Journal of Mathematical Research & Exposition
Vol.24, No.4, 627–648, November, 2004

Curvilinear Paths with Nonmonotonic Inexact Line Search Technique for Unconstrained Optimization *

ZHU De-tong
(Dept. of Math., Shanghai Normal University, Shanghai 200234, China)

Abstract: In this paper we modify approximate trust region methods via three preconditional curvilinear paths for unconstrained optimization. To easily form preconditional curvilinear paths within the trust region subproblem, we employ the stable Bunch-Parlett factorization method of symmetric matrices and use the unit lower triangular matrix as a preconditioner of the optimal path and modified gradient path. In order to accelerate the preconditional conjugate gradient path, we use preconditioner to improve the eigenvalue distribution of Hessian matrix. Based on the trial steps produced by the trust region subproblem along the three curvilinear paths providing a direction of sufficient descent, we mix a strategy using both trust region and nonmonotonic line search techniques which switch to back tracking steps when a trial step is unacceptable. Theoretical analysis is given to prove that the proposed algorithms are globally convergent and have a local superlinear convergent rate under some reasonable conditions. The results of the numerical experiment are reported to show the effectiveness of the proposed algorithms.

Key words: Curvilinear paths; preconditioner; trust region methods; nonmonotonic technique.

Classification: AMS(2000) 90C30, 65K05/CLC number: O221.2

Document code: A **Article ID**: 1000-341X(2004)04-0627-22

1. Introduction

In this paper we consider the nonlinear unconstrained minimization problems

$$\min f(x), \tag{1.1}$$

where $f(x) : \Re^n \to \Re^1$ is continuously differentiable. The trust region method is a very popular way for unconstrained minimization to assure global convergence. Many different versions have been suggested in using trust region strategy. An attractive idea is to solve

*Received date: 2002-03-11
Foundation item: Supported by National Science Foundation of China (10071050), Science Foundation of Shanghai Technical Sciences Committee (02ZA14070), Science Foundation of Shanghai Education Committee (02DK06)
Biography: ZHU De-tong (1954-), male, Ph.D., Professor.

the trust region subproblem approximately along a curve originating at x_k within the trust region. The various curvilinear path trust region algorithms are available (see [2], [5] and [15], for example). Bulteau and Vial proposed in [2] three curvilinear paths that are named as the conjugate gradient path, optimal path and modified gradient path within the trust regions. The conjugate gradient path make use of the eigenvalue distribution of Hessian matrix of the quadratic model function, then the convergence rate will depend on the eigenvalue and the calculation may be expensive and not stationary. Therefore, we aim to choose a preconditioner such that the eigenvalues of preconditional Hessian are more favorable for the convergence theory. The optimal path and modified gradient path can be expressed by the eigenvalues and eigenvectors of the Hessian matrix of the quadratic model function. However, a calculation of the full eigensystem of a symmetric matrix is usually time-consuming, and the optimal path and modified gradient path algorithms generally are impractical. Recently Zhang and Xu in [17] employ the stable Bunch-Parlett factorization method to factorize the Hessian to form a scaling optimal path within the trust region for unconstrained optimization.

In trust region algorithms it is sometimes helpful to include a preconditioner which is diagonal and fixed for the variables. In this paper, this idea of the stable Bunch-Parlett factorization motives to extend to preconditional optimal path and preconditional modified gradient path which can be easily formulated from the full eigensystem of the block matrix that is very easy to calculate. Meantime, we improve the eigenvalue in the preconditional conjugate gradient path by constructing a symmetric and positive define matrix as a preconditioner, such as the incomplete Cholesky approach or the stable Bunch-Parlett factorization. Another preconditional method may arise in linear constrained optimization solved by using affine scaling interior algorithms. On the other hand, we also noticed that Nocedal and Yuan[11] suggested a combination of the trust region and line search methods. In particular, in order to avoid expensive computation at the acceptable successful trial step, we shall show that the trial step on the three preconditional curvilinear paths should provide a direction of sufficient descent so that it can be accepted by employing the back tracking technique although the trial step maybe unsuccessful. Another valuable idea is to abandon the traditional monotonic decreasing requirement for the sequence $\{f(x_k)\}$ of the objective values (see [4] and [9]), because monotonicity may cause a series of very small steps if the contours of the objective function f are a family of curves with large curvature.

The main purpose of this paper is to propose an approximate trust region method with three preconditional curvilinear paths by adopting nonmonotonic back tracking technique. The paper is organized as follows. In Section 2, we propose the characterizations and properties of the three preconditional curvilinear paths in the trust region subproblem. In Section 3, we describe the algorithms which combine the techniques of trust region, back tracking and nonmonotonic search, and partial update factorizing method of decompositions of the matrices. In Section 4, the weak global convergence of the proposed algorithms are established and some further convergence properties such as strong global convergence and superlinear convergence rate are discussed. Finally, the results of numerical experiments of the proposed algorithms are reported in Section 5.

2. Curvilinear paths

An important portion of the unconstrained minimization procedure will be concerned with the solution of the trust region subproblem of the quadratic model form

$$\begin{aligned} \min \quad & q_k(\delta) \stackrel{\text{def}}{=} f_k + (g^k)^T\delta + \tfrac{1}{2}\delta^T B_k \delta, \\ \text{s.t.} \quad & \|\delta\| \le \Delta_k, \end{aligned} \tag{2.1}$$

where $f_k = f(x_k)$, $g^k = \nabla f(x_k)$, $\delta = x - x_k$, B_k is either $\nabla^2 f(x_k)$ or its approximation, Δ_k is the trust region radius, and $\|\cdot\|$ is the 2-norm. Let δ_k be the solution of the subproblem. Then set the next step

$$x_{k+1} = x_k + \delta_k.$$

In trust region algorithms it is sometimes helpful to include a scaling matrix for the variables. In most cases, the scaling matrix is diagonal and fixed. We will employ the stable Bunch-Parlett factorization method of symmetric matrices to factorize the Hessian matrix of the quadratic model function. The factorization method (see [1]) factorizes the matrix B_k into the form

$$P_k B_k P_k^T = L_k D_k L_k^T, \tag{2.2}$$

where P_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. The elements of the matrices $\{L_k\}$ and $\{L_k^{-1}\}$ are bounded by two fixed positive constants independent of the matrix B_k, i.e., there exist positive constants $c_1 \le c_2$ such that for all k (see [17]),

$$c_1 \le \|L_k\| \le c_2. \tag{2.3}$$

In our preconditional curvilinear paths type of trust region algorithm, at kth iteration, the matrix $L_k^T P_k$ is used to scale the variables

$$\widehat{\delta} = L_k^T P_k \delta, \tag{2.4}$$

and the preconditional trust region subproblem takes the form

$$\min \widehat{q}_k(\widehat{\delta}) \stackrel{\text{def}}{=} f_k + (\widehat{g}^k)^T\widehat{\delta} + \frac{1}{2}\widehat{\delta}^T D_k \widehat{\delta}, \quad \|\widehat{\delta}\| \le \Delta_k, \tag{2.5}$$

where $\widehat{g}^k = L_k^{-1} P_k g^k$. Note that in this problem, $\widehat{\delta}$ rather than $\delta = P_k^T L_k^{-T}\widehat{\delta}$ is required to be within the trust region, which will further improve the efficiency of the calculation of the solution step. Based on solving the about trust region subproblem, we give the following lemma which is due to Sorensen's paper in [15].

Lemma 2.1 *$\widehat{\delta}_k$ is a solution to the subproblem (2.5) if and only if it is a solution to the following equations of the forms*

$$(D_k + \mu_k I)\,\widehat{\delta}_k = -\widehat{g}, \tag{2.6}$$

$$\mu_k(\|\widehat{\delta}_k\| - \Delta_k) = 0, \qquad \mu_k \geq 0, \tag{2.7}$$

and $D_k + \mu_k I$ is positive semidefinite.

Lemma 2.1 establishes the necessary and sufficient conditions concerning the pair μ_k, $\widehat{\delta}_k$ when $\widehat{\delta}_k$ solves (2.5). The preconditional optimal path is concerned with the solution of Systems (2.6) and (2.7).

The preconditional modified gradient path emerging at the current point x_k of a general continuously differentiable function f is the solution of the differentiable equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = -\nabla f(x(t)), \quad \text{and} \quad x(0) = x_k. \tag{2.8}$$

For general function f, the solution of (2.8) could be obtained by numerical integration methods. However, if f is quadratic, a closed form solution of (2.8) exists. It motivates to use the local quadratic approximate to f in a neighorhood of x_k. Let x_k be the kth iteration and

$$\widehat{q}_k(\widehat{\delta}(t)) \overset{\text{def}}{=} f_k + (\widehat{g})^T\widehat{\delta}(t) + \frac{1}{2}\widehat{\delta}(t)^T D_k \widehat{\delta}(t)$$

be the local quadratic approximation of f at x_k, where $\widehat{\delta}(t) = L_k^T P_k \delta(t)$, $\delta(t) = x(t) - x_k$, and D_k is given in (2.2). Then the solution of

$$\frac{\mathrm{d}\widehat{\delta}(t)}{\mathrm{d}t} = -\nabla q_k(\widehat{\delta}(t)), \quad \text{and} \quad \widehat{\delta}(0) = 0$$

is a valid approximation of the curvilinear paths of f, and thus provides a set of sensible candidates for a successor point $x_{k+1} = x_k + \widehat{\delta}(t_k)$ to x_k.

