Journal of Computational Mathematics, Vol.13, No.1, 1995, 20–31.

A NONMONOTONIC TRUST REGION TECHNIQUE FOR NONLINEAR CONSTRAINED OPTIMIZATION* 1)

Zhu De-tong
(*Shanghai Normal University, Shanghai, China*)

Abstract

In this paper, a nonmonotonic trust region method for optimization problems with equality constraints is proposed by introducing a nonsmooth merit function and adopting a correction step. It is proved that all accumulation points of the iterates generated by the proposed algorithm are Kuhn-Tucker points and that the algorithm is q-superlinearly convergent.

1. Introduction

Consider the nonlinear equality constrained optimization problem

$$\min f(x), \text{s.t. } c(x) = 01.1$$

where $f : R^n \to R^1$ and $c : R^n \to R^m$, $m \leq n$. Recently, reduced Hessian methods are proposed to solve this problem. Coleman and Conn[1], and Nocedal and Overton[6] proposed separately similar quasi-Newton methods using approximate reduced Hessian. However, such methods can not ensure global convergence and therefore are available only when the initial starts are good enough.

Two basic approaches, namely the line search and the trust region, have been developed in order to ensure global convergence towards local minima (see [4] and [5] for example). However, most of the methods based on these two approaches enforce a monotonic decrease of a certain merit function at each step, and this can considerably slow the convergence rate of the minimization process, especially in the presence of steep-sided valleys (see [4], [5]). More recently, the nonmonotonic line search technique for unconstrained optimization was proposed by Grippo, Lampariello and Lucidi[5]. Furthermore, the nonmonotonic technique has been developed into the trust region algorithm for unconstrained optimization[4]. The nonmonotonic idea motivates the study on the projected Hessian methods with trust region. In this paper, we describe and analyze improved projected methods with nonmonotonic trust region for problem (1.1),

* Received July 11, 1992.
1) The Project Supported by National Natural Science Foundation of China.

introducing a nondifferentiable penalty function as a merit function and employing a correction step which allows us to overcome phenomena similar to the Maratos effect.

Section 2 of this paper gives the improved projected Hessian method in association with the nonmonotonic trust region in more detail. In Section 3, we state the global convergence properties of the method, while in Section 4 we prove the local convergence rate of the algorithm.

2. Algorithm

We first introduce some standard notations. Let $\|\cdot\|$ be the Euclidean norm on R^n. Let $f : R^n \to R^1$ be twice continuously differentiable, with gradient $g : R^n \to R^1$ and Hessian matrix $\nabla^2 f$. Let $c : R^m \to R^n$ be the vector of twice continuously differentiable constrained function $c_i(x)$, for $i = 1, 2, \cdots, m$; with the gradient of $c_i(x)$ denoted by $a_i(x)$ and the Hessian matrix of $c_i(x)$ denoted by $\nabla^2 c_i(x)$.

In the sequel, we adopt the notations

$$f_k := f(x_k), \quad g_k := g(x_k), \quad H_k := H(x_k).$$

We first state the revised projected reduced Hession algorithms, in which, after a moving vector p_k is determined by using the two-sided projected Hessian technique of Nocedal and Overton [6], a correction step will also be taken to make the performance of the algorithm more satisfactory and to overcome the Maratos effect.

Let $A(x) = \nabla c(x)$ be the $n \times m$ matrix consisting of the column vectors $a_i(x)$, for $i = 1, \cdots, m$. Assume $A(x)$ has full column rank. Make a QR decomposition for $A(x)$:

$$A(x) = [Y(x), Z(x)]\,[\,R\,(x)0] = Y(x)R(x)2.1$$

where $[Y, Z]$ is an orthogonal matrix and $R(x) \in R^{m\times m}$ is a nonsingular upper triangular matrix. The columns of $Y(x) \in R^{n\times m}$ and $Z(x) \in R^{n\times t}$, $t = n - m$, form respectively a normalized basis of the range space $R(A(x))$ of $A(x)$ and the null space $N(A(x)^T)$ of $A(x)^T$, i.e.

$$A(x)^T Z(x) = 0.2.2$$

Clearly,

$$Y(x)^T Z(x) = 0,\ Y(x)^T Y(x) = I_m,\ Z(x)^T Z(x) = I_t,$$
$$Y(x)Y(x)^T + Z(x)Z(x)^T = I_n.2.3$$

