Mathematica Aeterna, Vol. 6, 2016, no. 3, 455 - 466

An improved nonmonotone feasible direction method

Ke Su
Shibo Tang

College of Mathematics and Information Science, Hebei University, Baoding, China

Abstract

In this paper, a new sequential quadratic programming (SQP) method of feasible directions is proposed and analyzed for nonlinear programming,where a feasible direction of descent can be derived from solving only one QP subproblem. The algorithm has no demond on initial point, moreover it avoids using a penalty function or a filter. So it is more flexible and easier to implement. To avoid Maratos effect, a revised direction is computed by solving a linear system. Under some reasonable conditions, the global convergence is shown.

Mathematics Subject Classification: 90C30.

Keywords: Constrained optimization, KKT point, Sequential quadratic programming, Global convergence.

1 Introduction

This paper considers the following nonlinear inequality constrained optimization problem

$$(P) \quad \min\ f(x)$$
$$s.t.\ g_j(x) \le 0,\ j = 1, 2, \cdots, m,$$

where $f : R^n \to R$ and $g(x) = (g_1(x), g_2(x), \cdots, g_m(x))^T : R^n \to R^m$ are second-order continuously differentiable.

Nonlinear inequality constrained optimization problem is an important research topic in mathematical programming fields. Many practical problems can be modeled as the nonlinear inequality constrained optimization problem[1-5]. There are various important methods for solving nonlinear constrained optimization problem, such as feasible direction algorithm, interior-point algorithm, sequential quadratic programming, and so on. Among these methods,the sequential quadratic programming (SQP) method has become one of the most effective methods for solving nonlinear constrained optimization problem.In each iteration,the following quadratic subproblem needs to be solved in SQP method. SQP is solved as follows:

min $\nabla f(x_k)^T d + \frac{1}{2} d^T H_k d$
$s.t.\ g_j(x) + \nabla g_j(x_k)^T d \leq 0\ j = 1, 2, \cdots, m,$

The SQP algorithm have two serious shortcomings. First, in order to obtain a search direction, one must solve one or more quadratic programming subproblems per iteration, and the computation amount of this type algorithm is very large. Second, the search direction obtained is not often feasible. Although some additional procedures can be used, they may make the algorithms more complex. Therefore, it is necessary to study some new approach to avoid these shortcomings[6-10].

After the direction d_k is determined the next task is to find a step size α_k along the search direction. The ideal line search rule is the exact one which satisfies: $f(x_k + \alpha_k d_k) = \min_{\alpha>0} f(x_k + \alpha d_k)$ in fact, the exact step size is difficult or even impossible to seek in practical computation, and thus many researchers constructed some inexact line search rule, such as Armijo rule, Goldtein rule, Wolfe rule and nonmontone line search[12].

In 1982, Chamberlain et al. proposed a watchdog technique for constrained optimization, in which some standard line search conditions were relaxed to overcome the Marotos effect[11]. Motivated by this idea, Grippo,Grippo, and Lucidi presented a nonmonotone Armijo-type line search technique for the Newton method[12]. The traditional line search rules require the function value descent monotonically at each iteration.It may considerably slow the rate of convergence in the intermediate stages of the minimization process, especially in the presence of the narrow curved valley. While the nonmontone line search rules are effective or even powerful at some iteration, especially when the iterates are trapped in a narrow curved valley of objective functions.Now we give a new nonmonotone line search proposed by Zhang and W.Hager as follows:[13]

Initialization : Choose starting guess x_0, and parameter $0 \leq \eta_{min} \leq \eta_{max} \leq 1$, $0 < \delta < 1 < \rho$ and $\mu > 0$. Set $C_0 = f(x_0)$, $Q_0 = 1$, and $k = 0$.

Line search update : Set $x_{k+1} = x_k + \alpha_k d_k$ Where α_k satisfies the nonmonotone Armijo condition:$\alpha_k = \bar{\alpha}_k\ \beta^{h_k}$.

$$f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k \nabla f(x_k) d_k, \tag{1}$$

Where $\bar{\alpha}_k > 0$ is the trial step, and h_k is the largest integer such that (1) holds and $\alpha_k \leq \mu$.

