第31卷第4期 上海师范大学学报(自然科学版) Vol. 31, No. 4
2002年12月 Journal of Shanghai Teachers University(Natural Sciences) Dec. 2002

A Two-piece Update Algorithm with Nonmonotonic Backtracking Technique for Constrained Optimization

GU Yi-ming, ZHU De-tong

(Mathematics and Sciences College, Shanghai Teachers University, Shanghai 200234, China)

Abstract: We propose a two-piece update of projected Hessian algorithm with trust region method for solving nonlinear equality constrained optimization problems. To deal with large problems, a two-piece update of two-side-reduced Hessian is used to replace the full Hessian matrix. By adopting the l_1 penalty function as the merit function, a nonmonotonic backtracking trust region strategy is suggested which does not require the merit function to its value in every iteration. A correction step is avoided to overcome the Maratos effect. The proposed algorithm which switchs to nonmonotonic trust region strategy possesses global convergence while maintaining one step Q-superlinear local convergence rates if at least one of the update formula is updated in each iteration.

Key words: nonmonotonic technique; two-piece update; superlinear convergence

CLC number: O221.2 **Document code**: A **Article ID**: 1000-5137(2002)04-0008-08

1 Introduction

Consider the following optimization problem with nonlinear equality constraints:

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & c(x) = 0 \end{aligned} \tag{1.1}$$

where $f(x): \mathbf{R}^n \to \mathbf{R}^1$ and $c(x): \mathbf{R}^n \to \mathbf{R}^m$, $m \leqslant n$. GURWITZ[2] proposed a two-piece update method of a projected Hessian matrix. The basic idea can be summarized as follows: let $g(x) = \nabla f(x) \in \mathbf{R}^n$, $A(x) = \nabla c(x) = [\nabla c_1(x), \cdots, \nabla c_m(x)] \in \mathbf{R}^{n \times m}$. Assuming $A(x)$ has full column rank, then a QR decomposition can be performed, that is,

$$A(x) = [Y(x), Z(x)] \begin{bmatrix} R(x) \\ 0 \end{bmatrix}, \tag{1.2}$$

where $[Y(x), Z(x)]$ is an orthogonal matrix, $R(x)$ is a nonsingular upper triangular matrix of order m, and $Z(x) \in \mathbf{R}^{n \times (n-m)}$. The column vectors of $Y(x)$ and $Z(x)$ form an orthonormal basis for the

Received date: 2002-09-10

Foundation item: Supported by the Science Grant of Shanghai Normal University(DQ17).

Biography: GU Yi-ming(1974-), male, lecturer, Mathematics and Sciences College, Shanghai Teachers University.

null space $\mathcal{N}(A(x)^T)$ and the range space $\mathcal{R}(A(x))$, respectively. Let

$$l(x,\lambda) = f(x) - \lambda^T c(x) \tag{1.3}$$

be the Lagrangian function of problem (1.1), where $\lambda(x)$ can be obtained by solving the upper triangular equation

$$R(x)\lambda(x) = Y(x)^T g(x). \tag{1.4}$$

Let $W(x,\lambda) = \nabla^2_{xx} l(x,\lambda)$ be the Hessian of the function $l(x,\lambda)$ with respect to x. For simplicity, let $f(x_k)$ denote f_k; $\nabla f(x_k)$ denote g_k, etc. In each iteration, let B_k denote the current approximation to $Z_k^T W_k Z_k$ and D_k denote the current approximation to $Z_k^T W_k Y_k$. The DFP or BFGS update will be used to compute B_{k+1} and the Broyden update will be used to obtain D_{k+1}. The Gurwitz's method solves the equation

$$R_k^T p_k^y = - c_k, \tag{1.5}$$

to obtain p_k^y, and then solves another equation

$$B_k p_k^z = - Z_k^T g^k - D_k p_k^y \tag{1.6}$$

to obtain p_k^z. Let

$$p_k = Z_k p_k^z + Y_k p_k^y. \tag{1.7}$$

Take

$$x_{k+1} = x_k + p_k. \tag{1.8}$$

The null-space step and the range-space step are respectively defined by

$$v_k = Y_k^T(x_{k+1} - x_k), \qquad s_k = Z_k^T(x_{k+1} - x_k). \tag{1.9}$$

GURWITZ suggested an alternative strategy. If the range-space step is large compared to the null-space step, then they silmply perform the update rule 1. Similarly, if the null-space step is large compared to the range-space step, then they silmply perform the update rule 2. The update rules 1 and 2 were given in [2].