Let

$$L(x, \lambda) = f(x) - \lambda^T c(x)2.4$$

be the Lagrangian of problem (1.1), where λ is the solution vector of the least-squares problem, called projective multiplier,

$$\min_\lambda \|A(x)\lambda - g(x)\|.$$

From (2.1), we have

$$\lambda(x) = (A(x)^T A(x))^{-1} A(x)^T \nabla f(x) = R(x)^{-1} Y(x)^T g(x). 2.4$$

Let

$$W(x,\lambda) = \nabla^2_{xx} L(x,\lambda)$$

be the Hessian of the function $L(x,\lambda)$ with respect to x.

The general projected Hessian proposed by Nocedal and Overton in [6] follows. In each iteration:

Solve the equations

$$R_k^T p_k^y = -c_k, 2.5 B_k p_k^z = -Z_k^T g_k 2.6$$

to obtain p_k^y and p_k^z, respectively. Let

$$p_k = Z_k p_k^z + Y_k p_k^y 2.7$$

and then take

$$x_{k+1} = x_k + p_k.$$

The computation will be terminated when $c_k = 0$ and $Z_k^T g_k = 0$. At this point, the Kuhn-Tucker condition is satisfied. The matrix B_k is updated using the BFGS or DFP formulas after each iteration (see [6]).

A principal distinction between the Nocedal-Overton method or the Coleman-Conn method, called the reduced Hessian, and the usual quasi-Newton methods, called the full Hessian, is that in the former the updating matrix $B \in R^{t\times t}$ is an approximation of the square matrix $Z(x)^T W(x,\lambda) Z(x)$ of order t, whereas in the latter the updating matrix approximates $W(x,\lambda)$. Nocedal and Overton proved the local convergence and two-step q-superlinear convergence rate for their methods in [6]. Coleman and Conn also gave the result of the local convergence and convergence rate in [1].

Now, we consider the nonmonotonic trust region technique for the improved projected reduced Hessian method, introducing a nondifferentiable penalty function as the merit function.

In each iteration, we first solve a trust region subproblem

$$(S_k) \qquad l\min\ (Z_k^T g_k)^T p^z + 12 (p^z)^T B_k (p^z), \text{s.t. } \ \|p^z\| \le \Delta_k$$

where Δ_k is the radius of the trust region. Let p_k^z be the solution of problem (S_k). Then, calculate

$$p_k^y = -\alpha_k R_k^{-T} c_k 2.8$$

where

$$\alpha_k = \{ 1 \text{ ,if } c_k = 0 \text{,} 1, \text{if } c_k \neq 0 \text{ and } \Delta_k \|c_k\| \|A_k (A_k^T A_k)^{-1}\| \ge 1, \Delta_k \|c_k\| \|A_k (A_k^T A_k)^{-1}\|, \text{otherwise.} 2.9$$

The definition of p_k^y is the same as in [8] and for its motivation, please refer to that paper.

Let

$$p_k = Y_k p_k^y + Z_k p_k^z \quad 2.10$$

which will be the main movement in any successful iteration. But to overcome the Maratos effect, a correction term d_k generated from the lower triangular equation

$$R_k^T h_k = -c(x_k + p_k) + (1 - \alpha_k) c_k \quad 2.11$$

and then

$$d_k = Y_k h_k \quad 2.12$$

needs to be introduced.

In order to decide whether to take

$$x_{k+1} = x_k + p_k + d_k \quad 2.13$$

we introduce the l_1-nondifferentiable exact penalty function

$$\phi(x, \rho) = f(x) + \rho \|c(x)\|_1 \quad 2.14$$

where

$$\|c(x)\|_1 = \sum_{i=1}^{m} |c_i(x)| \quad 2.15$$

and ρ is a penalty parameter which is updated by

$$\rho_k = \{ \rho_{k-1}, \text{if } \rho_{k-1} \geq \tau + \|\lambda_k\|_\infty, \max\{\rho_{k-1}, \|\lambda_k\|_\infty\} + \tau, \text{otherwise} \quad 2.16$$

where τ is a given positive constant.

