A family of improved secant methods via nonmonotone curvilinear paths technique for equality constrained optimization

Detong Zhu
Department of Mathematics, Shanghai Normal University, Shanghai 200234, People’s Republic of China

Received 5 June 1999; received in revised form 15 June 2000

Abstract

This paper presents a family of improved secant algorithms via two preconditional curvilinear paths, the preconditional modified gradient path and preconditional optimal path, for solving general nonlinear optimization problems with nonlinear equality constraints. We employ the stable Bunch–Parlett factorization method of symmetric matrices so that two preconditional curvilinear paths are very easily formed. The nonmonotone curvilinear search technique, by introducing a nonsmooth merit function and adopting a dogleg-typed movement, is used to speed up the convergence progress in the contours of objective function with large curvature. Global convergence of the proposed algorithms is obtained under some reasonable conditions. Furthermore, the dogleg-typed step overcomes the Maratos effect to bring the local superlinear convergence rate. The results of numerical experiments are reported to show the effectiveness of the proposed algorithms.

MSC: 90C30; 65K05

Keywords: Curvilinear paths; Factorization; Preconditioner; Nonmonotonic technique; Constrained optimization

1. Introduction

In this paper we consider the nonlinear equality constrained minimization problems

\[
\min f(x) \\
\text{s.t. } c(x) = 0,
\]

where \(f(x) : \mathbb{R}^n \rightarrow \mathbb{R}^1\) and \(c(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m, m \leq n\) are twice continuously differentiable. Among the most successful methods for solving problem (1.1) we find the reduced Hessian methods in

This work is supported by the National Science Foundation Grant of China, Science Foundation Grant of Shanghai Natural Sciences Committee and Science Technology Foundation of Shanghai Higher Education.

0377-0427/01/$- see front matter © 2001 Elsevier Science B.V. All rights reserved.
PH: S0377-0427(00)00576-8
successive quadratic programming (SQP) (see [4,5,9]), and secant methods (two-step algorithms) as defined in [6]. Compared with the widely reduced Hessian methods in which the orthonormal basis \((Z_k \in \mathbb{R}^{n \times t}, \text{where } t = n - m)\) for the tangent space of the constraints at the current point \(x_k\) changes continuously with \(k\), the secant methods have a main advantage which rests in the use of an orthogonal projection operator which is continuous. However, recent reports indicate that it might be difficult to find a basis \(Z_k\) which changes continuously with \(k\) (see, for example, [3]).

We now first state a family of the improved secant algorithms in which after the moving vectors \(s_k\) is determined by using the original secant algorithms (two-step algorithms) a correction step \(d_k\) will also be considered to make the performance of the algorithm more satisfactory and to overcome the Maratos effect. Let \(|| \cdot ||\) be the Euclidean norm on \(\mathbb{R}^n\). For simplicity, we denote \(f(x_k)\) by \(f_k\); \(\nabla f(x_k)\) by \(g_k\); and \(\nabla^2 f(x_k)\) by \(\nabla^2 f_k\), etc.

The general improved form of the algorithms follows, in each iteration

\[
\lambda_k = U(x_k, \lambda_{k-1}, B_k), \tag{1.2}
\]

\[
B_kw_k = -\nabla_x l(x_k, \lambda_k) \triangleq -\phi^k, \tag{1.3}
\]

\[
h_k = P(x_k, B_k)w_k, \tag{1.4}
\]

\[
v_k = -A(x_k)^\top c_k, \tag{1.5}
\]

\[
y_k = \nabla_x l(x_k + h_k, \lambda_k) - \nabla_x l(x_k, \lambda_k), \tag{1.6}
\]

\[
s_k = h_k + v_k, \tag{1.7}
\]

\[
d_k = -A(x_k)^\top c(x_k + s_k), \tag{1.8}
\]

\[
B_{k+1} = \text{DFP/BFGS}(h_k, y_k, B_k), \tag{1.9}
\]

\[
x_{k+1} = x_k + s_k + d_k. \tag{1.10}
\]

In the above algorithms, \(l(x, \lambda)\) is the Lagrangian function defined for \(x \in \mathbb{R}^n\) and \(\lambda \in \mathbb{R}^m\) by

\[
l(x, \lambda) = f(x) + \lambda^\top c(x). \tag{1.11}
\]

Compared with the above-improved algorithms, the original secant methods proposed in [6], did not use the correction step \(d_k\), i.e., \(d_k \equiv 0\). The dogleg-typed improvement \(d_k\) can not only overcome the Maratos effect, but also yields the one-step \(q\)-superlinear convergence (see Section 5). The matrix \(B_k\) can be secant updated formulas defined as stated in [6] using an exact or approximate Hessian or quasi-Newton formula after each iteration. Fontecilla assume that the matrix \(B_k\) for all \(k\) is positive definite. However, \(B_k\) may be indefinite and even the inverse \(B_k^{-1}\) may not exist. We use the general pseudo-inverse of \(B_k\), \(B_k^\dagger\) instead of \(B_k^{-1}\) in this paper.

The pseudo-inverse of \(A_k^\top\), \(A(x_k)^\dagger\), will be equal to either of the two,

\[
A_k^\dagger = A_k(A_k^\top A_k)^{-1} \tag{1.12}
\]

or

\[
A_k^\dagger = B_k^\dagger A_k(A_k^\top B_k^\dagger A_k)^\dagger, \tag{1.13}
\]
Table 1
Algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$U$</th>
<th>$P$</th>
<th>$A^\dagger$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alg 1</td>
<td>(1.17)</td>
<td>$I$</td>
<td>(1.12)</td>
</tr>
<tr>
<td>Alg 2</td>
<td>(1.17)</td>
<td>$I$</td>
<td>(1.13)</td>
</tr>
<tr>
<td>Alg 3</td>
<td>(1.16)</td>
<td>$P_B$</td>
<td>(1.12)</td>
</tr>
<tr>
<td>Alg 4</td>
<td>(1.16)</td>
<td>$P_B$</td>
<td>(1.13)</td>
</tr>
<tr>
<td>Alg 5</td>
<td>(1.18)</td>
<td>$P$</td>
<td>(1.12)</td>
</tr>
<tr>
<td>Alg 6</td>
<td>(1.18)</td>
<td>$P$</td>
<td>(1.13)</td>
</tr>
</tbody>
</table>

where choice of $A(x_k)$ in (1.8) corresponds to that in (1.5). The projection onto $\mathcal{N}(A(x)^\top)$, the null space $A(x)^\top$, in (1.4) can be either the orthogonal projection $P(x)$ given by

$$P(x) = I - A(x)[A(x)^\top A(x)]^{-1}A(x)^\top$$  \hspace{1cm} (1.14)

or the oblique projection onto $\mathcal{N}(A(x)^\top)$ is defined by

$$P_B(x) = I - B^\dagger A(x)[A(x)^\top B^\dagger A(x)]^\dagger A(x)^\top.$$ \hspace{1cm} (1.15)

The multiplier updates in (1.2) can be chosen from one of the following updates:

1. **Projection update:**
   $$\lambda_k^P = -(A_k^\top A_k)^{-1}A_k^\top g_k,$$  \hspace{1cm} (1.16)

2. **Null-space update:**
   $$\lambda_k^S = -(A_k^\top B_k^\dagger A_k)^\dagger A_k^\top B_k^\dagger g_k,$$ \hspace{1cm} (1.17)

3. **Newton update:**
   $$\lambda_k^N = (A_k^\top B_k^\dagger A_k)^\dagger (c_k - A_k^\top B_k^\dagger g_k).$$ \hspace{1cm} (1.18)

With choices for multiplier updates in (1.2), projection operators in (1.4), vertical steps in (1.5) and correction step in (1.8) we obtain the following six improved algorithms (see Table 1).

In [6], Fontecilla proved that the proposed algorithms were a local two-step superlinear convergence rate in which did not refer to the global convergence. It was proved that if the initial point $x_0$ is close enough to the solution of problem (1.1), $x_s$, and $B_k \approx W_s$, then the iterative sequence of $\{x_k\}$ approaches $x_s$, where $W(x, \lambda)$ is defined as

$$W(x, \lambda) = \nabla^2_{x\lambda} l(x, \lambda).$$ \hspace{1cm} (1.19)

Furthermore, under the assumption

$$\lim_{k \to \infty} \frac{\|P_k(B_k - W_s)h_k\|}{\|h_k\|} = 0$$ \hspace{1cm} (1.20)

and under some other conditions, the convergence is two-step $q$-superlinear, i.e.,

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x_s\|}{\|x_{k-1} - x_s\|} = 0.$$ \hspace{1cm} (1.21)

Two basic approaches, namely the line search and trust region, have been developed in order to ensure global convergence towards local minima. The classical line search approach for the
secant methods is a well accepted way for constrained minimization to assure global convergence. An interesting idea is to form various curves for globalizing results based on the solution of a system of differential equations. Two special curves that are named as the optimal path and modified gradient path have been suggested in trust region technique (see [2]). The two curvilinear paths can be expressed by the eigenvalues and eigenvectors of the exact or inexact Hessian matrix of the quadratic model function of the objective function. However, calculation of the full eigensystem of a symmetric matrix is usually time consuming, and the optimal path and modified gradient path algorithms are generally impractical.

The approximate updating of the Hessian matrix of the quadratic model function can be factorized by employing the stable Bunch–Parlett factorization method of symmetric matrices. In optimization algorithms it is sometimes helpful to include a scaling matrix for the variables. In this paper, we introduce the scaling matrix, named after preconditioner in [12], similar to the curvilinear paths in [2], to generate the preconditional modified gradient path and preconditional optimal path in which we use the line search instead of the trust region strategy. The preconditional curvilinear paths can be very easily formulated from the full eigensystem of the factorized diagonal blocks matrix which is very easy to calculate.

The most successful implementations of these techniques are invariably based on the enforcement of a monotonic decrease of the objective function values. However, computational experience observation are that enforcing monotonicity may have dangerous effects in the minimization of highly nonlinear functions with the presence of steep-sided valleys, since the search for lower function values may cause any minimization algorithm to be trapped in the bottom of a valley. Nonmonotonic line search technique which was proposed by Grippo, Lampariello and Lucidi (see [7]) for unconstrained optimization, does not require objective values to decrease after every iteration so that the effects can be avoided. The nonmonotonic idea motivates in connection with the preconditional curvilinear paths approach to globalizing constrained optimization. The resulting algorithms via two special curvilinear paths shall possess global convergence while introducing the nondifferentiable penalty function as merit function and adding a correction step to overcome the Maratos effect and to ensure to a superlinear convergence rate.

