Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Computational and Applied Mathematics 197 (2006) 270–281
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

A variant of SQP method for inequality constrained optimization and its global convergence[☆]

Jiangtao Mo[a,*,1], Kecun Zhang[a], Zengxin Wei[b]

[a] *School of Science, Xian Jiaotong University, Xian, Shanxi 710049, PR China*
[b] *School of Mathematics and Information Science, Guangxi University, Guangxi 530004, PR China*

Received 7 July 2005; received in revised form 1 November 2005

Abstract

In this paper, a variant of SQP method for solving inequality constrained optimization is presented. This method uses a modified QP subproblem to generate a descent direction as each iteration and can overcome the possible difficulties that the QP subproblem of the standard SQP method is inconsistency. Furthermore, the method can start with an infeasible initial point. Under mild conditions, we prove that the algorithm either terminates as KKT point within finite steps or generates an infinite sequence whose accumulation point is a KKT point or satisfies certain first-order necessary condition. Finally, preliminary numerical results are reported.

Keywords: SQP method; Inequality constrained optimization; Global convergence

1. Introduction

In this paper, we consider the nonlinear inequality constrained optimization problem

$$\begin{aligned} &\min \quad f(x) \\ &\text{s.t.} \quad c_i(x) \leqslant 0, \quad i \in I, \end{aligned} \tag{1}$$

where $I = \{1, \ldots, m\}$, $f: R^n \to R$ and $c_i: R^n \to R$, $i \in I$, are continuously differentiable functions.

The method of sequence quadratic programming (SQP) is an important method for solving problem (1). At each iteration, the standard SQP method generates a decent direction by solving the following quadratic programming subproblem

$$\begin{aligned} &\min \quad \nabla f(x_k)^{\mathrm{T}} d + \tfrac{1}{2} d^{\mathrm{T}} B_k d \\ &\text{s.t.} \quad c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d \leqslant 0, \quad i \in I. \end{aligned} \tag{2}$$

where B_k is Hessian of Lagrangian function associated with (1). With an appropriate merit function, line search procedure, Hessian approximation procedure, and (if necessary) Maratos avoidance scheme, the SQP iteration is

[☆] This work was supported by the Science Foundation of Guangxi, No. 0542043.
[*] Corresponding author.
E-mail addresses: mjt@gxu.edu.cn (J. Mo), zxwei@gxu.edu.cn (Z. Wei).
[1] Present address: School of Mathematics and Information Science, Guangxi University, Guangxi 530004, PR China.

0377-0427/$ - see front matter
doi:10.1016/j.cam.2005.11.004

well-known to be globally and locally superlinearly convergent [8]. But the SQP method may fail if the linear constraints in quadratic programming subproblems (2) is inconsistency.

Many efforts have been made to overcome the difficulties associated with the inconsistency of quadratic programming subproblem (2). For example, Powell [10] suggested to solving a modified subproblem at each iterate:

$$\begin{aligned} \min \quad & \nabla f(x_k)^{\mathrm{T}} d + \tfrac{1}{2} d^{\mathrm{T}} B_k d + \tfrac{1}{2}\delta_k (1-\mu)^2 \\ \text{s.t.} \quad & \mu_i c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d \leqslant 0, \quad i \in I. \end{aligned}$$

where

$$\mu_i = \begin{cases} 1, & c_i(x_k) < 0 \\ \mu, & c_i(x_k) \geqslant 0 \end{cases} \quad \text{and} \quad 0 \leqslant \mu \leqslant 1, \quad \delta_k > 0$$

is a penalty parameter. The computational investigation provided by Schittkowski [12,13] showed that this modification worked very well. However, this approach may not be the best one for it cannot cope with a simple example presented by Burke and Han [4] and Burke [3].

Another approach was proposed by Burke and Han [4] and Burke [3]. Their methods can converge to a point which meets a certain first-order necessary optimality condition even when problem (1) is infeasible. Liu and Yuan [9] presented a method of the same convergent property with Burke and Han's. Their method solves two subproblems, one is an unconstrained piecewise quadratic subproblem, the other is a quadratic subproblem. In [17], Zhou presented a modified SQP method. Their method solves two subproblem, one is a linear programming with bound constraint, the other is a quadratic subproblem.

Recently, Zhang and Zhang [16] proposed a robust SQP method for solving problem (1). Similar to Zhou's method, at each iteration, their method solves a linear programming and a quadratic programming subproblem and is implementable. Under certain conditions, their method is globally convergence and locally superlinearly convergence.

In this paper, we describe another implementable method that can cope with the infeasibility of QP subproblem. Specifically, given $x_k \in R^n$, a symmetric positive definite matrix B_k, we solve a QP subproblem $QP(x_k; B_k)$ with the following form:

$$\begin{aligned} \min_{d,z} \quad & z + \tfrac{1}{2} d^{\mathrm{T}} B_k d \\ \text{s.t.} \quad & \nabla f(x_k)^{\mathrm{T}} d \leqslant z, \\ & c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d \leqslant z, \quad i \in I. \end{aligned} \tag{3}$$

Note that $QP(x_k, B_k)$ is always feasible for $d = 0$ and $z = \max_{i \in I} \{c_i(x_k); 0\}$ satisfy the constraints of (3). Let (d_k, z_k) be the solution of $Q(x_k, B_k)$. If $d_k \neq 0$, then d_k is a decent direction of merit function. Under mild conditions, our algorithm is global convergent.

