

Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical Lagrangian function and a full Lagrange multiplier update

August 2, 2020

JEFFERSON G. MELO ¹ AND RENATO D.C. MONTEIRO ²

Abstract

This paper establishes the iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian (IAPIAL) method for solving linearly constrained smooth nonconvex composite optimization problems which is based on the classical Lagrangian function and, most importantly, performs a full Lagrangian multiplier update, i.e., no shrinking factor is incorporated on it. More specifically, each IAPIAL iteration consists of inexactly solving a proximal augmented Lagrangian subproblem by an accelerated composite gradient (ACG) method followed by a full Lagrange multiplier update. Under the assumption that the domain of the composite function is bounded and the problem has a Slater point, it is shown that IAPIAL generates an approximate stationary solution in at most $\mathcal{O}(\log(1/\rho)/\rho^3)$ ACG iterations, where $\rho > 0$ is the tolerance for both stationarity and feasibility. Finally, the above bound is derived without assuming that the initial point is feasible.

key words. Inexact proximal augmented Lagrangian method, linearly constrained smooth nonconvex composite programs, inner accelerated first-order methods, iteration-complexity.

AMS subject classifications. 47J22, 49M27, 90C25, 90C26, 90C30, 90C60, 65K10.

1 Introduction

This paper presents an inner accelerated proximal inexact augmented Lagrangian (IAPIAL) method for solving the linearly constrained smooth nonconvex composite optimization problem

$$\phi^* := \min\{\phi(z) := f(z) + h(z) : Az = b\}, \quad (1)$$

where $A : \mathbb{R}^n \rightarrow \mathbb{R}^l$ is a linear operator, $b \in \mathbb{R}^l$, $h : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is a closed proper convex function which is M_h -Lipschitz continuous on its domain, and f is a real-valued differentiable nonconvex function such that, for some scalars $L_f \geq m_f > 0$, f is m_f -weakly convex on the domain, $\text{dom } h$, of

¹Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II- Caixa Postal 131, CEP 74001-970, Goiânia-GO, Brazil. (E-mail: jefferson@ufg.br). This author was partially supported by CNPq grant 312559/2019-4 and FAPEG/GO.

²School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (email: monteiro@isye.gatech.edu). This author was partially supported by ONR Grant N00014-18-1-2077.

h (i.e., satisfies (13) below) and its gradient is L_f -Lipschitz on $\text{dom } h$. For a given tolerance pair $(\hat{\rho}, \hat{\eta}) \in \mathbb{R}_{++}^2$, its goal is to find a triple $(\hat{z}, \hat{w}, \hat{p})$ satisfying

$$\hat{w} \in \nabla f(\hat{z}) + \partial h(\hat{z}) + A^* \hat{p}, \quad \|\hat{w}\| \leq \hat{\rho}, \quad \|A\hat{z} - b\| \leq \hat{\eta}. \quad (2)$$

More specifically, the IAPIAL method is based on the augmented Lagrangian function $\mathcal{L}_c(z; p)$ defined as

$$\mathcal{L}_c(z; p) := f(z) + h(z) + \langle p, Az - b \rangle + \frac{c}{2} \|Az - b\|^2, \quad (3)$$

which has been thoroughly studied in the literature (see for example [3, 5, 26, 34, 47]). Roughly speaking, for a fixed stepsize $\lambda > 0$ and started from $(z_0, p_0) \in \text{dom } h \times \mathbb{R}^l$ with $p_0 = 0$, IAPIAL repeatedly performs the following iteration: given $(z_{k-1}, p_{k-1}) \in \text{dom } h \times \mathbb{R}^l$, it computes (z_k, p_k) as

$$z_k \approx \underset{z}{\text{argmin}} \left\{ \lambda \mathcal{L}_c(z, p_{k-1}) + \frac{1}{2} \|z - z_{k-1}\|^2 \right\} \quad (4)$$

$$p_k = p_{k-1} + c(Az_k - b), \quad (5)$$

where z_k in (4) is a suitable approximate solution of the underlying prox-AL subproblem (4). IAPIAL sets $\lambda = 1/(2m_f)$ which, due to the fact that f is m_f -weakly convex, guarantees that the objective function of (4) is strongly convex. It then approximately solves the corresponding subproblem (4) by a strongly convex version of an accelerated composite gradient (ACG) method (see for example [4, 37, 41]) to obtain z_k . The latter point is then used to construct a triple $(\hat{z}_k, \hat{w}_k, \hat{p}_k)$ and the IAPIAL method stops if it satisfies (2). Otherwise, an auxiliary novel test is performed to decide whether: i) c should be left unchanged, or; ii) c is updated as $c \leftarrow 2c$ and (z_k, p_k) either reset to (z_0, p_0) (cold restart) or to (z_k, p_0) (hybrid warm restart). The iteration described above is then repeated with the updated c and the new pair (z_k, p_k) .

Under the assumption that the domain of h is bounded, has nonempty interior, and (1) has a Slater point, i.e., a point $\bar{z} \in \text{int}(\text{dom } h)$ such that $A\bar{z} = b$, it is shown that the total ACG iteration complexity of IAPIAL is $\mathcal{O}(\max\{1/\hat{\rho}^3, 1/(\hat{\rho}^2\sqrt{\hat{\eta}})\} \log(\max\{1/\hat{\rho}^2, 1/\hat{\eta}\}))$ independently of whether the cold or the hybrid warm restart strategy is used. Since each ACG iteration requires $\mathcal{O}(1)$ resolvent evaluations of h and/or gradient evaluations of f , the previous complexity also bounds the number of h -resolvents and gradients evaluations of f performed by IAPIAL. It is worth mentioning that the latter result holds without assuming that the initial point $z_0 \in \text{dom } h$ is feasible, i.e., satisfies $Az_0 = b$.

Related works. The following paragraphs discusses related works in different settings of (1), namely: in the convex setting (i.e., both f and h convex), in the nonconvex setting (i.e., f nonconvex) with h convex and $A = 0$, and in the nonconvex setting with h convex and $A \neq 0$.

Convex setting. Iteration-complexity of quadratic penalty methods for solving (1) under the assumption that f is convex and h is an indicator function of a convex set was first analyzed in [25] and further studied in [2, 40]. Iteration-complexity of first-order augmented Lagrangian (AL) methods for solving the aforementioned class of convex problem was studied in [3, 26, 34, 35, 45, 50]. Iteration-complexity of inexact proximal point methods using accelerated composite gradient (ACG) algorithms to solve their prox-subproblems were considered in [9, 39, 21, 16, 15, 38] in the setting of convex-concave saddle point problems and monotone variational inequalities.

Nonconvex setting with h convex and $A = 0$. Algorithms in this setting have been studied for example in [7, 10, 11, 12, 27, 44]. More specifically, an AG framework in this setting was proposed and analyzed in [11]. Since then, other accelerated schemes in this setting have been proposed in

the literature (see for example [30, 32, 31, 7, 10, 12, 27, 44, 22]) for first-order methods and [7, 8, 43] for second-order methods.

Nonconvex setting with $A \neq 0$. Proximal quadratic penalty (PQP) type methods in this setting have been studied in [22, 23, 33]. Iteration-complexity of a PQP inexact proximal point method whose subproblems are inexactly solved by an ACG scheme was first considered in [22] and further explored in [23] where the authors propose a more computationally efficient variant which adaptively chooses the prox-stepsize λ . Paper [33] also studies an inexact PQP method and establishes an improved iteration-complexity bound under the assumption that $\text{dom } h$ is bounded and the Slater condition holds. Finally, [24] analyzed the iteration-complexity of a PQP based method for solving (1) under the assumption that $f(\cdot) = \max\{\Phi(\cdot, y) : y \in Y\}$ where Y is a compact convex set, $-\Phi(x, \cdot)$ is proper lower semi-continuous convex for every $x \in \text{dom } h$, and $\Phi(\cdot, y)$ is nonconvex differentiable on $\text{dom } h$ and its gradient is uniformly Lipschitz continuous on $\text{dom } h$ for every $y \in Y$.

Proximal augmented Lagrangian (PAL) type methods for solving (1) or a more general class of it have been studied, for example, in [14, 19, 36]. Paper [19] studies the iteration-complexity of a linearized PAL method to solve (1) under the strong assumption that $h = 0$. Paper [14] introduces a perturbed AL function for problem (1) and studies an unaccelerated PAL inexact proximal method, establishing an $\mathcal{O}(1/(\hat{\eta}^4 + \hat{\rho}^4))$ iteration-complexity under the condition that the initial point z_0 be feasible, i.e., $Az_0 = b$ and $z_0 \in \text{dom } h$. In [36], the authors analyze the iteration-complexity of an inexact proximal accelerated PAL method based on the aforementioned perturbed AL function, showing that a solution to (2) is obtained in at most $\mathcal{O}(\log(1/\hat{\eta})/(\hat{\eta}\hat{\rho}^2))$ ACG iterations and that the latter bound can be improved to $\mathcal{O}(\log(1/\hat{\eta})/(\sqrt{\hat{\eta}}\hat{\rho}^2))$ under additional mildly stronger assumptions.

Papers [51, 52] present a primal-dual first-order algorithm for solving (1) where h is the indicator function of a box (in [52]) or more generally a polyhedron (in [51]), and show that it solves (2) with $\hat{\rho} = \hat{\eta}$ in at most $\mathcal{O}(1/\hat{\rho}^2)$ iterations. Each iteration of the algorithm performs a projected gradient step applied to a prox AL-type function followed by a conservative update on the Lagrangian multiplier and the prox center.

Nonconvex setting with nonlinear constraints. Other methods related to the ones of the previous two paragraphs were studied in [28, 48] where the more general setting of (1) with $Az = b$ replaced by a smooth nonlinear constraint $g(x) = 0$ or $g(x) \leq 0$ is considered. More specifically, papers [28] and [48] both consider AL type methods which perform Lagrange multiplier updates only when the penalty parameter c increases (this contrasts with (4)-(5) where the update on p_k is performed at every step regardless of whether c increases or not). Since the number of possible updates on c , and hence p_k , is at most $\log(\eta^{-1})$, suitable bounds on $\|p_k\|$ can be easily obtained and their analyses are much closer to the ones for the aforementioned PQP type methods than to the ones in [36] and/or this work. Paper [29] studies a hybrid penalty/AL based method whose penalty iterations are the ones which guarantee its convergence and whose AL iterations are included with the purpose of improving its computational efficiency. Paper [6] considers a primal-dual proximal point scheme for computing approximate stationary solution to a constrained nonconvex composite optimization problem and analyzes its iteration-complexity under different assumptions. Finally, [20] considers a penalty-ADMM method which approximately solves (1) by solving an equivalent reformulation of it.

Organization of the paper. Subsection 1.1 provides some basic definitions and notation. Section 2 reviews an ACG method. Section 3 contains three subsections. The first one presents our main problem of interest, the assumptions made on it and an outline of IAPIAL. Subsection 3.2 states the S-IAPIAL method and its main iteration-complexity result. Subsection 3.3 states the IAPIAL method and establishes its iteration-complexity bound. Section 4 is devoted to the proof

of the iteration-complexity result of S-IAPIAL and some related technical results. Finally, the appendix contains some basic results.

1.1 Notation and basic definitions

This subsection presents notation and basic definitions used in this paper.

Let \mathfrak{R}_+ and \mathfrak{R}_{++} denote the set of nonnegative and positive real numbers, respectively, and let \mathfrak{R}^n denote the n -dimensional Hilbert space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. The smallest positive singular value of a nonzero linear operator $Q : \mathfrak{R}^n \rightarrow \mathfrak{R}^l$ is denoted by σ_Q^+ . For a given closed convex set $X \subset \mathfrak{R}^n$, its boundary is denoted by ∂X and the distance of a point $x \in \mathfrak{R}^n$ to X is denoted by $\text{dist}_X(x)$. For any $t > 0$, we let $\log_1^+(t) := \max\{\log t, 1\}$ and $\bar{B}(0, t) := \{z \in \mathfrak{R}^n : \|z\| \leq t\}$.

