

*Research Article*

# Generalized Quadratic Augmented Lagrangian Methods with Nonmonotone Penalty Parameters

**Xunzhi Zhu, Jinchuan Zhou, Lili Pan, and Wenling Zhao**

*Department of Mathematics, School of Science, Shandong University of Technology, Zibo 255049, China*

Correspondence should be addressed to Jinchuan Zhou, jinchuanzhou@163.com

Received 15 January 2012; Accepted 21 March 2012

Academic Editor: Ying U. Hu

Copyright © 2012 Xunzhi Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

For nonconvex optimization problem with both equality and inequality constraints, we introduce a new augmented Lagrangian function and propose the corresponding multiplier algorithm. New iterative strategy on penalty parameter is presented. Different global convergence properties are established depending on whether the penalty parameter is bounded. Even if the iterative sequence  $\{x^k\}$  is divergent, we present a necessary and sufficient condition for the convergence of  $\{f(x^k)\}$  to the optimal value. Finally, preliminary numerical experience is reported.

## 1. Introduction

This paper is concerned with the following nonlinear programming problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, l, \\ & x \in \Omega, \end{aligned} \tag{P}$$

where  $f, g_i : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $i = 1, \dots, m$  and  $h_j : \mathbb{R}^n \rightarrow \mathbb{R}$  for  $j = 1, \dots, l$  are all continuously differentiable functions, and  $\Omega$  is a nonempty and closed subset in  $\mathbb{R}^n$ . Denoted by  $X$  the feasible region and by  $X^*$  the solution set.

The classical Lagrangian function for this problem is

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^l \mu_j h_j(x), \quad (1.1)$$

where  $(\lambda, \mu) \in \mathbb{R}_+^m \times \mathbb{R}^l$ . It is well known that the above classic Lagrangian function would fail to find out the optimal solution of primal problem because a nonzero duality gap may be arisen for nonconvex programming. Augmented Lagrangians were proposed to overcome this difficulty. It combines the advantage from Lagrangian duality methods and penalty function methods; for example, it does not need to make the penalty parameter infinitely large, and hence successfully avoid ill-conditioning effects. In fact, if the penalty parameter is very large, the Lagrangian relaxation subproblem would be very difficult in the sense that the stabilized Newton step is not regarded as a candidate to be a decreasing direction.

Global convergence properties of augmented Lagrangian methods have been studied by many researchers [1–8]. It should be noted that the multiplier sequence generated by the algorithm may be unbounded. Several strategies have proposed to deal with these questions, such as adding constraint qualifications [5, 6] or via safeguarding projection [1, 2, 9, 10].

In this paper, for the optimization problem  $(P)$  with both equality and inequality constraints, we introduce a new class of augmented Lagrangian functions, which include the well-known quadratic augmented Lagrangian function as special case [11–13]. It should be noted that this function class is more general since we do not restrict  $\phi$  to be convex (see (2.1) below for the definition of  $\phi$ ). Convergence properties of the corresponding multiplier algorithm are studied. Specially, a new updated strategy on penalty parameter is proposed, that is, the penalty parameter  $c_k$  is improved only when the multipliers are too large, and the iterative point is far away from the feasible region. This strategy guarantees that the penalty parameter remains unchanged, provided that measure of infeasibility of iterative point and bounding of multipliers have enough progress (see (2.4)). Furthermore, we study the performance of the multiplier sequences generated by our algorithm. Finally, compared with [4, 9, 10], we further consider the case when  $\{x^k\}$  is divergent, in which a necessary and sufficient condition for  $\{f(x^k)\}$  converging to the optimal value is given as well.

The organization of this paper is as follows. In the next section, we propose the multiplier algorithm and study the global convergence properties. Preliminary numerical results are reported in Section 3. Conclusion is drawn in Section 4.

## 2. Multiplier Algorithms

A new generalized quadratic augmented Lagrangian function for  $(P)$  is defined as follows:

$$L(x, \lambda, \mu, c) := f(x) + \sum_{j=1}^l \mu_j h_j(x) + \frac{c}{2} \sum_{j=1}^l h_j^2(x) + \frac{1}{2c} \sum_{i=1}^m \left\{ \max^2\{0, \phi(cg_i(x)) + \lambda_i\} - \lambda_i^2 \right\}, \quad (2.1)$$

where  $(x, \lambda, \mu, c) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}_{++}$ , and  $\mathbb{R}_{++}$  denotes the all positive real scalars, that is,  $\mathbb{R}_{++} = \{a \in \mathbb{R} \mid a > 0\}$ . The function  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  involved in (2.1) satisfies the following properties:

- (A<sub>1</sub>) continuously differentiable and strictly increasing on  $\mathbb{R}$  with  $\phi(0) = 0$  and  $\phi(\alpha) \geq \alpha$  for  $\alpha \geq 0$ .

Particularly, if  $\phi(\alpha) = \alpha$  for  $\alpha \in \mathbb{R}$ , then  $L$  reduces to the quadratic augmented Lagrangian function. Compared with [9, 14, 15], an important point made above is that  $\phi$  is not required to be convex. Hence, it makes the augmented Lagrangian we introduce here more general.

