+1}\|^2 \right.$$  

$$+ \left( 4 \frac{\bar{\varsigma}_2}{\bar{\varsigma}_1} + 6 \sqrt{\frac{\bar{\varsigma}_2}{\bar{\varsigma}_1}} + 16 \right) K_1 \right].$$

Finally, using (26) and (27), we can obtain

$$D_{k+1} \leq \|g_{k+1}\|^2 s_k^T y_k \tau_k \left[ 2L_0 + \left( 4 \frac{L_0}{\bar{\varsigma}_1} + 4 \frac{\bar{\varsigma}_2}{\bar{\varsigma}_1} + 6 \sqrt{\frac{\bar{\varsigma}_2}{\bar{\varsigma}_1}} + 16 \right) \max\{\xi_2, \xi_3\} \right]$$

$$\leq L_1 \|g_{k+1}\|^2 s_k^T y_k \tau_k,$$  

(52)

where

$$L_1 \triangleq \left[ 2L_0 + \left( 4 \frac{L_0}{\bar{\varsigma}_1} + 4 \frac{\bar{\varsigma}_2}{\bar{\varsigma}_1} + 6 \sqrt{\frac{\bar{\varsigma}_2}{\bar{\varsigma}_1}} + 16 \right) \max\{\xi_2, \xi_3\} \right].$$

With (28), (50) and (52), the sufficient descent property of $d_{k+1}$ under this case can be established by

$$g_{k+1}^T d_{k+1} \leq -\frac{\varsigma_1}{L_1} \|g_{k+1}\|^2.$$  

**Case V (quadratic):** when $d_{k+1}$ is generated by (8) and (41), the proof can be found in [43], in which Dai and Kou proved that

$$g_{k+1}^T d_{k+1} \leq -\frac{1}{2\theta_2} \|g_{k+1}\|^2.$$  

(53)

Next, if $d_{k+1}$ is generated by (12) and (40), Yang et al. [13] have drawn that

$$g_{k+1}^T d_{k+1} \leq -\frac{1}{16\theta_2} \|g_{k+1}\|^2.$$  

(54)

Combining (54) with (53), we finally have

$$g_{k+1}^T d_{k+1} \leq -\frac{1}{16\theta_2} \|g_{k+1}\|^2.$$  

After the above discussion, we can prove that there indeed exists a constant $\varsigma_1$ such that

$$g_{k+1}^T d_{k+1} \leq -\varsigma_1 \|g_{k+1}\|^2,$$

where

$$\varsigma_1 = \min \left\{ \varsigma_2, \frac{1}{2}, \frac{1}{\theta_4}, \frac{\bar{\varsigma}_5}{L_1}, \frac{1}{16\theta_2} \right\}.$$  

The proof is complete.  

**Lemma 2.** Assume that $f$ satisfies Assumption 2. If the search direction $d_{k+1}$ is calculated by WYL_TSCO, then there exists a constant $\varsigma_2 > 0$ such that

$$\|d_{k+1}\| \leq \varsigma_2 \|g_{k+1}\|.$$  

(55)

**Proof.** Similar to Lemma 1, the proof is divided into several cases.
Case I: if $d_{k+1}$ is calculated by negative gradient, then there certainly holds
\[ \|d_{k+1}\| \leq \|s_{k+1}\|. \]

Case II: if the search direction is the HS direction, according to Assumption 2 and the condition (33), we have
\[
\begin{align*}
\|d_{k+1}\| &= \|s_{k+1} + \rho_k^{HS} d_k\| \\
&= \|s_{k+1}\| + \frac{\|s_{k+1}\| \|y_k\| \|d_k\|}{d_k^Ty_k} \\
&\leq \left(1 + \frac{L}{\delta_1}\right) \|s_{k+1}\|.
\end{align*}
\]

