BULL. AUSTRAL. MATH. SOC. 65H05,90C30
VOL. 74 (2006) [69–83]

AN ACTIVE SET SEQUENTIAL QUADRATIC PROGRAMMING ALGORITHM FOR NONLINEAR OPTIMISATION

QING-JIE HU, YUN-HAI XIAO AND Y. CHEN

In this paper, we have proposed an active set feasible sequential quadratic programming algorithm for nonlinear inequality constraints optimization problems. At each iteration of the proposed algorithm, a feasible direction of descent is obtained by solving a reduced quadratic programming subproblem. To overcome the Maratos effect, a higher-order correction direction is obtained by solving a reduced least square problem. The algorithm is proved to be globally convergent and superlinearly convergent under some mild conditions *without strict complementarity.*

1. INTRODUCTION

Consider the following nonlinear inequality constrained optimisation:

$$\text{(1.1)} \qquad \text{(P)} \qquad \begin{array}{ll} \min & f_0(x) \\ \text{such that} & f_j(x) \leqslant 0, \quad j \in I = \{1, 2, \ldots, m\}, \end{array}$$

where $m > 0$ and the functions f_0, $f_j (j \in I) : R^n \to \mathrm{R}$ are all continuously differentiable.

It is well known that the method of feasible directions is one of the important methods for solving the problem (P). Method of feasible directions was originally developed by Zoutendijk [**26**]. Since then, Topkis and Veinott [**23**] made a modification to Zoutendijk's method, which assured their algorithm converges to a Fritz-John point. Later, Pironneau and Polak [**18, 19**] amended Topkis and Veinott's method so that the normalisation set was replaced by adding the term $\|d\|^2/2$ into the objective function of the subproblem. Their algorithms were proved to be globally convergent and had a linear convergence rate. Cawood and Kostreva [**3, 4**] proposed a norm-relaxed method of feasible directions by generalising the Pironneau and Polak's method of feasible directions. They showed that this method is globally convergent under mild assumptions. Since early feasible direction methods only used the information of first derivatives, all method of feasible directions above have at most a linear convergence rate.

Sequential Quadratic Programming algorithms are widely acknowledged to be among the most successful algorithms for solving (P)(See[**21, 22, 24, 10, 20, 6, 17, 13, 12, 11,**

Received 23rd February, 2006

8]). For an excellent recent survey of sequential quadratic programming algorithms, and the theory behind them, see [**1**]. In [**17**], a variation on the standard sequential quadratic programming algorithms for solving (P) (which is called a feasible sequential quadratic programming algorithm) is proposed. It generates iterations lying within the feasible set of (P). Feasible sequential quadratic programming is proved to be globally convergent and superlinearly convergent under some mild assumptions. However, at each iteration, these algorithms require the solution of two QP subproblems and a linear least squares problem, or two linear systems of equations and a linear least squares problem. Clearly, their computational cost per single iteration is relatively high.

Recently, another type of feasible sequential quadratic programming algorithm [**15**] has been proposed. In this algorithm, the following QP subproblem is considered, for an iteration point x^k:

$$(1.2) \qquad \text{(QP)} \qquad \begin{array}{ll} \min\limits_{(z,d)} & z + \frac{1}{2} d^T H_k d \\ \text{such that} & \nabla f_0(x^k)^T d \leqslant z, \\ & f_j(x^k) + \nabla f_j(x^k)^T d \leqslant \sigma_k z, \quad j \in I, \end{array}$$

where H_k is a symmetric positive definite matrix and an approximation of the Lagrangian Hessian matrix for (P), and σ_k is a positive parameter. In [**15**], it is necessary to solve an equality constrained QP subproblem to update the parameter σ_k such that $\sigma_k = O(\|d_0^{k-1}\|^2)$. On the other hand, in order to accept the unit step size, a correction direction is obtained by solving another equality constrained QP subproblem. Furthermore, the algorithm is proved to be locally two-step superlinearly convergent under certain conditions. Reference [**14**] proposed a similar algorithm to solve the problem (P). It also needs to solve two QP subproblems with inequality constraints, and like [**15**], it is proved to be locally two-step superlinearly convergent. Furthermore, it is required that σ_k approaches zero fast enough as $d^k \to 0$, that is, $\sigma_k = o(\|d^k\|)$. In [**25**], Zhu proposed a similar algorithm. In his algorithm a feasible direction of descent is obtained by solving the QP subproblem (1.2). In order to avoid Marotos effect, a high-order revised direction is computed by solving a reduced linear system. Furthermore, it is proved to be globally convergent and superlinearly convergent under some certain conditions. Unlike [**14, 15**], no auxiliary problem need be computed to update σ_k. On the other hand, to obtain locally superlinear convergence, for the above-mentioned algorithms, the strict complementary condition is necessary.

In this paper, we have proposed an active set feasible sequential quadratic programming algorithm for nonlinear inequality constraints optimisation problems. At each iteration of the proposed algorithm, a feasible direction of descent is obtained by solving a reduced quadratic programming subproblem. To overcome the Maratos effect, a higher-order correction direction is obtained by solving a reduced least square problem. The algorithm is proved to be globally convergent and superlinearly convergent under some mild conditions *without strict complementarity.*

The remainder of this paper is organised as follows. The proposed algorithm is stated in Section 2. In Section 3 and Section 4, under some mild assumptions, we show that this algorithm is globally convergent and locally superlinear convergent, respectively. Finally, we give concluding remarks about the proposed algorithm.

2. Description of algorithm

We denote the feasible set X of (P) by

$$X = \{x \in R^n : f_i(x) \leqslant 0, i \in I\},$$

and, for a feasible point $x \in X$, define the active set by

$$I(x) = \{i \in I : f_i(x) = 0\}.$$

In this paper, we suppose that the feasible set X is not empty and the following basic hypothesis holds.

Assumption A_1. The gradient vectors $\{\nabla f_j(x), j \in I(x)\}$ are linearly independent for each feasible point $x \in X$.