Given  $(x, \lambda, \mu, c)$ , the Lagrangian relaxation subproblem associated with the augmented Lagrangian  $L$  is defined as:

$$\begin{aligned} \min \quad & L(x, \lambda, \mu, c), \\ \text{s.t.} \quad & x \in \Omega. \end{aligned} \tag{L_{\lambda, \mu, c}}$$

Its solution set is denoted by  $S^*(\lambda, \mu, c)$ . We always assume throughout this paper that the set  $S^*(\lambda, \mu, c)$  is nonempty, which is ensured if  $\Omega$  is compact, since  $f$ ,  $g_i$ , and  $h_j$  are all continuous. In addition, we assume that:

$$(A_2) \quad f \text{ is bounded from below, that is, } f_* = \inf_{x \in \mathbb{R}^n} f(x) > -\infty.$$

It is a rather mild assumption in optimization problem because otherwise the objective function  $f$  can be replaced by  $e^{f(x)}$ . Recall that a vector  $x^*$  is said to be a stationary point of (P) if it is a feasible point and there exist  $\lambda_i \in \mathbb{R}_+$  for all  $i = 1, \dots, m$  and  $\mu_j$  for  $j = 1, \dots, l$  such that the following system holds:

$$0 \in \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^l \mu_j \nabla h_j(x^*) + \mathcal{N}_\Omega(x^*), \tag{2.2}$$

where  $\mathcal{N}_\Omega(x^*)$  denotes the normal cone of  $\Omega$  at  $x^*$  [16, Chapter 6]. Let  $\Lambda(x^*)$  denote the collection of multipliers satisfying (2.2). The set of  $\Lambda(x^*)$  is larger than that of  $R$ -multiplier defined in [17], which reduces to the well-known Karush-Kuhn-Tucker (KKT) point of (P) when  $\Omega = \mathbb{R}^n$ , since the complementarity condition is required for KKT point. The following is the multiplier algorithm based on the generalized quadratic augmented Lagrangian  $L$ . One of its main feature is that the Lagrangian multipliers associated with equality and inequality constraints are not restricted to be bounded.

*Algorithm 2.1* (multiplier algorithm based on  $L$ ).

*Step 1.* (initialization)

Let  $M > 0$  and  $\{\varepsilon_k\}_{k=1}^\infty$  be an arbitrary positive sequence with zero as limit (i.e.,  $\varepsilon_k \rightarrow 0^+$ ). Select an initial point  $x^0 \in \mathbb{R}^n$ ,  $\lambda^0 \in \mathbb{R}_+^m$ ,  $\mu_0 \in \mathbb{R}^l$ , and  $c_0 > 0$ . Set  $k := 0$ .

*Step 2.* (estimating multipliers)

Compute

$$\begin{aligned} \lambda_i^{k+1} &= \max\left\{0, \phi\left(c_k g_i(x^k)\right) + \lambda_i^k\right\} \phi'\left(c_k g_i(x^k)\right), \quad \forall i = 1, \dots, m, \\ \mu_j^{k+1} &= \mu_j^k + c_k h_j(x^k), \quad \forall j = 1, \dots, l. \end{aligned} \tag{2.3}$$

Step 3. (updating penalty parameter)

Let

$$c_{k+1} = c_k, \quad \text{if } \left| \lambda_i^{k+1} \right| \leq M, \left| \mu_j^{k+1} \right| \leq M, \left| h_j(x^k) \right| \leq \varepsilon_k, g_i(x^k) \leq \varepsilon_k, \forall i, j \quad (2.4)$$

$$c_{k+1} = (k+1) \max \left\{ 1, \sum_{i=1}^m (\lambda_i^{k+1})^2, \sum_{j=1}^l (\mu_j^{k+1})^2 \right\}, \quad \text{otherwise.} \quad (2.5)$$

Step 4. (solving Lagrangian relaxation subproblem)

Find  $x^{k+1} \in S^*(\lambda^{k+1}, \mu^{k+1}, c_{k+1})$ .

Step 5. (cycling)

If  $x^{k+1} \in \Omega$  and  $(\lambda^{k+1}, \mu^{k+1}) \in \Lambda(x^{k+1})$ , then stop; otherwise, let  $k := k+1$  and go back to Step 2.

Consider the perturbation of (P); precisely, for any fixed  $\alpha \geq 0$ , define the perturbation value function as follows:

$$v(\alpha) = \inf \{ f(x) \mid x \in \Omega(\alpha) \}, \quad (2.6)$$

where

$$\Omega(\alpha) = \{ x \in \Omega \mid g_i(x) \leq \alpha, |h_j(x)| \leq \alpha, i = 1, \dots, m, j = 1, \dots, l \}. \quad (2.7)$$

It is clear that  $\Omega(0) = X$ . In practical implementation of the previous algorithm, we always choose the constant  $M$  large enough to strictly include  $\partial v(0)$ , the set of subdifferential of  $v$  at origin, since according to the strong duality theory, the optimal Lagrangian multipliers belong to  $\partial v(0)$ ; see [18] for the detailed discussion. In [5, 6], the authors require the convergence of  $\{\lambda^k/c_k\}$  to zero. It follows from (2.5) that this nice property holds automatically in our new strategy. In addition, (2.4) and (2.5) indicate that Lagrangian relaxation subproblem needs to place more emphasis on decreasing the constraint violations via increasing the penalty parameter, provided that the current point  $x^k$  is far away from feasible region.