Now, the idea of general curvilinear paths proposed by Bulteau and Vial (see [2]) motivates to form the two preconditional curvilinear paths, i.e., preconditional optimal path and preconditional modified gradient path, respectively. When the trust region radius Δ_k of the subproblem (2.5) varies in the interval $[0, +\infty)$, the solution points form the preconditional paths and emanate from the origin. In order to define those arcs in a closed form, we shall use the eigensystem decomposition of B. Since D is a block diagonal matrix with 1×1 and 2×2 diagonal blocks, without lost of generality, let $\varphi_1 \leq \varphi_2 \leq \cdots \leq \varphi_n$ be eigenvalues of D and u^1, $u^2, \cdots, u^n$ be corresponding orthonormal eigenvectors. We partition the set $\{1, \ldots, n\}$ into $\mathcal{I}^+$, $\mathcal{I}^-$ and $\mathcal{N}$ according to $\varphi_i > 0$, $\varphi_i < 0$ and $\varphi_i = 0$ for $i \in \{1, \ldots, n\}$, respectively. We now give two preconditional curvilinear paths.

Preconditional optimal path

The preconditional optimal path $\Gamma(\tau)$ can be expressed as

$$\Gamma(\tau) = \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)), \tag{2.9}$$

where

$$\Gamma_1(t_1(\tau)) = -\,[\,\sum_{i\in\mathcal{I}} \frac{t_1(\tau)}{\varphi_i t_1(\tau) + 1}\widehat{g}_i^k u^i + t_1(\tau)\sum_{i\in\mathcal{N}} \widehat{g}_i^k u^i\,],$$

$$\Gamma_2(t_2(\tau)) = t_2(\tau)u^1,$$

and

$$\begin{array}{lll} t_1(\tau)=\tau \text{ and } t_2(\tau)=0, & \text{if} & \tau<\frac{1}{T}, \\ t_1(\tau)=\frac{1}{T} \text{ and } t_2(\tau)=\tau-\frac{1}{T}, & \text{if} & \tau\geq\frac{1}{T}, \end{array}$$

$\mathcal{I}_k=\mathcal{I}_k^+\cup\mathcal{I}_k^-$, $\widehat{g}_i^k=(\widehat{g}^k)^T u^i$, $i=1,\cdots,n$, $\widehat{g}^k=\sum_{i=1}^n \widehat{g}_i^k u^i$, $T=\max\{0,-\varphi_1\}$ and $1/T$ is defined as $+\infty$ if $T=0$. It should be noted that $\Gamma_2(t_2(\tau))$ is defined only when D is indefinite and $\widehat{g}_i^k=0$ for all $i\in\{1,\cdots,n\}$ with $\varphi_i=\varphi_1<0$ and that for other cases, $\Gamma(\tau)$ is defined only for $0\leq\tau<\frac{1}{T}$, that is, $\Gamma(\tau)=\Gamma_1(t_1(\tau))$.

Preconditional modified gradient path

The preconditional modified gradient path can be given in the following closed form (refer to see [2]):

$$\Gamma(\tau)=\Gamma_1(t_1(\tau))+\Gamma_2(t_2(\tau)),\quad \tau\in[0,+\infty), \tag{2.10}$$

where if $\widehat{g}_i^k\neq 0$ for some $i\in\mathcal{I}^-\bigcup\mathcal{N}$, the term $\Gamma_2(t_2(\tau))$ is not relevant, that is, if $\widehat{g}_i^k\neq 0$ for some $i\in\mathcal{I}^-\bigcup\mathcal{N}$, 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

$$\Gamma_1(t_1(\tau))=\sum_{i\in\mathcal{I}}\frac{\exp\{-\varphi_i t_1(\tau)\}-1}{\varphi_i}\widehat{g}_i^k u^i-t_1(\tau)\sum_{i\in\mathcal{N}}\widehat{g}_i^k u^i,$$

with

$$\Gamma_2(t_2)=\begin{cases} t_2u^1, & \text{if } \varphi_1<0,\\ 0, & \text{if } \varphi_1\geq 0,\end{cases}\qquad t_1(\tau)=\begin{cases}\frac{\tau}{1-\tau}, & \text{if } \tau<1,\\ +\infty, & \text{if } \tau\geq 1,\end{cases}$$

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

Preconditional conjugate gradient path

Suppose that we apply a standard preconditional conjugate direction algorithm to the quadratic but not necessarily convex function $q_k(\delta)$, starting from x_0. We construct a symmetric and positive define matrix $M_k=C_k^TC_k$ with C_k as a preconditioner, such as the incomplete Cholesky approach or the stable Bunch-Parlett factorization $C_k=(L_k^TP_k)^{-1}$ where the process does not make use of C_k explicitly. Then we generate a sequence of $v_1,\cdots,v_{m+1}$ $(v_1=0)$ and a sequence of preconditional conjugate direction $d_1,\cdots,d_{m+1}$ $(d_1=s_0,\ s_0=M_k^{-1}g^k)$. For $i=1,2,\cdots,m$, perform the iteration of the preconditional conjugate gradient path

$$s_{i+1}=M_k^{-1}g_{i+1},\quad i=1,\cdots,m; \tag{2.11}$$

$$d_{i+1}=s_{i+1}+\beta_i d_i,\quad i=1,\cdots,m; \tag{2.12}$$

$$v_{i+1}=v_i+\gamma_i d_i,\quad i=1,\cdots,m; \tag{2.13}$$

$$d_i^TH_kd_i>0,\quad i=1,\cdots,m, \tag{2.14}$$

where for $i=1,\cdots,m$

$$g_{i+1}=\nabla q_k(v_{i+1})=H_kv_{i+1}+g^k,\ \beta_i=\frac{g_{i+1}^Tv_i}{d_i^TH_kd_i}>0 \text{ and } \gamma_i=\frac{g_i^Tv_i}{d_i^TH_kd_i}>0. \tag{2.15}$$

The procedure stops either because $g_{i+1} = 0$ or because $g_{i+1} \neq 0$ but $p_{i+1}^T H_k p_{i+1} \leq 0$. In the former case v_{i+1} is a critical point of q_k; in the latter d_{i+1} is a descent direction (see [2]). Now, we define the preconditional conjugate gradient path by

$$\Gamma_k(\tau) = \sum_{i=1}^{m} t_i(\tau) d_i - t_{m+1}(\tau) d_{m+1} \tag{2.16}$$

and

$$t_i(\tau) = \min\{\gamma_i, \max\{0, \tau - \sum_{j=1}^{i-1} \gamma_j\}\},$$

where the conjugate direction d_i and γ_i are defined by (2.11)–(2.17) which have the property that

$$q_k(v_i + \gamma_i d_i) = \min_{\gamma} \quad q_k(v_i + \gamma d_i).$$

In this formula we take $\sum_{j=1}^{i-1} \gamma_j = 0$ for $i = 1$.

Properties of preconditional curvilinear paths

It is well known from solving the trust region algorithms to obtain the global convergence of the proposed algorithm, it is sufficient to show at kth iteration the sufficient descent condition of the predicted reduction defined by

$$\text{Pred}(\delta_k) = f_k - q_k(\delta_k) = f_k - \widehat{q}_k(\widehat{\delta}_k),$$

where $\delta_k = P_k^T L_k^{-1} \widehat{\delta}_k$ and $\widehat{\delta}_k$ is obtained by the step $\widehat{\delta}_k$ from the preconditional curvilinear paths in trust region. In this paper we only discuss the properties of preconditional conjugate gradient path in detail and summarize the properties of the other two paths as the following two lemmas, whose proofs are similar to those in [18].

