ON THE COMPLEXITY OF AN AUGMENTED LAGRANGIAN METHOD FOR NONCONVEX OPTIMIZATION

G.N. GRAPIGLIA* AND Y. YUAN†

Abstract. In this paper we study the worst-case complexity of an inexact Augmented Lagrangian method for nonconvex inequality-constrained problems. Assuming that the penalty parameters are bounded, we prove a complexity bound of $O(|\log(\epsilon)|)$ iterations for the referred algorithm generate an $\epsilon$-approximate KKT point, for $\epsilon \in (0,1)$. When the penalty parameters are unbounded, we prove an iteration complexity bound of $O(\epsilon^{-(p+1)/p})$ and $O\left(\epsilon^{-\left(\frac{4}{p}+\frac{1}{p}\right)}\right)$, respectively, when suitable $p$-order methods ($p \geq 2$) are used to approximately solve the unconstrained subproblems at each iteration of our Augmented Lagrangian scheme.

Key words. nonlinear programming, augmented lagrangian methods, tensor methods, worst-case complexity.

1. Introduction. In this paper we consider the inequality-constrained optimization problem

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad f(x), \\
\text{s.t.} & \quad c_i(x) \geq 0, \quad i = 1, \ldots, m,
\end{align*}
\]

where $f, c_i : \mathbb{R}^n \to \mathbb{R}$ ($i = 1, \ldots, m$) are continuously differentiable functions, possibly nonconvex. Augmented Lagrangian Methods are among the most efficient schemes for nonconvex constrained optimization problems (see [16, 19, 1, 2]). The theoretical analysis of these methods usually focus on global convergence properties. More specifically, in the case of (1.1)-(1.2), for any starting pair $(x_0, \lambda^{(0)}) \in \mathbb{R}^n \times \mathbb{R}_+^m$, one tries to show that the corresponding sequence $\{(x_k, \lambda^{(k)})\}_{k \geq 0}$ generated by the Augmented Lagrangian method possess the following asymptotic property:

GC. Given $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that

\[
\|\nabla f(x_{k_0}) - \sum_{i=1}^m \lambda_i^{(k_0)} \nabla c_i(x_{k_0})\| \leq \epsilon \quad \text{and} \quad \|c^{(-)}(x_{k_0})\| \leq \epsilon,
\]

where $c^{(-)}_i(x) = \min\{c_i(x), 0\}$ for $i = 1, \ldots, m$.

In this paper we obtain worst-case complexity bounds for the first $k_0$ such that (1.3) holds, when $\{(x_k, \lambda^{(k)})\}_{k \geq 0}$ is generated by a certain Augmented Lagrangian method that allows inexact solutions of its subproblems. Assuming that the penalty parameters are bounded, we prove a complexity bound of $O(|\log(\epsilon)|)$ iterations for the referred algorithm generate an $\epsilon$-approximate KKT point, for $\epsilon \in (0,1)$. When

---

*Departamento de Matemática, Universidade Federal do Paraná, Centro Politécnico, Cx. postal 19.081, 81531-980, Curitiba, Paraná, Brazil (grapiglia@ufpr.br). This author was partially supported by the National Council for Scientific and Technological Development - Brazil (grant 406269/2016-5).
†State Key Laboratory of Scientific/Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Zhong Guan Cun Donglu 55, Beijing 100190, People’s Republic of China (yyx@lsec.cc.ac.cn). This author was partially supported by NSFC (grant 11688101)
the penalty parameters are unbounded, we prove an iteration complexity bound of $O\left(\epsilon^{-2/(\alpha-1)}\right)$, where $\alpha > 1$ controls the rate of increase of the penalty parameters. In light of these results, for linearly constrained problems we are able to obtain evaluation complexity bounds of $O\left((\log(\epsilon))^{2p} \epsilon^{-(p+1)/p}\right)$ and $O\left(\epsilon^{-(1/p-1)}\right)$, respectively, when suitable $p$-order methods ($p \geq 2$) are used to approximately solve the unconstrained subproblems at each iteration of our Augmented Lagrangian scheme.

So far, most of the complexity results for nonconvex constrained optimization were obtained for algorithms specially designed to deal with worst-case scenarios, resulting many times in short-step schemes (see, e.g., \cite{10, 3, 12}). We believe that this work is a step towards complexity estimates for practical Augmented Lagrangian methods, in the context of nonconvex optimization.

The paper is organized as follows. In section 2, we describe the algorithm and present an iteration complexity analysis. In section 3, we obtain worst-case evaluation complexity bounds for linearly constrained problems. Finally, in section 4 we compare our results with the existing literature and discuss extensions to quadratically constrained problems.

2. Algorithm and Iteration Complexity Analysis. The Lagrangian function $L(x, \lambda)$ and the Augmented Lagrangian function $P(x, \lambda, \sigma)$ associated to (1.1)-(1.2) are given respectively by:

\begin{equation}
L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x),
\end{equation}

and

\begin{equation}
P(x, \lambda, \sigma) = f(x) + \sum_{i=1}^{m} \left\{ \begin{array}{ll}
-\lambda_i c_i(x) + \frac{1}{2} \sigma c_i(x)^2 & \text{if } c_i(x) < \frac{\lambda_i}{\sigma}, \\
-\frac{1}{2} \lambda_i^2 / \sigma, & \text{otherwise,}
\end{array} \right.
\end{equation}

\begin{equation}
= f(x) + \frac{\sigma}{2} \sum_{i=1}^{m} \left[ \left( c_i(x) - \frac{\lambda_i}{\sigma} \right)^2 - \left( \frac{\lambda_i}{\sigma} \right)^2 \right],
\end{equation}

where $\lambda_i$ ($i = 1, \ldots, m$) are Lagrange multipliers, $\sigma$ is the penalty parameter, and $(\tau)^- = \min \{0, \tau\}$.

