Iteration complexity of a proximal augmented Lagrangian method for solving nonconvex composite optimization problems with nonlinear convex constraints

August 16, 2020 (Revised: October 5, 2020)

Weiwei Kong [1], Jefferson G. Melo [2], and Renato D.C. Monteiro [3]

Abstract

This paper proposes and analyzes a proximal augmented Lagrangian (NL-IAPIAL) method for solving smooth nonconvex composite optimization problems with nonlinear $\mathcal{K}$-convex constraints, i.e., the constraints are convex with respect to the order given by a closed convex cone $\mathcal{K}$. Each NL-IAPIAL iteration consists of inexactly solving a proximal augmented Lagrangian subproblem by an accelerated composite gradient (ACG) method followed by a Lagrange multiplier update. Under some mild assumptions, it is shown that NL-IAPIAL generates an approximate stationary solution of the constrained problem in $\mathcal{O}(\log(1/\rho)/\rho^3)$ inner iterations, where $\rho > 0$ is a given tolerance. Numerical experiments are also given to illustrate the computational efficiency of the proposed method.

Key words. inexact proximal augmented Lagrangian method, $\mathcal{K}$-convexity, nonlinear constrained smooth nonconvex composite programming, accelerated first-order methods, iteration complexity.

AMS subject classifications. 49M05, 49M37, 90C26, 90C30, 90C60, 65K05, 65K10, 68Q25, 65Y20.

1 Introduction

This paper presents an inner accelerated nonlinear proximal inexact augmented Lagrangian (NL-IAPIAL) method for solving the cone convex constrained nonconvex composite optimization (CCC-NCO) problem

$$\phi^* := \inf_{z\in\Re^n} \left\{\phi(z) := f(z) + h(z) : g(z) \preceq_{\mathcal{K}} 0\right\}, \tag{1}$$

where $\mathcal{K}$ is a closed convex cone, $g : \Re^n \mapsto \Re^\ell$ is a differentiable $\mathcal{K}$-convex function with a Lipschitz continuous gradient, h is a proper closed convex function with compact domain, f is a nonconvex

[1]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205 (email: wwkong@gatech.edu). This author acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number PGSD3-516700-2018].

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

[3]School 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.

differentiable function on the domain of h with a Lipschitz continuous gradient, and the relation $g(z) \preceq_{\mathcal{K}} 0$ means that $g(z) \in -\mathcal{K}$. More specifically, the NL-IAPIAL method is based on the generalized (see [21] and [28, Section 11.K]) augmented Lagrangian (AL) function

$$\mathcal{L}_c(z;p) := f(z) + h(z) + \frac{1}{2c}\left[\operatorname{dist}^2(p + cg(z), -\mathcal{K}) - \|p\|^2\right], \tag{2}$$

where $\operatorname{dist}(y, S)$ denotes the Euclidean distance between a point $y \in \Re^\ell$ and a set $S \subseteq \Re^\ell$, and it performs the following proximal point-type update to generate its k-th iterate: given (z_{k-1}, p_{k-1}) and (λ, c_k), compute

$$z_k \approx \operatorname*{argmin}_u \left\{\lambda \mathcal{L}_{c_k}(u; p_{k-1}) + \frac{1}{2}\|u - z_{k-1}\|^2\right\}, \tag{3}$$

$$p_k = \Pi_{\mathcal{K}^*}(p_{k-1} + c_k g(z_k)), \tag{4}$$

where $\mathcal{K}^*$ denotes the dual cone of $\mathcal{K}$, the function $\Pi_{\mathcal{K}^*}$ denotes the projection onto $\mathcal{K}^*$, and z_k is a suitable approximate solution of the composite problem underlying (3). Contributing to its namesake of an "inner" accelerated method, our proposed method employs an accelerated composite gradient (ACG) algorithm to obtain the aforementioned point z_k. Moreover, at the end of the k-th iteration above, it performs a novel test to decide whether or not c_k is left unchanged or doubled.

Under a generalized Slater assumption[4] and a suitable choice of the inputs (λ, c), it is shown that for any $(\hat\rho, \hat\eta) \in \Re^2_{++}$, the NL-IAPIAL method obtains a quadruple $(\hat z, \hat p, \hat w, \hat q)$ satisfying

$$\hat w \in \nabla f(\hat z) + \partial h(\hat z) + \nabla g(\hat z)\hat p, \quad \langle g(\hat z) + \hat q, \hat p\rangle = 0, \quad g(\hat z) + \hat q \preceq_{\mathcal{K}} 0, \quad \hat p \succeq_{\mathcal{K}^*} 0 \tag{5}$$

$$\|\hat w\| \le \hat\rho, \quad \|\hat q\| \le \hat\eta, \tag{6}$$

in $\mathcal{O}((\hat\eta^{-1/2}\hat\rho^{-2} + \hat\rho^{-3})\log(\hat\rho^{-1} + \hat\eta^{-1}))$ ACG iterations. Moreover, this complexity result is shown without requiring that the initial point z_0 (in the domain of h) be feasible with respect to the nonlinear constraint, i.e., $g(z_0) \preceq_{\mathcal{K}} 0$. A key fact about NL-IAPIAL is that its generated sequence of Lagrange multipliers is shown to be always bounded, and this conclusion strongly uses the fact that its constraint function g is $\mathcal{K}$-convex (although (1) is nonconvex due to the nonconvexity assumption on f).

Related Works. The literature of AL-based methods is quite vast, so we focus our attention on those dealing with iteration complexities. Since AL-based methods for the convex case have been extensively studied in the literature (see, for example, [15, 2, 24, 3, 16, 20, 21, 27, 33]), we focus on papers that deal with nonconvex problems. Moreover, we concentrate more on those dealing with proximal AL-based methods, i.e., the ones for which the "inner" subproblems are of (or close to) the form in (3).

This paragraph reviews papers that consider proximal AL (PAL) methods for solving instances of (1) in which $\mathcal{K} = \{0\}$, or equivalently, g is affine and the constraint $g(z) \preceq_{\mathcal{K}} 0$ is replaced by $g(z) = 0$. Paper [9] studies the iteration complexity of a linearized PAL method under the restrictive assumption that the component of the objective function of (1) is identically zero. On the other hand, the remaining papers in this paragraph consider the general case in which h is a closed convex function (cf. problem (1)). Paper [7] introduces a perturbed θ-AL function, which agrees with the classical one in (2) when $\theta = 0$, and studies a corresponding unaccelerated PAL method whose iteration complexity is $\mathcal{O}(\hat\eta^{-4} + \hat\rho^{-4})$ under the strong condition that the initial point z_0 is feasible. Paper [23] analyzes the iteration complexity of an inexact proximal accelerated

[4]See Subsection 3.1 for the discussion.

PAL method based on the aforementioned perturbed AL function and shows, regardless of whether the initial point is feasible, that a solution to (1) is obtained in $\mathcal{O}(\hat{\eta}^{-1}\hat{\rho}^{-2}\log\hat{\eta}^{-1})$ ACG iterations and that the latter bound can be improved to $\mathcal{O}(\hat{\eta}^{-1/2}\hat{\rho}^{-2}\log\hat{\eta}^{-1})$ under an additional Slater-like assumption. Both papers [7, 23] assume that $\theta \in (0, 1]$, and hence, their analyses do not apply to the classical PAL method. In fact, as θ approaches zero, the universal constants that appear in the complexity bounds obtained in [7, 23] diverge to infinity. Using a different approach, i.e., one that does not rely on a merit function, paper [22] establishes the iteration complexity of an "inner" accelerated PAL method based on the classical augmented Lagrangian function in (2) and the Lagrange multiplier update (4).

For the case where each component of g is $\mathcal{K}$-convex and $\mathcal{K} = \{0\} \times \Re^k_+$, i.e., the constraint is of the form $g(x) = 0$ and/or $g(x) \leq 0$, papers [17, 30] present PAL methods that perform Lagrange multiplier updates only when the penalty parameter is updated. Hence, if the penalty parameter is never updated (which usually happens when the initial penalty parameter is chosen to be sufficiently large), then these methods never perform Lagrange multiplier updates, and thus they behave more like penalty methods.

For the case where g is not necessarily $\mathcal{K}$-convex and $\mathcal{K} = \{0\}$, i.e., the constraint is of the form $g(x) = 0$, paper [32] analyzes the complexity of a PAL method for solving (1) under the strong assumption that: (i) the composite function h is identically zero; (ii) the smallest singular value of $\nabla g(x)$ is uniformly bounded away from zero everywhere; and, optionally, (iii) the initial point is feasible.

It is now worth discussing how NL-IAPIAL compares with the works [22, 9, 30, 17, 7, 32]. First, the IAPIAL method of [22] is designed to solve the special instance of (1) in which $\mathcal{K} = \{0\}$. Also, in contrast to the NL-IAPIAL method, IAPIAL sets p_k to p_0 every time the penalty parameter c_k is increased, and hence it is not a full warm-start PAL approach. Next, compared to [30, 17], the multiplier update in (4) is performed at every prox iteration, regardless of whether the penalty parameter is updated. Finally, unlike the methods in [7, 9, 32], which require the initial point z_0 to be feasible, i.e., $g(z_0) \preceq_{\mathcal{K}} 0$, NL-IAPIAL only requires z_0 to be in the domain of h.

We finally discuss other papers that have motivated this work or are tangentially related to it. Papers [12, 11, 13, 19] establish the iteration complexity of quadratic penalty-based methods for solving (1). Paper [5] considers a primal-dual proximal point scheme and analyzes its iteration-complexity under strong conditions on the initial point. Papers [34, 35] present a primal-dual first-order algorithm for solving (1) when h is the indicator function of a box (in [35]) or more generally a polyhedron (in [34]), and show that it obtains a solution as in (5)-(6) in $\mathcal{O}(\hat{\rho}^{-2})$ iterations when $\hat{\rho} = \hat{\eta}$. Paper [18] 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 [10] considers a penalty-ADMM method that solves an equivalent reformulation of (1).

Organization of the Paper. Subsection 1.1 provides some basic definitions and notation. Section 2 reviews an ACG variant. Section 3 contains two subsections. The first one describes the main problem of interest and the assumptions made on it. The second one, on the other hand, states the NL-IAPIAL method and its main iteration complexity. Section 4 is divided into two subsections. The first one establishes the boundedness of the sequence of Lagrange multipliers generated by NL-IAPIAL, whereas the second one is devoted to the proof of the main iteration complexity result. Section 5 presents numerical experiments that demonstrate the efficiency of the NL-IAPIAL method. The end of the paper contains several appendices. Appendix A describes some basic convex analysis results, Appendix B gives the proof of a technical result, namely, Proposition 3.3.

1.1 Basic Definitions and Notations

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

Let $\Re_+$ and $\Re_{++}$ denote the set of nonnegative and positive real numbers, respectively, and let $\Re^2_{++} := \Re_{++} \times \Re_{++}$. We denote by R^n an n-dimensional inner product space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. For a given closed convex set $X \subset \Re^n$, its boundary is denoted by ∂X and the distance of a point $x \in \Re^n$ to X is denoted by $\text{dist}(x, X)$. For any $t > 0$, we let $\log_1^+(t) := \max\{\log t, 1\}$. Using the asymptotic notation $\mathcal{O}$, we denote $\mathcal{O}_1(\cdot) = \mathcal{O}(1 + \cdot)$.

The domain of a function $h : \Re^n \to (-\infty, \infty]$ is the set $\text{dom}\, h := \{x \in \Re^n : h(x) < +\infty\}$. 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 $\Re^n$ is denoted by $\overline{\text{Conv}}\, \Re^n$. The ε-subdifferential of a proper function $h : \Re^n \to (-\infty, \infty]$ is defined by

$$\partial_\varepsilon h(z) := \{u \in \Re^n : h(z') \geq h(z) + \langle u, z' - z \rangle - \varepsilon, \quad \forall z' \in \Re^n\} \tag{7}$$

for every $z \in \Re^n$. The classical subdifferential, denoted by $\partial h(\cdot)$, corresponds to $\partial_0 h(\cdot)$. If ψ is a real-valued function which is differentiable at $\bar{z} \in \Re^n$, then its affine approximation $\ell_\psi(\cdot, \bar{z})$ at $\bar{z}$ is given by

$$\ell_\psi(z; \bar{z}) := \psi(\bar{z}) + \langle \nabla \psi(\bar{z}), z - \bar{z} \rangle \quad \forall z \in \Re^n. \tag{8}$$

For a closed convex cone $\mathcal{K} \subset \Re^l$, the dual cone $\mathcal{K}^*$ and polar cone $\mathcal{K}^\circ$ are defined as

$$\mathcal{K}^* := \left\{ y \in \Re^l : \langle y, x \rangle \geq 0, x \in \mathcal{K} \right\}, \quad \mathcal{K}^\circ := -\mathcal{K}^*. \tag{9}$$

For given $u, v \in \Re^l$ we say that $u \preceq_{\mathcal{K}} v$ ($v \succeq_{\mathcal{K}} u$) if and only if $v - u \in \mathcal{K}$. Moreover, if $\mathcal{K}$ has nonempty interior, then we say that $u \prec_{\mathcal{K}} v$ if and only if $v - u \in \text{int}\, \mathcal{K}$. A function $g : \Re^n \to \Re^\ell$ is said to be $\mathcal{K}$-convex if for every $t \in [0, 1]$, we have

$$g(tz' + [1-t]z) - tg(z') - [1-t]g(z) \preceq_{\mathcal{K}} 0 \quad \forall z, z' \in \Re^n. \tag{10}$$

It is well-known that if g is differentiable and $\mathcal{K}$-convex, then for every $z, z' \in \Re^n$, we have

$$g'(z)(z' - z) \preceq_{\mathcal{K}} g(z') - g(z). \tag{11}$$

2 Review of an ACG Algorithm

This section reviews an ACG algorithm invoked by NL-IAPIAL for solving the sequence of subproblems (3) 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 inexact solution of the subproblem.