The paper is organized as follows. In Section 2, we propose the characterizations and properties of the preconditional curvilinear paths. In Section 3, we describe the nonmonotonic improved secant algorithms via the preconditional curvilinear paths. In Section 4, the global convergence of the proposed algorithms is established while the superlinear convergence rate is discussed in Section 5. Finally, the results of numerical experiments of the proposed algorithms are reported in Section 6.

2. Preconditional curvilinear paths

We first give a brief description of preconditional curvilinear paths. The modified gradient path and optimal path emerging at \( x_k \) of a general continuously differentiable function is the solution of the differential equation (see [2]). In each iteration of the algorithms, we shall first solve system (1.3). Since solving this system exactly may require too many computations, we choose to solve this problem approximately for practical consideration. We can use the exact or approximate Hessian to update \( B_k \), even if \( B_k \) will be indefinite. Let

\[
\psi_h(\delta_h(\tau)) = \nabla_x I(x_k, \lambda_k)^T \delta_h(\tau) + \frac{1}{2} \delta_h(\tau)^T B_k \delta_h(\tau),
\]
then the solution of
\[
\frac{d\delta_k(\tau)}{d\tau} = -\nabla \psi_k(\delta_k(\tau)) \quad \text{and} \quad \delta_k(0) = 0
\]  
(2.1)
is a valid approximation of the curvilinear paths for the gradient of \( l(x_k + \delta_k(\tau), \lambda_k) \) and thus provides a set of sensible candidates for a successor point \( x_k + \delta_k(\tau) \) to \( x_k \) for the modified gradient path. The optimal path will be concerned with trust region equations (see [2]). It is sometimes helpful to include a scaling matrix for the variables which is diagonal and fixed in most cases. We will employ the stable Bunch–Parlett factorization method of symmetric matrices to factorize the Hessian matrix of the quadratic model function. The stable Bunch–Parlett factorization method (see [1]) factorizes the matrix \( B_k \) into the form
\[
E_k B_k E_k^T = L_k D_k L_k^T,
\]  
(2.2)
where \( E_k \) is a permutation matrix, \( L_k \) a unit lower triangular matrix and \( D_k \) a block diagonal matrix with \( 1 \times 1 \) and \( 2 \times 2 \) diagonal blocks. This factorization has following properties (see [1] or [13]).

\( D_k \) and \( B_k \) have the same inertia, that is, they have the same number of positive, zero and negative eigenvalues. Further, the pseudo-inverse of \( B_k \) is
\[
B_k^\dagger = (L_k^{-1} E_k)^{-1} (L_k^{-1} E_k)^T,
\]  
(2.3)
where \( (\phi_i^k)^\dagger = \begin{cases} (\phi_i^k)^{-1} & \text{if } \phi_i^k \neq 0, \\ 0 & \text{if } \phi_i^k = 0. \end{cases} \)

There exist positive constants \( \bar{c}_1 \) and \( \bar{c}_2 \) such that for all \( k \)
\[
||L_k|| \leq \bar{c}_1, \quad ||L_k^{-1}|| \leq \bar{c}_2.
\]  
(2.3)

Since \( D_k \) is a block diagonal matrix, its eigenvalues and orthonormal eigenvectors can be easily calculated (see [13]). In our preconditional curvilinear path type of local quadratic approximation of \( f \) at \( x_k \), the matrix \( L_k \) is used to scale the variables
\[
w = L_k^T E_k \tilde{w}
\]  
(2.4)
and the local quadratic approximation function at \( x_k \),
\[
\hat{\psi}_k(\tilde{w}) \equiv (\tilde{\phi}_k)^T \tilde{w} + \frac{1}{2}(\tilde{w})^T D_k \tilde{w},
\]  
(2.5)
where \( \tilde{\phi}_k = L_k^{-1} E_k \phi_k \). Note that \( \tilde{w} \) rather than \( w = E_k^T L_k^{-1} \tilde{w} \) is required at local quadratic function \( \hat{\psi}_k(\tilde{w}) \), which will further improve the efficiency of the calculation of the solution step.

Now, we employ preconditioner to form the two paths as described by Bulteau and Vial (see [2]), i.e., preconditional optimal path and preconditional modified gradient path, respectively.

When the parameter \( \tau \) varies in the interval \([0, +\infty)\), the solution points form the scaled paths and emanate from the current origin \( x_k \). In order to define those arcs in a closed form, we shall use the eigensystem decomposition of \( B_k \). Since \( D_k \) is a block diagonal matrix with \( 1 \times 1 \) and \( 2 \times 2 \) diagonal blocks, its eigenvalues \( \phi_1^k, \phi_2^k, \ldots, \phi_n^k \) are real numbers and these are corresponding orthonormal eigenvectors \( u_1^k, u_2^k, \ldots, u_n^k \). Without loss of generality, let \( \phi_1^k \leq \phi_2^k \leq \cdots \leq \phi_n^k \) be eigenvalues of \( D_k \) and \( u_1^k, u_2^k, \ldots, u_n^k \) be corresponding orthonormal eigenvectors. We partition the set \( \{1, \ldots, n\} \) into \( \mathcal{I}_k^+, \mathcal{I}_k^- \) and \( \mathcal{N}_k \) according to \( \phi_i^k > 0, \phi_i^k < 0 \) and \( \phi_i^k = 0 \) for \( i \in \{1, \ldots, n\} \), respectively. We now give two preconditional curvilinear paths.
2.1. Preconditional optimal path

The null step $h_k = P(x_k, B_k)w_k$ in the null space $\mathcal{N}(A(x_k)^T)$, where $w_k = E_k^TL_k^{-1}\tilde{w}_k$, is obtained by the projection $P(x_k, B_k)$ on the full step $\tilde{w}_k$. The preconditional optimal path $\Gamma(\tau)$ for the full step $\tilde{w}_k$ of the full space $\mathbb{R}^n$ can be expressed as

$$\Gamma(\tau) = \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)), \quad (2.6)$$

where

$$\Gamma_1(t_1(\tau)) = - \left[ \sum_{i \in \mathcal{J}_k} \frac{t_1(\tau)}{\phi_i^k} - \frac{t_1(\tau)}{\phi_i^k} \right],$$

$$\Gamma_2(t_2(\tau)) = t_2(\tau) u_k^i$$

and

$$t_1(\tau) = \tau \quad \text{and} \quad t_2(\tau) = 0, \quad \text{if} \quad \tau < \frac{1}{T_k},$$

$$t_1(\tau) = \frac{1}{T_k} \quad \text{and} \quad t_2(\tau) = \tau - \frac{1}{T_k}, \quad \text{if} \quad \tau \geq \frac{1}{T_k},$$

$\mathcal{J}_k = \{i | \phi_i^k \neq 0, \ i = 1, \ldots, n\}$, $\mathcal{N}_k = \{i | \phi_i^k = 0, \ i = 1, \ldots, n\}$, $\tilde{\phi}_i^k = (\phi_i^k)^T u_k^i$, $i = 1, \ldots, n$, $\tilde{\phi}_i^k = \sum_{i=1}^n \phi_i^k u_k^i$, $T_k = \max\{0, -\phi_i^k\}$ and $1/T_k$ is defined as $+\infty$ if $T_k = 0$. It should be noted that $\Gamma_2(t_2(\tau))$ is defined only when $D_k$ is indefinite and $\tilde{g}_i^k = 0$ for all $i \in \{1, \ldots, n\}$ with $\phi_i^k = \phi_i^k < 0$ which is referred to as hard case for unconstrained optimization and that for other cases, $\Gamma(\tau)$ is defined only for $0 \leq \tau < 1/T_k$, that is, $\Gamma(\tau) = \Gamma_1(t_1(\tau))$.

The preconditional optimal path $\Gamma^v(\tau)$ for the range step $v$ in the range space $\mathcal{R}(A(x_k))$ of $A(x_k)$ can be expressed as

$$\Gamma^v(\tau) = - \frac{\tau}{\tau + 1} A(x_k)^T c_k. \quad (2.7)$$

2.2. Preconditional modified gradient path

Similarly, the null step $h_k = P(x_k, B_k)w_k$ in the null space $\mathcal{N}(A(x_k)^T)$, where $\tilde{w}_k = E_k^TL_k^{-1}\tilde{w}_k$, is also obtained by the projection $P(x_k, B_k)$ on the full step $\tilde{w}_k$. The preconditional modified gradient path $\Gamma(\tau)$ for the full step $\tilde{w}_k$ in the full space $\mathbb{R}^n$ can be given in the following closed form which can be referred to [2]:

$$\Gamma(\tau) = \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)), \quad \tau \in [0, +\infty), \quad (2.8)$$

where if $\tilde{\phi}_i^k \neq 0$ for some $i \in \mathcal{J}_k \cup \mathcal{N}_k$, the term $\Gamma_2(t_2(\tau))$ is not relevant, that is, if $\tilde{\phi}_i^k \neq 0$ for some $i \in \mathcal{J}_k \cup \mathcal{N}_k$, then $\Gamma_2(t_2(\tau)) = 0$. For the path $\Gamma(\tau)$, the definitions of $\Gamma_1(t_1(\tau))$ and $\Gamma_2(t_2(\tau))$ are as follows:

$$\Gamma_1(t_1(\tau)) = \sum_{i \in \mathcal{J}_k} \exp\{-\phi_i^k t_1(\tau)\} - 1 \frac{\tilde{\phi}_i^k}{\phi_i^k} - t_1(\tau) \sum_{i \in \mathcal{N}_k} \tilde{\phi}_i^k u_k^i$$
with
\[
\Gamma_2(t_2) = \begin{cases} 
  t_2 u_k^1 & \text{if } \varphi_1^k < 0, \\
  0 & \text{if } \varphi_1^k \geq 0,
\end{cases}
\]
\[
t_1(\tau) = \begin{cases} 
  \tau & \text{if } \tau < 1, \\
  1 - \tau & \text{if } \tau \geq 1
\end{cases}
\]
and
\[
t_2(\tau) = \max\{\tau - 1, 0\}.
\]

The preconditional modified gradient path \(\Gamma(v)(\tau)\) for step \(v\) of the range space \(\mathcal{R}(A(x_k))\) of \(A(x_k)\) can be expressed as
\[
\Gamma(v)(t_1(\tau)) = (\exp\{-t_1(\tau)\} - 1) A(x_k)^\dagger c_k,
\]
where \(t_1(\tau)\) given same as above.

