

# A Modified Limited SQP Method For Constrained Optimization\*

Gonglin Yuan<sup>1</sup>, Sha Lu<sup>2</sup>, Zengxin Wei<sup>1</sup>

<sup>1</sup>Department of Mathematics and Information Science, Guangxi University, Nanning, China

<sup>2</sup>School of Mathematics Science, Guangxi Teacher's Education University, Nanning, China

E-mail: [glyuan@gxu.edu.cn](mailto:glyuan@gxu.edu.cn)

Received December 23, 2009; revised February 24, 2010; accepted March 10, 2010

## Abstract

In this paper, a modified variation of the Limited SQP method is presented for constrained optimization. This method possesses not only the information of gradient but also the information of function value. Moreover, the proposed method requires no more function or derivative evaluations and hardly more storage or arithmetic operations. Under suitable conditions, the global convergence is established.

**Keywords:** Constrained Optimization, Limited Method, SQP Method, Global Convergence

## 1. Introduction

Consider the constrained optimization problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & h_i(x) = 0, \quad i \in E \\ & g_j(x) \leq 0, \quad j \in I \end{aligned} \quad (1)$$

where  $f, h_i, g_j: R^n \rightarrow R$  are twice continuously differentiable,  $E = \{1, 2, \dots, m'\}$ ,  $I = \{m'+1, m'+2, \dots, m'+l\}$ ,  $l > 0$  is an integer. Let the Lagrangian function be defined by

$$L(x, \mu, \lambda) = f(x) + \mu^T g(x) + \lambda^T h(x) \quad (2)$$

where  $\mu$  and  $\lambda$  are multipliers. Obviously, the Lagrangian function  $L$  is a twice continuously differentiable function. Let  $S$  be the feasible point set of the problem (1). We define  $I^*$  to be the set of all the subscripts of those inequality constraints which are active at  $x^*$ , i.e.,  $I^* = \{i \mid i \in I \text{ and } g_i(x) = 0\}$ .

It is well known that the SQP methods for solving twice continuously differentiable nonlinear programming problems, are essentially Newton-type methods for finding Kuhn-Tucker points of nonlinear programming problems. These years, the SQP methods have been in vogue [1-8]: Powell [5] gave the BFGS-Newton-SQP method for the nonlinearly constrained optimization. He gave some sufficient conditions, under which SQP method would yield 2-step Q-superlinear convergence rate (assuming convergence) but did not show that his mod-

ified BFGS method satisfied these conditions. Coleman and Conn [2] gave a new local convergence quasi-Newton-SQP method for the equality constrained nonlinear programming problems. The local 2-step Q-superlinear convergence was established. Sun [6] proposed quasi-Newton-SQP method for general  $LC^1$  constrained problems. He presented the locally convergent sufficient conditions and superlinear convergent sufficient conditions. But he did not prove whether the modified BFGS-quasi-Newton-SQP method satisfies the sufficient conditions or not. We know that, the BFGS update exploits only the gradient information, while the information of function values of the Lagrangian function (2) available is neglected.

If  $x \in R^n$  holds, then the problem (1) is called unconstrained optimization problem (UNP). There are many methods [9-13] for the UNP, where the BFGS method is one of the most effective quasi-Newton method. The normal BFGS update exploits only the gradient information, while the information of function values available is neglected for UNP too. These years, lots of modified BFGS methods (see [14-19]) have been proposed for UNP. Especially, many efficient attempts have been made to modify the usual quasi-Newton methods using both the gradient and function values information (e.g. [19,20]). Lately, in order to get a higher order accuracy in approximating the second curvature of the objective function, Wei, Yu, Yuan, and Lian [18] proposed a new BFGS-type method for UNP, and the reported numerical results show that the average performance is better than that of the standard BFGS method. The superlinear convergence of this modified has been established for uniformly convex function. Its global convergence is established by Wei, Li, and Qi [20]. Motivated by their ideas, Yuan and Wei [21] presented a modified BFGS method

\*This work is supported by the Chinese NSF grants 10761001 and the Scientific Research Foundation of Guangxi University (Grant No. X081082), and Guangxi SF grants 0991028.

which can ensure that the update matrix are positive definite for the general convex functions. Moreover, the global convergence is proved for the general convex functions.

The limited memory BFGS (L-BFGS) method (see [22]) is an adaptation of the BFGS method for large-scale problems. The implementation is almost identical to that of the standard BFGS method, the only difference is that the inverse Hessian approximation is not formed explicitly, but defined by a small number of BFGS updates. It is often provided a fast rate of linear convergence, and requires minimal storage.

Inspired by the modified method of [21], we combine this technique and the limited memory technique, and give a limited SQP method for constrained optimization. The global convergence of the proposed method will be established for generally convex function. The major contribution of this paper is an extension of, based on the basic of the method in [21], the method for the UNP to constrained optimization problems. Unlike the standard SQP method, a distinguishing feature of our proposed method is that a triple  $\{s_i, y_i, A_i^*\}$  being stored, where  $s_i = x_{i+1} - x_i$ ,  $y_i = \nabla_x L(z_{i+1}) - \nabla_x L(z_i) + A_i^* s_i$ ,  $z_{i+1} = (x_{i+1}, \mu_{i+1}, \lambda_{i+1})$ ,  $z_i = (x_i, \mu_i, \lambda_i)$ ,  $\mu_i$  and  $\lambda_i$  are the multipliers which are according to the Lagrangian objective function at  $x_i$ , while  $\mu_{i+1}$  and  $\lambda_{i+1}$  are the multipliers which are according to the Lagrangian objective function at  $x_{i+1}$ , and  $A_i^*$  is a scalar related to Lagrangian function value. Moreover, a limited memory SQP method is proposed. Compared with the standard SQP method, the presented method requires no more function or derivative evaluations, and hardly more storage or arithmetic operations.

This paper is organized as follows. In the next section, we briefly review some modified method and the L-BFGS method for UNP. In Section 3, we describe the modified limited memory SQP algorithm for (2). The global convergence will be established in Section 4. In the last section, we give a conclusion. Throughout this paper,  $\|\cdot\|$  denotes the Euclidean norm of vectors or matrix.

## 2. Modified BFGS Update and the L-BFGS Update for UNP

We will state the modified BFGS update and the L-BFGS update for UNP in the following subsections, respectively.