For $x \in X$, we now give the "guessing" of the active set $I(x)$ in [**7**]:

$$A(x;\varepsilon) = \Big\{i : f_i(x) + \varepsilon\rho\big(x, \lambda(x)\big) \geqslant 0\Big\},$$

where ε is a nonnegative parameter and $\rho\big(x, \lambda(x)\big) = \sqrt{\| \Phi(x, \lambda) \|}$ with

$$\Phi\big(x, \lambda(x)\big) = \begin{pmatrix} \nabla_x L\big(x, \lambda(x)\big) \\ \min\{-f(x), \lambda(x)\} \end{pmatrix}, \qquad f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \cdots \\ f_m(x) \end{pmatrix},$$

$$\lambda(x) = -\Big(\nabla f(x)^T \nabla f(x) + \mathrm{diag}\big(f_i(x)\big)^2\Big)^{-1} \nabla f(x)^T \nabla f_0(x) \quad \text{(See [16])},$$
$$\nabla f(x) = \big(\nabla f_i(x), i \in I\big).$$

It is obvious that (x^*, λ^*) is a KKT point of (P) if and only if $\Phi(x^*, \lambda^*) = 0$ or $\rho(x^*, \lambda^*) = 0$. Facchinei et al [**7**] showed that if the second order sufficient condition and the Mangassarian-Fromovotz Constraint Qualification hold, then for any $\varepsilon > 0$, when x is sufficient close to x^*, $A(x;\varepsilon)$ is an exact identification of $I(x^*)$.

The following algorithm is proposed for solving (P).

ALGORITHM

Parameters r, $r_j > 0 (j \in I)$, $\tau \in (2, 3)$, $\sigma \in (0, 1)$, $\sigma_1 \in (0, 1)$, $\nu \in (0, 1)$, $\beta \in (0, 1)$, $\alpha \in \big(0, (1/2)\big)$.

Data Choose an initial feasible point $x^1 \in X$, a symmetric positive matrix H_1, $\sigma_1 > 0$ and $\varepsilon^0 \geqslant 0$. Set $k = 1$.

Step 1 Set $\varepsilon = \varepsilon^{k-1}$.

Step 2 Set $A^k(\varepsilon) = A(x^k, \varepsilon)$. If $\nabla f_{A^k(\varepsilon)}(x^k)$ is not of full rank, then set $\varepsilon := \sigma\varepsilon$, go to step 2. (Where $\nabla f_{A^k(\varepsilon)}(x^k) = \big(\nabla f_j(x^k),\ j \in A^k(\varepsilon)\big)$.)

Step 3 Set $\varepsilon^k = \varepsilon,\ A^k = A^k(\varepsilon)$.

Step 4 (Compute the search direction) For the current iteration point x^k, solve

$$(2.1)\qquad \text{(NQP)}\qquad \begin{aligned} &\min && rz + \frac{1}{2}d^T H_k d \\ &\text{such that} && \nabla f(x^k)^T d \leqslant rz, \\ &&& f_j(x^k) + \nabla f_j(x^k)^T d \leqslant r_j \sigma_k z, \quad j \in A^k, \end{aligned}$$

to obtain an optimal solution (z_k, d^k). Let $(u_0^k, u_{A^k}^k)$ be corresponding KKT multipliers. If $d^k = 0$, then x^k is a KKT point for (P) and stop; otherwise go to Step 5.

Step 5 Compute the higher-order direction $\widetilde{d}^k$ by solving the least square problem:

$$(2.2)\qquad \begin{aligned} \text{(LS)}\quad &\min && \frac{1}{2}\|d\|_{H_k}^2 \\ &\text{such that} && f_j(x^k + d^k) + \nabla f_j(x^k)^T d = -(1-\rho)\|d^k\|^\tau + \rho r_j \sigma_k^\nu z_k \|d^k\| \\ &&& + f_j(x^k) + \nabla f_j(x^k)^T d^k - r_j \sigma_k z_k, \quad j \in A^k, \end{aligned}$$

where

$$\rho = \begin{cases} 0, & \text{if } \|d^k\|^2 \geqslant -r_j \sigma_k z_k; \\ 1, & \text{if } \|d^k\|^2 < -r_j \sigma_k z_k. \end{cases}$$

If $\|\widetilde{d}^k\| > \|d^k\|$, set $\widetilde{d}^k = 0$.

Step 6 (Do curve search) Compute the step size λ_k, which is the first number λ of the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying

$$(2.3)\qquad f_0(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) \leqslant f_0(x^k) + \alpha\lambda \nabla f_0(x^k)^T d^k,$$

$$(2.4)\qquad f_j(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) \leqslant 0, \quad \forall j \in I.$$

Step 7 Set a new iteration point by $x^{k+1} = x^k + \lambda_k d^k + \lambda_k^2 \widetilde{d}^k$, $\sigma_{k+1} = \min\{\sigma_1, \|d^k\|^\nu\}$.

Step 8 Compute a new symmetric positive definite matrix H_{k+1}, set $k := k+1$, and go back to Step 1.

We now show that the proposed algorithm is well defined.

Lemma 2.1. *Let $x^k \in X$ and suppose that Assumption A_1 holds. Then Step 2 in the proposed algorithm can be finished in a finite number of computations.*

The proof is similar to the one of Lemma 1.1 and Lemma 2.8 in [**9**].

Lemma 2.2. *Suppose that H_k is symmetric positive definite, parameters r, $r_j(j \in A^k)$ are all positive and $\sigma_k \geqslant 0$. Then* (NQP) *always has a unique optimal solution.*

The proof is similar to the one of [**25**, Lemma 1].

Lemma 2.3. *Suppose that H_k is symmetric positive definite, then* (LS) *always has a unique optimal solution.*

This proof is easy by using the positive definiteness of H_k and the full rank property of $g_{A^k}(x^k)$.