For the actual change of $\phi(x, \rho)$ from x_k to $x_k + p_k + d_k$;

$$\Delta\phi_k = \phi(x_k + p_k + d_k;, \rho_k) - \phi(x_k, \rho_k), 2.17$$

we define its estimate value by

$$\Delta\varphi_k = (Z_k^T g_k)^T p_k^T + \frac{1}{2}(p_k^z)^T B_k p_k^z - \alpha_k(\lambda_k^T c_k + \rho_k \|c_k\|_1). 2.18$$

Relaxing the acceptability condition on $p_k + d_k$, we use the nonmonotonic trust region method. Set

$$\phi(x_{l(k)}, \rho_{l(k)}) = \max_{0 \leq j \leq m(k)} \{\phi(x_{k-j}, \rho_{k-j})\} \quad 2.19$$

and

$$\Delta\tilde{\phi}_k = \phi(x_k + p_k + d_k, \rho_k) - \phi(x_{l(k)}, \rho_{l(k)}) \quad 2.20$$

where $m(0) = 0$ and $0 \leq m(k) \leq \min\{m(k-1)+1, M\}$, $k \geq 1$, M is a nonnegative integer, and $l(k)$ satisfies

$$k - m(k) \leq l(k) \leq k.2.21$$

Now the complete algorithm can be stated as follows.

Nonmonotonic Trust Region Algorithm (NTRA).

Given $\mu \in (0,1)$, $\eta \in (\mu, 1)$ and r_0, r_1, r_2 which satisfy

$$0 < r_0 \leq r_1 < 1 \leq r_2.2.22$$

Step 0. Give the starting x_0, f_0 and g_0 and an initial trust region radius Δ_0 and B_0, an initial approximation the Hessian at the starting point.

Step 1. Solve the subproblem (S_k) and (2.8)–(2.9).

Step 2. If $p_k = 0$, then stop; otherwise go to the next step.

Step 3. Compute $d_k, \rho_k, \Delta\phi_k, \Delta\varphi_k, \Delta\tilde{\phi}_k$ and set

$$\eta_k = \frac{\Delta\varphi_k}{\Delta\varphi_k}, \quad \text{and} \quad \varepsilon_k = \frac{\Delta\tilde{\phi}_k}{\Delta\varphi_k}.2.23$$

Step 4. In the case where

$$\varepsilon_k \geq \mu 2.24$$

this iteration is said to be successful; set

$$x_{k+1} = x_k + p_k + d_k 2.25$$

and choose

$$\Delta_{k+1} \in [\Delta_k, r_2\Delta_k], \quad \text{if } \eta_k \geq \eta, 2.26$$

or

$$\Delta_{k+1} \in [r_1\Delta_k, \Delta_k), \quad \text{if } \eta_k \leq \eta.2.27$$

Otherwise, i.e. $\varepsilon_k < \mu$, the iteration is said to be unsuccessful, and let

$$x_{k+1} = x_k, 2.28 \Delta_{k+1} \in [r_0\Delta_k, r_1\Delta_k].2.29$$

Go to step 1.

Step 5. Using the BFGS or DFP formula to update B_k, set $k \leftarrow k+1$ and $m(k) \leq \min\{m(k-1)+1, M\}$, and return to step 1.

Comparing the usual trust region with the nonmonotonic trust region, when $M > 0$, we see the accepted step $p_k + d_k$ should guarantee a certain decrease of $\phi(x_k + p_k + d_k, \rho_k)$ compared with $\phi(x_{l(k)}, \rho_{l(k)})$. Therefore $\{\phi(x_k, \rho_k)\}$ may not be monotonically decreasing although $\Delta\phi_k < 0$ can be proved in the following section.

Furthermore, it is easy to see that Algorithm NTRA is the usual trust region algorithm when $M = 0$. So the usual trust region algorithm, the monotonic trust region algorithm, can be viewed as a special case of Algorithm NTRA.

3. Global Convergence

In order to discuss the global convergence properties of Algorithm NTRA, we should make the following assumptions.

Assumption A1. Sequences $\{x_k\}, \{x_k + p_k\}$ and $\{x_k + p_k + d_k\}$ are all contained in a convex compact set Ω, and $\nabla^2 f$ and $\nabla^2 c_i$ $(i = 1, \cdots, m)$ are Lipschitz continuous matrix functions on a convex compact Ω.

Assumption A2. Matrix $A(x)$ has full column-rank over Ω.