2.3. Properties of preconditional curvilinear paths

We will discuss the properties in each of the above two preconditional paths in detail and summarize as follows.

Lemma 2.1. Let the full step \(\tilde{w}_k(\tau)\) be obtained from the preconditional optimal path in the full space \(\mathbb{R}^n\) and the null step \(h_k(\tau) = P(x_k, B_k)w_k(\tau)\), where \(w_k(\tau) = E_k^\dagger L_k^{-1}\tilde{w}_k(\tau)\), be obtained by the projection \(P(x_k, B_k)\) in the null space \(\mathcal{N}(A(x_k)^\dagger)\). Then we have that the norm function of the path \(h_k(\tau)\) is monotonically increasing for \(t \in (0, +\infty)\). The relation function \(\check{\psi}_k(\tau) \triangleq (g_k^\dagger)^\dagger h_k(\tau)\) between gradient \(g_k\) of the objective function and the step \(h_k(\tau)\) satisfies a sufficiently descent direction and is monotonically decreasing for \(0 < \tau < 1/T\). Furthermore,
\[
\nabla_x l(x_k, \lambda_k)^\dagger \frac{dh_k(\tau)}{d\tau} = g_k^\dagger \frac{dh_k(\tau)}{d\tau} = \frac{d\check{\psi}_k(\tau)}{d\tau}
\]
\[
\rightarrow -||P(x_k, B_k)\phi^k||^2 = -||P(x_k, B_k)g_k||^2 \quad \text{as } \tau \rightarrow 0,
\]
and
\[
\lim_{\tau \rightarrow \infty} \Gamma_k^\dagger(\tau) = -A(x_k)^\dagger c_k.
\]
If \(B_k\) is positive, then
\[
\lim_{\tau \rightarrow \infty} \Gamma_k(\tau) = -D_k^{-1} \phi^k.
\]

Proof. Let the step \(\tilde{w}_k(\tau)\) be obtained from the preconditional optimal path. Since \(u^1, u^2, \ldots, u^n\) are orthonormal eigenvectors and \(P(x_k, B_k)\), which simplicity if it cannot be confusion, denoted by \(P_k\), is the projection onto the null space \(\mathcal{N}(A(x_k)^\dagger)\), it is obvious that \(\Gamma(\tau)\) is a continuous path, and
\[
||\Gamma(\tau)|| = \begin{cases} 
  ||\Gamma_1(\tau)||^2 & \text{if } \tau < \frac{1}{T}, \\
  ||\Gamma_1(\frac{1}{T})||^2 + ||\Gamma_2(t_2(\tau))||^2 & \text{if } \tau \geq \frac{1}{T}.
\end{cases}
\]
Let

\[ \psi(\tau) = ||\mathbf{I}_1(\tau)||^2 = \sum_{i \in \mathcal{I}_k} \left( \frac{\tau}{\phi_i^k \tau + 1} \right)^2 (\mathbf{\hat{\phi}}_i^k)^2 + \tau^2 \sum_{i \notin \mathcal{I}_k} (\mathbf{\hat{\phi}}_i^k)^2. \]

Then using the fact \( \phi_i^k \tau + 1 > 0 \) for all \( i \in \mathcal{I}_k \) when \( 0 < \tau < 1/T \), we have

\[ \psi'(\tau) = \sum_{i \in \mathcal{I}_k} 2 \frac{\tau}{(\phi_i^k \tau + 1)^3} (\mathbf{\hat{\phi}}_i^k)^2 + 2 \tau \sum_{i \notin \mathcal{I}_k} (\mathbf{\hat{\phi}}_i^k)^2 > 0 \]

which means that \( ||\mathbf{I}_1(\tau)|| \) is monotonically increasing for \( 0 < \tau < 1/T \). \( ||\mathbf{I}_2(t_2(\tau))||^2 = (\tau - 1/T)^2 \) is certainly increasing for \( \tau \geq 1/T \). Since \( ||h_k(\tau)|| = ||P(x_k, B_k)w_k(\tau)|| \) and \( w_k(\tau) = E^T_k L_k^{-1} \tilde{w}_k(\tau) \), as above we have that the norm function of the path is monotonically increasing for \( \tau \in (0, +\infty) \).

From (2.6) and \( 0 < \tau < 1/T \), we have that, noting \( P_k = P_{kg_k} \) and \( (P_{kg_k})_i = (P_k \phi_i^k)_i \), where \( (P_k \phi_i^k)_i \) is the \( i \)th component of the vector \( P_k \phi^k \),

\[ \hat{\psi}_k(\tau) = (g^k)^T P_k E^T_k L_k^{-1} \mathbf{I}_1(\tau) \]

\[ = -(P_k \phi^k)^T \left[ \sum_{i \in \mathcal{I}_k} \frac{\tau_i(\tau)}{\phi_i^k \tau + 1} (P_k \phi^k)_i u' + \tau \sum_{i \notin \mathcal{I}_k} (P_k \phi^k)_i u' \right] \]

\[ = - \sum_{i \in \mathcal{I}_k} \frac{\tau}{\phi_i^k \tau + 1} (P_k \phi^k)_i^2 - \tau \sum_{i \notin \mathcal{I}_k} (P_k \phi^k)_i^2. \]

(2.13)

Then we have that

\[ \frac{d\hat{\psi}_k(\tau)}{d\tau} = - \sum_{i \in \mathcal{I}_k} \frac{1}{\phi_i^k \tau + 1} (P_k \phi^k)_i^2 - \sum_{i \notin \mathcal{I}_k} (P_k \phi^k)_i^2 \]

(2.14)

which means that \( \hat{\psi}_k(\tau) \) is monotonically decreasing for \( 0 < \tau < 1/T \). Taking \( \tau \to 0 \), (2.10) is true.

Using the fact \( ||\mathbf{D}_k|| + T \geq \phi_i^k + T > 0 \) for all \( i \in \mathcal{I}_k \) when \( 0 < \tau < 1/T \), we have that

\[ \frac{\tau}{\phi_i^k \tau + 1} \geq \frac{1}{||\mathbf{D}_k|| + 1/\tau}. \]

Therefore,

\[ \hat{\psi}_k(\tau) = (g^k)^T h_k(\tau) \leq - \frac{1}{||\mathbf{D}_k|| + 1/\tau} \sum_{i \in \mathcal{I}_k} (P_k \phi^k)_i^2 \leq - \frac{||P_k \phi^k||^2}{||\mathbf{D}_k|| + 1/\tau}. \]

(2.15)

In hard case, i.e., \( \tau \geq 1/T \) and \( \Gamma(\tau) = \Gamma_1(1/T) + \Gamma_2(t_2(\tau)) \), we have that for all \( \frac{1}{\phi_i^k + T} \geq \frac{1}{||\mathbf{D}_k|| + T}, \quad i \in \mathcal{I}_k \) and \( \frac{1}{T} \geq \frac{1}{||\mathbf{D}_k|| + T} \).
Therefore, we have that, noting \( \nabla_x l(x_k, \lambda_k)^T h_k(\tau) = (g_k)^T h_k(\tau) \),

\[
\hat{\psi}_k(\tau) \leq - \frac{1}{||D_k|| + T} \sum_{i \in I_k} (P_k g_k)^T_i - \frac{1}{T} \sum_{i \in I_k} (P_k g_k)^2_i \\
\leq - \frac{||P_k g_k||^2}{||D_k|| + T}.
\] (2.16)

Eqs. (2.15) and (2.16) mean that \( h_k(\tau) \) satisfies a sufficiently descent direction.

If \( B_k \) is positive, then by the definition of the preconditional optimal path \( \Gamma_k(\tau) \) which can be expressed as

\[
\Gamma_k(\tau) = - \sum_{i \in I_k} \frac{t_i(\tau)}{\phi^T_i t_i(\tau)} \tilde{\phi}_i u_i \\
\rightarrow - \sum_{i \in I_k} \frac{1}{\phi^T_i} \tilde{\phi}_i u_i \quad \text{as} \quad \tau \rightarrow \infty \\
= - D_k^{-1} \hat{\phi}^k
\]

which means that (2.12) holds.

It is clear that (2.12) holds as

\[
\lim_{\tau \to \infty} \frac{\tau}{\tau + 1} = -1. \quad \square
\]

**Lemma 2.2.** Let the full step \( \tilde{w}_k(\tau) \) be obtained from the preconditional modified gradient path in the full space \( \mathbb{R}^n \) and the null step \( h_k(\tau) = P(x_k, B_k) w_k(\tau) \), where \( w_k(\tau) = E_k L_k^{-1} \tilde{w}_k(\tau) \), be obtained by the projection \( P(x_k, B_k) \) in the null space \( \mathcal{N}(A(x_k)^T) \). Then we have that the norm function of the path \( h_k(\tau) \) is monotonically increasing for \( \tau \in [0, +\infty) \). The relation function \( \tilde{\psi}_k(\tau) \triangleq (g_k)^T h_k(\tau) \) between gradient \( g_k \) of the objective function and the step \( h_k(\tau) \) satisfies a sufficiently descent direction and is monotonically decreasing for \( 0 < \tau < 1 \). Furthermore,

\[
\nabla_x l(x_k, \lambda_k)^T \frac{dh_k(\tau)}{d\tau} = (g_k)^T \frac{dh_k(\tau)}{d\tau} = \frac{d\tilde{\psi}_k(\tau)}{d\tau} \\
\rightarrow - ||P(x_k, B_k) g_k||^2 = - ||P(x_k, B_k) g_k||^2, \quad \text{as} \quad \tau \to 0,
\] (2.17)

and

\[
\lim_{\tau \to \infty} \Gamma_k^x(\tau) = - A(x_k)^T c_k. \quad (2.18)
\]

If \( B_k \) is positive, then

\[
\lim_{\tau \to \infty} \Gamma_k(\tau) = - D_k^{-1} \hat{\phi}^k. \quad (2.19)
\]

**Proof.** From the definition of the path and using the orthonormality of vectors \( u_i \), we have

\[
||\Gamma_k(\tau)||^2 = \begin{cases} 
||\Gamma_1(\tau)||^2 & \text{if } \tau < 1, \\
||\Gamma_1(\tau)||^2 + ||\Gamma_2(t_2(\tau))||^2 & \text{if } \tau \geq 1.
\end{cases}
\]
\[ \psi_i(\tau) = ||\Gamma_i(\tau)||^2 = \sum_{i \in \mathcal{I}_k} \left( \frac{\exp\{-\phi_i^k t_i(\tau)\} - 1}{\phi_i^k} \right)^2 (\hat{\phi}_i^k)^2 + \left( \frac{\tau}{1-\tau} \right)^2 \sum_{i \in \mathcal{N}_k} (\hat{\phi}_i^k)^2. \]

