International Journal of Applied Mathematical Research, 3 (3) (2014) 336-347
©Science Publishing Corporation
www.sciencepubco.com/index.php/IJAMR
doi: 10.14419/ijamr.v3i3.3121
Research Paper

A new SQP algorithm and numerical experiments for nonlinear inequality constrained optimization problem

Mahdi Sohrabi-Haghighat *, Maryam Ghasemi

Department of Mathematics, Arak University, Arak, Iran
**Corresponding author Email: m-haghighat@araku.ac.ir*

Abstract

In this paper, a new algorithm based on SQP method is presented to solve the nonlinear inequality constrained optimization problem. As compared with the other existing SQP methods, per single iteration, the basic feasible descent direction is computed by solving at most two equality constrained quadratic programming. Furthermore, there is no need for any auxiliary problem to obtain the coefficients and update the parameters. Under some suitable conditions, the global and superlinear convergence are shown.

Keywords: *Global convergence, Inequality constrained optimization, Nonlinear programming problem, SQP method, Superlinear convergence rate.*

1. Introduction

In this paper, we consider the following nonlinear problem (NLP),

$$\begin{array}{ll} \min & f(x), \\ s.t. & g_j(x) \leq 0, \quad j = 1, 2, ..., m, \end{array} \tag{1}$$

where $f, g_j : \Re^n \to \Re (j = 1 \sim m)$ are continuously differentiable functions. The method of feasible directions (MFD) is widely acknowledged for solving problem (1). SQP algorithms generate iteratively the main search direction d_0 by solving the following quadratic programing subproblem:

$$\begin{array}{ll} \min & \nabla f(x)^T d + \frac{1}{2} d^T H d, \\ s.t. & g_j(x) + \nabla g_j(x)^T d \leq 0, \quad j = 1, 2, ..., m, \end{array} \tag{2}$$

where $H \in \Re^{n \times n}$ is a symmetric positive definite matrix.
Denote the feasible set of (1) by

$$X = \{x \in \Re^n | g_j(x) \leq 0, j = 1, 2, ..., m\}.$$

In [14], feasible sequential quadratic programing (FSQP) algorithms are presented to construct a feasible points sequence of problem (1). As it is pointed out in [11], feasible iterates are desirable for algorithm and its application because

- the QP subproblems (2) are always consistent, i.e., a feasible solution to (2) always exists;
- the objective function may be used directly as a merit function in the line search;
- the optimization process may be stopped after a few iterations, yielding a feasible point.

So it is very important to develop the FSQP methods, see [3,15,9]. Generally, in order to obtain the restoring feasible descent direction which assures the global convergence and the high-order direction which avoids Maratos effect, the FSQP algorithm requires solving two or three QP subproblem like (2) with inequality constraints in single iteration. In [23], a feasible descent direction is obtained only by solving one QP problem.

In addition, it is also a hot topic to solve the QP problem like (2) in the field of optimization. There exist a lot of algorithms to solve step by step a series of corresponding QP problem with only equality constraints to obtain the optimum solution to the QP subproblem (2). Obviously, it is simpler to solve the equality constrained QP problem than to solve the QP problem with inequality constraints.

In [20], another SQP algorithm is presented to solve general nonlinear programs with mixed equality and inequality constraints. Compared with most conventional SQP methods, this algorithm is merely necessary to solve QP subproblems with only equality constraints, and it is a superior numerical method due to its new computation of penalty weights and adequate efficient numerical experiments.

For the problem (1) without equality constraints, the algorithm in [20] defines an exact penalty function

$$\Phi(x;u) = f(x) + \sum_{j=1}^{m} u_j \max\{0, g_j(x)\}$$

and obtains the error d_0 in the KKT conditions by solving the following quadratic problem with only equality constraints:

$$\begin{array}{lll} \min & \nabla f(x)^T d + \frac{1}{2} d^T H d, & \\ s.t. & \nabla g_j(x)^T d = 0, & j \in A \subseteq \{1, 2, ..., m\}, \end{array} \tag{3}$$

where the so-called working set $A \subseteq \{1, 2, ..., m\}$ is suitably determined. Then, the search direction d and the multiplier estimates λ are obtained by solving another QP subproblem with only equality constraints:

$$\begin{array}{lll} \min & \nabla f(x)^T d + \frac{1}{2} d^T H d, & \\ s.t. & h_j(x) + \nabla g_j(x)^T d = 0, & j \in A, \end{array} \tag{4}$$

where $h_j(x)$ is defined according to the multiplier vector of the subproblem (3). If $d = 0$ and $\lambda \geq 0$, the algorithm stops. Otherwise, given new penalty weights $\tilde{u}$, the stepsize σ is obtained by combining backtracking with interpolation. Finally, a better point $\bar{x}$ is generated

$$\bar{x} = x + \sigma d.$$

However, unlike conventional SQP methods, from (4), it cannot guarantee that the corresponding approximating multipliers are nonnegative during iterations, that is to say, it holds that x is a KKT point of (1), only when $d = 0$ and $\lambda \geq 0$. Thereby, if $d = 0$, but $\lambda \geq 0$ does not hold, the algorithm in [20] will not implement successfully, because there is no method to remedy this bad case.

In this paper, we modify the feasible sequential equality constrained programming algorithm, presented in [24], with less complexity and computations to solve the nonlinear inequality constrained programming problem. In [24], in order to have a feasible descent direction d, of the objective function, it is necessary to solve some systems of linear equations that are defined by the corresponding multiplier vector and parameters μ_j^k. In addition, μ_j^k is updated in each iteration by the parameter $\bar{\mu}$. Furthermore, for obtaining the rate of convergence, it is necessary that the corresponding multiplier vector u^* according to KKT point x^* satisfies $u_j^* \leq \bar{\mu}$, $j = 1 \sim m$. Here, without this assumption, a new algorithm is proposed, in which, basic feasible descent direction is computed by solving at most two equality constrained quadratic programing. But, here, unlike [24], we do not compute any auxiliary problem to obtain d. Under some suitable assumptions, global convergence is obtained. Using the techniques given in [23], superlinear convergence rate is shown.