Lemma 2.2 *Let the step $\widehat{\delta}_k$ in trust region subproblem be obtained from the preconditional optimal path. Then the norm function of the path is monotonically increasing for $\tau \in (0, +\infty)$, and there exists $\tau_k \in (0, +\infty)$ such that the point $\Gamma(\tau_k)$ on the path with $\|\Gamma(\tau_k)\| = \Delta_k$ satisfies the systems (2.6)–(2.7) with $\mu_k \geq 0$ given as follows*

$$\mu_k = \frac{1}{\tau_k} \quad \text{as} \quad \tau_k < \frac{1}{T_k}, \tag{2.18}$$

$$\mu_k = \frac{1}{T_k}, \; t_2(\tau_k) = \tau_k - \frac{1}{T_k} \quad \text{as} \quad \tau_k \geq \frac{1}{T_k}, \tag{2.19}$$

where $T_k = \max\{0, -\varphi_1\}$. Furthermore, the predicted reduction $\widehat{q}_k(\widehat{\delta}_k)$ and the descent direction $(g^k)^T \delta_k$, respectively, satisfy the sufficient descent conditions

$$\text{Pred}(\delta_k) = f_k - \widehat{q}_k(\widehat{\delta}_k) \geq \omega_1 \|\widehat{g}^k\| \min\{\Delta_k, \frac{\|\widehat{g}^k\|}{\|D_k\|}\}; \tag{2.20}$$

$$-(g^k)^T \delta_k = -(\widehat{g}^k)^T \widehat{\delta}_k \geq \omega_2 \|\widehat{g}^k\| \min\{ \Delta_k, \frac{\|\widehat{g}^k\|}{\|D_k\|} \} \tag{2.21}$$

for all $\widehat{g}^k$, D_k and Δ_k, where $\omega_1 \ge \omega_2 > 0$ are some constants independent of k.

Lemma 2.3 *Let the step $\widehat{\delta}_k$ in trust region subproblem be obtained from the preconditional modified gradient path. Then we have that the norm function of the path is monotonically increasing for $\tau \in (0, +\infty)$, and there exists $\tau_k \in (0, +\infty)$ such that the point $\Gamma(\tau_k)$ on the path with $\|\Gamma(\tau_k)\| = \Delta_k$. Further, the predicted reduction $\widehat{q}_k(\widehat{\delta}_k)$ and the descent direction $(\widehat{g}^k)^T\widehat{\delta}_k$ also satisfy the sufficient descent conditions (2.20) and (2.21), respectively.*

Lemma 2.4 *Let the step $\widehat{\delta}_k$ in trust region subproblem be obtained from the preconditional conjugate gradient path. Then we have that the norm function of the path is monotonically increasing for $\tau \in (0, +\infty)$, and there exists $\tau_k \in (0, +\infty)$ such that the point $\Gamma(\tau_k)$ on the path with $\|\Gamma(\tau_k)\| = \Delta_k$. Further, the predicted reduction $\widehat{q}_k(\widehat{\delta}_k)$ and the descent direction $(\widehat{g}^k)^T\widehat{\delta}_k$ also satisfy the sufficient descent conditions (2.20) and (2.21), respectively.*

Proof Consider the following subproblem with preconditioner C_k,

$$\begin{aligned} \min \quad & \varphi_k(\delta) = f_k + (g^k)^T\delta + \frac{1}{2}\delta^T H_k \delta, \\ \text{s.t.} \quad & \|C_k^{-1}\delta\| \le \Delta_k, \qquad \delta \in \Gamma_k. \end{aligned} \tag{2.22}$$

We omit the kth iteration. If $\sum_{j=1}^{i-1}\gamma_j < \tau \le \sum_{j=1}^{i}\gamma_j \ \ (i \le m)$, and

$$\Gamma(\tau) = \sum_{j=1}^{i-1}\gamma_j d_j + (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i, \tag{2.23}$$

which imply $0 < \tau - \sum_{j=1}^{i-1}\gamma_j \le \gamma_i$, we have

$$t_i(\tau) = \min\{\gamma_i, \max\{0, \tau - \sum_{j=1}^{i-1}\gamma_j\}\} = \min\{\gamma_i, \tau - \sum_{j=1}^{i-1}\gamma_j\} = \tau - \sum_{j=1}^{i-1}\gamma_j.$$

for $k < i$,

$$t_k(\tau) = \min\{\gamma_k, \max\{0, \tau - \sum_{j=1}^{k-1}\gamma_j\}\} = \min\{\gamma_k, \tau - \sum_{j=1}^{k-1}\gamma_j\} = \gamma_k.$$

For $i < k \le m$, $t_k(\tau) = \min\{\gamma_k, 0\} = 0$, so (2.23) holds. When $\tau > \sum_{j=1}^{m}\gamma_j$, we get $\Gamma(\tau) = \sum_{j=1}^{m}\gamma_j d_j$. So the norm function

$$\vartheta(\tau) = \|\sum_{j=1}^{i-1}\gamma_j d_j\|^2 + 2(\tau - \sum_{j=1}^{i-1}\gamma_j)\sum_{j=1}^{i-1}\gamma_j d_j^T d_i + (\tau - \sum_{j=1}^{i-1}\gamma_j)^2\|d_i\|^2.$$

By induction, we get $d_i^T d_j > 0 \ (j < i)$. From $\gamma_j > 0 \ \ (j \le m)$ and $d_i^T d_j > 0 \ (j < i)$, we get that

$$\vartheta'(\tau) = 2\sum_{j=1}^{i-1}\gamma_j d_j^T d_i + 2(\tau - \sum_{j=1}^{i-1}\gamma_j)\|d_i\|^2 > 0,$$

which means that $\|\Gamma(\tau)\|$ is monotonically increasing on $\Re^+$.

When $\sum_{j=1}^{i-1}\gamma_j < \tau \le \sum_{j=1}^{i}\gamma_j$ $(i \le m)$, and noting $d_1 = -g^k$, we have that

$$(g^k)^T\Gamma(\tau) = -d_1^T[\sum_{j=1}^{i-1}\gamma_j d_j + (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i] = -\sum_{j=1}^{i-1}\gamma_j d_j^T d_1 - (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i^T d_1.$$

and from (2.11)–(2.14),

$$\begin{aligned}\Gamma(\tau)^T H_k\Gamma(\tau) &= [\sum_{j=1}^{i-1}\gamma_j d_j^T H_k + (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i^T H_k][\sum_{j=1}^{i-1}\gamma_j d_j + (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i].\\ &= \sum_{j=1}^{i-1}\gamma_j^2 d_j^T H_k d_j + (\tau - \sum_{j=1}^{i-1}\gamma_j)^2 d_i^T H_k d_i.\end{aligned}$$

Therefore,

$$\begin{aligned}\varphi(\Gamma(\tau)) &\stackrel{\text{def}}{=} f_k + (g^k)^T\Gamma(\tau) + \frac{1}{2}\Gamma(\tau)^T H_k\Gamma(\tau)\\ &= f_k - \sum_{j=1}^{i-1}\gamma_j d_j^T d_1 - (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i^T d_1 +\\ &\quad \frac{1}{2}\sum_{j=1}^{i-1}\gamma_j^2 d_j^T H_k d_j + \frac{1}{2}(\tau - \sum_{j=1}^{i-1}\gamma_j)^2 d_i^T H_k d_i.\end{aligned}$$

As $\tau - \sum_{j=1}^{i-1}\gamma_j \le \gamma_i$, we get

$$\begin{aligned}\frac{\mathrm{d}\varphi(\Gamma(\tau))}{\mathrm{d}\tau} &= -d_i^T d_1 + (\tau - \sum_{j=1}^{i-1}\gamma_j)d_i^T H_k d_i\\ &\le -d_i^T d_1 + \gamma_i d_i^T H_k d_i\\ &= -d_i^T d_1 - \frac{g_i^T d_i}{d_i^T H_k d_i} d_i^T H_k d_i\\ &= -d_i^T(d_1 + g_i)\\ &= -d_i^T(-g_1 + g_1 + \sum_{j=1}^{i-1}\gamma_j H_k d_j)\\ &= 0,\end{aligned}$$

which implies that $\varphi(\Gamma(\tau))$ monotonically increasing on $\tau \in (0, \infty)$. Therefore, we have that

$$\begin{aligned}\text{Pred}(\delta_k) &= f_k - \varphi_k(\delta_k)\\ &\ge f_k - \inf\{\varphi(\Gamma(\tau)) \mid \|C_k^{-1}\Gamma(\tau)\| \le \Delta_k\}\\ &\ge f_k - \inf_{\tau \le \gamma_1}\{\varphi(\Gamma(\tau)) \mid \|C_k^{-1}\Gamma(\tau)\| \le \Delta_k\}.\end{aligned}$$

When $\tau \le \gamma_1$, since $\Gamma(\tau) = -\tau M_k^{-1} g^k$, we have from $M_k = C_k^T C_k$, $C_k \stackrel{\text{def}}{=} (L_k P_k)^{-1}$ and $\widehat{g}^k = L_k^{-1} P_k g_k$ that

$$\begin{aligned}\text{Pred}(\delta_k) &\ge \max_{\tau \le \min\{\gamma_1,\ \frac{\Delta_k}{\|\widehat{g}^k\|}\}} \{-\frac{\tau^2}{2}(g^k)^T H_k g^k + \tau \|g^k\|^2\} \\ &= \max_{\tau \le \min\{\gamma_1,\ \frac{\Delta_k}{\|\widehat{g}^k\|}\}} \{-\frac{\tau^2}{2}(\widehat{g}^k)^T D_k \widehat{g}^k + \tau \|\widehat{g}^k\|^2\}.\end{aligned}$$

