Annals of Operations Research 103, 175–191, 2001

A Conic Trust-Region Method for Nonlinearly Constrained Optimization *

WENYU SUN ** wenyus@hotmail.com
School of Mathematics and Computer Science, Nanjing Normal University, Nanjing 210097, China and Postgraduate Program in Computing Science, Pontificia Universidade Catolica do Parana, Curitiba, PR, 80215-901, Brazil

YA-XIANG YUAN yyx@lsec.cc.ac.cn
LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, Beijing 100080, China

Abstract. Trust-region methods are powerful optimization methods. The conic model method is a new type of method with more information available at each iteration than standard quadratic-based methods. Can we combine their advantages to form a more powerful method for constrained optimization? In this paper we give a positive answer and present a conic trust-region algorithm for non-linearly constrained optimization problems. The trust-region subproblem of our method is to minimize a conic function subject to the linearized constraints and the trust region bound. The use of conic functions allows the model to interpolate function values and gradient values of the Lagrange function at both the current point and previous iterate point. Since conic functions are the extension of quadratic functions, they approximate general nonlinear functions better than quadratic functions. At the same time, the new algorithm possesses robust global properties. In this paper we establish the global convergence of the new algorithm under standard conditions.

Keywords: trust-region method, conic model, constrained optimization, nonlinear programming

AMS subject classification: 65K05, 90C30

1. Introduction

In this paper we consider a general optimization problem with nonlinear equality constraints

$$\min_{x\in\mathbb{R}^n} \ f(x) \tag{1.1}$$

$$\text{s.t.} \quad c(x) = 0, \tag{1.2}$$

* This work was supported by the National Natural Science Foundation of China.
** Corresponding author.

where $c(x) = (c_1(x), \ldots, c_m(x))^{\mathrm{T}}$, $f(x)$, $c_i(x)$, $i = 1, \ldots, m$, are twice continuously differentiable. The Lagrangian function for problem (1.1), (1.2) is defined as

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

where λ_i for $i = 1, \ldots, m$ are Lagrange multipliers. We use the notation $g(x) = \nabla f(x)$ and $A(x) = [a_1(x), \ldots, a_m(x)] = [\nabla c_1(x), \ldots, \nabla c_m(x)]$ which is an $n \times m$ matrix. The constraint gradients $\nabla c_i(x)$ are assumed to be linearly independent for all x. Throughout this paper we define $A(x_k) = A_k$, $g(x_k) = g_k$, $c(x_k) = c_k$ for the kth iteration. The sequential quadratic programming method for (1.1), (1.2) computes a search direction by minimizing a quadratic model of the Lagrangian subject to the linearized constraints. That is, at the kth iteration, the following subproblem:

$$\min_{d \in \mathbb{R}^n} \; g_k^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} B_k d \tag{1.4}$$

$$\text{s.t.} \quad A_k^{\mathrm{T}} d + c_k = 0 \tag{1.5}$$

is solved to obtain a search direction d_k, where x_k is the current iterate point and B_k is symmetric and an approximation to the Hessian $\nabla_{xx} L(x_k, \lambda_k)$ of the Lagrangian of problem (1.1), (1.2). The next iteration has the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.6}$$

where $\alpha_k > 0$ is a step length which satisfies some line search conditions (see Han [14], Powell [21], Powell and Yuan [23], Yuan and Sun [33]).

Trust-region methods for problem (1.1), (1.2) have been studied by many researchers, including Vardi [30], Byrd, Schnabel and Schultz [3], Toint [29], Zhang and Zhu [34], Powell and Yuan [23] and El-Alem [8], etc. It is especially worth mentioning that the book of Conn, Gould and Toint [5] is an excellent and comprehensive one on trust-region methods. Trust-region methods are robust, can be applied to ill-conditioned problems and have strong global convergence properties. Another advantage of trust-region methods is that there is no need to require the approximate Hessian of the trust-region subproblem to be positive definite. For unconstrained problems, Nocedal and Yuan [18] show that a trust-region trial step is always a descent direction for any approximate Hessian. It is well known that for line search methods one generally has to assume the approximate Hessian to be positive definite in order to ensure that the search direction is a descent direction.

The collinear scaling of variables and conic model method for unconstrained optimization have been first studied by Davidon [6]. Sorensen [26] published detailed results on a class of conic model method and proved that a particular member of this class has the Q-superlinear convergence. Ariyawansa [1] modified the derivation of Sorensen [26] and established the duality between the collinear scaling BFGS and DFP methods. Ariyawansa and Lau [2] derived the collinear scaling Broyden's family and established its superlinear convergence results. Sheng [25] studied further the interpolation properties of conic model method. Sun [27] analysed several non-quadratic model methods and pointed out that the collinear scaling method is one of nonlinear scaling methods with scale invariance. A typical conic model for unconstrained optimization is

$$\psi_k(s) = \frac{g_k^{\mathrm{T}} s}{1 - h_k^{\mathrm{T}} s} + \frac{1}{2} \frac{s^{\mathrm{T}} B_k s}{(1 - h_k^{\mathrm{T}} s)^2}, \tag{1.7}$$

which is an approximation to $f(x_k + s) - f(x_k)$ and coincides up to first order with the objective function, where B_k is an approximate Hessian of $f(x)$ at x_k. The vector h_k is the associated vector for the collinear scaling in the kth iteration, and it is normally called the horizontal vector. If $h_k = 0$, the conic model reduces to a quadratic model.

