第32卷第3期 上海师范大学学报(自然科学版) Vol. 32, No. 3
2003年9月 Journal of Shanghai Normal University (Natural Sciences) Sep . 2003

A Scaling Trust Region Interior Point Algorithm for Linear Constrained Optimization Subject to Bounds on Variables

GU Yi-ming, ZHU De-tong

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

Abstract: In this paper we propose a scaling trust region interior point algorithm for linear constrained optimization subject to bounds on variables. The proposed algorithm is globally convergent and locally fast convergent rate even if conditions are reasonable. The results of numerical experiment are reported to show the effectiveness of the proposed algorithm.

Key words: trust region; constrained optimization; interior point

CLC number: O221.2 **Document code**: A **Article ID**: 1000-5137(2003)03-0013-08

In this paper we consider the trust region interior point algorithm for solving the linear equality constrained optimization problem subject to bounds on variables:

$$\begin{aligned} &\min \quad f(x) \\ &\text{s t} \quad Ax = b,\ l \leqslant x \leqslant u. \end{aligned} \tag{1.1}$$

where f: $\mathbf{R}^n \to \mathbf{R}$ is smooth nonlinear function, not necessarily convex, $A \in \mathbf{R}^{m\times n}$ is a matrix, $b \in \mathbf{R}^m$ is a vector, the vectors l, $u \in \{\mathbf{R} \cup \{\pm\infty\}\}^n$, $l < u$.

There are quite a few articles proposing sequential convex quadratic programming methods with trust region idea in [1], [2]. These resulting methods generate sequences of points in the interior of the feasible set. At first we define the scaling matrix $D_k \overset{\text{def}}{=} \operatorname{diag}\{D_k^1, D_k^2, \cdots, D_k^n\}$ at the k-th iteration with the component D_k^i defined as follows

$$D_k^i \overset{\text{def}}{=} \begin{cases} x_k^i - l_i & \text{if } l_i > -\infty, \text{ and } u_i = +\infty \\ u_i - x_k^i & \text{if } l_i = -\infty, \text{ and } u_i < +\infty \\ \min\{x_k^i - l_i, u_i - x_k^i\} & \text{if } l_i > -\infty, \text{ and } u_i < +\infty \\ 1 & \text{if } l_i = -\infty, \text{ and } u_i = +\infty \end{cases} \tag{1.2}$$

where x_k^i, l_i and u_i are the i-th components of the vectors x_k, l and u, respectively. The matrix B_k is a symmetric approximation of the Hessian $\nabla^2 f(x_k)$ of the objective function in which B_k assumed to be positive semidefinite in [1]. A search direction at x_k by solving the trust region convex quadratic programming:

Received date: 2002-9-1

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

Biography: GU Yi-ming (1974-), male, lecturer, Mathematics and Sciences College, Shanghai Normal University; ZHU De-tong (1954-), male, doctor, professor, Mathematics and Sciences College, Shanghai Normal University.

$$\begin{aligned} \min \quad & \varphi(d) \overset{\text{def}}{=} g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s t} \quad & Ad = 0, \\ & \| D_k^{-1} d \| \leq \Delta_k, \quad l < x_k + d < u. \end{aligned} \tag{1.3}$$

where $g_k = \nabla f(x_k)$, $d = x - x_k$, $\varphi(d)$ is the local quadratic approximation of $f(x)$ and Δ_k is the trust region radius. Let d_k be the solution of the subproblem.

A solution that minimizes the model function within the trust region is solved as a trial step. If the actual reduction achieved on the objective function f at this point is satisfactory comparing with the reduction predicted by the quadratic model, the point is accepted as a new iteration; otherwise the trust region radius is adjusted and hence the procedure is repeated. The paper is organized as follows. In Section 2, we describe the trust region interior point algorithm with scaling matrix. In Section 3, weak global convergence of the proposed algorithm is established. Some further strong global convergence and local convergent rate are discussed in Section 4. Finally, the results of numerical experiments are reported in Section 5.

