JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Journal of Computational and Applied Mathematics 136 (2001) 73–97

www.elsevier.com/locate/cam

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: 90 C 30; 65 K 05

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

1. Introduction

In this paper we consider the nonlinear equality constrained minimization problems

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0, \end{aligned} \tag{1.1}$$

where $f(x) : \mathbb{R}^n \to \mathbb{R}^1$ and $c(x) : \mathbb{R}^n \to \mathbb{R}^m$, $m \leqslant n$ 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 Technical Sicences Committee and Science Technology Foundation of Shanghai Higher Education.

PII: 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 ($Z_k \in \mathbb{R}^{n\times t}$, 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 Eudidean norm on $\mathbb{R}^n$. For simplicity, we denote $f(x_k)$ by f_k; $\nabla f(x_k)$ by g^k; and $\nabla^2_{xx} f(x_k)$ by $\nabla^2 f_k$, etc.

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

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

$$B_k w_k = -\nabla_x l(x_k, \lambda_k) \triangleq -\phi^k, \tag{1.3}$$

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

$$v_k = -A(x_k)^\dagger c_k, \tag{1.5}$$

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

$$s_k = h_k + v_k, \tag{1.7}$$

$$d_k = -A(x_k)^\dagger c(x_k + s_k), \tag{1.8}$$

$$B_{k+1} = \mathrm{DFP/BFGS}(h_k, y_k, B_k), \tag{1.9}$$

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

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

$$l(x, \lambda) = f(x) + \lambda^{\mathrm{T}} c(x). \tag{1.11}$$

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^{T}, $A(x_k)^\dagger$, will be equal to either of the two,

$$A_k^\dagger = A_k (A_k^{\mathrm{T}} A_k)^{-1} \tag{1.12}$$

or

$$A_{B_k}^\dagger = B_k^\dagger A_k (A_k^{\mathrm{T}} B_k^\dagger A_k)^\dagger, \tag{1.13}$$

Table 1
Algorithms

Algorithm	U	P	$A^{\dagger}$
Alg 1	(1.17)	I	(1.12)
Alg 2	(1.17)	I	(1.13)
Alg 3	(1.16)	P_B	(1.12)
Alg 4	(1.16)	P_B	(1.13)
Alg 5	(1.18)	P	(1.12)
Alg 6	(1.18)	P	(1.13)

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

$$P(x) = I - A(x)[A(x)^{\mathrm{T}}A(x)]^{-1}A(x)^{\mathrm{T}} \tag{1.14}$$

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

$$P_B(x) = I - B^{\dagger}A(x)[A(x)^{\mathrm{T}}B^{\dagger}A(x)]^{\dagger}A(x)^{\mathrm{T}}. \tag{1.15}$$

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

(1) *Projection update*:

$$\lambda_k^{\mathrm{P}} = -(A_k^{\mathrm{T}}A_k)^{-1}A_k^{\mathrm{T}}g_k, \tag{1.16}$$

(2) *Null-space update*:

$$\lambda_k^{\mathrm{S}} = -(A_k^{\mathrm{T}}B_k^{\dagger}A_k)^{\dagger}A_k^{\mathrm{T}}B_k^{\dagger}g_k, \tag{1.17}$$

(3) *Newton update*:

$$\lambda_k^{\mathrm{N}} = (A_k^{\mathrm{T}}B_k^{\dagger}A_k)^{\dagger}(c_k - A_k^{\mathrm{T}}B_k^{\dagger}g_k). \tag{1.18}$$

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_*, and $B_k \approx W_*$, then the iterative sequence of $\{x_k\}$ approaches x_*, where $W(x,\lambda)$ is defined as

$$W(x,\lambda) = \nabla_{xx}^2 l(x,\lambda). \tag{1.19}$$

Furthermore, under the assumption

$$\lim_{k\to\infty} \frac{\|P_k(B_k - W_*)h_k\|}{\|h_k\|} = 0 \tag{1.20}$$

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

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_{k-1} - x_*\|} = 0. \tag{1.21}$$

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 differentiable 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

$$\psi_k(\delta_k(\tau)) = \nabla_x l(x_k,\lambda_k)^{\mathrm{T}}\delta_k(\tau) + \tfrac{1}{2}\delta_k(\tau)^{\mathrm{T}}B_k\delta_k(\tau),$$

then the solution of

$$\frac{\mathrm{d}\delta_k(\tau)}{\mathrm{d}\tau} = -\nabla\psi_k(\delta_k(\tau)) \quad \text{and} \quad \delta_k(0) = 0 \tag{2.1}$$

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

$$E_kB_kE_k^{\mathrm{T}} = L_kD_kL_k^{\mathrm{T}}, \tag{2.2}$$

where E_k is a permutation matrix, L_k a unit lower triangular matrix and D_k a block diagonal matrix with 1×1 and 2×2 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

$$B_k^{\dagger} = L_k^{-\mathrm{T}}D_k^{\dagger}L_k^{-1},$$

where $D_k^{\dagger} = \mathrm{diag}\{(\varphi_i^k)^{\dagger}\}$ and

$$(\varphi_i^k)^{\dagger} = \begin{cases} (\varphi_i^k)^{-1} & \text{if } \varphi_i^k \neq 0, \\ 0 & \text{if } \varphi_i^k = 0. \end{cases}$$

There exist positive constants $\bar{c}_1$ and $\bar{c}_2$ such that for all k

$$\|L_k\| \leqslant \bar{c}_1, \qquad \|L_k^{-1}\| \leqslant \bar{c}_2. \tag{2.3}$$

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

$$w = L_k^{\mathrm{T}}E_k\tilde{w} \tag{2.4}$$

and the local quadratic approximation function at x_k,

$$\hat{\psi}_k(\tilde{w}) \equiv (\tilde{\phi}^k)^{\mathrm{T}}\tilde{w} + \tfrac{1}{2}(\tilde{w})^{\mathrm{T}}D_k\tilde{w}, \tag{2.5}$$

where $\tilde{\phi}^k = L_k^{-1}E_k\phi^k$. Note that $\tilde{w}$ rather than $w = E_k^{\mathrm{T}}L_k^{-\mathrm{T}}\tilde{w}$ 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 τ 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 1×1 and 2×2 diagonal blocks, its eigenvalues $\varphi_1^k,\varphi_2^k,\ldots,\varphi_n^k$ are real numbers and these are corresponding orthonormal eigenvectors $u_k^1,u_k^2,\ldots,u_k^n$. Without loss of generality, let $\varphi_1^k\leqslant\varphi_2^k\leqslant\cdots\leqslant\varphi_n^k$ be eigenvalues of D_k and $u_k^1,u_k^2,\ldots,u_k^n$ be corresponding orthonormal eigenvectors. We partition the set $\{1,\ldots,n\}$ into $\mathscr{I}_k^+$, $\mathscr{I}_k^-$ and $\mathscr{N}_k$ according to $\varphi_i^k>0$, $\varphi_i^k<0$ and $\varphi_i^k=0$ 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 $\mathcal{N}(A(x_k)^{\mathrm{T}})$, where $w_k = E_k^{\mathrm{T}} L_k^{-1} \tilde{w}_k$, 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

$$\Gamma(\tau) = \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)), \tag{2.6}$$

where

$$\Gamma_1(t_1(\tau)) = -\left[\sum_{i\in\mathcal{I}_k} \frac{t_1(\tau)}{\varphi_i^k t_1(\tau)+1} \tilde{\phi}_i^k u_k^i + t_1(\tau) \sum_{i\in\mathcal{N}_k} \tilde{\phi}_i^k u_k^i\right],$$

$$\Gamma_2(t_2(\tau)) = t_2(\tau) u_k^1$$

and

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

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

$\mathcal{I}_k = \{i \mid \varphi_i^k \neq 0,\ i = 1,\ldots,n\}$, $\mathcal{N}_k = \{i \mid \varphi_i^k = 0\ i = 1,\ldots,n\}$, $\tilde{\phi}_i^k = (\tilde{\phi}^k)^{\mathrm{T}} u_k^i$, $i = 1,\ldots,n$, $\tilde{\phi}^k = \sum_{i=1}^n \tilde{\phi}_i^k u_k^i$, $T_k = \max\{0, -\varphi_1^k\}$ 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 $\varphi_i^k = \varphi_1^k < 0$ which is referred to as *hard case* for unconstrained optimization and that for other cases, $\Gamma(\tau)$ is defined only for $0 \leqslant \tau < 1/T_k$, 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

