Journal of Computational Mathematics, Vol.21, No.2, 2003, 207–220.

A TRUST-REGION ALGORITHM FOR NONLINEAR INEQUALITY CONSTRAINED OPTIMIZATION*[1)]

Xiaojiao Tong
(*Department of Mathematics and Computer, Changsha University of Electric Power, Changsha 410077*)

Shuzi Zhou
(*College of Mathematics and Econometrics, Hunan University, Changsha 410082*)

Abstract

This paper presents a new trust-region algorithm for n-dimension nonlinear optimization subject to m nonlinear inequality constraints. Equivalent KKT conditions are derived, which is the basis for constructing the new algorithm. Global convergence of the algorithm to a first-order KKT point is established under mild conditions on the trial steps, local quadratic convergence theorem is proved for nondegenerate minimizer point. Numerical experiment is presented to show the effectiveness of our approach.

Key words: Inequality constrained optimization; Trust-region method; Global convergence; Local quadratic convergence.

1. Introduction

In this paper, we study the following nonlinear inequality constrained optimization problem:

$$\begin{cases} min & f(x) \\ s.t & H(x) \leq 0, \end{cases} \tag{1.1}$$

where $H(x) = (h_1(x), h_2(x), \cdots, h_m(x))^T$, f(x) and $h_i(x)$, $i \in I = \{1, 2, \cdots, m\}$, are $R^n \to R$ twice continuously differentiable. We assume $m \leq n$ in this paper, which is important for our argument.

Trust-region algorithms are very efficient for solving nonlinear equality constrained problems. (see, [1], [2], [5], [11], for example). However, for nonlinear inequality constrained optimization problem, the results about trust region method are very few, (see [4], [6], [9], [12], [7], for example), and there are still some unsolved problems now. The paper [7] deals with inequality using slack variables and finally only discusses equality constraints and bound constraints. The convergence to so-called φ-stationary point has been proved. The paper [9] presents trust region method for an arbitrary closed set and proves the global convergence theorem. But it is very difficult to solve the subproblems arisen in the algorithm of [9]. Very general problems have been discussed in [12]. The basic idea of [12] is to reduce the smooth constrained optimization problem into a nonsmooth unconstrained problem by using l_∞ exact penalty function and then to solve the nonsmooth problem by trust region method. Global convergence of the method has been proved under the assumption that the penalty parameter is bounded. When the penalty parameter tends to infinity, the method of [12] is still convergent, but the limit is not the KKT point of the original problem. [3] discusses an interior Newton method, [4] gives a trust region approaches for nonlinear optimization only for a special case, that is the optimization problem

---

* Received February 23, 2000.

1) The work was partly supported by the Scientific Research Foundation of Hunan Provincial Education Department, No. 02B021, and the National Natural Science Foundation of China, No. 10171008.

with bounded constraints. [6] extends the method of [3] and [4] to the problem with bound constraints for partial variables and equality constraints.

This paper presents a new trust region method for nonlinear optimization with inequality constraints. We change the problem into an equivalent problem with equality constraints and non-negative constraints by using slack variables. Then we derive a new equivalent KKT conditions which is the basis for constructing our algorithm. The subproblems in the algorithm can be solved by the method proposed in [5] and [6]. We have proved that at least one accumulation point of the new algorithm is a first- order KKT point. The local quadratic convergence for nondegenerate minimizer point has been shown.

The problem proposed by this paper is different from [6], Assume p is the number of bound constraints. [6] requires that $m \leq n$ and $p = n - m$. Hence the problem of [6] can be reduced to an optimization problem with simply bound constraints. In our paper, by introducing slack variables problem (1.1) is changed into problem (2.1), where the number p of bound constraints is m. So our problem is different from the problem in [6].

The paper is organized as follows. In section 2, we derive an equivalent first-order KKT condition; In section 3, we discuss the Newton's method of the KKT equations; We present a method to compute trial step in section 4; In section 5, the new trust-region algorithm is formulated; The global convergence theorem of the algorithm is given in section 6; Section 7 makes local analysis for the algorithm; The last section is numerical test.

In this paper, the vector and matrix norms used are l_2 norm, subscripted indices k represents the evaluation of a function at a particular point. For example, f_k represents $f(x_k)$ and so on.

2. Optimality conditions

By introducing slack variables $s \in R^m$, (1.1) is transformed equivalently to the following problem for the variables $x \in R^n$ and $s \in R^m$:

$$\begin{cases} min & f(x) \\ s.t & H(x) + s = 0, \quad s \geq 0. \end{cases} \tag{2.1}$$

Denote $u = (x^T, s^T)^T \in R^{m+n}, C(u) = H(x) + s \in R^m$,

$$l(x, s, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i (h_i(x) + s_i),$$

$$A(x) = (\nabla h_1(x), \nabla h_2(x), \cdots, \nabla h_m(x)) \in R^{n \times m},$$

$J(u) = (A^T(x), I_m) \in R^{m \times (n+m)}$, which is Jacobian of $C(u)$. $I_m \in R^{m \times m}$is unit matrix.

A point $u^* = ((x^*)^T, (s^*)^T)^T$ satisfies the first -order KKT conditions of problems (2.1) if there exist $\lambda^* \in R^m$ such that:

$$\begin{cases} H(x^*) + s^* = 0, \quad s^* \geq 0, \\ \nabla f(x^*) + A(x^*)\lambda^* = 0, \\ s_i^* > 0 \quad \Longrightarrow \lambda_i^* = 0; \quad s_i^* = 0 \quad \Longrightarrow \lambda_i^* \geq 0 \quad i \in I. \end{cases} \tag{2.2}$$

We assume $m \leq n, rankA(x) = m$ in this paper. So the constrained qualification is satisfied. From QR factorization of $A(x)$

$$A(x) = (Y(x), Z(x)) \begin{pmatrix} R(x) \\ 0 \end{pmatrix}, \tag{2.3}$$

(2.2) is equivalent to

$$\begin{aligned} & H(x^*) + s^* = 0, \; s^* \geq 0, \; Z(x^*)^T \nabla f(x^*) = 0, \\ & s_i^* > 0 \;\; \Longrightarrow [-R(x^*)^{-1} Y(x^*)^T \nabla f(x^*)]_i = 0, \\ & s_i^* = 0 \;\; \Longrightarrow [-R(x^*)^{-1} Y(x^*)^T \nabla f(x^*)]_i \geq 0. \end{aligned} \tag{2.4}$$

Denote $\nabla F(u) = (\nabla f(x)^T, 0_m^T)^T \in R^{n+m}$ and introduce a matrix $D(u) \in R^{n\times n}$ as follows:

$$D(u) = \begin{pmatrix} \tilde{D}(u) & 0 \\ 0 & I_{n-m} \end{pmatrix} \tag{2.5}$$

where $\tilde{D}(u) \in R^{m\times m}$ is a diagonal matrix, and its diagonal elements are

$$(\tilde{D}(u))_{ii} = \begin{cases} 1, & if [-R(x)^{-1}Y(x)^T\nabla f(x)]_i < 0, \\ \sqrt{s_i}, & if [-R(x)^{-1}Y(x)^T\nabla f(x)]_i \geq 0 \end{cases}$$

Introduce another matrix $W(u) \in R^{(n+m)\times n}$:

$$W(u) = \begin{pmatrix} -Y(x)R(x)^{-T} & Z(x) \\ I_m & 0_{m\times(n-m)} \end{pmatrix}$$

Obviously, $rankW(u) = n$,$J(u)W(u) = 0$, the columns of W(u) form a basis for the null space of $J(u)$. Hence, we have the following proposition.