2 Algorithm

In this section we describe a trust region interior point method for linear equality constrained optimization (1.1) subject to box constraints on variables. The method involves choosing a scaling matrix D_k and a quadratic model of the objective function. We motivate our choice of scaling matrix by examining the optimality conditions for (1.1).

Let the Lagrangian function of problem (1.1)

$$l(x,\lambda,\mu,\nu) = f(x) + \lambda^T Ax - \mu^T(x - l) - \nu^T(u - x), \tag{2.1}$$

where the Lagrangian multipliers $\lambda \in \mathbf{R}^m$, $0 \leq \mu$, $\nu \in \mathbf{R}^n$. Optimality conditions for problem (1.1) are well established. Assuming feasibility, first-order necessary conditions for x_* to be a local minimizer are that there exist $\mu_*, \nu_* \in \mathbf{R}^n$ and $\lambda_* \in \mathbf{R}^m$ such that

$$\begin{aligned} & g_* - A^T\lambda_* - \mu_* + \nu_* = 0,\ \mu_* \geq 0,\ \nu_* \geq 0, \\ & Ax_* = b,\ l \leq x_* \leq u, \\ & \mu_*^T(l - x_*) = 0,\ \ \nu_*^T(u - x_*) = 0. \end{aligned} \tag{2.2}$$

Now we state the scaling trust region interior point algorithm for optimization problem (1.1).

Initialization step

Choose parameters $0 < \bar{\eta} < \eta_1 < \eta_2 < 1$, $0 < \gamma_0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $\epsilon \geq 0$. 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 strictly feasible point $x_0 \in \text{int}(\Omega) \overset{\text{def}}{=} \{x \mid Ax = b,\ l < x < u\}$. 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 $\| P_k \hat{g}_k \| \leq \epsilon$, stop with the approximate solution x_k, here the projection map $P_k \overset{\text{def}}{=} I - \hat{A}_k^T(\hat{A}_k\hat{A}_k^T)^{-1}\hat{A}_k$ of the null subspace of $\mathcal{N}(\hat{A}_k)$ with $\hat{A}_k \overset{\text{def}}{=} AD_k$, $\hat{g}_k \overset{\text{def}}{=} D_k g_k$.

(3) Solve subproblem

$$(S_k) \quad \begin{aligned} \min \quad & \psi(d) \overset{\text{def}}{=} g_k^T d + \frac{1}{2} d^T B_k d \\ \text{s.t.} \quad & Ad = 0,\quad l < x_k + d < u, \\ & \| D_k^{-1} d \| \leq \Delta_k. \end{aligned} \tag{2.3}$$

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

(4) Compute

$$\mathrm{Pred}(d_k) = -\psi_k(d_k), \quad \mathrm{Ared}(d_k) = f(x_k) - f(x_k + d_k). \tag{2.4}$$

Further, set $\rho_k = \mathrm{Ared}(d_k) / \mathrm{Pred}(d_k)$.

(5) If $\rho_k < \bar{\eta}$, then the iteration is said to be unsuccessful. Let $\Delta_k \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$, and go to step 3. Otherwise, that is, $\rho_k \geq \bar{\eta}$ is said to be successful iteration, then go to the next step.

(6) Set $x_{k+1} = x_k + d_k$, and take

$$\Delta_{k+1} = \begin{cases} [\gamma_1 \Delta_k, \gamma_2 \Delta_k], & \text{if } \rho_k \leq \eta_1, \\ (\gamma_2 \Delta_k, \Delta_k] & \text{if } \eta_1 < \rho_k < \eta_2, \\ (\Delta_k, \min\{\gamma_3 \Delta_k, \Delta_{\max}\}], & \text{if } \rho_k \geq \eta_2. \end{cases}$$

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

(7) Update B_k to obtain B_{k+1}. Then set $k \leftarrow k + 1$ and go to step 2.

