Hokkaido Mathematical Journal Vol. 30 (2001) p. 113–136

Global convergence of a trust-region algorithm for inequality constrained optimization

Xiaojiao TONG and Shuzi ZHOU

(Received July 14, 1999; Revised February 18, 2000)

Abstract. This paper presents a trust-region algorithm for n-dimensional nonlinear optimization subject to m nonlinear inequality constraints. Equivalent KKT conditions are derived, being 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. Condition $m\leq n$ is required.

Key words: inequality constrained optimization; nonlinear optimization; trust-region method; Global convergence.

1. Introduction

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

$$\begin{aligned}&\text{minimize } f(x)\\ &\text{subject to } H(x)\leq 0,\end{aligned}\tag{1.1}$$

where $H(x)=(h_1(x),h_2(x),\ldots,h_m(x))^T$, $f(x)$ and $h_i(x)$ $i\in I=\{1,2,\ldots,m\}$ are $R^n\rightarrow R$ 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 optimization problems. Many authors have studied the trust-region algorithm for solving equality constrained problems. (see [1], [2], [4], [5], [10], for example). However, for nonlinear inequality constrained optimization problems, the results about trust region methods are very few. (see [3], [6], [9], [7], for example). The paper [7] considers problems with equality constraints and bound constraints. Therefore, adding slack variables, any problem of the form (1.1) falls under the framework of [7], independently of the number of inequality constraints. Under regularity assummptions, some subsequence of the algorithm defined in [7] converges to a KKT point of the original

1991 Mathematics Subject Classification : 65K05, 90C30, 49M37.

problem. The paper [9] presents a trust region method for an arbitrary closed set and prove a 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 the trust region method. The 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. The work [3] discusses a interior trust region approach for nonlinear optimization only for a special case, that is the optimization problem with bounded constraints. [6] extends the method of [3] 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 new equivalent KKT conditions, which are the basis for constructing our algorithm. The subproblems in the algorithm can be solved by the method proposed in [4] and [6]. We have proved that at least one accumulation point of the new algorithm is KKT point.

The problem proposed by this paper is different from [4] and [6]. Assume p is the number of bound constraints, the paper [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 can not be reduced as in [6]. The paper [4] only considers equality constrained problems.

The paper is organized as follows. In Section 2, we derive equivalent first-order conditions; In Section 3, we present a method to compute trial step; In Section 4, the new trust-region algorithm is formulated; Section 5 gives a global convergence theorem of the algorithm; The numerical example is given at last section.

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 $f(x_k)$, l_k represents $l(x_k, s_k, \lambda_k)$.

2. Optimality conditions

By introducing slack variables $s \in R^m$, we obtain the following problem for the variables $x \in R^n$ and $s \in R^m$, which is equivalent to (1.1).

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; H(x) + s = 0, \quad s \geq 0. \end{aligned} \tag{2.1}$$

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

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

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

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

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

$$\begin{aligned} &\nabla f(x^*) + A(x^*)\lambda^* = 0, \quad \lambda^* - \mu^* = 0, \\ &H(x^*) + s^* = 0, \\ &\mu_i^* s_i^* = 0, \;\; (i \in I), \quad \mu^* \geq 0; \;\; s^* \geq 0. \end{aligned} \tag{2.2}$$

We assume in this paper that rank $A(x) = m$. So the constraint qualification is satisfied.

(2.2) can be rewritten as follows

$$\begin{aligned} &H(x^*) + s^* = 0, \quad s^* \geq 0, \\ &\nabla f(x^*) + A(x^*)\lambda^* = 0, \\ \\ &s_i^* > 0 \;\; \Rightarrow \lambda_i^* = 0 \;\; i \in I, \\ &s_i^* = 0 \;\; \Rightarrow \lambda_i^* \geq 0 \;\; i \in I. \end{aligned} \tag{2.3}$$

Since rank $A(x) = m$, we can make QR factorization of $A(x)$ as follows

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

where $Y(x) \in R^{n \times m}$, the columns of $Y(x)$ form a orthonormal basis of the space of $A(x)$; $Z(x) \in R^{n \times (n-m)}$, the columns of $Z(x)$ form an orthonormal bases for the null space of $A(x)$; $R(x) \in R^{m\times m}$ is an $m \times m$ nonsingular upper triangular matrix. It is easy to see that

$$\begin{aligned}&Y(x)^T Y(x) = I_m,\ Z(x)^T Z(x) = I_{n-m},\\ &Y(x)Y(x)^T + Z(x)Z(x)^T = I_n.\end{aligned}$$

Since $\operatorname{rank} A(x^*) = m$, the equation $\nabla f(x^*) + A(x^*)\lambda^* = 0$ can be written as:

$$\begin{aligned}&Y(x^*)^T \nabla f(x^*) + R(x^*)\lambda^* = 0,\\ &Z(x^*)^T \nabla f(x^*) = 0,\end{aligned}$$

that is:

$$\begin{aligned}&\lambda^* = -R(x^*)^{-1} Y(x^*)^T \nabla f(x^*),\\ &Z(x^*)^T \nabla f(x^*) = 0.\end{aligned} \tag{2.4}$$

Conversely, if (2.4) holds then

$$\begin{aligned}&\nabla f(x^*) + A(x^*)\lambda^*\\ &\quad = (Y(x^*)Y(x^*)^T + Z(x^*)Z(x^*)^T)\nabla f(x^*) + Y(x^*)R(x^*)\lambda^*\\ &\quad = Y(x^*)[Y(x^*)\nabla f(x^*) + R(x^*)\lambda^*] + Z(x^*)Z(x^*)^T \nabla f(x^*) = 0.\end{aligned}$$

Hence (2.3) is equivalent to

$$\begin{aligned}&H(x^*) + s^* = 0,\quad s^* \geq 0,\\ &Z(x^*)^T \nabla f(x^*) = 0,\end{aligned} \tag{2.5}$$

$$\begin{aligned}&s_i^* > 0\ \Rightarrow [-R(x^*)^{-1} Y(x^*)^T \nabla f(x^*)]_i = 0,\\ &s_i^* = 0\ \Rightarrow [-R(x^*)^{-1} Y(x^*)^T \nabla f(x^*)]_i \geq 0.\end{aligned}$$

We 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:

$$D(u) = \begin{pmatrix} \tilde{D}(u) & 0\\ 0 & I_{n-m}\end{pmatrix}, \tag{2.6}$$

where $\tilde{D}(u) \in R^{m\times m}$ is the following diagonal matrix:

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

Then last two equations of (2.5) can be written as:

$$\tilde{D}(u^*)[-R(x^*)^{-1}Y(x^*)^T\nabla f(x^*)] = 0.$$

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

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

It is obvious that $\operatorname{rank} W(u) = n$, $J(u)W(u) = 0$, and the columns of $W(u)$ form a basis for the null space of $J(u)$. Then we have the following proposition:

Proposition 2.1 *The point $u^* = ((x^*)^T, (s^*)^T)^T$ is a first order KKT point of* (1.1) *if and only if u^* satisfies*:

$$\begin{aligned} &H(x^*) + s^* = 0, \\ &s^* \geq 0, \\ &D(u^*)W(u^*)^T\nabla F(u^*) = 0. \end{aligned} \tag{2.7}$$

Remark 2.1 At point $[-R(x)^{-1}Y(x)^T\nabla f(x)]_i = 0$, matrix $D(u)$ is usually discontinuous, but $D(u)W(u)^T\nabla F(u)$ is still continuous.

3. Trial steps

Constrained optimization problem is often solved by SQP trust region algorithms. For problem (2.1), at kth iteration, we have u_k and need to compute trial step d_k. The trial step is computed by solving the following trust region subproblem:

$$\begin{aligned} &\text{minimize } \ l_k + \nabla_u l_k^T d + \frac{1}{2} d^T B_k d \\ &\text{subject to } \ C_k + J_k^T d = 0, \quad s_k + (d)_s \geq 0, \quad \|d\| \leq \Delta_k, \end{aligned}$$

where $B_k \in R^{(n+m)\times(n+m)}$ is a symmetric matrix and the Hessian of the Lagrangian at (u_k, λ_k) or an approximation to it, Δ_k is the trust region radius at kth iteration,

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

However, this approach may lead to inconsistent constraints if $C_k \neq 0$. To overcome this difficulty, similarly to [4], the trial step d_k is determined as $d_k = d_k^n + W_k d_k^t$, where d_k^n is the quasi-normal component, $d_k^t \in R^n$, $W_k d_k^t$ is the tangential component with respect to the null space of the J_k. We require $s_k > 0$, $s_k + (d_k)_s > 0$. If d_k is accepted, we set $u_{k+1} = u_k + d_k$.

We give now the method to compute d_k^n and d_k^t.

3.1. The quasi-normal component

The quasi-normal component d_k^n is related to the trust-region subproblem as follows:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|C_k + J_k d^n\|^2 \\ &\text{subject to} \quad \|d^n\| \leq \tau \Delta_k, \end{aligned} \tag{3.1}$$

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

In order to keep $s_k > 0$ for all k, we require that d^n has the form:

$$d^n = \begin{pmatrix} (d^n)_x \\ 0 \end{pmatrix}.$$

Then (3.1) can be rewritten as:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|C_k + A_k^T (d^n)_x\|^2 \\ &\text{subject to} \quad \|(d^n)_x\| \leq \tau \Delta_k. \end{aligned} \tag{3.2}$$

As in most trust region algorithms, we do not have to solve (3.2) exactly and only have to compute d_k^n satisfying the following conditions: there exist constants κ_1 and β_1 such that

$$\|d_k^n\| = \|(d_k^n)_x\| \leq \kappa_1 \|C_k\|, \tag{3.3}$$

$$\|C_k\|^2 - \|C_k + A_k^T (d_k^n)_x\|^2 \geq \beta_1 [\|C_k\|^2 - \|C_k + A_k^T v_k^n\|^2], \tag{3.4}$$

where $v_k^n \in R^n$ is the solution of the following problem

$$\begin{cases} \min \dfrac{1}{2}\|C_k + A_k^T v^n\|^2 \\ \text{subject to } \ \|v^n\| \le \tau\Delta_k, \ v^n \in \text{span}\{-A_k C_k\}. \end{cases}$$

Condition (3.4) is called a fraction of Cauchy decrease condition (see [4], for example), which is easily satisfied. In [4] algorithms has been provided algorithms to compute d_k^n satisfying (3.3).

3.2. The tangential component

Denote $q_k(d) = l_k + \nabla_u l_k^T d + \frac{1}{2} d^T B_k d$, which is a quadratic approximation of $l(x, s, \lambda)$ at point (x_k, s_k, λ_k).

Denote

$$d_k^t = \begin{pmatrix} \bar{d}_k^t \\ \hat{d}_k^t \end{pmatrix},$$

where $\bar{d}_k^t \in R^m$, $\hat{d}_k^t \in R^{n-m}$. Then

$$\begin{aligned} d_k &= \begin{pmatrix} (d_k)_x \\ (d_k)_s \end{pmatrix} = d_k^n + W_k d_k^t \\ &= \begin{pmatrix} (d_k^n)_x \\ 0 \end{pmatrix} + \begin{pmatrix} -Y_k R_k^{-T} & Z_k \\ I_m & 0_{m\times(n-m)} \end{pmatrix} \begin{pmatrix} \bar{d}_k^t \\ \hat{d}_k^t \end{pmatrix}, \end{aligned}$$

which implies $\bar{d}_k^t = (d_k)_s$. Hence

$$d_k^t = \begin{pmatrix} (d_k)_s \\ \hat{d}_k^t \end{pmatrix}.$$

Now we can write subproblem for d_k^t as follows:

$$\begin{aligned} &\text{minimize } \ q_k(d_k^n + W_k d^t) = q_k(d_k^n) + \bar{g}_k^T d^t + \frac{1}{2}(d^t)^T W_k^T B_k W_k d^t \\ &\text{subject to } \ \|\bar{D}_k^{-1} d^t\| \le \Delta_k, \ \ (d)_s \ge -\sigma_k s_k, \end{aligned} \tag{3.5}$$

where $\sigma_k \in [\sigma, 1)$, $\sigma \in (0, 1)$ is a constant and

$$\bar{g}_k = W_k^T[\nabla_u l_k + B_k d_k^n] = W_k^T(\nabla F_k + J_k^T \lambda_k + B_k d_k^n)$$

$$= W_k^T(\nabla F_k + B_k d_k^n) = \begin{pmatrix} \bar{g}_{k,1} \\ \bar{g}_{k,2} \end{pmatrix}, \tag{3.6}$$

$$\bar{D}_k = \begin{pmatrix} \hat{D}_k & 0 \\ 0 & I_{n-m} \end{pmatrix}, \quad (\hat{D}_k)_{ii} = \begin{cases} 1, & \text{if } (\bar{g}_{k,1})_i < 0, \\ (s_k)_i, & \text{if } (\bar{g}_{k,1})_i \geq 0, \end{cases} \tag{3.7}$$

