A MODIFIED SQP METHOD FOR INEQUALITY
CONSTRAINED OPTIMIZATION WITHOUT
STRICT COMPLEMENTARITY

ZHIJUN LUO
The Department of Mathematics & Applied Mathematics
Humanities & Science and Technology Institute of Hunan
Loudi, 417000
P. R. China
e-mail: ldlzj123@163.com

Abstract

In this paper, a new modified SQP method is proposed to solve the inequality con-
strained optimization. This algorithm starts from an arbitrary initial point and
adjusts penalty parameter automatically. A descent direction is obtained by solving
only one modified QP subproblem. In order to avoid Maratos effect, a height-order
correction direction is computed by an explicit formula. Thanks to the height-order
technique, under mild conditions without strict complementarity, the global and
local superlinear convergence properties are obtained.

1. Introduction

In this paper, we consider the following inequality constrained
nonlinear programming problem:

\[ \min f(x) \quad (1.1) \]
\[ \text{s.t. } g_j(x) \leq 0, \ j \in I = \{1, 2, \ldots, m\}, \]

2000 Mathematics Subject Classification: 90C30, 65K05.

Keywords and phrases: inequality constrained optimization, SQP algorithm, global
convergence, superlinear convergence.

Received January 3, 2009

© 2009 Scientific Advances Publishers
where \( f : \mathbb{R}^n \to \mathbb{R} \), and \( g_j : \mathbb{R}^n \to \mathbb{R}, (j \in I) \) are continuously differentiable.

The method of sequence quadratic programming (SQP) is an important method for solving problem (1.1). Because of its superlinear convergence rate, it is a topic of much active research [3] - [7]. However, the traditional SQP algorithms have two serious shortcomings: (1) SQP algorithms require that the relate QP subproblem must be consistency. (2) There exists Matatos effect. Many efforts have been made to overcome the shortcomings through modifying the quadratic subproblem and the direction \( d \). Mo et al. [3] proposed a variant of SQP for solving problem (1.1). Their methods solve only one quadratic programming subproblem at each iteration. However, they have not analyzed the superlinear convergence rate. In [7], Zhu presented an efficient SQP for problem (1.1). The basic feasible descent direction \( d \) is computed by solving the QP\((x^k, H_k)\) problem as follows:

\[
\begin{align*}
\min \quad & z + \frac{1}{2} d^T H_k d \\
\text{s.t.} \quad & \nabla f(x^k)^T d \leq z \\
& g_j(x^k) + \nabla g_j(x^k)^T d \leq \eta_k z, \quad j \in I,
\end{align*}
\]

(1.2)

where \( H \) is a symmetric positive definite matrix, and \( \eta \) is nonnegative scalar. However, the superlinear convergence properties of these proposed SQP algorithms, depend strictly on the strict complementarity, which is rather strong and difficult for testing. Recently, In [2], in the process of the iteration of their algorithms, the search direction is generated by solving only one convex QP and two explicit computation formulas, the iterative points are all feasible and the objective function value is monotone decreasing. Under weaker assumptions without the strict complementarity, the algorithm is proved to possess global convergence, strong convergence and superlinear convergence.

In this paper, another new modified SQP method is proposed based on the subproblem proposed in [7]. But, unlike [7], we employ the penalty
A MODIFIED SQP METHOD FOR INEQUALITY ...  111

function associated with (1.1) as a merit function. Hence, the algorithm
can start from an arbitrary initial point in order to avoid Maratos effect,
a height-order correction direction is computed by an explicit formula.
Hence, under weaker assumptions without the strict complementarity,
the algorithm is proved to possess global convergence, strong convergence
and superlinear convergence.

The paper is organized as follows. The algorithm is presented in
Section 2. In Section 3, the global convergence results of the algorithm
are proved. While the superlinear convergence rate is analyzed in Section
4.