Remark In the subproblem (S_k), $\psi_k(d)$ is a local quadratic model of the objective function f around x_k. A candidate iterative direction d is generated by minimizing $\psi_k(d)$ along the interior points of the feasible set within the ellipsoidal ball centered at x_k with radius Δ_k. A key property of this transformation in (S_k) is that $D_k^{-1} d_k$ is at least unit distance from all bounds in the scaled coordinates, i e, an arbitrary step $D_k^{-1} d_k$ to the point $x_k + d_k$ does not violate any bound if $d_k^T D_k^{-2} d_k < 1$. To see this, first observe that $d_k^T D_k^{-2} d_k = \sum_{i=1}^{n} (d_k^i / D_k^i)^2 < 1$, which implies $|d_k^i| < D_k^i$, for $i = 1, \cdots, n$, where d_k^i is the i-th component of d_k, D_k^i is defined given in (1.3). We discuss the correlation of l_i, u_i, x_k^i and d_k^i in the following cases, respectively.

(i) If $l_i > -\infty$ and $u_i = +\infty$, then $|d_k^i| < D_k^i = x_k^i - l_i$. So, we have $l_i < x_k^i + d_k^i < u_i$.

(ii) If $l_i = -\infty$ and $u_i < +\infty$, then $|d_k^i| < D_k^i = u_i - x_k^i$. We have $l_i < x_k^i + d_k^i < u_i$ as well.

(iii) If $l_i > -\infty$ and $u_i < +\infty$, then $|d_k^i| < D_k^i = \min\{x_k^i - l_i, u_i - x_k^i\}$. We can obtain $l_i < x_k^i + d_k^i < u_i$ same as (i) and (ii).

(iv) If $l_i = -\infty$ and $u_i = +\infty$, then $|d_k^i| < D_k^i = 1$. Clearly, we can get $l_i < x_k^i + d_k^i < u_i$.

Therefore, no matter what case is the sign of d_k^i, the inequality $l_i < x_k^i + d_k^i < u_i$ holds.

3 Global convergence

Assumption 1 The objective function $f(x): \mathbf{R}^n \to \mathbf{R}^1$ is twice continuously differentiable and bounded blew on $\mathbf{R}^n$.

Throughout this section we assume that Assumption 1 holds. Given $x_0 \in \mathrm{int}(\Omega)$, the algorithm generates a sequence $\{x_k\} \subset \mathbf{R}^n$. In our analysis, we denote the level set of f by

$$\mathscr{L}(x_0) = \{ x \in \mathbf{R}^n \mid f(x) \leq f(x_0), Ax = b, l \leq x \leq u \}.$$

The following assumption is commonly used in convergence analysis of most methods for linear equality constrained optimization.

Assumption 2 Sequence $\{x_k\}$ generated by the algorithm is contained in a compact set $\mathscr{L}(x_0)$ on $\mathbf{R}^n$. Matrix A has full row-rank m.

Assumption 3 $D_k B_k D_k$ is bounded, i e, there exists $r > 0$ such that $\| D_k B_k D_k \| \leq r$, $\forall\ k$.

Assumption 4 $D(x) \nabla^2 f(x) D(x)$ is bounded, i e, there exists $\hat{r} > 0$ such that $\| D(x) \nabla^2 f(x) D(x) \| \leq \hat{r}$, here $D(x) \overset{\mathrm{def}}{=} \mathrm{diag}\{D^1(x), D^2(x), \cdots, D^n(x)\}$. Furthermore, $D^i(x)$ has the similar definition as (1.2).

The following lemma establishes the necessary and sufficient condition concerning μ_k, d_k, and λ_k when d_k solves (2.3), which is similar to proof of Lemma 3.1 in [8] (also [6], [7]).

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

$$\begin{aligned} &(B_k + \mu_k D_k^{-2}) d_k = -g_k + A^T \lambda_k \\ &A\, d_k = 0,\ l \leqslant x_k + d_k \leqslant u \\ &\mu_k(\Delta_k^2 - d_k^T D_k^{-2} d_k) = 0 \end{aligned} \tag{3.2}$$

holds and $B_k + \mu_k D_k^{-2}$ is positive semidefinite in $\mathcal{N}(A)$.

Similar to Proofs of Lemma 3.3 in [8], we can also obtain following lemma.