In what follows, $c_i^-(x) = (c_i(x))^-$ for $i = 1, \ldots, m$. Let us consider the following Augmented Lagrangian Method. I is a slight variation of Algorithm 10.4.1 in \cite{20}, with a different rule for updating the penalty parameter. Moreover, we allow inexact solutions of the subproblems.
Algorithm 1. Augmented Lagrangian Method

**Step 0.** Given a feasible point \( x_0 \in \mathbb{R}^n \) of (1.2), \( \lambda^{(0)} \in \mathbb{R}_+^m \), \( \alpha > 1 \), \( \sigma^{(0)} > 0 \) and \( \gamma, \epsilon \in (0,1) \), set \( k := 0 \).

**Step 1.** Find an approximate solution \( x_{k+1} \) to

\[
\min_{x \in \mathbb{R}^n} P(x, \lambda^{(k)}, \sigma^{(k)}),
\]

such that

\[
P(x_{k+1}, \lambda^{(k)}, \sigma^{(k)}) \leq \min \{ P(x_k, \lambda^{(k)}, \sigma^{(k)}), P(x_0, \lambda^{(k)}, \sigma^{(k)}) \},
\]

and

\[
\|\nabla_x P(x_{k+1}, \lambda^{(k)}, \sigma^{(k)})\|_2 \leq \epsilon.
\]

**Step 2.** Set

\[
\sigma^{(k+1)} = \begin{cases} 
\sigma^{(k)}, & \text{if } \|c(\lambda^{(k)} - \sigma^{(k)})(x_{k+1})\|_2 \leq \gamma\|c(\lambda^{(k)})(x_k)\|_2, \\
\max \{ (k+1)^\alpha, \sigma^{(k)} \}, & \text{otherwise},
\end{cases}
\]

and

\[
\lambda^{(k+1)}_i = \max \left\{ \lambda^{(k)}_i - \sigma^{(k)}c_i(x_{k+1}), 0 \right\} \text{ for } i = 1, \ldots, m.
\]

**Step 3.** Set \( k := k + 1 \) and go back to Step 1.

In this section, our goal is to establish iteration complexity bounds for Algorithm 1, that is, upper bounds on the number of iterations necessary to generate a pair \( (x_k, \lambda^{(k)}) \) such that

\[
\|\nabla_x L(x_k, \lambda^{(k)})\|_2 \leq \epsilon \quad \text{and} \quad \|c(\lambda^{(k)})(x_k)\|_2 \leq \epsilon,
\]

for a given \( \epsilon \in (0,1) \). For that, we need three auxiliary results. The first one states that the augmented Lagrangian function is bounded from above by the objective function on points of the feasible set.

**Lemma 2.1.** Let \( \sigma > 0 \), \( \lambda \in \mathbb{R}_+^m \) and \( \bar{x} \in \mathbb{R}^n \) be a feasible point of (1.2). Then,

\[
P(\bar{x}, \lambda, \sigma) \leq f(\bar{x}).
\]

**Proof.** By assumptions, for \( i = 1, \ldots, m \), we have \( c_i(\bar{x}) \geq 0 \) and \( \frac{\lambda_i}{\sigma} \geq 0 \). Thus,

\[
c_i(\bar{x}) - \frac{\lambda_i}{\sigma} \geq \frac{\lambda_i}{\sigma}.
\]

Consequently,

\[
0 \geq \left( c_i(\bar{x}) - \frac{\lambda_i}{\sigma} \right)_- \geq \frac{\lambda_i}{\sigma},
\]

which shows that

\[
\left( c_i(\bar{x}) - \frac{\lambda_i}{\sigma} \right)_-^2 \leq \left( \frac{\lambda_i}{\sigma} \right)^2, \quad i = 1, \ldots, m.
\]
Then, (2.10) follows from the above inequality and (2.3). □

The next lemma provides an upper bound for the constraint violation on infeasible points.

**Lemma 2.2.** Let $\sigma > 0$, $\lambda \in \mathbb{R}^m_+$ and $\bar{x} \in \mathbb{R}^n$ be a feasible point of (1.2). If $x^+$ is an infeasible point of (1.2) such that

$$P(x^+, \lambda, \sigma) \leq P(\bar{x}, \lambda, \sigma),$$

then

$$\frac{1}{2} \left( \sum_{i=1}^{m} \frac{\lambda_i^2}{\sigma} \right) + f(\bar{x}) - f(x^+) \geq \sigma \frac{1}{2} \|c(-)(x^+)^2\|_2.$$

**Proof.** From (2.11), Lemma 2.1 and (2.3), it follows that

$$0 \leq P(\bar{x}, \lambda, \sigma) - P(x^+, \lambda, \sigma)$$

$$\leq f(\bar{x}) - f(x^+) - \sigma \frac{1}{2} \sum_{i=1}^{m} \left[ (c_i(x^+) - \frac{\lambda_i}{\sigma})^2 - \left( \frac{\lambda_i}{\sigma} \right)^2 \right]$$

$$= f(\bar{x}) - f(x^+) - \sigma \frac{1}{2} \sum_{i=1}^{m} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2 + \sigma \frac{1}{2} \sum_{i=1}^{m} \left( \frac{\lambda_i}{\sigma} \right)^2.$$

Therefore,

$$\frac{1}{2} \left( \sum_{i=1}^{m} \frac{\lambda_i^2}{\sigma} \right) + f(\bar{x}) - f(x^+) \geq \sigma \frac{1}{2} \sum_{i=1}^{m} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2.$$

Let $J = \{j \in \{1, \ldots, m\} | c_j(x^+) < 0\}$. Since $x^+$ is an infeasible point to (1.2), we have $J \neq \emptyset$. Thus,

$$\frac{1}{2} \left( \sum_{i=1}^{m} \frac{\lambda_i^2}{\sigma} \right) + f(\bar{x}) - f(x^+) \geq \sigma \frac{1}{2} \sum_{i=1}^{m} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2 \geq \sigma \frac{1}{2} \sum_{i \in J} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2 \geq \sigma \frac{1}{2} \sum_{i \in J} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2 \geq \sigma \frac{1}{2} \sum_{i \in J} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2.$$