2. Description of Algorithm

Let

\[ \Phi(x) = \max_{j \in I} \{ g_j(x), 0 \}. \]

The direction derivative along \( d \in \mathbb{R}^n \) of \( \Phi(x) \) is

\[ \Phi'(x; d) = \max_{j \in I_0(x)} \{ \nabla g_j(x)^T d \}, \]

where \( I_0(x) = \{ j \in I : g_j(x) = \Phi(x) \} \). Generally speaking, \( \Phi'(x; d) \) is not
continuous. In [3], Mo used the following continuous approximation of
\( \Phi'(x; d) : \)

\[ \Phi^*(x; d) = \max_{j \in I_0(x)} \{ g_j(x) + \nabla g_j(x)^T d, 0 \} - \Phi(x). \]

Lemma 2.1 [6]. (1) For any \( x, d \in \mathbb{R}^n \), we have \( \Phi^*(x; d) \geq \Phi'(x; d) \),
and there exists \( \delta > 0 \), such that \( \Phi^*(x; kd) = \Phi'(x; kd) \), \( \forall k \in [0, \delta] \);

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

In order to obtain the global and local superlinear convergence
properties, we employ the penalty function associated with (1.1) as a
merit function, i.e.,

\[ \Psi_\sigma(x) = f(x) + \sigma \Phi(x), \]
where \( \sigma > 0 \) is the penalty parameter. Then the approximation directional derivatives of \( \Psi_\sigma(x) \) is

\[
\theta_\sigma(x; d) = \nabla f(x)^T d + \sigma \Phi^*(x; d).
\]

To overcome the Maratoes effect, a suitable auxiliary direction must be adopted. In this paper, we use an explicit auxiliary formula to compute a auxiliary direction \( \tilde{d}^k \) like [2]:

\[
\tilde{d}^k = -N_k(N_k^T N_k)^{-1} \left( \|d^k\|^2 + |\eta_k z_k|^2 \|d^k\| \|\nu + \tilde{g}^k\| \right),
\]

where \( e = (1, 1, \cdots, 1)^T \in R^m \), \( \tau \in (2, 3) \), \( \xi \in (0, 1) \),

\[ N_k := N(x^k) = (\nabla g_j(x^k), j \in I_0(x)), \]

\[
\tilde{g}_j^k = g_j(x^k + d^k) - g_j(x^k) - \nabla g_j(x^k)^T d^k.
\]

Now we state our algorithm as follows.

**Algorithm A:** **Step 0.** Given initial point \( x^0 \in R^n \), a symmetric positive definite matrix \( H_0 \in R^{n \times n} \). Choose parameters \( \beta \in (0, \frac{1}{2}) \), \( \sigma_0 > 0 \), \( \gamma \in (0, 1) \), \( \tau \in (2, 3) \), \( \eta_0 = \rho > 0 \). Set \( k = 0 \);

**Step 1.** Compute \((d^k, z_k)\) by solving \( QP(x^k, H_k) \) subproblem (1.2). If \( d^k = 0 \), then STOP;

**Step 2.** If \( \theta_{\sigma_k}(x^k; d^k) \leq -d^k^T H_k d^k \), let \( \sigma_{k+1} = \sigma_k \); Otherwise, let

\[
\sigma_{k+1} = \max \left\{ \frac{\nabla f(x)^T d + d^k^T H_k d^k}{-\Phi^*(x^k; d^k)}, 2\sigma_k \right\},
\]

**Step 3.** Compute \( \tilde{d}^k \) by (2.1). If \( \|\tilde{d}^k\| > \|d^k\| \), then set \( \tilde{d}^k = 0 \).

**Step 4.** Compute \( t_k \), the first number \( t \) in the sequence \( \{1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \cdots \} \) satisfying:
\[ \Psi_{\sigma_k+1}(x^k + t d^k + t^2 \tilde{d}^k) \leq \Psi_{\sigma_k+1}(x^k) + t \beta \sigma_{\sigma_k+1}(x^k; d^k). \] (2.3)

**Step 5.** Generate \( H_{k+1} \). Set \( x^{k+1} = x^k + t d^k + t^2 \tilde{d}^k \), \( \eta_{k+1} = \min \{ \rho, \|d^k\|^\gamma \} \). Set \( k := k + 1 \). Go back to Step 1.

### 3. Global Convergence of Algorithm

In this section, we analyze the global convergence of the Algorithm. The following general assumptions are true throughout this paper.

**H 3.1.** The sequences \( \{x^k\} \) and \( \{(d^k, z_k)\} \) are uniformly bounded.

**H 3.2.** The functions \( f, g_j, j \in I \) are two-times continuously differentiable.

**H 3.3.** \( \forall x \in R^n \), the set of vectors \( \{\nabla g_j(x) : j \in I_0(x)\} \) is linearly independent.

**H 3.4.** There exist \( a, b > 0 \), such that \( a \|d\|^2 \leq d^T H_k d \leq b \|d\|^2 \), for all \( k \in R \) and \( d \in R^n \).