Proposition 2.1 *The point $u^* = ((x^*)^T, (s^*)^T)^T$ is a first order KKT point of (2.1) if and only if u^* satisfies*

$$\begin{cases} C(u^*) = 0, \quad s^* \geq 0, \\ D(u^*)^2 W(u^*)^T \nabla F(u^*) = 0. \end{cases} \tag{2.6}$$

Definition 2.1 *A point $u \in R^{n+m}$ is nondegenerate for problem (1.1) if $[-R(x)^{-1}Y(x)^T\nabla f(x)]_i = 0$ implies $s_i > 0$ for all $i \in \{1, 2, \cdots, m\}$.*

Define two diagonal matrixs $E(u) \in R^{n\times n}$ and $\tilde{E}(u) \in R^{m\times m}$ by

$$E(u) = \begin{pmatrix} \tilde{E}(u) & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.7}$$

where the diagonal elements of $\tilde{E}(u)$ are given by

$$(\tilde{E}(u))_{ii} = \begin{cases} [-R(x)^{-1}Y(x)^T\nabla f(x)]_i, & \text{if } [-R(x)^{-1}Y(x)^T\nabla f(x)]_i > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

From the definitions of $D(u)$ and $W(u)$, we have the following proposition.

Proposition 2.2 *If a nondegenerate point u^* satisfies the second-order sufficient KKT condition of (2.1),then*

$$D(u^*)W(u^*)^T\nabla_u^2 l(x^*, s^*, \lambda^*)W(u^*)D(u^*) + E(u^*), \tag{2.9}$$

is positive definite, where $\lambda^* = -R(x^*)^{-1}Y(x^*)^T\nabla f(x^*)$.

Proof. From the second-order sufficient condition $\nabla_u^2 l(x^*, s^*, \lambda^*)$ is positive definite on

$$\Omega^* = \{d \in R^{n+m} | d^T\nabla_u C(u^*) = 0, d^T e_i = 0, i \in I(s^*)\}, I(s^*) = \{i | s_i^* = 0\}.$$

Let $d = W(u^*)D(u^*)d^t, d^t \in R^n$. It is easy to prove that $d \in \Omega^*$. Hence, $\forall d^t \in R^n$

$$(d^t)^T D(u^*)W(u^*)^T\nabla_u^2 l(x^*, s^*, \lambda^*)W(u^*)D(u^*)d^t \geq 0, \tag{2.10}$$

where the equality holds only for $d = 0$. It follows from nondegenerate condition that $(d^t)^T E(u^*)\ d^t > 0$ for $d = 0$ and $d^t \neq 0$, which combing with (2.10) implies (2.9). □

3. Newton's interior-point method

In this section we consider Newton's interior-point method to solve (2.6), i.e.,

$$\begin{aligned} &C(u) = 0, \\ &D(u)^2 W(u)^T \nabla F(u) = 0, \end{aligned} \tag{3.1}$$

with $s > 0$.

Since $rank(W(u)) = n$, we make a factorization of d_u by

$$d_u = d^n + W(u)d^t \tag{3.2}$$

with $d^n = ((d^n)_x^T, 0^T)^T \in R^{n+m}, (d^n)_x \in R^n, d^t \in R^n$. We require that s is strictly feasible. By using the similar analysis in [3], [4] and [6], combining with (3.2), $J(u)W(u) = 0, J(u) = (A(x)^T, I)$, we obtain the system of equation for Newton-like step $d_u = d^n + W(u)d^t$ by

$$\begin{aligned} A(x)^T(d^n)_x &= -C(u), \\ [\bar{D}(u)W(u)^T \nabla_u^2 l(x,s,\lambda) W(u) \bar{D}(u) + E(u)] \bar{D}^{-1}(u) d^t &= -\bar{D}(u)\bar{g}(u), \end{aligned} \tag{3.3}$$

where $E(u)$ is defined by (2.7).

$$\bar{g} = W(u)^T[\nabla_u^2 l(x,s,\lambda)d^n + \nabla F(u)] = \begin{pmatrix} \bar{g}_1(u) \\ \bar{g}_2(u) \end{pmatrix}$$

and $\bar{g}_1(u) \in R^m, \bar{g}_2(u) \in R^{n-m}$. $\bar{D}(u) \in R^{n \times n}$ and $\hat{D}(u)$ are diagonal matrices

$$\bar{D}(u) = \begin{pmatrix} \hat{D}(u) & 0 \\ 0 & I_{n-m} \end{pmatrix} \qquad (\hat{D}(u))_{ii} = \begin{cases} 1, & \text{if } (\bar{g}_1(u))_i < 0, \\ \sqrt{s_i}, & \text{if } (\bar{g}_1(u))_i \geq 0. \end{cases} \tag{3.4}$$

4. Trial steps

(3.3) is a basis that we make a trust region algorithm. At each iteration k, Δ_k is a trust region radius, we compute the trial step d_k. If d_k is accepted, we set $u_{k+1} = u_k + d_k$, where

$$u_k = \begin{pmatrix} x_k \\ s_k \end{pmatrix}, \quad d_k = \begin{pmatrix} (d_k)_x \\ (d_k)_s \end{pmatrix}.$$

We require $s_k > 0, s_{k+1} > 0$. The trial step d_k is determined by $d_k = d_k^n + W_k d_k^t$, where d_k^n is the quasi-normal component, $W_k d_k^t$ is the tangential component with respect to the null space of J_k and $d_k^t \in R^n$.

4.1. The quasi-normal component

In order to keep $s_k > 0$, the quasi-normal component is chosen by $d_k^n = \binom{(d_k^n)_x}{0}$ and related to the trust-region subproblem

$$\begin{cases} min \ \frac{1}{2}\|C_k + A_k^T(d^n)_x\|^2 \\ s.t \quad \|(d^n)_x\| \leq \tau\Delta_k. \end{cases} \tag{4.1}$$

where $\tau \in (0,1)$ is a constant independent of k, $J_k = (A_k^T, I_m)$

Same as in ([5] and [6]), (4.1) is not to be solved exactly, it is only required that

$$\|d_k^n\| = \|(d_k^n)_x\| \leq \kappa_1 \|C_k\|, \tag{4.2}$$

$$\|C_k\|^2 - \|C_k + J_k d_k\|^2 = \|C_k\|^2 - \|C_k + A_k^T(d_k^n)_x\|^2 \geq \kappa_2 \|C_k\| min\{\kappa_3\|C_k\|, \tau\Delta_k\}, \tag{4.3}$$

where $\kappa_1, \kappa_2, \kappa_3$ are positive constants. (4.3) just is a weaker form of Cauchy decrease condition for (4.1) (see [6]). [5] has provided algorithms to compute d_k^n satisfying (4.2) and (4.3).

