A trust region interior point algorithm for infinite dimensional nonlinear programming

Chunfa Li · Shiqin Xu · Xue Yang

Received: 1 February 2007 / Revised: 26 November 2007 / Published online: 8 April 2008
© The Author(s) 2008

Abstract A trust region interior point algorithm for infinite dimensional nonlinear problem, which is motivated by the application of black-box approach to the distributed parameter system optimal control problem with equality and inequality constraints on states and controls, and with bounds on the controls is formulated. By introducing a proper functional which is analogous to the Lagrange function, both equality and inequality constraints can be treated identically and the first order optimality condition is given, then based on the works of Coleman, Ulbrich and Heinkenschloss, the trust region interior point algorithm which is employed to solve the optimization problem under consideration is presented.

Keywords Infinite dimensional nonlinear programming · Trust region method · Interior point method · Optimal control

Mathematics Subject Classification (2000) 49N35 · 49N10

1 Introduction

Problems in various areas of applications of optimization where differential equations are involved lead to formulations as infinite dimensional optimization problems. This

C. Li (✉) · S. Xu · X. Yang
School of Management, Tianjin University of Technology, 263 Hong Qi Nan Road, Tianjin,
Nan Kai District, 300191, People's Republic of China
e-mail: chunfali@163.com

S. Xu
e-mail: xsq0215@126.com

X. Yang
e-mail: ykxcfo-2008@163.com

class of problems belongs to optimal control problems, parameter identification problems and optimal shape design problems. In this paper, we discuss a class of nonlinear optimization problems in function space where the solution is restricted by pointwise upper and lower bounds and by finitely many equality and inequality constraints of functional type, which arises when the black-box approach is applied to optimal control problems governed by differential equations. It is well known the infinite dimensional optimization problem is studied much less comparing to the finite dimensional setting. Maurer and Zowe [1] presented an abstract optimality condition for infinite dimensional programming problems with cone constraints. Further Maurer investigated the first and second order sufficient conditions [2]. But all the mentioned works do not involve the algorithm to solve this kind of problems.

Trust region approaches for unconstrained and constrained optimization have been proven to be very successful both theoretically and practically at the finite dimensional setting. Different strategies combining with trust region method are developed rapidly to deal with various optimization problems with special structure. Minimization problems with simple bound constraints in R^n space form an important and common class of problems [3, 4]. Chen and Han [5] give a new non-monotone trust region algorithm using the active strategy to deal with these problems. The trust region interior point algorithm for bound constrained is introduced by Coleman and Li [4], the idea of this algorithm is based on the reformulation of the KKT necessary optimality conditions as a system of nonlinear equations using affine scaling matrix, the advantage of this approach is that the scaling matrix is determined by the distance of the iterate to the bounds and by the direction of the gradient, the nonlinear system is then solved by an affine scaling interior method. These methods enjoy strong theoretical convergence properties as well as a good numerical behavior. It is well known that the values of control function are restricted by a bounded constraints, according to this special structure, the trust region interior point method introduced by Coleman and Li had been applied to solve a discretized optimal control problem with bound constraints on the control by Dennis et al. [6]. But the infinite dimensional cases are studied much less comparing to the finite dimensional setting. Kelley and Scachs [7] and Toint [8] investigated the infinite dimensional optimization problem with simple bound constraints, they combined the trust region strategy with the projected gradient method to solve the optimal control problem with bound constraints only on controls in L^2 Hilbert space. In a recent paper, Ulbrich, Ulbrich and Heinkenschloss [9] extended the ideas introduced by Coleman and Li [4] to the corresponding infinite dimensional problem with simple bound constraints only. The trust region method was also be employed to solve a simple distributed parameter identification problem by Wang and Yuan [10], but they did not include constraint for the identified parameters. Besides the above works mentioned, the infinite dimensional optimization problem dealt with trust region strategy has seldom been, as far as we know, reported by the open literature.

Of course, the numerical solution of the infinite dimensional problems requires a discretization and allows the application of the previous mentioned algorithms to the resulting finite dimensional problems. We know that the convergence speed of iterative methods applied to a discretized problem may depend strongly on their convergence rate for the underlying infinite-dimensional problem, and although the algorithm in infinite dimensional spaces is only conceptual and not implementable, it can

serve as a tool to predict the convergence behavior of finite dimensional implementations. If the algorithm can be applied to the infinite dimensional problem and the convergence can be proved asymptotically in the infinite dimensional setting, then the same convergence behavior can be expected if the algorithm is applied to the finite dimensional discretized problems.

In this paper, we consider a trust region interior point algorithm for infinite dimensional nonlinear problem which is motivated by optimal control problems with equality and inequality constraints on states and controls, and with bounds on the controls, our algorithm is based on the application of Newton-like iteration to affine scaling formulation of the first order necessary optimality condition, and then we use trust-region interior point technique to guarantee global convergence and to handle the bound constraints.

The main idea of the algorithm presented in this paper comes from the trust region interior-point algorithm for bound constrained problem in R^n introduced by Coleman and Li [4], and it is an extension of the previous works by Ulbrich and Heinkenschloss [9]. However, we consider here the much more difficult issue incorporating the entire structure into an algorithm that handles equality and inequality constraints of functional type.

This paper is organized as follows. In Sect. 2, we formulate the general infinite dimensional nonlinear programming problem, and introduce a proper functional which is analogous to the Lagrange function and an affine scaling functional. Based on these two functionals we present in detail the first-order necessary conditions which will be needed to build the algorithm. In Sect. 3, the Newton-like strategy is used to deal with the above optimality conditions. Finally, the trust-region subproblems are built and the trust region interior point algorithm is presented.

2 Formulation of problem and optimality condition

Optimal control problem can be formulated as an abstract optimization problem in infinite dimensional space [11]. In this paper we will study the following optimization problem, which is motivated by the application of black-box approach to the optimal control problem with equality and inequality constraints on states and controls, and with bounds on the controls. Then the problem under consideration can be formulated as follows:

$$(\mathbf{P}) \quad \begin{cases} \min & J(u) \\ \text{s.t.} & G_j(u) = 0 \quad 1 \leq j \leq m_1 \\ & G_j(u) \leq 0 \quad m_1 + 1 \leq j \leq m \\ & u \in \mathcal{U}_{ad} \end{cases}$$

where $\mathcal{U}_{ad}$ is a nonempty, closed, convex and bounded subset of $L^\infty(X)$, and $J, G_j : L^2(X) \rightarrow R^1$ are given functionals with differentiability properties to be fixed later. For convenience and simplicity, we consider only the following simpler model throughout this paper.

