第31卷第1期 上海师范大学学报(自然科学版) Vol. 31, No. 1
2002年3月 Journal of Shanghai Teachers University (Natural Sciences) Mar. 2002

Projected Quasi-Newton Algorithm with Nonmonotone Trust Region for Equality Constrained Optimization Problems

QIAN Chun-qing, FU Jun, ZHU De-tong

(Mathematical and Science College, Shanghai Teachers University, Shanghai 200234, China)

Abstract: This paper gives a decompose of projected quasi-Newton algorithm in association with nonmonotone trust region for solving nonlinear equality constrained optimization problems. The proposed method is globally convergent even if conditions are mild. In order to assure local superlinear rate and obtain other convergence properties, a second order correction step which brings the iterates closer to the feasible set is described. The correction step allows to prove that the proposed algorithm is also locally superlinear convergent.

Key words: Trust region; nonmonotone technique, constrained optimization

CLC number: O241.7 **Document code**: A **Article ID**: 1000-5137(2002)01-0024-07

1 Introduction

In this paper we analyze successive quadratic programming methodsfor solving the equality constrained optimization problem.

$$\min \quad f(x) \qquad \text{subject to } c(x) = 0. \tag{1.1}$$

where f: $R^n \to R$ and c: $R^n \to R^m$ are smooth nonlinear function. There are quite a few articles proposing sequential quadratic programming methods (SQP). These methods generate a search direction at x_k by solving the quadratic programming:

$$\min \quad g_k^T d + \frac{1}{2} d^T B_k d \qquad \text{subject to } c_k + A_k^T d = 0. \tag{1.2}$$

where g_k is the gradient of f at x_k. $A_k = A(x_k) = [\nabla c_1(x_k), \ldots, \nabla c_m(x_k)]$ is the $n \times m$ Jacobi matrix of $c(x)$ at x_k and L is the Lagrangian function defined for by

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda^i c_i(x). \tag{1.3}$$

Let B_k is a matrix that approximates the Hessian of the Lagrangian function.

$$W_k = W(x_k, \lambda_k) = \nabla^{2x} L(x_k, \lambda_k) = \nabla^2 f(x_k) - \sum_{i=1}^{m} \lambda_k^i \nabla^2 c_i(x_k). \tag{1.4}$$

Received date: 2001-05-16

Foundation item: Supported by Science Foundatia Grant of Shanghai Teachers University (DQ17).

Biography: QIAN Chun-qing (1965-), male, lecture, Mathematical and Science College, Shanghai Teachers University.

COLEMAN and CONN[3], NOCEDAL and OVERTON[6] proposed separately similar quasi-Newton methods using approximate reduced Hessian. By constructing an orthonormal basis ($Z_k \in R^{n\times(n-m)}$) for the tangent space of the constrains at the current point x_k, they use the fact that the matrix $Z_k^TH_kZ_k \in R^{(n-m)\times(n-m)}$ is positive definite. However, recent reports indicate that it might be difficult to find a basis Z_k which changes continuously with k (see, for example [2]). Two basic approaches, namely the line search and the trust region, have been developed to ensure global convergence towards local minima (see [1] and [4]). However, all versions of these approaches enforce a monotone decrease of acertain merit function at each step. Recently, the nonmonotonic line search technique for solving unconstrained optimization is proposed by GRIPPO et al in [4]. The nonmonotonic idea motivates to further study the projected quasi-Newton methods with trust region.

In this paper, we describe and analyze the projected quasi-Newton methods associated with nonmonotone trust region for problem (1.1), we suggest to solve two subproblems rather than solving the continuous Z_k with k. Section 2 presents the decompose of projected quasi-Newton method in association with nonmonotone trust region technique in detail. In section 3, we prove the global convergence properties of the proposed algorithm, while in section 4, we devote the local convergence rate of the algorithm.

2 Algorithm

We first introduce some standard notations for this paper.