where $\bar{g}_{k,1} \in R^m$, $\bar{g}_{k,2} \in R^{n-m}$, $\hat{D}_k \in R^{m \times m}$ is a diagonal matrix. Again we do not need to compute exactly d_k^t, and only required d_k^t to satisfy the following fraction of Cauchy decrease condition:

$$q_k(d_k^n) - q_k(d_k^n + W_k d_k^t) \geq \beta_2[q_k(d_k^n) - q_k(d_k^n + W_k v_k^d)], \tag{3.8}$$

where $\beta_2 > 0$ is a constant, $v_k^d \in R^n$ is a solution of the following problem:

$$\begin{aligned} &\text{minimize} \quad q_k(d_k^n + W_k v) \\ &\text{subject to} \quad \|\bar{D}_k^{-1} v\| \leq \Delta_k, \ v \in \text{span}\{-\bar{D}_k^2 \bar{g}_k\}, \ \bar{v} \geq -\sigma_k s_k, \end{aligned} \tag{3.9}$$

where $\bar{v} \in R^m$, $\hat{v} \in R^{n-m}$, $v^T = (\bar{v}^T, \hat{v}^T)$.

Remark 3.1 It is obvious that $\bar{v}_k^d + s_k > 0$.

Remark 3.2 Inexact solution of (3.5) satisfying (3.8) exists, for example, we can choice $d_k^t = v_k^d$. In Section 6 we will introduce an algorithm to solve subproblem (3.5).

3.3. Calculation of Lagrange multiplier λ_{k+1}

From (2.4) we have the following formula for calculating Lagrange multiplier:

$$\lambda_{k+1} = -R(x_k + (d_k)_x)^{-1} Y(x_k + (d_k)_x)^T \nabla f(x_k + (d_k)_x). \tag{3.10}$$

3.4. Choice of merit function

We use the augmented Lagrangian as a merit function:

$$\Phi(x, s, \lambda; \rho) = f(x) + \sum_{i=1}^{m} \lambda_i (h_i(x) + s_i) + \rho \|C(u)\|^2, \tag{3.11}$$

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

At kth iteration, the actual reduction is defined by

$$\begin{aligned}\text{ared}(d_k;\rho_k) &= \Phi(x_k, s_k, \lambda_k; \rho_k) \\ &\quad - \Phi(x_k + (d_k)_x, s_k + (d_k)_s, \lambda_{k+1}; \rho_k), \end{aligned} \tag{3.12}$$

and the predicted reduction is defined by

$$\begin{aligned}\text{pred}(d_k;\rho_k) &= \Phi(x_k, s_k, \lambda_k; \rho_k) \\ &\quad - [q_k(d_k) + \Delta\lambda_k^T(J_k d_k + C_k) + \rho_k\|J_k d_k + C_k\|^2] \\ &= q_k(0) - q_k(d_k) - \Delta\lambda_k^T(J_k d_k + C_k) \\ &\quad + \rho_k[\|C_k\|^2 - \|J_k d_k + C_k\|^2], \end{aligned} \tag{3.13}$$

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

4. Statement of algorithm

Algorithm 4.1

Step 0. Choose $u_0 = (x_0^T, s_0^T)^T$, $x_0 \in R^n$, $s_0 \in R^m$, $s_0 > 0$, $\Delta_0 > 0$ and $\lambda_0 \in R^m$, $\rho_{-1} \geq 1$, symmetric $B_0 \in R^{(n+m)\times(n+m)}$, $a_1, \tau \in (0,1)$, $\bar{\rho} > 0$, $\Delta_{\max} \geq \Delta_{\min} > 0$, $\eta \in (0,1)$. $k := 0$.

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

Step 2. Compute d_k^n satisfying (3.3) and (3.4); Compute d_k^t satisfying (3.8); $d_k := d_k^n + W_k d_k^t$.

Step 3. Compute λ_{k+1} by (3.10), $\Delta\lambda_k := \lambda_{k+1} - \lambda_k$.

Step 4. Compute $\text{pred}(d_k;\rho_{k-1})$.
If $\text{pred}(d_k;\rho_{k-1}) \geq \frac{\rho_{k-1}}{2}[\|C_k\|^2 - \|J_k d_k + C_k\|^2]$ then set $\rho_k = \rho_{k-1}$; Otherwise set

$$\rho_k = \frac{2[q_k(d_k) - q_k(0) + \Delta\lambda_k^T(J_k d_k + C_k)]}{\|C_k\|^2 - \|J_k d_k + C_k\|^2} + \bar{\rho}. \tag{4.1}$$

Step 5. Compute $\text{ared}(d_k;\rho_k)$, $\text{pred}(d_k;\rho_k)$. If

$$\frac{\text{ared}(d_k;\rho_k)}{\text{pred}(d_k;\rho_k)} < \eta,$$

then $\Delta_k := a_1\|d_k\|$ 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 4.1 From Step 4 we have:

$$\rho_k \geq \rho_{k-1} \geq 1, \tag{4.2}$$

$$\text{pred}(d_k; \rho_k) \geq \frac{\rho_k}{2}[\|C_k\|^2 - \|J_k d_k + C_k\|^2]. \tag{4.3}$$

In fact, if $\rho_k = \rho_{k-1}$ then (4.2) and (4.3) are obvious. If ρ_k is updated by (4.1) then (4.2) is obvious and we have

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

$$q_k(0) - q_k(d_k) - \Delta\lambda_k^T(J_k d_k + C_k) \geq -\frac{\rho_k}{2}[\|C_k\|^2 - \|J_k d_k + C_k\|^2],$$

which combining with (3.13) yields (4.3).

5. Global convergence

5.1. Assumptions of global convergence

In order to establish the global convergence of Algorithm 4.1, we need some assumptions (compare with [4–6]).

AS.1 For all $k, u_k, u_k + d_k \in \Omega \subset R^{n+m}$,

$$\Omega = \begin{pmatrix} \Omega_x \\ \Omega_s \end{pmatrix}, \ \Omega_x \in R^n, \ \Omega_s = R^m_+,$$

where 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 $x \in \Omega_x$, $\operatorname{rank} A(x) = m$.