Case III (conic): if $d_{k+1}$ is calculated by (8) and (29), from [52] we conclude that
\[
\|d_{k+1}\| \leq \left(\frac{10\xi_3^2 + 5\xi_1^2 + 5m}{\bar{\rho}_0 \xi_1^2}\right) \|s_{k+1}\|,
\]

where $m = 2n_0 / \bar{\rho}_0$ with
\[ n_0 = \max\left\{4 - 3\bar{\rho}_0, 1 + \frac{2\xi_2}{\xi_1}\right\}, \quad \xi = \max\{\xi_2, \xi_3, \xi_4\}.
\]

Case IV (conic): if $d_{k+1}$ is formed by (12) and (16), we have
\[
\|d_{k+1}\| = \|s_{k+1} + \mu_{k+1}s_k + v_{k+1}y_k\| \\
\leq \frac{1}{\rho_{k+1}} (\|s_{k+1}\| + \|s_k\| + |q_3||y_k^2|),
\]
so in order to prove Lemma 2 in this case, we first need to obtain a lower bound of $D_{k+1}$. Combining (23), (25) and (51), and the value range of $\xi_k$, we can derive
\[
D_{k+1} = s_k^Ty_k \tau_k M_k (\xi_k \max\{K, N_k, n_k\} - N_k) \\
\geq s_k^Ty_k \tau_k \rho_0 \frac{6}{5} \max\{K, N_k, n_k\} - N_k \\
\geq s_k^Ty_k \tau_k \rho_0 \frac{K}{5}.
\]

Then using the above inequality, we can transform (56) into
\[
\begin{align*}
\|d_{k+1}\| &\leq \frac{5}{\rho_0 K s_k^Ty_k \tau_k} (\|s_{k+1}\| + \|s_k\| + |q_3||y_k^2|) \\
&\leq \frac{5}{\rho_0 K} \|s_{k+1}\|^2 g_k^2 + 4 \frac{\|s_{k+1}\| \|y_k\| \|y_k^2\|^2}{s_k^Ty_k \tau_k} + 2 \frac{\|s_{k+1}\| \|s_k\| \|y_k\| \|s_k\|^2}{s_k^Ty_k} \\
&\quad + 4 \frac{\|s_{k+1}\| \|s_k\| \|y_k\| \|y_k^2\| \|\omega_k\|}{s_k^Ty_k \tau_k} + 2 \frac{\|s_{k+1}\| \|y_k\|^2 \|\omega_k\|}{s_k^Ty_k} + \frac{\rho_{k+1} \|s_k\|^2}{s_k^Ty_k} \\
&\quad + \frac{\|s_k\|^2 \omega_k^2}{s_k^Ty_k \tau_k} + 2 \frac{\rho_{k+1} \|s_k\| \|y_k\| \|y_k^2\|^2}{s_k^Ty_k \tau_k} + \frac{\rho_{k+1} \|y_k^2\|^2}{s_k^Ty_k} \\
&\leq \frac{5\|s_{k+1}\| \|s_{k+1}\|^2}{K} \left(\frac{2 + 2\xi_2}{\xi_1^0} + 6 \frac{2 + 2\xi_2}{\xi_1} + 4 \frac{M_0}{\xi_1} + 2\right) \left(\frac{5\xi_2^2 + 30\xi_1^3 \xi_2 + 10\xi_1^4 \sqrt{\xi_0 \xi_2} + (60 + 20M_0)\xi_1 + 20\xi_2 + 20}{\xi_1^0}\right)
\end{align*}
\]
Since (26) implies that $K_1 \geq \xi_1$, we finally obtain an upper bound of $d_{k+1}$, that is

$$
\|d_{k+1}\| \leq \left( \frac{25\xi_1^2 + 5\xi_2^2 + 30\xi_1\xi_2 + 10\xi_1\sqrt{\xi_1\xi_2} + (60 + 20M_0)\xi_1 + 20\xi_2 + 20}{\rho_0 \xi_1^3} \right) \|g_{k+1}\|.
$$

For convenience, we define

$$
L_2 \triangleq \left( \frac{25\xi_1^2 + 5\xi_2^2 + 30\xi_1\xi_2 + 10\xi_1\sqrt{\xi_1\xi_2} + (60 + 20M_0)\xi_1 + 20\xi_2 + 20}{\rho_0 \xi_1^3} \right),
$$

namely, $\|d_{k+1}\| \leq L_2 \|g_{k+1}\|$.