GURWITZ[2] discussed only the local Q-superlinear of this original algorithm. It was proved that if the initial point x_0 is close enough to the solution of the problem (1.1), x_*, and $B_k \approx Z_*^T W_* Z_* \stackrel{\text{def}}{=} H_*$, $D_k \approx Z_*^T W_* Y_* \stackrel{\text{def}}{=} G_*$, then the iterative sequence of $\{x_k\}$ approaches x_*. Furthermore, under some reasonable conditions, the convergence is at least two-step Q-superlinear. If at least one of the matrices B_k or D_k is updated in each iteration, then the rate of convergence is one-step Q-superlinear.

Trust region method and line search are two very popular ways for minimization to assure global convergence. By using the l_1 penalty function as a merit function, the resulting algorithm with trust region strategy shall possess global convergence while maintaining superlinear local convergence rates under some reasonable conditions.

In this paper we shall extend the nonmonotonic backtracking trust region typed method to two-piece update algorithm for constrained optimization problems, giving the oretic analysis. In order to overcome the Maratos effect and ensure a superlinear convergence rate, a correction step in each iteration is added to the algorithm in [2]. The paper is outlined as follows. In section 2, we state a revised two-piece update of projected Hessian algorithm with backtracking trust region method. The global convergence of the algorithm is proved in section 3; while the local convergence rates of the roposed algorithm is presented in section 4.

2 Algorithm

The trust region algorithm for solving the problem (1.1) is simple to explain and motivate. In a current iteration x_k of the algorithm we shall solve a quadratic subproblem

$$\min \quad g_k^T p + \frac{1}{2} p^T W_k p \tag{2.1}$$

$$\text{subject to} \quad A_k^T p + c_k = 0 \tag{2.2}$$

$$\|p\| \leqslant \Delta_k \tag{2.3}$$

where Δ_k is called the trust region radius. It motivates to solve the linear constraint (1.5) at the GURWITZ's method. This is done by defining a relaxation parameter $\mu_1 \in (\frac{1}{2},1)$ and computing a step p_k^y that solves the vertical (or normal) subproblem

$$\min \quad \|R_k^T p^y + c_k\|$$

$$\text{subject to} \quad \|p^y\| \leqslant \mu_1 \Delta_k. \tag{2.4}$$

The algorithm is designed so that the full step p need not move any closer to the feasible manifold than $Y_k p_k^y$ does, so we next reformulate (1.6) as

$$\min \quad q_k(p^z) \overset{\text{def}}{=} (Z_k^T g_k + D_k p_k^y)^T p^z + \frac{1}{2}(p^z)^T B_k p^z$$

$$\text{subject to} \quad \|p^z\| \leqslant \mu_2 \Delta_k \tag{2.5}$$

where $\mu_2 \in (\frac{1}{2},1)$ also gives a relaxation parameter. Let the solution of the horizontal (or tangential) subproblem be denoted by p_k^z. Define the total trust region algorithm as

$$p_k = Z_k p_k^z + Y_k p_k^y. \tag{2.6}$$

We then set

$$x_{k+1} = x_k + p_k, \tag{2.7}$$

provided x_{k+1} gives a reduction in the merit function; otherwise, the trust region is reduced and a new trial step is computed. The merit function is described as follows.

In order to decide the acceptance of the new point at each iteration, and to adjust the trust region radius, introducing a merit function is necessary. Here we choose the l_1 exact penalty function as the merit function,

$$\varphi(x) = f(x) + \sum_{i=1}^{m} r_i |c_i(x)|, \tag{2.8}$$

where r_i is the i-th components of the vector r.

We now describe the complete algorithm.

