Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2011, Article ID 902131, 14 pages
doi:10.1155/2011/902131

Research Article

New Convergence Properties of the Primal Augmented Lagrangian Method

Jinchuan Zhou, Xunzhi Zhu, 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 23 August 2011; Revised 25 November 2011; Accepted 26 November 2011

Academic Editor: Simeon Reich

New convergence properties of the proximal augmented Lagrangian method is established for continuous nonconvex optimization problem with both equality and inequality constrains. In particular, the multiplier sequences are not required to be bounded. Different convergence results are discussed dependent on whether the iterative sequence $\{x^k\}$ generated by algorithm is convergent or divergent. Furthermore, under certain convexity assumption, we show that every accumulation point of $\{x^k\}$ is either a degenerate point or a KKT point of the primal problem. Numerical experiments are presented finally.

1. Introduction

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

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

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

Augmented Lagrangian algorithms are very popular tools for solving nonlinear programming problems. At each outer iteration of these methods, a simpler optimization problem is solved, for which efficient algorithms can be used, especially when the problems are large. The most famous augmented Lagrangian algorithm based on the Powell-Hestenes-Rockafellar [1–3] formula has been successfully used for defining practical nonlinear programming algorithms [4–7]. At each iteration, a minimization problem with simple constraints is approximately solved whereas Lagrange multipliers and penalty parameters are updated in the master routine. The advantage of the Augmented Lagrangian approach over other methods is that the subproblems can be solved using algorithms that can deal with a very large number of variables without making use of factorization of matrices of any kind.

An indispensable assumption in the most existing global convergence analysis for augmented Lagrangian methods is that the multiplier sequence generated by the algorithms is bounded. This restrictive assumption confines applications of augmented Lagrangian methods in many practical situation. The important work on this direction includes [8], where global convergence of modified augmented Lagrangian methods for nonconvex optimization with equality constraints was established; and Andreani et al. [4] and Birgin et al. [9] investigated the augmented Lagrangian methods using safeguarding strategies for nonconvex constrained problems. Recently, for inequality-constrained global optimization, Luo et al. [10] established the convergence properties of the primal-dual method based on four types of augmented Lagrangian functions without the boundedness assumption of the multiplier sequence. More information can be found in [5, 11, 12].

In this paper, for the optimization problem (P) with both equality and inequality constraints, we further study the convergence property of the proximal Lagrangian method without requiring the boundedness of multiplier sequences. The main contribution of this paper lies in the following three aspects. First, more general constraints are considered, without restricting only inequality constraints as in [10, 13] and requiring boundedness of X as in [9]. Second, an essential assumption on the global convergence properties given in [4–7, 9, 10] is that the iterative sequence $\{x^k\}$ must be convergent in advance; here, we further discuss the case when $\{x^k\}$ is divergent and develop a necessary and sufficient condition for $\{f(x^k)\}$ converging to the optimal value of primal problem. Third, the definition of *degeneration* in [9, 10] is extended from inequality constraint to both inequality and equality constraints.

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

2. Multiplier Algorithms

The primal augmented Lagrangian function for (P) is

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

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

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

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

Given $\varepsilon \geq 0$, then the ε-optimal solution set of $(L_{\lambda,\mu,c})$, denoted by $S^*(\lambda, \mu, c, \varepsilon)$, is defined as

$$\left\{ x \in \Omega \mid L(x, \lambda, \mu, c) \leq \inf_{x \in \Omega} L(x, \lambda, \mu, c) + \varepsilon \right\}. \tag{2.2}$$

If Ω is closed and bounded, then the global optimal solution of $(L_{\lambda,\mu,c})$ exists. However, if Ω is unbounded, then $(L_{\lambda,\mu,c})$ maybe unsolvable. To overcome this difficultly, we assume throughout this paper that f is bounded on Ω from below, that is,

$$f_* := \inf_{x \in \Omega} f(x) > -\infty. \tag{2.3}$$

This assumption is rather mild in optimization programming, because otherwise the objective function f can be replaced by $e^{f(x)}$. It ensures that the ε-optimal solution set with $\varepsilon > 0$ always exists, since $L(x, \lambda, \mu, c)$ is bounded from below by (2.1) and (2.3).

