Journal of Global Optimization **11:** 193–205, 1997.

A Modified SQP Method and Its Global Convergence

GUANGLU ZHOU
Institute of Operations Research, Qufu Normal University, Qufu, Shandong, 273165, P. R. China

(Received: 26 January 1995; accepted: 8 April 1996)

Abstract. The sequential quadratic programming method developed by Wilson, Han and Powell may fail if the quadratic programming subproblems become infeasible or if the associated sequence of search directions is unbounded. In [1], Han and Burke give a modification to this method wherein the QP subproblem is altered in a way which guarantees that the associated constraint region is nonempty and for which a robust convergence theory is established. In this paper, we give a modification to the QP subproblem and provide a modified SQP method. Under some conditions, we prove that the algorithm either terminates at a Kuhn–Tucker point within finite steps or generates an infinite sequence whose every cluster is a Kuhn–Tucker point. Finally, we give some numerical examples.

Key words: Nonlinear programming, SQP method, pseudo directional derivatives.

1. Introduction

We consider the following nonlinear programming problem

$$\begin{aligned} &\min f(x), \\ &\text{s.t. } \; g(x) \leqslant 0, \end{aligned} \tag{1.1}$$

where function $f : R^n \longrightarrow R^1$ and $g : R^n \longrightarrow R^m$ are all continuously differentiable. The SQP method generates a sequence x^k converging to the desired solution by means of solving the quadratic programming problem

$$\begin{aligned} \min \, & \nabla f(x)^T d + \frac{1}{2} d^T H d, \\ \text{s.t. } \; & g(x) + g'(x) d \leqslant 0, \\ & d \in R^n, \end{aligned} \tag{1.2}$$

iteratively, where $H \in R^{n \times n}$ is symmetric positive definite. The iteration then has the form

$$x^{k+1} := x^k + \lambda_k d_k,$$

where d_k solves (1.2) and λ_k is a step length chosen to reduce the value of a merit function for (1.1).

The SQP method may fail if the quadratic programming subproblems (1.2) become infeasible or if the associated sequence of search directions is unbounded. In [1], Han and Burke give a modification to this method wherein the QP subproblem (1.2) is altered in a way which guarantees that the associated constraint region is nonempty for each $x \in R^n$ and for which a reasonably robust convergence theory is established.

Our method is similar to the method of Burke and Han [1] in that it can overcome some difficulties associated with the infeasibility of the QP subproblems (1.2). In this paper, we give a modification to (1.2) and provide a modified SQP method. Under some conditions, we prove that the algorithm either terminates at a Kuhn–Tucker point within finite steps or generates an infinite sequence whose every cluster is a Kuhn–Tucker point.

In [10], Martin proved that every Kuhn–Tucker point of (1.1) is a global minimum of (1.1) if and only if Problem (1.1) is KT-invex. Therefore, if Problem (1.1) is KT-invex, then the proposed algorithm in this paper either terminates at a global minimum of (1.1) within finite steps or generates an infinite sequence whose every cluster is a global minimum of (1.1) under some conditions.

This paper is organized as follows. In Section 2, the concept of pseudo directional derivatives is given. Section 3 gives some lemmas. In Section 4, we discuss the modified QP subproblems. In Section 5 the proposed algorithm is stated. The global convergence theory for the method is presented in Section 6, and some numerical examples are given in the last section.

The notation that we employ is standard. However, a partial list of definitions is provided for the reader's convenience.

(1) $f'(x;d) := \lim_{\lambda\downarrow 0}(f(x+\lambda d)-f(x))/\lambda$

(2) $g'(x)$ is the Frechet derivative of g at x.

(3) Let $\|\cdot\|_\infty$ denote the maximum norm on R^n, i.e. $\|x\|_\infty := \max\{|x_j| : j = 1,2,\cdots,n\}$

(4) Let $M=\{1,2,\cdots,m\}$, $N=\{1,2,\cdots,n\}$, $e=(1,1,\cdots,\ 1)^T \in R^n$.

2. Continuous Approximation of Directional Derivatives

Let

$$g_0(x)=0,$$

$$\Phi(x)=\max\{g_j(x) : j\in M\cup\{0\}\}. \tag{2.1}$$

Then the directional derivatives of $\Phi(x)$ in any direction $d\in R^n$ is

$$\Phi'(x;d)=\max_{j\in I_0(x)}\{\bigtriangledown g_j(x)^T d\}, \tag{2.2}$$

where $I_0(x) = \{j : g_j(x) = \Phi(x), \quad j \in M \cup \{0\}\}$.