**Case V (quadratic):** if $d_{k+1}$ is formed by (8) and (41), it is similar to Lemma 4 in Li, Liu and Liu [42]. According to their proof, we can deduce

$$
\|d_{k+1}\| \leq \frac{20}{\sigma_1} \|g_{k+1}\|.
$$

If $d_{k+1}$ is calculated by (12) and (40), it follows from [13] that

$$
\|d_{k+1}\| \leq \frac{750}{\sigma_1} \|g_{k+1}\|.
$$

Since $\sigma_1 > 0$, we finally have

$$
\|d_{k+1}\| \leq \frac{750}{\sigma_1} \|g_{k+1}\|.
$$

According to the above analysis, the proof of Lemma 2 is completed by setting

$$
\varsigma_2 = \max \left\{ \left( \frac{10\xi_2^2 + 5\xi_1 + 5m}{\rho_0 \xi_1^2} \right), 1, 1 + \frac{L}{\sigma_1}, L_2, \frac{750}{\sigma_1} \right\}.
$$

\[\Box\]

4.2. Convergence Analysis

In this subsection, we will give the global convergence of the presented algorithm for general functions.

**Lemma 3.** Suppose $\alpha_k$ satisfies the line search conditions (5) and (6), and $f$ satisfies Assumption 2, then

$$
\alpha_k \geq \frac{(1 - \sigma)|g_k^T d_k|}{L \|d_k\|^2}. \quad (57)
$$

**Proof.** According to the line search condition (5), we can easily obtain

$$(\sigma - 1)g_k^T d_k \leq (g_{k+1}^T - g_k)^T d_k = y_k^T d_k \leq \|y_k\| \|d_k\| \leq L \|s_k\| \|d_k\| = \alpha_k L \|d_k\|^2,$$

because $\sigma - 1 < 0$ and $g_k^T d_k < 0$, it is obvious that

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

\[\Box\]

**Theorem 1.** Suppose the objective function $f$ satisfies Assumptions 1 and 2, and sequence $\{x_k\}$ is generated by WYL_TSCO, then we have

$$
\lim_{k \to \infty} \|g_k\| = 0. \quad (58)
$$
Proof. According to (57), (49) and (55), it follows
\[
\delta a_k g_k^T d_k \leq -\frac{(1 - \sigma)\delta (g_k^T d_k)^2}{L} \leq -\frac{(1 - \sigma)\delta_{\xi_1}^2}{L\xi_2} ||g_k||^2 = -H||g_k||^2,
\]
where \(H \triangleq \frac{(1 - \sigma)\delta_{\xi_1}^2}{L\xi_2}\), then combining (6), we have
\[
f_{k+1} \leq C_k + \delta a_k g_k^T d_k \leq C_k - H||g_k||^2. \tag{59}
\]

With the reasonable assumptions that \(Q_0 = 1\), \(Q_1 = 2\), \(Q_2 = 3\), \(Q_3 = 4\), \(Q_4 = 5\) and \(\eta_0 = \eta_1 = \eta_2 = \eta_3 = \eta_4 = 1\) according to the update formulas (7) and (48), we can acquire that
\[
Q_{k+1} = 1 + \sum_{j=0}^{k} \prod_{i=0}^{j} \eta_{k-i}, \quad k = 0, 1, 2, \ldots
\]
Combining (48), we can derive the general formula of \(Q_{k+1}\),
\[
Q_{k+1} = \begin{cases} 
1 + \eta^{k/n} + n \sum_{i=1}^{k/n} \eta^i, & \text{mod } (k, n) = 0, \\
1 + \text{mod}(k, n) + \eta^{\lfloor k/n \rfloor} + n \sum_{i=1}^{\lfloor k/n \rfloor} \eta^i, & \text{mod } (k, n) \neq 0,
\end{cases}
\]
where \(\lfloor \cdot \rfloor\) denotes the floor function. Thus, it is easy to obtain an upper bound of \(Q_{k+1}\),
\[
Q_{k+1} \leq 1 + \text{mod}(k, n) + \eta^{\lfloor k/n \rfloor + 1} + n \sum_{i=1}^{\lfloor k/n \rfloor + 1} \eta^i
\]
\[
\leq 1 + n + (n + 1) \sum_{i=1}^{\lfloor k/n \rfloor + 1} \eta^i
\]
\[
\leq 1 + n + (n + 1) \sum_{i=1}^{k+1} \eta^i
\]
\[
\leq (1 + n) \sum_{i=0}^{k+1} \eta^i
\]
\[
\leq \frac{1 + n}{1 - \eta}
\]
\[
= W, \tag{60}
\]
where \(W = \frac{1 + n}{1 - \eta}\).

