ITERATION-COMPLEXITY OF AN INEXACT PROXIMAL ACCELERATED AUGMENTED LAGRANGIAN METHOD FOR SOLVING LINEARLY CONSTRAINED SMOOTH NONCONVEX COMPOSITE OPTIMIZATION PROBLEMS

JEFFERSON G. MELO *, RENATO D.C. MONTEIRO †, AND HAIRONG WANG †

Abstract. This paper proposes and establishes the iteration-complexity of an inexact proximal accelerated augmented Lagrangian (IPAAL) method for solving linearly constrained smooth nonconvex composite optimization problems. Each IPAAL iteration consists of inexactly solving a proximal augmented Lagrangian subproblem by an accelerated composite gradient (ACG) method followed by a suitable Lagrange multiplier update. It is shown that IPAAL generates an approximate stationary solution in at most $\mathcal{O}(\log(1/\rho)/\rho^3)$ ACG iterations, where $\rho > 0$ is the given tolerance. It is also shown that the previous complexity bound can be sharpened to $\mathcal{O}(\log(1/\rho)/\rho^{2.5})$ under additional mildly stronger assumptions. The above bounds are derived assuming that the initial point is neither feasible nor the domain of the composite term of the objective function is bounded. Some preliminary numerical results are presented to illustrate the performance of the IPAAL method.

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

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

1. Introduction.

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

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

where $A : \Re^n \mapsto \Re^l$ is a linear operator, $b \in \Re^l$, $h : \Re^n \to (-\infty, \infty]$ is a closed proper convex function, and f is a real-valued differentiable (possibly nonconvex) function whose gradient is L–Lipschitz and which, for some $0 < m \leq L$, satisfies

$$f(u) \geq f(z) + \langle \nabla f(z), u - z\rangle - \frac{m}{2}\|u - z\|^2 \quad \forall\, z, u \in \mathrm{dom}\ h. \tag{2}$$

For a given tolerance pair $(\hat\rho, \hat\eta) \in \Re^2_{++}$, its goal is to find a triple $(\hat z, \hat p, \hat v)$ satisfying

$$\hat v \in \nabla f(\hat z) + \partial h(\hat z) + A^*\hat p, \qquad \|\hat v\| \leq \hat\rho, \qquad \|A\hat z - b\| \leq \hat\eta. \tag{3}$$

More specifically, the θ-IPAAL method is based on the θ-augmented Lagrangian (θ-AL) function $\mathcal{L}_c^\theta(z;p)$ defined as

$$\mathcal{L}_c^\theta(z;p) := f(z) + h(z) + (1-\theta)\langle p, Az - b\rangle + \frac{c}{2}\|Az - b\|^2, \tag{4}$$

where $\theta \geq 0$ is a given parameter. Note that when $\theta = 0$, $\mathcal{L}_c^\theta(\cdot,\cdot)$ reduces to the well known quadratic augmented Lagrangian function which has been thoroughly studied in the literature (see for example [3, 5, 24, 28, 39]). Moreover, when $\theta = 1$, $\mathcal{L}_c^\theta(\cdot,\cdot)$ does not depend on p and reduces to the quadratic penalty function frequently used in penalty methods for solving (1). Roughly speaking, for a fixed stepsize $\lambda > 0$, the static version of θ-IPAAL method repeatedly performs the following iteration: given $(z_{k-1}, p_{k-1}) \in \mathrm{dom}\, h \times \Re^l$, it computes (z_k, p_k) as

$$z_k \approx \mathrm{argmin}_z \left\{\lambda \mathcal{L}_c^\theta(z, p_{k-1}) + \frac{1}{2}\|z - z_{k-1}\|^2\right\} \tag{5}$$

$$p_k = (1-\theta)p_{k-1} + c(Az_k - b). \tag{6}$$

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

†School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. (E-mails: monteiro@isye.gatech.edu and hairongwhr@gatech.edu). The work of Renato D.C. Monteiro was partially supported by ONR Grant N00014-18-1-2077.

