Journal of Computational Mathematics, Vol.18, No.2, 2000, 133–140.

SUBSPACE SEARCH METHOD FOR A CLASS OF LEAST SQUARES PROBLEM*[1)]

Zi-Luan Wei
(*Institute of Computational Mathematics and Scientific / Engineerin Computing, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100080, China*)

Abstract

A subspace search method for solving a class of least squares problem is presented in the paper. The original problem is divided into many independent subproblems, and a search direction is obtained by solving each of the subproblems, as well as a new iterative point is determined by choosing a suitable steplength such that the value of residual norm is decreasing. The convergence result is also given. The numerical test is also shown for a special problem.

Key words: Subspace search method, A class of least squqres problem, Convergence analysis.

1. Introduction

Now the least squares problem is considered as follows:

$$Min\ r(x,y) = \frac{1}{2}\|Ax + By - b\|^2 \quad s.t.\ \ x\ \geq 0 \tag{1.1}$$

where $A \in R^{m\times t}$, $B \in R^{m\times q}$, and $b \in R^m$ are given constant matrices and vectors, respectively.

These problems arise in many areas of applications, such as scientific and engineering computing, physics, statistics, fitted curve, economic, mathematical programming, social science, and as a component part of some large computation problem, as an example, a nonlinear least squares problem is approximated locally by using of various linearization schemes.

This problem often is to analyze and solve a systems of linear algebraic equations, which may be overdetermined, underdetermined , or exactly determined, and may or may not be consistent with linear and inequality constraints. Many successful algorithms for solving this problems have been studied during past decades, and the decomposition method (QR) is a popular approach to use for solving the problems. G.H.Golub has contributed many significant ideas and algorithms relating to this problems in both theoretical and practice, and a practical and useful numerical methods for solving this problems were stated in [2] in details. A different techniques have also developed for the problems.

* Received August 12, 1996.

1) This research supported partially by The National Natural Science Foundation of China–19771079 and 19601035, and State Key Laboratory of Scientific and Engineering Computing..

In this paper, we present a subspace search method for solving the problem (1.1). The main steps of the algorithm are to divide the problem (1.1) into independent subproblems at an initial feasible point and solve each of these subproblems to obtain a search direction, and then to determine a new feasible iterative point by choosing a suitable steplength such that the value of the residual norm is decreasing. the convergence of the algorithm is proved under certain assumptions. The main feature of the algorithm is that large scale problom (1.1) can be transformed into many small independent subproblems, and all the subproblems can be solved simultaneously.

This paper is organized as follows. in section 2 we describe the algorithm. The convergence results are proved under certain assumptions in section 3, A modification algorithm and numerical test is also given in section 4.

2. Derivation of the Algorithm

This section deals with the basic ideas that how to construct the subspace serach algorithm for solving the problem (1.1). Without loss of generality, assume that vector $x \in R^t$ and $y \in R^q$ can be divided into $(x_1^T, x_2^T, \cdots, x_{l_1}^T)^T$ and $(y_1^T, y_2^T, \cdots, y_{l_2}^T)^T$, and $x_i \in R^{n_i}$, $y_j \in R^{n_j}$, respectively, and that $\sum_{i=1}^{l_1} n_i = t, \sum_{j=1}^{l_2} n_j = q$. Accordingly, matrix A and B can be also divided into following form $A = (A_1, A_2, \cdots, A_{l_1}) \quad B = (B_1, B_2, \cdots, B_{l_2})$, where $A_i \in R^{m \times n_i}$ $(i = 1, 2, \cdots, l_1)$, $B_j \in R^{m \times n_j}$ $(j = 1, 2, \cdots, l_2)$. Therefore, the function $r(x, y)$ can be expressed as following form

$$r(x,y) = \frac{1}{2}\| \sum_{i=1}^{l_1} A_i x_i + \sum_{j=1}^{l_2} B_j y_j - b\|^2 \tag{2.1}$$

