Journal of Computational Mathematics, Vol.19, No.3, 2001, 241–258.

CURVILINEAR PATHS AND TRUST REGION METHODS WITH NONMONOTONIC BACK TRACKING TECHNIQUE FOR UNCONSTRAINED OPTIMIZATION*[1)]

De-tong Zhu
(*Department of Mathematics, Shanghai Normal University, Shanghai 200234, China*)

Abstract

In this paper we modify type approximate trust region methods via two curvilinear paths for unconstrained optimization. A mixed strategy using both trust region and line search techniques is adopted which switches to back tracking steps when a trial step produced by the trust region subproblem is unacceptable. We give a series of properties of both optimal path and modified gradient path. The global convergence and fast local convergence rate of the proposed algorithms are established under some reasonable conditions. A nonmonotonic criterion is used to speed up the convergence progress in some ill-conditioned cases.

Key words: Curvilinear paths, Trust region methods, Nonmonotonic technique, Unconstrained optimization.

1. Introduction

Trust region method is a well-accepted technique in nonlinear optimization to assure global convergence. One of the advantages of the model is that it does not require the objective function to be convex. Many different versions have been suggested in using trust region technique. For each iteration, suppose a current iterate point, a local quadratic model of the function and a trust region with center at the point and a certain radius are given. A point that minimizes the model function within the trust region is solved as a trial point. If the actual reduction achieved on the function f at this point is satisfactory comparing with the reduction predicted by the quadratic model, the point is accepted as a new iterate, then the trust region radius is adjusted and the procedure is repeated. Otherwise, the trust region radius should be reduced and a new trial point needs to be determined. Recently Bulteau and Vial proposed in [1] curvilinear paths with trust region method for unconstrained optimization and a main feature of which is instead of minimizing a quadratic function within the whole trust region which is a hyperball, it only minimizes the function over a simple curvilinear path inside the trust region. In other words, their method is an approximate trust region method via curvilinear paths.

It is also noticed that Nocedal and Yuan [9] suggested a combination of the trust region and line search method. The motivation is intuitive. As we know, in traditional trust region method, after solving a subproblem we need to use some criterion to check if the trial step is acceptable. If not, the subproblem must be resolved with a reduced trust region radius. It is possible that the trust region subproblem needs to be resolved many times before obtaining an acceptable solution, and hence the total computation for completing one iteration might be expensive. A plausible remedy is that at an unsuccessful trial step we switch to the line search

* Received March 12, 1998.
[1)]This project is supported by the Chinese National Science Foundation Grant 10071050, the Science and Technology Foundation of Shanghai Higher Education.

technique by employing the back tracking steps. Of course the prerequisite for being able to making this shift is that although the trial step is unacceptable as next iterative point, it should provide a direction of sufficient descent.

Another valuable idea is to abandon the traditional monotonic decreasing requirement for the sequence $\{f(x_k)\}$ of the objective values (see [3] and [7]), because monotonicity may cause a series of very small steps if the contours of objective function f are a family of curves with large curvature.

The main purpose of this paper is to modify and improve the curvilinear path type approximate trust region method by adopting the above ideas: back tracking and nonmonotonic search. In particular, we shall show that the trial step generated by their subproblem produces a sufficiently descent direction. We shall focus on unconstrained optimization

$$\min_{x\in\Re^n} f(x).$$

Both theoretical analysis and numerical experiment will be undertaken for the improved algorithms.

The paper is organized as follows. In section 2, we give expressions to the curvilinear path model trust region steps and propose the characterizations and properties of the curvilinear paths quadratic subproblem. In Section 3, we describe the algorithm which combines the techniques of trust region, back tracking and nonmonotonic search. In section 4, weak global convergence of the proposed algorithm is established. Some further convergence properties such as strong global convergence and superlinear convergence rate are discussed in section 5. Finally, the results of numerical experiments are reported in section 6.

2. Curvilinear Paths

In trust region algorithms, an important portion of the unconstrained minimization procedure will be concerned with the solution of a subproblem of the form

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

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, $q_k(\delta)$ is the local quadratic approximation of f and Δ_k is the trust region radius and $\|\cdot\|$ throughout is 2-norm. Let δ_k be the solution of the subproblem. Then set next step

$$x_{k+1} = x_k + \delta_k. \tag{2.2}$$

Based on solving the about trust region subproblem, we given the following Lemmas which are due to Sorensen paper in [13].

Lemma 2.1. *s is a solution to the subproblem (2.1) in which x_k is given by x if and only if s is a solution to the following equations of the forms*

$$(B + \mu I)\ s = -g(x) \tag{2.3}$$

$$\mu(\|s\|^2 - \Delta^2) = 0, \quad \mu \ge 0. \tag{2.4}$$

and $B + \mu I$ is positive semidefinite.

Lemma 2.1 establishes the necessary and sufficient conditions concerning the pair $\mu,\ s$ when s solves (2.1). The next results are immediate consequences of lemma 2.1.

Lemma 2.2. *Let $\mu \in \Re^1,\ s \in \Re^n$ be a solution to the following equations of the forms*

$$(B + \mu I)\ s = -g(x) \tag{2.5}$$

and $B + \mu I$ is positive semidefinite. Then we have that
(1) *if $\mu = 0$ and $\|\delta\| \le \Delta$ then s solves the subproblem (2.1);*
(2) *if $\|\delta\| = \Delta$ then s solves*

$$\min q_x(\delta) \quad \text{subject to} \quad \|\delta\| = \Delta;$$

(3) *if* $\mu \geq 0$ *and* $\|\delta\| = \Delta$ *then* s *solves the subproblem (2.1).*
Further, if, in fact, $B + \mu I$ *is positive definite. then* s *is unique in each of the cases (1), (2) and (3).*

Now, we form the two paths as Bulteau and Vial (see [1]), i.e., optimal path and modified gradient path, respectively.

When the trust region radius Δ_k of problem (2.1) varies in the interval $[0, +\infty)$, the solution points form the scaled paths and emanate from the origin. In order to definte those arcs in a closed form, we shall use the eigensystem decomposition of B. Since B is sysmetric matrix, its eigenvalues $\phi_1, \phi_2, \cdots, \phi_n$ are real number and these are corresponding orthonormal eigenvectors $u^1,\ u^2, \cdots, u^n$. Without lost of generality, let $\phi_1 \leq \phi_2 \leq \cdots \leq \phi_n$ be eigenvalues of B 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 $\phi_i > 0,\ \phi_i < 0$ and $\phi_i = 0$ for $i \in \{1, \ldots, n\}$, respectively. We now give two curvilinear paths.