Algorithm

(1) Choose parameters $0 < \bar{\eta} < \eta_1 < \eta_2 < 1$, $\mu_1, \mu_2 \in (\frac{1}{2},1)$, $\Delta_{\max} > 0$, $0 < \gamma_1 < 1 < \gamma_2$, $\varepsilon \geqslant 0$, $\beta \in (0, \frac{1}{2})$, $\alpha \in (0,1)$, $0 < \tau_1 \leqslant \tau_2 \leqslant \frac{1}{4}$, an integer $M \geqslant 0$ and $\rho > 0$. Pick a starting point x_0, an initial positive definite matrix B_0, an initial trust region radius $\Delta_0 \in (0, \Delta_{\max})$, and a positive penalty weight vector $r_0 \in \mathbf{R}^m$. Let $t(0) = 0$ and set $k = 0$.

(2) Calculate f_k, g_k, c_k, and A_k. Make a QR decomposition of A_k to get Y_k, Z_k, and R_k.

(3) If $\|c_k\| + \|Z_k^T g_k\| \leqslant \varepsilon$, stop. Otherwise, go to the next step.

(4) Compute the multiplier

$$\lambda_k = R_k^{-1} Y_k^T g_k, \tag{2.9}$$

and solve the vertical subproblem (2.4) and the horizontal subproblem (2.5) to the solutions p_k^y and p_k^z, respectively. Then compute p_k given in (1.7) and hence obtain the following equation

$$R_k^T d_k = - c(x_k + p_k) + c_k + R_k^T p_k^y \tag{2.10}$$

to get the solution d_k.

(5) Let

$$r_{(k+1)_i} = \begin{cases} r_{k_i}, & \text{if } r_{k_i} \geqslant \Psi_i^k + \rho, \\ \max\{r_{k_i}, \Psi_i^k\} + \rho, & \text{otherwise}, \end{cases} \tag{2.11}$$

where $\Psi_i^k = \max\{|\lambda_{k_i}| + |((p_k^z)^T D_k R_k^{-T})_i|, \|R_k^{-1}\| |(Y_k^T g_k + D_k^T p_k^z)_i|\}$ and r_{k_i}, λ_{k_i} and $(a)_i$ are the i-th components of the vectors r_k, λ_k and a, respectively.

(6) Set $\varphi_k(x_{l(k)}) = \max\limits_{0\leqslant j\leqslant t(k)} \varphi_k(x_{k-j})$. Test the line search condition

$$\varphi_k(x_k + \alpha_k(p_k + Y_k d_k)) \leqslant \varphi_k(x_{l(k)}) + \eta\alpha_k D\varphi_k(x_k; p_k), \tag{2.12}$$

where $D\varphi_k(x_k;p_k)$ is the directional derivative of the merit function φ_k in the direction p_k.

(7) If the line search condition (2.12) is not satisfied, choose a new $\alpha_k \to \alpha\alpha_k$ and go to step 6; otherwise set

$$x_{k+1} = x_k + \alpha_k(p_k + Y_k d_k). \tag{2.13}$$

(8) Calculate the predicted reduction,

$$\text{Pred}_k(p_k) = -g_k^T p_k - \frac{1}{2}(p_k^z)^T B_k p_k^z + \sum_{i=1}^m r_{(k+1)_i}\{|c_i(x_k)| - |\nabla c_i(x_k)^T p_k + c_i(x_k)|\}. \tag{2.14}$$

(9) Compute

$$\widehat{\text{Ared}}_k(p_k) = \varphi_k(x_{l(k)}) - \varphi_k(x_{k+1}), \tag{2.15}$$

where

$$\varphi_k(x) = f(x) + \sum_{i=1}^m r_{(k+1)_i}|c_i(x)|. \tag{2.16}$$

(10) Further, compute

$$\bar{\theta}_k = \frac{\widehat{\text{Ared}}_k(p_k)}{\text{Pred}_k(p_k)}. \tag{2.17}$$

If $\bar{\theta}_k < \bar{\eta}$, then set

$$\Delta_k \leftarrow \gamma_1 \Delta_k,$$

return to step 4. Otherwise, go to the next step.

(11) Set

$$\Delta_{k+1} = \begin{cases} \min\{\gamma_2\Delta_k, \Delta_{\max}\}, & \text{if } \bar{\theta}_k \geqslant \eta_2 \\ \Delta_k, & \text{if } \eta_2 > \bar{\theta}_k > \eta_1 \\ \gamma_1\Delta_k & \text{otherwise.} \end{cases} \tag{2.18}$$

(12) Choose $t(k+1) \leqslant \min\{t(k)+1, M\}$. Obtain B_{k+1} by updating B_k using the DFP formula or the BFGS formula if and only if update rule 1 holds. And obtain D_{k+1} by updating D_k using the Broyden formula if and only if update rule 2 holds. Set $k \leftarrow k+1$ and return to step 2.

