Adaptive inexact fast augmented Lagrangian methods for constrained convex optimization

Andrei Patrascu¹ · Ion Necoara¹ · Quoc Tran-Dinh²

Received: 12 May 2015 / Accepted: 22 February 2016 / Published online: 11 March 2016 © Springer-Verlag Berlin Heidelberg 2016

Abstract In this paper we study two inexact fast augmented Lagrangian algorithms for solving linearly constrained convex optimization problems. Our methods rely on a combination of the excessive-gap-like smoothing technique introduced in Nesterov (SIAM J Optim 16(1):235–249, 2005) and the general inexact oracle framework studied in Devolder (Math Program 146:37–75, 2014). We develop and analyze two augmented based algorithmic instances with constant and adaptive smoothness parameters, and derive a total computational complexity estimate in terms of projections on a simple primal feasible set for each algorithm. For the constant parameter algorithm we obtain the overall computational complexity of order \(O\left(\frac{1}{\epsilon^{5/4}}\right)\), while for the adaptive one we obtain \(O\left(\frac{1}{\epsilon}\right)\) total number of projections onto the primal feasible set in order to achieve an \(\epsilon\)-optimal solution for the original problem.

The research leading to these results has received funding from: UEFISCDI Romania, PN II-RU-TE, project MoCOBiDS, No. 176/01.10.2015; Sectorial Operational Programme Human Resources Development 2007–2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/134398; NAFOSTED, Vietnam.

Andrei Patrascu
andrei.patrascu@acse.pub.ro

Ion Necoara
ion.necoara@acse.pub.ro

Quoc Tran-Dinh
quoctd@email.unc.edu

1 Automatic Control and Systems Engineering Department, University Politehnica Bucharest, 060042 Bucharest, Romania

2 Department of Statistics and Operations Research, University of North Carolina at Chapel Hill (UNC), Chapel Hill, USA
Keywords Inexact augmented Lagrangian · Smoothing techniques · Primal-dual fast gradient · Excessive gap · Computational complexity

1 Introduction

The class of large-scale linearly constrained convex optimization problems provides a powerful tool to model and formulate several practical applications in modern engineering, signal and image processing, machine learning, statistics, and economical applications, see, e.g., [4,5,16]. Various prominent problems in such fields can be cast into this class of optimization problems, such as linear (distributed) model predictive control [16,20], network utility maximization [4,5], and compressed sensing [2,30]. These models on the one hand present a new opportunity to efficiently handle recent applications, but on the other hand, raise many new challenges to the existing numerical methods and computational power.

While classical optimization approaches are no longer suitable to handle modern applications, recent developments in first-order methods provides a practical tool to solve these problems much more efficiently. Among various first-order frameworks, the primal-dual first-order methods become an emerging approach to handle large-scale constrained convex problems, which has recently gained a great attention due to its low complexity-per-iteration and its flexibility for handling linear operators, complex constraints and non-smoothness. In convex optimization, primal-dual methods are rooted from the min-max and saddle point theorem, where strong duality is key to guarantee the zero duality gap between the primal and dual problems. Based on this principle, several first-order primal-dual frameworks have been introduced in the literature such as primal-dual subgradient methods, primal-dual splitting techniques, alternating direction optimization methods, Chambolle–Pock’s algorithm, Arrow–Hurwicz’s scheme, primal-dual hybrid gradient, and primal-dual smoothing algorithms [3,6–8,10,13,15,19,22].

Amongst these methods, the augmented Lagrangian (AL) framework is perhaps the most popular one, which has been extensively studied, e.g., in [2,11,12,14,20,27,28]. This method has proved its power in practice when attacking large-scale applications with complex constraints in different fields. It is well known that the AL smoothing technique is a multi-stage strategy implying successive computations of solutions for certain primal-dual subproblems. In general, these subproblems do not have closed form solutions, which makes the AL strategies inherently related to inexact first-order algorithms. In these settings, most of the complexity results regarding AL and fast AL algorithms are given under inexact first-order information [1,14,20,28]. To our best knowledge, the complexity estimates for the fast AL methods have been studied only in [20,28]. Unfortunately, in these papers only outer complexity estimates are provided, while the complexity-per-iteration is elusive. Consequently, the overall computational complexity estimate of these algorithms remains unknown. Specially, in some situations, we can choose the smoothness parameters in order to perform only one outer iteration, see, e.g., [1,14], but the overall complexity estimate is still high due to the high complexity-per-iteration (inner iterations). Trading off these quantities is a crucial question in the implementation of AL methods.
In this paper we aim at improving the overall computational complexity estimates for the fast AL methods by exploiting the inexact oracle framework developed in [9]. We combine this inexact oracle concept and the excessive gap-like fast gradient algorithms introduced in [25] to design new algorithms that achieve better overall computational complexity while preserving the same key features as in existing AL methods.

1.1 Our contributions

We develop two inexact fast augmented Lagrangian (IFAL) algorithms corresponding to a constant (or nonadaptive) and an adaptive smoothness parameter update strategy. We analyze the overall computational complexity of these algorithms, combining both the outer and inner loops. Using the inexact oracle framework [9], our approach allows us to obtain clean and intuitive complexity results and, moreover, it facilitates the derivation of the overall computational complexity of our methods.

(a) For the basic IFAL method with constant smoothness parameter, we derive \( O\left(\frac{1}{\sqrt{\epsilon}}\right) \) outer complexity estimates corresponding to simple inner accuracy updates. We also derive the overall computational complexity, which requires \( O\left(\frac{1}{\epsilon^{5/4}}\right) \) number of projections onto a primal, simple feasible set, in order to achieve an \( \epsilon \)-optimal solution in the sense of the objective residual and the feasibility violation.

(b) Then, we show that for an optimal choice of the smoothness parameter, we need to perform only one outer iteration. Based on this result, we introduce an adaptive IFAL method with variable smoothness parameters for which we can prove an overall computational complexity of order \( O\left(\frac{1}{\epsilon}\right) \) projections onto a primal, simple feasible set.

(c) We show that our adaptive variant of the inexact fast AL method is implementable, i.e., it is based on computable stopping criteria, which provides a concrete condition to terminate the algorithm. Moreover, we also compare our results with other existing overall complexity estimates for AL methods in the literature, and high-light the advantageous features of our methods.

1.2 Paper organization

The rest of this paper is organized as follows. In Sect. 2 we define our optimization model and recall some preliminary concepts related to duality and inexact oracles. In Sect. 3 we develop the inexact fast AL method with constant smoothness parameter and analyze its computational complexity. To improve this complexity, in Sect. 3 we study an adaptive parameter variant and provide its complexity estimate. Finally, in Sect. 5, we compare our results with other existing complexity results on the AL method in the literature.
1.3 Notations

We work with the Euclidean space $\mathbb{R}^n$ composed by column vectors. For any $u, v \in \mathbb{R}^n$ we denote the inner product $\langle u, v \rangle := u^T v$, and the Euclidean norm $\|u\| := \sqrt{\langle u, u \rangle}$. For any matrix $G$ we denote by $\|G\|$ its spectral norm.