Recall that a vector x^* is said to be a KKT point of (P) if there exist $\lambda_i^* \geq 0$ for each $i = 1, \ldots, m$ and μ_j^* for each $j = 1, \ldots, l$ such that

$$0 \in \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x) + \sum_{j=1}^{l} \mu_j \nabla h_j(x) + \mathcal{N}_\Omega(x^*), \quad \lambda_i^* g_i(x^*) = 0, \quad \forall i = 1, \ldots, m, \tag{2.4}$$

where $\mathcal{N}_\Omega(x^*)$ denotes the normal cone of Ω at x^*. The collection set of all λ^* and μ^* satisfying (2.4) is denoted by $\Lambda(x^*)$.

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

Algorithm 2.1 (Multiplier algorithm based on L).

Step 1. Select an initial point $x^0 \in \mathbb{R}^n$, $\lambda^0 \geq 0$, $\mu_0 \in \mathbb{R}$, $c_0 > 0$, and $\varepsilon_0 \geq 0$. Set $k := 0$.

Step 2. Compute

$$\lambda_i^{k+1} = \max\left\{0, c_k g_i\left(x^k\right) + \lambda_i^k\right\}, \quad \forall i = 1, \ldots, m, \tag{2.5}$$

$$\mu_j^{k+1} = \mu_j^k + c_k h_j\left(x^k\right), \quad \forall j = 1, \ldots, l, \tag{2.6}$$

$$\varepsilon_{k+1} = \frac{\varepsilon_k}{k+1}, \tag{2.7}$$

$$c_{k+1} \geq (k+1) \max \left\{ 1, \sum_{i=1}^{m} \left(\lambda_i^{k+1}\right)^2, \sum_{j=1}^{l} \left(\mu_j^{k+1}\right)^2 \right\}. \tag{2.8}$$

Step 3. Find $x^{k+1} \in S^*(\lambda^{k+1}, \mu^{k+1}, c_{k+1}, \varepsilon_{k+1})$;

Step 4. 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 2.

The iterative formula for ε_{k+1} given in (2.7) is just used to guarantee its convergence to zero. In fact, in the practical numerical experiment, we can choose $\varepsilon_{k+1} = \varepsilon_k / c_k$ to improve the convergence of the algorithm. The following lemma gives the relationship between the penalty parameter c_k and the multipliers λ^k and μ^k.

Lemma 2.2. *Let (λ^k, μ^k, c_k) be given as in Algorithm 2.1, then the following terms*

$$\frac{\lambda^k}{c_k}, \frac{\mu^k}{c_k}, \frac{\left(\lambda^k\right)^2}{c_k}, \frac{\left(\mu^k\right)^2}{c_k} \tag{2.9}$$

all approach to zero as $k \to \infty$.

Proof. This follows immediately from (2.8). □

For establishing the convergence property of Algorithm 2.1, we first consider the perturbation analysis of (P). Given $\alpha \geq 0$, define the perturbation of feasible region as

$$X(\alpha) = \left\{ x \in \Omega \mid g_i(x) \leq \alpha, \left|h_j(x)\right| \leq \alpha, \ i = 1, \ldots, m, \ j = 1, \ldots, l \right\}, \tag{2.10}$$

and the perturbation of level set as

$$L(\alpha) = \left\{ x \in \Omega \mid f(x) \leq v(0) + \alpha \right\}. \tag{2.11}$$

It is clear that $X(0)$ coincides with the feasible set of (P). The corresponding perturbation function is given as

$$v(\alpha) = \inf\left\{ f(x) \mid x \in X(\alpha) \right\}. \tag{2.12}$$

The following result shows that the perturbation value function is upper semicontinuous at zero.

Lemma 2.3. *The perturbation function v is upper semicontinuous at zero from right.*

Proof. Since $X(0) \subset X(\alpha)$ for any $\alpha \geq 0$, then $v(\alpha) \leq v(0)$ by definition (2.12). This implies that $\limsup_{\alpha \to 0^+} v(\alpha) \leq v(0)$. □

Lemma 2.4. *Let $(\lambda^k, \mu^k, c_k, \varepsilon_k)$ be given as in Algorithm 2.1. For any $\varepsilon > 0$, one has*

$$S^*\left(\lambda^k, \mu^k, c_k, \varepsilon_k\right) \subseteq \left\{x \in \Omega \mid L\left(x, \lambda^k, \mu^k, c_k\right) \leq v(0) + \varepsilon\right\}, \tag{2.13}$$

whenever k is sufficiently large.