Remark 1 An important feature of this algorithm is that we decide whether accept $x_k + p_k$ or $x_k + p_k + Y_k d_k$ as x_{k+1} by ratio $\bar{\theta}_k$. Therefore, it is allowable that $\varphi_k(x_{k+1}) > \varphi_k(x_k)$ and thus this is a non-monotonic sequential technique. It is easy to see that the usual monotone algorithms can be viewed as a special case of the proposed algorithms when $M = 0$.

3 Convergence Analysis

We make the following assumption in this section.

Assumption A1 The sequence of $\{x_k\}$, $\{x_k + p_k\}$, and $\{x_k + p_k + Y_k d_k\}$ generated by the algorithms are contained in a compact set $\mathscr{X}$; $f(x)$ and $c(x)$ are twice continuously differentiable on $\mathscr{X}$; the matrix $A(x)$ has full column rank over $\mathscr{X}$; thus the matrix $R(x) \in \mathbf{R}^{m\times m}$ in (1.2) and its inverse $R(x)^{-1}$ are defined and continuous on $\mathscr{X}$; $\{B_k\} \subset \mathbf{R}^{m\times m}$ are bounded symmetric matrices, and $\{D_k\} \subset \mathbf{R}^{m\times(n-m)}$ are bounded matrices.

According to the assumption, there exist some positive constants τ, w, b and b' such that

$$\zeta_k \stackrel{\text{def}}{=} \|R_k^{-1}\| \leqslant \tau, \quad b_k \stackrel{\text{def}}{=} \|B_k\| \leqslant b, \quad \|D_k\| \leqslant b', \quad \|W(x_k,\lambda_k)\| \leqslant w, \ \forall \ k. \tag{3.1}$$

The following two properties have been proved (see Lemma 3.1 in [8]and Lemma 3.5 in [7]).

Lemma 3.1 Under the Assumption A1, the correction vector d_k given in (2.10) satisfies

$$\|d_k\| = O(\|p_k\|^2). \tag{3.2}$$

Lemma 3.2 For any k, the condition

$$\sum_{i=1}^{m} r_{(k+1)_i}\{|c_i(x_k)| - |\nabla c_i(x_k)^T p_k + c_i(x_k)|\} - g_k^T Y_k p_k^y + (p_k^z)^T D_k p_k^y \geqslant \rho \min\{\|c_k\|, \frac{\mu_1\Delta_k}{\zeta_k}\}. \tag{3.3}$$

Let p_k^z denote the solution of the horizontal subproblem (2.5). It is easy to obtain

$$-(Z_k^T g_k + D_k p_k^y)^T p_k^z - \frac{1}{2}(p_k^z)^T B_k p_k^z \geqslant \frac{1}{2}\|Z_k^T g_k + D_k p_k^y\| \min\{\mu_2\Delta_k, \frac{\|Z_k^T g_k + D_k p_k^y\|}{\|B_k\|}\}. \tag{3.4}$$

By the definition of $\mathrm{Pred}_k(p_k)$ and the update formula (2.10) of penalty parameter, we get from (3.3) and (3.4) that

$$\mathrm{Pred}_k(p_k) \geqslant \frac{1}{2}\|Z_k^T g_k + D_k p_k^y\| \min\{\mu_2\Delta_k, \frac{\|Z_k^T g_k + D_k p_k^y\|}{\|B_k\|}\} + \rho \min\{\|c_k\|, \frac{\mu_1\Delta_k}{\zeta_k}\}. \tag{3.5}$$

Similar to Lemma 3.4 and Corollary 3.5 in [6], we can also obtain that

Lemma 3.3 There exists a positive constant l such that for all k

$$|\mathrm{Ared}_k(p_k) - \mathrm{Pred}_k(p_k)| \leqslant l\|p_k\|^2 \leqslant l(\mu_1^2 + \mu_2^2)\Delta_k^2. \tag{3.6}$$