2 Problem formulation and preliminaries

We consider the following linearly constrained convex optimization problem:

$$f^* = \min_{u \in \mathbb{R}^n} \left\{ f(u) : Gu + g = 0, u \in \mathcal{U} \right\}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper, closed and convex function, $\mathcal{U} \subseteq \mathbb{R}^n$ is a nonempty, closed, convex and simple set, $G \in \mathbb{R}^{m \times n}$, and $g \in \mathbb{R}^m$. Associated with (1), we can define the dual problem as:

$$d^* = \max_{x \in \mathbb{R}^m} d(x), \tag{2}$$

where $d(x) = \min_{u \in \mathcal{U}} \{ f(u) + \langle x, Gu + g \rangle \}$ is the dual function. We use $\mathcal{U}^*$ to denote the optimal solution set of (1), which is assumed to be nonempty. We also denote by $\mathcal{X}^*$ the solution set of the dual problem (2).

The goal of this paper is to develop new inexact Augmented Lagrangian methods equipped with rigorous convergence guarantees for approximating a solution of (1). Our approach requires the following blanket assumptions which are assumed to be valid throughout the paper without recalling them again.

**Assumption 1**

(i) The solution set $\mathcal{U}^*$ is nonempty. The feasible set $\mathcal{U}$ is bounded and simple, i.e. diameter $D_\mathcal{U} = \max_{u, v \in \mathcal{U}} \|u - v\|$ is finite and the projection onto $\mathcal{U}$ can be computed in a closed form or in polynomial time.

(ii) Either $\text{dom}(f) \cap \mathcal{U}$ is a polyhedral set or the following Slater condition holds:

$$\{u : Gu + g = 0\} \cap \text{ri}(\mathcal{U} \cap \text{dom}(f)) \neq \emptyset,$$

where $\text{ri}(\cdot)$ is the relative interior of a given set $\cdot$.

(iii) The objective function $f$ is smooth and its gradient is Lipschitz continuous on its domain with the Lipschitz constant $L_f > 0$, i.e.:

$$\|\nabla f(u) - \nabla f(v)\| \leq L_f \|u - v\|, \quad \forall u, v \in \text{dom}(f).$$

Under Assumption A.1, the dual solution set $\mathcal{X}^*$ of (2) is also nonempty and bounded. In addition, the strong duality holds, i.e., $f^* = d^*$.

If $f$ is strongly convex, then it is well-known that the Lagrangian dual function associated with the linear constraint $Gu + g = 0$ has Lipschitz gradient. This setting has been extensively studied in the literature, see, e.g., [16,17,28]. Thus, in the rest
of our paper we only assume that \( f \) is convex, but not necessarily strongly convex. Due to the constraint, it is clear that the dual function \( d \) defined by (2) is, in general, nonsmooth and concave, which induces the difficulties in the application of usual first-order methods to the dual problem (2). Therefore, various (dual) subgradient schemes have been developed, with the iteration complexities of order \( O(\varepsilon^{-2}) \) \cite{21,26} in order to achieve an \( \varepsilon \)-optimal solution for the original problem. Under additional mild assumptions, we aim at improving the iteration complexity required for solving linearly constrained convex optimization problems of the form (1). Our approach relies on a combination between smoothing techniques and duality theory \cite{19,25,29}.

First, we briefly present the Lagrangian duality framework as follows. The Lagrange and dual functions associated to (1) are defined as

\[
L(u, x) = f(u) + \langle x, Gu + g \rangle \quad \text{and} \quad d(x) = \min_{u \in \mathcal{U}} L(u, x).
\]

From Assumption A.1(ii), it follows that solving the convex problem (1) is equivalent to solving the dual formulation, i.e.:

\[
d^* = \max_{x \in \mathbb{R}^m} d(x) = \max_{u \in \mathcal{U}} \min_{x \in \mathbb{R}^m} L(u, x) = f^*.
\]

The goal of this paper is to find an approximate primal solution for (1) up to a given accuracy \( \epsilon > 0 \) in the following sense:

**Definition 1** Given a desired accuracy \( \epsilon > 0 \), the point \( u_\epsilon \in \mathcal{U} \) is called an \( \epsilon \)-optimal solution for the primal problem (1) if it satisfies:

\[
f(u_\epsilon) - f^* \leq \epsilon \quad \text{and} \quad \|Gu_\epsilon + g\| \leq \epsilon.
\]

This set of optimality criteria has been adopted by Rockafellar in [27] in the context of classical augmented Lagrangian methods. Moreover, it has also been used by Nesterov in [23] for analyzing his primal-dual subgradient methods. Clearly, once we have an \( \epsilon \)-feasible point, i.e., \( \|Gu_\epsilon + g\| \leq \epsilon \), then we can also obtain a lower bound on \( f(u_\epsilon) - f^* \). Indeed, we have the relation:

\[
f^* = \min_{u \in \mathcal{U}} f(u) + \langle x^*, Gu + g \rangle \leq f(u_\epsilon) + \|x^*\|\|Gu_\epsilon + g\|,
\]

and thus \( f(u_\epsilon) - f^* \geq -\|x^*\|\|Gu_\epsilon + g\| \). Moreover, from a practical point of view, it is sufficient to find an \( \epsilon \)-feasible point \( u_\epsilon \) satisfying \( f(u_\epsilon) - f^* \leq \epsilon \).

Given the max–min formulation (3), one can intuitively consider a double smoothing of the convex–concave Lagrangian function. Regarding the dual function, one of the most widely known smoothing strategies for obtaining an approximate smooth dual function with Lipschitz continuous gradient is the augmented Lagrangian (AL) smoothing \cite{1,20,27,28}. Thus, we combine the AL technique and a smooth approximation of the primal function to define

\[
\mathcal{L}_{\mu \rho}(u, x) = f(u) + \langle x, Gu + g \rangle + \frac{\rho}{2}\|Gu + g\|^2 - \frac{\mu}{2}\|x\|^2,
\]
where $\mu, \rho > 0$ are two smoothness parameters. Clearly, $\mathcal{L}_{\mu\rho}(u, x) \to \mathcal{L}(u, x)$ as $\mu, \rho$ tend to $0^+$, and $\mathcal{L}_{\mu\rho}(\cdot, x)$ has Lipschitz continuous gradient with the Lipschitz constant $L_L = L_f + \rho \|G\|^2$ for any fixed $x$. Based on this approximation of the true Lagrangian function we also define two smooth approximations of the primal and dual functions $f$ and $d$, respectively:

$$
\begin{align*}
    d_\rho(x) &= \min_{u \in \mathcal{U}} \mathcal{L}_{0\rho}(u, x) \\
    f_\mu(u) &= \max_{x \in \mathbb{R}^n} \mathcal{L}_{\mu0}(u, x)
\end{align*}
$$