Lemma 3.2 Let the step d_k be the solution of the trust region subproblem, then there exists τ such that the step d_k satisfies the following sufficient descent condition.

$$\text{Pred}\,(d_k) \geqslant \tau \, \| P_k \hat{g}_k \| \, \min\, \{1, \Delta_k, \frac{\| P_k \hat{g}_k \|}{\| D_k B_k D_k \|}\} \tag{3.3}$$

for all g_k, B_k and Δ_k, where $P_k = I - \hat{A}_k^T (\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k$ with $\hat{A}_k = AD_k$ and $\hat{g}_k = D_k g_k$.

Lemma 3.3 Under the Assumptions 1 ~ 2. If $\| P_k \hat{g}_k \| \neq 0$, via the finite iterations, we affirmably obtain a d_k such that $\rho_k = \text{Ared}(d_k)/\text{Pred}(d_k) \geqslant \bar{\eta}$. In this case the step is accepted. Thus we take $x_{k+1} = x_k + d_k$ as a new iteration.

Proof From Lemma 3.2 and Assumption 2, let $\varepsilon = \| P_k \hat{g}_k \|$, when $\Delta_k \leqslant \min\{\frac{\varepsilon}{\|D_k B_k D_k\|}, 1\}$, we have

$$\text{Pred}_k(d_k) \geqslant \tau \, \| P_k \hat{g}_k \| \min\{1, \Delta_k, \frac{\| P_k \hat{g}_k \|}{\| D_k B_k D_k \|}\} \geqslant \tau \varepsilon \Delta_k$$

Again, we can obtain

$$\begin{aligned} | f(x_k + d_k) - f(x_k) - \varphi(d_k) | &= | \frac{1}{2} d_k^T (\nabla^2 f(x_k + \xi d_k) - B_k) d_k | \\ &\leqslant \frac{1}{2} \| D_k^{-1} d_k \|^2 \, \| D_k \nabla^2 f(x_k + \xi d_k) D_k - D_k B_k D_k \| \\ &\leqslant \frac{1}{2} (r + \hat{r}) \Delta_k^2. \end{aligned}$$

where $\xi \in [0,1]$. Thus, $\lim_{\Delta_k \to 0} | \rho_k - 1 | = \lim_{\Delta_k \to 0} \frac{| \text{Ared}_k(d_k) - \text{Pred}_k(d_k) |}{| \text{Pred}_k(d_k) |} = 0$. This shows that there exists Δ_ϵ, such that when $\Delta_k \leqslant \Delta_\epsilon$, $\rho_k \geqslant \bar{\eta}$, which means that the successful step is obtained in the finite iterations. □

Similar to proof of Lemma 4.1 in [8], we can obtain the following lemma.

Lemma 3.4 Assume that Assumptions 1 ~ 4 hold. If there exists $\epsilon > 0$ such that

$$\| P_k \hat{g}_k \| \geqslant \epsilon \tag{3.4}$$

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

$$\Delta_k \geqslant \alpha, \ \forall\, k. \tag{3.5}$$

where $\alpha = \min\{\frac{2\hat{\tau}\varepsilon\gamma_0(1 - \bar{\eta})}{r + \hat{r}}, \frac{\varepsilon\gamma_1}{r}\}$, and $\hat{\tau} = \min\{1, \tau\}$.

Theorem 3.5 Assume that Assumptions 1 ~ 4 hold. Let $\{x_k\} \subseteq \mathbf{R}^n$ be a sequence generated by the algorithm. Then

$$\liminf_{k \to \infty} \| P_k \hat{g}_k \| = 0. \tag{3.6}$$

Proof If the conclusion (3.6) is not true, there for all sufficiently large k, $\| P_k \hat{g}_k \| \geqslant \epsilon$ for some $\epsilon > 0$. Along with (3.3) and Lemma 3.4, for a successful interaction with index k, we can get that

$$f(x_k) - f(x_{k+1}) \geqslant \bar{\eta}\mathrm{Pred}(d_k) \geqslant \tau\bar{\eta}\epsilon\min\{1,\Delta_k,\frac{\epsilon}{r}\},$$