In general, $\Phi'(x; d)$ is not continuous. In [4], M. S. Bazaraa provides the following continuous approximation of $\Phi'(x; d)$

$$\Phi^*(x; d) = \max_{j \in I_0(x)} \{g_j(x) + \nabla g_j(x)^T d\} - \Phi(x), \tag{2.3}$$

$\Phi^*(x; d)$ are called *pseudo directional derivatives* of $\Phi(x)$ at x in the direction d. It can be proven that $\Phi^*(x; d)$ is a continuous function on $R^n \times R^n$.

LEMMA 2.1[6]. *For any $x, d \in R^n$, we have*

$$\Phi^*(x; d) \geqslant \Phi'(x; d) \tag{2.4}$$

and there exist $\delta > 0$ such that

$$\Phi^*(x; td) = \Phi'(x; td), \ \ \forall t \in [0, \delta]. \tag{2.5}$$

LEMMA 2.2[6]. *For any $x \in R^n$, $\Phi^*(x; \cdot)$ is a convex function on R^n.*

3. Some Lemmas

Let

$$\Psi(x) = \max\{g_j(x) : j \in M\}. \tag{3.1}$$

For all x, d in R^n, let $\Psi^*(x; d)$ denote the following first-order approximation to $\Psi(x + d)$:

$$\Psi^*(x; d) = \max\{g_j(x) + \nabla g_j(x)^T d : j \in M\}. \tag{3.2}$$

Let the functions $\Psi(x, \sigma)$, $\Psi^0(x, \sigma) : R^n \times R \longrightarrow R$ be defined, for all $\sigma > 0$, by

$$\Psi(x, \sigma) = \min\{\Psi^*(x; d) : \|d\|_\infty \leqslant \sigma\} \tag{3.3}$$

$$\Psi^0(x, \sigma) = \max\{\Psi(x, \sigma), 0\}\,. \tag{3.4}$$

REMARK. (3.3) is equivalent to the following linear programming, which we denote by $LP(x, \sigma)$

$$\min\{z : g_j(x) + \nabla g_j(x)^T d \leqslant z, j \in M, \|d\|_\infty \leqslant \sigma\}.$$

Let

$$\theta(x, \sigma) \ = \ \Psi(x, \sigma) - \Psi(x), \tag{3.5}$$

$$\theta^0(x, \sigma) \ = \ \Psi^0(x, \sigma) - \Psi(x). \tag{3.6}$$

Let the set F be defined by

$$F = \{x : g_j(x) \leqslant 0, \quad j \in M\} = \{x : \Psi(x) \leqslant 0\} \tag{3.7}$$

and let F^c denote the complement of F, i.e.

$$F^c = \{x : \Psi(x) > 0\} \tag{3.8}$$

DEFINITION 3.1[1]. The Mangasarian–Fromowitz constraint qualification (MFCQ) is said to be satisfied at a point $x \in R^n$, with respect to the underlying constraint system $g(x) \leqslant 0$, if there is a $z \in R^n$ such that

$$\bigtriangledown g_i(x)^T z < 0, \;\; i \in \{i : g_i(x) \geqslant 0, \;\; i \in M\}.$$

LEMMA 3.1. *For all x in F^c, if the MFCQ is satisfied at x. Then for all $\sigma > 0$, we have*

$$\theta(x, \sigma) < 0.$$

Proof. Let

$$I(x) = \{i : g_i(x) \geqslant 0, i \in M\}.$$

For all $x \in F^c$ and $\sigma > 0$, by Definition 3.1, there exists $d \in R^n$ and $\|d\|_\infty \leq \sigma$ such that

$$g_i(x) + \bigtriangledown g_i(x)^T d < g_i(x), \;\; i \in I(x),$$

$$g_i(x) + \bigtriangledown g_i(x)^T d < 0, \quad i \in M \setminus I(x).$$

So

$$\Psi^*(x; d) < \Psi(x).$$

Hence

$$\Psi(x, d) < \Psi(x)\,, \quad \text{i.e} \quad \theta(x, d) < 0. \qquad \square$$

LEMMA 3.2[3]. $\Psi(x, \sigma) : R^n \times R^+ \longrightarrow R$ *is continuous.*

COROLLARY 3.3. $\Psi^0(x, \sigma), \;\; \theta(x, \sigma)$ and $\theta^0(x, \sigma)$ *are all continuous on* $R^n \times R^+$.