**Lemma 2.2.** *Let  $\{x^k\}$  be the iterative sequence generated by Algorithm 2.1, and let  $x^*$  be one of its accumulate points. If  $\{c_k\}$  is bounded, then  $x^*$  is a stationary point of (P).*

*Proof.* If the algorithm is terminated finitely, then the result follows immediately according to the stop criterion in Step 5. On the other hand, if the condition (2.5) occurs infinitely, then  $c_k \geq k$ , and hence  $c_k$  must be unbounded. Therefore, by hypothesis we knew that only condition (2.4) occurs when  $k$  is large enough. Because  $g_i(x^k) \leq \varepsilon_k$  and  $|h_j(x^k)| \leq \varepsilon_k$  by (2.4), taking the limit yields  $g_i(x^*) \leq 0$  and  $h_j(x^*) = 0$ , since  $\varepsilon_k$  converges to zero. This shows the feasibility of  $x^*$ . Meanwhile, we know from (2.4) that  $\{\lambda_i^k\}$  and  $\{\mu_j^k\}$  are bounded. Hence, we can assume without loss of generality that  $\lambda_i^k \rightarrow \lambda_i^*$  and  $\mu_j^k \rightarrow \mu_j^*$ . Since  $x^k$  is an optimal solution of

minimizing  $L(x, \lambda^k, \mu^k, c_k)$  over  $\Omega$  by Step 4, then according to the well-known optimality conditions we have

$$-\nabla_x L(x^k, \lambda^k, \mu^k, c_k) \in \mathcal{N}_\Omega(x^k), \quad (2.8)$$

which together with the formula of (2.3) and (2.4) yields

$$-\nabla f(x^k) - \sum_{j=1}^l \mu_j^{k+1} \nabla h_j(x^k) - \sum_{i=1}^m \lambda_i^{k+1} \nabla g_i(x^k) \in \mathcal{N}_\Omega(x^k). \quad (2.9)$$

Taking limits gives us

$$-\nabla f(x^*) - \sum_{j=1}^l \mu_j^* \nabla h_j(x^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) \in \mathcal{N}_\Omega(x^*), \quad (2.10)$$

where we have used the upper semicontinuity of normal cone [16, Chapter 6]. This completes the proof.  $\square$

Now, we turn our main concern to the case when  $\{c_k\}$  is unbounded.

**Theorem 2.3.** *Let  $\{x^k\}$  be the iterative sequence generated by Algorithm 2.1, and let  $x^*$  be one of its accumulate points. If  $\{c_k\}$  is unbounded, then  $x^*$  is an optimal solution of (P).*

*Proof.* It is clear that if  $\{c_k\}$  is unbounded, then the iteration formula (2.5) occurs infinitely. Hence, for simplification, we assume that (2.5) always happens as  $k$  sufficiently large, and that  $x^k$  converges to  $x^*$  (not resorting to a subsequence). Taking into account of (2.5), the terms  $\lambda^k/c_k$ ,  $\mu^k/c_k$ ,  $(\lambda^k)^2/c_k$ ,  $(\mu^k)^2/c_k$  are all converges to zero. This fact will be used in the following analysis.

Let us first show that  $x^*$  is a feasible point of (P). We argue it by contradiction. Note first that

$$\begin{aligned} L(x^k, \lambda^k, \mu^k, c_k) &= \inf \left\{ L(x, \lambda^k, \mu^k, c_k) \mid x \in \Omega \right\} \\ &\leq \inf \left\{ L(x, \lambda^k, \mu^k, c_k) \mid x \in \Omega(0) \right\} \\ &= \inf \left\{ f(x) + \frac{1}{2c_k} \sum_{i=1}^m \left\{ \max^2 \left\{ 0, \phi(c_k g_i(x)) + \lambda_i^k \right\} - (\lambda_i^k)^2 \right\} \mid x \in \Omega(0) \right\} \\ &\leq \inf \{ f(x) \mid x \in \Omega(0) \} \\ &= v(0), \end{aligned} \quad (2.11)$$

where the third and fourth steps are due to the feasibility of  $x \in X$ . This implies that

$$\limsup_{k \rightarrow \infty} L(x^k, \lambda^k, \mu^k, c_k) \leq v(0) < +\infty. \quad (2.12)$$

Now, let us consider the following two cases.

*Case 1.* There exists an index  $j_0$  such that  $|h_{j_0}(x^*)| > 0$ . Due to the continuity of  $h_{j_0}$ , we must have  $|h_{j_0}(x^k)| > (1/2)|h_{j_0}(x^*)| > 0$  for  $k$  is large enough. Therefore,

$$\begin{aligned} L(x^k, \lambda^k, \mu^k, c_k) &\geq f_* + \sum_{j=1}^l \mu_j^k h_j(x^k) + \frac{c_k}{2} \sum_{j=1}^l h_j^2(x^k) - \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \\ &= f_* + \frac{c_k}{2} \sum_{j=1}^l \left\{ \left( h_j(x^k) + \frac{\mu_j^k}{c_k} \right)^2 - \frac{(\mu_j^k)^2}{c_k^2} \right\} - \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \\ &\geq f_* + \frac{c_k}{2} \left( h_{j_0}(x^k) + \frac{\mu_{j_0}^k}{c_k} \right)^2 - \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 - \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2. \end{aligned} \quad (2.13)$$

Noticing that  $\mu_{j_0}^k/c_k \rightarrow 0$  as mentioned above and  $|h_{j_0}(x^k)| \geq (1/2)|h_{j_0}(x^*)|$ , we know that as  $k$  is large enough

$$\left| h_{j_0}(x^k) + \frac{\mu_{j_0}^k}{c_k} \right| \geq |h_{j_0}(x^k)| - \frac{|\mu_{j_0}^k|}{c_k} \geq \frac{1}{4} |h_{j_0}(x^*)|, \quad (2.14)$$

which further implying

$$\left( h_{j_0}(x^k) + \frac{\mu_{j_0}^k}{c_k} \right)^2 > \frac{1}{16} |h_{j_0}(x^*)|^2. \quad (2.15)$$

This, together with (2.13), yields  $v(0) = +\infty$  since  $c_k$  approaches to  $\infty$ , which contradicts (2.12).