Cost update : Choose $\eta_k \in [\eta_{min}, \eta_{max}]$, and set

$$Q_{k+1} = \eta_k Q_k + 1, C_{k+1} = (\eta_k Q_k C_k + f(k+1))/Q_{k+1} \tag{2}$$

Replace k by $k+1$ and return to the convergence test.

Observe that C_{k+1} is a convex combination of C_k and $f(x_{k+1})$.Since $C_0 = f(x_0)$,it follows that C_k is a convex combination of the function values $f(x_0)$, $f(x_1)$,$f(x_2)$,$\cdots$,$f(x_k)$.The choice of η_k controls the degree of nonmonotonicity. If $\eta_k = 0$ for each k, then the line is the usual monotone Armijo line search. If $\eta_k = 1$ for each k, then $C_k = A_k$, where

$$A_k = \frac{1}{k+1}\sum_{i=0}^{k} f_i, f_i = f(x_i), \tag{3}$$

is the average function value. the scheme with $C_k = A_k$ was suggested to us by Yu-hong Dai. In [14], the possibility of comparing the current function value with an average was analyzed.

Motivated by the above techniques, we propose a modified SQP method based on the subproblem proposed in [7].This method has no requirement on initial point and only need to solve one QP subproblem.Under some reasonable conditions,we prove its global convergence.

This paper is organized as follow. Some notions and symbols are given in section 2. In section 3, we describe the algorithm.The global convergence theory for the method is presented in section 4.

2 Preliminary Notes

The symbols and notions we used in this paper are standard as following:

(1)Directional derivative

$$f'(x,d) = \lim_{\lambda\to 0}\frac{f(x+\lambda d) - f(x)}{\lambda} \tag{4}$$

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

(3)$\|x\|_\infty = \max\{|x|_j\, j = 1, 2, \ldots, n\}$;

(4)Let $I = \{1, 2, \cdots, m\}$, the set of active indices at a point x is defined as $A(x) = \{j \in I \mid g_j(x) = 0\}$.

Furthermore for convenience,we list some signs and lemmas as follows:

$$\Phi(x) = \max\{0, g_j(x) : j \in I\}. \tag{5}$$

$$\Psi(x) = \max\{g_j(x) : j \in I\}. \tag{6}$$

For $\forall x, d \in R^n$, let $\Psi^*(x; d)$ be the first order approximation to $\Psi(x + d)$, namely $\Psi^*(x; d) = \max\{g_j(x) + \nabla g_j(x)^T d : j \in I\}$

For $\forall\sigma>0$,function $\Psi(x, \sigma), \Psi^0(x, \sigma) : R^n \times R^+ \to R$ are defined as follows:

$$\Psi(x, \sigma) = min\{\Psi^*(x; d) : \|d\| \leq \sigma\}, \tag{7}$$

$$\Psi^0(x,\sigma) = max\{\Psi(x,\sigma), 0\}. \tag{8}$$

Remark 2.1 (7)equals to the following linear programming:

$$LP(x,\sigma) : min\{z : g_j(x) + \nabla g_j(x)^T d \le z, j \in I, \|d\| \le \sigma\}. \tag{9}$$

Denote

$$\theta(x,\sigma) = \Psi(x,\sigma) - \Psi(x), \tag{10}$$

$$\theta^0(x,\sigma) = \Psi^0(x,\sigma) - \Psi(x), \tag{11}$$

$$F = \{x : g_j(x) \le 0; j \in I\} = \{x : \Psi(x) \le 0\}. \tag{12}$$

$$F^c = \{x : \Psi(x) > 0\}. \tag{13}$$

Definition 2.1 *[6].Mangasarian-Fromotz constraint qualification (MFCQ) is said to be satisfied by $g(x) \le 0$ at x if $\exists z \in R^n$ such that $\nabla g_j(x)^T z < 0$,$\forall j \in \{j \in I | g_j(x) \ge 0\}$.*

Lemma 2.1 [7]. $\forall x \in F^c$, MFCQ is satisfied at x, then $\theta(x,o) < 0 (\sigma > 0)$.