Lemma 2.4. *Suppose that the conditions in Lemma 2.2 are satisfied and (z_k, d^k) is an optimal solution of* (2.1). *Then*

(i) $rz_k + 1/2(d^k)^T H_k d^k \leqslant 0$ and $z_k \leqslant 0$;

(ii) $z_k = 0 \Longleftrightarrow d^k = 0 \Longleftrightarrow x^k$ *is a KKT point for* (P);

(iii) $z_k < 0 \Longrightarrow d^k$ *is a feasible direction of descent for* (P) *at point x^k.*

The proof is similar to the one of [**25**, Lemma 2].

Lemma 2.5. *The line search in Step 6 of the proposed algorithm yields a stepsize $\lambda_k = \beta^j$ for some finite $j = j(k)$.*

It is not difficult to finish the proof of this lemma.

3. Global Convergence

In this section, we analyse the global convergence of the proposed algorithm. The following assumptions are necessary.

Assumption A_2. The sequence $\{x^k\}$, which is generated by the proposed algorithm, is bounded.

Assumption A_3. There exist a, $b > 0$ such that $a\|d\|^2 \leqslant d^T H_k d \leqslant b\|d\|^2$ for all k and all $d \in R^n$.

We suppose that x^* is a given accumulation point of $\{x^k\}$. In view of A^k and J_k being a subset of the finite and fixed set I, respectively, there exist an infinite index set K such that

$$\lim_{k\in K} x^k = x^*, \quad A^k \equiv A, \quad J_k \equiv J, \ \ \forall k \in K, \tag{3.1}$$

where

$$J_k = \{j \in A^k : f_j(x^k) + \nabla f_j(x^k)^T d^k = r_j \sigma_k z_k\}.$$

Lemma 3.1. *Suppose that Assumptions A_2 and A_3 hold. Then the sequences $\{d^k : k \in K\}, \{z_k : k \in K\}$ and $\{\tilde{d}^k : k \in K\}$ are all bounded.*

PROOF: Firstly, from $\nabla f_0(x^k) \to \nabla f_0(x^*)$, $k \in K$, there exists a constant $c_0 > 0$ such that $\| \nabla f_0(x^k) \| \leqslant c_0$, $\forall k \in K$. Furthermore, from Lemma 2.4, formulas (2.1) and Assumption A_3, one has

$$\begin{aligned} 0 \geqslant r z_k + \frac{1}{2}(d^k)^T H_k d^k &\geqslant \nabla f_0(x^k)^T d^k + \frac{a}{2}\|d^k\|^2 \\ &\geqslant -\|\nabla f_0(x^k)\| \cdot \|d^k\| + \frac{a}{2}\|d^k\|^2 \\ &\geqslant -c_0\|d^k\| + \frac{a}{2}\|d^k\|^2, \qquad k \in K. \end{aligned}$$

This shows that $\{d^k : k \in K\}$ is bounded.

Secondly, the boundedness of $\{z_k : k \in K\}$ follows from the boundedness of $\{d^k : k \in K\}$ as well as the following inequalities

$$0 \geqslant z_k \geqslant \frac{1}{r}\nabla f_0(x^k)^T d^k \geqslant -\frac{1}{r}\|\nabla f_0(x^k)\| \cdot \|d^k\| \geqslant -\frac{c_0}{r}\|d^k\|, \quad k \in K. \tag{3.2}$$

Lastly, the boundedness of $\{\widetilde{d}^k : k \in K\}$ follows immediately from the boundedness of $\{d^k : k \in K\}$. ∎

We know that the KKT conditions of (NQP) can be formulated as follows:

$$H_k d^k + u_0^k \nabla f(x^k) + \sum_{j \in A^k} u_j^k \nabla f_j(x^k) = 0, \quad u_{A^k}^k = (u_j^k, j \in A^k), \tag{3.3}$$

$$r = r u_0^k + \sum_{j \in A^k} u_j^k r_j \sigma_k, \quad u_j^k \geqslant 0, j \in A^k, u_0^k \geqslant 0, \tag{3.4}$$

$$0 \leqslant u_0^k \perp (r z_k - \nabla f_0(x^k)^T d^k) \geqslant 0, \tag{3.5}$$

$$0 \leqslant u_j^k \perp (r_j \sigma_k z_k - f_j(x^k) - \nabla f_j(x^k)^T d^k) \geqslant 0, \quad j \in A^k, \tag{3.6}$$

where the notation $x \perp y$ means $x^T y = 0$.

LEMMA 3.2.

(i) *The multiplier sequence $\{u_0^k\}_{k=0}^{\infty}$ is bounded.*

(ii) *Let multiplier vector $u^k = (u_{A^k(\varepsilon)}^k, 0_{I \setminus A^k(\varepsilon)}) = (u_{J_k}^k, 0_{I \setminus J_k})$. If $\lim_{k \in K} x^k = x^*$ and $\lim_{k \in K} d^k = 0$, then $\{u^k : k \in K\}$ is bounded under Assumptions A_1, A_2 and A_3.*

PROOF: (i) From the KKT condition (3.4), one has

$$r = r u_0^k + \sum_{j \in A^k} u_j^k r_j \sigma_k \geqslant r u_0^k, \quad 0 \leqslant u_0^k \leqslant 1.$$

(ii) Suppose by contradiction that the given statement is not true, then there exists an infinite index $K' \subseteq K$ such that $\| u^k \| = \| u_J^k \| \to \infty$, $k \in K'$. Therefore, dividing (3.3) by $\|u_J^k\|$ to yield

$$\frac{1}{\|u_J^k\|} H_k d^k + \frac{u_0^k}{\| u_J^k \|} \nabla f_0(x^k) + \sum_{j \in J} \frac{u_j^k}{\| u_J^k \|} \nabla f_j(x^k) = 0. \tag{3.7}$$

Noting that the sequence $\{u_J^k/\|u_J^k\| : k \in K'\}$ is bounded with norm one, we can assume without loss of generality that

$$\frac{u_j^k}{\|u_J^k\|} \to \overline{u}_j, \quad k \in K', \quad j \in J, \quad 0 \leqslant (\overline{u}_j,\ j \in J) \neq 0. \tag{3.8}$$

Thus, passing to the limit $k \in K'$ and $k \to \infty$ in (3.7), and taking into account Assumption A_3 as well as the given conditions, we have

$$\sum_{j\in J} \overline{u}_j \nabla f_j(x^*) = 0. \tag{3.9}$$

On the other hand, from the given conditions, one has $J \subseteq I(x^*)$, so we can construct a contradiction from (3.8), (3.9) and Assumption A_1. Therefore the boundedness of $\{u^k : k \in K\}$ is shown. ∎

Lemma 3.3. *Suppose that $\{x^k\}$ is a sequence generated by the proposed algorithm, $\lim_{k\in K} x^k = x^*$ and $\lim_{k\in K} d^k = 0$ hold. Then x^* is a KKT point of* (P).