Now we analyze the properties of function $r(x, y)$ in the neighbourhood of a given initial vector $(\bar{x}, \bar{y})$. Assume that (x, y) is in the neighbourhood of $(\bar{x}, \bar{y})$ and let

$$x = \bar{x} + (x - \bar{x}) \quad y = \bar{y} + (y - \bar{y}) \tag{2.2}$$

Substituting (2.2) into (2.1),and it is easy to derive that

$$\begin{aligned} r(x,y) = \frac{1}{2}[\sum_{i=1}^{l_1} \|\bar{r} - A_i(x_i - \bar{x}_i)\|^2 + \sum_{j=1}^{l_2} \|\bar{r} - B_j(y_j - \bar{y}_j)\|^2 - (l_1 + l_2 - 1)\bar{r}] + \\ \frac{1}{2}\hat{r}(x,y) = \sum_{i=1}^{l_1} \xi_i(x_i, \bar{x}, \bar{y}) + \sum_{j=1}^{l_2} \zeta_j(y_j, \bar{x}, \bar{y}) - \frac{l_1 + l_2 - 1}{2}\bar{r} + \frac{1}{2}\hat{r}(x,y) \end{aligned} \tag{2.3}$$

where $\bar{r} = r(\bar{x}, \bar{y})$, and $\hat{r}(x, y)$ is an error of order $O(\|x - \bar{x}\|^2 + \|y - \bar{y}\|^2)$. Namely,

$$\begin{aligned} \hat{r}(x,y) = & \sum_{i=1}^{l_1} \sum_{j \neq i}^{l_1} (x_i - \bar{x}_i) A_i^T A_j (x_j - \bar{x}_j) + \sum_{i=1}^{l_2} \sum_{j \neq i}^{l_2} (y_i - \bar{y}_i) B_i^T B_j (y_j - \bar{y}_j) \\ & + 2\sum_{i=1}^{l_1} \sum_{j=1}^{l_2} (x_i - \bar{x}_i) A_i^T B_j (y_j - \bar{y}_j) \end{aligned} \tag{2.4}$$

and

$$\xi_i(x_i, \bar{x}, \bar{y}) = \frac{1}{2}\|\bar{r} - A_i(x_i - \bar{x}_i)\|^2, \qquad i = 1, 2, \cdots, l_1 \tag{2.5}$$

$$\zeta_j(y_j, \bar{x}, \bar{y}) = \frac{1}{2}\|\bar{r} - B_j(y_j - \bar{y}_j)\|^2 \qquad j = 1, 2, \cdots, l_2 \tag{2.6}$$

Let

$$\varphi(x, y, \bar{x}, \bar{y}) = \sum_{i=1}^{l_1} \xi_i(x_i, \bar{x}, \bar{y}) + \sum_{j=1}^{l_2} \zeta_j(y_j, \bar{x}, \bar{y}) - \frac{l_1 + l_2 - 1}{2}\bar{r} \tag{2.7}$$

be a function defined on $X_1 \times X_2 \times \cdots, \times X_{l_1} \times R^q$, where $X_i = \{x_i \in R^{n_i}, x_i \ \geq 0\}(i = 1, 2, \cdots, l_1)$. It follows from (2.7) that the function $\varphi(x, y, \bar{x}, \bar{y})$ is separable, and it is easy to verify that $\bar{r} = \varphi(\bar{x}, \bar{y}, \bar{x}, \bar{y})$, and the gradients of $r(x, y)$ and $\varphi(x, y, \bar{x}, \bar{y})$ are equal at $(\bar{x}, \bar{y})$. Assume the (x, y) belongs to a neighbourhood of some reference point $(\bar{x}, \bar{y})$, then the function $\varphi(x, y, \bar{x}, \bar{y})$ can be refered to an approximate locally expression of function $r(x, y)$ at $(\bar{x}, \bar{y})$, and

$$r(x, y) - \varphi(x, y, \bar{x}, \bar{y}) = \hat{r}(x, y) \tag{2.8}$$

From (2.8), it is easy to see that as long as δ is sufficiently small the difference between $r(x, y)$ and $\varphi(x, y, \bar{x}, \bar{y})$ can be contained in a given range for all $(x, y) \in N(\bar{x}, \bar{y}, \delta)$, where $N(\bar{x}, \bar{y}, \delta)$ denotes the Euclidean ball about $(\bar{x}, \bar{y})$ with radius δ. As a result, $r(x, y)$ can be approximated by the function $\varphi(x, y, \bar{x}, \bar{y})$ with an error $\frac{1}{2}\hat{r}(x, y)$ in the neighbourhood of $(\bar{x}, \bar{y})$. Thus, the original problem (1.1) can be replaced locally by the following problem

$$min \quad \varphi(x, y, \bar{x}, \bar{y}) \qquad s.t. \ \ x \ \geq 0 \tag{2.9}$$