LEMMA 3.4. *For all x in F^c, $\sigma > 0$, if $\theta(x, \sigma) < 0$, then $\theta^0(x, \sigma) < 0$.*

Proof. For all $x \in F^c$, we have $\Psi(x) > 0$. By (3.4), (3.5) and (3.6), we have

$$\begin{aligned}\theta^0(x,\sigma) &= \Psi^0(x,\sigma) - \Psi(x)\\ &= \max\{\Psi(x,\sigma) - \Psi(x), -\Psi(x)\}\\ &= \max\{\theta(x,\sigma), -\Psi(x)\}\\ &< 0\,.\end{aligned}$$

4. The Modified SQP Subproblems

Given $x \in R^n$ and $\sigma > 0$, we define $D(x,\sigma,\beta)$ to be the set

$$\begin{aligned}&D(x,\sigma,\beta)\\ &\quad = \{d \in R^n : g_j(x) + \bigtriangledown g_j(x)^T d \le \Psi^0(x,\sigma), \quad j \in M, \quad \|d\|_\infty \leqslant \beta\}\end{aligned}$$

where $\beta > \sigma$. If $d^* \in R^n$ is the solution to $LP(x,\sigma)$, then $d^* \in D(x,\sigma,\beta)$. So $D(x,\sigma,\beta)$ is nonempty. We now describe the modification to the subproblem (1.2). The subproblem (1.2) is simply replaced by the convex program $Q(x,H,\sigma,\beta)$

$$\begin{aligned}&\min \bigtriangledown f(x)^T d + \frac{1}{2} d^T H d,\\ &\text{s.t.}\quad g_j(x) + \bigtriangledown g_j(x)^T d \leqslant \Psi^0(x,\sigma), \qquad j \in M,\\ &\qquad\ \|d\|_\infty \leqslant \beta.\end{aligned}$$

These convex programs have the following properties.

LEMMA 4.1. *Let $x \in R^n$, $0 < \sigma < \beta$, and $H \in R^{n\times n}$ be symmetric and positive definite. If the MFCQ is satisfied at x, then*
(1) The convex program $Q(x,H,\sigma,\beta)$ has a unique solution d where d satisfies the following K-T conditions: There exist vectors $U = (u_1,u_2,\cdots,u_m)^T$, $V = (v_1,v_2,\cdots,v_n)^T$ and $L = (l_1,l_2,\cdots,l_n)^T$ such that

(a) $g_j(x) + \bigtriangledown g_j(x)^T d \leqslant \Psi^0(x,\sigma),\ j \in M,\ \|d\|_\infty \leqslant \beta,$

(b) $U \geqslant 0,\ V \geqslant 0,\ L \geqslant 0,$

(c) $\bigtriangledown f(x) + Hd + g'(x)^T U + V - L = 0,$

(d) $\sum_{j=1}^n u_j(g_j(x) + \bigtriangledown g_j(x)^T d - \Psi^0(x,\sigma)) = 0,$

$V^T(d - \beta e) = 0,\ L^T(-d - \beta e) = 0.$

(2) If $d = 0$ is the solution to $Q(x,H,\sigma,\beta)$, then x is a K-T point of (1.1).

Proof. (1) Since H is symmetric and positive definite, this follows from the elementary theory of convex programming.
(2) Suppose that $\Psi^0(x,\sigma) > 0$, we have that $x \in F^c$. By Lemma 3.1, we have $0 \notin D(x,\sigma,\beta)$, which contradicts that $d = 0$.
Hence

$$\Psi^0(x,\sigma) = 0.$$

By (1), we have that x is a K-T point of (1.1). $\square$

LEMMA 4.2. *For all* $x \in F^c$, $0 < \sigma \leqslant \beta$. *If the MFCQ is satisfied at* x, *then for all* $d \in D(x, \sigma, \beta)$, *we have*