### 2.1. Modified BFGS Update

Quasi-Newton methods for solving UNP often need to update the iterate matrix  $B_k$ . In tradition,  $\{B_k\}$  satisfies

the following quasi-Newton equation:

$$B_{k+1} S_k = \delta_k \quad (3)$$

where  $S_k = x_{k+1} - x_k$ ,  $\delta_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ . The very famous update  $B_k$  is the BFGS formula

$$B_{k+1} = B_k - \frac{B_k S_k S_k^T B_k}{S_k^T B_k S_k} + \frac{\delta_k \delta_k^T}{S_k^T \delta_k} \quad (4)$$

Let  $H_k$  be the inverse of  $B_k$ , then the inverse update formula of (4) method is represented as

$$\begin{aligned} H_{k+1} &= H_k - \frac{\delta_k^T (S_k - H_k \delta_k) S_k S_k^T}{(\delta_k^T S_k)^2} + \\ &\frac{(S_k - H_k \delta_k) S_k^T + S_k (S_k - H_k \delta_k)^T}{(\delta_k^T S_k)^2} \\ &= (I - \frac{S_k \delta_k^T}{\delta_k^T S_k}) H_k (I - \frac{\delta_k S_k^T}{\delta_k^T S_k}) + \frac{S_k S_k^T}{\delta_k^T S_k}, \end{aligned} \quad (5)$$

which is the dual form of the *DFP* update formula in the sense that  $H_k \leftrightarrow B_k$ ,  $H_{k+1} \leftrightarrow B_{k+1}$ , and  $s_k \leftrightarrow y_k$ . It has been shown that the BFGS method is the most effective in quasi-Newton methods from computation point of view. The authors have studied the convergence of  $f$  and its characterizations for convex minimization [23-27]. Our pioneers made great efforts in order to find a quasi-Newton method which not only possess global convergence but also is superior than the BFGS method from the computation point of view [15-17,20,28-31]. For general functions, it is now known that the BFGS method may fail for non-convex functions with inexact line search [32], Mascarenhas [33] showed that the non-convergence of the standard BFGS method even with exact line search. In order to obtain a global convergence of BFGS method without convexity assumption on the objective function, Li and Fukushima [15,16] made a slight modification to the standard BFGS method. Now we state their work [15] simply. Li and Fukushima (see [15]) advised a new quasi-Newton equation with the following form  $B_{k+1} S_k = \delta_k^{1*}$ , where  $\delta_k^{1*} = \delta_k + t_k \|g_k\| S_k$ ,

$t_k > 0$  is determined by  $t_k = 1 + \max\{-\frac{S_k^T \delta_k}{\|S_k\|^2}, 0\}$ . Under appropriate conditions, these two methods [15,16] are globally and superlinearly convergent for nonconvex minimization problems.

In order to get a better approximation of the objective function Hessian matrix, Wei, Yu, Yuan, and Lian (see [18]) also proposed a new quasi-Newton equation:

$$B_{k+1} (2) S_k = \delta_k^{2*} = \delta_k + A_k (3) S_k, \text{ where}$$

$$A_k (3) = \frac{2[f(x_k) - f(x_k + \alpha_k d_k)] + [\nabla f(x_k + \alpha_k d_k) + \nabla f(x_k)]^T S_k}{\|S_k\|^2}.$$

Then the new BFGS update formula is

$$B_{k+1}(2) = B_k(2) - \frac{B_k(2)S_k S_k^T B_k(2)}{S_k^T B_k(2)S_k} + \frac{\delta_k^{2*} \delta_k^{2*T}}{S_k^T \delta_k^{2*}}. \quad (6)$$

Note that this quasi-Newton formula (6) contains both gradient and function value information at the current and the previous step. This modified BFGS update formula differs from the standard BFGS update, and a higher order approximation of  $\nabla^2 f(x)$  can be obtained (see [18,20]).

It is well known that the matrix  $B_k$  are very important for convergence if they are positive definite [24,25]. It is not difficult to see that the condition  $S_k^T \delta_k^{2*} > 0$  can ensure that the update matrix  $B_{k+1}(2)$  from (6) inherits the positive definiteness of  $B_k(2)$ . However this condition can be obtained only under the objective function is uniformly convex. If  $f$  is a general convex function, then  $S_k^T \delta_k^{2*}$  and  $S_k^T \delta_k$  may equal to 0. In this case, the positive definiteness of the update matrix  $B_k$  can not be sure. Then we conclude that, for the general convex functions, the positive definiteness of the update matrix  $B_k$  generated by (4) and (6) can not be satisfied.

In order to get the positive definiteness of  $B_k(2)$  based on the definition of  $\delta_k^{2*}$  and  $\delta_k$  for the general convex functions, Yuan and Wei [21] give a modified BFGS update, *i. e.*, the modified update formula is defined by

$$B_{k+1}(3) = B_k(3) - \frac{B_k(3)S_k S_k^T B_k(3)}{S_k^T B_k(3)S_k} + \frac{\delta_k^{3*} \delta_k^{3*T}}{\delta_k^{3*T} S_k}, \quad (7)$$

where  $\delta_k^{3*} = \delta_k + A_k S_k$ ,  $A_k = \max\{A_k(3), 0\}$ . Then the corresponding quasi-Newton equation is

$$B_{k+1}(3)S_k = \delta_k^{3*} \quad (8)$$

which can ensure that the condition  $S_k^T \delta_k^{3*} > 0$  holds for the general convex function  $f$  (see [21] in detail). Therefore, the update matrix  $B_{k+1}$  from (8) inherits the positive definiteness of  $B_k$  for the general convex function.

## 2.2. Limited Memory BFGS-Type Method

The limited memory BFGS (L-BFGS) method (see [22]) is an adaptation of the BFGS method for large-scale problems. In the L-BFGS method, matrix  $H_k$  is obtained by updating the basic matrix  $H_0(\tilde{m} > 0)$  times using BFGS formula with the previous  $\tilde{m}$  iterations. The standard BFGS correction (5) has the following

form

$$H_{k+1} = V_k^T H_k V_k + \rho_k S_k S_k^T \quad (9)$$

where  $\rho_k = \frac{1}{S_k^T \delta_k}$ ,  $V_k = I - \rho_k \delta_k S_k^T$ ,  $I$  is the unit matrix. Thus,  $H_{k+1}$  in the L-BFGS method has the following form:

$$\begin{aligned} H_{k+1} &= V_k^T H_k V_k + \rho_k S_k S_k^T \\ &= V_k^T [V_{k-1}^T H_{k-1} V_{k-1} + \rho_{k-1} S_{k-1} S_{k-1}^T] V_k + \rho_k S_k S_k^T \\ &= \dots \\ &= [V_k^T \dots V_{k-\tilde{m}+1}^T] H_{k-\tilde{m}+1} [V_{k-\tilde{m}+1} \dots V_k] \\ &\quad + \rho_{k-\tilde{m}+1} [V_{k-1}^T \dots V_{k-\tilde{m}+2}^T] S_{k-\tilde{m}+1} S_{k-\tilde{m}+2}^T [V_{k-\tilde{m}+2} \dots V_{k-1}] \\ &\quad + \dots \\ &\quad + \rho_k S_k S_k^T. \end{aligned} \quad (10)$$

## 3. Modified SQP Method

In this section, we will state the normal SQP method and the modified limited memory SQP method, respectively.

