Multiplier Algorithm Based on A New Augmented Lagrangian Function

XUNZHI ZHU
Department of Mathematics
School of Science
Shandong University of Technology
CHINA
zhxzh1997@126.com

JINGYONG TANG
College of Mathematics and
Information Science
Xinyang Normal University
CHINA
tangjingyong926@163.com

JIAN WANG
Department of Mathematics
School of Science
Shandong University of Technology
CHINA
wjzhenhua@126.com

SHUJIANG DING
Department of Mathematics
School of Science
Shandong University of Technology
CHINA
dsjzb@sdut.edu.cn

Abstract: In this paper, for nonconvex optimization problem with both equality and inequality constrains, we introduce a new augmented Lagrangian function and propose the corresponding multiplier algorithm. The global convergence is established without requiring the boundedness of multiplier sequences. In particular, if the algorithm terminates in finite steps, then we obtain a KKT point of the primal problem; otherwise the iterative sequence $\{x^k\}$ generalized by algorithm converges to optimal solutions. Even if $\{x^k\}$ is divergent we also present a necessary and sufficient condition for the convergence of $\{f(x^k)\}$ to the optimal value. Finally, preliminary numerical experience is reported.

Key–Words: constrained global optimization, multiplier algorithm, global convergence

1 Introduction

Multipliers methods, also called augmented Lagrangian methods, have enjoyed a long and successful history as a tool for solving the nonlinear, convex or nonconvex optimization problems. Since its first proposal independently by Hestenes [10] and Powell [19] in 1969 and its comprehensive study by Rockafellar [20, 21, 22] and Bertsekas [3], multiplier methods have received much interest. These methods have been studied and applied to general mathematical programming problems involving various classes of constraints, such as equality constraints [3, 10, 19], inequality constraints, and complementarity constraints; see [3, 15, 20, 21].

In this paper, we consider the following nonlinear programming problem

$$
\begin{aligned}
(P) \qquad & \min\ f(x) \\
& \text{s.t. } g_i(x) \le 0, \quad i = 1, \cdots, m; \\
& \qquad h_j(x) = 0, \quad j = 1, \cdots, l;
\end{aligned}
$$

where $f, g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \cdots, m$ and $h_j : \mathbb{R}^n \to \mathbb{R}$ for $j = 1, \cdots, l$ are all twice differentiable functions. Denoted by X the feasible region and by X^* the solution set. (P) is called nonconvex optimization problem if $f(x)$ or X is nonconvex.

The standard Lagrangian function for this problem is defined as

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

where $(\lambda, \mu) \in \mathbb{R}_+^m \times \mathbb{R}^l$

Recall that a vector x is said to be a KKT point of (P) if there exist $\lambda_i \in \mathbb{R}_+$ for all $i = 1, \cdots, m$ and μ_j for $j = 1, \cdots, l$ such that the following system holds

$$
\begin{aligned}
& \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{l} \mu_j \nabla h_j(x) = 0, \\
& \lambda_i g_i(x) = 0, \quad \text{for all } i = 1, \cdots m,
\end{aligned}
$$

where the second condition is referred as the well-known *complementarity condition.* For notational simplification, the multipliers satisfying the above is denoted by $\Lambda(x)$.

The Lagrangian dual problem (D) is presented below.

$$(D) \qquad \max\ \theta(\lambda,\mu) \quad \text{s.t.}\ \lambda \geq 0$$

where $\theta(\lambda,\mu) = \inf_{x\in\mathbb{R}^n} L(x,\lambda,\mu)$.

It is well known that for any feasible solution x and $(\lambda,\mu) \in \mathbb{R}^m_+ \times \mathbb{R}^l$, the weak duality relation $f(x) \geq \theta(\lambda,\mu)$ always holds. Under suitable convexity assumptions, the strong duality theorem shows that the primal and dual problems have equal optimal objective values. Consequently, the classical Lagrangian method based on the above dual formulation has been widely applied to convex optimization due to the zero duality gap between the primal and dual problems. However, for a nonconvex constrained optimization problem a nonzero duality gap may be arisen by using the above Lagrangian. To overcome this drawback, various augmented Lagrangians have been introduced. The strong duality properties and exact penalization of different types of augmented Lagrangians or nonlinear Lagrangians have been studied by many researchers (see e.g., [1, 4, 5, 6, 11, 13, 15, 16, 18, 24, 25, 27]).

An indispensable assumption in the most existing global convergence analysis for augmented Lagrangian methods is that the multiplier sequence generated in the algorithms is bounded. This restrictive assumption confines applications of augmented Lagrangian methods in many practical situation. The important work on this direction includes [4, 8, 9, 12], where global convergence of modified augmented Lagrangian methods for nonconvex optimization with equality constraints was established. Recently, in the context of inequality-constrained global optimization, Luo et.al [14] proved the convergence properties of the primal-dual method based on four types of augmented Lagrangian functions without the boundedness assumption of the multiplier sequence.

In this paper, for the optimization problem (P) with both equality and inequality constraints, we introduce a new augmented Lagrangian function, which includes the well-known essentially quadratic augmented Lagrangian function in [21], and its structures are simpler and easier to solve than those in [11, 23]. It should be mentioned that it cannot be derived from the generalized augmented Lagrangian functions in [11, 23]. Based on this new augmented Lagrangian, we propose the corresponding multiplier algorithm and establish its global convergence properties. Specially, let $\{x^k\}$ be the iterative sequence generalized by algorithm. Then every limit point of $\{x^k\}$ is the optimal solution of (P). Compared with [14], 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. Finally, under Mangasarian-Fromovitz constraint qualification, we show that $\{x^k\}$ converge to a KKT point of (P).