$$\mathcal{U}_{ad} = \{u \in L^2(X) \mid a(x) \leq u(x) \leq b(x), \text{ a.e } x \in X; a(x), b(x) \in L^\infty(X)\}$$

We introduce the following notations. $\mathcal{L}(Y, Z)$ is the space of linear bounded operators from a Banach space Y into a Banach space Z . By $\|\cdot\|_q$ we denote the norm of the Lebesgue space $L^q(X)$, $1 \leq q \leq \infty$, and we write $(\cdot, \cdot)_2$ for the inner product of the Hilbert space $H = L^2(X)$. For $(v, w) \in (L^q(X), L^q(X)^*)$, with $L^q(X)^*$ denoting the dual space of $L^q(X)$, we use the canonical dual pairing $\langle v, w \rangle = \int_X v(x)w(x)d\mu(x)$ for which, if $q < \infty$, the dual space $L^q(X)^*$ is given by $L^{q'}(X)$, $\frac{1}{q} + \frac{1}{q'} = 1$ (in the case $q = 1$ this means $q' = \infty$). Especially, if $q = 2$, we have $L^2(X)^* = L^2(X)$ and $\langle \cdot, \cdot \rangle$ coincides with $(\cdot, \cdot)_2$.

Finally, set $U' = L^{p'}(X)$, $\frac{1}{p} + \frac{1}{p'} = 1$, which is the same as U^* if $p < \infty$. Moreover, it is easily seen that $w \rightarrow \langle \cdot, w \rangle$ defines a linear norm preserving injection from $L^1(X)$ to $L^\infty(X)^*$. Therefore, we may always interpret U' as subspace of U^* . Then we can get the following chain of continuous imbeddings:

$$V \hookrightarrow U \hookrightarrow H = H^* \hookrightarrow U' \hookrightarrow U^* \hookrightarrow V^*$$

Throughout this paper, we will work with differentiability in the Fréchet sense. We can write $\nabla J(u), \nabla G_j(u) \in U^*$, $j \in \{1, 2, \dots, m\}$ for the gradient and $\nabla^2 J(u), \nabla^2 G_j(u) \in \mathcal{L}(U, U^*)$, $j \in \{1, 2, \dots, m\}$ for the second Fréchet derivative of J, G_j at $u \in \mathcal{U}_{ad}$ respectively if they exist. The $\|\cdot\|_\infty$ -interior of $\mathcal{U}_{ad}$ is denoted by $\mathcal{U}_{ad}^o$:

$$\mathcal{U}_{ad}^o = \bigcup_{\delta > 0} \mathcal{U}^\delta, \mathcal{U}^\delta = \{u \in U \mid a(x) + \delta \leq u(x) \leq b(x) - \delta, x \in X\}$$

Definition 2.1 A point $\bar{u} \in \mathcal{U}_{ad}$ is a local solution of **(P)** if, for every feasible point u of **(P)**, there exist a real number of $r > 0$ such that $\|u - \bar{u}\|_{L^\infty(X)} \leq r$, we have

$$J(\bar{u}) \leq J(u)$$

In order to discuss the optimality conditions of problem **(P)**, introduce the following notations and assumptions.

Firstly, for arbitrary $\epsilon > 0$, denote set of ϵ -inactive constraints by

$$X_\epsilon = \{x \in X \mid a(x) + \epsilon \leq \bar{u}(x) \leq b(x) - \epsilon\}$$

Then make the following regularity assumption which is equivalent to the independence of the derivatives $\{G_j\}_{j \in I_0}$ in $L^\infty(X_\epsilon)$.

(A1) $\exists \epsilon_{\bar{u}} > 0$ and $\{h_j\}_{j \in I_0} \subset L^\infty(X)$, with $\text{supp } h_j \subset X_{\epsilon_{\bar{u}}}$ such that $G'_i(\bar{u})h_j = \delta_{ij}$, $i, j \in I_0$ where

$$I_0 = \{j \leq m \mid G_j(\bar{u}) = 0\}$$

I_0 is the set of indices corresponding to active constraints, we also denote the set of non-active constraints by I_-

$$I_- = \{j \leq m \mid G_j(\bar{u}) < 0\}$$

Under the regularity assumption, we can derive the following first order necessary conditions for optimality satisfied by $\bar{u}$ which is completely analogous to the finite dimensional problem with simple bounds.

Theorem 2.1 Assume that (A1) holds, let $\bar{u}$ be local optimal solution of problem (P), the objective functional J and the constraint functionals $\{G_j\}_{j=1}^m$ are of class C^1 in a neighborhood of $\bar{u}$. Then there exists real numbers $\{\bar{\lambda}_j\}_{j=1}^m \subset \mathbb{R}$, $\bar{\mu}_b \geq 0$, $\bar{\mu}_a \geq 0$, such that

$$\begin{cases} (\nabla J(\bar{u}) + \sum_{j=1}^m \bar{\lambda}_j \nabla G_j(\bar{u}) + \bar{\mu}_b - \bar{\mu}_a)h = 0 & \forall h \in L^\infty(X) \\ (a(x) - \bar{u})\bar{\mu}_a + (\bar{u} - b(x))\bar{\mu}_b = 0 \\ \bar{\mu}_a \geq 0, \bar{\mu}_b \geq 0 \\ \bar{\lambda}_j \geq 0, m_1 + 1 \leq j \leq m, \bar{\lambda}_j = 0 & \text{if } j \in I_- \\ \bar{\lambda}_j G_j(\bar{u}) = 0 \end{cases} \quad (1)$$

Now introduce the following functional

$$L(u, \lambda, \vartheta) = J(u) + \sum_{j=1}^{m_1} \lambda_j G_j(u) + \sum_{j=m_1+1}^m \frac{1}{2} \vartheta_j^2 G_j(u) \quad (2)$$

We can obtain the following first order optimality necessary condition of (P).

Theorem 2.2 Assume that (A1) holds, J and $\{G_j\}_{j=1}^m$ are of class C^1 in a neighborhood of $\bar{u}$. Then there exist real numbers $\{\bar{\lambda}_j\}_{j=1}^{m_1}$ and $\{\bar{\vartheta}_j\}_{j=m_1+1}^m$ such that

$$\left\langle \nabla J(\bar{u}) + \sum_{j=1}^{m_1} \bar{\lambda}_j \nabla G_j(\bar{u}) + \sum_{j=m_1+1}^m \frac{1}{2} \bar{\vartheta}_j^2 \nabla G_j(\bar{u}), u - \bar{u} \right\rangle \geq 0, \quad \forall u \in \mathcal{U}_{ad} \quad (3)$$

For the proof of this theorem, we can referred to Clark [12].

Be similar to Theorem 2.1, we can obtain an equivalent first order necessary optimality conditions of problem (P).