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

2.2. Preconditional modified gradient path

Similarly, the null step $h_k = P(x_k, B_k)w_k$ in the null space $\mathcal{N}(A(x_k)^{\mathrm{T}})$, where $w_k = E_k^{\mathrm{T}} L_k^{-1} \tilde{w}_k$, 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 $\tilde{w}$ in the full space $\mathbb{R}^n$ can be given in the following closed form which can be referred to [2]:

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

where if $\tilde{\phi}_i^k \neq 0$ for some $i \in \mathcal{I}_k^- \cup \mathcal{N}_k$, the term $\Gamma_2(t_2(\tau))$ is not relevant, that is, if $\tilde{\phi}_i^k \neq 0$ for some $i \in \mathcal{I}_k^- \cup \mathcal{N}_k$, 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:

$$\Gamma_1(t_1(\tau)) = \sum_{i\in\mathcal{I}_k} \frac{\exp\{-\varphi_i^k t_1(\tau)\} - 1}{\varphi_i^k} \tilde{\phi}_i^k u_k^i - t_1(\tau) \sum_{i\in\mathcal{N}_k} \tilde{\phi}_i^k u_k^i$$

with

$$\Gamma_2(t_2) = \begin{cases} t_2 u_k^1 & \text{if } \varphi_1^k < 0, \\ 0 & \text{if } \varphi_1^k \geqslant 0, \end{cases}$$

$$t_1(\tau) = \begin{cases} \dfrac{\tau}{1-\tau} & \text{if } \tau < 1, \\ +\infty & \text{if } \tau \geqslant 1 \end{cases}$$

and

$$t_2(\tau) = \max\{\tau - 1, 0\}.$$

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

$$\Gamma^v(t_1(\tau)) = (\exp\{-t_1(\tau)\} - 1)A(x_k)^{\dagger} c_k, \tag{2.9}$$

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 $w_k(\tau) = E_k^{\mathrm{T}} L_k^{-1} \tilde{w}_k(\tau)$, be obtained by the projection $P(x_k, B_k)$ in the null space $\mathscr{N}(A(x_k)^{\mathrm{T}})$. Then we have that the norm function of the path $h_k(\tau)$ is monotonically increasing for $\tau \in (0, +\infty)$. The relation function $\hat{\psi}_k(\tau) \triangleq (g^k)^{\mathrm{T}} h_k(\tau)$ between gradient g^k of the objective function and the step $h_k(\tau)$ satisfies a sufficiently descent direction and is monotonically discreasing for $0 < \tau < 1/T$. Furthermore,*

$$\nabla_x l(x_k, \lambda_k)^{\mathrm{T}} \frac{\mathrm{d}h_k(\tau)}{\mathrm{d}\tau} = g_k^{\mathrm{T}} \frac{\mathrm{d}h_k(\tau)}{\mathrm{d}\tau} = \frac{\mathrm{d}\hat{\psi}_k(\tau)}{\mathrm{d}\tau}$$
$$\to -||P(x_k, B_k)\phi^k||^2 = -||P(x_k, B_k)g^k||^2 \quad \text{as } \tau \to 0, \tag{2.10}$$

and

$$\lim_{\tau \to \infty} \Gamma_k^v(\tau) = -A(x_k)^{\dagger} c_k. \tag{2.11}$$

If B_k is positive, then

$$\lim_{\tau \to \infty} \Gamma_k(\tau) = -D_k^{-1} \tilde{\phi}^k. \tag{2.12}$$

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 $\mathscr{N}(A(x_k)^{\mathrm{T}})$, it is obvious that $\Gamma(\tau)$ is a continuous path, and

$$||\Gamma(\tau)||^2 = \begin{cases} ||\Gamma_1(\tau)||^2 & \text{if } \tau < \frac{1}{T}, \\ ||\Gamma_1(\frac{1}{T})||^2 + ||\Gamma_2(t_2(\tau))||^2 & \text{if } \tau \geqslant \frac{1}{T}. \end{cases}$$

Let

$$\psi(\tau)=||\Gamma_1(\tau)||^2=\sum_{i\in\mathscr{I}_k}\left(\frac{\tau}{\varphi_i^k\tau+1}\right)^2(\tilde{\phi}_i^k)^2+\tau^2\sum_{i\in\mathscr{N}_k}(\tilde{\phi}_i^k)^2.$$

Then using the fact $\varphi_i^k\tau+1>0$ for all $i\in I_k$ when $0<\tau<1/T$, we have

$$\psi'(\tau)=\sum_{i\in\mathscr{I}_k}2\frac{\tau}{(\varphi_i^k\tau+1)^3}(\tilde{\phi}_i^k)^2+2\tau\sum_{i\in\mathscr{N}_k}(\tilde{\phi}_i^k)^2>0$$

which means that $||\Gamma_1(\tau)||$ is monotonically increasing for $0<\tau<1/T$. $||\Gamma_2(t_2(\tau))||^2=(\tau-1/T)^2$ is certainly increasing for $\tau\geqslant 1/T$. Since $||h_k(\tau)||=||P(x_k,B_k)w_k(\tau)||$ and $w_k(\tau)=E_k^{\mathrm{T}}L_k^{-1}\tilde{w}_k(\tau)$, 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 $P_k\phi^k=P_kg^k$ and $(P_kg^k)_i=(P_k\phi^k)_i$ where $(P_kg^k)_i$ is the ith component of the vector P_kg^k,

$$\begin{aligned}\hat{\psi}_k(\tau)&=(g^k)^{\mathrm{T}}P_kE_k^{\mathrm{T}}L_k^{-1}\Gamma_1(\tau)\\&=-(P_kg^k)^{\mathrm{T}}\left[\sum_{i\in\mathscr{I}_k}\frac{t_1(\tau)}{\varphi_i^k\tau+1}(P_kg^k)_iu^i+\tau\sum_{i\in\mathscr{N}_k}(P_kg^k)_iu^i\right]\\&=-\sum_{i\in\mathscr{I}_k}\frac{\tau}{\varphi_i^k\tau+1}(P_kg^k)_i^2-\tau\sum_{i\in\mathscr{N}_k}(P_kg^k)_i^2.\end{aligned}\tag{2.13}$$

Then we have that

$$\frac{\mathrm{d}\hat{\psi}_k(\tau)}{\mathrm{d}\tau}=-\sum_{i\in\mathscr{I}_k}\frac{1}{(\varphi_i^k\tau+1)^2}(P_kg^k)_i^2-\sum_{i\in\mathscr{N}_k}(P_kg^k)_i^2\tag{2.14}$$