Case 2. There exists an index  $i_0$  such that  $g_{i_0}(x^*) > 0$ . Similarly, due to the continuity of  $g_{i_0}$ , we know that  $g_{i_0}(x^k) \geq (1/2)g_{i_0}(x^*) > 0$  as  $k$  is sufficiently large. Therefore,

$$\begin{aligned} L(x^k, \lambda^k, \mu^k, c_k) &\geq f_* + \frac{1}{2c_k} \sum_{i=1}^m \left\{ \max^2 \left\{ 0, \phi(c_k g_i(x^k)) + \lambda_i^k \right\} - (\lambda_i^k)^2 \right\} - \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 \\ &\geq f_* + \frac{1}{2c_k} \max^2 \left\{ 0, c_k g_{i_0}(x^k) + \lambda_{i_0}^k \right\} - \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 - \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 \quad (2.16) \\ &\geq f_* + \frac{c_k}{8} g_{i_0}^2(x^*) - \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 - \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2, \end{aligned}$$

where the second inequality comes from the fact  $\phi(a) \geq a$  for all  $a \geq 0$  by  $(A_1)$ , and the last inequality follows from the nonnegativity of  $\lambda_i^k$  by (2.3), since  $\phi'(a) \geq 0$  for all  $a \in \mathbb{R}$  due to the monotonicity of  $\phi$ . Invoking (2.5) and taking limits in the above inequality yields  $v(0) = +\infty$ , which contradicts (2.12). So far, we have established the feasibility of  $x^*$ .

It remains to show that  $x^*$  is an optimal solution. In fact, notice that

$$\begin{aligned} f(x^k) &= L(x^k, \lambda^k, \mu^k, c_k) - \frac{1}{2c_k} \sum_{i=1}^m \left\{ \max^2 \left\{ 0, \phi(c_k g_i(x^k)) + \lambda_i^k \right\} - (\lambda_i^k)^2 \right\} \\ &\quad - \sum_{j=1}^l \mu_j^k h_j(x^k) - \frac{c_k}{2} \sum_{j=1}^l h_j^2(x^k) \\ &\leq L(x^k, \lambda^k, \mu^k, c_k) + \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \quad (2.17) \\ &\leq v(0) + \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2, \end{aligned}$$

where the second inequality is due to (2.11). Taking limits in the above yields

$$f(x^*) = \lim_{k \rightarrow \infty} f(x^k) \leq \limsup_{k \rightarrow \infty} \left\{ v(0) + \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \right\} = v(0), \quad (2.18)$$

where the last step comes from the fact that  $(\mu_j^k)^2/c_k$  and  $(\lambda_i^k)^2/c_k$  converge to zero (see (2.5)). This means that  $x^*$  must be an optimal solution of  $(P)$ . The proof is complete.  $\square$

Finally, let us deal with the performance of multiplier sequences  $\{\lambda^k\}$  and  $\{\mu^k\}$  generated by our proposed algorithm.

**Theorem 2.4.** Let  $\{x^k\}$  be an iterative sequence generated by Algorithm 2.1, and let  $x^*$  be one of its accumulate points. The following statements hold:

(i) if both  $\{\sum_{i=1}^m \lambda_i^k\}$  and  $\{\sum_{j=1}^l \mu_j^k\}$  are bounded, then we have

$$-\nabla f(x^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^l \mu_j^* \nabla h_j(x^*) \in \mathcal{N}_\Omega(x^*), \quad (2.19)$$

where  $\lambda_i^*$  and  $\mu_j^*$  are accumulate points of  $\{\lambda_i^k\}$  and  $\{\mu_j^k\}$ ;

(ii) if either  $\{\sum_{i=1}^m \lambda_i^k\}$  or  $\{\sum_{j=1}^l \mu_j^k\}$  is unbounded, then we have

$$-\sum_{j=1}^l \mu_j^* \nabla h_j(x^*) - \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) \in \mathcal{N}_\Omega(x^*), \quad (2.20)$$

where  $\lambda_i^*$  and  $\mu_j^*$  are accumulate points of  $\{\lambda_i^k/T^k\}$  and  $\{\mu_j^k/T^k\}$  with

$$T^k = \sum_{i=1}^m \lambda_i^k + \sum_{j=1}^l \mu_j^k. \quad (2.21)$$

*Proof.* The proof is divided into the following two cases.

*Case 1.* Both  $\{\sum_{i=1}^m \lambda_i^k\}$  and  $\{\sum_{j=1}^l \mu_j^k\}$  are bounded. In this case, we can obtain that  $x^*$  is a stationary point of (P) by following almost the same argument as in Lemma 2.2.