There is no doubt that problem (2.9) can be transformed into minimizing the subproblem in each subspace as follows because $\varphi(x, y, \bar{x}, \bar{y})$ is separable function.

$$(P_1) \quad min \ \xi_i(x_i, \bar{x}, \bar{y}) = \frac{1}{2}\|\bar{r} - A_i(x_i - \bar{x}_i)\|^2 \ \ s.t. \ x_i \ \geq 0 \quad i = 1, 2, \cdots, l_1$$

$$(P_2) \quad min \ \ \zeta_j(y_j, \bar{x}, \bar{y}) = \frac{1}{2}\|\bar{r} - B_j(y_j - \bar{y}_j)\|^2 \quad j = 1, 2, \cdots, l_2$$

Suppose that $(\hat{x}, \hat{y})$ is an optimal solution to the problem (P_1) and (P_2), Obviously, $(\hat{x} - \bar{x}, \hat{y} - \bar{y})$ is a descent direction of $\varphi(x, y, \bar{x}, \bar{y})$ at $(\bar{x}, \bar{y})$, As a result, a new iterative point $x = \bar{x} + \alpha(\hat{x} - \bar{x})$, $y = \bar{y} + \alpha(\hat{y} - \bar{y})$ can be generated such that $r(x, y) < r(\bar{x}, \bar{y})$ if a suitable steplength $\alpha > 0$ is chosen. So a subspace search algorithm for solving the problem (1.1) can be constructed as follows.

Algorithm A:

Let $x^0 \geq 0, y^0$ be a given initial point, $\varepsilon > 0$ be some prescribed accuracy, and $\alpha \in (0, \ \bar{\delta}\ )$ $(\bar{\delta} = \frac{1}{l_1 + l_2 - 1})$, $k := 0$, and $(\ x^{k+1}, y^{k+1}\ )$ is obtained by the following steps.

(i) Let $\bar{x} = x^k, \bar{y} = y^k$, and solve the problem (P_1) and (P_2), and obtain an optimal solution $(\ \hat{x}^k, \hat{y}^k\ )$.

(ii) If $max_i \|A_i(\hat{x}_i^k - x_i^k)\ | \ \leq \varepsilon$, and $max_j \|B_j(\hat{y}_j^k - y_j^k)\| \ \leq \varepsilon$, then stop, and $(\hat{x}^k, \hat{y}^k)$ is an approximate solution for the problem (1.1), Otherwise, generate a new iterative point

$$x^{k+1} = x^k + \alpha(\hat{x}^k - x^k) \quad y^{k+1} = y^k + \alpha(\hat{y}^k - y^k)$$

let $k := k + 1$, and return to (i).

It follows from the definition of algorithm A that the main computational work is from solving problems (P_1) and (P_2) at each iteration.

3. Convergence Results

This section analyzes the convergence of the algorithm A. We prove that sequence x^k, y^k generated by the algorithm A converges to an optimal solution x^*, y^* of problem (1.1) under certain assumptions. Several lemmas are introduced in order to prove the convergent conclusion.

Lemma 3.1. *Suppose that $A \in R^{m\times t}, B \in R^{m\times q}, b \in R^m$ is given, then x^*, y^* is a minimizer of problem (1.1) if and only if there exists a $s^* \in R^t, s^* \geq 0$ such that x^*, y^* and s^* satisfy following conditions:*

$$A^T r^* = s^*, \quad (x^*)^T s^* = 0, \quad s^* \geq 0, \quad x^* \geq 0, \qquad B^T r^* = 0$$

where $r^* = b - Ax^* - By^*$. The above conditions can be also rewritten as follows

$$A_i^T r^* = s_i^*, \quad (x_i^*)^T s_i^* = 0, \quad s_i^* \geq 0, \quad x_i^* \geq 0, \qquad i = 1, 2, \cdots, l_1, \tag{3.1}$$

$$B_j^T r^* = 0 \qquad j = 1, 2, \cdots, l_2 \tag{3.2}$$

As is well known, r^* is called residual vector.