Di and Sun [7] present a trust-region method based on conic model for unconstrained optimization where the trust-region subproblem has the form

$$\min \ \psi_k(s) = \frac{g_k^{\mathrm{T}} s}{1 - h_k^{\mathrm{T}} s} + \frac{1}{2} \frac{s^{\mathrm{T}} B_k s}{(1 - h_k^{\mathrm{T}} s)^2} \tag{1.8}$$

$$\text{s.t.} \quad \|D_k s\| \leqslant \Delta_k, \tag{1.9}$$

where D_k is a scaling matrix and Δ_k is the trust-region radius. Han, Han and Sun [13] presented a different kind of the conic trust region method for unconstrained optimization which is self-adjustable between conic model and quadratic model. In fact, since our assumption (2.38) below implies that the conic model always remains uniformly bounded over the trust-region, and as this model coincides up to first order with the objective function, the conic trust-region subproblem fits exactly in the framework of [5]. The conic model methods are the generalization of the quadratic model methods. They have several advantages. First, if the objective function has strong non-quadratic behavior or its curvature changes severely, the quadratic model methods often produce a poor prediction of the minimizer of the function. In this case, conic model approximates the objective function better than a quadratic, because it has more freedom in the model. Second, the quadratic model does not take into account the information concerning the function value in the previous iteration which is useful for algorithms. However, the conic model possesses richer interpolation information and satisfies four interpolation conditions of the function values and the gradient values at the current and the previous points. Using these rich interpolation information may improve the performance of the algorithms. Third, the conic model method has the similar global and local convergence properties as the quadratic model method. Finally, the initial and limited numerical results provided in [7,13,17] etc. show that the conic model method gives improvement over the quadratic model method. For example, in [7] we reported the detailed numerical results on 19 standard test functions described in [16] and compared our algorithm with SUMSL algorithm (see [10]). The experiment indicates that the performance of our algorithm is satisfactory and potentially competitive. In [13], an eigensolution-based method under the framework of conic trust-region with self-adjust strategy is presented. The elementary numerical result and some comparisons with the Newton's method and quasi-Newton method show that the values of the objective functions of this class of methods descend faster than other methods. In [17] a derivative-free method under the framework of conic trust-region is considered. The numerical results of some middle-scale problems also show that this class of methods is competitive.

In this paper we like to study the combination of trust-region techniques and conic model methods for optimization problem with equality constraints (1.1), (1.2). The aim of this paper is to discuss how to construct the conic trust-region subproblem and present a conic trust-region algorithm, and how to establish global convergence of the conic trust-region method in the non-linear constrained case. In this work, for the linearized constraints, we apply the null-space techniques, which have been used for line search type algorithms (see [9]) and trust-region algorithms (for example, see [3]). By using null-space techniques, we can rewrite the trust-region subproblems for constrained problems as subproblems for unconstrained problems, which can be solved easily (see [7]).

Throughout this paper, we use $\|\cdot\|$ for the 2-norm.

The organization of this paper is as follows. In the following section, the derivation of our method and a description of our algorithm are presented. The global convergence properties are studied in section 3. A short discussion is given in section 4.

2. The algorithm

The trial step of a trust-region algorithm is usually obtained by solving a trust-region subproblem. Because of the nice theoretical properties and performance of the sequential quadratic programming methods, it is natural to consider the subproblem that is similar to the subproblem (1.4), (1.5). In this paper we study the extension where the quadratic objective function in (1.4) is replaced by a conic function (1.7). To construct a trust-region subproblem, we must have a trust-region bound constraint:

$$\|s\| \leqslant \Delta_k, \tag{2.1}$$

where $\Delta_k > 0$ is the trust-region radius at the kth iteration. It is easy to see that there is a possibility that the linearized constraints (1.5) may have no solutions in the trust-region (2.1). To overcome this difficulty, we use a relaxed version of the linearized constraint as Byrd, Schnabel and Schultz [3] proposed. That is, at the kth iteration, the trial step s_k is computed by solving the following conic model trust-region subproblem

$$\min\ \psi_k(s) = \frac{g_k^{\mathrm{T}} s}{1 - h_k^{\mathrm{T}} s} + \frac{1}{2}\frac{s^{\mathrm{T}} B_k s}{(1 - h_k^{\mathrm{T}} s)^2} \tag{2.2}$$

$$\text{s.t.}\quad A_k^{\mathrm{T}} s + \theta_k c_k = 0, \tag{2.3}$$

$$\|s\| \leqslant \Delta_k, \tag{2.4}$$

where B_k denotes a symmetric approximation to the Hessian of the Lagrangian of problem (1.1), (1.2), $\theta_k \in (0, 1]$ is a relaxation parameter and $h_k \in \mathbb{R}^n$ is a horizontal vector. θ_k is so chosen that the feasible set of (2.3) and (2.4) is not empty. Geometrically speaking, the role of θ_k is to compress the feasible area of each constraints of (1.5) to the direction of origin (for example, see Omojokun [19], Zhang and Zhu [34]).

Assume that A_k has full column rank, and there exist an orthogonal matrix Q_k and a nonsingular upper triangular matrix R_k such that

$$A_k = Q_k R_k = \begin{bmatrix} Q_k^{(1)} & Q_k^{(2)} \end{bmatrix} \begin{bmatrix} R_k^{(1)} \\ 0 \end{bmatrix} = Q_k^{(1)} R_k^{(1)}. \tag{2.5}$$

The above relation implies that (2.3) can be rewritten as

$$\left(R_k^{(1)}\right)^{\mathrm{T}} \left(Q_k^{(1)}\right)^{\mathrm{T}} s = -\theta_k c_k. \tag{2.6}$$

Therefore the feasible point for (2.3) can be presented by

$$s = -\theta_k Q_k^{(1)} \left(R_k^{(1)}\right)^{-\mathrm{T}} c_k + Q_k^{(2)} u \tag{2.7}$$

for any $u \in \mathbb{R}^{n-m}$, since $Q_k^{(2)} u$ lies in the null space of A_k^{T}.

In order to ensure that s lies in the trust-region, we choose θ_k so that the norm of the first term of the right-hand side of (2.7) is at most $\tau \Delta_k$, where $\tau \in (0, 1)$ is a given constant. Using the notation

$$b_k = -Q_k^{(1)} \left(R_k^{(1)}\right)^{-\mathrm{T}} c_k, \tag{2.8}$$

we can let

$$\theta_k = \min\left[1, \frac{\tau \Delta_k}{\|b_k\|}\right] \tag{2.9}$$

which is the largest number θ in (0, 1] satisfying $\|\theta b_k\| \leqslant \tau \Delta_k$.