### 3.1. Normal SQP Method

The first-order Kuhn-Tucker condition of (2) is

$$\begin{cases} \nabla f(x^*) + \mu^T \nabla g(x^*) + \lambda^T \nabla h(x^*) = 0, \\ g(x^*) \leq 0, \mu_j \geq 0, \mu_j g_j(x^*) = 0, \text{ for } j \in I, \\ h(x^*) = 0. \end{cases} \quad (11)$$

The system (11) can be represented by the following system:

$$H(z) = 0, \quad (12)$$

where  $z = (z, \mu, \lambda) \in S$  and  $H: R^{n+m'+l'} \rightarrow R^{n+m'+l'}$  is defined by

$$H(z) = \begin{pmatrix} \nabla f(x) + \mu^T \nabla g(x) + \lambda^T \nabla h(x) \\ \min\{-g(x), \mu\} \\ h(x) \end{pmatrix}. \quad (13)$$

Since  $\nabla f$ ,  $\nabla g$ , and  $\nabla h$  are continuously differentiable functions, it is obviously that  $H(z)$  is continuously differentiable function. Then, for all  $d \in R^{n+m'+l'}$ , the directional derivative  $H'(z:d)$  of the function  $H(z)$  exists. Denote the index sets by

$$\alpha(z) = \{i \mid \mu_i > -g_i(x)\} \quad (14)$$

$$\text{and } \beta(z) = \{i \mid \mu_i \leq -g_i(x)\}. \quad (15)$$

Under the complementary condition, it is clearly that  $\alpha(z)$  is an index set of strongly active inequality constraints, and  $\beta(z)$  is an index set of weakly active and inactive inequality constraints. In terms of these sets, the directional derivative along the direction  $d = (d_x, d_\mu, d_\lambda)$  is given as follows

$$H'(z : d) = \begin{pmatrix} Gd \\ (-\nabla g_i^T d_x)_{i \in \alpha(z)} \\ \min\{d_{\mu_i}, (-\nabla g_i^T d_x)\}_{i \in \beta(z)} \\ \nabla h(x)^T d_x \end{pmatrix}, \quad (16)$$

where  $G$  is a matrix which elements are the partial derivatives of  $\nabla_x L(z)$  to  $d_x, d_\mu, d_\lambda$ , respectively. If  $\min\{d_{\mu_i}, (-\nabla g_i^T d_x)\}_{i \in \beta(z)} = d_{\mu_i}$  holds, then the set

$$W(z) = \begin{pmatrix} V & \nabla g_\alpha(x) & \nabla g_\beta(x) & \nabla h(x) \\ \nabla g_\alpha(x)^T & 0 & 0 & 0 \\ 0 & 0 & I_\beta & 0 \\ \nabla h(x)^T & 0 & 0 & 0 \end{pmatrix}. \quad (17)$$

By (33) in [6], we know than the system

$$W_k d_k = -H(z_k), \quad (18)$$

where  $d_k = (d_{x_k}, d_{\mu_k}, d_{\lambda_k})$  and  $W_k = W(z_k)$ , define the Kuhn-Tucker condition of problem (2), which also defines the Kuhn-Tucker condition of the following quadratic programming  $QP(z_k, V_k)$ :

$$\begin{aligned} \min \quad & \nabla f(x_k)^T s + \frac{1}{2} s^T V_k s, \\ \text{s.t.} \quad & g_\alpha(x_k) + \nabla g_\alpha(x_k)^T s = 0, \\ & g_\beta(x_k) + \nabla g_\beta(x_k)^T s \leq 0, \\ & h(x_k) + \nabla h(x_k)^T s = 0, \end{aligned} \quad (19)$$

where  $s = x - x_k, V_k = \nabla_{xx}^2 L(z_k)$ .

Generally, suppose that  $B_k(1)$  is an estimate of  $V_k$  and  $B_k(1)$  can be updated by BFGS method of quasi-Newton formula

$$B_{k+1}(1) = B_k(1) - \frac{B_k(1)s_k s_k^T B_k(1)}{s_k^T B_k(1)s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \quad (20)$$

where  $s_k = x_{k+1} - x_k, y_k = \nabla_x L(z_{k+1}) - \nabla_x L(z_k), z_{k+1} = (x_{k+1}, \mu_{k+1}, \lambda_{k+1}), z_k = (x_k, \mu_k, \lambda_k), \mu_k$  and  $\lambda_k$  are the multipliers which are according to the Lagrangian objective function at  $x_k$ , while  $\mu_{k+1}$  and  $\lambda_{k+1}$  are the multipliers which are according to the Lagrangian objective function at  $x_{k+1}$ . Particularly, when we use the update formula (20) to (19), the above quadratic programming problem can be written as  $QP(z_k, B_k)$ :

$$\begin{aligned} \min \quad & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k(1)s, \\ \text{s.t.} \quad & g_\alpha(x_k) + \nabla g_\alpha(x_k)^T s = 0, \\ & g_\beta(x_k) + \nabla g_\beta(x_k)^T s \leq 0, \\ & h(x_k) + \nabla h(x_k)^T s = 0. \end{aligned} \quad (21)$$

Suppose that  $(s, \mu, \lambda)$  is a Kuhn-Tucker triple of the sub problem  $QP(z_k, B_k)$ , therefore, it is obviously that  $s = 0$  if  $(x_k, \mu_k, \lambda)$  is a Kuhn-Tucker triple of (2).

### 3.2. Modified Limited Memory SQP Method

The normal limited memory BFGS formula of quasi-Newton-SQP method with  $H_k$  for constrained optimization (2) is defined by

$$\begin{aligned} H_{k+1} &= V_k^T H_k V_k + \rho_k s_k s_k^T \\ &= V_k^T [V_{k-1}^T H_{k-1} V_{k-1} + \rho_{k-1} s_{k-1} s_{k-1}^T] V_k + \rho_k s_k s_k^T \\ &= \dots \\ &= [V_k^T \dots V_{k-\tilde{m}+1}^T] H_{k-\tilde{m}+1} [V_{k-\tilde{m}+1} \dots V_k] \\ &\quad + \rho_{k-\tilde{m}+1} [V_{k-1}^T \dots V_{k-\tilde{m}+2}^T] s_{k-\tilde{m}+1} s_{k-\tilde{m}+2}^T [V_{k-\tilde{m}+2} \dots V_{k-1}] \\ &\quad + \dots \end{aligned} \quad (22)$$

where  $\rho_k = \frac{1}{s_k^T y_k}, V_k = I - \rho_k y_k s_k^T, I$  is the unit

matrix. To maintain the positive definiteness of the limited memory BFGS matrix, some researchers suggested to discard correction  $\{s_k, y_k\}$  if  $s_k^T y_k > 0$  does not hold (e.g. [34]). Another technique was proposed by Powell [35] in which  $y_k$  is defined by

$$y = \begin{cases} y_k, & \text{if } s_k^T y_k \geq 0.2s_k^T B_k s_k, \\ \theta_k y_k + (1 - \theta_k) B_k s_k, & \text{otherwise,} \end{cases}$$

where  $\theta_k = \frac{0.8s_k^T B_k s_k}{s_k^T B_k s_k - s_k^T y_k}, B_k^{-1} = H_k$  of (22). How-

ever, if the Lagrangian objective function  $L(x, \mu, \lambda)$  is a general convex function, then  $s_k^T y_k$  may equal to 0. In this case, the positive definiteness of the update matrix  $H_k$  of (22) can not be sure.