Proof: Taking into account the boundedness of $\{u_0^k : k \in K\}$, $\{u^k : k \in K\}$ and $\{\sigma_k\}$, we can assume without loss of generality that

$$u^k = (u_j^k, j \in I) \to u^* = (u_j^*, j \in I), \quad u_0^k \to u_0^*, \quad \sigma_k \to \sigma_*, \quad k \in K.$$

Moreover, the fact $\lim_{k\in K} x^k = x^*$ and $\lim_{k\in K} d^k = 0$ implies $\lim_{k\in K} z^k = 0$. Thus, passing to the limit $k \in K$ and $k \to \infty$ in (3.3)—(3.6) and the given conditions, we obtain

$$\begin{gathered} u_0^*\nabla f_0(x^*) + \sum_{j\in J} u_j^*\nabla f_j(x^*) = 0, \\ u_j^* f_j(x^*) = 0, \quad u_j^* \geqslant 0, \quad f_j(x^*) \leqslant 0, \quad j \in A, \\ r = ru_0^* + \sigma_* \sum_{j\in J} r_j u_j^*, \quad u_0^* \geqslant 0. \end{gathered} \tag{3.10}$$

From the third formula of (3), we know that $(u_0^*, u_J^*) \neq 0$, furthermore, $u_0^* > 0$ from Assumption A_1, which together with (3) shows that $\big(x^*, (u^*/u_0^*)\big)$ is a KKT pair of (P). The proof is complete. ∎

Based on Lemma 3.1, Lemma 3.2 and Lemma 3.3, we now can present the global convergence theorem of the proposed algorithm as follows.

Theorem 3.1. *Suppose that Assumptions A_1, A_2 and A_3 hold, then the proposed algorithm either stops at a KKT point x^k for problem* (P) *in a finite number of steps or generates an infinite sequence $\{x^k\}$ of points such that each accumulation point x^* is a KKT point for problem* (P). *Furthermore, there exists an infinite index K such that the sequence $\{u^k/u_0^k : k \in K\}$ converges to a KKT multiplier associated with x^* and $\lim_{k\in K} u_0^k > 0$.*

PROOF: The proof is similar to the one in [25, Theorem 1].

The first statement is obvious. Thus, assume that the proposed algorithm generated an infinite sequence $\{x^k\}$ and (3.1)holds. The cases $\sigma_* = 0$ and $\sigma_* > 0$ are considered, separately.

A. $\sigma_* = 0$. From step 7, there exists an infinite index set $K_1 \subseteq K$ such that $\lim_{k \in K_1} d^{k-1} = 0$. By step 7, it holds that

$$\| x^k - x^{k-1} \| \leqslant t_k \| d^{k-1} \| + t_k^2 \| \widetilde{d}^{k-1} \| \leqslant 2t^k \| d^{k-1} \| \longrightarrow 0, \quad k \in K_1.$$

So, the fact that $\lim_{k \in K_1} x^k = x^*$ implies that $\lim_{k \in K_1} x^{k-1} = x^*$. Moreover, we know that x^* is a KKT point for problem (P) from Lemma 3.3.

B. $\sigma_* > 0$. Obviously, it is sufficient to show $\lim_{k \in K} d^k = 0$. For this, we suppose by contradiction that $\lim_{k \in K} d^k \neq 0$, then there exist an infinite subset $K' \subseteq K$ and a constant $\Delta > 0$ such that $\| d^k \| \geqslant \Delta$ holds for all $k \in K'$. The remainder proof is divided into two steps as follows, and we always assume that $k \in K'$ is sufficient large and $\lambda > 0$ is sufficient small.

a. Show that there exists a constant $\overline{\lambda} > 0$ such that the step size $\lambda_k \geqslant \overline{\lambda}$ for $k \in K'$.

$$\begin{aligned}
f_0(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) - f_0(x^k) - \alpha\lambda \nabla f_0(x^k)^T d^k& \\
&= \nabla f_0(x^k)^T(\lambda d^k + \lambda^2 \widetilde{d}^k) - \alpha\lambda \nabla f_0(x^k)^T d^k + o(\lambda) \\
&\leqslant \lambda(1-\alpha)\nabla f_0(x_k)^T d^k + o(\lambda) \\
&\leqslant \lambda(1-\alpha) r z_k + o(\lambda) \\
&\leqslant -\frac{1}{2}\lambda(1-\alpha)(d^k)^T H_k d^k + o(\lambda) \\
&\leqslant -\frac{1}{2} a\lambda(1-\alpha)\|d^k\|^2 + o(\lambda) \\
&\leqslant -\frac{1}{2} a\lambda(1-\alpha)\Delta^2 + o(\lambda).
\end{aligned}$$

The last inequality above shows that (2.3) holds for $k \in K'$ and $\lambda > 0$ small enough.

Analyse (2.4): if $j \notin I(x^*)$, that is, $f_j(x^*) < 0$. from the continuity of $f_j(x)$ and the boundedness of $\{d^k : k \in K\}$ and $\{\widetilde{d}^k : k \in K\}$, we know $f_j(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) \leqslant 0$ holds for $k \in K'$ large enough and $\lambda > 0$ small enough.