Summing all the successful iterations from 0 to k, we have

$$f(x_0) - f(x_{k+1}) = \sum_{i=0}^{k}[f(x_i) - f(x_{i+1})] \geqslant \sum_{i=0}^{k}\tau\bar{\eta}\epsilon\min\{1,\Delta_k,\frac{\epsilon}{r}\},$$

Since $f(x_k)$ is bounded blew on $\mathbf{R}^n$, we obtain that $f(x_0) - f(x_{k+1})$ is bounded. It implies that

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

which contradicts (3.5) in Lemma 3.4. Hence the conclusion of the theorem is true. □

4 Local Convergence

Theorem 3.5 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 and the local convergent rate, but it requires more assumptions.

We denote the sets of active constraints by

$$I(x) \overset{\text{def}}{=} \{i \mid x^i = l_i,\ i = 1,\cdots,n\},\quad Q(x) \overset{\text{def}}{=} \{i \mid x^i = u_i,\ i = 1,\cdots,n\}.$$

To any $I(x) \cup Q(x) \subset \{1,\cdots,n\}$ we associate the optimization subproblem

$(P)_{I\cup Q}$ $\min f(x)$; s t $Ax = b$, $(x - l)_{I(x)} = 0$ or $(x - u)_{Q(x)} = 0$.

Assumption 5 For all $I(x) \cup Q(x) \subset \{1,\cdots,n\}$, the first order optimality system associated to $(P)_{I\cup Q}$ has no nonisolated solutions.

Assumption 6 The constraints of (1.1) are qualified in the sense that $(A^T\lambda)_i = 0$, $\forall\ i \notin I(\bar{x}) \cup Q(\bar{x})$ implies that $\lambda = 0$.

Assuming that $(\bar{\lambda},\bar{\mu},\bar{\nu})$ is associated with a unique pair $\bar{x}$ which satisfies Assumption 5. Define the set of strictly active constraints as

$$J_I(\bar{x}) = \{i \mid \bar{\mu}_i > 0,\ i = 1,\cdots,n\},\ J_Q(\bar{x}) = \{i \mid \bar{\nu}_i > 0,\ i = 1,\cdots,n\},$$

and the extended critical cone as

$$T(\bar{x}) = \{d \in \mathbf{R}^n \mid Ad = 0, d^i = 0, i \in J_I(\bar{x}) \cup J_Q(\bar{x})\}$$

Assumption 7 The solution x_* of problem (1.1) satisfies the strong second order condition, that is, there exists $\alpha > 0$ such that

$$p^T\nabla^2 f(x_k)p \geqslant \alpha\|p\|^2,\ \forall\ p \in T(\bar{x}). \tag{4.1}$$

Given $d \in \mathcal{N}(A)$, the null space of A, we define d^Y, d^N as the orthogonal projection of d onto T and N, where N is the orthogonal complement of T in $\mathcal{N}(A)$, i e, $N = \{z \in \mathcal{N}(A) \mid z^Td = 0,\ \forall\ d \in T(\bar{x})\}$, which means that $d = d^Y + d^N$ and $\|d\|^2 = \|d^Y\|^2 + \|d^N\|^2$.

Similar to Lemma 4.1 in [8], we obtain the following lemma.

Lemma 4.1 Assume that Assumptions 5 ~ 7 hold. If x_k is sufficiently close to x_* then

$$d^T(B_k + 2\mu_k D_k^{-2})d_k \geqslant \frac{\alpha}{2}\|d_k\|^2 + \kappa\|d_k^N\|^2, \tag{4.2}$$

$$d^T(B_k + \mu_k D_k^{-2})d_k \geqslant \frac{\alpha}{2}\|d_k\|^2 + \kappa_1\|d_k^N\|^2. \tag{4.3}$$

where $\kappa > 0$, $\kappa_1 > 0$, and α given in (4.1), multiplier μ_k given in (3.1).