The paper is organized as follows. In section 2, we propose the multiplier algorithm and study the global convergence properties. Some preliminary numerical results are reported in Section 3. The conclusion is drawn in Section 4.

2 Multiplier Algorithms

A new *generalized essential quadratic augmented Lagrangian function* for (P) is defined as follows,

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

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, i.e., $\mathbb{R}_{++} = \{a \in \mathbb{R} \mid a > 0\}$.

The function $\phi : \mathbb{R} \to \mathbb{R}$ involved in (1) needs to satisfy the following properties:

(A_1) ϕ is continuously differentiable and convex on $\mathbb{R}$;

(A_2) $\phi(0) = 0$, $\phi'(0) = 1$;

(A_3) $\lim_{t\to-\infty}\phi(t) = -\infty$.

If, in particular, $\phi(t) = t$, then L reduces to the essential quadratic augmented Lagrangian function introduced by Rockafellar, which has been studied by many authors; see [23] for more information. Note that the above three conditions ensure that ϕ is increasing over $\mathbb{R}$. In fact, it only needs to show that $\phi'(t) > 0$ for all t. If not, we can find t_0 such that $\phi'(t_0) \leq 0$. Since $\phi'(0) > 0$ by (A_2), then according to mean value theorem there exists t_1 satisfying $\phi'(t_1) = 0$, which in turn implies that t_1 is a global minimization of ϕ over $\mathbb{R}$, i.e., $\inf_{t\in\mathbb{R}}\phi(t) = \phi(t_1) > -\infty$, contradicting the condition (A_3).

Given (x,λ,μ,c), the *Lagrangian relaxation problem* associated with the augmented Lagrangian L is defined as

$$(L_{\lambda,\mu,c}) \qquad \min\ L(x,\lambda,\mu,c) \quad \text{s.t.}\ x \in \mathbb{R}^n.$$

Its solution set is denoted by $S^*(\lambda,\mu,c)$.

Throughout this paper we always assume that f is bounded from below, i.e.,

$$f_* = \inf_{x\in\mathbb{R}^n} f(x) > -\infty.$$

This assumption is rather mild in optimization problem, because otherwise the objective function f can be replaced by $e^{f(x)}$.

The multiplier algorithm based on the generalized essential quadric augmented Lagrangian L is proposed below. One of its main feature is that the Lagrangian multipliers associated with equality and inequality constraints are not restricted to be bounded, which make the algorithm applicable for many problems in practice.

Algorithm 1. (Multiplier algorithm based on L)*:*

Step 0. Select an initial point $x^0 \in \mathbb{R}^n$, $\lambda^0 \geq 0$, $\mu^0 \in \mathbb{R}$ and $c_0 > 0$. Set $k := 0$,

Step 1. Compute

$$\lambda_i^{k+1} = \max\{0, \phi(c_k g_i(x^k)) + \lambda_i^k\}\phi'(c_k g_i(x^k)),\quad \forall i = 1, \cdots, m, \tag{2}$$

$$\mu_j^{k+1} = \mu_j^k + c_k h_j(x^k),\quad \forall j = 1, \cdots, l, \tag{3}$$

$$c_{k+1} \geq (k+1)\max\{1, \sum_{i=1}^m (\lambda_i^{k+1})^2, \sum_{j=1}^l (\mu_j^{k+1})^2\}, \tag{4}$$

Step 2. Find $x^{k+1} \in S^*(\lambda^{k+1}, \mu^{k+1}, c_{k+1})$,

Step 3. If $x^{k+1} \in X$ and $(\lambda^{k+1}, \mu^{k+1}) \in \Lambda(x^{k+1})$, then STOP; otherwise, let $k := k+1$ and go back to Step 1.

Remark 2. *If we consider ϵ-global minimization of the subproblems, then only the global convergence property can be established. But, a natural question is when the multiplier sequences λ^k and μ^k are bounded, whether their accumulate points, say λ^* and μ^*, must be KKT multipliers of x^*. To answer this question, we considers exact global minimization of the subproblems.*

From (4), it is easy to see that

$$\frac{\lambda^k}{c_k},\ \frac{\mu^k}{c_k},\ \frac{(\lambda^k)^2}{c_k},\ \frac{(\mu^k)^2}{c_k} \to 0.\ (k\to\infty) \tag{5}$$

For establishing the convergence property of Algorithm 1, we first consider the perturbation analysis of (P). Given $\alpha \geq 0$, define the perturbation of feasible region as

$$X(\alpha) = \left\{ x \in \mathbb{R}^n \ \middle| \ \begin{array}{l} |g_i(x) \leq \alpha, |h_j(x)| \leq \alpha, \\ i = 1, \cdots, m, j = 1, \cdots, l \end{array} \right\},$$

the corresponding perturbation function as

$$v(\alpha) = \inf\{f(x) \mid x \in X(\alpha)\},$$

and the perturbation of level set as

$$L(\alpha) = \{x \in \mathbb{R}^n \mid f(x) \leq v(0) + \alpha\}.$$

It is clear that $X(0)$ coincides with the feasible set of (P).