For $i \in J$, we have

$$c_i(x^+) - \frac{\lambda_i}{\sigma} \leq c_i(x^+) < 0,$$

which gives

$$\sum_{i \in J} \left( c_i(x^+) - \frac{\lambda_i}{\sigma} \right)^2 \geq \sum_{i \in J} c_i(x^+)^2.$$

Thus, combining (2.13) and (2.14), we get

$$\frac{1}{2} \left( \sum_{i=1}^{m} \frac{\lambda_i^2}{\sigma} \right) + f(\bar{x}) - f(x^+) \geq \sigma \frac{1}{2} \sum_{i=1}^{m} c_i(x^+)^2 = \sigma \frac{1}{2} \sum_{i=1}^{m} \min \{c_i(x^+), 0\}^2$$

$$= \sigma \frac{1}{2} \sum_{i=1}^{m} c_i^-(x^+)^2$$

$$= \sigma \frac{1}{2} \|c^-(x^+)^2\|_2.$$
that is, (2.12) holds. □

The following lemma specializes the upper bound from Lemma 2.2 to points generated by Algorithm 1.

Lemma 2.3. Let the sequence \( \{(x_k, \lambda^{(k)}, \sigma^{(k)})\}_{k \geq 0} \) be generated by Algorithm 1. If \( f(x) \) is bounded from below by \( f_{low} \), then

\[
\begin{align*}
(2.15) & \quad k \left[ \mu^{(0)}_2 + 4(f(x_0) - f_{low}) \right] \geq \sigma^{(k)} \|c(-)(x_{k+1})\|_2^2, \quad \forall k \geq 1,
\end{align*}
\]

where

\[
(2.16) \quad \mu^{(k)}_i = \frac{\lambda^{(k)}_i}{\sqrt{\sigma^{(k)}}}, \quad i = 1, \ldots, m.
\]

Proof. Since \( x_0 \) is a feasible point of (1.2), it follows from (2.5), Lemma 2.2, (2.16) and the bound \( f(x_{k+1}) \geq f_{low} \) that

\[
(2.17) \quad \frac{1}{2} \mu^{(k)}_2 + f(x_0) - f_{low} \geq \frac{\sigma^{(k)}}{2} \|c(-)(x_{k+1})\|_2^2.
\]

Now, let us compute an upper bound for \( \|\mu^{(k)}_2\|_2 \). From (2.16), (2.7), (2.8), (2.2), Lemma 2.1 and (2.5), we have

\[
\begin{align*}
\|\mu^{(k+1)}_2\|_2^2 &= \sum_{i=1}^m \frac{[\lambda^{(k+1)}_i]^2}{\sigma^{(k+1)}} \\
&\leq \sum_{i=1}^m \frac{[\lambda^{(k)}_i]^2}{\sigma^{(k)}} \\
&= \sum_{i=1}^m \frac{[\lambda^{(k)}_i]^2}{\sigma^{(k)}} + \sum_{i=1}^m \left( \frac{\max \{\lambda^{(k)}_i - \sigma^{(k)} c_i(x_{k+1})\}^2}{\sigma^{(k)}} - [\lambda^{(k)}_i]^2 \right) \\
&= \|\mu^{(k)}_2\|_2^2 + 2 \sum_{i=1}^m \left( \frac{\lambda^{(k)}_i c_i(x_{k+1}) + \frac{1}{2} \sigma^{(k)} c_i(x_{k+1})^2}{\sigma^{(k)}} \right), \quad \text{if } c_i(x_{k+1}) < \frac{\lambda^{(k)}_i}{\sigma^{(k)}} \\
&\leq \|\mu^{(k)}_2\|_2^2 + 2 \left[ P(x_{k+1}, \lambda^{(k)}, \sigma^{(k)}) - f(x_{k+1}) \right] \\
&\leq \|\mu^{(k)}_2\|_2^2 + 2 \left[ P(x_{k+1}, \lambda^{(k)}, \sigma^{(k)}) - f(x_{k+1}) \right] - 2 \left[ P(x_0, \lambda^{(k)}, \sigma^{(k)}) - f(x_0) \right] \\
&= \|\mu^{(k)}_2\|_2^2 + 2 \left[ P(x_{k+1}, \lambda^{(k)}, \sigma^{(k)}) - P(x_0, \lambda^{(k)}, \sigma^{(k)}) \right] + 2 \left[ f(x_0) - f(x_{k+1}) \right] \\
&\leq \|\mu^{(k)}_2\|_2^2 + 2 \left[ f(x_0) - f_{low} \right].
\end{align*}
\]

Therefore, by induction,

\[
(2.18) \quad \|\mu^{(k)}_2\|_2^2 \leq \|\mu^{(0)}_2\|_2^2 + 2 (f(x_0) - f_{low}) k
\]

Finally, combining (2.17) and (2.18) we get (2.15). □

Now, we will analyse the iteration-complexity of Algorithm 1 considering separately the following cases:

(i) \( \{\sigma^{(k)}\}_{k \geq 0} \) is bounded from above by \( \sigma_{max} \).
(ii) \( \lim_{k \to +\infty} \sigma^{(k)} = +\infty \).
First, note that, by (2.1), (2.8), (2.2) and (2.6), we have

\[
\|\nabla_x L(x_k, \lambda^{(k)})\|_2 = \left\| \nabla f(x_k) - \sum_{i=1}^{m} \lambda_i^{(k)} \nabla c_i(x_k) \right\|_2 \\
= \left\| \nabla f(x_k) - \sum_{i=1}^{m} \max \left\{ \lambda_i^{(k-1)} - \sigma^{(k-1)} c_i(x_k), 0 \right\} \nabla c_i(x_k) \right\|_2 \\
= \left\| \nabla_x P(x_k, \lambda^{(k-1)}, \sigma^{(k-1)}) \right\|_2 \\
\leq \epsilon,
\]

for all \( k \geq 1 \). Thus, to bound the number of iterations necessary to ensure (2.9), we only need to bound the number of iterations in which we have

\[
\|c^{(-)}(x_k)\|_2 > \epsilon.
\]

The theorem below gives an upper bound of \( O(\log(\epsilon^{-1})) \) iterations in case (i).