4.2. the tangential component

Denote $d^t = ((\bar{d}^t)^T, (\hat{d}^t)^T)^T$, where $\bar{d}^t \in R^m, \hat{d}^t \in R^{n-m}$. We have $\bar{d}^t = d_s$ since $d = d^n + W_k d^t = ((d_x^T, d_s^T)^T$ we have $\bar{d}^t = d_s$. From (3.3) the trust region subproblem of tangential component is

$$\begin{cases} min & \bar{g}_k^T d^t + \frac{1}{2}(d^t)^T(W_k^T B_k W_k + E_k \bar{D}_k^{-2})d^t \\ s.t & \|\bar{D}_k^{-1} d^t\| \leq \Delta_k, \end{cases} \tag{4.4}$$

where $B_k \in R^{(n+m)\times(n+m)}$ is a symmetric matrix, which is an approximation to the Hessian matrix $\nabla_u^2 l(x_k, s_k, \lambda_k)$. Note that if trust restriction is inactive, the solution of (4.4) is the solution of (3.3) with $B_k = \nabla_u^2 l_k$. In order to require $s_k + (d_k)_s > 0$, we choose $\sigma_k \in (\sigma, 1], \sigma \in (0, 1)$ and compute it with

$$\bar{d}_k^t = (d_k)_s \geq -\sigma_k s_k, \tag{4.5}$$

(4.5) can be satisfied by scaling technique, for example, let

$$\tau_k = \sigma_k min\{1, min\{\frac{-(s_k)_i}{(d_k)_{si}}, (d_k)_{si} < 0\}\}, \tag{4.6}$$

where $(d_k)_{si}$ expresses the ith component of $(d_k)_s$.

Denote

$$\begin{aligned} q_k(d) &= l_k + \nabla_u l_k^T d + \frac{1}{2} d^T B_k d, \\ \psi(d) &= q_k(d_k^n + W_k d) + \frac{1}{2} d^T (E_k \bar{D}_k^{-2}) d, \end{aligned} \tag{4.7}$$

then (4.4) can be rewritten as

$$\begin{cases} min & \psi_k(d^t) \\ s.t & \|\bar{D}_k^{-1} d^t\| \leq \Delta_k. \end{cases} \tag{4.8}$$

Again we do not need to compute d_k^t exactly, only compute d_k^t with $\bar{d}_k^t = (d_k)_s \geq -\sigma_k s_k$ and satisfy a fraction Cauchy decrease for the subproblem (4.8), i.e.,

$$\psi_k(0) - \psi_k(d_k^t) \geq \beta[\psi_k(0) - \psi_k(v_k^d)], \tag{4.9}$$

where $\beta \in (0, 1)$ is a constant independent of k, $v_k^d \in R^n$ is a solution of the following problem:

$$\begin{cases} min \ \psi_k(v^d) \\ s.t \quad \|\bar{D}_k^{-1} v^d\| \leq \Delta_k, v^d \in Span\{-\bar{D}_k^2 \bar{g}_k\}, \quad \bar{v}^d \geq -\sigma_k s_k, \end{cases} \tag{4.10}$$

where $\bar{v^d} \in R^m, \hat{v}^d \in R^{n-m}, v^d = ((\bar{v}^d)^T, (\hat{v}^d)^T)^T$. It is obvious that $s_k + \bar{v}_k^d > 0$.

4.3. Calculate Lagrange multiplier λ_{k+1} and choose merit function

From (2.3) we have the following formula for calculating Lagrange multiplier

$$\lambda_{k+1} = -R(x_k + (d_k)_x)^{-1} Y(x_k + (d_k)_x)^T \nabla f(x_k + (d_k)_x) \tag{4.11}$$

We use augmented Lagrangian as a merit function:

$$\Phi(x, s, \lambda; \rho) = f(x) + \sum_{i=1}^{m} \lambda_i (h_i(x) + s_i) + \rho \|C(u)\|^2, \tag{4.12}$$

where $\rho > 0$ is a penalty parameter.

At kth iteration, the actual reduction and predicted reduction are defined by

$$ared(d_k; \rho_k) = \Phi(x_k, s_k, \lambda_k; \rho_k) - \Phi(x_k + (d_k)_x, s_k + (d_k)_s, \lambda_{k+1}; \rho_k), \tag{4.13}$$

$$pred(d_k; \rho_k) = q_k(0) - q_k(d_k) - \Delta\lambda_k^T (J_k d_k + C_k) + \rho_k [\|C_k\|^2 - \|J_k d_k + C_k\|^2], \tag{4.14}$$

where $\Delta\lambda_k = \lambda_{k+1} - \lambda_k$.

5. Statement of algorithm

Algorithm 5.1.

Step 0. Set $u_0 = (x_0^T, s_0^T)^T, s_0 > 0, \Delta_0 > 0$; set $\rho_{-1} \geq 1, B_0 \in R^{(n+m)\times(n+m)}$ is a symmetric matrix, $a_1, \tau \in (0,1), \bar{\rho} > 0, \Delta_{max} \geq \Delta_{min} > 0, \eta \in (0,1)$ and $\varepsilon > 0$,k:=0.

Step 1. If $\|C_k\| + \|D_k W_k^T \nabla F_k\| \leq \varepsilon$, then stop.

Step 2. Compute d_k^n to satisfy (4.2) and (4.3); compute d_k^t to satisfy (4.5) and (4.9); $d_k := d_k^n + W_k d_k^t$.

Step 3. Compute $\lambda_k, \lambda_{k+1}, \Delta\lambda_k := \lambda_{k+1} - \lambda_k$.

Step 4. Compute $pred(d_k; \rho_{k-1})$.

If $pred(d_k; \rho_{k-1}) \geq \frac{\rho_{k-1}}{2}[\|C_k\|^2 - \|J_k d_k + C_k\|^2]$ then $\rho_k := \rho_{k-1}$;

otherwise

$$\rho_k := \frac{2[q_k(d_k) - q_k(0) + \Delta\lambda_k^T(J_k d_k + C_k)]}{\|C_k\|^2 - \|J_k d_k + C_k\|^2} + \bar{\rho}. \tag{5.1}$$

Step 5. Compute $ared(d_k;\rho_k), pred(d_k;\rho_k)$. If $\frac{ared(d_k;\rho_k)}{pred(d_k;\rho_k)} < \eta$ then $\Delta_k := a_1 \max\{\|d_k^n\|, \|\bar{D}_k^{-1}\ d_k^t\|\}$ and goto step 2

otherwise $\Delta^k := \Delta_k$. Choose Δ_{k+1} such that $\Delta_{max} \geq \Delta_{k+1} \geq max\{\Delta_{min}, \Delta^k\}$.