Let us define the optimal solutions of the two previous optimization problems:

$$
\begin{align*}
    u_\rho(x) &= \arg\min_{u \in \mathcal{U}} \mathcal{L}_{0\rho}(u, x) \quad \text{and} \quad x_\mu(u) &= \arg\max_{x \in \mathbb{R}^n} \mathcal{L}_{\mu0}(u, x) \left(= \frac{1}{\mu} (Gu + g)\right).
\end{align*}
$$

Clearly, both functions $f_\mu$ and $d_\rho$ are smooth approximations of $f$ and $d$, respectively. In particular, we observe that the smoothed primal function $f_\mu$ has Lipschitz continuous gradient with the Lipschitz constant $L_{f_\mu} = L_f + \mu \|G\|^2$. Moreover, the smoothed dual function $d_\rho$ is concave and its gradient $\nabla d_\rho(x) = Gu_\rho(x) + g$ is Lipschitz continuous with the Lipschitz constant $L_{d_\rho} = 1/\rho$. We emphasize again that, in most practical cases, $u_\rho(x)$ cannot be computed exactly, but within a pre-specified accuracy, which leads us to the study of the inexact oracle framework introduced in [9] for analyzing our inexact first-order algorithms. Recall that a smooth function $\phi : Q \to \mathbb{R}$ is equipped with a first-order $(\delta, L)$-oracle if for any $y \in Q$ we can compute $(\phi_{\delta,L}(y), \nabla \phi_{\delta,L}(y)) \in \mathbb{R} \times \mathbb{R}^n$ such that the following bounds hold on $\phi$ (so-called inexact descent lemma) [9]:

$$
0 \leq \phi(x) - (\phi_{\delta,L}(y) + \langle \nabla \phi_{\delta,L}(y), x - y \rangle) \leq \frac{L}{2} \|x - y\|^2 + \delta, \quad \forall x, y \in Q.
$$

If $\tilde{u}_\rho(x) \in \mathcal{U}$ denotes the inexact solution of the $u$-inner subproblem satisfying

$$
0 \leq \mathcal{L}_{0\rho}(\tilde{u}_\rho(x), x) - d_\rho(x) \leq \delta, \quad (5)
$$

then using the notation $\nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(x), x) = Gu_\rho(x) + g$, we have the following important auxiliary result. A similar result can be found in [18], but for completeness we also provide the proof here.

Lemma 1 Let $\delta > 0$ and $\tilde{u}_\rho(x) \in \mathcal{U}$ satisfy (5). Then, for all $x, y$, we have

$$
0 \leq \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle - d_\rho(x) \leq L_{d_\rho} \|x - y\|^2 + 2\delta. \quad (6)
$$

Proof For the left hand side inequality of (5), we observe that:

$$
\mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle = \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), x) \\
\geq \min_{u \in \mathcal{U}} \mathcal{L}_{0\rho}(u, x) = d_\rho(x).
$$
For the right hand side inequality of (5), note that for any fixed \( u \in \mathcal{U} \) the function \( h(x) = \mathcal{L}_{0\rho}(u, x) - d_\rho(x) \) has Lipschitz gradient with the Lipschitz constant \( L_{d_\rho} \) and \( h(x) \geq 0 \) for all \( x \in \mathbb{R}^m \). Therefore, we have:

\[
h(x) - \min_{x \in \mathbb{R}^m} h(x) \geq \frac{1}{2L_{d_\rho}} \| \nabla h(x) \|^2 = \frac{1}{2L_{d_\rho}} \| \nabla_x \mathcal{L}_{0\rho}(u, x) - \nabla d_\rho(x) \|^2,
\]

for all \( u \in \mathcal{U} \). Taking \( u = \tilde{u}_\rho(x) \) and using the definition (5) of \( \tilde{u}_\rho(x) \), we have \( h(x) - \min_{x \in \mathbb{R}^m} h(x) = h(x) \leq \delta \) and obtain the following approximate gradient relation:

\[
\| \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(x), x) - \nabla d_\rho(x) \| \leq \sqrt{2\delta L_{d_\rho}}. \tag{7}
\]

From the Lipschitz continuity of \( \nabla d_\rho(\cdot) \), (5) and (7), we have that for any \( x, y \in \mathbb{R}^m \) the following relations holds:

\[
d_\rho(x) \geq d_\rho(y) + \langle \nabla d_\rho(y), x - y \rangle - \frac{L_{d_\rho}}{2} \| x - y \|^2 \tag{5}
\]

\[
\geq \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla d_\rho(y), x - y \rangle - \frac{L_{d_\rho}}{2} \| x - y \|^2 - \delta
\]

\[
= \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle - \frac{L_{d_\rho}}{2} \| x - y \|^2 - \delta
\]

\[
+ \langle \nabla d_\rho(y) - \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle
\]

\[
\geq \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle - \frac{L_{d_\rho}}{2} \| x - y \|^2 - \delta
\]

\[
- \| \nabla d_\rho(y) - \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) \| \| x - y \|
\]

\[
\geq \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle - \frac{L_{d_\rho}}{2} \| x - y \|^2 - \delta
\]

\[
- \sqrt{2\delta L_{d_\rho}} \| x - y \|.
\]

On the other hand, for any positive pair of constants \( (t, \alpha) \) we have: \( t \leq \frac{t^2}{2\alpha} + \frac{\alpha}{2} \). Thus, taking \( t = \sqrt{2 \delta L_{d_\rho}} \| x - y \| \) and \( \alpha = 2 \delta \) in the previous inequalities, we obtain the right hand side inequality of the theorem.

The relation (6) implies that \( d_\rho \) is equipped with a \((2\delta, 2L_{d_\rho})\)-oracle with \( \phi_{d_\rho}(x) = \mathcal{L}_{0\rho}(\tilde{u}_\rho(x), x) \) and \( \nabla \phi_{d_\rho}(x) = \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(x), x) = G\tilde{u}_\rho(x) + g \). It is important to note that the analysis considered in [20,28,29] requires to solve the inner problem with higher accuracy of order \( O(\delta^2) \), i.e.:

\[
\mathcal{L}_{0\rho}(\tilde{u}_\rho(x), x) - d_\rho(x) \leq O(\delta^2),
\]

in order to ensure the bounds on \( d_\rho(x) \) as

\[
\mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(y), y), x - y \rangle - d_\rho(x)
\]

\[
\leq \frac{L_{d_\rho}}{2} \| x - y \|^2 + \left( 1 + \sqrt{2L_{d_\rho}D_{\mathcal{U}}} \right) \delta.
\]
where $D_U$ is the diameter of the bounded convex domain $U$. It is obvious that our approach in this paper is less conservative, requiring to solve the $u$-inner subproblem with less accuracy than in [20,28,29]. As we will see in the sequel, this will also have an impact on the total complexity of our methods compared to those in the previous papers.

3 Inexact fast augmented Lagrangian method

In this section we propose an augmented Lagrangian smoothing strategy which is similar to the excessive gap technique introduced in [25,28,29]. Existing excessive gap strategies are based on primal-dual fast gradient methods, which maintain at each iteration some excessive gap inequality. Using this inequality, the convergence of the outer loop of the algorithm is naturally determined.