Let $\|\cdot\|$ be the Euclidean norm on R^n. Let $f: R^n \to R$ be twice continuously differentiable, with gradient $g: R^n \to R$ and Hessian matrix $\nabla^2 f$. Let $c: R^n \to R^m$ be the vector of twice continuouslydifferentiable constraint functions $c_i(x)$, for $i = 1, \ldots, m$, with the Hessian matrix of $c_i(x)$ denoted by $\nabla^2 c_i(x)$, and let $A(x)$ be the n by m matrix consisting of the column vectors $a_i(x)$, for $i = 1, \ldots, m$. Denote the first-order Lagrange multiplier estimates by $\lambda(x) = [A(x)^TA(x)]^{-1}A(x)^Tg(x)$.

The projection matrix $P(x) = I - A(x)[A(x)^TA(x)]^{-1}A(x)^T$.

In each iteration, we first solve the subproblem

$$(S_k) \quad \min g_k^Td + \frac{1}{2}d^TB_kd \quad \text{subject to } A_k^Td = 0, \ \|d\| \leqslant \Delta_k.$$

where Δ_k is a trust region radius. Let d_k be the solution of the subproblem (S_k), let s_k be the solution of the second subproblem (P_k).

$$(P_k) \quad \min\|c_k + A_k^Ts\| \quad \text{subject to} \quad \|s\| \leqslant \Delta_k.$$

In order to decide whether we should take $x_{k+1} = x_k + p_k$, where $p_k = d_k + s_k$, we adapt the l_1-norm nondifferentiable exact penalty function $\varphi(x, \rho) = f(x) + \rho\|c(x)\|_1$, set

$$\rho_k = \begin{cases} \rho_{k-1}, & \text{if } \rho_{k-1} \geqslant \max\{\|\lambda_k\|_\infty, (\|g_k\| + \tau)\cdot\|A_k(A_k^TA_k)^{-1}\|\} + \tau \\ \max\{\rho_{k-1}, \|\lambda_k\|_\infty, \dfrac{\|g_k\| + \tau}{\alpha}\cdot\|A_k(A_k^TA_k)^{-1}\|\} + \tau, & \text{otherwise.} \end{cases} \tag{2.1}$$

where τ is a positive constant and constant α satisfies $\|c(x)\|_1 \geqslant \alpha\|c(x)\|$.

The actual reduction in the penalty function in going from x_k to x_{k+1} is thus given by

$$\text{Ared}_k(p_k) = f_k - f(x_k + p_k) + \rho_k\{\|c_k\|_1 - \|c(x_k + p_k)\|_1\}. \tag{2.2}$$

Using these same approximations we can compute a prediction of the reduction

$$\text{Pred}_k(p_k) = -g_k^Td_k - \frac{1}{2}d_k^TB_kd_k - g_k^Ts_k + \rho_k\{\|c_k\|_1 - \|A_k^Ts_k + c_k\|_1\}. \tag{2.3}$$

Relaxing the acceptability condition on p_k, we set

$$\varphi(x_{l(k)}, \rho_{l(k)}) = \max_{0 \leqslant j \leqslant m(k)} \{\varphi(x_{k-j}, \rho_{k-j})\}, \tag{2.4}$$

$$\overline{\mathrm{Ared}}_k(p_k) = \varphi(x_{l(k)}, \rho_{l(k)}) - \varphi(x_k + p_k, \rho_k). \tag{2.5}$$

where $m(0) = 0$ and $0 \leqslant m(k) \leqslant \min\{m(k-1)+1, M\}$, $k \geqslant 1$, and M is a nonnegative integer. It is clearly to see $k - m(k) \leqslant l(k) \leqslant k$.

Algorithm

Given $\mu \in (0,1)$, $\eta \in (\mu, 1)$, γ_0, γ_1 and γ_2 which must satisfy $0 < \gamma_0 \leqslant \gamma_1 < 1 < \gamma_2$.

Step 0: the starting point x_0, f_0 and g_0 are given, as well as an initial trust region radius $\Delta_0 > 0$ and B_0, an initial approximation to the Hessian at the starting point. Set $k = 0$.