Theorem 4.2 Assume that Assumptions 5 ~ 7 hold. Let $\{x_k\}$ be a sequence generated by the algorithm. Then $d_k \to 0$. Furthermore, if x_k is close enough to x_*, and x_* is a strict local minimum of the problem (1.1), then $x_k \to x_*$.

Proof We can deduce that

$$| f(x_k + d_k) - f(x_k) - \psi_k(d_k) | = | \frac{1}{2} d_k^T (\nabla^2 f(x_k) - B_k) d_k + o(\|d_k\|^2) | = o(\|d_k\|^2).$$

By (3.2), we get that

$$g_k^T d_k = - d_k^T (B_k + \mu_k D_k^{-2}) d_k - \lambda_k^T A d_k = - d_k^T (B_k + \mu_k D_k^{-2}) d_k \leqslant - \frac{\alpha}{2} \| d_k \|^2 - \kappa_1 \| d_k^N \|^2.$$

From (2.4) and (4.2), for a large enough k,

$$\mathrm{Pred}(d_k) = d_k^T (B_k + \mu_k D_k^{-2}) d_k - \frac{1}{2} d_k^T B_k d_k = \frac{1}{2} d_k^T (B_k + 2\mu_k D_k^{-2}) d_k \geqslant \frac{\alpha}{4} \|d_k\|^2 + \frac{\kappa}{2} \|d_k^N\|^2.$$

where κ given in (4.2). As for a large enough k, we obtain that

$$\rho_k = \frac{f_k - f(x_k + d_k)}{\mathrm{Pred}(d_k)} = 1 + \frac{f_k - f(x_k + d_k) + \psi_k(d_k)}{\mathrm{Pred}(d_k)} \geqslant 1 - \frac{o(\|d_k\|^2)}{\frac{\alpha}{4} \| d_k \|^2 + \frac{\kappa}{2} \|d_k^N\|^2} \geqslant \bar{\eta}.$$

According to the acceptance rule in step 4, we have

$$f(x_k) - f(x_{k+1}) \geqslant \bar{\eta} \mathrm{Pred}(d_k) \geqslant \frac{\alpha\bar{\eta}}{4} \|d_k\|^2 + o(\|d_k\|^2). \tag{4.4}$$

Clearly, we know

$$\lim_{k \to \infty} [f(x_k) - f(x_{k+1})] = 0. \tag{4.5}$$

(4.4) and (4.5) imply that $\lim_{k \to \infty} \|d_k\| = 0$.

Assume that there exists a limit point x_* which is a local minimum of f. As the Assumption 3 necessarily holds in a neighborhood of x_*, then x_* is the only limit point $\{x_k\}$ in some neighborhood $\mathcal{N}(x_*;\delta)$ of x_*, where $\delta > 0$ is a small constant. On the other hand, we know that $f(x_k) \geqslant f(x_{k+1})$. So, we define $\hat{f} = \inf\{f(x) ; x \in \mathcal{N}(x_*;\delta) \backslash \mathcal{N}(x_*;\delta/2)\}$. Because x_* is a strick local minimum of the problem (1.1), we may assume $\hat{f} > f_*$. Now, assuming that $f(x_k) \leqslant \hat{f}$ and $x_k \in \mathcal{N}(x_*;\delta/2)$, it follows that $f(x_{k+1}) \leqslant \hat{f}$, $\| d_k \| \leqslant \delta/2$ and $x_{k+1} \in \mathcal{N}(x_*;\delta)$; using the definition of $\hat{f}$, we find that $x_{k+1} \in \mathcal{N}(x_*;\delta/2)$. This implies that sequence $\{x_k\}$ remains in $\mathcal{N}(x_*;\delta/2)$. Hence, $x_k \to x_*$. It means that the conclusion of the theorem is true.