Define

$$\widetilde{\Delta}_k = \sqrt{\Delta_k^2 - \theta_k^2 \|b_k\|^2}. \tag{2.10}$$

The above definition and our choice of θ_k imply that $\widetilde{\Delta}_k \geqslant \sqrt{1-\tau^2}\Delta_k > 0$. Now, the subproblem (2.2)–(2.4) can be written to

$$\min\ g_k^{\mathrm{T}} w + \frac{1}{2} w^{\mathrm{T}} B_k w \tag{2.11}$$

$$\text{s.t.}\quad w = \frac{\theta_k b_k + Q_k^{(2)} u}{1 - h_k^{\mathrm{T}}(\theta_k b_k + Q_k^{(2)} u)}, \tag{2.12}$$

$$\|u\| \leqslant \widetilde{\Delta}_k. \tag{2.13}$$

One can see that w is a collinear scaling of s, with

$$w = \frac{s}{1 - h_k^{\mathrm{T}} s}, \qquad s = \frac{w}{1 + h_k^{\mathrm{T}} w}. \tag{2.14}$$

Equality (2.12) indicates the relationship between the scaling variable w and the reduced variable u, which is similar to the subproblem (49) for unconstrained optimization in [7] with the reduced variable u instead of the variable d. Therefore problem (2.11)–(2.13) can be solved by the technique given by Di and Sun [7].

There are several mathematically equivalent reformulations of the conic trust-region method. These different reformulations, with different matrix factorizations, lead to different reduced subproblems which are worthy of study in numerical optimization. For example, by means of collinear scaling (2.14), we can directly rewrite (2.2)–(2.4) as

$$\min\ g_k^{\mathrm{T}} w + \frac{1}{2} w^{\mathrm{T}} B_k w \tag{2.15}$$

$$\text{s.t.}\quad \overline{A}_k^{\mathrm{T}} w = -\theta_k c_k, \tag{2.16}$$

$$s = \frac{w}{1 + h_k^{\mathrm{T}} w}, \tag{2.17}$$

$$\|s\| \leqslant \Delta_k, \tag{2.18}$$

where

$$\overline{A}_k = A_k + \theta_k h_k c_k^{\mathrm{T}}, \tag{2.19}$$

which is a rank-one modification of A_k. If $\overline{A}_k$ has full column rank, there exist orthogonal matrix $\overline{Q}_k$ and nonsingular upper triangular matrix $\overline{R}_k$ such that

$$\overline{A}_k = \overline{Q}_k \overline{R}_k = \begin{bmatrix} \overline{Q}_k^{(1)} & \overline{Q}_k^{(2)} \end{bmatrix} \begin{bmatrix} \overline{R}_k^{(1)} \\ 0 \end{bmatrix} = \overline{Q}_k^{(1)} \overline{R}_k^{(1)}. \tag{2.20}$$

The above relation implies that (2.16) can be rewritten as

$$\big(\overline{R}_k^{(1)}\big)^{\mathrm{T}} \big(\overline{Q}_k^{(1)}\big)^{\mathrm{T}} w = -\theta_k c_k. \tag{2.21}$$

Therefore the feasible point for (2.16) can be presented by

$$w = -\theta_k \overline{Q}_k^{(1)} \big(\overline{R}_k^{(1)}\big)^{-\mathrm{T}} c_k + \overline{Q}_k^{(2)} z = \theta_k \overline{b}_k + \overline{Q}_k^{(2)} z \tag{2.22}$$

for any $z \in \mathbb{R}^{n-m}$, where $\overline{b}_k = -\overline{Q}_k^{(1)} (\overline{R}_k^{(1)})^{-\mathrm{T}} c_k$. Therefore subproblem (2.15)–(2.18) can be further rewritten as the following reduced subproblem

$$\min\ \widetilde{g}_k^{\mathrm{T}} z + \frac{1}{2} z^{\mathrm{T}} \widetilde{B}_k z \tag{2.23}$$

$$\text{s.t.}\quad \|s\| \leqslant \Delta_k, \tag{2.24}$$

$$s = \frac{\theta_k \overline{b}_k + \overline{Q}_k^{(2)} z}{1 + \theta_k h_k^{\mathrm{T}} \overline{b}_k + h_k^{\mathrm{T}} \overline{Q}_k^{(2)} z}, \tag{2.25}$$

where

$$\widetilde{g}_k = \big(\overline{Q}_k^{(2)}\big)^{\mathrm{T}} \big(g_k + \theta_k B_k \overline{b}_k\big) \tag{2.26}$$

and

$$\widetilde{B}_k = \big(\overline{Q}_k^{(2)}\big)^{\mathrm{T}} B_k \overline{Q}_k^{(2)}. \tag{2.27}$$

Here (2.25) indicates the relationship between the variable s and the reduced scaling variable z.

It should be pointed out that, in subproblem (2.23)–(2.25) for z, we use the decomposition (2.20), where $\overline{Q}_k^{(2)}z$ is the component in the null space of $\overline{A}_k^{\mathrm{T}} = (A_k+\theta_k h_k c_k^{\mathrm{T}})^{\mathrm{T}}$; in subproblem (2.11)–(2.13) for u, we use the decomposition (2.5), where $Q_k^{(2)}u$ is the component in the null space of A_k^{T}. Since our aim in this paper is to present the conic trust-region subproblem for solving equality constrained optimization and establish its convergence properties, we will not pursue this line about several reformulations of the problem and their numerical comparison.

By the way, in computations we only require the decomposition for A_k or $\overline{A}_k$. In fact, the decomposition of $\overline{A}_k$ can be easily obtained from that of A_k since they are related by a rank-one update matrix, and vice versa. Details can be found in Golub and Van Loan [12], Gill, Murray and Wright [11] or He and Sun [15].

The merit function we apply is the L_1 exact penalty function

$$\phi(x) = f(x) + \sigma \left\|c(x)\right\|_1, \tag{2.28}$$

where $\sigma > 0$ is a penalty parameter. As pointed out by Coleman and Conn [4], this function has the advantage that, for σ sufficiently large, any strong local minimizer of (1.1), (1.2) is a local minimizer of $\phi(x)$. It is found that this function is very convenient to be used as a merit function to force global convergence in line search type algorithms (for example, see [14]). Other merit functions include the L_∞ exact penalty function [32] and Fletcher's differentiable exact penalty function [23].