2.1. Optimal path

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

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

where

$$\begin{aligned} \Gamma_1(t_1(\tau)) &= -\,[\sum_{i\in\mathcal{I}} \frac{t_1(\tau)}{\phi_i t_1(\tau) + 1} g_i^k u^i + t_1(\tau) \sum_{i\in\mathcal{N}} g_i^k u^i\ ], \\ \Gamma_2(t_2(\tau)) &= t_2(\tau) u^1, \end{aligned}$$

and

$$t_1(\tau) = \begin{cases} \tau, & \text{if } \ \tau < \frac{1}{T}, \\ \frac{1}{T}, & \text{if } \ \tau \geq \frac{1}{T}, \end{cases}$$

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

$\mathcal{I} = \{\ i \mid \phi_i \neq 0,\ i = 1, \cdots, n\}$, $\mathcal{N} = \{\ i \mid \phi_i = 0,\ i = 1, \cdots, n\}$, $g_i^k = (g^k)^T u^i,\ i = 1, \cdots, n,\ \ g^k = \sum_{i=1}^n g_i^k u^i$, $T = \max\{0, -\phi_1\}$ and $1/T$ is defined as $+\infty$ if $T = 0$. It should be noted that $\Gamma_2(t_2(\tau))$ is defined only when B is indefinite and $g_i^k = 0$ for all $i \in \{1, \cdots, n\}$ with $\phi_i = \phi_1 < 0$ which is referred to **hard case** (see [8]) for unconstrained optimization and that for other cases, $\Gamma(\tau)$ is defined only for $0 \leq \tau < \frac{1}{T}$, that is, $\Gamma(\tau) = \Gamma_1(t_1(\tau))$.

2.2. Modified gradient path

The modified gradient path can be given in the following closed form which can refer to see [1]:

$$\Gamma(\tau) = \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)), \quad \tau \in [0, +\infty) \tag{2.7}$$

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

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

with

$$\Gamma_2(t_2) = \begin{cases} t_2 u^1, & \text{if } \ \phi_1 < 0, \\ 0, & \text{if } \ \phi_1 \geq 0, \end{cases}$$

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

and

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

2.3. Properties of the curvilinear paths

It is well known from solving the trust region algorithms in order to the global convergence of the proposed algorithm, it is a sufficient condition to show that at k-th iteration the predicted reduction defined by

$$\mathrm{Pred}(\delta_k) = f_k - q_k(\delta_k)$$

which is obtained by the step δ_k from the curvilinear paths in trust region, satisfies the following sufficient descent condition

$$\mathrm{Pred}(\delta_k) \geq \widehat{\omega}\|g^k\| \min\{\Delta_k, \frac{\|g^k\|}{\|B_k\|}\} \tag{2.8}$$

for all g^k, B_k and Δ_k, where $\widehat{\omega} > 0$ is a constant independent k. We can obtain this result in each of the above two paths given by Bulteau and Vial (see [1]). In order to discuss the properties in detail, we will summarize as follows.

Lemma 2.3. *Let the step δ_k in trust region be obtained from the optimal path. Then we have that the norm function of the path is monotonically increasing for $\tau \in (0, +\infty)$, and there exist τ^* such that the point $\Gamma(\tau^*)$ on the path with*

$$\|\Gamma(\tau^*)\| = \Delta_k$$

satisfies the following system

$$(B_k + \mu_k I)\Gamma(\tau^*) = -g^k, \tag{2.9}$$

where $\mu_k \geq 0$ given as follows and

$$\mu_k = 1/t_1(\tau^*) \quad \text{as } \tau^* < 1/T \tag{2.10}$$

$$\mu_k = \frac{1}{T}, \quad t_2(\tau^*) = \tau^* - \frac{1}{T} \quad \text{as } \tau^* \geq 1/T \tag{2.11}$$

where $T = \max\{0, -\phi_1\}$. Futhermore, the predicted reduction $\mathrm{Pred}(\delta_k)$ satisfies the sufficient descent condition (2.8).

Proof. Let the step δ_k be obtained from the optimal path. It is obvious that $\Gamma(\tau)$ is a continuous path, since $u^1,\ u^2, \cdots, u^n$ are orthonormal eigenvectors.

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

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

Let

$$\psi(\tau) = \|\Gamma_1(\tau)\|^2 = \sum_{i\in\mathcal{I}} (\frac{\tau}{\phi_i\tau + 1})^2 (g_i^k)^2 + \tau^2 \sum_{i\in\mathcal{N}} (g_i^k)^2.$$

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

$$\psi'(\tau) = \sum_{i\in\mathcal{I}} \frac{2\tau}{(\phi_i\tau + 1)^3}(g_i^k)^2 + 2\tau \sum_{i\in\mathcal{N}} (g_i^k)^2 > 0.$$

Thus, $\|\Gamma_1(\tau)\|$ is monotonically increasing for $0 < \tau < \frac{1}{T}$. Since $\|\Gamma_2(t_2(\tau))\|^2 = (\tau - \frac{1}{T})^2$, $\|\Gamma_2(t_2(\tau))\|$ is certainly increasing for $\tau \geq \frac{1}{T}$. As above we have that the norm function of the path is monotonically increasing for $\tau \in (0, +\infty)$.

This monotonically increasing property ensures that for any given $\Delta_k \in (0, +\infty)$, when $\|\Gamma(\tau)\| \geq \Delta_k$ there is a unique point on the path $\Gamma(\tau)$, $\Gamma(\tau^*)$ say, such that

$$\|\Gamma(\tau^*)\| = \Delta_k \tag{2.12}$$

and τ^* can be uniquely determined by the equation

$$\|\Gamma(\tau)\| = \Delta_k. \tag{2.13}$$

If B_k is positive definite or positive semi-definite, then the path is finite and $\Gamma(\tau^*) = \Gamma_1(t_1(\tau^*))$

and $\tau^* \in (0, +\infty)$ as $T = 0$. Since

$$\begin{aligned}&(B_k + \mu_k I)\Gamma_1(t_1(\tau^*))\\ =\;& -\sum_{i=1}^{n}(\phi_i + \mu_k)u^i {u^i}^T[\sum_{i\in\mathcal{I}} \frac{t_1(\tau^*)}{\phi_i t_1(\tau^*) + 1} g_i^k u^i + t_1(\tau^*)\sum_{i\in\mathcal{N}} g_i^k u^i]\\ =\;& -[\sum_{i\in\mathcal{I}}(\phi_i + \mu_k)\frac{t_1(\tau^*)}{\phi_i t_1(\tau^*) + 1} g_i^k u^i + \mu t_1(\tau^*)\sum_{i\in\mathcal{N}} g_i^k u^i],\end{aligned} \tag{2.14}$$

it is clear that system

$$(B_k + \mu_k I)\Gamma(\tau^*) = -g^k$$

is satisfied with $\mu_k = 1/t_1(\tau^*)$. If B_k is indefinite, then the path is infinite. If $\tau^* < 1/T$, then $\Gamma(\tau^*) = \Gamma_1(t_1(\tau^*))$ and it follows from (2.14) that system

$$(B_k + \mu_k I)\Gamma(\tau^*) = -g^k$$

is also satisfied with $\mu_k = 1/t_1(\tau^*)$. Note that in this case $B_k + \mu_k I$ is positive definite, since $\mu_k > T = -\phi_1$. If $\tau^* \geq 1/T$ then $\Gamma(\tau^*) = \Gamma(\frac{1}{T}) + \Gamma(t_2(\tau^*))$.

The above property ensure that,

$$(B_k + \mu_k I)\Gamma(\tau^*) \quad = \quad -g^k \tag{2.15}$$

which means that $\|\Gamma(\tau^*)\| = \Delta_k$ and (2.9) hold. μ_k is given by

$$\mu_k = \begin{cases} \frac{1}{\tau^*}, & \text{if} \quad \tau < \frac{1}{T}, \\ \frac{1}{T}, & \text{if} \quad \tau \geq \frac{1}{T}. \end{cases} \tag{2.16}$$

By Lemma 2.1 establishing the necessary and sufficient conditions solving subproblem (2.1) and Powell in [10] establishing along the predicted reduction which is the sufficient descent direction, we have that the conclusion of the theorem holds.