Lemma 3. *For any λ, μ, and $c > 0$, one has*

$$S^*(\lambda, \mu, c) \subseteq \{x \in \mathbb{R}^n | L(x, \lambda, \mu, c) \leq v(0)\}.$$

Proof. For any $\bar{x} \in S^*(\lambda, \mu, c)$, we have

$$\begin{aligned}
& L(\bar{x}, \lambda, \mu, c) = \inf\{L(x, \lambda, \mu, c) | x \in \mathbb{R}^n\} \\
\leq\ & \inf\{L(x, \lambda, \mu, c) | x \in X(0)\} \\
=\ & \inf\{f(x) + \frac{1}{2c}\sum_{i=1}^m \{\max{}^2\{0, \phi(cg_i(x))\} \\
& \quad +\lambda_i\} - \lambda_i^2 | \ x \in X(0)\} \\
\leq\ & \inf\{f(x) | x \in X(0)\} \\
=\ & v(0),
\end{aligned}$$

where the second inequality uses the fact $\phi(cg_i(x)) \leq 0$ for all $x \in X(0)$, since ϕ is increasing as previous discussion. □

Lemma 4. *Let (λ^k, μ^k, c_k) be given as in Algorithm 1. For any $\epsilon > 0$, one has*

$$\{x \in \mathbb{R}^n | L(x, \lambda^k, \mu^k, c_k) \leq v(0)\} \subseteq X(\epsilon) \cap L(\epsilon).$$

whenever k is sufficiently large.

Proof. We prove this result by contradiction. Suppose that we can find an $\epsilon_0 > 0$ and an infinite subsequence $K \subseteq \{1, 2, \cdots\}$ such that

$$z^k \in \{x \in \mathbb{R}^n | L(x, \lambda^k, \mu^k, c_k) \leq v(0)\}, \quad \forall k \in K, \tag{6}$$

but

$$z^k \notin X(\epsilon_0) \quad or \quad z^k \notin L(\epsilon_0), \quad \forall k \in K.$$

Consider the following cases:

Case 1. $z^k \notin X(\epsilon_0)$, $k \in K$.

$$\begin{aligned}
& v(0) \geq L(z^k, \lambda^k, \mu^k, c_k) \qquad\qquad (7) \\
=\ & f(z^k) + \frac{1}{2c_k}\sum_{i=1}^m \left\{\max{}^2\{0, \phi(c_k g_i(z^k)) + \lambda_i^k\}\right. \\
& \left. -(\lambda_i^k)^2\right\} + \sum_{j=1}^l \mu_j^k h_j(z^k) + \frac{c_k}{2}\sum_{j=1}^l h_j^2(z^k).
\end{aligned}$$

Since $z^k \notin X(\epsilon_0)$, it needs to consider the following two subcases.

Subcase 1. There exist an index j_0 and an infinite subsequence $K_0 \subseteq K$ such that $|h_j(z^k)| \geq \epsilon_0$. It then follows from (7) that

$$\begin{aligned}
v(0) &\geq f_* + \sum_{j=1}^{l} \mu_j^k h_j(z^k) + \frac{c_k}{2}\sum_{j=1}^{l} h_j^2(z^k) \\
&\quad -\frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\
&= f_* + \frac{c_k}{2}\sum_{j=1}^{l}\{\big(h_j(z^k)+\frac{\mu_j^k}{c_k}\big)^2 - \frac{(\mu_j^k)^2}{c_k^2}\} \\
&\quad -\frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\
&= f_* + \frac{c_k}{2}\sum_{j\neq j_0}\big(h_j(z^k)+\frac{\mu_j^k}{c_k}\big)^2 \\
&\quad +\frac{c_k}{2}\{\big(h_{j_0}(z^k)+\frac{\mu_{j_0}^k}{c_k}\big)^2 - \frac{(\mu_{j_0}^k)^2}{c_k^2}\} \\
&\quad -\frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 - \frac{c_k}{2}\sum_{j\neq j_0}\frac{(\mu_j^k)^2}{c_k^2} \\
&\geq f_* - \frac{c_k}{2}\sum_{j\neq j_0}\frac{(\mu_j^k)^2}{c_k^2} - \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2, \\
&\quad +\frac{c_k}{2}\{\big(h_{j_0}(z^k)+\frac{\mu_{j_0}^k}{c_k}\big)^2 - \frac{(\mu_{j_0}^k)^2}{c_k^2}\} \\
&= f_* - \frac{1}{2}\sum_{j\neq j_0}\frac{(\mu_j^k)^2}{c_k} - \frac{1}{2}\sum_{i=1}^{m}\frac{(\lambda_j^k)^2}{c_k} \\
&+ \frac{c_k}{2}\{\big(h_{j_0}(z^k)+\frac{\mu_{j_0}^k}{c_k}\big)^2 - \frac{(\mu_{j_0}^k)^2}{c_k^2}\}. \qquad (8)
\end{aligned}$$

Using (5) and the fact $|h_{j_0}(z^k)| \geq \epsilon_0$ yield

$$\big(h_{j_0}(z^k)+\frac{\mu_{j_0}^k}{c_k}\big)^2 - \frac{(\mu_j^k)^2}{c_k} \geq \frac{1}{2}\epsilon_0^2,$$

whenever k is sufficiently large. This, together with (8), yields $v(0) = +\infty$ by taking $k \in K_0$ approaches to ∞, which leads to a contradiction.

Subcase 2. There exists an index i_0 and an infinite subsequence $K_0 \subseteq K$ such that $g_{i_0}(z^k) > \epsilon_0$. It follows from (7) that