The plan of this paper is as follows: in section 2, the algorithm is proposed. In section 3, we show that the algorithm is globally convergent. The superlinear convergence rate is analyzed in section 4. In section 5, some numerical experiments are implemented. Because of similarity to paper [24], the proofs of some theorems and lemmas have been eliminated. So for reader's comfort, we preserved the structure of [24] in our paper.

2. Description of algorithm

Let $I(x)$ denote the active constraints set of (1.1):

$$I(x) = \{j \in I | g_j(x) = 0\}, \quad I = \{1, 2, ..., m\}. \tag{5}$$

Now, the algorithm for solving Problem (1.1) can be presented as follows.

Algorithm A. **Step 0.** Initialization: Given a starting point $x^0 \in X$, and an initial symmetric positive definite matrix $H_0 \in \Re^{n\times n}$. Choose parameter $\varepsilon_0,\ \nu \in (0,1),\ \alpha \in (0, \frac{1}{2}),\ \tau \in (2,3), \rho > 0$. Let $k = 0$;
Step 1. Computation of an approximate active set J_k:
Step 1.1. Let $i = 0,\ \varepsilon_{k,i} = \varepsilon_0$;
Step 1.2. Let

$$J_{k,i} = \{j \in I | -\varepsilon \le g_j(x^k) \le 0\}, \quad A_{k,i} = (\nabla g_j(x^k), j \in J_{k,i}). \tag{6}$$

If $A_{k,i}$ is not of full column rank, then set $i = i+1,\ \varepsilon_{k,i} = \frac{1}{2}\varepsilon_{k,i-1}$ and go to Step 1.2; Otherwise, let $J_k = J_{k,i},\ A_k = A_{k,i},\ i_k = i$, and go to Step 2.
Step 2. Computation of the vector a^k which is important to the criterion of the KKT point:
Step 2.1. Reorder the rows of A_k as follows:

$$A_k \triangleq \begin{pmatrix} A_k^1 \\ A_k^2 \end{pmatrix},$$

where the invertible matrix A_k^1, is the matrix whose rows are $|J_k|$ linearly independent rows of A_k, and A_k^2 is the matrix whose rows are the remaining $n - |J_k|$ rows of A_k. Correspondingly, let $\nabla f(x^k)$ be decomposed as $\nabla f_1(x^k)$ and $\nabla f_2(x^k)$, i.e,

$$\nabla f(x^k) \triangleq \begin{pmatrix} \nabla f_1(x^k) \\ \nabla f_2(x^k) \end{pmatrix}.$$

Step 2.2. Solve the following system of linear equations:

$$A_k^1 u = -\nabla f_1(x^k). \tag{7}$$

Let $a^k = (a_j^k, j \in J_k) \in \Re^{|J_k|}$ be the unique solution;
Step 3. Computation of the direction d_0^k: Solve the following equality constrained QP subproblem at x^k:

$$\begin{array}{ll} \min & \nabla f(x^k)^T d + \frac{1}{2} d^T H_k d \\ s.t. & g_j(x^k) + \nabla g_j(x^k)^T d = -\min\{0, a_j^k\} \quad j \in J_k. \end{array} \tag{8}$$

Let d_0^k be the KKT point of (8) and $b^k = (b_j^k, j \in J_k)$ be the corresponding multiplier vector. If $d_0^k = 0$, STOP. Otherwise, Let

$$\eta_k = \min\{\|d_0^k\|^\nu, \rho\},$$

continue.
Step 4. Computation of the feasible direction with descent d^k which guarantees the global convergence:
Step 4.1. Solve the following constrained QP subproblem at x^k :

$$\begin{array}{ll} \min & z + \frac{1}{2} d^T H_k d, \\ s.t. & \nabla f(x^k)^T d = z \\ & \nabla g_j(x^k)^T d = \eta_k z \quad j \in J_k. \end{array} \tag{9}$$

Let d^k be the KKT point of (9) and $\lambda^k = (\lambda_j^k, j \in J_k)$ be the corresponding multiplier vector.
If $d^k \neq 0$ go to Step 5.
Step 4.2. If for all $j \in J_k,\ \lambda_j^k \ge 0$ and the complementary condition holds, such that,

$$g_j(x^k) \cdot \lambda_j^k = 0 \quad for\ all\ j \in J_k,$$

then STOP and x^k is the KKT point of the main problem.
Step 4.3. Else if, there is some multiplier $\lambda_j^k < 0$ then set

$$j' = \arg\min\{\lambda_j^k | \lambda_j^k < 0\}.$$

and delete j' from the working set J_k and resolve the subproblem (9), it is proved that $d^k \neq 0$. In this case go to Step 5.

Step 4.4. Else remove the indices from the working set, whose corresponding constraints are not satisfying the complementarity condition and resolve the subproblem (9) with the new working set J_k.

Let d^k be the KKT point of this subproblem and $\lambda^k = (\lambda_j^k, j \in J_k)$ be the corresponding multiplier vector. If $d^k \neq 0$ then go to step 5.

If $d^k = 0$ and for all $j \in J_k$, $\lambda_j^k \geq 0$ then STOP and x^k is the KKT point of the main problem. Else go to Step 4.3.

Step 5. Computing of the high-order revised direction $\tilde{d}^k$ which avoids Maratos effect.

Step 5.1. Obtain d_1^k by solving the following $|J_k| \times |J_k|$ system of linear equations.

$$(A_k^1)^T d_1^k = -\psi_k e - g^k, \tag{10}$$

where

$$\psi_k = \max\{\|d^k\|^\tau, -\eta_k z_k \|d^k\|\}, \quad e = (1, ..., 1)^T \in \Re^{|J_k|}, \quad g^k = (g_j(x^k + d^k), j \in J_k).$$

Let d_1^k be the solution.