Lemma 2.4. *Let the step δ_k in trust region be obtained from the modified gradient path. Then we have that the norm function of the path is monotonically increasing for $\tau \in (0, +\infty)$, and the predicted reduction* $\text{Pred}(\delta_k)$ *satisfies the sufficient descent condition (2.8).*

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

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

Let

$$\psi_1(\tau) = \|\Gamma_1(\tau)\|^2 = \sum_{i\in\mathcal{I}}(\frac{\exp\{-\phi_i t_1(\tau)\} - 1}{\phi_i})^2 (g_i^k)^2 + \tau^2 \sum_{i\in\mathcal{N}} (g_i^k)^2.$$

Then using the fact $1 - \tau > 0$, we have

$$\psi_1'(\tau) = \sum_{i\in\mathcal{I}} \frac{2\exp\{-\phi_i t_1(\tau)\}}{(1-\tau)^2}(\frac{1 - \exp\{-\phi_i t_1(\tau)\}}{\phi_i})(g_i^k)^2 + 2\tau \sum_{i\in\mathcal{N}} (g_i^k)^2 > 0.$$

Thus, $\|\Gamma_1(\tau)\|$ is monotonically increasing for $\tau < 1$. Since $\|\Gamma_2(t_2(\tau))\|^2 = (\tau - 1)^2$, $\|\Gamma_2(t_2(\tau))\|$ is certainly increasing for $\tau \geq 1$. As above we have that the norm function of the modified gradient path is monotonically increasing for $\tau \in (0, +\infty)$.

By lemma 6.1 of Buliteau and Vial in [1], we have that the predicted reduction satisfies the sufficient descent condition (2.8).

The following lemmas show the relation between the gradient g^k of the objective function and the step δ_k generated from the two paths, i.e., optimal path and modified gradient path, respectively. We can see from the lemmas that the direction of the trial step generated in the two paths is a sufficiently descent direction.

Lemma 2.5. *At the k-th iteration, let δ_k be generated from the optimal path in trust region subproblem then*

$$(g^k)^T \delta_k \leq -\omega_1 \|g^k\| \min\{\ \Delta_k,\ \frac{\|g^k\|}{\|B_k\|}\ \} \tag{2.17}$$

where $\omega_1 > 0$ is a constant.

Proof. Since δ_k is generated from the optimal path, it ensures that,

$$(B_k + \mu_k I)\delta_k = -g^k. \tag{2.18}$$

From

$$\mu_k \delta_k = -g^k - B_k \delta_k,$$

we take norm in the above equation and obtain

$$\mu_k \|\delta_k\| \le \|g^k\| + \|B_k\|\|\delta_k\|. \tag{2.19}$$

And note $\|\delta_k\| = \|\Gamma(\tau_k)\| = \Delta_k$,

$$0 \le \mu_k \le \frac{\|g^k\|}{\|\delta_k\|} + \|B_k\| \le \frac{\|g^k\|}{\Delta_k} + \|B_k\|. \tag{2.20}$$

From (2.10) and $0 < \tau_k < \frac{1}{T}$, we have that from (2.11),

$$\begin{aligned} \widehat{\psi}(\tau) &= (g^k)^T \Gamma_1(\tau) \\ &= -(g^k)^T [\sum_{i \in \mathcal{I}} \frac{t_1(\tau)}{\phi_i \tau + 1} g_i^k u^i + \tau \sum_{i \in \mathcal{N}} g_i^k u^i] \\ &= -\sum_{i \in \mathcal{I}} \frac{\tau}{\phi_i \tau + 1} (g_i^k)^2 - \tau \sum_{i \in \mathcal{N}} (g_i^k)^2. \end{aligned} \tag{2.21}$$

Then we have that

$$\frac{d\widehat{\psi}(\tau)}{d\tau} = -\sum_{i \in \mathcal{I}} \frac{1}{(\phi_i \tau + 1)^2} (g_i^k)^2 - \sum_{i \in \mathcal{N}} (g_i^k)^2 \le 0 \tag{2.22}$$

which means that $\widehat{\psi}(\tau)$ is monotonically discreasing for $0 < \tau < \frac{1}{T}$. Thus using the fact $\phi_i \tau_k + 1 > 0$ for all $i \in \mathcal{I}$ when $0 < \tau_k < \frac{1}{T}$, we have that

$$\|B_k\|\tau_k + 1 \ge \phi_i \tau_k + 1.$$

So, as $\mu_k = \frac{1}{\tau_k}$

$$\frac{\tau_k}{\phi_i \tau_k + 1} \ge \frac{\tau_k}{\|B_k\|\tau_k + 1} = \frac{1}{\|B_k\| + \mu_k}.$$

From (2.21), we have that from (2.20),

$$\begin{aligned} & (g^k)^T \delta_k \\ \le & -\frac{1}{\|B_k\| + \mu_k} \sum_{i \in \mathcal{I}} (g_i^k)^2 - \frac{1}{\mu_k} \sum_{i \in \mathcal{N}} (g_i^k)^2 \\ \le & -\frac{\|g^k\|^2}{\|B_k\| + \mu_k} \\ \le & -\frac{\|g^k\|^2}{2\|B_k\| + \frac{\|g^k\|}{\Delta_k}} \\ \le & -\frac{\|g^k\|^2}{2 \max\{\|B_k\|, \frac{\|g^k\|}{\Delta_k}\}} \\ \le & -\frac{1}{2}\|g^k\| \min\{\frac{\|g^k\|}{\|B_k\|}, \Delta_k\}. \end{aligned} \tag{2.23}$$

In hard case, i.e., $\tau_k \ge 1/T$ then $\Gamma(\tau_k) = \Gamma(\frac{1}{T}) + \Gamma(t_2(\tau_k))$. Since $g_i^k = 0$ for all i with $\phi_i = \phi_1$, we have that

$$\begin{aligned} \Gamma(\tau_k) &= \Gamma_1(\frac{1}{T}) + \Gamma(t_2(\tau_k)) \\ &= -\sum_{i \in \mathcal{I}, i \ne 1} \frac{T}{\tau + T} g_i^k u^i - \frac{1}{T} \sum_{i \in \mathcal{N}, i \ne 1} g_i^k u^i. \end{aligned}$$

From $\mu_k = \frac{1}{T}$, we have that for all

$$\frac{1}{\phi_i + T} \geq \frac{1}{\|B_k\| + \mu_k}.$$

Hence, we can obtain that

$$(g^k)^T \delta_k = (g^k)^T \Gamma(\frac{1}{T}) = (g^k)^T \Gamma_1(\frac{1}{T}) \leq -\frac{1}{2}\|g^k\| \min\{\frac{\|g^k\|}{\|B_k\|}, \Delta_k\}. \tag{2.24}$$

From (2.22) and (2.23), and taking $\omega_1 = \frac{1}{2}$ we have that (2.16) holds.