Proof. For any given ε, it follows from (2.7) and Lemma 2.4 that when k is large enough, we have

$$\frac{1}{2c_k}\sum_{i=1}^{m}\left(\frac{\lambda_i^k}{c_k}\right)^2 + \frac{1}{2c_k}\sum_{i=1}^{l}\left(\frac{\mu_i^k}{c_k}\right)^2 + \varepsilon_k \leq \varepsilon. \tag{2.14}$$

Therefore, for $\overline{x} \in S^*(\lambda^k, \mu^k, c_k, \varepsilon_k)$,

$$\begin{aligned} L\left(\overline{x}, \lambda^k, \mu^k, c_k\right) &\leq \inf\left\{L\left(x, \lambda^k, \mu^k, c_k\right) \mid x \in \Omega\right\} + \varepsilon_k \\ &\leq \inf\left\{L\left(x, \lambda^k, \mu^k, c_k\right) \mid x \in X(0)\right\} + \varepsilon_k \\ &\leq \inf\{f(x) \mid x \in X(0)\} + \frac{1}{2c_k}\sum_{i=1}^{m}\left(\frac{\lambda_i^k}{c_k}\right)^2 + \frac{1}{2c_k}\sum_{i=1}^{l}\left(\frac{\mu_i^k}{c_k}\right)^2 + \varepsilon_k \\ &\leq v(0) + \varepsilon. \end{aligned} \tag{2.15}$$

□

Lemma 2.5. *Let (λ^k, μ^k, c_k) be given as in Algorithm 2.1. For any $\varepsilon > 0$, one has*

$$\left\{x \in \Omega \mid L\left(x, \lambda^k, \mu^k, c_k\right) \leq v(0) + \varepsilon\right\} \subseteq X(\varepsilon). \tag{2.16}$$

whenever k is sufficiently large.

Proof. We prove this result by the way of contradiction. Suppose that we can find an $\varepsilon_0 > 0$ and an infinite subsequence $K \subseteq \{1, 2, \ldots\}$ such that

$$z^k \in \left\{x \in \Omega \mid L\left(x, \lambda^k, \mu^k, c_k\right) \leq v(0) + \varepsilon\right\}, \quad \forall k \in K, \tag{2.17}$$

but

$$z^k \notin X(\varepsilon_0), \quad \forall k \in K. \tag{2.18}$$

It follows from (2.17) that

$$\begin{aligned} v(0) + \varepsilon &\geq L\left(z^k, \lambda^k, \mu^k, c_k\right) \\ &= f\left(z^k\right) + \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(h_j\left(z^k\right) + \frac{\mu_j^k}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, g_i\left(z^k\right) + \frac{\lambda_i^k}{c_k}\right\}^2\right]. \end{aligned} \tag{2.19}$$

Since $z^k \notin X(\varepsilon_0)$, it needs to consider the following two cases.

Case 1. There exist an index j_0 and an infinite subsequence $K_0 \subseteq K$ such that $|h_{j_0}(z^k)| > \varepsilon_0$. It then follows from (2.19) that

$$\begin{aligned} v(0) + \varepsilon &\geq f_* + \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(h_j\left(z^k\right) + \frac{\mu_j^k}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, g_i\left(z^k\right) + \frac{\lambda_i^k}{c_k}\right\}^2\right] \\ &\geq f_* + \frac{c_k}{2}\left(h_{j_0}\left(z^k\right) + \frac{\mu_{j_0}^k}{c_k}\right)^2. \end{aligned} \tag{2.20}$$

Using Lemma 2.2 and the fact that $|h_{j_0}(z^k)| \geq \varepsilon_0$ gives us

$$\left(h_{j_0}(z^k) + \frac{\mu_{j_0}^k}{c_k}\right)^2 \geq \frac{1}{2}\varepsilon_0, \tag{2.21}$$

whenever k is sufficiently large. This, together with (2.20), yields $v(0) = +\infty$ by taking $k \in K_0$ approaching to ∞, which leads to a contradiction.