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

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

Hence there are constants $\nu_i > 0$, $(i = 1, 2, \ldots, 9)$ independent of k and u, such that:

$$\begin{aligned} &\|\nabla f(x)\| \le \nu_1,\ \|Y_k\| \le \nu_2,\ \|R_k^{-1}\| \le \nu_3, \\ &\|R_k\| \le \nu_4,\ \|R_k\| \le \nu_5,\ \|W_k\| \le \nu_6, \\ &\|\bar{D}_k\| \le \nu_7,\ \|B_k\| \le \nu_8,\ \|\lambda_k\| \le \nu_9. \end{aligned} \tag{5.1}$$

From now on we assume that above assumptions hold.

5.2. Intermedium results

Lemma 5.1 *Assume d_k is computed by algorithm* 4.1. *Then for each k we have*

$$\|C_k\|^2 - \|J_k d_k + C_k\|^2 \ge \kappa_2 \|C_k\| \min\{\tau\Delta_k, \kappa_3\|C_k\|\}, \tag{5.2}$$

where κ_2, κ_3 are positive constants independent of k.

Proof. We have $d_k = d_k^n + W_k d_k^t$ and

$$A_k^T (d_k^n)_x = (A_k^T, I_m) \begin{pmatrix} (d_k^n)_x \\ 0 \end{pmatrix} + J_k W_k d_k^t = J_k d_k.$$

Then $\|C_k\|^2 - \|J_k d_k + C_k\|^2 = \|C_k\|^2 - \|A_k^T (d_k^n)_x + C_k\|^2$. From reference [9] and global assumptions we have

$$\|C_k\|^2 - \|A_k^T (d_k^n)_x + C_k\|^2 \ge \frac{1}{2}\|A_k C_k\| \min\Big\{\tau\Delta_k, \frac{\|A_k C_k\|}{\|A_k A_k^T\|}\Big\},$$

$$\|A_k C_k\| \ge \frac{\|C_k\|}{\nu_2\nu_3}, \quad \|A_k A_k^T\| \le \nu_2^2\nu_4^2.$$

Hence

$$\|C_k\|^2 - \|J_k d_k + C_k\|^2 \ge \kappa_2 \|C_k\| \min\{\tau\Delta_k, \kappa_3\|C_k\|\},$$

where $\kappa_2 = \frac{1}{2\nu_2\nu_3}$, $\kappa_3 = \frac{1}{\nu_2^3\nu_3\nu_4^2}$. The proof is completed. □

Lemma 5.2 *Assume d_k^t is an approximate solution of* (3.5) *and satifies* (3.8). *Then we have*

$$q_k(d_k^n) - q_k(d_k^n + W_k d_k^t) \ge \kappa_4 \|\bar{D}_k \bar{g}_k\| \min\{\kappa_5 \|\bar{D}_k \bar{g}_k\|, \kappa_6\Delta_k\}, \tag{5.3}$$

where κ_4, κ_5, κ_6 are positive constants independent of k.

Proof. Denote

$$\hat{g}_k = \bar{D}_k \bar{g}_k = \begin{pmatrix} \hat{g}_{k,1} \\ \hat{g}_{k,2} \end{pmatrix}.$$

Define $\psi : R^+ \to R$ as follows:

$$\begin{aligned} \psi(t) &= q_k\left(d_k^n - tW_k\bar{D}^2 \frac{\bar{g}_k}{\|\bar{D}_k\bar{g}_k\|}\right) - q_k(d_k^n) \\ &= -t\|\hat{g}_k\| + \frac{1}{2}t^2 r_k, \end{aligned}$$

where $r_k = \hat{g}_k^T \hat{B}_k \hat{g}_k / \|\hat{g}_k\|^2$, $\hat{B}_k = \bar{D}_k W_k^T B_k W_k \bar{D}_k$.

It is obvious that

$$v = -t\bar{D}_k^2 \frac{\bar{g}_k}{\|\bar{D}_k\bar{g}_k\|} = -t\bar{D}_k \frac{\hat{g}_k}{\|\hat{g}_k\|} \in \text{span}\{-\bar{D}_k^2 \bar{g}_k\},$$

we denote

$$v = \begin{pmatrix} \bar{v} \\ \hat{v} \end{pmatrix} = -\frac{t}{\|\hat{g}_k\|} \begin{pmatrix} \hat{D}_k \hat{g}_{k,1} \\ \hat{g}_{k,2} \end{pmatrix},$$

where $\bar{v} \in R^m$, $\hat{v} \in R^{n-m}$. Then $\bar{v} = -t\hat{D}_k \hat{g}_{k,1} / \|\hat{g}_k\|$.

We require v to satisfy feasible conditions of (3.9), i.e.

$$\text{(i) } \|\bar{D}_k^{-1} v\| \le \Delta_k, \quad \text{(ii) } \bar{v} \ge -\sigma_k s_k.$$

It is easy to show that the two conditions above are satisfied for $t \in [0, T_k]$, where $T_k = \min\{\Delta_k, \sigma_k \min\{\|\hat{g}_k\|/(\hat{g}_k)_i, (\hat{g}_k)_i > 0\}\}$.

Let t_k^* be the minimum point of ψ in $[0, T_k]$.

If $t_k^* \in (0, T_k)$ then it is obvious $0 < r_k \le \|\hat{B}_k\|$, and

$$\psi(t_k^*) = -\frac{\|\hat{g}_k\|^2}{2r_k} \le -\frac{\|\hat{g}_k\|^2}{2\|\hat{B}_k\|}. \tag{5.4}$$

Assume $t_k^* = T_k$. Then for case $r_k > 0$, we have $\|\hat{g}_k\|/r_k \ge T_k$, i.e. $r_k T_k \le \|\hat{g}_k\|$; For case $r_k \le 0$, we have also $r_k T_k \le \|\hat{g}_k\|$. Hence, for $t_k^* = T_k$ we have

$$\psi(t_k^*) = \psi(T_k) = -T_k\|\hat{g}_k\| + \frac{r_k}{2}T_k^2 \le -\frac{1}{2}T_k\|\hat{g}_k\|,$$

which combining with (5.4) yields

$$-\psi(t_k^*) \geq \frac{1}{2}\|\hat{g}_k\| \min\left\{ \frac{\|\hat{g}_k\|}{\|\hat{B}_k\|}, T_k \right\}.$$

Because of

$$T_k \geq \min\{\Delta_k, \sigma_k\} \geq \Delta_k \min\left\{1, \frac{\sigma}{\Delta_{\max}}\right\} = \kappa_6 \Delta_k,$$

$$\|\hat{B}_k\| \leq \nu_6^2 \nu_7^2 \nu_8,$$

we conclude that

$$-\psi(t_k^*) \geq \frac{1}{2}\|\bar{D}_k \bar{g}_k\| \min\{\kappa_5 \|\bar{D}_k \bar{g}_k\|, \kappa_6 \Delta_k\},$$

where $\kappa_5 = 1/\nu_6^2 \nu_7^2 \nu_8$, $\kappa_6 = \min\{1, \sigma/\Delta_{\max}\}$ are constants independent of k.