Theorem 2.3 Assume that the regularity condition (A1) holds, let $\bar{u}$ be the local optimal solution of (P). Then there exist real numbers $\{\bar{\lambda}_j\}_{j=1}^{m_1}$, $\{\bar{\vartheta}_j\}_{j=m_1+1}^m$, $\bar{\mu}_b \geq 0$ and $\bar{\mu}_a \geq 0$ such that

$$\begin{cases} (\nabla J(\bar{u}) + \sum_{j=1}^{m_1} \bar{\lambda}_j \nabla G_j(\bar{u}) + \sum_{j=m_1+1}^m \frac{1}{2} \bar{\vartheta}_j^2 \nabla G_j(\bar{u}) \\ \quad + \bar{\mu}_b - \bar{\mu}_a)h = 0, & \forall h \in L^\infty(X) \\ (\bar{u} - a(x))\bar{\mu}_a + (b(x) - \bar{u})\bar{\mu}_b = 0 \\ \bar{\mu}_a \geq 0, \bar{\mu}_b \geq 0 \\ G_j(\bar{u}) = 0, & j \in \{1, 2, \dots, m_1\} \\ \bar{\vartheta}_j G_j(\bar{u}) = 0, & j \in \{m_1 + 1, \dots, m\} \end{cases} \quad (4)$$

The proofs of preceding Theorems 2.1 and 2.3 are almost similar to the finite dimensional situation.

In order to use the Coleman-Li affine scaling method and the interior-point method to solve the problem (P), we have to impose some additional assumptions on the objective functional $J(u)$ and the constraint functionals $G_j(u)$, $j \in \{1, 2, \dots, m\}$.

(A2) J, G_j are twice continuously Fréchet differentiable on $\mathcal{N}$ respectively with derivatives $\nabla J : \mathcal{N} \rightarrow L^{p'}$, $\nabla^2 J : \mathcal{N} \rightarrow \mathcal{L}(L^2, L^{p'})$; $\nabla G_j : \mathcal{N} \rightarrow L^{p'}$, $\nabla^2 G_j : \mathcal{N} \rightarrow \mathcal{L}(L^2, L^{p'})$, $\frac{1}{p} + \frac{1}{p'}$, where $\mathcal{N}$ is an open neighborhood $\mathcal{N} \subset \mathcal{U}_{ad}$; Moreover, there is a constant $C > 0$ such that $\|\nabla J(u)\|_\infty < C$, $\|\nabla G_j(u)\|_\infty < C$ for all $u \in \mathcal{U}_{ad}$ and there exist functions $f, g_j \in L^2(X)$, $j \in \{1, 2, \dots, m\}$, such that for every $h \in L^\infty(X)$,

$$\begin{aligned}\nabla J(\bar{u})h &= \int_X f(\bar{u})(x)h(x)d\mu(x) \quad \text{and} \quad \nabla G_j(\bar{u})h \\ &= \int_X g_j(\bar{u})(x)h(x)d\mu(x) \quad j \in \{1, 2, \dots, m\}\end{aligned}$$

(A3) If $\{h_k\}_{k=1}^\infty \subset L^\infty(X)$ is bounded, $h \in L^\infty(X)$, $h_k(x) \rightarrow h(x)$ a.e. in X , then

$$\left(\nabla^2 J(\bar{u}) + \sum_{j=1}^m \bar{\mu}_j \nabla^2 G_j(\bar{u}) \right) h_k^2 \rightarrow \left(\nabla^2 J(\bar{u}) + \sum_{j=1}^m \bar{\mu}_j \nabla^2 G_j(\bar{u}) \right) h^2$$

Define

$$d(\bar{u})(x) = f(\bar{u})(x) + \sum_{j=1}^{m_1} \bar{\lambda}_j g_j(\bar{u})(x) + \sum_{j=m_1+1}^m \frac{1}{2} \bar{\vartheta}_j^2 g_j(\bar{u})(x) \quad (5)$$

According to the above assumptions and notations, another first order optimality condition of problem **(P)** can be presented as follows.

Theorem 2.4 Assume that conditions (A1)–(A3) are valid, and let $\bar{u}$ be a local solution of problem **(P)**, $J(u)$ and $\{G_j(u)\}_{j=1}^m$ are Fréchet differentiable at $\bar{u}$ with $\nabla J(\bar{u}) \in U'$, $\{\nabla G_j(u)\}_{j=1}^m \in U'$. Then

$$d(\bar{u})(x) = \begin{cases} = 0 & \text{for a.e. } x \in X \text{ where } a(x) < \bar{u} < b(x) \\ \leq 0 & \text{for a.e. } x \in X \text{ where } \bar{u}(x) = b(x) \\ \geq 0 & \text{for a.e. } x \in X \text{ where } \bar{u}(x) = a(x) \end{cases} \quad (6)$$

Proof Since $\bar{u}$ is a local solution, according to Theorem 2.2 and the assumption (A2), we have

$$\begin{aligned}& \left\langle \nabla J(\bar{u}) + \sum_{j=1}^{m_1} \bar{\lambda}_j \nabla G_j(\bar{u}) + \sum_{j=m_1+1}^m \frac{1}{2} \bar{\vartheta}_j^2 \nabla G_j(\bar{u}), u - \bar{u} \right\rangle \\ &= \int_X \left(f(\bar{u})(x) + \sum_{j=1}^{m_1} \bar{\lambda}_j g_j(\bar{u})(x) + \sum_{j=m_1+1}^m \frac{1}{2} \bar{\vartheta}_j^2 g_j(\bar{u})(x) \right) (u - \bar{u}) d\mu(x) \\ &= \int_X d(\bar{u})(x) (u - \bar{u}) d\mu(x) \geq 0 \quad \text{for all } a(x) \leq \bar{u}(x) \leq b(x)\end{aligned}$$

and then

$$d(\bar{u})(x)(u - \bar{u}) \geq 0 \quad \text{for a.e. } x \in X \text{ where } a(x) \leq \bar{u}(x) \leq b(x) \quad (7)$$

Obviously, if $a(x) < \bar{u}(x) < b(x)$, then we can conclude that $d(\bar{u})(x) = 0$; if $\bar{u} = a(x)$, then by (7) $d(\bar{u})(x) \geq 0$; analogously if $\bar{u} = b(x)$, then $d(\bar{u})(x) \leq 0$. $\square$

Now define a scaling functional which is assumed to satisfy

$$q(u)(x) = \begin{cases} = 0 & \text{if } u(x) = a(x) \text{ and } d(u)(x) \geq 0 \\ = 0 & \text{if } u(x) = b(x) \text{ and } d(u)(x) \leq 0 \\ > 0 & \text{otherwise} \end{cases} \quad (8)$$

for all $x \in X$.