The QP subproblem which is similar to $QP(x_k, B_k)$ has recently been used in the constrained optimization by Birge et al. [2], Lawrence and Tits [8], Chen and Kostreva [5] and Kostreva [7]. They introduced the right-hand side constraint perturbation in (3) subproblem and used it to obtain a feasible direction. But, in this paper, our goal is to compute a descent direction even if the constraints in (2) is inconsistent.

The paper is organized as follows. Our algorithm is presented in Section 2. In Section 3, the global convergence results of the algorithm are proved. Some preliminary numerical results are reported in Section 4. Finally, the conclusions are given in Section 5.

2. The algorithm

In this section we define our SQP method for inequality constrained optimization. In our approach, the algorithm can start at any point $x \in R^n$.

In order to obtain the global convergence, we employ the penalty function associated with (1) as a merit function, i.e.,

$$\Psi_\sigma(x) = f(x) + \sigma \Phi(x),$$

where $\sigma > 0$ is the penalty parameter and

$$\Phi(x) = \max_{i\in I} \{c_i(x), 0\}. \tag{4}$$

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

$$\Phi'(x; d) = \max_{i\in I_0(x)} \{\nabla c_i(x)^{\mathrm{T}} d\},$$

where $I_0(x) = \{i \in I : c_i(x) = \Phi(x)\}$.

In general, $\Phi'(x; d)$ is not continuous. In [1], Bazaraa used the following continuous approximation of $\Phi'(x; d)$:

$$\Phi^*(x; d) = \max_{i\in I_0(x)} \{c_i(x) + \nabla c_i(x)^{\mathrm{T}} d, 0\} - \Phi(x).$$

Then the approximation directional derivatives of $\Psi_\sigma(x)$ is

$$\theta_\sigma(x; d) = \nabla f(x)^{\mathrm{T}} d + \sigma\Phi^*(x; d).$$

Unlike Zhou's [17] and Zhang and Zhang's [16] method, our algorithm solves only one quadratic programming subproblem $QP(x_k, B_k)$ at each iteration. Let (d_k, z_k) be the solution of (3). If $d_k = 0$, we have $z_k = 0$ and x_k is a KKT point of (1) (see Lemma 2 below). If $d_k \neq 0$, d_k is a descent direction of $\Psi(x_k, \sigma_{k+1})$ for sufficiently large σ_{k+1}.

The object of updating the penalty parameter σ_k is to force d_k to be a descent direction of merit function $\Psi_{\sigma_{k+1}}(x)$ at x_k. Thus, at the kth iteration we let σ_k unchanged if d_k is a descent direction; Otherwise, σ_k is increased in the following way:

$$\sigma_{k+1} = \max\left\{\frac{\nabla f(x_k)^{\mathrm{T}} d_k + d_k^{\mathrm{T}} B_k d_k}{-\Phi^*(x_k, d_k)}, 2\sigma_k\right\}, \tag{5}$$

Now we can state our algorithm as follows.

Algorithm 1 (*A SQP algorithm for inequality constrained optimization*). *Step* 0: Given initial point $x_0 \in R^n$, a symmetric positive definite matrix $B_0 \in R^{n\times n}$, some scalars $\omega \in (0, 1)$, $\beta \in (0, 1/2)$ and initial penalty parameter $\sigma_0 > 0$. Set $k := 0$.

Step 1: Compute (d_k, z_k) by solving subproblem (3). If $d_k = 0$ then stop;

Step 2: If $\theta_{\sigma_k}(x_k, d_k) \leqslant -d_k^{\mathrm{T}} B_k d_k$, let $\sigma_{k+1} = \sigma_k$; Otherwise, σ_{k+1} is updated by (5).

Step 3: Compute t_k the first number t in the sequence $\{1, \omega, \omega^2, \ldots\}$ satisfying

$$\Psi_{\sigma_{k+1}}(x_k + td_k) - \Psi_{\sigma_{k+1}}(x_k) \leqslant t\beta\theta_{\sigma_{k+1}}(x_k; d_k).$$

Set $x_{k+1} = x_k + t_k d_k$;

Step 4: Generate B_{k+1}. Set $k := k + 1$ and go back to Step 1.

3. Global convergence

In this section we establish the global convergence of Algorithm 1. We make a few assumptions that will be in force throughout.

Assumption 1. The sequences $\{x_k\}$ and $\{(d_k, z_k)\}$ are uniformly bounded.

Assumption 2. The functions f, c_i, $i \in I$ are twice continuously differentiable.

Assumption 3. For all $x \in R^n$, the set of vectors $\{\nabla c_i(x) : i \in I_0(x)\}$ is linearly independent, where $I_0(x) = \{i : c_i(x) = \Phi(x)\}$.

Assumption 4. There exists constants $0 < M_1 \leqslant M_2$ such that

$$M_1\|d\|^2 \leqslant d^{\mathrm{T}} B_k d \leqslant M_2\|d\|^2, \quad \forall d \in R^n \text{ and } k = 1, 2, \ldots .$$

First, we give some definition of stationary points as follows.