Assumption A3. $Y(x), Z(x)$ and $R(x)$, obtained by a QR decomposition of $A(x)$, are Lipschitz continuous matrix functions on Ω.

Assumption A4. There is a $b > 0$ such that, for each matrix B_k,

$$v^T B_k v \leq b\|v\|^2, \quad \forall v \in R^t,\ \forall k.3.1$$

According to the assumptions, there exists an r such that

$$\|R_k^{-1}\| \leq r, \quad \forall k.3.2$$

The following properties have been proved in [8].

Lemma 3.1.

$$\|d_k\| = O(\|p_k\|^2),\ \textit{and}\ \|d_k\| = O(\Delta_k^2)3.3$$

(*see Lemma* 3.1 *in* [8]).

Lemma 3.2.

$$\alpha_k\|c_k\|_1 \geq \min\{\|c_k\|, \frac{\Delta_k}{r}\}, \quad \textit{for all } k3.4$$

(*see Lemma* 3.3 *in* [8]).

Lemma 3.3.

$$|\Delta\phi_k - \Delta\phi_k| = O(\|p_k\|^2).3.5$$

Proof. The proof is similar to that of Lemma 3.2 in [8].

Corollary 3.4. There exists a positive constant L such that, for all k,

$$|\Delta\phi_k - \Delta\varphi_k| \leq L\Delta_k^2.3.6$$

Proof. By (3.5) and the definitions of p_k^z and p_k^y,

$$\|p_k\|^2 = \|p_k^z\|^2 + \|p_k^y\|^2 \leq 2\Delta_k^2.3.7$$

Lemma 3.5.

$$\Delta\varphi_k \leq -\frac{1}{2}\|Z_k^T g_k\| \min\Big\{\Delta_k, \frac{\|Z_k^T g_k\|}{b}\Big\} - \tau \min\Big\{\|c_k\|, \frac{\Delta_k}{r}\Big\}3.8$$

(*see Lemma* 3.5 *in* [8]).

Define

$$K_\varepsilon = \{x\|c(x)\| + \|Z(x)^T g(x)\| \leq \varepsilon\}.3.9$$

Lemma 3.6. *For any given $\varepsilon > 0$, there exists $\bar{\Delta} > 0$ such that, for any $x \in X \backslash K_\varepsilon$, with a trust region radius $\Delta_k \leq \bar{\Delta}$ (2.4) holds, i.e.*

$$\xi_k \geq \mu.$$

Proof. By the definitions of $\Delta\varphi_k$ and $\Delta\tilde{\phi}_k$ in (2.17) and (2.20), we have

$$\Delta\phi_k = \phi(x_k + p_k + d_k, \rho_k) - \phi(x_k, \rho_k) \leq \phi(x_k + p_k + d_k, \rho_k) - \phi(x_{l(k)}; \rho_{l(k)}) = \Delta\tilde{\phi}_k.3.10$$

Therefore, from (3.8), $\Delta\phi_k < 0$, we have

$$\xi_k = \frac{\Delta\tilde{\phi}_k}{\Delta\varphi_k} \geq \frac{\Delta\phi_k}{\Delta\varphi_k} = \eta_k.3.11$$

Moreover, by Lemma 3.6 of [8], $\eta_k \geq \mu$ when $\Delta_k \leq \bar{\Delta}$. This implies that (2.24) is true.

Theorem 3.7. *Under assumptions* A1–A4, *assume the sequence $\{x_k\}$ generated by Algorithm NTRA. Then,*

$$\lim_{k\to\infty} \inf\{\|Z_k^T g_k\| + \|c_k\|\} = 0, 3.12$$

i.e. every limit point of $\{x_k\}$ is a Kuhn-Tucker point of problem (1.1).

Proof. According to (2.24) and (3.8), we have

$$\phi(x_k+p_k+d_k, \rho_k)-\phi(x_{l(k)}, \rho_{l(k)}) \leq \mu\Delta\varphi_k \leq -\frac{\mu}{2}\|Z_k^T g_k\| \min\left\{\|\Delta_k\|, \frac{\|Z_k^T g_k\|}{b}\right\}-\mu\tau \min\left\{\|c_k\|, \frac{\Delta_k}{r}\right\}.3.13$$

Similarly to the proof of the theorem of [8], we have that the sequence $\{\phi(x_{l(k)}, \rho_{l(k)})\}$ is nonincreasing for large enough k.