Lemma 3.4 Let p_k be generated by the proposed algorithm. Then

$$D\varphi_k(x_k; p_k) \leqslant -\frac{1}{4}\|Z_k^T g_k + D_k p_k^y\| \min\{\mu_2\Delta_k, \frac{\|Z_k^T g_k + D_k p_k^y\|}{\|B_k\|}\} - \rho \min\{\|c_k\|, \frac{\mu_1}{\zeta_k}\} \tag{3.7}$$

Proof From $c_i(x_k + \tau p_k) = c_i(x_k) + \nabla c_i(x_i)^T p_k + O(\|p_k\|^2)$, $i = 1,2,\cdots,m$ we get

$$\lim_{\tau\to 0} \frac{|c_i(x_k + \tau p_k)| - |c_i(x_k)|}{\tau} = - |\nabla c_i(x_k)^T p_k|, \ i = 1,2,\cdots,m.$$

Consider the vertical subproblem (2.4). Set $\bar{p}_k^y = - R_k^T c_k$.

(1) If $\|\bar{p}_k^y\| \leqslant \mu_1\Delta_k$, and hence $c_k + R_k^T \bar{p}_k^y = 0$, then $\bar{p}_k^y$ is a solution of the subproblem (2.4). So $p_k^y = \bar{p}_k^y$ and we have $|\nabla c_i(x_k)^T p_k| = |c_i(x_k)|$.

(2) If $\|\bar{p}_k^y\| > \mu_1\Delta_k$, then $\frac{\mu_1\Delta_k \bar{p}_k^y}{\|\bar{p}_k^y\|}$ is a feasible solution of the subproblem (2.4). We have

$$\begin{aligned} - |\nabla c_i(x_k)^T p_k| &\leqslant - |c_i(x_k)| + |\nabla c_i(x_k)^T p_k + c_i(x_k)| \\ &\leqslant - \frac{\mu_1\Delta_k}{\|R_k^{-1}\|\|c_k\|}|c_i(x_k)|, \quad i = 1,2,\cdots,m. \end{aligned} \tag{3.8}$$

Since p_k^z is the horizontal subproblem (2.5), it is not difficult to obtain

$$(Z_k^T g_k + D_k p_k^y)^T p_k^z \leqslant -\frac{1}{4}\|Z_k^T g_k + D_k p_k^y\| \min\{\mu_2\Delta_k, \frac{\|Z_k^T g_k + D_k p_k^y\|}{\|B_k\|}\}.$$

The directional derivative $D\varphi_k(x_k;p_k)$ of the merit function $\varphi_k(x)$ at the point x_k along the direction p_k exists and

$$D\varphi_k(x_k;p_k) = g_k^T p_k - \sum_{i=1}^{m} r_{(k+1)_i} |\nabla c_i(x_k)^T p_k|$$

$$= (Z_k^T g_k + D_k p_k^y)^T p_k^z + g_k^T Y_k p_k^y - \sum_{i=1}^{m} r_{(k+1)_i} |\nabla c_i(x_k)^T p_k| - (p_k^y)^T D_k p_k^z. \tag{3.9}$$

(i) $p_k^y = \bar{p}_k^y$, by the definition of the multiplier λ_k in (2.9), we have that

$$g_k^T Y_k p_k^y = \lambda_k R_k^T p_k^y \leqslant \|\lambda_k\| \cdot \|R_k^T p_k^y\| \leqslant \|\lambda_k\| \cdot \|c_k\|,$$

and

$$|(D_k p_k^y) p_k^z| \leqslant \|(p_k^z)^T D_k R_k^{-T}\|_1 \cdot \|c_k\|.$$

Hence, the r_{k+1} updated in (2.11) implies that the conclusion of the lemma is true.

(ii) $\|\bar{p}_k^y\| > \mu_1 \Delta_k$. Further

$$|g_k^T Y_k p_k^y + (D_k p_k^y)^T p_k^z| \leqslant \|Y_k^T g_k + (p_k^z)^T D_k\|_1 \cdot \|p_k^y\| \leqslant \mu_1 \|Y_k^T g_k + (p_k^z)^T D_k\|_1 \Delta_k.$$

which means that the conclusion of the lemma is also true.