Definition 1. A point $x \in R^n$ is called

(1) a strong stationary point of (1) if x is feasible and there exists scalars λ_i, $i \in I$, satisfying

$$\nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0,$$
$$\lambda_i c_i(x) = 0, \quad \lambda_i \geqslant 0, \quad i \in I,$$

(2) a weak stationary point of (1) if x is feasible and there exists an infeasible sequence x_k converging to x such that

$$\lim_{k\to\infty} \frac{\max_{d\in D(x_k)} \max_{i\in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d; 0\}}{\Phi(x_k)} = 1, \tag{6}$$

where $D(x_k) = \{d : \nabla f(x_k)^{\mathrm{T}} d \leqslant 0\}$.

It should be noted that there are some difference between our definition and that of Yuan [15], Liu and Yuan [9], Zhang and Zhang [16]. A strong stationary point defined above is precisely a Krush–Kuhn–Tucker (KKT) point of (1). As for weak stationary point, we can prove the following lemma.

Lemma 1. *If $x \in R^n$ is a weak stationary point, then there exists $\lambda_0 \in R$ and $\lambda \in R^m$ such that*

$$\lambda_0 \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0, \tag{7}$$

$$\lambda_i \geqslant 0, \quad i \in I, \tag{8}$$

holds.

Proof. Suppose that $d(x)$ minimizes the constrained problem $P(x)$:

$$\min_{d\in R^n} \quad \tfrac{1}{2} d^{\mathrm{T}} B d - \max_{i\in I} \{c_i(x) + \nabla c_i(x)^{\mathrm{T}} d; 0\}$$
$$\text{s.t.} \quad \nabla f(x)^{\mathrm{T}} d \leqslant 0.$$

at the iteration point x, where B is any positive definite matrix. Then, the first-order optimality condition at x gives that

$$Bd - \nabla c(x)\mu(x) + v(x)\nabla f(x) = 0, \tag{9}$$

$$v(x)\nabla f(x)^{\mathrm{T}} d(x) = 0, \quad v(x) \geqslant 0, \tag{10}$$

$$\mu(x) \in \partial u(x). \tag{11}$$

where $u(x) = \max_{i\in I} \{c_i(x) + \nabla c_i(x)^{\mathrm{T}} d; 0\}$ and $\nabla c(x) = (\nabla c_1(x), \ldots, \nabla c_m(x))$. It follows from (11) that $\mu_i(x) \geqslant 0$ for all $i \in I$.

Now suppose that x is a weak stationary point, $\{x_k\}_{k\in K}$ is a subsequence converging to x. Suppose that $d(x_k)$ is a solution of problem $P(x_k)$, then (9)–(11) holds at x_k and

$$\max_{d\in D(x_k)} \max_{i\in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d(x_k); 0\} - \Phi(x_k) \geqslant \tfrac{1}{2} d(x_k)^{\mathrm{T}} B d(x_k) \geqslant 0,$$

where $D(x_k)=\{d:\nabla f(x_k)^{\mathrm{T}}d\leqslant 0\}$. This, together with (6), implies that

$$\lim_{k\in K,k\to\infty}\frac{d(x_k)^{\mathrm{T}}Bd(x_k)}{\Phi(x_k)}=0.$$

Thus,

$$\lim_{k\in K,k\to\infty}\|d(x_k)\|=0.$$

Since $\{(\mu(x_k),v(x_k))\}$ is bounded, there is a cluster (μ^*,v^*) of $\{(\mu(x_k),v(x_k))\}_{k\in K}$ such that $\mu_i^*\geqslant 0$ for $i\in I$ and $v^*\geqslant 0$. Taking limit for $k\to\infty$ and $k\in K$ in (9), we have

$$\nabla c(x)\mu^*-v^*\nabla f(x)=0.$$

We see that (7) and (8) hold with $\lambda_0=-v^*$ and $\lambda_i=\mu_i^*$ for $i\in I$. This completes our proof. □

Next we prove some important properties of subproblem (3).

Lemma 2. *Suppose that $x\in R^n$, B is positive definite matrix and (d,z) is the solution of $QP(x,B)$.*

(1) *The following inequality holds*:

$$z\leqslant\Phi(x)-\tfrac{1}{2}d^{\mathrm{T}}Bd. \tag{12}$$

(2) *If $d=0$, then $z=0$ and x is a strong stationary point of problem* (1).

Proof. (1) Since $\hat{d}=0$ and $\hat{z}=\max_{i\in I}\{c_i(x);0\}$ is a feasible point of $QP(x,B)$, from the optimality of (d,z), we have

$$z+\tfrac{1}{2}d^{\mathrm{T}}Bd\leqslant\hat{z}=\max_{i\in I}\{c_i(x),0),$$

which, together with (4), implies that (12) holds.