We define the actual reduction in the merit function by

$$\mathrm{Ared}_k = f(x_k) - f(x_k+s_k) + \sigma_k\big(\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}}s_k\right\|_1\big), \tag{2.29}$$

where s_k is a trial step computed by the algorithm at x_k and σ_k is a penalty parameter in the current iteration. Correspondingly, the predicted reduction is defined as

$$\mathrm{Pred}_k = -\frac{g_k^{\mathrm{T}}s_k}{1-h_k^{\mathrm{T}}s_k} - \frac{1}{2}\frac{s_k^{\mathrm{T}}B_k s_k}{(1-h_k^{\mathrm{T}}s_k)^2} + \sigma_k\big(\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}}s_k\right\|_1\big). \tag{2.30}$$

We choose

$$\sigma_k = \sigma_{k-1} \tag{2.31}$$

if

$$\psi_k(0) - \psi_k(s_k) \geqslant -\frac{\sigma_{k-1}}{2}\big(\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}}s_k\right\|_1\big), \tag{2.32}$$

otherwise we let

$$\sigma_k = \max\left[2\sigma_{k-1}, \frac{2(\psi_k(s_k) - \psi_k(0))}{\|c_k\|_1 - \|c_k + A_k^{\mathrm{T}}s_k\|_1}\right]. \tag{2.33}$$

Therefore the following inequality

$$\mathrm{Pred}_k \geqslant \frac{1}{2}\sigma_k\big(\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}}s_k\right\|_1\big) \tag{2.34}$$

holds for all k. The ratio of the actual reduction and the predicted reduction is

$$\rho_k = \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}, \tag{2.35}$$

which plays an important role in choosing the next iterate point and updating the new trust-region. For a given constant $\mu \in [0, 1)$, if $\rho_k > \mu$, the step s_k is accepted; otherwise, it is rejected and we reduce the trust-region radius and compute a new trial step s_k. The trust-region scheme is essentially similar to Conn, Could and Toint [5], Powell [20], Powell and Yuan [23], Di and Sun [7], Yuan and Sun [33] and Yuan [31].

In the following, we give a description of our algorithm.

Algorithm 2.1 (Conic trust-region algorithm for nonlinearly constrained optimization).

Step 0. Input data to be given are starting point $x_1 \in \mathbb{R}^n$, $B_1 \in \mathbb{R}^{n\times n}$, $\Delta_1 > 0$, $\sigma_0 > 0$, $\tau > 0$, and $\varepsilon > 0$. Choose $\mu \in [0, 1)$, $\eta \in (\mu, 1)$, $0 < \xi_1 < 1 < \xi_2$. Choose $h_1 \in \mathbb{R}^n$. Set $k = 1$.

Step 1. Compute $f(x_k)$, g_k and Q_k, R_k.
If $\|(Q_k^{(2)})^{\mathrm{T}} g_k\| \leqslant \varepsilon$ and $\|c_k\| \leqslant \varepsilon$, stop.

Step 2. Compute b_k from (2.8). Compute θ_k from (2.9).

Step 3. Solve (2.11)–(2.13) obtaining u_k;
Set $s_k = \theta_k b_k + Q_k^{(2)} u_k$.

Step 4. Compute σ_k from (2.31)–(2.33) and compute

$$\rho_k = \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k}.$$

Step 5. Set

$$x_{k+1} = \begin{cases} x_k + s_k & \text{if } \rho_k > \mu, \\ x_k & \text{otherwise;} \end{cases} \tag{2.36}$$

Choose the new trust-region bound satisfying

$$\Delta_{k+1} = \begin{cases} \max\left[\Delta_k, \xi_2\|s_k\|\right] & \text{if } \rho_k \geqslant \eta; \\ \xi_1\|s_k\| & \text{otherwise.} \end{cases} \tag{2.37}$$

Step 6. Generate B_{k+1}; generate h_{k+1}.

Step 7. $k := k + 1$, go to step 1.

In the algorithm above, instead of our original updating rule, the updating rule (2.37) is given according to the suggestion of a referee's report and has become more common. Such a rule guarantees that the trust-region radius remains bounded above since, as convergence occurs, $s_k \to 0$. Note that this is useful in the convergence analysis below because lemmas 3.2 and 3.4, theorems 3.5 and 3.6 all require the trust-region radius to be uniformly bounded.

How to choose the scaling vector h_k is one of the key issues of a conic model method. In general, h_{k+1} and B_{k+1} are so chosen that certain generalized quasi-Newton equations are satisfied, which means that the conic model function interpolates both the function values and the gradient values of the objective function at x_k and x_{k+1}. In [7,13], we have given some choices of h_k. More details on several updating formulae for h_{k+1} and B_{k+1} can be found, for example, in Davidon [6], Sorensen [26], Ariyawansa [1], Di and Sun [7] and Han, Han and Sun [13]. In this paper, we do not study any specific updating formulae for h_{k+1} and B_{k+1}. Instead, we assume that they are generalized by some interpolating conditions. The conditions that we assume for proving global convergence are that the matrices B_k are uniformly bounded and

$$\exists \delta_2 \in (0, 1)\colon\ \|h_k\|\Delta_k < \delta_2, \tag{2.38}$$

which ensures that the conic model function $\psi_k(s)$ is bounded over the trust-region $\{s \mid \|s\| \leqslant \Delta_k\}$. We would like to reiterate the fact that our algorithm reduces to a quadratic model based algorithm if $h_k = 0$ for all k. Note that, under the smoothness assumptions taken in this paper, the objective function is locally convex quadratic around a local minimizer. It means that choosing $h_k \simeq 0$ asymptotically is suitable when x_k is near the minimizer. This strategy has been employed in [7,13].

3. Global convergence

In this section, we establish the convergence results of our algorithm given in the previous section.

First we have the following lemmas:

Lemma 3.1. If $\{(A_k^{\mathrm T} A_k)^{-1}\}$ is uniformly bounded, then there exists a positive constant δ_1 such that

$$\theta_k\|b_k\| \leqslant \min\bigl(\tau\Delta_k, \delta_1\|c_k\|\bigr) \tag{3.1}$$

for all k.

Proof. The boundedness of $\{(A_k^{\mathrm T} A_k)^{-1}\}$ implies that $\{\|(R_k^{(1)})^{-1}\|\}$ is also uniformly bounded. Therefore there exists a positive constant δ_1 such that $\|(R_k^{(1)})^{-1}\| \leqslant \delta_1$ for all k. Hence the lemma follows from the definitions of θ_k and b_k in (2.8) and (2.9) respectively, where $Q_k^{(1)}$ is orthogonal in columns. □

Lemma 3.2. Let s_k be the solution of the subproblem (2.2)–(2.4). If $\{x_k\}$ and $\{(A_k^{\mathrm T} A_k)^{-1}\}$ are uniformly bounded, and if the quantities $\|h_k\|\Delta_k$ are bounded above by a constant $\delta_2 \in (0, 1)$, then there exist positive constants δ_3 and δ_4 such that

$$\psi_k(0) - \psi_k(s_k) \geqslant \delta_3 \left\| \left(Q_k^{(2)}\right)^{\mathrm{T}} g_k \right\| \min\left[\Delta_k, \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{\|B_k\|}\right] - \delta_4\left(1 + \|B_k\|\right) \min\left[\Delta_k, \|c\|\right] \tag{3.2}$$

for all k.