Moreover, we obtain

$$\phi(x_{l(k)}, \rho_{l(k)}) \leq \phi(x_{l(l(k)-1)}, \rho_{l(l(k)-1)})-\frac{\mu}{2}\|Z_{l(k)-1}^T g_{l(k)-1}\| \min\left\{\Delta_{l(k)-1}\frac{\|Z_{l(k)-1}^T g_{l(k)-1}\|}{b}\right\}-\mu\tau \min\left\{\|c_{l(k)-1}, \frac{\Delta_{l(k)-1}}{r}\right.$$

for all $k > M$.

If the conclusion (3.12) is not true, there is some $\varepsilon > 0$ such that

$$\|c_k\| + \|Z_k^T g_k\| \geq 2\varepsilon, \quad k = 1, 2, \cdots, 3.15$$

and hence either

$$\|c_k\| \geq \varepsilon, \quad k = 1, 2, \cdots,$$

or

$$\|Z_k^T g_k\| \geq \varepsilon, \quad k = 1, 2, \cdots.3.16$$

Therefore, we have either

$$\phi(x_{l(k)}, \rho_{l(k)}) \leq \phi(x_{l(l(k)-1)}, \rho_{l(l(k)-1)}) - \frac{\mu}{2}\varepsilon \min\left\{\Delta_{l(k)-1}, \frac{\varepsilon}{b}\right\}3.17$$

or

$$\phi(x_{l(k)}, \rho_{l(k)}) \leq \phi(x_{l(l(k)-1)}, \rho_{l(l(k)-1)}) - \mu\tau \min\Big\{\varepsilon, \frac{\Delta_{l(k)-1}}{r}\Big\}.3.18$$

Since $\{\phi(x_{l(k)}, \rho_{l(k)})\}$ is nonincreasing and ρ_k is bounded for large k, from (3.17) and (3.18) we have

$$\lim_{k\to\infty} \Delta_{l(k)-1} = 0.3.19$$

Let

$$\hat{l}(k) = l(k+M+2).3.20$$

Similarly to the proof of the theorem in [5], by induction we obtain

$$\lim_{k\to\infty} \Delta_{\hat{l}(k)-j} = 03.21$$

for any given $j \geq 1$.

From (2.21),

$$k+2 = (k+M+2) - M \leq (k+M+2) - m(k+M+2) \leq l(k+M+2) = \hat{l}(k).$$

Then, taking $j = \hat{l}(k) - k \geq 1$, we have

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

On the other hang, if $\|c_k\| \geq \varepsilon$, from Lemma 3.5, when $\Delta_k \leq r\varepsilon$,

$$\Delta\varphi_k \leq -\frac{\tau\Delta_k}{r}; 3.23$$

if $\|Z_k^T g_k\| \geq \varepsilon$, then from Lemma 3.5, when $\frac{\varepsilon}{b} \geq \Delta_k$,

$$\Delta\varphi_k \leq -\frac{1}{2}\varepsilon\Delta_k.3.24$$

As above, when $\Delta_k \leq \min\{\frac{\varepsilon}{b}, r\varepsilon\}$ we have

$$\Delta\varphi_k \leq -\Delta_k \min\left\{\frac{1}{2}\varepsilon, \frac{\tau}{r}\right\} = -\tau_1\Delta_k.3.25$$

From (3.6), as $\Delta_k \to 0$,

$$|(\eta_k - 1)| = \left(\frac{\Delta\phi_k - \Delta\varphi_k}{\Delta\varphi_k}\right) \leq \frac{L\Delta_k^2}{\tau_1\Delta_k} \to 0.3.26$$

This implies $\eta_k \to 1$, i.e. for large k, $\eta_k \geq \eta$. This implies that the trust region radius will be bounded away from 0, which contradicts (3.22).