$$\begin{aligned}
v(0) &\geq f_* + \sum_{j=1}^{l} \mu_j^k h_j(z^k) + \frac{c_k}{2}\sum_{j=1}^{l} h_j^2(z^k) \\
&\quad +\frac{1}{2c_k}\sum_{i=1}^{m}\max{}^2\{0, \phi(c_k g_i(z^k)) + \lambda_i^k\} \\
&\quad -\frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\
&= f_* + \frac{c_k}{2}\sum_{j=1}^{l}\{\big(h_j(z^k)+\frac{\mu_j^k}{c_k}\big)^2 - \frac{(\mu_j^k)^2}{c_k^2}\} \\
&\quad +\frac{1}{2c_k}\{\max{}^2\{0, \phi(c_k g_{i_0}(z^k)) + \lambda_{i_0}^k\} - (\lambda_{i_0}^k)^2\} \\
&\quad +\frac{1}{2c_k}\sum_{i\neq i_0}\max{}^2\{0, \phi(c_k g_i(z^k)) + \lambda_i^k\} \\
&\quad -\frac{1}{2c_k}\sum_{i\neq i_0}(\lambda_i^k)^2 \\
&\geq f_* - \frac{c_k}{2}\sum_{j=1}^{l}\frac{(\mu_j^k)^2}{c_k^2} + \frac{1}{2c_k}(c_k)^2 g_{i_0}^2(z^k) \\
&\quad +g_{i_0}(z^k)\lambda_{i_0}^k - \frac{1}{2c_k}\sum_{i\neq i_0}(\lambda_i^k)^2 \\
&\geq f_* + \frac{c_k}{2}\epsilon_0^2 + \epsilon_0\lambda_{i_0}^k \\
&\quad -\frac{1}{2c_k}\sum_{j=1}^{l}\frac{(\mu_j^k)^2}{c_k^2} - \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\
&\geq f_* + \frac{c_k}{2}\epsilon_0^2 - \frac{1}{2c_k}\sum_{j=1}^{l}\frac{(\mu_j^k)^2}{c_k^2} - \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2,
\end{aligned}$$

where the second inequality comes from the convex inequality $\phi(a) \geq \phi(0) + \phi'(0)a = a$ for all $a \in \mathbb{R}$ and the last inequality follows from the nonnegativity of λ by (2). Taking limits in the above inequality yields $v(0) = +\infty$, which is a contradiction

Case 2. $z^k \notin L(\epsilon_0), \forall k \in K$.

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

$$\leq \quad v(0) + \frac{1}{2c_k}\sum_{j=1}^{l}(\mu_j^k)^2 + \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2. \quad (9)$$

When k is large enough, it follows from (5) that

$$\frac{1}{c_k}\sum_{j=1}^{l}(\mu_j^k)^2 + \frac{1}{c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \leq \epsilon_0,$$

which together with (9) means $f(z^k) \leq v(0) + \epsilon_0$, i.e., $z^k \notin L(\epsilon_0)$, this is a contradiction. The proof is complete. □

With these preparation, the global convergence property of Algorithm 1 can be given, which shows that if the algorithm terminates in finite steps, then we obtain a KKT point of (P); otherwise every limit point of $\{x^k\}$ would be an optimal solution of (P).

Theorem 5. *Let $\{x^k\}$ be the iterative sequence generated by Algorithm 1. Then If $\{x^k\}$ is terminated in finite steps, then we get the KKT point of (P); otherwise, every limit point of $\{x^k\}$ belongs to X^*.*

Proof. According to the construction of Algorithm 1, the first part is clear. For the second case, it follows from Lemma 3 that when k is large enough we have

$$S^*(x^k, \lambda^k, \mu^k) \subseteq X(\epsilon) \cap L(\epsilon).$$

Thus,

$$x^k \in X(\epsilon) \cap L(\epsilon). \quad (10)$$

Note that $X(\epsilon)$ and $L(\epsilon)$ are closed, due to the continuity of f, g_i $(i = 1, \cdots, m)$ and h_j $(j = 1, \cdots, l)$. Taking the limit in (10) yields $x^* \in X(\epsilon) \cap L(\epsilon)$, which further shows that $x^* \in X(0) \cap L(0)$, since $\epsilon > 0$ is arbitrary, i.e., $x^* \in X^*$. The proof is complete. □

The foregoing result is applicable to the case when $\{x^k\}$ at least has an accumulation point. However, a natural question arises: how does the algorithm perform as $\{x^k\}$ is divergent? The following theorem gives an answer.

Theorem 6. *Let $\{x^k\}$ be an iterative sequence generalized by Algorithm 1. Suppose that $\lim_{\epsilon\to 0^+} v(\epsilon) > -\infty$, then the following statements are equivalent:*

(1) $\lim_{k\to\infty} f(x^k) = v(0)$;

(2) $v(\alpha)$ is continuous at $\alpha = 0$.

Proof. $(2) \Rightarrow (1)$. According to the proof of Theorem 5 (cf. (10)), we know that

$$v(\epsilon) \leq f(x^k) \leq v(0) + \epsilon, \quad (11)$$

whenever k is sufficiently large. Since $v(\alpha)$ is continuous at $\alpha = 0$, taking the lower limitation in (11) yields

$$\begin{aligned} v(0) &= \lim_{\epsilon\to 0^+} v(\epsilon) \leq \liminf_{k\to\infty} f(x^k) \\ &\leq \limsup_{k\to\infty} f(x^k) \leq v(0), \end{aligned}$$

i.e.,

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

$(1) \Rightarrow (2)$. Note that $v(\epsilon) \leq v(0)$ for all $\epsilon \geq 0$, since $X(0) \subseteq X(\alpha)$, then we must have $\lim_{\epsilon\to 0^+} v(\epsilon) < v(0)$. Suppose on the contrary that v is not continuous at zero from right, then there must exist $\delta_0 > 0$ as $\epsilon > 0$ is small enough such that

$$v(\epsilon) \leq v(0) - \delta_0, \quad \forall j. \quad (12)$$

For any given c_k, we can choose a ϵ_k satisfying

$$\epsilon_k c_k \to 0 \quad \text{as } k \to \infty,$$

In addition, let $z^k \in X(\epsilon_k)$ with $f(z^k) \leq v(\epsilon_k) + \frac{\delta_0}{2}$, which further implies $f(z^k) \leq v(0) - \frac{\delta_0}{2}$ by (12). Therefore,