**Definition 3.1.** A point \( x \in R^n \) is called

1. a **strong stationary point** of (1.1) if \( x \) is feasible and there exists scalars \( \lambda_j, j \in I \), satisfying

   \[ \nabla f(x) + \sum_{j=1}^m \lambda_j \nabla g_j(x) = 0, \quad \lambda_j g_j(x) = 0, \quad \lambda_j \geq 0, \quad j \in I. \]

2. a **weak stationary point** of (1.1) if \( x \) is feasible and there exists an infeasible sequence \( \{x^k\} \) converging to \( x \) such that

   \[ \lim_{k \to \infty} \frac{\max_{d \in D(x^k)} G(x^k; d)}{\Phi(x^k)} = 1, \]
where \( D(x^k) = \{ d | \nabla f(x^k)^T d \leq 0 \} \), \( G(x^k; d) = \max_{j \in I} \{ g_j(x) + \nabla g_j(x)^T d, 0 \} \).

It is easy to see that a strong stationary point defined as above is precisely a KKT point of (1.1). As for weak stationary point is like a F-J point of (1.1).

From the above assumptions, it is easy to obtain that the matrix \( N_k^T N_k \) is nonsingular and positive definite. Moreover, there exists a constant \( c > 0 \) such that \( \| (N_k^T N_k)^{-1} \| \leq c \). Furthermore, from Lemma 2.1 and the above assumptions, it is easy to see that the Algorithm A is well defined. According to [3], similarly, we can obtain the following results:

**Lemma 3.1.** Suppose that \( H_k \) is positive definite matrix and \((d^k, z_k)\) be the solution of problem \( QP(x^k, H_k) \).

1. The following inequality holds:
   \[
   z_k \leq \frac{1}{\eta} \Phi(x^k) - \frac{1}{2} d^k T H_k d^k.
   \] (3.1)

2. If \( x^k \to \hat{x} \) and \( H_k \to \hat{H} \), as \( k \to \infty \), then \( \{ (d^k, z_k) \} \to (\hat{d}, \hat{z}) \), where \((\hat{d}, \hat{z})\) is the unique solution of problem \( QP(\hat{x}, \hat{H}) \).

3. If \( K \) is an infinite index set such that \( \{d^k\}_{k \in K} \to 0 \), then, all accumulation points of \( \{x^k\}_{k \in K} \) are strong stationary point of (1.1).

**Lemma 3.2.** If \( \sigma_k \to \infty \), then \( \lim_{k \to \infty} \Phi(x^k) = 0 \). If \( \sigma_k = \sigma > 0 \) for all \( k \) large enough, \( \{x^k\} \) is an infinite sequence and \( \{x^k\}_{k \in K} \) is a convergent subsequence. Then \( d^k \to 0 \) as \( k \to \infty \) and \( k \in K \).

**Theorem 3.1.** Let \( \{x^k\} \) be an infinite sequence generated by the Algorithm. Then any accumulation point of \( \{x^k\} \) is either a strong stationary point or a weak stationary of (1.1).
Proof. Suppose $K$ is an infinite index set such that $x^k \to \hat{x}$, $k \in K$, $k \to \infty$. Let $(d^k, z^k)$ be the solution of $\text{QP}(x^k, H_k)$. If there exists an infinite index set $K_1 \subset K$ such that $d^k \to 0$, $k \to \infty$, $k \in K_1$, then it follows from Lemma 3.1 that $\hat{x}$ is a strong stationary point of (1.1). Now we suppose that there exists a constant $c_0$ such that

$$\|d^k\| \geq c_0, \quad (3.2)$$

for $k \in K$ and $k$ large enough. In view of Lemma 3.2, it holds that $\sigma_k \to \infty$, as $k \to \infty$. We will show that $\hat{x}$ is a weak stationary point. If it is not true, there exists a constant $\xi_1 > 0$ such that for $k$ large enough,

$$\max_{d \in D(x^k)} G(x^k; d) \leq \Phi(x^k) - \xi_1. \quad (3.3)$$

Suppose that $\hat{d}^k \in D(x^k)$ such that

$$G(x^k; \hat{d}^k) = \max_{d \in D(x^k)} G(x^k; d) \quad (3.4)$$

Since $\sigma_k \to \infty$, it follows from (3.1), H 3.4, (3.2) and Lemma 3.2 that for $k$ large enough,