Whether there exists a limited memory SQP method which can ensure that the update matrix are positive definite for general convex Lagrangian objective function  $L(x, \mu, \lambda)$ . This paper gives a positive answer. Let

$$\tilde{A}_k = \frac{2[L(z_k) - L(z_{k+1})] + [\nabla_x L(z_{k+1}) + \nabla_x L(z_k)]^T s_k}{\|s_k\|^2}.$$

Considering the discussion of the above section, we discuss  $\tilde{A}_k$  for general convex Lagrangian objective function

$L(x, \mu, \lambda)$  in the following cases to state our motivation.

case i: If  $\tilde{A}_k > 0$ , we have

$$s_k^T (y_k + \tilde{A}_k s_k) = s_k^T y_k + \tilde{A}_k \|s_k\|^2 > s_k^T y_k \geq 0. \quad (23)$$

case ii: If  $\tilde{A}_k < 0$ , we get

$$\begin{aligned} 0 > \tilde{A}_k &= \frac{2[L(z_k) - L(z_{k+1})] + [\nabla_x L(z_{k+1}) + \nabla_x L(z_k)]^T s_k}{\|s_k\|^2} \\ &\geq \frac{-2\nabla_x L(z_{k+1})s_k + [\nabla_x L(z_{k+1}) + \nabla_x L(z_k)]^T s_k}{\|s_k\|^2} \\ &= \frac{-s_k^T y_k}{\|s_k\|^2}, \end{aligned} \quad (24)$$

which means that  $s_k^T y_k > 0$  holds. Then we present our modified limited memory SQP formula

$$\begin{aligned} H_{k+1}^* &= V_k^{*T} H_k^* V_k^* + \rho_k^* s_k s_k^T \\ &= V_k^{*T} [V_{k-1}^{*T} H_{k-1}^* V_{k-1}^* + \rho_{k-1}^* s_{k-1} s_{k-1}^T] V_k^* + \rho_k^* s_k s_k^T \\ &= \dots \\ &= [V_k^{*T} \dots V_{k-\tilde{m}+1}^{*T}] H_{k-\tilde{m}+1}^* [V_{k-\tilde{m}+1}^* \dots V_k^*] \\ &\quad + \rho_{k-\tilde{m}+1}^* [V_{k-1}^{*T} \dots V_{k-\tilde{m}+2}^{*T}] s_{k-\tilde{m}+1} s_{k-\tilde{m}+1}^T [V_{k-\tilde{m}+2}^* \dots V_{k-1}^*] \\ &\quad + \dots \\ &\quad + \rho_k^* s_k s_k^T \end{aligned} \quad (25)$$

where  $\rho_k^* = \frac{1}{s_k^T y_k^*}$ ,  $V_k^* = I - \rho_k^* y_k^* s_k^T$ , and

$y_k^* = y_k + \max\{\tilde{A}_k, 0\} s_k$ . It is not difficult to see that the modified limited memory SQP formula (25) contains both the gradient and function value information of Lagrangian function at the current and the previous step if  $\tilde{A}_k > 0$  holds.

Let  $B_k^*$  be the inverse of  $H_k^*$ . More generally, suppose that  $B_k^*$  is an estimate of  $V_k$ . Then the above quadratic programming problem (19) can be written as  $QP(z_k, B_k^*)$ :

$$\begin{aligned} \min \quad & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k^* s, \\ \text{s.t.} \quad & g_\alpha(x_k) + \nabla g_\alpha(x_k)^T s = 0, \\ & g_\beta(x_k) + \nabla g_\beta(x_k)^T s \leq 0, \\ & h(x_k) + \nabla h(x_k)^T s = 0. \end{aligned} \quad (26)$$

Suppose that  $(s, \mu, \lambda)$  is a Kuhn-Tucker triple of the subproblem  $QP(z_k, B_k^*)$ , therefore, it is obviously that  $s = 0$  if  $(x_k, \mu_k, \lambda)$  is a Kuhn-Tucker triple of (2).

Now we state our algorithm as follows.

### Modified limited memory SQP algorithm 1 for (2) (M-L-SQP-A1)

**Step 0:** Star with an initial point  $z_0 = (x_0, \mu_0, \lambda_0)$  and an estimate  $H_0^*$  of  $V_0 = \nabla_{xx}^2 L(z_0)$ ,  $H_0^*$  is a symmetric and positive definite matrix, positive constants  $0 < \delta < \sigma < 1, m_0 > 0$  is a positive constant. Set  $k = 0$ ;

**Step 1:** For given  $z_k$  and  $H_k^*$ , solve the subproblem

$$\begin{aligned} \min \quad & \nabla f(x_k)^T s + \frac{1}{2} s^T H_k^{*-1} s, \\ \text{s.t.} \quad & g_\alpha(x_k) + \nabla g_\alpha(x_k)^T s = 0, \\ & g_\beta(x_k) + \nabla g_\beta(x_k)^T s \leq 0, \\ & h(x_k) + \nabla h(x_k)^T s = 0, \end{aligned} \quad (27)$$

and obtain the unique optimal solution  $d_k$ ;

**Step 2:**  $\alpha_k$  is chosen by the modified weak Wolfe-Powell (MWWP) step-size rule

$$L(z_k + \alpha_k d_k) \leq L(z_k) + \delta \alpha_k \nabla_x L(z_k)^T d_k, \quad (28)$$

and

$$\nabla_x L(z_k + \alpha_k d_k)^T d_k \geq \sigma \nabla_x L(z_k)^T d_k, \quad (29)$$

then let  $x_{k+1} = x_k + \alpha_k d_k$ .

**Step 3:** If  $z_{k+1}$  satisfies a prescribed termination criterion (18), stop. Otherwise, go to step 4;

**Step 4:** Let  $\tilde{m} = \min\{k+1, m_0\}$ . Update  $H_0^*$  for  $\tilde{m}$  times to get  $H_{k+1}^*$  by formula (25).

**Step 5:** Set  $k = k+1$  and go to step 1.

Clearly, we note that the above algorithm is as simple as the limited memory SQP method, form storage and cost point of a view at each iteration.

In the following, we assume that the algorithm updates  $B_k^*$ -the inverse of  $H_k^*$ . The M-L-SQP-A1 with Hessian approximation  $B_k^*$  can be stated as follows.

### Modified limited memory SQP algorithm 2 for (2) (M-L-SQP-A2)

**Step 0:** Star with an initial point  $z_0 = (x_0, \mu_0, \lambda_0)$  and an estimate  $B_0^*$  of  $V_0 = \nabla_{xx}^2 L(z_0)$ ,  $B_0^*$  is a symmetric and positive definite matrix, positive constants  $0 < \delta < \sigma < 1, m_0 > 0$  is a positive constant. Set  $k = 0$ ;

**Step 1:** For given  $z_k$  and  $B_k^*$ , solve the subproblem  $QP(z_k, B_k^*)$  and obtain the unique optimal solution  $d_k$ ;

**Step 2:** Let  $\tilde{m} = \min\{k+1, m_0\}$ . Update  $B_k^*$  with the triples  $\{s_i, y_i, A_i^*\}_{i=k-\tilde{m}+1}^k, i.e.,$  for  $l = k - \tilde{m} + 1, \dots, k$ , compute

$$B_k^{l+1*} = B_k^{l*} - \frac{B_k^{l*} s_l s_l^T B_k^{l*}}{s_l^T B_k^{l*} s_l} + \frac{y_l^* y_l^{*T}}{y_l^{*T} s_l}, \quad (30)$$

where  $s_l = x_{l+1} - x_l$ ,  $y_l^* = y_l + A_l^* s_l$  and  $B_k^{k-\bar{m}+1*}$  for all  $k$ .