(2) Since (d,z) is the solution of $QP(x,B)$, there exists $v\in R$ and $\mu\in R^m$ such that

$$v\nabla f(x)+\sum_{i=1}^{m}\mu_i\nabla c_i(x)=0, \tag{13}$$

$$1-v-\sum_{i=1}^{m}\mu_i=0, \tag{14}$$

$$vz=0,\quad v\geqslant 0, \tag{15}$$

$$\mu_i[c_i(x)-z]=0,\quad \mu_i\geqslant 0,\ \ i\in I, \tag{16}$$

$$0\leqslant z,\quad c_i(x)\leqslant z,\ \ i\in I. \tag{17}$$

By the definition of $\Phi(x)$ and (17), $\Phi(x)\leqslant z$. Then, from (12) and $d=0$, $\Phi(x)=z$. From (16), it follows that

$$\mu_i=0,\quad \forall i\notin I_0(x)=\{i\in I:c_i(x)=\Phi(x)\}.$$

Hence, it follows from Assumption 3, (13) and (14) that

$$v>0.$$

This, together with (15), implies that $z=0$. Then, from (16) and (17),

$$\mu_ic_i(x)=0,\quad \mu_i\geqslant 0,\ \ i\in I,$$
$$c_i(x)\leqslant 0,\quad i\in I.$$

Therefore, by (13), x is a strong stationary point of problem (1) with $\lambda_j = \mu_i/v$, $i \in I$.

The next two lemmas establish that the line search in Step 3 of Algorithm 1 is well defined.

Lemma 3 (*Xue Guoliang [14]*). (1) *For any* $x, d \in R^n$, *we have*

$$\Phi^*(x; d) \geqslant \Phi'(x; d)$$

and there exist $\delta > 0$ *such that*

$$\Phi^*(x; td) = \Phi'(x; td) \quad \forall t \in [0, \delta];$$

(2) *For any* $x \in R^n$, $\Phi^*(x; .)$ *is a convex function on* R^n.

Lemma 4. *Suppose that* $d_k \neq 0$. *Then the line search in Step* 3 *of Algorithm* 1 *is well defined*, *i.e.*, *Step* 3 *yields a step* $t_k = \omega^j$ *for some finite* $j = j(k)$.

Proof. It follows from Step 2 of Algorithm 1 and Assumption 4 that

$$\theta_{\sigma_{k+1}}(x_k; d_k) \leqslant -d_k^{\mathrm{T}} B_k d_k \leqslant -M_1 \|d_k\|^2 < 0. \tag{18}$$

Now we prove that the line search in Step 3 of Algorithm 1 is well defined. If it is not true, the following inequality

$$\frac{\Psi_{\sigma_{k+1}}(x_k + \omega^j d_k) - \Psi_{\sigma_{k+1}}(x_k)}{\omega^j} > \beta\theta_{\sigma_{k+1}}(x_k; d_k)$$

holds for some k and $j = 1, 2, \ldots$. Taking limit for $j \to \infty$, we have

$$\nabla f(x_k)^{\mathrm{T}} d_k + \sigma_{k+1}\Phi'(x_k; d_k) \geqslant \beta\theta_{\sigma_{k+1}}(x_k; d_k).$$

Then it follows from Lemma 3(1) and the definition of $\theta_{\sigma_{k+1}}(x_k; d_k)$ that

$$(1-\beta)\theta_{\sigma_{k+1}}(x_k; d_k) \geqslant 0,$$

which contradicts (18) for $\beta \in (0, \frac{1}{2})$. This completes our proof. □

The previous lemmas imply that Algorithm 1 is well defined. In fact, if Algorithm 1 generates a finite sequence $x_1, \ldots, x_t$, then by Lemma 2, x_t is a strong stationary. Now we suppose that the algorithm never satisfies the termination condition, i.e., the algorithm generates a infinite sequence $\{x_k\}$.

Lemma 5. *Let* (d_k, z_k) *be the solution of problem* $Q(x_k; B_k)$. *If*

$$x_k \to \hat{x} \quad \text{and} \quad B_k \to \hat{B} \quad \text{as } k \to \infty,$$

then $\{(d_k, z_k)\}$ *converges to* $(\hat{d}, \hat{z})$, *where* $(\hat{d}, \hat{z})$ *is the unique solution of problem* $Q(\hat{x}; \hat{B})$.

Proof. Let $\{(d_k, z_k)\}_{k\in K}$ be any subsequence converging to (d_0, z_0). Since (d_k, z_k) is the solution of problem $Q(x_k, B_k)$, from the first-order optimization condition, there exist $v_k \in R^1$ and $\mu_k \in R^m$ such that

$$B_k d_k + v_k \nabla f(x_k) + \sum_{i=1}^{m} (\mu_k)_i \nabla c_i(x_k) = 0, \tag{19}$$

$$1 - v_k - \sum_{i=1}^{m} (\mu_k)_i = 0, \tag{20}$$

$$v_k[\nabla f(x_k)^{\mathrm{T}} d_k - z_k] = 0, \quad v_k \geqslant 0, \tag{21}$$

$$(\mu_k)_i[c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d_k - z_k] = 0, \quad (\mu_k)_i \geqslant 0, \;\; i \in I, \tag{22}$$

$$\nabla f(x_k)^{\mathrm{T}} d_k \leqslant z_k, \quad c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d_k \leqslant z_k, \quad i \in I. \tag{23}$$

Note that (20) implies that $\{v_k\}$ and $\{\mu_k\}$ are bound. Without loss of generality, we may assume that

$$v_k \to v_0 \geqslant 0, \quad \mu_k \to \mu_0 \geqslant 0 \quad \text{as } k \to \infty \text{ and } k \in K$$