Let $j \in I(x^*)$, that is, $f_j(x^*) = 0$. Then $j \in A^k$ from (3.1). So by Taylor expansion and (2.1), we have

$$\begin{aligned}
f_j(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) &= f_j(x^k) + \lambda \nabla f_j(x^k)^T d^k + o(\lambda) \\
&\leqslant f_j(x^k) + \lambda(r_j \sigma_k z_k - f_j(x^k)) + o(\lambda) \\
&= (1-\lambda) f_j(x^k) + \lambda r_j \sigma_k z_k + o(\lambda) \\
&\leqslant \lambda r_j \sigma_k z_k + o(\lambda).
\end{aligned}$$

Therefore, from (2.1) and Assumption A_3, we have

$$\begin{aligned} f_j(x^k + \lambda d^k + \lambda^2 \widetilde{d}^k) &\leqslant \lambda r_j \sigma^* \Big(-\frac{1}{2r} (d^k)^T H_k d^k \Big) + o(\lambda) \\ &\leqslant -\lambda r_j \sigma^* \frac{1}{2r} a \parallel d^k \parallel^2 + o(\lambda) \\ &\leqslant -\lambda r_j \sigma^* \frac{1}{2r} a \Delta^2 + o(\lambda). \end{aligned}$$

Thus, from the inequality above, we can conclude the search inequality (2.4) holds for $k \in K'$ large enough and $\lambda > 0$ small enough.

Summarising the analysis above, we conclude that there exists $\overline{\lambda} > 0$ such that $\lambda_k \geqslant \overline{\lambda}$ for all $k \in K'$.

b. Use $\lambda_k \geqslant \overline{\lambda} > 0$ to bring a contradiction. From (2.3), (2.1) and Assumption A_3, we have

$$\begin{aligned} f_0(x^{k+1}) &\leqslant f_0(x^k) + \alpha \lambda_k \nabla f_0(x^k)^T d^k \leqslant f_0(x^k) + \alpha \lambda_k r z_k \\ &\leqslant f_0(x^k) - \frac{1}{2} \alpha \lambda_k (d^k)^T H_k d^k \leqslant f_0(x^k) - \frac{1}{2} \alpha \lambda_k a \parallel d^k \parallel^2, \qquad \forall k. \end{aligned}$$

Therefore the sequence $\{f_0(x^k)\}$ is decreasing. Furthermore combining $\lim_{k \in K'} f_0(x^k) = f_0(x^*)$, one knows $\lim_{k \to \infty} f_0(x^k) = f_0(x^*)$. On the other hand, one also has

$$f_0(x^{k+1}) \leqslant f_0(x^k) - \frac{1}{2} a \alpha \overline{\lambda} \Delta^2, \qquad \forall k \in K'.$$

Passing to the limit $k \in K'$ and $k \to \infty$ in the inequality above, we have $-(1/2) a \alpha \overline{\lambda} \Delta^2 \geqslant 0$, which is a contradiction. So, $d^* = 0$. According to Lemma 3.3, x^* is a KKT point for problem (P). ∎

4. Rate of convergence

In this section, we shall analyse the convergence rate of the proposed algorithm. For this, the following further hypothesis is necessary.

ASSUMPTION A_4. (i) The functions $f_j(x) (j \in I)$ are all second-order continuously differentiable.

(ii) The sequence $\{x^k\}$ generated by the algorithm possesses an accumulation point x^* such that KKT pair (x^*, u^*) satisfies the strong second-order sufficiency conditions, that is,

$$(4.1) \qquad d^T \nabla^2_{xx} L(x^*, u^*) d > 0, \quad \forall d \in \Omega \stackrel{\text{def}}{=} \big\{ d \in R^n : d \neq 0, \nabla f_{I^+}(x^*)^T d = 0 \big\},$$

where

$$(4.2) \qquad L(x, u) = f_0(x) + \sum_{j \in I} u_j f_j(x), \qquad I^+ = \{ j \in I : u_j^* > 0 \}.$$

https://doi.org/10.1017/S0004972700035577 Published online by Cambridge University Press

Lemma 4.1.

(i) *Suppose that Assumptions* A_1, A_2 *hold. Then* $\lim_{k\to\infty} d^k = 0$, $\lim_{k\to\infty} \widetilde{d}^k = 0$, $\lim_{k\to\infty} z_k = 0$, $\lim_{k\to\infty} \sigma_k = 0$, *and* $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$.

(ii) *If Assumptions* A_1, A_2 *and* A_3 *are satisfied, then* $\lim_{k\to\infty} x^k = x^*$.

Proof: (i) Similar to the proof [**25**, Lemma 4.2]. We have that $\lim_{k\to\infty} d^k = 0$, $\lim_{k\to\infty} z^k = 0$. Furthermore, it is easy to conclude that $\lim_{k\to\infty} \widetilde{d}^k = 0$ from Step 5 and $\lim_{k\to\infty} \sigma^k = 0$ from Step 7.

On the other hand, from the conclusion above, one has

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = \lim_{k\to\infty} \|\lambda_k d^k + \lambda_k^2 \widetilde{d}^k\| \leqslant \lim_{k\to\infty} \left(\|d^k\| + \|\widetilde{d}^k\|\right) = 0.$$

(ii) Under Assumption A_4 (ii), one can conclude that the given limit point x^* is an isolated KKT point of (1.1)(See [**11**, Theorem 1.2.5]), therefore x^* is an isolated accumulation point of $\{x^k\}$ from Theorem 3.1, and this together with $\lim_{k\to\infty} \|x^{k+1}-x^k\| = 0$ shows that $\lim_{k\to\infty} x^k = x^*$. The proof is finished. ∎

Lemma 4.2. *Under all the above-mentioned assumptions, when k is sufficiently large, the matrix*

$$M_k \overset{\text{def}}{=} \begin{pmatrix} H_k & \nabla f_{A^k}(x^k) \\ \nabla f_{A^k}(x^k)^T & 0 \end{pmatrix}.$$

is nonsingular, furthermore, there exists a constant $C > 0$ *such that* $\|M_k^{-1}\| \leqslant C$.