Lemma 3.2. *Suppose that $\varphi(x, y, \bar{x}, \bar{y})$ is defined by (2.7), and that (x^*, y^*) is an minimizer of $\varphi(x, y, \bar{x}, \bar{y})$ on $X_1 \times X_2 \times \cdots \times X_{l_1} \times R^q$, then*

$$\varphi(x^*, y^*, \bar{x}, \bar{y}) \leq \varphi(\bar{x}, \bar{y}, \bar{x}, \bar{y}) - \frac{1}{2}(\sum_{i=1}^{l_1} \|A_i d_i\|^2 + \sum_{j=1}^{l_2} \|B_j \bar{d}_j\|^2) \tag{3.3}$$

where $X_i = \{x_i \in R^{n_i}, \quad x_i \geq 0\}$, $d_i = x_i^* - \bar{x}_i$, $(i = 1, 2, \cdots, l_1)$, $\bar{d}_j = y_j^* - \bar{y}_j$, $(j = 1, 2, \cdots, l_2)$.

Proof. It follows from (2.5) and (2.6), $\xi_i(x_i, \bar{x}, \bar{y})$ $(i = 1, 2, \cdots, l_1)$ and $\zeta_j(y_j, \bar{x}, \bar{y})$ $(j = 1, 2, \cdots, l_2)$ are convex function, so is $\varphi(x, y, \bar{x}, \bar{y})$. Assume that (x^*, y^*) is an optimal solution for probem (2.9), as $\varphi(x, y, \bar{x}, \bar{y})$ is a separable function, so (x^*, y^*) can be obtained from problem (P_1) and (P_2). A necessary and sufficient condition for x_i^* and y_j^* being a solution of (P_1) and (P_2) is

$$-A_i^T[\bar{r} - A_i(x_i^* - \bar{x}_i)] = s_i^*, \quad s_i^* \geq 0, \quad (x_i^*)^T s_i^* = 0 \quad i = 1, 2, \cdots, l_1 \tag{3.4}$$

$$B_j^T[\bar{r} - B_j(y_j^* - \bar{y}_j)] = 0, \quad j = 1, 2, \cdots, l_2. \tag{3.5}$$

Let $d_i = x_i^* - \bar{x}_i, \bar{d}_j = y_j^* - \bar{y}_j$, and from the formula of inner product, it is easy to show that

$$(A_i d_i)^T (A_i d_i - \bar{r}) = -\bar{x}_i^T s_i^*, \qquad i = 1, 2, \cdots, l_1 \tag{3.6}$$

$$(B_j \bar{d}_j)^T (B_j \bar{d}_j - \bar{r}) = 0, \qquad j = 1, 2, \cdots, l_2. \tag{3.7}$$

On the other hand, $\varphi(x, y, \bar{x}, \bar{y})$ is a convex quadratic function, so one can easily derive that

$$\varphi(x^*,y^*,\bar{x},\bar{y}) = \varphi(\bar{x},\bar{y},\bar{x},\bar{y}) + d^T\nabla\varphi(\bar{x},\bar{y},\bar{x},\bar{y}) + \frac{1}{2}d^T\nabla^2\varphi(\bar{x},\bar{y},\bar{x},\bar{y})d$$

$$=\varphi(\bar{x},\bar{y},\bar{x},\bar{y}) - \sum_{i=1}^{l_1} d_i^T A_i\bar{r} - \sum_{j=1}^{l_2} \bar{d}_j^T B_j\bar{r} + \frac{1}{2}[\sum_{i=1}^{l_1} d_i^T A_i^T A_i d_i + \sum_{j=1}^{l_2} \bar{d}_j^T B_j^T B_j \bar{d}_j$$

$$=\varphi(\bar{x},\bar{y},\bar{x},\bar{y}) + \sum_{i=1}^{l_1}(A_i d_i)^T(A_i d_i - \bar{r}) + \sum_{j=1}^{l_2}(B_j\bar{d}_j)^T(B_j\bar{d}_j - \bar{r}) -$$

$$\frac{1}{2}(\sum_{i=1}^{l_1}\|A_i d_i\|^2 + \sum_{j=1}^{l_2}\|B_i\bar{d}_j\|^2) \tag{3.8}$$

By (3.6)-(3.8), it is straightforward to obtain

$$\varphi(x^*,y^*,\bar{x},\bar{y}) \le \varphi(\bar{x},\bar{y},\bar{x},\bar{y}) - \frac{1}{2}[\sum_{i=1}^{l_1}\|A_i d_i\|^2 + \sum_{j=1}^{l_2}\|B_j\bar{d}_j\|^2] - \sum_{i=1}^{l_1}\bar{x}_i^T s_i^*$$

This implies that (3.3) holds since $\bar{x}^T s^* \ge 0$, which proves the conclusion of the lemma.