where z_k in (5) should be understood as a suitable approximate solution of the underlying θ-prox-AL subproblem. To complete the above outline of the θ-IPAAL method, we now describe how z_k is computed without elaborating on its inexactness. It can be easily seen that (2) implies that the objective function of (5) is strongly convex whenever $\lambda < 1/m$. The θ-IPAAL method then sets $\lambda = \tau/m$ for some suitably chosen $\tau \in (0,1)$ and then approximately solves the corresponding subproblem (5) by a strongly convex version of an accelerated composite gradient (ACG) method (see for example [4, 30, 35]) to obtain z_k. Also, it is shown that each pair (z_k, p_k) obtained in the above manner can be refined to a triple $(\hat{z}, \hat{v}, \hat{p}) = (\hat{z}_k, \hat{v}_k, \hat{p}_k)$ satisfying the inclusion in (3) and the static θ-IPAAL method is then stopped whenever the first inequality in (3) is also satisfied. Finally, the static θ-IPAAL method is shown to satisfy the following properties: 1) it stops in $\mathcal{O}(\sqrt{c}\log(c)/\hat{\rho}^2)$ ACG iterations; 2) for every $k \ge 1$, the refined iterate $\hat{z}_k$ satisfies $\|A\hat{z}_k - b\| = \mathcal{O}(1/\sqrt{c})$.

Observe that property 2) guarantees that $\hat{z}_k$ is a near feasible point, i.e., satisfies the second inequality in (3), only when c is sufficiently large. Based on this remark, a dynamic version of the θ-IPAAL method for finding a triple $(\hat{z}, \hat{v}, \hat{p})$ satisfying (3) is also considered. More specifically, it chooses an initial penalty parameter c and it repeatedly: a) invokes the static θ-IPAAL with the current c to obtain a triple $(\hat{z}, \hat{v}, \hat{p})$ satisfying the inclusion and the first inequality in (3); and b) doubles c whenever the second inequality in (3) is violated, until a triple $(\hat{z}, \hat{v}, \hat{p})$ satisfying (3) is obtained. It is then shown that the ACG iteration-complexity of this dynamic variant is $\mathcal{O}([1/(\hat{\eta}\hat{\rho}^2)]\log(1/\hat{\eta}))$. It is also shown that the previous complexity can be sharpened to $\mathcal{O}([1/(\sqrt{\hat{\eta}}\hat{\rho}^2)]\log(1/\hat{\eta}))$ under the mildly stronger assumptions that: a) $\operatorname{int}(\operatorname{dom} h) \cap \{z : Az = b\}$ is nonempty; b) for some $\bar{c} \ge 0$, the quadratic penalty function $\mathcal{L}^1_{\bar{c}}$ has bounded level set, and; c) h belongs to a special class of closed convex functions which contains all indicator functions of closed convex sets. Finally, it is worth emphasizing that all the results mentioned above are derived without assuming that the initial point $z_0 \in \operatorname{dom} h$ is feasible, i.e., satisfies $Az_0 = b$.

Related works. We first discuss papers dealing with related algorithms for solving the convex version of (1) and other related monotone problems. Iteration-complexity analysis of quadratic penalty methods for solving (1) under the assumption that f is convex and h is a convex indicator function was first studied in [23] and further explored in [2, 33]. Iteration-complexity of first-order augmented Lagrangian methods for solving the latter class of linearly constrained convex programs was studied in [3, 24, 28, 29, 38, 41]. Inexact proximal point methods using accelerated gradient algorithms to solve their prox-subproblems were previously considered in [9, 16, 15, 19, 31] in the setting of convex-concave saddle point problems and monotone variational inequalities.