The proof of Lemma 4.2 is similar to in [**12**, Lemma 2.2] or in [**11**, Lemma 2.2.2], and is omitted.

Lemma 4.3. *Suppose that Assumptions* A_1, A_2 *and* A_3 *hold. Then*

$$|z_k| = O\left(\|d^k\|\right), \quad \|\widetilde{d}^k\| = O\Big(\max\{\|d^k\|^2, -r_j\sigma_k z_k, j \in A_k\}\Big) = o\left(\|d^k\|\right), \tag{4.3}$$

$$I^+ \subseteq J_k \subseteq I(x^*) = A^k. \tag{4.4}$$

Proof: Firstly, from the first inequality constraint of (2.1), we have

$$-\left\|\nabla f_0(x^k)\right\| \cdot \|d^k\| \leqslant r z_k, \quad |z_k| \leqslant \frac{1}{r}\left\|\nabla f_0(x^k)\right\| \cdot \|d^k\|.$$

So it is not difficult to verify that $|z_k| = O\left(\|d^k\|\right)$.

Secondly, we shall show the second equation of (4.3). In view of (LS) being equivalent to solve the following system of linear equations:

$$\begin{pmatrix} H_k & \nabla f_{A^k}(x^k) \\ \nabla f_{A^k}(x^k)^T & 0 \end{pmatrix} \begin{pmatrix} \widetilde{d}^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} 0 \\ -(1-\rho)\|d^k\|^\tau e_{A^k} + \rho\sigma_k^\nu z_k \|d^k\| r_{A_k} + \widetilde{f}_{A^k} - \sigma_k z_k r_{A^k} \end{pmatrix}$$

https://doi.org/10.1017/S0004972700035577 Published online by Cambridge University Press

where

$$\widetilde{f}_{A^k} = -f_{A^k}(x^k + d^k) + f_{A^k}(x^k) + \nabla f_{A^k}(x^k)^T d^k,$$
$$r_{A_k} = (r_j, j \in A_k)^T,$$
$$e_{A^k} = (1, \dots, 1)^T \in R^{|A^k|}.$$

By using the Taylor expansion, one has that $\widetilde{f}_{A^k} = O(\|d^k\|^2)$. So it is not difficult to verify that

$$\|\widetilde{d}^k\| = O\Big(\max\{\|d^k\|^2, -r_j\sigma_k z_k, j \in A_k\}\Big) = o(\|d^k\|).$$

From Lemma 4.1, Lemma 4.2, $\tau \in (2,3)$ and $|z_k| = O(\|d^k\|)$.

To show the relationship (4.4), one first gets $J_k \subseteq I(x^*)$ from $\lim_{k\to\infty}(x^k, d^k, z_k, \sigma_k) = (0,0,0,0)$. From [7, Theorem 2.3 and Theorem 3.7], we know that $I(x^*) = A^k$ under Assumptions A_1, A_4(ii). Furthermore, one has $\lim_{k\to\infty} \lambda^k_{I^+} = \lambda^*_{I^+} > 0$ from Theorem 3.1, so $\lambda^k_{I^+} > 0$ and $I^+ \subseteq J_k$ holds for k large enough. The proof is complete. ∎

Lemma 4.4. *Suppose that Assumptions $A_1, A_2$and A_3 hold. Then $\{u^k/u^k_0\}$ converges to the KKT multiplier associated with x^* for* (P) *and* $\lim_{k\to\infty} u^k_0 = 1$.

Proof: Using $\lim_{k\to\infty}(x^k, d^k, \widetilde{d}^k, z_k) = (0,0,0,0)$ (see Lemma 4.1), and a proof similar to that of Lemma 3.2(ii), we can conclude that the entire sequence $\{u^k\}$ is bounded. Noting that the KKT multiplier associated with x^* for (P) is unique we can conclude that, for any infinite index subset K, the sequence $\{u^k/u^k_0 : k \in K\}$ possesses an accumulation point $\widetilde{u}^*$ such that $(x^*, \widetilde{u}^*)$ is a KKT pair for (P). Therefore, the sequence $\{u^k/u^k_0\}$ converges to a KKT multiplier associated with x^* for (P). Lastly, from (3.4), the boundedness of $\{u^k\}$ and Lemma 4.1(i), one has

$$\lim_{k\to\infty} u^k_0 = \lim_{k\to\infty}\left(1 - \frac{1}{\tau}\sum_{j\in J_k} u^k_j r_j \sigma_k\right) = 1.$$

The proof is complete. ∎

To ensure the step size $\lambda_k \equiv 1$ for k large enough, an additional assumption as follows is necessary.

Assumption A_5. Suppose that

$$\left\|(\nabla^2_{xx} L(x^k, (u^k_{J_k}/u^k_0) - H_k)d^k\right\| = o(\|d^k\|),$$

where

$$L\Big(x, \frac{u^k_{J_k}}{u^k_0}\Big) = f_0(x) + \sum_{j\in J_k} \frac{u^k_j}{u^k_0} f_j(x).$$

Remark 1. This assumption is similar to the well-known Dennis–More Assumption [2] that guarantees superlinear convergence for quasi-Newton methods.

Lemma 4.5. *Suppose that Assumptions A_1, A_2, A_3, A_4 and A_5 hold. Then the step size of the proposed algorithm always equals one, that is, $\lambda_k \equiv 1$, if k is sufficiently large.*

Proof: We know that it is sufficient to verify that (2.3) and (2.4) hold for $\lambda = 1$, and the statement "k large enough" will be omitted in the following discussion.

We first prove (2.4) holds for $\lambda = 1$. For $j \notin I(x^*)$, that is, $f_j(x^*) < 0$, in view of $(x^k, d^k, \widetilde{d}^k) \to (x^*, 0, 0)(k \to \infty)$, we can conclude $f_j(x^k + d^k + \widetilde{d}^k) \leqslant 0$ holds.