As space is limited, we only set out the main results and omit the resemble proofs. Similar to the proofs of Lemma 3.7 and Lemma 3.8 in [6], we can also obtain the following lemma and theorem, respectively.

Lemma 3.5 Suppose that assumption A1 holds. If there exists $\varepsilon > 0$ such that

$$\|c(x_k)\| + \|Z_k^T g_k\| \geqslant \varepsilon, \tag{3.10}$$

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

$$\Delta_k \geqslant \alpha, \ \forall \ k. \tag{3.11}$$

Theorem 3.6 Suppose that assumption A1 holds. Let $\{x_k\} \subset \mathbf{R}^n$ be a sequence generated by the proposed algorithm. Then

$$\liminf_{k \to \infty} \{\|c(x_k)\| + \|Z_k^T g_k\|\} = 0. \tag{3.12}$$

4 Local Convergence Rates

In order to get the local convergence rate of the proposed algorithm, more assumptions are needs

Assumption A2 x_* is a K-K-T point of the problem (1.1), that is, there is a vector $\lambda_* \in \mathbf{R}^m$ such that

$$c_* = 0 \ \text{and} \ g_* - A_* \lambda_* = 0.$$

Assumption A3 There exists a constant $\tau > 0$ such that

$$d^T (Z_*^T W_* Z_*) d \geqslant \tau \|d\|^2, \ \forall \ d \in \mathbf{R}^{n-m}.$$

Assumption A4

$$\lim_{k \to \infty} \frac{\|(B_k - H_*) Z_k^T (x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0, \tag{4.1}$$

$$\lim_{k \to \infty} \frac{\|(D_k - G_*) Z_k^T (x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0. \tag{4.2}$$

Lemma 4.1 Assume that Assumptions A2～A4 hold. If $x_{k+1} = x_k + p_k + Y_k d_k$, then

$$|\mathrm{Pred}_k(p_k) - \mathrm{Ared}_k(p_k)| = o(\|p_k^z\|^2 + \|p_k^y\|). \tag{4.3}$$

Proof Similar to the proof of Lemma 4.2 in [8]. So (4.3) is true.

Lemma 4.2 Assume that Assumptions A2～A4 hold. Then

$$\mathrm{Pred}_k(p_k) \geqslant \frac{\tau}{4} \|p_k^z\|^2 + \rho \min\{\|p_k^y\|, \frac{\|p_k^y\|}{\tau}\} + o(\|p_k^z\| \cdot \|p_k\|); \tag{4.4}$$

$$D\varphi_k(x_k; p_k) \leqslant -\frac{\tau}{8} \|p_k\|^2 - \rho \min\{\|p_k^y\|, \frac{\|p_k^y\|}{\tau}\} + o(\|p_k^z\| \cdot \|p_k\|). \tag{4.5}$$

Proof The solution p_k^z in the subproblem (2.5) satisfies that there exists a $\mu_k \geqslant 0$ such that

$$(B_k + \mu_k I) p_k^z = -(Z_k^T g_k + D_k p_k^y)$$

and from Assumption A4, for large k

$$(Z_k^T g_k + D_k p_k^y)^T p_k^z + \frac{1}{2}(p_k^z)^T B_k p_k^z \leqslant -\frac{1}{2}(p_k^z)^T B_k p_k^z - \mu_k \|p_k^z\|^2 \leqslant -\frac{\tau}{4}\|p_k^z\|^2 + o(\|p_k^z\| \cdot \|p_k\|). \quad (4.6)$$

Further,

$$(Z_k^T g_k)^T p_k^z = (Z_k^T g_k + D_k p_k^y)^T p_k^z - (p_k^y)^T D_k p_k^z$$

$$\leqslant -(p_k^z)^T B_k p_k^z - (p_k^y)^T D_k p_k^z \leqslant -\frac{\tau}{2}\|p_k^z\|^2 + O(\|p_k^z\| \cdot \|p_k^y\|).$$

As we know that similar to the proof of Lemma 3.4,

$$\mathrm{Pred}_k(p_k) \geqslant \frac{\tau}{4}\|p_k^z\|^2 + \rho \min\{\|c_k\|, \frac{\mu_1 \Delta_k}{2}\} + o(\|p_k^z\| \cdot \|p_k\|), \quad (4.7)$$

$$D\varphi_k(x_k; p_k) \leqslant -\frac{\tau}{2}\|p_k^z\|^2 - \rho \min\{\|c_k\|, \frac{\mu_1 \Delta_k}{2}\} + o(\|p_k^z\| \cdot \|p_k\|) + o(\|p_k^y\|). \quad (4.8)$$