Lemma 2.6. *At the k-th iteration, let δ_k be generated from the modified gradient path in trust region subproblem then the condition (2.17) is also satisfied, i.e.,*

$$(g^k)^T \delta_k \leq -\omega_1 \|g^k\| \min\{\ \Delta_k,\ \frac{\|g^k\|}{\|B_k\|}\ \}$$

where $\omega_1 > 0$ is a constant.

Proof. We difine the value of

$$\widehat{q}_k(\delta_k) \stackrel{\text{def}}{=} (g^k)^T \delta_k + \frac{1}{2}\delta_k^T B_k \delta_k,$$

$\widehat{q}_k$ along the modified gradient path is given by

$$\widehat{q}_k(\Gamma_1(t_1(\tau))) = \sum_{i \in \mathcal{I}} \frac{\exp\{-2\phi_i t_1(\tau)\} - 1}{2\phi_i}(g_i^k)^2 - t_1(\tau) \sum_{i \in \mathcal{N}} (g_i^k)^2, \tag{2.25}$$

for

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

By theorem 4.2 in [1], we have that there exists $\omega > 0$ such that

$$\widehat{q}_k(\Gamma_1(t_1(\tau))) = -\text{Pred}(\Gamma_1(t_1(\tau))) \leq -\omega \|g^k\| \min\{\Delta_k, \frac{\|g^k\|}{\|B_k\|}\}. \tag{2.26}$$

It is clear to see that if ϕ_i^k ($i = 1, \ldots, n$) is the eigenvalues of the B_k, then $\frac{1}{2}\phi_i^k$ ($i = 1, \ldots, n$) is the eigenvalues of the $\frac{1}{2}B_k$. Therefore, by the definition of the modified gradient path and taking the $\frac{1}{2}B_k$ in (2.25), we have that

$$\begin{aligned} & (g^k)^T \delta_k \\ = \ & (g^k)^T \Gamma_1(t_1(\tau)) \\ = \ & \sum_{i \in \mathcal{I}} \frac{\exp\{-\phi_i^k t_1(\tau)\} - 1}{\phi_i^k}(g_i^k)^2 - t_1(\tau) \sum_{i \in \mathcal{N}} (g_i^k)^2 \\ \leq \ & -\omega \|g^k\| \min\{\Delta_k, \frac{\|g^k\|}{\frac{1}{2}\|B_k\|}\}. \end{aligned} \tag{2.27}$$

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

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

we get that

$$
\begin{aligned}
& (g^k)^T \delta_k \\
= \; & (g^k)^T \Gamma_1(t_1(\tau)) + \Gamma_2(t_2(\tau)) \\
= \; & (g^k)^T \Gamma_1(t_1(\tau)) + (g^k)^T \Gamma_2(t_2(\tau)) \\
= \; & -\sum_{i \in \mathcal{I}^+} \frac{1}{\phi_i^k}(g_i^k)^2 + (\tau - 1) g_1^k \\
= \; & -\sum_{i \in \mathcal{I}^+} \frac{1}{\phi_i^k}(g_i^k)^2 \qquad (2.28)\\
\leq \; & -\frac{1}{\max\{\phi_j^k, j \in \mathcal{I}^+\}} \sum_{i \in \mathcal{I}^+} (g_i^k)^2 \\
\leq \; & -\frac{\|g^k\|^2}{\|B_k\|}.
\end{aligned}
$$

By (2.27)–(2.28), taking $\omega_1 = \min\{1, 2\omega\}$, we have that (2.17) holds.

3. Algorithm

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

Initialization step

Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0,1)$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $\epsilon > 0$ and positive integer M.

Let $m(0) = 0$. Choose a symmetric matrix B_0. Select an initial trust region radius $\Delta_0 > 0$ and a maximal trust region radius $\Delta_{\max} \geq \Delta_0$, give a starting point $x_0 \in R^n$. Set $k = 0$, go to the main step.

Main Step

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

2. If $\|g^k\| \leq \epsilon$ or $f_k - f_{k+1} \leq \epsilon \max\{1, |f_k|\}$, stop with the approximate solution x_k.

3. Form various scaled paths Γ_k, the optimal path or modified gradient path.

4. Solve subproblem

$$
(S_k) \qquad \begin{aligned} & \min \quad \psi_k(\delta) \stackrel{\text{def}}{=} (g^k)^T \delta + \frac{1}{2}\delta^T B_k \delta \\ & \text{s.t.} \quad \|\delta\| \leq \Delta_k, \; \delta \in \Gamma_k. \end{aligned}
$$

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

5. 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

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

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

$$\widehat{\rho}_k = \frac{\widehat{\text{Ared}}(h_k)}{\text{Pred}(h_k)}, \tag{3.6}$$

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 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.

Remark 2. In the subproblem (S_k), $\psi_k(\delta)$ is a local quadratic model of the objective function f around x_k. A candidate iterative direction δ is generated by minimizing $\psi_k(\delta)$ along the curve paths Γ_k within the ball centered at x_k with radius Δ_k. As being proved in [1], 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, problem (S_k) can be solved with great ease.

Remark 3. Note that 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 $x_k + h_k$ until the criterion is satisfied.

Remark 4. Comparing usual monotone technique with nonmonotonic technique, when $M \ge 1$, the accepted step h_k only guarantees that $f(x_k + h_k)$ is smaller than $f(x_{l(k)})$. Therefore, generally $f(x_k)$ is no longer monotonically decreasing. Furthermore, it is easy to see that the proposed algorithm becomes the usual monotone algorithm when $M = 0$. So, the usual monotone algorithm can be viewed as a special case of the proposed algorithm.

4. Convergence Analysis

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

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

The following assumptions are commonly used in convergence analysis of most methods for unconstrained optimization.

Assumption A1. B_k is bounded, i.e.,

$$\|B_k\| \le b,\ \forall k,$$

and

$$\|\nabla^2 f(x)\| \le \widehat{b},\ \forall x \in \mathcal{L}(x_0).$$

Lemma 4.1. *Assume that assumption A1 holds. If there exists* $\epsilon > 0$ *such that*

$$\|g^k\| \geq \epsilon \tag{4.1}$$

for all k*, then there is* $\widehat{\alpha} > 0$ *such that*

$$\Delta_k \geq \widehat{\alpha}, \ \forall k \tag{4.2}$$

where

$$\widehat{\alpha} \ \equiv \ \min\{\frac{\bar{\omega}\epsilon(1-\eta_2) \ \sigma\gamma_1}{b+\widehat{b}}, \ \frac{\bar{\omega}\epsilon\gamma_1(1-\beta)}{\bar{b}}\},$$

$\bar{\omega} = \min\{1, \omega_1, \widehat{\omega}\}$ *and* $\bar{b} = \max\{b, \widehat{b}\}$.