Then $(\bar{u}, \bar{\lambda}, \bar{\vartheta})$ satisfy first order necessary conditions for optimality of problem **(P)**, if and only if that

$$\begin{cases} q^t(\bar{u})d(\bar{u}) = 0 \\ G_j(\bar{u}) = 0, & j \in \{1, 2, \dots, m_1\} \\ \bar{\vartheta}_j G_j(\bar{u}) = 0, & j \in \{m_1 + 1, \dots, m\} \end{cases} \quad (9)$$

where $t > 0$ is arbitrary.

Remark 2.1 Since $q^t, t > 0$, also satisfy (8), we may restrict ourselves to the case $t = 1$. First of all, assume that Theorem 2.3 holds. If $a(x) < \bar{u}(x) < b(x)$, then in terms of the complementary conditions, we can derive that $\bar{\mu}_a = \bar{\mu}_b = 0$ which implies $d(\bar{u})(x) = 0$ for a.e. $x \in X$; if $\bar{u}(x) = a(x)$, then $\bar{\mu}_b = 0$ and $\bar{\mu}_a \geq 0$, we can conclude that $d(\bar{u})(x) \geq 0$ for a.e. $x \in X$; Finally, in the case where $\bar{u}(x) = b(x)$, we can derive that $d(\bar{u})(x) \leq 0$. On the other hand, let $q(\bar{u})d(\bar{u}) = 0$ hold. If we assume that $d(\bar{u}) < 0$, which must result in $q(\bar{u}) = 0$, according to the definition of $q(u)(x)$, it can be shown that $\bar{u} = b(x)$, then let $\bar{\mu}_a = 0$ and $\bar{\mu}_b = -d(\bar{u})(x) > 0$; Analogously, assume that $d(\bar{u})(x) > 0$, we can derive that $\bar{u}(x) = a(x)$, hence, let $\bar{\mu}_b = 0$ and $\bar{\mu}_a = d(\bar{u})(x) > 0$. Finally if $d(\bar{u})(x) = 0$, then let $\bar{\mu}_a = \bar{\mu}_b = 0$, therefore, the equivalence is valid.

Remark 2.2 According to the above conclusions, if we can figure out all the points $(\bar{u}, \bar{\lambda}, \bar{\vartheta})$ which satisfy the system of equation (9), then the local solutions of **(P)** must belong to the set which consists of the first component $\bar{u}$ of all $(\bar{u}, \bar{\lambda}, \bar{\vartheta})$ satisfying (9).

3 Newton step and affine scaling functional

In this section, we concentrate our efforts on solving the nonlinear system (9) by means of Newton method. The bound constraints on u are enforced by a scaling of Newton step, but there exists a difficulty pointed out in [9]. In general it is not possible to find a functional q satisfying the condition (8) that depends smoothly on u , hence

it is crucial to choose the affine scaling functional q in such a way that q is as smooth as possible in a neighborhood of a KKT-point of problem **(P)**. Based on this idea and referred to [13], define an affine scaling functional as follows:

$$q_I(u)(x) = \begin{cases} u(x) - a(x) & \text{if } d(u)(x) > 0 \text{ or } d(u)(x) = 0, u(x) \leq \frac{1}{2}(a(x) + b(x)) \\ b(x) - u(x) & \text{if } d(u)(x) < 0 \text{ or } d(u)(x) = 0, u(x) > \frac{1}{2}(a(x) + b(x)). \end{cases} \quad (10)$$

It is easily to verify that q_I satisfies (8) as mentioned in [13] that the functional q_I is discontinuous at a KKT-point $\bar{u}$ since $|q_I(u) - q_I(\bar{u})| = |b(x) - a(x) - |u - \bar{u}||$ on $\{x \in X \mid q(\bar{u})(x)d(\bar{u})(x)\} < 0$. It can be deduced that

$$q'_I(u)(x) = \begin{cases} 1 & \text{if } d(u)(x) > 0 \text{ or } d(u)(x) = 0 \text{ and } u(x) \leq \frac{1}{2}(a(x) + b(x)) \\ -1 & \text{if } d(u)(x) < 0 \text{ or } d(u)(x) = 0 \text{ and } u(x) > \frac{1}{2}(a(x) + b(x)). \end{cases} \quad (11)$$

As it is well known that it is indispensable to find a suitable substitute for the derivation of $q^t d$, when the Newton method is employed to solve the non-smooth nonlinear system (9) with $t = 1$. Such as mentioned in [14], we can choose an approximate derivative $Q_u(u)w \in L(U, U')$ to replace the generally nonexistence derivative of $u \in \mathcal{U}_{ad} \rightarrow U'$ at u , here and in the equality, the linear operator $Q(u)$ denotes the pointwise multiplication operator associated with $q(u)$, i.e.,

$$Q(u) : v \rightarrow q(u)v.$$

Now we discuss this problem mentioned above. Obviously, it can be shown that a Newton step in (u, λ, ϑ) on nonlinear system (9) is given by

$$\begin{cases} (Q(u)(\nabla^2 J(u) + \sum_{j=1}^{m_1} \lambda_j \nabla^2 G_j(u) \sum_{j=m_1+1}^m \frac{1}{2} \vartheta_j^2 \nabla^2 G_j(u)) \\ \quad + Q_u(u)d(u)I) \Delta u \\ \quad + Q(u) \sum_{j=1}^{m_1} g_j(u) \Delta \lambda_j + q(u) \sum_{j=m_1+1}^m \vartheta_j g_j(u) \Delta \vartheta_j \\ = -Q(u)d(u) \\ g_j(u) \Delta u = -G_j(u) & j = 1, 2, \dots, m_1 \\ \vartheta_j g_j(u) \Delta u + G_j(u) \Delta \vartheta_j = -\vartheta_j G_j(u) & j = m_1 + 1, \dots, m \end{cases} \quad (12)$$

Multiplying throughout on the left of the first equation by Q^{-1} , the above system can be equivalently written as

$$\begin{cases} (\nabla^2 J(u) + \sum_{j=1}^{m_1} \lambda_j \nabla^2 G_j(u) + \sum_{j=m_1+1}^m \frac{1}{2} \vartheta_j^2 \nabla G_j^2(u) \\ \quad + Q^{-\frac{1}{2}} Q_u(u)d(u) Q^{-\frac{1}{2}}) \Delta u \\ \quad + \sum_{j=1}^{m_1} g_j(u) \Delta \lambda_j + \sum_{j=m_1+1}^m \vartheta_j g_j(u) \Delta \vartheta_j = -d(u) \\ g_j(u) \Delta u = -G_j(u) & j = 1, 2, \dots, m_1 \\ \vartheta_j g_j(u) \Delta u + G_j(u) \Delta \vartheta_j = -\vartheta_j G_j(u) & j = m_1 + 1, \dots, m \end{cases} \quad (13)$$