We have proved that the improved projected reduced Hessian method with nonmonotonic trust region technique is globally convergent. The advantage of our choice is relaxing some trust region conditions to accept steps. The cost is that the convergence result is

$$\lim_{k\to\infty} \inf\{\|Z_k^T g_k\| + \|c_k\|\} = 03.27$$

instead of

$$\lim_{k\to\infty} \{\|Z_k^T g_k\| + \|c_k\|\} = 0.3.28$$

4. Local Convergence Rate

In the following, we further discuss the local convergence rate of the improved algorithm with nonmonotonic trust region technique. We further take

Assumption A5.

$$x_k \to x_* 4.1$$

where x_* is a Kuhn-Tucker point of problem (1.1), i.e. there exists a vector $\lambda_* \in R^m$ such that

$$\nabla_* l(x_*, \lambda_*) = g_* + A_* \lambda_* = 0.4.2$$

Assumption A6. There is $\alpha_1 > 0$ such that

$$p^T W_* p \geq \alpha_1 \|p\|^2, \quad \text{when} \quad A_*^T p = 04.3$$

where

$$W_* = W|x_*, \lambda_*|$$

is a Hessian matrix of the Lagrangian of problem (1.1) at x_*.

Assumption A7.

$$\lim_{k\to\infty} \frac{\|(B_k - Z_*^T W_* Z_*)p_k^z\|}{\|p_k\|} = 0.4.4$$

The following two properties were proved in [8].

Lemma 4.1.

$$\Delta\phi_k - \Delta\varphi_k = o(\|p_k^z\|^2) + o(\|p_k^y\|).4.5$$

Lemma 4.2.

$$\Delta\varphi_k \leq -\frac{\alpha_1}{2}\|p_k^z\|^2 - \frac{\tau}{r}\|p_k^y\| 4.6$$

(*See Lemma* 4.2 *and Lemma* 4.3 *in* [8]).

Lemma 4.3. *There is $\bar{\Delta} > 0$ such that, for sufficiently large k,*

$$\Delta_k \geq \bar{\Delta}.4.7$$

Proof. By the previous two lemmas,

$$|(\eta_k - 1)| = \left|\frac{(\Delta\phi_k - \Delta\varphi_k)}{(\Delta\varphi_k)}\right| \leq \frac{(o(\|p_k^z\|^2) + o(\|p_k^y\|))}{\frac{\alpha_1}{2}\|p_k^z\|^2 + \frac{\tau}{r}\|p_k^y\|} \to 04.8$$

as $p_k^z \to 0$ and $p_k^y \to 0$. Thus, by the proposed algorithm, for large k, $\Delta_{k+1} \geq \Delta_k$, and (4.7) holds.

The theorem implies that, as $p_k^z \to 0$ and for large k, the constraint of the subproblem (S_k) is inactive, which means that p_k^z is the solution of the equation

$$B_k p_k^z = -Z_k^T g_k 4.9$$

and $c_k \to c_* = 0$, which ensures $\alpha_k = 1$ for large k. So

$$p_k^y = -R_k^{-T} c_k 4.10$$

and h_k is the solution of equation

$$R_k h_k = -c(x_k + p_k) 4.11$$

where

$$p_k = Y_k p_k^y + Z_k p_k^z .4.12$$

In addition,

$$d_k = Y_k h_k .4.13$$

Set

$$x_{k+1} = x_k + p_k + d_k .4.14$$

Theorem 4.4. *Under assumptions* A1–A7, *the algorithm is two-step q-superlinearly convergent, i.e.*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_{k-1} - x_*\|} = 0.4.15$$

Furthermore, sequence $\{x_k + p_k\}$ *is one-step q-superlinearly convergent, i.e.*

$$\lim_{k\to\infty} \frac{\|x_k + p_k - x_*\|}{\|x_{k-1} + p_{k-1} - x_*\|} = 0.4.16$$

Proof. The proof of (4.15) can be found in Theorem 4.1 of [8]. Now we prove (4.16).

It is easy to prove that

$$\|d_k\| = O(\|x_k + p_k - x_*\|), 4.17 \|x_{k+1} - x_*\| = O(\|x_k + p_k - x_*\|).4.18$$

By [6], we obtain

$$\|p_k\| = O(\|x_k - x_*\|), 4.19 \|x_k + p_k - x_*\| = O(\|x_k - x_*\|).4.20$$

And hence

$$A_k^T(x_k+p_k-x_*) = A_k^T(x_k-x_*)-(c_k-c_*) = (A_k-A_*)^T(x_k-x_*)+o(\|x_k-x_*\|) = o(\|x_k-x_*\|) = o(\|x_{k-1}+p_{k-1}-x_*\|).4$$