*Case 2.* Either  $\{\sum_{i=1}^m \lambda_i^k\}$  or  $\{\sum_{j=1}^l \mu_j^k\}$  is divergent. According to the updated strategy on penalty parameter  $c_k$ , we know that (2.5) must occur as  $k$  sufficiently large. Hence,  $\{c_k\}$  is unbounded in this case. Since  $\{x^k\}$  is a global optimal solution of  $L(x, \lambda^k, \mu^k, c_k)$  over  $\Omega$  by Step 4 in Algorithm 2.1, applying the optimality condition to the augmented Lagrangian relaxation problem  $L_{(\lambda^k, \mu^k, c_k)}$  yields

$$-\nabla_x L(x^k, \lambda^k, \mu^k, c_k) \in \mathcal{N}_\Omega(x^k), \quad (2.22)$$

that is,

$$-\nabla f(x^k) - \sum_{i=1}^m \lambda_i^{k+1} \nabla g_i(x^k) - \sum_{j=1}^l \mu_j^{k+1} \nabla h_j(x^k) \in \mathcal{N}_\Omega(x^k). \quad (2.23)$$

Since  $\lambda_i^{k+1}/T^{k+1} \in [0, 1]$  and  $\mu_j^{k+1}/T^{k+1} \in [0, 1]$  are bounded, we can assume without loss of generality that

$$\frac{\lambda_i^{k+1}}{T^{k+1}} \rightarrow \lambda_i^*, \quad \frac{\mu_j^{k+1}}{T^{k+1}} \rightarrow \mu_j^*, \quad i = 1, \dots, m, \quad j = 1, \dots, l. \quad (2.24)$$



Clearly,  $\lambda_i^*$  and  $\mu_j^*$  are not all zeros. Dividing by  $T^{k+1}$  in both sides of (2.23) and using the fact that  $\mathcal{N}_\Omega$  is cone, we have

$$-\frac{1}{T^{k+1}} \nabla f(x^k) - \sum_{i=1}^m \frac{\lambda_i^{k+1}}{T^{k+1}} \nabla g_i(x^k) - \sum_{j=1}^l \frac{\mu_j^{k+1}}{T^{k+1}} \nabla h_j(x^k) \in \mathcal{N}_\Omega(x^k). \quad (2.25)$$

Taking limits in the above yields

$$-\sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) - \sum_{j=1}^l \mu_j^* \nabla h_j(x^*) \in \mathcal{N}_\Omega(x^*). \quad (2.26)$$

This establishes the result as desired.  $\square$

Now, a natural question arises: how does the algorithm perform if  $\{x^k\}$  is divergent? The following theorem gives the answer.

**Theorem 2.5.** *Let  $\{x^k\}$  be an iterative sequence generated by Algorithm 2.1. If  $\{c_k\}$  is unbounded, then the following statements are equivalent:*

(a)  $f(x^k)$  converges to the optimal value, that is,

$$\lim_{k \rightarrow \infty} f(x^k) = v(0); \quad (2.27)$$

(b)  $v(\alpha)$  is lower semicontinuous at  $\alpha = 0$  from right, that is,  $\liminf_{\alpha \rightarrow 0^+} v(\alpha) \geq v(0)$ ;

(c)  $v(\alpha)$  is continuous at  $\alpha = 0$  from right, that is,  $\lim_{\alpha \rightarrow 0^+} v(\alpha) = v(0)$ .

*Proof.* (a)  $\Rightarrow$  (b). Suppose on the contrary that  $v$  is not lower semicontinuous at zero from right, then there must exist  $\delta_0 > 0$  and  $\xi_t \rightarrow 0^+$  (as  $t \rightarrow \infty$ ) such that

$$v(\xi_t) \leq v(0) - \delta_0, \quad \forall t. \quad (2.28)$$

For any given  $k$ , since  $\xi_t \rightarrow 0$  we can choose a subsequence  $t_k$  satisfying

$$\xi_{t_k} c_k \rightarrow 0 \quad \text{as } k \rightarrow \infty, \quad (2.29)$$

which together with the continuity of  $\phi$  implies that

$$\phi(c_k \xi_{t_k}) \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (2.30)$$

In addition, pick  $z^k \in \Omega(\xi_{t_k})$  with  $f(z^k) \leq v(\xi_{t_k}) + (\delta_0/2)$ . Then  $f(z^k) \leq v(0) - (\delta_0/2)$  by (2.28).

Therefore,

$$\begin{aligned}
f(x^k) &= L(x^k, \lambda^k, \mu^k, c_k) - \frac{1}{2c_k} \sum_{i=1}^m \left\{ \max^2 \{0, \phi(c_k g_i(x^k)) + \lambda_i^k\} - (\lambda_i^k)^2 \right\} \\
&\quad - \sum_{j=1}^l \mu_j^k h_j(x^k) - \frac{c_k}{2} \sum_{j=1}^l h_j^2(x^k) \\
&\leq \inf_{x \in \Omega} L(x, \lambda^k, \mu^k, c_k) - \frac{c_k}{2} \sum_{j=1}^l \left\{ \left( h_j(x^k) + \frac{\mu_j^k}{c_k} \right)^2 - \frac{(\mu_j^k)^2}{c_k^2} \right\} + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \\
&\leq \inf_{x \in \Omega} L(x, \lambda^k, \mu^k, c_k) + \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \\
&\leq f(z^k) + \sum_{j=1}^l \mu_j^k h_j(z^k) + \frac{c_k}{2} \sum_{j=1}^l h_j^2(z^k) \\
&\quad + \frac{1}{2c_k} \sum_{i=1}^m \left\{ \max^2 \{0, \phi(c_k g_i(z^k)) + \lambda_i^k\} - (\lambda_i^k)^2 \right\} + \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \\
&= f(z^k) + \frac{c_k}{2} \sum_{j=1}^l \left\{ \left( h_j(z^k) + \frac{\mu_j^k}{c_k} \right)^2 - \frac{(\mu_j^k)^2}{c_k^2} \right\} \\
&\quad + \frac{1}{2c_k} \sum_{i=1}^m \left\{ \max^2 \{0, \phi(c_k g_i(z^k)) + \lambda_i^k\} - (\lambda_i^k)^2 \right\} + \frac{1}{2c_k} \sum_{j=1}^l (\mu_j^k)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\lambda_i^k)^2 \\
&\leq v(0) - \frac{\delta_0}{2} + \frac{c_k}{2} \sum_{j=1}^l \left( \xi_{t_k} + \frac{|\mu_j^k|}{c_k} \right)^2 + \frac{1}{2c_k} \sum_{i=1}^m (\phi(c_k \xi_{t_k}) + \lambda_i^k)^2,
\end{aligned} \tag{2.31}$$