Now we adopt the following notations, let

$$\begin{aligned} B &= S + \sum_{j=1}^{m_1} \lambda_j T_j + \sum_{j=m_1+1}^m \frac{1}{2} \vartheta_j^2(u) T_j, & C &= Q^{-\frac{1}{2}}(u) Q_u(u) d(u) Q^{-\frac{1}{2}}(u) \\ D_j &= g_j(u) \quad j = 1, 2, \dots, m_1, & E_j &= \vartheta_j g_j(u) \quad j = m_1 + 1, \dots, m \\ D &= (D_1, D_2, \dots, D_{m_1}), & E &= (E_{m_1+1}, E_{m_1+1}, \dots, E_m) \\ G^\lambda &= (G_1, G_2, \dots, G_{m_1}), & G^\vartheta &= (G_{m_1+1}, G_{m_1+2}, \dots, G_m) \\ \lambda &= (\lambda_1, \lambda_2, \dots, \lambda_{m_1})^T, & \vartheta &= (\vartheta_{m_1+1}, \vartheta_{m_1+2}, \dots, \vartheta_m)^T \end{aligned}$$

$$M = \begin{pmatrix} B + C & D & E \\ D^T & 0 & 0 \\ E^T & 0 & \text{diag}(G^\vartheta) \end{pmatrix}$$

$$s = \begin{pmatrix} \Delta u \\ \Delta \lambda \\ \Delta \vartheta \end{pmatrix} \quad l = \begin{pmatrix} u \\ \lambda \\ \vartheta \end{pmatrix} \quad \xi = \begin{pmatrix} d(u) \\ (G^\lambda)^T \\ \text{diag}(G^\vartheta) \vartheta \end{pmatrix}$$

where $S, T_j (1, 2, \dots, m)$ denote the symmetric approximation of $\nabla^2 J(u), \nabla^2 G_j(u)$ respectively, and their corresponding norms $\|S\|_{\mathcal{L}(U, U)}, \|T_j\|_{\mathcal{L}(U, U)}$ are uniformly bounded respectively.

We are now able to formulate the following Newton-like iteration for the solution of system of equation (9). Give $l_k = (u^k, \lambda^k, \vartheta^k)^T \in U \times R^{m_1} \times R^{m-m_1}$, here $u_k \in \mathcal{U}_{ad}^0$, the new iterate $l_{k+1} = l_k + s_k \in \mathcal{U}_{ad}^0 \times R^{m_1} \times R^{m-m_1}$ can be solved by

$$M_k s_k = -\xi_k \quad (14)$$

Introducing the scaled step $\Delta \hat{u}_k = Q_k^{-\frac{1}{2}} \Delta u_k$, thus it is easily to get the following equivalent system.

$$\begin{cases} (Q_k^{\frac{1}{2}} B_k Q_k^{\frac{1}{2}} + Q_u^k d_k) \Delta \hat{u}_k + Q_k^{\frac{1}{2}} D^k \Delta \lambda^k \\ \quad + Q_k^{\frac{1}{2}} E^k \Delta \vartheta^k = -Q_k^{\frac{1}{2}} d_k \\ Q_k^{\frac{1}{2}} g_j^k \Delta \hat{u}_k = -G_j^k & j = 1, 2, \dots, m_1 \\ \vartheta_j^k Q_k^{\frac{1}{2}} g_j^k \Delta \hat{u}_k + G_j^k \Delta \vartheta_j^k = -\vartheta_j^k G_j^k & j = m_1 + 1, \dots, m \end{cases} \quad (15)$$

Denote that $\hat{D}^k = Q_k^{\frac{1}{2}} D^k, \hat{E}^k = Q_k^{\frac{1}{2}} E^k, \hat{d}_k = Q_k^{\frac{1}{2}} d^k, \hat{Q}_u^k = Q_u^k Q_k^{-\frac{1}{2}}, \hat{B}_k = Q_k^{\frac{1}{2}} B_k Q_k^{\frac{1}{2}}$. Then the system of equations (15) can be formulated by the following form

$$\hat{M}_k \hat{s}_k = -\hat{\xi}_k \quad (16)$$

where

$$\hat{M}_k = \begin{pmatrix} \hat{B}_k + \hat{Q}_u^k \hat{d}_k & \hat{D}^k & \hat{E}^k \\ (\hat{D}^k)^T & 0 & 0 \\ (\hat{E}^k)^T & 0 & \text{diag}(G^{\vartheta^k}) \end{pmatrix}$$

$$\hat{s}_k = \begin{pmatrix} \Delta \hat{u}_k \\ \Delta \lambda^k \\ \Delta \vartheta^k \end{pmatrix}, \quad \hat{\xi}_k = \begin{pmatrix} \hat{d}_k \\ (G^{\lambda^k})^T \\ \text{diag}(G^{\vartheta^k}) \vartheta^k \end{pmatrix}$$

Obviously, since $\hat{M}_k$ is symmetric, $\hat{s}_k$ is a solution of (16) if and only if it is a stationary point of the quadratic functional

$$\hat{\psi}_k(\hat{s}) = \langle \hat{s}, \hat{\xi}_k \rangle + \frac{1}{2} \langle \hat{s}, \hat{M}_k \hat{s} \rangle. \quad (17)$$

Remark 3.1 Using the freedom provided by (8) to choose a proper affine scaling functional $q(u)$, we can prove that $\hat{\psi}_k(\hat{s})$ is convex and admits a global minimum at $\hat{s} = 0$ if $u_k = \bar{u}$ is the local solution of (P), and $\lambda^k = \bar{\lambda}$, $\vartheta^k = \bar{\vartheta}$ are the corresponding multipliers.

4 Formulation of the algorithm

In the previous section, we have transformed the first optimality necessary condition of (P) into a hybrid Newton equation which can be solved by classical Newton method. It is known that the Newton method has high rate of convergence, but it is guaranteed to converge only in the local vicinity of solution. One of the possibilities to enlarge the region of its convergence is line search algorithm, and such a algorithm obtains a search direction in each iteration, and search along this directions to obtain a better point. The search direction is a descent direction, normally computed by solving a subproblem that approximates the original optimization problem near the current iterate. In this paper, we use a relatively new strategy—trust region algorithm, the key content of a trust region algorithm are how to compute the trust region trial step and how to decide whether a trial step be accepted or not.

To globalize the iterate, the corresponding trust region subproblem based on the above Newton steps should be constructed. Before doing this, we present the following assumption and adopt some notations.