Step 6. $x_{k+1} := x_k + (d_k)_x; s_{k+1} := s_k + (d_k)_s$, compute B_{k+1}, k:=k+1 goto step 1.

Remark 5.1. From step 4 we easy prove the following statements:

$$\rho_k \geq \rho_{k-1} \geq 1, \tag{5.2}$$

$$pred(d_k;\rho_k) \geq \frac{\rho_k}{2}[\|C_k\|^2 - \|J_k d_k + C_k\|^2]. \tag{5.3}$$

6. Global convergence

6.1. Assumptions of global convergence

In order to establish the global convergence of Algorithm 5.1, we need some assumptions (compare with [5]-[6]).

AS.1 For all $k, u_k, u_k + d_k \in \Omega \subset R^{n+m}$, where $\Omega = \binom{\Omega_x}{\Omega_s}$, $\Omega_x \in R^n$, $\Omega_s = R_+^m$ and Ω_x is an open convex set of R^n.

AS.2 $f(x), h_i(x) (i \in I)$ are twice continuously differentiable on Ω_x

AS.3 For any $x \in \Omega_x, rank A(x) = m$.

AS.4 $f(x), \nabla f(x), \nabla^2 f(x), h_i(x), Y(x), Z(x), R(x), R^{-1}(x), \nabla^2 h_i(x)$ are uniformly bounded in Ω_x.

AS.5 $\{B_k\}, \{s_k\}$ are bounded.

From the assumptions above, there exist constants $\nu_i > 0, (i = 1, 2, \cdots, 10)$ independent of k and u such that

$$\begin{cases} \|\nabla f(x)\| \leq \nu_1, \|Y_k\| \leq \nu_2, \|R_k^{-1}\| \leq \nu_3, \|R_k\| \leq \nu_4, \|Z_k\| \leq \nu_5, \\ \|W_k\| \leq \nu_6, \|\bar{D}_k\| \leq \nu_7, \|B_k\| \leq \nu_8, \|\lambda_k\| \leq \nu_9, \|J_k\| \leq \nu_{10}. \end{cases} \tag{6.1}$$

From now on we assume that above assumptions hold.

6.2. Intermedium results

Lemma 6.1. *Assume that d_k is computed by algorithm 5.1. Then for each step d_k we have*

$$max\{\|d_k^n\|, \|\bar{D}_k^{-1} d_k^t\|\} \geq \frac{1}{1 + \nu_6\nu_7}\|d_k\|. \tag{6.2}$$

Proof. (6.2) follows from

$$\|d_k\| = \|d_k^n + W_k d_k^t\| \le \|d_k^n\| + \nu_6\|d_k^t\| \le \|d_k^n\| + \nu_6\nu_7\|\bar{D}_k^{-1} d_k^t\|.$$

□

Lemma 6.2. *Supposed that d_k^t is an approximate solution of (4.8) and satisfies (4.5) and (4.9). Then we have*

$$q_k(d_k^n) - q_k(d_k^n + W_k d_k^t) \ge \kappa_4\|\bar{D}_k\bar{g}_k\| min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\}, \tag{6.3}$$

where $\kappa_4, \kappa_5, \kappa_6$ are positive constants independent of k.

Proof. The proof is similar to the proof of Lemma 6.2 in [6]. □

Lemma 6.3. *Assume that d_k is computed by Algorithm 5.1. Then*

$$\begin{aligned} pred(d_k;\rho) \quad & \ge \kappa_4\|\bar{D}_k\bar{g}_k\| min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\} - \kappa_7\|C_k\| \\ & +\rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2]. \end{aligned} \tag{6.4}$$

Proof. This result is proved directly from (4.14), $d_k = d_k^n + W_k d_k^t$, global assumptions, (4.2) and Lemma 6.2 with $\kappa_7 = [\nu_1 + (\nu_2\nu_4 + 1)\nu_9 + \frac{1}{2}\nu_8\tau\Delta_{max}]\kappa_1 + 2\nu_9$. □

Lemma 6.4. *There is a positive constant κ_8 independent of k such that*

$$|ared(d_k;\rho_k) - pred(d_k;\rho_k)| \le \kappa_8\rho_k\|d_k\|^2. \tag{6.5}$$

Proof. The lemma is proved by mean-value theorem. □

6.3. Global convergence

Lemma 6.5. *If $\|C_k\| \le \alpha\Delta_k$, $\|\bar{D}_k\bar{g}_k\| + \|C_k\| > \epsilon$ and α is a constant such that*

$$\alpha \le min\{\frac{\epsilon}{3\Delta_{max}}, \frac{\kappa_4\epsilon}{3\kappa_7} min\{\frac{2\kappa_5\epsilon}{3\Delta_{max}}, \kappa_6\}\}, \tag{6.6}$$

then

$$pred(d_k;\rho) \ge \frac{\kappa_4}{2}\|\bar{D}_k\bar{g}_k\| min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\} + \rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2], \tag{6.7}$$

$$pred(d_k;\rho) \ge \kappa_9\Delta_k, \tag{6.8}$$

$$\rho_k = \rho_{k-1}, \tag{6.9}$$

where κ_9 is a constant independent of k.

Proof. From $\|\bar{D}_k\bar{g}_k\| + \|C_k\| \ge \epsilon$ and (6.6) we have $\|C_k\| \le \frac{\epsilon}{3}, \|\bar{D}_k\bar{g}_k\| > \frac{2}{3}\epsilon$. Then it follows from Lemma 6.3 and (6.6) that

$$\begin{aligned} pred(d_k;\rho) \ge & \frac{\kappa_4}{2}\|\bar{D}_k\bar{g}_k\| min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\} + \frac{1}{3}\epsilon\kappa_4\Delta_k min\{\frac{2\kappa_5\epsilon}{3\Delta_{max}}, \kappa_6\} \\ & - \kappa_7\alpha\Delta_k + \rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2] \\ & \ge \frac{\kappa_4}{2}\|\bar{D}_k\bar{g}_k\| min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\} + \rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2]. \end{aligned}$$

(6.7) is proved. From (4.3) and (6.7) we have (6.8) with $\kappa_9 = \frac{\kappa_4\epsilon}{3} min\{\frac{2\kappa_5\epsilon}{\Delta_{max}}, \kappa_6\}$. Moreover, (6.7) implies that $pred(d_k;\rho_{k-1}) \ge \rho_{k-1}[\|C_k\|^2 - \|J_k d_k + C_k\|^2]$. Then from step 4 of Algorithm 5.1 it follows (6.9). □

Theorem 6.1. *The algorithm is valid, i.e., the interior loop (step2-step5) can be ended in finite times at each iteration.*

Proof. Let the index of interion loop be i. Then corresponding values are

$$\Delta_{k,i}, d_{k,i}, \rho_{k,i}, pred(d_{k,i};\rho_{k,i}), ared(d_{k,i};\rho_{k,i}).$$