Proof. We first prove that if

$$\Delta_k \leq \frac{\bar{\omega}\epsilon(1-\beta)}{\bar{b}}, \tag{4.3}$$

then $\lambda_k = 1$ must satisfy the condition (3.1) in step 5, i.e.,

$$f(x_k + \delta_k) \leq f(x_{l(k)}) + \beta (g^k)^T \delta_k. \tag{4.4}$$

If the above formula is not true, we have

$$f(x_k + \delta_k) > f(x_{l(k)}) + \beta (g^k)^T \delta_k \geq f(x_k) + \beta (g^k)^T \delta_k. \tag{4.5}$$

Because $f(x)$ is twice continuously differentiable, we have

$$f(x_k + \delta_k) - f(x_k) = (g^k)^T \delta_k + \frac{1}{2}\delta_k^T \nabla^2 f(x_k + \xi\delta_k)\delta_k,$$

where $\xi \in [0, 1]$. Hence, (4.5) implies that

$$(1-\beta)(g^k)^T \delta_k + \frac{1}{2}\delta_k^T \nabla^2 f(x_k + \xi\delta_k)\delta_k > 0,$$

from which we obtain

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

By (4.1) and (4.3),

$$-\bar{\omega}\epsilon(1-\beta)\min\{\Delta_k, \frac{\epsilon}{b}\} + \widehat{b}\Delta_k^2 > \ 0. \tag{4.6}$$

Since

$$\Delta_k \leq \frac{\bar{\omega}\epsilon(1-\beta)}{\bar{b}} \leq \frac{\epsilon}{b},$$

we have

$$[-\bar{\omega}\epsilon(1-\beta) + \widehat{b}\Delta_k]\Delta_k > \ 0.$$

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

$$\bar{\omega}\epsilon(1-\beta) < \widehat{b}\Delta_k \leq \bar{b}\Delta_k.$$

which contradicts (4.3).

From the above we see that if (4.3) holds, then the step size $\lambda_k = 1$, i.e., $h_k = \delta_k$ and hence

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

We now assume, without loss of generality, that

$$\epsilon \leq \min\{1, \frac{\bar{\omega}\gamma_1\sigma(1-\eta_2)}{b+\widehat{b}}, \frac{\bar{\omega}(1-\beta)\gamma_1}{\bar{b}}\}. \tag{4.7}$$

We know that

$$\begin{aligned} & |f(x_k + \delta_k) - f(x_k) - \psi_k(\delta_k)| \\ \leq \ & \frac{1}{2}\|\delta_k\|^2 \|\nabla^2 f(x_k + \xi\delta_k) - B_k\| \\ \leq \ & \frac{1}{2}(b + \widehat{b})\Delta_k^2, \end{aligned} \tag{4.8}$$

where $\xi \in [0,1]$.

We now assume that there exists a k such that

$$\Delta_k < \min\{\frac{\bar{\omega}\epsilon(1-\eta_2)\sigma\gamma_1}{b+\widehat{b}},\ \frac{\bar{\omega}\epsilon\gamma_1(1-\beta)}{\bar{b}}\}, \tag{4.9}$$

from which we shall derive a contradiction.

Let t be the first iteration number such that the above inequality holds. As

$$\Delta_{t-1} \ \le \frac{\Delta_t}{\gamma_1} \ \le \frac{\bar{\omega}\epsilon(1-\beta)}{\bar{b}},$$

(4.3) holds for $k=t-1$, and hence $\lambda_{t-1}=1$, i.e., $\delta_{t-1}=h_{t-1}$. As

$$b\Delta_{t-1} \le (b+\widehat{b})\Delta_{t-1} \le (b+\widehat{b})\frac{\Delta_t}{\gamma_1} \le \bar{\omega}\epsilon(1-\eta_2)\sigma \le \epsilon\sigma, \tag{4.10}$$

and by (2.8), we obtain that

$$\text{Pred}(h_{t-1}) = -\psi_{t-1}(\delta_{t-1}) \ge \widehat{\omega}\epsilon\min\{\Delta_{t-1}, \frac{\epsilon\sigma}{b}\} \ge \bar{\omega}\epsilon\Delta_{t-1}. \tag{4.11}$$

Set

$$\rho_k = \frac{f(x_k)-f(x_k+h_k)}{\text{Pred}(h_k)}, \tag{4.12}$$

then, by (4.8) and (4.11), we have

$$\begin{aligned} |\rho_{t-1}-1| &\le \frac{|f(x_{t-1}+h_{t-1})-f(x_{t-1})+\text{Pred}(h_{t-1})|}{\text{Pred}(h_{t-1})} \\ &\le \frac{1}{2}\frac{(b+\widehat{b})\Delta_{t-1}^2}{\bar{\omega}\epsilon\Delta_{t-1}} \\ &= \frac{1}{2}\frac{(b+\widehat{b})\Delta_{t-1}}{\bar{\omega}\epsilon} \\ &\le 1-\eta_2. \end{aligned} \tag{4.13}$$

This implies that $\rho_{t-1} \ge \eta_2$. Therefore,

$$\widehat{\rho}_{t-1} \ge \rho_{t-1} \ge \eta_2.$$

By the updating rule for the trust region radius Δ_k in the step 6, we have

$$\Delta_t \ \ge \ \Delta_{t-1},$$

which implies that

$$\Delta_{t-1} \le \Delta_t \le \min\{\frac{\bar{\omega}\epsilon(1-\eta_2)\sigma\gamma_1}{b+\widehat{b}},\ \frac{\bar{\omega}\epsilon\gamma_1(1-\beta)}{\bar{b}}\}. \tag{4.14}$$

This contradicts the assumption that t is the first index with (4.9) holding.

Hence, (4.9) never holds for any k, i.e., the conclusion of the lemma is true.

We are now ready to state one of our main results.

Theorem 4.2. *Assume that assumption A1 holds. Let $\{x_k\} \in \Re^n$ be a sequence generated by the algorithm. Then*

$$\liminf_{k\to\infty} \|g^k\| = 0. \tag{4.15}$$

Proof. According to the acceptance rule in step 5, we have

$$f(x_{l(k)}) - f(x_k+\lambda_k\delta_k) \ge -\lambda_k\beta(g^k)^T\delta_k. \tag{4.16}$$

Taking into account that $m(k+1) \le m(k)+1$, and $f(x_{k+1}) \le f(x_{l(k)})$, we have

$$\begin{aligned} f(x_{l(k+1)}) &= \max_{0\le j\le m(k+1)}\{f(x_{k+1-j})\} \\ &\le \max_{0\le j\le m(k)+1}\{f(x_{k+1-j})\} \\ &= \max\{f(x_{l(k)}), f(x_{k+1})\} \\ &= f(x_{l(k)}). \end{aligned}$$

This means that the sequence $\{f(x_{l(k)})\}$ is nonincreasing for all k, and therefore $\{f(x_{l(k)})\}$ is convergent.