From the subproblem (2.4), we know

$$\|p_k^y\| \leqslant \|R_k^{-1}\| [\|c_k\| + \|R_k^T p_k^y + c_k\|] \leqslant \bar{\tau}\|c_k\|.$$

The above inequality together with (4.7) and (4.8) imply that the lemma holds.

Theorem 4.3 Assume that Assumptions A2~A4 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Then

$$\lim_{k\to\infty}\{\|Z(x_k)^T g(x_k)\| + \|c(x_k)\|\} = 0. \quad (4.9)$$

Proof Similar to the proof of Theorem 4.2 in [6], we have that (4.9) holds.

Remark 2 Theorem 3.6 indicates that at least one limit point of $\{x_k\}$ is a stationary point. In this, Theorem 4.3 extends to a stronger result, that is, all the limit points are K-K-T points.

Now, we discuss the convergence rate for the proposed algorithm when B_k is positive definite.

Theorem 4.4 Under the Assumptions A2~A4, the proposed algorithm is at least two-step Q-superlinearly convergent, i. e.,

$$\lim_{k\to\infty}\frac{\|x_{k+1} - x_*\|}{\|x_{k-1} - x_*\|} = 0. \quad (4.10)$$

Furthermore, if at least one of the matrices B_k and D_k is updated as GURWITZ's alternative strategy given in [6] at each iteration, then the rate of convergence is one-step Q-superlinear, i. e.,

$$\lim_{k\to\infty}\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \quad (4.11)$$

Proof It is similar to the proof of Theorem 4.4 in [6]. So (4.10) and (4.11) hold.

References:

[1] CELIS M R, DENNIS J E, TAPIA R A. A Trust Region Strategy for Nonlinear Equality Constrained Optimization[J]. Numerical Optimization, SIAM Philadelphia, 1985, 71-82.

[2] GURWITZ C B. Local Convergence of Two-piece Update of a Projected Hessian Matrix[J]. SIAM J Optim, 1994(4): 461-485.

[3] NOCEDAL J, OVERTON M L. Projected Hessian Updating Algorithms for Nonlinearly Constrained Optimization[J]. SIAM J Numer Anal, 1985, 22: 821-850.

[4] NOCEDAL J, YUAN Y. Combining Trust Region and Line Search Techniques[J]. Advances in Nonlinear

Programming, Kluwer, Dordvecht, 1998, 153-175.

[5] POWELL M J D, YUAN Y. A Recursive Quadratic Programming Algorithm that Uses Differentiable Exact Penalty Functions[J]. Mathematical Programming, 1986, 35: 265-278.

[6] ZHU D. Trust Region Methods with Nonmonotone Backtracking Technique for Constrained Optimization Problems[J]. Optimization, 1999, 46: 81-113.

[7] ZHU D. A Two-piece Update of Projected Hessian Algorithm with Nonmontonic Trust Region Method for Constrained Optimization[J]. To appear in Acta Mathematicate Applicate Sinica, 2002.

[8] ZHANG J, ZHU D. A Trust Region Typed Dogleg Method for Nonlinear Optimization[J]. Optimization, 1990, 21: 543-557.

约束优化的两块校正非单调回代法

顾益明，朱德通
（上海师范大学 数理信息学院，上海 200234）

摘要：提出一种二块校正既约 Hessian 方法的非单调信赖域回代算法来解决约束优化问题. 一般采用二块校正的双边既约 Hesse 阵方法代替完全 Hesse 阵方法处理大规模问题. 为了获得算法的整体收敛性，引入非光滑的 l_1 罚函数作为势函数. 在每次回代中不必使罚函数都单调递减，以使能克服高度非线性情况下的峡谷状态，同时采用二阶校正步计算能避免 Maratos 效应. 只要每一步迭代至少运用两种校正规则之一，算法就能保持一步局部 Q-超线性收敛速率.

关键词：非单调技术；二块校正；超线性收敛