Step 5.2. According to the transformation of A_k, define

$$\tilde{d}^k \triangleq \begin{pmatrix} d_1^k \\ 0 \end{pmatrix},$$

so, it holds that

$$A_k^T \tilde{d}^k = (A_k^1)^T d_1^k + (A_k^2)^T 0 = (A_k^1)^T d_1^k.$$

If $\|\tilde{d}^k\| > \|d^k\|$, set $\tilde{d}^k = 0$;

Step 6. The line search: Compute t_k, the first number t in the sequence $\{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, ...\}$ is satisfying

$$f(x^k + td^k + t^2\tilde{d}^k) \leq f(x^k) + \alpha t \nabla f(x^k)^T d^k, \tag{11}$$

$$g_j(x^k + td^k + t^2\tilde{d}^k) \leq 0, \quad j \in I \tag{12}$$

Step 7. Update:

Obtain H_{k+1} by updating the positive definite matrix H_k using some quasi-Newton formula.

Set

$$x^{k+1} = x^k + td^k + t^2\tilde{d}^k.$$

Set $k = k + 1$, Go back to Step 1.

Remarks.

- Firstly, in Step 1.2, 4.3 and 4.4 if $J_k = \emptyset$, then x^k is a strict feasible interior point of main problem, in addition, A_k and a^k will have no definitions. Here, Algorithm A will be very simple, since Steps 2 and 3 will not proceed.
- Secondly, if $m \gg n$, that is to say, the number of constraints is much greater than the variable dimension, then $|J_k| = n$ (Obviously, according to H 3.3, it is impossible to have $|J_k| > n$, else the number of linear independent vectors would be greater than n in the space $\Re^n$.) Here $A_k \in \Re^{n\times n}$ is nonsingular, and we might as well denote that J_k is the set of some indices such that $A_k = (\nabla g_j(x^k), j \in J_k)$ and $|J_k| = n$. So, it is not necessary to decompose A_k and $\nabla f(x^k)$, and the high-order revised direction $\tilde{d}^k$ in step 5.2 is equal to d_1^k.

3. Global convergence of algorithm

In this section, at first, we show that Algorithm A given in section 2 is well-defined. To do this, we make the following general assumptions and let them hold throughout the paper:

H 3.1. The feasible set is nonempty, i.e, $X \neq \emptyset$;

H 3.2. The functions $f(x)$, $g_j(x)(j = 1 \sim m)$ are two-times continuously differentiable;

H 3.3. $\forall x \in X$, the vectors $\{\nabla g_j(x), j \in I(x)\}$ are linearly independent.

H 3.4. $\{x^k\}$ is bounded, which is the sequence generated by the algorithm.

Lemma 3.1 *For any iteration, there is no infinite cycle in Step 1. Moreover, if* $\{x^k\}_{k\in K} \to x^*$ *then there exists a constant* $\bar{\varepsilon} > 0$*, such that* $\varepsilon_{k,i_k} \geq \bar{\varepsilon}$*, for* $k \in K$*,* k *large enough.*

Proof. The proof of this lemma is similar to the proof of Lemma 3.1 in [24].

Lemma 3.2 (z_k, d^k) *is the unique solution of the QP subproblem (9) at* x^k*, and* $\{d_0^k, d^k, z_k, a^k, b^k\}$ *is bounded.*

Proof. According to H 3.3-H 3.4, it is obvious that the claim holds.

Lemma 3.3 *For QP subproblem (9) at* x^k*, it holds that,*
(1) $z_k + \frac{1}{2}(d^k)^T H_k d^k \leq 0$.
(2) $z_k = 0 \Leftrightarrow d^k = 0 \Leftrightarrow z_k + \frac{1}{2}(d^k)^T H_k d^k = 0$.
(3) If $d^k \neq 0$*, then* $z_k < 0$*, and* d^k *is a feasible direction of descent at* x^k*.*

Proof. The proof of this lemma is similar to the proof of Lemma 3.4 in [4].

Lemma 3.4 *If* $d^k = 0$ *then* x^k *is a KKT point of main problem or there is a feasible direction of descent at* x^k *that obtained by resolving the subproblem corresponding to the working set* J_k*.*

proof. According to subproblem (9) and its KKT conditions, we have,

$$\begin{pmatrix} 1 \\ H_k d^k \end{pmatrix} + \hat{\mu}_k \begin{pmatrix} -1 \\ \nabla f(x^k) \end{pmatrix} + \sum_{j\in J_k} \hat{\lambda}_j^k \begin{pmatrix} -\eta_k \\ \nabla g_j(x^k) \end{pmatrix} = 0, \tag{13}$$

since $d^k = 0$, it holds that

$$\hat{\mu}_k \nabla f(x^k) + \sum_{j\in J_k} \hat{\lambda}_j^k \nabla g_j(x^k) = 0, \tag{14}$$

$$1 - \hat{\mu}_k - \sum_{j\in J_k} \eta_k \hat{\lambda}_j^k = 0. \tag{15}$$

We obtain $\hat{\mu}_k$ such that $\hat{\mu}_k \geq 0$. If $\hat{\mu} < 0$ then denote $\hat{\mu}_k = -\hat{\mu}_k$, $\hat{\lambda}_j^k = -\hat{\lambda}_j^k$. From the equation (14), it is obvious that $\hat{\mu}_k \neq 0$. Therefore from the equation (14) we obtain

$$\nabla f(x^k) + \sum_{j\in J_k} \frac{\hat{\lambda}_j^k}{\hat{\mu}_k} \nabla g_j(x^k) = 0,$$