By (3.1) and (2.17), for all $k > M$,

$$\begin{aligned} & f(x_{l(k)}) \\ = \; & f(x_{l(k)-1} + \lambda_{l(k)-1}\delta_{l(k)-1}) \\ \le \; & \max_{0\le j\le m(l(k)-1)} \{f(x_{l(k)-j-1})\} + \lambda_{l(k)-1}\beta (g^{l(k)-1})^T \delta_{l(k)-1} \qquad (4.17) \\ \le \; & \max_{0\le j\le m(l(k)-1)} \{f(x_{l(k)-j-1})\} - \lambda_{l(k)-1}\beta\omega_1 \|g^{l(k)-1}\| \min\{\Delta_{l(k)-1}, \frac{\|g^{l(k)-1}\|}{\|B_{l(k)-1}\|}\}. \end{aligned}$$

If the conclusion of the theorem is not true, then there exists some $\epsilon > 0$ such that

$$\|g^k\| \ge \epsilon, \; k = 1, 2, \cdots. \qquad (4.18)$$

Therefore, we have that

$$f(x_{l(k)}) \le f(x_{l(l(k)-1)}) - \lambda_{l(k)-1}\beta\omega_1\epsilon \min\{\Delta_{l(k)-1}, \frac{\epsilon}{b}\}. \qquad (4.19)$$

As $\{f(x_{l(k)})\}$ is convergent, we obtain from (4.19) that

$$\lim_{k\to\infty} \lambda_{l(k)-1} \; \Delta_{l(k)-1} = 0. \qquad (4.20)$$

This, by $\|\delta_k\| \le \Delta_k$, implies that

$$\lim_{k\to\infty} \lambda_{l(k)-1} \; \|\delta_{l(k)-1}\| = 0. \qquad (4.21)$$

(4.20) means that either

$$\liminf_{k\to\infty} \lambda_{l(k)-1} = 0, \qquad (4.22)$$

or

$$\liminf_{k\to\infty} \Delta_{l(k)-1} = 0. \qquad (4.23)$$

For $k > M$, we have

$$k - M \le k - m(k) \le l(k) \le k,$$

and hence

$$0 \le k - l(k) \le M.$$

By the updating formula of Δ_k, for all j,

$$\gamma_1^j \Delta_k \le \Delta_{k+j} \le \gamma_2^j \Delta_k,$$

so that

$$\gamma_1^{M+1} \Delta_{l(k)-1} \le \Delta_k \le \gamma_2^{M+1} \Delta_{l(k)-1}.$$

If (4.23) holds, then

$$\liminf_{k\to\infty} \Delta_k = 0,$$

which contradicts (4.2) in Lemma 4.1.

If (4.22) holds, by (4.21), following the way used in [7], we can prove by induction that

$$\lim_{k\to\infty} \|\delta_{l(k)-j}\| = 0, \qquad (4.24)$$

and

$$\lim_{k\to\infty} f(x_{l(k)-j}) = \lim_{k\to\infty} f(x_{l(k)})$$

for any positive integer j. Furthermore, as $k \ge l(k) \ge k - M$, from

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

and (4.24), it can be derived that

$$\lim_{k\to\infty} f(x_{l(k)}) = \lim_{k\to\infty} f(x_k). \qquad (4.25)$$

By the rule for accepting the step h_k,

$$\begin{aligned} f(x_{k+1}) - f(x_{l(k)}) &\le \beta\lambda_k (g^k)^T\delta_k \\ &\le -\lambda_k\beta\omega_1\|g^k\|\min\{\Delta_k, \frac{\|g^k\|}{\|B_k\|}\} \\ &\le -\lambda_k\beta\omega_1\epsilon\min\{\Delta_k, \frac{\epsilon}{b}\}. \end{aligned} \tag{4.26}$$

By Lemma 4.1, (4.25) and (4.26) mean that

$$\lim_{k\to\infty}\lambda_k = 0.$$

The acceptance rule (3.1) means that, for large enough k,

$$f(x_k + \frac{\lambda_k}{\omega}\delta_k) - f(x_k) \ge f(x_k + \frac{\lambda_k}{\omega}\delta_k) - f(x_{l(k)}) > \beta\frac{\lambda_k}{\omega}(g^k)^T\delta_k. \tag{4.27}$$

Since

$$f(x_k + \frac{\lambda_k}{\omega}\delta_k) - f(x_k) = \frac{\lambda_k}{\omega}(g^k)^T\delta_k + o(\frac{\lambda_k}{\omega}\|\delta_k\|),$$

we have

$$(1-\beta)\frac{\lambda_k}{\omega}(g^k)^T\delta_k + o(\frac{\lambda_k}{\omega}\|\delta_k\|) \ge 0. \tag{4.28}$$

Dividing (4.28) by $\frac{\lambda_k}{\omega}\|\delta_k\|$ and noting that $1-\beta > 0$ and $(g^k)^T\delta_k \le 0$, we obtain

$$\lim_{k\to\infty}\frac{(g^k)^T\delta_k}{\|\delta_k\|} = 0. \tag{4.29}$$

From

$$(g^k)^T\delta_k \le -\omega_1\|g^k\|\min\{\Delta_k, \frac{\|g^k\|}{\|B_k\|}\} \le -\omega_1\epsilon\min\{\Delta_k, \frac{\epsilon}{b}\} \tag{4.30}$$

and $\|\delta_k\| \le \Delta_{\max}$, we have that

$$\lim_{k\to\infty}\frac{\Delta_k}{\|\delta_k\|} = 0, \tag{4.31}$$

which contradicts $\Delta_k \ge \|\delta_k\|$ and hence the conclusion of the theorem is true.

5. Further Convergence Properties

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 more assumptions.

Assumption A2. There exists $\hat{\tau} > 0$ such that

$$\|\delta_k\| \le \hat{\tau}\|g^k\|. \tag{5.1}$$

Assumption A3.

$$\lim_{k\to\infty}\frac{\|(B_k - H_k)\delta_k\|}{\|\delta_k\|} = 0, \tag{5.2}$$

where $H_k = \nabla^2 f(x_k)$.

Theorem 5.1. *Assume that assumptions A2 and A3 hold. Let $\{x_k\}$ be a sequence generated by the algorithm. Then*

$$\lim_{k\to\infty}\|g^k\| = 0. \tag{5.3}$$

Proof. Assume that there are an $\epsilon_1 \in (0,1)$ and a subsequence $\{g^{m_i}\}$ of $\{g^k\}$ such that for all $m_i,\ i = 1, 2, \cdots$

$$\|g^{m_i}\| \ge \epsilon_1. \tag{5.4}$$

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

$$\|g^k\| \ge \epsilon_2, \quad \text{for } m_i \le k < l_i, \tag{5.5}$$

and

$$\|g^{l_i}\| \le \epsilon_2 \tag{5.6}$$

for an $\epsilon_2 \in (0, \epsilon_1)$.

By (5.1), we get

$$\begin{aligned}(g^k)^T\delta_k &\le -\omega_1\|g^k\|\min\{\Delta_k, \frac{\|g^k\|}{\|B_k\|}\} \\ &\le -\frac{\omega_1}{\hat{\tau}}\|\delta_k\|\min\{\|\delta_k\|, \frac{\|\delta_k\|}{\hat{\tau}b}\} \\ &= -\hat{\omega}_1\|\delta_k\|^2, \end{aligned}\tag{5.7}$$

where $\hat{\omega}_1 = \frac{\omega_1}{\hat{\tau}}\min\{1, \frac{1}{\hat{\tau}b}\}$. (4.17) and (5.7) mean that

$$f(x_{l(k)}) \le f(x_{l(l(k)-1)}) - \lambda_{l(k)-1}\beta\hat{\omega}_1\|\delta_{l(k)-1}\|^2. \tag{5.8}$$

Similar to the proof of the theorem 4.2, we can obtain (4.25), i.e.,

$$\lim_{k\to\infty} f(x_{l(k)}) = \lim_{k\to\infty} f(x_k). \tag{5.9}$$

By the accepting rule of the step δ_k,

$$f(x_{k+1}) - f(x_{l(k)}) \le \beta\lambda_k(g^k)^T\delta_k \le -\lambda_k\beta\hat{\omega}_1\|\delta_k\|^2. \tag{5.10}$$