The domain of a function $h : \mathfrak{R}^n \rightarrow (-\infty, \infty]$ is the set $\text{dom } h := \{x \in \mathfrak{R}^n : h(x) < +\infty\}$. Moreover, h is said to be proper if $\text{dom } h \neq \emptyset$. The set of all lower semi-continuous proper convex functions defined in \mathfrak{R}^n is denoted by $\text{Conv}(\mathfrak{R}^n)$. The ε -subdifferential of a proper function $h : \mathfrak{R}^n \rightarrow (-\infty, \infty]$ is defined by

$$\partial_\varepsilon h(z) := \{u \in \mathfrak{R}^n : h(z') \geq h(z) + \langle u, z' - z \rangle - \varepsilon, \quad \forall z' \in \mathfrak{R}^n\} \quad (6)$$

for every $z \in \mathfrak{R}^n$. The classical subdifferential, denoted by $\partial h(\cdot)$, corresponds to $\partial_0 h(\cdot)$. Recall that, for a given $\varepsilon \geq 0$, the ε -normal cone of a closed convex set C at $z \in C$, denoted by $N_C^\varepsilon(z)$, is defined as

$$N_C^\varepsilon(z) := \{\xi \in \mathfrak{R}^n : \langle \xi, u - z \rangle \leq \varepsilon, \quad \forall u \in C\}.$$

If ψ is a real-valued function which is differentiable at $\bar{z} \in \mathfrak{R}^n$, then its affine approximation $\ell_\psi(\cdot, \bar{z})$ at \bar{z} is given by

$$\ell_\psi(z; \bar{z}) := \psi(\bar{z}) + \langle \nabla \psi(\bar{z}), z - \bar{z} \rangle \quad \forall z \in \mathfrak{R}^n. \quad (7)$$

2 An accelerated gradient method

This subsection reviews the ACG variant invoked by the IAPIAL method for solving the sequence of subproblems (4) which arise during its implementation. It also describes a bound on the number of ACG iterations performed in order to obtain a certain type of approximate solution of the subproblem.

Consider the following composite optimization problem

$$\min\{\psi(x) := \psi_s(x) + \psi_n(x) : x \in \mathfrak{R}^n\} \quad (8)$$

where the following conditions are assumed to hold:

- (A1)** $\psi_n : \mathfrak{R}^n \rightarrow (-\infty, +\infty]$ is a proper, closed and μ -strongly convex function with $\mu \geq 0$;
- (A2)** ψ_s is a convex differentiable function on $\text{dom } \psi_n$ and there exists $M_s > 0$ satisfying $\psi_s(u) - \ell_{\psi_s}(u; x) \leq M_s \|u - x\|^2/2$ for every $x, u \in \text{dom } \psi_n$ where $\ell_{\psi_s}(\cdot; \cdot)$ is defined in (7).

We refer to a pair (ψ_s, ψ_n) satisfying (8), **(A1)** and **(A2)** as a composite structure of ψ .

We now state the aforementioned ACG variant for solving (8). We remark that other ACG variants such as the ones in [1, 16, 42, 41, 49] could also have been used in the development of the IAPIAL method.

ACG Method

(0) Let a pair of functions (ψ_s, ψ_n) satisfying **(A1)** and **(A2)** and an initial point $x_0 \in \text{dom } \psi_n$ be given, and set $y_0 = x_0$, $A_0 = 0$, $\Gamma_0 \equiv 0$ and $j = 0$;

(1) compute

$$\begin{aligned}
 A_{j+1} &= A_j + \frac{\mu A_j + 1 + \sqrt{(\mu A_j + 1)^2 + 4M_s(\mu A_j + 1)A_j}}{2M_s}, \\
 \tilde{x}_j &= \frac{A_j}{A_{j+1}}x_j + \frac{A_{j+1} - A_j}{A_{j+1}}y_j, \quad \Gamma_{j+1} = \frac{A_j}{A_{j+1}}\Gamma_j + \frac{A_{j+1} - A_j}{A_{j+1}}\ell_{\psi_s}(\cdot, \tilde{x}_j), \\
 y_{j+1} &= \operatorname{argmin}_y \left\{ \Gamma_{j+1}(y) + \psi_n(y) + \frac{1}{2A_{j+1}}\|y - y_0\|^2 \right\}, \\
 x_{j+1} &= \frac{A_j}{A_{j+1}}x_j + \frac{A_{j+1} - A_j}{A_{j+1}}y_{j+1};
 \end{aligned}$$

(2) compute

$$\begin{aligned}
 u_{j+1} &= \frac{y_0 - y_{j+1}}{A_{j+1}}, \\
 \eta_{j+1} &= \psi(x_{j+1}) - \Gamma_{j+1}(y_{j+1}) - \psi_n(y_{j+1}) - \langle u_{j+1}, x_{j+1} - y_{j+1} \rangle;
 \end{aligned}$$

(3) set $j \leftarrow j + 1$ and go to (1).

Some remarks about the ACG method follow. First, the main core and usually the common way of describing an iteration of the ACG method is as in step 1. Second, the extra sequences $\{u_j\}$ and $\{\eta_j\}$ computed in step 2 will be used to develop a stopping criterion for the ACG method when it is called as a subroutine for solving the subproblems of S-IAPIAL in Subsection 3.2. Third, the ACG method in which $\mu = 0$ is a special case of a slightly more general one studied by Tseng in [49] (see Algorithm 3 of [49]). The analysis of the general case of the ACG method in which $\mu \geq 0$ was studied in [16, Proposition 2.3]. The sequence $\{A_k\}$ has the following increasing property:

$$A_j \geq \frac{1}{M_s} \max \left\{ \frac{j^2}{4}, \left(1 + \sqrt{\frac{\mu}{4M_s}} \right)^{2(j-1)} \right\}. \quad (9)$$

The next proposition summarizes the main properties of the ACG method that will be need in our analysis.

Proposition 2.1. *Let $\{(A_j, x_j, u_j, \eta_j)\}$ be the sequence generated by the ACG method applied to (8), where (ψ_s, ψ_n) is a given pair of data functions satisfying **(A1)** and **(A2)** with $4M_s \geq \mu > 0$. Then, the following statements hold:*

a) for every $j \geq 1$, we have $u_j \in \partial_{\eta_j}(\psi_s + \psi_n)(x_j)$;

b) for any $\tilde{\sigma} > 0$, the ACG method obtains a triple $(x, u, \eta) = (x_j, u_j, \eta_j)$ satisfying

$$u \in \partial_\eta(\psi_s + \psi_n)(x) \quad \|u\|^2 + 2\eta \leq \tilde{\sigma}^2 \|x_0 - x + u\|^2 \quad (10)$$

in at most

$$\left\lceil 1 + \sqrt{\frac{M_s}{\mu}} \log_1^+ \left((1 + \tilde{\sigma}^{-1}) \sqrt{2M_s} \right) \right\rceil \quad (11)$$

iterations.

Proof: The inclusion in (a) follows immediately from [22, Proposition 8(c)]. The proof of (b) follows from the first statement in [22, Lemma 9] and by noting that if j is larger than or equal to the number in (11), then $A_j \geq 2(1 + \tilde{\sigma}^{-1})^2$, in view of (9). It should be noted that the parameter $\tilde{\sigma}$ in (10) corresponds to $\sqrt{\sigma}$ in [22, Proposition 8 and Lemma 9]. ■

3 The IAPIAL method

This section is divided into three subsections. The first one discusses the problem of interest, describes the main assumptions made on it, and outlines the IAPIAL method. Subsection 3.2 presents the S-IAPIAL method and its main iteration-complexity result. Subsection 3.3 presents the IAPIAL method and its overall ACG iteration-complexity result.

3.1 Problem of interest, assumptions and IAPIAL outline

This subsection describes the problem of interest, the assumptions made on it, and the type of approximate stationary solution we are interested in computing for it. It also provides an outline for the IAPIAL methods discussed in detail in Subsections 3.2 and 3.3.

The main problem of interest in this paper is (1) where $f, h : \mathfrak{R}^n \rightarrow (-\infty, \infty]$, $A : \mathfrak{R}^n \rightarrow \mathfrak{R}^l$ and $b \in \mathfrak{R}^l$ satisfy the following assumptions:

- (B1) A is a nonzero linear operator;
- (B2) $h \in \overline{\text{Conv}}(\mathfrak{R}^n)$ is L_h -Lipschitz continuous on $\mathcal{H} := \text{dom } h$;
- (B3) the diameter $D := \sup\{\|z - z'\| : z, z' \in \mathcal{H}\}$ of \mathcal{H} is finite and there exists $\nabla_f \geq 0$ such that $\|\nabla f(z)\| \leq \nabla_f$ for every $z \in \mathcal{H}$;
- (B4) there exists $\bar{z} \in \text{int}(\mathcal{H})$ such that $A\bar{z} = b$;
- (B5) f is nonconvex and differentiable on \mathcal{H} , and there exist $L_f \geq m_f > 0$ such that, for all $z, z' \in \mathcal{H}$,

$$\|\nabla f(z') - \nabla f(z)\| \leq L_f \|z' - z\|, \quad (12)$$

$$f(z') - \ell_f(z'; z) \geq -\frac{m_f}{2} \|z' - z\|^2. \quad (13)$$

Some comments about assumptions (B1)-(B5) are in order. First, it is shown in Lemma A.3 that $\partial_\varepsilon h(z) \subset \bar{B}(0, L_h) + N_{\mathcal{H}}^\varepsilon(z)$ for every $z \in \mathcal{H}$. This inclusion will be used to bound the sequence of Lagrangian multipliers generated by the IAPIAL method. Second, it is well known that (12) implies that $|f(z') - \ell_f(z'; z)| \leq L_f \|z' - z\|^2/2$ for every $z, z' \in \mathcal{H}$, and hence that (13) holds with

$m_f = L_f$. However, better iteration-complexity bounds can be derived when a scalar $m_f < L_f$ satisfying (13) is available. Third, (13) implies that the function $f(\cdot) + m_f \|\cdot\|^2/2$ is convex on \mathcal{H} . Moreover, since f is nonconvex on \mathcal{H} in view of **(B5)**, the smallest m_f satisfying (13) is positive. Finally, the existence of a scalar ∇_f as in **(B3)** is actually not an extra assumption since, using (12) and the boundedness of \mathcal{H} in **(B3)**, it can be easily seen that for any $y \in \mathcal{H}$, the scalar $\nabla_f = \nabla_{f,y} := \|\nabla f(y)\| + L_f D$ majorizes $\|\nabla f(z)\|$ for any $z \in \mathcal{H}$.

It is well known that, under some mild conditions, if \bar{z} is a local minimum of (1), then there exists $\bar{p} \in \mathfrak{R}^l$ such that (\bar{z}, \bar{p}) is a stationary solution of (1), i.e.,

$$0 \in \nabla f(\bar{z}) + \partial h(\bar{z}) + A^* \bar{p}, \quad A \bar{z} - b = 0. \quad (14)$$

The main complexity results of this paper are stated in terms of the following notion of approximate stationary solution which is a natural relaxation of (14).