$$\Phi^*(x; d) \leqslant \theta^0(x, \sigma) < 0.$$

Proof. For all $x \in F^c, \Phi(x) = \Psi(x)$. For all $d \in D(x, \sigma, \beta)$, we have

$$\begin{aligned} \Phi^*(x; d) &= \max_{j \in I(x)} \{g_j(x) + \bigtriangledown g_j(x)^T d\} - \Psi(x) \\ &\leqslant \Psi^0(x, \sigma) - \Psi(x) \\ &= \theta^0(x, \sigma) < 0\,, \end{aligned}$$

where $I(x) = \{j : g_j(x) = \Psi(x),\ \ j \in M\}$. $\square$

LEMMA 4.3. *For all* $x \in F,\ \ 0 < \sigma \leqslant \beta,\ \ d \in D(x, \sigma, \beta)$, *we have*

$$\Phi^*(x; d) = 0.$$

Proof. For all $x \in F$, we have that $\Psi^0(x, \sigma) = 0$ and $\Phi(x) = 0$. For all $d \in D(x, \sigma, \beta)$, we have

$$g_j(x) + \bigtriangledown g_j(x)^T d \leqslant 0, j \in M.$$

Hence

$$\Phi^*(x; d) = \max_{j \in I_0(x)} \{g_j(x) + \bigtriangledown g_j(x)^T d\} - \Psi(x) = 0$$

where $I_0(x) = \{j : g_j(x) = 0, j \in M \cup \{0\}\}$. $\square$

5. Algorithm

Now we state the basic algorithm as follows.

Algorithm A.

Initialization: Choose $x_0 \in R^n, \alpha_0 > 0, \delta > 0, 0 < \sigma_l < \sigma_r < \overline{\beta}, \sigma_0 \in [\sigma_l, \sigma_r], \beta_0 \in (\sigma_0, \overline{\beta}]$, Σ a compact set of symmetric and positive definite matrices, $H_0 \in \Sigma$.

Have $(x_i, \alpha_i, H_i, \sigma_i, \beta_i)$, obtain $(x_{i+1}, \alpha_{i+1}, H_{i+1}, \sigma_{i+1}, \beta_{i+1})$ as follows:

(1) Compute $\Psi(x_i, \sigma_i), \Psi^0(x_i, \sigma_i)$.

(2) Let d_i be the solution to the convex program $Q(x, H_i, \sigma_i, \beta_i)$. If $d_i = 0$, stop.

(3) If $\bigtriangledown f(x_i)^T d_i + \alpha_i \Phi^*(x_i; d_i) \leq -d_i^T H_i d_i$, set $\alpha_{i+1} := \alpha_i$; otherwise set

$$\alpha_{i+1} := \max\left\{\frac{\bigtriangledown f(x_i)^T d_i + d_i^T H_i d_i}{-\Phi^*(x_i; d_i)}, 2\alpha_i\right\}.$$

(4) Set $x_{i+1} := x_i + \lambda_i d_i$, where $\lambda_i \in [0, \delta]$ and satisfying

$$P_{\alpha_{i+1}}(x_{i+1}) \leqslant \min_{0 \leqslant \lambda \leqslant \delta} P_{\alpha_{i+1}}(x_i + \lambda d_i) + \epsilon_i,$$

where $\{\epsilon_i\}$ is a sequence of nonnegative numbers satisfying

$$\sum_{i=0}^{\infty} \epsilon_i < \infty.$$

(5) Choose $H_{i+1} \in \Sigma, \ \ \sigma_{i+1} \in [\sigma_l, \sigma_r], \ \ \beta_{i+1} \in (\sigma_{i+1}, \overline{\beta}]$.

REMARKS. (1) The procedure for choosing the step length in step (4) of the Algorithm A was introduced in [2].

(2) The merit function in step (4) is

$$P_\alpha(x) := f(x) + \alpha\Phi(x).$$

(3) In step (5) one is allowed to adjust the parameters σ_i and β_i iteratively. Therefore it is possible to incorporate a trust region like strategy. However, our proof theory does not allow the radius of these trust regions to either decrease to zero or become unbounded.

6. Global Convergence

In this section we establish the global convergence of Algorithm A.