Case 2. There exist an index i_0 and an infinite subsequence $K_0 \subseteq K$ such that $g_{i_0}(z^k) > \varepsilon_0$. It follows from (2.19) that

$$\begin{aligned} v(0) + \varepsilon &\geq f_* + \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(h_j\left(z^k\right) + \frac{\mu_j^k}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, g_i\left(z^k\right) + \frac{\lambda_i^k}{c_k}\right\}^2\right] \\ &\geq f_* + \frac{c_k}{2}\left[\sum_{i \neq i_0}\max\left\{0, g_i\left(z^k\right) + \frac{\lambda_i^k}{c_k}\right\}^2 + \max\left\{0, g_{i_0}\left(z^k\right) + \frac{\lambda_{i_0}^k}{c_k}\right\}^2\right] \\ &\geq f_* + \frac{c_k}{2}\max\left\{0, g_{i_0}\left(z^k\right) + \frac{\lambda_{i_0}^k}{c_k}\right\}^2 \\ &\geq f_* + \frac{\varepsilon_0 c_k}{4}, \end{aligned} \tag{2.22}$$

where the last step is due to Lemma 2.2, since $g_{i_0}(z^k) > \varepsilon_0$ and $\lambda_{i_0}^k/c_k \to 0$. Taking limits in the above inequality yields $v(0) = +\infty$, which is a contradiction. This completes the proof. □

Lemma 2.6. *Let (λ^k, μ^k, c_k) be given as in Algorithm 2.1. For any $\varepsilon > 0$, one has*

$$\left\{x \in \Omega \mid L\left(x, \lambda^k, \mu^k, c_k\right) \leq v(0) + \varepsilon\right\} \subseteq L(\varepsilon), \quad \forall k = 1, 2, \ldots. \tag{2.23}$$

Proof. For an arbitrarily $\overline{x} \in \{x \in \Omega \mid L(x, \lambda^k, \mu^k, c_k) \le v(0) + \varepsilon\}$, we have

$$\begin{aligned} f(\overline{x}) &= L\left(\overline{x}, \lambda^k, \mu^k, c_k\right) - \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(h_j(\overline{x}) + \frac{\mu_j^k}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, g_i(\overline{x}) + \frac{\lambda_i^k}{c_k}\right\}^2\right] \\ &\le v(0) + \varepsilon. \end{aligned} \tag{2.24}$$

The proof is complete. □

With these preparation, the global convergence property of Algorithm 2.1 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 the optimal solution of (P).

Theorem 2.7. *Let $\{x^k\}$ be the iterative sequence generated by Algorithm 2.1. Then if $\{x^k\}$ is terminated in finite steps, then one gets a KKT point of (P); otherwise, every limit point of $\{x^k\}$ belongs to X^*.*

Proof. According to the construction of Algorithm 2.1, the first part is clear. It remains to prove the second part. Let $\varepsilon > 0$ be given. It follows from Lemmas 2.4–2.6 that when k is large enough, we have

$$\begin{aligned} S^*\left(\lambda^k, \mu^k, c_k, \varepsilon_k\right) &\subseteq \left\{x \in \Omega \mid L\left(x, \lambda^k, \mu^k, c_k\right) \le v(0) + \varepsilon\right\} \\ &\subseteq X(\varepsilon) \cap L(\varepsilon). \end{aligned} \tag{2.25}$$

Thus,

$$x^k \in X(\varepsilon) \cap L(\varepsilon). \tag{2.26}$$

Note that $X(\varepsilon)$ and $L(\varepsilon)$ are closed, due to the continuity of f, g_i for all $i = 1, \ldots, m$ and h_j for all $j = 1, \ldots, l$ and the closeness of Ω. Taking the limit in (2.26) yields $x^* \in X(\varepsilon) \cap L(\varepsilon)$, which further shows that $x^* \in X(0) \cap L(0)$, since $\varepsilon > 0$ is arbitrary, that is, $x^* \in X^*$. 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 2.8. *Let $\{x^k\}$ be an iterative sequence generated by Algorithm 2.1. Then,*

$$\lim_{k \to \infty} f\left(x^k\right) = v(0) \tag{2.27}$$

if and only if $v(\alpha)$ is lower semicontinuous at $\alpha = 0$ from right.