We note that C_k, J_k, W_k are not changed in interior loop. The proof is by contradiction. If $i \to +\infty$ then from algorithm we have $\Delta_{k,i} \to 0$ and

$$|\frac{ared(d_{k,i};\rho_{k,i})}{prad(d_{k,i};\rho_{k,i})} - 1| > 1 - \eta. \tag{6.10}$$

Case(i), $\|C_k\| \neq 0$. From (5.3) and (4.3) we know

$$pred(d_{k,i};\rho_{k,i}) \geq \frac{\rho_{k,i}}{2}[\|C_k\|^2 - \|C_k + J_k d_{k,i}\|^2] \geq \frac{\rho k, i}{2}\kappa_2\|C_k\| min\{\tau, \frac{\kappa_3\|C_k\|}{\Delta_{max}}\}\Delta_{k,i}, \tag{6.11}$$

which combining with Lemma 6.4 to yield

$$|\frac{ared(d_{k,i};\rho_{k,i})}{pred(d_{k,i};\rho_{k,i})} - 1| \leq \frac{2\kappa_8}{\kappa_2\|C_k\| min\{\tau, \frac{\kappa_3\|C_k\|}{\Delta_{max}}\}}\Delta_{k,i} \to 0. \quad (i \to \infty)$$

This contradicts with (6.10).

Case(ii), $\|C_k\| = 0$. It follows from (4.2), definition of $\bar{g}$ and (3.8) that $d^n_{k,i} = 0$, and

$$\bar{g}_k = \begin{pmatrix} \bar{g}_{k,1} \\ \bar{g}_{k,2} \end{pmatrix} = \begin{pmatrix} -R_k^{-1}Y_k^T\nabla f_k \\ Z_k^T\nabla f_k \end{pmatrix},$$

then $\hat{D}_k = \tilde{D}_k, D_k = \bar{D}_k, \|D_k W_k^T \nabla F_k\| = \|\bar{D}_k\bar{g}_k\|$. Therefore, it follows from Lemma 6.4 and lemma 6.5 that $\rho_{k,i} = \rho_{k-1}$ and

$$|\frac{ared(d_{k,i};\rho_{k-1})}{pred(d_{k,i};\rho_{k-1})} - 1| \leq \frac{\kappa_8\rho_{k-1}\|d_{k,i}\|^2}{\kappa_9\Delta_{k,i}} \to 0,$$

which contradicts with (6.10) again. The proof is completed. □

Lemma 6.6. *If $\|\bar{D}_k\bar{g}_k\| + \|C_k\| > \epsilon$ for all k, then the sequence $\{\rho_k\}$ and $\Phi(x_k, s_k, \lambda_k; \rho_k)$ are bounded. Furthermore, there exists a constant Δ^* independent of k such that*

$$\Delta^k \geq \Delta^*, \tag{6.12}$$

where Δ^k is the accepted radius of trust region method at kth iteration.

Proof. See Lemma 7.11, Lemma 7.12 and Lemma 8.2 of paper [5]. □

Theorem 6.2. *The sequences of $\{u_k\}$ generated by Algorithm 5.1 satisfies*

$$\lim_{k\to+\infty} inf[\|D_k W_k^T\nabla F_k\| + \|C_k\|] = 0. \tag{6.13}$$

Proof. First we proof

$$\lim_{k\to+\infty} inf[\|\bar{D}_k\bar{g}_k\| + \|C_k\|] = 0. \tag{6.14}$$

The proof of (6.14) is by contradiction. Suppose for all k $\|\bar{D}_k\bar{g}_k\| + \|C_k\| > \epsilon$. We discuss two cases: (i) $\|C_k\| \leq \alpha\Delta^k; (ii)\|C_k\| > \alpha\Delta^k$, where α is defined by (6.6).

Case (i). From Lemma 6.5 and Lemma 6.6 we have $pred(d_k;\rho_k) \geq \kappa_9\Delta^k \geq \kappa_9\Delta^*$.

Case (ii). From (5.2), (5.3), (4.3) and Lemma 6.6 we follow

$$pred(d_k;\rho_k) \geq \frac{1}{2}\kappa_2\|C_k\| min\{\tau\Delta^k, \kappa_3\|C_k\|\} \geq \frac{\kappa_2}{2}\alpha min\{\tau, \kappa_3\alpha\}(\Delta^*)^2.$$

Denote $K_{14} = min\{\kappa_9\Delta^*, \frac{\kappa_2}{2}\alpha min\{\tau, \kappa_3\alpha\}(\Delta^*)^2\}$. Then for two cases we have $pred(d_k;\rho_k)$ $\geq K_{14}$. From Algorithm 5.1 we obtain that for each k

$$\Phi_k - \Phi_{k+1} \geq \eta pred(d_k;\rho_k) \geq \eta K_{14}. \tag{6.15}$$

On the other hand, $\{\Phi(x_k, s_k, \lambda_k;\rho_k)\}$ is bounded and descent, this implies $\Phi_k - \Phi_{k+1} \to 0, (k \to \infty)$, which contradicts with (6.15). Then (6.14) holds.

Now we prove (6.13). Let $\lim_{k\in K_1}[\|\bar{D}_k\bar{g}_k\| + \|C_k\|] = 0$. We have $\lim_{k\in K_1}\|C_k\| = \lim_{k\in K_1}\|d_k^n\| = 0$ by (4.2). Because of global assumptions and the expression of $\bar{g}_k$ we have

$$\lim_{k\in K_1}\|\bar{D}_k\bar{g}_k\| = \lim_{k\in K_1}\|\bar{D}_k W_k^T\nabla F_k\| = 0. \tag{6.16}$$

Next we only show that $\bar{D}_k$ can be replaced by D_k. From expressions (2.5) and (3.4) we need only to consider $\tilde{D}_k$ and $\hat{D}_k$. Divide B_k by

$$B_k = \begin{pmatrix} B_{k1} & B_{k2} \\ B_{k3} & B_{k4} \end{pmatrix}$$

where $B_{k1} \in R^{n\times n}, B_{k2} \in R^{n\times m}, B_{k3} \in R^{m\times n}, B_{k4} \in R^{m\times m}$. Given $i \in \{1, 2, \cdots, m\}$. Assume there exists $\epsilon_1 > 0$ such that for all $k \in K_1$ holds

$$|((\tilde{D}_k - \hat{D}_k)(-R_k^{-1}Y_k^T\nabla f_k))_i| > \epsilon_1. \tag{6.17}$$

It is obvious that $(-R_k^{-1}Y_k^T\nabla f_k)_i \not\to 0$. Then there exist $\epsilon_2 > 0$ and $K_2 \subset K_1$ such that for $k \in K_2$, $|(-R_k^{-1}Y_k^T\nabla f_k)_i| > \epsilon_2$. There are two cases for k:

$$(i)(-R_k^{-1}Y_k^T\nabla f_k)_i > \epsilon_2 > 0; \quad (ii)(-R_k^{-1}Y_k^T\nabla f_k)_i < -\epsilon_2 < 0.$$