Proof. Define

$$s_k(t) = \theta_k b_k - t Q_k^{(2)} \left(Q_k^{(2)}\right)^{\mathrm{T}} g_k, \tag{3.3}$$

where θ_k is defined in (2.9). Then we can see that $s_k(t)$ is feasible of (2.3), (2.4) for all $t \in [0, \widetilde{\Delta}_k / \|(Q_k^{(2)})^{\mathrm{T}} g_k\|]$. Therefore, from the definitions of s_k and $s_k(t)$, we have that

$$\psi_k(0) - \psi_k(s_k) \geqslant \psi_k(0) - \psi_k\left(s_k(t)\right) \tag{3.4}$$

for all $t \in [0, \widetilde{\Delta}_k / \|(Q_k^{(2)})^{\mathrm{T}} g_k\|]$. Using $h_k^{\mathrm{T}} s_k(t) \leqslant \|h_k\| \Delta_k < \delta_2 < 1$, Cauchy–Schwartz inequality and $\|Q_k^{(2)}\| = 1$, we get

$$\begin{aligned}
\psi_k(0) - \psi_k\left(s_k(t)\right) &= -\frac{\theta_k g_k^{\mathrm{T}} b_k}{1 - h_k^{\mathrm{T}} s_k(t)} + t \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2}{1 - h_k^{\mathrm{T}} s_k(t)} \\
&\quad - \frac{\theta_k^2 b_k^{\mathrm{T}} B_k b_k}{2(1 - h_k^{\mathrm{T}} s_k(t))^2} + t \frac{\theta_k b_k^{\mathrm{T}} B_k Q_k^{(2)} (Q_k^{(2)})^{\mathrm{T}} g_k}{(1 - h_k^{\mathrm{T}} s_k(t))^2} \\
&\quad - \frac{t^2 g_k^{\mathrm{T}} Q_k^{(2)} (Q_k^{(2)})^{\mathrm{T}} B_k Q_k^{(2)} (Q_k^{(2)})^{\mathrm{T}} g_k}{2(1 - h_k^{\mathrm{T}} s_k(t))^2} \\
&\geqslant -\frac{\theta_k \|b_k\| \|g_k\|}{1 - \delta_2} + t \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2}{1 + \delta_2} \\
&\quad - \frac{\|\theta_k b_k\| \|B_k\| (\|\theta_k b_k\| + \widetilde{\Delta}_k)}{(1 - \delta_2)^2} - \frac{t^2 \|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2 \|B_k\|}{2(1 - \delta_2)^2} \\
&= -\frac{\theta_k \|b_k\| \|g_k\|}{1 - \delta_2} - \frac{\|\theta_k b_k\| \|B_k\| (\|\theta_k b_k\| + \widetilde{\Delta}_k)}{(1 - \delta_2)^2} \\
&\quad + \left[t \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2}{1 + \delta_2} - t^2 \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2 \|B_k\|}{2(1 - \delta_2)^2} \right]
\end{aligned} \tag{3.5}$$

for all $t \in [0, \widetilde{\Delta}_k / \|(Q_k^{(2)})^{\mathrm{T}} g_k\|]$. By calculus and (2.10), we have that

$$\begin{aligned}
&\max_{0 \leqslant t \leqslant \widetilde{\Delta}_k / \|(Q_k^{(2)})^{\mathrm{T}} g_k\|} \left\{ t \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2}{1 + \delta_2} - \frac{t^2 \|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2 \|B_k\|}{2(1 - \delta_2)^2} \right\} \\
&\geqslant \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2}{2(1 + \delta_2)} \min\left[\frac{\widetilde{\Delta}_k}{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}, \frac{(1 - \delta_2)^2}{(1 + \delta_2) \|B_k\|} \right] \\
&\geqslant \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{2(1 + \delta_2)} \min\left[\Delta_k, \frac{(1 - \delta_2)^2 \|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{(1 + \delta_2) \|B_k\|} \right] - \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{2(1 + \delta_2)} \left(\theta_k \|b_k\|\right).
\end{aligned} \tag{3.6}$$

Hence it follows from (3.4), (3.5) and (3.6) that

$$\begin{aligned}\psi_k(0)-\psi_k(s_k) &\geqslant \max_{0\leqslant t\leqslant \widetilde{\Delta}_k/\|(Q_k^{(2)})^{\mathrm{T}} g_k\|} \left|\psi_k(0)-\psi_k\big(s_k(t)\big)\right| \\ &\geqslant \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{2(1+\delta_2)} \min\left[\Delta_k, \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|(1-\delta_2)}{(1+\delta_2)\|B_k\|}\right] \\ &\quad - \frac{\theta_k\|b_k\|[\|B_k\|(\theta_k\|b_k\|+\widetilde{\Delta}_k)+2\|g_k\|]}{(1-\delta_2)^2}. \end{aligned} \tag{3.7}$$

The boundedness of x_k, the rule (2.37) and inequality (3.1) imply that $\theta_k\|b_k\|+\widetilde{\Delta}_k$ and $\|g_k\|$ are bounded above uniformly. This observation and inequality (3.7) show that there exist positive constants δ_3 and δ_4 such that (3.2) holds for all k. For instance, if $\theta_k\|b_k\|+\widetilde{\Delta}_k \leqslant \kappa_1$ and $\|g_k\| \leqslant \kappa_2$ for all sufficiently large k, the values

$$\delta_3 = \frac{1-\delta_2}{2(1+\delta_2)^2} \quad \text{and} \quad \delta_4 = \frac{\min(\kappa_1, 2\kappa_2)\min(\tau, \delta_1)}{(1-\delta_2)^2}$$

are acceptable. □

Lemma 3.3. If $\{(A_k^{\mathrm{T}} A_k)^{-1}\}$ is uniformly bounded, there exists a positive constant δ_5 such that

$$\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}} s_k\right\|_1 \geqslant \delta_5 \min\left\{\|c_k\|, \Delta_k\right\}. \tag{3.8}$$