Proof. We first show the sufficiency. According to the proof of Theorem 2.7 (recall (2.26)), we know that

$$v(\varepsilon) \le f\left(x^k\right) \le v(0) + \varepsilon, \tag{2.28}$$

whenever k is sufficiently large. Since $v(\alpha)$ is lower semicontinuous at $\alpha = 0$ from right, taking the lower limitation in (2.28) yields

$$\begin{aligned} v(0) &\le \liminf_{\varepsilon \to 0^+} v(\varepsilon) \le \liminf_{k \to \infty} f\left(x^k\right) \\ &\le \limsup_{k \to \infty} f\left(x^k\right) \le v(0), \end{aligned} \tag{2.29}$$

that is,

$$\lim_{k \to \infty} f\left(x^k\right) = v(0). \tag{2.30}$$

We now show the necessity. Suppose on the contrary that v is not lower semicontinuous at zero from right, then there exist $\delta_0 > 0$ and $\varepsilon_j \to 0^+$ (as $j \to \infty$) such that

$$v(\varepsilon_j) \le v(0) - \delta_0, \quad \forall j = 1, 2, \dots, m. \tag{2.31}$$

For any given k, since $\varepsilon_j \to 0$ we can choose a subsequence j_k satisfying

$$\varepsilon_{j_k} c_k \longrightarrow 0 \quad \text{as } k \longrightarrow \infty. \tag{2.32}$$

In addition, let $z^k \in X(\varepsilon_{j_k})$ with $f(z^k) \le v(\varepsilon_{j_k}) + \delta_0/2$, which further implies $f(z^k) \le v(0) - \delta_0/2$ by (2.31). Therefore,

$$\begin{aligned} f\left(x^k\right) &= L\left(x^k, \lambda^k, \mu^k, c_k\right) - \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(h_j\left(x^k\right) + \frac{\mu_j^k}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, g_i\left(x^k\right) + \frac{\lambda_i^k}{c_k}\right\}^2\right] \\ &\le \inf_{x\in\Omega} L\left(x, \lambda^k, \mu^k, c_k\right) + \varepsilon_k \\ &\le f\left(z^k\right) + \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(h_j\left(z^k\right) + \frac{\mu_j^k}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, g_i\left(z^k\right) + \frac{\lambda_i^k}{c_k}\right\}^2\right] + \varepsilon_k \\ &\le v(0) - \frac{\delta_0}{2} + \frac{c_k}{2}\left[\sum_{j=1}^{l}\left(\varepsilon_{j_k} + \frac{\left|\mu_j^k\right|}{c_k}\right)^2 + \sum_{i=1}^{m}\max\left\{0, \varepsilon_{j_k} + \frac{\lambda_i^k}{c_k}\right\}^2\right] + \varepsilon_k, \end{aligned} \tag{2.33}$$

where the last step is due to the fact $|h_j(z^k)| \le \varepsilon_{j_k}$ and $g_i(z^k) \le \varepsilon_{j_k}$ since $z^k \in X(\varepsilon_{j_k})$. Taking limits in both sides of (2.33) and using (2.7), (2.27), and Lemma 2.2, we get

$$v(0) = \lim_{k\to\infty} f\left(x^k\right) \le v(0) - \frac{\delta_0}{2}, \tag{2.34}$$

which leads to a contradiction. The proof is complete. □

Note that in many practical cases, the set Ω typically stands for a more simple constraint, for example, a box or a bounded polytope [7]. Hence, we conclude this paper by considering the case of Ω is a bounded, closed, and convex subset of $\mathbb{R}^n$. In this case, the global optimal solution of the augmented Lagrangian relaxation problem always exists. Hence, we choose $\varepsilon_0 = 0$ in Step 1 of Algorithm 2.1, which in turn implies that $\varepsilon_k = 0$ for all k by (2.7). First, however, we need to extend the definition of degenerate from inequality constraint as in [10] to both inequality and equality constraints.

Definition 2.9. A point $x^* \in X$ is said to be degenerate if there exists $\lambda^* \in \mathbb{R}^m_+$ and $\mu^* \in \mathbb{R}^l$ such that

$$\sum_{i\in I(x^*)} \lambda_i^* + \sum_{j=1}^{l} \left|\mu_j^*\right| > 0, \qquad \mathcal{P}_\Omega\left[x^* - \sum_{j=1}^{l} \mu_j^* \nabla h_j(x^*) - \sum_{i\in I(x^*)} \lambda_i^* \nabla g_i(x^*)\right] = x^*, \tag{2.35}$$

where $\mathcal{P}_\Omega(x)$ denotes the projection of x onto Ω and $I(x^*) = \{i \mid g_i(x^*) = 0,\ i = 1,\dots,m\}$.