By defining $\lambda_j^k = \frac{\hat{\lambda}_j^k}{\hat{\mu}_k}$ we have

$$\nabla f(x^k) + \sum_{j\in J_k} \lambda_j^k \nabla g_j(x^k) = 0.$$

Now, we consider three cases:
i) If for all $j \in J_k, \lambda_j^k \geq 0$ and

$$\lambda_j^k g_j(x^k) = 0 \quad for\ all\ j \in J_k$$

Then for all $j \in I \setminus J_k$ we define $\lambda_j^k = 0$ and x^k is the KKT point of main subproblem.
ii) If there are some $\lambda_j^k < 0$, then according to Step 4.3, we remove the constraint j' from J_k and resolve the subproblem corresponding to the working set J_k.

$$\begin{array}{ll} \min & z_k + \frac{1}{2}(d^k)^T H_k d^k \\ s.t. & \nabla f(x^k)^T d^k = z_k, \\ & \nabla g_j(x^k)^T d^k = \eta_k z_k \quad j \in J_k \setminus \{j'\} \equiv \hat{J}_k. \end{array}$$

Let (d^k, z_k) be the KKT point of this subproblem, then we have

$$\tilde{\mu}_k \begin{pmatrix} \nabla f(x^k) \\ -1 \end{pmatrix} + \sum_{j\in \hat{J}_k} \tilde{\lambda}_j^k \begin{pmatrix} \nabla g_j(x^k) \\ -\eta_k \end{pmatrix} + \begin{pmatrix} H_k d^k \\ 1 \end{pmatrix} = 0 \tag{16}$$

Given $\eta_k \geq 0$ and suppose $d = 0$ is obtained from the above subproblem. We define the set

$$N^k(\eta_k) \triangleq \left\{ \begin{pmatrix} \nabla f(x^k) \\ -1 \end{pmatrix}, \begin{pmatrix} \nabla g_j(x^k) \\ -\eta_k \end{pmatrix},\ j \in J_k \cup \{j'\} \right\},$$

and show this set is linearly independent. Now suppose the claim does not hold; i.e., suppose there exist scalars $v^j, j \in \{0\} \cup \hat{J}_k$, not all zero, such that

$$v^0 \begin{pmatrix} \nabla f(x^k) \\ -1 \end{pmatrix} + \sum_{j \in \hat{J}_k} v^j \begin{pmatrix} \nabla g_j(x^k) \\ -\eta_k \end{pmatrix} = 0. \tag{17}$$

In view of assumption H 3.3, $v^0 \neq 0$ and the scalars v^j are unique modulo a scaling factor. This uniqueness, the fact that $d^k = 0$, and the first scalar equations in the optimality conditions (13) imply that $\hat{\mu}_k = 1$ and

$$\hat{\lambda}_j^k = \begin{cases} \frac{v^j}{v^0}, & j \in J_k, \\ 0 & else \end{cases}$$

are KKT multipliers for subproblem (9). Thus, in view of (13),

$$\eta_k \sum_{j \in J_k} \frac{v^j}{v^0} = 0.$$

But this contradicts (17), which gives

$$\eta_k \sum_{j \in J_k} \frac{v^j}{v^0} = -1;$$

hence $N^k(\eta_k)$ is linearly independent.
Now by subtracting (13) from (16) we obtain

$$(\tilde{\mu}_k - \hat{\mu}_k) \begin{pmatrix} \nabla f(x^k) \\ -1 \end{pmatrix} + \sum_{j \in \hat{J}_k} (\tilde{\lambda}_j^k - \hat{\lambda}_j^k) \begin{pmatrix} \nabla g_j(x^k) \\ -\eta_k \end{pmatrix} - \hat{\lambda}_{j'}^k \begin{pmatrix} \nabla g_{j'}(x^k) \\ -\eta_k \end{pmatrix} + \begin{pmatrix} H_k d^k \\ 0 \end{pmatrix} = 0. \tag{18}$$

and multiplying by $\begin{pmatrix} d^k \\ z_k \end{pmatrix}^T$ we have,

$$\hat{\lambda}_{j'}^k \begin{pmatrix} d^k \\ z_k \end{pmatrix}^T \begin{pmatrix} \nabla g_{j'}(x^k) \\ -\eta_k \end{pmatrix} = ((d^k)^T \ z_k) \begin{pmatrix} H_k d^k \\ 0 \end{pmatrix} = (d^k)^T H_k d^k \geq 0,$$

since H_k is positive definite, we have $(d^k)^T H_k d^k = 0$ only if $d^k = 0$. But if $d^k = 0$, then by substituting into (18) and using linear independence of $N^k(\eta_k)$ we have that $\hat{\lambda}_{j'} = 0$, which contradicts our choice of j'. Hence we conclude that $(d^k)^T H_k d^k > 0$, and since $\hat{\lambda}_{j'}^k < 0$ by assumption, it follows immediately that

$$\begin{pmatrix} \nabla g_{j'}(x^k) \\ -\eta_k \end{pmatrix}^T \begin{pmatrix} d^k \\ z_k \end{pmatrix} < 0.$$

This shows that $d^k \neq 0$ and d^k is a feasible direction of descent at x^k.
iii) If some of the constraints are not satisfying the complementarity condition, we remove them from the working set and resolve the subproblem (9) with the new working set J_k.
Let d^k be the KKT point of this subproblem, and $\lambda^k = (\lambda_j^k, j \in J_k)$ be the corresponding multiplier vector. If $d^k \neq 0$ then d^k is a feasible direction of descent at x^k.
In summary, if $d^k = 0$ and for all $j \in J_k,\ \lambda_j^k \geq 0$ then from the above discussion, x^k is the KKT point of the main problem.

Otherwise if $d^k = 0$ and there exist $j \in J_k$ such that $\lambda_j^k < 0$ or some of the constraints are not satisfying the complementarity, then according to cases (ii) and (iii) we obtain a feasible direction of descent at x^k.

Lemma 3.5 *The line search in Step 6 yields a stepsize $t_k = (\frac{1}{2})^i$ for some finite $i = i(k)$.*

Proof. According to (3) of lemma 3.3, Lemma 3.4 and assumption H3.2, it is easy to finish the proof.