In this paper, we use an excessive gap-like inequality, which holds at each outer iteration of the proposed algorithm. Given a dual smoothness parameter $\rho$, an inner accuracy $\delta$ and an outer accuracy $\epsilon$, we further develop an Inexact Fast Augmented Lagrangian (IFAL) algorithm for solving (1) as follows.

**Algorithm IFAL ($\rho, \epsilon$)**

**Initialization:** Give $u^0 \in U$, $x^0 \in \mathbb{R}^m$, $\mu_0 > 0$ and $\{\tau_k\}_{k \geq 0}$ a step size sequence

**Iterations:** For $k = 0, 1, \ldots$, perform the following steps:

1. $\hat{x}^k = (1 - \tau_k)x^k + \frac{\tau_k}{\mu_k}(Gu^k + g)$.
2. Find $\tilde{u}_\rho(\hat{x}^k)$ such that $\mathcal{L}_0(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k) - d_\rho(\hat{x}^k) \leq \delta_k$.
3. $u^{k+1} = (1 - \tau_k)u^k + \tau_k\tilde{u}_\rho(\hat{x}^k)$.
4. $x^{k+1} = \hat{x}^k + \rho(G\tilde{u}_\rho(\hat{x}^k) + g)$.
5. Set $\mu_{k+1} = (1 - \tau_k)\mu_k$.
6. If a stopping criterion holds then **terminate** and **return** $(u^k, x^k)$.

End

Here, $\tau_k \in (0, 1)$ is a step size, which will be updated later. To analyze the convergence of IFAL($\rho, \epsilon$), as in [25,28,29], we define the smoothed duality gap:

$$\Delta_k = f_{\mu_k}(u^k) - d_\rho(x^k).$$

Based on this smoothed duality gap $\Delta_k$, we further provide a descent inequality, which will facilitate the derivation of a simple inner accuracy update and of the total complexity estimate of IFAL($\rho, \epsilon$).

**Lemma 2** Let $\rho, \epsilon > 0$ and $\{(u^k, x^k)\}_{k \geq 0}$ be the sequences generated by the IFAL($\rho, \epsilon$) algorithm. If the parameter $\tau_k \in (0, 1)$ satisfies $\frac{\tau_k^2}{1 - \tau_k} \leq \mu_k\rho$ for all $k \geq 0$, then the following excessive gap inequality holds:

$$\Delta_{k+1} \leq (1 - \tau_k)\Delta_k + 2\delta_k. \quad (8)$$
Proof First, by the strong convexity of $\|\cdot\|^2$, for any $x \in \mathbb{R}^m$, we have

$$
\Delta_k = \max_{x \in \mathbb{R}^m} \left\{ f(u^k) + \langle x, Gu^k + g \rangle - \frac{\mu_k}{2} \|x\|^2 \right\} - d_\rho(x^k)
$$

$$
\geq f(u^k) + \langle x, Gu^k + g \rangle - \frac{\mu_k}{2} \|x\|^2 + \frac{\mu_k}{2} \|x - x_{\mu_k}(u^k)\|^2 - d_\rho(x^k).
$$

(9)

Second, given any $x \in \mathbb{R}^m$, from the definition of $\mathcal{L}_{0\rho}$, we can show that

$$
\mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k), x - \hat{x}^k \rangle = f(\tilde{u}_\rho(\hat{x}^k)) + \langle x, \tilde{G}\tilde{u}_\rho(\hat{x}^k) + g \rangle
$$

$$
+ \frac{\rho}{2} \|\tilde{G}\tilde{u}_\rho(\hat{x}^k) + g\|^2.
$$

(10)

Multiplying (9) by $1 - \tau_k$ and (10) by $\tau_k$, then combining the results we obtain

$$(1 - \tau_k)\Delta_k \geq (1 - \tau_k) \left( f(u^k) + \langle x, Gu^k + g \rangle - \frac{\mu_k}{2} \|x\|^2 + \frac{\mu_k}{2} \|x - x_{\mu_k}(u^k)\|^2
$$

$$
- d_\rho(x^k) \right) + \tau_k \left( f(\tilde{u}_\rho(\hat{x}^k)) + \langle x, \tilde{G}\tilde{u}_\rho(\hat{x}^k) + g \rangle + \frac{\rho}{2} \|\tilde{G}\tilde{u}_\rho(\hat{x}^k) + g\|^2
$$

$$
- \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k) - \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k), x - \hat{x}^k \rangle \right).
$$

Using the convexity of $f$, we further derive this estimate for any $x \in \mathbb{R}^m$ as

$$(1 - \tau_k)\Delta_k \geq f((1 - \tau_k)u^k + \tau_k\tilde{u}_\rho(\hat{x}^k)) + \langle x, G \left( (1 - \tau_k)u^k + \tau_k\tilde{u}_\rho(\hat{x}^k) \right) + g \rangle
$$

$$
- \frac{\mu_{k+1}}{2} \|x\|^2 - (1 - \tau_k)d_\rho(x^k) + \frac{\mu_{k+1}}{2} \|x - x_{\mu_k}(u^k)\|^2
$$

$$
- \tau_k \left( \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k) + \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), x - \hat{x}^k) \rangle \right)
$$

Next, using $u^{k+1}$ from Step 3 and (6) into the last estimate we get

$$(1 - \tau_k)\Delta_k \geq f(u^{k+1}) + \langle x, Gu^{k+1} + g \rangle - \frac{\mu_{k+1}}{2} \|x\|^2 - \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k)
$$

$$
- \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k), (1 - \tau_k)x^k + \tau_kx - \hat{x}^k \rangle + \frac{\mu_{k+1}}{2} \|x - x_{\mu_k}(u^k)\|^2
$$

$$
\geq f(u^{k+1}) + \langle x, Gu^{k+1} + g \rangle - \frac{\mu_{k+1}}{2} \|x\|^2 - \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k)
$$

$$
- \max_{z := (1 - \tau_k)x^k + \tau_kx} \left\{ \langle \nabla_x \mathcal{L}_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k), z - \hat{x}^k \rangle - \frac{\mu_{k+1}}{2\tau_k} \|z - \hat{x}^k\|^2 \right\}.
$$

Now, by the assumption that $\frac{\tau_k^2}{1 - \tau_k} \leq \mu_{k}\rho$ and the right hand side of the inexact descent lemma (6) we therefore derive this estimate to
\[(1 - \tau_k)\Delta_k \geq f(u^{k+1}) + \langle x, Gu^{k+1} + g \rangle - \frac{\mu_{k+1}}{2} \|x\|^2 - d_\rho(x^{k+1}) - 2\delta_k, \quad \forall x \in \mathbb{R}^m.\]

Finally, by choosing \(x = \frac{1}{\mu_{k+1}}(Gu^{k+1} + g)\), we obtain (8). \(\square\)

We note that similar excessive gap inequalities have been proved in [25, 28, 29], which were then used to analyze the convergence rate of excessive gap algorithms. However, the main results in [25, 28, 29] only concern with the outer iteration-complexity and have not taken into account the necessary inner computational effort for finding \(\tilde{u}_\rho(\cdot)\), since it is very difficult to estimate this quantity using the approaches presented in those papers.