where the last step is due to the fact  $|h_j(z^k)| \leq \xi_{t_k}$  and  $g_i(z^k) \leq \xi_{t_k}$  since  $z^k \in \Omega(\xi_{t_k})$  and  $\phi$  is nondecreasing by (A<sub>1</sub>). Taking the limits in both sides of (2.31) and using (2.27) and (2.30) yield

$$v(0) = \lim_{k \rightarrow \infty} f(x^k) \leq v(0) - \frac{\delta_0}{2}, \tag{2.32}$$

which leads to a contradiction.

(b)  $\Rightarrow$  (a). Since  $x^k$  converges to  $x^*$ , it then follows from Theorem 2.3 that  $x^*$  is an optimal solution of (P). Hence, for any arbitrary  $\varepsilon$ , we have  $f(x^k) \leq f(x^*) + \varepsilon = v(0) + \varepsilon$  and

$x^k \in \Omega(\varepsilon)$  (since  $g_i(x^k) \rightarrow g_i(x^*) \leq 0 < \varepsilon$  and  $|h_j(x^k)| \rightarrow |h_j(x^*)| = 0 < \varepsilon$  as  $k$  is sufficiently large). The latter further implies that  $v(\varepsilon) \leq f(x^k)$  by definition (2.6), that is,

$$v(\varepsilon) \leq f(x^k) \leq v(0) + \varepsilon. \quad (2.33)$$

Since  $v(\varepsilon)$  is lower semicontinuous at  $\varepsilon = 0$  from right by hypothesis, taking the lower limitation in (2.33) yields

$$v(0) \leq \liminf_{\varepsilon \rightarrow 0^+} v(\varepsilon) \leq \liminf_{k \rightarrow \infty} f(x^k) \leq \limsup_{k \rightarrow \infty} f(x^k) \leq v(0), \quad (2.34)$$

that is,

$$\lim_{k \rightarrow \infty} f(x^k) = v(0). \quad (2.35)$$

(c)  $\Rightarrow$  (b). It is clear.

(b)  $\Rightarrow$  (c). It follows by invoking the fact that  $v(\alpha) \leq v(0)$ , since  $\Omega(0) \subset \Omega(\alpha)$  for all  $\alpha \geq 0$ . Then  $\limsup_{\alpha \rightarrow 0^+} v(\alpha) \leq v(0)$ , that is,  $v$  is upper semicontinuous at origin from right. This together with the lower semicontinuous of  $v$  at origin by hypothesis yields the desired result.  $\square$

### 3. Preliminary Numerical Reports

To give some insight into the behavior of our proposed algorithm presented in this paper, we solve two problems by letting  $\phi$  take the following different functions:

- (1)  $y_1 = \phi_1(\alpha) = \alpha$ ;
- (2)  $y_2 = \phi_2(\alpha) = e^\alpha - 1$ ;
- (3)  $y_3 = \phi_3(\alpha) = \begin{cases} \alpha^2 + \alpha, & \alpha \geq 0, \\ \alpha & \text{otherwise;} \end{cases}$
- (4)  $y_4 = \phi_4(\alpha) = \begin{cases} \alpha^3 + \alpha, & \alpha \geq 0, \\ -\alpha^2 & \text{otherwise.} \end{cases}$

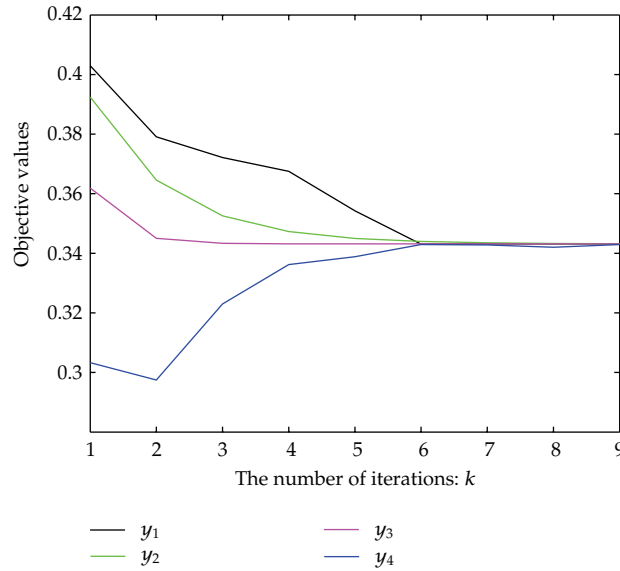
The computer codes were written in Matlab 7.0, and the test was done at a PC of Pentium 4 with 2.8GHz CPU and 1.99GB memory. We implement our algorithm to solve the following programming problems, where Example 3.1 is obtained by adding an equality constraint in Example 6.3 in [19], and Example 3.2 comes from [20]. The corresponding numerical results are reported below, where  $k$  is the number of iterations,  $c_k$  is the penalty parameter,  $x^k$  is iterative point, and  $f(x^k)$  is the objective function value.