Consider the following composite optimization problem

$$\min\ \{\psi(x) := \psi_s(x) + \psi_n(x) : x \in \Re^n\}, \tag{12}$$

where the following conditions are assumed to hold:

(A1) $\psi_n \in \overline{\text{Conv}}\, \Re^n$ is a 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

$$\psi_s(u) - \ell_{\psi_s}(u; x) \leq \frac{M_s}{2}\|u - x\|^2 \quad \forall x, u \in \text{dom}\, \psi_n$$

where $\ell_{\psi_s}(\cdot\,;\cdot)$ is defined in (8).

The ACG algorithm, given $(x_0, \tilde{\sigma}) \in \mathcal{H} \times \Re_{++}$, inexactly solves (12) by obtaining a triple (x, u, η) satisfying

$$u \in \partial_\eta(\psi_s + \psi_n)(x) \quad \|u\|^2 + 2\eta \le \tilde{\sigma}^2 \|x_0 - x + u\|^2. \tag{13}$$

With this in mind, we now state the titular algorithm of this section.

ACG Algorithm

Input: a quadruple $(M_s, \mu, \psi_s, \psi_n)$ satisfying assumptions (A1)–(A2), an initial point $x_0 \in \operatorname{dom} \psi_n$, and a scalar $\tilde{\sigma} > 0$;
Output: a triple (x, u, η) satisfying (13);

(0) set $y_0 = x_0$, $A_0 = 0$, $\Gamma_0 \equiv 0$ and $j = 1$;

(1) compute

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

(2) compute

$$u_j = \frac{y_0 - y_j}{A_j}, \quad \eta_j = \psi(x_j) - \Gamma_j(y_j) - \psi_n(y_j) - \langle u_j, x_j - y_j \rangle,$$

and set $(x, u, \eta) = (x_j, u_j, \eta_j)$;

(3) if (x, u, η) satisfies the inequality in (13) then stop and output (x, u, η); otherwise, update $j \leftarrow j + 1$ and go to step 1.

Some remarks about the ACG algorithm are in order. First, step 1 is the most common way of describing an iteration of an accelerated gradient-based algorithm. Second, the auxiliary sequences $\{u_j\}_{j\ge1}$ and $\{\eta_j\}_{j\ge1}$ generated by step 2 are used as a stopping criterion (see step 3) for the ACG method when it is called as a subroutine for solving the subproblems of NL-IAPIAL in Section 3. Third, the ACG algorithm in which $\mu = 0$ is a special case of a slightly more general one studied by Tseng in [31] (see Algorithm 3 of [31]). The analysis of the general case of the ACG algorithm in which $\mu \ge 0$ was studied in [8, Proposition 2.3]. Finally, the elements of the sequence $\{A_k\}_{k\ge1}$ have the following lower bound:

$$A_j \ge \frac{1}{M_s} \max\left\{ \frac{j^2}{4}, \left(1 + \sqrt{\frac{\mu}{4M_s}}\right)^{2(j-1)} \right\}. \tag{14}$$

The next proposition, whose proof can be found in [22], summarizes the main properties of the ACG algorithm that will be needed in the analysis of the NL-IAPIAL method.

Proposition 2.1. *Let $\{(A_j, x_j, u_j, \eta_j)\}$ be the sequence generated by the ACG algorithm applied to (12). 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 (12) in at most

$$\left\lceil 1 + \sqrt{\frac{M_s}{\mu}} \log_1^+ \left([1 + \tilde{\sigma}^{-1}] \sqrt{2M_s} \right) \right\rceil \tag{15}$$

iterations.

Before proceeding to the development of the NL-IAPIAL method, we remark that other ACG-type algorithms, such as the ones in [1, 8, 26, 25, 31], could have been used in place of the one presented in this section.

3 The NL-IAPIAL Method

This section consists of two subsections. The first one precisely describes the problem of interest and its assumptions, whereas the second one presents the NL-IAPIAL method and its corresponding iteration complexity.

3.1 Problem of Interest

This subsection presents the main problem of interest and the assumptions underlying it.

Consider problem (1) where $\mathcal{K}$ is a nonempty closed convex cone, and the functions f, g and h satisfy the following assumptions:

(B1) $h \in \overline{\text{Conv}}\ \Re^n$ is K_h-Lipschitz continuous for some $K_h > 0$, and $\mathcal{H} := \text{dom}\, h$ is compact with diameter $D_h := \sup_{z',z\in\mathcal{H}} \|z' - z\| < \infty$;

(B2) f is a nonconvex function which is differentiable on $\mathcal{H}$, and there exists $(m_f, L_f) \in \Re^2_{++}$ such that, for every $z', z \in \mathcal{H}$, we have

$$-\frac{m_f}{2}\|z' - z\|^2 \leq f(z') - f(z) - \langle \nabla f(z), z' - z \rangle, \tag{16}$$

$$\|\nabla f(z') - \nabla f(z)\| \leq L_f \|z' - z\|; \tag{17}$$

(B3) $g : \Re^n \mapsto \Re^\ell$ is $\mathcal{K}$-convex and differentiable, and there exists $L_g > 0$ such that

$$\|\nabla g(z') - \nabla g(z)\| \leq L_g \|z' - z\| \quad \forall z', z \in \Re^n;$$

(B4) there exist $\bar{z} \in \text{int}\, \mathcal{H}$ and $\tau \in (0, 1]$ such that $g(\bar{z}) \preceq_{\mathcal{K}} 0$ and

$$\max \{\|\nabla g(z)p\|, |\langle p, g(\bar{z})\rangle|\} \geq \tau \|p\| \quad \forall z \in \mathcal{H},\ \forall p \succeq_{\mathcal{K}^*} 0. \tag{18}$$

We now make some comments about the above assumptions. First, it is well-known that (17) 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 (16) holds with $m_f = L_f$. However, better iteration-complexity bounds can be derived when a scalar $m_f < L_f$ satisfying (16) is available. Second, (16) implies that the function $f + m_f\|\cdot\|^2/2$ is convex on $\mathcal{H}$.

Moreover, since f is nonconvex on $\mathcal{H}$, in view of (B2), the smallest m_f satisfying (16) is positive. Third, since $\mathcal{H}$ is compact by (B1), the image of any continuous $\Re^k$-valued function is bounded. In view of this observation, we introduce the useful notation for any continuously differentiable function $\Psi : \Re^n \mapsto \Re^k$:

$$B_\Psi^{(0)} := \sup_{z\in\mathcal{H}} \|\Psi(z)\| < \infty, \quad B_\Psi^{(1)} := \sup_{z\in\mathcal{H}} \|\nabla\Psi(z)\| < \infty. \tag{19}$$

Finally, the conditions in (B4) can be viewed as a generalization of a Slater-like assumption with respect to g, as shown in Proposition 3.1 below.

Proposition 3.1. *(Slater-like Assumption) Assume that the constraint $g(z) \preceq_{\mathcal{K}} 0$ is of the form*

$$g_\iota(z) \preceq_{\mathcal{J}} 0 \quad g_e(z) = 0 \tag{20}$$

where $\mathcal{J} \subseteq \Re^s$ is a closed convex cone, $g_\iota : \Re^n \to \Re^s$ is continuously differentiable, and $g_e : \Re^n \to \Re^t$ is an onto affine map (and hence $g = (g_\iota, g_e)$ and $\mathcal{K} = \mathcal{J} \times \{0\}$). Assume also that there exists $\bar z \in \mathcal{H}$ such that $g_\iota(\bar z) \prec_{\mathcal{J}} 0$ and $g_e(\bar z) = 0$. Then, there exists $\tau > 0$ such that $(\bar z, \tau)$ satisfies (18). If, in addition, $\bar z \in \operatorname{int}\mathcal{H}$, then $(\bar z, \tau)$ satisfies (B4).

Proof. Since g_e is affine and onto, its gradient matrix $G_e := \nabla g_e$ is independent of z and has full column rank. Hence, there exists $\tau_e > 0$ such that

$$\|G_e p_e\| \ge \tau_e \|p_e\|_1 \quad \forall p_e \in \Re^s. \tag{21}$$

On the other hand, the assumption that $g_\iota(\bar z) \prec_{\mathcal{J}} 0$, and Lemma A.2 with $\mathcal{K} = \mathcal{J}$ and $x = -g_\iota(\bar z) \in \mathcal{J}$, imply that there exists $\tau_\iota > 0$ such that

$$-\langle p_\iota, g_\iota(\bar z)\rangle \ge \tau_\iota \|p_\iota\| \quad \forall p_\iota \in \mathcal{J}^*.$$

Using the previous inequality and the fact that $\|\nabla g_\iota(z)\|$ is bounded on $\mathcal{H}$, we conclude that there exists $\gamma > 0$ such that

$$-\|\nabla g_\iota(z) p_\iota\| - 2\gamma \langle p_\iota, g_\iota(\bar z)\rangle \ge [2\gamma\tau_\iota - \|\nabla g_\iota(z)\|] \cdot \|p_\iota\| \ge \tau_\iota \|p_\iota\|_1 \quad \forall z \in \mathcal{H}. \tag{22}$$

Relations (21), (22), and the reverse triangle inequality, then imply that for every $z \in \mathcal{H}$,

$$\begin{aligned}
&\|\nabla g(z) p\| - 2\gamma \langle p, g(\bar z)\rangle = \|\nabla g_\iota(z) p_\iota + G_e p_e\| - 2\gamma \langle p_\iota, g_\iota(\bar z)\rangle \\
&\ge \|G_e p_e\| - \|\nabla g_\iota(z) p_\iota\| - 2\gamma \langle p_\iota, g_\iota(\bar z)\rangle \ge \tau_e \|p_e\|_1 + \tau_\iota \|p_\iota\|_1 \ge \tau_c \|p\|_1 \ge \tau_c \|p\|,
\end{aligned}$$

where $\tau_c := \min\{\tau_e, \tau_\iota, 1\}$. It is now straightforward to see that the above inequality yields inequality (18) with $\tau = \tau_c/(1+2\gamma) \in (0, 1]$. The last part of the proposition now follows from the statement of assumption (B4) and the previous conclusion. □

Some comments about Proposition 3.1 are in order. First, the assumption that g_ι is $\mathcal{J}$-convex and g_e is affine implies that g is $\mathcal{K}$-convex. Second, the Slater condition is in regards to a single point $\bar z \in \mathcal{H}$, as opposed to condition (18) which involves inequality (18) at all pairs $(z, p) \in \mathcal{H} \times \mathcal{K}^*$. Third, (B4) can be replaced by the Slater-like assumption of Proposition 3.1 since the former is implied by the latter. Actually, a slightly more involved analysis can be done to show that the assumption that g_e is onto (which is part of the assumption of Proposition 3.1) can be removed at the expense of obtaining a weaker version of (B4), namely: inequality (18) holds for every pair

$(z,p) \in \mathcal{H} \times (\mathcal{J}^* \times \operatorname{Im} \nabla g_e)$, instead of $(z,p) \in \mathcal{H} \times (\mathcal{J}^* \times \Re^t) = \mathcal{H} \times \mathcal{K}^*$. Finally, since the analysis of this paper can be easily adapted to this slightly weaker version of (B4), the Slater-like condition of Proposition 3.1 without g_e assumed to be onto (or equivalently, ∇g_e to have full column rank) can be used in place of (B4) in order to guarantee that all of the results derived in this paper for NL-IAPIAL hold.

We now briefly discuss the notion of an approximate solution of (1) that is sought after by the NL-IAPIAL method. It is well-known that a necessary condition for a point z^* to be a local minimum of (1) is that there exists a multiplier $p^* \in \mathcal{K}^*$ that satisfies the conditions

$$0 \in \nabla f(z^*) + \partial h(z^*) + \nabla g(z^*)p^*, \quad \langle g(z^*), p^* \rangle = 0, \quad g(z^*) \preceq_{\mathcal{K}} 0, \quad p^* \succeq_{\mathcal{K}^*} 0. \tag{23}$$

In view of this fact, we say that a triple $(\hat{z}, \hat{p}, \hat{w})$ is a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1) if it satisfies conditions (5)-(6), which can be seen as a relaxation of (23) for any $(\hat{\rho}, \hat{\eta}) \in \Re^2_{++}$. It is also worth mentioning that the last three conditions in (23) (resp. (5)) are equivalent[5] to the inclusion $g(z^*) \in N_{\mathcal{K}^*}(p^*)$ (resp. the inequality $\operatorname{dist}(g(\hat{z}), N_{\mathcal{K}^*}(\hat{p})) \leq \hat{\eta}$).

We end this subsection by stating a technical result which describes some properties about the smooth part of the Lagrangian in (2).