Step 1: Solving the two subproblems (S_k) and (P_k) respectively, obtain steps d_k and s_k.

Step 2: If $d_k = 0$ and $s_k = 0$, then stop; else go to next step.

Step 3: Compute $p_k = d_k + s_k$ and ρ_k given by (2.1), Compute $\mathrm{Ared}_k(p_k)$, $\overline{\mathrm{Ared}}_k(p_k)$ and $\mathrm{Pred}_k(p_k)$ given by (2.2), (2.5) and (2.3), respectively.

Further, set

$$\eta_k = \frac{\mathrm{Ared}_k(p_k)}{\mathrm{Pred}_k(p_k)}, \text{ and } \xi_k = \frac{\overline{\mathrm{Ared}}_k(p_k)}{\mathrm{Pred}_k(p_k)}. \tag{2.6}$$

Step 4: In the case, $\xi_k \geqslant \mu$, The iteration is said to be successful: set $x_{k+1} = x_k + p_k$, and choose $\Delta_{k+1} \in [\Delta_k, \gamma_2\Delta_k]$, if $\eta_k \geqslant \eta$; or $\Delta_{k+1} \in [\gamma_1\Delta_k, \Delta_k)$, if $\eta_k < \eta$, Otherwise, i e $\xi_k < \mu$, The iteration is said to be unsuccessful, let $x_k \leftarrow x_k$ and $\Delta_k \in [\gamma_0\Delta_k, \gamma_1\Delta_k)$, Go to step 1.

Step 5: Using the quasi-Newton updating formulae to update B_k for B_{k+1}, set $m(k+1) \leqslant \min\{m(k)+1, M\}$ and $k \leftarrow k+1$, return to step 1.

This is a theoretical algorithm and many details should be added to specify a practical numerical procedure. In particular, as to the stopping criterion of the algorithm, in practical use we can check the condition $\|d_k\| + \|s_k\| \leqslant \varepsilon$ for a small convergent tolerance $\varepsilon > 0$.

3 Global convergence

We make the following assumptions in the section.

Assumption 1 Sequence $\{x_k\}$ generated by the algorithm is contained in a compact set X on R^n.

Assumption 2 Matrix $A(x)^T = \nabla c(x)$ has full column-rank on X. There are constants $\beta' \geqslant \beta > 0$ such that $\beta \leqslant \|A(x)^T A(x)\| \leqslant \beta'$, $\forall x \in X$.

Lemma 3.1 d_k is a solution of subproblem (S_k) if and only if there exist $\mu_k \geqslant 0$, $\lambda_k \in R^m$ such that

$$(B_k + \mu_k I)d_k = -g_k + A_k\lambda_k$$

$$A_k^T d_k = 0, \quad \mu_k(\Delta_k - \|d_k\|) = 0, \tag{3.1}$$

holds and $B_k + \mu_k I$ is positive semidefinite in $N(A_k^T)$.

Lemma 3.2 Assumptions 1~2 hold. If $d_k = 0$ and $s_k = 0$, then x_k is a Kuhn-Tucker point of the Problem (1.1).

Lemma 3.3 If x_k is a Kuhn-Tucker point of the Problem (1.1), and B_k is positive definite in $N(A_k^T)$, then$d_k = 0$ is the solution of (S_k) and $s_k = 0$ is the solution of (P_k).

Lemma 3.4 Let $b_k = \max\{\|P_i B_i P_i\| \mid i = 1, \ldots, k\} + 1$, we have that

$$-g_k^T d_k - \frac{1}{2} d_k^T B_k d_k \geqslant \frac{1}{2}\|P_k g_k\| \min\{\Delta_k, \frac{\|P_k g_k\|}{b_k}\}. \tag{3.2}$$

Lemma 3.5 If ρ_k is updating by (2.1), we have that

$$\rho_k\{\|c_k\| - \|A_k^T s_k + c_k\|\} - g_k^T s_k \geqslant \tau \min\{\alpha\|c_k\|, \Delta_k\}, \tag{3.3}$$

where s_k solves (P_k) and constant α satisfies $\|c(x)\|_1 \geqslant \alpha\|c(x)\|$.