**Theorem 2.4.** Let the sequence \( \{x_k\}_{k \geq 0} \) be generated by Algorithm 1 such that

\[
\|c^{(-)}(x_k)\|_2 > \epsilon, \quad \text{for } k = 1, \ldots, T.
\]

If \( f(x) \) is bounded from below by \( f_{\text{low}} \) and \( \{\sigma^{(k)}\}_{k \geq 0} \) is bounded from above by \( \sigma_{\text{max}} \), then

\[
T \leq \frac{1}{\sigma_{\text{max}}} + 2 + \frac{\frac{1}{2} \log \left( \|\mu^{(0)}\|_2^2 + 4 [f(0) - f_{\text{low}}] \right) + \log(\epsilon^{-1})}{\log(\gamma^{-1})}
\]

where \( \mu^{(0)} \) is defined by (2.16) and \( \alpha \) and \( \gamma \) are the parameters in (2.7).

**Proof.** Consider the set

\[
\mathcal{U} = \left\{ j \in \mathbb{N} | \|c^{(-)}(x_{j+1})\|_2 > \gamma \|c^{(-)}(x_j)\|_2 \right\}.
\]

Given \( k \in \mathcal{U} \), we must have \( k \leq \frac{1}{\sigma_{\text{max}}} \), since otherwise, by (2.7) we would have \( \sigma^{(k+1)} \geq (k+1)^{\alpha} > \sigma_{\text{max}} \), contradicting our assumption on the penalty parameters.

Let us denote \( \bar{k} = \max \{j \mid j \in \mathcal{U} \} \). Suppose that \( T > \bar{k} + 2 \) and define \( s = T - (\bar{k} + 2) \). Then,

\[
\{ \bar{k} + 2, \ldots, \bar{k} + 2 + s \} = \{ \bar{k} + 2, \ldots, T \} \subset \mathbb{N} - \mathcal{U},
\]

and, consequently,

\[
\epsilon < \|c^{(-)}(x_T)\|_2 = \|c^{(-)}(x_{\bar{k}+2+s})\|_2 \leq \gamma^{s} \|c^{(-)}(x_{\bar{k}+2})\|_2.
\]

Since \( \bar{k} \in \mathcal{U} \), it follows that \( \sigma^{(k+1)} \geq (\bar{k} + 1)^{\alpha} \). Thus, by Lemma 2.3 we have

\[
(\bar{k} + 1)^{\alpha} \|c^{(-)}(x_{\bar{k}+2})\|_2^2 \leq \left[ \|\mu^{(0)}\|_2^2 + 4 (f(0) - f_{\text{low}}) \right] (\bar{k} + 1)
\]

\[
\Rightarrow \|c^{(-)}(x_{\bar{k}+2})\|_2^2 \leq \left[ \|\mu^{(0)}\|_2^2 + 4 (f(0) - f_{\text{low}}) \right] \frac{(\bar{k} + 1)}{(\bar{k} + 1)^{\alpha}}
\]

\[
\leq \left[ \|\mu^{(0)}\|_2^2 + 4 (f(0) - f_{\text{low}}) \right]
\]
Thus, $T \leq (2.25)$ where $\mu (2.24)$ 

Combining (2.22) and (2.23), we obtain
\[
(2.23) = \Rightarrow \| \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \|^2 .
\]

By (2.24), this means that $\| f(x_0) - f_{low} \| \|^2$. Therefore, $\| \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \|^2 \geq \epsilon_1$.

\[
\Rightarrow s \log (\gamma^{-1}) < \frac{1}{2} \log (\| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \) + \log (\epsilon^{-1})
\]

\[
\Rightarrow s \leq \frac{1}{2} \log (\| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \) + \log (\epsilon^{-1})
\]

Therefore,
\[
T = k + 2 + s \leq \frac{1}{2} \sigma_{max} + 2 + \frac{1}{2} \log (\| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \) + \log (\epsilon^{-1}),
\]

and the proof is complete. □

The next theorem establishes an upper bound of $O \left( \epsilon^{-\frac{2}{\alpha}} \right)$ iterations in case (ii).

**THEOREM 2.5.** Let the sequence $\{x_k\}_{k \geq 0}$ be generated by Algorithm 1 such that
\[
\| c^{(-)}(x_k) \|_2 > \epsilon, \quad \text{for } k = 1, \ldots, T.
\]

If $f(x)$ is bounded from below by $f_{low}$ and $\lim_{k \to +\infty} \sigma^{(k)} = +\infty$, then
\[
T \leq 4 + \left( \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \right) \frac{1}{\epsilon} \epsilon^{-\frac{2}{\alpha}} + \frac{1}{2} \log (\| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \) + \log (\epsilon^{-1})
\]

where $\mu (0)$ is defined by (2.16) and $\alpha$ and $\gamma$ are the parameters in (2.7).

**Proof.** Again, consider the set $\mathcal{U}$ defined in (2.21). Given $k \in \mathcal{U}$, it follows from (2.7) and Lemma 2.3 that
\[
(k + 1) \left( \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \right) \geq (k + 1) \alpha \| c^{(-)}(x_{k+2}) \|_2^2
\]

\[
\Rightarrow \| c^{(-)}(x_{k+2}) \|_2 \leq \left( \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \right) \frac{1}{(k + 1) \alpha^{-1}}.
\]

Thus,
\[
k \in \mathcal{U} \text{ and } k + 1 \geq \left( \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \right) \frac{1}{\epsilon} \epsilon^{-\frac{2}{\alpha}} \Rightarrow \| c^{(-)}(x_{k+2}) \|_2 \leq \epsilon.
\]

By (2.24), this means that
\[
\hat{k} = \max \{ k \in \{1, \ldots, T - 1 \} \mid k \in \mathcal{U} \} \leq \left( \| \mu (0) \|_2^2 + 4 \| f(x_0) - f_{low} \| \right) \frac{1}{\epsilon} \epsilon^{-\frac{2}{\alpha}}.
\]
Suppose that $T - 2 > \tilde{k} + 2$ and define $s = (T - 2) - (\tilde{k} + 2)$. Then,

$$\{\tilde{k} + 2, \ldots, \tilde{k} + 2 + s\} = \{\tilde{k} + 2, \ldots, T - 2\} \subset \mathbb{N} - \mathcal{U},$$

and so

$$\epsilon < \|c^{(-)}(x_{T-2})\|_2 = \|c^{(-)}(x_{\tilde{k}+2+s})\|_2 \leq \gamma^s \|c^{(-)}(x_{\tilde{k}+2})\|_2.$$  