Then using the fact \( \tau > 0 \) and \( \exp\{-\phi_i^k t_i(\tau)\} \leq 1 \), we have

\[ \psi_i'(\tau) = \sum_{i \in \mathcal{I}_k} \frac{2\exp\{-\phi_i^k t_i(\tau)\}}{(1-\tau)^2} (1 - \exp\{-\phi_i^k t_i(\tau)\})(\hat{\phi}_i^k)^2 + \frac{2\tau}{(1-\tau)^2} \sum_{i \in \mathcal{N}_k} (\hat{\phi}_i^k)^2 > 0. \]

Thus, \( ||\Gamma_i(\tau)||^2 \) is monotonically increasing for \( 0 < \tau < 1 \). Since \( ||\Gamma_2(t_2(\tau))||^2 = (\tau - 1)^2 \), \( ||\Gamma_2(t_2(\tau))|| \) is certainly increasing for \( \tau \geq 1 \). Since \( ||h_k(\tau)|| = ||P(x_k, B_k)w_k(\tau)|| \) and \( w_k(\tau) = E_k^T L_k^{-1} \tilde{w}_k(\tau) \), as above we have that the norm function of the preconditional modified gradient path is monotonically increasing for \( \tau \in [0, +\infty) \).

If \( \tau \geq 1 \) then \( t_i(\tau) = +\infty \), that is, \( \hat{\phi}_i^k = 0 \) \( \forall i \in \mathcal{I}_k \cup \mathcal{N}_k \), the term \( \Gamma_2(t_2(\tau)) \) is relevant. In that case, by

\[ \lim_{\tau \to \infty} \frac{\exp\{-\phi_i^k t_i\} - 1}{\phi_i^k} = -\frac{1}{\phi_i^k}, \quad \text{if} \quad \phi_i^k > 0, \]

we get that, noting \( \nabla_i l(x_k, \hat{x}_k)^T h_k(\tau) = (g_i^k)^T h_k(\tau) \) and \( P_k \phi_i^k = P_k g_i^k \),

\[ (g_i^k)^T h_k(\tau) = (g_i^k)^T P_k E_k^T L_k^{-1} \{ \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)) \} \]

\[ = -\sum_{i \in \mathcal{I}_k} \frac{1}{\phi_i^k} (P_k g_i^k)^2 + (\tau - 1)(P_k g_i^k)^1 \]

\[ = -\sum_{i \in \mathcal{I}_k} \frac{1}{\phi_i^k} (P_k g_i^k)^2 \]

\[ \leq -\frac{||P_k g_i^k||^2}{||D_k||} \]  \hspace{1cm} (2.20)

which means that \( h_k(\tau) \) satisfies a sufficiently descent direction.

If \( 0 < \tau < 1 \), then we have that

\[ \tilde{\psi}_k(\tau) = (g_i^k)^T P_k E_k^T L_k^{-1} \Gamma_1(t_1(\tau)) \]

\[ = \sum_{i \in \mathcal{I}_k} \exp\{-\phi_i^k t_i(\tau)\} \frac{1}{\phi_i^k} (P_k g_i^k)^2 - t_i(\tau) \sum_{i \in \mathcal{N}_k} (P_k g_i^k)^2. \]  \hspace{1cm} (2.21)

Then we have that

\[ \frac{d\tilde{\psi}_k(\tau)}{d\tau} = -\sum_{i \in \mathcal{I}_k} \frac{1}{(1-\tau)^2} \exp\{-\phi_i^k t_i(\tau)\} (P_k g_i^k)^2 - \frac{1}{(1-\tau)^2} \sum_{i \in \mathcal{N}_k} (P_k g_i^k)^2 < 0 \]  \hspace{1cm} (2.22)

which means that \( \tilde{\psi}_k(\tau) \) is monotonically decreasing for \( 0 < \tau < 1 \). Taking \( \tau \to 0 \), (2.17) is true.
If $B_k$ is positive, then by the definition of the preconditional modified gradient path $\Gamma(\tau)$ which can be expressed as

$$
\Gamma_k(\tau) = \sum_{i \in J_k} \exp\{-\varphi^k_i t^k_{i}(\tau)\} - 1 \phi^k_i u^i,
$$

as $\tau \to 1$

$$
= -D_k^{-1} \phi^k.
$$

which means that (2.18) holds.

It is clear that (2.19) holds as $t^k_{i}(\tau) \to +\infty$, if $\tau \geq 1$, and hence

$$
\lim_{\tau \to \infty} \exp\{-t^k_{i}(\tau)\} = 0.
$$

3. Algorithms

In this section we describe nonmonotone improved secant algorithms via preconditional curvilinear paths. In order to decide the acceptance of the new point at each iteration, it is necessary to introduce a merit function. We denote the merit functions $\vartheta(\cdot, \rho)$ and its directional derivative in the direction $d$, by $D\vartheta(\cdot, \rho; d)$, where $\rho$ is the penalty parameter. Here we choose the $l_1$ exact penalty function as the merit function $\vartheta(\cdot, \rho)$,

$$
\vartheta(x, \rho) = f(x) + \sum_{i=1}^{m} \rho_i |c_i(x)|, \quad (3.1)
$$

where $\rho_i$ is the $i$th component of the penalty vector $\rho$ which is updated and given as follows:

$$
\rho_{(k+1)} = \begin{cases} 
\rho_k & \text{if } \rho_k \geq |\Psi_k| + \kappa, \\
\max\{\rho_k, |\Psi_k|\} + \kappa & \text{otherwise},
\end{cases} \quad (3.2)
$$

where $\kappa > 0$ is the given constant, $\rho_k$ is the $i$th component of the vector $\rho_k$, and $\Psi_k = \max\{\lambda^p_k, \lambda^S_k\}$ with $\lambda^p_k$ and $\lambda^S_k$ being the $i$th component of the vectors $\lambda^p_k$ and $\lambda^S_k$, respectively.

Since the functions $c_i(x_k + s_k(\tau))$, $i = 1, 2, \ldots, m$, are differentiable, we have that

$$
c_i(x_k + s_k(\tau)) = c_i(x_k) + \tau \nabla c_i(x_k)^T s'_k(\xi \tau),
$$

where $\xi \in (0, 1)$. If $s_k(\tau)$ is obtained by the preconditional optimal path, then

$$
\nabla c_i(x_k)^T s'_k(\tau) = -\frac{1}{(1 + \tau)^2} \nabla c_i(x_k)^T A(x_k)^T c(x_k) = -\frac{1}{(1 + \tau)^2} c_i(x_k)
$$

and if $s_k(\tau)$ is obtained by the preconditional modified gradient path, then

$$
\nabla c_i(x_k)^T s'_k(\tau) = -\frac{\exp\{-t^k_{i}(\tau)\}}{(1 - \tau)^2} \nabla c_i(x_k)^T A(x_k)^T c_k = -\frac{\exp\{-t^k_{i}(\tau)\}}{(1 - \tau)^2} c_i(x_k).
$$

From above

$$
\lim_{\tau \to 0} \frac{|c_i(x_k + p_k(\tau))| - |c_i(x_k)|}{\tau} = -|c_i(x_k)|, \quad i = 1, 2, \ldots, m,
$$

where $p_k(\tau)$ is obtained by the preconditional optimal path.
which implies that its directional $|c_i(x)|$ derivative in the direction $s_i'(0)$, we have that
\[ D[|c_i(x_i; s_i'(0))|] = -|c_i(x_i)|, \quad i = 1, 2, \ldots, m. \]

Using (1.12), (1.13), (2.11) and (2.18), we obtain
\[ g_k^Tv_k(\tau) = -\bar{z}_k(\tau)g_kA(x_k)^Tc_k = -\bar{z}_k(\tau)\lambda_k^Tc_k, \]
where
\[ \bar{z}_k(\tau) = \begin{cases} \frac{1}{(\tau+1)^2} & \text{if preconditional optimal path,} \\ \exp\{-\tilde{t}_k(\tau)\} & \text{if preconditional modified gradient path} \end{cases} \quad (3.3) \]
and
\[ \lambda_k = \begin{cases} \lambda_k^P A(x_k)^T & \text{given in (1.12),} \\ \lambda_k^S A(x_k)^T & \text{given in (1.13).} \end{cases} \quad (3.4) \]

Hence from (3.1), we obtain that from $\bar{z}_k(0) = 1$,
\[ D\partial_k(x_k, \rho; s_i'(0)) = (g_k^T)^2h_k(0) - \sum_{i=1}^{m} \rho_{(k+1)}|c_i(x_k)| + \lambda_k c_i(x_k). \quad (3.5) \]

The difficulty with nondifferentiable penalty function $\partial(\cdot, \rho)$ by caused by the Maratos effect (see [14]). In order to overcome the Maratos effect and to ensure a superlinear convergence rate, the algorithm is added a correction step denoted in the correction vector $d_k(\tau)$ where the form is given by solving the following equation:
\[ d(\tau) = -A(x_k)^Tc(x_k + s_k(\tau)), \quad (3.6) \]
where $s_k(\tau) = h_k(\tau) + v_k(\tau)$. The effect of this correction step, which is normal to the constraints, is to decrease the quantity $||c(x)||$ so that it is in the order of $||x_k - x^*||^2$. This means that the merit function will then be decreased at the point $x_k + s_k(\tau) + z_k(\tau)d_k(\tau)$, as we will show below. Set then
\[ x_{k+1} = x_k + s_k(\tau_k) + z_k(\tau_k)d_k(\tau_k), \quad (3.7) \]
where $z_k(\tau)$ is given in (3.12) and $\tau_k$ satisfies the acceptance criterion (3.11).

Now, we describe the improved secant algorithms via nonmonotone preconditional curvilinear paths.

Initialization step: Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0, 1)$, and positive integer $M$. Let $m(0) = 0$. Choose a symmetric matrix $B_0$, give a starting point $x_0 \in \mathbb{R}^n$ and a positive penalty weight vector $\rho_0 \in \mathbb{R}^m$. Set $k = 0$, go to the main step.