*Example 3.1.* Consider the following nonconvex programming:

$$\begin{aligned} \min \quad & (x_1 + x_2 - 2)^2 + (x_1 - x_2)^2 + 30(\min\{0, x_1 - x_2\})^2 \\ \text{s.t.} \quad & x_1 + x_2 - \sqrt{2} \leq 0, \\ & x_1^2 + 5x_1x_2 - 3 = 0. \end{aligned} \quad (3.1)$$

**Table 1:** Result of Example 3.1.

$\phi_i(s)$	$k$	$c_k$	$x^k$	$f(x^k)$
$\phi_1(s)$	1	2.5147	(0.6722, 0.6722)	0.4029
	4	2.5147	(0.6950, 0.6950)	0.3675
	6	2.5147	(0.7071, 0.7071)	0.3431
$\phi_2(s)$	1	10.4139	(0.6612, 0.6612)	0.3924
	3	10.4139	(0.6981, 0.6981)	0.3526
	5	10.4139	(0.7071, 0.7071)	0.3432
$\phi_3(s)$	1	17.5462	(0.6517, 0.6517)	0.3676
	3	17.5462	(0.6992, 0.6992)	0.3450
	5	17.5462	(0.7071, 0.7071)	0.3432
$\phi_4(s)$	1	17.5462	(0.6388, 0.6388)	0.3033
	3	17.5462	(0.7273, 0.7273)	0.3230
	5	17.5462	(0.7071, 0.7071)	0.3431



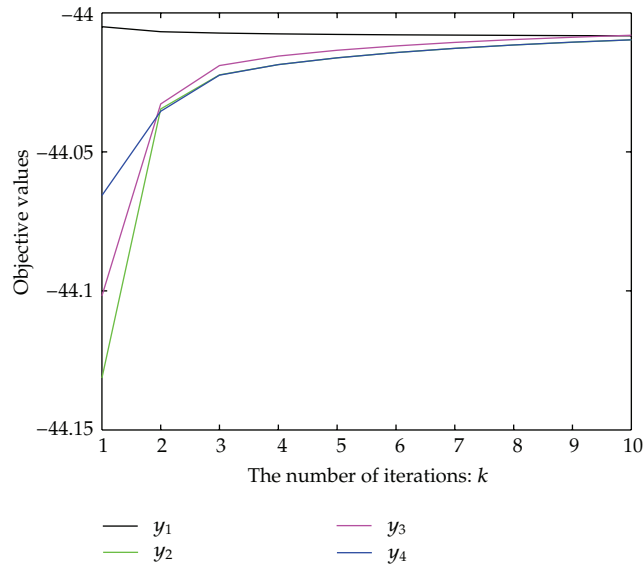
**Figure 1:** Result of Example 3.1.

Example 3.2.

$$\begin{aligned}
 \min \quad & f(x) = -5(x_1 + x_2) + 7(x_4 - 3x_3) + x_1^2 + x_2^2 + 2x_3^2 + x_4^2 \\
 \text{s.t.} \quad & \sum_{i=1}^4 x_i^2 + x_1 - x_2 + x_3 - x_4 - 8 \leq 0, \\
 & x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leq 0, \\
 & 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 - 5 = 0.
 \end{aligned} \tag{3.2}$$

**Table 2:** Result of Example 3.2.

$\phi_i(s)$	$k$	$c_k$	$x^k$	$f(x^k)$
$\phi_1(s)$	1	8.4614	(-0.0072, 0.9963, 1.9907, -1.0058)	-43.8610
	3	24.2390	(-0.0011, 0.9997, 1.9998, -1.0033)	-44.0076
	9	48.5455	(-0.0012, 0.9997, 1.9998, -1.0036)	-44.0083
$\phi_2(s)$	1	8.4430	(0.0017, 1.0017, 2.0063, -1.0073)	-44.1314
	3	24.7705	(-0.0028, 0.9992, 1.9994, -1.0086)	-44.0186
	9	48.7133	(-0.0014, 0.9996, 1.9997, -1.0044)	-44.0097
$\phi_3(s)$	1	3.8876	(0.0025, 1.0017, 2.0054, -1.0027)	-44.1018
	3	9.8546	(-0.0027, 0.9992, 1.9994, -1.0085)	-44.0190
	7	48.2683	(0.0005, 1.0002, 2.0006, -0.9987)	-44.0050
$\phi_4(s)$	1	7.7614	(0.0025, 1.0007, 2.0032, -1.0019)	-44.0657
	3	24.6380	(-0.0023, 0.9993, 1.9995, -1.0072)	-44.0186
	7	48.5943	(-0.0012, 0.9997, 1.9998, -1.0036)	-44.0081

**Figure 2:** Result of Example 3.2.

In the implementation of our algorithm, we use the BFGS quasi-Newton method with a mixed quadratic and cubic line search procedure to solve the Lagrangian subproblem:

- (i) for Example 3.1, the initial data are  $x^0 = (1, 1)^T$ ,  $M = 1000$ ,  $e_k = 1/2^k$ ,  $\lambda^0 = 1$ , and  $\mu^0 = 1$ ,  $c_0 = 1$ ;  $\{c_k\}$  remains unchanged because  $M$  is taken large enough to ensure the validity of (2.4); the obtained solution is a stationary point with the corresponding multipliers  $\lambda^* = 1.172$  and  $\mu^* = 0$ , and the inequality is active;
- (ii) for Example 3.2, the result starts from the nonfeasible point  $x^0 = (2, 0, -2, 1)^T$ ,  $M = 100$ ,  $e_k = 1/2^k$ ,  $\lambda^0 = (1, 1)^T$ ,  $\mu^0 = 1$ , and  $c_0 = 1$ ; the obtained solution is also a stationary point with the corresponding multipliers  $\lambda^* = (1, 0)^T$  and  $\mu^* = 2$ , and the first inequality constraint is active.