Since $\tilde{k} \in \mathcal{U}$, it follows that $\sigma(\tilde{k} + 1) \geq (\tilde{k} + 1)^\alpha$. Thus, using Lemma 2.3 as in the proof of Theorem 2.4, we obtain

$$\|c^{(-)}(x_{\tilde{k}+2})\|_2 \leq \left(\|\mu(0)\|_2^2 + 4[f(x_0) - f_{low}]\right)^{\frac{1}{2}}.$$  

Combining (2.26) and (2.27), we obtain

$$s \leq \frac{1}{2} \log\left(\frac{\|\mu(0)\|_2^2 + 4[f(x_0) - f_{low}]}{\log(\gamma^{-1})}\right).$$

Therefore,

$$T - 2 = \tilde{k} + 2 + s \leq \left(\|\mu(0)\|_2^2 + 4[f(x_0) - f_{low}]\right)^{\frac{1}{2}} \epsilon^{\frac{2}{\alpha - 1}} + 2 \frac{1}{2} \log\left(\frac{\|\mu(0)\|_2^2 + 4[f(x_0) - f_{low}]}{\log(\gamma^{-1})}\right) + \log(\epsilon^{-1}),$$

which gives (2.25). \(\Box\)

In summary, Theorem 2.4 means that if the sequence of penalty parameters is bounded from above, then Algorithm 1 takes at most $O(\log(\epsilon^{-1}))$ iterations to generate $(x_k, \lambda^{(k)})$ satisfying (2.9). On the other hand, if the sequence of penalty parameters is unbounded, Theorem 2.5 gives an upper bound of $O(\epsilon^{-\frac{2}{\alpha - 1}})$ iterations, where $\alpha > 1$ defines the increase of the penalty parameter. As expected (see [1], p. 104), the bigger is $\alpha$, the more aggressive is the update of the penalty parameter, and the smaller is the number of iterations needed for Algorithm 1 to find an $\epsilon$-approximate KKT point.

3. Evaluation Complexity for Linearly Constrained Problems. At each iteration of Algorithm 1, subproblem (2.4) must be approximately solved. In general, this is done by applying an iterative optimization method for unconstrained problems. On its turn, this unconstrained method requires a certain number of calls of the oracle$^1$, which is proportional to the number of inner iterations. Therefore, a full estimation of the worst-case complexity of Algorithm 1 must also take into account the iteration complexity of the auxiliary unconstrained method used to solve the subproblems. With this goal in mind, now we shall consider the following special case of (1.1)-(1.2):

\begin{align*}
\text{(3.1)} & \quad \min_{x \in \mathbb{R}^n} f(x), \\
\text{(3.2)} & \quad \text{s.t. } a_i^T x - b_i \geq 0, \ i = 1, \ldots, m,
\end{align*}

$^1$By calls of the oracle we mean the joint computation of $f$, $c_i$ ($i = 1, \ldots, m$) and their derivatives.
where $f : \mathbb{R}^n \to \mathbb{R}$ is $p$-times differentiable ($p \geq 3$), possibly nonconvex, $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for $i = 1, \ldots, m$. Given $\lambda \in \mathbb{R}^n_+$ and $\sigma > 0$, the corresponding augmented Lagrangian function to (3.1)-(3.2) is

$$P(x, \lambda, \sigma) = f(x) + \sum_{i=1}^{m} \left\{ \begin{array}{ll} -\lambda_i (a_i^T x - b_i) + \frac{1}{2} \sigma (a_i^T x - b_i)^2 \ & \text{if } a_i^T x - b_i < \frac{\lambda_i}{\sigma}, \\ -\frac{1}{2} \lambda_i^2 / \sigma, \ & \text{otherwise}, \end{array} \right.$$ 

For each $i \in \{1, \ldots, n\}$, let us denote by $a_{ij}$ the $j$th component of vector $a_i$.

**Lemma 3.1.** Given $f : \mathbb{R}^n \to \mathbb{R}$, $\lambda \in \mathbb{R}^n_+$ and $\sigma > 0$, let $P(x, \lambda, \sigma)$ be defined by (3.3). If $f$ is $p$-times differentiable with $p \geq 3$, then

$$D^p P(x, \lambda, \sigma) = D^p f(x), \ \forall x \in \mathbb{R}^n,$$

that is, the $p$th derivative of $P(\cdot, \lambda, \sigma)$ with respect to $x$ is equal to the $p$th derivative of $f(\cdot)$.

**Proof.** Indeed, from (3.3), a direct calculation shows that

$$\frac{\partial P}{\partial x_j} (x, \lambda, \sigma) = \frac{\partial f}{\partial x_j} (x) + \sum_{i=1}^{m} \left\{ \begin{array}{ll} -\lambda_i a_{ij} + \sigma (a_i^T x - b_i) a_{ij} \ & \text{if } a_i^T x - b_i < \frac{\lambda_i}{\sigma}, \\ 0, \ & \text{otherwise}, \end{array} \right.$$ 

Then,

$$\frac{\partial^2 P}{\partial x_k \partial x_j} (x, \lambda, \sigma) = \frac{\partial^2 f}{\partial x_k \partial x_j} (x) + \sum_{i=1}^{m} \left\{ \begin{array}{ll} \sigma a_{ik} a_{ij} \ & \text{if } a_i^T x - b_i < \frac{\lambda_i}{\sigma}, \\ 0, \ & \text{otherwise}, \end{array} \right.$$ 

which gives

$$\frac{\partial^3 P}{\partial x_i \partial x_k \partial x_j} (x, \lambda, \sigma) = \frac{\partial^3 f}{\partial x_i \partial x_k \partial x_j} (x).$$

Thus,

$$D^3 P(x, \lambda, \sigma) = D^3 f(x), \ \forall x \in \mathbb{R}^n,$$

and, therefore, (3.4) holds. □

From Lemma 3.1 it follows that $\nabla^2 P(\cdot, \lambda, \sigma)$ is $L_2$-Lipschitz continuous when $\|D^3 f(x)\| \leq L_2$ for all $x \in \mathbb{R}^n$. In this case, one can minimize $P(\cdot, \lambda, \sigma)$ by using a second-order method $M_2$. More specifically, let us consider the following assumptions on $f$ and $M_2$:

**A1.** There exist $L_2 > 0$ and $f_{low} \in \mathbb{R}$ such that $\|D^3 f(x)\| \leq L_2$ and $f(x) \geq f_{low}$ for all $x \in \mathbb{R}^n$.