Definition 3.1. *Given a tolerance pair $(\hat{\rho}, \hat{\eta}) \in \mathfrak{R}_{++} \times \mathfrak{R}_{++}$, a triple $(\hat{z}, \hat{p}, \hat{w}) \in \mathcal{H} \times \mathfrak{R}^l \times \mathfrak{R}^n$ is said to be a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1) if it satisfies (2).*

We have outlined the IAPIAL method in Section 1 as one which repeatedly performs the following iteration: i) find a pair (z_k, w_k) as in (4)-(5) and; ii) perform an auxiliary test to decide whether to update c as $c \leftarrow 2c$ or leave it as is. Hence, the IAPIAL iterations can be grouped into cycles where each cycle consists of all iterations with a common penalty parameter value. Subsection 3.2 considers a single cycle of IAPIAL, referred there as S-IAPIAL, and studies its main properties. More specifically, Theorem 3.3 shows that: i) there exists a threshold value

$$\bar{c} = \Theta \left(\max \left\{ \frac{1}{\hat{\rho}^2}, \frac{1}{\hat{\eta}} \right\} \right) \quad (15)$$

such that S-IAPIAL with penalty parameter c satisfying $\bar{c} \leq c = \Theta(\bar{c})$ obtains a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1) in at most $\mathcal{O}(\max\{1/\hat{\rho}^3, 1/(\sqrt{\hat{\eta}}\hat{\rho}^2)\} \log_1^+(\max\{1/\hat{\eta}, 1/\hat{\rho}^2\}))$ ACG iterations; ii) regardless of the value of c , the number of iterations of S-IAPIAL (or within an IAPIAL cycle) is still under control due to the use of the auxiliary test for stopping S-IAPIAL (or as a rule for terminating the IAPIAL cycle). Even though i) implies that S-IAPIAL can solve (2) under the aforementioned condition on c (e.g., with $c = \bar{c}$), the constant \bar{c} is hard or impossible to compute exactly. IAPIAL, outlined in Section 1 and further studied in Subsection 3.3 is essentially a dynamic version of S-IAPIAL in which a line search on c is introduced. More specifically, Subsection 3.3 states and analyzes the IAPIAL method with all its cycles juxtaposed next to one another instead of just a single cycle as in S-IAPIAL. It is shown in Theorem 3.4 that IAPIAL obtains a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1) within the same iteration-complexity as that of S-IAPIAL with its constant penalty parameter c satisfying $\bar{c} \leq c = \Theta(\bar{c})$ where \bar{c} is as in (15).

The desired approximate solution z_k as in (4) is computed by applying the ACG method of Section 2 to subproblem (4). More specifically, applying the ACG method of Section 2 with initial point $x_0 = z_{k-1}$ and composite structure (ψ_s, ψ_n) given by

$$\psi_s = \lambda \left(f + \langle p_{k-1}, A \cdot -b \rangle + \frac{c}{2} \|A \cdot -b\|^2 \right) + \frac{1}{4} \|\cdot - z_{k-1}\|^2, \quad \psi_n = \lambda h + \frac{1}{4} \|\cdot - z_{k-1}\|^2, \quad (16)$$

a triple $(z_k, v_k, \varepsilon_k)$ satisfying

$$v_k \in \partial_{\varepsilon_k} \left(\lambda \mathcal{L}_c(\cdot, p_{k-1}) + \frac{1}{2} \|\cdot - z_{k-1}\|^2 \right) (z_k), \quad \|v_k\|^2 + 2\varepsilon_k \leq \tilde{\sigma}^2 \|r_k\|^2 \quad (17)$$

where r_k is defined as

$$r_k := v_k + z_{k-1} - z_k, \quad (18)$$

is guaranteed to be obtained in a number of iterations bounded by (11) with $M_s := \lambda(L_f + c\|A\|^2) + 1/2$ and $\mu := 1/2$, as long as ψ_s is convex. In view of the above definition of ψ_s , a sufficient condition for it to be convex is that $\lambda \leq 1/(2m_f)$. IAPIAL then sets $\lambda = 1/(2m_f)$ since the larger the stepsize λ is, the better its overall performance is. Clearly, (ψ_s, ψ_n) is a composite structure of the objective function of (4). Moreover, if $(z_k, v_k, \varepsilon_k)$ satisfies (17) with $\tilde{\sigma} = 0$, then z_k is the exact solution of (4). Hence, a triple $(z_k, v_k, \varepsilon_k)$ satisfying (17) can be viewed as an approximate solution of (4).

The goal of IAPIAL (and S-IAPIAL with large c) is to find a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1). The following technical result, whose proof is presented in Appendix, shows that: i) a triple $(z_k, v_k, p_k, \varepsilon_k)$ satisfying (17) can be used to construct a triple $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ satisfying the inclusion in (2) with $(\hat{z}, \hat{p}, \hat{w}) = (\hat{z}_k, \hat{p}_k, \hat{w}_k)$, and; ii) $\|\hat{w}_k\|$ can be nicely bounded in terms of $\|r_k\|$ and ε_k , where r_k is as in (18).

Proposition 3.2. *Assume that $(\lambda, z_{k-1}, p_{k-1}) \in \mathfrak{R}_{++} \times \mathfrak{R}^n \times \mathfrak{R}^l$ and the iterate $(z_k, v_k, \varepsilon_k)$ satisfies (17). Moreover, define the function g_k , the residual w_k , and the scalar δ_k as*

$$g_k := f + \langle p_{k-1}, A \cdot -b \rangle + \frac{c}{2} \|A \cdot -b\|^2, \quad (19)$$

$$w_k := \frac{1}{\lambda} [r_k + (\lambda L_c + 1)(z_k - \hat{z}_k)], \quad \delta_k := \frac{\varepsilon_k}{\lambda}, \quad (20)$$

where $L_c := \lambda(L_f + c\|A\|^2)$, r_k is as in (18), and

$$\hat{z}_k := \operatorname{argmin}_u \left\{ \lambda [\langle \nabla g_k(z_k), u - z_k \rangle + h(u)] - \langle r_k, u - z_k \rangle + \frac{\lambda L_c + 1}{2} \|u - z_k\|^2 \right\}. \quad (21)$$

Then, the following statements hold:

a) the quadruple $(z_k, w_k, p_k, \delta_k)$ satisfies

$$w_k \in \nabla f(z_k) + \partial_{\delta_k} h(z_k) + A^* p_k, \quad (22)$$

$$\lambda \|w_k\| \leq \left(1 + \tilde{\sigma} \sqrt{\lambda L_c + 1}\right) \|r_k\|, \quad \delta_k \leq \frac{\tilde{\sigma}^2 \|r_k\|^2}{2\lambda}, \quad (23)$$

where p_k is as in (5).

b) \hat{z}_k and the pair (\hat{w}_k, \hat{p}_k) defined as

$$\hat{w}_k := w_k + \nabla g_k(\hat{z}_k) - \nabla g_k(z_k), \quad \hat{p}_k := p_{k-1} + cA(\hat{z}_k - b) \quad (24)$$

satisfy

$$\hat{w}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^* \hat{p}_k, \quad (25)$$

$$\lambda \|\hat{w}_k\| \leq \left(1 + 2\tilde{\sigma} \sqrt{\lambda L_c + 1}\right) \|r_k\|, \quad \|\hat{z}_k - z_k\| \leq \frac{\tilde{\sigma}^2 \|r_k\|}{\sqrt{\lambda L_c + 1}}. \quad (26)$$

We now make a few remarks about Proposition 3.2. First, the residual w_k in (20) is not used in the description of the methods of this section but it, together with its bound in (23), plays an important role in their analysis. Second, the size of the residual \hat{w}_k on the other hand is used to

monitor the termination of both methods. Third, it is shown that the residual and the feasibility gap sequences $\{\hat{w}_k\}$ and $\{\|A\hat{z}_k - b\|\}$ generated by S-IAPIAL with penalty parameter c satisfy

$$\min_{i \leq k} \|\hat{w}_i\| = \Theta\left(\frac{1}{\sqrt{k}} + \frac{1}{\sqrt{c}}\right), \quad \|A\hat{z}_k - b\| = \Theta\left(\frac{1}{c}\right). \quad (27)$$

Fourth, as IAPIAL forces c to grow, it follows from the third remark and the inclusion in (25) that $(\hat{z}_k, \hat{p}_k, \hat{w}_k)$ will eventually become a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1).

3.2 The S-IAPIAL method

This subsection describes S-IAPIAL, which is essentially the general IAPIAL method outlined in Section 1 (see the paragraph containing (4)-(5)) and Subsection 3.1 (see the paragraph containing (17)-(18)) looked from the perspective of a single cycle, during which the penalty parameter c is kept fixed.

We start by formally describing the S-IAPIAL method.

S-IAPIAL

- 0) Let scalars $\nu > 0$ and $\sigma \in (0, 1/\sqrt{2}]$, initial point $z_0 \in \mathcal{H}$, tolerance pair $(\hat{\rho}, \hat{\eta}) \in \mathfrak{R}_{++} \times \mathfrak{R}_{++}$, and penalty parameter $c > 0$ be given; set $k = 1$, $p_0 = 0$, and

$$\lambda := \frac{1}{2m_f}, \quad L_c := L_f + c\|A\|^2, \quad \sigma_c := \min\left\{\frac{\nu}{\sqrt{\lambda L_c + 1}}, \sigma\right\}; \quad (28)$$

- 1) apply the ACG method with inputs $x_0 := z_{k-1}$, $(M_s, \mu) := (\lambda L_c + 1/2, 1/2)$, and (ψ_s, ψ_n) as in (16), to obtain a triple $(z_k, v_k, \varepsilon_k) \in \mathcal{H} \times \mathfrak{R}^n \times \mathfrak{R}_+$ satisfying (17) with $\tilde{\sigma} = \sigma_c$;
- 2) compute $(\hat{z}_k, \hat{w}_k, \hat{p}_k)$ as in (21) and (24); if $\|\hat{w}_k\| \leq \hat{\rho}$ and $\|A\hat{z}_k - b\| \leq \hat{\eta}$, then stop with success and return $(\hat{z}, \hat{w}, \hat{p}) = (\hat{z}_k, \hat{w}_k, \hat{p}_k)$; else, go to step 3;
- 3) if $k \geq 2$ and

$$\Delta_k := \frac{1}{k-1} [\mathcal{L}_c(z_1, p_1) - \mathcal{L}_c(z_k, p_k)] \leq \frac{\lambda(1 - \sigma^2)\hat{\rho}^2}{4(1 + 2\nu)^2}, \quad (29)$$

then stop with failure, declare c small and return $(\hat{z}, \hat{w}, \hat{p}) = (\hat{z}_k, \hat{w}_k, \hat{p}_k)$; otherwise, set

$$p_k = p_{k-1} + c(Az_k - b) \quad (30)$$

and $k \leftarrow k + 1$, and go to step 1.

The S-IAPIAL method performs two types of iterations, namely, the outer ones indexed by k and the ACG ones performed during its calls to the ACG method in step 1.

We now make a few remarks about S-IAPIAL. First, the scalar λ defined in step 0 ensures that the prox augmented Lagrangian subproblem (4) is strongly convex. Second, the scalars M_s and μ in step 1 are the Lipschitz constant and the strongly convexity parameter of $\nabla\psi_s$ and ψ_n , respectively, where (ψ_s, ψ_n) is as in (16). Third, (30) is performing a full Lagrange multiplier update in that the multiplier factor of the second term $c(Az_k - b)$ of its right hand side is equal to one. Fourth, Theorem 3.3(c) shows that for sufficiently large penalty parameter c , the S-IAPIAL method successfully stops in step 2, and its output is a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of

(1). Fifth, the stopping criterion in step 3 is a way of detecting that the penalty parameter is small. In this case, S-IAPIAL fails to obtain a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1) but gains the valuable information that the current value of c is small. The general method IAPIAL discussed in the next subsection then uses this information to increase c and restart S-IAPIAL with the new value of c and with the initial point z_0 either set to be the same as in the previous S-IAPIAL call, i.e., z_0 is kept constant (cold S-IAPIAL restart), or set to be equal to z_k where z_k is the iterate computed in step 3 of S-IAPIAL just before it failed (warm S-IAPIAL restart).

We are now ready to state the main result about S-IAPIAL whose proof is given at the end of Section 4.