Lemma 3.2. *Define the function*

$$\widetilde{\mathcal{L}}_c(z;p) := f(z) + \frac{1}{2c}\left[\operatorname{dist}^2(p + cg(z), -\mathcal{K}) - \|p\|^2\right] \quad \forall (z,p,c) \in \Re^n \times \Re^\ell \times \Re_{++}. \tag{24}$$

Then, for every $c > 0$ and $p \in \Re^\ell$, the following properties hold:

(a) $\widetilde{\mathcal{L}}_c(\cdot;p)$ is convex, differentiable, and its gradient is given by

$$\nabla_z \widetilde{\mathcal{L}}_c(z;p) = \nabla f(z) + \nabla g(z) \Pi_{\mathcal{K}^*}(p + cg(z)) \quad \forall z \in \Re^n;$$

(b) $\nabla_z \widetilde{\mathcal{L}}_c(\cdot;p)$ is $\widetilde{L}$-Lipschitz continuous where

$$\widetilde{L} = \widetilde{L}(c,p) := L_f + L_g\|p\| + c\left(B_g^{(0)} L_g + [B_g^{(1)}]^2\right), \tag{25}$$

and the quantities L_f, L_g, and $(B_g^{(0)}, B_g^{(1)})$ are as in (B2), (B3), and (19), respectively.

Proof. We first state that the case of $f \equiv 0$ and $L_f = 0$ has been previously shown in [21, Proposition 5] under the condition that $B_g^{(1)}$ is a Lipschitz constant of g. Hence, in view of assumption (B2) and the definition of $\mathcal{L}_c$, it suffices to verify the aforementioned condition. Indeed, using the Mean Value Inequality and the definition of $B_g^{(1)}$ in (19) we have that

$$\|g(z') - g(z)\| \leq \sup_{t \in [0,1]} \|\nabla g(tz' + [1-t]z)\| \cdot \|z' - z\| \leq B_g^{(1)} \|z' - z\| \quad \forall z', z \in \mathcal{H},$$

and hence that g is $B_g^{(1)}$-Lipschitz continuous. □

[5] See, for example, [28, Example 11.4] with $\bar{x} = g(z^*)$ and $\bar{v} = p^*$.

3.2 The NL-IAPIAL Method and Its Iteration Complexity

This subsection presents the NL-IAPIAL method and its corresponding iteration complexity.

Before presenting the method, we give a short, but precise, outline of its key steps as well as a description of how its key iterates are generated. Recall from the introduction that the NL-IAPIAL method, for obtaining a $(\hat\rho,\hat\eta)$-approximate solution as in (5)-(6), is an iterative method which, at its k-th step, computes its next iterate (z_k,p_k) using the updates in (3) and (4). More specifically, given (λ,σ) and (z_{k-1},p_{k-1}), and using the function $\widetilde{L}$ in (25) to define

$$L_{k-1}^{\psi} := \lambda\widetilde{L}(c_k,p_{k-1})+1, \quad \sigma_{k-1} := \frac{\sigma}{\sqrt{L_{k-1}^{\psi}}}, \tag{26}$$

the iterate z_k is obtained by calling the ACG algorithm of Section 2 with inputs

$$\begin{aligned} \psi_s &= \lambda\widetilde{\mathcal{L}}_{c_k}(\cdot;p_{k-1})+\frac{1}{4}\|\cdot-z_{k-1}\|^2, \quad \psi_n=\lambda h+\frac{1}{4}\|\cdot-z_{k-1}\|^2,\\ M_s &= \lambda L_{k-1}^{\psi}-\frac{1}{2}, \quad \mu=\frac{1}{2}, \quad x_0=z_{k-1}, \quad \tilde\sigma=\sigma_{k-1}, \end{aligned} \tag{27}$$

so that z_k, together with the other ACG outputs v_k and ε_k, satisfy

$$v_k\in\partial_{\varepsilon_k}\left(\lambda\mathcal{L}_{c_k}(\cdot;p_{k-1})+\frac{1}{2}\|\cdot-z_{k-1}\|^2\right)(z_k), \quad \|v_k\|^2+2\varepsilon_k\le\sigma_{k-1}^2\|v_k+z_{k-1}-z_k\|^2. \tag{28}$$

Using this triple (z_k,v_k,ε_k), and the available data $(\lambda,z_{k-1},p_{k-1},\sigma_{k-1})$, it then generates a refined quadruple $(\hat z,\hat p,\hat w,\hat q)=(\hat z_k,\hat p_k,\hat w_k,\hat q_k)$ satisfying all the conditions in (5). If this quadruple also satisfies the bounds in (6), then the method stops and outputs this refined solution. Otherwise, p_k is computed according to (4) and the method continues to the $(k+1)$-th iteration.

The next result, whose proof can be found in Appendix B, describes how the quadruple $(\hat z_k,\hat p_k,\hat w_k,\hat q_k)$ is generated and presents some of its key properties.

Proposition 3.3. *Suppose $(\lambda,z_{k-1},p_{k-1},\sigma_{k-1})$ and (z_k,v_k,ε_k) satisfy (28) and define the auxiliary quantities*

$$\delta_k := \frac{1}{\lambda}\varepsilon_k, \quad r_k := v_k+z_{k-1}-z_k, \quad w_k := \frac{1}{\lambda}\left[r_k+L_{k-1}^{\psi}(z_k-\hat z_k)\right], \tag{29}$$

where

$$\hat z_k := \operatorname*{argmin}_u\left\{\lambda\left[\left\langle\nabla_z\widetilde{\mathcal{L}}_{c_k}(z_k;p_{k-1}),u-z_k\right\rangle+h(u)\right]-\langle r_k,u-z_k\rangle+\frac{L_{k-1}^{\psi}}{2}\|u-z_k\|^2\right\}, \tag{30}$$

the function $\widetilde{\mathcal{L}}_{c_k}$ is as in (24), and the constant L_{k-1}^{ψ} is as in (26). Moreover, consider the triple $(\hat p_k,\hat q_k,\hat w_k)$ given by

$$\begin{aligned} \hat p_k &:= \Pi_{\mathcal{K}^*}\left(p_{k-1}+c_kg(\hat z_k)\right),\\ \hat q_k &:= \frac{1}{c_k}(p_{k-1}-\hat p_k),\\ \hat w_k &:= w_k+\nabla_z\widetilde{\mathcal{L}}_{c_k}(\hat z_k;p_{k-1})-\nabla_z\widetilde{\mathcal{L}}_{c_k}(z_k;p_{k-1}). \end{aligned} \tag{31}$$

Then, the following properties hold for every $k\ge1$:

(a) the triple (z_k, w_k, δ_k), together with p_k given by (4), satisfy

$$
\begin{aligned}
w_k &\in \nabla f(z_k) + \partial_{\delta_k} h(z_k) + \nabla g(z_k) p_k, \\
\|w_k\| &\le \frac{1}{\lambda}\left(1 + \sigma_{k-1}\sqrt{L^{\psi}_{k-1}}\right)\|r_k\|, \quad \delta_k \le \frac{1}{2\lambda}\sigma_{k-1}^2 \|r_k\|^2;
\end{aligned}
\tag{32}
$$

(b) the quadruple $(\hat{z}, \hat{p}, \hat{w}, \hat{q}) := (\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)$ satisfies (5) and

$$
\|\hat{w}_k\| \le \frac{1}{\lambda}\left(1 + 2\sigma_{k-1}\sqrt{L^{\psi}_{k-1}}\right)\|r_k\|, \quad \|\hat{q}_k\| \le \frac{B_g^{(1)}\sigma_{k-1}}{\sqrt{L^{\psi}_{k-1}}}\|r_k\| + \frac{1}{c_k}\|p_k - p_{k-1}\|, \tag{33}
$$

where $B_g^{(1)}$ is given by (19).

Two comments about Proposition 3.3 are in order. First, the relations in (a) will be used to establish the boundedness of the sequence $\{p_k\}$. Second, as we will see, NL-IAPIAL computes a sequence $\{(\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)\}$ as in Proposition 3.3. In view of (b), the quadruple $(\hat{z}, \hat{p}, \hat{w}, \hat{q}) := (\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)$ always satisfies (5). Hence, in order to obtain a $(\hat{\rho}, \hat{\eta})$-approximate solution of (1), it remains only to guarantee that condition (6) will eventually be satisfied. The inequalities in (33) will be essential to show the latter fact.

We now state the NL-IAPIAL method in its entirety.

NL-IAPIAL Method

Input: a function triple (f, g, h) satisfying assumptions (B1)–(B4), a stepsize $\lambda \le 1/(2m_f)$, a scalar $\sigma \in (0, 1/\sqrt{2}]$, a penalty parameter $c_1 > 0$, an initial pair $(z_0, p_0) \in \mathcal{H} \times \Re^l$, and a tolerance pair $(\hat{\rho}, \hat{\eta}) \in \Re^2_{++}$;
Output: a triple $(\hat{z}, \hat{p}, \hat{w}, \hat{q})$ satisfying (5)-(6);

(0) set $c = c_1$, $\hat{k} = 0$, and $k = 1$;

(1) let L^{ψ}_{k-1}, $\widetilde{\mathcal{L}}_{c_k}$, and σ_{k-1} be as in (26) and use the ACG algorithm with inputs $(M_s, \mu, \psi_s, \psi_n)$, x_0, and $\tilde{\sigma}$ given by (27) to obtain a triple $(z_k, v_k, \varepsilon_k)$ satisfying (28);

(2) compute the refined quadruple $(\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)$ according to (31); if $(\hat{w}, \hat{q}) := (\hat{w}_k, \hat{q}_k)$ satisfies (6) then stop and output $(\hat{z}, \hat{p}, \hat{w}, \hat{q}) = (\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)$; otherwise, go to step 3;

(3) compute the multiplier p_k according to (4);

(4) if $k > \hat{k} + 1$ and

$$
\Delta_k := \frac{1}{k - \hat{k} - 1}\left[\mathcal{L}_{c_k}(z_{\hat{k}+1}, p_{\hat{k}+1}) - \mathcal{L}_{c_k}(z_k, p_k)\right] \le \frac{\lambda(1-\sigma^2)\hat{\rho}^2}{4(1+2\sigma)^2} \tag{34}
$$

then set $c_{k+1} = 2c_k$ and $\hat{k} = k$; otherwise, set $c_{k+1} = c_k$;

(5) update $k \leftarrow k + 1$, and go to step 1.

Some remarks about the NL-IAPIAL method are in order. First, step 1 performs the call to the ACG algorithm that is described in the second paragraph of this subsection. Second, it performs two kinds of iterations, namely, the ones indexed by k and the ones performed by the ACG algorithm every time it is called in step 1. Let us refer to the former as "outer" iterations and the latter as "inner" (or ACG) iterations. Third, its input z_0 can be any element in the domain of h and does not necessarily need to be a point satisfying the constraint $g(z_0) \preceq_{\mathcal{K}} 0$. Finally, in view of Proposition 3.3(b) and the remark following it, NL-IAPIAL stops in step 2 if and only if the quadruple $(\hat{z}, \hat{p}, \hat{w}, \hat{q})$ is a $(\hat{\rho}, \hat{\eta})$-approximate solution of (1).

We will now give some intuition about step 3 of NL-IAPIAL but first introduce the notion of a cycle. Define the l-th cycle $\mathcal{K}_l$ as the l-th set of consecutive indices k for which c_k remains constant, i.e.,

$$\mathcal{C}_l := \{k : c_k = \tilde{c}_l := 2^{l-1} c_1\}. \tag{35}$$

For every $l \geq 1$, we let k_l denote the largest index in $\mathcal{C}_l$. Hence,

$$\mathcal{C}_l = \{k_{l-1} + 1, \ldots, k_l\} \quad \forall l \geq 1$$

where $k_0 := 0$. Clearly, the different values of $\hat{k}$ that arise in step 4 are exactly the indices in the index set $\{k_l : l \geq 0\}$. Moreover, in view of the test performed in step 4, we have that $k_l - k_{l-1} \geq 2$ for every $l \geq 1$, or equivalently, every cycle contains at least two indices. While generating the indices in the l-th cycle, if an index $k \geq k_{l-1} + 2$ satisfying (34) is found, k becomes the last index k_l in the l-th cycle and the $(l+1)$-th cycle is started at iteration $k_l + 1$ with the penalty parameter set to $\tilde{c}_{l+1} = 2\tilde{c}_l$, where $\tilde{c}_l$ is as in (35).