LEMMA 6.1. *Let $d \in R^n$ be the solution to $Q(x, H, \sigma, \beta)$ for some $x \in R^n$ and some symmetric and positive definite matrix $H \in R^{n\times n}$. Then the directional derivative $P'_\alpha(x; d)$ satisfies the inequality*

$$\begin{aligned} P'_\alpha(x; d) &\leqslant \bigtriangledown f(x)^T d + \alpha\Phi^*(x; d) \\ &\leqslant -d^T H d - \left(\sum_{j=1}^{m} u_j\right) \theta^0(x, \sigma) + \alpha\Phi^*(x; d), \end{aligned} \tag{6.1}$$

where $U = (u_1, u_2, \cdots, u_n)^T$ is the Lagrange multiplier of $Q(x, H, \sigma, \beta)$.

Proof. If $d = 0$, the result holds, trivially. Thus, suppose that $d \neq 0$, by Lemma 2.1, we have

$$\begin{aligned} P'_\alpha(x; d) &= \bigtriangledown f(x)^T d + \alpha\Phi'(x; d) \\ &\leqslant \bigtriangledown f(x)^T d + \alpha\Phi^*(x; d). \end{aligned}$$

By Lemma 4.1, we have

$$\bigtriangledown f(x) = -[Hd + g'(x)^T U + V - L].$$

Hence,

$$\begin{aligned} P'_\alpha(x;d) &\leqslant -d^T Hd+\sum_{j=1}^m u_j(g_j(x)-\Psi^0(x,\sigma))-\beta(V+L)^T e+\alpha\Psi^*(x;d)\\ &\leqslant -d^T Hd+\left(\sum_{j=1}^m u_j\right)(\Psi(x)-\Psi^0(x,\sigma))+\alpha\Phi^*(x;d)\\ &= -d^T Hd-\left(\sum_{j=1}^m u_j\right)\theta^0(x,\sigma)+\alpha\Phi^*(x;d). \end{aligned}$$ □

REMARK 6.1. By Lemma 6.1 and observations made in the previous section, step (1), (2) and (3) of the Algorithm A assure us that

$$\begin{aligned} P'_{\alpha_{i+1}}(x;d) &\leqslant \nabla f(x_i)^T d_i+\alpha_i\Phi^*(x_i,\sigma_i)\\ &\leqslant -d_i^T H_i d_i\\ &< 0. \end{aligned}$$

It is not difficult to verify that the criteria for specifying λ_i in step (4) are consistent.

THEOREM 6.1. *Suppose that the MFCQ is satisfied at $x_0 \in R^n$. Let $\sigma_l > 0$ and set $F := \{x : g(x) \leqslant 0\}$. Then there is a neighborhood $N(x_0)$ of x_0 such that*

(1) the MFCQ is satisfied at every point in $N(x_0)$,

(2) if $x_0 \in F$, then $\Psi^0(x,\sigma) = 0$ for all $x \in N(x_0), \sigma \geqslant \sigma_l$, and

$$\frac{\theta^0(x,\sigma)}{\Phi^*(x;d)} \leqslant 1,$$

for all $x \in N(x_0)\setminus F,\ \ \sigma \geqslant \sigma_l$, where d is a solution of $Q(x,H,\sigma,\beta)$,

(3) if $x_0 \in F$, then

$$\sup\left\{\sum_{j=1}^m u_j : H\in\Sigma,\ \ x\in N(x_0),\ \ \sigma\in[\sigma_l,\sigma_r],\ \ \beta\in(\sigma,\overline{\beta}]\right\}<\infty,$$

where $\Sigma \subset R^{n\times n}$ is any compact set of symmetric positive definite matrices and $0 < \sigma_l < \sigma_r < \overline{\beta}$.

Proof. The proof of this theorem is similar to that of Theorem 5.1 in [1]. □

COROLLARY 6.1. *Let $x_0 \in R^n$ be such that $g(x_0) \leqslant 0$, and the MFCQ is satisfied at x_0. Also let $0 < \sigma_l < \sigma_r < \overline{\beta}$ and let Σ be a nonempty compact set of*

$n \times n$ symmetric positive definite matrices. There then is a neighborhood U of x_0 and a constant $K \geqslant 0$ such that

$$0 \leqslant \frac{\nabla f(x)^T d + d^T H d}{-\Phi^*(x;d)} \leqslant \frac{\left(\sum_{j=1}^m u_j\right) \theta^0(x,\sigma)}{\Phi^*(x;d)} \leq K,$$

for all $(x,\sigma,\beta,H) \in U \times \Gamma(\sigma_l,\sigma_r,\overline{\beta}) \times \Sigma$ where $\Gamma(\sigma_l,\sigma_r,\overline{\beta}) := \{(\sigma,\beta) : \sigma \in [\sigma_l,\sigma_r], \beta \in (\sigma,\overline{\beta}]\}$.