We now discuss papers dealing with related algorithms for solving (1) when f is nonconvex and $A = 0$, i.e., the unconstrained version of (1). Paper [11] proposed an accelerated gradient framework to solve an unconstrained problem with better iteration-complexity than the usual composite gradient method. Since then, many authors have proposed other accelerated frameworks for solving the unconstrained counterpart of (1) under different assumptions on the functions f and h (see, for example, [7, 10, 12, 25, 37]). In particular, by exploiting the lower curvature m, [7, 10, 37] proposed some algorithms which improve the iteration-complexity bound of [11] in terms of the dependence on the Lipschitz constant. Finally, there has been a growing interest in the iteration-complexity of methods for solving optimization problems using second order information (see, for example, [7, 8, 32, 36]).

There are only a few papers analyzing iteration-complexity of quadratic penalty and/or augmented Lagrangian type methods for solving (1) in its general form, i.e., $A \neq 0$ and f nonconvex. This paragraph discusses the quadratic penalty type methods while the one below discusses the augmenetd Lagrangian type methods. Quadratic penalty type methods were studied in [20, 22, 27]. More specifically, paper [20] proposed a quadratic penalty accelerated inexact proximal point (QP-AIPP) method which can be viewed as an instance of the θ-IPAAL method outlined above with $\theta = 1$, and hence with (4) being the usual quadratic penalty function for (1). The QP-AIPP and the θ-IPAAL methods share similar ACG iteration-complexity bounds. More specifically, the ACG iteration-complexity of the QP-AIPP method is $\mathcal{O}(1/(\hat{\rho}^2\hat{\eta}))$ which is similar (up to a logarithm multiplicative term) to the first bound derived for the θ-IPAAL method mentioned above. Paper [22] proposes a more computationally efficient variant of QP-AIPP, namely, R-QP-AIPP, which adaptively chooses the prox-stepsize λ in a more aggressive manner and as a result generates possibly nonconvex subproblems (5) which are tentatively solved by a standard strongly convex version of an ACG variant. More recently, an alternative proximal quadratic penalty method for (1) whose subproblems are also approximately solved by a strongly convex version of an ACG variant is studied in [27]. An $\mathcal{O}(\log(1/\hat\eta)/(\hat\rho^2\sqrt{\hat\eta}))$ ACG-iteration-complexity is established for the method under the assumption that $\operatorname{dom} h$ is bounded and the Slater condition mentioned above holds. Finally, paper [21] studies the complexity of a quadratic penalty 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 $x \in \operatorname{dom} h$, and $\Phi(\cdot, y)$ is nonconvex differentiable on $\operatorname{dom} h$ and its gradient is uniformly Lipschitz continuous on $\operatorname{dom} h$ for every $y \in Y$.

Paper [17] studies the iteration-complexity of a linearized version of the augmented Lagrangian method to solve (1) but assumes the strong condition (among a few others) that $h = 0$, which most important problems arising in applications do not satisfy. Paper [14] studies an unaccelerated augmented Lagrangian inexact proximal method for (1) based on the θ-AL function (4) with θ chosen in $(0, 1]$ and establishes an $\mathcal{O}(1/(\hat\eta^4 + \hat\rho^4))$ iteration-complexity where each iteration exactly solves a subproblem of the form (5) except that the function f that appears (4) is replaced by its linearization at z_{k-1} and the prox term $\|z - z_{k-1}\|^2$ is replaced by $\|z - z_{k-1}\|^2_{B^*B}$ for some matrix B such that $B^*B + A^*A - I$ is positive semidefinite. In contrast to the θ-IPAAL method studied in this paper, their method requires the restrictive condition that its initial point x_0 be feasible, i.e., satisfy $Ax_0 = b$ and $x_0 \in \operatorname{dom} h$.