From the definition of v_k^d we have

$$\begin{aligned} q_k(d_k^n) - q_k(d_k^n + W_k v_k^d) &\geq -\psi(t_k^*) \\ &\geq \frac{1}{2}\|\bar{D}_k \bar{g}_k\| \min\{\kappa_5 \|\bar{D}_k \bar{g}_k\|, \kappa_6 \Delta_k\}, \end{aligned}$$

which combining with (3.8) implies (5.3) with $\kappa_4 = \beta_2/2$. □

Lemma 5.3 *Assume d_k is computed by algorithm. Then*

$$\begin{aligned} \text{pred}(d_k; \rho) \geq\ & \kappa_4 \|\bar{D}_k \bar{g}_k\| \min\{\kappa_5 \|\bar{D}_k \bar{g}_k\|, \kappa_6 \Delta_k\} - \kappa_7 \|C_k\| \\ & + \rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2]. \end{aligned} \tag{5.5}$$

Proof. From (3.13) and $d_k = d_k^n + W_k d_k^t$ we have

$$\begin{aligned} \text{pred}(d_k, \rho) =\ & [q_k(0) - q_k(d_k^n) - \Delta\lambda_k^T (J_k d_k + C_k)] \\ & + [q_k(d_k^n) - q_k(d_k^n + W_k d_k^t)] \\ & + \rho[\|C_k\|^2 - \|J_k d_k + C_k\|^2]. \end{aligned}$$

By using global assumptions, condition (3.3) and Lemma 5.1 we obtain

$$\begin{aligned} & q_k(0) - q_k(d_k^n) - \Delta\lambda_k^T (J_k d_k + C_k) \\ & = -\nabla_u l_k^T d_k^n - \frac{1}{2}(d_k^n)^T B_k d_k^n - \Delta\lambda_k^T (J_k d_k + C_k) \\ & \geq -\left[\|\nabla_u l_k^T\| + \frac{1}{2}\|B_k\| \|d_k^n\|\right] \|d_k^n\| - \|\Delta\lambda_k\| \|C_k\| \geq -\kappa_7 \|C_k\|, \end{aligned}$$

where $\kappa_7 = [\nu_1 + (\nu_2\nu_4 + 1)\nu_9 + \frac{1}{2}\nu_8\tau\Delta_{\max}]\kappa_1 + 2\nu_9$ is a constant independent of k. Then (5.5) follows from Lemma 5.2. □

Lemma 5.4 ([1]) *There is a positive constant κ_8 independent of k such that*

$$|\operatorname{ared}(d_k;\rho_k) - \operatorname{pred}(d_k;\rho_k)| \leq \kappa_8\rho_k\|d_k\|^2. \tag{5.6}$$

5.3. Global convergence

Lemma 5.5 *If $\|C_k\| \leq \alpha\Delta_k$, $\|\bar{D}_k\bar{g}_k\| + \|C_k\| > \epsilon$ and*

$$\alpha \leq \min\left\{\frac{\epsilon}{3\Delta_{\max}}, \frac{\kappa_4\epsilon}{3\kappa_7}\min\left\{\frac{2\kappa_5\epsilon}{3\Delta_{\max}}, \kappa_6\right\}\right\}. \tag{5.7}$$

Then

$$\begin{aligned}\operatorname{pred}(d_k;\rho_k) \geq{}& \frac{\kappa_4}{2}\|\bar{D}_k\bar{g}_k\|\min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\}\\ &+ \rho_k[\|C_k\|^2 - \|J_kd_k + C_k\|^2],\end{aligned} \tag{5.8}$$

$$\operatorname{pred}(d_k;\rho_k) \geq \kappa_9\Delta_k, \tag{5.9}$$

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

where κ_9 is a constant independent of k.

Proof. From $\|\bar{D}_k\bar{g}_k\| + \|C_k\| \geq \epsilon$ and (5.7) we have $\|C_k\| \leq \frac{\epsilon}{3}$, $\|\bar{D}_k\bar{g}_k\| > \frac{2}{3}\epsilon$. Then it follows from Lemma 5.3 and (5.7) that

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

(5.8) is proved.

From Lemma 5.1 and (5.8) we have

$$\operatorname{pred}(d_k;\rho) \geq \frac{\kappa_4}{2}\|\bar{D}_k\bar{g}_k\|\min\{\kappa_5\|\bar{D}_k\bar{g}_k\|, \kappa_6\Delta_k\}$$

$$\geq \frac{\kappa_4\epsilon}{3}\min\left\{\frac{2\kappa_5\epsilon}{\Delta_{\max}}, \kappa_6\right\}\Delta_k = \kappa_9\Delta_k,$$

where $\kappa_9 = \frac{\kappa_4\epsilon}{3}\min\{\frac{2\kappa_5\epsilon}{\Delta_{\max}}, \kappa_6\}$ is a constant. (5.9) is proved.

(5.8) implies that

$$\text{pred}(d_k; \rho_{k-1}) \geq \rho_{k-1}[\|C_k\|^2 - \|J_k d_k + C_k\|^2].$$

Then from Step 4 of the algorithm we know $\rho_k = \rho_{k-1}$. □

Theorem 5.6 *The algorithm is valid,that is, interior loop (Step 2–Step 5) can be ended in finite times at each iteration.*

Proof. Let the index of interion loop be i and the correspondent values be

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

The values 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

$$\left|\frac{\text{ared}(d_{k,i}; \rho_{k,i})}{\text{pred}(d_{k,i}; \rho_{k,i})} - 1\right| > 1 - \eta. \tag{5.10}$$

We consider two cases:

Case (i): Assume $\|C_k\| \neq 0$. Then from (4.3) and Lemma 5.1 we know

$$\begin{aligned}\text{pred}(d_{k,i}; \rho_{k,i}) &\geq \frac{\rho_{k,i}}{2}[\|C_k\|^2 - \|C_k + J_k d_{k,i}\|^2] \\ &\geq \frac{\rho_{k,i}}{2}\kappa_2\|C_k\|\min\{\tau\Delta_{k,i}, \kappa_3\|C_k\|\} \\ &\geq \frac{\rho_{k,i}}{2}\kappa_2\|C_k\|\min\left\{\tau, \frac{\kappa_3\|C_k\|}{\Delta_{\max}}\right\}\Delta_{k,i},\end{aligned}$$

which combining with Lemma 5.4 yields

$$\begin{aligned}\left|\frac{\text{ared}(d_{k,i}; \rho_{k,i})}{\text{pred}(d_{k,i}; \rho_{k,i})} - 1\right| &\leq \frac{\kappa_8\rho_{k,i}\|d_{k,i}\|^2}{\frac{\rho_{k,i}}{2}\kappa_2\|C_k\|\min\{\tau, \frac{\kappa_3\|C_k\|}{\Delta_{\max}}\}\Delta_{k,i}} \\ &\leq \frac{2\kappa_8}{\kappa_2\|C_k\|\min\{\tau, \frac{\kappa_3\|C_k\|}{\Delta_{\max}}\}}\Delta_{k,i} \to 0.\end{aligned}$$

It contradicts with (5.10).

Case (ii): Assume $\|C_k\| = 0$. Then it follows from (3.3), (2.6) and (3.7) that $d^n_{k,i} = 0$, and

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

$\hat{D}_k = \tilde{D}_k$, $D_k = \bar{D}_k$, $\|D_k W_k^T \nabla F_k\| = \|\bar{D}_k \bar{g}_k\|$.

Therefore, it follows from Lemma 5.4 and Lemma 5.5 that $\rho_{k,i} = \rho_{k-1}$ and

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

which contradicts with (5.10) again. The proof is completed. □

Lemma 5.7 *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.*

Proof. See Lemma 7.11 and Lemma 7.12 of paper [4]. □

Lemma 5.8 *If* $\|\bar{D}_k \bar{g}_k\| + \|C_k\| > \epsilon$ *for all* k *then there exists a constant* Δ^* *independent of* k *such that*

$$\Delta^k \geq \Delta^*, \tag{5.11}$$

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

Proof. We consider the kth iteration. Denote by $\Delta_k = \Delta_{k,0}$ the starting radius. Let j be the number of interior loop, $d_{k,j} = d_k$ is the accepted trial step.

(1) If $j = 0$ then

$$\Delta^k = \Delta_{k,0} = \Delta_k \geq \Delta_{\min}. \tag{5.12}$$

(2) If $j \geq 1$ then discuss three cases:

(i) $\|C_k\| > \alpha \Delta_{k,i}$ for all $i = 0, 1, 2, \ldots, j$.

(ii) There exists a largest index $l < j$ such that $\|C_k\| > \alpha \Delta_{k,i}$ for $i = l+1, \ldots, j$ holds.

(iii) $\|C_k\| = 0$.

Case (i): In this case $i = 0, 1, \ldots, j-1$ correspond reject steps. From

(4.3), Lemma 5.1 and Lemma 5.4 we have

$$1-\eta \le \left| \frac{\text{ared}(d_{k,i};\rho_{k,i})}{\text{pred}(d_{k,i};\rho_{k,i})} - 1 \right| \le \frac{2\kappa_8 \|d_{k,i}\|}{\kappa_2 \|C_k\| \min\{\tau, \alpha\kappa_3\}}$$

$$\begin{aligned} \|d_{k,i}\| &\ge \frac{\kappa_2 \|C_k\| \min\{\tau, \alpha\kappa_3\}(1-\eta)}{2\kappa_8} \\ &\ge \frac{\kappa_2 \alpha \Delta_{\min} \min\{\tau, \alpha\kappa_3\}(1-\eta)}{2\kappa_8}. \end{aligned}$$

Hence

$$\Delta^k = \Delta_{k,j} = a_1 \|d_{k,j-1}\| \ge K_{11}, \tag{5.13}$$

where

$$K_{11} = \frac{a_1 \kappa_2 \alpha \Delta_{\min} \min\{\tau, \alpha\kappa_3\}(1-\eta)}{2\kappa_8}.$$

Case (ii): For $i = 0, 1, \ldots, l$ holds $\|C_k\| \le \alpha\Delta_{k,i}$. From Lemma 5.5 we know $\text{pred}(d_{k,i};\rho_{k,i}) \ge \kappa_9 \Delta_{k,i}$. Since $d_{k,i}$ is rejected step, it follows from Lemma 5.4 and Lemma 5.7 that

$$1-\eta < \left| \frac{\text{ared}(d_{k,i};\rho_{k,i})}{\text{pred}(d_{k,i};\rho_{k,i})} - 1 \right| \le \frac{\kappa_8 \rho^* \|d_{k,i}\|}{\kappa_9},$$

$$\|d_{k,i}\| \ge \frac{\kappa_9(1-\eta)}{\kappa_8 \rho^*}, \tag{5.14}$$

where ρ^* is an upper bound of $\{\rho_k\}$.

For $i = l+1, \ldots, j$ holds $\|C_k\| > \alpha\Delta_{k,i}$. If $j = l+1$, then from the way of updating the trust-region radius, we have $\Delta^k = \Delta_{k,j} = a_1\|d_{k,l}\|$. If $j > l+1$, then same as in Case (i) we have for rejected step $d_{k,i}$

$$\begin{aligned} \|d_{k,i}\| &\ge \frac{\kappa_2 \|C_k\| \min\{\tau, \alpha\kappa_3\}(1-\eta)}{2\kappa_8} \\ &\ge \frac{\kappa_2 \alpha \min\{\tau, \alpha\kappa_3\}(1-\eta)}{2\kappa_8} \Delta_{k,l+1} \\ &= \frac{\kappa_2 \alpha \min\{\tau, \alpha\kappa_3\}(1-\eta)}{2\kappa_8} a_1 \|d_{k,l}\|, \end{aligned}$$

and

$$\Delta^k = a_1\|d_{k,j-1}\| \geq \frac{a_1^2\kappa_2\alpha\min\{\tau,\alpha\kappa_3\}(1-\eta)}{2\kappa_8}\|d_{k,l}\|.$$

Denote

$$K_{12} = \min\Big\{a_1, \frac{a_1^2\kappa_2\alpha\min\{\tau,\alpha\kappa_3\}(1-\eta)}{2\kappa_8}\Big\}.$$