$$\begin{aligned} & f(x^k) = L(x^k, \lambda^k, \mu^k, c_k) \\ & -\frac{1}{2c_k}\sum_{i=1}^{m}\{\max{}^2\{0, \phi(c_k g_i(x^k)) + \lambda_i^k\} - (\lambda_i^k)^2\} \\ & -\sum_{j=1}^{l}\mu_j^k h_j(x^k) - \frac{c_k}{2}\sum_{j=1}^{l} h_j^2(x^k) \\ & \leq \inf_{x\in\mathbb{R}^n} L(x, \lambda^k, \mu^k, c_k) - \frac{c_k}{2}\sum_{j=1}^{l}\{\big(h_j(x^k) \\ & +\frac{\mu_j^k}{c_k}\big)^2 - \frac{(\mu_j^k)^2}{c_k^2}\} + \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\ & \leq \inf_{x\in\mathbb{R}^n} L(x, \lambda^k, \mu^k, c_k) \\ & +\frac{1}{2c_k}\sum_{j=1}^{l}(\mu_j^k)^2 + \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\ & \leq f(z^k) + \sum_{j=1}^{l}\mu_j^k h_j(z^k) + \frac{c_k}{2}\sum_{j=1}^{l} h_j^2(z^k) \\ & +\frac{1}{2c_k}\sum_{i=1}^{m}\{\max{}^2\{0, \phi(c_k g_i(z^k) + \lambda_i^k)\} - (\lambda_i^k)^2\} \\ & +\frac{1}{2c_k}\sum_{j=1}^{l}(\mu_j^k)^2 + \frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \end{aligned}$$

$$
\begin{aligned}
&= f(z^k) + \frac{c_k}{2}\sum_{j=1}^{l}\big\{\big(h_j(z^k)+\frac{\mu_j^k}{c_k}\big)^2 - \frac{(\mu_j^k)^2}{c_k^2}\big\} \\
&+\frac{1}{2c_k}\sum_{i=1}^{m}\{\max{}^2\{0,\phi(c_k g_i(z^k)+\lambda_i^k)\}-(\lambda_i^k)^2\} \\
&+\frac{1}{2c_k}\sum_{j=1}^{l}(\mu_j^k)^2+\frac{1}{2c_k}\sum_{i=1}^{m}(\lambda_i^k)^2 \\
&\le v(0)-\frac{\delta_0}{2}+\frac{1}{2c_k}\sum_{j=1}^{l}\big(\epsilon_{j_k}+\frac{|\mu_j^k|}{c_k}\big)^2 \\
&+\frac{1}{2c_k}\sum_{i=1}^{m}\big(c_k\epsilon_{j_k}+\lambda_i^k\big)^2, \qquad (13)
\end{aligned}
$$

where the last step is due to the fact $|h_j(z^k)| \le \epsilon_k$ and $g_i(z^k) \le \epsilon_k$ since $z^k \in X(\epsilon_k)$ and ϕ is increasing. Taking the limits in both sides of (13) and using Theorem 6 we get

$$v(0)=\lim_{k\to\infty} f(x^k) \le v(0)-\frac{\delta_0}{2}$$

which leads to a contradiction. The proof is complete. $\square$

At a point x^*, denote by

$$I(x^*)=\{i \mid g_i(x^*)=0, i=1,\cdots,m\}$$

the active inequality constraints. The Mangasarian-Fromovitz constraint qualification is that $\nabla h_j(x^*)$ for $j=1,\cdots,l$ are linearly independent and there exists $h\in \mathbb{R}^n$ such that

$$
\begin{aligned}
&\langle \nabla g_i(x^*), h\rangle < 0, \quad \forall i\in I(x^*) \\
\text{and}\quad &\langle \nabla h_j(x^*), h\rangle = 0, \quad \forall j=1,\cdots,l. \qquad (14)
\end{aligned}
$$

The linear independent constraint qualification is that $\nabla h_j(x^*)$ for $j=1,\cdots,l$ and ∇g_i for $i\in I(x^*)$ are linearly independent.

Theorem 7. *Let $\{x^k\}$ be the iterative sequence generated by Algorithm 1. Then*

1. *If $\lim_{k\to\infty} x^k = x^*$ and the M-F constraint qualification is satisfied at x^*, then $\{\lambda^k,\mu^k\}$ is bounded and any of its limit points, say (λ^*,μ^*), satisfies that (x^*,λ^*,μ^*) belongs to the KKT system of (P).*

2. *If the linearly independent constraint qualification holds at x^*, we further obtain that the multiplier sequence $\{\lambda_i^k\}$ and $\{\mu_j^k\}$ are convergent.*

Proof. If $\lim_{k\to\infty} x^k = x^*$, then it follows from Theorem 5 that $x^*\in X^*$. If $i\notin I(x^*)$, then $g_i(x^k)<0$ as k large enough. This means the existence of $\epsilon_0>0$ such that $g_i(x^k)\le -\epsilon_0$ whenever k is sufficiently large. Therefore,

$$\lim_{k\to\infty} c_k g_i(x^k) = -\infty. \qquad (15)$$

Taking into account of the properties (A1), (A2) and Step 1 in Algorithm 1, we obtain

$$
\begin{aligned}
\lambda_i^{k+1} &= \max\{0,\phi(c_k g_i(x^k))+\lambda_i^k\}\phi'(c_k g_i(x^k)) \\
&\le \max\{0,\lambda_i^k\}=\lambda_i^k\le\cdots\le\lambda_i^0, \qquad (16)
\end{aligned}
$$

where the second equation comes from the nonnegativity of λ_i according to the construction in Algorithm 1. This justifies the boundedness $\{\lambda_i^k\}$ for $i\notin I(x^*)$. Taking limit in (16) and using (15) and the fact $\lim_{k\to\infty}\phi(c_k g_i(x^k))=-\infty$ by A3 yield

$$\lim_{k\to\infty}\lambda_i^k = 0, \quad \forall i\notin I(x^*). \qquad (17)$$

We now show that λ_i^k for $i\in I(x^*)$ and μ_j^k for $j=1,2,\cdots,l$ are bounded as well. If this is not true, then we can find an infinite subsequence $K\subseteq\{1,2,\cdots,\}$ such that

$$T_k := \sum_{i\in I(x^*)}\lambda_i^k+\sum_{j=1}^{l}|\mu_j^k|\to+\infty, \quad k\to\infty. \qquad (18)$$