In (19)–(23), taking limit for $k \to \infty$ and $k \in K$, we have

$$\hat{B} d_0 + v_0 \nabla f(\hat{x}) + \sum_{i=1}^{m} (\mu_0)_i \nabla c_i(\hat{x}) = 0,$$

$$1 - v_0 - \sum_{i=1}^{m} (\mu_0)_i = 0,$$

$$v_0[\nabla f(\hat{x})^{\mathrm{T}} d_0 - z_0] = 0, \quad v_0 \geqslant 0,$$

$$(\mu_0)_i[c_i(\hat{x}) + \nabla c_i(\hat{x})^{\mathrm{T}} d_0 - z_0] = 0, \quad (\mu_0)_i \geqslant 0, \;\; i \in I,$$

$$\nabla f(\hat{x})^{\mathrm{T}} d_0 \leqslant z_0, \quad c_i(\hat{x}) + \nabla c_i(\hat{x})^{\mathrm{T}} d_0 \leqslant z_0, \quad i \in I,$$

which implies that (d_0, z_0) is a KKT point of $Q(\hat{x}, \hat{B})$. From Assumption 4, $\hat{B}$ is positive definite matrix. Then it follows from Lemma 2 that $(d_0, z_0) = (\hat{d}, \hat{z})$.

Lemma 6. *Let d_k be solution of $QP(x_k, B_k)$. If K is an infinite index set such that $\{d_k\}_{k\in K}$ converges to zero, then all accumulation points of $\{x_k\}_{k\in K}$ are strong stationary point of* (1).

Proof. By Assumption 1, there is an infinite index set $K' \subseteq K$ such that $\{x_k\}_{k\in K'}$ converges to $\hat{x}$.

Let (d_k, z_k) be the solution of $QP(x_k, B_k)$. From Assumptions 1 and 4, without loss of generality, we assume that

$$B_k \to \hat{B}, \quad d_k \to \hat{d}, \quad z_k \to \hat{z}, \quad \text{as } k \to \infty \text{ and } k \in K'.$$

where $\hat{B}$ is positive definite matrix. It follows from Lemma 5 that $(\hat{d}, \hat{z})$ is a solution of $QP(\hat{x}, \hat{B})$. But, by the hypothesis of the lemma, $\hat{d} = 0$. This, together with Lemma 2, implies that $\hat{x}$ is a strong stationary point of (1).

Lemma 7. *If $\sigma_k \to \infty$, then*

$$\lim_{k\to\infty} \Phi(x_k) = 0.$$

Proof. It follows from Lemma 4.2 of [15] that $\lim_{k\to\infty} \Phi(x_k)$ exists, if $\sigma_k \to \infty$. By the definition of $\Phi(x)$, we have $\lim_{k\to\infty} \Phi(x_k) \geqslant 0$.

Suppose $\lim_{k\to\infty} \Phi(x_k) > 0$. Then there exists a constant $c_1 > 0$ such that for k large enough

$$\Phi(x_k) > c_1. \tag{24}$$

We consider two cases separately.

Case 1: There exists a positive constant $c_2 > 0$ such that, for k large enough,

$$\|d_k\| > c_2. \tag{25}$$

By the definition of Φ^* and $I_0(x_k) \subseteq I$,

$$\begin{aligned}\Phi^*(x_k; d_k) &\leqslant \max_{i\in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d_k, 0\} - \Phi(x_k)\\ &\leqslant \max\left\{-\tfrac{1}{2} d_k^{\mathrm{T}} B_k d_k, -\Phi(x_k)\right\}.\end{aligned}$$

The last inequality follows from (12). Hence, from Assumption 4, (24) and (25), the following inequality holds for k large enough,

$$\Phi^*(x_k; d_k) \leqslant -c_3,$$

where $c_3 = \min\{\frac{1}{2}M_1 c_2^2, c_1\} > 0$. The above inequality, $\sigma_k \to \infty$, (12), Assumptions 1, 2 and 4 imply that the following inequality

$$\begin{aligned}\theta_{\sigma_k}(x_k, d_k) + d_k^{\mathrm{T}} B_k d_k &\leqslant z_k + d_k^{\mathrm{T}} B_k d_k + \sigma_k \Phi^*(x_k; d_k)\\ &\leqslant \Phi(x_k) + \tfrac{1}{2} d_k^{\mathrm{T}} B_k d_k - c_3\sigma_k < 0\end{aligned}$$

holds for k large enough. From Step 2 of Algorithm 1, we have

$$\sigma_{k+1} = \sigma_k$$

for k large enough, which contradicts the fact $\sigma_k \to \infty$.

Case 2: Case 1 does not hold. Thus there exists an infinite index set K such that $\{d_k\}_{k\in K}$ converges to zero. Since K is an infinite index set, by Assumption 1, $\{x_k\}_{k\in K}$ has at least one accumulation point. Let $K' \subseteq K$ such that $\{x_k\}_{k\in K'}$ converges to $\hat{x}$. By Lemma 6, $\hat{x}$ is a strong stationary point of (1). Hence, $\Phi(\hat{x}) = 0$, which implies that

$$\Phi(x_k) < c_1$$

holds for $k \in K'$ large enough. This contradicts (24).