If $\gamma_1 \le \frac{\Delta_k}{\|\widehat{g}^k\|}$, then $\tau = \frac{\|\widehat{g}^k\|^2}{(\widehat{g}^k)^T D_k \widehat{g}^k} = \gamma_1$ is the solution of the above subproblem, and

$$\begin{aligned}\text{Pred}(\delta_k) &\ge -\frac{(\widehat{g}^k)^T D_k \widehat{g}^k}{2}\left(\frac{\|\widehat{g}^k\|^2}{(\widehat{g}^k)^T D_k \widehat{g}^k}\right)^2 + \frac{\|\widehat{g}^k\|^4}{(\widehat{g}^k)^T D_k \widehat{g}^k} \\ &= \frac{\|\widehat{g}^k\|^4}{2(\widehat{g}^k)^T D_k \widehat{g}^k} \ge \frac{\|\widehat{g}^k\|^2}{2\|D_k\|} \\ &\ge \frac{1}{2}\|\widehat{g}^k\| \min\{\Delta_k,\ \frac{\|\widehat{g}^k\|}{\|D_k\|}\}.\end{aligned}$$

If $\gamma_1 > \frac{\Delta_k}{\|\widehat{g}^k\|}$, i.e., $\frac{(\widehat{g}^k)^T D_k \widehat{g}^k}{\|\widehat{g}^k\|^2} < \frac{\|\widehat{g}^k\|}{\Delta_k}$, then $\tau = \frac{\Delta_k}{\|\widehat{g}^k\|}$ is the solution of the above subproblem, and

$$\begin{aligned}\text{Pred}(\delta_k) &\ge -\Delta_k^2 \frac{(\widehat{g}^k)^T D_k \widehat{g}^k}{2\|\widehat{g}^k\|^2} + \Delta_k \|\widehat{g}^k\| \\ &\ge -\Delta_k^2 \frac{\|\widehat{g}^k\|}{2\Delta_k} + \Delta_k \|\widehat{g}^k\| = \frac{\Delta_k \|\widehat{g}^k\|}{2} \\ &\ge \frac{1}{2}\|\widehat{g}^k\| \min\{\Delta_k,\ \frac{\|\widehat{g}^k\|}{\|D_k\|}\}.\end{aligned}$$

From the above two cases, we get that

$$\text{Pred}(\delta_k) \ge \frac{1}{2}\|\widehat{g}^k\| \min\{\Delta_k, \frac{\|\widehat{g}^k\|}{\|D_k\|}\}.$$

On the other hand, noting the definition of $\text{Pred}(\delta_k)$ and $d_j^T H_k d_j > 0 \ \ (j = 1, \cdots, i-1)$, we have

$$\begin{aligned}(g^k)^T \delta_k = (\widehat{g}^k)^T \widehat{\delta}_k &= -\text{Pred}(\delta_k) - \frac{1}{2}\delta_k^T H_k \delta_k \\ &= -\text{Pred}(\delta_k) - \frac{1}{2}\sum_{j=1}^{i-1} \gamma_j^2 d_j^T H_k d_j - \frac{1}{2}(\tau - \sum_{j=1}^{i-1} \gamma_j)^2 d_i^T H_k d_i \\ &\le -\text{Pred}(\delta_k) \\ &\le -\frac{1}{2}\|\widehat{g}^k\| \min\{\Delta_k, \frac{\|\widehat{g}^k\|}{\|D_k\|}\}.\end{aligned}$$

This completes the proof.

The above lemmas show the relation between the gradient $\widehat{g}^k$ of the objective function and the step $\widehat{\delta}_k$ generated from the various paths which are a sufficiently descent direction.

3. Algorithms

In this section we describe a method combining nonmonotonic line search technique with an approximate trust region algorithm, which uses preconditional curvilinear paths instead of a minimization in the whole trust region.

Initialization step Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0,1)$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $\varepsilon > 0$ and positive integer M. Let $m(0) = 0$. Choose a symmetric matrix B_0. Select an initial trust region radius Δ_0 and a maximal trust region radius $\Delta_{\max}$ such that $\Delta_{\max} \geq \Delta_0 > 0$, and give a starting point $x_0 \in \Re^n$. Set $k = 0$, then go to the main step.

Main Step

1. Evaluate $f_k = f(x_k)$, $g^k = \nabla f(x_k)$.

2. If $\|g^k\| \leq \varepsilon$, stop with the approximate solution x_k.

3. Factorize B_k into the form (2.2) and calculate the eigenvalues and orthonormal eigenvectors of D_k. Form three preconditional paths Γ_k, the preconditional conjugate gradient path, the preconditional optimal path or the preconditional modified gradient path.

4. Solve the subproblem via the preconditional optimal path or the preconditional modified gradient path

$$(S_k) \begin{array}{ll} \min & \psi_k(\delta) = \widehat{\psi}_k(\widehat{\delta}) \stackrel{\text{def}}{=} (\widehat{g}^k)^T \widehat{\delta} + \frac{1}{2} \widehat{\delta}^T D_k \widehat{\delta} \\ \text{s.t.} & \|\widehat{\delta}\| \leq \Delta_k, \ \widehat{\delta} \in \Gamma_k. \end{array}$$

   Denote by $\widehat{\delta}_k$ the solution of the subproblem (S_k). Solve the subproblem via the preconditional conjugate gradient path

$$(S_k') \begin{array}{ll} \min & \psi_k(\delta) \stackrel{\text{def}}{=} (g^k)^T \delta + \frac{1}{2} \delta^T H_k \delta, \\ \text{s.t.} & \|C_k^{-1} \delta\| \leq \Delta_k, \quad \delta \in \Gamma_k. \end{array}$$

   Denote by δ_k the solution of the subproblem (S_k').

5. Let $\delta_k = P_k^T L_k^{-1} \widehat{\delta}_k$, and $\widehat{\delta}_k$ be solution of the subproblem (S_k) or δ_k be solution of the subproblem (S_k'). Choose $\lambda_k = 1, \ \omega, \ \omega^2, \ \cdots$ until the following inequality is satisfied

$$f(x_k + \lambda_k \delta_k) \leq f(x_{l(k)}) + \lambda_k \beta (g^k)^T \delta_k, \tag{3.1}$$

   where $f(x_{l(k)}) = \max_{0 \leq j \leq m(k)} \{f(x_{k-j})\}$.

6. Set

$$h_k = \lambda_k \delta_k, \tag{3.2}$$

$$x_{k+1} = x_k + h_k. \tag{3.3}$$

Calculate

$$\mathrm{Pred}(h_k) = -\psi_k(h_k), \tag{3.4}$$

$$\widehat{\mathrm{Ared}}(h_k) = f(x_{l(k)}) - f(x_k + h_k), \tag{3.5}$$

$$\rho_k = \frac{\mathrm{Ared}(h_k)}{\mathrm{Pred}(h_k)}, \tag{3.6}$$

and take

$$\Delta_{k+1} = \begin{cases} [\gamma_1 \Delta_k,\ \gamma_2 \Delta_k], & \text{if } \widehat{\rho}_k \le \eta_1, \\ (\gamma_2 \Delta_k,\ \Delta_k], & \text{if } \eta_1 < \widehat{\rho}_k < \eta_2, \\ (\Delta_k,\ \min\{\gamma_3 \Delta_k, \Delta_{\max}\}], & \text{if } \widehat{\rho}_k \ge \eta_2. \end{cases}$$

Calculate $f(x_{k+1})$ and g^{k+1}.

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 As shown below, the preconditional curvilinear paths can be generated by employing general symmetric matrices which may be indefinite. In each iteration the algorithm solves only one trust region subproblem. If the solution δ_k fails to meet the acceptance criterion (3.1) (take $\lambda_k = 1$), then we turn to line search, i.e., retreat from $x_k + \delta_k$ until the criterion is satisfied. It is easy to see the usual monotone algorithm can be viewed as a special case of the proposed algorithm in the case $M = 0$.

Remark 2 In the subproblems (S_k) and (S'_k), a candidate iterative direction $\widehat{\delta}$ is generated by minimizing $\psi_k(\widehat{\delta})$ along the curve paths Γ_k within the ball centered at x_k with radius Δ_k. As being proved in [2], moving along these Γ_k with x_k as the starting point, the distance to x_k is increasing, but the value of $\psi_k(\delta)$ is decreasing. Therefore, the subproblem (S_k) can be solved with great ease.

Since the preconditional optimal path and preconditional modified gradient path algorithms generally require both calculation of the full eigensystem of the symmetric matrix B_k and repeated decompositions of the matrices, forming these paths are usually time-consuming and impractical. We suggest another way of updating the approximate Hessian matrix B_k that under reasonable conditions preserves uniform boundedness of $\{\|B_k\|\}$ and, in addition, incorporates second-order information using the decomposing technique.