Since $x^{k-1}\in S^*(\lambda^{k-1},\mu^{k-1},c_k)$ by Algorithm 1, according to the well-known of optimality conditions we must have

$$\nabla_x L(x^{k-1},\lambda^{k-1},\mu^{k-1},c_k)=0,$$

i.e.,

$$\nabla f(x^{k-1})+\sum_{i=1}^{m}\lambda_i^k\nabla g_i(x^{k-1})+\sum_{j=1}^{l}\mu_j^k\nabla h_i(x^{k-1})=0. \qquad (19)$$

Since $\frac{\lambda_i^k}{T_k}$ and $\frac{|\mu_j^k|}{T_k}$ are bounded, we can assume without loss of generality that

$$\frac{\lambda_i^k}{T_k}\to\tilde{\lambda}_i^*, \quad \forall i\in I(x^*)$$

and

$$\frac{|\mu_j^k|}{T_k}\to\tilde{\mu}_i^*, \ \forall j=1,2,\cdots,l.$$

Since

$$\sum_{i\in I(x^*)}\frac{\lambda_i^k}{T_k}+\sum_{j=1}^{l}\frac{|\mu_j^k|}{T_k}=1, \quad \forall k \qquad (20)$$

taking limit with respect to $k \in K$ in equation (20) gives us

$$\sum_{i\in I(x^*)} \tilde{\lambda}_i^* + \sum_{j=1}^{l} \tilde{\mu}_i^* = 1,$$

which implies that $\tilde{\lambda}_i^*$ for $i \in I(x^*)$ and $\tilde{\mu}_j^*$ for $j = 1, \cdots, m$ are not all zero. Dividing on both sides of (19) by T_k, taking limit with respect to $k \in K$, and using (17) and (18), we get

$$\sum_{i\in I(x^*)} \tilde{\lambda}_i^k \nabla g_i(x^*) + \sum_{j=1}^{l} \tilde{\mu}_j^k \nabla h_j(x^*) = 0. \quad (21)$$

We claim that $\tilde{\lambda}_i^*$ for $i \in I(x^*)$ are not all zero. In fact, if $\tilde{\lambda}_i^* = 0$ for all $i \in I(x^*)$, then (21) takes the form

$$\sum_{j=1}^{l} \tilde{\mu}_j^k \nabla h_j(x^*) = 0,$$

which implies $\tilde{\mu}_j^* = 0$ for $j = 1, \cdots, m$, since $\nabla h_j(x^*)$ are linear independent. This contradicts the fact that $\tilde{\lambda}_i^*$ for $i \in I(x^*)$ and $\tilde{\mu}_j^*$ for $j = 1, \cdots, m$ are not all zero, as shown above.

Multiplying h in both sides of (21) yields

$$\begin{aligned} 0 &= \sum_{i\in I(x^*)} \tilde{\lambda}_i^k \langle \nabla g_i(x^*), h\rangle + \sum_{j=1}^{l} \tilde{\mu}_j^k \langle \nabla h_j(x^*), h\rangle \\ &= \sum_{i\in I(x^*)} \tilde{\lambda}_i^k \langle \nabla g_i(x^*), h\rangle < 0, \end{aligned}$$

where the inequality comes from (14) and the fact that $\tilde{\lambda}_i^*$ for $i \in I(x^*)$ are not all zero. This leads to a contradiction. Therefore, we establish the boundedness of $\{\lambda^k, \mu^k\}$. Let (λ^*, μ^*) be a limit point of $\{\lambda^k, \mu^k\}$. It follows from (17) that $\lambda^* = 0$ for $i \notin I(x^*)$ and from (19) that

$$\nabla f(x^*) + \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{l} \mu_j^* \nabla h_i(x^*) = 0.$$

This establishes Part (a). Part (b) can be proved in the same vein, just noting that in the presence of linearly independence constraint qualification, the Lagrangian multiplier must be unique. This together with the boundedness of $\{\lambda_i^k\}$ for $i = 1, \cdots, m$ and $\{\mu_j^k\}$ for $j = 1, \cdots, l$ ensure the convergence of these sequences to the unique accumulation point. □

3 Numerical Reports

To give some insight into the behavior of our proposed algorithm presented in this paper, we solve a nonconvex programming problems by letting ϕ take the following different functions:

1. $\phi_1(\alpha) = \alpha$,
2. $\phi_2(\alpha) = \ln(\frac{1+e^\alpha}{2}) + \frac{1}{2}\alpha$.
3. $\phi_3(\alpha) = \sinh\alpha = \frac{e^\alpha - e^{-\alpha}}{2}$.
4. $\phi_4(\alpha) = \begin{cases} -\frac{1}{4}\ln(-2\alpha) - \frac{3}{8}, & \alpha \le -\frac{1}{2}, \\ \alpha + \frac{1}{2}\alpha^2 & otherwise; \end{cases}$

In the following, we solve the following three nonlinear programming by choosing ϕ to take the above four different functions. The test was done at a PC of Pentium 4 with 2.8GHz CPU and 1.99GB memory, and the preliminary numerical experience are reported below, where k is the number of iterations, c_k is the penalty parameter, λ^k and μ^k is multipliers, x^k is iterative point found by the algorithm, and $f(x^k)$ is the objective value. In the implementation of our algorithm, we use the BFGS Quasi-Newton method [2] with a mixed quadratic and cubic line search procedure to solve the Lagrangian subproblem.