Lemma 2.2 [7]. $\Psi(x,\sigma), \Psi^0(x,\sigma), \theta(x,\sigma), \theta^0(x,\sigma)$ are all continuous on $R^n \times R^+$.

Lemma 2.3 [7]. $\forall x \in F^c$, if $\theta(x,\sigma) < 0$, then $\theta^0(x,\sigma) < 0$.

3 Main Results

Given $x \in R^n, \sigma > 0. D(x,\sigma)$ is defined as the following set:

$D(x,\sigma) = \{d | g_j(x) + \nabla g_j(x)^T d \le \Psi^0(x,\sigma), j \in I\}$

If d^* is the solution of $LP(x,\sigma)$, the $d^* \in D(x,\sigma)$, hence $D(x,\sigma)$ is nonempty. The quadratic subproblem is replaced by the following convex programming problem:

$Q(x_k, H_k, \sigma_k) : \min \ \nabla f(x_k)^T d + \frac{1}{2} d^T H_k d$

$s.t. \ g_j(x) + \nabla g_j(x_k)^T d \le \Psi^0(x_k, \sigma_k) \ j \in L_k,$

where L_k is the set of approximate active indices of the point x_k. Clearly, by the above statement, the convex programming $Q(x_k, H_k, \sigma_k)$ is unique. The convex programming problem has the following properties.

Theorem 3.1 [7].Suppose that $x_k \in R^n, H_k \in R^{n\times n}$ is a symmetric positive definite matrix. If MFCQ is satisfied at x_k then

(1) The convex programming problem $Q(x_k, H_k, \sigma_k)$ has a unique solution d_k which satisfies KKT conditions, i.e. there exist vectors $\lambda^k = (\lambda_j^k \ , j \in L_k \bigcup E)$ such that

$(a)\ g_j(x_k) + \nabla g_j(x_k)^T d_k \le \Psi^0(x_k, \sigma_k), \ j \in L_k;$

$(b)\ \lambda_j^k \ge 0, \ j \in L_k;$

$(c)\ \nabla f(x_k) + H_k d_k + B_k \lambda^k = 0, B_k = (\nabla g_j(x_k), j \in L_k);$

(*d*) $\lambda_j^k(g_j(x_k)+\nabla g_j(x_k)^T d_k)=0, j\in L_k$;

(2)If $d_k=0$ is the solution of $Q(x_k,H_k,\sigma_k)$,then x_k is a KKT point of problem (P)

Lemma 3.1 [7] $\forall x\in F, d\in D(x,\sigma)$, then $\Psi^*(x;d)=0$.

Now,the algorithm for solving problem (P) can be stated as follows.

Algorithm 3.1

step 0 : Given $x_0\in R^n$, $H_0=I$. $k=0$, $\varepsilon_0>0$, $\sigma_r>\sigma_l>0$. $\sigma_0\in[\sigma_l,\sigma_r]$, $0\le\eta_{min}\le\eta_{max}\le 1$, $0<\delta<1$, $0<\rho<1, \mu\ge 0$ .Set $C_0=f(x_0), Q_0=1, k=0$.

step 1 : Computation of an 'active' constraint set L_k:

S1.1 Let $i=0, \varepsilon_{k,i}=\varepsilon_0$;

S1.2 Set

$L_{k,i}=\{j\in I|-\varepsilon_{k,i}\le g_j(x_k)-\Phi(x_k)\le 0\}$,

$B_{k,i}=(\nabla g_j(x_k), j\in L_{k,i})$.

If $det(B_{k,i}^T B_{k,i})\ge\varepsilon_{k,i}$, let $L_k=L_{k,i}, B_k=B_{k,i}, i_k=i$, go to step 2;

S1.3 Set $i=i+1, \varepsilon_{k,i}=\frac{1}{2}\varepsilon_{k,i-1}$, and go to S1.2 (inner loop).

step 2 : Computation of the direction d_k: Compute $\Psi(x,\sigma), \Psi^0(x,\sigma)$,let d_k be the solution of convex programming problem $Q(x_k,H_k,\sigma_k)$. If $d_k=0$,then x_k is a KKT point of problem (P);

step 3 : Let $\hat{d}_k$ be the least norm solution of the following linear equation system:

$g_j(x_k+d_k)+\nabla g_j(x_k)^T d=0,\quad j\in L_k$. and if the above linear equation system is inconsistent or $\|\hat{d}_k\|>\|d_k\|$,then let $\hat{d}_k=0$;

step 4 : let $d_k=d_k+\hat{d}_k$, If

$$f(x_k+\alpha_k d_k)\le C_k+\delta\alpha_k\nabla f(x_k)^T d_k \tag{14}$$

$$g_j(x_k+\alpha_k d_k)\le 0,\quad j\in I \tag{15}$$

hold, then go to step 7.

Where $\alpha_k=\bar{\alpha}_k\beta^{h_k}$, $\bar{\alpha}_k>0$ is the trial step, and h_k is the largest integer such that (14) holds.

step 5 : Computation of direction q_k:

Let B_k^1 be the matrix whose rows are $|L_k|$ linearly independent row of B_k,and B_k^2 be the matrix whose rows are the remaining $n-|L_k|$ rows of B_k.We might denote $B_k=\begin{pmatrix}B_k^1\\B_k^2\end{pmatrix}$

Like B_k,we might as well let $\nabla f(x_k)=\begin{pmatrix}\nabla f_1(x_k)\\\nabla f_2(x_k)\end{pmatrix}$.

Compute $\rho_k=-\nabla f(x_k)^T d_k,\ \pi_k=-(B_k^1)^{-1}\nabla f_1(x_k),\ q_k=\rho_k(d_k+\overline{d}_k)$,

$$\widetilde{d}_k=\frac{-\rho_k((B_k^1)^{-1})^T e}{1+2|e^T\pi_k|} \tag{16}$$

,

where $\overline{d}_k = \begin{pmatrix} \widetilde{d}_k \\ 0 \end{pmatrix}$,
$e = (1, 1, \ldots, 1)^T \in R^{|L_k|}$.

step 6 : Compute α_k such that:

$$f(x_k + \alpha_k q_k) \leq C_k + \delta\alpha_k \nabla f(x_k)^T q_k \tag{17}$$

$$g_j(x_k + \alpha_k q_k) \leq 0, \quad j \in I \tag{18}$$

Where $\alpha_k = \bar{\alpha}_k \beta^{h_k}$, $\bar{\alpha}_k > 0$ is the trial step, and h_k is the largest integer such that (17) holds .

Let $d_k = q_k$;

step 7 : Update: Choose $\sigma_{k+1} \in [\sigma_l, \sigma_r], x_{k+1} = x_k + \alpha_k d_k, k = k + 1$,

$$H_{k+1} = H_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k}, \tag{19}$$

where $y_k = \nabla f(x_{k+1}) - \nabla f(x_k),\ s_k = x_{k+1} - x_k$, and go to step 1.

Remark 3.1 H_{k+1} can be obtained by iterative formula.

Remark 3.2 In this paper,the penalty function and filter are avoided.

Remark 3.3 When the solution of $Q(x_k, H_k, \sigma_k)$ is unacceptable,it generates revised direction by solving a system of linear equation,which takes full advantage of good property of d_k.

4 Global convergence of algorithm

To obtain the global convergent properties, we always assume that the following assumptions hold.

Assumptions

A4.1 The objective function f and the constraint functions $g_j (j \in I)$ are twice continuously differentiable.

A4.2 For any $x \in R^n$,the vectors $\{\nabla g_j(x), j \in I(x)\}$ are linerly independent,where $I(x) = \{j \in I | g_j(x) = \Phi(x)\}$.

A4.3 The iterate $\{x_k\}$ and the multiplier λ^k remain in a closed,bounded convex subsets $S \subset R^n$.

A4.4 There exists two constants $0<a \leq b$ such that $a\|d\|^2 \leq d^T H_k d \leq b\|d\|^2$, for all k, for all $d \in R^n$.