Proof Set $s_k^* = -A_k(A_k^T A_k)^{-1}c_k$, we consider two cases:

(1) If $\|s_k^*\| \leqslant \Delta_k$, then $s_k = s_k^*$ is the solution of (P_k) for $A_k^T s_k + c_k = 0$. We have

$$\rho_k\{\|c_k\|_1 - \|A_k^T s_k + c_k\|_1\} - g_k^T s_k \geqslant \rho_k \|c_k\|_1 + \lambda_k^T c_k \geqslant \tau\alpha\|c_k\|. \tag{3.4}$$

(2) If $\|s_k^*\| > \Delta_k$, then $(\frac{\Delta_k}{\|s_k^*\|})\|s_k^*\|$ is the feasible solution of the subproblem (P_k). So

$$\|c_k\|_1 - \|c_k + A_k^T s_k\|_1 \geqslant \|c_k\|_1 - \|c_k + (\frac{\Delta_k}{\|s_k^*\|})A_k^T s_k^*\|_1 \geqslant \frac{\alpha\Delta_k}{\|A_k(A_k^T A_k)^{-1}\|}. \tag{3.5}$$

Therefore, we have that, using (2.1)

$$\rho_k\{\|c_k\|_1 - \|A_k^T s_k + c_k\|_1\} - g_k^T s_k \geqslant \frac{\alpha\rho_k\Delta_k}{\|A_k(A_k^T A_k)^{-1}\|} - \|g_k\|\Delta_k \geqslant \tau\Delta_k. \tag{3.6}$$

According to (3.4) and (3.6), (3.3) is true. □

Combining the above two lemmas, we now state our main result on the model decrease at iteration k, for $\mathrm{Pred}_k(p_k)$.

Lemma 3.6 Under assumptions A1~A2, we have that

$$\mathrm{Pred}_k(p_k) \geqslant \frac{1}{2}\|P_k g_k\| min\{\Delta_k, \frac{\|P_k g_k\|}{b_k}\} + \tau\min\{\alpha\|c_k\|, \Delta_k\}. \tag{3.7}$$

Proof Combining (3.2) and (3.3), (2.3) implies that (3.7) is true. □

Lemma 3.7 Suppose that the assumptions A1~A2 hold. There exists $\bar{b}_k > 0$ such that

$$\bar{b}_k = \max\{\|\nabla^2 f(\xi_{0,j})\|, \|\nabla^2 c_i(\xi_{i,j})\| \mid i = 1, 2, \dots, m,\ j = 1, 2, \dots, k\} + 1$$

where $\xi_{i,j}$, $i = 0, 1, \dots, m$ belongs to the line segment from x_j to $x_j + p_j$ contained in X. Then we have that

$$|\mathrm{Ared}_k(p_k) - \mathrm{Pred}_k(p_k)| \leqslant [b_k + (1 + m\rho_k)\bar{b}_k]\Delta_k^2. \tag{3.8}$$

Proof By the definitions of Ared_k and Pred_k, we have that

$$|\mathrm{Ared}_k(p_k) - \mathrm{Pred}_k(p_k)| \leqslant \frac{1}{2}(|d_k^T B_k d_k| + |p_k^T \nabla^2 f(\xi_{0,k}) p_k|) +$$
$$\rho_k\left|\sum_{i=1}^{m}|c_i(x_k) + \nabla c_i(x_k)^T p_k| + \frac{1}{2}\sum_{i=1}^{m}|p_k^T \nabla^2 c_i(\xi_{i,k}) p_k| - \|A_k^T s_k + c_k\|_1\right| \leqslant$$
$$[b_k + (1 + m\rho_k)\bar{b}_k]\Delta_k^2. \tag{3.9}$$

□

Because assumptions A1~A2 hold, there exists $\bar{\rho} > 0$ such that

$$\rho_k = \bar{\rho} \quad \text{for large enough } k. \tag{3.10}$$

Assumption 3 There exists $b > 0$ such that $\max\{b_k, \bar{b}_k\} \leqslant b$, $\forall\ k$. And $\nabla^2 f$ and $\nabla^2 c_i$, $i = 1, 2, \dots, m$ are Lipschitz continuous matrix functions on the set X.