Lemma 3.3. *With the same assumptions as in lemma 3.2, let* $u = \bar{x}+\alpha d, v = \bar{y}+\alpha\bar{d}$ *, then* $u \ge 0$ *, and*

$$\varphi(u,v,\bar{x},\bar{y}) \le \varphi(\bar{x},\bar{y},\bar{x},\bar{y}) - \frac{\alpha}{2}(\sum_{i=1}^{l_1}\|A_i d_i\|^2 + \sum_{j=1}^{l_2}\|B_j\bar{d}_j\|^2) \tag{3.9}$$

where $d = x^* - \bar{x}, \bar{d} = y^* - \bar{y}, \alpha \in (0,1)$.

From the assumptions and the convexity of function $\varphi(u,v,\bar{x},\bar{y})$, it is straightforward to show that (3.9) is true, and it is easy to verify that u is greater than zero.

Theorem 3.4. *Suppose that* $x^0 \ge 0, y^0$ *is an initial point,* $\alpha \in (0,\bar{\delta})$ *, and that sequence {* $\hat{x}^k, \hat{y}^k$ *} and {* x^k, y^k *} are generated by the algorithm A. Then either there exists an integer* $k_0 > 0$ *such that* $\hat{x}^{k_0} = x^{k_0}, \hat{y}^{k_0} = y^{k_0}$ *((* $\hat{x}^{k_0}$, $\hat{y}^{k_0}$ *) is an optimal solution of problem (1.1)) or*

$$\lim_{k\to\infty} A_i(\hat{x}_i^k - x_i^k) = 0 \quad i = 1,\cdots,l_1, \quad \lim_{k\to\infty} B_j(\hat{y}_j^k - y_j^k) = 0 \quad j = 1,\cdots,l_2, \tag{3.10}$$

and the accumulation point of { x^k *,*y^k *} is a solution of the problem (1.1).*

Proof. If there exists an integer $k_0 > 0$ such that $\hat{x}^{k_0} = x^{k_0}, \hat{y}^{k_0} = y^{k_0}$, then one can verify that ($\hat{x}^{k_0}$, $\hat{y}^{k_0}$) satisfeis conditions (3.1)-(3.2), so ($\hat{x}^{k_0}$, $\hat{y}^{k_0}$) is an optimal solution of problem (1.1), Otherwise, it follows from the definition $\varphi(x,y,\bar{x},\bar{y})$ and (2.3) that the following relationship holds for an arbitrary $\bar{x} \ge 0, \bar{y}$

$$|r(x,y) - \varphi(x,y,\bar{x},\bar{y})| = \frac{1}{2}|\hat{r}(x,y)| \tag{3.11}$$

By the Cauchy-Schwarz inequality, it is easy to derive that

$$\frac{1}{2}|\hat{r}(x,y)| \le \frac{l_1+l_2-1}{2}(\sum_{i=1}^{l_1}\|A_i d_i\|^2 + \sum_{j=1}^{l_2}\|B_j\bar{d}_j\|^2) \tag{3.12}$$

Assume that (x,y) belongs to the neighbourhood of $(\bar{x},\bar{y})$, then (3.11) and the inequalities (3.12) gives

$$r(x,y) \le \varphi(x,y,\bar{x},\bar{y}) + \frac{l_1+l_2-1}{2}(\sum_{i=1}^{l_1}\|A_i d_i\|^2 + \sum_{j=1}^{l_2}\|B_j\bar{d}_j\|^2) \tag{3.13}$$

Let $\hat{x}, \hat{y}$ is an minimizer of (2.9) with the starting point $(\bar{x}, \bar{y})$, and assume that $u = \bar{x} + \alpha(\hat{x} - \bar{x}), v = \bar{y} + \alpha(\hat{y} - \bar{y})$ and substitute them into (3.13), we have

$$r(u,v) \leq \varphi(u,v,\bar{x},\bar{y}) + \frac{l_1 + l_2 - 1}{2}\alpha^2 (\sum_{i=1}^{l_1} \|A_i d_i\|^2 + \sum_{j=1}^{l_2} \|B_j \bar{d}_j\|^2) \tag{3.14}$$

where $d = \hat{x} - \bar{x}, \bar{d} = \hat{y} - \bar{y}$ Substituting (3.9) into (3.14) yields

$$r(u,v) \leq \varphi(\bar{x},\bar{y},\bar{x},\bar{y}) - \frac{\alpha}{2}[1 - (l_1 + l_2 - 1)\alpha](\sum_{i=1}^{l_1} \|A_i d_i\|^2 + \sum_{j=1}^{l_2} \|B_j \bar{d}_j\|^2)$$