Proof. If $\|b_k\| \leqslant \tau\Delta_k$, we have that $\theta_k = 1$. Then if follows from (2.7), (2.8) that

$$\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}} s_k\right\|_1 = \|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}} b_k\right\|_1 = \|c_k\|_1 \geqslant \|c_k\|. \tag{3.9}$$

If $\|b_k\| > \tau\Delta_k$, we have, by first using (2.7)–(2.9) and then (2.8) and (2.5), that

$$\begin{aligned}\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}} s_k\right\|_1 &= \|c_k\|_1 - \left\|c_k + \theta_k A_k^{\mathrm{T}} b_k\right\|_1 \\ &= \theta_k\|c_k\|_1 = \frac{\tau\Delta_k\|c_k\|_1}{\|b_k\|_2} \geqslant \frac{\tau}{\|(R_k^{(1)})^{-1}\|}\Delta_k. \end{aligned} \tag{3.10}$$

Thus, the lemma follows from the inequalities (3.9), (3.10). □

Using the above three lemmas, we can show that σ_k is bounded above.

Lemma 3.4. If $\{x_k\}$, $\{\|B_k\|\}$ and $\{(A_k^{\mathrm{T}} A_k)^{-1}\}$ are uniformly bounded and if the quantities $\|h_k\|\Delta_k$ are bounded above by a constant $\delta_2 \in (0, 1)$, then there exists an integer k^* such that

$$\sigma_k = \sigma_{k^*} \tag{3.11}$$

for all $k \geqslant k^*$.

Proof. If the lemma is not true, there exists a subsequence $\{x_{k_i}\}$ such that

$$\lim_{i\to\infty} \sigma_{k_i} = \infty, \tag{3.12}$$

and using (2.32)

$$\psi_{k_i}(0) - \psi_{k_i}(s_{k_i}) < -\frac{1}{2}\sigma_{k_i}\left(\|c_{k_i}\|_1 - \left\|c_{k_i} + A_{k_i}^{\mathrm{T}} s_{k_i}\right\|_1\right) \tag{3.13}$$

for all i. The above inequality and lemma 3.3 imply that

$$\psi_{k_i}(0) - \psi_{k_i}(s_{k_i}) < -\frac{1}{2}\sigma_{k_i}\delta_5 \min\left\{\|c_{k_i}\|, \Delta_{k_i}\right\}. \tag{3.14}$$

Now, the inequality (3.14) and lemma 3.2 give that

$$\frac{1}{2}\sigma_{k_i}\delta_5 \leqslant \delta_4\left(1 + \|B_{k_i}\|\right). \tag{3.15}$$

Then (3.15) and (3.12) contradict the boundedness of $\|B_k\|$. Therefore, the lemma is true. □

Theorem 3.5. Under the conditions of lemma 3.4,

$$\lim_{k\to\infty} \|c_k\| = 0. \tag{3.16}$$

Proof. If the theorem is not true, there exists a constant $\delta_6 > 0$ such that

$$\limsup_{k\to\infty} \|c_k\| = \delta_6 > 0. \tag{3.17}$$

Let $\widehat{x}$ be an accumulation point of $\{x_k\}$ satisfying $\|c(\widehat{x})\| = \delta_6$. There exists a positive constant δ_7 such that

$$\left\|c(x)\right\| \geqslant \delta_6/2 \quad \text{for all } \left\|x - \widehat{x}\right\| \leqslant \delta_7. \tag{3.18}$$

Define the set

$$K(\delta) = \left\{k \mid \|c_k\| \geqslant \delta\right\}. \tag{3.19}$$

Then, using (2.34) and lemma 3.3, for sufficiently large $k \in K(\delta_6/4)$ we have

$$\begin{aligned} \mathrm{Pred}_k &\geqslant \frac{1}{2}\sigma_k\left(\|c_k\|_1 - \left\|c_k + A_k^{\mathrm{T}} s_k\right\|_1\right) \geqslant \frac{1}{2}\sigma_k\delta_5 \min\left[\|c_k\|, \Delta_k\right] \\ &\geqslant \frac{1}{2}\delta_5\sigma_k \min\left[\frac{\delta_6}{4}, \Delta_k\right]. \end{aligned} \tag{3.20}$$

Define

$$K_0 = \left\{k \,\middle|\, \frac{\mathrm{Ared}_k}{\mathrm{Pred}_k} \geqslant \eta\right\}. \tag{3.21}$$

The boundedness of $\{x_k\}$ and the previous lemma imply that

$$\sum_{k\in K_0\cap K(\delta_6/4)} \mathrm{Pred}_k < \infty, \tag{3.22}$$

which, together with (3.20) implies that

$$\sum_{k\in K_0\cap K(\delta_6/4)} \Delta_k < \infty. \tag{3.23}$$

Therefore there exists an integer $\widehat{k}$ such that

$$\sum_{k\in K_0\cap K(\delta_6/4),\ k\geqslant\hat{k}} \Delta_k < \frac{\delta_7(1-\xi_1)}{4}, \tag{3.24}$$

where $0 < \xi_1 < 1$ is defined in algorithm 2.1. Because $\widehat{x}$ is an accumulation point of $\{x_k\}$, there exists $\overline{k} > \widehat{k}$ such that

$$\|x_{\overline{k}} - \widehat{x}\| < \frac{\delta_7}{4}. \tag{3.25}$$

By induction, we can use inequalities (3.18), (3.24) and (3.25) to prove that

$$k \in K\left(\frac{\delta_6}{2}\right) \tag{3.26}$$

for all $k \geqslant \overline{k}$. For $k = \overline{k}$, (3.26) follows from (3.25) and (3.18). Now assume that (3.26) is true for all $k = \overline{k}, \ldots, i$, we need to show that it is also true for $k = i+1$. From (3.24) and the updating rule of algorithm 2.1 that $\Delta_{k+1} \leqslant \xi_1\Delta_k$ for all $k \notin K_0$, we have that

$$\sum_{j=\overline{k}}^{i}\Delta_j = \sum_{j=\overline{k},\ j\in K_0}^{i} \Delta_j + \sum_{j=\overline{k},\ j\notin K_0}^{i} \Delta_j \leqslant \left(1+\frac{\xi_1}{1-\xi_1}\right) \sum_{j=\overline{k},\ j\in K_0}^{i} \Delta_j \leqslant \frac{\delta_7}{4}. \tag{3.27}$$

Therefore

$$\|x_{i+1} - \widehat{x}\| \leqslant \|x_{\overline{k}} - \widehat{x}\| + \sum_{j=\overline{k}}^{i} \Delta_j \leqslant \frac{\delta_7}{2}. \tag{3.28}$$

Thus, it follows from (3.18) and (3.19) that

$$x_{i+1} \in K\left(\frac{\delta_6}{2}\right). \tag{3.29}$$

By induction, we see that (3.26) holds for all $k \geqslant \overline{k}$.