Lemma 3.6 *For the QP subproblem (9) x^k, if $d_0^k = 0$, then x^k is a KKT point of main problem. If $d_0^k \neq 0$ then d^k computed in Step 4 is a feasible direction of descent for the main problem at x^k or $d^k = 0$ and x^k is the KKT point for the main problem at x^k.*

Proof. The proof of the first claim is similar to the proof of Lemma 3.3 in [24] and the second claim according to Lemma 3.4 is obvious.
The above discussion shows that the Algorithm A is well defined.
In the sequel, the global convergence of Algorithm A is shown. For this purpose we make the following additional assumption.
H 3.5. There exist $a, b > 0$, such that $a\|d\|^2 \leq d^T H_k d \leq b\|d\|^2$, for all k and all $d \in \Re^n$.
Since there are only finite choices for set $J_k \subset I$, from H 3.4, H 3.5 and Lemma 3.2, we might as well assume that there exist a subsequence K, such that

$$x^k \to x^*, \quad H_k \to H_*, \quad d_0^k \to d_0^*, \quad d^k \to d^*, \quad z_k \to z_*,$$

$$a^k \to a^*, \quad b^k \to b^* \quad J_k \equiv J \neq \emptyset, \quad k \in K, \tag{19}$$

where J is a constant set.

Theorem 3.7 *The algorithm either stops at the KKT point x^k of the problem (1.1) in finite number of steps, or generates an infinite sequance $\{x^k\}$ any accumulation point of which is a KKT point of the problem (1.1).*

Proof. The first statement is easy to show, since the only stopping point is in Step 3 and Step 4. Thus, assume that the algorithm generates an infinite sequence $\{x^k\}$ and (19) holds. According to Lemma 3.3, it is only necessary to prove that $d_0^* = 0$.
Suppose by contradiction that $d_0^* \neq 0$. Then, it is easy to see that d^* is the sole solution of the following quadratic subproblem:

$$\begin{array}{ll} \min & z + \frac{1}{2} d^T H_* d, \\ s.t. & \nabla f(x^*)^T d = z \\ & \nabla g_j(x^*)^T d = \eta_* z \quad j \in I(x^*), \end{array}$$

so, imitating the proof of Lemma 3.4, it is obvious that d^* is well defined, and it holds that

$$\nabla f(x^*)^T d^* < 0, \quad \nabla g_j(x^*)^T d^* < 0, \quad j \in I(x^*) \subseteq J. \tag{20}$$

Thus, from (20), it is easy to see that the stepsize t_k obtained in Step 6 are bounded away from zero on K, i.e.,

$$t_k \geq t_* = \inf\{t_k, k \in K\} > 0, k \in K. \tag{21}$$

In addition, from (11) and Lemma 3.6, it is obvious that $\{f(x^k)\}$ is monotonously decreasing. So according to assumption H 3.2, the fact $\{x^k\}_K \to x^*$ implies that

$$f(x^k) \to f(x^*), \quad k \to \infty.$$

So, from (11), (20) and (21), it holds that

$$0 = \lim_{k \in K} (f(x^{k+1}) - f(x^k)) \leq \lim_{k \in K} (\alpha t_k \nabla f(x^k)^T d^k) \leq \frac{1}{2} \alpha t_* f(x^*)^T d^* < 0,$$

which is a contradiction. Thus, x^* is a KKT point of (1.1).

4. The rate of convergence

Now we discuss the convergent rate of the algorithm, and prove that sequence $\{x^k\}$ generated by the algorithm is one-step superlinearly convergent. For this purpose, we add some stronger regularity assumptions.
H 4.1. The second-order sufficiency conditions with strict complementary slackness are satisfied at the KKT point x^* and the corresponding multiplier vector u^*.
H 4.2. $H_k \to H_*, \quad k \to \infty$.
H 4.3. Let

$$\|P_k(H_k - \nabla^2_{xx} \ell(x^*, u^*) d^k\| = o(\|d^k\|),$$

where

$$P_k = I_n - A_k(A_k^T A_k)^{-1}A_k^T, \quad \nabla^2_{xx}\ell(x^*, u^*) = \nabla^2 f(x^*) + \sum_{j\in I} u_j^* \nabla^2 g_j(x^*).$$

Under assumption H 4.1, we know that the KKT point x^* is isolated.
Meanwhile, if x^* is an interior point, then for k large enough, Algorithm A will eternally become the quasi Newton method to unconstrained optimization. Obviously, this method is superlinearly convergent. In the sequel, we might as well assume that x^* is not an interior point of (1.1), i.e., $I(x^*) \neq \emptyset$.

Lemma 4.1 *It holds, for k large enough, that*

$$J_k \equiv I(x^*) \triangleq I_*, \quad d_0^k \to 0, \quad a^k \to u_{I_*} = (u_j^*, j \in I_*), \quad b^k \to (u_j^*, j \in I_*),$$

$$\mu_k \to 1 \quad z_k = O(\|d^k\|).$$

Proof. The proof of this lemma is similar to the proof of Lemma 4.1 in [24] and Lemma 4.3 in [23].

Lemma 4.2 *For k large enough, (d^k, λ^k) obtained from step 4 satisfies*

$$\|d^k\| \sim \|d_0^k\|, \quad \lambda^k \to u_{I_*}^*.$$

Proof. From $a^k \to u_{I_*} = (u_j^*, j \in I_*)$ and the conditions with strict complementary, we know, for k large enough, that $a_j^k > 0$, $g_j(x^k) = 0,\ j \in I_*$, thereby, from (8), it holds, for k large enough, that

$$\begin{array}{l} H_k d_0^k + A_k b^k + \nabla f(x^k) = 0 \\ \qquad \nabla g_j(x^k)^T d_0^k = 0 \quad j \in I_*, \end{array} \tag{22}$$

From (9), it holds, for k large enough, that

$$\begin{array}{l} H_k d^k + A^k \lambda^k + \nabla f(x^k) = 0 \\ \qquad \nabla g_j(x^k)^T d^k = \eta_k z_k \quad j \in I_*. \end{array} \tag{23}$$