Main Step:

1. Calculate $f_k$, $g_k$, $c_k$, and $A(x_k)^T$. Compute the multiplier $\lambda_{k+1}$ can be chosen from one of the following updates: (1) projection update (1.16), (2) null-space update (1.17), and (3) Newton update (1.18).
2. If $||c_k|| + ||P_k g_k|| \leq \varepsilon$ (where $\varepsilon$ is a given small constant), stop.
3. Factorize $B_k$ into the form (2.2) using stable Bunch–Parlett factorization method, that is, $E_kB_kE_k^T = L_kD_kL_k^T$. Calculate the eigenvalues and orthonormal eigenvectors of $D_k$. From either the preconditional optimal path or preconditional modified gradient path, we obtain $\Gamma_k(\tau)$, and $\Gamma_k^*(\tau)$, respectively.
4. Denote by $\tilde{w}_k$, $v_k$ the steps of the preconditional curvilinear paths $\Gamma_k(\tau)$, $\Gamma^*_k(\tau)$, respectively. Compute

$$w_k(\tau) = E_k^T L_k^{-1} \tilde{w}(\tau),$$

$$h_k(\tau) = P(x_k B_k) w_k(\tau),$$

$$s_k(\tau) = h_k(\tau) + v_k(\tau),$$

then solve (3.6) to obtain $d_k(\tau)$. Update the penalty vector $\rho$ by using the following formula given in this section. Set

$$p_k(\tau) = s_k(\tau) + z_k(\tau) d_k(\tau).$$

5. Choose

$$\tau_k = \infty, \omega^{-n}, \omega^{-(n-1)}, \ldots,$$

until the following inequality is satisfied:

$$\vartheta(x_k + p_k(\tau_k), \rho_k) \leq \vartheta(x_{\ell(k)}, \rho_{\ell(k)}) + \beta z_k(\tau_k) D \vartheta(x_k, \rho_k; s_k'(0)),$$

where $\vartheta(x_{\ell(k)}, \rho_{\ell(k)}) = \max_{0 \leq j \leq m(k)} \{\vartheta(x_{k-j}, \rho_{k-j})\}$ and

$$z_k(\tau) = \begin{cases} \frac{\tau}{\tau + 1}, & \text{if preconditional optimal path,} \\ 1 - \exp\{-t_1(\tau)\}, & \text{if preconditional modified gradient path.} \end{cases}$$

6. Set

$$x_{k+1} = x_k + p_k(\tau_k).$$

7. Take $m(k + 1) = \min\{m(k) + 1, M\}$, and update $B_k$ to obtain $B_{k+1}$. Then set $k \leftarrow k + 1$ and go to step 2.

**Remark 1.** Note that in each iteration if the solution $p_k(\infty)$ fails to meet the acceptance criterion (3.11) (take $\tau_k = \infty$), then we turn to curvilinear search, i.e., retreat from $x_k + p_k(\tau_k)$ until the criterion is satisfied.

**Remark 2.** Generally, $\{\vartheta(x_k + p_k(\tau_k), \rho_k)\}$ is nonmonotonically decreasing so that the proposed algorithm becomes the usual monotone algorithm when $M = 0$.

**Remark 3.** As shown below, the preconditional curvilinear paths can be generated by employing general symmetric matrices which may be indefinite. Thus the matrix $B_k$ at step 7 can be produced from evaluating the exact Hessian matrix $B_k = \nabla^2 f(x_k)$, or using an approximate Hessian. The factorization of matrix may not be required and only computation of eigenvalues of matrix $B_k$ is necessary at each iteration.
4. Convergence analysis

In order to discuss the convergence properties of the proposed algorithms, we should make the following assumptions in this section.

Assumption A1. The sequences \( \{x_k\}, \{x_k + s_k\} \) and \( \{x_k + s_k + d_k\} \) generated by the algorithms are contained in a compact convex set \( \Omega \); \( \nabla^2 f(x) \) and \( \nabla^2 c_i(x) (i = 1, \ldots, m) \) are Lipschitz continuous matrix functions on \( \Omega \).

Assumption A2. The matrix \( A(x) \) has full column rank over \( \Omega \).

Assumption A3. Norm sequences \( \{|A(x_k)^\dagger|\} \) is bounded for all \( k \), that is, there is a constant \( \beta_1 > 0 \) such that
\[ ||A(x_k)^\dagger|| \leq \beta_1, \quad \forall k, \quad (4.1) \]
where \( A(x_k)^\dagger \) is given in (1.12) and (1.13).

Lemma 4.1. Under the Assumption A3, we obtain
\[ d_k(\tau) = O(||s_k(\tau)||^2). \quad (4.2) \]

Proof. Since \( A_k^T h_k(\tau) = 0 \) and \( s_k(\tau) = h_k(\tau) + v_k(\tau) \), it means
\[ A_k^T s_k(\tau) = A_k^T (h_k(\tau) + v_k(\tau)) = A_k^T v_k(\tau) = -c_k. \quad (4.3) \]

Therefore,
\[ ||c(x_k + s_k(\tau))|| = ||c_k + A_k^T s_k(\tau)|| + O(||s_k(\tau)||^2) = O(||s_k(\tau)||^2). \quad (4.4) \]

Using (4.1), (3.6) implies that (4.2) holds. \( \square \)

Under the above assumptions, we can state the following result, whose proof can be found in [14].

Lemma 4.2. \( P(x_k, B_k)g^k = 0 \) and \( c(x_k) = 0 \) if and only if \( x_k \) is a Kuhn–Tucker point of problem (1.1).

To establish the convergence properties of the proposed algorithm, we are now ready to employ one of our main lemma.

Lemma 4.3. Assume that Assumptions A1–A3 holds. Let \( \{x_k\} \in \mathbb{R}^n \) be a sequence generated by the algorithm using the \( l_1 \) exact penalty function \( \vartheta(x_k, \rho_k) \) as merit function with (3.2) updating the penalty vector \( \rho_k \), we have that
\[ D\vartheta(x_k, \rho; s'_k(0)) \leq -\omega_k(\tau)||P(x_k, B_k)g^k||^2 - \kappa \sum_{i=1}^m |c_i(x_k)|, \quad (4.5) \]
where
\[
\omega_k(\tau) = \begin{cases} 
\min_{i \in J_k} \{\frac{1}{(\phi_i^k \tau + 1)}\}, & \text{if preconditional optimal path,} \\
\frac{1}{(1-\tau)^2} \min_{i \in J_k} \{\exp\{-\phi_i^k t_i(\tau)\}, 1\}, & \text{if preconditional modified gradient path.}
\end{cases}
\]

Furthermore, there exists \( \kappa_1 > 0 \) such that \( 0 < \tau_k < 1/T \).
\[
D\vartheta(x_k, \rho; s'_k(0)) \leq -\omega_k^1(\tau)\|h_k(\tau)\|^2 - \kappa_1\|v_k(\tau)\|, \tag{4.6}
\]
where \( \omega_k^1(\tau) = \hat{\omega}_k(\tau)\omega_k(\tau) \) and
\[
\hat{\omega}_k(\tau) = \begin{cases} 
\min_{i \in J_k} \{\frac{1}{(\phi_i^k \tau + 1)}\}, & \text{if optimal path,} \\
\min_{i \in J_k} \{\exp\{-\phi_i^k t_i(\tau)\} - 1\}, & \text{if modified gradient path.}
\end{cases}
\]

**Proof.** By (2.14) and (2.22), let \( h_k(\tau) \) be generated by either the preconditional optimal path or the preconditional modified gradient path, we have that
\[
(P_k g_k)(h_k)(0) \leq -\omega_k(\tau)\|P(x_k, B_k)g^k\|^2.
\]
From (3.2) updating the penalty vector \( \rho_k \), we have that (4.5) holds.

Let the step \( \Gamma_k(\tau_k) \) be obtained from the preconditional optimal path. Then we have that the norm function of \( P_k \Gamma_k(\tau) = h_k(\tau) \) with \( 0 < \tau_k < 1/T \) has
\[
\|h_k(\tau)\|^2 = \|P_k \Gamma_k(\tau)\|^2 = \sum_{i \in J_k} \left(\frac{\tau}{\phi_i^k \tau + 1}\right)^2 (P_k g^k)_i^2 + \tau^2 \sum_{i \in J_k} (P_k g^k)_i^2
\]
\[
\leq \max_{i \in J_k} \left\{\left(\frac{\tau}{\phi_i^k \tau + 1}\right)^2, \tau^2\right\} \|P_k g^k\|^2.
\]

Let the step \( \Gamma_k(\tau_k) \) be obtained from the preconditional modified gradient path. Then similar to the above, we have the norm function of the path
\[
\|h_k(\tau)\|^2 = \|P_k \Gamma_k(\tau)\|^2 \leq \max_{i \in J_k} \left\{\left(\frac{\exp\{-\phi_i^k t_i(\tau)\}}{\phi_i^k} - 1\right)^2, \left(\frac{\tau}{1-\tau}\right)^2\right\} \|P_k g^k\|^2. \tag{4.8}
\]

Since \( \|c_k\| \geq \|\tilde{c}_k\| \), \( 0 < \tilde{c}_k(\tau) \leq 1 \) and \( \|v_k(\tau)\| \leq \beta_1 \tilde{c}_k(\tau)\|c_k\| \leq \beta_1 \|c_k\| \), where \( \beta_1 \) is given in (4.1), we further conclude that taking \( \kappa_1 = \kappa/\beta_1 \), (4.6) holds. \( \Box \)

**Lemma 4.4.** Assume that Assumptions A1–A3 hold. Let \( \{\rho_k\} \subset \mathbb{R}^m \) be a penalty vector sequence generated by the algorithm using the updating formula. Then there exists a vector \( \bar{\rho} > 0 \) such that for all large \( k \), \( \rho_k \leq \bar{\rho} \), i.e., the sequence \( \{\rho_k\} \) is uniformly bounded from above.

**Proof.** Let the penalty vector \( \rho_k \) be a sequence generated by the algorithm using the \( l_1 \) exact penalty function \( \hat{\vartheta}(x_k, \rho_k) \) as merit function with (3.1). Referring to Assumptions A1–A3, \( \|\hat{\lambda}_k\| \) is uniformly
bounded for all \( x_k \). By (3.2) it follows that the sequence \( \{\rho_k\} \) is uniformly bounded from above, that is, there exists a vector \( \tilde{\rho} > 0 \) such that \( \rho_k \leq \tilde{\rho} \). \( \square \)

To establish the convergence properties of the proposed algorithm, we are now ready to state our main theorem.