Then for $i = l+1,\dots,j$ we have

$$\Delta^k = \Delta_{k,j} \geq K_{12}\|d_{k,l}\|.$$

Since $d_{k,l}$ is rejected step, which satisfies (5.14), we have for Case (ii)

$$\Delta^k \geq K_{12}\frac{\kappa(1-\eta)}{\kappa_8\rho^*}. \tag{5.15}$$

Case (iii): Similar to the Proof of Theorem 5.6 we have $\|\bar{D}_k\bar{g}_k\| > \epsilon$. Then from Lemma 5.5 we have $\rho_{k,i} = \rho_{k-1}$ and

$$\mathrm{pred}(d_{k,i};\rho_{k,i}) \geq \frac{\kappa_4}{2}\epsilon\min\Big\{\frac{\kappa_5\epsilon}{\Delta_{\max}},\kappa_6\Big\}\Delta_{k,i}.$$

For rejected step $d_{k,i}$, we have

$$\left|\frac{\mathrm{ared}(d_{k,i};\rho_{k-1})}{\mathrm{pred}(d_{k,i},\rho_{k-1})} - 1\right| > 1-\eta.$$

Hence combination of Lemma 5.4 and Lemma 5.7 implies that

$$\|d_{k,i}\| \geq \frac{(1-\eta)\kappa_4\epsilon\min\{\frac{\kappa_5\epsilon}{\Delta_{\max}},\kappa\}}{2\kappa_8\rho^*}.$$

Therefore, we have

$$\Delta^k = a_1\|d_{k,j-1}\| \geq K_{13}, \tag{5.16}$$

where

$$K_{13} = \frac{a_1(1-\eta)\kappa_4\epsilon\min\{\frac{\kappa_4\epsilon}{\Delta_{\max}},\kappa_6\}}{2\kappa_8\rho^*}.$$

Finally, it follows from (5.12), (5.13), (5.15) and (5.16) that (5.11) holds with $\Delta^* = \min\{\Delta_{\min}, K_{11}, K_{12}\frac{\kappa_9(1-\eta)}{\kappa_8\rho^*}, K_{13}\}$. The proof is completed. □

Theorem 5.9 *The sequences of* $\{u_k\}$ *generated by Algorithm* 4.1 *satisfies*

$$\lim_{k\to+\infty}\inf[\|D_k W_k^T \nabla F_k\| + \|C_k\|] = 0. \tag{5.17}$$

Proof. First we proof

$$\lim_{k\to+\infty}\inf[\|\bar{D}_k \bar{g}_k\| + \|C_k\|] = 0. \tag{5.18}$$

The proof of (5.18) is by contradiction. Suppose for all k

$$\|\bar{D}_k \bar{g}_k\| + \|C_k\| > \epsilon.$$

We discuss two cases: (i) $\|C_k\| \le \alpha\Delta^k$; (ii) $\|C_k\| > \alpha\Delta^k$, where α is defined by (5.7).

Case (i): From Lemma 5.5 and Lemma 5.8 we have

$$\mathrm{pred}(d_k;\rho_k) \ge \kappa_9\Delta^k \ge \kappa_9\Delta^*.$$

Case (ii): From (4.2), (4.3), Lemma 5.1 and Lemma 5.8 we have

$$\begin{aligned}\mathrm{pred}(d_k;\rho_k) &\ge \frac{\rho_k}{2}[\|C_k\|^2 - \|J_k d_k + C_k\|^2] \\ &\ge \frac{1}{2}\kappa_2\|C_k\|\min\{\tau\Delta^k, \kappa_3\|C_k\|\} \\ &\ge \frac{\kappa_2}{2}\alpha\min\{\tau, \kappa_3\alpha\}(\Delta^*)^2.\end{aligned}$$

Denote $K_{14} = \min\{\kappa_9\Delta^*, \frac{\kappa_2}{2}\alpha\min\{\tau,\kappa_3\alpha\}(\Delta^*)^2\}$. Then for two cases

$$\mathrm{pred}(d_k;\rho_k) \ge K_{14}.$$

Hence, for all k we have

$$\Phi_k - \Phi_{k+1} \ge \eta\,\mathrm{pred}(d_k;\rho_k) \ge \eta K_{14},$$

which contradicts with boundedness of $\Phi(x_k, s_k, \lambda_k; \rho_k)$. Then (5.18) holds.

Next we prove (5.17).

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 (3.3). Because of the global assumptions and the expression of $\bar{g}_k$ we have

$$\lim_{k\in K_1}\|\bar{D}_k\bar{g}_k\| = \lim_{k\in K_1}\|\bar{D}_k W_k^T\nabla F_k\| = 0. \tag{5.19}$$

Next we only show that $\bar{D}_k$ can be replaced by D_k. From expressions (2.6) and (3.7) we need only to consider $\tilde{D}_k$ and $\hat{D}_k$. Denote

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

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 $i \in \{1, 2, \ldots, m\}$. Assume there exists $\epsilon_1 > 0$ such that for all $k \in K_1$ holds

$$|((\tilde{D}_k - \hat{D}_k)(-R_k^{-1}Y_k^T\nabla f_k))_i| > \epsilon_1. \tag{5.20}$$

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

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

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

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

Hence $(\hat{D}_k)_{ii} = (s_k)_i = (\tilde{D}_k)_{ii}$. Similarly for Case (ii) we have $(\tilde{D}_k)_{ii} = (\hat{D}_k)_{ii} = 1$ for sufficiently large $k \in K_2$.

Therefore $\lim_{k\in K_2} |(\tilde{D}_k - \hat{D}_k)_{ii}| = 0$. It contradicts with (5.20). So we obtain

$$\liminf_k |[(\tilde{D}_k - \hat{D}_k)(-R_k^{-1}Y_k^T\nabla f_k)]_i| = 0.$$

Hence we have

$$\begin{aligned} &\liminf_k \|(\tilde{D}_k - \hat{D}_k)(-R_k^{-1}Y_k^T\nabla f_k)\| \\ &\quad = \liminf_k \|(D_k - \bar{D}_k)W_k^T\nabla F_k\| = 0, \end{aligned}$$

which combining with (5.19) yields

$$\liminf_k \|D_k W_k^T \nabla F_k\| = 0,$$

and

$$\liminf_k [\|D_k W_k^T \nabla F_k\| + \|C_k\|] = 0.$$

The theorem has been proved. □

Remark 5.1 Theorem 5.9 and Proposition 2.1 show that there exists at least one accumulation point of $\{x_k\}$ generated by the Algorithm 4.1, which is a KKT point of (1.1).

6. Numerical example

First we give a method to solve subproblems of (3.2) and (3.5) respectively to satisfy (3.3), (3.4) and (3.8).