(A4) If approximation to the Hessian matrices is used, then we require that all of them be uniformly bounded, namely, the symmetric approximation S^k, T_j^k , ($j = 1, 2, \dots, m$) of $\nabla^2 J^k, \nabla^2 G_j^k$ are uniformly bounded in norms of $\|\cdot\|_{\mathcal{L}(U, U')}$ respectively.

Let

$$\begin{aligned} \hat{N}_k &= \begin{pmatrix} \hat{B}_k + \hat{Q}_u^k \hat{d}_k & \hat{E}^k \\ (\hat{E}^k)^T & \text{diag}(G^{\vartheta^k}) \end{pmatrix}, & \Delta \hat{v}_k &= \begin{pmatrix} \Delta \hat{u}_k \\ \Delta \vartheta^k \end{pmatrix} \\ \hat{\eta}^k &= \begin{pmatrix} \hat{d}_k \\ \text{diag}(G^{\vartheta^k}) \vartheta^k \end{pmatrix} \\ \hat{N}_k &= \begin{pmatrix} \text{diag}(\hat{D}^k) & 0 \\ 0 & 0 \end{pmatrix}, & \hat{\xi}^k &= \begin{pmatrix} (G^{\lambda^k})^T \\ 0 \end{pmatrix} \end{aligned}$$

Then the system of equations (16) can be reformulated as the following form:

$$\begin{cases} \hat{N}_k \Delta \hat{v}_k + \hat{\bar{D}}^k \Delta \lambda^k = -\hat{\eta}^k \\ (\hat{\bar{D}}^k)^T \Delta \hat{v}_k = -\hat{\xi}^k \end{cases} \quad (18)$$

Now the above system of equations (18) suggests the following trust region SQP subproblem in terms of the scaled step $\Delta \hat{v}_k$.

$$\begin{aligned} \min \quad & \hat{\varphi}_k(\Delta \hat{v}) = \langle \hat{\eta}^k, \Delta \hat{v} \rangle + \frac{1}{2} \langle \Delta \hat{v}, \hat{N}_k \Delta \hat{v} \rangle \\ \text{s.t.} \quad & (\hat{\bar{D}}^k)^T \Delta \hat{v} + \hat{\xi}^k = 0 \\ & \|\Delta \hat{v}\|_p \leq \Delta_k \\ & u_k + Q_k^{\frac{1}{2}} \Delta \hat{u} \in \mathcal{U}_{ad} \end{aligned} \quad (19)$$

where $\Delta \hat{v} = (\Delta \hat{u}, \Delta \vartheta) \in U \times R^{m-m_1}$, and we define the norm of the space $L^2(X) \times R^{m-m_1}$ as $\|\Delta \hat{v}\|_p = \|\Delta \hat{u}\|_p + \|\Delta \vartheta\|_2$.

Subproblem (19) is equivalent to the following problem in the original variable space.

$$\begin{aligned} \min \quad & \varphi_k(\Delta u, \Delta \vartheta) = \langle d_k, \Delta u \rangle + \frac{1}{2} \langle \Delta u, (B_k + Q_k^{-1} Q_u^k d_k) \Delta u \rangle + \langle E^k \Delta \vartheta, \Delta u \rangle \\ & + (\vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta + \frac{1}{2} \Delta \vartheta^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta \\ \text{s.t.} \quad & \langle D^k, \Delta u \rangle = -G^{\lambda^k} \\ & \|Q_k^{\frac{1}{2}} \Delta u\|_p + \|\Delta \vartheta\|_2 \leq \Delta_k, \quad u_k + \Delta u \in \mathcal{U}_{ad} \end{aligned} \quad (20)$$

Obviously the above subproblem does not actually give us a step in $\lambda \in R^{m_1}$, so we update $\lambda \in R^{m_1}$ using a projection formula. Since we know that at optimality

$$Q(\bar{u})d(\bar{u}) = 0$$

namely

$$Q(\bar{u}) \left(f(\bar{u}) + D\bar{\lambda} + \frac{1}{2} E\bar{\vartheta} \right) = 0$$

where these notations $f(\bar{u})$, D , E are consistent with the ones in above section.

Taking λ at each iteration to be the least squares solution of

$$Q(u)D\lambda = - \left(Q(u)f(u) + \frac{1}{2} Q(u)E\vartheta \right)$$

or in other words,

$$\lambda = - \left[(Q(u)D)^T (Q(u)D) \right]^{-1} (Q(u)D)^T \left(Q(u)f(u) + \frac{1}{2} Q(u)E\vartheta \right) \quad (21)$$

In the rest of this section, we will formulate the algorithm. First of all, for the update of the trust-region radius Δ_k and the acceptance of the step, we use a very common strategy. We know whether a step is accepted or not is based on the accuracy of the actual decrease in a merit functional which is relative to the predicted decrease. Now introduce the following merit functional.

$$\mathcal{L}(u, \lambda, \vartheta; \rho) = J(u) + \sum_{j=1}^{m_1} \lambda_j G_j(u) + \frac{1}{2} \sum_{j=m_1+1}^m \vartheta_j^2 G_j(u) + \rho \sum_{j=1}^{m_1} \|G_j(u)\|_p^2 \quad (22)$$

where ρ is a positive number bounded below and is updated at each iteration.

Then the actual reduction of the merit functional in moving from $(u_k, \lambda^k, \vartheta^k)$ to $(u_k + \Delta u_k, \lambda^{k+1}, \vartheta^k + \Delta \vartheta^k)$ is defined to be

$$ared_k = \mathcal{L}(u_k, \lambda^k, \vartheta^k; \rho_k) - \mathcal{L}(u_k + \Delta u_k, \lambda^{k+1}, \vartheta^k + \Delta \vartheta^k; \rho_k)$$

and this can be written as

$$\begin{aligned} ared_k &= L(u_k, \lambda^k, \vartheta^k) - L(u_k + \Delta u_k, \lambda^{k+1}, \vartheta^k + \Delta \vartheta^k) - \sum_{j=1}^{m_1} \Delta \lambda_j^k G_j(u_k + \Delta u_k) \\ &\quad + \rho_k \sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k + \Delta u_k)\|_p^2 \right) \end{aligned} \quad (23)$$