From (3.10) and Assumption A3, we have that there exists $L > 0$ such that (3.8) can be rewritten as follows $|\mathrm{Ared}_k(p_k) - \mathrm{Pred}_k(p_k)| \leqslant L\Delta_k^2$.

Theorem 3.8 Assume that assumptions A1~A3 hold. Let sequence $\{x_k\}$ be generated by the algorithm, then

$$\liminf_{k\to\infty}\{\|P_k g_k\| + \|c_k\|\} = 0. \tag{3.11}$$

Proof According to (3.7), we have

$$\varphi(x_k + p_k, \rho_k) - \varphi(x_{l(k)}, \rho_{l(k)}) \leqslant -\mu\mathrm{Pred}_k(P_k) \leqslant -\frac{\mu}{2}\|P_k g_k\|\min\{\Delta_k, \frac{\|P_k g_k\|}{b}\} - \mu\tau\min\{\Delta_k, \alpha\|c_k\|\}. \tag{3.12}$$

From (3.10), there is $\bar{\rho} > 0$ such that $\rho_k \equiv \bar{\rho}$, for large enough k.

Taking into account that $m(k+1) \leqslant m(k) + 1$, we have

$$\varphi(x_{l(k+1)},\rho_{l(k+1)})=\max_{0\leqslant j\leqslant m(k+1)}\{\varphi(x_{k+1-j},\rho_{k+1-j})\}\leqslant\max_{0\leqslant j\leqslant m(k)+1}\{\varphi(x_{k+1-j},\rho_{k+1-j})\}=\varphi(x_{l(k)},\rho_{l(k)}) \quad (3.13)$$

for large enough k. Moreover, we can obtain that for all $k>M$ and large enough k,

$$\varphi(x_{l(k)},\rho_{l(k)})\leqslant\varphi(x_{l(l(k)-1)},\rho_{l(l(k)-1)})-$$

$$\frac{\mu}{2}\|P_{l(k)-1}g_{l(k)-1}\|\min\{\Delta_{l(k)-1},\frac{\|P_{l(k)-1}g_{l(k)-1}\|}{b}\}-\mu\tau\min\{\Delta_{l(k)-1},a\|c_{l(k)-1}\|\}. \quad (3.14)$$

If the conclusion (3.11) is not true, there exists an $\varepsilon>0$ such that $\|c_k\|+\|P_kg_k\|\geqslant2\varepsilon$, $k=1,2,\ldots$, $\|c_k\|+\|P_kg_k\|\geqslant2\varepsilon$, $k=1,2,\ldots$, and hence either $\|c_k\|\geqslant\varepsilon$, $k=1,2,\ldots$ or $\|P_kg_k\|\geqslant$, $k=1,2,\ldots$

Therefore, we have that (3.14) can be written as follows, either

$$\varphi(x_{l(k)},\rho_{l(k)})=\varphi(x_{l(l(k)-1)},\rho_{l(l(k)-1)})-\frac{\mu}{2}\varepsilon\min\{\Delta_{l(k)-1},\frac{\varepsilon}{b}\}. \quad (3.15)$$

or

$$\varphi(x_{l(k)},\rho_{l(k)})\leqslant\varphi(x_{l(l(k)-1)},\rho_{l(l(k)-1)})-\mu\tau\min\{a\varepsilon,\ \Delta_{l(k)-1}\}. \quad (3.16)$$