Theorem 3.3. *Assume that conditions (B1)–(B5) hold. Define*

$$\phi_* := \inf_{z \in \mathfrak{R}^n} \phi(z), \quad \bar{d} := \text{dist}_{\partial\mathcal{H}}(\bar{z}), \quad (31)$$

$$R_* := \phi^* - \phi_* + \frac{D^2}{\lambda}, \quad \kappa_0 := \frac{1}{\sigma_A^+} \left[2(L_h + \nabla_f)D + \left(\frac{2(1+\nu)}{1-\sigma} + \frac{\sigma^2}{2(1-\sigma)^2} \right) \frac{D^2}{\lambda} \right], \quad (32)$$

where ϕ^* is as in (1), D and ∇_f are as in (B3), and \bar{z} is as in (B4). Then, the following statements about the S-IAPIAL method hold:

a) at each outer iteration, it performs at most

$$\left\lceil 1 + \sqrt{T_c} \log_1^+ \left(\frac{2T_c}{\min\{\nu, \sigma\}} \right) \right\rceil \quad (33)$$

ACG iterations, where $T_c := 2\lambda L_c + 1$ and L_c is as in (28);

b) it performs a finite number of outer iterations which is bounded by

$$\left\lceil 1 + \frac{4(1+2\nu)^2}{(1-\sigma^2)\lambda\hat{\rho}^2} \left(3R_* + \frac{\kappa_0^2}{2\bar{d}^2c} \right) \right\rceil; \quad (34)$$

c) if the penalty parameter c is such that

$$c \geq \bar{c} := \max \left\{ \frac{\kappa_1}{\hat{\rho}^2}, \frac{\kappa_2}{\hat{\eta}} \right\}, \quad (35)$$

where

$$\kappa_1 := \frac{32(1+2\nu)^2 \kappa_0^2}{\lambda(1-\sigma^2)\bar{d}^2}, \quad \kappa_2 := \left(\frac{2\kappa_0}{\bar{d}} + \frac{\nu D}{\lambda\|A\|(1-\sigma)} \right), \quad (36)$$

then: i) every iterate $(\hat{z}_k, \hat{w}_k, \hat{p}_k)$ generated by S-IAPIAL satisfies $\|A\hat{z}_k - b\| \leq \hat{\eta}$ and the inclusion in (25); and ii) S-IAPIAL successfully stops in its step 2 with a triple $(\hat{z}, \hat{w}, \hat{p}) = (\hat{z}_k, \hat{w}_k, \hat{p}_k)$ which is a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1).

We now make some remarks about the complexities described in Theorem 3.3 regarding their dependencies on the penalty parameter c and the tolerances $\hat{\rho}$ and $\hat{\eta}$. First, since $T_c = \mathcal{O}(c)$, Theorem 3.3(a) shows that the ACG method invoked in step 2 performs at most $\mathcal{O}(\sqrt{c} \log_1^+(c))$ iterations to compute a triple $(z_k, v_k, \varepsilon_k)$ satisfying (17). Second, Theorem 3.3(b) shows that S-IAPIAL performs $\mathcal{O}((1+c^{-1})/\hat{\rho}^2)$ outer iterations. Third, it follows from Theorem 3.3 that, under the condition that $\bar{c} \leq c = \Theta(\bar{c})$, S-IAPIAL obtains a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1) in at most

$$\mathcal{O} \left(\max \left\{ \frac{1}{\hat{\rho}^3}, \frac{1}{\sqrt{\hat{\eta}}\hat{\rho}^2} \right\} \log_1^+ \left(\max \left\{ \frac{1}{\hat{\eta}}, \frac{1}{\hat{\rho}^2} \right\} \right) \right)$$

ACG iterations.

3.3 The IAPIAL method

This subsection describes the IAPIAL method and presents the main result of this paper which establishes its ACG iteration-complexity.

The statement of IAPIAL below makes use of S-IAPIAL presented in Subsection 3.2. More specifically, it consists of repeatedly invoking S-IAPIAL with $c = c_\ell := c_1 2^{\ell-1}$ where c_1 is an initial choice for the penalty parameter and ℓ is the S-IAPIAL call count.

IAPIAL method

- (0) Let a pair of scalars $(\nu, \sigma) \in \mathfrak{R}_{++} \times (0, 1/\sqrt{2}]$, and a pair of tolerances $(\hat{\rho}, \hat{\eta}) \in \mathfrak{R}_{++} \times \mathfrak{R}_{++}$ be given, choose an initial penalty parameter $c_1 > 0$ and set $c = c_1$;
 - (1) choose an initial point $z_0 \in \mathcal{H}$ and let $(\hat{z}, \hat{w}, \hat{p})$ be the output obtained by the S-IAPIAL method with input $(z_0, \nu, \sigma, c, \hat{\rho}, \hat{\eta})$;
 - (2) if $\|\hat{w}\| \leq \hat{\rho}$ and $\|A\hat{z} - b\| \leq \hat{\eta}$, stop and output $(\hat{z}, \hat{w}, \hat{p})$; otherwise, set $c \leftarrow 2c$ and return to step 1.
-

We now make a remark about IAPIAL. The point z_0 chosen in step 1 as being the initial point for S-IAPIAL can always be chosen the same point (cold start) or a varying point. In the latter case, a common sense approach (hybrid warm start) is to choose z_0 to be the point \hat{z} output at the previous iteration.

The following result establishes the overall ACG iteration-complexity for the IAPIAL method to obtain a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1).

Theorem 3.4. *Assume that conditions (B1)–(B5) of Subsection 3.1 hold. Then, the IAPIAL method obtains a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution $(\hat{z}, \hat{w}, \hat{p})$ of problem (1) in at most*

$$\mathcal{O} \left(\left[\sqrt{\frac{T_1}{m_f}} + \sqrt{\frac{L_f}{m_f}} \log_1^+ \left(\frac{T_1}{c_1} \right) \right] \left[1 + \frac{m_f}{\hat{\rho}^2} \max \left\{ R_*, \frac{\kappa_0^2}{d^2 c_1} \right\} \right] \log_1^+(T_2) \right) \quad (37)$$

ACG iterations, where R_* and κ_0 are as in (32), and

$$T_1 := \max \left\{ c_1, \frac{m_f \kappa_0^2}{d^2 \hat{\rho}^2}, \frac{\kappa_0}{d \hat{\eta}} \right\}, \quad T_2 := \frac{1}{m_f} (L_f + T_1 \|A\|^2). \quad (38)$$

Proof: First note that the l th loop of the IAPIAL method invokes S-IAPIAL with penalty parameter $c = c_l$ where $c_l := 2^{l-1} c_1$, for every $l \geq 1$. Hence, in view of Theorem 3.3(c) and the stopping criterion in step 2 of IAPIAL, we conclude that it obtains a $(\hat{\rho}, \hat{\eta})$ -approximate solution of (1) in at most \bar{l} iterations, where

$$\bar{l} := \min \{ l : c_l \geq \bar{c} \} \quad (39)$$

and \bar{c} is as in (35). Moreover, it follows by Theorem 3.3(a) and (b) that the total number of ACG iterations performed by IAPIAL is bounded by

$$\left(\sum_{l=1}^{\bar{l}} \left[1 + \sqrt{T_{c_l}} \log_1^+ \left(\frac{2T_{c_l}}{\min\{\nu, \sigma\}} \right) \right] \right) \left[1 + \frac{4(1+2\nu)^2}{(1-\sigma^2)\lambda \hat{\rho}^2} \left(3R_* + \frac{\kappa_0^2}{2d^2 c_1} \right) \right], \quad (40)$$

where $T_{c_l} = 2\lambda L_{c_l} + 1$. In view of the above definition of c_l and (39), we have

$$c_l \leq \max \{c_1, 2\bar{c}\}, \quad \forall l = 1, \dots, \bar{l}. \quad (41)$$

Hence, it follows from the definition of L_{c_l} (see (28) with $c = c_l$) that

$$T_{c_l} = 2\lambda (L_f + c_l \|A\|^2) + 1 \leq 2\lambda (L_f + \max \{c_1, 2\bar{c}\} \|A\|^2) + 1 = \mathcal{O}(T_2), \quad (42)$$

where the last relation is due to the definitions of λ , \bar{c} , T_1 , and T_2 given in (28), (35), and (38). It also follows from the definitions of c_l , T_{c_l} , and L_{c_l} that

$$\begin{aligned} \sum_{l=1}^{\bar{l}} \sqrt{T_{c_l}} &= \sum_{l=1}^{\bar{l}} \sqrt{[2\lambda (L_f + 2^{l-1} c_1 \|A\|^2) + 1]} \leq \sum_{l=1}^{\bar{l}} \left(\sqrt{2\lambda L_f + 1} + \sqrt{\lambda c_1} \|A\| \sqrt{2^l} \right) \\ &\leq \bar{l} \sqrt{2\lambda L_f + 1} + 8\sqrt{\lambda c_1} \|A\|. \end{aligned}$$

From the above inequalities, (41) and the definition of \bar{l} in (39), we have

$$\begin{aligned} \sum_{l=1}^{\bar{l}} \sqrt{T_{c_l}} &\leq 8\sqrt{2\lambda} \|A\| \sqrt{\max \{c_1, 2\bar{c}\}} + \sqrt{2\lambda L_f + 1} \log_1^+ \left(\frac{2 \max \{c_1, 2\bar{c}\}}{c_1} \right) \\ &= \mathcal{O} \left(\sqrt{\frac{T_1}{m_f}} + \sqrt{\frac{L_f}{m_f}} \log_1^+ \left(\frac{T_1}{c_1} \right) \right), \end{aligned}$$

where the last relation is due to the definitions of λ , \bar{c} , and T_1 given in (28), (35), and (38), respectively. Hence, (37) follows by combining the latter estimate, definition of λ , (40), (42), and by noting that $\sqrt{T_{c_l}} \log_1^+(t) \geq 1$ for all $t > 0$ and $l = 1, \dots, \bar{l}$. \blacksquare

We now make some remarks about Theorem 3.4. First, its iteration-complexity does not depend on how z_0 is selected in step 0. As a consequence, it applies to both the cold start and the hybrid warm start approaches mentioned above. Second, it follows from Theorem 3.4 that the ACG iteration-complexity of IAPIAL expressed only in terms of the tolerance pair $(\hat{\rho}, \hat{\eta})$ and assuming $c_1 = \mathcal{O}(1)$ is

$$\mathcal{O} \left(\max \left\{ \frac{1}{\hat{\rho}^3}, \frac{1}{\sqrt{\hat{\eta}} \hat{\rho}^2} \right\} \log_1^+ \left(\max \left\{ \frac{1}{\hat{\eta}}, \frac{1}{\hat{\rho}^2} \right\} \right) \right)$$

and hence, its ACG iteration complexity is essentially the same as that of S-IAPIAL with a large penalty parameter $c = \Theta(\bar{c})$, where \bar{c} is as in (35).

4 Technical Results and Proof of Theorem 3.3

This section provides the proof of Theorem 3.3 regarding the behavior of S-IAPIAL.

The first lemma below describes how the Lagrangian function $\mathcal{L}_c(\cdot, \cdot)$ varies from one iteration to the next one.

Lemma 4.1. *Let $\{(z_k, v_k, p_k, \varepsilon_k)\}$ be generated by S-IAPIAL, let $\{r_k\}$ be as in (18), and define $\{\Delta p_k\}$ as*

$$\Delta p_k := p_k - p_{k-1}, \quad \forall k \geq 1. \quad (43)$$

Then, the following inequality holds for every $k \geq 1$:

$$\|r_k\|^2 \leq \frac{2\lambda}{1 - \sigma_c^2} \left(\mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_k) + \frac{1}{c} \|\Delta p_k\|^2 \right). \quad (44)$$

Proof: Using (30) and the definitions of \mathcal{L}_c and Δp_k given in (3) and (43), respectively, we have

$$\mathcal{L}_c(z_k, p_k) - \mathcal{L}_c(z_k, p_{k-1}) = \langle \Delta p_k, Az_k - b \rangle = \left\langle \Delta p_k, \frac{p_k - p_{k-1}}{c} \right\rangle = \frac{1}{c} \|\Delta p_k\|^2. \quad (45)$$

Now, it follows from (17), (18), and the Cauchy-Schwarz inequality that

$$\begin{aligned} \lambda \mathcal{L}_c(z_k, p_{k-1}) - \lambda \mathcal{L}_c(z_{k-1}, p_{k-1}) &\leq -\frac{1}{2} \|z_k - z_{k-1}\|^2 + \langle v_k, z_k - z_{k-1} \rangle + \varepsilon_k \\ &= -\frac{1}{2} \|v_k + z_k - z_{k-1}\|^2 + \frac{\|v_k\|^2}{2} + \varepsilon_k \leq -\frac{1 - \sigma_c^2}{2} \|r_k\|^2, \end{aligned}$$

which implies that

$$\frac{1 - \sigma_c^2}{2\lambda} \|r_k\|^2 \leq \mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_{k-1}).$$

The inequality in (44) then follows by combining the latter inequality with (45). \blacksquare

Recall that Proposition 3.2(b) implies that the triple $(\hat{z}, \hat{w}, \hat{p}) = (\hat{z}_k, \hat{w}_k, \hat{p}_k)$ satisfies the inclusion in (2). The following result gives a preliminary bound on the residual \hat{w}_k , which measures the quality of the above triple as a $(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1) (see Definition 3.1).