**A2.** If $A1$ holds and $P(\cdot, \lambda, \sigma)$ in (3.3) is below bounded by $P_{low}$, method $M_2$, with starting point $\tilde{x}$, can find an $\epsilon$-approximate stationary point of $P(\cdot, \lambda, \sigma)$ in at most

$$C_{M_2} \sqrt{L_2} (P(\tilde{x}, \lambda, \sigma) - P_{low}) \epsilon^{-\frac{3}{2}}$$

iterations, where $C_{M_2}$ is a positive constant that depends only on the method $M_2$. 
An important class of unconstrained methods that satisfy A2 is the class of methods based on the cubic regularization of Newton’s method (see, e.g., [18, 8, 4, 13]). The next lemma establishes an upper bound on the number of iterations for these methods to compute approximate solutions of (2.4) satisfying (2.5)-(2.6).

**Lemma 3.2.** Let \( \{(x_j, \lambda^{(j)}, \sigma^{(j)})\}_{j=0}^k \) be generated by Algorithm 1 applied to (3.1)-(3.2), with \( k \geq 0 \). Suppose that a monotone method \( M_2 \) is applied to minimize \( P(...) \) with starting point \( (3.5) \hat{x}_{k,0} \). If \( f \) satisfies A1 and method \( M_2 \) satisfies A2, then \( M_2 \) takes at most \( O(k\epsilon^{-\frac{3}{2}}) \) iterations to generate \( \hat{x}_{k,\ell} \) such that
\[
\|\nabla_x P(\hat{x}_{k,\ell}, \lambda^{(k)}, \sigma^{(k)})\|_2 \leq \epsilon.
\]

**Proof.** By (2.3), (2.16), A1 and (2.18), we have
\[
P(x, \lambda^{(k)}, \sigma^{(k)}) \geq f(x) - \sigma^{(k)} \sum_{i=1}^m \left( \frac{\lambda^{(k)}}{\sigma^{(k)}} \right)^2
= f(x) - \frac{1}{2} \|\mu^{(k)}\|_2^2, \quad \forall x \in \mathbb{R}^n
\geq f_{low} - \frac{1}{2} \|\mu^{(0)}\|_2^2 - (f(x_0) - f_{low}) k,
\]
which gives the lower bound
\[
P_{low} = f_{low} - \frac{1}{2} \|\mu^{(0)}\|_2^2 - (f(x_0) - f_{low}) k,
\]
for \( P(x, \lambda^{(k)}, \sigma^{(k)}) \). On the other hand, it follows from (3.5) and Lemma 2.1 that
\[
P(\hat{x}_{k,0}, \lambda^{(k)}, \sigma^{(k)}) \leq P(x_0, \lambda^{(k)}, \sigma^{(k)}) \leq f(x_0).
\]
Thus, by A2, (3.6) and (3.7), we conclude that method \( M_2 \) takes at most
\[
C_{M_2} \sqrt{L_2} \left( f(x_0) - f_{low} + \frac{1}{2} \|\mu^{(0)}\|_2^2 + |f(x_0) - f_{low}| k \right) \epsilon^{-\frac{3}{2}}
\]
iterations to generate an \( \epsilon \)-approximate stationary point of \( P(...) \).

Now, combining Theorems 2.4 and 2.5 with Lemma 3.2, we can obtain worst-case complexity bounds for the total number of inner iterations performed in Algorithm 1 to find an \( \epsilon \)-approximate KKT point of (3.1)-(3.2).

**Theorem 3.3.** Suppose that Algorithm 1 is applied to solve (3.1)-(3.2) with \( f \) satisfying A1. Moreover, assume that at each iteration of Algorithm 1, a monotone method \( M_2 \) satisfying A2 is used to approximately solve (2.4) with starting point \( \hat{x}_{k,0} \) given in (3.5). Then, the following statements are true:

(a) If \( \left\{ \sigma^{(k)} \right\}_{k \geq 0} \) is bounded from above by \( \sigma_{max} \), then Algorithm 1 takes at most
\[
O \left( \left\lfloor \log(\epsilon) \right\rfloor^2 \epsilon^{-\frac{3}{2}} \right)
\]
ininner iterations of \( M_2 \) to generate an \( \epsilon \)-approximate KKT point of (3.1)-(3.2).
(b) If \( \lim_{k \to +\infty} \sigma^{(k)} = +\infty \), then Algorithm 1 takes at most \( O\left(\epsilon^{-\left(\frac{4 \alpha - 1}{2}\right)}\right) \) inner iterations of \( M_2 \) to generate an \( \epsilon \)-approximate KKT point of (3.1)-(3.2).

Proof. By Theorems 2.4 and 2.5, there exists an iteration number \( T \) such that
\[
\|\nabla_x L(x_{T+1}, \lambda^{(T+1)})\|_2 \leq \epsilon \quad \text{and} \quad \|c^{(\cdot)}(x_{T+1})\|_2 \leq \epsilon.
\]
By Lemma 3.2, the total number of inner iterations of \( M_2 \) performed until iteration \( T + 1 \) of Algorithm 1 is proportional to
\[
\left(\sum_{k=1}^{T+2} k\right) \epsilon^{-\frac{3}{2}} = O\left( (T + 2)^2 \epsilon^{-\frac{3}{2}} \right).
\]
Thus, (a) and (b) follow directly from the upper bounds on \( T \) given in (2.20) and (2.25), respectively.