Considering the quadratic model $\tilde{\mathcal{L}}_k$ of $\mathcal{L}(u_k, \lambda^k, \vartheta^k; \rho_k) - \mathcal{L}(u_k + \Delta u_k, \lambda^{k+1}, \vartheta^k + \Delta \vartheta^k; \rho_k)$. Since

$$\begin{aligned} &\mathcal{L}(u_k, \lambda^k, \vartheta^k; \rho_k) - \mathcal{L}(u_k + \Delta u_k, \lambda^{k+1}, \vartheta^k + \Delta \vartheta^k; \rho_k) \\ &= L(u_k, \lambda^k, \vartheta^k) - L(u_k + \Delta u_k, \lambda^k, \vartheta^k + \Delta \vartheta^k) \\ &\quad + \left[L(u_k + \Delta u_k, \lambda^k, \vartheta^k + \Delta \vartheta^k) - L(u_k + \Delta u_k, \lambda^{k+1}, \vartheta^k + \Delta \vartheta^k) \right] \\ &\quad + \rho_k \sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k + \Delta u_k)\|_p^2 \right) \end{aligned}$$

then

$$\begin{aligned} \tilde{\mathcal{L}}_k &= - \left(\left\langle \nabla J(u_k) + \sum_{j=1}^{m_1} \lambda_j^k \nabla G_j(u_k) + \frac{1}{2} \sum_{j=m_1+1}^m (\vartheta_j^k)^2 \nabla G_j(u_k), \Delta u_k \right\rangle \right. \\ &\quad + \sum_{j=m_1+1}^m (\vartheta_j^k G_j(u_k)) \Delta \vartheta_j^k + \frac{1}{2} \left\langle \Delta u_k, (\nabla^2 J(u_k) + \sum_{j=1}^{m_1} \lambda_j^k \nabla^2 G_j(u_k)) \right. \\ &\quad + \frac{1}{2} \sum_{j=m_1+1}^m (\vartheta_j^k)^2 \nabla^2 G_j(u_k)) \Delta u_k \left. \right\rangle + \sum_{j=m_1+1}^m \langle \vartheta_j^k \nabla G_j(u_k) \Delta \vartheta_j^k, \Delta u_k \rangle \\ &\quad + \frac{1}{2} \sum_{j=m_1+1}^m \Delta \vartheta_j^k G_j(u_k) \Delta \vartheta_j^k \left. \right) \end{aligned}$$

$$\begin{aligned}
& - \sum_{j=1}^{m_1} \Delta \lambda_j^k G_j(u_k + \Delta u_k) + \rho_k \sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k + \Delta u_k)\|_p^2 \right) \\
& = - \left(\left\langle \nabla J(u_k) + D^k \lambda^k + \frac{1}{2} E^k \vartheta^k, \Delta u_k \right\rangle + (\vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta^k \right. \\
& \quad \left. + \frac{1}{2} \langle \Delta u_k, B_k \Delta u_k \rangle + \langle E^k \Delta \vartheta^k, \Delta u_k \rangle + \frac{1}{2} (\Delta \vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta^k \right) \\
& \quad - \sum_{j=1}^{m_1} \Delta \lambda_j^k (G_j(u_k) + \nabla G_j(u_k + \Delta u_k)) \\
& \quad + \rho_k \sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k + \Delta u_k)\|_p^2 \right) \\
& = - \left(\langle d_k, \Delta u_k \rangle + \frac{1}{2} \langle \Delta u_k, B_k \Delta u_k \rangle + \langle E^k \Delta \vartheta^k, \Delta u_k \rangle + (\vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta^k \right. \\
& \quad \left. + \frac{1}{2} (\Delta \vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta^k \right) - \left(G^{\lambda^k} + \langle D^k, \Delta u_k \rangle \right) \Delta \lambda^k \\
& \quad + \rho_k \sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k) + \langle \nabla G_j(u_k), \Delta u_k \rangle\|_p^2 \right)
\end{aligned}$$

Based on the above result and the subproblem (20), the predicted reduction has the following form:

$$\begin{aligned}
pred_k & = -\varphi_k(\Delta u_k, \Delta \vartheta_k) + \frac{1}{2} \langle \Delta u_k, (Q_k^{-1} d_k) \Delta u_k \rangle - (G^{\lambda^k} + \langle D^k, \Delta u_k \rangle) \Delta \lambda^k \\
& \quad + \rho_k \sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k) + \langle \nabla G_j(u_k), \Delta u_k \rangle\|_p^2 \right)
\end{aligned}$$

Then define the decrease ratio as

$$\gamma_k = \frac{ared_k}{pred_k}.$$

Now adopt the following strategy which is similar to that in [15] for the update of the trust-region radius.

Algorithm 4.1 (Update of the trust-region radius Δ_k)

Give the constants: $0 < \beta_1 < \beta_2 < \beta_3$, and $0 < \alpha_1 < 1 < \alpha_2 < \alpha_3$.

1. If $\gamma_k \leq \beta_1$, then reduce the trust-region radius, and choose $\Delta_{k+1} \in (0, \alpha \Delta_k)$.
2. If $\beta_1 < \gamma_k < \beta_2$, then accept the step $u_{k+1} = u_k + \Delta u_k$, and choose $\Delta_{k+1} \in [\alpha_1 \Delta_k, \Delta_k]$
3. If $\beta_2 \leq \gamma_k < \beta_3$, then accept the step $u_{k+1} = u_k + \Delta u_k$, and choose $\Delta_{k+1} \in [\Delta_k, \alpha_2 \Delta_k]$

4. If $\gamma \geq \beta_3$, then accept the step $u_{k+1} = u_k + \Delta u_k$, and increase the trust-region radius, choose $\Delta_{k+1} \in [\alpha_2 \Delta_k, \alpha_3 \Delta_k]$.

For updating the penalty parameter ρ_k at each iteration, we use a scheme proposed by El-Alem [16] which was employed to deal with the equality constraints in finite dimensional setting.

Algorithm 4.2 (Update of the penalty parameter)

Step 1. Initialization

Given ρ_{-1} and choose a small constant $\hat{\rho} > 0$;

- Step 2. At the current iterate u_k , after Δu_k has been chosen. Set $\rho_k = \rho_{k-1}$.
If $\text{pred}_k \geq \frac{1}{2} \rho_k \sum_{j=1}^{m_1} (\|G_j(u_k)\|_p^2 - \|G_j(u_k) + \langle \nabla G_j(u_k), \Delta u_k \rangle\|_p^2)$,
then $\rho_{k+1} = \rho_k$; else

$$\begin{aligned} \rho_{k+1} = & \frac{1}{2} \rho_k \left(\sum_{j=1}^{m_1} \left(\|G_j(u_k)\|_p^2 - \|G_j(u_k) + \langle \nabla G_j(u_k), \Delta u_k \rangle\|_p^2 \right) \right)^{-1} \\ & \times \left(\langle d_k, \Delta u_k \rangle + \frac{1}{2} \langle \Delta u_k, B_k \Delta u_k \rangle + \langle E^k \Delta \vartheta^k, \Delta u_k \rangle \right. \\ & \left. + (\vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta^k + \frac{1}{2} (\Delta \vartheta^k)^T \text{diag}(G^{\vartheta^k}) \Delta \vartheta^k \right) + \hat{\rho} \end{aligned}$$

As mentioned in [16], the initial choice of the penalty parameter ρ_{-1} is arbitrary. However, it should be chosen such that it is consistent with the scale of the problem under consideration. Assuming that the trust region subproblem (20) is solvable and all the approximations to the Hessian matrixes or operators of the functionals J and G_j are uniformly bounded. Now we present the trust region interior point algorithm for solving problem (P), and this algorithm is similar to that introduced by Dennis, El-Alem and Maciel [17] for the equality constrained optimization in the finite dimensional setting.