Example 8. *[26]*

$$\begin{aligned} \min\ & 4x_1^3 - 3x_1 + 6x_2 \\ \text{s.t.}\ & x_1 + x_2 \le 4 \\ & x_1 - 4x_2 \le 0 \\ & 2x_1^2 + x_2 = 5 \\ & x_1 \ge 0, x_2 \ge 0 \end{aligned}$$

The global optimal solution is $x^* = (\frac{\sqrt{5}}{2}, 5 - \sqrt{5})$ with optimal value $f(x^*) = 18.8197$. Numerical results are reported in Table 1, where the initial data are $x^0 = (0, 0)$, $c_0 = 1$, and $(\lambda^0, \mu^0) = (1, 1, 1)$.

Example 9. *Beale problem [7]*

$$\begin{aligned} \min & \left\{ \begin{array}{c} -8x_1 - 6x_2 - 4x_3 + 2x_1^2 + 2x_2^2 \\ +x_3^2 + 2x_1x_2 + 2x_1x_3 + 9 \end{array} \right\} \\ \text{s.t.}\ & x_1 + x_2 + 2x_3 \le 3 \\ & x_i \ge 0, i = 1, 2, 3. \end{aligned}$$

The global optimal solution is $x^* = (\frac{4}{3}, \frac{7}{9}, \frac{4}{9})$ with optimal value $f(x^*) = \frac{1}{9}$. Numerical results are reported in Table 2, where the initial data are $x^0 = (0.5, 0.5, 0.5)$, $c_0 = 1$, and $(\lambda^0, \mu^0) = (1, 1, 1, 1)$.

Table 1: Result of Example 8

$\phi_i(s)$	k	c_k	x^k	$f(x^k)$
$\phi_1(s)$	1	1.0000	(1.1178, 2.7018)	18.4441
	3	143.9365	(1.1179, 2.7643)	18.8203
	7	252.0040	(1.1180, 2.7640)	18.8197
$\phi_2(s)$	1	1.0000	(1.1159, 0.2790)	3.8845
	3	2.0000	(1.1178, 2.7653)	18.8251
	4	107.1985	(1.1180, 2.7640)	18.8197
$\phi_3(s)$	1	1.0000	(1.1150, 0.2779)	3.8672
	3	2.0000	(1.1181, 2.7639)	18.8196
	4	143.9365	(1.1180, 2.7640)	18.8197
$\phi_4(s)$	1	1.0000	(1.1162, 0.2784)	3.8843
	2	2.0000	(1.1179, 2.7639)	18.8179
	3	107.2257	(1.1180, 2.7640)	18.8197

Example 10. *[17]*

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

The global optimal solution is $x^* = (0,1,2,-1)$ with optimal value $f(x^*) = -44$. Numerical results are reported in Table 3, where the initial data are $x^0 = (2,2,2,2)$, $c_0 = 1$, and $(\lambda^0,\mu^0) = (1,1,1)$.

4 Conclusion

A new generalized essentially quadratic augmented Lagrangian function and the corresponding multiplier algorithm are proposed. Particularly, the global convergence property is established without requiring the boundedness of Lagrangian multiplier sequence. In addition, a necessary and sufficient condition for the convergence of $f(x_k)$ to the optimal value is given. This guarantees that the algorithm can be applicable for many problems in practice. Our numerical reports indicate that the generalization of ϕ from linear function to nonlinear function is not merely generalization for its own sake but also can obtain better convergence performance. As our future work, one of the interesting and important topics is whether these nice properties could be extended to more general cone programming, such as second-order cone programming and semidefinite programming.

Table 2: Result of Example 9

$\phi_i(s)$	k	c_k	x^k	$f(x^k)$
$\phi_1(s)$	1	1.0000	(1.5452, 0.6365, 0.0907)	0.2977
	2	1.0000	(1.2725, 0.8180, 0.5457)	0.0744
	4	10.0000	(1.3259, 0.7824, 0.4563)	0.1065
	7	100.0000	(1.3333, 0.7777, 0.4443)	0.1111
$\phi_2(s)$	1	1.0000	(1.5408, 0.6389, 0.0980)	0.2932
	2	1.0000	(1.3689, 0.7540, 0.3835)	0.1369
	4	10.0000	(1.3344, 0.7775, 0.4429)	0.1116
	7	10.0000	(1.3333, 0.7777, 0.4443)	0.1111
$\phi_3(s)$	1	1.0000	(1.5134, 0.6578, 0.1458)	0.2626
	2	1.0000	(1.3606, 0.7598, 0.3978)	0.1306
	4	1.0000	(1.3335, 0.7775, 0.4430)	0.1118
	6	10.0000	(1.3333, 0.7777, 0.4443)	0.1111
$\phi_4(s)$	1	1.0000	(1.5445, 0.6369, 0.0918)	0.2970
	2	1.0000	(1.3720, 0.7518, 0.3997)	0.1291
	3	1.0000	(1.3346, 0.7775, 0.4433)	0.1114
	4	1.0000	(1.3333, 0.7777, 0.4443)	0.1111

Acknowledgements: Research of this paper is supported by National Natural Science Foundation of China (11271233), Shandong Province Natural Science Foundation (ZR2012AM016), and Basic and Frontier Technology Research Project of Henan Province (142300410318).