If \( D^p f(\cdot) \) is Lipschitz continuous for \( p \geq 3 \), it follows from Lemma 3.1 that \( D^p P(\cdot, \lambda, \sigma) \) is also Lipschitz continuous. In this case, one can minimize \( P(\cdot, \lambda, \sigma) \) using a tensor method \( M_p \). More specifically, let us consider now the following assumptions on \( f \) and \( M_p \):

A1'. There exists \( f_{\text{low}} \in \mathbb{R}^n \) such that \( f(x) \geq f_{\text{low}} \) for all \( x \in \mathbb{R}^n \) and, for some \( p \geq 3 \), \( D^p f(\cdot) \) is \( L_p \)-Lipschitz continuous, i.e.,
\[
\|D^p f(x) - D^p f(y)\| \leq L_p \|x - y\|, \quad \forall x, y \in \mathbb{R}^n.
\]

A2'. If A1' holds and \( P(\cdot, \lambda, \sigma) \) in (3.3) is below bounded by \( P_{\text{low}} \), method \( M_p \), with starting point \( \tilde{x}_0 \), can find an \( \epsilon \)-approximate stationary point of \( P(\cdot, \lambda, \sigma) \) in at most
\[
C_{M_p} L_p \left( P(\tilde{x}, \lambda, \sigma) - P_{\text{low}} \right) \epsilon^{-\frac{p+1}{p}}
\]
iterations, where \( C_{M_p} \) is a positive constant that depends only on the method \( M_p \).

Recently, many tensor methods satisfying A2' have been proposed (see, e.g., [4, 11, 14]). The use of these tensor methods to approximately solve (2.4) give us complexity bounds better than the ones obtained in Theorem 3.3, as we can see below.

**Theorem 3.4.** Suppose that Algorithm 1 is applied to solve (3.1)-(3.2) with \( f \) satisfying A1'. Moreover, assume that at each iteration of Algorithm 1, a monotone method \( M_p \) satisfying A2' is used to approximately solve (2.4) with starting point \( \tilde{x}_{k,0} \) given in (3.5). Then, the following statements are true:

(a) If \( \{\sigma^{(k)}\}_{k \geq 0} \) is bounded from above by \( \sigma_{\text{max}} \), then Algorithm 1 takes at most \( O\left( \log(\epsilon)\right) \) inner iterations of \( M_p \) to generate an \( \epsilon \)-approximate KKT point of (3.1)-(3.2).

(b) If \( \lim_{k \to +\infty} \sigma^{(k)} = +\infty \), then Algorithm 1 takes at most \( O\left( \epsilon^{-\left(\frac{4 \alpha - 1}{2}\right)}\right) \) inner iterations of \( M_p \) to generate an \( \epsilon \)-approximate KKT point of (3.1)-(3.2).

Proof. As in Lemma 3.2, we conclude that, at the \( k \)th iteration, method \( M_p \) takes at most \( O\left( k^{\frac{-p+1}{p}} \right) \) iterations to generate \( \tilde{x}_{k,\ell} \) such that
\[
\|\nabla_x P(\tilde{x}_{k,\ell}, \lambda^{(k)}, \sigma^{(k)})\|_2 \leq \epsilon.
\]
Thus, the total number of inner iterations of $\mathcal{M}_p$ performed until iteration $T + 1$ is proportional to

$$\left( \sum_{k=1}^{T+2} k \right) \epsilon^{-\frac{p+1}{p}} = \mathcal{O}\left( (T+2)\epsilon^{-\frac{p+1}{p}} \right).$$

Thus, (a) and (b) follow directly from the upper bounds on $T$ given in (2.20) and (2.25).

4. Discussion.

4.1. Related Literature. [17] proposed a Primal-Dual Algorithm (Prox-PDA) for the linearly constrained problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

s.t. $a_i^T x - b_i = 0, \ i = 1, \ldots, m.$

It was shown that Prox-PDA enjoys an iteration complexity of $\mathcal{O}(\epsilon^{-1})$. Therefore, even when the penalty parameters in Algorithm 1 are unbounded, the iteration complexity bound of $\mathcal{O}(\epsilon^{-\frac{2}{p-1}})$ given by Theorem 2.5 is better than the one proved in [17], if we take $\alpha > 3$.

[3] proposed a Two-Phases Algorithm (FTarget) for the general constrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$

(4.1) s.t. $c_i(x) = 0, \ i = 1, \ldots, m_e,$

(4.2) $c_i(x) \geq 0, \ i = m_e + 1, \ldots, m.$

(4.3)

Under suitable smoothness conditions, they proved that if the subproblems of FTarget are approximately solved by a $p$-order method $\mathcal{M}_p$ satisfying A2 (if $p = 2$) or A2' (if $p \geq 3$), then FTarget can find an $\epsilon$-approximate KKT point of (4.1)-(4.3) performing at most $\mathcal{O}\left( \epsilon^{-\frac{2(p+1)}{p}} \right)$ problem’s evaluations. Thus, even when the penalty parameters in Algorithm 1 are unbounded, our evaluation complexity bound of $\mathcal{O}\left( \epsilon^{-\frac{3}{2}} \right)$ for problem (3.1)-(3.2) is better than the one proved in [3], if we take $\alpha > 1 + 4p$.