According to the above derivations and $\bar{r} = \varphi(\bar{x},\bar{y},\bar{x},\bar{y})$, it is straightforward to show that

$$r(u,v) \leq r(\bar{x},\bar{y}) - \frac{\alpha}{2}[1 - (l_1 + l_2 - 1)\alpha](\sum_{i=1}^{l_1} \|A_i d_i\|^2 + \sum_{j=1}^{l_2} \|B_j \bar{d}_j\|^2)$$

Obviously, if (u,v),$(\bar{x},\bar{y})$ and $(\hat{x},\hat{y})$ are chosen as (x^{k+1}, y^{k+1}), (x^k, y^k) and $(\hat{x}^k, \hat{y}^k)$ in Algorithm A, respectively, then the sequence $r(x^k, y^k)$ is convergent for every $\alpha \in (0, \bar{\delta})$. This impleis that (3.10) hold.

Let x^*, y^* be a limit of a convergent subsequence of the sequence x^k, y^k, and let $\hat{x}^*, \hat{y}^*$ be an accumulation point of the corresponding subsequence of $\hat{x}^k, \hat{y}^k$, By (3.10), it is obvious that $x^* = \hat{x}^*, y^* = \hat{y}^*$. Hence

$$\nabla\varphi(x^*, y^*, \bar{x}, \bar{y}) = \nabla\varphi(\hat{x}^*, \hat{y}^*, \bar{x}, \bar{y}) = \nabla r(x^*, y^*)$$

From (3.1)-(3.2), the necessary and sufficient conditions of optimality for problem (1.1) hold at x^*, y^* , which proves the second part of the conclusion of the theorem.

4. Modification of the Algorithm and Applications

This section discusses a possibility to modify the algorithm A. As is well known, using linear function $\varphi(x,y,\bar{x},\bar{y})$ to approximate objective function $r(x,y)$ to find a direction of movement, it is similar to the steepest descent method, and the steepest descent algorithm usually works quite well during the early stages of computational process. However, as a stationary is approached, it often occurs poor convergence. The main reason is that the term of the second order is essentially ignored. Therefore, we should expect that the direction generated at late stages will be updated by adding the second order to $\varphi(x,y,\bar{x},\bar{y})$ for improving the convergence. On the other hand, based on the convergence analysis in previous section, it is necessary to choose the steplength $\alpha \in (\ 0,\ \bar{\delta}\ )$ for the convergence of the algorithm A, this can not guarantee that a maximun descent of the value of residual norm is obtained at each iteration. Therefore, it is possible to choose a steplength in algorithm A such that the value of residual norm can decrease as much as possible at each iteration.

Assume that $\hat{x}, \hat{y}$ is generated by the step (i) of the algorithm A with starting point $\bar{x}, \bar{y}$, it is desirable to create a new point by using the following expression

$$x_i = \bar{x}_i + \lambda_i(\hat{x}_i - \bar{x}_i) = \bar{x}_i + \lambda_i d_i \quad y_j = \bar{y}_j + \lambda_j(\hat{y}_j - \bar{y}_j) = \bar{y}_j + \lambda_j \bar{d}_j \tag{4.1}$$

the set of parameters λ_i ($i = 1, 2, \cdots, l_1, j = l_1 + 1, l_1 + 2, \cdots, l_1 + l_2$) needs to be determined such that the new iterative point $x \geq 0$, and the value of the residual norm $r(x,y)$ is decreasing, that is, $r(x,y) < \bar{r}$. Clearly, the parameters $\lambda_i \in$ (0,1) ($i = 1, 2, \cdots, l_1$) can guarantee the feasibility of x, It follows from (3.1), it is easy to see that

$$r(\bar{x}_1, \cdots, \bar{x}_i + d_i, \cdots, \bar{x}_{l_1}, \bar{y}) - \bar{r} = -\bar{x}_i^T s_i^* - \frac{1}{2}\|A_i d_i\|^2 \tag{4.2}$$

Similarly, we have

$$r(\bar{x}, \bar{y}_1, \cdots, \bar{y}_j + \bar{d}_j, \cdots, \bar{y}_{l_2},) - \bar{r} = -\frac{1}{2}\|B_j \bar{d}_j\|^2 \tag{4.3}$$

If we choose the values of λ in (4.1) as follows

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{l_1+l_2} = \frac{1}{l_1 + l_2} \tag{4.4}$$