$$z_k \leq \frac{1}{\eta_k} \Phi(x^k) - \frac{1}{2} d^{kT} H_k d^k = -\frac{1}{2} d^{kT} H_k d^k \leq -\xi_2 < 0,$$

where $\xi_2 > 0$ is a constant. Thus, from the first constraint of (1.2), we have $\nabla f(x^k)^T d^k \leq 0$, i.e., $d^k \in D(x^k)$. Hence, from (3.4), we obtain

$$G(x^k; d^k) \leq G(x^k; \hat{d}^k).$$

Then, H 3.1, Lemma 3.2 and $\sigma_k \to \infty$ imply that inequality

$$\theta_{\sigma_k}(x^k; d^k) + d^{kT} H_k d^k \leq z_k + d^{kT} H_k d^k + \theta_{\sigma_k}(G(x^k; d^k) - \Phi(x^k))$$

$$\leq \frac{1}{2} d^{kT} H_k d^k - \xi \sigma_k < 0,$$
holds for $k$ large enough, which contradicts the parameter updating procedure in the Step 2 of the Algorithm. So, it follows that $\hat{x}$ is a weak stationary point of (1.1).

4. Superlinear Convergence of Algorithm

In this section, we further discuss the superlinear convergence properties of the proposed algorithm under some mild conditions without the strict complementarity assumption. For this reason, we add the following additional hypothesis is necessary.

**H 4.1.** The KKT pair $(x^*, u^*)$ satisfies the strongly second-order sufficient conditions, i.e.,

$$d^T\nabla^2_{xx}L(x^*, u^*)d > 0, \quad \forall d \in \Omega \triangleq \{d \in R^n | d \neq 0, \nabla g_{I^*}(x^*)^Td = 0\},$$

where

$$L(x, u) = f(x) + \sum_{I} u_j g_j(x), \quad I^+_x = \{j \in I : u_j^* > 0\}.$$

**Lemma 4.1** [3]. For $k$ large enough, $\sigma_k \equiv \sigma > 0$.

Due to Lemma 4.1, we always assume that $\sigma_k \equiv \sigma$ for all $k$ the rest of this paper. Thereby, from H 4.1 and Proposition 4.1 in [4] we can get that $x_k \to x^*, k \to \infty$. Furthermore, according to Lemma 4.1 and Lemma 3.2, for $k$ large enough, it holds that

$$d^k \to 0, z_k \to 0, \eta_k \to 0, \nu_k \to 1, z_k = O(\|d^k\|). \quad (4.1)$$

**Lemma 4.2.** For $k$ large enough, it holds that

1). $L_k = I_0(x^*) \Delta \bigtriangleup I^*_x, \quad \mu_k \to u^*$, where,

$I_0(x^*) = \{j \in I \mid g_j(x^*) = \Phi(x^*)\}, \quad L_k = \{j \in I \mid g_j(x^k) + \nabla g_j(x^k)^Td^k = \eta_k z_k\}.$

2). $\tilde{d}^k$ obtained by Step 3 satisfies that $\|\tilde{d}^k\| = O(\|\tilde{d}^k\|)$.

3). $I^+_x \subseteq L_k.$
Proof. 1). In view of Lemma 3.1 and (4.1), we have $\Phi(x^k) = \eta_k z_k \to 0, \ k \to \infty$. It implies that $I_s = I(x^*) = \{j \in I \ | \ g_j(x^*) = 0\}$. From Lemma 4.3 in [7], we can get $L_k = I(x^*)$. Then, it is easy to see that $L_k = I_0(x^*) \supseteq I_s$. The rest proof of the Lemma, see Lemma 4.3 in [7].

2). From the definition of $\tilde{d}^k$, it is easy to get the result.

3). Taking into account

$$(x^k, \eta_k, d^k, z_k) \to (x^*, 0, 0, 0), \ k \to +\infty.$$  

We obtain $L_k \subseteq I(x^*)$. On the other hand, for $j \in I^+_s$, we get $\mu_j^k \to u_j^*$ $> 0$ from 1), then $I^+_s \subseteq L_k$ for $k$ large enough.

To ensure the unit step size can be accepted, we make another assumption.