A partial update method is suggested for the approximate Hessian matrix B_k. That is, instead of using exact updating matrix in the computations, we may use an approximate updating matrix that satisfies the inexact Newton methods (see [19]).

4. Convergence analyses

Throughout this section we assume that $f : \Re^n \to \Re^1$ is twice continuously differentiable and bounded from below. Given $x_0 \in \Re^n$, the algorithm generates a sequence $\{x_k\} \subset \Re^n$. In our analysis, we denote the level set of f by

$$\mathcal{L}(x_0) = \{\ x \in \Re^n \mid f(x) \le f(x_0)\ \}.$$

An estimate is expressed in terms of the function

$$w_k(f, x_k, \delta_k) = \frac{f(x_k + \delta_k) - f_k - (g^k)^T \delta_k}{\|\delta_k\|^2}.$$

Since f is twice differentiable, the mean-value theorem implies that

$$w_k(f, x_k, \delta_k) = 2\int_0^1 \int_0^1 \tau_2 \frac{\delta_k^T \nabla^2 f(x_k + \tau_1\tau_2\delta_k)\delta_k}{\|\delta_k\|^2} \mathrm{d}\tau_1 \mathrm{d}\tau_2. \tag{4.1}$$

It is also easy to verify that there exists a finite constant $c_3 \ge 1$ such that

$$w_k(f, x_k, \delta_k) \le c_3, \quad \forall\ k, \quad \forall\ \delta_k \in \{\delta_k | x_k + \delta_k \in \mathcal{L}(x_0)\}. \tag{4.2}$$

Set

$$b_k \stackrel{\text{def}}{=} \max_{i=1,2,\cdots,k} \{\|D_k\|\} + 1, \tag{4.3}$$

$$\widehat{b}_k \stackrel{\text{def}}{=} \max_{i=1,2,\cdots,k} \{|w_i(f, x_i, \delta_i)|\} + 1. \tag{4.4}$$

Lemma 4.1 *If there exists $\varepsilon > 0$ such that*

$$\|g^k\| \ge \varepsilon \tag{4.5}$$

for all k, then there is an $\varepsilon_0 > 0$ such that

$$(c_2^2 b_k + \widehat{b}_k)\Delta_k \ge \varepsilon_0, \ \forall\ k, \tag{4.6}$$

where c_2 is given in (2.3). In fact, $\varepsilon_0 = \min\{\omega_2(1-\beta), \omega_1(1-\eta_2)\}\varepsilon\gamma_1$.

Proof We first show that there is an $\varepsilon_1 > 0$ such that if $\lambda_k = 1$ does not satisfy the condition (3.1) in step 5, then

$$\widehat{b}_k \Delta_k \ge \varepsilon_1. \tag{4.7}$$

Assume, on the contrary, that there is infinite subsequence $\mathcal{K}$ such that $\{\widehat{b}_k\Delta_k\}$ converges to zero for $k \in \mathcal{K}$. If (3.1) is not true at $\lambda_k = 1$, we have

$$f(x_k + \delta_k) > f(x_{l(k)}) + \beta(g^k)^T\delta_k \ge f(x_k) + \beta(g^k)^T\delta_k. \tag{4.8}$$

Because $f(x)$ is continuously differentiable, we have that from (4.4)

$$f(x_k + \delta_k) - f(x_k) - (g^k)^T\delta_k \le |\ f(x_k + \delta_k) - f(x_k) - (g^k)^T\delta_k\ | \le \widehat{b}_k\|\delta_k\|^2.$$

Inequality (4.8) implies that

$$(\beta - 1)(g^k)^T\delta_k \le \widehat{b}_k\|\delta_k\|^2.$$

Hence,

$$(1-\beta)(g^k)^T\delta_k + \widehat{b}_k\|\delta_k\|^2 \ge (1-\beta)(g^k)^T\delta_k > 0.$$

Note that $\{b_k\Delta_k\}$ converges to zero for $k \in \mathcal{K}$, by (4.5) and (4.7),

$$[-\omega_2\varepsilon(1-\beta) + \widehat{b}_k\Delta_k]\Delta_k \ge -\omega_2\varepsilon(1-\beta)\min\{\Delta_k, \frac{\varepsilon}{b_k}\} + \widehat{b}_k\Delta_k^2 > 0.$$

This means that, by $\Delta_k > 0$,

$$\omega_2\varepsilon(1-\beta) < \widehat{b}_k\Delta_k, \tag{4.9}$$

which contradicts that $\{\widehat{b}_k\Delta_k\}$ converges to zero for $k \in \mathcal{K}$ since $\widehat{b}_k \ge 1$. From the above we see that if (4.7) does not hold when $\varepsilon_1 = \varepsilon\omega_2(1-\beta)$, then the step size $\lambda_k = 1$, i.e., $h_k = \delta_k$ and hence $x_{k+1} = x_k + \delta_k$.

We know that

$$|f(x_k+\delta_k) - f(x_k) - \psi_k(\delta_k)| \le \frac{1}{2}\|\delta_k\|^2\|w_k(f, x_k, \delta_k) - B_k\| \le \frac{1}{2}(b_kc_2^2 + \widehat{b}_k)\Delta_k^2. \tag{4.10}$$

Set $b_k\Delta_k \le \varepsilon$. We obtain that from (2.20)

$$\text{Pred}(h_k) = -\psi_k(\delta_k) \ge \omega_1\varepsilon\min\{\Delta_k, \frac{\varepsilon}{b_k}\} \ge \omega_1\varepsilon\Delta_k. \tag{4.11}$$

Set $\rho_k = \frac{f(x_k)-f(x_k+h_k)}{\text{Pred}(h_k)}$. By (4.10) and (4.11), with

$$\Delta_k \le \frac{\omega_1\varepsilon(1-\eta_2)\gamma_1}{(b_kc_2^2 + \widehat{b}_k)}, \tag{4.12}$$

we have that

$$\begin{aligned}|\rho_k - 1| &\le \frac{|f(x_k+h_k) - f(x_k) + \text{Pred}(h_k)|}{|\text{Pred}(h_k)|}\\ &\le \frac{1}{2}\frac{(b_kc_2^2+\widehat{b}_k)\Delta_k^2}{\omega_1\varepsilon\Delta_k} \le 1-\eta_2.\end{aligned}$$

This implies that $\rho_k \ge \eta_2$. Therefore, $\widehat{\rho}_k \ge \rho_k \ge \eta_2$. By the updating rule for the trust region radius Δ_k in the step 6, we have $\Delta_{k+1} \ge \Delta_k$ when $\Delta_k \le \frac{\omega_1\varepsilon(1-\eta_2)\gamma_1}{(b_kc_2^2+\widehat{b}_k)}$, that is, $\varepsilon_1 = \omega_1\varepsilon(1-\eta_2)\gamma_1$. Hence, the conclusion of the lemma is true. In fact, from (4.7), (4.9) and (4.12), we have that $\varepsilon_0 = \min\{\omega_2(1-\beta), \omega_1(1-\eta_2)\}\gamma_1\varepsilon$. □

We are ready to state one of our main results of which the proof is similar to those in [18].

Theorem 4.2 *Let $\{x_k\} \in \Re^n$ be a sequence generated by the algorithm. Then*

$$\liminf_{k\to\infty}\|g^k\| = 0. \tag{4.14}$$

Theorem 4.2 indicates that at least one limit point of $\{x_k\}$ is a stationary point. In this section we shall first extend this theorem to a stronger result, but it requires the following assumptions.

Assumption A1. There exists $b > 0$ such that

$$b_k \leq b. \tag{4.15}$$

Assumption A2. Let $H_k = \nabla^2 f(x_k)$,

$$\lim_{k\to\infty} \frac{\|(B_k - H_k)\delta_k\|}{\|\delta_k\|} = 0. \tag{4.16}$$

Theorem 4.3 *Assume that the assumptions A1 and A2 hold and $H_* = \nabla^2 f(x_*)$ is positive definite. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Then*

$$\lim_{k\to\infty} \|g^k\| = 0. \tag{4.17}$$

Proof It is clear to see that since (2.3) holds, (4.17) holds if and only if

$$\lim_{k\to\infty} \|\widehat{g}^k\| = 0.$$

Assume that there are an $\varepsilon_1 \in (0,1)$ and a subsequence $\{\widehat{g}^{m_i}\}$ of $\{\widehat{g}^k\}$ such that for all m_i, $i = 1, 2, \cdots$,

$$\|\widehat{g}^{m_i}\| \geq \varepsilon_1. \tag{4.18}$$

Theorem 4.2 guarantees the existence of another subsequence $\{\widehat{g}^{l_i}\}$ such that

$$\|\widehat{g}^k\| \geq \varepsilon_2, \quad \text{for } m_i \leq k < l_i, \tag{4.19}$$

and

$$\|\widehat{g}^{l_i}\| \leq \varepsilon_2 \tag{4.20}$$

for an $\varepsilon_2 \in (0, \varepsilon_1)$.