Denote

$$d^k = d_0^k + \Delta d^k, \quad \lambda^k = b^k + \Delta\lambda^k.$$

According to (22) and (23), it holds that

$$\begin{array}{l} H_k \Delta d^k + A_k \Delta\lambda^k = 0 \\ \nabla g_j(x^k)^T \Delta d^k = \eta_k z_k \quad j \in I_*, \end{array}$$

i.e.,

$$\begin{array}{l} (\Delta d^k)^T H_k \Delta d^k + (A_k^T \Delta d^k)^T \Delta\lambda^k = 0 \\ \qquad A_k^T \Delta d^k = z_k \|d_0^k\|^\nu e \quad j \in I_*. \end{array} \tag{24}$$

Since $z_k = O(\|d^k\|)$ so, it is easy that $\|\Delta d^k\|^2 = \|d^k\|\|d_0^k\|^\nu$ and $\|d^k\|^2 \le \|d^k\|\|d_0^k\|^\nu + \|d_0^k\|^2$ and from Lemma 4.1 thereby, we have

$$\|d^k\| \sim \|d_0^k\|, \quad d^k \to 0, \quad k \to \infty. \tag{25}$$

From the KKT conditions for the subproblem (9) we have

$$H_k d^k + A_k \lambda^k + \mu_k \nabla f(x^k) = 0.$$

and according to Lemma 4.1 and (25) it holds that

$$A_k \lambda^k + \nabla f(x^k) = 0,$$

and imitating the proof of Lemma 4.1, it holds that $\lambda^k \to u_{I_*}^*,\ k \to \infty$.

Lemma 4.3 *for k large enough, $\tilde{d}^k$ obtained from Step 5 satisfies:*

$$\|\tilde{d}^k\| = O(\max\{\|d^k\|^2, -\eta_k z_k\}) = o(\|d^k\|).$$

Proof. By expanding $g_j(x^k + d^k)$ around x^k, we have

$$g_j(x^k + d^k) = g_j(x^k) + \nabla g_j(x^k)^T d^k + O(\|d^k\|^2).$$

and from (8) it holds that

$$g_j(x^k + d^k) = -\nabla g_j(x^k)^T d_0^k - \min\{0, a_j\} + \nabla g_j(x^k)^T d^k + O(\|d^k\|^2).$$

From $a^k \to u_{I_*} = (u_j^*, j \in I_*)$ and the conditions with strict complementarity, we know, for k large enough, that $a_j^k > 0, j \in I_*$, thereby from above, it holds, for k large enough, that

$$g_j(x^k + d^k) = \nabla g_j(x^k)^T (d^k - d_0^k) + O(\|d^k\|^2).$$

and from (24)and (25) we have

$$g_j(x^k + d^k) = \eta_k z_k + O(\|d^k\|^2).$$

Definitions of ψ_k and Lemma 4.1 implies that $\psi_k = o(\|d^k\|^2)$. So, from (10), it is clear that

$$\|d_1^k\| = O(\max\{\|d^k\|^2, -\eta_k z_k\}) = o(\|d^k\|).$$

Thereby, we have

$$\|\tilde{d}^k\| = O(\max\{\|d^k\|^2, -\eta_k z_k\}) = o(\|d^k\|).$$

Denote

$$\nabla_{xx}^2 \ell(x^k, \mu_k, \lambda^k) = \nabla^2 f(x^k) + \sum_{j \in I_*} \frac{\lambda_j^k}{\mu_k} \nabla^2 g_j(x^k)$$

Due to Lemma 4.1, Lemma 4.2 and H 4.4, it holds that

$$\|P_k(H_k - \nabla_{xx}^2 \ell(x^k, \mu_k, \lambda^k))d^k\| = o(\|d^k\|). \tag{26}$$

Lemma 4.4 *For k large enough, the step $t_k \equiv 1$ is accepted by the line search.*

Proof. The proof of this lemma is similar to the proof of Lemma 4.5 in [23] with some little differences since the QP (9) is different from that of [23].

Theorem 4.5 *Under all above-mentioned assumptions, the algorithm is superlinearly convergent, i.e., the sequence $\{x^k\}$ generated by the algorithm satisfies $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$.*

Proof. The proof is analogous to Theorem 4.1 in [5] but with some technical differences since the QP (9) is different from that of [5].
Denote

$$\nabla_{xx}^2 \ell(x^*, u^*) = \nabla^2 f(x^*) + \sum_{j \in I} u_j^* \nabla^2 g_j(x^*) = \nabla^2 f(x^*) + \sum_{j \in I_*} u_j^* \nabla^2 g_j(x^*).$$

From the KKT conditions for the subproblem (9), the fact $J_k \equiv I_*$ implies that

$$\begin{aligned} \mu_k \nabla f(x^k) + H_k d^k + A_k \lambda_{I_*}^k &= 0, \\ A_k^T d^k &= \eta_k z_k, \quad j \in I_*. \end{aligned} \tag{27}$$

So, from definition of $P_k (P_k A_k = 0)$ and (27), it holds that

$$P_k H_k d^k = -\mu_k P_k \nabla f(x^k) \quad = \quad -\mu_k P_k (\nabla f(x^k) + A_k u_{I_*}^*) = -\mu_k P_k \nabla_{xx}^2 \ell(x^*, u^*)(x^k - x^*) + O(\|x^k - x^*\|^2).$$

i.e.,

$$P_k \nabla_{xx}^2 \ell(x^*, u^*)(x^k + d^k - x^*) = -P_k(\frac{1}{\mu_k} H_k - \nabla_{xx}^2 \ell(x^*, u^*))d^k + O(\|x^k - x^*\|^2).$$