The following result, whose proof is deferred to Section 4, describes the inner iteration complexity of NL-IAPIAL. Its iteration-complexity bound is expressed in terms of the following parameters:

$$\kappa_0 := 2\left[K_h + B_f^{(1)}\right] D_h + \left[\frac{\sigma^2}{2(1-\sigma)^2} + 2\left(\frac{1+\sigma}{1-\sigma}\right)\right] \frac{D_h^2}{\lambda}, \tag{36}$$

$$\kappa_1 := L_f + \frac{\kappa_0 L_g}{\min\{1, \bar{d}\}\tau}, \qquad \kappa_2 := B_g^{(0)} L_g + [B_g^{(1)}]^2, \tag{37}$$

$$\kappa_3 := \frac{16(1+2\sigma)^2 \max\{\|p_0\|, \kappa_0\}^2}{\lambda(1-\sigma^2)\min\{1, \bar{d}^2\}\tau^2}, \quad \kappa_4 := \frac{\sigma D_h}{\lambda(1-\sigma) B_g^{(1)}} + \frac{2\max\{\|p_0\|, \kappa_0\}^2}{\min\{1, \bar{d}\}\tau}, \tag{38}$$

where $(B_g^{(0)}, B_g^{(1)}, B_f^{(1)})$, (K_h, D_h), L_f, L_g, and $(\tau, \bar{z})$ are as in (19), (B1), (B2), (B3), and (B4), respectively, the quantities σ, (z_0, p_0), and λ are as in the input of the NL-IAPIAL, and $\bar{d}$ is defined as

$$\bar{d} := \operatorname{dist}(\bar{z}, \partial \mathcal{H}). \tag{39}$$

Theorem 3.4. *Consider the parameters λ, c_1, and $(\hat{\rho}, \hat{\eta})$ as described in the input of NL-IAPIAL method and the scalars κ_i, $i = 0, \ldots, 4$, as in (36), (37), and (38). Moreover, define $\mathcal{T}$ and R_ϕ as*

$$\mathcal{T} = \mathcal{T}(\hat{\rho}, \hat{\eta}) := [\lambda(\kappa_1 + c_1\kappa_2) + 1] \frac{\max\{c_1, 2\bar{c}\}}{c_1}, \qquad R_\phi := \phi^* - \phi_* + \frac{D_h^2}{\lambda}, \tag{40}$$

where ϕ^* and D_h are as in (1) and (B1), respectively, and $\bar{c}$ and ϕ_* are given by

$$\bar{c} = \bar{c}(\hat{\rho}, \hat{\eta}) := \frac{\kappa_3}{\hat{\rho}^2} + \frac{\kappa_4}{\hat{\eta}}, \qquad \phi_* := \inf_{z \in \Re^n} \phi(z). \tag{41}$$

Then, the NL-IAPIAL method obtains a $(\hat{\rho}, \hat{\eta})$-approximate solution of (1) within a number of inner iterations bounded by

$$\mathcal{O}_1\left(\frac{R_\phi}{\lambda\hat{\rho}^2}\sqrt{\mathcal{T}}\log_1^+\mathcal{T}\right). \tag{42}$$

It is worth mentioning that the iteration complexity bound in Theorem 3.4, in terms of the tolerance pair $(\hat{\rho}, \hat{\eta})$, is

$$\mathcal{O}_1\left(\left[\frac{1}{\sqrt{\hat{\eta}}\cdot\hat{\rho}^2}+\frac{1}{\hat{\rho}^3}\right]\log_1^+\left(\frac{1}{\hat{\eta}}+\frac{1}{\hat{\rho}^2}\right)\right).$$

4 Proof of Theorem 3.4

This section presents several technical results that are needed to establish Theorem 3.4. It is divided into two subsections. The first one presents a bound on the sequence of multipliers $\{p_k\}$ generated by the NL-IAPIAL, while the second one is devoted to proving Theorem 3.4.

4.1 Bounding on the Sequence of Lagrangian Multipliers

The goal of this subsection is to show that the sequence of multipliers $\{p_k\}$ generated by NL-IAPIAL is bounded.

The first result presents a key inclusion and some basic bounds on several residuals generated by NL-IAPIAL.

Lemma 4.1. *Let (r_k, w_k, δ_k) be as in Proposition 3.3, and consider D_h and τ as in assumptions (B1) and (B4), respectively. Moreover, define*

$$\xi_k = w_k - \nabla f(z_k) - \nabla g(z_k)p_k. \tag{43}$$

Then, it holds that

$$\xi_k \in \partial_{\delta_k} h(z_k), \quad \|r_k\| \le \frac{D_h}{1-\sigma}, \quad \delta_k \le \frac{1}{2\lambda}\left(\frac{\sigma D_h}{1-\sigma}\right)^2, \quad \|w_k\| \le \frac{(1+\sigma)D_h}{\lambda(1-\sigma)}. \tag{44}$$

Proof. Using Proposition 3.3(a) and the definition of ξ_k yields the required inclusion. On the other hand, the definitions of σ_{k-1} in (26), the inequality in (28), and the fact that $z_k, z_{k-1} \in \mathcal{H}$ imply that

$$\|r_k\| = \|v_k + z_{k-1} - z_k\| \le \|v_k\| + D_h \le \sigma_{k-1}\|r_k\| + D_h \le \sigma\|r_k\| + D_h,$$

which, after a simple re-arrangement, yields the desired bound on $\|r_k\|$. Consequently, the definition of δ_k, the aforementioned bound on $\|r_k\|$, the inequality in (28), and the definition of σ_{k-1} gives the bound on δ_k. Finally, the definitions of w_k and σ_{k-1} together with Proposition 3.3(a) yield

$$\|w_k\| \le \frac{1}{\lambda}\left(1+\sigma_{k-1}\sqrt{L_{k-1}^{\psi}}\right)\|r_k\| = \frac{1+\sigma}{\lambda}\|r_k\|,$$

which, combined with the previous bound on $\|r_k\|$, gives the desired bound on $\|w_k\|$. □

The next result presents some important properties about the iterates generated by NL-IAPIAL.

Lemma 4.2. *Let $\{(z_k, p_k, c_k)\}$ be generated by NL-IAPIAL and define, for every $k \geq 1$,*

$$s_k := \Pi_{-\mathcal{K}}(p_{k-1} + c_k g(z_k)). \tag{45}$$

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

$$p_{k-1} + c_k g(z_k) = p_k + s_k, \quad \langle p_k, s_k \rangle = 0, \quad (p_k, s_k) \in \mathcal{K}^* \times (-\mathcal{K}), \tag{46}$$

$$\mathcal{L}_{c_k}(z_k, p_{k-1}) = \phi(z_k) + \frac{1}{2c_k}\left(\|p_k\|^2 - \|p_{k-1}\|^2\right). \tag{47}$$

Proof. The two identities in (46) follow from the definitions of p_k and s_k in (4) and (45), respectively, the fact that $(\mathcal{K}^*)^\circ = -\mathcal{K}$, and [29, Exercise 2.8] with $\mathcal{K} = \mathcal{K}^*$ and $x = p_{k-1} + c_k g(z_k)$. In view of the definitions of $\mathcal{L}_c$ in (2) and s_k in (45), we have

$$\mathcal{L}_{c_k}(z_k, p_{k-1}) = \phi(z_k) + \frac{1}{2c_k}\left[\|p_{k-1} + c_k g(z_k) - s_k\|^2 - \|p_{k-1}\|^2\right]$$

which, in view of the first identity in (46), immediately implies (47). □

The following technical result, whose proof can be found in [22, Lemma 4.7], plays an important role in the proof of Proposition 4.4 below.

Lemma 4.3. *Let h be a function as in (B1). Then, for every $u, z \in \mathcal{H}$, $\delta > 0$, and $\xi \in \partial_\delta h(z)$, we have*

$$\|\xi\|\text{dist}(u, \partial\mathcal{H}) \leq [\text{dist}(u, \partial\mathcal{H}) + \|z - u\|]\, K_h + \langle \xi, z - u \rangle + \delta,$$

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

We are now ready to prove the main result of this subsection, namely, that the sequence $\{p_k\}_{k\geq 1}$ is bounded.

Proposition 4.4. *Consider the sequence $\{(p_k, c_k)\}$ generated by NL-IAPIAL and let κ_0, τ, and $\bar{d}$ be as in (36), (B4) and (39), respectively. Then, the following statements hold:*

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

$$\min\{1, \bar{d}\}\tau\|p_k\| + \frac{\|p_k\|^2}{c_k} \leq \kappa_0 + \frac{1}{c_k}\langle p_k, p_{k-1}\rangle;$$

(b) for every $k \geq 0$, we have

$$\|p_k\| \leq C_0 := \frac{\max\{\|p_0\|, \kappa_0\}}{\min\{1, \bar{d}\}\tau}. \tag{48}$$

Proof. (a) Let $\bar{z}$ be as in (B4). It follows from Lemma 4.1 that $\xi_k \in \partial_{\delta_k} h(z_k)$ for every $k \geq 1$. Hence, assumption (B1), the fact that $\bar{d} \leq D_h$ and $z_k \in \mathcal{H}$, Lemma 4.3 with $\xi = \xi_k$, $z = z_k$, $u = \bar{z}$ and $\delta = \delta_k$, and the bound on $\|\delta_k\|$ in Lemma 4.1, imply that

$$\bar{d}\|\xi_k\| \leq 2D_h K_h + \frac{\sigma^2 D_h^2}{2\lambda(1-\sigma)^2} + \langle \xi_k, z_k - \bar{z}\rangle. \tag{49}$$

On the other hand, using the assumption that g is $\mathcal{K}$-convex (see (B3)), the fact that $p_k \in \mathcal{K}^*$, the definition of ξ_k in (43), the bound on $\|w_k\|$ in Lemma 4.1, and the Cauchy-Schwarz inequality, we conclude that

$$\begin{aligned}
\langle \xi_k, z_k - \bar{z}\rangle &= \langle w_k - \nabla f(z_k) - \nabla g(z_k)p_k, z_k - \bar{z}\rangle \\
&= \langle w_k - \nabla f(z_k), z_k - \bar{z}\rangle + \langle p_k, g'(z_k)(\bar{z} - z_k)\rangle \\
&\le \langle w_k - \nabla f(z_k), z_k - \bar{z}\rangle + \langle p_k, g(\bar{z}) - g(z_k)\rangle \\
&\le B_f^{(1)} D_h + \frac{(1+\sigma)D_h^2}{\lambda(1-\sigma)} + \langle p_k, g(\bar{z}) - g(z_k)\rangle
\end{aligned} \tag{50}$$

where $B_f^{(1)}$ is as in (19). Now, defining

$$\kappa := \left[2K_h + B_f^{(1)}\right] D_h + \left[\frac{\sigma^2}{2(1-\sigma)^2} + \frac{1+\sigma}{1-\sigma}\right]\frac{D_h^2}{\lambda}, \tag{51}$$

and using (49), (50), together with the relations in (46), we conclude that

$$\bar{d}\|\xi_k\| - \langle p_k, g(\bar{z})\rangle \le \kappa - \langle p_k, g(z_k)\rangle = \kappa - \frac{1}{c_k}\langle p_k, s_k + p_k - p_{k-1}\rangle = \kappa - \frac{\|p_k\|^2}{c_k} + \frac{1}{c_k}\langle p_k, p_{k-1}\rangle$$

where s_k is as in (45). Noting that the definition of ξ_k and the reverse triangle inequality yield

$$\|\xi_k\| = \|\nabla f(z_k) - w_k + \nabla g(z_k)p_k\| \ge -\|\nabla f(z_k) - w_k\| + \|\nabla g(z_k)p_k\|,$$

it follows that

$$\bar{d}\|\nabla g(z_k)p_k\| - \langle p_k, g(\bar{z})\rangle \le \kappa - \frac{\|p_k\|^2}{c_k} + \frac{1}{c_k}\langle p_k, p_{k-1}\rangle + \bar{d}\|\nabla f(z_k) - w_k\|. \tag{52}$$

Using now the triangle inequality, assumption (B4), (51), (52), the fact that $\bar{d} \le D_h$, and the definition of κ_0 in (36), we finally conclude that

$$\min\{1, \bar{d}\}\tau\|p_k\| + \frac{\|p_k\|^2}{c_k} \le \kappa + B_f^{(1)} D_h + \frac{(1+\sigma)D_h^2}{\lambda(1-\sigma)} + \frac{1}{c_k}\langle p_k, p_{k-1}\rangle = \kappa_0 + \frac{1}{c_k}\langle p_k, p_{k-1}\rangle.$$

(b) This statement is proved by induction. Since $\tau \le 1$, inequality (48) trivially holds for $k = 0$. Assume that (48) holds with $k = i - 1$ for some $i \ge 1$. This assumption, together with the bound obtained in the latter result and the Cauchy-Schwarz inequality, then imply that

$$\begin{aligned}
\left(\min\{1, \bar{d}\}\tau + \frac{\|p_i\|}{c_i}\right)\|p_i\| &\le \kappa_0 + \frac{\|p_i\| \cdot \|p_{i-1}\|}{c_i} \le \kappa_0 + \frac{\|p_i\| C_0}{c_i} \\
&\le \left(\min\{1, \bar{d}\}\tau + \frac{\|p_i\|}{c_i}\right) C_0,
\end{aligned}$$

which implies that $\|p_i\| \le C_0$. Then, (48) also holds with $k = i$ and hence, by induction, we conclude that (48) holds for the whole sequence $\{p_k\}$. $\square$

4.2 Proving Theorem 3.4

The main goal of this subsection is to present the proof of Theorem 3.4.

The proof of Theorem 3.4 requires several technical results. The first one characterizes the change in the augmented Lagrangian between consecutive iterations of the NL-IAPIAL method.

Lemma 4.5. *The sequence $\{(z_k, p_k)\}$ generated by NL-IAPIAL satisfies the relations*

$$\mathcal{L}_{c_k}(z_k, p_k) \le \mathcal{L}_{c_k}(z_k, p_{k-1}) + \frac{1}{c_k}\|p_k - p_{k-1}\|^2, \tag{53}$$

$$\mathcal{L}_{c_k}(z_k, p_k) \le \mathcal{L}_{c_k}(z_{k-1}, p_{k-1}) - \left(\frac{1-\sigma^2}{2\lambda}\right)\|r_k\|^2 + \frac{1}{c_k}\|p_k - p_{k-1}\|^2, \tag{54}$$

for every $k \ge 1$, where (σ, λ) is given by the input of NL-IAPIAL and r_k is as in (29).