We now prove that the solution $\widehat{\delta}_k$ in the three proconditional curvilinear paths satisfies

$$\tilde{q}_k(\delta_k) \stackrel{\text{def}}{=} (g^k)^T\delta_k + \delta_k^T B_k \delta_k = (\widehat{g}^k)^T\widehat{\delta}_k + (\widehat{\delta}_k)^T D_k \widehat{\delta}_k \leq 0. \tag{4.21}$$

For the preconditional optimal path, since (2.6) holds, we have that

$$\tilde{q}_k(\delta_k) = (\widehat{g}^k)^T\widehat{\delta}_k + (\widehat{\delta}_k)^T D_k \widehat{\delta}_k = -\mu_k\|\widehat{\delta}_k\|^2,$$

where $\mu_k \geq 0$. It implies that (4.21) holds.

From the definition of the preconditional modified gradient path, we have for $\tau < 1$ that

$$\begin{aligned}
\tilde{q}_k(\Gamma_1(t_1(\tau))) &= \sum_{i\in\mathcal{I}_k} \frac{\exp\{-2\varphi_i^k t_1(\tau)\} - \exp\{-\varphi_i^k t_1(\tau)\}}{2\varphi_i^k}(\widehat{g}_i^k)^2 - \frac{t_1(\tau)}{2}\sum_{i\in\mathcal{N}_k}(\widehat{g}_i^k)^2 \\
&= \sum_{i\in\mathcal{I}_k} \frac{\exp\{-\varphi_i^k t_1(\tau)\}(\exp\{-\varphi_i^k t_1(\tau)\} - 1)}{2\varphi_i^k}(\widehat{g}_i^k)^2 - \frac{t_1(\tau)}{2}\sum_{i\in\mathcal{N}_k}(\widehat{g}_i^k)^2 \\
&\leq 0,
\end{aligned}$$

since $\frac{\exp\{-\varphi_i^k t_1(\tau)\}-1}{2\varphi_i^k} \leq 0$.

If $\tau \geq 1$, then $t_1(\tau) = +\infty$, that is, $\hat{g}_i^k = 0$, $\forall i \in \mathcal{I}_k^- \bigcup \mathcal{N}_k$, the term $\Gamma_2(t_2(\tau))$ is relevant. In the case, by

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

we get that

$$\tilde{q}_k(\Gamma_1(t_1(\tau)) + \Gamma_2(t_1(\tau))) = -\sum_{i\in\mathcal{I}_k^+} \frac{1}{2\varphi_i^k}(\hat{g}_i^k)^2 \leq 0.$$

From the above inequities, we have that (4.21) holds.

From the definition of preconditional conjugate gradient path, we have that when $\sum_{j=1}^{i-1}\gamma_j < \tau \leq \sum_{j=1}^{i}\gamma_j$ $(i \leq m)$ and noting $d_1 = -g^k$,

$$\begin{aligned}\tilde{q}_k(\Gamma(\tau)) &= (g^k)^T\Gamma(\tau) + \Gamma(\tau)^T H_k \Gamma(\tau) \\ &= -\sum_{j=1}^{i-1}\gamma_j d_j^T d_1 - (\tau - \sum_{j=1}^{i-1}\gamma_j) d_i^T d_1 + \sum_{j=1}^{i-1}\gamma_j^2 d_j^T H_k d_j + (\tau - \sum_{j=1}^{i-1}\gamma_j)^2 d_i^T H_k d_i \\ &\leq 0,\end{aligned}$$

where the last inequality is deduced by $\tau - \sum_{j=1}^{i-1}\gamma_j < \gamma_i$ and (2.11)-(2.15). So, (4.21) holds.

By (4.16) and (4.21), we get

$$\begin{aligned}(g^k)^T\delta_k &\leq -(\hat{\delta}_k)^T D_k \hat{\delta}_k = -\delta_k^T B_k \delta_k = -\delta_k^T H_k \delta_k + o(\|\delta_k\|^2) \\ &\leq -\nu\|\delta_k\|^2 + o(\|\delta_k\|^2) \leq -\frac{\nu\|\hat{\delta}_k\|^2}{c_1} + o(\|\delta_k\|^2),\end{aligned} \tag{4.22}$$

where $\nu > 0$ is the minimum eigenvalue of the matrix H_* since H_* is positive definite.

Since the sequence $\{f(x_{l(k)})\}$ is nonincreasing for all k, (4.22) means that

$$f(x_{l(k)}) \leq f(x_{l(l(k)-1)}) - \lambda_{l(k)-1}\frac{\nu}{c_1}\|\hat{\delta}_{l(k)-1}\|^2 + o(\|\hat{\delta}_{l(k)-1}\|^2). \tag{4.23}$$

Similar to the proof of the Theorem 4.2 in [18], we can prove that

$$\lim_{k\to\infty}\|\hat{\delta}_k\| = 0. \tag{4.24}$$

For large enough i and $m_i \leq k < l_i$, we have

$$\begin{aligned}f(x_k + \delta_k) &= f(x_k) + (g^k)^T\delta_k + o(\|\delta_k\|) \\ &\leq f(x_{l(k)}) + \beta(g^k)^T\delta_k + (1-\beta)(g^k)^T\delta_k + o(\|\delta_k\|).\end{aligned}$$

From (4.25), for large enough i and $m_i \leq k < l_i$, we have

$$(1-\beta)(g^k)^T\delta_k + o(\|\delta_k\|) \leq 0.$$

Hence (4.25) means that the step size $\lambda_k = 1$, i.e., $h_k = \delta_k$ for large enough i and $m_i \le k < l_i$.

By the assumption A2, we know that

$$\begin{aligned}
&|f(x_k+\delta_k) - f(x_k) - \psi_k(\delta_k)| \\
&= |[(g^k)^T\delta_k + \frac{1}{2}\delta_k^T H_k\delta_k + o(\|\delta_k\|^2)] - [(g^k)^T\delta_k + \frac{1}{2}\delta_k^T B_k\delta_k]| \\
&= o(\|\delta_k\|^2).
\end{aligned}$$

From (2.8), similar to the proof of (4.22), for large enough i, $m_i \le k < l_i$,

$$\mathrm{Pred}(\delta_k) = -(\widehat{g}^k)^T\widehat{\delta}_k - \frac{1}{2}(\widehat{\delta}_k)^T D_k\widehat{\delta}_k \ge \frac{1}{2}(\widehat{\delta}_k)^T D_k\widehat{\delta}_k \ge \frac{\nu\|\widehat{\delta}_k\|^2}{2c_1} + o(\|\widehat{\delta}_k\|^2). \tag{4.27}$$

As $h_k = \delta_k$, for large $i, m_i \le k < l_i$, we obtain that

$$\begin{aligned}
\widehat{\rho}_k \ge \rho_k &= \frac{f_k - f(x_k+h_k)}{\mathrm{Pred}(h_k)} = 1 + \frac{f_k - f(x_k+\delta_k) + \psi_k(\delta_k)}{\mathrm{Pred}(h_k)} \\
&\ge 1 - \frac{2c_1\ o(\|\widehat{\delta}_k\|^2)}{\|\widehat{\delta}_k\|^2 + o(\|\widehat{\delta}_k\|^2)} \ge \eta_2.
\end{aligned} \tag{4.28}$$

This means that for large i, $m_i \le k < l_i$,

$$f_k - f(x_k+h_k) \ge \eta_2\mathrm{Pred}(h_k) \ge \eta_2\frac{\nu\|h_k\|^2}{4c_1}.$$

Therefore, we can deduce that, for large i,

$$\|x_{m_i} - x_{l_i}\|^2 \le \frac{4c_1}{\nu\eta_2}\sum_{k=m_i}^{l_i-1}[f(x_k) - f(x_k+h_k)] = \frac{4c_1}{\nu\eta_2}(f_{m_i} - f_{l_i}). \tag{4.29}$$

Inequality (4.29) means that $f_{m_i} - f_{l_i}$ tends to zero as i tends to infinity and hence $\|x_{m_i} - x_{l_i}\|$ tends to zero as i tends to infinity. By the continuity of the gradients $g(x)$ and $\widehat{g}(x)$, we thus deduce that $\|\widehat{g}^{m_i} - \widehat{g}^{l_i}\|$ also tends to zero as i tends to infinity. However, this is impossible because of the definitions of $\{m_i\}$ and $\{l_i\}$, which imply that the triangle inequality to show

$$\|\widehat{g}^{m_i} - \widehat{g}^{l_i}\| \ge \|\widehat{g}^{m_i}\| - \|\widehat{g}^{l_i}\| \ge \varepsilon_1 - \varepsilon_2, \tag{4.30}$$

Therefore (4.30) contradicts $\|\widehat{g}^{m_i} - \widehat{g}^{l_i}\| \to 0$ as $i \to \infty$. This implies that (4.18) is not true, and hence the conclusion of the theorem holds. □

