Far East J. Appl. Math. **28**(2) (2007), 243-250

AN SQP ALGORITHM WITH TRUST REGION FOR NONLINEAR CONSTRAINED OPTIMIZATION

XIAOHUI YANG, ZHIBIN ZHU and QINGGAN TANG

College of Mathematics and Computational Science
Guilin University of Electronic Technology
Guilin, 541004, P. R. China
e-mail: 04271106@mailbox.gliet.cn

Abstract

In this paper, an improved trust region algorithm is presented to solve nonlinear equality and inequality constraints optimization by combining with the SQP methods. Under some suitable conditions, the global convergence and surperlinear convergence are obtained.

1. Introduction

We consider the following nonlinear programming with inequality and equality constraints:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_j(x) \le 0,\ j \in L_1 = \{1, \ldots, m_e\}, \\ & f_j(x) = 0,\ j \in L_2 = \{m_{e+1}, \ldots, m\}, \end{aligned} \tag{1.1}$$

where $x \in R^n$, $f_j : R^n \to R$. X denotes the feasible set.

Trust region methods have been proved theoretically and practically to be effective for unconstrained and equality constrained optimization

2000 Mathematics Subject Classification: 90C30, 65K05.

Keywords and phrases: nonlinear constrained optimization, trust region method, SQP methods, global convergence.

Received October 31, 2006

problems. However, relatively, there are few trust region methods which were proved to be efficient for general constrained optimization (1.1). We propose a modified version of the feasible trust region algorithm in [3] for solving nonlinear programs with mix equality and inequality constraints.

In this paper, we consider a feasible decent SQP methods in [2] combining with trust region techniques in [3, 6]. The main features of the proposed algorithm are summarized as follows: (a) We obtain an equivalent QP subproblem. It is different from that in [1, 7] which dispenses with slack variables, and we use a merit function to convert the equality constraints into the inequality constraints. (b) Motivated by the ideas in [3], we combined trust region technique with the generalized gradient projection, and its trust region is a general compact set containing the origin as an interior point.

We introduce the penalty function which is given by Mayne and Polak in [5].

$$\begin{aligned} \min \quad & F_{C_k}(x) \triangleq f_0(x) - C_k \sum_{j \in L_2} f_j(x) \\ \text{s.t.} \quad & f_j(x) \le 0,\ j \in L \triangleq L_1 \cup L_2, \end{aligned} \tag{1.2}$$

where penalty parameter $C_k > 0$, and the feasible set of (1.2) $X^+ = \{x \in R^n \mid f_j(x) \le 0,\ j \in L\}$.

2. Description of Algorithm

Some basic assumptions are given as follows:

H2.1. The feasible set is nonempty, and the functions $f_j(x)$ are two-times continuously differentiable.

H2.2. The vectors $\{\nabla f_j(x),\ j \in J(x)\}$ are linearly independent, here $x \in X^+$.

Denote active set $J(x) = \{j \in L_1 \mid f_j(x) = 0\} \cup L_2 \triangleq J_1(x) \cup L_2$, and make some definitions as follows:

$f(x) = (f_1(x), \ldots, f_m(x))^T$, $f_J(x) = (f_j(x),\ j \in J)$,

$$g_j(x) = \nabla f_j(x),\ j \in L, \quad g_J(x) = (g_j(x) = \nabla f_j(x),\ j \in J),$$

$$L_k = \{j \in L \mid f_j(x^k) + g_j(x^k)^T d^k = 0\},$$

$$J(x^k, \varepsilon) = \{j \in L_1 \mid -\varepsilon_{k,i} \le f_j(x^k) \le 0\} \cup L_2 \triangleq J_1(x^k, \varepsilon) \cup L_2,$$

$$N_k = g_{J_k}(x^k),\ \ u(x^k) = -(N_k^T N_k)^{-1} N_k^T g_0(x^k) = (u_j(x^k),\ j \in J_k). \quad (2.1)$$

Lemma 2.1. *If parameter* $C_k > |u_j(x^k)|,\ j \in L_2,$ *then* (x^k, u^k) *is the K-T point pair of the problem* (1.1), *if and only if* (x^k, λ_j^k) *is the K-T point pair of the problem* (1.2), *and it holds that if* $j \in J_1(x^k, \varepsilon_k),\ u_j^k = \lambda_j^k,$ *and if* $j \in L_2,\ u_j^k = \lambda_j^k - C_k.$

Proof. See the proof of Lemma 2.2 in [4].