Proof. Let s_k be as in (45). Using (47), the definition of $\mathcal{L}_c$ in (2), the fact that $s_k \in -\mathcal{K}$ and $p_{k-1} + c_k g(z_k) = p_k + s_k$ in view of (46), we have that

$$\begin{aligned}
\mathcal{L}_{c_k}(z_k, p_k) - \mathcal{L}_{c_k}(z_k, p_{k-1}) &= \mathcal{L}_{c_k}(z_k, p_k) - \phi(z_k) - \frac{1}{2c_k}\left(\|p_k\|^2 - \|p_{k-1}\|^2\right) \\
&= \frac{1}{2c_k}\left(\mathrm{dist}^2(p_k + c_k g(z_k), -\mathcal{K}) - \|p_k\|^2\right) - \frac{1}{2c_k}\left(\|p_k\|^2 - \|p_{k-1}\|^2\right) \\
&\le \frac{1}{2c_k}\left(\|p_k + c_k g(z_k) - s_k\|^2 - \|p_k\|^2\right) - \frac{1}{2c_k}\left(\|p_k\|^2 - \|p_{k-1}\|^2\right) \\
&= \frac{1}{2c_k}\left(\|2p_k - p_{k-1}\|^2 - 2\|p_k\|^2 + \|p_{k-1}\|^2\right),
\end{aligned}$$

which immediately implies (53). Now, in view of the definition of the ε-subdifferential given in (7) and the fact that $(z_k, v_k, \varepsilon_k)$ satisfies both the inclusion and the inequality in (28), we conclude that

$$\begin{aligned}
&\lambda\mathcal{L}_{c_k}(z_k, p_{k-1}) - \lambda\mathcal{L}_{c_k}(z_{k-1}, p_{k-1}) \le -\frac{1}{2}\|z_k - z_{k-1}\|^2 + \langle v_k, z_k - z_{k-1}\rangle + \varepsilon_k \\
&= -\frac{1}{2}\|v_k + z_k - z_{k-1}\|^2 + \frac{1}{2}\|v_k\|^2 + \varepsilon_k \le -\left(\frac{1-\sigma_{k-1}^2}{2}\right)\|r_k\|^2 \le -\left(\frac{1-\sigma^2}{2}\right)\|r_k\|^2,
\end{aligned} \tag{55}$$

where the last inequality follows from the fact that $\sigma_{k-1} \le \sigma$ in view of (26). Inequality (54) now follows by combining (53) with (55). $\square$

Recall that the l-th cycle $\mathcal{C}_l$ of NL-IAPIAL is defined in (35). The next results present some properties of the iterates generated during an NL-IAPIAL cycle. The first one below establishes an upper bound on the augmented Lagrangian function along the iterates within a NL-IALPIAL cycle.

Lemma 4.6. *Consider the sequence $\{(z_k, p_k)\}_{k\in\mathcal{C}_l}$ generated during the l-th cycle of NL-IAPIAL. Then, for every $k \in \mathcal{C}_l$, we have*

$$\mathcal{L}_{\tilde{c}_l}(z_k, p_k) \le R_\phi + \phi_* + \frac{4C_0^2}{\tilde{c}_l}, \tag{56}$$

where ϕ_*, $\tilde{c}_l$, R_ϕ, and C_0 are as in (1), (35), (40), and (48), respectively.

Proof. First note that for any $k \in \mathcal{C}_l$, we have $c_k = \tilde{c}_l = 2^{l-1}c_1$. Moreover, $(\lambda, z_k, v_k, \varepsilon_k, \sigma_{k-1})$ satisfies the inclusion and the inequality in (28). Hence, it follows from Lemma A.3 with $s = 1$, $\tilde{\sigma} = \sigma_{k-1}$ and $\tilde{\phi} = \lambda\mathcal{L}_{\tilde{c}_l}(\cdot, p_{k-1})$, and assumption (B1) that for every $z \in \mathcal{H}$,

$$\lambda\mathcal{L}_{\tilde{c}_l}(z_k, p_{k-1}) + \frac{1-2\sigma_{k-1}^2}{2}\|r_k\|^2 \leq \lambda\mathcal{L}_{\tilde{c}_l}(z, p_{k-1}) + \|z - z_{k-1}\|^2 \leq \lambda\mathcal{L}_{\tilde{c}_l}(z, p_{k-1}) + D_h^2 \tag{57}$$

where r_{k_0} is as in (29) with $k = k_0$. Now, observe that the definitions of σ, σ_{k-1} and L_{k-1}^{ψ} (see NL-IAPIAL input and (26)) imply that $\sigma_{k-1} \leq \sigma \in (0, 1/\sqrt{2}]$ and that the definition of $\mathcal{L}_c$ in (2) implies that $\mathcal{L}_{\tilde{c}_l}(z, p_{k-1}) \leq \phi(z)$ for every $z \in \mathcal{F} := \{z \in \mathcal{H} : g(z) \preceq_{\mathcal{K}} 0\}$. Using then the definition of ϕ^* given in (1), the aforementioned observations, and the minimization of the right hand side of (57) with respect to $z \in \mathcal{F}$, we get

$$\mathcal{L}_{\tilde{c}_l}(z_k, p_{k-1}) \leq \phi^* + \frac{D_h^2}{\lambda} = R_\phi + \phi_*$$

where the last equality is due to the definition of R_ϕ in (40). Combining the above inequality, (53) and the fact that $\|u + u'\|^2 \leq 2\|u\|^2 + 2\|u'\|^2$, we have

$$\begin{aligned}\mathcal{L}_{\tilde{c}_l}(z_k, p_k) &\leq \mathcal{L}_{\tilde{c}_l}(z_k, p_{k-1}) + \frac{1}{\tilde{c}_l}\|p_k - p_{k-1}\|^2 \\ &\leq \mathcal{L}_{\tilde{c}_l}(z_k, p_{k-1}) + \frac{2}{\tilde{c}_l}\left(\|p_k\|^2 + \|p_{k-1}\|^2\right) \leq R_\phi + \phi_* + \frac{4C_0^2}{\tilde{c}_l},\end{aligned}$$

and hence the conclusion of the lemma follows. □

The next result presents some bounds on the sequences $\{\|r_k\|\}_{k\in\mathcal{C}_l}$ and $\{\Delta_k\}_{k\in\mathcal{C}_l}$.

Lemma 4.7. *Let $\{(z_k, v_k, \varepsilon_k)\}_{k\in\mathcal{C}_l}$ be generated during the l-th cycle of NL-IAPIAL and consider $\{r_k\}_{k\in\mathcal{C}_l}$ and $\{\Delta_k\}_{k\in\mathcal{C}_l}$ as in (26) and step 4, respectively. Then, for every $k \in \mathcal{C}_l$ such that $k \geq k_{l-1} + 2$, we have*

$$\min_{k_{l-1}+2\leq j\leq k}\|r_j\|^2 \leq \frac{2\lambda}{1-\sigma^2}\left(\Delta_k + \frac{4C_0^2}{\tilde{c}_l}\right), \tag{58}$$

$$\Delta_k \leq \frac{1}{k - k_{l-1} - 1}\left(R_\phi + \frac{9C_0^2}{2\tilde{c}_l}\right), \tag{59}$$

where C_0 is as in (48).

Proof. Relations (48), (54), the fact that $c_k = \tilde{c}_l$ for every $k \in \mathcal{C}_l$, and the inequality $\|p_k - p_{k-1}\|^2 \leq 2\|p_k\|^2 + 2\|p_{k-1}\|^2$, imply that for any $k \in \mathcal{C}_l$ such that $k \geq k_{l-1} + 2$ the following inequalities hold:

$$\begin{aligned}&\frac{(1-\sigma^2)(k - k_{l-1} - 1)}{2\lambda}\min_{k_{l-1}+2\leq j\leq k}\|r_j\|^2 \leq \frac{(1-\sigma^2)}{2\lambda}\sum_{j=k_{l-1}+2}^{k}\|r_j\|^2 \\ &\leq \mathcal{L}_{\tilde{c}_l}(z_{k_{l-1}+1}, p_{k_{l-1}+1}) - \mathcal{L}_{\tilde{c}_l}(z_k, p_k) + \frac{1}{\tilde{c}_l}\sum_{j=k_{l-1}+2}^{k}\|p_j - p_{j-1}\|^2 \\ &\leq \mathcal{L}_{\tilde{c}_l}(z_{k_{l-1}+1}, p_{k_{l-1}+1}) - \mathcal{L}_{\tilde{c}_l}(z_k, p_k) + \frac{4(k - k_{l-1} - 1)C_0^2}{\tilde{c}_l},\end{aligned}$$

and hence that (58) holds, in view of the definition of Δ_k in step 4 of NL-IAPIAL. Now, in view of the definitions of $\mathcal{L}_c$ and ϕ_* given in (2) and (41), respectively, we have

$$\mathcal{L}_{\tilde{c}_l}(z_k, p_k) = \phi(z_k) + \frac{1}{2\tilde{c}_l}\left[\mathrm{dist}^2(p_k + \tilde{c}_l g(z_k), -\mathcal{K}) - \|p_k\|^2\right] \geq \phi_* - \frac{\|p_k\|^2}{2\tilde{c}_l}.$$

It follows from the above inequality, (56) with $k = k_{l-1} + 1$, and the definition of Δ_k in step 4 of NL-IAPIAL that

$$\Delta_k \leq \frac{1}{k - k_{l-1} - 1}\left(R_\phi + \phi_* + \frac{4C_0^2}{\tilde{c}_l} + \frac{\|p_k\|^2}{2\tilde{c}_l} - \phi_*\right),$$

which proves (59) in view of (48). □

The next technical lemma presents some additional properties of the refined iterates generated by NL-IAPIAL.

Lemma 4.8. *Consider the sequences $\{(z_k, p_k, v_k, \varepsilon_k)\}_{k\in\mathcal{C}_l}$ and $\{(\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)\}_{k\in\mathcal{C}_l}$ generated during the l-th cycle of NL-IAPIAL. Then, the following statements hold:*

(a) for every $k \in \mathcal{C}_l$, the quadruple $(\hat{z}_k, \hat{p}_k, \hat{w}_k, \hat{q}_k)$ satisfies (33);

(b) for every $k \in \mathcal{C}_l$ and $k \geq k_{l-1} + 2$, there exists an index $i \in \{k_{l-1} + 2, \ldots, k\}$ such that

$$\|\hat{w}_i\|^2 \leq \frac{2(1+2\sigma)^2 R_\phi}{\lambda(1-\sigma^2)(k - k_{l-1} - 1)} + \frac{\kappa_3}{2\tilde{c}_l}, \quad \|\hat{q}_i\| \leq \frac{\kappa_4}{\tilde{c}_l}, \tag{60}$$

where R_ϕ is as in (40) and (κ_3, κ_4) is as in (38).

Proof. (a) In view of step 1 of NL-IAPIAL, we have that $(\lambda, z_{k-1}, p_{k-1}, \sigma_{k-1})$, and $(z_k, v_k, \varepsilon_k)$ satisfy (28). Hence, it follows from Proposition 3.3(b) and (c) that $(\hat{z}_k, \hat{w}_k, \hat{p}_k, \hat{q}_k)$ computed in step 2 of NL-IAPIAL satisfies (33).

(b) Let $k \in \mathcal{C}_l$ such that $k \geq k_{l-1} + 2$. In view of Lemma 4.7, there exists an index $i \in \{k_{l-1} + 2, \ldots, k\}$ such that

$$\|r_i\|^2 \leq \frac{2\lambda}{1-\sigma^2}\left[\frac{R_\phi}{k - k_{l-1} - 1} + \frac{4C_0^2}{\tilde{c}_l}\right]. \tag{61}$$

The bound on $\|\hat{w}_i\|^2$ now follows from combining (61), the first inequality in (33), the definitions of κ_3 and C_0 in (38) and (48), and the fact that $\sigma_{k-1}(L^\psi_{k-1})^{1/2} = \sigma$, in view of (26).

Now, recall that for any $k \in \mathcal{C}$ $c_k = \tilde{c}_l$. Hence, in view of the second inequality in (33), (48), the triangle inequality for norms, and the facts that $\sigma_{k-1}(L^\psi_{k-1})^{1/2} = \sigma$ and $L^\psi_{k-1} \geq \lambda\tilde{c}_l[B_g^{(1)}]^2$ (see (26)), we have

$$\begin{aligned}\|\hat{q}_i\| &\leq \frac{B_g^{(1)}\sigma_{k-1}}{\sqrt{L^\psi_{k-1}}}\|r_i\| + \frac{1}{\tilde{c}_l}\left(\|p_i\| + \|p_{i-1}\|\right) \leq \frac{B_g^{(1)}\sigma}{L^\psi_{k-1}}\|r_i\| + \frac{2C_0}{\tilde{c}_l} \\ &\leq \frac{\sigma D_h}{\lambda(1-\sigma)B_g^{(1)}\tilde{c}_l} + \frac{2C_0}{\tilde{c}_l} = \frac{\kappa_4}{\tilde{c}_l},\end{aligned} \tag{62}$$

where the last relation is due to the definitions of κ_4 and C_0 in (38) and (48), respectively. □

The next result establishes some bounds on the number of inner and outer iterations performed during an NL-IAPIAL cycle. It also shows that if the penalty parameter is sufficiently large, then NL-IAPIAL finds a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1).