A4.5 f(x) is bounded above on the level set $\ell = \{x | f(x) \leq f(x_0)\}$

A4.6 In some neighborhood Ω of ℓ ,f is continuously differentiable,and its gradient $\nabla f(x)$ is Lipschitz continuous, namely, there exists a constant ℓ such that

$$\|\nabla f(x) - \nabla f(x_k)\| \le \ell \|x - x_k\| \tag{20}$$

Lemma 4.1 For any iterate k ,the index i_k defined in step 1 is finite,which means that the inner loop S1.1-S1.3 terminates in finite number of times.

Proof Suppose by a contradiction that Algorithm 3.1 will run infinitely between step 1.2 and step 1.3,so we have

$$det(B_{k,i}^T B_{k,i}) < \frac{1}{2}\varepsilon_0, \tag{21}$$

By the definition of $L_{k,i}$,we can see that $L_{k,i+1} \subseteq L_{k,i}$.And there are only finite possible subset of I,so we have $L_{k,i+1} \equiv L_{k,i}$,for large enough i. We denote it by L_k^*,now letting $i \to \infty$,then we obtain

$$det(B_{L_k^*}^T B_{L_k^*}) = 0 \quad and \quad L_k^* = I(x_k), \tag{22}$$

which contradicts the Assumption A4.2.

Lemma 4.2 If $d_k \neq 0$,then it holds

$\nabla f(x_k)^T d_k < 0, \quad \nabla f(x_k)^T q_k \le \frac{1}{2}\rho_k^2 < 0.$

$\nabla g_j(x_k)^T d_k = 0, \quad \nabla g_j(x_k)^T q_k \le -\frac{\rho_k^2}{1+2|e^T \pi_k|} < 0.$

The proof is similar to Lemma 2.2 [7].

From Lemma 4.2 we know the search in step 6 will be successful.

By the above statement ,we see that Algorithm 3.1 is implementable.Now,we turn to prove the global convergence of Algorithm 3.1.

Theorem 4.1 [7].Assume that MFCQ is satisfied at $x_0 \in R^n$.Let $\sigma_l > 0$ and $F = \{x | g(x) \le 0\}$,then there exists a neighbor $N(x_0)$ such that

(1)MFCQ is satisfied at any point in $N(x_0)$.

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

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

$$sup\{\sum_{j=1}^{m} \mu_j : H \in \Sigma, x \in N(x_0), \sigma \in [\sigma_l, \sigma_r]\} < \infty, \tag{23}$$

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

Lemma 4.3 If $\nabla f(x_k) d_k \le 0$ for each k, then for the iterates generated by step4 or step6, we have $f(x_k) \le C_k \le A_k$ for each k. Moreover, if $\nabla f(x_k) d_k < 0$ and $f(x_k)$ is bounded from below, then there exists α_k satisfying Armijo conditions of the line search update.

Proof: Defining $D_k : R \to R$ by

$$D_k(t) = \frac{tC_{k-1} + f_k}{t+1}, \tag{24}$$

we have

$$D_k^{'}(t) = \frac{C_{k-1} - f_k}{(t+1)^2}, \tag{25}$$

Since $\nabla f(x_k)d_k \leq 0$, it follows from (14) that $f_k \leq C_{k-1}$, which implies that $D_k^{'} \geq 0$ for all $t \geq 0$, Hence, D_k is nondecreasing, and $f_k = D_k(0) \leq D_k(k)$ for all $t \geq 0$. in particular, taking $t = \eta_{k-1}Q_{k-1}$ gives

$$f_k = D_k(0) \leq D_k(\eta_{k-1}Q_{k-1}) = C_k. \tag{26}$$

the upper bound $C_k \leq A_k$ is proved by induction. For $k = 0$ this holds by initialization $C_0 = f(x_0)$. Now assume that $C_j \leq A_j$ for all $0 \leq j < k$. by (14), the initialization $Q_0 = 1$, and the fact that $\eta_k \in [0, 1]$, we have

$$Q_{j+1} = 1 + \sum_{i=0}^{j} \prod_{m=0}^{i} \eta_{j-m} \leq j + 2. \tag{27}$$

Since D_k is monotone nondecreasing, (27) imply that

$$C_k = D_k(\eta_{k-1}Q_{k-1}) = D_k(Q_k - 1) \leq D_k(k). \tag{28}$$

By the induction step,

$$D_k(k) = \frac{kC_{k-1} + f_k}{k+1} \leq \frac{kA_{k-1} + f_k}{k+1} = A_k, \tag{29}$$