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 $||D_k||+T\geqslant\varphi_i^k+T>0$ for all $i\in\mathscr{I}_k$ when $0<\tau<1/T$, we have that

$$\frac{\tau}{\varphi_i^k\tau+1}\geqslant\frac{1}{||D_k||+1/\tau}.$$

Therefore,

$$\begin{aligned}\hat{\psi}_k(\tau)&=(g^k)^{\mathrm{T}}h_k(\tau)\\&\leqslant-\frac{1}{||D_k||+1/\tau}\sum_{i\in\mathscr{I}_k}(P_kg^k)_i^2\\&\leqslant-\frac{||P_kg^k||^2}{||D_k||+1/\tau}.\end{aligned}\tag{2.15}$$

In hard case, i.e., $\tau\geqslant 1/T$ and $\Gamma(\tau)=\Gamma_1(1/T)+\Gamma_2(t_2(\tau))$, we have that for all

$$\frac{1}{\varphi_i^k+T}\geqslant\frac{1}{||D_k||+T},\quad i\in\mathscr{I}_k\ \text{ and }\ \frac{1}{T}\geqslant\frac{1}{||D_k||+T}.$$

Therefore, we have that, noting $\nabla_x l(x_k,\lambda_k)^{\mathrm{T}} h_k(\tau) = (g^k)^{\mathrm{T}} h_k(\tau)$,

$$\hat{\psi}_k(\tau) \leqslant -\frac{1}{\|D_k\|+T}\sum_{i\in\mathscr{I}_k}(P_k g^k)_i^2 - \frac{1}{T}\sum_{i\in\mathscr{N}_k}(P_k g^k)_i^2$$
$$\leqslant -\frac{\|P_k g^k\|^2}{\|D_k\|+T}. \tag{2.16}$$

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

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

$$\Gamma_k(\tau) = -\sum_{i\in\mathscr{I}_k}\frac{t_1(\tau)}{\varphi_i^k t_1(\tau)+1}\tilde{\phi}_i^k u^i$$
$$\to -\sum_{i\in\mathscr{I}_k}\frac{1}{\varphi_i^k}\tilde{\phi}_i^k u^i \quad \text{as } \tau\to\infty$$
$$= -D_k^{-1}\tilde{\phi}^k$$

which means that (2.12) holds.

It is clear that (2.12) holds as

$$\lim_{\tau\to\infty} -\frac{\tau}{\tau+1} = -1. \qquad \square$$

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 $w_k(\tau) = E_k^{\mathrm{T}} L_k^{-1}\tilde{w}_k(\tau)$, be obtained by the projection $P(x_k,B_k)$ in the null space $\mathscr{N}(A(x_k)^{\mathrm{T}})$. Then we have that the norm function of the path $h_k(\tau)$ is monotonically increasing for $\tau\in[0,+\infty)$. The relation function $\bar{\psi}_k(\tau) \triangleq (g^k)^{\mathrm{T}} h_k(\tau)$ 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,*

$$\nabla_x l(x_k,\lambda_k)^{\mathrm{T}}\frac{\mathrm{d}h_k(\tau)}{\mathrm{d}\tau} = (g^k)^{\mathrm{T}}\frac{\mathrm{d}h_k(\tau)}{\mathrm{d}\tau} = \frac{\mathrm{d}\bar{\psi}_k(\tau)}{\mathrm{d}\tau}$$
$$\to -\|P(x_k,B_k)\phi^k\|^2 = -\|P(x_k,B_k)g^k\|^2, \quad \textit{as } \tau\to 0, \tag{2.17}$$

and

$$\lim_{\tau\to\infty}\Gamma_k^v(\tau) = -A(x_k)^{\dagger} c_k. \tag{2.18}$$

If B_k is positive, then

$$\lim_{\tau\to\infty}\Gamma_k(\tau) = -D_k^{-1}\tilde{\phi}^k. \tag{2.19}$$

Proof. From the definition of the path and using the orthonormality of vectors u^i, we have

$$\|\Gamma_k(\tau)\|^2 = \begin{cases} \|\Gamma_1(\tau)\|^2 & \text{if } \tau<1,\\ \|\Gamma_1(\frac{1}{T})\|^2 + \|\Gamma_2(t_2(\tau))\|^2 & \text{if } \tau\geqslant 1.\end{cases}$$

Let

$$\psi_1(\tau) = \|\Gamma_1(\tau)\|^2 = \sum_{i\in\mathscr{I}_k}\left(\frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}\right)^2(\tilde{\phi}_i^k)^2 + \left(\frac{\tau}{1-\tau}\right)^2\sum_{i\in\mathscr{N}_k}(\tilde{\phi}_i^k)^2.$$

Then using the fact $\tau > 0$ and $\exp\{-\varphi_i^k t_1(\tau)\}\leqslant 1$, we have

$$\psi_1'(\tau) = \sum_{i\in\mathscr{I}_k}\frac{2\exp\{-\varphi_i^k t_1(\tau)\}}{(1-\tau)^2}(1-\exp\{-\varphi_i^k t_1(\tau)\})(\tilde{\phi}_i^k)^2 + \frac{2\tau}{(1-\tau)^2}\sum_{i\in\mathscr{N}_k}(\tilde{\phi}_i^k)^2 > 0.$$

Thus, $\|\Gamma_1(\tau)\|$ 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 $\tau\geqslant 1$. Since $\|h_k(\tau)\| = \|P(x_k, B_k)w_k(\tau)\|$ and $w_k(\tau) = E_k^{\mathrm{T}}L_k^{-1}\tilde{w}_k(\tau)$, as above we have that the norm function of the preconditional modified gradient path is monotonically increasing for $\tau\in[0,+\infty)$.

If $\tau\geqslant 1$ then $t_1(\tau)=+\infty$, that is, $\tilde{\phi}_i^k = 0\ \forall\ i\in\mathscr{I}_k^-\cup\mathscr{N}_k$, the term $\Gamma_2(t_2(\tau))$ is relevant. In that case, by

$$\lim_{t\to\infty}\frac{\exp\{-\varphi_i^k t\}-1}{\varphi_i^k} = -\frac{1}{\varphi_i^k}, \quad \text{if } \varphi_i^k > 0,$$

we get that, noting $\nabla_x l(x_k,\lambda_k)^{\mathrm{T}} h_k(\tau) = (g^k)^{\mathrm{T}} h_k(\tau)$ and $P_k\phi^k = P_k g^k$,

$$\begin{aligned}(g^k)^{\mathrm{T}} h_k(\tau) &= (g^k)^{\mathrm{T}} P_k E_k^{\mathrm{T}} L_k^{-1}\{\Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau))\}\\ &= -\sum_{i\in\mathscr{I}_k^+}\frac{1}{\varphi_i^k}(P_k g^k)_i^2 + (\tau-1)(P_k g^k)_1\\ &= -\sum_{i\in\mathscr{I}_k^+}\frac{1}{\varphi_i^k}(P_k g^k)_i^2\\ &\leqslant -\frac{\|P_k g^k\|^2}{\|D_k\|}\end{aligned} \tag{2.20}$$

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

If $0<\tau<1$, then we have that

$$\begin{aligned}\bar{\psi}_k(\tau) &= (g^k)^{\mathrm{T}} P_k E_k^{\mathrm{T}} L_k^{-1}\Gamma_1(t_1(\tau))\\ &= \sum_{i\in\mathscr{I}_k}\frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}(P_k g^k)_i^2 - t_1(\tau)\sum_{i\in\mathscr{N}_k}(P_k g^k)_i^2.\end{aligned} \tag{2.21}$$

Then we have that

$$\frac{\mathrm{d}\bar{\psi}_k(\tau)}{\mathrm{d}\tau} = -\sum_{i\in\mathscr{I}_k}\frac{1}{(1-\tau)^2}\exp\{-\varphi_i^k t_1(\tau)\}(P_k g^k)_i^2 - \frac{1}{(1-\tau)^2}\sum_{i\in\mathscr{N}_k}(P_k g^k)_i^2 < 0 \tag{2.22}$$

which means that $\bar{\psi}_k(\tau)$ is monotonically discreasing for $0<\tau<1$. Taking $\tau\to 0$, (2.17) is true.