Since $\{\varphi(x_{l(k)},\rho_{l(k)})\}$ is nonincreasing for large enough k, so $\lim_{k\to\infty}\Delta_{l(k)-1}=0$. For $k>M$, as $k-M\leqslant k-m(k)\leqslant l(k)\leqslant k$ by updating formula of Δ_k, we have that for any j, $0\leqslant\Delta_{k+j}\leqslant\gamma_2^j\Delta_k$, then $0\leqslant\Delta_k\leqslant\gamma_2^{M+1}\Delta_{l(k)-1}$, which means $\lim_{k\to\infty}\Delta_k=0$. On the other hand, if $\|c_k\|\geqslant\varepsilon$, from lemma 3.6 when $\Delta_k\leqslant a\varepsilon$, then $\mathrm{Pred}_k(p_k)\geqslant\tau\Delta_k$; if $\|P_kg_k\|\geqslant\varepsilon$, also from lemma 3.6, when $\frac{\varepsilon}{b}\geqslant\Delta_k$, we have $\mathrm{Pred}_k(p_k)\geqslant\frac{\varepsilon}{2}\Delta_k$. As above, we have that if $\Delta_k\leqslant\min\{a\varepsilon,\frac{\varepsilon}{b}\}$, then $\mathrm{Pred}_k(P_k)\geqslant\bar\tau\Delta_k$, where $\bar\tau=\min\{\frac{\varepsilon}{2},\tau\}$. When $\Delta_k\to0$,

$$|\eta_k-1|=\frac{|\mathrm{Pred}_k(p_k)-\mathrm{Ared}_k(p_k)|}{|\mathrm{Pred}_k(p_k)|}\leqslant\frac{L\Delta_k^2}{\bar\tau\Delta_k}\to0. \quad (3.17)$$

This implies $\eta_k\to1$, i e for large enough k, $\eta_k\geqslant\eta$. It also implies that trust region radius will be bounded away from zero, which contradicts $\lim_{k\to\infty}\Delta_k=0$. □

4 Local Convergence

In the following paragraph, we further discuss the local convergence of the proposed algorithm. Further, we assume that

Assumption 4 x_* is a Kuhn-Tucker Point of problem (1.1), i e. there exists a vector $\lambda_*\in R^m$ such that $g_*-A_*\lambda_*=0$ and $c_*=0$.

Assumption 5 There exists τ_1 such that

$$d^TW_*d\geqslant\tau_1\|d\|^2,\ \text{when } A_*^Td=0. \quad (4.1)$$

where $W_*=W(x_*,\lambda_*)=\nabla^2f(x_*)-\sum_{i=1}^m\lambda_*^i\nabla^2c_i(x_*)$ is the Hessian matrix of the Lagrangian function of problem (1.1) at x_*.

Assumption 6

$$\lim_{k\to\infty}\frac{\|P_k(B_k-W_k)d_k\|}{\|d_k\|}=0. \quad (4.2)$$

From (4.2), it is easy to see that, since $P_kd_k=d_k$ we have $d_k^TB_kd_k=d_k^TW_kd_k+o(\|d_k\|^2)$.

We introduce a correction vector v_k to improve local convergence rate of the algorithm.

$$v_k=-A_k(A_k^TA_k)^{-1}[c(x_k+p_k)-c_k-A_k^Ts_k]. \quad (4.3)$$

Therefore, in the algorithm, we take $x_{k+1}=x_k+p_k+v_k$ instead of $x_{k+1}=x_k+p_k$, and revise the step of the algorithm.

Lemma 4.1 $\|v_k\|=O(\|p_k\|^2)$.

Corollary 4.2 $\|v_k\| = O(\Delta_k^2)$.

By lemma 4.1 or corollary 4.2, the algorithm which is revised by a correction step will also keep global convergence properties of the original algorithm in section 2. We needn't reprove here these because the proofs are similar to these of the original algorithm.

As explained in the last section, we also assume $\rho_k \equiv \bar{\rho}$ for all large enough k.