Note that M-L-SQP-A1 and M-L-SQP-A2 are mathematically equivalent. In the next section, we will establish the global convergence of M-L-SQP-A2.

### 4. Convergence analysis of M-L-SQP-A2

Let  $x^*$  be a local optimal solution and  $z^* = (x^*, \mu^*, \lambda^*)$  be the corresponding Kuhn-Tucker triple of problem (1). In order to get the global convergence of M-L-SQP-A2, the following assumptions are needed.

**Assumption A.** 1)  $f, h_i$  and  $g_i$  are twice continuously differentiable functions for all  $x \in S$  and  $S$  is bounded.

2)  $\{\nabla h_i(x^*), i \in E\} \cup \{\nabla g_i(x^*), j \in I^*\}$  are positive linear independence.

3) (Strict complementarity) For  $j \in I^*$ ,  $\mu_j^* > 0$ .

(iv)  $s^T V s > 0$  for all  $s \neq 0$  with  $\nabla h_i(x^*)^T s = 0, i \in E$  and  $\nabla g_i(x^*)^T s = 0, j \in I^*$ , where  $V = \nabla_{xx}^2 L(z^*)$ .

(v)  $\{z_k\}$  converges to  $z^*$  where  $\nabla_x L(z^*) = 0$ .

(vi) The Lagrangian function  $L(z)$  is convex for all  $z \in S$ .

Assumption A(vi) implies that there exists a constant  $H > 0$  such that

$$\|V\| \leq H, \quad \forall z \in S. \quad (31)$$

Due to the strict complementary Assumption A(3), at a neighborhood of  $z^*$ , the method (26) is equivalent to the following equality constrained quadratic programming:

$$\begin{aligned} \min \quad & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k^* s, \\ \text{s.t.} \quad & g_\alpha(x_k) + \nabla g_\alpha(x_k)^T s = 0, \\ & h(x_k) + \nabla h(x_k)^T s = 0. \end{aligned} \quad (32)$$

Without loss of generality for the locally convergent analysis, we may discuss that there are only active constraints in (2). Then (18) becomes the following system with  $B_k^*$  instead of  $V_k$ :

$$\begin{pmatrix} B^* & \nabla g_\alpha(x) & \nabla h(x) \\ \nabla g_\alpha(x)^T & 0 & 0 \\ \nabla h(x)^T & 0 & 0 \end{pmatrix} \begin{pmatrix} d_{x_k} \\ d_{\mu_k} \\ d_{\lambda_k} \end{pmatrix} = - \begin{pmatrix} \nabla_x L(z_k) \\ g_\alpha(x_k) \\ h(x_k) \end{pmatrix} = -H(z_k) \quad (33)$$

In the case of only considering active constraints, we

can suppose that

$$W_k = \begin{pmatrix} V_k & \nabla g_\alpha(x) & \nabla h(x) \\ \nabla g_\alpha(x)^T & 0 & 0 \\ \nabla h(x)^T & 0 & 0 \end{pmatrix} \quad (34)$$

And

$$D_{H,K} = \begin{pmatrix} B_k^* & \nabla g_\alpha(x) & \nabla h(x) \\ \nabla g_\alpha(x)^T & 0 & 0 \\ \nabla h(x)^T & 0 & 0 \end{pmatrix}, \quad (35)$$

when  $B_k^*$  is close to  $V_k$ ,  $D_{H,K}$  is close to  $W_k$ .

**Lemma 4.1** Let Assumption A hold. Then there exists a positive number  $M_1$  such that

$$\frac{\|y_k^*\|^2}{s_k^T y_k^*} \leq M_1, \quad k = 0, 1, 2, \dots.$$

**Proof.** By Assumption A, then there exists a positive number  $M_0$  such that (see [36])

$$\frac{\|y_k\|^2}{s_k^T y_k} \leq M_0, \quad k \geq 0. \quad (36)$$

Since the function  $L(x)$  is convex, then we have  $L(z_{k+1}) - L(z_k) \geq \nabla_x L(z_k)^T s_k$  and  $L(z_k) - L(z_{k+1}) \geq -\nabla_x L(z_{k+1})^T s_k$ , the above two inequalities together with the definition of  $\tilde{A}_k$  imply that

$$|\tilde{A}_k| \leq \frac{|s_k^T y_k|}{\|s_k\|^2}. \quad (37)$$

Using the definition of  $y_k^*$ , we get

$$s_k^T y_k^* = s_k^T y_k + \max\{\tilde{A}_k, 0\} \geq s_k^T y_k \quad (38)$$

and

$$\|y_k^*\| \leq \|y_k\| + \|\max\{\tilde{A}_k, 0\} s_k\| \leq \|y_k\| + \|y_k\| \leq 2\|y_k\|, \quad (39)$$

where the second inequality of (39) follows (37). Combining (38), (39), and (36), we obtain:

$$\frac{\|y_k^*\|^2}{s_k^T y_k^*} \leq \frac{4\|y_k\|^2}{s_k^T y_k} \leq 4M_0.$$

Let  $M_1 = 4M_0$ , we get the conclusion of this lemma. The proof is complete.

**Lemma 4.2** Let  $B_k$  is generated by (30). Then we have

$$\det(B_{k+1}^*) = \det(B_k^{k-\bar{m}+1*}) \prod_{l=k-\bar{m}+1}^k \frac{s_l^T y_l^*}{s_l^T B_l^* s_l}, \quad (40)$$

where  $\det(B_k^*)$  denotes the determinant of  $B_k^*$ .

**Proof.** To begin with, we take the determinant in both sides of (20)

$$\begin{aligned}
 \det(B_{k+1}(1)) &= \det(B_k(1) \left( I - \frac{s_k s_k^T B_k(1)}{s_k^T B_k(1) s_k} + \frac{B_k(1)^{-1} y_k y_k^T}{s_k^T y_k} \right)) \\
 &= \det(B_k(1)) \det \left( I - \frac{s_k s_k^T B_k(1)}{s_k^T B_k(1) s_k} + \frac{B_k(1)^{-1} y_k y_k^T}{s_k^T y_k} \right) \\
 &= \det(B_k(1)) \left[ \left( 1 - s_k^T \frac{B_k(1) s_k}{s_k^T B_k(1) s_k} \right) \left( 1 + (B_k(1)^{-1} y_k)^T \frac{y_k}{y_k^T s_k} \right) \right. \\
 &\quad \left. - \left( -s_k^T \frac{y_k}{y_k^T s_k} \right) \left( \frac{B_k(1) s_k}{s_k^T B_k(1) s_k} \right)^T B_k(1)^{-1} y_k \right] \\
 &= \det(B_k(1)) \frac{y_k^T s_k}{s_k^T B_k(1) s_k},
 \end{aligned}$$

where the third equality follows from the formula (see, e.g., [37] **Lemma 7.6**)

$$\det(I + u_1 u_2^T + u_3 u_4^T) = (1 + u_1^T u_2)(1 + u_3^T u_4) - (u_1^T u_4)(u_2^T u_3).$$

Therefore, there is also a simple expression for the determinant of (30)

$$\det(B_{k+1}^*) = \det(B_k^{k-\tilde{m}+1*}) \prod_{l=k-\tilde{m}+1}^k \frac{s_l^T y_l^*}{s_l^T B_l^* s_l}.$$

Then we complete the proof.