In the sequel, based on our approach, we provide the total computational complexity of the IFAL algorithm (including inner complexity) for attaining an \(\epsilon\)-optimal solution of problem (1). First, we notice that if we assume that \(\mu_0 = \frac{4}{\rho}\), then, by taking into account that \(\mu_k = \mu_0 \prod_{j=0}^{k}(1 - \tau_j)\), a simple choice of the sequence \(\{\tau_k\}_{k \geq 0}\) satisfying \(\frac{\tau_k^2}{1 - \tau_k} \leq \mu_k \rho\) is given by \(\tau_k = \frac{2}{k + 3}\).

**Theorem 1** Let \(\rho, \epsilon > 0, \mu_0 = \frac{4}{\rho}\) and \(\tau_k = \frac{2}{k + 3}\). Let \((u^k, x^k)_{k \geq 0}\) be the sequences generated by the IFAL(\(\rho, \epsilon\)) algorithm. If we choose the inner accuracy \(\delta_k = \frac{\epsilon}{2(k+3)}\), then the following estimates on the objective residual and the feasibility violation hold:

\[
\begin{align*}
\left\{ \begin{array}{l}
f(u^k) - f^* \leq \frac{\Delta_0}{(k+1)(k+2)} + \frac{\epsilon}{2}, \\
\|Gu^k + g\| \leq \frac{16\|x^*\|^2 + 4\sqrt{\rho\Delta_0}}{\rho(k+1)(k+2)} + \frac{2\sqrt{\epsilon}}{\sqrt{\rho(k+1)}}.
\end{array} \right.
\end{align*}
\]

**Proof** From (8), we can derive

\[
\Delta_{k+1} \leq \Delta_0 \prod_{j=0}^{k}(1 - \tau_j) + 2\delta_k + 2 \sum_{i=1}^{k} \delta_{k-i} \prod_{j=k-i+1}^{k}(1 - \tau_j)
= \Delta_0 \prod_{j=0}^{k}(1 - \tau_j) + 2 \prod_{j=0}^{k}(1 - \tau_j) \left( \sum_{i=0}^{k} \frac{\delta_i}{\prod_{s=0}^{i}(1 - \tau_s)} \right).
\]

(12)

By the choice of \(\tau_k\), we can show that \(\prod_{j=0}^{k}(1 - \tau_j) = \prod_{j=1}^{k+1} \frac{j+1}{j+3} = \frac{2}{(k+2)(k+3)}\), we can further bound the cumulative error as follows:

\[
2 \prod_{j=0}^{k}(1 - \tau_j) \left( \sum_{i=0}^{k} \frac{\delta_i}{\prod_{s=0}^{i}(1 - \tau_s)} \right) = \frac{4}{(k+2)(k+3)} \sum_{i=0}^{k} \frac{(i+2)(i+3)\delta_i}{2} \\
= \frac{\epsilon}{2} \frac{(k+2)(k+3) - 1}{(k+2)(k+3)} \leq \frac{\epsilon}{2}.
\]

\(\copyright\) Springer
Using this bound and the weak duality theorem $f^* \geq d_\rho(x)$ for any $x \in \mathbb{R}^m$, (12) implies

$$f(u^k) - f^* \leq f_{\mu_k}(u^k) - d_\rho(x^k) = \Delta_k \leq \frac{\Delta_0}{(k+1)(k+2)} + \frac{\epsilon}{2},$$

which is the first estimate in (11).

On the other hand, using the KKT condition of (1) we have $\langle \nabla f(u^*), u^k - u^* \rangle \geq 0$ for $u^k \in \mathcal{U}$. This implies $\langle \nabla f(u^*), u^k - u^* \rangle \geq \langle x^*, G(u^k - u^*) \rangle = \langle x^*, Gu^k + b \rangle \geq -\|x^*\|\|Gu^k + b\|$. Using this estimate, we can derive

$$f_{\mu_k}(u^k) - f^* = f(u^k) + \frac{1}{2\mu_k} \|Gu^k + g\|^2 - f^*$$

$$\geq \frac{1}{2\mu_k} \|Gu^k + g\|^2 + \langle \nabla f(u^*), u^k - u^* \rangle$$

$$\geq \frac{1}{2\mu_k} \|Gu^k + g\|^2 - \|x^*\|\|Gu^k + g\|.$$

(14)

Combining (13) and (14), we obtain

$$\|Gu^k + g\| \leq 2\mu_k\|x^*\| + \left[ \frac{2\mu_k\Delta_0}{(k+1)(k+2)} + \mu_k\epsilon \right]^{1/2}$$

$$\leq \frac{4\mu_0\|x^*\| + 2\sqrt{\mu_0\Delta_0}}{(k+1)(k+2)} + \sqrt{\frac{\mu_0\epsilon}{k+1}},$$

which is indeed the second estimate of (11).

Now, we provide the overall computational complexity of the \textbf{IFAL}(\rho, \epsilon) algorithm in terms of the number of projections onto the primal set $\mathcal{U}$.

\textbf{Theorem 2} Let $\rho, \epsilon > 0$, $\mu_0 = \frac{4}{\rho}$ and $\tau_k = \frac{2}{k+3}$. Assume that at each outer iteration $k$ of the \textbf{IFAL}(\rho, \epsilon) algorithm, Nesterov’s optimal method [24] is called for computing an approximate solution $\tilde{u}_\rho(\hat{x}^k)$ of the inner subproblem (4) such that $L_0(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k) - d_\rho(\hat{x}^k) \leq \delta_k \left( = \frac{\epsilon}{2(k+3)} \right)$. Then, for attaining an $\epsilon$-optimal solution of (1) in the sense of Definition 1, the \textbf{IFAL}(\rho, \epsilon) algorithm needs to perform at most

$$\left\lceil \frac{16\gamma^{3/2}D_{\mathcal{U}}}{\epsilon^{5/4}} \frac{\sqrt{2(L_f + \rho\|G\|^2)}}{\sqrt{L_f}} \right\rceil$$

projections onto the simple set $\mathcal{U}$, where the constant $\gamma$ is given by

$$\gamma = \max \left\{ \sqrt{2\Delta_0}, \sqrt{8/\rho}, \left( \frac{32\|x^*\|}{\rho} + 8 \sqrt{\frac{\Delta_0}{\rho}} \right)^{1/2} \right\}.$$
Proof From Theorem 1 we observe that, for attaining an \( \epsilon \)-optimal solution, the number of outer iterations \( N_{\text{out}} \) must satisfy

\[
N_{\text{out}} \leq \frac{1}{\sqrt{\epsilon}} \max \left\{ \sqrt{2\Delta_0}, 2\sqrt{\mu_0}, \left( 8\mu_0 \|x^*\| + 4\sqrt{\mu_0} \Delta_0 \right)^{1/2} \right\} = \gamma / \sqrt{\epsilon}.
\]