LEMMA 6.2. *Let $x_k \longrightarrow \overline{x}$, $H_k \longrightarrow \overline{H}$, $\sigma_k \longrightarrow \overline{\sigma}$, $\beta_k \longrightarrow \overline{\beta}$. Then $d_k \longrightarrow \overline{d}$, where d_k is a solution to $Q(x_k,H_k,\sigma_k,\beta_k)$ and $\overline{d}$ is a solution to $Q(\overline{x},\overline{H},\overline{\sigma},\overline{\beta})$.*

Proof. Suppose that $\{d_k\}$ does not converge to $\overline{d}$, then there exists a subsequence $\{d_s\} \subset \{d_t\}$ that converges to $d' \neq \overline{d}$. By Corollary 3.3, we have

$$\Psi^0(x_s,\sigma_s) \longrightarrow \Psi^0(\overline{x},\overline{\sigma}) \quad (s \longrightarrow \infty).$$

For all $d \in D(\overline{x},\overline{\sigma},\overline{\beta})$, there exists $d_m \in D(x_s,\sigma_s,\beta_s)$ such that

$$d_m \longrightarrow d \qquad (m \longrightarrow \infty).$$

Since d_s is a solution to $Q(x_s,H_s,\sigma_s,\beta_s)$, we have

$$\nabla f(x_s)^T d_s + \frac{1}{2} d_s^T H_s d_s \leq \nabla f(x_s)^T d_m + \frac{1}{2} d_m^T H_s d_m.$$

Let $s \longrightarrow +\infty$, $m \longrightarrow +\infty$, we have

$$\nabla f(\overline{x})^T d' + \frac{1}{2} d'^T \overline{H} d' \leqslant \nabla f(\overline{x})^T d + \frac{1}{2} d^T \overline{H} d.$$

This contradicts that $\overline{d}$ is the single solution to $Q(\overline{x},\overline{H},\overline{\sigma},\overline{\beta})$.

THEOREM 6.2. *Assume that the the MFCQ is satisfied. Then, any sequence x_k generated from the algorithm A either terminates at a Kuhn–Tucker point of (1.1) or any accumulation point is a Kuhn–Tucker point of (1.1).*

Proof. If the sequence $\{x_k\}$ terminates at $\overline{x}$ finitely, by Lemma 4.1 , $\overline{x}$ is a K-T point of (1.1). Thus we assume that $\{x_k\}$ is an infinite sequence. Let $\overline{x}$ be a cluster point of $\{x_k\}$. There is no loss of generality in assuming $x_k \longrightarrow \overline{x}$, $H_k \longrightarrow \overline{H}$, $\sigma_k \longrightarrow \overline{\sigma}$, $\beta_k \longrightarrow \overline{\beta}$. By Corollary 6.1, there is a constant $\alpha \geqslant 0$ such that

$$\alpha_k \leqslant \alpha \qquad k = 1, 2, \cdots.$$

There is no loss of generality in assuming $\alpha_k = \alpha$ for all $k = 1, 2, \cdots$. Let d_k be a solution to $Q(x_k,H_k,\sigma_k,\beta_k)$ and $\overline{d}$ a solution to $Q(\overline{x},\overline{H},\overline{\sigma},\overline{\beta})$, by Lemma 6.2, we have $d_k \longrightarrow d$. If $\overline{d} = 0$, then $\overline{x}$ is a K-T point of (1.1) by Lemma 6.2. Suppose that

$$\overline{d} \neq 0.$$

Let $\overline{\lambda} \in [0, \delta]$ be chosen such that

$$P_\alpha(\overline{x} + \overline{\lambda}\, \overline{d}) = \min_{0 \le \lambda \leqslant \delta} P_\alpha(\overline{x} + \lambda \overline{d}).$$

By Remark 6.1, we have

$$P_\alpha(\overline{x} + \overline{\lambda}\, \overline{d}) < P_\alpha(\overline{x}).$$