**Lemma 4.3** Let Assumption A hold. Then there exists a positive constant  $\delta_1$  such that

$$\|s_k\| \geq \delta_1 \eta_k, \text{ where } \eta_k = \frac{-\nabla_x L(z_k)^T d_k}{\|d_k\|}.$$

**Proof.** By Assumption A, we have

$$\begin{aligned}
 (\nabla_x L(z_{k+1}) - \nabla_x L(z_k))^T d_k &= \\
 \alpha_k d_k^T \int_0^1 V(z_k + t\alpha_k d_k) dt &\leq \alpha_k \|d_k\|^2 (H+1).
 \end{aligned}$$

On the other hand, using (29), we get

$$(\nabla_x L(z_{k+1}) - \nabla_x L(z_k))^T d_k \geq -(1-\sigma) \nabla_x L(z_k)^T d_k.$$

Therefore,  $\|s_k\| \geq \frac{1-\sigma}{H+1} \eta_k$ , let  $\delta_1 = \frac{1-\sigma}{H+1}$ . The proof is complete.

Using Assumption A, it is not difficult to get the following lemma.

**Lemma 4.4** Let Assumption A hold. Then the sequence  $\{L(z_k)\}$  monotonically decreases, and  $z_k \in S$  for all  $k \geq 0$ . Moreover,

$$\sum_{k=0}^{\infty} (-\alpha_k \nabla_x L(z_k)^T d_k) < +\infty.$$

Similar to Lemma 2.6 in [38], it is not difficult to get the following lemma. Here we also give the proof process.

**Lemma 4.5** If the sequence of nonnegative numbers  $m_k$  ( $k = 0, 1, \dots$ ) satisfy

$$\prod_{j=0}^k m_j \geq c_1^k, \quad c_1 > 0, \quad k = 1, 2, \dots, \quad (41)$$

then  $\limsup_k m_k > 0$ .

**Proof.** We will get this result by contradiction. Assume that  $\limsup_k m_k = 0$ , then, for  $0 < \varepsilon_1 < c_1$ , there exists  $k_1 > 0$ , such that  $m_k < \varepsilon_1$  for all  $k \geq k_1$ . Hence, for all  $k > k_1$ ,

$$\begin{aligned}
 c_1^k &\leq \prod_{j=0}^{k_1-1} m_j \prod_{j=k_1}^k \varepsilon_1 \\
 +\infty &= \limsup_k \left( \frac{c_1}{\varepsilon_1} \right)^k \leq \left( \prod_{j=0}^{k_1-1} m_j \right) \varepsilon_1^{1-k_1} < +\infty,
 \end{aligned}$$

which is a contradiction, thus,  $\limsup_k m_k > 0$ .

**Lemma 4.6** Let  $\{x_k\}$  be generated by M-L-SQP-A2 and Assumption A hold. If  $\liminf_{k \rightarrow \infty} \|\nabla_x L(z_k)\| > 0$ , then, there exists a constant  $\varepsilon_0 > 0$  such that

$$\prod_{j=0}^k \eta_j \geq (\varepsilon_0)^{k+1}, \text{ for all } k \geq 0.$$

**Proof.** Assume that  $\liminf_{k \rightarrow \infty} \|\nabla_x L(z_k)\| > 0$ , i.e., there exists a constant  $c_2 > 0$  such that

$$\|\nabla_x L(z_k)\| \geq c_2, \quad k = 0, 1, 2, \dots \quad (42)$$

Now we prove that the update matrix  $B_{k+1}^*$  will always be generated by the update formula (30), i.e.,  $B_{k+1}^*$  inherits the positive definiteness of  $B_k^*$  or  $s_k^T y_k^* > 0$  always holds. For  $k = 0$ , this conclusion holds at hand. For all  $k \geq 1$ , assume that  $B_k^*$  is positive definite. We will deduce that  $s_k^T y_k^* > 0$  always holds from the following three cases.

Case 1. If  $\tilde{A}_k > 0$ . By the definition of  $y_k^*$  and Assumption A, we have

$$s_k^T y_k^* = s_k^T y_k + \max\{\tilde{A}_k, 0\} > s_k^T y_k \geq 0.$$

Case 2. If  $\tilde{A}_k < 0$ . By the definition of  $y_k^*$ , (24), and Assumption A, we get  $s_k^T y_k^* = s_k^T y_k > 0$ .

Case 3. If  $\tilde{A}_k = 0$ . By the definition of  $y_k^*$ , (29), Assumption A,  $d_k = -B_k^{*-1} \nabla_x L(z_k)$ , and the positive definiteness of  $B_k^*$ , we obtain

$$s_k^T y_k^* = s_k^T y_k \geq -(1-\sigma) \alpha_k d_k^T \nabla_x L(z_k) = (1-\sigma) \alpha_k d_k^T B_k^* d_k > 0,$$

So, we have  $s_k^T y_k^* > 0$ , and  $B_{k+1}^*$  will be generated by the update formula (30). Thus, the update matrix  $B_{k+1}^*$  will always be generated by the update formula (30).

Taking the trace operation in both sides of (30), we get

$$\begin{aligned}
 \text{Tr}(B_{k+1}^*) &= \\
 \text{Tr}(B_k^{k-\tilde{m}+1*}) - \sum_{l=k-\tilde{m}+1}^k \frac{\|B_l^* s_l\|^2}{s_l^T B_l^* s_l} + \sum_{l=k-\tilde{m}+1}^k \frac{\|y_l^*\|^2}{s_l^T y_l^*}, &\quad (43)
 \end{aligned}$$

where  $Tr(B_k^*)$  denotes the trace of  $B_k^*$ . Repeating this trace operation, we have

$$\begin{aligned} Tr(B_{k+1}^*) &= Tr(B_k^{k-\tilde{m}+1*}) - \sum_{l=k-\tilde{m}+1}^k \frac{\|B_l^* s_l\|^2}{s_l^T B_l^* s_l} + \sum_{l=k-\tilde{m}+1}^k \frac{\|y_l^*\|^2}{s_l^T y_l^*} \\ &= \dots \\ &= Tr(B_0^*) - \sum_{l=0}^k \frac{\|B_l^* s_l\|^2}{s_l^T B_l^* s_l} + \sum_{l=0}^k \frac{\|y_l^*\|^2}{s_l^T y_l^*}. \end{aligned} \tag{44}$$

Combining (42), (44),  $d_k = -B_k^{*-1} \nabla_x L(z_k)$ , and Lemma 4.1, we obtain

$$Tr(B_{k+1}^*) \leq Tr(B_0^*) - \sum_{l=0}^k \frac{c_2^2}{\nabla_x L(z_j)^T H_j^* \nabla_x L(z_j)} + (k+1)M_1. \tag{45}$$