Relation (28) and (29),we have the upper bound of C_k. In fact, when $\alpha_k = 0$, $f(x_k + \alpha_k d_k) = f(x_k)$ there must exist a sufficient small α_k

$$f(x_k + \alpha_k d_k) \leq f_k + \delta\alpha_k \nabla f(x_k)d_k, \tag{30}$$

because of $\nabla f(x_k)d_k \leq 0$ and $0 < \delta < 1$. what is more, $f(x_k) \leq C(x_k)$, So we have

$$f(x_k + \alpha_k d_k) \leq C_k + \delta\alpha_k \nabla f(x_k)d_k. \tag{31}$$

Lemma 4.4 As we are known, α_k is generated by step4 or step6, so α_k satisfies (1), If assumption A4.5 and A4.6 holds, we have

$$\alpha_k \geq \frac{2\beta(1-\delta)|g_k^T d_k|}{\ell\|d_k\|^2}. \tag{32}$$

Proof: by algorithm 3.1 we have $\alpha_k \leq 1$, and $\alpha_k = \bar{\alpha}_k\beta^{h_k}$, h_k is the smallest integer such that (1) holds. since we have

$$f(x_k + \frac{\alpha_k}{\beta}d_k) > C_k + \delta\frac{\alpha_k}{\beta}g_k^T d_k \geq f(x_k) + \delta\frac{\alpha_k}{\beta}g_k^T d_k \tag{33}$$

while ∇f is Lipschitz continuous,

$f(x_k + \alpha d_k) - f(x_k) = \alpha g_k^T d_k + \int_0^\alpha [\nabla f(x_k + td_k) - \nabla f(x_k)]^T d_k dt$
$\leq \alpha g_k^T d_k + \int_0^\alpha tL\|d_k\|^2 dt$
$= \alpha g_k^T d_k + \frac{1}{2} L\alpha^2 \|d_k\|^2.$
Combining this with (33) we can proves (32) holds.

Theorem 4.2 Suppose the Algorithm 3.1 is applied to problem (P), and the AssumptionsA4.1-A4.4 hold. Let $\{x_k\}$ be the sequence of iterates produced by the algorithm. then there are two following possible cases:

(A) The iteration terminates at a KKT point.

(B) Any accumulation point of $\{x_k\}$ is a KKT point of problem (P).

Proof

(A) It is evident according to the above statement.

(B) By the construction of Algorithm 3.1, there are two cycles between step1 and step 7, one generates $\{x_k\}$ with the form $x_{k+1} = x_k + \alpha_k d_k$, the other generates with the form $x_{k+1} = x_k + \alpha_k q_k$. We prove that the claim according the two cases.

Case I: Suppose there are infinite point gotten by the relation $x_{k+1} = x_k + \alpha_k d_k$, by Assumption A4.3, there must exists a point x^* such that $x_k \to x^*(k \in K)$,where K is a infinite index set.Also, by the algorithm, we can obtain that x^* is a feasible point and $\Psi^0(x_k, \sigma_k) \to 0, (k \in K)$.

Suppose x^* is not a KKT point, let $K_1 = \{k \in K | \nabla f(x_k)^T d_k > -\frac{1}{2} d_k^T H_k d_k\} \subset K$.

(i) K_1 is an infinite index set.

If $\lim\limits_{k \in K_1, k \to \infty} \|d_k\| = 0$,then it is easy to see that x^* is a KKT point.It is a contradiction.So,without loss of generality,we suppose that $\|d_k\| \geq \varepsilon$ for $k \in K_1$.

By the algorithm and Assumption A4.4,we can assume $\exists k_0$, for $k > k_0, k \in K_1$, it holds

$$g(x_k) \leq \frac{a\varepsilon^2}{2M} \leq \frac{a\|d_k\|^2}{2M} \leq \frac{d_k^T H_k d_k}{2M}. \tag{34}$$

While by KKT condition of the problem (P), we have $\nabla f(x_k) + h_k d_k + B_k \lambda_k = 0, B_k = (\nabla g_j(x_k),\ j \in L_k)$.