Lemma 4.2. *Consider the sequences $\{(z_k, v_k, p_k, \varepsilon_k)\}$ and $\{(\hat{z}_k, \hat{w}_k)\}$ generated by S-IAPIAL and define*

$$C_1 := \frac{2(1 + 2\nu)^2}{1 - \sigma^2}, \quad (46)$$

where ν and σ are as in step 0 of S-IAPIAL. Then, the following statements hold:

a) for every $k \geq 1$, we have

$$\|\hat{w}_k\|^2 \leq \frac{C_1}{\lambda} \left(\mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_k) + \frac{1}{c} \|\Delta p_k\|^2 \right), \quad (47)$$

where Δp_k is as in (43).

b) for every $k \geq 2$, we have

$$\min_{i \leq k} \|\hat{w}_i\|^2 \leq \frac{C_1}{\lambda} \left(\Delta_k + \frac{4}{c(k-1)} \sum_{i=1}^k \|p_i\|^2 \right) \quad (48)$$

where Δ_k is as in (29).

Proof: a) It follows from step 1 of S-IAPIAL that $(\lambda, z_{k-1}, p_{k-1})$ and $(z_k, v_k, \varepsilon_k)$ satisfy (17) with $\tilde{\sigma} = \sigma_c$. Hence, it follows from Proposition 3.2(b) that the triple $(\hat{z}_k, \hat{w}_k, \hat{p}_k)$ computed in step 2 of satisfies, in particular, the first inequality in (26) with $\tilde{\sigma} = \sigma_c$. This conclusion, together with the definition of σ_c in (28) and inequality (44), then imply that

$$\|\hat{w}_k\|^2 \leq \frac{(1 + 2\nu)^2 \|r_k\|^2}{\lambda^2} \leq \frac{2(1 + 2\nu)^2}{\lambda(1 - \sigma_c^2)} \left(\mathcal{L}_c(z_{k-1}, p_{k-1}) - \mathcal{L}_c(z_k, p_k) + \frac{1}{c} \|\Delta p_k\|^2 \right),$$

and hence that (47) holds, in view of the definition of C_1 in (46) and the fact that $\sigma_c \leq \sigma$.

b) Summing inequality (47) from $k = 2$ to $k = k$, and using the definition of Δ_k given in (29), we obtain

$$(k-1) \min_{i \leq k} \|\hat{w}_i\|^2 \leq \sum_{i=2}^k \|\hat{w}_i\|^2 \leq \frac{C_1}{\lambda} \left((k-1)\Delta_k + \sum_{i=2}^k \frac{\|\Delta p_i\|^2}{c} \right).$$

Inequality (48) now follows from the previous inequality and the fact that the definition of Δp_i in (43) implies that $\|\Delta p_i\|^2 = \|p_i - p_{i-1}\|^2 \leq 2\|p_i\|^2 + 2\|p_{i-1}\|^2$. \blacksquare

We now provide an outline of the technical results that follow in view of bound (48). The preliminary bound (48) on $\min_{i \leq k} \|\hat{w}_i\|^2$ is the sum of two terms, one of which depends on Δ_k . The goal of the next two results, i.e., Lemmas 4.3 and 4.4, is to obtain a more explicit bound on Δ_k , which essentially implies that Δ_k behaves as $\mathcal{O}(1/k)$ as long as $\|p_k\|$ can be shown to be bounded. Proposition 4.9, which is established with the help of the technical results preceding it, namely, Lemmas 4.5-4.8, then show that $\|p_k\|$ is indeed bounded. Note that bound (48) and the above two observations will actually imply that $\min_{i \leq k} \|\hat{w}_i\|^2$ behaves as $\mathcal{O}(c^{-1} + k^{-1})$, as has already been previewed in (27).

The following result shows that the value $\mathcal{L}_c(z_1, p_1)$ of the Lagrangian function $\mathcal{L}_c(\cdot, \cdot)$ at the first iterate (z_1, p_1) can be majorized by a scalar which does not depend on c . This fact is not immediately apparent from the definition of $\mathcal{L}_c(\cdot, \cdot)$ and plays an important role in showing that S-IAPIAL or IAPIAL can start from an arbitrary (and hence infeasible) point in \mathcal{H} .

Lemma 4.3. *The first quadruple $(z_1, v_1, p_1, \varepsilon_1)$ generated by S-IAPIAL satisfies*

$$\mathcal{L}_c(z_1, p_1) \leq 3R_* + \phi_*, \quad (49)$$

where ϕ_* and R_* are as in (31) and (32), respectively.

Proof: The fact that $(z_1, v_1, \varepsilon_1)$ satisfies (17) with $\tilde{\sigma} = \sigma_c$ and $k = 1$, Lemma A.2 with $s = 1$ and $\tilde{\phi} = \lambda \mathcal{L}_c(\cdot, p_0)$, and condition **(B3)**, imply that for every $z \in \mathcal{H}$,

$$\lambda \mathcal{L}_c(z_1, p_0) + \frac{1 - 2\sigma_c^2}{2} \|r_1\|^2 \leq \lambda \mathcal{L}_c(z, p_0) + \|z - z_0\|^2 \leq \lambda \mathcal{L}_c(z, p_0) + D^2,$$

where r_k is as in (18). Using the definition of ϕ^* in (1), the fact that $1 - 2\sigma_c^2 \geq 1 - 2\sigma^2 \geq 0$ due to step 0 of IAPIAL, and the fact that the definition of \mathcal{L}_c in (3) implies that $\mathcal{L}_c(z, p_0) = (f + h)(z)$ for every $z \in \mathcal{F} := \{z \in \mathcal{H} : Az = b\}$, we then conclude from the above inequality, as z varies in \mathcal{F} , that

$$\mathcal{L}_c(z_1, p_0) \leq \phi^* + \frac{D^2}{\lambda} = R_* + \phi_*,$$

where the last equality is due to the definition of R_* in (32). The above inequality together with the fact that $p_0 = 0$, update (30) with $k = 1$, and the definitions of \mathcal{L}_c and ϕ_* as in (3) and (31), respectively, then imply that

$$\mathcal{L}_c(z_1, p_1) = \mathcal{L}_c(z_1, p_0) + c\|Az_1 - b\|^2 = 3\mathcal{L}_c(z_1, p_0) - 2(f + h)(z_1) \leq 3(R_* + \phi_*) - 2\phi_*,$$

and hence that the lemma holds. \blacksquare

The following result shows that $\Delta_k = \mathcal{O}(1/k)$ as long as $\|p_k\| = \mathcal{O}(1)$.

Lemma 4.4. *Let $\{(z_k, v_k, p_k, \varepsilon_k)\}$ be generated by S-IAPIAL and consider the sequence $\{\Delta_k\}$ as in (29). Then, the following statements hold:*

a) for every $k \geq 1$, we have

$$\mathcal{L}_c(z_k, p_k) \geq \phi_* - \frac{\|p_k\|^2}{2c}, \quad (50)$$

where ϕ_* is as in (31);

b) for every $k \geq 2$, we have

$$\Delta_k \leq \frac{1}{k-1} \left(3R_* + \frac{\|p_k\|^2}{2c} \right), \quad (51)$$

where R_* is as in (32).

Proof: (a) Using the definitions of \mathcal{L}_c and ϕ_* given in (3) and (31), respectively, we have

$$\begin{aligned} \mathcal{L}_c(z_k, p_k) &= (f + h)(z_k) + \langle p_k, Az_k - b \rangle + \frac{c}{2} \|Az_k - b\|^2 \\ &\geq \phi_* + \frac{1}{2} \left\| \frac{p_k}{\sqrt{c}} + \sqrt{c}(Az_k - b) \right\|^2 - \frac{1}{2c} \|p_k\|^2, \end{aligned}$$

and hence that (50) holds.

(b) This statement follows from (49), (50), and the definition of Δ_k in (29). \blacksquare

The next technical results (i.e., Lemmas 4.5-4.8) develop the necessary tools for showing in Proposition 4.9 that the sequence $\{p_k\}$ is bounded. The first one gives some straightforward bounds among the different quantities involved in the analysis of S-IAPIAL.

Lemma 4.5. *Let $\{(z_k, v_k, p_k, \varepsilon_k)\}$ be generated by S-IAPIAL and let $\{r_k\}$ be as in (18). Then, the following inequalities hold for every $k \geq 1$,*

$$\|r_k\| \leq \frac{D}{1-\sigma}, \quad \|v_k\| \leq \frac{\sigma D}{1-\sigma}, \quad \varepsilon_k \leq \frac{\sigma^2 D^2}{2(1-\sigma)^2} \quad (52)$$

where D is as in **(B3)** and σ is as in step 0 of S-IAPIAL.

Proof: First note that, in view of step 1 of S-IAPIAL, $(\lambda, z_{k-1}, p_{k-1})$ and $(z_k, v_k, \varepsilon_k)$ satisfy (17) with $\tilde{\sigma} = \sigma_c$. Hence, using the inequality in (17), the definition of r_k given in (18), the triangle inequality, the first condition in **(B3)**, and the fact that $\sigma_c \leq \sigma$, we have

$$\|r_k\| - D \leq \|r_k\| - \|z_k - z_{k-1}\| \leq \|v_k\| \leq \sigma \|r_k\|, \quad \varepsilon_k \leq \frac{\sigma^2 \|r_k\|^2}{2}. \quad (53)$$

The first inequality in (52) immediately follows from the first setting of inequalities in (53). The last two inequalities in (52) follow from the first inequality in (52) and the last two inequalities in (53). \blacksquare

The next result defines a slack $\xi_k \in \partial_{\delta_k} h(z_k)$ which realizes the inclusion in (22) and gives a preliminary bound on $\|p_k\|$ in terms of $\|\xi_k\|$.

Lemma 4.6. *Consider the sequence $\{(z_k, v_k, p_k, \varepsilon_k)\}$ generated by S-IAPIAL and the sequence $\{(w_k, \delta_k)\}$ as in (20), and define*

$$\xi_k := w_k - \nabla f(z_k) - A^* p_k \quad (54)$$

for every $k \geq 1$. Then, the following statements hold:

a) for every $k \geq 1$, we have

$$\xi_k \in \partial_{\delta_k} h(z_k), \quad \|w_k\| \leq \frac{(1+\nu)D}{\lambda(1-\sigma)}, \quad \delta_k \leq \frac{\sigma^2 D^2}{2\lambda(1-\sigma)^2} \quad (55)$$

where D is as in **(B3)**, and ν and σ are as in step 0 of S-IAPIAL;

b) for every $k \geq 1$, we have

$$\sigma_A^+ \|p_k\| \leq \|\xi_k\| + \nabla_f + \frac{(1+\nu)D}{\lambda(1-\sigma)}, \quad (56)$$

where σ_A^+ is defined in Subsection 1.1 and ∇_f is as in **(B3)**.

Proof: (a) It follows from step 1 of S-IAPIAL that $(\lambda, z_{k-1}, p_{k-1})$ and $(z_k, v_k, \varepsilon_k)$ satisfy (17) with $\tilde{\sigma} = \sigma_c$. Hence, it follows from Proposition 3.2(a) that (22) and (23) hold. The inclusion in (55) follows from (22) and the definition of ξ_k in (54). The inequalities in (55) follow from the inequalities in (23) with $\tilde{\sigma} = \sigma_c$, the definition of σ_c in (28), and the first inequality in (52).