Theorem 2.10. *Suppose that Ω is a bounded, closed, and convex set of $\mathbb{R}^n$. Let $\varepsilon_0 = 0$ and $\{x^k\}$ be the iterative sequence generated by Algorithm 2.1. Then, every accumulation point of $\{x^k\}$, say x^*, is either a degenerate or a KKT point of (P).*

Proof. Noting that $\varepsilon_k = 0$ for all k by $\varepsilon_0 = 0$ and (2.7), then $\{x^k\}$ is a global optimal solution of $L(x, \lambda^k, \mu^k, c_k)$ by Step 3 in Algorithm 2.1. Applying the well-known optimality condition of optimization problem to the augmented Lagrangian relaxation problem $(L_{\lambda,\mu,c})$ yields

$$-\nabla_x L\left(x, \lambda^k, \mu^k, c_k\right) \in \mathcal{N}_\Omega\left(x^k\right), \tag{2.36}$$

where $\mathcal{N}_\Omega(x^k)$ is the normal cone of Ω at x^k. This together with (2.5) and (2.6) means that

$$\mathcal{P}_\Omega\left[x^k - \nabla f\left(x^k\right) - \sum_{j=1}^{l} \mu_j^{k+1} \nabla h_j\left(x^k\right) - \sum_{i=1}^{m} \lambda_i^{k+1} \nabla g_i\left(x^k\right)\right] = x^k, \tag{2.37}$$

where we have used the basic property of normal cone of convex set. Let $\mathcal{K}$ be an infinite subsequence in $\{1, 2, \dots\}$ such that $\{x^k\}_{\mathcal{K}} \to x^* \in \Omega$. Consider now the following two cases.

Case 1. Either $\{\lambda^{k+1}\}_{\mathcal{K}}$ or $\{\mu^{k+1}\}_{\mathcal{K}}$ is unbounded. In this case, we must have

$$T^k := \sum_{i=1}^{m}\lambda_i^{k+1} + \sum_{j=1}^{l}\left|\mu_j^{k+1}\right| \longrightarrow \infty, \quad k \longrightarrow \infty,\ k \in \mathcal{K}. \tag{2.38}$$

Since $0 \le \lambda_i^{k+1}/T^k \le 1$ and $0 \le \mu_i^{k+1}/T^k \le 1$ are bounded, we can assume by passing a subsequence if necessary that

$$\frac{\lambda_i^{k+1}}{T^k} \longrightarrow \lambda_i^*, \qquad \frac{\mu_j^{k+1}}{T^k} \longrightarrow \mu_j^*, \quad i = 1,\dots,m, \quad j = 1,\dots,l. \tag{2.39}$$

Clearly, λ_i^* and μ_j^* are not all zeros. On the other hand, since $\mathcal{N}_\Omega(x^*)$ is cone, then it follows from (2.36) that

$$-\frac{1}{T^k}\nabla_x L\left(x, \lambda^k, \mu^k, c_k\right) \in \mathcal{N}_\Omega\left(x^k\right), \tag{2.40}$$

from which and using the basic property of normal cone of convex set, we further have

$$\mathcal{P}_\Omega\left[x^k - \frac{1}{T^k}\left(\nabla f\left(x^k\right) + \sum_{j=1}^{l}\mu_j^{k+1}\nabla h_j\left(x^k\right) + \sum_{i=1}^{m}\lambda_i^{k+1}\nabla g_i\left(x^k\right)\right)\right] = x^k. \tag{2.41}$$

Since $x^k \to x^*$ and $T^k \to \infty$ as $k \in \mathcal{K} \to \infty$, we obtain from (2.39) and (2.41)

$$\mathcal{P}_\Omega\left[x^* - \sum_{j=1}^{l}\mu_j^*\nabla h_j(x^*) - \sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)\right] - x^* = 0, \tag{2.42}$$

where we have used the continuity of the projection operator.