□

Similar to prove Theorem 4 in [8], we can obtain

Theorem 4.3 Assume that assumptions 3 ~5 hold. Let $\{x_k\}$ be a sequence generated by the algorithm. Then

$$\lim_{k \to \infty} \| P_k \hat{g}_k \| = 0. \tag{4.6}$$

Theorem 4.4 Let $\{x_k\}$ be a sequence propose by the algorithm. Assume that $\{B_k\}$ is bounded, then

(i) any limit point x_* of $\{x_k\}$ is a solution of the first-order optimality system associated to problem $(P)_{I \cup Q(x_*)}$, that is,

$$\begin{aligned} &\nabla f(x_*) - A^T \lambda_* - \mu_* + \nu_* = 0, \ Ax_* = b, \\ &(x_*)_{I(x_*)} = l_{I(x_*)};\ \mu_*^i = 0,\ i \notin I(x_*), \\ &(x_*)_{Q(x_*)} = u_{Q(x_*)};\ \nu_*^i = 0,\ i \notin Q(x_*) \end{aligned} \tag{4.7}$$

(ii) if Assumptions 3 ~4 and (4.7) hold, then x_* satisfies the first-order optimality system of (1.1); i e, there exist $\lambda_* \in \mathbf{R}^m$, $\mu_*, \nu_* \in \mathbf{R}^n$ such that

$$\begin{aligned} &\nabla f(x_*) - A^T \lambda_* - \mu_* + \nu_* = 0, \\ &Ax_* = b,\ l \leqslant x_* \leqslant u \\ &\mu_*^T (x_* - l) = 0,\ \nu_*^T (u - x_*) = 0. \end{aligned} \tag{4.8}$$

Proof First we can prove that the sequence $\{x_k\}$ satisfies the following conditions: (i) $d_k \to 0$, (ii)

$\mu_k \to 0$, (iii) $D_k[\nabla f_k - A^T\lambda_k] \to 0$.

Indeed, From Theorem 4.2, we have obtain that (i) holds. Since $f(x)$ is twice continuously differentiable we have that

$$f(x_k + d_k) = f(x_k) + g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k + o(\|d_k\|^2). \tag{4.9}$$

Furthermore, by (3.1) we have

$$g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k = -\mu_k d_k^T D_k^{-2} d_k. \tag{4.10}$$

Hence we obtain that

$$f(x_k) - f(x_k + d_k) = \mu_k d_k^T D_k^{-2} d_k + o(\|d_k\|^2) = \mu_k \Delta_k^2 + o(\|d_k\|^2).$$

Due to the fact that $\{f(x_k)\}$ is bounded from below, we deduce that $\mu_k \Delta_k^2 \to 0$ as $k \to \infty$. It implies that either

$$\lim_{k\to\infty} \mu_k = 0 \tag{4.11}$$

or

$$\liminf_{k\to\infty} \Delta_k = 0. \tag{4.12}$$

Now we prove that if $\mu_k \geqslant \delta > 0$ for all k sufficiently large, then a contradiction will come into being. From (4.9) we have that when x_k is close enough to x_* ,

$$| f(x_k + d_k) - f(x_k) - \psi_k(d_k) | = o(\|d_k\|^2). \tag{4.13}$$

Again, we have

$$\mathrm{Pred}(d_k) = \frac{1}{2} d_k^T (B_k + \mu_k D_k^{-2}) d_k + \frac{\mu_k}{2} d_k^T D_k^{-2} d_k \geqslant \frac{\mu_k}{2} d_k^T D_k^{-2} d_k \geqslant \frac{\mu_k}{2} \Delta_k^2 \geqslant \frac{\delta}{2} \Delta_k^2. \tag{4.14}$$

By (4.13) ~ (4.14), we have that

$$| \rho_k - 1 | = \frac{\mathrm{Ared}(d_k) - \mathrm{Pred}(d_k)}{\mathrm{Pred}(d_k)} = \frac{o(\|d_k\|^2)}{\delta_k \Delta_k^2 / 2} \to 0 \quad \text{as} \quad \Delta_k \to 0$$

That means that sequence $\{\rho_k\}$ converges to unity and not decreased for sufficiently large k , and hence bounded away from zero. So, (4.12) cannot hold.