On the other hand, at each outer iteration \( k \), Assumption A.1(iii) implies that Nesterov’s optimal method [24] applied on the inner subproblem (4) performs

\[
N_{\text{in}}^k = 2D_U \sqrt{\frac{L_f + \rho \|G\|^2}{\delta_k}} = 2D_U \sqrt{\frac{2(L_f + \rho \|G\|^2)(k + 3)}{\epsilon}}
\]

inner iterations (i.e., projections onto \( \mathcal{U} \)). Using this estimate we can easily derive the total number of projections onto the simple set \( \mathcal{U} \) necessary for attaining an \( \epsilon \)-optimal point as

\[
\sum_{k=0}^{N_{\text{out}}} N_{\text{in}}^k = 2D_U \left[ \frac{2(L_f + \rho \|G\|^2)}{\epsilon} \right]^{1/2} \sum_{k=0}^{N_{\text{out}}} (k + 3)^{1/2} \leq 2D_U \left[ \frac{2(L_f + \rho \|G\|^2)}{\epsilon} \right]^{1/2} (N_{\text{out}}^3 + 3)^{3/2} \leq 16\gamma^{3/2} D_U \left[ \frac{2(L_f + \rho \|G\|^2)}{\epsilon} \right]^{1/2} \epsilon^{5/4},
\]

which proves the theorem. \( \square \)

Remark 1 We observe that \( \Delta_0 \leq \tilde{\Delta}_0 \), where \( \tilde{\Delta}_0 = f_{\mu_0}(u^0) - \mathcal{L}_{0\rho}(\tilde{u}_\rho(x^0), x^0) + \delta_0 \) can be computed explicitly. Then, using this upper bound in our estimates, makes the \( \text{IFAL}(\rho, \epsilon) \) algorithm implementable, i.e., the algorithm is terminated within an \( \epsilon \)-optimal solution at the outer iteration \( N_{\text{out}} \) provided that the following two computable conditions hold:

\[
\|Gu^k + g\| \leq \epsilon \quad \text{and} \quad N_{\text{out}} \geq \sqrt{\frac{2\tilde{\Delta}_0}{\epsilon}}.
\]

Moreover, as suggested, e.g., in [28, 29], if we choose the initial points \( u^0 = \tilde{u}_\rho(0) \) and \( x^0 = \frac{1}{\mu_0}(Gu^0 + g) \), then we can guarantee that \( \Delta_0 \leq \delta_0 \).

4 Adaptive inexact fast augmented Lagrangian method

We analyze the overall computational complexity of the \( \text{IFAL}(\rho, \epsilon) \) algorithm for an optimal choice of the smoothness parameter \( \rho \) and then we introduce an adaptive variant of this method which preserves the same computational complexity (up to a logarithmic factor), but is fully implementable in practice.
First, assume that we adopt the initialization strategy suggested in Remark 1 such that $\Delta_0 \leq \delta_0$. Therefore, using this strategy and the previous assumption that $\delta_k = \frac{\epsilon}{2(k+3)}$, the outer complexity can be estimated by

$$N_{\text{out}} \leq \frac{1}{\sqrt{\epsilon}} \max \left\{ \sqrt{\frac{\epsilon}{3}}, \sqrt{\frac{8}{\rho}}, \left( \frac{32 \|x^*\|}{\rho} + 8 \sqrt{\frac{\epsilon}{6 \rho}} \right)^{1/2} \right\} =: \tilde{\gamma} \sqrt{\epsilon}. \quad (16)$$

Note that the variation of the smoothing parameter $\rho$ induces a trade-off between the number of the outer iterations and the complexity of the inner subproblem, i.e., for a sufficiently large $\rho$, we have a single outer iteration, but a complex inner subproblem. The next result provides an optimal choice for $\rho$ (up to a constant factor) such that the best total complexity is achieved. For simplicity of the exposition, we can assume that $\|x^*\| \geq 1$ without loss of generality.

**Theorem 3** Let $\rho, \epsilon > 0$, $\mu_0 = \frac{4}{\rho}$ and $\tau_k = \frac{2}{k+3}$. Assume that, at each outer iteration $k$ of the IFAL($\rho, \epsilon$) algorithm, Nesterov’s optimal method [24] is called for computing an approximate solution $\tilde{u}_\rho(\hat{x}^k)$ of the inner subproblem (4) such that

$$L_{0\rho}(\tilde{u}_\rho(\hat{x}^k), \hat{x}^k) - d_\rho(\hat{x}^k) \leq \delta_k \left( = \frac{\epsilon}{2(k+3)} \right).$$

Then, by choosing the smoothness parameter $\rho$ as

$$\rho = \frac{64\|x^*\|^2}{\epsilon},$$

for attaining an $\epsilon$-optimal solution of (1), the IFAL($\rho, \epsilon$) algorithm needs to perform at most

$$\left\lceil \sqrt{\frac{6L_f D_{\text{U}}^2}{\epsilon}} + \frac{4\sqrt{6}D_{\text{U}} \|G\| \|x^*\|}{\epsilon} \right\rceil$$

projections onto the simple primal feasible set $\mathcal{U}$.

**Proof** First, we observe that for $\rho \geq \frac{64\|x^*\|^2}{\epsilon}$, the IFAL($\rho, \epsilon$) algorithm performs a single outer iteration. Indeed, from (16) one can obtain $\tilde{\gamma} \leq \sqrt{\epsilon}$ provided that

$$\rho \geq \frac{8}{\epsilon} \quad \text{and} \quad \frac{32\|x^*\|}{\rho} + 8 \sqrt{\frac{\epsilon}{6 \rho}} \leq \epsilon.$$

It can be seen that for any $\rho$ such that $\rho \geq \frac{64\|x^*\|^2}{\epsilon}$, it satisfies the above two conditions. In this case, when a single outer iteration is sufficient, the total computational complexity is given by the inner complexity estimate (15) as

$$\sqrt{\frac{6L_f D_{\text{U}}^2}{\epsilon}} + \frac{4\sqrt{6}D_{\text{U}} \|G\| \|x^*\|}{\epsilon}.$$
On the other hand, if \( \rho < \frac{64\|x^*\|^2}{\epsilon} \), observing that \( \frac{64\|x^*\|^2}{\epsilon} \) majorizes all terms in (16), then we can further bound \( N^{\text{out}} \) as follows:

\[
N^{\text{out}} \leq \frac{8\|x^*\|}{\sqrt{\rho\epsilon}}. \tag{17}
\]

Using the same inner complexity estimate (15) of Nesterov’s optimal method [24], the total computational complexity is given by

\[
\sum_{k=0}^{N^{\text{out}}+1} N^{\text{in}}_k \leq 2D\mathcal{U} \left[ \frac{2(L_f + \rho\|G\|^2)}{\epsilon} \right]^{1/2} (N^{\text{out}} + 4)^{3/2} \leq 2(N^{\text{out}} + 4)^{3/2} \sqrt{\frac{2L_f D^2}{\epsilon} + 2D\mathcal{U}\|G\|(N^{\text{out}} + 4)^{3/2} \sqrt{\frac{2\rho}{\epsilon}}.}
\]