**Theorem 4.5.** Suppose that Assumptions A1–A3 hold. Let \( \{x_k\} \subset \mathbb{R}^n \) be a sequence generated by the algorithm. Then

\[
\liminf_{k \to \infty} \{|P(x_k, B_k)g_k| + |c(x_k)||\} = 0. \tag{4.9}
\]

Further, no limit of the sequence \( \{x_k\} \) is a local maximum of the merit function \( \vartheta(x, \rho) \).

**Proof.** Eq. (4.9) holds if and only if the following equation holds:

\[
\liminf_{k \to \infty} \{|P(x_k, B_k)g_k|^2 + |c(x_k)||\} = 0.
\]

If the conclusion of the theorem is not true, without loss of generality, then assume that there exists some \( \varepsilon > 0 \) such that

\[
|c_k| + |P_kg^k| \geq \varepsilon, \quad k = 1, 2, \ldots .
\]

According to the acceptance rule (3.11) in step 5, we have

\[
\vartheta(x_{l(k)}, \rho_{l(k)}) - \vartheta(x_k + p_k(\tau_k), \rho_k) \leq \beta z_k(\tau_k)D\vartheta(x_k, \rho_k; s_k'(0)). \tag{4.10}
\]

Taking into account that \( m(k + 1) \leq m(k) + 1 \), and from Lemma 4.4 there exists \( \tilde{\rho} > 0 \) such that for all large \( k \), \( \rho_k \equiv \tilde{\rho} \), we have that \( \vartheta(x_{k+1}, \rho_{k+1}) \leq \vartheta(x_{l(k)}, \rho_{l(k)}) \). Similar to the proof of theorem in [7], we have that the sequence \( \{\vartheta(x_{l(k)}, \rho_{l(k)})\} \) is nonincreasing for all large \( k \), and therefore \( \{\vartheta(x_{l(k)}, \rho_{l(k)})\} \) is convergent.

By (3.11) and (4.5), for all \( k > M \),

\[
\vartheta(x_{l(k)}, \rho_{l(k)})
\]

\[
= \vartheta(x_{l(k)-1} + p_{l(k)-1}(\tau_{l(k)-1}), \rho_{l(k)-1})
\]

\[
\leq \max_{0 \leq j \leq m(l(k)-1)} \{\vartheta(x_{l(k)-1} - j, \rho_{l(k)-1} - j)\} + \beta z_{l(k)-1}(\tau_{l(k)-1})D\vartheta(x_{l(k)-1} - j, \rho_{l(k)-1} - j; s_{l(k)-1}'(0))
\]

\[
\leq \max_{0 \leq j \leq m(l(k)-1)} \{\vartheta(x_{l(k)-1} - j, \rho_{l(k)-1} - j)\}
\]

\[
- \beta \omega_{l(k)-1}(\tau_{l(k)-1})z_{l(k)-1}(\tau_{l(k)-1})\{||P_{l(k)-1}g^{l(k)-1}||^2 + \kappa||c_{l(k)-1}||\}. \tag{4.11}
\]

As \( \{\vartheta(x_{l(k)}, \rho_{l(k)})\} \) is convergent, and \( \omega_{l(k)-1}(\tau_{l(k)-1}) \neq 0 \) we obtain that from (4.10) and (4.11)

\[
\lim_{k \to \infty} \vartheta_{l(k)-1}(\tau_{l(k)-1})[||P_{l(k)-1}g^{l(k)-1}||^2 + \kappa||c_{l(k)-1}||] = 0. \tag{4.12}
\]

This, by the definitions of preconditional curvilinear paths (2.7)–(2.8) and (2.9)–(2.10), as well as \( ||c_k|| + ||P_kg^k|| \geq \varepsilon \), implies that

\[
\lim_{k \to \infty} \vartheta_{l(k)-1}(\tau_{l(k)-1}) = 0. \tag{4.13}
\]
By definition (3.12) of \( \tau_{l(k)-1} \), (4.13) means that
\[
\liminf_{k \to \infty} \tau_{l(k)-1} = 0. \tag{4.14}
\]
From the definitions of preconditional curvilinear paths (2.6)--(2.7) and (2.8)--(2.9), this implies that
\[
\lim_{k \to \infty} ||p_{l(k)-1}(\tau_{l(k)-1})|| = 0 \tag{4.15}
\]
and
\[
\lim_{k \to \infty} \vartheta(x_{l(k)-j}, \rho_{l(k)-j}) = \lim_{k \to \infty} \vartheta(x_{l(k)}, \rho_{l(k)}) \tag{4.16}
\]
for any positive integer \( j \). Furthermore, as large \( k \geq l(k) \geq k - M \), from
\[
x_{l(k)} = x_{k-M} + p_{k-M-1}(\tau_{k-M-1}) + \cdots + p_{l(k)-1}(\tau_{l(k)-1})
\]
and (4.16), it can be derived that
\[
\lim_{k \to \infty} \vartheta(x_{l(k)}, \rho_{l(k)}) = \lim_{k \to \infty} \vartheta(x_{k}, \rho_{k}). \tag{4.17}
\]
By the rule for accepting the step \( p_k(\tau_k) \), (4.1) and (4.3),
\[
\vartheta(x_{k+1}, \rho_{k+1}) - \vartheta(x_{l(k)}, \rho_{l(k)}) \leq \beta g_k(\tau_k) D\vartheta(x_k, \rho_k; s_k'(0)) \leq -\beta \omega_k(\tau_k) g_k(\tau_k) \{||P_k g_k||^2 + \kappa ||c(x_k)||\}. \tag{4.18}
\]
Eqs. (4.17) and (4.18) mean that
\[
\lim_{k \to \infty} \vartheta(x_{k}) = 0, \quad \text{and hence} \quad \lim_{k \to \infty} \tau_{k} = 0.
\]
Therefore,
\[
\lim_{k \to \infty} ||p_k(\tau_k)|| = 0 \tag{4.19}
\]
which establishes
\[
\lim_{k \to \infty} ||x_{k+1} - x_k|| = 0. \tag{4.20}
\]
From \( \tau_k \to 0 \), as \( k \to \infty \), the acceptance rule (3.11) means that, for large enough \( k \), without loss of generality, \( p_k \equiv \bar{\rho} \).
\[
\vartheta \left( x_k + p_k \left( \frac{\tau_k}{\omega} \right), \bar{\rho} \right) - \vartheta(x_k, \bar{\rho}) \geq \vartheta \left( x_k + p_k \left( \frac{\tau_k}{\omega} \right), \bar{\rho} \right) - \vartheta(x_{l(k)}, \bar{\rho})
\]
\[
> \beta g_k \left( \frac{\tau_k}{\omega} \right) D\vartheta(x_k, \rho_k; s_k'(0)). \tag{4.21}
\]
For the composite function \( \vartheta(x_k + p_k(\bar{\tau}_k), \bar{\rho}) \) about \( \tau \), we have that noting \( \vartheta_k(\bar{\tau}_k) d\bar{\tau}_k(0) = O(\tau_k/\omega) d\bar{\tau}_k(0) \),
\[
\vartheta \left( x_k + p_k \left( \frac{\tau_k}{\omega} \right), \bar{\rho} \right) = \vartheta(x_k, \bar{\rho}) + \frac{\tau_k}{\omega} \frac{d\vartheta(x_k + p_k(\tau_k), \bar{\rho})}{d\tau} \bigg|_{\tau=0} + o \left( \frac{\tau_k}{\omega} \right)
\]
\[
= \vartheta(x_k, \bar{\rho}) + \frac{\tau_k}{\omega} D\vartheta(x_k, \bar{\rho}; p_k'(0)) + o \left( \frac{\tau_k}{\omega} \right)
\]
\[
= \vartheta(x_k, \bar{\rho}) + \frac{\tau_k}{\omega} D\vartheta(x_k, \bar{\rho}; s_k'(0)) + o \left( \frac{\tau_k}{\omega} \right). \tag{4.22}
\]
From (4.21) and (4.22), we have
\[
\left[ \frac{\tau_k}{\omega} - \beta z_k \left( \frac{\tau_k}{\omega} \right) \right] D\theta(x_k, \tilde{p}; s'_k(0)) + o \left( \frac{\tau_k}{\omega} \right) \geq 0. \tag{4.23}
\]
Dividing (4.23) by $\frac{\tau_k}{\omega}$ and noting that $1 - \beta > 0$, $z_k(\frac{\tau_k}{\omega}) \div (\frac{\tau_k}{\omega}) \to 1$ as $\tau_k \to 0$, and $D\theta(x_k, \tilde{p}; s'_k(0)) \leq 0$, we obtain
\[
\lim_{k \to \infty} D\theta(x_k, \tilde{p}; s'_k(0)) = 0. \tag{4.24}
\]
From (4.5), (4.24) means that, from $\frac{\tau_k}{\omega} \to 0$ as $k \to \infty$,
\[
\lim_{k \to \infty} \left\{ \| P_k g^k \|^2 + \| c_k \| \right\} = \lim_{k \to \infty} D\theta(x_k, \tilde{p}; s'_k(0)) = 0 \tag{4.25}
\]
which implies that (4.9) is true.

Assume that there exists a limit point $x_*$ which is a local maximum of $f$, let $\{x_k\}_k$ be a subsequence of $\{x_k\}$ converging to $x_*$. Then the limit point $x_*$ is also a local maximum of the merit function $\theta(x, \tilde{p})$, for large $\tilde{p}$.

As $k \geq (k) \geq k - M$, for any $k$ there exists a point $x(k)$ such that
\[
\lim_{k \to \infty} \| x(k) - x_k \| = 0, \tag{4.26}
\]
so that we can obtain
\[
\lim_{k \to \infty} \| x(k) - x_* \| \leq \lim_{k \to \infty} \| x_* - x_k \| + \lim_{k \to \infty} \| x(k) - x_k \| = 0. \tag{4.27}
\]
This means that also the subsequence $\{x_k\}_k$ converges to $x_*$. On the other hand, we have that the sequence $\{\theta(x(k), \rho(k))\}$ is nonincreasing for all large $k$, and therefore $\{\theta(x(k), \rho(k))\}$ is convergent, noting $c_* = 0$, i.e.,
\[
\lim_{k \to \infty} \{\theta(x(k), \rho(k))\} = \theta(x_*, \tilde{p}) = f(x_*) \tag{5.1}
\]
and $\{\theta(x(k), \rho(k))\} \geq f(x_*)$ for all large $k$. Moreover, we can find a sufficiently large index $k$ such that
\[
f(x_*) \leq \theta(x_{(k+M)}, \rho_{(k+M)}) < \theta(x(k), \rho(k)) \tag{4.28}
\]
so that we conclude that, in any neighborhood of $x_*$, there exists a point $x(k)$ with $k \in \mathcal{N}$ such that $f(x_*) < \theta(x(k), \tilde{p})$ holds. This contradicts the assumption that $x_*$ is a local maximum of the merit function $\theta(x, \tilde{p})$, for large $\tilde{p}$. It means that the conclusion of the theorem is true. □

5. Local convergence rates

Theorem 4.5 indicates that at least one limit point of $\{x_k\}$ is a stationary point. Next, we show that the convergence rate is superlinear convergence for the algorithm when $B_k$ is positive definite. It requires the following assumptions.