(b) Using **(B4)**, the fact that $p_0 = 0$ (see step 0 of S-IAPIAL), and the update formula (30), it is easy to see that $\{p_k\} \subset A(\mathbb{R}^n)$. Using Lemma A.1, relation (54), the triangle inequality, the second condition in **(B3)**, and the first inequality in (55), we conclude that

$$\sigma_A^+ \|p_k\| \leq \|A^* p_k\| \leq \|\xi_k\| + \|\nabla f(z_k)\| + \|w_k\| \leq \|\xi_k\| + \nabla_f + \frac{(1+\nu)D}{\lambda(1-\sigma)}, \quad \forall k \geq 1,$$

and hence that (56) holds. \blacksquare

The next technical result essentially allows us to obtain a preliminary bound on $\|\xi_k\|$ under assumption **(B4)**.

Lemma 4.7. *Let h be a function as in **(B2)**. Then, for every $z, z' \in \mathcal{H}$, $\varepsilon > 0$, and $\xi \in \partial_\varepsilon h(z)$, we have*

$$\|\xi\| \text{dist}_{\partial\mathcal{H}}(z') \leq (\text{dist}_{\partial\mathcal{H}}(z') + \|z - z'\|) L_h + \langle \xi, z - z' \rangle + \varepsilon,$$

where $\partial\mathcal{H}$ denotes the boundary of \mathcal{H} .

Proof: Let $\varepsilon > 0$, $z, z' \in \mathcal{H}$ and $\xi \in \partial_\varepsilon h(z)$ be given. It follows from the Lipschitz continuity of h in **(B2)** combined with the equivalence between (a) and (d) of Lemma A.3 that there exist $\xi_1 \in \bar{B}(0, L_h)$ and $\xi_2 \in N_{\mathcal{H}}^\varepsilon(z)$ such that $\xi = \xi_1 + \xi_2$. Clearly, it follows from the definitions of $\bar{B}(0, L_h)$ and $N_{\mathcal{H}}^\varepsilon(z)$ in Subsection 1.1 that

$$\|\xi_1\| \leq L_h, \quad \mathcal{H} \subset H_- := \{u \in \mathbb{R}^n : \langle \xi_2, u - z \rangle - \varepsilon \leq 0\}.$$

Using the last inclusion and the fact that $z' \in \mathcal{H}$, we easily see that

$$\text{dist}_{\partial\mathcal{H}}(z') \|\xi_2\| \leq \text{dist}_{\partial H_-}(z') \|\xi_2\| = \langle \xi_2, z - z' \rangle + \varepsilon.$$

The last inequality, the fact that $\xi = \xi_1 + \xi_2$, the triangle inequality, and the Cauchy-Schwarz inequality, then imply that

$$\begin{aligned} \text{dist}_{\partial\mathcal{H}}(z') \|\xi\| &\leq \text{dist}_{\partial\mathcal{H}}(z') \|\xi_1\| + \text{dist}_{\partial\mathcal{H}}(z') \|\xi_2\| \leq \text{dist}_{\partial\mathcal{H}}(z') \|\xi_1\| + \langle \xi_2, z - z' \rangle + \varepsilon \\ &= \text{dist}_{\partial\mathcal{H}}(z') \|\xi_1\| - \langle \xi_1, z - z' \rangle + \langle \xi_2, z - z' \rangle + \varepsilon \\ &\leq (\text{dist}_{\partial\mathcal{H}}(z') + \|z - z'\|) \|\xi_1\| + \langle \xi_2, z - z' \rangle + \varepsilon, \end{aligned}$$

which combined with the fact that $\|\xi_1\| \leq L_h$ shows that the conclusion of the lemma holds. \blacksquare

The next result establishes an important inequality which allow us to relate the size of p_k with that of p_{k-1} (see (57) below).

Lemma 4.8. Consider the sequence $\{(z_k, v_k, p_k, \varepsilon_k)\}$ generated by S-IAPIAL and the sequence $\{\xi_k\}$ as in (54), and let \bar{d} and κ_0 be as in (31) and (32), respectively. Then, for every $k \geq 1$, we have

$$\frac{\|p_k\|^2}{c\sigma_A^+} + \bar{d}\|p_k\| \leq \frac{1}{c\sigma_A^+} \langle p_k, p_{k-1} \rangle + \kappa_0, \quad (57)$$

where σ_A^+ is defined in Subsection 1.1.

Proof: Let \bar{z} be as in (B4) and recall that $\bar{d} = \text{dist}_{\partial\mathcal{H}}(\bar{z})$ in view of (31). Also, note that $\xi_k \in \partial_{\delta_k} h(z_k)$ for every $k \geq 1$, in view of Lemma 4.6(a). Hence, it follows from Lemma 4.7 with $\xi = \xi_k$, $z = z_k$, $z' = \bar{z}$ and $\varepsilon = \delta_k$, assumption (B3), and the last inequality in (52), that

$$\bar{d}\|\xi_k\| \leq (\bar{d} + \|z_k - \bar{z}\|)L_h + \langle \xi_k, z_k - \bar{z} \rangle + \delta_k \leq (\bar{d} + D)L_h + \langle \xi_k, z_k - \bar{z} \rangle + \frac{\sigma^2 D^2}{2\lambda(1-\sigma)^2},$$

which combined with (56) and the fact that $\bar{d} \leq D$ imply that

$$\bar{d}\sigma_A^+\|p_k\| \leq (2L_h + \nabla_f)D + \left[\frac{1+\nu}{1-\sigma} + \frac{\sigma^2}{2(1-\sigma)^2} \right] \frac{D^2}{\lambda} + \langle \xi_k, z_k - \bar{z} \rangle. \quad (58)$$

On the other hand, using (54), the Cauchy-Schwarz inequality, the triangle inequality, and the fact that $A\bar{z} = b$ in view of (B4), we obtain

$$\begin{aligned} \langle \xi_k, z_k - \bar{z} \rangle &= \langle w_k - \nabla f(z_k) - A^*p_k, z_k - \bar{z} \rangle \leq (\|w_k\| + \|\nabla f(z_k)\|)\|z_k - \bar{z}\| - \langle p_k, Az_k - b \rangle \\ &\leq \frac{(1+\nu)D^2}{\lambda(1-\sigma)} + \nabla_f D - \frac{1}{c} \langle p_k, p_k - p_{k-1} \rangle \end{aligned}$$

where the last inequality is due to (B3), (30), and the inequality in (55). Hence, (57) easily follows by combining the above inequality with (58) and by using the definition of κ_0 given in (32). ■

We observe that (57) holds under the weaker assumption that $\bar{z} \in \mathcal{H}$ and $A\bar{z} = b$. However, when $\bar{z} \in \partial\mathcal{H}$, the scalar \bar{d} which appears in (57) becomes zero. Under the stronger assumption (B4), and hence $\bar{d} > 0$, the following result establishes the boundedness of the sequence of Lagrange multipliers $\{p_k\}$.

Proposition 4.9. The sequence $\{p_k\}$ generated by S-IAPIAL satisfies, for every $k \geq 0$,

$$\|p_k\| \leq \frac{\kappa_0}{\bar{d}}, \quad (59)$$

where \bar{d} and κ_0 are as in (31) and (32), respectively.

Proof: The proof is done by induction on k . Since $p_0 = 0$ and $\kappa_0 \geq 0$, (59) trivially holds for $k = 0$. Assume now that (59) holds with $k = k - 1$ for some $k \geq 1$. This assumption together with (57) and the Cauchy-Schwarz inequality, then imply that

$$\left(\frac{\|p_k\|}{c\sigma_A^+} + \bar{d} \right) \|p_k\| \leq \frac{\|p_k\|\|p_{k-1}\|}{c\sigma_A^+} + \kappa_0 \leq \frac{\|p_k\|\kappa_0}{c\sigma_A^+\bar{d}} + \kappa_0 = \left(\frac{\|p_k\|}{c\sigma_A^+} + \bar{d} \right) \frac{\kappa_0}{\bar{d}}, \quad (60)$$

and hence that $\|p_k\| \leq \kappa_0/\bar{d}$. We have thus proved that (59) holds for every $k \geq 0$. ■

Lemma 4.10. *Consider the sequences $\{(z_k, v_k, p_k, \varepsilon_k)\}$ and $\{(\hat{z}_k, \hat{w}_k)\}$ generated by S-IAPIAL and let $\{\Delta_k\}$ be as in (29). Then, the following inequalities hold:*

$$\Delta_k \leq \frac{1}{k-1} \left(3R_* + \frac{\kappa_0^2}{2d^2c} \right), \quad \min_{i \leq k} \|\hat{w}_i\|^2 \leq \frac{C_1 \Delta_k}{\lambda} + \frac{\kappa_1}{2c}, \quad \forall k \geq 2, \quad (61)$$

$$\hat{w}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^* \hat{p}_k, \quad \|A\hat{z}_k - b\| \leq \frac{\kappa_2}{c}, \quad \forall k \geq 1, \quad (62)$$

where R_* and κ_0 are as in (32), κ_1 and κ_2 are as in (36), and D and C_1 are as in (B3) and (46), respectively.

Proof: The first inequality in (61) follows immediately by combining (51) and (59). The second inequality in (61) follows from the definition of κ_1 in (36), inequalities (48) and (59), and the fact that $k/(k-1) \leq 2$ for all $k \geq 2$. Now, since, in view of step 1 of S-IAPIAL, the triples $(\lambda, z_{k-1}, p_{k-1})$ and $(z_k, v_k, \varepsilon_k)$ satisfy (17) with $\tilde{\sigma} = \sigma_c$, it follows from Proposition 3.2(b) that the triple $(\hat{z}_k, \hat{w}_k, \hat{p}_k)$ computed in step 2 of S-IAPIAL satisfies the inclusion in (25) and the inequalities in (26) with $\tilde{\sigma} = \sigma_c$. Hence, the inclusion in (62) follows immediately from the inclusion in (25). Moreover, using the second inequality in (26) with $\tilde{\sigma} = \sigma_c$, the triangle inequality, and the definitions of σ_c , p_k , and Δp_k given in (28), (30), and (43), respectively, we have

$$\begin{aligned} \|A\hat{z}_k - b\| &\leq \|Az_k - b\| + \|A\| \|\hat{z}_k - z_k\| \leq \|Az_k - b\| + \frac{\sigma_c \|A\| \|r_k\|}{\sqrt{\lambda L_c + 1}} \\ &\leq \frac{\|\Delta p_k\|}{c} + \frac{\nu \|A\| \|r_k\|}{\lambda L_c + 1} \leq \frac{\|p_k\| + \|p_{k-1}\|}{c} + \frac{\nu D}{\lambda c \|A\| (1 - \sigma)}. \end{aligned}$$

where the last inequality is due to the first inequality in (52) and the fact that $L_c \geq c\|A\|^2$ (see the definition of L_c in (28)). Hence, the inequality in (62) follows from the above inequalities, inequality (59), and the definition of κ_2 given in (36). \blacksquare

Now we are ready to present the proof of Theorem 3.3.

Proof of Theorem 3.3: (a) First note that S-IAPIAL invokes in its step 1 the ACG method with (ψ_s, ψ_n) as in (16) and $(M_s, \mu) := (T_c/2, 1/2)$ where $T_c = 2\lambda L_c + 1$ and L_c is as in (28). Hence, the statement in (a) follows from the above observations, the definition of σ_c given in (28), Proposition 2.1 with $\tilde{\sigma} = \sigma_c$, and the fact that

$$\left(1 + \frac{1}{\sigma_c}\right) \sqrt{2M_s} \leq \frac{2\sqrt{\lambda L_c + 1}}{\min\{\nu, \sigma\}} \sqrt{T_c} \leq \frac{2T_c}{\min\{\nu, \sigma\}}.$$

(b) Let \bar{k} denote the number in (34) and assume that S-IAPIAL has not stopped before the \bar{k} -th iteration or in step 2 of the \bar{k} -th iteration. We will show that it must stop at step 3 of the \bar{k} -th iteration. Indeed, in view of the first inequality in (61) and the definition of \bar{k} , we obtain

$$\Delta_{\bar{k}} \leq \frac{1}{\bar{k}-1} \left(3R_* + \frac{\kappa_0^2}{2d^2c} \right) \leq \frac{(1 - \sigma^2)\lambda\hat{\rho}^2}{4(1 + 2\nu)^2} \quad (63)$$

and hence the stopping criterion in step 3 is satisfied at the \bar{k} -th iteration.