When \(k \geq 5\), combining (7), (59) and (60), we obtain
\[
C_{k+1} = C_k + \frac{f_{k+1} - C_k}{Q_{k+1}} \leq C_k - \frac{H}{W} ||g_k||^2,
\]
which means
\[
\frac{H}{W} ||g_k||^2 \leq C_k - C_{k+1}. \tag{61}
\]

When \(k < 5\), (6) and (46) imply that
\[
C_{k+1} \leq f_{k+1} + 0.9(C_k - f_{k+1}) = C_k + 0.1(f_{k+1} - C_k) < C_k.
\]

Hence \(C_k\) is monotonically decreasing. According to Lemma 1.1 in [15], we have \(f_{k+1} \leq C_{k+1}\) for each \(k \geq 5\), and (46) indicates that \(f_{k+1} \leq C_{k+1}\) for each \(k < 5\). Thus, with the assumption that \(f\) is bounded below, \(C_k\) is certainly bounded from below.
Summing up the above analysis and (61), we can finally obtain that

\[ \sum_{k=0}^{\infty} \frac{H}{W} \| g_k \|^2 < \infty, \]

which implies

\[ \lim_{k \to \infty} \| g_k \|^2 = 0. \]

The proof is completed. \( \square \)

5. Numerical Results

In this section, the results of the numerical experiments are shown below. The unconstrained test functions were taken from [62] with the given initial points. To prove the efficiency of the proposed WYL_TSCO algorithm, we compare its numerical performance with SMCG_Conic and TSCG. The performance profile proposed by Dolan and Moré [63] is used to evaluate the performance of these methods. The dimension of the test functions is 10,000. All the programs were written in C code.

SMCG_Conic is also a subspace minimization conjugate gradient algorithm based on the conic model, and is a pioneering one that combines the subspace technique with a conic model to seek the search direction. The numerical experiments in [52] show that the performance of SMCG_Conic is very efficient. Since the biggest difference between WYL_TSCO and SMCG_Conic is the dimension of the used subspace, the comparison between WYL_TSCO and SMCG_Conic can not only reflect the high efficiency of our algorithm, but also reveal the influence on the numerical result generated by the change of dimension for the adopted subspace in the subspace minimization conjugate gradient algorithm.

As for TSCG, it is an efficient subspace minimization conjugate gradient method, and it successfully applies the idea of subspace method and WYL method which are employed in our algorithm as well. However, TSCG only uses the quadratic model to approximate the objective function. Thus, the comparison between the numerical performance of WYL_TSCO and TSCG can reveal the influence of the approximation model in the subspace method.

For the initial stepsize of the first iteration, we adopt the adaptive strategy used in [42]. The other parameters of WYL_TSCO are selected as follows.

\[ \begin{align*}
\epsilon &= 10^{-6}, \quad \epsilon_1 = 10^{-3}, \quad \epsilon_2 = 10^{-4}, \quad \delta = 0.001, \quad \sigma = 0.9999, \quad \lambda_{\min} = 10^{-30}, \quad \lambda_{\max} = 10^{30}, \\
\lambda_1 &= 0.15 \times 10^2, \quad \lambda_2 = 8.5 \times 10^4, \quad \lambda_3 = 4 \times 10^8, \quad \lambda_4 = 6.5 \times 10^9, \quad \lambda_5 = 0.1, \quad \lambda_0 = 0.3, \\
\rho_0 &= 0.9, \quad \theta_1 = 10^{-2}, \quad \theta_2 = 10^4, \quad \theta_3 = 10^{-2}, \quad \theta_4 = 10^{-4}, \quad \lambda_1 = 10^{-7}, \quad \lambda_2 = 0.05.
\end{align*} \]