(5.9) and (5.10) imply that

$$\lim_{k\to\infty} \lambda_k \ \|\delta_k\|^2 = 0. \tag{5.11}$$

Assume that there exists a subsequence $\mathcal{K} \subseteq \{k\}$ such that

$$\lim_{k\to\infty, k\in\mathcal{K}} \|\delta_k\| > 0. \tag{5.12}$$

Then (5.11) means

$$\lim_{k\to\infty, k\in\mathcal{K}} \lambda_k = 0. \tag{5.13}$$

Similar to (4.29), we can prove that

$$\lim_{k\to\infty, k\in\mathcal{K}} \frac{(g^k)^T\delta_k}{\|\delta_k\|} = 0. \tag{5.14}$$

From (5.7), we have that

$$\lim_{k\to\infty, k\in\mathcal{K}} \|\delta_k\| = 0, \tag{5.15}$$

which contradicts (5.12). Therefore, we know that

$$\lim_{k\to\infty} \|\delta_k\| = 0. \tag{5.16}$$

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

$$\begin{aligned} f(x_k+\delta_k) &= f(x_k) + (g^k)^T\delta_k + o(\|\delta_k\|) \\ &\le f(x_{l(k)}) + \beta(g^k)^T\delta_k + (1-\beta)(g^k)^T\delta_k + o(\|\delta_k\|). \end{aligned}\tag{5.17}$$

Note that

$$(g^k)^T\delta_k \le -\omega_1\|g^k\|\min\{\Delta_k, \frac{\|g^k\|}{\|B_k\|}\} \le -\omega_1\epsilon_2\min\{\|\delta_k\|, \frac{\epsilon_2}{b}\}. \tag{5.18}$$

From (5.16) and (5.18), for large enough i and $m_i \le k < l_i$, we have

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

Hence (5.17) 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 (5.2), we know that

$$\begin{aligned} & |f(x_k+\delta_k) - f(x_k) - \psi_k(\delta_k)| \\ =\ & |((g^k)^T\delta_k + \frac{1}{2}\delta_k^T H_k\delta_k + o(\|\delta_k\|^2)) - ((g^k)^T\delta_k + \frac{1}{2}\delta_k^T B_k\delta_k)| \\ =\ & o(\|\delta_k\|^2). \end{aligned}\tag{5.19}$$

From (2.8) and (5.1), for large enough $i,\ m_i \le k < l_i$,

$$\begin{aligned}\mathrm{Pred}(\delta_k) &\ge \widehat{\omega}\|g^k\|\min\{\Delta_k, \frac{\sigma\|g^k\|}{b}\}\\ &\ge \frac{\widehat{\omega}}{\widehat{\tau}}\|\delta_k\|\min\{\|\delta_k\|, \frac{\sigma\|\delta_k\|}{\widehat{\tau}b}\}\\ &\ge \widehat{\omega}_2\|\delta_k\|^2, \qquad (5.20)\end{aligned}$$

where $\widehat{\omega}_2 = \frac{\widehat{\omega}}{\tau}\min\{1, \frac{\sigma}{\tau b}\}$. As $\delta_k = h_k$, for large $i, m_i \le k < l_i$, we obtain that

$$\begin{aligned}\widehat{\rho}_k &\ge \rho_k = \frac{f_k - f(x_k + h_k)}{\mathrm{Pred}(h_k)}\\ &= 1 + \frac{f_k - f(x_k+\delta_k) + \psi_k(\delta_k)}{\mathrm{Pred}(h_k)}\\ &\ge 1 - \frac{o(\|\delta_k\|^2)}{\widehat{\omega}_2\|\delta_k\|^2} \qquad (5.21)\\ &\ge \eta_2.\end{aligned}$$

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

$$f_k - f(x_k + h_k) \ge \eta_2 \mathrm{Pred}(h_k) \ge \eta_2\widehat{\omega}_2\|h_k\|^2.$$

Therefore, we can deduce that, for large i,

$$\begin{aligned}&\|x_{m_i} - x_{l_i}\|^2\\ \le\ &\sum_{k=m_i}^{l_i-1}\|x_{k+1} - x_k\|^2\\ =\ &\sum_{k=m_i}^{l_i-1}\|\delta_k\|^2 \qquad (5.22)\\ \le\ &\frac{1}{\eta_2\widehat{\omega}_2}\sum_{k=m_i}^{l_i-1}[f(x_k) - f(x_k+h_k)]\\ =\ &\frac{1}{\eta_2\widehat{\omega}_2}(f_{m_i} - f_{l_i}).\end{aligned}$$

(5.22) and (5.9) mean that for large i, we have

$$\|x_{m_i} - x_{l_i}\| \le \frac{\epsilon_1}{2L},$$

where L is the Lipschitz constant of $g(x)$ in $\mathcal{L}(x_0)$. We then use the triangle inequality to show

$$\begin{aligned}\|g^{m_i}\| &\le \|g^{m_i} - g^{l_i}\| + \|g^{l_i}\|\\ &\le L\|x_{m_i} - x_{l_i}\| + \epsilon_2. \qquad (5.23)\end{aligned}$$

We choose $\epsilon_2 = \frac{\epsilon_1}{4}$, then (5.23) contradicts (5.4). This implies that (5.4) is not true, and hence the conclusion of the theorem holds.

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

Theorem 5.2. *If B_k is eventually positive definite and the assumptions A2 and A3 hold, then $\{x_k\}$ converges to x^* superlinearly, that is,*

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

Proof. At the k-th iteration, let δ_k be generated from the optimal path in trust region subproblem. When B_k is positive definite, we have that from (2.18),

$$(g^k)^T\delta_k + \delta_k^T B_k\delta_k = -\mu_k\|\delta_k\|^2 \le 0.$$

Let the step δ_k in trust region be obtained from the modified gradient path, Since B_k is positive definite, which means that $\mathcal{I}^- \cup \mathcal{N} = \phi$, we have that from (2.25) and (2.27),

$$\begin{aligned}
& (g^k)^T\delta_k + \delta_k^T B_k \delta_k \\
= \ & 2[(g^k)^T\delta_k + \frac{1}{2}\delta_k^T B_k\delta_k] - (g^k)^T\delta_k \\
= \ & 2\sum_{i\in\mathcal{I}} \frac{\exp\{-2\phi_i t_1(\tau)\} - 1}{2\phi_i}(g_i^k)^2 - \sum_{i\in\mathcal{I}} \frac{\exp\{-\phi_i t_1(\tau)\} - 1}{\phi_i}(g_i^k)^2 \\
\leq \ & 0.
\end{aligned}$$

Because $f(x)$ is twice continuously differentiable, we have that, noting (5.2) and from above,

$$\begin{aligned}
f(x_k + \delta_k) &= f(x_k) + (g^k)^T\delta_k + \frac{1}{2}\delta_k^T H_k \delta_k + o(\|\delta_k\|^2) \\
&\leq f(x_{l(k)}) + \beta(g^k)^T h_k + (\frac{1}{2} - \beta)(g^k)^T h_k + \frac{1}{2}[(g^k)^T\delta_k + \delta_k^T B_k \delta_k] \\
&\quad + \frac{1}{2}\delta_k^T (H_k - B_k)\delta_k + o(\|\delta_k\|^2) \qquad (5.24) \\
&\leq f(x_{l(k)}) + \beta(g^k)^T\delta_k - (\frac{1}{2} - \beta)\widehat{\omega}_2\|\delta_k\|^2 + o(\|\delta_k\|^2) \\
&\leq f(x_{l(k)}) + \beta(g^k)^T\delta_k,
\end{aligned}$$

where the last two inequalities hold because of $\beta < \frac{1}{2}$ and (5.7).