(c) Item (i) follows immediately from (62) and condition (35) on penalty parameter c . This item together with Definition 3.1 imply that, for any given iteration k , S-IAPIAL stops at its corresponding step 2 if and only if \hat{w}_k satisfies $\|\hat{w}_k\| \leq \hat{\rho}$, in which case the triple $(\hat{z}_k, \hat{w}_k, \hat{p}_k)$ is a

$(\hat{\rho}, \hat{\eta})$ -approximate stationary solution of (1). To prove ii), assume for contradiction that S-IAPIAL stops in step 3 (instead of step 2) of some iteration k . In view of (29), the definition of C_1 in (46), and the above observation, this means that

$$\frac{C_1 \Delta_k}{\lambda} \leq \frac{\rho^2}{2}, \quad \min_{i \leq k} \|\hat{w}_i\| > \hat{\rho}.$$

Using condition (35) on c , we easily see that these relations contradict the second inequality in (61). Hence, ii) follows. \blacksquare

A Appendix: Basic auxiliary results

This section presents four auxiliary results which are used in the complexity analysis of IAPIAL. This section also contains the proof of Proposition 3.2.

The following basic result is used in Lemma 4.6. Its proof can be found, for instance, in [13, Lemma 1.4].

Lemma A.1. *Let $A : \mathfrak{R}^n \rightarrow \mathfrak{R}^l$ be a non-zero linear operator and let σ_A^+ denote its smallest positive singular value. Then, for every $u \in A(\mathfrak{R}^n)$, we have $\sigma_A^+ \|u\| \leq \|A^* u\|$.*

The next result, whose proof can be found in [36, Lemma A.2], is used in the proof of Lemma 4.3.

Lemma A.2. *Let proper function $\tilde{\phi} : \mathfrak{R}^n \rightarrow (-\infty, \infty]$, scalar $\tilde{\sigma} \in (0, 1)$ and $(z_0, z_1) \in \mathfrak{R}^n \times \text{dom } \tilde{\phi}$ be given, and assume that there exists (v_1, ε_1) such that*

$$v_1 \in \partial_{\varepsilon_1} \left(\tilde{\phi} + \frac{1}{2} \|\cdot - z_0\|^2 \right) (z_1), \quad \|v_1\|^2 + 2\varepsilon_1 \leq \tilde{\sigma}^2 \|v + z_0 - z_1\|^2. \quad (64)$$

Then, for every $z \in \mathfrak{R}^n$ and $s > 0$, we have

$$\tilde{\phi}(z_1) + \frac{1}{2} [1 - \tilde{\sigma}^2(1 + s^{-1})] \|v_1 + z_0 - z_1\|^2 \leq \tilde{\phi}(z) + \frac{s+1}{2} \|z - z_0\|^2.$$

The following result derives several characterizations of condition **(B2)**, and shows that it implies an important inclusion that is used in the proof of Lemma 4.7.

Lemma A.3. *Let $h \in \overline{\text{Conv}}(\mathfrak{R}^n)$ and $L_h \geq 0$ be given. Then, the following statements are equivalent:*

a) *for every $z, z' \in \mathcal{H}$, we have*

$$h(z') \leq h(z) + L_h \|z' - z\|;$$

b) *for every $z, z' \in \mathcal{H}$, we have*

$$h'(z; z' - z) \leq L_h \|z' - z\|;$$

c) *for every $z, z' \in \mathcal{H}$ and $s \in \partial h(z)$, we have*

$$\langle s, z' - z \rangle \leq L_h \|z' - z\|;$$

d) *for every $z \in \mathcal{H}$, we have*

$$\partial h(z) \subset \bar{B}(0; L_h) + N_{\mathcal{H}}(z);$$

e) for every $z \in \mathcal{H}$, we have

$$\partial h(z) \cap \bar{B}(0; L_h) \neq \emptyset.$$

Moreover, either of the above conditions implies that:

i) \mathcal{H} is closed;

ii) for any $z \in \mathcal{H}$ and $\varepsilon \geq 0$, we have

$$\partial_\varepsilon h(z) \subset \bar{B}(0; L_h) + N_{\mathcal{H}}^\varepsilon(z).$$

Proof: [a) \Rightarrow b)] This statement follows from the fact that $h(z') - h(z) \geq h'(z; z' - z)$ for every $z, z' \in \mathcal{H}$ (see [46, Theorem 23.1]).

[b) \Rightarrow c)] This statement follows from the fact that $h'(z; z' - z) \geq \langle s, z' - z \rangle$ for every $z, z' \in \mathcal{H}$ and $s \in \partial h(z)$, (see [46, Theorem 23.2]).

[c) \Rightarrow d)] Letting $T_{\mathcal{H}}(z) = \text{cl}(\mathbb{R}_+ \cdot (\mathcal{H} - z))$ and $N_{\mathcal{H}}(z)$ denote the tangent cone and normal cone of \mathcal{H} at z , respectively, and letting $S := \bar{B}(0; L_h) + N_{\mathcal{H}}(z)$, we easily see that c) is equivalent to

$$\langle s, \cdot \rangle \leq L_h \|\cdot\| + I_{T_{\mathcal{H}}(z)}(\cdot) = \sigma_{\bar{B}(0; L_h)}(\cdot) + \sigma_{N_{\mathcal{H}}(z)}(\cdot) = \sigma_S(\cdot) \quad \forall s \in \partial h(z),$$

where the first equality follows in view of the discussion in page 115 of [46] and [17, Example 2.3.1 combined with Proposition 5.2.4], the last equality is due to [46, Corollary 16.4.1]. Since the above hold for every $s \in \partial h(z)$, we conclude that $\sigma_{\partial h(z)} \leq \sigma_S$. Since both $\partial h(z)$ and S are closed, it follows from [46, Corollary 13.1.1] that $\partial h(z) \subset S = \bar{B}(0; L_h) + N_{\mathcal{H}}(z)$.

[d) \Rightarrow e)] Assume that d) holds. We will first show that e) holds for every $z \in \text{ri } \mathcal{H}$. Indeed, assume that $z \in \text{ri } \mathcal{H}$. This implies that $N_{\mathcal{H}}(z)$ is a subspace, namely, the one orthogonal to the subspace parallel to the affine hull of \mathcal{H} . It follows from d) that there exists $s \in \partial h(z)$ and $n \in N_{\mathcal{H}}(z)$ such that $\|s - n\| \leq L_h$. Since $N_{\mathcal{H}}(z)$ is a subspace, it follows that $-n \in N_{\mathcal{H}}(z)$. The claim now follows by the observation that $s \in \partial f(z)$ and $-n \in N_{\mathcal{H}}(z)$ immediately implies that $s - n \in \partial f(z)$. We will now show that e) also holds for every $z \in \text{rbd } \mathcal{H}$. Indeed, assume that $z \in \text{rbd } \mathcal{H}$. Then, due to [17, Proposition 2.1.8], there exists $\{z_k\} \subset \text{ri } \mathcal{H}$ such that z_k converges to z as $k \rightarrow \infty$. Since e) holds for every $z \in \text{ri } \mathcal{H}$ and $\{z_k\} \subset \text{ri } \mathcal{H}$, we conclude that for every k , there exists $s_k \in \partial h(z_k)$ such that $\|s_k\| \leq L_h$. Hence, by Bolzano-Weisstrass' theorem, there exists a subsequence $\{s_k\}_{k \in \mathcal{K}}$ converging to some s , which clearly satisfies $\|s\| \leq L_h$. Using the fact that $\{(z_k, s_k)\}_{k \in \mathcal{K}} \in \text{Gr}(\partial h)$ and $\{(z_k, s_k)\}_{k \in \mathcal{K}}$ converges to (z, s) , and the fact that $h \in \overline{\text{Conv}}(\mathfrak{R}^n)$ implies that the set $\text{Gr}(\partial h)$ is closed, we then conclude that $(z, s) \in \text{Gr}(\partial h)$, i.e., $s \in \partial h(z)$. We have thus shown that e) holds for every $z \in \text{rbd } \mathcal{H}$ as well.

[e) \Rightarrow a)] Let $z, z' \in \mathcal{H}$ be given and assume that e) holds. Then, there exists $s' \in \partial h(z')$ such that $\|s'\| \leq L_h$. Hence, we have

$$h(z) - h(z') \geq \langle s', z - z' \rangle \geq -\|s'\| \|z' - z\| \geq -L_h \|z' - z\|.$$

Hence, a) follows.

[a) \Rightarrow i)] Assume that $\{z_k\} \subset \mathcal{H}$ converges to z . The fact that $h \in \overline{\text{Conv}}(\mathfrak{R}^n)$ and the assumption that (a) holds imply that

$$h(z) \leq \liminf_{k \rightarrow +\infty} h(z_k) \leq \liminf_{k \rightarrow +\infty} (h(z_1) + L_h \|z_k - z_1\|) = h(z_1) + L_h \|z - z_1\| < +\infty,$$

and hence that $z \in \mathcal{H}$. We have thus shown that \mathcal{H} is closed.

[a) \Rightarrow ii)] Let $z \in \mathcal{H}$ and $\varepsilon \geq 0$ be given and assume that a) holds. Consider the function ϕ_z defined as

$$\phi_z(z') := h(z) + L_h \|z' - z\| + I_{\mathcal{H}}(z') \quad \forall z' \in \mathfrak{R}^n.$$

Clearly, $\phi_z(z) = h(z)$ and $\phi_z \geq h$ in view of a). Using these two observations and the definition of the ε -subdifferential given in (6), we immediately see that $\partial_\varepsilon h(z) \subset \partial_\varepsilon \phi_z(z)$. On the other hand, using the ε -subdifferential rule for the sum of two convex functions (see [18, Theorem 3.1.1]), we have that

$$\partial_\varepsilon \phi_z(z) \subset \partial_\varepsilon (L_h \|\cdot - z\|)(z) + \partial_\varepsilon I_{\mathcal{H}}(z) = \partial_\varepsilon (L_h \|\cdot\|)(0) + N_{\mathcal{H}}^\varepsilon(z).$$

where the last equality is due to the the affine composition rule for the ε -subdifferential (see [18, Theorem 3.2.1]) and the fact that $N_{\mathcal{H}}^\varepsilon(\cdot) = \partial_\varepsilon I_{\mathcal{H}}(\cdot)$. The implication now follows from the above two inclusions and the fact that $\partial_\varepsilon (L_h \|\cdot\|)(0) = \bar{B}(0; L_h)$. \blacksquare

We observe that a) of Lemma A.3 is the same as condition **(B2)**. Conditions b) to e) are all equivalent to a), and hence **(B2)**. The implication a) \Rightarrow ii) is the one that is used in the proof of Lemma 4.7.

The following result is used to prove Proposition 3.2.