Together with (34),we obtain that for all $k \in K_1, k > k_0$, it holds

$\nabla f(x_k)^T d_k = -d_k^T H_k d_k - d_k^t B_k \lambda^k = -d_k^T H_k d_k - (\lambda^k)^T g(x_k)$
$\leq -d_k^T H_k d_k + \|\lambda_k\|_\infty g(x_k) \leq M g(x_k) - d_k^T H_k d_k$
$\leq -\frac{1}{2} d_k^T H_k d_k$

which contradicts the definition of K_1.

(ii) K_1 is a finite index set.

That means it hold $\nabla f(x_k)^T d_k \leq -\frac{1}{2} d_k^T H_k d_k$ for large enough k.

We have

$f_{k+1} \le C_k + \delta\alpha_k \nabla f(x_k)^T d_k \le C_k - \frac{1}{2}\delta\alpha_k d_k^T H_k d_k$
$\le C_k - \frac{\delta\beta(1-\delta)|\nabla f(x_k)^T d_k| d_k^T H_k d_k}{\ell \|d_k\|^2}$
$\le C_k - \frac{\delta\beta(1-\delta)(d_k^T H_k d_k)^2}{2\ell\|d_k\|^2} \le C_k - \frac{\delta\beta(1-\delta)(a\|d_k\|^2)^2}{2\ell\|d_k\|^2}$

$= C_k - \frac{a^2\delta\beta(1-\delta)}{2\ell}\|d_k\|^2$

$C_{k+1} = \frac{\eta_k Q_k C_k + f_{k+1}}{Q_{k+1}} \le C_{k+1} = \frac{\eta_k Q_k C_k + C_k - \frac{a^2\delta\beta(1-\delta)}{2\ell}\|d_k\|^2}{Q_{k+1}}$
$= C_k - \frac{a^2\delta\beta(1-\delta)}{2\ell}\frac{\|d_k\|^2}{Q_{k+1}}$

Since f is bounded from below and $f_k \le C_k$ for all k, we conclude that C_k is bounded from below.Then

$$\sum_{k=i_0}^{\infty} \frac{\|d_k\|^2}{Q_{k+1}} < \infty. \tag{35}$$

because of $\eta_{max} < 1$,then by (27) we have

$$Q_{k+1} = 1 + \sum_{j=0}^{k}\prod_{i=0}^{j}\eta_{k-i} \le 1 + \sum_{j=0}^{k}\eta_{max}^{j+1} \le \sum_{j=0}^{\infty}\eta_{max}^{j} = \frac{1}{1-\eta_{max}}. \tag{36}$$

then

$$\lim_{k\to\infty}\|d_k\| = 0. \tag{37}$$

Case II: Suppoe there are infinite points gotten from the relation $x_{k+1} = x_k + \alpha_k q_k$.

Suppose also by contradiction that $\|d_k\| > \varepsilon, k \in K$.

By Lemma 4.2,we have $\nabla f(x_k)^T q_k \le \frac{1}{2}\rho_k^2 < 0$.Then we have

$f_{k+1} \le C_k + \delta\alpha_k \nabla f(x_k)^T q_k \le C_k - \frac{1}{2}\delta\alpha_k\rho_k^2 = C_k - \frac{1}{2}\delta\alpha_k(\nabla f(x_k)^T d_k)^2$

$\le C_k - \frac{\delta\beta(1-\delta)|\nabla f(x_k)^T d_k|(\nabla f(x_k)^T d_k)^2}{\ell\|d_k\|^2} \le C_k + \frac{\delta\beta(1-\delta)(\nabla f(x_k)^T d_k)^3}{\ell\|d_k\|^2}$

$\le C_k - \frac{\delta\beta(1-\delta)}{8\ell\|d_k\|^2}(d^T H_k d_k)^3 \le C_k - \frac{\delta\beta(1-\delta)}{8\ell\|d_k\|^2}(a\|d_k\|^2)^3 = C_k - \frac{a^3\delta\beta(1-\delta)}{8\ell}\|d_k\|^4$