Therefore, the lemma is proved. □

Lemma 8. *Suppose that $\sigma_k = \sigma > 0$ for all k large enough, $\{x_k\}$ is an infinite sequence and $\{x_k\}_{k\in K}$ is a convergent subsequence. Then $d_k \to 0$ as $k \to \infty$ and $k \in K$.*

Proof. Without loss of generality, assume that $\sigma_k = \sigma$ for all k. We proceed by contradiction. Suppose that there is an infinite subset $K' \subset K$ and a positive constant η_0 such that

$$\|d_k\| \geqslant \eta_0, \quad k \in K'. \tag{26}$$

It follows from the definition of $\Phi^*(x; d)$ that, for any $t \in (0, 1]$,

$$\Phi^*(x_k, td_k) - \Phi(x_k) \leqslant t(\Phi^*(x_k, d_k) - \Phi(x_k)).$$

By Assumption 4, we have

$$\begin{aligned}&\Psi_{\sigma_{k+1}}(x_k + td_k) - \Psi_{\sigma_{k+1}}(x_k) - t\beta\theta_{\sigma_{k+1}}(x_k; d_k)\\ &\quad\leqslant t(1-\beta)\theta_{\sigma_{k+1}}(x_k; d_k) + t^2\eta_1\|d_k\|^2,\end{aligned} \tag{27}$$

where $\eta_1 > 0$ is a constant. From Step 2 of Algorithm 1 and (26), there exists a $\eta_1 > 0$ such that

$$\theta_{\sigma_{k+1}}(x_k; d_k) \leqslant -\eta_1 < 0 \quad \text{for } k \in K'.$$

This, together with (27), implies that, for all $k \in K'$, there exists a $t_0 \in (0, 1)$ such that

$$\Psi_{\sigma_{k+1}}(x_k + td_k) - \Psi_{\sigma_{k+1}}(x_k) \leqslant t\beta\theta_{\sigma_{k+1}}(x_k; d_k) \quad \text{for } t \in [0, t_0],$$

By Step 3 of Algorithm 1, we have $t_k \geqslant t_0$ for all $k \in K'$. Thus,

$$\Psi_{\sigma_{k+1}}(x_k + t_k d_k) - \Psi_{\sigma_{k+1}}(x_k) \leqslant -t_0\beta\eta_1 < 0 \quad \text{for } k \in K'.$$

This implies that

$$\Psi_{\sigma_{k+1}}(x_k + t_k d_k) \to -\infty \quad \text{as } k \to \infty \text{ and } k \in K',$$

which contradicts the assumption that $\{x_k\}$ is bounded. Hence, there is not K' and η_0 satisfying (26) and the lemma is proved. □

We now state and prove the main result of this section.

Theorem 1. *Let $\{x_k\}$ be an infinite sequence generated by Algorithm* 1. *Then any accumulation point of $\{x_k\}$ is either a strong stationary point or a weak stationary point of* (1).

Proof. Suppose K is an infinite index set such that $x_k \to \hat{x}$ as $k \to \infty$ and $k \in K$. Let (d_k, z_k) be the solution of $QP(x_k, B_k)$. If there exists an infinite index set $K' \subset K$ such that

$$d_k \to 0 \quad \text{as } k \to \infty \text{ and } k \in K',$$

then it follows from Lemma 6 that $\hat{x}$ is a strong stationary point of (1).

Now suppose that there exists a constant c_0 such that

$$\|d_k\| \geqslant c_0 \tag{28}$$

for $k \in K$ and k large enough. In the view of Lemma 8, it follows that

$$\sigma_k \to \infty \quad \text{as } k \to \infty.$$

We will show that $\hat{x}$ is a weak stationary point. Proceeding by contradiction, suppose that there exists a constant $\eta_1 > 0$ such that for k large enough,

$$\max_{d \in D(x_k)} \max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d; 0\} \leqslant \Phi(x_k) - \eta_1, \tag{29}$$

where $D(x_k) = \{d : \nabla f(x_k)^{\mathrm{T}} d \leqslant 0\}$.

Suppose that $\hat{d}_k$ belongs to $D(x_k)$ such that

$$\max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} \hat{d}_k; 0\} = \max_{d \in D(x_k)} \max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d; 0\}. \tag{30}$$

Since $\sigma_k \to \infty$, it follows from (12), Assumption 4, (28) and Lemma 7 that for k large enough,

$$z_k \leqslant -\eta_2,$$

where $\eta_2 > 0$ is a constant. Thus, from the first constraint of (3), we have

$$\nabla f(x_k)^{\mathrm{T}} d_k \leqslant 0,$$

i.e., $d_k \in D(x_k)$ for k large enough. Hence, from (30), we have

$$\max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d_k; 0\} \leqslant \max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} \hat{d}_k; 0\}.$$

Then Assumption 1, Lemma 7 and $\sigma_k \to \infty$ imply that inequality

$$\begin{aligned}
\theta_{\sigma_k}(x_k, d_k) + d_k^{\mathrm{T}} B_k d_k &\leqslant z_k + d_k^{\mathrm{T}} B_k d_k + \sigma_k \left(\max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} d_k; 0\} - \Phi(x_k) \right) \\
&\leqslant \Phi(x_k) + \tfrac{1}{2} d_k^{\mathrm{T}} B_k d_k \\
&\quad + \sigma_k \left(\max_{i \in I} \{c_i(x_k) + \nabla c_i(x_k)^{\mathrm{T}} \hat{d}_k; 0\} - \Phi(x_k) \right) \\
&\leqslant \Phi(x_k) + \tfrac{1}{2} d_k^{\mathrm{T}} B_k d_k - \eta \sigma_k < 0,
\end{aligned}$$

holds for k large enough, which contradicts the parameter updating procedure in Algorithm 1. Hence, it follows that $\hat{x}$ is weak stationary point of (1).