If B_k is positive, then by the defintion of the preconditional modified gradient path $\Gamma(\tau)$ which can be expressed as

$$\begin{aligned}\Gamma_k(\tau) &= \sum_{i\in\mathscr{I}_k} \frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}\tilde{\phi}_i^k u^i \\ &\to -\sum_{i\in\mathscr{I}_k} \frac{1}{\varphi_i^k}\tilde{\phi}_i^k u^i \quad \text{as } \tau\to 1 \\ &= -D_k^{-1}\tilde{\phi}^k,\end{aligned}$$

which means that (2.18) holds.

It is clear that (2.19) holds as $t_1(\tau)\to+\infty$, if $\tau\geqslant 1$, and hence

$$\lim_{\tau\to\infty}\exp\{-t_1(\tau)\}=0. \qquad \square$$

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 ρ is the penalty parameter. Here we choose the l_1 exact penalty function as the merit function $\vartheta(\cdot,\rho)$,

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

where ρ_i is the ith component of the penalty vector ρ which is updated and given as follows:

$$\rho_{(k+1)_i}=\begin{cases}\rho_{k_i} & \text{if } \rho_{k_i}\geqslant|\Psi_{k_i}|+\kappa,\\ \max\{\rho_{k_i},|\Psi_{k_i}|\}+\kappa & \text{otherwise,}\end{cases} \tag{3.2}$$

where $\kappa>0$ is the given constant, ρ_{k_i} is the ith component of the vector ρ_k, and $\Psi_{k_i}=\max\{\lambda_{k_i}^{\mathrm{P}},\lambda_{k_i}^{\mathrm{S}}\}$ with $\lambda_{k_i}^{\mathrm{P}}$ and $\lambda_{k_i}^{\mathrm{S}}$ being the ith component of the vectors λ_k^{P} and λ_k^{S}, respectively.

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

$$c_i(x_k+s_k(\tau))=c_i(x_k)+\tau\nabla c_i(x_k)^{\mathrm{T}}s_k'(\xi\tau),$$

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

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

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

$$\nabla c_i(x_k)^{\mathrm{T}}s_k'(\tau)=-\frac{\exp\{-t_1(\tau)\}}{(1-\tau)^2}\nabla c_i(x_k)^{\mathrm{T}}A(x_k)^{\dagger}c_k=-\frac{\exp\{-t_1(\tau)\}}{(1-\tau)^2}c_i(x_k).$$

From above

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

which implies that its directional $|c_i(x)|$ derivative in the direction $s_k'(0)$, we have that

$$D\{|c_i(x_k; s_k'(0))|\} = -|c_i(x_k)|, \quad i = 1, 2, \ldots, m.$$

Using (1.12), (1.13), (2.11) and (2.18), we obtain

$$g_k^{\mathrm{T}} v_k'(\tau) = -\bar{\alpha}_k(\tau) g_k A(x_k)^{\dagger} c_k = -\bar{\alpha}_k(\tau) \lambda_k^{\mathrm{T}} c_k,$$

where

$$\bar{\alpha}_k(\tau) = \begin{cases} \frac{1}{(\tau+1)^2} & \text{if preconditional optimal path,} \\ \frac{\exp\{-t_1(\tau)\}}{(\tau-1)^2} & \text{if preconditional modified gradient path} \end{cases} \tag{3.3}$$

and

$$\lambda_k = \begin{cases} \lambda_k^{\mathrm{P}} & A(x_k)^{\dagger} \text{ given in (1.12),} \\ \lambda_k^{\mathrm{S}} & A(x_k)^{\dagger} \text{ given in (1.13).} \end{cases} \tag{3.4}$$

Hence from (3.1), we obtain that from $\bar{\alpha}_k(0) = 1$,

$$D\vartheta_k(x_k, \rho; s_k'(0)) = (g^k)^{\mathrm{T}} h_k'(0) - \sum_{i=1}^{m} \{\rho_{(k+1)_i} |c_i(x_k)| + \lambda_{k_i} c_i(x_k)\}. \tag{3.5}$$

The difficulty with nondifferentiable penalty function $\vartheta(\cdot, \rho)$ 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:

$$d(\tau) = -A(x_k)^{\dagger} c(x_k + s_k(\tau)), \tag{3.6}$$

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 $||x_k - x_*||^3$. This means that the merit function will then be decreased at the point $x_k + s_k(\tau) + \alpha_k(\tau) d_k(\tau)$, as we will show below. Set then

$$x_{k+1} = x_k + s_k(\tau_k) + \alpha_k(\tau_k) d_k(\tau_k), \tag{3.7}$$

where $\alpha_k(\tau)$ is given in (3.12) and τ_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 $A(x_k)^{\dagger}$. Compute the multiplier λ_{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 $||c_k|| + ||P_k g^k|| \leqslant \varepsilon$ (where ε is a given small constant), stop.
3. Factorize B_k into the form (2.2) using stable Bunch–Parlett factorization method, that is, $E_k B_k E_k^{\mathrm{T}} = L_k D_k L_k^{\mathrm{T}}$. 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 $\Gamma_k^v(\tau)$, respectively.

4. Denote by $\tilde{w}_k$, v_k the steps of the preconditional curvilinear paths $\Gamma_k(\tau)$, $\Gamma_k^v(\tau)$, respectively. Compute

$$w_k(\tau) = E_k^{\mathrm{T}} L_k^{-\mathrm{T}} \tilde{w}(\tau), \tag{3.8}$$

$$h_k(\tau) = P(x_k B_k) w_k(\tau), \tag{3.9}$$

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

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

$$p_k(\tau) = s_k(\tau) + \alpha_k(\tau) d_k(\tau). \tag{3.10}$$

5. Choose

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

until the following inequality is satisfied:

$$\vartheta(x_k + p_k(\tau_k), \rho_k) \leqslant \vartheta(x_{l(k)}, \rho_{l(k)}) + \beta\alpha_k(\tau_k) D\vartheta(x_k, \rho_k; s_k'(0)), \tag{3.11}$$

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

$$\alpha_k(\tau) = \begin{cases} \frac{\tau}{\tau+1}, & \text{if preconditional optimal path,} \\ 1-\exp\{-t_1(\tau)\} & \text{if preconditional modified gradient path.} \end{cases} \tag{3.12}$$

6. Set

$$x_{k+1} = x_k + p_k(\tau_k). \tag{3.13}$$

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 compution 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 Ω; $\nabla^2 f(x)$ and $\nabla^2 c_i(x)$ $(i=1,\dots,m)$ are Lipschitz continuous matrix functions on Ω.

Assumption A2. The matrix $A(x)$ has full column rank over Ω.