Now from (3.26) and (3.27), by setting $i \to \infty$, we can see that

$$\sum_{k=\overline{k}}^{\infty} \Delta_k \leqslant \frac{\delta_7}{4}, \tag{3.30}$$

which implies

$$\lim_{k\to\infty} x_k = \widehat{x}, \qquad \lim_{k\to\infty} \Delta_k = 0. \tag{3.31}$$

It can be seen that

$$\text{Ared}_k = \text{Pred}_k + \text{o}\big(\|s_k\|\big) = \text{Pred}_k + \text{o}(\Delta_k) \tag{3.32}$$

which, together with (3.31) and (3.20), gives

$$\lim_{k\to\infty} \rho_k = \lim_{k\to\infty} \frac{\text{Ared}_k}{\text{Pred}_k} = 1. \tag{3.33}$$

The above limit shows that $\Delta_{k+1} \geqslant \Delta_k$ for all sufficiently large k, which contradicts (3.31). This completes our proof. □

Theorem 3.6. Under the conditions of lemma 3.4,

$$\liminf_{k\to\infty} \big\|\big(Q_k^{(2)}\big)^{\mathrm{T}} g_k\big\| = 0. \tag{3.34}$$

Proof. Assume that the theorem is false, hence there exists a positive constant δ_8 such that

$$\big\|\big(Q_k^{(2)}\big)^{\mathrm{T}} g_k\big\| \geqslant \delta_8 \tag{3.35}$$

for all k. If

$$\delta_4\big(1 + \|B_k\|\big)\|c_k\| > \delta_3 \big\|\big(Q_k^{(2)}\big)^{\mathrm{T}} g_k\big\| \frac{\Delta_k}{4}, \tag{3.36}$$

then, by using (2.34) and lemma 3.3, we have

$$\begin{aligned} \text{Pred}_k &\geqslant \frac{1}{2}\sigma_k\big(\|c_k\|_1 - \|c_k + A_k^{\mathrm{T}} s_k\|_1\big) \geqslant \frac{1}{2}\sigma_k\delta_5 \min\big[\Delta_k, \|c_k\|\big] \\ &\geqslant \frac{1}{2}\sigma_k\delta_5 \min\left[\frac{\delta_3\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{4\delta_4(1 + \|B_k\|)}, 1\right]\Delta_k. \end{aligned} \tag{3.37}$$

If inequality (3.36) fails, the following relation follows from (2.30) and (3.2) that

$$\text{Pred}_k \geqslant \psi_k(0) - \psi_k(s_k) \geqslant \frac{1}{2}\delta_3 \big\|\big(Q_k^{(2)}\big)^{\mathrm{T}} g_k\big\| \min\left[\Delta_k, \frac{\|(Q_k^{(2)})^{\mathrm{T}} g_k\|}{\|B_k\|}\right] \tag{3.38}$$

when k is large, because

$$\delta_4\big(1 + \|B_k\|\big)\|c_k\| < \frac{1}{2}\frac{\delta_3\|(Q_k^{(2)})^{\mathrm{T}} g_k\|^2}{\|B_k\|} \tag{3.39}$$

for all large k. Relations (3.37) and (3.38) tell us that there exists a positive constant δ_9 such that

$$\text{Pred}_k \geqslant \delta_9 \min[1, \Delta_k] \tag{3.40}$$

for all large k. From inequality (3.40), we can derive a contradiction similarly as in the proof of the previous theorem. This shows that (3.34) holds. □

Remarks.

(1) If the uniform boundedness assumption $\|B_k\| \leqslant M$ is replaced by

$$\|B_k\| \leqslant M_1 + M_2 \sum_{i=1}^{k} \Delta_i \quad \text{or} \quad \|B_k\| \leqslant M_1 + M_2 k,$$

where M_1 and M_2 are positive constants, we can also get the above result in which the proof is parallel to Powell [20,22] and Yuan [31].

(2) Theorem 3.6 guarantees, under the conditions of lemma 3.4, that the sequence $\{x_k\}$ has limit points and one of them is first-order critical. Along the line of proofs of theorems 6.4.6 and 15.4.11 of [5], the above result implies that all the limit points are first-order critical. This means that the stronger result

$$\lim_{k\to\infty} \left(Q_k^{(2)}\right)^{\mathrm{T}} g_k = 0 \tag{3.41}$$

holds.

4. Discussion

We have presented a conic trust-region algorithm for nonlinearly constrained optimization. Though the algorithm studied here is only for equality constrained problems, it can be extended to general inequality constraints. We have established the global convergence of the algorithm. Regarding the local convergence rate, we can prove that the method is superlinearly convergent if we assume that x_k converges to a second order sufficient point x^* and if B_k converges to the Hessian of the Lagrangian at the solution x^* in the following sense

$$\lim_{k\to\infty} \frac{\|(Q_k^{(2)})^{\mathrm{T}}(B_k - \nabla_{xx}^2 L(x^*, \lambda^*))s_k\|}{\|s_k\|} = 0 \tag{4.1}$$

which is also said to be projected Dennis–Moré condition, where $L(x, \lambda)$ is the Lagrangian function and λ^* is the corresponding multiplier at the solution x^*. The detailed discussion about the local convergence rate will be given in a separate paper.

Obviously, the conic trust-region method combines the advantages of both conic model method and trust-region method. It is a different type of model in a trust-region framework. We think that conic trust-region method has great potential from both the theoretical point of view and the numerical point of view (also see [7,13,17,26]). This class of methods is still in the infancy and a lot of topics and issues wait to be resolved. Therefore, further research into this class of algorithms, regarding theory and numerical experiments, is worthwhile.