Thus, by the convexity of $r(x,y)$, it is easy to verify that

$$\begin{aligned} & r(x,y) - \bar{r} = r(\bar{x}_1 + \lambda_1 d_1, \cdots, \bar{x}_{l_1} + \lambda_{l_1} d_{l_1}, \bar{y}_1 + \lambda_{l_1+1}\bar{d}_1, \cdots, \bar{y}_{l_2} + \lambda_{l_1+l_2}\bar{d}_{l_2}) - \bar{r} \\ & \leq \lambda_1[r(\bar{x}_1 + \lambda_1 d_1, \bar{x}_2, \cdots, \bar{x}_{l_1}, \bar{y}) - \bar{r}] + \cdots + \lambda_{l_1}[r(\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_{l_1} + \lambda_{l_1} d_{l_1}, \bar{y}) - \bar{r}] \\ & \lambda_{l_1+1}[r(\bar{x}, \bar{y}_1 + \lambda_{l_1+1}\bar{d}_1, \bar{y}_2, \cdots, \bar{y}_{l_2}) - \bar{r}] + \cdots + \lambda_{l_1+l_2}[r(\bar{x}, \bar{y}_1, \bar{y}_2, \cdots, \bar{y}_{l_2} + \lambda_{l_1+l_2}\bar{d}_{l_2}) - \bar{r}] \\ & \leq -\frac{1}{2(l_1+l_2)}(\sum_{i=1}^{l_1}\|A_i d_i\|^2 + \sum_{j=1}^{l_2}\|B_j \bar{d}_j\|^2) \end{aligned} \tag{4.5}$$

which implies that the value of $r(x,y)$ is decreasing. The above analysis leads to modify the algorithm A. Clearly, if the term of approximate second order is added to $\xi_i(x_i, \bar{x}, \bar{y})$ and $\zeta_j(y_j, \bar{x}, \bar{y})$ and solve (P_1) and (P_2) to generate search direction in step (i), and (4.4) is used to replace α in step (ii) of algorithm A. Then, a modification algorithm $\bar{A}$ is obtained immediately, and a similar convergence result can be described as follows.

Theorem 4.1. *Suppose that $x^0 \geq 0, y$ is an initial point and sequence $\hat{x}^k, \hat{y}^k$ and x^k, y^k is generated by the algorithm $\bar{A}$, then there exists an integer $k_0 \geq 0$ such that $\hat{x}^{k_0} = x^k, \hat{y}^{k_0} = y^k$ (x^k, y^k is a solution of the problem (1.1)) or*

$$\lim_{k\to\infty} A_i(\hat{x}_i^k - x_i^k) = 0 \quad i = 1, \cdots, l_1, \quad \lim_{k\to\infty} B_j(\hat{y}_j^k - y_j^k) = 0 \quad j = 1, \cdots, l_2, \tag{4.6}$$

and the accumulation point of $\{ x^k, y^k \}$ is a solution of the problem (1.1).

Now we consider the following linear programming problem

$$Min\ z = c^T x \quad s.t. \quad Ax = \bar{b}, \quad x \geq 0 \tag{4.7}$$

where $A \in R^{m\times n}$, $c \in R^n$, $\bar{b} \in R^m$ are given real constant matrix and vectors, respectively.

As is well known, if there are $y \in R^m$ and $s \in R^n$, and s is negative vector such that

$$s + A^T y - c = 0, \quad s \geq 0 \tag{4.8a}$$

$$Ax - \bar{b} = 0, \quad x \geq 0 \tag{4.8b}$$

$$c^T x - \bar{b}^T y = 0, \tag{4.8c}$$

hold, then x is a solution of the (4.7) problem. On the other hand, it follows from the definition of the problem (1.1) that the formula (4.8) can be written its equivalent form as follows

$$Min\ r(x,s,y) = \frac{1}{2}\|A_1 x + A_2 s + By - b\|^2 \quad s.t. \quad x \geq 0, \quad s \geq 0. \tag{4.9}$$