Theorem 4.3 means $\|P_k \hat{g}_k\| \to 0$ which implies $P_k \hat{g}_k = [I - \hat{A}_k^T (\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k] \hat{g}_k = D_k[\nabla f_k - A^T \lambda_k]$ because of $\hat{A}_k = AD_k$, $\hat{g}_k = D_k g_k$ and $\lambda_k = (\hat{A}_k^T \hat{A}_k)^{-1} \hat{A}_k^T \hat{g}_k$. By the conditions (i), (ii), (iii), similar to the proofs of Theorem 2.2 in [1], we can also obtain that the conclusions of the theorem are true. □

The conclusions of Theorem 4.4 are the same as theorem 2.2 in [1] under the same as assumptions in [1].

We now discuss the rate of local convergence for the proposed algorithm.

Theorem 4.5 Assume that Assumptions 5 ~ 7 hold. For sufficiently k , then the trust region constraint is inactive, that is, there exists a $\hat{\Delta} > 0$ such that $\Delta_k \geqslant \hat{\Delta}$.

Due to the trust region radius is inactive, the rate of local convergence for the proposed algorithm relies on the Hessian of objective function at x_* and the local convergence rate of d_k. If d_k becomes the quasi-Newton step, then the sequence $\{x_k\}$ generated by the algorithm converges x_* superlinear.

5 Numerical Experiments

Numerical experiments on the scaling trust region interior point algorithm in this paper have been performed on a Pentium 4 personal computer. In this section we present the numerical results. In order to check effectiveness of the method, we select the parameters as following: $\epsilon = 10^{-6}$, $\bar{\eta} = 0.01$, $\eta_1 = 0.1$, $\eta_2 = 0.75$, $\gamma_0 = 0.1$, $\gamma_1 = 0.2$, $\gamma_2 = 0.5$, $\gamma_3 = 2$, $\Delta_{max} = 10$, $\Delta_0 = 1$.

Table 1 Experimental Results

Problem Name	n	ITR	NF	NG
HS28	3	7	9	7
HS48	5	4	5	4
HS49	5	36	38	36
HS51	5	3	4	3
HS73	5	14	15	14

The experiments are carried out on 5 standard test problems quoted from [5] (HS: the problem from Hock and Schitkowski[5]). ITR, NF and NG stand for the numbers of iterations, function evaluations and gradient evaluations, respectively.

References:

[1] BONNANS J F, POLA C. A trust region interior point algorithm for linear constrainede optimization[J]. SIAM J Optimization, 1997, 7(3):717-731.

[2] BONNANS F, COLEMAN T. Combining trust region and affine scaling for linearly constrained nonconvex minimization [J]. Advances in Nonlinear Programming (Y Yuan ed) Kluwer Academic Publishers, 1998: 219-250.

[3] DENNIS J E, VICENTE L N. On the convergence theory of trust-region-based algorithms for equality-constrained optimization[J]. SIAM Journal on Optimization, 1997(7): 527-550.

[4] DENNIS J E, SCHNABLE R B. Numerical methods for unconstrained optimization and nonlinear equations[M]. Prentice Hall, New Jersey, 1983.

[5] HOCK W, SCHITTKOWSKI K. Test examples fro nonlinear programming code, Lecture Notes in Economics and Mathematical Systems 187[M]. Springer-Verlag, Berlin, 1981.

[6] MORE J J, SORENSEN D C. Computing a trust region step[J]. SIAM Journal on Science and Statistical Computing, 1983(4): 553-572.

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

[8] ZHU D T. Nonmonotonic back tracking trust region interior point algorithm for linear constrained optimization[J]. J of Computational and Applied Mathematics, 2003, 155: 285-305.

[9] ZHU D T. Curvilinear paths and trust region methods with nonmonotonic back tracking technique for unconstrained optimization[J]. J of Computational Mathematic, 2001, 19(3):241-258.

有界变量与线性等式约束优化的信赖域内点算法

顾益明, 朱德通

(上海师范大学 数理信息学院, 上海 200234)

摘要: 提出一种既有界变量又有线性等式约束的非线性优化问题的信赖域内点算法,在合理的条件下所提供的算法不仅具有整体收敛性而且保持局部收敛速率。数值计算结果说明算法的有效性。

关键词: 信赖域;约束优化;内点