For case(i) we have $(\tilde{D}_k)_{ii} = (\sqrt{s_k})_i$. From $\lim_{k\in K_2}\|d_k^n\| = 0$ and global assumptions we have for sufficiently large $k \in K_2$

$$(\bar{g}_{k,1})_i = [-R_k^{-1}Y_k^T\nabla f_k + (-R_k^{-1}Y_k^T B_{k1} + B_{k3})(d_k^n)_x]_i > \frac{\epsilon_2}{2} > 0.$$

Hence $(\hat{D}_k)_{ii} = (\sqrt{s_k})_i = (\tilde{D}_k)_{ii}$. Similarly for case (ii) we have for sufficiently large $k \in K_2$ $(\tilde{D}_k)_{ii} = (\hat{D}_k)_{ii} = 1$. Therefore $\lim_{k\in K_2}|(\tilde{D}_k - \hat{D}_k)_{ii}| = 0$. It contradicts with (6.17). Thus it follows

$$\lim_k inf|[(\tilde{D}_k - \hat{D}_k)(-R_k^{-1}Y_k^T\nabla f_k)]_i| = 0 = \lim_k inf\|(D_k - \bar{D}_k)W_k^T\nabla F_k\| = 0,$$

which combining with (6.16) to yield $\lim_k inf\|D_k W_k^T\nabla F_k\| = 0$. Then we have (6.13). □

7. Local analysis

We study the local rate of convergence of Algorithm 5.1 in this section. It will be shown that asymptotically the trust region will be inactive under some conditions. By combining with Newton's step in our algorithm (see CS.3 below), the fast local rate of convergence will be maintained. We add the following local assumptions.

CS.1 The second derivaties of f and $h_i(i = 1, 2, \cdots, m)$ are Lipschitz continuous in Ω_x and Hessian matrix is exact, i.e., $B_k = \nabla_u^2 l(x_k, s_k, \lambda_k)$ with Lagrange multiplier $\lambda_k = -R_k^{-1}Y_k^T\nabla f_k$ for all k.

CS.2 The quasi-normal component is computed by

$$d_k^n = \begin{pmatrix} -\alpha_k Y_k R_k^{-T} C_k \\ 0 \end{pmatrix}, \tag{7.1}$$

where

$$\alpha_k = \begin{cases} 1, & \text{if } \| - Y_k R_k^{-T} C_k\| \le \tau\Delta_k; \\ \frac{\tau\Delta_k}{\|-Y_k R_k^{-T} C_k\|}, & \text{if } \| - Y_k R_k^{-T} C_k\| > \tau\Delta_k. \end{cases}$$

CS.3 If the solution of (3.3) for d^t exists with $\|\bar{D}_k^{-1} d^t\| \le \Delta_k$, which is denoted by $(d_k^t)^N$, we choose tangential component as $\tau_k^N (d_k^t)^N$, where $\tau_k^N = \sigma_k \min\{1, \min\{\frac{-(s_k)_i}{(\bar{d}_k^t)_i^N}, (\bar{d}_k^t)_i^N < 0\}\}$ and

$$(d_k^t)^N = \begin{pmatrix} (\bar{d}_k^t)^N \\ (\hat{d}_k^t)^N \end{pmatrix},\ (\bar{d}_k^t)^N \in R^m,\ (\hat{d}_k^t)^N \in R^{n-m},$$

CS.4 $u_k \to u^*$ and u^* is a nondegenerate point satisfying second-order sufficient conditions. The parameter σ_k is chosen such that $\sigma_k \ge \sigma$ and $|\sigma_k - 1| = O(\|\bar{D}_k \bar{g}_k\|)$.

Remark 7.1. It is easy proved that quasi-normal component (7.1) satisfies (4.1)-(4.3). Furthermore, it is obvious that $s_k + \tau_k^N (\bar{d}_k^t)^N > 0$.

Remark 7.2. Combining with proposition 2.1 and CS.4 we have

$$C(u^*) = 0, D(u^*)^2 W(u^*) \nabla F(u^*) = 0, -R(x^*)^{-1} Y(x^* \nabla f(x^*) \ge 0. \tag{7.2}$$

Moreover, by introducing a finite set of functions and using a similar way of analysis in [3], we can obtain that, unter local assumptions, there exists a constant $\bar{\gamma} > 0$ independent of k such that for sufficiently large k,

$$\lambda_{min}(\bar{D}_k W_k^T B_k W_k \bar{D}_k + E_k) \ge \bar{\gamma}. \tag{7.3}$$

See [3] for more details.

Lemma 7.1. *If d_k^n is chosen by (7.1), then there exists a constant $\kappa_{10} > 0$ such that*

$$\|d_k^n\| \le \kappa_{10} \|d_k\|. \tag{7.4}$$

Proof. From (7.1) and the relationship of $d_k, d_k^n, d_k^t, \hat{d}_k^t$ we follow that

$$\|d_k\|^2 \ge \|(d_k)_x\|^2 \ge \|\hat{d}_k^t\|^2, , \quad \|d_k^t\|^2 = \|(d_k)_s\|^2 + \|\hat{d}_k^t\|^2 \le 2\|d_k\|^2.$$

Then

$$\|d_k^n\| \le \|d_k\| + \|W_k d_k^t\| \le (1 + \sqrt{2}\nu_6)\|d_k\| = \kappa_{10}\|d_k\|,$$

where $\kappa_{10} = 1 + \sqrt{2}\nu_6$. □

Lemma 7.2. *There exists constant $\kappa_{11} > 0, \kappa_{12} > 0$ such that*

$$\|ared(d_k; \rho) - pred(d_k; \rho)\| \le \kappa_{11}\|d_k\|^3 + \rho\kappa_{12}\|C_k\|\|d_k\|^2. \tag{7.5}$$

Proof. It follows from mean-value theorem.

Lemma 7.3. *For sufficient large k the predicted decrease satisfies*

$$\begin{aligned} pred(d_k; \rho) \ &\ge -\kappa_{13}\|d_k\|\|C_k\| + \kappa_{14}(\|d_k\| - 2\kappa_1\kappa_{10}\|C_k\|)\|d_k\| \\ &+ \rho(\|C_k\|^2 - \|J_k d_k + C_k\|^2), \end{aligned} \tag{7.6}$$

where $\kappa_{13} > 0, \kappa_{14} > 0$ are constants independent of k.