Using (17) for the second term on the right hand side of the above estimate, and optimizing over \( \rho \) we obtain that, for the optimal parameter \( \rho^* = \frac{\|x^*\|^2}{\epsilon} \in \left(0, \frac{64\|x^*\|^2}{\epsilon}\right) \), the necessary number of outer iterations is constant. We can conclude that the choice \( \frac{64\|x^*\|^2}{\epsilon} \) is optimal up to a constant factor. \( \square \)

If one knows a priori an upper bound on \( \|x^*\| \), then the previous result indicates that a proper choice of \( \rho \) determines that a single outer iteration of the \( \text{IFAL}(\rho, \epsilon) \) algorithm to be sufficient to attain \( \epsilon \)-primal solution and the overall computational complexity is of order \( O\left(\frac{1}{\epsilon}\right) \), which is better than \( O\left(\frac{1}{\epsilon^{5/4}}\right) \) for any \( \rho > 0 \). However, in practice \( \|x^*\| \) is unknown and thus the optimal smoothness parameter \( \rho \) cannot be computed. In order to cope with this problem, we further provide an implementable adaptive variant of the \( \text{IFAL}(\rho, \epsilon) \) algorithm, which preserves the same total complexity given in Theorem 3 (up to a logarithmic factor). The Adaptive Inexact Fast augmented Lagrangian (A-IFAL) algorithm relies on a search procedure which is often used for penalty and augmented Lagrangian methods in the case when a bound on the optimal Lagrange multipliers is unknown (see, e.g., [14]).

**A-IFAL(\( \rho_0, \epsilon \)) Algorithm**

**Initialization:** Choose \( \rho_0, \epsilon > 0, \mu_0 = \frac{4}{\rho_0} \) and \((u^0, x^0)\) such that \( \Delta_0 \leq \delta_0 \).

**Iterations:** For \( k = 1, 2, \ldots \), perform:

1. Starting from \((u^{k-1}, x^{k-1})\), run a single iteration of the \( \text{IFAL}(\rho_k, \epsilon) \) algorithm and obtain the output: \((u^k, x^k)\).

2. If \( \|Gu^k + g\| \leq \epsilon \) then **terminate**. Otherwise, update \( \rho_{k+1} = 2\rho_k \).

**End**

We notice that the **A-IFAL(\( \rho, \epsilon \))** algorithm can be considered as a variant of the **IFAL(\( \rho, \epsilon \))** algorithm with a variable increasing smoothness parameter \( \rho \) and a constant inner accuracy \( \delta = \frac{\epsilon}{5} \). The following result provides the overall complexity of **A-IFAL(\( \rho, \epsilon \))** necessary for attaining an \( \epsilon \)-optimal solution.
Theorem 4 Let \(\{(u^k, x^k)\}_{k \geq 0}\) be the sequence generated by the A-IFAL(\(\rho, \epsilon\)) algorithm. Then, this algorithm needs to perform at most
\[
\log_2 \left( \frac{64\|x^*\|^2}{\epsilon \rho_0} \right) \sqrt{\frac{24L_f D^2 \|G\| \|x^*\|}{\epsilon}} + \frac{96\sqrt{3}D_{U\ell}\|G\|\|x^*\|}{\epsilon}
\]
projections onto the primal feasible set \(U\) to attain an \(\epsilon\)-solution of (1).

Proof We observe that the maximum number of outer iterations performed by the A-IFAL(\(\rho, \epsilon\)) algorithm is given by
\[
N_{\text{out}} \leq \log_2 \left( \frac{64\|x^*\|^2}{\epsilon \rho_0} \right) \sqrt{\frac{24L_f D^2 \|G\| \|x^*\|}{\epsilon}} + \frac{96\sqrt{3}D_{U\ell}\|G\|\|x^*\|}{\epsilon}
\]
which proves the bound in Theorem 4.

It is important to note that the A-IFAL(\(\rho, \epsilon\)) algorithm has the same computational complexity as the IFAL(\(\rho^*, \epsilon\)) algorithm, up to a logarithmic factor. Moreover, both algorithms are implementable, i.e., they can be terminated based on verifiable stopping criteria and their parameters are easy to compute.

5 Comparison with other augmented Lagrangian complexity results

In this section we compare the computational complexity and other features of the IFAL/A-IFAL algorithm with some existing related works and the complexity results of AL methods.

Given \(x \in U\) and \(r > 0\), we use the notations \(B_r(x) = \{y \in \mathbb{R}^n | \|y\| \leq r\}\) and \(N_{U\ell}(x) = \{y \in \mathbb{R}^n | \langle y, z - x \rangle \leq 0 \ \forall z \in U\}\). Then, in [14], the authors analyzed the classical AL method for the same class of problems (1). They developed an implementable variant of the classical AL method to obtain an \(\epsilon\) - suboptimal primal-dual pair \((u_\epsilon, x_\epsilon)\) satisfying the following criteria:
\[
\nabla f(u_\epsilon) + G^T x_\epsilon \in -N_{U\ell}(u_\epsilon) + \mathcal{B}_\epsilon(0) \quad \text{and} \quad \|Gu_\epsilon + g\| \leq \epsilon.
\]

The authors provided their own iteration-complexity analysis for the augmented Lagrangian method using an inexact dual gradient algorithm. Without any artificial
perturbation on the problem, they shown that it is necessary to perform $O\left(\frac{1}{\varepsilon^{1/4}}\right)$ projections onto the simple primal set $U$ in order to obtain a primal-dual pair satisfying (18). One main disadvantage of the method in [14] is that, for some $\rho > 0$, it requires \textit{a priori} a pre-specified number of the outer iterations in order to compute the inner accuracy and to terminate the algorithm. Moreover, to satisfy our $\epsilon$-optimality definition, an average primal iterate $u^k = \frac{1}{k} \sum_{i=0}^{k} u^i$ must be computed (see e.g. [20]). Furthermore, for any fixed $\rho$, $O\left(\frac{\|x^*\|}{\rho \varepsilon}\right)$ outer iterations are required, and the inner accuracy must be chosen as $\delta_k = O\left(\frac{\varepsilon^3}{\|x^*\|}\right)$. In this case, the method in [14] requires:

$$\left[\frac{2^7 D_u L f^{1/2} \|x^*\|^{3/2}}{\rho^2 \varepsilon^{5/2}} + \frac{2^7 \|G\| D_u \|x^*\|^{3/2}}{\rho^{5/2} \varepsilon^{5/2}}\right]$$

(19)
total projections onto the set $U$, provided that $\rho \leq \frac{2^4 \sqrt{\|x^*\|}}{5 \epsilon}$. However, we observe that for $\rho = O\left(\|x^*\|^2 / \epsilon\right)$, we obtain from (19) an $O\left(1/\epsilon\right)$ overall computational complexity, while for an arbitrary constant parameter $\rho$, the complexity estimates are much worse than our results given in this paper.