For $j \in I(x^*) = A^k(\varepsilon)$, from Taylor expansion, (2.1), (2.2) and formula (4.3), we have

$$
\begin{aligned}
&(4.5)\quad f_j(x^k + d^k + \widetilde{d}^k) = f_j(x^k + d^k) + \nabla f_j(x^k + d^k)^T \widetilde{d}^k + O(\|\widetilde{d}^k\|^2)\\
&(4.6)\quad = f_j(x^k + d^k) + \nabla f_j(x^k)^T \widetilde{d}^k + O\big(\|d^k\| \, \big\|\widetilde{d}^k\big\|\big) + O(\|\widetilde{d}^k\|^2)\\
&(4.7)\quad = -(1-\rho)\|d^k\|^\tau + \rho r_j \sigma_k^\nu z_k \|d^k\| + f_j(x^k) + \nabla f_j(x^k)^T d^k - r_j \sigma_k z_k\\
&(4.8)\quad + O\Big(\max\{\|d^k\|^3, -r_j \sigma_k z_k \|d^k\|\}\Big)\\
&(4.9)\quad \leqslant -(1-\rho)\|d^k\|^\tau + \rho r_j \sigma_k^\nu z_k \|d^k\| + O\Big(\max\{\|d^k\|^3, -r_j \sigma_k z_k \|d^k\|\}\Big).
\end{aligned}
$$

Therefore we have from (4.5) and the value of ρ

$$
f_j(x^k + d^k + \widetilde{d}^k) \leqslant -(1-\rho)\|d^k\|^\tau + \rho r_j \sigma_k^\nu z_k \|d^k\| + O\Big(\max\{\|d^k\|^3, -r_j \sigma_k z_k \|d^k\|\}\Big) < 0.
$$

This shows that (2.4) holds for $\lambda = 1$.

The next objective is to show (2.3) holds for $\lambda = 1$.

From Taylor expansion and taking into account relationship (4.3), we have

$$
\begin{aligned}
(4.10)\quad \omega_k &\overset{\text{def}}{=} f_0(x^k + d^k + \widetilde{d}^k) - f_0(x^k) - \alpha \nabla f_0(x^k)^T d^k\\
&= \nabla f_0(x^k)^T (d^k + \widetilde{d}^k) + \frac{1}{2}(d^k)^T \nabla^2_{xx} f_0(x^k) d^k - \alpha \nabla f_0(x^k)^T d^k + o(\|d^k\|^2).
\end{aligned}
$$

On the other hand, from the KKT condition of (2.1) and formula (4.3), one has

$$
\begin{aligned}
(4.11)\quad u_0^k \nabla f_0(x^k)^T (d^k + \widetilde{d}^k) &= -(d^k)^T H_k d^k - \sum_{j \in J_k} u_j^k \nabla f_j(x^k)^T (d^k + \widetilde{d}^k) + o(\|d^k\|^2).\\
u_0^k \nabla f_0(x^k)^T d^k &= -(d^k)^T H_k d^k - \sum_{j \in J_k} u_j^k \nabla f_j(x^k)^T d^k\\
&= -(d^k)^T \nabla f_k d^k + \sum_{j \in J_k} u_j^k f_j(x^k) - \sum_{j \in J_k} u_j^k r_j \sigma_k z_k.
\end{aligned}
$$

Again, from the third equation of (4.5) and Taylor expansion, we have

$$
f_j(x^k) + \nabla f_j(x^k)^T (d^k + \widetilde{d}^k) + \frac{1}{2}(d^k)^T \nabla^2_{xx} f_j(x^k) d^k = o(\|d^k\|^2), \quad j \in J_k.
$$

Thus

$$(4.12)\quad -\sum_{j\in J_k} u_j^k \nabla f_j(x^k)^T (d^k+\tilde{d}^k) = \sum_{j\in J_k} u_j^k f_j(x^k) + \frac{1}{2}(d^k)^T\Big(\sum_{j\in J_k} u_j^k \nabla_{xx}^2 f_j(x^k)\Big) d^k + o(\|d^k\|^2).$$

Substituting (4.12) into (4.11), one has

$$(4.13)\quad u_0^k \nabla f_0(x^k)^T (d^k+\tilde{d}^k) = -(d^k)^T H_k d^k + \sum_{j\in J_k} u_j^k f_j(x^k) + \frac{1}{2}(d^k)^T\Big(\sum_{j\in J_k} u_j^k \nabla_{xx}^2 f_j(x^k)\Big) d^k + o(\|d^k\|^2).$$

Substituting (4.13) and the third equation of (4.11) into (4.10), we obtain

$$\begin{aligned}\omega_k &= \frac{1}{u_0^k}(\alpha-1)(d^k)^T H_k d^k + \frac{1}{2}(d^k)^T \nabla_{xx}^2 L(x^k, \frac{u^k}{u_0^k}) d^k \\ &\quad + (1-\alpha)\sum_{j\in J_k} \frac{u_j^k}{u_0^k} f_j(x^k) + \alpha \sum_{j\in J_k} \frac{u_j^k}{u_0^k} r_j \sigma_k z_k + o(\|d^k\|^2) \\ &\leqslant \Big(\frac{1}{u_0^k}(\alpha-1)+\frac{1}{2}\Big) a\|d^k\|^2 + \frac{1}{2}(d^k)^T\Big(\nabla_{xx}^2 L\big(x^k, \frac{u^k}{u_0^k}\big) - B_k\Big) d^k \\ &\quad + (1-\alpha)\sum_{j\in J_k} \frac{u_j^k}{u_0^k} f_j(x^k) + \alpha \sum_{j\in J_k} \frac{u_j^k}{u_0^k} r_j \sigma_k z_k + o(\|d^k\|^2).\end{aligned}$$

So, using Assumption A_5 and the given conditions, one has

$$\omega_k \leqslant \Big(\frac{1}{u_0^k}(\alpha-1)+\frac{1}{2}\Big) a\|d^k\|^2 + o(\|d^k\|^2).$$

Therefore, according to $\alpha \in (0,(1/2))$, $u_0^k \to 1$, we know (2.3) holds for $\lambda = 1$. The whole proof is finished. ∎

Let

$$R_k = (\nabla f_j(x^k), j \in J_k), \quad P_k = E_n - R_k(R_k^T R_k)^{-1} R_k^T.$$

To discuss the convergence rate of the proposed algorithm, we give a lemma as follows.