The above preliminary numerical result as shown in Tables 1 and 2 and Figures 1 and 2 indicates that Algorithm 2.1 may have better numerical performance, when  $\phi$  is nonlinear (see  $\phi_2$  and  $\phi_3$ ), and  $\phi$  is nonconvex (see  $\phi_4$ ) because the decreasing steep is more “quickly” in the beginning of the iterative, which leads to the requirement of fewer iterative steps. Our next research topic is to further study the performance of our algorithm for solving various practical questions.

## 4. Conclusions

We introduce a new class of generalized quadratic augmented Lagrangian function with nonconvex functions. Global convergence properties of the corresponding multiplier algorithm are proposed without requiring the boundedness of Lagrangian multiplier sequence. Penalty parameter is improved only if multiplier sequences are too large and the iterative point is far away from feasible region. Necessary and sufficient conditions for the convergence of  $\{f(x^k)\}$  to the optimal value are established as well.

## Disclosure

The research of J. Zhou and X. Zhu was partly supported by National Natural Science Foundation of China (11101248), Shandong Province Natural Science Foundation (ZR2010AQ026), and Young Teacher Support Program of Shandong University of Technology. The research of L. Pan was partly supported by National Natural Science Foundation of China (11171247 and 71101140). The research of W. Zhao was partly supported by National Natural Science Foundation of China (10971118).

## Acknowledgments

The authors appreciate very much the constructive and excellent comments from reviewers, which greatly improved the presentation of the paper. In particular, one of reviewers drew our attention to the paper [17], where the detailed relationship between various major constraint qualifications is given.

## References

- [1] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, “On augmented Lagrangian methods with general lower-level constraints,” *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1286–1309, 2007.
- [2] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt, “Augmented Lagrangian methods under the constant positive linear dependence constraint qualification,” *Mathematical Programming*, vol. 111, no. 1-2, pp. 5–32, 2008.
- [3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, NY, USA, 1982.
- [4] E. G. Birgin, C. A. Floudas, and J. M. Martínez, “Global minimization using an augmented Lagrangian method with variable lower-level constraints,” *Mathematical Programming*, vol. 125, no. 1, pp. 139–162, 2010.
- [5] A. R. Conn, N. I. M. Gould, and P. L. Toint, “A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds,” *SIAM Journal on Numerical Analysis*, vol. 28, no. 2, pp. 545–572, 1991.

- [6] A. R. Conn, N. Gould, A. Sartenaer, and Ph. L. Toint, "Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints," *SIAM Journal on Optimization*, vol. 6, no. 3, pp. 674–703, 1996.
- [7] R. M. Lewis and V. Torczon, "A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds," *SIAM Journal on Optimization*, vol. 12, no. 4, pp. 1075–1089, 2002.
- [8] A. Rubinov and X. Yang, *Lagrange-Type Functions in Constrained Non-Convex Optimization*, vol. 85, Kluwer Academic, Dordrecht, The Netherlands, 2003.
- [9] H. Z. Luo, X. L. Sun, and D. Li, "On the convergence of augmented Lagrangian methods for constrained global optimization," *SIAM Journal on Optimization*, vol. 18, no. 4, pp. 1209–1230, 2007.
- [10] H. Z. Luo, X. L. Sun, and Y. F. Xu, "Convergence properties of modified and partially-augmented Lagrangian methods for mathematical programs with complementarity constraints," *Journal of Optimization Theory and Applications*, vol. 145, no. 3, pp. 489–506, 2010.
- [11] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, pp. 303–320, 1969.
- [12] M. J. D. Powell, "A Method for Nonlinear Constraints in Minimization Problems," in *Optimization*, R. Fletcher, Ed., pp. 283–298, Academic Press, New York, NY, USA, 1969.
- [13] R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming," *SIAM Journal on Control and Optimization*, vol. 12, pp. 268–285, 1974.
- [14] X. L. Sun, D. Li, and K. I. M. McKinnon, "On saddle points of augmented Lagrangians for constrained nonconvex optimization," *SIAM Journal on Optimization*, vol. 15, no. 4, pp. 1128–1146, 2005.
- [15] C.-Y. Wang and D. Li, "Unified theory of augmented Lagrangian methods for constrained global optimization," *Journal of Global Optimization*, vol. 44, no. 3, pp. 433–458, 2009.
- [16] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317, Springer, Berlin, Germany, 1998.
- [17] D. P. Bertsekas and A. E. Ozdaglar, "Pseudonormality and a Lagrange multiplier theory for constrained optimization," *Journal of Optimization Theory and Applications*, vol. 114, no. 2, pp. 287–343, 2002.
- [18] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, NY, USA, 2000.
- [19] K. L. Teo, X. Q. Yang, and L. S. Jennings, "Computational discretization algorithms for functional inequality constrained optimization," *Annals of Operations Research*, vol. 98, pp. 215–234, 2000.
- [20] R. Y. K. Fung, J. Tang, and D. Wang, "Extension of a hybrid genetic algorithm for nonlinear programming problems with equality and inequality constraints," *Computers & Operations Research*, vol. 29, no. 3, pp. 261–274, 2002.