Assumption A3. Norm sequences $\{||A(x_k)^\dagger||\}$ is bounded for all k, that is, there is a constant $\beta_1>0$ such that

$$||A(x_k)^\dagger||\leqslant\beta_1,\quad \forall k, \tag{4.1}$$

where $A(x_k)^\dagger$ is given in (1.12) and (1.13).

Lemma 4.1. *Under the Assumption* A3, *we obtain*

$$d_k(\tau)=\mathrm{O}(||s_k(\tau)||^2). \tag{4.2}$$

Proof. Since $A_k^{\mathrm{T}}h_k(\tau)=0$ and $s_k(\tau)=h_k(\tau)+v_k(\tau)$, it means

$$A_k^{\mathrm{T}}s_k(\tau)=A_k^{\mathrm{T}}(h_k(\tau)+v_k(\tau))=A_k^{\mathrm{T}}v_k(\tau)=-c_k. \tag{4.3}$$

Therefore,

$$||c(x_k+s_k(\tau))||=||c_k+A_k^{\mathrm{T}}s_k(\tau)||+\mathrm{O}(||s_k(\tau)||^2)=\mathrm{O}(||s_k(\tau)||^2). \tag{4.4}$$

Using (4.1), (3.6) implies that (4.2) holds. □

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* ρ_k, *we have that*

$$D\vartheta(x_k,\rho;s_k'(0))\leqslant-\omega_k(\tau)||P(x_k,B_k)g^k||^2-\kappa\sum_{i=1}^m|c_i(x_k)|, \tag{4.5}$$

where

$$\omega_k(\tau)=\begin{cases}\min\limits_{i\in\mathscr{I}_k}\{\frac{1}{(\varphi_i^k\tau+1)^2},1\} & \textit{if preconditional optimal path},\\ \frac{1}{(1-\tau)^2}\min\limits_{i\in\mathscr{I}_k}\{\exp\{-\varphi_i^k t_1(\tau),1\} & \textit{if preconditional modified gradient path}.\end{cases}$$

Furthermore, there exists $\kappa_1>0$ *such that* $0<\tau_k<1/T$.

$$D\vartheta(x_k,\rho;s_k'(0))\leqslant-\omega_k^1(\tau)\|h_k(\tau)\|^2-\kappa_1\|v_k(\tau)\|, \tag{4.6}$$

where $\omega_k^1(\tau)=\hat{\omega}_k(\tau)\omega_k(\tau)$ *and*

$$\hat{\omega}_k(\tau)=\begin{cases}\frac{\min_{i\in\mathscr{I}_k}\{(\varphi_i^k\tau+1)^2,1\}}{\tau^2} & \textit{if optimal path},\\ \min\limits_{i\in\mathscr{I}_k}(\frac{\varphi_i^k}{\exp\{-\varphi_i^k t_1(\tau)\}-1})^2,\frac{(1-\tau)^2}{\tau^2}\} & \textit{if modified gradient path}.\end{cases} \tag{4.7}$$

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

$$(P_kg_k)^{\mathrm{T}}(h_k)'(0)\leqslant-\omega_k(\tau)\|P(x_k,B_k)g^k\|^2.$$

From (3.2) updating the penalty vector ρ_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

$$\begin{aligned}\|h_k(\tau)\|&=\|P_k\Gamma_1(\tau)\|^2\\&=\sum_{i\in\mathscr{I}_k}\left(\frac{\tau}{\varphi_i^k\tau+1}\right)^2(P_kg^k)_i^2+\tau^2\sum_{i\in\mathscr{N}_k}(P_kg^k)_i^2\\&\leqslant\max_{i\in\mathscr{I}_k}\left\{\left(\frac{\tau}{\varphi_i^k\tau+1}\right)^2,\tau^2\right\}\|P_kg^k\|^2.\end{aligned}$$

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

$$\|h_k(\tau)\|=\|P_k\Gamma_1(\tau)\|^2\leqslant\max_{i\in\mathscr{I}_k}\left\{\left(\frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}\right)^2,\left(\frac{\tau}{1-\tau}\right)^2\right\}\|P_kg^k\|^2. \tag{4.8}$$

Since $\|c_k\|_1\geqslant\|c_k\|$, $0<\bar{\alpha}_k(\tau)\leqslant1$ and $\|v_k'(\tau)\|\leqslant\beta_1\bar{\alpha}_k(\tau)\|c_k\|\leqslant\beta_1\|c_k\|$ where β_1 is given in (4.1), we further conclude that taking $\kappa_1=\kappa/\beta_1$, (4.6) holds. □

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 fomular. Then there exists a vector* $\bar{\rho}>0$ *such that for all large* k, $\rho_k\leqslant\bar{\rho}$, *i.e., the sequence* $\{\rho_k\}$ *is uniformly bounded from above.*

Proof. Let the penalty vector ρ_k be a sequence generated by the algorithm using the l_1 exact penalty function $\vartheta(x_k,\rho_k)$ as merit function with (3.1). Referring to Assumptions A1–A3, $\|\lambda_k\|$ 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 $\bar{\rho} > 0$ such that $\rho_k \leqslant \bar{\rho}$. $\square$

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:

$$\liminf_{k\to\infty} \{||P(x_k, B_k)g_k||^2 + ||c(x_k)||_1\} = 0.$$

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

$$||c_k|| + ||P_k g^k|| \geqslant \varepsilon, \quad k = 1, 2, \dots .$$

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

$$\vartheta(x_{l(k)}, \rho_{l(k)}) - \vartheta(x_k + p_k(\tau_k), \rho_k) \leqslant \beta\alpha_k(\tau_k) D\vartheta(x_k, \rho_k; s'_k(0)). \tag{4.10}$$

Taking into account that $m(k+1) \leqslant m(k) + 1$, and from Lemma 4.4 there exists $\bar{\rho} > 0$ such that for all large k, $\rho_k \equiv \bar{\rho}$, we have that $\vartheta(x_{k+1}, \rho_{k+1}) \leqslant \vartheta(x_{l(k)}, \rho_{l(k)})$. 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 $k > M$,

$$\begin{aligned}
&\vartheta(x_{l(k)}, \rho_{l(k)}) \\
&\quad = \vartheta(x_{l(k)-1} + p_{l(k)-1}(\tau_{l(k)-1}), \rho_{l(k)-1}) \\
&\quad \leqslant \max_{0\leqslant j\leqslant m(l(k)-1)} \{\vartheta(x_{l(k)-j-1}, \rho_{l(k)-j-1})\} + \beta\alpha_{l(k)-1}(\tau_{l(k)-1}) D\vartheta(x_{l(k)-1}, \rho_{l(k)-1}; s'_{l(k)-1}(0)) \\
&\quad \leqslant \max_{0\leqslant j\leqslant m(l(k)-1)} \{\vartheta(x_{l(k)-j-1}, \rho_{l(k)-j-1})\} \\
&\qquad -\beta\omega_{l(k)-1}(\tau_{l(k)-1})\alpha_{l(k)-1}(\tau_{l(k)-1})\{||P_{l(k)-1} g^{l(k)-1}||^2 + \kappa||c_{l(k)-1}||\}.
\end{aligned} \tag{4.11}$$

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)

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

This, by the definitions of preconditional curvilinear paths (2.7)–(2.8) and (2.9)–(2.10), as well as $\{||c_k|| + ||P_k g^k||\} \geqslant \varepsilon$, implies that

$$\lim_{k\to\infty} \alpha_{l(k)-1}(\tau_{l(k)-1}) = 0. \tag{4.13}$$

By definition (3.12) of $\alpha_{l(k)-1}(\tau_{l(k)-1})$, (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 $k \geqslant l(k) \geqslant k - M$, from

$$x_{l(k)} = x_{k-M-1} + p_{k-M-1}(\tau_{k-M-1}) + \cdots + p_{l(k)-1}(\tau_{l(k)-1})$$