Lemma A.4. *Let functions $g, h : \mathfrak{R}^n \rightarrow (-\infty, \infty]$ be such that $h \in \overline{\text{Conv}}(\mathfrak{R}^n)$ and g is differentiable on \mathcal{H} and its gradient is L_g -Lipschitz continuous on \mathcal{H} . Let $(z^-, z, v, \varepsilon) \in \mathfrak{R}^n \times \mathcal{H} \times \mathfrak{R}^n \times \mathfrak{R}_+$ be such that*

$$v \in \partial_\varepsilon \left(g + h + \frac{1}{2} \|\cdot - z^-\|^2 \right) (z) \quad (65)$$

and compute

$$\hat{z} := \operatorname{argmin}_u \left\{ \langle \nabla \hat{g}(z), u - z \rangle + \frac{L_g + 1}{2} \|u - z\|^2 + h(u) \right\}, \quad (66)$$

$$w := v + z^- - z + (L_g + 1)(z - \hat{z}), \quad (67)$$

where

$$\hat{g} := g + \frac{1}{2} \|\cdot - z^-\|^2 - \langle v, \cdot \rangle. \quad (68)$$

Then the following statements hold:

a) the triple (z, w, ε) satisfies

$$w \in \nabla g(z) + \partial_\varepsilon h(z), \quad (69)$$

$$\|w\| \leq \|v + z^- - z\| + \sqrt{2(L_g + 1)\varepsilon}; \quad (70)$$

b) \hat{z} and \hat{w} defined by

$$\hat{w} := w + \nabla g(\hat{z}) - \nabla g(z) \quad (71)$$

satisfies

$$\hat{w} \in \nabla g(\hat{z}) + \partial h(\hat{z}), \quad (72)$$

$$\|\hat{w}\| \leq \|v + z^- - z\| + 2\sqrt{2(L_g + 1)\varepsilon}, \quad \|\hat{z} - z\| \leq \sqrt{2(L_g + 1)^{-1}\varepsilon}. \quad (73)$$

Proof: (a) First note that the pair of functions (g, h) and the scalar $M := L_g$ satisfy the assumptions of [22, Lemma 20]. Note also that (z^-, z, w, ε) where $w := v$ satisfies the relations in (92) of [22] with

$$\lambda = 1, \rho = \|z^- - z + v\|,$$

and \hat{g} , \hat{z} , and w defined in (68)–(67) correspond to f , z_f , and v defined in (93)–(94) of [22], respectively. Hence, the inclusion in (69) and the inequality in (70) follow from [22, Lemma 20(a)] and the fact that $0 \leq \delta_f \leq \varepsilon$ implies $\partial_{\delta_f} h(z) \subset \partial_\varepsilon h(z)$, see (6).

(b) The inclusion in (72) and the first inequality in (73) follow immediately from [22, Lemma 20(b)] and by noting that \hat{w} in (67) corresponds to v_f defined in (97) of [22]. The second inequality in (73) is not stated in [22, Lemma 20], but it is in its proof. More specifically, the aforementioned inequality follows from (99) of [22] and by noting that q_f defined in (87) corresponds to $(L_g + 1)(z - \hat{z})$. ■

Next we present the proof of Proposition 3.2.

Proof of Proposition 3.2:

First note that the assumptions of Lemma A.4 is satisfied with $g := \lambda g_k$, $h := \lambda h$, $L_g := \lambda L_c$, and $(z^-, z, v, \varepsilon) := (z_{k-1}, z_k, v_k, \varepsilon_k)$, in view of the assumptions of the proposition, **(B2)**, **(B5)**, definition of \mathcal{L}_c in (3), and the inclusion in (17). Hence, we have $\nabla \hat{g}(z_k) = \lambda \nabla g_k(z_k) - r_k$ where \hat{g} is as in (68) and r_k is as in (18). As a consequence, \hat{z} , w , and \hat{w} computed in (66), (67), and (71) correspond, respectively, to \hat{z}_k , λw_k , and $\lambda \hat{w}_k$ defined in (21), (20), and (24). It follows from the above observations and the conclusions of Lemma A.4 that the inclusions in (22) and (25) follow from (69) and (72), in view of the definition (6), the definition of δ_k given in (20), the definitions of p_k and \hat{p}_k given in (5) and (24), and the fact that $\nabla g_k(\cdot) = \nabla f(\cdot) + A^*(p_{k-1} + c(A \cdot - b))$. Moreover, the inequality in (23) and both inequalities in (26) follow from the ones in (70) and (72), the definition of r_k in (18) and the inequality in (17). The second inequality in (23) follows from the definition of δ_k given in (20) and the inequality in (17). ■

References

- [1] H. Attouch and J. Peypouquet. The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$. *SIAM J. Optim.*, 26(3):1824–1834, 2016.
- [2] N.S. Aybat and G. Iyengar. A first-order smoothed penalty method for compressed sensing. *SIAM J. Optim.*, 21(1):287–313, 2011.
- [3] N.S. Aybat and G. Iyengar. A first-order augmented Lagrangian method for compressed sensing. *SIAM J. Optim.*, 22(2):429–459, 2012.
- [4] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- [5] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York, 1982.
- [6] D. Boob, Q. Deng, and G. Lan. Stochastic first-order methods for convex and nonconvex functional constrained optimization. *Available on arXiv:1908.02734*, 2019.
- [7] Y. Carmon, J. C. Duchi, O. Hinder, and A. Sidford. Accelerated methods for nonconvex optimization. *SIAM J. Optim.*, 28(2):1751–1772, 2018.

- [8] C. Cartis, N. Gould, and P. Toint. On the complexity of steepest descent, Newton’s and regularized Newton’s methods for nonconvex unconstrained optimization problems. *SIAM J. Optim.*, 20(6):2833–2852, 2010.
- [9] Y. Chen, G. Lan, and Y. Ouyang. Optimal primal-dual methods for a class of saddle point problems. *SIAM J. Optim.*, 24(4):1779–1814, 2014.
- [10] D. Drusvyatskiy and C. Paquette. Efficiency of minimizing compositions of convex functions and smooth maps. *Math. Program.*, 178:503–558, 2019.
- [11] S. Ghadimi and G. Lan. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Math. Program.*, 156:59–99, 2016.
- [12] S. Ghadimi, G. Lan, and H. Zhang. Generalized uniformly optimal methods for nonlinear programming. *J. Scientific Comp.*, 79(3):1854–1881, Jun 2019.
- [13] M.L.N. Gonçalves, J.G. Melo, and R.D.C. Monteiro. Convergence rate bounds for a proximal admm with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *Pac. J. Optim.*, 15(3):379–398, 2019.
- [14] D. Hajinezhad1 and M. Hong. Perturbed proximal primal–dual algorithm for nonconvex non-smooth optimization. *Math. Program.*, 176:207–245, 2019.
- [15] Y. He and R.D.C. Monteiro. Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player Nash equilibrium problems. *SIAM J. Optim.*, 25(4):2182–2211, 2015.
- [16] Y. He and R.D.C. Monteiro. An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. *SIAM J. Optim.*, 26(1):29–56, 2016.
- [17] J.B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms I*. Springer, Berlin, 1993.
- [18] J.B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms II*. Springer, Berlin, 1993.
- [19] M. Hong. Decomposing linearly constrained nonconvex problems by a proximal primal dual approach: algorithms, convergence, and applications. *Available on arXiv:1604.00543*, 2016.
- [20] B. Jiang, T. Lin, S. Ma, and S. Zhang. Structured nonconvex and nonsmooth optimization algorithms and iteration complexity analysis. *Comput. Optim. Appl.*, 72(3):115–157, 2019.
- [21] O. Kolossoski and R.D.C. Monteiro. An accelerated non-Euclidean hybrid proximal extragradient-type algorithm for convex-concave saddle-point problems. *Optim. Methods Softw.*, 32(6):1244–1272, 2017.
- [22] W. Kong, J.G. Melo, and R.D.C. Monteiro. Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs. *SIAM J. Optim.*, 29(4):2566–2593, 2019.
- [23] W. Kong, J.G. Melo, and R.D.C. Monteiro. An efficient adaptive accelerated inexact proximal point method for solving linearly constrained nonconvex composite problems. *Comput. Optim. Appl.*, 76(2):305–346, 2019.

- [24] W. Kong and R.D.C. Monteiro. An accelerated inexact proximal point method for solving nonconvex-concave min-max problems. *Available on arXiv:1905.13433v2*, 2019.
- [25] G. Lan and R.D.C. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. *Math. Program.*, 138(1):115–139, Apr 2013.
- [26] G. Lan and R.D.C. Monteiro. Iteration-complexity of first-order augmented Lagrangian methods for convex programming. *Math. Program.*, 155(1):511–547, Jan 2016.
- [27] H. Li and Z. Lin. Accelerated proximal gradient methods for nonconvex programming. *Adv. Neural Inf. Process. Syst.*, 28:379–387, 2015.
- [28] Z. Li, P.-Y. Chen, S. Liu, S. Lu, and Y. Xu. Rate-improved inexact augmented Lagrangian method for constrained nonconvex optimization. *Available on arXiv:2007.01284*, 2020.
- [29] Z. Li and Y. Xu. First-order inexact augmented Lagrangian methods for convex and nonconvex programs: nonergodic convergence and iteration complexity. *Personal technical report*, 2019.
- [30] J. Liang and R.D.C. Monteiro. A Doubly Accelerated Inexact Proximal Point Method for Nonconvex Composite Optimization Problems. *Available on arXiv:1811.11378v2*, 2018.
- [31] J. Liang and R.D.C. Monteiro. An average curvature accelerated composite gradient method for nonconvex smooth composite optimization problems. *Available on arXiv:1909.04248v3*, 2019.
- [32] J. Liang, R.D.C. Monteiro, and C.-K. Sim. "a fista-type accelerated gradient algorithm for solving smooth nonconvex composite optimization problems. *Available on arXiv:1905.07010v2*, 2019.
- [33] Q. Lin, R. Ma, and Y. Xu. Inexact proximal-point penalty methods for non-convex optimization with non-convex constraints. *Available on Arxiv:1908.11518*, 2019.
- [34] Y.F. Liu, X. Liu, and S. Ma. On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. *Math. Oper. Res.*, 44(2):632–650, 2019.
- [35] Z. Lu and Z. Zhou. Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. *Available on arXiv:1803.09941*, 2018.
- [36] J.G. Melo, R.D.C. Monteiro, and H. Wang. Iteration-complexity of an inexact proximal accelerated augmented Lagrangian method for solving linearly constrained smooth nonconvex composite optimization problems. *Available on arXiv:2006.08048*.
- [37] R.D.C. Monteiro, Ortiz, and Benar F. Svaiter. An adaptive accelerated first-order method for convex optimization. *Comput. Optim. Appl.*, 64:31–73, 2016.
- [38] R.D.C. Monteiro and B.F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM J. Optim.*, 22(3):914–935, 2012.
- [39] R.D.C. Monteiro and B.F. Svaiter. An accelerated hybrid proximal extragradient method for convex optimization and its implications to second-order methods. *SIAM J. Optim.*, 23(2):1092–1125, 2013.

- [40] I. Necoara, A. Patrascu, and F. Glineur. Complexity of first-order inexact Lagrangian and penalty methods for conic convex programming. *Optim. Methods Softw.*, pages 1–31, 2017.
- [41] Y. E. Nesterov. *Introductory lectures on convex optimization : a basic course*. Kluwer Academic Publ., 2004.
- [42] Y.E. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, pages 1–37, 2012.
- [43] Y.E. Nesterov and B.T. Polyak. Cubic regularization of newton method and its global performance. *Math. Program.*, 108(1):177–205, 2006.
- [44] C. Paquette, H. Lin, D. Drusvyatskiy, J. Mairal, and Z. Harchaoui. Catalyst for gradient-based nonconvex optimization. In *AISTATS 2018-21st International Conference on Artificial Intelligence and Statistics*, pages 1–10, 2018.
- [45] A. Patrascu, I. Necoara, and Q. Tran-Dinh. Adaptive inexact fast augmented Lagrangian methods for constrained convex optimization. *Optim. Lett.*, 11(3):609–626, 2017.
- [46] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [47] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.
- [48] M. Sahin, A. Eftekhari, A. Alacaoglu, F. Latorre, and V Cevher. An inexact augmented Lagrangian framework for nonconvex optimization with nonlinear constraints. *Available on arXiv:1906.11357*, 2019.
- [49] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. <http://www.mit.edu/~dimitrib/PTseng/papers.html>, 2008.
- [50] Y. Xu. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. *Math. Program.*, 2019.
- [51] J. Zhang and Z.-Q. Luo. A global dual error bound and its application to the analysis of linearly constrained nonconvex optimization. *Available on arXiv:2006.16440*, 2020.
- [52] J. Zhang and Z.-Q. Luo. A proximal alternating direction method of multiplier for linearly constrained nonconvex optimization. *Available on arXiv:2006.16440*, 2020.