References:

[1] R. Andreani, E. G.Birgin, J. M. Martínez, On Augmented Lagrangian methods with general lower-level constraints, *SIAM J. Optim.* 18, 2007, pp. 1286-1309.

[2] M. S. Bazaraa, H. D. Sherali, C. M. Shetty, Nonlinera programming theory and algorithms, New York: John Wiley and Sons, 1993.

[3] D. P. Bertsekas, Constrained Optimization and Lagrange Multipliers Method, Academic Press, Inc., 1982.

[4] E. G. Birgin, C. A. Floudas, J. M. Martínez, Global minimization using an Augmented Lagrangian method with variable lower-level constraints, *Math. Program.*, 125, 2010, pp. 139-162.

[5] R. S. Burachik, A. M. Rubinov, On the absence of duality gap for Lagrange-type functions, *J. Industrial Manage. Optim.*, 1, 2005, pp. 33-38.

[6] R. S. Burachik, A. M.Rubinov, Abstract convexity and augmented lagrangians, *SIAM J. Optim.* 18, 2007, pp. 413-436.

[7] C. Charalambous, Nonlinear least path optimization and nonlinear programming, *Math. Programming*, 12, 1977, pp. 195-225.

Table 3: Result of Example 10

ϕ_i	k	c_k	x^k	$f(x^k)$
ϕ_1	3	8.4614	(-0.0069, 0.9967, 1.9901, -1.0061)	-43.8571
	6	24.2390	(-0.0011, 0.9996, 1.9998, -1.0034)	-44.0077
	11	48.5455	(-0.0001, 1.0000, 2.0000, -1.0000)	-43.9995
ϕ_2	3	8.4430	(0.0017, 1.0017, 2.0063, -1.0073)	-44.1314
	6	24.7705	(-0.0033, 0.9997, 1.9994, -1.0081)	-44.0152
	11	48.7133	(0.0000, 1.0000, 2.0000, -1.0000)	-44.0000
ϕ_3	1	3.8876	(0.1147, 0.5510, 1.9888, -1.0459)	-43.0933
	6	9.8546	(-0.0045, 0.9969, 2.0025, -0.9983)	-43.9926
	8	48.2683	(0.0000, 1.0000, 2.0000, -1.0000)	-44.0000
ϕ_4	1	7.7614	(0.0049, 1.0029, 2.0091, -1.0005)	-44.1538
	4	24.6380	(-0.0031, 0.9995, 1.9997 -1.0069)	-44.0135
	9	48.5943	(0.0000, 1.0000, 2.0000, -1.0000)	-44.0000

[8] A. R.Conn, N. I. M.Gould, P. L. Toint, A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM J. Numer. Anal.* 28, 1991, pp. 545-572.

[9] A. R. Conn, N. I. M.Gould, A.Sartenaer, P.L.Toint, onvergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints *SIAM J. Optim.* 6, 1996, pp. 674-703.

[10] M. R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.* 4, 1969, pp. 303-320.

[11] X. X. Huang, X. Q. Yang, A unified augmented Lagrangian approach to duality and exact penalization, *Math. Oper. Res.* 28, 2003, pp. 524-532.

[12] R. M. Lewis, V. Torczon, A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds, *SIAM J. Optim.*, 12, 2002, pp. 1075-1089.

[13] D. Li, X. L. Sun, Local convexification of the Lagrangian function in nonconvex optimization, *J. Optim. Theory Appl.*, 104, 2000, pp. 109-120.

[14] H. Z. Luo, X. L. Sun, D. Li, On the convergence of augmented Lagrangian methods for constrained global optimization, *SIAM J. Optim.*, 18, 2007, pp. 1209-1230.

[15] H. Z. Luo, X. L. Sun, Y. F. Xu, Convergence properties of modified and partially-augmented Lagrangian methods for mathematical programs with complementarity contraints, *J. Optim. Theory Appl.*, 145, 2010, pp. 489-506.

[16] H. Z. Luo, G. Mastroeni, H. X. Wu, Separation Approach for Augmented Lagrangians in Constrained Nonconvex Optimization, *J. Optim. Theory Appl.* 144, 2010, pp. 275-290.

[17] G. D. Pillo and L. Grippo, A New Augmented Lagrangian Function for Inequality Constraints in Nonlinear Programming Problems, *JOTA*. 36,1982, pp. 495- 519.

[18] R. A. Polyak, Nonlinear rescaling vs. smoothing technique in convex optimization, *Math. Program.* 92, 2002, pp. 197-235.

[19] M. J. D.Powell, *A Method for Nonlinear Constraints in Minimization Problems*, (R.Fletcher, ed.) Academic Press, New York, 1969.

[20] R. T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, *J. Optim. Theory Appl.* 12, 1973 pp. 555-562.

[21] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM. J. Control Optim.* 12, 1974, pp. 268-285.

[22] R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming *Math. Oper. Res.* 1,1976,pp. 97-116.

[23] R. T. Rockafellar, J. B. Wets: Variational Analysis. Springer, Berlin, 1998.

[24] A. M. Rubinov, X. Q. Yang, Lagrange-type Functions in Constrained Non-Convex Optimization, Kluwer Academic Publishers, Dordrecht, 2003.

[25] X. L. Sun, D. Li, K. I. M. McKinnon, On saddle points of augmented Lagrangians for constrained nonconvex optimization *SIAM J. Optim.* 15 , 2005 pp. 1128-1146.

[26] H. P. Williams, Model building in mathematical programming, 3rd ed., New York: Wiley, 1990.

[27] X. Q. Yang, X. X. Huang, A nonlinear Lagrangian approach to constrained optimization problems, *SIAM J. Optim.* 11, 2001, pp. 1119-1144.