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),

$$\begin{aligned}\vartheta(x_{k+1}, \rho_{k+1}) - \vartheta(x_{l(k)}, \rho_{l(k)}) &\leqslant \beta\alpha_k(\tau_k) D\vartheta(x_k, \rho_k; s_k'(0)) \\ &\leqslant -\beta\omega_k(\tau_k)\alpha_k(\tau_k)\{||P_k g^k||^2 + \kappa||c(x_k)||\}.\end{aligned} \tag{4.18}$$

Eqs. (4.17) and (4.18) mean that

$$\lim_{k\to\infty} \alpha_k(\tau_k) = 0, \quad \text{and hence} \quad \lim_{k\to\infty} \tau_k = 0.$$

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, $\rho_k \equiv \bar{\rho}$.

$$\begin{aligned}\vartheta\left(x_k + p_k\left(\frac{\tau_k}{\omega}\right), \bar{\rho}\right) - \vartheta(x_k, \bar{\rho}) &\geqslant \vartheta\left(x_k + p_k\left(\frac{\tau_k}{\omega}\right), \bar{\rho}\right) - \vartheta(x_{l(k)}, \bar{\rho}) \\ &> \beta\alpha_k\left(\frac{\tau_k}{\omega}\right) D\vartheta(x_k, \bar{\rho}; s_k'(0)).\end{aligned} \tag{4.21}$$

For the composite function $\vartheta(x_k + p_k(\frac{\tau}{\omega}), \bar{\rho})$ about τ, we have that noting $\alpha_k(\frac{\tau_k}{\omega})d_k'(0) = O(\tau_k/\omega)d_k'(0)$,

$$\begin{aligned}\vartheta\left(x_k + p_k\left(\frac{\tau_k}{\omega}\right), \bar{\rho}\right) &= \vartheta(x_k, \bar{\rho}) + \frac{\tau_k}{\omega}\left.\frac{\mathrm{d}\vartheta(x_k + p_k(\tau_k), \bar{\rho})}{\mathrm{d}\tau}\right|_{\tau=0} + o\left(\frac{\tau_k}{\omega}\right) \\ &= \vartheta(x_k, \bar{\rho}) + \frac{\tau_k}{\omega} D\vartheta(x_k, \bar{\rho}; p_k'(0)) + o\left(\frac{\tau_k}{\omega}\right) \\ &= \vartheta(x_k, \bar{\rho}) + \frac{\tau_k}{\omega} D\vartheta(x_k, \bar{\rho}; s_k'(0)) + o\left(\frac{\tau_k}{\omega}\right).\end{aligned} \tag{4.22}$$

From (4.21) and (4.22), we have

$$\left[\frac{\tau_k}{\omega} - \beta\alpha_k\left(\frac{\tau_k}{\omega}\right)\right] D\vartheta(x_k, \bar{\rho}; s_k'(0)) + \mathrm{o}\left(\frac{\tau_k}{\omega}\right) \geqslant 0. \tag{4.23}$$

Dividing (4.23) by τ_k/ω and noting that $1-\beta>0$, $\alpha_k(\tau_k/\omega) \div (\tau_k/\omega) \to 1$ as $\tau_k \to 0$, and $D\vartheta(x_k, \bar{\rho}; s_k'(0)) \leqslant 0$, we obtain

$$\lim_{k\to\infty} D\vartheta(x_k, \bar{\rho}; s_k'(0)) = 0. \tag{4.24}$$

From (4.5), (4.24) means that, from $\tau_k \to 0$ as $k \to \infty$,

$$\lim_{k\to\infty} \{||P_k g^k||^2 + ||c_k||\} = \lim_{k\to\infty} D\vartheta(x_k, \bar{\rho}; s_k'(0)) = 0 \tag{4.25}$$

which implies that (4.9) is true.

Assume that there exists a limit point x_* which is a local maximum of f, let $\{x_k\}_{\mathscr{K}}$, be a subsequence of $\{x_k\}$ converging to x_*. Then the limit point x_* is also a local maximum of the merit function $\vartheta(x, \bar{\rho})$, for large $\bar{\rho}$.

As $k \geqslant l(k) \geqslant k - M$, for any k there exists a point $x_{l(k)}$ such that

$$\lim_{k\to\infty} ||x_{l(k)} - x_k|| = 0, \tag{4.26}$$

so that we can obtain

$$\lim_{k\in\mathscr{K}, k\to\infty} ||x_{l(k)} - x_*|| \leqslant \lim_{k\in\mathscr{K}, k\to\infty} ||x_* - x_k|| + \lim_{k\in\mathscr{K}, k\to\infty} ||x_{l(k)} - x_k|| = 0. \tag{4.27}$$

This means that also the subsequence $\{x_k\}_{\mathscr{K}}$ converges to x_*.

On the other hand, 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, noting $c_* = 0$, i.e.,

$$\lim_{k\to\infty} \{\vartheta(x_{l(k)}, \rho_{l(k)})\} = \vartheta(x_*, \bar{\rho}) = f(x_*)$$

and $\{\vartheta(x_{l(k)}, \rho_{l(k)})\} \geqslant f(x_*)$ for all large k. Moreover, we can find a sufficiently large index k such that

$$f(x_*) \leqslant \vartheta(x_{l(k+M)}, \rho_{l(k+M)}) < \vartheta(x_{l(k)}, \rho_{l(k)}) \tag{4.28}$$

so that we conclude that, in any neighborhood of x_*, there exists a point $x_{l(k)}$ with $k \in \mathscr{K}$ such that $f(x_*) < \vartheta(x_{l(k)}, \bar{\rho})$ holds. This contradicts the assumption that x_* is a local maximum of the merit function $\vartheta(x, \bar{\rho})$, for large $\bar{\rho}$. 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

$$p^{\mathrm{T}}(Z_*^{\mathrm{T}} W_* Z_*)p \geqslant \bar{\tau}||p||^2, \quad \forall p \in \mathbb{R}^t \tag{5.2}$$

i.e., the second-order sufficient condition holds at x_*. Further, there exists a constant $\bar{\omega} > 0$ such that $||W(x)|| \leqslant \bar{\omega}$ $\forall x \in \Omega$, where $W(x)$ given in (1.19).

Assumption A6. $x_k \to x_*$.