Lemma 4.9. *Let R_ϕ and $\bar{c}(\hat{\rho}, \hat{\eta})$ be as in (40) and (41), respectively, and define*

$$\bar{L}_c^\psi := \lambda(\kappa_1 + c\kappa_2) + 1, \tag{63}$$

where (κ_1, κ_2) is as in (37). Then, the following statements about NL-IAPIAL hold:

(a) every outer iteration k within the l-th cycle performs $\mathcal{O}_1((\bar{L}_{\tilde{c}_l}^\psi)^{1/2} \log_1^+(\bar{L}_{\tilde{c}_l}^\psi))$ inner iterations;

(b) every cycle performs $\mathcal{O}_1\left(R_\phi/[\lambda\hat{\rho}^2]\right)$ outer iterations;

(c) if $\tilde{c}_l \geq \bar{c}(\hat{\rho}, \hat{\eta})$ then NL-IAPIAL must stop in the l-th cycle with a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1).

Proof. (a) First note that within the l-th cycle, $c_k = \tilde{c}_l$. Hence, in view of (48) and the definitions of L_{k-1}^ψ, $(\kappa_1,\, \kappa_2)$, and $\bar{L}_c^\psi$ given in (26), (37), and (63), we have

$$\begin{aligned} L_{k-1}^\psi &= \lambda\left[L_f + L_g\|p_{k-1}\| + c_k\left(B_g^{(0)}L_g + [B_g^{(1)}]^2\right)\right] + 1 \\ &\leq \lambda\left[L_f + L_g C_0 + \tilde{c}_l\left(B_g^{(0)}L_g + [B_g^{(1)}]^2\right)\right] + 1 = \lambda(\kappa_1 + \kappa_2\tilde{c}_l) + 1 = \bar{L}_{\tilde{c}_l}^\psi. \end{aligned}$$

Now, since NL-IAPIAL invokes in its step 1 the ACG algorithm with inputs $(M_s, \mu, \psi_s, \psi_n), x_0$, and $\tilde{\sigma}$ given by (27), the statement in (a) follows from Proposition 2.1, the above estimate and the fact that

$$\sqrt{\frac{M_s}{\mu}} \log^+\left[\left(1 + \tilde{\sigma}^{-1}\right)\sqrt{2M_s}\right] \leq \sqrt{2L_{k-1}^\psi}\log^+\left(\frac{2\sqrt{L_{k-1}^\psi}}{\sigma}\sqrt{2L_{k-1}^\psi}\right) \leq \sqrt{2L_{k-1}^\psi}\log^+\left(\frac{3L_{k-1}^\psi}{\sigma}\right).$$

(b) Fix a cycle l. It follows from (59) that, for every $k \in \mathcal{C}_l$, $k \geq k_{l-1} + 2$,

$$\Delta_k \leq \frac{1}{k - k_{l-1} - 1}\left(R_\phi + \frac{9C_0^2}{2\tilde{c}_l}\right).$$

Hence, since $\tilde{c}_l \geq c_1$, it is easy to see that if k satisfies

$$k > k_{l-1} + 1 + \frac{4(1+2\sigma)^2}{\lambda(1-\sigma^2)\hat{\rho}^2}\left(R_\phi + \frac{9C_0^2}{2c_1}\right)$$

then the condition in step 4 holds, ending the l-th cycle. Since the cycle starts at $k_{l-1}+1$, statement (b) follows immediately from the above bound.

(c) From the definition of $\bar{c}(\cdot,\cdot)$ in (41) and the fact that $\tilde{c}_l \geq \bar{c}(\cdot,\cdot)$, we have

$$\tilde{c}_l \geq \frac{\kappa_3}{\hat{\rho}^2}, \quad \tilde{c}_l \geq \frac{\kappa_4}{\hat{\eta}}, \tag{64}$$

where κ_3 and κ_4 are as in (38). Now, let $\bar{k} \geq k_{l-1} + 2$ be the smallest index such that

$$\frac{2(1+2\sigma)^2 R_\phi}{\lambda(1-\sigma^2)(\bar{k} - k_{l-1} - 1)} \leq \frac{\hat{\rho}^2}{2}. \tag{65}$$

Hence, in view of (64), (65), and Lemma 4.8(b), there exists an index $i \in \{k_{l-1}+2,\ldots,\bar{k}\}$ such that

$$\|\hat{w}_i\| \le \hat{\rho}, \quad \|\hat{q}_i\| \le \hat{\eta}$$

which implies that NL-IAPIAL must stop at iteration i, in view of step 4. Hence, the proof of the statement in (c) follows. □

We are now ready give the proof of Theorem 3.4.

Proof of Theorem 3.4. First, recall that in the l-th cycle of NL-IAPIAL, we have $c_k = \tilde{c}_l = 2^{l-1}c_1$, for every $l \ge 1$, see (35). Now, let $\bar{l}$ be the first index l such that $\tilde{c}_l \ge \bar{c}$, where $\bar{c}$ is as in (41). Hence, in view of Lemma 4.9(c), we see that NL-IAPIAL obtains a $(\hat{\rho},\hat{\eta})$-approximate solution of (1) within the $\bar{l}$-th cycle. Moreover, it follows by Lemma 4.9(a) and (b) that the total number of inner iterations performed by NL-IAPIAL is $\mathcal{O}_1(T_I)$ where

$$T_I := \frac{R_\phi}{\lambda\hat{\rho}^2}\sum_{l=1}^{\bar{l}}\sqrt{\bar{L}^{\psi}_{\tilde{c}_l}}\log_1^+(\bar{L}^{\psi}_{\tilde{c}_l}) \tag{66}$$

and $\bar{L}^{\psi}_{\tilde{c}_l}$ is as in (63). Since c_k is doubled every time the cycle is changed, we have in view of the definitions of $\tilde{c}_l$ and $\bar{l}$ that

$$\tilde{c}_l \le \max\{c_1, 2\bar{c}\}, \qquad \forall l = 1,\ldots,\bar{l}. \tag{67}$$

Hence, it follows from (63) that

$$\bar{L}^{\psi}_{\tilde{c}_l} = \lambda(\kappa_1+\tilde{c}_l\kappa_2)+1 \le [\lambda(\kappa_1+c_1\kappa_2)+1]\frac{\max\{c_1,2\bar{c}\}}{c_1}. \tag{68}$$

Moreover, (63), (67), and the fact that $\tilde{c}_l = 2^{l-1}c_1$, also imply that

$$\begin{aligned}\sum_{l=1}^{\bar{l}}\sqrt{\bar{L}^{\psi}_{\tilde{c}_l}} &= \sum_{l=1}^{\bar{l}}\sqrt{\lambda(\kappa_1+\tilde{c}_l\kappa_2)+1} \le \sqrt{\lambda(\kappa_1+c_1\kappa_2)+1}\sum_{l=1}^{\bar{l}}\sqrt{2}^{l-1} \\ &\le 8\sqrt{\lambda(\kappa_1+c_1\kappa_2)+1}\left(\frac{\tilde{c}_{\bar{l}}}{c_1}\right)^{1/2} \\ &\le 8\sqrt{\lambda(\kappa_1+c_1\kappa_2)+1}\left(\frac{\max\{c_1,2\bar{c}\}}{c_1}\right)^{1/2}.\end{aligned}$$

Hence, (42) then follows by combining (40), (66), (68), and the above inequalities. □

5 Numerical Experiments

This section presents numerical experiments that highlight the performance of NL-IAPIAL against two other benchmark methods for solving nonconvex composite problems with convex constraints. It contains three subsections. The first one presents the results on a set of linearly-constrained problems, the second one presents the results on a set of nonlinear cone-constrained problems, while the third one contains a summary and some comments.

Before proceeding, we describe the implementation details of the NL-IAPIAL and two other benchmark algorithms. Addition details about the experiments are then provided after these descriptions. It is worth mentioning that the code for generating the results of this section is available online[6].

[6]See the examples in `./tests/benchmarks/` from the GitHub repository https://github.com/wwkong/nc_opt/.

The NL-IAPIAL implementation differs from its description in Section 3 in four ways. First, it modifies the parameter $\tilde{\sigma}$ that is given to the ACG algorithm in its step 1. More specifically, instead of choosing $\tilde{\sigma} = \sigma_{k-1}$ at the k-th iteration, the implementation chooses $\tilde{\sigma} = \min\{\nu/(L_{k-1}^{\psi})^{1/2}, \sigma\}$ for some $\nu > 0$ and $\sigma \in (0, 1/\sqrt{2}]$. Second, in view of the first modification, it replaces condition (34) with the modified condition

$$\Delta_k \leq \frac{\lambda(1-\sigma^2)\hat{\rho}^2}{4(1+2\nu)^2},$$

where ν is as in the first modification. Third, it replaces the ACG algorithm with a more computationally efficient ACG variant. More specifically, this variant generates its iterates according to the ACG algorithm described in [14, Section 5], but keeps the stopping criterion of the ACG algorithm in Section 2. Fourth, instead of doubling c_k in its step 4 when Δ_k is sufficiently small, the implementation chooses to quintuple c_k. Finally, regarding (σ, ν) and the other hyperparameters of the method, the NL-IAPIAL implementation chooses

$$c_1 = \max\left\{1, \frac{L_f}{\left[B_g^{(1)}\right]^2}\right\}, \quad \lambda = \frac{1}{2m}, \quad \sigma = \frac{1}{\sqrt{2}}, \quad \nu = \sqrt{\sigma\left(\lambda L_f + 1\right)}, \quad p_0 = 0.$$

While we do not show how these modifications affect the convergence of this NL-IAPIAL variant, we do note that they can be analyzed using the techniques of this paper and those in [22].

The first of the two benchmark methods is an implementation of the QP-AIPP method of [11], with four modifications. First, instead of performing the prescribed number of ACG iterations for each penalty prox subproblem, it stops the ACG call when the ACG inequality

$$\|u_j\| + 2\eta_j \leq \sigma\|x_j - x_0 + u_j\|^2$$

is satisfied and returns the triple (x_j, u_j, η_j). Second, if $(\hat{z}_\ell, \hat{v}_\ell)$ is the iterate generated by the ℓ-th AIPP call, where $\hat{z}_0 = z_0$, then it sets the starting point of $(\ell+1)$-th AIPP call to be $\hat{z}_\ell$ rather than z_0, i.e., it implements a warm-start strategy for the starting point. Third, it replaces the ACG algorithm in the QP-AIPP method with the ACG variant described in the previous paragraph, i.e., the one used for the NL-IAPIAL implementation. Fourth, it chooses the first penalty parameter to be $c_0 = \max\left\{1, \hat{c} + L_f/[B_g^{(1)}]^2\right\}$. With regards to the other hyperparameters of the method, the QP-AIPP implementation chooses $\hat{c} = 0$ and $\sigma = 0.3$.

The second of the two benchmark methods is an implementation of the iALM method of [17]. Regarding the hyperparameters of the method, the implementation chooses

$$\sigma = 5, \quad \beta_0 = \max\left\{1, \frac{L_f}{\left[B_g^{(1)}\right]^2}\right\}, \quad w_0 = 1, \quad \boldsymbol{y}^0 = 0, \quad \gamma_k = \frac{(\log 2)\|c(x^1)\|}{(k+1)\left[\log(k+2)\right]^2},$$

for every $k \geq 1$. Moreover, the starting point given to the k-th APG call is set to be $\boldsymbol{x}^{k-1}$ which is the prox center for the k-th prox subproblem. For problems of the form

$$\min_x \left\{f(x) + h(x) : c(x) \preceq_{\mathcal{K}} 0\right\},$$

where $\mathcal{K}$ is a pointed convex cone, the method attempts to solve the equivalent problem with equality constraints under an additional slack variable. More specifically, it introduces an additional slack variable s, changes the problem to be

$$\min_{x,s} \left\{f(x) + h(x) : c(x) + s = 0, s \succeq_{\mathcal{K}} 0\right\},$$

and adds the an indicator function of the constraint $s \succeq_{\mathcal{K}} 0$ to the nonsmooth component of the composite objective function.

We now describe some common details about the experiments. First, throughout this section, we denote I to be the identity matrix, $\mathbb{S}^n$ to be the set of symmetric n-by-n matrices, $\mathbb{S}^n_+$ to be the set of positive semidefinite matrices in $\mathbb{S}^n$, and δ_S to be the set-dependent indicator function that is 1 if its argument is in the set S and $+\infty$ otherwise. Second, given a tolerance pair $(\hat{\rho}, \hat{\eta}) \in \Re^2_{++}$, a pointed convex cone $\mathcal{K}$, and $z_0 \in \operatorname{dom} h$, all of the methods attempt to find a quadruple $(\hat{z}, \hat{p}, \hat{w}, \hat{q})$ satisfying

$$\begin{gathered} \hat{w} \in \nabla f(\hat{z}) + \partial h(\hat{z}) + \nabla g(\hat{z})\hat{p}, \quad \langle g(\hat{z}) + \hat{q}, \hat{p} \rangle = 0, \quad g(\hat{z}) + \hat{q} \preceq_{\mathcal{K}} 0, \quad \hat{p} \succeq_{\mathcal{K}^*} 0, \\ \|\hat{w}\| \leq \frac{\hat{\rho}}{1 + \|\nabla f(z_0)\|}, \quad \|\hat{q}\| \leq \frac{\hat{\eta}}{1 + \|g(z_0) - \Pi_{-\mathcal{K}}(g(z_0))\|}. \end{gathered} \tag{69}$$

Third, the iterations counted in these experiments are the number of ACG iterations needed to obtain a quadruple satisfying (69). Moreover, these iterations include those that are performed by parameter line searches, such as in the ACG variant used by the NL-IAPIAL and QP-AIPP implementations. Fourth, the bold numbers in each of the tables of this section indicate the method that performed the most efficiently for a given metric, e.g., runtime or iteration count. Fifth, each method is run with a time limit of 10000 seconds. If a method does not terminate with a solution for a particular problem instance, we do not report its iteration count at the point of termination and the runtime for that instance is marked with a [*] symbol. Finally, all algorithms described at the beginning of this section are implemented in MATLAB 2020a and are run on Linux 64-bit machines each containing Xeon E5520 processors and at least 8 GB of memory.