Therefore, according to $\|\tilde{d}^k\| = o(\|d^k\|)$ one gets

$$P_k \nabla_{xx}^2 \ell(x^*, u^*)(x^k + d^k + \tilde{d}^k - x^*) = -P_k(\frac{1}{\mu_k} H_k - \nabla_{xx}^2 \ell(x^*, u^*))d^k + O(\|x^k - x^*\|^2) + o(\|d^k\|). \tag{28}$$

Consider $j \in I_*$, expanding $g_j(x^k + d^k + \tilde{d}^k)$ around $x^k + d^k$ we obtain

$$g_j(x^k + d^k + \tilde{d}^k) = g_j(x^k + d^k) + \nabla g_j(x^k + d^k)^T \tilde{d}^k + O(\|\tilde{d}^k\|) = g_j(x^k + d^k) + \nabla g_j(x^k)^T \tilde{d}^k + O(\|d^k\|.\|\tilde{d}^k\|).$$

From (10), we get

$$A_k^T \tilde{d}^k = (A_k^1)^T d_1^k = \psi_k e - g^k,$$

i.e.,

$$\nabla g_j(x^k)^T \tilde{d}^k = -g_j(x^k + d^k) - \psi_k, \quad j \in I_*.$$

Thereby, according to Lemma 4.3, it holds that

$$g_j(x^k + d^k + \tilde{d}^k) = -\psi_k + O(\max\{\|d^k\|^2, -\eta_k z_k \|d^k\|\}). \tag{29}$$

From (29) and expanding $g_j(x^k)$ around x^*, we obtain

$$A_k^T(x^k - x^*) + O(\|x^k - x^*\|^2) + A_k^T(d^k + \tilde{d}^k) + \frac{1}{2}(d^k)^T \nabla^2 g_j(x^k) d^k = O(\|d^k\|^2),$$

i.e.,

$$A_k^T(x^k + d^k + \tilde{d}^k - x^*) = O(\|x^k - x^*\|^2) + o(\|d^k\|). \tag{30}$$

Thereby, from (28) and (30), it holds that

$$\begin{pmatrix} P_k \nabla_{xx}^2 \ell(x^*, u^*) \\ A_k^T \end{pmatrix} (x^{k+1} - x^*) = \begin{pmatrix} -P_k \left(\frac{1}{\mu_k} H_k - \nabla_{xx}^2 \ell(x^*, u^*) \right) d^k \\ o(\|d^k\|) \end{pmatrix} + O(\|x^k - x^*\|^2).$$

While, it is not difficult to show that the matrix $G_k = \begin{pmatrix} P_k \nabla_{xx}^2 \ell(x^*, u^*) & A_k \\ A_k^T & 0 \end{pmatrix}$ is nonsingular. For nonsingularity, we show that the system of $G_k \begin{pmatrix} y \\ \bar{y} \end{pmatrix} = 0$ have unique solution zero. So

$$A^T y = 0, P_k \nabla_{xx}^2 \ell(x^*, u^*) y + A_k \bar{y} = 0.$$

From of definition of P_k and $A^T y = 0$ we have $y^T P_k = y^T$ and multiply above equation to y^T, we obtain

$$y^T P_k \nabla_{xx}^2 \ell(x^*, u^*) u + y^T A_k \bar{y} = 0$$

which implies that $y^T \nabla_{xx}^2 \ell(x^*, u^*) y = 0$. And from positive definite of $\nabla_{xx}^2 \ell(x^*, u^*)$, we have that $y = 0$ and from the full column rank of A_k we have $\bar{y} = 0$. Therefore G_k is nonsingular. So, according to (26) and $\mu_k \to 1$, we get

$$\|x^{k+1} - x^*\| = o(\|d^k\|) + O(\|x^k - x^*\|^2).$$

Thereby we have,

$$\begin{aligned}
\|x^{k+1} - x^*\| &= O(\|x^k - x^*\|^2) + o(\|d^k\|) \\
&= O(\|x^k - x^*\|^2) + o(\|d^k + \tilde{d}^k\|) \\
&= O(\|x^k - x^*\|^2) + \frac{o(\|d^k + \tilde{d}^k)}{\|d^k + \tilde{d}^k\|}(\|(x^{k+1} - x^*) - (x^k - x^*)\|) \\
&\leq O(\|x^k - x^*\|^2) + \frac{o(\|d^k + \tilde{d}^k)}{\|d^k + \tilde{d}^k\|}(\|x^{k+1} - x^*\| - \|x^k - x^*\|) \\
&= o(\|x^{k+1} - x^*\|) + o(\|x^k - x^*\|)
\end{aligned}$$

So, it holds that $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$.

Table 1: The detailed information of numerical experiments

No.	NIT	EQPs	NF	FV	$\|d_0^k\|$
012	22	44	22	−29.9999991789812	$8.4568e-09$
024	17	34	17	−0.98766543423900	$3.7665e-10$
029	18	36	18	−22.6274571000400	$7.1230e-09$
030	22	44	22	1.00000764398900	$5.3546e-09$
033	35	70	35	−4.5867557899000	$5.9646e-09$
034	30	60	60	−0.8434000000000	$3.9765e-10$
035	09	18	09	0.11100026545000	$7.1418e-09$
036	100	200	100	−3270.6543343000	$5.3464e-09$
037	100	200	100	−3443.9678909000	$2.7645e-09$
043	44	88	44	−43.678899886500	$6.5112e-10$
045	90	180	90	1.10009873456800	$8.5954e-09$
065	14	28	28	0.95352921678900	$5.8765e-10$
076	40	80	40	−4.6837650017800	$4.2315e-09$
083	70	140	70	−29980.999997500	$5.9012e-09$
084	70	140	70	6954034.76688900	$3.6721e-10$
100	53	106	53	680.626765545400	$3.0897e-09$
118	101	202	101	664.897553232100	$4.0234e-09$