Assumption A7.

$$\lim_{k\to\infty} \frac{||P_k(B_k - W_*)h_k||}{||h_k||} = 0. \tag{5.3}$$

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 $P(x_k, B_k)E_k^{\mathrm{T}} L_k^{-1} \Gamma_k(\tau_k)$ satisfies*

$$\hat{q}_k(h_k(\tau_k)) \stackrel{\text{def}}{=} (g^k)^{\mathrm{T}} h_k(\tau_k) + h_k(\tau_k)^{\mathrm{T}} B_k h_k(\tau_k) \leqslant 0. \tag{5.4}$$

Proof. At the kth 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

$$\begin{aligned} B_k P_k E_k^{\mathrm{T}} L_k^{-1} \Gamma_k(\tau_k) &= B_k P_k E_k^{\mathrm{T}} L_k^{-1} \Gamma_1(t_1(\tau_k)) \\ &= -\sum_{i=1}^{n} \varphi_i^k u_k^i (u_k^i)^{\mathrm{T}} \left[\sum_{i \in \mathscr{I}_k} \frac{t_1(\tau_k)}{\varphi_i^k t_1(\tau) + 1} (P_k g^k)_i u_k^i \right] \\ &= -\sum_{i \in \mathscr{I}_k} \frac{\varphi_i^k t_1(\tau_k)}{\varphi_i^k t_1(\tau_k) + 1} (P_k g^k)_i u_k^i. \end{aligned} \tag{5.5}$$

Hence, we have that

$$\begin{aligned} &(g^k)^{\mathrm{T}} h_k(\tau_k) + h_k(\tau_k)^{\mathrm{T}} B_k h_k(\tau_k) \\ &\quad = -\sum_{i \in \mathscr{I}_k} \frac{\tau_k}{\varphi_i^k \tau_k + 1} (P_k g^k)_i^2 + \sum_{i \in \mathscr{I}_k} \frac{\varphi_i^k \tau_k^2}{(\varphi_i^k \tau_k + 1)^2} (P_k g^k)_i^2 \\ &\quad = -\sum_{i \in \mathscr{I}_k} \frac{\tau_k}{(\varphi_i^k \tau_k + 1)^2} (P_k g^k)_i^2 \\ &\quad \leqslant 0. \end{aligned} \tag{5.6}$$

Let the step $\tilde{w}_k$ be obtained from the preconditional modified gradient path. Since B_k is positive definite, it means that $\mathscr{I}_k^- \cup \mathscr{N}_k = \phi$. We define the value of

$$\tilde{q}_k(h_k) \stackrel{\text{def}}{=} (g^k)^{\mathrm{T}} h_k + \tfrac{1}{2} h_k^{\mathrm{T}} B_k h_k,$$

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]),

$$\tilde{q}_k(P_k\Gamma_1(t_1(\tau))) = \sum_{i\in\mathscr{I}_k} \frac{\exp\{-2\varphi_i^k t_1(\tau)\}-1}{2\varphi_i^k}(P_k g^k)_i^2. \tag{5.7}$$

It is clear to see that

$$(g^k)^{\mathrm{T}}h_k = (g^k)^{\mathrm{T}}[P_k E_k^{\mathrm{T}} L_k^{-1}\Gamma_1(t_1(\tau))] = \sum_{i\in\mathscr{I}_k} \frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}(P_k g^k)_i^2. \tag{5.8}$$

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

$$\begin{aligned}
&(g^k)^{\mathrm{T}}h_k + h_k^{\mathrm{T}}B_k h_k\\
&\quad = 2[(g^k)^{\mathrm{T}}h_k + \tfrac{1}{2}h_k^{\mathrm{T}}B_k h_k] - (g^k)^{\mathrm{T}}h_k\\
&\quad = 2\sum_{i\in\mathscr{I}_k} \frac{\exp\{-2\varphi_i^k t_1(\tau)\}-1}{2\varphi_i^k}(P_k g^k)_i^2 - \sum_{i\in\mathscr{I}_k} \frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}(P_k g^k)_i^2\\
&\quad = \sum_{i\in\mathscr{I}_k} \exp\{-\varphi_i^k t_1(\tau)\}\frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{\varphi_i^k}(P_k g^k)_i^2\\
&\quad \leqslant 0.
\end{aligned} \tag{5.9}$$

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

Lemma 5.2. *Let $\{x_k\} \subset \mathbb{R}^n$ be a sequence generated by the algorithm using the l_1 penalty function $\vartheta(x_k, \rho_k)$ 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 $||c(x_k+s_k)|| = \mathrm{O}(||s_k||^2)$, we have that

$$\begin{aligned}
g_k^{\mathrm{T}}d_k &= \lambda_{k+1}^{\mathrm{T}}c(x_k+s_k) - \xi_k c_k^{\mathrm{T}}(A_k^{\mathrm{T}}B_k^{-1}A_k)^{-1}c(x_k+s_k)\\
&= \lambda_{k+1}^{\mathrm{T}}c(x_k+s_k) + \mathrm{o}(||c_k||)\\
&= \frac{1}{2}s_k^{\mathrm{T}}\left(\sum_{i=1}^m \lambda_{k+1,i}\nabla^2 c_i(x_k)\right)s_k + \mathrm{o}(||c_k||) + \mathrm{o}(||h_k||^2)\\
&= \frac{1}{2}h_k^{\mathrm{T}}\left(\sum_{i=1}^m \lambda_{k+1,i}\nabla^2 c_i(x_k)\right)h_k + \mathrm{o}(||v_k||) + \mathrm{o}(||h_k||^2),
\end{aligned} \tag{5.10}$$

where $\xi_k = 0$ or 1 and λ_k given in (1.16), (1.17) or (1.18).

Since, $||W(x)||\leqslant\bar{\omega}\ \forall x\in\Omega$ and noting (5.10),

$$\begin{aligned}
&f(x_k+s_k+d_k)\\
&\quad=f_k+g_k^{\mathrm T}(v_k+h_k+d_k)+\tfrac12(s_k+d_k)^{\mathrm T}(\nabla^2 f_k)(s_k+d_k)+\mathrm{o}(||v_k||)+\mathrm{o}(||s_k||^2)\\
&\quad=f_k+g_k^{\mathrm T}v_k+g_k^{\mathrm T}h_k+\tfrac12 h_k^{\mathrm T}(\nabla^2 f_k)h_k\\
&\qquad+\tfrac12 h_k^{\mathrm T}\left(\sum_{i=1}^m \lambda_{k+1,i}\nabla^2 c_i(x_k)\right)h_k+\mathrm{o}(||v_k||)+\mathrm{o}(||s_k||^2)\\
&\quad=f_k+g_k^{\mathrm T}h_k+g_k^{\mathrm T}v_k+\tfrac12 h_k^{\mathrm T}W_kh_k+\mathrm{o}(||v_k||)+\mathrm{o}(||h_k||^2),
\end{aligned}\tag{5.11}$$

$$\begin{aligned}
c(x_k+s_k+d_k)&=c(x_k+s_k)+A(x_k+s_k)^{\mathrm T}d_k+\mathrm{o}(||d_k||)\\
&=[A(x_k+s_k)-A_k]^{\mathrm T}d_k+\mathrm{o}(||d_k||)\\
&=\mathrm{o}(||h_k||^2)+\mathrm{o}(||c_k||)
\end{aligned}\tag{5.12}$$

and

$$D\vartheta(x_k,\rho_k;s_k)=g_k^{\mathrm T}v_k+g_k^{\mathrm T}h_k-\rho_k||c_k||_1\tag{5.13}$$

we have that, noting (5.4) and B_k being positive definite, i.e., $(g^k)^{\mathrm T}h_k\leqslant -h_k^{\mathrm T}B_kh_k\leqslant 0$

$$\begin{aligned}
&\vartheta(x_k+s_k+d_k,\rho_k)-\vartheta(x_k,\rho_k)-\beta D\vartheta(x_k,\rho_k;s_k)\\
&\quad=(\tfrac12-\beta)g_k^{\mathrm T}h_k+\tfrac12 h_k^{\mathrm T}W_kh_k+(1-\beta)g_k^{\mathrm T}v_k+\tfrac12 g_k^{\mathrm T}h_k\\
&\qquad-(1-\beta)\rho_k||c_k||+\mathrm{o}(||c_k||)+\mathrm{o}(||h_k||^2)+\mathrm{o}(||s_k||^2)\\
&\quad\leqslant-(\tfrac12-\beta)h_k^{\mathrm T}B_kh_k-\tfrac12 h_k^{\mathrm T}(B_k-W_k)h_k+(1-\beta)g_k^{\mathrm T}v_k\\
&\qquad-(1-\beta)\rho_k||c_k||+\mathrm{o}(||c_k||)+\mathrm{o}(||h_k||^2)+\mathrm{o}(||s_k||^2)\\
&\quad\leqslant-(\tfrac12-\beta)h_k^{\mathrm T}B_kh_k+(1-\beta)g_k^{\mathrm T}v_k\\
&\qquad-(1-\beta)\rho_k||c_k||+\mathrm{o}(||c_k||)+\mathrm{o}(||h_k||^2),
\end{aligned}\tag{5.14}$$