$C_{k+1} = \frac{\eta_k Q_k C_k + f_{k+1}}{Q_{k+1}} \le C_{k+1} = \frac{\eta_k Q_k C_k + C_k - \frac{a^3\delta\beta(1-\delta)}{8\ell}\|d_k\|^4}{Q_{k+1}} = C_k - \frac{a^3\delta\beta(1-\delta)}{8\ell}\frac{\|d_k\|^4}{Q_{k+1}}$

Since f is bounded from below and $f_k \le C_k$ for all k, we conclude that C_k is bounded from below.Then

$$\sum_{k=i_0}^{\infty} \frac{\|d_k\|^4}{Q_{k+1}} < \infty. \tag{38}$$

because of $\eta_{max} < 1$,then by (27) we have

$$Q_{k+1} = 1 + \sum_{j=0}^{k} \prod_{i=0}^{j} \eta_{k-i} \leq 1 + \sum_{j=0}^{k} \eta_{max}^{j+1} \leq \sum_{j=0}^{\infty} \eta_{max}^{j} = \frac{1}{1-\eta_{max}}. \tag{39}$$

then

$$\lim_{k\to\infty} \|d_k\| = 0. \tag{40}$$

ACKNOWLEDGEMENTS. This work is supported by the National Science Fpundation of Hebei Province (No. A2014201033) and the Key Research Foundation of Education Bureau of Hebei Province (No. ZD2015069)

References

[1] P.T. Boggs, J.W. Tolle, P. Wang, On the local convergence of quasi-newton methods for constrained optimization, SIAM J. Control Optim. 20 (1982) 161C171.

[2] J.F. Bonnons, E.R. Painer, A.L. Titts, J.L. Zhou, Avoiding the Maratos effect by means of nonmontone linesearch, inequality constrained problems-feasible iterates, SIAM J. Numer. Anal. 29 (1992) 1187C1202.

[3] S.P. Han, Superlinearly convergence variable metric algorithm for general nonlinear programming problems, Math. Program. 11 (1976) 263C282. uses differentiable exact penalty function[J].1986.

[4] M.J.D. Powell, A fast algorithm for nonlinear constrained optimization calculations, in: G.A. Waston (Ed.), Numerical Analysis, Springer-Verlag, Berlin, 1978, pp. 144C157.

[5] M.J.D. Powell, Variable metric methods for constrained optimization, in: Bachen et al. (Eds.), Mathematical Programming C The State of Art, Springer-Verlag, Berlin, 1982.

[6] J.V. Burke, S.P. Han, A robust SQP method, Math. Program. 43 (1989) 277C303.

[7] G.L. Zhou, A modified SQP method and its global convergence, J. Global Optim. 11 (1997) 193C205.

[8] J.L. Zhang, X.S. Zhang, A modified SQP method with nonmonotone line-search technique, J. Global Optim. 21 (2001) 201C218.

[9] Z.B. Zhu, K.C. Zhang, J.B. Jian, An improved SQP algorithm for inequality constrained optimization, Math. Meth. Oper. Res 58 (2003) 271C282.

[10] Che Jianren, Su Ke. A modified SQP method and its global convergence,Applied Mathematics and Computation. 2, (2007), 945-951.

[11] Chamberlain, R.M, Powell, M.J.D, Lemarechal, C, and Pedersen, H.C,The watchdog technique for forcing convergence in algorithm for constrained optimization,Math. Program. Stud.16, 1982, 1-17.

[12] L. Grippo, F. Lampariello, and S. Ludidi, A nonmonotone line search technique for Newtons method, SIAM J. Numer. Anal. 23, 1986, 707-716.

[13] Hongchao Zhang and William W. Hager. A nonmonotone line search technique and its application to unconstrained optimization, SIAM J. Optim. 14, 2004, 1043-1056.

[14] L. Grippo, F. Lampariello, and S. Lucidi, A class of nonmonotone stabilization methods in unconstrained optimization, numer. Math., 59, 1991, 779-805.

Received: June, 2016