The quasi-normal component d_k^n is computed by the following formulas:

$$d_k^n = \begin{pmatrix} -\alpha_k Y_k R_k^{-T} C_k \\ 0 \end{pmatrix}, \tag{6.1}$$

where

$$\alpha_k = \begin{cases} 1, & \text{if } \|-Y_k R_k^{-T} C_k\| \le \tau\Delta_k, \\ \dfrac{\tau\Delta_k}{\|-Y_k R_k^{-T} C_k\|}, & \text{otherwise.} \end{cases} \tag{6.2}$$

We apply the basic conjugate-gradient algorithm proposed by Steihang and Toint to solve the problem (3.5) and modify it to incorporate the constraint $(d)_s \ge -\sigma_k s_k$.

Algorithm 6.1

Step 1. Set $(d^t)^0 = 0 \in R^n$, $r_0 = -\bar{g}_k$, $q_0 = \bar{D}_k r_0$, $d_0 = q_0$, $1 >> \epsilon_1 > 0$.

Step 2. For $i = 1, 2, \ldots$ do

(1) set

$$(d^t)^i = \begin{pmatrix} (d_1^t)^i \\ (d_2^t)^i \end{pmatrix}, \quad d_i = \begin{pmatrix} d_i^1 \\ d_i^2 \end{pmatrix},$$

where $(d_1^t)^i \in R^m$, $d_i^1 \in R^m$.

Compute $\gamma_i = \frac{r_i^T q_i}{d_i^T (W_k^T B_k W_k) d_i}$.

(2) Compute $\tau_i = \max\{\tau > 0 : \|\bar{D}_k^{-1}((d^t)^i + \tau d_i)\| \le \Delta_k; (d_1^t)^i + \tau d_i^1 \ge -\sigma_k s_k\}$.

(3) If $\gamma_i \le 0$ or $\gamma_i > \tau_i$, then set $d_k^t = (d^t)^i + \tau_i d_i$ goto Step 3; Otherwise set $(d^t)^{i+1} = (d^t)^i + \gamma_i d_i$.

(4) Update the residuals: $r_{i+1} = r_i - \gamma_i(W_k^T B_k W_k)d_i, q_{i+1} = \bar{D}_k^2 r_{i+1}$.

(5) Check truncation criteria: If $\sqrt{\frac{r_{i+1}^T q_{i+1}}{r_0^T q_0}} \leq \epsilon_1$, set $d_k^t = (d^t)^{i+1}$ goto Step 3.

(6) Compute $\alpha_i = \frac{r_{i+1}^T q_{i+1}}{r_i^t q_i}$ and set $d_{i+1} = q_{i+1} + \alpha_i d_i$.

Step 3. Compute $d_k = d_k^n + W_k d_k^t$.

A Matlab subroutine is programming to test Algorithm 4.1. We choose parameters of Algorithm 4.1 as $\tau = 0.8$, $a_1 = 0.5$, $\eta = 0.01$, $\Delta_{\max} = 10$, $\Delta_{\min} = 0.01$, $\bar{\rho} = 0.01$, $\rho_{-1} = 3$.

The chosen test example is problem 43 of [7],

$$\begin{aligned} \text{minimize} \quad & f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\ \text{subject to} \quad & x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 - 8 \leq 0, \\ & x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leq 0, \\ & 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5 \leq 0. \end{aligned}$$

The optimal point of the example is $x^* = (0, 1, 2, -1)$,and optimal value $f(x^*) = -44$. We choose $\sigma_k = 0.995$ for all k. There are two choices for B_k,

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

The result is reported in Table 6.1.

Table 6.1.

(x_0, s_0)	$(1,1,1,1,1,1,1)$		$(1.5,1.5,1.5,1.5,1,1,1)$		$(2,2,2,2,2,2,2)$	
B_k	$B_k^{(1)}$	$B_k^{(2)}$	$B_k^{(1)}$	$B_k^{(2)}$	$B_k^{(1)}$	$B_k^{(2)}$
k	64	85	104	85	118	154
x_k	x^*	x^*	x^*	x^*	x^*	x^*
$f(x_k)$	-44.0000	-44.0000	-44.0000	-44.0000	-44.0000	-44.0000
res 1	4.3844E-12	1.6429E-12	2.1909E-13	1.0529E-15	2.0241E-13	1.7198E-13
res 2	8.2805E-06	8.9343E-05	9.7256E-05	8.9088E-05	9.6992E-05	9.4903E-05

where res 1 $= \|C_k\|$, res 2 $= \|D_k W_k^T \nabla F_k\|$.

From Table 6.1 we can see that the calculated result is coincident with the theoretical analysis.

References

[1] El-Alem M., *A global convergence theory for the Celis-Dennis-Tapia trust-region algorithm for constrained optimization.* SIAM J. Numer. Anal. **28** No. 1 (1991), 266–290.

[2] El-Alem M., *A robust trust-region algorithm with nonmonotonic penalty parameter scheme for constrained optimization.* SIAM J. Optim. **5** No. 2 (1995), 348–378.

[3] Coleman T.F. and Li Y., *An interior trust region approach for nonlinear minimization subject to bounds.* SIAM J. Optim. **6** No. 2 (1996), 418–445.

[4] Dennis J.E., El-Alem M. and Maciel M.C., *A global convergence theory for general trust-region-based algorithm for equality constrained optimization.* SIAM J. Optim. **7** No. 1 (1997), 177–207.

[5] Dennis J.E. and Vicente L.N., *On the convergence theory of trust-region-based algorithm for equality-constrained optimization.* SIAM J. Optim. **7** No. 4 (1997), 927–950.

[6] Dennis J.E., Heinkenschloss M. and Vicente L.N., *Trust-region interior-point SQP algorithms for a class of nonlinear programming problems.* SIAM J. Control Optim. **36** No. 5 (1998), 1750–1794.

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

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

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

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

[11] Powell M.J.D., *Convergence properties of a class of minimization algorithms, in Nonlinear Programming 2.* O. Mangasarian, R. Meyer and S. Robinson, eds, Academic Press, New York, 1975, 1–27.

[12] Yuan Y.X., *On the convergence of a new trust region algorithm.* Numerische Mathematik **70** No. 4 (1995), 515–539.

Xiaojiao Tong
Department of Mathematics
Hunan University
Changsha 410082
P.R. China
E-mail: luoket@public.cs.hn.cn

Shuzi Zhou
Department of Mathematics
Hunan University
Changsha 410082
P.R. China