where $A_1^T = [0_1, A^T, c] \in R^{n\times(n+m+1)}$, $A_2^T = [I, 0_2, 0_3] \in R^{n\times(n+m+1)}$, $B^T = [A, 0_4, -\bar{b}] \in R^{n\times(n+m+1)}$, $0_1 \in R^{n\times n}$, $0_2 \in R^{n\times m}, 0_3 \in R^{n\times 1}, 0_4 \in R^{m\times m}$, and $I \in R^{n\times n}$ is an itentity matrix, and $b^T = (c^T, \bar{b}^T, 0)$. If we set $A = [\ 1\ 1\ 1\ 1\ 0\ 0\ ;\ -1\ 2\ -2\ 0\ 1\ 0\ ;\ 2\ 1\ 0\ 0\ 0\ 1\ ] \in R^{3\times 6}, b = (\ 4\ 6\ 5\ )^T, c = (\ -1\ -2\ -1\ 0\ 0\ 0\ )^T$, then $x = (\ 0\ \ 3.5\ \ 0.5\ \ 0\ \ 0\ \ 1.5\ )^T, y = (\ 1.5\ \ 0.25\ \ 0\ )^T, s = (\ 0.25\ \ 0\ \ 1.5\ \ 0\ \ 0.25\ \ 0\ )^T$ is a solution of the problem.

The algorithm A was written by the Mathlab and run on the work station of the State Key Laboratory of Scientific and Engineering Computing. The initial point is chosen as $x = s = e, y = 0(where\ \ e = (\ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ )^T)$, and the results were shown in the Table 1 and Table 2 (The second-order term was used to update in the step (i)).

Table 1 ($k = 200$).

α	$r(x,y,s)$	$\Vert x^k - x^*\Vert_\infty$	$\Vert s^k - s^*\Vert_\infty$	$\Vert y^k - y^*\Vert_\infty$	$\vert c^T(x^k - x^*)\vert$
0.20	0.0029	0.0505	0.0058	0.0024	0.0072
0.25	0.0015	0.0276	0.0028	0.0012	0.0039
0.30	7.1800×10^{-4}	0.0146	0.0013	5.482×10^{-4}	0.0021
0.35	3.6895×10^{-4}	0.0080	6.564×10^{-4}	2.664×10^{-4}	0.0011
0.40	2.0477×10^{-4}	0.0047	3.464×10^{-4}	1.390×10^{-4}	6.708×10^{-4}
0.45	1.2352×10^{-4}	0.0030	1.982×10^{-4}	7.872×10^{-5}	4.220×10^{-4}
0.50	8.0040×10^{-5}	0.0020	1.229×10^{-4}	4.038×10^{-5}	2.844×10^{-4}

Table 2 ($\alpha = 0.45$).

k	$r(x,y,s)$	$\Vert x^k - x^*\Vert_\infty$	$\Vert s^k - s^*\ \vert_\infty$	$\Vert y^k - y^*\Vert_\infty$	$\vert c^T(x^k - x^*)\vert$
100	0.0110	0.2640	0.0174	0.0069	0.0377
125	0.0036	0.0855	0.0057	0.0023	0.0122
150	0.0012	0.0278	0.0019	7.404×10^{-4}	0.0040
175	3.789×10^{-4}	0.0091	6.080×10^{-4}	2.414×10^{-4}	0.0013
200	1.235×10^{-4}	0.0030	1.982×10^{-4}	7.872×10^{-5}	$4.220\ \times10^{-4}$
225	4.026×10^{-5}	9.628×10^{-4}	6.461×10^{-5}	2.566×10^{-5}	$1.376\ \times10^{-4}$
250	1.312×10^{-5}	3.139×10^{-4}	2.106×10^{-5}	8.365×10^{-6}	$4.484\ \times10^{-5}$

References

[1] G.H. Golub, C.V. Loan, Matrix Computations, Johns Hopkins University Press, 1983.

[2] C.L. Lawson, R.J. Hanson, Solving Least Squares Problems, Prentice-Hall, EnglewoodCliffs. NJ., 1975.

[3] G. Stephaopoulos, W. Westerberg, The use of Hestenes' method of multipliers to resolve dual gap in engineering system optimization, *J. Optim. Theory Appl.,* **15** (1975), 285-309.

[4] R. Fletcher, Practical Methods of Optimization, 2-nd ed., John Wiley, Chichester, New York, 1987.

[5] O.L. Mangasarian, Parallel gradient distribution in unconstrained optimization, *SIAM J. Control and Optimization,* **4** (1995), 1916-1925.

[6] Z.L. Wei, Subspace search method for quadratic programming with box constraints, *JCM,* **21** (1999), 307-314.

[7] Y. Yuan, Numerical Methods for Nonlinear Programming, Shanghai Scientific and Technical Publishers, 1993 (in Chinese).