Moreover,

$$c_k = c(x_{k-1}+p_{k-1}+d_{k-1}) = x(x_{k-1}+p_{k-1})A(x_{k-1}+p_{k-1})^T d_{k-1}+O(\|d_{k-1}\|^2) = [A(x_{k-1}+p_{k-1})-A_{k-1}]^T d_{k-1}+O(\|$$

In [6], Nocedal and Overton proved that

$$Z_k^T W_*(x_k - x_*) = Z_k^T g_k + O(\|x_k - x_*\|^2).4.23$$

Let

$$P_k = I - A_k(A_k^T A_k)^{-1} A_k^T, \quad Q_k = A_k(A_k^T A_k)^{-1} A_k^T.4.24$$

Then,

$$Q_k(x_k - x_*) = Q_k(x_k + p_k - x_*) - Q_k p_k = o(\|x_{k-1} + p_{k-1} - x_*\|) - Y_k p_k^y = o(\|x_{k-1} + p_{k-1} - x_*\|) + A_k(A_k^T A_k)^{-1} c_k = o(\|x$$

Since

$$Z_k^T W_* P_k(x_k + p_k - x_*) = -Z_k^T g_k + Z_k^T W_*[(x_k - x_*) - Q_k(x_k - x_*)] + o(\|p_k\|^2), 4.26$$

we have

$$Z_k^T(x_k + p_k - x_*) = o(\|x_{k-1} + p_{k-1} - x_*\|), 4.27$$

and (4.21) can be written as

$$Y_k^T(x_k + p_k - x_*) = o(\|x_{k-1} + p_{k+1} - x_*\|).4.28$$

Combining (4.27) and (4.28) and noting

$$\|x_k + p_k - x_*\| \le \|Z_k^T(x_k + p_k - x_*)\| + \|Y_k^T(x_k + p_k - x_*)\|,$$

we obtain

$$\|x_k + p_k - x_*\| = o(\|x_{k-1} + p_{k-1} - x_*\|),$$

i.e., (4.16) is true.

Remark. In the proof of Lemma 4.3, (4.8) indicates that

$$\eta_k \to 1, 4.29$$

which implies that, when k is large enough, $\{\phi(x_k, \rho_k)\}$ is nonmonotonically decreasing. So the same results of the superlinear convergence rate for the usual trust region in [8] and Algorithm NTRA remain valid.

In the above, we have studied the convergence properties of the projected reduced Hessian methods with nonmonotonic trust region technique, introducing the nondifferentiable merit function, for nonlinear equality constrained optimization problems. The proposed algorithm still needs a lot of numerical tests so that it can be applied in practice.

References

[1] T.F. Coleman and A.R. Conn, On the local convergence of a quasi-Newton method for the nonlinear programming problem, *SIAM J. Numer. Anal.,* 21 (1984), 755–769.

[2] R.H. Byrd, R.B. Schnable and G.A. Shity, A trust region algorithm for nonlinearly constrained optimization, *SIAM J. Numer. Anal.,* 24 (1987), 1152–1170.

[3] M.R. Celis, J.E. Dennis and R.A. Tapia, A trust region strategy for nonlinear equality constrained optimization, in P. Boggs et al. (ed.), Numerical Optimization, SIAM, Philadelphia, 1985.

[4] N.Y. Deng, Y. Xiao and F.J. Zhou, A nonmonotone trust region algorithm, to appear.

[5] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone line search technique for Newton's methods, *SIAM J. Numer. Anal.,* 23 (1986), 401–419.

[6] J. Nocedal and M.L. Overton, Projected Hessian updating algorithms for nonlinearly constrained optimization, *SIAM J. Numer. Anal.,* 22 (1985), 821–850.

[7] M.J.D. Powell and Y. Yuan, A trust region algorithm for equality constrained optimization, Report, DAMTP 1986 INAZ, Cambridge, England, 1986.

[8] J. Zhang and D. Zhu, A trust region Degleg method for nonlinear optimization, *Optimization,* 21 (1990), 543–557.

[9] J. Zhang, D. Zhu and S. Hou, Some improved two-sided projected quasi-Newton algorithms and their convergence, I–II, *Acta Mathematicae Applieatae Sinica,* 5 (1989), 33-59.