Using  $B_{k+1}^*$  is positive definite, we have  $Tr(B_{k+1}^*) > 0$ . By (45), we obtain

$$\sum_{l=0}^k \frac{c_2^2}{\nabla_x L(z_j)^T H_j^* \nabla_x L(z_j)} \leq \frac{Tr(B_0^*) + (k+1)M_1}{c_2^2} \tag{46}$$

and

$$Tr(B_{k+1}^*) \leq Tr(B_0^*) + (k+1)M_1. \tag{47}$$

By the geometric-arithmetic mean value formula we get

$$\prod_{j=0}^k \nabla_x L(z_j)^T H_j^* \nabla_x L(z_j) \geq \left[ \frac{(k+1)c_2^2}{Tr(B_0^*) + (k+1)M_1} \right]^{k+1}. \tag{48}$$

Using Lemma 4.2, (30), and (38), we have

$$\begin{aligned} \det(B_{k+1}^*) &= \det(B_k^{k-\tilde{m}+1*}) \prod_{l=k-\tilde{m}+1}^k \frac{s_l^T y_l^*}{s_l^T B_l^* s_l} \\ &\geq \det(B_k^{k-\tilde{m}+1*}) \prod_{l=k-\tilde{m}+1}^k \frac{s_l^T y_l}{s_l^T B_l^* s_l} \\ &\geq \det(B_k^{k-\tilde{m}+1*}) \prod_{l=k-\tilde{m}+1}^k \frac{1-\sigma}{\alpha_l} \\ &\geq \dots \\ &\geq \det(B_0^*) \prod_{j=0}^k \frac{1-\sigma}{\alpha_j}. \end{aligned}$$

This implies

$$\frac{\det(B_0^*)}{\det(B_{k+1}^*)} \leq \prod_{j=0}^k \frac{\alpha_j}{1-\sigma}. \tag{49}$$

By using the geometric-arithmetic mean value formula again, we get

$$\det(B_{k+1}^*) \leq \left[ \frac{Tr(B_{k+1}^*)}{n} \right]^n. \tag{50}$$

Using (47), (49) and (50), we obtain

$$\begin{aligned} \prod_{j=0}^k \frac{\alpha_j}{1-\sigma} &\geq \frac{\det(B_0^*) n^n}{[Tr(B_0^*) + (k+1)M_1]^n} \\ &\geq \frac{1}{k+1} \frac{\det(B_0^*) n^n}{[Tr(B_0^*) + M_1]^n} \\ &\geq \left( \frac{1}{\exp(n)} \right)^{k+1} \min \left\{ \frac{\det(B_0^*) n^n}{[Tr(B_0^*) + M_1]^n}, 1 \right\} \\ &\geq \min \left\{ \frac{\det(B_0^*) n^n}{[Tr(B_0^*) + M_1]^n}, 1 \right\} \\ &\geq C_3^{k+1} \end{aligned} \tag{51}$$

where  $c_3 \geq \left( \frac{1}{\exp(n)} \right) \min \left\{ \frac{\det(B_0^*) n^n}{[Tr(B_0^*) + M_1]^n}, 1 \right\}$ . Let

$$\cos \theta_j = \frac{-\nabla_x L(z_j)^T d_j}{\|\nabla_x L(z_j)\| \|d_j\|}.$$

Multiplying (48) with (51), for all  $k \geq 0$ , we get

$$\begin{aligned} \prod_{j=0}^k \|s_k\| \|\nabla_x L(z_j)\| \cos \theta_j &\geq c_3^{k+1} \left[ \frac{(k+1)c_2^2}{Tr(B_0^*) + (k+1)M_1} \right]^{k+1} \\ &\geq \left[ \frac{c_3 c_2^2}{Tr(B_0^*) + M_1} \right]^{k+1}. \end{aligned} \tag{52}$$

According to Lemma 4.4 and Assumption A we know that there exists a constant  $M'_2 > 0$

such that

$$\|s_k\| = \|x_{k+1} - x_k\| \leq \|x_{k+1}\| + \|x_k\| \leq 2M'_2. \tag{53}$$

Combining the definition of  $\eta_k$  and (53), and noting that  $\|\nabla_x L(z_j)\| \cos \theta_j = \eta_j$ , we get for all  $k \geq 0$ ,

$$\prod_{j=0}^k \eta_j \geq \left[ \frac{c_3 c_2^2}{(Tr(B_0^*) + M_1) 2M'_2} \right]^{k+1} = \varepsilon_0^{k+1}.$$

The proof is complete.

Now we establish the global convergence theorem for M-L-SQP-A2.

**Theorem 4.1** Let Assumption (i) hold and let the sequence  $\{z_k\}$  be generated by M-L-SQP-A2. Then we have

$$\liminf_{k \rightarrow \infty} \|\nabla_x L(z_k)\| = 0. \tag{54}$$

**Proof.** By Lemma 4.3 and (28), we get

$$\begin{aligned} L(z_{k+1}) &\leq L(z_k) - \delta \|s_k\| \eta_k \\ &\leq L(z_k) - \delta \tilde{\delta} \eta_k^2. \end{aligned} \tag{55}$$

By (55), we have  $\sum_{k=0}^{\infty} \eta_k^2 < +\infty$ , this implies that



$$\lim_{k \rightarrow \infty} \eta_k = 0. \quad (56)$$

Therefore, relation (54) can be obtained from (56) and Lemma 4.6 directly.

## 5. Conclusion

For further research, we should study the properties of the modified limited memory SQP method under weak conditions. Moreover, numerical experiments for practically constrained problems should be done in the future.