Set

$$\beta = P_\alpha(\overline{x}) - P_\alpha(\overline{x} + \overline{\lambda}\, \overline{d}).$$

Since

$$x_k + \overline{\lambda} d_k \longrightarrow \overline{x} + \overline{\lambda}\, \overline{d},$$

it follows that, for sufficiently large k, we have

$$P_\alpha(x_k + \overline{\lambda} d_k) + \beta/2 < P_\alpha(\overline{x}). \tag{6.2}$$

However, by

$$P_\alpha(x_{k+1}) < P_\alpha(x_k) + \epsilon_k, \quad \sum_{i=k}^{\infty} \epsilon_i < \beta/2,$$

for sufficiently large k we have

$$\begin{aligned} P_\alpha(\overline{x}) &< P_\alpha(x_{k+1}) + \sum_{i=k+1}^{\infty} \epsilon_k \\ &\leqslant \min_{0 \leqslant \lambda \leqslant \beta} P_\alpha(x_k + \lambda d_k) + \epsilon_k + \sum_{i=k+1}^{\infty} \epsilon_i \\ &< P_\alpha(x_k + \overline{\lambda} d_k) + \beta/2, \end{aligned}$$

which contradicts (6.2). Hence,

$$\overline{d} = 0,$$

and $\overline{x}$ is a Kuhn–Tucker point of (1.1). □

DEFINITION 6.1[10]. Problem (1.1) is said to be KT-invex, if there exists a function $h : R^n \times R^n \to R^n$ such that $x, u \in F$, then

(1) $f(x) - f(u) - \bigtriangledown f(u)^T h(x, u) \geqslant 0$,

(2) if $g_i(u) = 0, i = 1, 2, \cdots, m$, then $-\bigtriangledown g_i(u)^T h(x, u) \geqslant 0$,

where F is the feasible set of Problem (1.1).

In [10], Martin proved that every Kuhn–Tucker point of Problem (1.1) is a global minimum of (1.1) if and only if Problem (1.1) is KT-invex. Therefore we have:

COROLLARY 6.2. *Assume that the MFCQ is satisfied. If Problem (1.1) is KT-invex, then any sequence $\{x_k\}$ generated from the Algorithm A either terminates at a global minimum of (1.1) or any accumulation point is a global minimum of (1.1).*

7. Some Discussions and Numerical Examples

In this section we discuss further refinements of the algorithm proposed above to accommodate practical calculations, and give some numerical examples to show the success of the proposed method.

(1) Updating of H_i is most effectively done by the quasi-Newton methods. The matrix H_i is intended to be an approximation of the Hessian of the Lagrangian

$$L(x, \lambda) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x)$$

at the point $(x, \lambda) = (x_i, \lambda_i)$. The matrix H_i is updated by the BFGS formula [9].

(2) If $\|d_i\|$ is sufficiently small, the current point x_i is considered to be a Kuhn–Tucker point of (1.1), and the algorithm stops in step 2.

(3) An equality constraint $h(x) = 0$ exists in the original problem, it is most easily handled as two corresponding inequalities $h(x) \leqslant 0$ and $h(x) \geqslant 0$, and we can apply the above algorithm.

(4) An example is given in the following in order to demonstrate situations in which the algorithm proposed in this paper succeeds while the SQP method developed by Wilson, Han and Powell can fail if the initial value of x is set to 3.

$$\begin{aligned} &\min x, \\ &\text{s.t. } x \leqslant 1, \\ &\qquad x^2 \geqslant 0. \end{aligned}$$

NUMERICAL EXAMPLES. Finally we show the behavior of Algorithm A on some typical test problems. In the experiments below, the algorithm parameters were set as follows: $\alpha_0 = 100, \delta = 1, \sigma_l = 1, \sigma_r = 2, \overline{\beta} = 3$ and $H_0 = I \in R^{n\times n}$. A C test program of Algorithm A with BFGS update was written and applied to the following problems.