5. Numerical experiments and conclusions

In this section, we carry out some limited numerical experiments based on the algorithm was presented in section 2. In the implementations we set $\nu = 0.5,\ \alpha = 0.25,\ \beta = 0.4,\ \tau = 2.25,\ \rho = 1$ and $H_0 = I$. The $n \times n$ unit matrix H_k is updated by the BFGS formula [18]. Of course, the Hessian matrix of the objective function of subproblem (9) is an $(n+1) \times (n+1)$ matrix

$$G_k = \begin{pmatrix} H_k & 0 \\ 0 & 0 \end{pmatrix}.$$

Obviously, G_k is singular and in the implementation, we define

$$G_k = \begin{pmatrix} H_k & 0 \\ 0 & \delta_k \end{pmatrix},\ \delta_k = \max\{\min\{\eta_k, 0.5\}, 10^{-4}\},$$

therefore, the QP subproblem (9) is replaced by the following QP subproblem:

$$\begin{array}{ll} \min & z + \frac{1}{2} d^T H_k d + \delta_k z^2, \\ s.t. & \nabla f(x^k)^T d = z \\ & \nabla g_j(x^k)^T d^k = \eta_k z \qquad j \in J_k. \end{array}$$

However, the fact $z_k = O(\|d^k\|) \to 0$ implies that this replacement would not have an effect on the global convergence and the superlinear convergence rate of the proposed algorithm.

In the implementation, the stopping criterion of Step 3 and Step 4 is changed to

If $\|d^k\| \leq 10^{-8}$, STOP.

Following [11], this algorithm has been tested on some standard problems in [8] and a feasible initial point is provided for each problem. The results are summarized in Table 1. For each problem, No. is the number of the test problem in [8], NIT is the number of iterations, EQPs is the number of the total times solving the quadratic programing subproblem with only equality constraints, NF is the number of evaluations of the objective functions, and FV is the final value of the objective function. Comparison the computation results in Table 1 and the computation results of the algorithm given in [11] shows our algorithm is more efficient on these numerical problems.

References

[1] M. S. Bazaraa, C. M. Shetly, Nonlinear Programing Theory and Algorithms, *AWiley*, New York, 1979.

[2] J. Birge, L. Qi, Z. Wei,A variant of the Topkis-Veinott method for solving inequality constrained optimization problems, *J. Appl. Math. Optim.* 41 (2000)309-330.

[3] J. F. Bonnans, E. R. Panier,A. L. Tits, J. Zhou, Avoiding the Maratos effect by means of a nonmonotone line search II: Inequality constrained problems-feasible iterate, *SIAM J. Numer. Anal.* 29 (1992) 1187-1202.

[4] X. B. Chen, M. M. Kostreva, A generalization of the norm-relaxed method of feasible directions, *Appl. Math. Comput.* 102 (1999) 257-272.

[5] F. Facchinei, S. Lucidi,Quadratically and superlinearly convergent for the solution of inequality constrained otimization problem, *JOTA* 85 (1995) 265-289.

[6] M. Fukushima, A successive quadratic programming algorithm with global and superlinear convergence properties, *Math. Programming* 35 (1986) 253-264.

[7] S. P. Han,Superlinearly convergent variable metric algorithm for general nonlinear programming problems, *Math. Programming* 11 (1976) 263-282.

[8] W. Hock, K. Schittkowski,Test Examples for Nonlinear Programming Codes, *Lecture Notes in Economics and Mathematical Systems, vol. 187, Springer,* Berlin, 1981.

[9] M. M. Kostreva, X. Chen,A superlinearly convergent method of feasible directions, *Appl. Math. Comput.* 116 (2000) 245-255.

[10] CT. Lawrence, A. L. Tits, Nonlinear equality constraints in feasible sequential quadratic programming, *Optim. Methods Softw.* 6 (1996) 265-282.

[11] C. T. Lawrence, A. L. Tits, A computationally efficient feasible sequential quadratic programming algorithm, *SIAM J. Optim.* 11 (2001) 1092-1118.

[12] N. Maratos, Exact penalty functions for finite dimensional and control optimization problems, *Ph. D. Thesis, University of Science and Technology,* London, 1978.

[13] D. Q. Mayne, E. Polak, A superlinearly convergent algorithm for constrained optimization problems, *Math. Programming* 16 (1982) 45-61.

[14] E. R. Panier, A. L. Tits, A superlinearly convergent feasible method for the solution of inequality constrained optimization problems, *SIAM J. Control Optim.* 25 (1987) 934-950.

[15] E. R. Panier, A. L. Tits, On combining feasibility, descent and superlinear convergence in inequality constrained optimization, *Math. Programing* 59 (1993) 261-276.

[16] E. R. Panier, A. L. Tits, J. N. Herskovits, A QP-free global convergent, locally superlinearly convergent algorithm for inequality constrained optimization, *SIAM J. Control Optim.* 26 (1988) 788-811.

[17] O. Pironneau, E. Polak, On the rate of convergence of certain methods for centers, *Math. Programming* 2 (1972) 230-257.

[18] M. J. D. Powell, Y. Yuan, A fast algorithm for nonlinearly constrained optimization calculations, in: G. A. Waston (Ed.), *Numerical Analysis, Springer,* Berlin, 1978.

[19] M. J. D. Poweell, Y. Yuan, A recursive programming algorithm that uses differentiable exact penalty function, *Math. Programming* 35 (1986) 265-278.

[20] P. Spellucci, An SQP method for general nonlinear programs using only equality constrained subproblems, *Math. Programming* 82 (1998) 413-448.

[21] D. M. Topkis, A. F. Veinott, On the convergence of some feasible direction algorithms for nonlinear programming, *SIAM J. Control* 5 (1967) 268-279.

[22] G. Zoutendijk, Methods of Feasible Directions, *Elsevier,* Amsterdam, 1960.

[23] Z. B. Zhu, An efficient sequential quadratic programming algorithm for nonlinear programming, *Journal of Computational and Applied Mathematics,* 175 (2005) 447-464.

[24] Z. Zhu, A sequential equality constrained quadratic programming algorithm for inequality constrained optimization, *Journal of Computational and Applied Mathematics,* 212 (2008) 112-125.