Let Ω be a general compact set containing the origin as an interior point, and $r \in R$. Denote $r\Omega = \{r\omega \mid \omega \in \Omega\}$. For the current iteration point $x^k,\ r_k > 0$ and Lagrange Hessian symmetric positive definite matrix H_k, we consider the following QP subproblem with trust region:

$$\min \quad q_k(d) = f_0(x) + \nabla F_{C_k}(x^k)^T d + \frac{1}{2} d^T H_k d$$

$$\text{s.t.} \quad f_j(x^k) + g_j(x^k)^T d \le 0,\ j \in J_k = J(x^k, \varepsilon),\ d \in r_k\Omega. \quad (2.2)$$

According to the definition of the K-T point, it is easy to obtain the following result.

Lemma 2.2. *If* $d_0^k = 0,$ *then* x^k *is the K-T point of the problem* (2.2). *Furthermore, if* $C_k > |u_j(x^k)|,\ j \in L_2,$ *then* x^k *is the K-T point of the problem* (1.1).

The formal structure of the algorithm is as follows:

Algorithm

Step 0. Initialization and date: Given parameters $\sigma > 2,\ \tau \in (2, 3),$ $\gamma,\ \psi > 0,\ \delta,\ \alpha,\ \beta \in (0, 1),\ \delta < \gamma,\ \psi$ and δ small enough, $\gamma_1 > \psi,\ \xi$ and C small enough and large enough positive, Ω a general compact set containing the origin as an interior point $e = (1, \ldots, 1)^T \in R^n$. Given a starting point $x^1 \in X$ and an initial symmetric positive definite matrix H_1, set $k = 1$, $\varepsilon > 0$.

Step 1. Computation of an approximate active set J_k:

(i) Let $i = 0$, $\varepsilon_{k,i} = \varepsilon_0$;

(ii) Set

$$J_{k,i} = \{j \in I \mid -\varepsilon_{k,i} \leq f_j(x^k) < 0\},\ A_{k,i} = (g_j(x^k),\ j \in J_{k,i}). \tag{2.3}$$

If $A_{k,i}$ is not of full rank, then set $i = i + 1$, $\varepsilon_{k,i} = \frac{1}{2}\varepsilon_{k,i-1}$, and go to (ii). Otherwise, let $J_k = J_{k,i}$, $A_k = A_{k,i}$, $i_k = i$, and go to step 2.

Step 2. Update parameter C_k:

$$t_k = \max\{|u_j(x^k)| \, j \in L_2\} + C_\varepsilon,\ C_k = \begin{cases} \max\{t_k, C_{k-1}\}, & C_{k-1} < t_k, \\ C_{k-1}, & C_{k-1} \geq t_k. \end{cases}$$

Step 3. Obtain the solution d^k by solving the problem (2.2).

Step 4. Compute Δq_k, we define the predict reduction as follows:

$$\Delta q_k = q_k(0) - q_k(d^k) = -\nabla F_{C_k}(x^k)^T d^k - \frac{1}{2}(d^k)^T H_k d^k. \tag{2.4}$$

If $\Delta q_k = 0$, then x^k is K-T point of the problem (1.1), STOP. Otherwise, compute the height order modified direction $\tilde{d}^k$ and ΔF_{0k} as follows:

$$D_k = D(x^k) = diag(f_j(x^k)^2),\ j \in L, \quad Q_k = Q(x^k) = (N_k^T N_k + D_k)^{-1} N_k^T,$$

$$\beta_k = (\beta_j^k,\ j \in L),\ e^k = (e_j^k,\ j \in L), \quad \tilde{d}^k = Q_k^T(-\|d^k\|^T e^k - \beta_j^k),$$

$$\beta^k = (f_j(x^k + d^k),\ j \in J_k;\ 0,\ j \in L \setminus J_k), \quad e^k = (1,\ j \in J_k;\ 0,\ j \in J \setminus J_k). \tag{2.5}$$

Define the actual reduction as follows:

$$\Delta F_{0k} = F_{C_k}(x^k) - F_{C_k}(x^k + d^k + \tilde{d}^k). \tag{2.6}$$

Step 5. If

$$\|H_k d^k\| \leq C\|d^k\|^{\frac{1}{2}}, \; f_j(x^k + d^k + \tilde{d}^k) \leq 0, \; \forall j \in L, \tag{2.7}$$

$$\Delta q_k \geq \xi\|d^k\|^{\sigma}, \; S_k = \frac{\Delta F_{0k}}{\Delta q_k} \geq \delta, \tag{2.8}$$

set

$$x^{k+1} = x^k + d^k + \tilde{d}^k, \; \gamma_{k+1} = \begin{cases} 2\gamma_k, & S_k > \gamma, \\ \gamma_k, & S_k \leq \gamma, \end{cases} \tag{2.9}$$

and obtain the positive definite matrix H_{k+1} by updating H_k using quasi-Newton formulas. Set $k := k+1$, and go back to Step 1. If (2.7) and (2.8) are not satisfied, then go to Step 6.