Lemma 4.3 Assume that assumptions A1～A6 hold, then

$$|\mathrm{Ared}_k(p_k + v_k) - \mathrm{Pred}_k(p_k)| = o(\|d_k\|^2 + \|s_k\|). \tag{4.4}$$

Lemma 4.4 Assume that assumptions A1～A6 hold. Then there exists $\xi > 0$ such that for large enough k,

$$\mathrm{Pred}_k(p_k) \geqslant \xi(\|d_k\|^2 + \|s_k\|). \tag{4.5}$$

Theorem 4.5 Under assumptions A1～A6. For any large enough k, we have

$$\eta_k \geqslant \eta. \tag{4.6}$$

Proof Similar to the proof of (3.14) in theorem 3.8, using (4.5), for large k, we have

$$\varphi(x_{l(k)}, \rho_{l(k)}) \leqslant \varphi(x_{l(l(k)-1)}, \rho_{l(l(k)-1)}) - \xi(\|d_{l(k)-1}\|^2 + \|s_{l(k)-1}\|). \tag{4.7}$$

By induction, similar to the proof of the theorem in [4], we have $\lim_{k\to\infty}\{\|d_k\|^2 + \|s_k\|\} = 0$. From lemma 4.4 and lemma 4.3, we get that

$$|\eta_k - 1| = \frac{|\mathrm{Pred}_k(p_k) - \mathrm{Ared}_k(p_k)|}{|\mathrm{Pred}_k(p_k)|} \to 0, \quad k \to +\infty. \tag{4.8}$$

So (4.6) is true. □

Theorem 4.6 Under assumptions A1～A6, we have

$$\lim_{k\to\infty}\{\|P_k g_k\| + \|c_k\|\} = 0. \tag{4.9}$$

Proof Assume a subsequence $\{m_i\}$ of this subsequence satisfies

$$\|P_{m_i} g_{m_i}\| + \|c_{m_i}\| \geqslant 2\varepsilon_1, \tag{4.10}$$

for an $\varepsilon_1 \in (0,1)$. Theorem 3.8 guarantees the existence of another subsequence $\{l_i\}$ such that $\|P_{l_i+1} g_{l_i+1}\| + \|c_{l_i+1}\| < \varepsilon(\|P_{m_i} g_{m_i}\| + \|c_{m_i}\|)$ and $\|P_k g_k\| + \|c_k\| > \varepsilon(\|P_{m_i} g_{m_i}\| + \|c_{m_i}\|)$, $m_i \leqslant k \leqslant l_i$ for any ε. So at least one of $\|c_k\| \geqslant \varepsilon_1\varepsilon$ and $\|P_k g_k\| \geqslant \varepsilon_1\varepsilon$ holds for $m_i \leqslant k \leqslant l_i$. Applying (3.6) and (4.6), we obtain that either $\mathrm{Ared}_k(p_k) \geqslant \frac{1}{2}\eta\varepsilon_1\varepsilon\min\{\frac{\varepsilon_1\varepsilon}{b}, \Delta_k\}$ or $\mathrm{Ared}_k(p_k) \geqslant \eta\tau\min\{\alpha\varepsilon_1\varepsilon, \Delta_k\}$. No matter whether cases hold, we have $\lim_{k\to\infty}\Delta_k = 0$ which implies $\mathrm{Ared}_k(p_k) \geqslant \min\{\frac{1}{2}\eta\varepsilon_1\varepsilon, \eta\tau\}\Delta_k = \bar{\tau}\Delta_k$, where let $\bar{\tau} = \min\{\frac{1}{2}\eta\varepsilon_1\varepsilon, \eta\tau\}$. We use the triangle inequality to show that

$$\|P_{m_i} g_{m_i}\| + \|c_{m_i}\| \leqslant (\|P_{m_i} g_{m_i} - P_{l_i+1} g_{l_i+1}\| + \|c_{m_i} - c_{l_i+1}\|) + \varepsilon(\|P_{m_i} g_{m_i}\| + \|c_{m_i}\|). \tag{4.11}$$