If $i \notin I(x^*)$, then $g_i(x^*) < 0$. Since $c_k \to \infty$, we have $c_k g_i(x^k) \to -\infty$ as $k \in \mathcal{K} \to \infty$. Using (2.5) and Lemma 2.2, we obtain

$$\lim_{k\to\infty, k\in\mathcal{K}}\lambda_i^{k+1} = \lim_{k\to\infty, k\in\mathcal{K}}\max\left\{0, g_i\left(x^k\right) + \frac{\lambda_i^k}{c_k}\right\}c_k = 0, \quad \forall i \notin I(x^*), \tag{2.43}$$

which, together with (2.39), implies that $\lambda_i^* = 0$ for all $i \notin I(x^*)$. Therefore, we obtain from (2.42) that

$$\mathcal{P}_\Omega\left[x^* - \sum_{j=1}^{l}\mu_j^*\nabla h_j(x^*) - \sum_{i\in I(x^*)}\lambda_i^*\nabla g_i(x^*)\right] - x^* = 0. \tag{2.44}$$

So x^* is degenerate.

Case 2. Both $\{\lambda^{k+1}\}_{\mathcal{K}}$ and $\{\mu^{k+1}\}_{\mathcal{K}}$ are bounded. In this case, we can assume without loss of generality that

$$\lim_{k\to\infty,k\in\mathcal{K}} \lambda_i^k = \lambda_i^* \geq 0, \qquad \lim_{k\to\infty,k\in\mathcal{K}} \mu_i^k = \mu_i^*. \tag{2.45}$$

Taking limits in (2.37) gives rise to

$$\mathcal{P}_\Omega\left[x^* - \nabla f(x^*) - \sum_{j=1}^{l}\mu_j^*\nabla h_j(x^*) - \sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*)\right] = x^*, \tag{2.46}$$

which is equivalent to

$$-\nabla f(x^*) - \sum_{j=1}^{l}\mu_j^*\nabla h_j(x^*) - \sum_{i=1}^{m}\lambda_i^*\nabla g_i(x^*) \in \mathcal{N}_\Omega(x^*). \tag{2.47}$$

We claim that x^* is a feasible point. In fact, if $g_i(x^*) > 0$ for some i, then $c_k g_i(x^k) \to \infty$ as $k \in \mathcal{K}_1 \to \infty$. From (2.5), we must have $\lambda_i^{k+1} \to \infty$, contradicting the boundedness of $\{\lambda_i^{k+1}\}_{k\in\mathcal{K}}$. Note that (2.6) can be rewritten as

$$\frac{\mu_j^{k+1}}{c_k} = \frac{\mu_j^k}{c_k} + h_j\left(x^k\right). \tag{2.48}$$

Taking limits in both sides and using the boundedness of $\{\mu_i^{k+1}\}_{k\in\mathcal{K}}$, we obtain that $h_j(x^*) = 0$ for all $j = 1, 2, \ldots, l$. Thus, x^* is a feasible solution of (P) as claimed.

If $i \notin I(x^*)$, that is, $g_i(x^*) < 0$, then following almost the same argument as in Case 1, we can show that $\lambda_i^* = 0$ (cf. (2.43)). Therefore,

$$g_i(x^*) \leq 0, \quad \lambda_i^* g_i(x^*) = 0, \quad i = 1, \ldots, m; \quad h_j(x^*) = 0, \quad j = 1, 2, \ldots, l. \tag{2.49}$$

This together with (2.47) implies that x^* is a KKT point of (P) and λ^*, μ^* are the corresponding Lagrangian multipliers. □

3. Numerical Reports

To give some insight into the behavior of our proposed algorithm presented in this paper, we solve the following nonlinar programming problems. The test was done at a PC of Pentium 4 with 2.8 GHz CPU and 1.99 GB memory, and the computer codes were written in MATLAB 7.0. Numerical results are reported in Tables 1–4, where k is the number of iterations, c_k is the penalty parameter, x^k is iterative point found by the algorithm, and $f(x^k)$ is the objective value.

Table 1: Result of Example 3.1.

k	ϵ^k	c_k	x^k	$f(x^k)$
1	1.0000	6.2500	(0.2424, 0.2424, 0.4661)	0.3348
3	0.1667	40.0541	(0.2299, 0.2299, 0.4438)	0.3027

Table 2: Result of Example 3.2.

k	ϵ^k	c_k	x^k	$f(x^k)$
1	1	6.2500	(0.1643, 0.1643, 1.0000)	0.1581
3	1.667	42.6030	(0.1350, 0.1350, 1.0000)	0.1076