Step 6. If $\gamma_k < \psi$, go to Step 7. Otherwise, set $\gamma_k := \frac{1}{4}\gamma_k$, and go back to Step 1.

Step 7. Compute a feasible decent direction $\tilde{d}^k$:

$$\pi^k = -Q_k g_0(x^k), \; P_k = P(x^k) = E_n - N_k Q_k,$$

$$\rho_k = \rho(x^k) = \frac{\|P_k g_0(x^k)\|^2 + \omega_k}{1 + |e^T \pi(x^k, C_k)|},$$

$$\omega_k = \omega(x^k) = \sum_{j \in L_1} \max\{-\pi_j^k, \pi_j^k f_j(x^k)^2\} - \sum_{j \in L_2} (\pi_j^k + C_k) f_j(x^k), \tag{2.10}$$

$$V^k = (V_j^k, j \in L), \; V_j^k = \begin{cases} -1 - \rho_k, & j \in L_1, \pi_j^k < 0, \\ f_j(x^k)^2 - \rho_k, & j \in L_1, \pi_j^k \geq 0, \\ -f_j(x^k) - \rho_k, & j \in L_2. \end{cases} \tag{2.11}$$

If $\rho_k = 0$, then x^k is K-T point of the problem (1.1), STOP. Otherwise, compute

$$\tilde{d}^k = \rho_k \left\{ -P_k g_0(x^k) + Q_k^T \left(v^k + \frac{\rho_k}{2(1 + |e^T \pi^k|)} e \right) \right\}. \tag{2.12}$$

Step 8. Compute λ^k the first number λ in the sequence $\{1, \beta, \beta^2, \beta^3, ...\}$ satisfying:

$$F_{C_k}(x^k + \lambda\tilde{d}^k) \leq F_{C_k}(x^k) + \lambda\alpha\nabla F_{C_k}(x^k)^T\tilde{d}^k,$$

$$f_j(x^k + \lambda\tilde{d}^k) \leq 0, \ \forall j \in L. \tag{2.13}$$

Step 9. Set

$$x^{k+1} = x^k + \lambda^k\tilde{d}^k, \ \gamma_{k+1} = \gamma_k. \tag{2.14}$$

Set $k := k + 1,$ and go back to Step 1.

3. Convergence of Algorithm

Lemma 3.1. (i) *Feasible point* x^k *is a K-T point of the problem* (1.1), *if and only if* $\rho_k = 0$;

(ii) $g_j(x^k)^T\tilde{d}^k \leq -\frac{1}{2}\rho_k^2, \ j \in J(x^k) \cup \{0\}$;

(iii) *If* $\{x^k\}$ *is bounded, then there exists* $C_0 > 0$, *such that* $g_j(x^k)^T\tilde{d}^k \leq -C_0\|\tilde{d}^k\|^2, \ j \in J(x^k) \cup \{0\}$.

Lemma 3.2. *If sequence* $\{x^k\}$ *is bounded, then there exists* $k_0 > 0$, *such that, for all* $k \geq k_0$, $C_k \equiv C_{k_0}$. (We denote $C_k \equiv C$ in the remainder of this paper.)

Theorem 3.1. *Under assumptions* H2.1, H2.2, *the algorithm either stops at a K-T point* x^k *of the problem* (1.1) *in finite iterations or generates an infinite sequence* $\{x^k\}$, *whose any accumulation point* x^* *is a K-T point of the problem* (1.1).