Proof. The predicted decrease can be written as

$$pred(d_k;\rho) = [q_k(0) - q_k(d_k^n) - \Delta\lambda_k^T(J_k d_k + C_k)] - [q_k(d_k^n) - q_k(d_k^n + W_k d_k^t)] + \rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2].$$

From (7.1) it follows $\nabla_u l_k^T d_k^n = 0$, which combining with lemma 7.1 and $\|\Delta\lambda_k\| \le \kappa\|d_k\|$ to yield

$$\begin{aligned} &q_k(0) - q_k(d_k^n) - \Delta\lambda_k^T(J_k d_k + C_k) \\ =& -\frac{1}{2}(d_k^n)^T B_k d_k^n - \Delta\lambda_k(J_k d_k + C_k) \ge -\kappa_{13}\|d_k\|\|C_k\|, \end{aligned} \tag{7.7}$$

where $\kappa > 0$ is a constant, $\kappa_{13} = \frac{1}{2}\nu_8\kappa_1\kappa_{10} + \kappa$. From $\psi_k(0) - \psi(d_k^t) \ge 0$ we have

$$-\bar{g}_k^T d_k^t \ge \frac{1}{2}(d_k^t)^T(W_k^T B_k W_k + E_k\bar{D}_k^{-2})d_k^t.$$

Then for sufficiently large k, it follows from (7.3) that

$$\|\bar{D}_k\bar{g}_k\| \ge \frac{\bar{\gamma}}{2\nu_7}\|d_k^t\|. \tag{7.8}$$

On the other hand, from lemma 7.1 and (4.2) we have

$$\|d_k\|^2 \le 2\|d_k^n\|^2 + 2\nu_6\|d_k^t\|^2 \le 2\kappa_1\kappa_{10}\|C_k\|\|d_k\| + 2\nu_6\|d_k^t\|^2.$$

Hence, we obtain

$$\|d_k^t\|^2 \ge \frac{1}{2\nu_6}[\|d_k\| - 2\kappa_1\kappa_{10}\|C_k\|]\|d_k\|, \tag{7.9}$$

it follows from (6.3), $\Delta_k \ge \frac{\|d_k^t\|}{\nu_7}$ and (7.8)-(7.9) that

$$q_k(d_k^n) - q_k(d_k^n + W_k d_k^t) \ge \frac{\kappa_4\bar{\gamma}}{4\nu_6\nu_7} min\{\frac{\kappa_5\bar{\gamma}}{2\nu_7}, \frac{\kappa_6}{\nu_7}\}(\|d_k\| - 2\kappa_1\kappa_{10})\|d_k\|,$$

which combining with (7.7) to yield (7.6) with $\kappa_{14} = \frac{\kappa_4\bar{\gamma}}{4\nu_6\nu_7} min\{\frac{\kappa_5\bar{\gamma}}{2\nu_7}, \frac{\nu_6}{\nu_7}\}$. □

Lemma 7.4. *For sufficiently large k, $\{\rho_k\}$ is bounded, the trust radius Δ_k is uniform bounded away from zero and eventually all the iterations will be successful.*

Proof. First we prove the boundedness of $\{\rho_k\}$. We consider two cases (i) $\|C_k\| \le \bar{\alpha}\|d_k\|$; (ii) $\|C_k\| > \bar{\alpha}\|d_k\|$, where

$$\bar{\alpha} \le \frac{\kappa_{14}}{2(\kappa_{13} + 2\kappa_1\kappa_{10}\kappa_{14})}, \tag{7.10}$$

Case (i). From Lemma 7.3 we have

$$\begin{aligned} pred(d_k;\rho) \ &\ge \tfrac{\kappa_{14}}{2}\|d_k\|^2 + [\tfrac{\kappa_{14}}{2}\|d_k\| - (\kappa_{13} + 2\kappa_1\kappa_{10}\kappa_{14})\|C_k\|]\|d_k\| \\ &+\rho(\|C_k\|^2 - \|J_k d_k + C_k\|^2). \end{aligned} \tag{7.11}$$

From $\|C_k\| \le \bar{\alpha}\|d_k\|$ we obtain $pred(d_k;\rho) \ge \frac{\rho}{2}(\|C_k\|^2 - \|J_k d_k + C_k\|^2)$, then $\rho_k = \rho_{k-1}$ since Algorithm 5.1.

Case (ii). If ρ_k is increased then

$$\begin{aligned} &\frac{\rho_k}{2}(\|C_k\|^2 - \|J_k d_k + C_k\|^2) \\ =& q_k(d_k) - q_k(0) + \Delta\lambda_k^T(J_k d_k + C_k) + \frac{\bar{\rho}}{2}(\|C_k\|^2 - \|J_k d_k + C_k\|^2). \end{aligned}$$

From (7.7),(4.3) and $\|d_k\| \le (\tau + \nu_6\nu_7)\Delta_k$ we derive that

$$\frac{\rho_k\kappa_2}{2}\{\frac{\tau}{\tau+\nu_6\nu_7}, \kappa_3\bar{\alpha}\}\|d_k\|\|C_k\| \le (\kappa_{13}+\nu_{10})\|C_k\|\|d_k\|,$$

which combining with $\|C_k\| > \bar{\alpha}\|d_k\|$ to yield

$$\rho_k \le \frac{2(\kappa_{13}+\bar{\rho}\nu_{10})}{\kappa_2 min\{\frac{\tau}{\tau+\nu_6\nu_7}, \kappa_3\bar{\alpha}\}}.$$

Hence, for two cases $\{\rho_k\}$ is bounded. We denote $\rho_k \le \rho^*$.

Next we prove

$$\lim_{k\to+\infty}\frac{ared(d_k;\rho_k)}{pred(d_k;\rho_k)} = 1. \tag{7.12}$$

For case (i), $\|C_k\| \le \bar{\alpha}\|d_k\|$, from (7.11), $pred(d_k;\rho_k) \ge \frac{\kappa_{14}}{2}\|d_k\|^2$. It follows from Lemma 7.2 and $\|d_k\| \to 0, \|C_k\| \to 0$ that

$$|\frac{ared(d_k;\rho_k)}{pred(d_k;\rho_k)} - 1| \le \frac{\kappa_{11}\|d_k\| + \rho^*\kappa_{12}\|C_k\|}{0.5\kappa_{14}} \to 0, (k\to\infty).$$

Therefore (7.12) holds. For case (ii), because of $\|C_k\| > \bar{\alpha}\|d_k\|$, from (5.3) and (4.3) we have

$$pred(d_k;\rho_k) \ge \frac{\rho_k}{2}\kappa_2\bar{\alpha}\min\{\frac{\tau}{\tau+\nu_6\nu_7}, \kappa_3\bar{\alpha}\}\|d_k\|^2.$$

We prove (7.12) since Lemma 7.2 and $\rho_k \ge 1$. From Algorithm 5.1 we know that Δ_k is uniform bounded away from zero and every iteration will be successful for sufficiently large k. □

Lemma 7.5. *For sufficiently large k we have $\alpha_k = 1$ in (7.1) and*

$$|\tau_k^N - 1| = O(\|u_k - u^*\|). \tag{7.13}$$

Proof. This lemma is proved in a similar way of the proof of Lemma 12 in [3] and Corollary 9.4 in [6]. □

Lemma 7.6. *For sufficient large k we have that $\|\bar{D}_k^{-1}(d_k^t)^N\| \le \Delta_k$ and $\tau_k^N(d_k^t)^N$ satisfies (4.9).*

Proof From (7.3) we know $\|(\bar{D}_kW_k^TB_kW_k\bar{D}_k + E_k)^{-1}\| \le 1/\bar{\gamma}$. Then we obtain that $\|\bar{D}_k^{-1}(d_k^t)^N\| \le \Delta_k$ since $(d_k^t)^N = (-\bar{D}_k^2W_kB_kW_k + E_k)^{-1}\bar{D}_k^2\bar{g}_k$, $\lim_{k\to\infty}\|\bar{D}_k\bar{g}_k\| = 0$ and Lemma 7.4.