5.1 Linearly-Constrained Problems

This subsection examines the performance of the NL-IAPIAL method on a set of linearly-constrained nonconvex composite problems.

5.1.1 Nonconvex QSDP

Given a pair of dimensions $(\ell, n) \in \mathbb{N}^2$, a scalar pair $(\alpha, \beta) \in \Re^2_{++}$, linear operators $\mathcal{A} : \mathbb{S}^n_+ \mapsto \Re^\ell$, $\mathcal{B} : \mathbb{S}^n_+ \mapsto \Re^n$, and $\mathcal{C} : \mathbb{S}^n_+ \mapsto \Re^\ell$ defined pointwise by

$$[\mathcal{A}(Z)]_i = \langle A_i, Z \rangle, \quad [\mathcal{B}(Z)]_j = \langle B_j, Z \rangle, \quad [\mathcal{C}(Z)]_i = \langle C_i, Z \rangle,$$

for matrices $\{A_i\}_{i=1}^\ell, \{B_j\}_{j=1}^n, \{C_i\}_{i=1}^\ell \subseteq \Re^{n \times n}$, positive diagonal matrix $D \in \Re^{n \times n}$, and a vector pair $(b, d) \in \Re^\ell \times \Re^\ell$, this sub-subsection considers the following nonconvex quadratic semidefinite programming (QSDP) problem:

$$\begin{aligned} \min_z \ & -\frac{\alpha}{2}\|D\mathcal{B}(Z)\|^2 + \frac{\beta}{2}\|\mathcal{C}(Z) - d\|^2 \\ \text{s.t. } & \mathcal{A}(Z) = b, \\ & Z \in P^n, \end{aligned}$$

where $P^n = \{Z \in \mathbb{S}^n_+ : \operatorname{tr} Z = 1\}$ denotes the n-dimensional spectraplex.

We now describe the experiment parameters for the instances considered. First, the dimensions were set to $(\ell, n) = (10, 50)$ and only 1% of the entries of A_i, B_j, and C_i were set to be nonzero. Second, the entries of A_i, $, B_j$, C_i, and d (resp. D) were generated by sampling from the uniform distribution $\mathcal{U}[0,1]$ (resp. $\mathcal{U}\{1,...,1000\}$). Third, the vector b was set to $b = \mathcal{A}(I/n)$. Fourth, the initial starting point z_0 was chosen to be a random point in $\mathbb{S}^n_+$. More specifically, three unit vectors $v_1, v_2, v_3 \in \Re^n$ and three scalars $e_1, e_2, e_3 \in \Re_+$ were first generated by sampling vectors $\tilde{v}_i \sim \mathcal{U}^n[0,1]$ and scalars $\tilde{d}_i \sim \mathcal{U}[0,1]$ and setting $v_i = \tilde{v}_i/\|\tilde{v}_i\|$ and $e_i = \tilde{e}_i/(\sum_{j=1}^3 e_j)$ for $i = 1, 2, 3$. The starting point was then set to $z_0 = \sum_{i=1}^3 e_i v_i v_i^T$. Fifth, with respect to the termination criterion (69) the relevant functions and quantities were

$$f(z) = -\frac{\alpha}{2}\|D\mathcal{B}(Z)\|^2 + \frac{\beta}{2}\|\mathcal{C}(Z) - d\|^2, \quad h(z) = \delta_{P^n}(z), \quad g(z) = \mathcal{A}(Z) - b,$$
$$\mathcal{K} = \{0\}, \quad \hat{\rho} = 10^{-4}, \quad \hat{\eta} = 10^{-4}.$$

Sixth, using the fact that $\|Z\|_F \leq 1$ for every $Z \in P_n$, the constant hyperparameters for the NL-IAPIAL and iALM methods were

$$L_g = 0, \quad B_g^{(1)} = \|\mathcal{A}\|, \quad L_j = 0, \quad \rho_j = 0, \quad B_j = \|A_j\|_F$$

for $1 \leq j \leq \ell$. Finally, each problem instance considered was based on a specific pair (m, L_f) for which the scalar pair (α, β) is selected so that $L_f = \lambda_{\max}(\nabla^2 f)$ and $-m = \lambda_{\min}(\nabla^2 f)$.

We now present the experiment tables for this set of problem instances.

(m, L_f)		Iteration Count			Runtime		
m	L_f	iALM	QP-AIPP	NL-IAPIAL	iALM	QP-AIPP	NL-IAPIAL
10^1	10^2	65780	5548	**366**	485.30	74.70	**5.99**
10^1	10^3	34629	2490	**217**	235.58	29.69	**3.18**
10^1	10^4	54469	4145	**644**	366.25	52.10	**7.89**
10^1	10^5	136349	18218	**2181**	1211.00	279.61	**36.75**
10^1	10^6	371276	84797	**13931**	2166.40	813.79	**140.63**

Table 1: Iteration and Runtimes for Nonconvex QSDP Problems

5.2 Nonlinear Cone-Constrained Problems

This subsection examines the performance of NL-IAPIAL on a set of nonlinear cone-constrained nonconvex composite problems. Since the QP-AIPP method is only designed for problems with linear constraints, we do not include it as a benchmark method in this subsection.

5.2.1 Nonconvex QC-QSDP

Given a dimension pair $(\ell, n) \in \mathbb{N}^2$, matrices $P, Q, R \in \Re^{n \times n}$, and the quantities (α, β), $\mathcal{B}$, $\mathcal{C}$, $\{B_j\}_{j=1}^n, \{C_i\}_{i=1}^\ell$, D, d as in Sub-subsection 5.1.1, this sub-subsection considers the nonconvex quadratically constrained QSDP (QC-QSDP) problem:

$$\begin{aligned}
\min_Z \; & -\frac{\alpha}{2}\|D\mathcal{B}(Z)\|^2 + \frac{\beta}{2}\|\mathcal{C}(Z) - d\|^2 \\
\text{s.t.} \; & \frac{1}{2}(PZ)^* PZ + \frac{1}{2}Q^* QZ + \frac{1}{2}ZQ^* Q \preceq R^* R, \\
& 0 \leq \lambda_i(Z) \leq \frac{1}{\sqrt{n}}, \qquad i \in \{1, ..., n\}, \\
& Z \in \mathbb{S}^n_+,
\end{aligned}$$

where $\lambda_i(Z)$ denotes i^{th} largest eigenvalue of Z and the constraint $M \preceq 0$ is equivalent to $-M \in \mathbb{S}^n_+$.

We now describe the experiment parameters for the instances considered. First, the dimensions were set to $(\ell, n) = (10, 50)$. Second, the quantities $\mathcal{B}$, $\mathcal{C}$, D, and d were generated in the same way as in Sub-subsection 5.1.1. On the other hand, the matrix R was set to be I/n and the entries of matrices P and Q were sampled from the uniform distributions $\mathcal{U}[0, 1/\sqrt{n}]$ and $\mathcal{U}[0, 1/n]$, respectively. Third, the initial starting point z_0 was set to be the zero matrix. Fourth, with respect to the termination criterion (69) the relevant functions and quantities were

$$f(z) = -\frac{\alpha}{2}\|D\mathcal{B}(Z)\|^2 + \frac{\beta}{2}\|\mathcal{C}(Z) - d\|^2, \quad h(z) = \delta_S(z),$$
$$g(z) = \frac{1}{2}(PZ)^*PZ + \frac{1}{2}Q^*QZ + \frac{1}{2}ZQ^*Q \preceq R^*R,$$
$$\mathcal{K} = \mathbb{S}^n_+, \quad \hat{\rho} = 10^{-3}, \quad \hat{\eta} = 10^{-3},$$

where $S = \{Z \in \mathbb{S}^n_+ : 0 \le \lambda_i(Z) \le 1/\sqrt{n}, i = 1, ..., n\}$. Fifth, using the fact that $\|Z\|_F \le 1$ for every $Z \in S$, the constant hyperparameters for the NL-IAPIAL and iALM methods were

$$L_g = \|P\|_F^2, \quad B_g^{(1)} = \frac{1}{2}\|P\|_F^2 + \|Q\|_F^2,$$
$$L_{ij} = \left|[P^*P]_{ij}\right|, \quad \rho_{ij} = 0, \quad B_j = \frac{1}{2}\left|[P^*P]_{ij}\right| + \left|[Q^*Q]_{ij}\right|,$$

for $1 \le i, j \le n$. Finally, like in Sub-subsection 5.1.1, each problem instance considered was based on a specific pair (m, L_f) for which the scalar pair (α, β) is selected so that $L_f = \lambda_{\max}(\nabla^2 f)$ and $-m = \lambda_{\min}(\nabla^2 f)$.

We now present the experiment tables for this set of problem instances.

(m, L_f)		Iteration Count			Runtime		
m	L_f	iALM	QP-AIPP	NL-IAPIAL	iALM	QP-AIPP	NL-IAPIAL
10^1	10^2	261486	-	**134740**	2196.30	-	**1941.00**
10^1	10^3	318159	-	**7447**	2649.80	-	**100.12**
10^1	10^4	686484	-	**27235**	5306.20	-	**309.10**
10^1	10^5	-	-	**140028**	10000.00*	-	**1607.20**
10^1	10^6	-	-	**482578**	10000.00*	-	**6312.00**

Table 2: Iteration and Runtimes for Nonconvex QC-QSDP Problems

5.3 Summary and Comments

This subsection gives a summary and some comments about the numerical experiments.

The most efficient method overall is NL-IAPIAL. For the linearly constrained problems, the second most efficient one is the QP-AIPP method. Compared to the iALM method, both NL-IAPIAL and QP-AIPP methods perform significantly fewer iterations to reach an approximate solution as in (69). On the other hand, for the problem in Sub-subsection 5.2.1, NL-IAPIAL requires more time per iteration than iALM.

We now give some reasons for the results. Before proceeding, we first remark that iALM, NL-IAPIAL, and QP-AIPP are all methods that solve a sequence of convex prox subproblems using an ACG algorithm. First, we saw that QP-AIPP performs fewer iterations than iALM because it uses a weaker termination criterion when solving its prox subproblems. Second, we conjecture that NL-IAPIAL performs fewer iterations than QP-AIPP and iALM because it performs its Lagrange multiplier update more often. Finally, for the problem in Sub-subsection 5.2.1, we saw that NL-IAPIAL uses more time per iteration than iALM because the ACG iterations in NL-IAPIAL contain more floating point operations than the corresponding ACG iterations in iALM.

It is worth mentioning that a more computationally efficient variant of the QP-AIPP method is presented in [12], in which the prox stepsize parameter λ is chosen adaptively and the corresponding prox subproblems are not necessarily convex. Since an analogous extension for augmented Lagrangian-based methods is not immediate, we leave the development of an adaptive NL-IAPIAL variant for a future work.

Appendix A Convex Analysis

This appendix presents some results from convex analysis.

The first result presents some well-known (see, for example, [4] and [28, Example 11.4]) properties about the projection and distance functions over a closed convex set.

Lemma A.1. *Let $\mathcal{K} \subseteq \Re^n$ be a closed convex cone. Then the following properties hold:*

(a) for every $u, z \in \Re^n$, we have $\|\Pi_S(u) - \Pi_S(u)\| \le \|u - z\|$;

(b) the function $d(\cdot) := \mathrm{dist}^2(\cdot, S)/2$ is differentiable, and its gradient is given by

$$\nabla d(u) = u - \Pi_S(u) \in N_S(\Pi_S(u)) \quad \forall u \in \Re^n; \tag{70}$$

(c) it holds that $u \in N_{\mathcal{K}^*}(p)$ if and only if $\langle u, p\rangle = 0$, $u \in -\mathcal{K}$, and $p \in \mathcal{K}^*$.

The next result presents a well-known fact (see, for example, [6, Sub-subsection 2.13.2]) about closed convex cones.

Lemma A.2. *For any closed convex cone $\mathcal{K}$, we have that $x \in \operatorname{int} \mathcal{K}$ if and only if*

$$\inf_p \left\{\langle p, x\rangle : p \in \mathcal{K}^*,\ \|p\| = 1\right\} > 0. \tag{71}$$

The below technical result presents a fact about approximate subdifferentials, and its proof can be found, for example, in [22, Lemma A.2].

Lemma A.3. *Let proper function $\tilde\phi : \Re^n \to (-\infty, \infty]$, scalar $\tilde\sigma \in (0,1)$ and $(z_0, z_1) \in \Re^n \times \operatorname{dom} \tilde\phi$ be given, and assume that there exists (v_1, ε_1) such that*

$$v_1 \in \partial_{\varepsilon_1}\left(\tilde\phi + \frac{1}{2}\|\cdot - z_0\|^2\right)(z_1), \quad \|v_1\|^2 + 2\varepsilon_1 \le \tilde\sigma^2\|v + z_0 - z_1\|^2. \tag{72}$$

Then, for every $z \in \Re^n$ and $s > 0$, we have

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

Appendix B Proof of Proposition 3.3

This appendix gives the proof of Proposition 3.3.