Assumption A4. $x_*$ is a K-K-T point of problem (1.1), i.e., there is a vector $\lambda_* \in \mathbb{R}^m$ such that
\[
c_* = 0 \quad \text{and} \quad g_* - A_* \lambda_* = 0. \tag{5.1}
\]
Assumption A5. There exists a constant $\bar{\tau} > 0$ such that
\[
p^T(Z^TWZ_p)p \geq \bar{\tau}\|p\|^2, \quad \forall p \in \mathbb{R}^l
\] (5.2)
i.e., the second-order sufficient condition holds at $x_*$. Further, there exists a constant $\tilde{\sigma} > 0$ such that $\|W(x)\| \leq \tilde{\sigma} \quad \forall x \in \Omega$, where $W(x)$ given in (1.19).

Assumption A6. $x_k \rightarrow x_*$.

Assumption A7.
\[
\lim_{k \rightarrow \infty} \frac{\|P_h(B_k - W_s)h_k\|}{\|h_k\|} = 0.
\] (5.3)
In fact, (5.3) is a sufficient condition of the secant methods for two-step $q$-superlinear convergence (see [6,10], for instant).

Lemma 5.1. If $B_k$ is eventually positive definite and Assumptions A1–A7 hold, then $h_k(\tau)$ generated from the two projection curvilinear paths $P(x_k, B_k)E_k^TL_k^{-1}\Gamma_k(\tau_k)$ satisfies
\[
\hat{q}_k(h_k(\tau_k)) \equiv (g^k)^T h_k(\tau_k) + h_k(\tau_k)^TB_k h_k(\tau_k) \leq 0.
\] (5.4)

Proof. At the $k$th iteration, let $\Gamma_k(\tau_k)$ be generated from the preconditional optimal path. When $B_k$ is positive definite, by the definition of the preconditional optimal path $\Gamma(\tau)$, it can be expressed as
\[
B_k P_hE_k^T L_k^{-1}\Gamma_k(\tau_k) = B_k P_hE_k^T L_k^{-1}\Gamma_1(t_1(\tau_k)) = -\sum_{i=1}^n \frac{\phi_i^k u_i^k(u_i^k)^T}{\phi_i^k t_1(t_1(\tau_k) + 1)(P_k g^k)u_i^k}.
\] (5.5)
Hence, we have that
\[
(g^k)^T h_k(\tau_k) + h_k(\tau_k)^T B_k h_k(\tau_k)
\]
\[
= -\sum_{i \in \mathcal{I}_k} \frac{\tau_k}{\phi_i^k t_1(t_1(\tau_k) + 1)(P_k g^k)u_i^2} + \frac{\phi_i^k \tau_k^2}{(\phi_i^k t_1(\tau_k) + 1)^2 (P_k g^k)^2}
\]
\[
= -\sum_{i \in \mathcal{I}_k} \frac{\tau_k}{(\phi_i^k t_1 + 1)^2 (P_k g^k)^2}
\]
\[
\leq 0.
\] (5.6)
Let the step $\tilde{w}_k$ be obtained from the preconditional modified gradient path. Since $B_k$ is positive definite, it means that $\mathcal{S}_k \cup \mathcal{N}_k = \phi$. We define the value of
\[
\tilde{q}_k(h_k) \equiv (g^k)^T h_k + \frac{1}{2} h_k^T B_k h_k,
\]
where noting \( h_k = h_k(\tau_k) \), for simplicity, \( \tilde{q}_k \) along the precoditional modified gradient path is given by (see, (4.6) in [2]),

\[
\tilde{q}_k(P_k \Gamma_1(t_1(\tau))) = \sum_{i \in \mathcal{I}_k} \exp\left\{\frac{-2 \phi_i^k t_1(\tau)}{\varphi_i^k}\right\} - \frac{1}{\varphi_i^k}(P_k g^k_i)^2. \tag{5.7}
\]

It is clear to see that

\[
(g^k)^T h_k = (g^k)^T [P_k E^T_k L_k^{-1} \Gamma_1(t_1(\tau))] = \sum_{i \in \mathcal{I}_k} \exp\left\{\frac{-\phi_i^k t_1(\tau)}{\varphi_i^k}\right\} - \frac{1}{\varphi_i^k}(P_k g^k_i)^2. \tag{5.8}
\]

We have that from (5.7) and (5.8),

\[
(g^k)^T h_k + h_k^T B_k h_k = 2[(g^k)^T h_k + \frac{1}{2} h_k^T B_k h_k] - (g^k)^T h_k
\]

\[
= 2 \sum_{i \in \mathcal{I}_k} \exp\left\{\frac{-2 \phi_i^k t_1(\tau)}{\varphi_i^k}\right\} - \frac{1}{\varphi_i^k}(P_k g^k_i)^2 - \sum_{i \in \mathcal{I}_k} \exp\left\{\frac{-\phi_i^k t_1(\tau)}{\varphi_i^k}\right\} - \frac{1}{\varphi_i^k}(P_k g^k_i)^2
\]

\[
= \sum_{i \in \mathcal{I}_k} \exp\left\{\frac{-\phi_i^k t_1(\tau)}{\varphi_i^k}\right\} \exp\left\{\frac{-\phi_i^k t_1(\tau)}{\varphi_i^k}\right\} - \frac{1}{\varphi_i^k}(P_k g^k_i)^2
\]

\[
\leq 0. \tag{5.9}
\]

From (5.6) and (5.9), we have that the conclusion of the theorem is true. \( \square \)

**Lemma 5.2.** Let \( \{x_k\} \subset \mathbb{R}^n \) be a sequence generated by the algorithm using the \( l_1 \) penalty function \( \vartheta(x_k, p_k) \) as merit function. If \( B_k \) is eventually positive definite and Assumptions H1–H7 hold, then for sufficiently large \( k \), \( \tau_k \equiv \infty \) in (3.11).

**Proof.** We denote \( \tilde{w}_k(\infty) \) by \( \tilde{w}_k \), whenever it does not lead to confusion for simplicity. Similarly, \( h_k(\infty) \) is denoted by \( h_k \), and \( v_k(\infty) \) by \( v_k \). By the definition of \( d_k \) and \( ||c(x_k + s_k)|| = O(||s_k||^2) \), we have that

\[
g_k^T d_k = \lambda_k^T A_k c(x_k + s_k) - \varepsilon_k c_k^T (A_k^T B_k^{-1} A_k)^{-1} c(x_k + s_k)
\]

\[
= \lambda_k^T A_k c(x_k + s_k) + o(||c_k||)
\]

\[
= \frac{1}{2} s_k^T \left( \sum_{i=1}^m \lambda_{k+1,i} \nabla^2 c_i(x_k) \right) s_k + o(||c_k||) + o(||h_k||^2)
\]

\[
= \frac{1}{2} h_k^T \left( \sum_{i=1}^m \lambda_{k+1,i} \nabla^2 c_i(x_k) \right) h_k + o(||v_k||) + o(||h_k||^2), \tag{5.10}
\]

where \( \xi_k = 0 \) or 1 and \( \lambda_k \) given in (1.16), (1.17) or (1.18).
Since, \(|W(x)| \leq \tilde{\omega} \forall x \in \Omega\) and noting (5.10),
\[
f(x_k + s_k + d_k) = f_k + g_k^T v_k + h_k + (s_k + d_k)^T (\nabla^2 f_k)(s_k + d_k) + o(||v_k||) + o(||s_k||^2)
\]
\[
= f_k + g_k^T v_k + h_k + \frac{1}{2} h_k^T (\nabla^2 f_k) h_k + \frac{1}{2} h_k^T (\nabla^2 f_k) h_k \\
+ \frac{1}{2} h_k^T \left( \sum_{i=1}^m \dot{\ell}_{k+1,i} \nabla^2 c_i(x_k) \right) h_k + o(||v_k||) + o(||s_k||^2)
\]
\[
= f_k + g_k^T h_k + g_k^T v_k + h_k + \frac{1}{2} h_k^T W_k h_k + o(||v_k||) + o(||h_k||^2)
\]
\[
c(x_k + s_k + d_k) = c(x_k + s_k) + A(x_k + s_k)^T d_k + o(||d_k||)
\]
\[
= [A(x_k + s_k) - A_k]^T d_k + o(||d_k||)
\]
\[
= o(||h_k||^2) + o(||c_k||)
\]
we have that, noting (5.4) and \(B_k\) being positive definite, i.e., \((g_k^k)^T h_k \leq -h_k^T B_k h_k \leq 0\)
\[
\vartheta(x_k + s_k + d_k, \rho_k) - \vartheta(x_k, \rho_k) - \beta D \vartheta(x_k, \rho_k; s_k)
\]
\[
= (\frac{1}{2} - \beta) g_k^T h_k + \frac{1}{2} h_k^T W_k h_k + (1 - \beta) g_k^T v_k + \frac{1}{2} g_k^T h_k \\
- (1 - \beta) \rho_k ||c_k|| + o(||c_k||) + o(||h_k||^2) + o(||s_k||^2)
\]
\[
\leq - (\frac{1}{2} - \beta) h_k^T B_k h_k - \frac{1}{2} h_k^T (B_k - W_k) h_k + (1 - \beta) g_k^T v_k
\]
\[
- (1 - \beta) \rho_k ||c_k|| + o(||c_k||) + o(||h_k||^2) + o(||s_k||^2)
\]
\[
\leq - (\frac{1}{2} - \beta) h_k^T B_k h_k + (1 - \beta) g_k^T v_k
\]
\[
- (1 - \beta) \rho_k ||c_k|| + o(||c_k||) + o(||h_k||^2),
\]
the last inequality is reduced by (5.3). Further, we have that
\[
g_k^T v_k = \begin{cases} 
(\dot{\ell}_k^p)^T c_k & \text{for ALG 1, ALG 3, and ALG 5,} \\
(\dot{\ell}_k^c)^T c_k & \text{for ALG 2, ALG 4, and ALG 6,} 
\end{cases}
\]
\[
\leq \begin{cases} 
||\dot{\ell}_k^p||_{\infty} \cdot ||c_k||_1 & \text{for ALG 1, ALG 3, and ALG 5,} \\
||\dot{\ell}_k^c||_{\infty} \cdot ||c_k||_1 & \text{for ALG 2, ALG 4, and ALG 6,} 
\end{cases}
\]
\[
\leq ||\Psi_k||_{\infty} \cdot ||c_k||_1.
\]
From (5.2) and (5.3), we have for large \(k\)
\[
\frac{\tau}{2} ||h_k||^2 \leq h_k^T B_k h_k.
Therefore, we obtain that, for $k$ large enough and by (3.2) and (5.14)