4. Numerical results

To show the behavior of Algorithm 1, numerical tests are performed on some small size problems previously used in related literatures.

A Matlab subroutine is programmed and used to solve the problems. The standard internal function QP in Optimization Toolbox is used to solve subproblem (3) in our algorithm.

The first problem was considered in [11,9] to illustrate the fact that the algorithm in [11] terminated at an approximate Kuhn–Tucker point which was not the approximate minimum point. The second one is from Zhou [17].

Example 1 (*Sahba [11], Liu and Yuan [9], Zhang and Zhang [16]*).

$$\begin{aligned} \min \quad & f(x) = x_1 x_2 \\ \text{s.t.} \quad & c(x) \leqslant 0, \end{aligned}$$

where $c(x) \in R^5$ and $c_1(x) = \sin x_1$, $c_2(x) = -\cos x_1$, $c_3(x) = x_1^2 + x_2^2 - \pi/2$, $c_4(x) = -x_1 - \pi$, $c_5(x) = -x_2 - \pi/2$.

On this problem, staring at the point $(0, 5)^{\mathrm{T}}$, Algorithm 1 terminates after 14 iterations at the point $(-0.8862, 0.8862)^{\mathrm{T}}$ which is the approximate minimum point. A full step of one is taken at each iteration. This result is similar to that of Liu and Yuan [9] and Zhang and Zhang [16]. But the algorithm in [11] terminated at the point $(0, -1.25331)^{\mathrm{T}}$ which is an approximate Kuhn–Tucker point and not the approximate minimum point.

Table 1
Numerical results of Algorithm 1

Prob	n, m	x_0	$\Phi(x_0)$	Ni	Nf	$f(x_k)$	$\|d_k\|$	$\Phi(x_k)$
hs003	2, 1	(10, 1)	0	36	37	2.11973459e − 010	5.29e − 007	0
		(8, −2)	2	36	37	1.43306083e − 010	6.18e − 007	0
hs012	2, 1	(0, 0)	0	15	16	−2.99999960e + 001	3.96e − 007	0
		(8, −6)	267	12	14	−3.00000032e + 001	4.46e − 007	1.93e − 005
hs031	3, 7	(1, 1, 1)	0	82	83	6.00004323e + 000	9.59e − 007	0
		(1.5, 0.5, 3)	2	117	118	6.00004520e + 000	9.83e − 007	0
hs034	3, 5	(0, 1.05, 2.9)	0	17	21	−8.34031999e − 001	6.44e − 007	0
		(3, 3, 3)	17.09	21	24	−8.34032019e − 001	6.16e − 007	0
hs035	3, 1	(0.5, 0.5, 0.5)	0	11	12	1.11111360e − 001	6.47e − 007	0
		(3, 3, 3)	9	15	16	1.11111207e − 001	2.44e − 007	0
hs065	3, 7	(4, 4, 4)	0	6	7	9.53529029e − 001	2.67e − 007	0
		(−5, 5, 0)	2	48	49	9.53528897e − 001	1.73e − 007	0
hs066	3, 8	(0, 1.05, 2.9)	0	15	16	5.18176876e − 001	9.95e − 007	0
		(1, 1, 1)	1.72	20	21	5.18162862e − 001	6.62e − 007	1.03e − 006
hs113	10, 8	(2, 3, 5, 5, 1, 2, 7, 3, 6, 10)	0	73	74	3.04820825e + 001	8.76e − 007	0
		(5, 5, 5, 5, 5, 5, 5, 5, 5, 5)	116	82	148	3.04820766e + 001	7.87e − 007	0
s215	2, 2	(1, 1)	0	20	21	9.53674316e − 007	6.74e − 007	0
		(−1, 2)	1	23	24	7.07746851e − 007	5.00e − 007	0
s225	2, 5	(1.5, 1.5)	0	54	55	2.00001291e + 000	9.13e − 007	0
		(3, 1)	2	60	61	2.00001277e + 000	9.03e − 007	0
s226	2, 4	(0.8, 0.25)	0	13	14	−4.99999922e − 001	3.27e − 007	0
		(3, −3)	17	15	16	−5.00000187e − 001	5.90e − 007	1.12e − 006
s227	2, 2	(0.5, 0.5)	0	35	36	1.00000355e + 000	8.37e − 007	0
		(2, 8)	62	26	27	1.00000332e + 000	7.84e − 007	0
s230	2, 2	(1, 1)	0	19	20	3.75001602e − 001	8.01e − 007	0
		(0, 0)	1	18	19	3.74998625e − 001	6.87e − 007	2.75e − 006