An evaluation complexity bound of $\mathcal{O}(\epsilon^{-3/2})$ was proved by [15] for a second-order Interior Trust-Region Point Algorithm designed to solve problems of the form

$$\min_{x \in \mathbb{R}^n} f(x),$$

s.t. $a_i^T x - b_i = 0, \ i = 1, \ldots, m$

$x_j \geq 0, \ j = 1, \ldots, n,$

where $f$ is not necessarily twice differentiable on the boundary of the feasible region. A similar bound of $\mathcal{O}(\epsilon^{-3/2})$ was also proved by [5] for an algorithm to solve (3.1)-(3.2) based on active-set strategies. On the other hand, [8] proved an evaluation complexity bound of $\mathcal{O}(\epsilon^{-\frac{p+1}{p}})$ for a $p$-order tensor method ($p \geq 2$) adapted to solve constrained problems of the form

$$\min_{x \in \mathbb{R}^n} f(x),$$

s.t. $x \in \mathcal{F},$
where \( \mathcal{F} \subset \mathbb{R}^n \) is closed, convex and non-empty. When the penalty parameters in Algorithm 1 are bounded, the evaluation complexity bounds of \( O \left( \frac{|\log(\epsilon)|^2 \epsilon^{-\frac{1}{p+1}}}{p} \right) \) given in Theorems 3.3 and 3.4 (for \( p \geq 2 \)) are slightly worse than the bounds mentioned above. It is worth to mention that a complexity bound of the same order can be obtained even when the penalty parameters are unbounded. For that, it is enough to replace (2.7) by the following update rule:

\[
\sigma^{(k+1)} = \begin{cases} 
\sigma^{(k)}, & \text{if } \|c^{(-)}(x_{k+1})\|_2 \leq \gamma \|c^{(-)}(x_k)\|_2, \\
\max \{4^{k+1}, \sigma^{(k)}\}, & \text{otherwise.}
\end{cases}
\]

Indeed, using (4.4) in the proof of Theorem 2.5, we can obtain \( \tilde{k} \leq O \left( \log(\epsilon^{-1}) \right) \) which gives an iteration complexity bound of \( O \left( \log(\epsilon^{-1}) \right) \). This corresponds to an evaluation complexity bound of \( O \left( \frac{|\log(\epsilon)|^2 \epsilon^{-\frac{1}{p+1}}}{p} \right) \) for problem (3.1)-(3.2) when the penalty parameters are unbounded. However, (4.4) increases the penalty parameters much more aggressively than (2.7), and can lead to a premature ill-conditioning of (2.4). Therefore, despite the improved worst-case complexity, it is unlikely that (4.4) will give an efficient algorithm in practice.

Finally, for \( p = 2 \) our evaluation complexity bound of \( O \left( \epsilon^{-\left(\frac{n+1}{2} + \frac{1}{2}\right)} \right) \) is better than the bounds of \( O(\epsilon^{-2}) \) proved for the first-order schemes proposed by [7] and by [9], if \( \alpha > 9 \). This is not surprising, since our result is obtained using a second-order method to solve the subproblems.

4.2. Nonconvex Problems with Quadratic Constraints. Consider now the following problem

\[
\begin{align*}
(4.5) & \quad \min_{x \in \mathbb{R}^n} f(x), \\
(4.6) & \quad \text{s.t. } x^T P_i x + q_i^T x + r_i \geq 0, \quad i = 1, \ldots, m
\end{align*}
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \) is four times differentiable, possibly nonconvex, \( P_i \in \mathbb{R}^{n \times n} \), \( q_i \in \mathbb{R}^n \) and \( r_i \in \mathbb{R} \), for \( i = 1, \ldots, m \). Let \( P(x, \lambda, \sigma) \) denote the augmented Lagrangian function corresponding to (4.5)-(4.6). Then, as in Lemma 3.1, one can see that \( D^4 P(x, \lambda, \sigma) = D^4 f(x) \) for all \( x \in \mathbb{R}^n \). Thus, if \( D^4 f(x) \) is bounded, it follows that \( D^3 P(x, \lambda, \sigma) \) is Lipschitz continuous, and we can apply Algorithm 1 to (4.5)-(4.6), solving (2.4) with a third-order tensor method. Specifically, if A1’ and A2’ hold for \( p = 3 \), then (as in Theorem 3.4) we conclude that Algorithm 1 applied to (4.5)-(4.6) takes at most \( O \left( \frac{|\log(\epsilon)|^2 \epsilon^{-\frac{1}{2}}}{2} \right) \) inner iterations of \( M_3 \) if \( \{\sigma^{(k)}\} \) is bounded, and \( O \left( \epsilon^{-\left(\frac{n+1}{2} + \frac{1}{2}\right)} \right) \) inner iterations of \( M_3 \) if \( \{\sigma^{(k)}\} \) is unbounded. In this case, a possible choice for \( M_3 \) satisfying A2’ is given by [6].

5. Conclusion. In this paper, we have studied the worst-case complexity of an inexact Augmented Lagrangian method for inequality-constrained optimization problems. For the case in which the penalty parameters are bounded, we established a complexity bound of \( O(|\log(\epsilon)|) \) iterations for the referred algorithm generate an \( \epsilon \)-approximate KKT point, for \( \epsilon \in (0, 1) \). For the case in which the penalty parameters are unbounded, we proved an iteration complexity bound of \( O(\epsilon^{-2/(\alpha-1)}) \), where \( \alpha > 1 \) controls the rate of increase of the penalty parameters. In the particular class of linearly constrained problems, these bounds yield to evaluation complexity bounds of \( O(|\log(\epsilon)|^2 \epsilon^{-(p+1)/p}) \) and \( O \left( \epsilon^{-\left(\frac{n+1}{p} + \frac{1}{p}\right)} \right) \), respectively, when appropriate \( p \)-order
methods \((p \geq 2)\) are used to approximately solve the unconstrained subproblems at each iteration.

A key point in the Augmented Lagrangian method considered in this work is that it requires the feasibility of the starting point \(x_0\), which may be difficult to compute for general nonconvex constraints. Moreover, the assumption that the objective function is bounded from below is also crucial in our complexity analysis. Up to now, it is not clear how these restrictions can be avoided. Another natural question is whether our analysis can be adapted in order to cover possibly less aggressive update rules for the penalty parameters, such as

\[
\sigma^{(k+1)} = \begin{cases} 
\sigma^{(k)}, & \text{if } \|c^-(x_{k+1})\|_2 \leq \gamma \|c^-(x_k)\|_2, \\
\alpha \sigma^{(k)}, & \text{otherwise.}
\end{cases}
\]

The authors are planning to address these and other interesting questions in their future research.

REFERENCES