Algorithm 4.3 (Trust-region interior-point algorithm)

Given $u_0 \in \mathcal{U}_{ad}^0$, $\vartheta^0 \in R^{m-m_1}$, $\lambda^0 \in R^{m_1}$, $\Delta_0 > 0$, $\rho_{-1} > 0$, $B_0 \in \mathcal{L}(U, U')$, Q_0 .

Choose $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \varepsilon$ such that $0 < \beta_1 < \beta_2 < \beta_3$, $0 < \alpha_1 < 1 < \alpha_2 < \alpha_3$ and $\varepsilon > 0$;

Let $k = 0$.

- Step 1. Compute the gradients $f(u_k)$, $g_j(u_k)$, ($j = 1, 2, \dots, m$), $d(u_k)$, the scaling q_k and its associated scaling multiplication operator Q^k .

- Step 2. If the following condition is satisfied, then terminate the algorithm and let $u^* = u_k$.

$$\|Q^k d_k\|_{p'} + \sum_{j=1}^{m_1} \|G_j(u_k)\|_p + \sum_{j=m_1+1}^m \|\vartheta_j^k G_j(u_k)\|_p \leq \varepsilon \quad \left(\frac{1}{p} + \frac{1}{p'} = 1 \right)$$

- Step 3. Solve the trust-region subproblem (20) to determinate the trial step $(\Delta u_k, \Delta \vartheta^k)$
- Step 4. Calculate the value of $\lambda^k \in R^{m_1}$ based on the formula (21) and set $\Delta \lambda^k = \lambda^{k+1} - \lambda^k$;
- Step 5. Update the penalty parameter ρ_{k-1} to obtain ρ_k by using Algorithm 4.2;
- Step 6. Evaluate the step and update the trust-region radius Δ_k using Algorithm 4.1, if the step is accepted, then update B_k , set $k = k + 1$, and go to Step 1, else go to Step 3.

The algorithm presented in this paper is based on the reformulation of the first order necessary optimality condition of the problem **(P)** using the affine scaling method [4], and this algorithm is only a model algorithm which refers to infinite dimensional optimization.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Maurer, H., Zowe, J.: First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. Program.* **16**, 98–110 (1979)
2. Maurer, H.: First and second order sufficient optimality conditions in mathematical programming and optimal control. *Math. Program. Stud.* **4**, 163–177 (1981)
3. Coleman, T.F., Li, Y.: On the convergence of interior reflective Newton methods for nonlinear minimization subject to bounds. *Math. Program.* **69**, 187–224 (1994)
4. Coleman, T.F., Li, Y.: An interior point trust-region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.* **6**, 418–445 (1996)
5. Dennis, J.E., Heinkenschloss, M., Vicente, L.N.: Trust-region interior-point algorithms for a class of nonlinear programming problems. *SIAM J. Control Optim.* **36**, 1750–1794 (1998)
6. Chen, Z.W., Han, J.Y., Xu, D.C.: A non-monotone trust region method for nonlinear programming with simple bound constraints. *Appl. Math.* **43**, 63–85 (2000)
7. Kelley, C.T., Sachs, E.W.: A trust region method for parabolic boundary control problems. *SIAM J. Optim.* **9**, 1064–1081 (1999)
8. Toint, P.L.: Global convergence of a class of trust region methods for non-convex minimization in Hilbert space. *SIAM J. Numer. Anal.* **8**, 231–252 (1988)
9. Ulbrich, M., Ulbrich, S., Heinkenschloss, M.: Global convergence of trust region interior point algorithms for infinite-dimensional non-convex minimization subject to pointwise bounds. *SIAM J. Control Optim.* **37**, 731–764 (1999)
10. Wang, Y., Yuan, Y.: A trust region method for solving distributed parameter identification problems. *J. Comput. Math.* **20**, 575–582 (2002)
11. Cases, E., Tröltzsch, F.: Second order necessary and sufficient optimality conditions for optimization problems and applications to control theory. *SIAM J. Optim.* **13**, 406–431 (2002)
12. Clark, F.: A new approach to Lagrange multipliers. *Math. Oper. Res.* **1**, 165–174 (1976)
13. Ulbrich, M., Ulbrich, S.: Superlinear convergence of affine-scaling interior-point newton methods for infinite-dimensional nonlinear problems with pointwise bounds. *SIAM J. Control Optim.* **38**, 1938–1984 (2000)
14. Dennis, J.E., Vicente, L.W.: Trust-region interior-point algorithms for minimization problems with simple bounds. In: Fisher, H., Riedmüller, B., Schäffler, S. (eds.) *Applied Mathematics and Parallel Computing, Festschrift for Klaus Ritter*, pp. 97–107. Springer, Berlin (1996)
15. Vincent, L.N.: On interior-point Newton algorithms for discretized optimal control problems with state constraints. *Optim. Method. Softw.* **8**, 249–275 (1998)

16. El-Alem, M.: Global convergence without the assumption of linear independence for a trust-region algorithm for constrained optimization. *J. Optim. Theory Appl.* **87**, 563–577 (1995)
17. Dennis, J., El-Alem, M., Maciel, M.C.: A global convergence theory for general trust-region-based algorithms for equality constrained optimization. *SIAM J. Optim.* **7**, 177–207 (1997)

Chunfa Li received his master degree and Ph.D. degree in Operational Research and Control Theory from Dalian University of Technology in 1998 and 2003, respectively, and now is a postdoctoral research fellow in management engineering and science of Tianjin university, meanwhile he is an associate professor in Tianjin university of technology. His research interests cover optimization, optimal control theory and its applications.

Shiqin Xu received her bachelor degree in Information and Computing Science from Hebei Normal University in 2007. At present, she is a graduate student in Management Science and Engineering of Tianjin University of Technology. Her research interests cover inventory control theory, optimization, reverse logistics and its applications.

Xue Yang received his bachelor degree in Information Management and System from Hebei University in 2007. Now he is a graduate student in Management Science and Engineering of Tianjin University of Technology. His research interests cover optimal control, optimization theories and its applications.