LEMMA 4.6. *Under all the above-mentioned assumptions, when k is sufficiently large, the matrix*

$$D_k \stackrel{\text{def}}{=} \begin{pmatrix} P_k \nabla_{xx}^2 L(x^*, u^*) \\ R_k^T \end{pmatrix}$$

is of full rank.

The proof of Lemma 4.6 is similar to that in [**12**, Lemma 2.2] or [**11**, Lemma 2.2.2], and is omitted.

https://doi.org/10.1017/S0004972700035577 Published online by Cambridge University Press

THEOREM 4.1. *Under all above-mentioned assumptions, the algorithm is superlinearly convergent. that is, the sequence $\{x^k\}$ generated by the algorithm satisfies*

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

The proof is similar to the one in [**25**, Theorem 4.1], and is omitted.

5. Concluding remarks

In this paper, we have presented an active set feasible sequential quadratic programming algorithm for optimisation problems with nonlinear inequality constraints. Because of introduction of the active set technique, the size of the QP subproblem is reduced. To overcome the Maratos effect, a higher-order correction direction is obtained by solving a reduced least square problem. The algorithm is proved to be globally convergent and superlinearly convergent under some mild conditions *without strict complementarity*. Thus, the results show that the global convergence and superlinearly convergence are still guaranteed by deleting some "redundant" constraints.

References

[1] P.T. Boggs and J.W. Tolle, 'Sequential quadratic programming', *Acta Numer.* (1995), 1–51.

[2] C.G. Broyden, J.E. Dennis and J.J. More, 'On the local and superlinear convergence of quasi-Newton methods', *J. Inst. Math. Appl.* **12** (1973), 223–245.

[3] M.E. Cawood and M.M. Kostreva, 'Norm-relaxed method of feasible directions for solving nonlinear programming problems', *J. Optim. Theory. Appl.* **83** (1994), 311–320.

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

[5] J.F.A. De, O. Pantoja and D.Q. Mayne, 'Exact penalty function algorithm with simple updating of the penalty parameter', *J. Optim. Theory. Appl.* **69** (1991), 441–467.

[6] F. Facchinei, 'Robust recursive quadratic programming algorithm model with global and superlinear convergence properties', *J. Optim. Theory. Appl.* **92** (1997), 543–579.

[7] F. Facchinei, A. Fischer and C. Kanzow, 'On the accurate identification of active constraints', *SIAM J. Optim.* **9** (1999), 14–32.

[8] F. Facchinei and S. Lucidi, 'Quadratically and superlinearly convergent algorithms for the solution of inequality constrained optimization problems', *J. Optim. Theory. Appl.* **85** (1995), 265–289.

[9] Z.Y. Gao, G.P. He and F. Wu, 'A method of sequential systems of linear equations with arbitrary initial point', *Sci. China Ser. A* **27** (1997), 24–33.

[10] S.P. Han, 'A globally convergent method for nonlinear programming', *J. Optim. Theory. Appl.* **22** (1977), 297–309.

[11] J.B. Jian, *Researches on superlinearly and quadratically convergent algorithm for nonlinearly constrained optimization*, Ph.D.Thesis (School of Xi'an Jiaotong University., Xi'an, China, 2000).

[12] J.B. Jian, 'Two extension models of SQP and SSLE algorithms for optimization and their superlinear and quadrtical convergence', *Appl. Math. J. Chinese Univ. Ser. A* **16** (2001), 435–444.

[13] J.B. Jian, K.C. Zhang and S.J. Xue, 'A superlinearly and quadratically convergent SQP type feasible method for constrained optimization', *Appl. Math. J. Chinese Univ. Ser. B* **15** (2000), 319–331.

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

[15] C.T. Lawrence. and A.L.Tits, 'A computationally efficient feasible sequential quadratic programming algorithm', *SIAM J. Optim.* **11** (2001), 1092–1118.

[16] S. Lucidi, 'New results on a continuously differentiable penalty function', *SIAM J. Optim.* **2** (1992), 558–574.

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

[18] O. Pironneau and E. Polak, 'On the rate of convergence of certain methods of centers', *Math. Program.* **2** (1972), 230–257.

[19] O. Pironneau and E. Polak, 'Rate of convergence of a class of methods of feasible directions', *SIAM J. Numer. Anal.* **10** (1973), 161–173.

[20] M.J.D. Powell, 'A fast algorithm for nonlinearly constrained optimization calculations', in *Numerical analysis*, Lecture Notes in Math. **630**, (Proc. 7th Biennial Conf., Univ. Dundee, Dundee, 1977) (Springer–Verlag, Berlin, 1978), pp. 144–157.

[21] L. Qi and Y.F. Yang, 'A Globally and superlinearly convergent SQP algorithm for nonlinear constrained optimization', AMR00/7, Applied Mathematics Report, University of New South Wales, Sydney, March 2000.

[22] P. Spellucci, 'A new technique for inconsistent QP probrems in the SQP methods', *Math. Methods. Oper. Res.* **47** (1998), 355–400.

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

[24] S.J. Wright, 'Superlinear convergence of a stabilized SQP method to a degenerate solution', *Comput. Optim. Appl.* **11** (1998), 253–275.

[25] Z. Zhu, 'An efficient sequential quadratic programming algorithm for nonlinear programming', *J. Comput. Appl. Math.* **175** (2005), 447–464.

[26] G. Zoutendijk, *Methods of feasible directions* (Elsevier, Amsterdam, 1960).

Institute of Applied Mathematics
Hunan University
410082 Changsha
Peoples Republic China
and
Department of Information
Hunan Business College
410205 Changsha
Peoples Republic of China
e-mail: hqj0525@126.com.cn

Institute of Applied Mathematics
Hunan University
410082 Changsha
Peoples Republic China

Department of Information
Hunan Business College
410205 Changsha
Peoples Republic of China