**H 4.2.** Let

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

where

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

$$A_k = (\nabla g_j(x^k)), j \in L_k = \{j \in I \ | \ g_j(x^k) + \nabla g_j(x^k)^T d^k = \eta_k z_k\}.$$  

Denote

$$\nabla^2_{xx} L(x^k, \nu_k, \mu_k) = \nabla^2 f(x^k) + \sum_{j \in I_s} \frac{\mu_j^k}{\nu_k} \nabla^2 g(x^k).$$

Due to (4.1), Lemma 4.2 and H 4.2, it holds that

$$\|P_k(H_k - \nabla^2_{xx} L(x^k, \nu_k, \mu_k))d^k\| = o(|d^k|). \quad (4.2)$$
Lemma 4.3. For $k$ large enough, the step $t_k = 1$ is accepted by the Step 4.

Proof. Since $\|d^k\|\to 0$ as $k \to \infty$ and $\|\tilde{d}^k\| = o(\|d^k\|)$, then

\[
s \triangleq \psi_\sigma(x^k + d^k + \tilde{d}^k) - \psi_\sigma(x^k) - \beta_\alpha(x^k; d^k) \leq (1 - \beta)\theta_\sigma(x^k; d^k) + \nabla f(x^k)^T \tilde{d}^k + \frac{1}{2} (d^k)^T \nabla^2 f(x^k) d^k + o(\|d^k\|^2).
\]

So, we have

\[
\nabla f(x^k) = -\frac{1}{\nu_k} \left[ H_k d^k + N_k u^k \right].
\]

From the definition of $\tilde{d}^k$ and Lemma 4.2, we have

\[-N_k^T \tilde{d}^k = \left( (\|d^k\|^2 + |\eta_k z_k|\|d^k\|) e + \tilde{g}^k \right) = o(\|d^k\|^2) + \frac{1}{2} (d^k)^T \nabla^2 g_j(x^k) d^k,
\]

then

\[
\nabla f(x^k)^T \tilde{d}^k = -\frac{1}{\nu_k} [d^k]^T H_k \tilde{d}^k + \mu^k N_k^T \tilde{d}^k]
\]

\[= o(\|d^k\|^2) + \sum_{j \in L_h} \frac{\mu_j}{\nu_k} \left[ \frac{1}{2} (d^k)^T \nabla^2 g_j(x^k) d^k \right].
\]

Then, it is easy to see that

\[s \leq (1 - \beta)\theta_\sigma(x^k; d^k) + \frac{1}{2} d^k^T \nabla^2 f(x^k, v_k, \mu_k) d^k + o(\|d^k\|^2)
\]

\[\leq (\beta - \frac{1}{2} \lambda_\sigma H_k d_k - \frac{1}{2} d^k^T (H_k - \nabla^2 f(x^k, v_k, \mu_k)) d^k + o(\|d^k\|^2).
\]

Denote $A_\sigma = (\nabla g_j(x^k), j \in I_\sigma)$, $P_\sigma = I_n - A_\sigma (A_\sigma^T A_\sigma)^{-1} A_\sigma^T$, then $P_k \to P_\sigma$. Let

\[d^k = P_\sigma d^k + y_k.
\]
\[ y_k = A_s(A_s^TA_s)^{-1}A_s^Td_k \]
\[ = A_s(A_s^TA_s)^{-1}(A_s^TA_k - A_k^T)d_k + A_s(A_s^TA_s)^{-1}A_s^Td_k, \]
then, according to the facts \( L_k = I_s = I_0(x^*) \), we have
\[
\|y_k\| = o(\|d_k\|) + O\left( \sum_{j \in I_s} g_j^2(x^k) \right)^{\frac{1}{2}} + O\left( \sum_{I_s} \| \eta_k z_k^2 \|^{\frac{1}{2}} \right).
\]
Thereby, it follows from (4.1), (4.2), and Lemma 4.2, that for \( k \) large enough
\[
s \leq (\beta - \frac{1}{2})d_k^TH_kd_k - \frac{1}{2}(P_*d_k + y_k)^T(H_k - \nabla x^2 L(x^k, \nu_k, \mu_k))d_k + o(\|d_k\|^2)
\]
\[ = (\beta - \frac{1}{2})d_k^TH_kd_k + o(\|d_k\|^2). \]
Hence, according to \( \beta \in (0, \frac{1}{2}) \), we get that \( s \leq 0 \), i.e., \( t_k = 1 \) for all \( k \) large enough.

Moreover, in view of Theorem 5.2 in [1], we may obtain the following theorem:

**Theorem 4.1.** Under all above-mentioned assumptions, the algorithm is superlinearly convergent, i.e., the sequence \( \{x^k\} \) generated by the algorithm satisfies
\[
\|x^{k+1} - x^*\| = o(\|x^k - x^*\|).
\]

**References**