Theorem 4.4 *Assume that Assumptions A1 and A2 hold. If the matrices D_k satisfy the following condition*

$$\varphi_1[D_k] \le \tau_1\varphi_1[\nabla^2 f(x_k)], \text{ when } \varphi_1[\nabla^2 f(x_k)] < 0, \tag{4.31}$$

where τ_1 is some positive constant and $\varphi_1[B]$ is the minimum eigenvalue of the symmetric matrix B, then $\nabla^2 f(x_*)$ is positive semi-definite where x_* is a limit point of $\{x_k\}$ (second *order stationary point convergence).*

Proof Because $f(x)$ is twice continuously differentiable, we have that, noting (4.16),

$$\begin{aligned} f(x_k+\delta_k) =& f(x_k)+(g^k)^T\delta_k+\frac{1}{2}\delta_k^T H_k\delta_k+o(\|\delta_k\|^2) \\ \le& f(x_{l(k)})+\beta(g^k)^T\delta_k+(\frac{1}{2}-\beta)(g^k)^T\delta_k+ \\ &\frac{1}{2}[(g^k)^T\delta_k+\delta_k^T B_k\delta_k]+\frac{1}{2}\delta_k^T(H_k-B_k)\delta_k+o(\|\delta_k\|^2) \\ \le& f(x_{l(k)})+\beta(g^k)^T\delta_k-(\frac{1}{2}-\beta)\widehat{\omega}_2\|\delta_k\|^2+o(\|\delta_k\|^2) \\ \le& f(x_{l(k)})+\beta(g^k)^T\delta_k, \end{aligned} \tag{4.32}$$

where the last two inequalities hold because of $\beta<\frac{1}{2}$ and (4.21). By the above inequality, we know that

$$x_{k+1}=x_k+\delta_k,$$

which implies that for large enough k, the step size $\lambda_k=1$, i.e., $h_k=\delta_k$.

Suppose to the contrary that $\varphi_1[\nabla^2 f(x_*)]\le -2\varepsilon$ for some positive constant ε. By the condition of the theorem, there exists K such that if $k\ge K$, $\varphi_1[D_k]\le \tau_1\varphi_1[\nabla^2 f(x_k)]\le -\tau_1\varepsilon$. Similar to those in [2], we can also obtain that there exists $\tau>0$ such that for large enough k,

$$-\psi(\delta_k)=-[(\widehat{g}^k)^T\widehat{\delta}_k+\frac{1}{2}(\widehat{\delta}_k)^T D_k\widehat{\delta}_k]\ge -\frac{\tau}{2}\varphi_1[D_k]\|\widehat{\delta}_k\|^2.$$

Hence,

$$-\psi(\delta_k)\ge -\frac{\tau}{2}\varphi_1[D_k]\|\widehat{\delta}_k\|^2\ge \frac{\tau_1\tau\varepsilon}{2}\|\widehat{\delta}_k\|^2. \tag{4.33}$$

Since D_k is not positive definite, the Newton method or the quasi-Newton step is never taken for $k\ge K$. From Lemma 2.2, Lemma 2.3 and Lemma 2.4, we get that the step $\widehat{\delta}_k$ obtained from the preconditional optimal path, the preconditional modified gradient path or the preconditional conjugate gradient path satisfies $\|\widehat{\delta}_k\|=\Delta_k$.

By the assumption A2 and (4.33), we can obtain

$$\begin{aligned} |\rho_k-1| =& \left|\frac{[(g^k)^T h_k+\frac{1}{2}h_k^T B_k h_k]-[(g^k)^T h_k+\frac{1}{2}h_k^T H_k h_k+o(\|h_k\|^2)]}{\mathrm{Pred}(h_k)}\right| \\ =& \left|\frac{o(\|h_k\|^2)}{\mathrm{Pred}(h_k)}\right|\le \frac{2|o(\|\widehat{\delta}_k\|^2)|}{\tau_1\tau\varepsilon\|\widehat{\delta}_k\|^2}\ \to\ 0. \end{aligned}$$

This, in turn, implies that for sufficiently large $k\ge K$ and sufficiently small Δ_k, Δ_k cannot be decreased further. Thus, the updating rules for the trust region radius then prevent Δ_k from tending to zero. But from (4.33) and $\|\widehat{\delta}_k\|=\Delta_k$, we get

$$f_k-f(x_k+h_k)\ge \eta_2\mathrm{Pred}(h_k)\ge \frac{1}{2}\eta_2\tau_1\tau\varepsilon\Delta_k^2,$$

and since f is bounded below and $f_k-f(x_k+h_k)$ converges to 0, we know Δ_k converges 0, which is a contradiction. Hence $\varphi_1[\nabla^2 f(x_*)]\ge 0$. This concludes the proof of the theorem.

We now discuss the convergence rate for the algorithm when B_k is positive definite.

Theorem 4.5 *If B_k is eventually positive definite and Assumptions A1 and A2 hold, then $\{x_k\}$ converges to x^* superlinearly:*

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

Proof We can prove that (4.32) also holds for large enough k, then the step size $\lambda_k = 1$, i.e., $h_k = \delta_k$.

Similar to the proof of (4.27), we can prove that (4.27) also holds for large enough k. By the assumption A2, we can obtain

$$\begin{aligned}\rho_k - 1 &= \frac{[(g^k)^T h_k + \frac{1}{2}h_k^T B_k h_k] - [(g^k)^T h_k + \frac{1}{2}h_k^T H_k h_k + o(\|h_k\|^2)]}{\mathrm{Pred}(h_k)} \\ &= \frac{o(\|h_k\|^2)}{\mathrm{Pred}(h_k)} \to 0. \end{aligned} \tag{4.34}$$

Conclusions (4.27) and (4.34) mean that when $\|\delta_k\| \to 0$, $\rho_k \to 1$. Hence there exists $\widehat{\Delta} > 0$ such that when $\|\delta_k\| \le \widehat{\Delta}$, $\rho_k \ge \eta_2$, and therefore, $\Delta_{k+1} \ge \Delta_k$. As $h_k \to 0$, there exists an index K' such that $\|\widehat{\delta}_k\| \le \widehat{\Delta}$ whenever $k \ge K'$. Thus $\Delta_k \ge \Delta_{K'}$, $\forall k \ge K'$.

On the other hand, as $\widehat{g}^k \to \widehat{g}^* = 0$, we can obtain that Assumption A8 in [2] for the preconditional conjugate gradient path, the preconditional optimal path or the preconditional modified gradient path holds, which ensures

$$\widehat{\delta}_k = \lim_{\tau\to+\infty} \Gamma_k = -D_k^{-1}\widehat{g}^k.$$

The step size $\lambda_k = 1$ for large enough k means that

$$h_k = \delta_k = -P_k^T L_k^{-T} D_k^{-1} L_k^{-1} P_k g^k = -B_k^{-1} g^k.$$

Therefore, the algorithm becomes the Newton method or the quasi-Newton method. As in this case assumption A2 is a sufficient condition for superlinear convergence, the theorem is proved. □

5. Numerical experiments

Numerical experiments on the preconditional optimal path and preconditional modified gradient path algorithms with the nonmonotonic back tracking technique given in this paper have been performed on an IBM 586 personal computer. In this section we present the numerical results. We compare with different nonmonotonic parameters $M = 0$, $M = 4$ and $M = 8$, respectively, for the proposed algorithms. The monotonic algorithms are realized by taking $M = 0$. In order to check effectiveness of the back tracking technique, we select the same parameters as used in [7]. The selected parameter values are: $\widehat{\eta} = 0.01$, $\eta_1 = 0.001$, $\eta_2 = 0.75$, $\gamma_1 = 0.5$, $\gamma_2 = 2$, $\Delta_{\max} = 10$, and initially $\Delta_0 = 1$. The computation terminates when one of the following stopping criterions is satisfied: $\|g^k\| \le 10^{-6}$, or $f_k - f_{k+1} \le 10^{-8}\max\{1, |f_k|\}$.

The experiments are carried out on 15 standard test problems which are quoted from [14]. Besides the recommended starting points in [14], denoted by x_{0a}, we also test these methods with another set of starting points x_{0b}. The computational results for $B_k = H_k$, the real Hessian, are presented at the following table, where POPPATH and PMGPATH denote respectively the preconditional optimal path algorithm and the preconditional modified gradient path algorthm proposed in this paper with nonmonotonic decreasing and back tracking techniques. NF and NG stand for the numbers of function evaluations and gradient evaluations respectively. NO stands for the number of iterations in which non-monotonic decreasing situation occurs, that is, the number of times $f_k - f_{k+1} < 0$. The number of iterations is not presented in the following table because it always equals NG.