SMCG_Conic and TSCG use the original parameters in their paper respectively. In addition to stopping when the stopping criterion \( \| g_k \|_\infty \leq \epsilon \) holds, the algorithm also stops when the number of iterations exceeds 200,000.

For the 80 test problems from [62], WYL_TSCO successfully solves 100% of them, while TSCG and SMCG_Conic both successfully solve 76 problems. We compare the numerical results of problems that are solved successfully by all the three algorithms, the number of which are 72. The numerical results of the three algorithms are presented in the following figures.

Figure 1 depicts the performance based on the number of iterations for the three methods. It shows that the performance of WYL_TSCO is similar to that of TSCG, but is a little inferior to SMCG_Conic’s only when \( \tau < 1.31 \). However, WYL_TSCO outperforms TSCG and SMCG_Conic when \( \tau > 1.31 \).

In Figure 2, we observe that WYL_TSCO and TSCG outperform SMCG_Conic on the number of function evaluations, and TSCG is slightly ahead of WYL_TSCO only in the case of \( \tau < 1.5 \).

Similar to Figure 1, Figure 3 illustrates that WYL_TSCO is at a little disadvantage compared with SMCG_Conic only when \( \tau < 1.318 \), but is competitive with TSCG.
Figure 1. Performance profile based on the number of iterations.

Figure 2. Performance profile based on the number of function evaluations.

For the CPU time, Figure 4 shows that WYL_TSCO has an improvement on TSCG and SMCG_Conic, which shows the high efficiency of WYL_TSCO.
For the 80 test problems, the numerical results show that WYL_TSCO and TSCG may perform a little worse than SMCG_Conic when \( \tau \) is very small, but they overall have significant improvements over SMCG_Conic, and WYL_TSCO also is superior to TSCG for almost all \( \tau \). Moreover, WYL_TSCO can solve more problems than TSCG and SMCG_Conic, which might show the advantage of the three-dimensional subspace method and conic...
model. To sum up, the WYL_TSCO is an efficient algorithm for solving the unconstrained optimization problem.

6. Conclusions

Combining the idea of WYL method, this paper proposes a three-dimensional subspace method based on conic models for unconstrained optimization (WYL_TSCO). The sufficient descent property of the search direction and the global convergence of this method are obtained under some suitable assumptions. The selection of an approximation model depends on whether certain criterions are satisfied. Furthermore, the strategies of initial stepsize and nonmonotone line search are exploited which are beneficial to the convergence and efficiency. According to the numerical results and theoretical analysis, WYL_TSCO is very competitive and promising.

Author Contributions: Conceptualization, S.Y. and J.X.; methodology, G.W.; software and validation, G.W. and M.P.; visualization and formal analysis, G.W.; writing—original draft preparation, G.W.; supervision, S.Y.; funding acquisition, S.Y. and J.X. All authors have read and agreed to the published version of the manuscript.

Funding: Chinese National Natural Science Foundation no. 12262002, Natural Science Foundation of Guangxi Province (CN) no. 2020GXNSFAA159014 and 2021GXNSFAA196076, the Program for the Innovative Team of Guangxi University of Finance and Economics, and Guangxi First-class Discipline Construction Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
13. Huo, J.; Yang, J.; Wang, G.; Yao, S. A class of three-dimensional subspace conjugate gradient algorithms for unconstrained optimization. Symmetry. 2022, 14, 80. [CrossRef]


34. Yuan, Y.X. Subspace methods for large scale nonlinear equations and nonlinear least squares. *Optim. Eng.* 2009, 10, 207–218. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.