Example 3.1 (see [14]). It holds that

$$\begin{aligned} \min \quad & 0.5(x_1+x_2)^2+50(x_2-x_1)^2+x_3^2+|x_3-\sin(x_1+x_2)| \\ \text{s.t.} \quad & (x_1-1)^2+(x_2-1)^2+(x_3-1)^2-1.5\le 0. \end{aligned} \tag{3.1}$$

Example 3.2 (see [14]). Consider

$$\begin{aligned} \min \quad & 0.5(x_1+x_2)^2+50(x_2-x_1)^2+\sin^2(x_1+x_2) \\ \text{s.t.} \quad & (x_1-1)^2+(x_2-1)^2+(x_3-1)^2-1.5\le 0. \end{aligned} \tag{3.2}$$

Example 3.3. It holds that

$$\begin{aligned} \min \quad & 1000-x_1^2-2x_2^2-x_3^2-x_1x_2-x_1x_3 \\ \text{s.t.} \quad & x_1^2+2x_2^2+x_3^2-25=0. \end{aligned} \tag{3.3}$$

Example 3.4. Consider

$$\begin{aligned} \min \quad & 1000-x_1^2-2x_2^2-x_3^2-x_1x_2-x_1x_3 \\ \text{s.t.} \quad & 8x_1+14x_2+7x_3-56=0 \\ & -x_i\le 0, \quad i=1,2,3. \end{aligned} \tag{3.4}$$

4. Conclusions

Augmented Lagrangian methods are useful tools for solving many practical nonconvex optimization problems. In this paper, new convergence property of proximal augmented Lagrangian algorithm is established without requiring the boundedness of multiplier sequences. It is proved that 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

Table 3: Result of Example 3.3.

k	ϵ^k	c_k	x^k	$f(x^k)$
1	1	576	(3.1094, 3.5073, −1.4777)	957.2349
2	0.5000	1.9679e + 008	(2.9094, 3.606, −1.8787)	956.9680

Table 4: Result of Example 3.4.

k	ϵ^k	c_k	x^k	$f(x^k)$
1	1	3025	(0.7101, 2.8042, 1.5490)	978.2781
2	0.5000	8.4469e + 007	(0.7208, 2.8092, 1.5579)	978.1236

optimal solution. 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. Moreover, under suitable assumptions, we show that every accumulation point of the iterative sequence generated by the algorithm is either a degenerate or is a KKT point of the primal problem. As our future work, one of the interesting and important topics is whether these nice properties could be extended to more general cone programming, for example, nonlinear semidefinite programming or second-order cone programming.

Acknowledgments

The authors would like to thank the referees for their valuable comments, which greatly improved the presentation of the paper. Research of the first author was partly supported by the National Natural Science Foundation of China (11101248, 11026047) and Shandong Province Natural Science Foundation (ZR2010AQ026). Research of the second author was partly supported by the National Natural Science Foundation of China (11171247). Research of the forth author was partly supported by the National Natural Science Foundation of China (10971118).

References

[1] M. R. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, pp. 303–320, 1969.

[2] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed., Academic Press, New York, NY, USA, 1969.

[3] R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming," *SIAM Journal on Control and Optimization*, vol. 12, pp. 268–285, 1974, Collection of articles dedicated to the memory of Lucien W. Neustad.

[4] 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.

[5] 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.

[6] E. G. Birgin and J. M. Martnez, "Augmented Lagrangian method with nonmonotone penalty parameters for constrained optimization," *Computational Optimization and Applications*. In press.

[7] E. G. Birgin, D. Fernandez, and J. M. Martnez, "On the boundedness of penalty parameters in an Augmented Lagrangian method with lower level constraints," *Optimization Methods and Software*. In press.

[8] A. R. Conn, N. Gould, A. Sartenaer, and P. 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.

[9] E. G. Birgin, R. A. Castillo, and J. M. Martínez, "Numerical comparison of augmented Lagrangian algorithms for nonconvex problems," *Computational Optimization and Applications*, vol. 31, no. 1, pp. 31–55, 2005.

[10] 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.

[11] X. X. Huang and X. Q. Yang, "A unified augmented Lagrangian approach to duality and exact penalization," *Mathematics of Operations Research*, vol. 28, no. 3, pp. 524–532, 2003.

[12] 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.

[13] 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.

[14] R. N. Gasimov, "Augmented Lagrangian duality and nondifferentiable optimization methods in nonconvex programming," *Journal of Global Optimization*, vol. 24, no. 2, pp. 187–203, 2002.