Proof. If the method stops at $\{x^k\}$, from the algorithm, it is easy to see that x^k is a K-T point of (1.1). Assume now that the algorithm generates an infinite sequence $\{x^k\}$, and x^* is a limit point, i.e., there exists $K(|K| = \infty)$, such that

$$\lim_{k \in K} x^k = x^*, \ J_k \equiv J_0, \ \forall k \in K.$$

There is one of the following cases obtained:

Case A. $x^{k+1} = x^k + d^k + \tilde{d}^k$ is generated by (2.9), $\forall k \in K$. Firstly, it is obvious that $J_k = J_0 \subseteq J(x^*)$. So, from K-T conditions, there exist multipliers λ_j^k such that

$$\nabla F_C(x^k) + H_k d_k + \sum_{j \in J_k} \lambda_j^k g_j(x^k) = 0. \tag{3.1}$$

By the first inequality of (2.7), we have, for $k \in K$, $k \to \infty$, that

$$\lambda_{J_k}^k = -(N_k^T N_k + D_k)^{-1} N_k^T (\nabla F_C(x^k) + H_k d_k)$$

$$\to -(N_*^T N_* + D_*)^{-1} N_*^T (\nabla F_C(x^*)) = \lambda_{J_0}^*.$$

So

$$\nabla F_C(x^*) + \sum_{j \in J_0} \lambda_{J_0}^* g_j(x^*) = 0, \ (x^*, \lambda^*), \ \lambda^* = \begin{cases} \lambda_{J_0}^*, & J_0, \\ 0, & L \setminus J_0, \end{cases} \tag{3.2}$$

then (x^*, λ^*) is K-T point pair of the problem (1.2). From Lemma 2.1 and the definition of the parameter C_k, it is easy to see that x^* is K-T point of the problem (1.1).

Case B. $x^{k+1} = x^k + \lambda^k \tilde{d}^k$ is generated by (2.14), $\forall k \in K$. Here, imitating the proof in [3], it holds that x^* is also a K-T point of the problem (1.1).

In order to obtain the superlinear convergence rate, we make the following additional assumption.

H3.1. Assume $\{x^k\}$ exists a limit point x^*, and $\lim_{k \to \infty} H_k = \nabla_{xx}^2 L$ (x^*, u^*) which is positive definite. The second-order sufficiency conditions with strict complementary slackness are satisfied at the K-T point pair (x^*, u^*).

Under the general analysis to SQP type methods, we have the following results.

Lemma 3.3. *Under all stated assumptions, it holds that* $x^k \to x^*$,

$d^k \to 0$. *For k large enough, there holds*

$$L_k = \{j \in L \mid f_j(x^k) + f_j(x^k)^T d^k = 0\} \equiv J_*, \ \|\widetilde{d}^k\| = O(\|d^k\|)^2. \qquad (3.3)$$

Lemma 3.4. *For k large enough, the trust region iteration of the algorithm is always successful, i.e., $x^{k+1} = x^k + d^k + \widetilde{d}^k$ is all generated by* (2.9), *the algorithm does not go to Step* 7, *Step* 8 *and Step* 9.

According to Lemma 3.4, it is easy to obtain the superlinear convergence rate.

Theorem 3.2. *Under all stated assumptions, the algorithm is superlinearly convergent, i.e., the sequence $\{x^k\}$ generated by the algorithm satisfies $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$.*

Acknowledgement

This work was supported by the National Natural Science Foundation (No. 10501009) of China.

References

[1] R. H. Byrd, N. I. M. Gould, J. Nocedal and R. A. Waltz, An algorithm for nonlinear optimization using linear programming and equality constrained subproblems, Math. Program. Ser. B 100(1) (2004), 27-48.

[2] R. Eliane and A. L. Tits, Superlinearly convergent feasible method for the solution of inequality constrained optimization problems, SIAM J. Control Optim. 25(4) (1987), 934-950.

[3] Jinbao Jian, A feasible trust region algorithm with fast convergence for nonlinear inequality constrained optimization, Math. Numerica/Sinica 24 (2002), 273-282.

[4] Jin-Bao Jian, Chun-Ming Tang, Qing-Jie Hu and Hai-Yan Zheng, A feasible descent SQP algorithm for general constrained optimization without strict complementarity, J. Comp. Appl. Math. 180 (2005), 391-412.

[5] D. Q. Mayen and E. Polak, Feasible directions algorithms for optimization problems with equality and inequality constraints, Math. Program. 11 (1976), 67-80.

[6] M. J. D. Powell and Y. Yuan, A trust region algorithm for optimization with equality constrained optimization, Math. Program. 49 (1991), 189-211.

[7] Xiaojiao Tong and Shuzi Zhou, A trust region algorithm for nonlinear constrained optimization problem, Math. Program. 24(3) (2004), 445-460.

■