It follows from (4.11) that

$$(1-\varepsilon)(\|P_{m_i} g_{m_i}\| + \|c_{m_i}\|) \leqslant \|P_{m_i} g_{m_i} - P_{l_i+1} g_{l_i+1}\| + \|c_{m_i} - c_{l_i+1}\| \leqslant$$

$$2L\|x_{m_i} - x_{l_i+1}\| \leqslant \frac{2L}{\bar{\tau}}[\varphi(x_{m_i}; \rho_{m_i}) - \varphi(x_{l_i+1}; \rho_{l_i+1})], \tag{4.12}$$

where L is the Lipschitz constant of $\|P(x)g(x)\|$ and $\|c(x)\|$. By $\varphi(x, \rho)$ being bound below we know that the last right-hand side of (4.12) approaches zero as i tends to infinity. This contradicts (4.10) and hence proves the theorem. □

Theorem 4.7 Assume that assumptions A1～A6 hold. Then the proposed algorithm is two-step Q-superlinear convergent, i.e.

$$\lim_{k\to\infty}\frac{\|x_{k+1} - x_*\|}{\|x_{k-1} - x_*\|} = 0. \tag{4.13}$$

Furthermore, sequence $\{x_k + p_k\}$ is one-step Q-superlinear convergent, i e

$$\lim_{k\to\infty} \frac{\| x_k + p_k - x_* \|}{\| x_{k-1} + p_{k-1} - x_* \|} = 0. \tag{4.14}$$

Proof Since (4.6) holds, it means that for large k, the trust region radius will be nondecreasing, i e $\Delta_{k+1} \geqslant \Delta_k$. Hence there exists $\delta > 0$ such that $\Delta_{k+1} \geqslant \delta$, for large k. As $d_k \to 0$, it means that for large k the trust region constraint of the subproblem (S_k) is inactive, i e $\| d_k \| < \Delta_k$. So $s_k \to 0$, and hence $\| s_k \| < \Delta_k$, for the subproblem (P_k) which means $s_k = - A_k(A_k^T A_k)^{-1} c_k$ and $v_k = - A_k(A_k^T A_k)^{-1} c(x_k + p_k)$. In addition $\xi_k \geqslant \eta_k \geqslant \mu$, by the rule $\xi_k \geqslant \mu$ in the step4 of the algorithm, we have $x_{k+1} = x_k + p_k + v_k$. To summarize above, similar to the proof of theorem 4.5 in [6], we have obtain that (4.13) and (4.14) hold. □

We have studied the convergence properties of the projected quasi-Newton algorithm in association with nonmonotone trust region technique for equality constrained optimization problems. We also feel that the proposed algorithm needs a lot of numerical testing so that it can be applied in practice. These researches will be studied in coming papers.

References:

[1] BOGGS P T, TOLLE J W, WANG P. On the local convergence of quasi-Newton methods for constrained optimization[J]. SIAM J Control Opt, 1982, 20 161-171.

[2] BYRD R H, SCHNABLE R B. Continuity of the null space basis and constrained optimization[J]. Math Prog, 1986, 35: 32-41.

[3] COLEMAN T F, CONN A R. On the local convergence of a quasi-Newton method for the nonlinear programming problem[J]. SIAM J Numer Anal, 1984, 21: 755-769.

[4] GRIPPO L, LAMPARIELLO, LUCIDI S. A Nonmonotone Line Search Technique for Newton's Methods[J]. SIAM J Numer Anal, 1986, 23: 707-716.

[5] NOCEDAL J, OVERTON M L. Projected Hessian update algorithms for nonlinear constrained optimization [J]. SIAM J Numer Anal, 1985, 22: 821-850.

[6] ZHANG J, ZHU D. Projected quasi-Newton algorithm with trust region for constrained optimization [J]. J Optim Thoery Appl, 1990, 67: 269-393.

等式约束优化的投影拟牛顿法的非单调信赖域算法

钱纯青, 傅 军, 朱德通

(上海师范大学 数理信息学院, 上海 200234)

摘要: 提供了分解投影拟牛顿法结合非单调信赖域算法求解非线性等式约束优化问题. 在合理的条件下, 证明了算法的整体收敛性. 通过引进二阶矫正步克服了 MARATOS 效应, 使算法保持了局部超线性收敛速率.

关键词: 信赖域; 非单调技术; 约束优化