Example 2 (*Zhou[17]*).

$$\min \quad f(x) = \sum_{i=1}^{3} x_i^2 x_{i+1}^2 + x_1^2 x_4^2$$

$$\text{s.t.} \quad c_1(x) = 4 - \sum_{i=1}^{4} x_i^2 \leqslant 0,$$

$$c_2(x) = 1 - \sum_{i=1}^{4} (-1)^{i+1} x_i \leqslant 0.$$

From the initial point $(2.5, 1.5, 0, 0)^T$, Algorithm 1 terminates after 12 iterations at the point $(3.1503, 0, 4.4661, 0)^T$ and the value of objective function is 0. This result is better than Zhou's [17]. In [17], Algorithm A terminated after 9 iterations at $(1.2508, 0.7500, 1.2492, 0.7450)^T$ and the value of objective function is 3.515627.

We also test some problems selected from [6]. In the tests, we choose initial parameters $\omega = 0.5$, $\beta = 0.25$, $\sigma = 1$ and $\varepsilon = 10^{-6}$. The initial Lagrangian Hessian estimate $B_0 = I$ and B_k is updated by the damped BGFS formula [10]. The stopping condition is $\|d_k\| \leqslant 10^{-6}$.

The numerical results are reported in Table 1, where the columns have the following meanings: Prob denotes the name of test problems in [6], n and m denote the number of variables and constraints, *Ni* and *Nf* denote the number of iterations and objective function and constraints evaluations, x_0 and x_k denote the initial and termination point, $\Phi(x_0)$ and $\Phi(x_k)$ denote the value of $\Phi(x)$ defined by (4) at x_0 and x_k. Note that x is feasible if and only if $\Phi(x) = 0$. For each problem, we test Algorithm 1 with two initial points, a feasible point and an infeasible point. The preliminary numerical results are encouraging. Algorithm 1 performances stable numerical results for feasible and infeasible initial points. Hence, this algorithm is an improvement of the classical SQP method.

5. Conclusions

A SQP method that can be applied to inequality constrained optimization problems has been presented. Global convergence has been shown under mild assumptions. There are two significant differences between the presented method and previously proposed SQP methods. One is that the quadratic subproblem (3) is always consistent. The other is that the present method can start from an infeasible initial point. Preliminary numerical results indicate that the presented method is viable and efficient.

References

[1] M.S. Bazaraa, J.J. Goode, An algorithm for solving linearly constraint minimax problems, European J. Oper. Res. 11 (1982) 158–166.
[2] J. Birge, L. Qi, Z. Wei, A variant of the Topkis–Veinott method for solving inequality constrained optimization problems, Appl. Math. Optim. 41 (2000) 309–330.
[3] J.V. Burke, A sequential quadratic programming method for potentially infeasible mathematical programs, J. Math. Anal. Appl. 139 (1989) 319–351.
[4] J.V. Burke, S.P. Han, A Robust SQP Method, Math. Programming 43 (1989) 277–303.
[5] X.B. Chen, M.M. Kostreva, A generalization of the norm-relaxed method of feasible directions, Appl. Math. Comput. 102 (1999) 257–272.
[6] W. Hock, K. Schittokowski, Test examples for nonlinear programming codes, Lecture Notes in Economics and Mathematical Systems, vol. 187, Springer, Berlin, Heidelberg, New York, 1981.
[7] M.M. Kostreva, X. Chen, A superlinearly convergent method of feasible directions, Appl. Math. Comput. 116 (2000) 231–244.
[8] C. Lawrence, A. Tits, A computationally efficient feasible sequential quadratic programming algorithm, Technical Research Report, T.R. 98-46, Institute of System Research of the University of Maryland, <http://techreports.isr.umd.edu>
[9] X.W. Liu, Y. Yuan, A robust algorithm for optimization with general equality and inequality constraints, SIAM J. Sci. Comput. 22 (2) (2000) 517–534.
[10] M.J.D. Powell, A fast algorithm for nonlinearly constrained optimization calculations, in: G.A. Watson (Ed.), Proceedings of 1977 Dundee Biennial Conference on Numerical Analysis, Springer, Berlin, 1978, pp. 144–157.
[11] M. Sahba, Globally convergent algorithm for constrained optimization problems, J. Optim. Theory Appl. 52 (1987) 291–309.
[12] K. Schittkowski, The nonlinear programming method of Wilson, Han and Powell with an augmented Lagrangian type line search function, Part 1: Convergence analysis, Numer. Math. 38 (1981) 83–114.

[13] K. Schittkowski, On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function, Math. Operations sch. u. Statist., Ser. Optim. 14 (1983) 197–216.

[14] X. Guoliang, Algorithm and their convergent properties for solving a class of linearly constraint minimax problems, Comn. Appl. Math. Comput. 5 (2) (1991) 11–18.

[15] Y. Yuan, On the convergence of the new trust region algorithm, Numer. Math. 70 (1995) 515–539.

[16] J. Zhang, X. Zhang, A robust SQP method for optimization with inequality constraints, J. Comput. Math. 21 (2) (2003) 247–256.

[17] G. Zhou, A modified SQP method and its global convergence, J. Global Optim. 11 (1997) 193–205.