the last inequality is reduced by (5.3). Further, we have that

$$\begin{aligned}
g_k^{\mathrm T}v_k&=\begin{cases}(\lambda_k^{\mathrm P})^{\mathrm T}c_k & \text{for ALG 1, ALG 3, and ALG 5,}\\ (\lambda_k^{\mathrm S})^{\mathrm T}c_k & \text{for ALG 2, ALG 4, and ALG 6,}\end{cases}\\
&\leqslant\begin{cases}||\lambda_k^{\mathrm P}||_\infty\cdot||c_k||_1 & \text{for ALG 1, ALG 3, and ALG 5,}\\ ||\lambda_k^{\mathrm S}||_\infty\cdot||c_k||_1 & \text{for ALG 2, ALG 4, and ALG 6,}\end{cases}\\
&\leqslant||\Psi_k||_\infty\cdot||c_k||_1.
\end{aligned}\tag{5.15}$$

From (5.2) and (5.3), we have for large k

$$\frac{\bar\tau}{2}||h_k||^2\leqslant h_k^{\mathrm T}B_kh_k.$$

Table 2
Experimental Results of the algorithm ALG 2

Problem name	$M=0$		$M=4$			$M=8$		
	NF	NG	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	38	21	34	16	1	34	16	1
HS028	58	31	58	31	0	58	31	0
HS060	8	7	8	7	0	8	7	0
SC220	153	115	129	98	3	139	103	4
SC252	54	48	65	46	2	68	46	2
SC235	31	16	29	15	1	29	15	1
SC219	69	56	57	49	3	54	48	4
SC216	79	68	76	59	1	76	55	1
PMGP								
HS006	14	10	14	10	0	14	10	0
HS026	16	16	14	14	1	14	14	1
HS027	39	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	142	102	132	107	4	112	98	5
SC252	57	50	68	54	2	68	54	2
SC235	31	16	29	15	1	29	15	1
SC219	62	49	55	44	2	46	35	3
SC216	77	58	71	51	1	71	51	1

Therefore, we obtain that, for k large enough and by (3.2) and (5.14)

$$\begin{aligned}
&\vartheta(x_k+s_k+d_k,\rho_k)-\vartheta(x_k,\rho_k)-\beta D\vartheta(x_k,\rho_k;s_k)\\
&\quad\leqslant -\left(\tfrac{1}{2}-\beta\right)\frac{\bar{\tau}}{2}\|h_k\|^2+(1-\beta)(\Psi_k-\rho_k)\|c_k\|_1+\mathrm{o}(\|c_k\|)+\mathrm{o}(\|h_k\|^2)\\
&\quad\leqslant -\left(\tfrac{1}{2}-\beta\right)\frac{\bar{\tau}}{2}\|h_k\|^2-(1-\beta)\kappa\|c_k\|_1+\mathrm{o}(\|c_k\|)+\mathrm{o}(\|h_k\|^2)\\
&\quad\leqslant 0
\end{aligned}\tag{5.16}$$

and hence

$$\begin{aligned}
\vartheta(x_k+s_k+d_k,\rho_k)-\vartheta(x_{l(k)},\rho_{l(k)}) &\leqslant \vartheta(x_k+s_k+d_k,\rho_k)-\vartheta(x_k,\rho_k)\\
&\leqslant \beta D\vartheta(x_k,\rho_k;s_k)
\end{aligned}\tag{5.17}$$

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
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.,

$$\lim_{k\to\infty} \frac{||x_{k+1} - x_*||}{||x_{k-1} - x_*||} = 0. \tag{5.18}$$

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

$$\lim_{k\to\infty} \frac{||x_k + s_k - x_*||}{||x_{k-1} + s_{k-1} - x_*||} = 0. \tag{5.19}$$

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

Problem name	$M=0$		$M=4$			$M=8$		
	NF	NG	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	38	21	34	16	1	35	17	1
HS028	58	31	58	31	0	58	31	0
HS060	8	7	8	7	0	8	7	0
SC220	142	103	128	91	3	119	91	4
SC252	54	48	65	46	2	68	46	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	58	47	51	40	1	51	40	1
HS028	65	42	65	42	0	65	42	0
HS060	8	7	8	7	0	8	7	0
SC220	145	113	136	93	4	91	77	7
SC252	59	52	69	56	2	69	56	2
SC235	31	16	29	15	1	29	15	1
SC219	62	49	57	49	2	46	35	3
SC216	79	64	72	52	1	72	52	1

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 $||P(x_k,B_k)g^k||+||c_k||\leqslant 10^{-6}$ and $|\vartheta(x_k,\rho_k)-\vartheta(x_{k+1},\rho_{k+1})|\leqslant 10^{-6}|\vartheta(x_k,\rho_k)|$.

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 Γ_k depend on the value of τ_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

[1] J.R. Bunch, B.N. Parlett, Direct method for solving symmetric indefinite systems of linear equations, SIAM J. Numer. Anal. 8 (1971) 639–655.
[2] J.P. Bulteau, J.Ph. Vial, Curvilinear path and trust region in unconstrained optimization, a convergence analysis, Math. Programming Study 30 (1987) 82–101.
[3] R.H. Byrd, R.B. Schnable, Continuity of the null space basis and constrained optimization, Math. Programming 35 (1986) 32–41.
[4] T.F. Coleman, A.R. Conn, On the local convergence of a quasi-Newton method for the nonlinear programming problem, SIAM J. Numer. Anal. 21 (1984) 755–769.
[5] R. Fletcher, Practical Methods of Optimization, Wiley, New York, 1984.
[6] R. Fontecilla, Local convergence of secant methods for nonlinear constrained optimization, SIAM J. Numer. Anal. 25 (1988) 692–712.
[7] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotonic line search technique for Newton's methods, SIAM J. Numer. Anal. 23 (1986) 707–716.
[8] W. Hock, K. Schittkowski, in: Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems, vol. 187, Springer, Berlin, 1981.
[9] J. Nocedal, M.L. Overton, Projected Hessian updating algorithms for nonlinearly constrained optimization, SIAM J. Numer. Anal. 22 (1985) 821–850.
[10] M.J.D. Powell, Y. Yuan, A recursive quadratic programming algorithm that uses differentiable exact penalty functions, Math. Programming 35 (1986) 265–278.
[11] K. Schittkowski, in: More Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems, vol. 282, Springer, Berlin, 1987.
[12] T. Stehaug, The conjugent method and trust regions in large scale optimization, SIAM J. Numer. Anal. 20 (1983) 626–637.
[13] J.Z. Zhang, C.X. Xu, A scaled optimal path trust region alogirthm, Research Report MA-96-13, City University of Hong Kong, 1996.
[14] J. Zhang, D. Zhu, A Projective quasi-Newton method for nonlinear optimization, J. Comput. Appl. Math. 53 (1990) 291–307.