We also note that the solution of (4.8) is $(d_k^t)^N$ when trust restraint is inactive. Combing with positive definite of $\bar{D}_kW_k^TB_kW_k\bar{D}_k + E_k$, $|\tau_k^N - 1| = O(\|u_k - u^*\|)$ and Lemma 7.4 we have

$$\begin{aligned}
&\psi_k(0) - \psi_k(\tau_k^N(d_k^t)^N)\\
=& -\tau_k^N\bar{g}_k^T(d_k^t)^N - \frac{1}{2}(\tau_k^N)^2[(d_k^t)^N]^T(W_k^TB_kW_k + E_k\bar{D}_k^{-2})(d_k^t)^N\\
&\ge \tau_k^N[\psi_k(0) - \psi_k((d_k^t)^N)] \ge \beta[\psi_k(0) - \psi_k(v_k^d)],
\end{aligned}$$

this means that $\tau_k^N(d_k^t)^N$ satisfies (4.9). □

Theorem 7.1. *Let $\{u_k\}$ be the iterative sequence generated by Algorithm 5.1. Under the global and the local assumptions, $\{u_k\}$ convergences to u^* quadratically.*

Proof. Denote $d_k^N = d_k^n + W_k(d_k^t)^N, d_k = d_k^n + \tau_k^N W_k(d_k^t)^N$, it follows from (7.13) that

$$\|d_k - d_k^N\| = |\tau_k^N - 1| \|W_k(d_k^t)^N\| = O(\|u_k - u^*\|^2).$$

From Theorem 11 of paper [3] we have

$$\|u_{k+1} - u^*\| = \|u_k + d_k - u^*\| = O(\|u_k - u^*\|^2).$$

□

8. Numerical example

A Matlab subroutine is programming to test Algorithm 5.1. The test problem comes from the Hock and Schittkowski collection in [8]. We use the formula (7.1) to compute the quasi-normal component and use modified conjugate-gradient algorithms to compute the tangential component satisfied conditions (4.8) and (4.9). The parameters of Algorithm 5.1 are chosen as

$$\tau = 0.8, a_1 = 0.5, \eta = 0.01, \Delta_{max} = 10, \Delta_{min} = 0.01, \bar{\rho} = 0.01, \rho_{-1} = 3.$$

There are two choices for B_k,

$$B_k^{(1)} = \begin{pmatrix} \nabla_x^2 l_k & 0 \\ 0 & 0 \end{pmatrix}, B_k^{(2)} = I_{(n+m)\times(n+m)}.$$

The example is problem 29 of [8], which has four optimal solutions

$$x^* = (a, b, c), \quad (a, -b, -c), \quad (-a, b, -c), \quad (-a, -b, c),$$

where $a = 4, b = 2\sqrt{2}, c = 2$, the optimal value of function is $f(x^*) = -16\sqrt{2}$. We choose $\sigma_k = 0.9995$ for all k. The result is reported in Table 8.1.

Table 8.1

(x_0, s_0)	$(1,1,1,1)$		$(2,2,2,2)$		$(1.5,1.5,1.5,1)$	
B_k	$B_k^{(1)}$	$B_k^{(2)}$	$B_k^{(1)}$	$B_k^{(2)}$	$B_k^{(1)}$	$B_k^{(2)}$
k	8	206	8	216	7	203
x_k	y^*	y^*	y^*	y^*	y^*	y^*
$f(x_k)$	-22.6274	-22.6274	-22.6274	-22.6274	-22.6274	-22.6274
res1	5.5396E-15	1.7366E-11	1.7804E-15	1.7366E-11	2.1354E-15	1.8034E-11
res2	5.9204E-07	9.5262E-06	1.6593E-07	9.5250E-06	7.6865E-07	9.7095E-06

Where (x_0, s_0) is start point, k is the number of iterations, $x_k = y^* = (4.0000, 2.8284, 2.0000)$ is approximate optimal point, $res1 = \|C_k\|$, $res2 = \|D_k W_k^T \nabla F_k\|$.

Remark 8.1. For our computation, when we choose different start point, for example, we choose start points as

$$(x_0, s_0) = (1, -1, -1, 1), (2, -2, -2, 2), (1.5, -1.5, -1.5, 1),$$

then $x_k = y^* = (4.0000, -2.8284, -2.0000)$ and other values in Table 8.1 are same, i.e., $\{x_k\}$ converges to the optimal point $(a, -b, -c)$. Other two optimal points have the same characterization.

From numerical example it shows that the calculated result is coincident with theoretical analysis.

References

[1] M. El-Alem, A global convergence theory for the Celis-Dennis -Tapia trust-region algorithm for constrained optimization, *SIAM J. Numer. Anal.*, **28**:1 (1991), 266–290.

[2] M. El-Alem, A robust trust-region algorithm with nonmonotic penalty parameter scheme for constrained optimization, *SIAM J. Optim.*, **5**:2 (1995), 348–378.

[3] T.F. Coleman, Y. Li, On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds, *Math. Prog.* **67**(1994), 189–224.

[4] T.F. Coleman and Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, *SIAM J. Optim.*, **6**:2 (1996), 418–445.

[5] J.E. Dennis, M. El-Alem and M.C. Maciel, A global convergence theory for general trust-region-based algorithm for equality constrained optimization, *SIAM J. Optim.*, **7**:1 (1997), 177–207.

[6] J.E. Dennis, M. Heinkenschloss, and L.N. Vicente, Trust- region interior-point SQP algorithms for a class of nonlinear programming problems, *SIAM J. Control Optim.*, **36**:5 (1998), 1750–1794.

[7] F.A.M. Gomes, M.C. Maciel, J.M. Martinez, Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameter, *Math. Prog.*, 84 (1999), 161–200.

[8] W.Hock and K.Schittkowski, Test examples for nonlinear programming codes, Lecture Notes in economics and Mathematical Systems 187, Springer-Verlag, Berlin, 1981.

[9] J.M. Martinez, S.A. Saitos, A trust -region strategy for minimization on arbitrary domains, *Math. Prog.* **68**(1995), 267–301.

[10] J. Nocedal and M.L. Overton, Projected Hessian updating algorithms for nonlinear constrained optimization, *SIAM J. Numer. Anal.*, **22**:5 (1985), 821–850.

[11] M.J.D. Powell and Y. Yuan, A trust region algorithm for equality constrained optimization, *Math. Prog.* **49**(1991), 189–211.

[12] Y.X. Yuan, On the convergence of a new trust region algorithm, *Numerische Mathematik,* **70**:4 (1995), 515–539.