It is important to note that although the inexact excessive gap methods introduced in [20,28,29] are similar to the IFAL($\rho, \varepsilon$) algorithm and the authors also provided an outer complexity estimate of order $O\left(\frac{1}{\varepsilon^{1/2}}\right)$ for a fixed $\rho$, the update rules for the inner accuracy in [20,28,29] induce difficulties in deriving the overall computational complexity. Moreover, assuming one implements the update rule of the inner accuracy $\delta_k$ given, e.g., by [28, Theorem 5.1], at each outer iteration, it is required a primal iterate $\tilde{u}_\rho(x)$ satisfying: $L_0(\tilde{u}_\rho, x) - d_\rho(x) \leq \frac{\rho \delta_k^2}{2}$. For a small constant $\rho$ and a high accuracy, the theoretical complexity estimate of such algorithms can be very pessimistic. Moreover, from our previous analysis we can conclude that for an adequate choice of the parameter $\rho$, the number of outer iterations is only 1, and therefore, the outer complexity estimates are irrelevant to the total complexity of the method.

Other recent complexity results concerning with the classical AL method were given, e.g., in [1,2,20]. For example, in [1], an adaptive classical AL method for cone constrained convex optimization models was analyzed. Unfortunately, this method is not entirely implementable since the stopping criteria cannot be verified, and the inner accuracy is constrained to be of order

$$\delta_k = O\left(\frac{1}{k^2 \beta^k}\right),$$

where $\beta > 1$. The authors in [1] showed that the outer complexity is of order $O\left(\log(1/\epsilon)\right)$ and thus the overall complexity is similar to the estimates given in our paper (up to a logarithmic factor). However, our IFAL($\rho, \varepsilon$) algorithm represents an accelerated augmented Lagrangian method based on the excessive gap technique, and, in addition, it can be easily implemented in practice, i.e., it can be terminated based on verifiable stopping criteria, and its parameters are easy to compute.
6 Concluding remarks

We have analyzed the iteration-complexity of the two inexact accelerated first-order augmented Lagrangian algorithms for solving a class of large scale linearly constrained convex optimization problems. By means of smoothing techniques and an excessive gap-like condition, we provided the estimates on the overall computational complexity of these algorithms. We have showed how to compute an optimal choice of the penalty parameter $\rho$, and compared our theoretical results with other existing results in the literature.

References

1. Aybat, N., Iyengar, G.: An augmented Lagrangian method for conic convex programming, working paper. arXiv:1302.6322 (2013)
2. Aybat, N., Iyengar, G.: A first-order augmented Lagrangian method for compressed sensing. SIAM J Optim 22, 429–459 (2012)
3. Bauschke, H., Combettes, P.: Convex analysis and monotone operators theory in Hilbert spaces. Springer, Verlag (2011)
4. Ben-Tal, A., Nemirovski, A.: Lectures on modern convex optimization: analysis, algorithms, and engineering applications, vol. 3. MPS/SIAM series on optimization, SIAM (2001)
5. Bertsekas, D.: Convex optimization theory. Athena Scientific (2009)
6. Briceno-Arias, L., Combettes, P.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM J Optim 21, 1230–1250 (2011)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J Math Imaging Vis 40, 120–145 (2011)
8. Combettes, P.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization 53, 475–504 (2004)
9. Devolder, O., Glineur, F., Nesterov, Y.: First-order methods of smooth convex optimization with inexact oracle. Math Program 146, 37–75 (2014)
10. Eckstein, J., Bertsekas, D.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math Program 55, 293–318 (1992)
11. He, B., Tao, M., Yuan, X.: Alternating direction method with Gaussian back substitution for separable convex programming. SIAM J Optim 22(2), 313–340 (2012)
12. He, B., Yang, H., Zhang, C.: A modified augmented Lagrangian method for a class of monotone variational inequalities. Eur J Oper Res 159(1), 35–51 (2004)
13. He, B., Yuan, X.: On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method. SIAM J Num Anal 50, 700–709 (2012)
14. Lan, G., Monteiro, R.: Iteration-complexity of first-order augmented Lagrangian methods for convex programming. Math Program 155(1–2), 511–547 (2016). doi:10.1007/s10107-015-0861-x
15. Li, X., Yuan, X.: A proximal strictly contractive Peaceman–Rachford splitting method for convex programming with applications to imaging. SIAM J Imaging Sci 8, 1332–1365 (2015)
16. Necoara, I., Nedelcu, V.: Rate analysis of inexact dual first order methods: application to dual decomposition. IEEE Trans Automa Control 59(5), 1232–1243 (2014)
17. Necoara, I., Patrascu, A.: Iteration complexity analysis of dual first order methods for conic convex programming, technical report. Opt Met Soft. arXiv:1409.1462 (2014)
18. Necoara, I., Patrascu, A., Glineur, F.: Complexity certifications of first order inexact Lagrangian and penalty methods for conic convex programming, Tech. Rep., Univ. Politehnica Bucharest, pp. 1–34 (2015)
19. Necoara, I., Suykens, J.: Application of a smoothing technique to decomposition in convex optimization. IEEE Trans Autom Control 53(11), 2674–2679 (2008)
20. Nedelcu, V., Necoara, I., Tran-Dinh, Q.: Computational complexity of inexact gradient augmented Lagrangian methods: application to constrained MPC. SIAM J Control Optim 52(5), 3109–3134 (2014)
21. Nedic, A., Ozdaglar, A.: Approximate primal solutions and rate analysis for dual subgradient methods. SIAM J Optim 19(4), 1757–1780 (2009)
22. Nemirovskii, A.: Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J Optim 15, 229–251 (2004)
23. Nesterov, Y.: New primal-dual subgradient methods for convex problems with functional constraints. http://lear.inrialpes.fr/workshop/osl2015/slides/osl2015_yurii.pdf (2015)
24. Nesterov, Y.: Introductory lectures on convex optimization: a basic course. Kluwer, Boston (2004)
25. Nesterov, Y.: Excessive gap technique in nonsmooth convex minimization. SIAM J Optim 16(1), 235–249 (2005)
26. Nesterov, Y.: Subgradient methods for huge-scale optimization problems. Math Program 146, 275–297 (2014)
27. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math Oper Res 1, 97–116 (1976)
28. Tran-Dinh, Q., Cevher, V.: A primal-dual algorithmic framework for constrained convex minimization, technical report. arXiv:1406.5403 (2014)
29. Tran-Dinh, Q., Necoara, I., Diehl, M.: Fast inexact distributed optimization algorithms for separable convex optimization. Optimization 65(2), 325–356 (2016)
30. Tran-Dinh, Q., Savorgnan, C., Diehl, M.: Combining Lagrangian decomposition and excessive gap smoothing technique for solving large-scale separable convex optimization problems. Comput Optim Appl 55(1), 75–111 (2013)