EXAMPLE 1.

$$\min f(x) = x - \frac{1}{2} + \frac{1}{2}\cos^2 x,$$

Table 1. Computational results for Example 2

k	0	1	2	3	4	8
x_1	2	1.5762621	1.3411281	1.2534775	1.2399127	1.2247615
x_2	2	1.5762621	1.3411281	1.2534775	1.2399127	1.2247615
x_3	2	1.5762621	1.3411281	1.2534775	1.2399127	1.2247615
x_4	2	1.5762621	1.3411281	1.2534775	1.2399127	1.2247615
f	32	24.693	12.9402	9.874752	9.454192	9.000488

Table 2. Computational results for Example 3

k	0	1	3	5	9
x_1	2.5	2.136197	1.397261	1.260670	1.250843
x_2	1.5	0.8936609	0.8288156	0.7419351	0.7500085
x_3	0	0.3638035	1.102739	1.23933	1.249157
x_4	0	0.6063391	0.6711844	0.958065	0.7499915
f	14.0625	5.476474	3.603782	3.516288	3.515627

$$\text{s.t. } x \geqslant 0.$$

$$x^* = 0, f(x^*) = 0.$$

If we use Algorithm A, then the solution can be obtained at the 2nd iteration under initial point $x_0 = 2$.

EXAMPLE 2.

$$\min f(x) = \sum_{i=1}^{4} x_i^2,$$
$$\text{s.t. } g(x) = 6 - \sum_{i=1}^{4} x_i^2 \leqslant 0.$$
$$x^* = (1.224745, 1.224745, 1.224745, 1.224745)^T,$$
$$f(x^*) = 9.$$

EXAMPLE 3.

$$\min f(x) = \sum_{i=1}^{3} x_i^2 x_{i+1}^2 + x_1 x_4,$$
$$\text{s.t. } g_1(x) = 4 - \sum_{i=1}^{4} x_i \leqslant 0,$$

Table 3. Computational results for Example 4

k	0	1	3	5	8
x_1	0	7.905995E-02	0	0	0
x_2	0.25	0.05	0	0	0
x_3	0	0.1581139	1.017791	1.767791	2
f	0.1666667	-0.13381	-1.017791	-1.767791	-2

$$g_2(x) = 1 - \sum_{i=1}^{4}(-1)^{i+1}x_i \leqslant 0.$$

$$x^* = (1.240023, 0.753253, 1.259977, 0.746746)^T,$$
$$f(x^*) = 3.515915.$$

EXAMPLE 4.

$$\min f(x) = \frac{4}{3}(x_1^2 - x_1x_2 + x_2^2)^{\frac{3}{4}} - x_3,$$
$$\text{s.t.} x \geqslant 0, x_3 \leqslant 2.$$

$$x^* = (0, 0, 2)^T, \quad f(x^*) = -2.$$

Acknowledgment

The author has been supported in part by the Natural Science Foundation of China. The author would like to thank the editor, Prof. Wang Changyu and Prof. Shi Zhenjun for their valuable suggestions.

References

1. Burke, J.V and Han, S.P. (1989), A Robust SQP Method, *Mathematical Programming* **43**, 277–303.
2. Han, S.P. (1977), A Globally Convergent Method for Nonlinear Programming, *Journal of Optimization Theory and Applications* **22**(3), 297–309.
3. Mayne, D.Q. and Sahba, M. (1985), An Efficient Algorithm for Solving Inequalities, *Journal of Optimization Theory and Applications* **45**(3), 407–424.
4. Bazaraa, M.S. and Goode, J.J. (1982), An Algorithm for Solving Linearly Constraint Minimax Problems, *European Journal of Operational Research* **11**, 158–166.
5. Rockafeller, R.T. (1970), *Convex analysis*, Princeton University Press, Princeton, NJ.
6. Xue Guoliang (1991), Algorithm and Their Convergent Properties for Solving a Class of Linearly Constraint Minimax Problems, *Comn. on Appl. Math. and comput*. **5**(2), 11–18.
7. Han, S.P. (1976), Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems, *Mathematical Programming* **11**, 263–282.
8. Yamashita, H.(1982), A Globally Convergent Constrained Quasi-Newton Method with an Lagrangian Type Penalty Function, *Mathematical Programming* **23**, 75–86.
9. Powell, M.J.D. (1978), A Fast Algorithm for Nonlinearly constrained optimization calculations, in G.A.Watson, (ed.), *Numerical Analysis*; Springer-Verlag, Berlin, 144–157.
10. Martin, D.H. (1985), The Essence of Invexity, *Journal of Optimization Theory and Applications* **47**(1), 65-76.