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., $h_k = \delta_k$.

By assumption A3, we can obtain

$$\begin{aligned}
\rho_k - 1 &= \frac{\mathrm{Ared}(h_k) - \mathrm{Pred}(h_k)}{\mathrm{Pred}(h_k)} \\
&= \frac{((g^k)^T h_k + \frac{1}{2}h_k^T B_k h_k) - ((g^k)^T h_k + \frac{1}{2}h_k^T H_k h_k + o(\|h_k\|^2))}{\mathrm{Pred}(h_k)} \\
&= \frac{o(\|h_k\|^2)}{|\mathrm{Pred}(h_k)|}. \qquad (5.25)
\end{aligned}$$

Assumption A2 and (2.8) deduce that

$$\begin{aligned}
\mathrm{Pred}(h_k) &\geq \widehat{\omega}\|g^k\| \cdot \min\{\Delta_k, \frac{\sigma\|g^k\|}{\|B_k\|}\} \\
&\geq \frac{\widehat{\omega}}{\widehat{\tau}}\|h_k\| \cdot \min\{\|h_k\|, \frac{\sigma\|h_k\|}{\widehat{\tau}b}\} \\
&= \widehat{\omega}_2\|\delta_k\|^2, \qquad (5.26)
\end{aligned}$$

where

$$\widehat{\omega}_2 = \frac{\widehat{\omega}}{\widehat{\tau}} \cdot \min\{1, \frac{\sigma}{\widehat{\tau}b}\}.$$

(5.25) and (5.26) mean that when $\|\delta_k\| \to 0$, $\rho_k \to 1$. Hence there exists $\widehat{\Delta} > 0$ such that when $\|\delta_k\| \leq \widehat{\Delta}$, $\rho_k \geq \eta_2$, and therefore, $\Delta_{k+1} \geq \Delta_k$. As $h_k \to 0$, there exists an index K' such that $\|\delta_k\| \leq \widehat{\Delta}$ whenever $k \geq K'$. Thus

$$\Delta_k \geq \Delta_{K'}, \ \forall k \geq K'.$$

On the other hand, as $g^k \to g^* = 0$, the two paths satisfies the assumption A8 in [1] which ensures

$$\lim_{\tau\to+\infty} \Gamma_k(\tau) = -B_k^{-1}g^k.$$

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

$$h_k = \delta_k = -B_k^{-1} g^k.$$

Therefore, the algorithm becomes the Newton method or the quasi-Newton method. As in this case assumption A3 is a sufficient condition for superlinear convergence, the theorem is proved.

6. Numerical Experiments

Numerical experiments on the nonmonotonic back tracking optimal path method 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 algorithm. A monotonic algorithm is realized by taking $M = 0$. In order to check effectiveness of the back tracking technique, we select the same parameters as used in [5]. The selected parameter values are: $\widehat{\eta} = 0.01, \eta_1 = 0.001, \eta_2 = 0.75$, $\gamma_1 = 0.2$, $\gamma_2 = 0.5$, $\gamma_3 = 2$, $\Delta_{\max} = 10$, $\beta = 0.2$, and initially $\Delta_0 = 1$. The computation terminates when one of the following stopping criterions is satisfied

$$\|g^k\| \leq 10^{-6},$$

or

$$f_k - f_{k+1} \leq 10^{-8} \max\{1, |f_k|\}.$$

Experimental Results

Problem	Initial	BTPATH							
		M =0		M=4			M=8		
Name	Point	NF	NG	NF	NG	NO	NF	NG	NO
Rosenbrock	x_{0a}	25	21	16	14	5	13	12	4
(C=100)	x_{0b}	14	12	7	7	1	7	7	1
Rosenbrock	x_{0a}	92	60	16	16	5	16	14	5
(C=10000)	x_{0b}	30	25	8	8	1	8	8	1
Rosenbrock	x_{0a}	249	214	26	24	5	16	14	5
(C=1000000)	x_{0b}	45	35	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}	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}	33	30	32	30	0	32	30	0

The experiments are carried out on 8 standard test problems which are quoted from [12]. Besides the recommended starting points in [12], 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 BTPATH denote it variation 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 nonmonotonic 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.

The results under BTPATH ($M = 0$) represent mixture of trust region and line search techniques via optimal path considered in this paper. Our curvilinear type of approximate trust region method is very easy to resolve the subproblem (S_k) with a reduced radius via optimal path. Indeed, the formulation of path Γ_k does not depend on the value of Δ_k, so that when the trust region is contracted and δ_k is outside the new region, we only need to set the point back 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.

References

[1] J.P. Bulteau, J.Ph. Vial, Curvilinear path and trust region in unconstrained optimization, a convergence analysis, Mathematical Programming Study, **30** (1987), 82-101.

[2] R.H. Byrd, R.B. Schnabel, G.A. Shultz, Approximate solution of the trust region problem by minimization over two-dimensional subspaces, Mathematical Programming, **40** (1988), 247-263.

[3] N.Y. Deng, Y. Xiao, F.J. Zhou, A nonmonotonic trust region algorithm, *Journal of Optimization Theory and Applications,* **76** (1993), 259-285.

[4] J.E. Dennis, J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.,* **28** (1974), 549-560.

[5] J.E.Jr. Dennis, R.B. Schnable, Numerical methods for unconstrained optimization and nonlinear equations, Prentice Hall, New Jersey, (1983).

[6] R. Fletcher, Practical Methods of Optimization, Unconstrained Optimization, Vol.1, John Wiley & Sons New York, 1980.

[7] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotonic line search technique for Newton's methods, *SIAM Journal on Numerical Analysis,* **23** (1986), 707-716.

[8] J.J. Moré, D.C. Sorensen, Computing a trust region step, *SIAM Journal on Science and Statistical Computing,* **4** (1983), 553-572.

[9] J. Nocedal, Y. Yuan, Combining trust region and line search techniques (1991), in Y. Yuan ed. Advances in Nonlinear Programming (Kluver, Dordvechat, 1998), 153-175.

[10] M.J.D. Powell, A hybrid method for nonlinear equations, In Numerical Methods for Nonlinear Algebraic Equations, edited by Ph. Rabonowitz, Gordon and Breach (1970), pp.87-114.

[11] M.J.D. Powell, On the global convergence of trust region algorithms for unconstrained minimization, Mathematical Programming, **29** (1984), 297-303.

[12] K. Schittkowski, More Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems, 282, Springer-Verlag, Berlin (1987).

[13] D.C. Sorensen, Newton's method with a model trust region modification, *SIAM J. Numer. Anal.,* **19** (1982), 409-426.

[14] T. Stehaug, The conjugent method and trust regions in large scale optimization, *SIAM J. Numer. Anal.,* **20** (1983), 626-637.