Finally, we want to mention that it may be interesting to study the more general trust-region model using the constraints (2.3), (2.4) and with an objective function $f_k(s)$ which is a general model function,

$$\begin{aligned} \min\ & f_k(s) \\ \text{s.t.}\ & A_k^{\mathrm{T}} s + \theta_k c_k = 0, \\ & \|s\| \leqslant \Delta_k \end{aligned} \tag{4.2}$$

for constrained optimization. It is potentially valuable because such a general approach will ask us to consider the properties of $f_k(s)$ that make the trust-region method work.

Acknowledgments

Many thanks to two anonymous referees for their various valuable suggestions and criticisms. In fact, they pointed out several errors and gave more detailed comments. They suggested making clear the proofs of the lemmas and theorems. Their comments improved the presentation of this paper. We thank one of referees for the important reference [5] and new updating rule (2.37). We also thank Professor Danny Ralph of Cambridge University for his helpful suggestions.

References

[1] K.A. Ariyawansa, Deriving collinear scaling algorithms as extensions of quasi-Newton methods and the local convergence of DFP and BFGS related collinear scaling algorithm, Math. Programming 49 (1990) 23–48.

[2] K.A. Ariyawansa and D.T.M. Lau, Local and Q-superlinear convergence of a class of collinear scaling algorithms that extends quasi-Newton methods with Broyden's bounded-ϕ class of updates, Optimization 23 (1992) 323–339.

[3] R.H. Byrd, R.B. Schnabel and G.A. Shultz, A trust region algorithm for nonlinear constrained optimization, SIAM J. Numer. Anal. 24 (1987) 1152–1170.

[4] T.F. Coleman and A.R. Conn, Second order condition for an exact penalty function, Math. Programming 19 (1980) 178–185.

[5] A.R. Conn, N.I.M. Could and Ph.L. Toint, *Trust Region Methods* (SIAM, Philadelphia, USA, 2000).

[6] W.C. Davidon, Conic approximation and collinear scaling for optimizers, SIAM J. Numer. Anal. 17 (1980) 268–281.

[7] S. Di and W. Sun, Trust region method for conic model to solve unconstrained optimization problems, Optimization Methods and Software 6 (1996) 237–263.

[8] M. El-Alem, A robust trust region algorithm with a nonmonotonic penalty parameter scheme for constrained optimization, SIAM J. Optimization 5 (1995) 348–378.

[9] R. Fletcher, *Practical Methods of Optimization*, 2nd edn. (Wiley, 1987).

[10] D.M. Gay, Algorithm 611, subroutines for unconstrained minimization using a model/trust region approach, ACM Trans. Math. Softw. 9 (1983) 503–524.

[11] P.E. Gill, W. Murray and M.H. Wright, *Numerical Linear Algebra and Optimization*, Vol. 1 (Addison-Wesley, Reading, MA, 1991).

[12] G.E. Golub and C.F. Van Loan, *Matrix Computations*, 3rd edn. (Johns Hopkins University Press, Baltimore, 1996).

[13] Q. Han, J. Han and W. Sun, A modified quasi-Newton method with collinear scaling for unconstrained optimization, J. Computational and Applied Mathematics, submitted.
[14] S.P. Han, A globally convergent method for nonlinear programming, J. Optimization Theory and Applications 22 (1977) 297–309.
[15] X. He and W. Sun, *Introduction to Generalized Inverses of Matrices* (Jiangsu Sci. & Tech. Publisher, Nanjing, 1991).
[16] J.J. Moré, B.S. Garbow and K.E. Hillstrom, Testing unconstrained optimization software, ACM Trans. Math. Software 7 (1981) 17–41.
[17] Q. Ni and S. Hu, A new derivative free algorithm based on conic interpolation model, Technical report, Faculty of Science, Nanjing University of Aeronautics and Astronautics (2001).
[18] J. Nocedal and Y. Yuan, Combining trust region and line search techniques, Report NAM 07, Department of EECS, Northwestern University (1991).
[19] E.O. Omojokun, Trust region algorithms for optimization with nonlinear equality and inequality constraints, Ph.D. Thesis, University of Colorado at Boulder (1989).
[20] M.J.D. Powell, Convergence properties of a class of minimization algorithms, in: *Nonlinear Programming* 2, eds. O.L. Mangasarian, R.R. Meyer and S.M. Robinson (Academic Press, New York, 1975).
[21] M.J.D. Powell, Variable metric methods for constrained optimization, in: *Mathematical Programming, The State of the Art*, eds. M.G.A. Bachem and B. Korte (Springer, Berlin, New York, 1983) pp. 283–311.
[22] M.J.D. Powell, On the global convergence of trust region algorithms for unconstrained minimization, Math. Programming 29 (1984) 297–303.
[23] M.J.D. Powell and Y. Yuan, A trust region algorithm for equality constrained optimization, Math. Programming 49 (1991) 189–211.
[24] L. Qi and W. Sun, An iterative method for the minimax problem, in: *Minimax and Applications*, eds. D.Z. Du and P.M. Pardalos (Kluwer Academic, Boston, 1995) pp. 55–67.
[25] S. Sheng, Interpolation by conic model for unconstrained optimization, Computing 54 (1995) 83–98.
[26] D.C. Sorensen, The q-superlinear convergence of a collinear scaling algorithm for unconstrained optimization, SIAM Journal of Numerical Analysis 17 (1980) 84–114.
[27] W. Sun, On nonquadratic model optimization methods, Asia and Pacific Journal of Operations Research 13 (1996) 43–63.
[28] W. Sun, R. Sampaio and J. Yuan, Trust region algorithm for nonsmooth optimization, Applied Mathematics and Computation 85 (1997) 109–116.
[29] Ph.L. Toint, Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space, IMA J. Numerical Analysis 8 (1988) 231–252.
[30] A. Vardi, A trust region algorithm for equality constrained minimization: convergence properties and implementation, SIAM J. Numer. Anal. 22 (1985) 575–591.
[31] Y. Yuan, Condition for convergence of trust region algorithm for nonsmooth optimization, Math. Programming 31 (1985) 220–228.
[32] Y. Yuan, On the convergence of a new trust region algorithm, Numer. Math. 70 (1995) 515–539.
[33] Y. Yuan and W. Sun, *Optimization Theory and Methods* (Science Press, Beijing, China, 1997).
[34] J.Z. Zhang and D.T. Zhu, Projected quasi-Newton algorithm with trust region for constrained optimization, J. Optimization Theory and Application 67 (1990) 369–393.