## 6. References

- [1] P. T. Boggs, J. W. Tolle and P. Wang, "On The Local Convergence of Quasi-Newton Methods for Constrained Optimization," *SIAM Journal on Control and Optimization*, Vol. 20, No. 2, 1982, pp. 161-171.
- [2] F. H. Clarke, "Optimization and Nonsmooth Analysis," Wiley, New York, 1983.
- [3] T. F. Coleman and A. R. Conn, "Nonlinear Programming Via Exact Penalty Function: Asymptotic Analysis," *Mathematical Programming*, Vol. 24, No. 1, 1982, pp. 123-136.
- [4] M. Fukushima, "A Successive Quadratic Programming Algorithm with Global and Superlinear Convergence Properties," *Mathematical Programming*, Vol. 35, No. 3, 1986, pp. 253-264.
- [5] M. J. D. Powell, "The Convergence of Variable Metric methods for Nonlinearly Constrained Optimization Calculations," In O. L. Mangasarian, R. R. Meyer and S. M. Robinson Eds., *Nonlinear Programming 3*, Academic Press, New York, 1978, pp. 27-63.
- [6] W. Sun, "Newton's Method and Quasi-Newton-SQP Method for General  $LC^1$  Constrained Optimization," *Applied Mathematics and Computation*, Vol. 92, No. 1, 1998, pp. 69-84.
- [7] F. C. Thomas and Q. C. Andrew, "On The Local Convergence of a Quasi-Newton Methods for The Nonlinear Programming Problem," *SIAM Journal on Numerical Analysis*, Vol. 21, No. 4, 1984, pp. 755-769.
- [8] Y. Yuan and W. Sun, "Theory and Methods of Optimization," Science Press of China, Beijing, 1999.
- [9] G. Yuan, "Modified Nonlinear Conjugate Gradient Methods with Sufficient Descent Property for Large-Scale Optimization Problems," *Optimization Letters*, Vol. 3, No. 1, 2009, pp. 11-21.
- [10] G. L. Yuan and X. W. Lu, "A New Line Search Method with Trust Region for Unconstrained Optimization," *Communications on Applied Nonlinear Analysis*, Vol. 15, No. 1, 2008, pp. 35-49.
- [11] G. L. Yuan and X. W. Lu, "A Modified PRP Conjugate Gradient Method," *Annals of Operations Research*, Vol. 166, No. 1, 2009, pp. 73-90.
- [12] G. Yuan, X. Lu and Z. Wei, "A Conjugate Gradient Method with Descent Direction for Unconstrained Optimization," *Journal of Computational and Applied Mathematics*, Vol. 233, No. 2, 2009, pp. 519-530.
- [13] G. L. Yuan and Z. X. Wei, "New Line Search Methods for Unconstrained Optimization," *Journal of the Korean Statistical Society*, Vol. 38, No. 1, 2009, pp. 29-39.
- [14] W. C. Davidon, "Variable Metric Methods for Minimization," *SIAM Journal on Optimization*, Vol. 1, No. 1, 1991, pp. 1-17.
- [15] D. H. Li and M. Fukushima, "A Modified BFGS Method and Its Global Convergence in Non-convex Minimization," *Journal of Computational and Applied Mathematics*, Vol. 129, No. 1-2, 2001, pp. 15-35.
- [16] D. H. Li and M. Fukushima, "On The Global Convergence of The BFGS Methods for Non-convex Unconstrained Optimization Problems," *SIAM Journal on Optimization*, Vol. 11, No. 4, 2000, pp.1054-1064.
- [17] M. J. D. Powell, "A New Algorithm for Unconstrained Optimization," In J. B. Rosen, O. L. Mangasarian and K. Ritter Eds., *Nonlinear Programming*, Academic Press, New York, 1970.
- [18] Z. Wei, G. Yu, G. Yuan and Z. Lian, "The Superlinear Convergence of A Modified BFGS-type Method for Unconstrained Optimization," *Computational Optimization and Applications*, 29(2004), pp. 315-332.
- [19] J. Z. Zhang, N. Y. Deng and L. H. Chen, "New Quasi-Newton Equation and Related Methods for Unconstrained Optimization," *Journal of Optimization Theory and Applications*, Vol. 102, No. 1, pp. 147-167.
- [20] Z. Wei, G. Li and L. Qi, "New Quasi-Newton Methods for Unconstrained Optimization Problems," *Applied Mathematics and Computation*, Vol. 175, No. 2, 2006, pp. 1156-1188.
- [21] G. L. Yuan and Z. X. Wei, "Convergence Analysis of A Modified BFGS Method on Convex Minimizations," *Computational Optimization and Applications*, doi: 10.1007/s10 589-008-9219-0.
- [22] R. H. Byrd, J. Nocedal and R. B. Schnabel, "Representations of Quasi-Newton Matrices and Their Use in Limited Memory Methods," *Mathematical Programming*, Vol. 63, No. 1-3, 1994, pp. 129-156.
- [23] C. G. Broyden, J. E. Dennis and J. J. Moré, "On The Local and Superlinear Convergence of Quasi-Newton Methods," *IMA Journal of Applied Mathematics*, Vol. 12 No. 3, 1973, pp. 223-246.
- [24] R. H. Byrd and J. Nocedal, "A Tool for The Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization," *SIAM Journal on Numerical Analysis*, Vol. 26, No. 3, 1989, pp. 727-739.
- [25] R. H. Byrd, J. Nocedal and Y. Yuan, "Global Convergence of A Class of Quasi-Newton Methods on Convex Problems," *SIAM Journal on Numerical Analysis*, Vol. 24, No. 5, 1987, pp.1171-1189.
- [26] J. E. Dennis and J. J. Moré, "A Characterization of Super-Linear Convergence and Its Application to Quasi-Newton Methods," *Mathematics of Computation*, Vol. 28, No. 126, 1974, pp. 549-560.
- [27] A. Griewank and Ph. L. Toint, "Local Convergence

- Analysis for Partitioned Quasi-Newton Updates,” *Numerische Mathematik*, Vol. 39, No. 3, 1982, pp. 429-448.
- [28] A. Perry, “A Class of Conjugate Algorithms with a Two Step Variable Metric Memory,” Discussion paper, No. 269, Center for Mathematical Studies in Economics and Management Science, Northwestern University, 1977.
- [29] D. F. Shanno, “On the Convergence of A New Conjugate Gradient Algorithm,” *SIAM Journal on Numerical Analysis*, 15, No. 6, 1978, pp. 1247-1257.
- [30] Z. Wei, L. Qi and X. Chen, “An SQP-type Method and Its Application in Stochastic Programming,” *Journal of Optimization Theory and Applications*, Vol. 116, No. 1, 2003, pp. 205-228.
- [31] G. L. Yuan and Z. X. Wei, “The Superlinear Convergence Analysis of A Nonmonotone BFGS Algorithm on Convex Objective Functions,” *Acta Mathematica Sinica*, Vol. 24, No. 1, 2008, pp. 35-42.
- [32] Y. Dai, “Convergence Properties of the BFGS Algorithm,” *SIAM Journal on Optimization*, Vol. 13, No. 3, 2003, pp. 693-701.
- [33] W. F. Mascarenhas, “The BFGS Method with Exact Line Searches Fals for Non-convex Objective Functions,” *Mathematical Programming*, Vo. 99 No. 1, 2004, pp. 49-61.
- [34] R. H. Byrd, P. Lu, J. Nocedal and C. Zhu, “A Limited Memory Algorithm for Bound Constrained Optimization,” *SIAM Journal on Scientific Computing*, Vol. 16, No. 5, 1995, pp. 1190-1208.
- [35] M. J. D. Powell, “A fast Algorithm for Nonlinear Constrained Optimization Calculations,” *Lecture Notes in Mathematics*, Vol. 630, Springer, Berlin, pp. 144-157.
- [36] M. J. D. Powell, “Some Properties of The Variable Metric Algorithm,” In F. A. Lootsma Ed., *Numerical methods for Nonlinear Optimization*, Academia Press, London, 1972.
- [37] J. E. Dennis and J. J. Moré, “Quasi-Newton Methods, Motivation and Theory,” *SIAM Review*, Vol. 19, No. 1, 1977, pp. 46-89.
- [38] J. Y. Han and G. H. Liu, “Global Convergence Analysis of a New Nonmonotone BFGS Algorithm on Convex Objective Functions,” *Computational Optimization and Applications*, Vol. 7, No. 3, 1997, pp. 277-289.