Problem	Initial	POPPATH							
		M=0		M=4			M=8		
Name	Point	NF	NG	NF	NG	NO	NF	NG	NO
Rosenbrock	x_{0a}	23	19	16	14	5	13	12	4
(C=100)	x_{0b}	12	10	7	7	1	7	7	1
Rosenbrock	x_{0a}	76	54	16	16	5	16	14	5
(C=10000)	x_{0b}	28	23	8	8	1	8	8	1
Rosenbrock	x_{0a}	215	199	27	25	4	18	16	5
(C=1000000)	x_{0b}	61	43	15	15	4	15	15	4
Freudenstein	x_{0a}	6	6	6	6	0	6	6	0
	x_{0b}	12	10	11	11	1	11	11	1
Cube	x_{0a}	30	23	9	9	2	9	9	2
	x_{0b}	21	18	11	11	3	11	11	3
Box	x_{0a}	17	17	17	17	0	17	17	0
	x_{0b}	13	11	15	15	1	15	15	1
Engvall	x_{0a}	17	16	17	16	0	17	16	0
	x_{0b}	20	18	19	19	1	19	19	1
Wood	x_{0a}	56	39	54	35	5	28	28	5
	x_{0b}	13	12	14	14	1	14	14	1
Powell	x_0	16	16	16	16	0	16	16	0
Davidon	x_0	11	11	12	12	1	12	12	1
Osborne	x_0	13	13	13	13	0	13	13	0
Biggs	x_{0a}	43	18	53	42	5	51	41	6
	x_{0b}	60	33	179	102	29	183	94	16
Banana	x_{0a}	27	20	19	18	1	16	16	2
(n=6)	x_{0b}	32	26	24	23	3	23	23	4
Banana	x_{0a}	34	27	21	21	2	21	21	2
(n=10)	x_{0b}	41	36	33	33	4	32	32	4
Banana	x_{0a}	45	35	45	35	0	45	35	0
(n=16)	x_{0b}	32	30	32	30	0	32	30	0

Table 1: Experimental Results of the Preconditional Optimal Path Algorithm

The results under POPPATH and PMGPATH represent mixture of trust region and line search techniques considered in this paper. Our preconditional curvilinear type of approximate trust region method is very easy to resolve the subproblem (S_k) with a reduced radius. Indeed, the formulation of curvilinear paths Γ_k do not depend on the value of Δ_k, so that when the trust region is contracted and $\widehat{\delta}_k$ is outside the new region, we only need to set the point back tracking along the same path until reaching the new boundary. The back tracking can outperform the traditional method when the trust region subproblem is solved accurately over the whole hyperball.

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.

Problem	Initial	PMGPATH							
		M=0		M=4			M=8		
Name	Point	NF	NG	NF	NG	NO	NF	NG	NO
Rosenbrock	x_{0a}	23	19	16	14	5	13	12	4
(C=100)	x_{0b}	13	11	7	7	1	7	7	1
Rosenbrock	x_{0a}	82	48	16	16	5	16	14	5
(C=10000)	x_{0b}	28	24	8	8	1	8	8	1
Rosenbrock	x_{0a}	223	201	28	26	4	16	14	5
(C=1000000)	x_{0b}	62	43	15	15	4	15	15	4
Freudenstein	x_{0a}	6	6	6	6	0	6	6	0
	x_{0b}	12	10	11	11	1	11	11	1
Cube	x_{0a}	30	23	9	9	2	9	9	2
	x_{0b}	21	18	11	11	3	11	11	3
Box	x_{0a}	17	17	17	17	0	17	17	0
	x_{0b}	13	11	15	15	1	15	15	1
Engvall	x_{0a}	17	16	17	16	0	17	16	0
	x_{0b}	20	18	19	19	1	19	19	1
Wood	x_{0a}	56	39	54	35	5	28	28	5
	x_{0b}	13	12	14	14	1	14	14	1
Powell	x_0	16	16	16	16	0	16	16	0
Davidon	x_0	11	11	12	12	1	12	12	1
Osborne	x_0	13	13	13	13	0	13	13	0
Biggs	x_{0a}	40	18	51	33	5	54	38	8
	x_{0b}	67	36	181	83	32	172	63	15
Banana	x_{0a}	27	20	19	18	1	16	16	2
(n=6)	x_{0b}	32	26	24	23	3	23	23	4
Banana	x_{0a}	30	24	21	21	2	21	21	2
(n=10)	x_{0b}	41	36	33	33	4	32	32	4
Banana	x_{0a}	45	35	45	35	0	45	35	0
(n=16)	x_{0b}	32	30	32	30	0	32	30	0

Table 2: Experimental Results of the Preconditional Modified Gradient Path Agorithm

Acknowledgements: The author would like to thank the referees for their detailed comments and helpful suggestions.

References:

[1] BUNCH J R, PARLETT B N. *Direct method for solving symmetric indefinite systems of linear equations* [J], SIAM Journal on Numerical Analysis, 1971, **8**: 639–655.

[2] BULTEAU J P, VIAL J Ph. *Curvilinear path and trust region in unconstrained optimization, a convergence analysis* [J]. Mathematical Programming Study, 1987, **30**: 82–101.

[3] BYRD R H, SCHNABEL R B, SHULTZ G A. *Approximate solution of the trust region problem by minimization over two-dimensional subspaces* [J]. Mathematical Programming, 1988, **40**: 247–263.

[4] DENG N Y, XIAO Y, ZHOU F J. *A nonmonotonic trust region algorithm* [J]. Journal of Optimization Theory and Applications, 1993, **76**: 259–285.

[5] DENNIS J E Jr, ECHEBEST N, GUARDARUCCI M T, et al. *A curvilinear search using tridiagonal secant updates for unconstrained optimization* [J]. SIAM Journal on Optimization, 1991, **1**: 333–357.

[6] DENNIS J E, MORÉ J J. *A characterization of superlinear convergence and its application to quasi-Newton methods* [J]. Math. Comp., 1974, **28**: 549–560.

[7] DENNIS J E Jr, SCHNABLE R B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* [M]. Prentice Hall, New Jersey, 1983.

[8] FLETCHER R. *Practical Methods of Optimization, Unconstrained Optimization* [M]. John Wiley & Sons New, York, 1980.

[9] GRIPPO L, LAMPARIELLO F, LUCIDI S. *A nonmonotonic line search technique for Newton's methods* [J]. SIAM Journal on Numerical Analysis, 1986, **23**: 707–716.

[10] MORÉ J J, SORENSEN D C. *Computing a trust region step* [J]. SIAM Journal on Science and Statistical Computing, 1983, **4**: 553–572.

[11] NOCEDAL J, YUAN Y. *Combining trust region and line search techniques* [J]. in Y.Yuan ed. Advances in Nonlinear Programming (Kluwer, Dordvecht), 1998, 153–175.

[12] POWELL M J D. *A hybrid method for nonlinear equations* [J]. In Numerical Methods for Nonlinear Algebraic Equations, Edited by Ph. Rabonowitz, Gordon and Breach, 1970, 87–114.

[13] POWELL M J D. *On the global convergence of trust region algorithms for unconstrained minimization* [J]. Mathematical Programming, 1984, **29**: 297–303.

[14] SCHITTKOWSKI K. *More Test Examples for Nonlinear Programming Codes* [M]. Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1987, 29.

[15] SORENSEN D C. *Newton's method with a model trust region modification* [J]. SIAM J. Numer. Anal., 1982, **19**: 409–426.

[16] STEHAUG T. *The conjugent method and trust regions in large scale optimization* [J]. SIAM J. Numer. Anal., 1983, **20**: 626–637.

[17] ZHANG J Z, XU C X. *A scaled optmal path trust region alogirthm* [J]. J. Optim. Theory Appl., 1999, **102**: 127–146.

[18] ZHU D. Curvilinear paths and trust region methods with nonmonotonic back tracking technique for unconstrained optimization [J]. J. Computational Mathematics, 2001, **19**: 241–258.

[19] ZHU D. *Nonmonotonic preconditional curvilinear path algorithms for unconstrained optimization* [J]. A J. of Chinese Universities, Numerical Mathematics, 2003, **12**: 99–210.

非单调技术预条件弧线路径信赖域解无约束优化

朱 德 通

(上海师范大学数学系， 上海 200234)

摘 要: 本文提供修正近似信赖域类型路经三类预条件弧线路径方法解无约束最优化问题. 使用对称矩阵的稳定 Bunch-Parlett 易于形成信赖域子问题的弧线路径，使用单位下三角矩阵作为最优路径和修正梯度路径的预条件因子. 运用预条件因子改进 Hessian 矩阵特征值分布加速预条件共轭梯度路径收敛速度. 基于沿着三类路径信赖域子问题产生试探步，将信赖域策略与非单调线搜索技术相结合作为新的回代步. 理论分析证明在合理条件下所提供的算法是整体收敛性，并且具有局部超线性收敛速率，数值结果表明算法的有效性.

关键词: 弧线路径；预条件因子；信赖域；非单调技术.