Before giving the proof of Proposition 3.3, we first present a technical result that describes some properties of a composite gradient step.

Lemma B.1. *Let $\psi_n \in \overline{\text{Conv}}\ \Re^n$, $z \in \text{dom}\,\psi_n$, and ψ_s be a differentiable function on* $\text{dom}\,h$ *which satisfies*

$$\psi_s(u) - \ell_{\psi_s}(u; z) \le \frac{L}{2}\|u - z\|^2$$

for some $L \ge 0$ and every $u \in \text{dom}\,\psi_s$. Moreover, define the quantities

$$\hat{z} := \underset{u}{\text{argmin}} \left\{ \ell_{\psi_s}(u; z) + \psi_n(u) + \frac{L}{2}\|u - z\|^2 \right\}, \quad \hat{e} := L(z - \hat{z}), \tag{73}$$

$$\zeta := \psi_n(z) - \psi_n(\hat{z}) - \langle \hat{e} - \nabla\psi_s(z), z - \hat{z} \rangle, \quad \Delta := (\psi_n + \psi_s)(z) - (\psi_n + \psi_s)(\hat{z}). \tag{74}$$

Then, the following statements hold:

(a) it holds that

$$\hat{e} \in \nabla\psi_s(z) + \partial_\zeta \psi_n(z), \quad \zeta \ge 0, \quad \zeta + \frac{1}{2L}\|\hat{e}\|^2 \le \Delta;$$

(b) $\hat{e} \in \nabla\psi_s(z) + \partial\psi_n(\hat{z})$;

(c) for any $\varepsilon > 0$ satisfying $0 \in \partial_\varepsilon(\psi_s + \psi_n)(z)$, we have $\Delta \le \varepsilon$.

Proof. The proof of parts (a) and (b) can be found, for example, in [11, Lemma 19]. To show part (c), observe that the assumed inclusion, (7), and the definition of Δ yield

$$\varepsilon \ge (\psi_n + \psi_s)(z) - (\psi_n + \psi_s)(\hat{z}) = \Delta.$$

□

We are now ready to prove Proposition 3.3 from the main body of the paper.

Proof of Proposition 3.3. (a) First, let $\widetilde{\mathcal{L}}_{c_k}$ and L^ψ_{k-1} be as in (24) and (26), respectively, and define

$$\psi_s := \lambda\widetilde{\mathcal{L}}_{c_k}(\cdot; p_{k-1}) - \langle v_k, \cdot \rangle + \frac{1}{2}\|\cdot - z_{k-1}\|^2, \quad \psi_n := \lambda h. \tag{75}$$

Note that, in view of assumption (B1), Lemma 3.2, and the definition of L^ψ_{k-1} in (26), the functions pair (ψ_s, ψ_n) defined above satisfies the assumptions of Lemma B.1 with $L = L^\psi_{k-1}$. Note also that the element $\hat{z}$ computed according to (73) corresponds to $\hat{z}_k$ computed as in (30), in view of the definition of r_k given in (29). Moreover, since $(\lambda, z_{k-1}, p_{k-1}, \sigma_{k-1})$ and $(z_k, v_k, \varepsilon_k)$ satisfy (28), it can be easily seen that the inclusion in Lemma B.1(c) holds with $\varepsilon = \varepsilon_k$, and hence, it follows by combining Lemma B.1(a) and (c) that

$$L^\psi_{k-1}(z_k - \hat{z}_k) \in \nabla\psi_s(z_k) + \partial_{\varepsilon_k}\psi_n(z_k), \quad L^\psi_{k-1}\|z_k - \hat{z}_k\| \le \sqrt{2\varepsilon_k L^\psi_{k-1}}, \tag{76}$$

where the above inclusion we used that $\partial_\zeta\psi_n(z_k) \subset \partial_{\varepsilon_k}\psi_n(z_k)$, in view of relation (7) and the fact that $\zeta \le \varepsilon_k$. Now note that the definition of r_k in (29) and Lemma 3.2(a) imply that

$$\begin{aligned} \nabla\psi_s(u) &= \lambda\nabla_z\widetilde{\mathcal{L}}_{c_k}(u; p_{k-1}) - r_k \\ &= \lambda\left[\nabla f(u) + \nabla g(u)\Pi_{\mathcal{K}^*}(p_{k-1} + c_k g(u))\right] - r_k \quad \forall u \in \Re^n. \end{aligned} \tag{77}$$

Hence, in view of (4), (7), the definitions of w_k and δ_k given in (29), the definition of ψ_n in (75), the inclusion in (76), and (77), we have

$$
\begin{aligned}
w_k = \frac{1}{\lambda}\left[r_k + L^{\psi}_{k-1}(z_k - \hat{z}_k)\right] &\in \frac{1}{\lambda}\left[r_k + \nabla\psi_s(z_k) + \partial_{\varepsilon_k}\psi_n(z_k)\right] \\
&= \nabla f(z_k) + \nabla g(z_k)\Pi_{\mathcal{K}^*}(p_{k-1} + c_k g(z_k)) + \partial_{\varepsilon_k/\lambda} h(z_k) \\
&= \nabla f(z_k) + \nabla g(z_k) p_k + \partial_{\delta_k} h(z_k),
\end{aligned}
$$

which proves the inclusion in (32). We now show that the inequalities in (32) hold. The bound on δ_k in (32) follows immediately from the inequality in (28) and the definitions of δ_k and r_k given in (29). Now, it follows from the inequality in (28), the definitions of w_k and r_k in (29), the triangle inequality for norms, and the inequality in (76) that

$$
\begin{aligned}
\lambda\|w_k\| = \|r_k + L^{\psi}_{k-1}(z_k - \hat{z}_k)\| &\le \|r_k\| + L^{\psi}_{k-1}\|z_k - \hat{z}_k\| \\
&\le \|r_k\| + \sqrt{2\varepsilon_k L^{\psi}_{k-1}} \le \left(1 + \sigma_{k-1}\sqrt{L^{\psi}_{k-1}}\right)\|r_k\|,
\end{aligned} \tag{78}
$$

which immediately implies the desired bound on $\|w_k\|$.

(b) We first show that the inclusion in (33) holds. Using (77), Lemma B.1(b), the definition of $\hat{z}_k$ in (30), Lemma 3.2(b), and the definitions of w_k and $(\hat{w}_k, \hat{p}_k)$ in (29) and (31), respectively, we have

$$
\begin{aligned}
\hat{w}_k &= \frac{1}{\lambda}\left[r_k + L^{\psi}_{k-1}(z_k - \hat{z}_k)\right] + \left[\nabla_z \widetilde{\mathcal{L}}_{c_k}(\hat{z}_k; p_{k-1}) - \nabla_z \widetilde{\mathcal{L}}_{c_k}(z_k; p_{k-1})\right] \\
&\in \frac{1}{\lambda}\left[r_k + \nabla\psi_s(z_k) + \partial\psi_n(\hat{z}_k)\right] + \left[\nabla_z \widetilde{\mathcal{L}}_{c_k}(\hat{z}_k; p_{k-1}) - \nabla_z \widetilde{\mathcal{L}}_{c_k}(z_k; p_{k-1})\right] \\
&= \nabla_z \widetilde{\mathcal{L}}_{c_k}(\hat{z}_k; p_{k-1}) + \partial h(\hat{z}_k) = \nabla f(\hat{z}_k) + \nabla g(\hat{z}_k)\Pi_{\mathcal{K}^*}(p_{k-1} + c_k g(\hat{z}_k)) + \partial h(\hat{z}_k) \\
&= \nabla f(\hat{z}_k) + \nabla g(\hat{z}_k)\hat{p}_k + \partial h(\hat{z}_k),
\end{aligned}
$$

which is the desired inclusion in (33). We now show that the bound on $\|\hat{w}_k\|$ in (33) holds. Using the definition of $\hat{w}_k$ in (31), Lemma 3.2(b), the definition of L^{ψ}_{k-1}, the inequality in (28) together with the definition of r_k given in (29), the inequality in (76), the triangle inequality for norms, and (78), we have

$$
\begin{aligned}
\lambda\|\hat{w}_k\| &\le \lambda\|w_k\| + \lambda\|\nabla_z \widetilde{\mathcal{L}}_{c_k}(\hat{z}_k; p_{k-1}) - \nabla_z \widetilde{\mathcal{L}}_{c_k}(z_k; p_{k-1})\| \\
&\le \left(1 + \sigma_{k-1}\sqrt{L^{\psi}_{k-1}}\right)\|r_k\| + L^{\psi}_{k-1}\|\hat{z}_k - z_k\| \le \left(1 + 2\sigma_{k-1}\sqrt{L^{\psi}_{k-1}}\right)\|r_k\|,
\end{aligned}
$$

which immediately implies the desired bound on $\|\hat{w}_k\|$.

To show the bound on $\hat{q}_k$, we first use the definitions of $\hat{p}_k$ and p_k, the definition of $B^{(1)}_g$ given by (19), the last two inequalities in (78), the Mean Value Inequality, and Lemma A.1(a) to obtain

$$
\begin{aligned}
\frac{1}{c_k}\|\hat{p}_k - p_k\| &= \frac{1}{c_k}\left\|\Pi_{\mathcal{K}^*}\left(p_{k-1} + c_k g(\hat{z}_k)\right) - \Pi_{\mathcal{K}^*}\left(p_{k-1} + c_k g(z_k)\right)\right\| \le \frac{1}{c_k}\|c_k g(\hat{z}_k) - c_k g(z_k)\| \\
&\le \sup_{t\in[0,1]}\|\nabla g(t\hat{z}_k + [1-t]z_k)\| \cdot \|\hat{z}_k - z_k\| \le B^{(1)}_g\|\hat{z}_k - z_k\| \le \frac{B^{(1)}_g \sigma_{k-1}}{\sqrt{L^{\psi}_{k-1}}}\|r_k\|.
\end{aligned}
$$

Hence, using the triangle inequality for norms and the definition of $\hat{q}_k$ given in (31), we have

$$
\|\hat{q}_k\| = \frac{1}{c_k}\|\hat{p}_k - p_{k-1}\| \le \frac{1}{c_k}\|\hat{p}_k - p_k\| + \frac{1}{c_k}\|p_k - p_{k-1}\| \le \frac{B^{(1)}_g \sigma_{k-1}}{\sqrt{L^{\psi}_{k-1}}}\|r_k\| + \frac{1}{c_k}\|p_k - p_{k-1}\|.
$$

For the remaining relations in (33), first observe that Lemma A.1(b) with $u = p_{k-1} + c_k g(\hat{z}_k)$ and the definitions of $\hat{q}_k$ and $\hat{p}_k$ in (31) imply that

$$g(\hat{z}_k) + \hat{q}_k = \frac{1}{c_k}\left[p_{k-1} + c_k g(\hat{z}_k) - \hat{p}_k\right] \in N_{\mathcal{K}^*}(\hat{p}_k). \tag{79}$$

Combining the above relations and Lemma A.1(c) with $u = g(\hat{z}_k) + q_{k-1}$ and $p = \hat{p}_k$, we conclude that the remaining relations in (33) hold.

□

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. *First-order methods in optimization.* SIAM, 2017.

[5] D. Boob, Q. Deng, and G. Lan. Stochastic first-order methods for convex and nonconvex functional constrained optimization. *Available on arXiv:1908.02734*, 2019.

[6] M. Dattorro and J. Dattorro. *Convex Optimization & Euclidean Distance Geometry.* Meeboo Publishing, 2005.

[7] D. Hajinezhad and M. Hong. Perturbed proximal primal–dual algorithm for nonconvex nonsmooth optimization. *Math. Program.*, 176:207–245, 2019.

[8] 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.

[9] M. Hong. Decomposing linearly constrained nonconvex problems by a proximal primal dual approach: algorithms, convergence, and applications. *Available on arXiv:1604.00543*, 2016.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] W. Kong and R.D.C. Monteiro. Accelerated inexact composite gradient methods for nonconvex spectral optimization problems. *arXiv e-prints*, page arXiv:2007.11772, July 2020.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] Z. Lu and Z. Zhou. Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. *Available on arXiv:1803.09941*, 2018.

[22] J.G. Melo and R.D.C Monteiro. Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical lagrangian function and a full Lagrange multiplier update. *arXiv preprint arXiv:2008.00562*, 2020.

[23] 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*, 2020.

[24] 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.

[25] Y. E. Nesterov. *Introductory lectures on convex optimization : a basic course*. Kluwer Academic Publ., 2004.

[26] Y.E. Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, pages 1–37, 2012.

[27] 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.

[28] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Springer, Berlin, 1998.

[29] A. Ruszczynski. *Nonlinear optimization*. Princeton university press, 2011.

[30] 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.

[31] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. `http://www.mit.edu/~dimitrib/PTseng/papers.html`, 2008.

[32] Y. Xie and S.J. Wright. Complexity of proximal augmented lagrangian for nonconvex optimization with nonlinear equality constraints. *arXiv preprint arXiv:1908.00131*, 2019.

[33] Y. Xu. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. *Math. Program.*, 2019.

[34] 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.

[35] J. Zhang and Z.-Q. Luo. A proximal alternating direction method of multiplier for linearly constrained nonconvex optimization. *Available on arXiv:2006.16440*, 2020.