\[
\vartheta(x_k + s_k + d_k, \rho_k) - \vartheta(x_k, \rho_k) - \beta D \vartheta(x_k, \rho_k; s_k) \\
\leq - (\frac{1}{2} - \beta) \frac{\tau}{2} ||h_k||^2 + (1 - \beta)(\psi_k - \rho_k)||c_k||1 + o(||c_k||) + o(||h_k||^2) \\
\leq - (\frac{1}{2} - \beta) \frac{\tau}{2} ||h_k||^2 - (1 - \beta)\kappa||c_k||1 + o(||c_k||) + o(||h_k||^2) \\
\leq 0
\]

and hence

\[
\vartheta(x_k + s_k + d_k, \rho_k) - \vartheta(x_{(i)}, \rho_{(i)}) \\
\leq \vartheta(x_k + s_k + d_k, \rho_k) - \vartheta(x_k, \rho_k) \\
\leq \beta D \vartheta(x_k, \rho_k; s_k)
\]

then the step length $\tau_k \equiv \infty$, for sufficiently large $k$. This means that the theorem is true. □

**Theorem 5.3.** Under Assumptions A1−A7, the sequence $\{x_k\}$ of points generated by the improved secant algorithms with the nonmonotone acceptance criterion (3.11) is two-step $q$-superlinear.
Table 3
Experimental results of the algorithm ALG 3

| Problem name | $M = 0$ | | $M = 4$ | | $M = 8$ |
|--------------|---------|---|---|---|---|---|---|---|---|
|               | NF | NG | NF | NG | NO | NF | NG | NO | NF | NG | NO |
| POPP          |     |     |     |     |     |     |     |     |     |     |     |
| HS006        | 14  | 10 | 14  | 10 | 0  | 14  | 10 | 0  |     |     |     |
| HS026        | 16  | 16 | 14  | 14 | 1  | 14  | 14 | 1  |     |     |     |
| HS027        | 36  | 20 | 32  | 15 | 1  | 34  | 16 | 1  |     |     |     |
| HS028        | 57  | 31 | 57  | 31 | 0  | 57  | 31 | 0  |     |     |     |
| HS060        | 8   | 7  | 8   | 7  | 0  | 8   | 7  | 0  |     |     |     |
| SC220        | 146 | 115| 132 | 98 | 3  | 121 | 93 | 4  |     |     |     |
| SC252        | 56  | 49 | 66  | 48 | 2  | 69  | 52 | 2  |     |     |     |
| SC235        | 31  | 16 | 29  | 15 | 1  | 29  | 15 | 1  |     |     |     |
| SC219        | 63  | 51 | 56  | 47 | 3  | 51  | 42 | 4  |     |     |     |
| SC216        | 76  | 57 | 71  | 51 | 1  | 71  | 51 | 1  |     |     |     |
| PMGP          |     |     |     |     |     |     |     |     |     |     |     |
| HS006        | 14  | 10 | 14  | 10 | 0  | 14  | 10 | 0  |     |     |     |
| HS026        | 16  | 16 | 14  | 14 | 1  | 14  | 14 | 1  |     |     |     |
| HS027        | 38  | 22 | 35  | 17 | 1  | 35  | 17 | 1  |     |     |     |
| HS028        | 61  | 32 | 61  | 32 | 0  | 61  | 32 | 0  |     |     |     |
| HS060        | 8   | 7  | 8   | 7  | 0  | 8   | 7  | 0  |     |     |     |
| SC220        | 144 | 112| 134 | 95 | 4  | 91  | 76 | 7  |     |     |     |
| SC252        | 57  | 50 | 68  | 54 | 2  | 68  | 54 | 2  |     |     |     |
| SC235        | 31  | 16 | 29  | 15 | 1  | 29  | 15 | 1  |     |     |     |
| SC219        | 62  | 49 | 57  | 49 | 2  | 46  | 35 | 3  |     |     |     |
| SC216        | 77  | 58 | 71  | 51 | 1  | 71  | 51 | 1  |     |     |     |

convergence, i.e.,

$$\lim_{k \to \infty} \frac{||x_{k+1} - x^*||}{||x_k - x^*||} = 0. \quad (5.18)$$

Furthermore, the sequence $\{x_k + s_k\}$ of points generated by the improved secant algorithms is $q$-superlinear convergence, i.e.,

$$\lim_{k \to \infty} \frac{||x_k + s_k - x^*||}{||x_k + s_{k-1} - x^*||} = 0. \quad (5.19)$$

**Proof.** From Theorem 5.2, we get that $\tau_k \equiv \infty$ for all large enough $k$, then

$$x_{k+1} = x_k + s_k + d_k$$

where $s_k$ and $d_k$ are given by (3.9) and (3.6), that is, it means that $h_k$, $v_k$, $s_k$, and $d_k$ are generated by (1.4), (1.5), (1.7) and (1.8), respectively. From Fontecilla [6], we have that (5.18) is true. Furthermore, similar to the proof of Theorem 4.11 in [14], we can prove that (5.19) is also true. □
### Table 4
Experimental results of the algorithm ALG 6

<table>
<thead>
<tr>
<th>Problem name</th>
<th>$M = 0$</th>
<th></th>
<th>$M = 4$</th>
<th></th>
<th>$M = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NF</td>
<td>NG</td>
<td>NF</td>
<td>NG</td>
<td>NO</td>
</tr>
<tr>
<td>POPP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS006</td>
<td>14</td>
<td>10</td>
<td>14</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>HS026</td>
<td>16</td>
<td>16</td>
<td>14</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>HS027</td>
<td>38</td>
<td>21</td>
<td>34</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>HS028</td>
<td>58</td>
<td>31</td>
<td>58</td>
<td>31</td>
<td>0</td>
</tr>
<tr>
<td>HS060</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>SC220</td>
<td>142</td>
<td>103</td>
<td>128</td>
<td>91</td>
<td>3</td>
</tr>
<tr>
<td>SC252</td>
<td>54</td>
<td>48</td>
<td>65</td>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td>SC235</td>
<td>31</td>
<td>16</td>
<td>29</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>SC219</td>
<td>63</td>
<td>51</td>
<td>56</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>SC216</td>
<td>76</td>
<td>57</td>
<td>71</td>
<td>51</td>
<td>1</td>
</tr>
<tr>
<td>PMGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS006</td>
<td>14</td>
<td>10</td>
<td>14</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>HS026</td>
<td>16</td>
<td>16</td>
<td>14</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>HS027</td>
<td>58</td>
<td>47</td>
<td>51</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>HS028</td>
<td>65</td>
<td>42</td>
<td>65</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>HS060</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>SC220</td>
<td>145</td>
<td>113</td>
<td>136</td>
<td>93</td>
<td>4</td>
</tr>
<tr>
<td>SC252</td>
<td>59</td>
<td>52</td>
<td>69</td>
<td>56</td>
<td>2</td>
</tr>
<tr>
<td>SC235</td>
<td>31</td>
<td>16</td>
<td>29</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>SC219</td>
<td>62</td>
<td>49</td>
<td>57</td>
<td>49</td>
<td>2</td>
</tr>
<tr>
<td>SC216</td>
<td>79</td>
<td>64</td>
<td>72</td>
<td>52</td>
<td>1</td>
</tr>
</tbody>
</table>

6. Numerical experiments

Numerical experiments on the improved secant algorithms ALG 2, ALG 3 and ALG 6 with the nonmonotonic preconditional optimal path (POPP) and preconditional modified gradient path (PMGP), respectively, have been performed on an IBM 586 personal computer. We compare with different nonmonotonic parameters $M = 0, 4$ and 8, respectively, for the proposed algorithms. The monotonic algorithms are realized by taking $M = 0$. In order to check the effectiveness of the nonmonotonic technique, the selected parameter values are: $\beta = 0.1$, $\omega = 0.2$, $\kappa = 0.5$, $\rho_0 = 10$. The computation terminates when one of the following stopping criterions is satisfied: $||P(x_k, B_k)g_k|| + ||c_k|| \leq 10^{-6}$ and $||\vartheta(x_k, \rho_k) - \vartheta(x_{k+1}, \rho_{k+1})|| \leq 10^{-6}||\vartheta(x_k, \rho_k)||$.

Our preconditional curvilinear method is very easy to be resolved, since the full eigensystem of the matrix $D$ is very easy to calculate. Indeed, the formulation of curvilinear paths $I_k$ depend on the value of $\tau_k$, we only need to set the point back along the same curvilinear path until (3.11) is satisfied. The experiments are carried out on 10 standard test problems which are quoted from [8,11] (HS: the problems from Hock and Schittkowski [8], and SC: from Schittkowski [11]). The computational results for $B_k = H_k$, the real Hessian, are presented in Tables 2–4. POPP and PMGP with ALG 2, 3, 6 denote, respectively, the secant algorithms ALG 2, 3, 6 via the preconditional optimal path algorithm and the preconditional modified gradient path algorithm proposed with $l_1$ penalty function.
NF and NG stand for the numbers of function evaluations and gradient evaluations, respectively. NO stands for the number of iterations in which nonmonotonic decreasing situation occurs, that is, the number of times $\vartheta(x_k, \rho_k) < \vartheta(x_{k+1}, \rho_{k+1})$. The number of iterations is not presented in Table 4 because it always equals NG.

The last three parts of the table, under the headings of $M = 0$, 4 and 8, respectively, show that for most test problems the nonmonotonic technique does bring in some noticeable improvement.

Acknowledgements

The author would like to thank the two anonymous referees for their helpful comments and suggestions.

References