We now discuss three other alternative works dealing with complexity of first-order algorithms for solving (1). The first one [18] presents a penalty ADMM approach which introduces an artificial variable y in (1) and then penalizes y to obtain the penalized problem

$$\min\left\{ f(z) + h(z) + \frac{c}{2}\|y\|^2 : Ax + y = b \right\}, \tag{7}$$

which is then solved by a two-block ADMM. Since (7) satisfies the assumption that its y-block objective function component has Lipschitz continuous gradient everywhere and its y-block coefficient matrix is the identity, an iteration-complexity of the two-block ADMM for solving (7), and hence (1), can be established. More specifically, it has been shown in Remark 4.3 of [18] that the overall number of composite gradient steps performed by the aforementioned two-block ADMM penalty scheme to obtain a triple $(\hat z, \hat v, \hat p)$ satisfying (3) is bounded by $\mathcal{O}(\hat\rho^{-6})$ under the assumptions that $\hat\eta = \hat\rho$, the level sets of $f + h$ are bounded and the initial triple (z_0, y_0, p_0) is such that $(y_0, p_0) = (0, 0)$ and z_0 is feasible.

The second one [26] studies a hybrid penalty based and 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. More specifically, the latter ones are performed: 1) at the initial stage of the method in order to provide an initial value for the constant multiplier used during the penalty phase; 2) at the final stage of the method in order to provide a better final multiplier estimate.

Finally, the third one [6] studies a primal-dual proximal point type method for computing approximate stationary solution to a constrained smooth nonconvex composite optimization problem and establishes its iteration-complexity bounds under different sets of assumptions.

Organization of the paper. Subsection 1.1 provides some basic definitions and notation. Section 2 contains two subsections. The first one presents our main problem of interest and the assumptions made on it. It also presents a refinement procedure used in the θ-IPAAL method. The second subsection reviews an ACG variant which will be used to approximately solve θ-prox-AL subproblems of the θ-IPAAL method. Section 3 contains two subsections. The first one states the static θ-IPAAL method as well as its iteration-complexity bounds. The second subsection is devoted to the study of a dynamic variant of the static θ-IPAAL method. The iteration-complexity analysis of this dynamic scheme is also presented in this subsection. Section 4 contains the proofs of two main results of this paper, namely, Theorems 4 and 5. It is divided into two subsections. The proof of Theorem 4 is given in the first subsection while the one of Theorem 5 is given in the second subsection. Section 5 contains the proof of an auxiliary technical result. Secion 6 is devoted to some preliminary numerical results. Section 7 presents some concluding remarks. Finally, some basic auxiliary results are considered in Appendix **??**.

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

Let $\Re^n$ denote the n-dimensional Euclidean space with inner product and associated norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. We use $\Re^{l\times n}$ to denote the set of all $l \times n$ matrices. The image space of a matrix

$Q \in \Re^{l\times n}$ is defined as $\mathrm{Im}(Q) := \{Qx : x \in \Re^n\}$ and $\mathcal{P}_Q$ denotes the Euclidean projection onto $\mathrm{Im}\,(Q)$. The smallest positive eigenvalue of $(Q^*Q)^{1/2}$ is denoted by $\sigma^+(Q)$. If Q is a symmetric and positive semidefinite matrix, the seminorm induced by Q on $\Re^n$, denoted by $\|\cdot\|_Q$, is defined as $\|\cdot\|_Q := \langle Q(\cdot), \cdot\rangle^{1/2}$. The distance of a point x to a closed convex set X is denoted by $\mathrm{dist}_X(x)$. For $t > 0$, define $\log_1^+(t) := \max\{\log t, 1\}$.

The domain of a function $h : \Re^n \to (-\infty, \infty]$ is the set $\mathrm{dom}\, h := \{x \in \Re^n : h(x) < +\infty\}$. Moreover, h is said to be proper if $h(x) < \infty$ for some $x \in \Re^n$. The set of closed proper convex functions defined in $\Re^n$ is denoted by $\overline{\mathrm{Conv}}\,(\Re^n)$. The ε-subdifferential of a function $h \in \overline{\mathrm{Conv}}\,(\Re^n)$ is defined by

$$\partial_\varepsilon h(z) := \{u \in \Re^n : h(z') \geq h(z) + \langle u, z' - z\rangle - \varepsilon, \quad \forall z' \in \Re^n\} \tag{8}$$

for every $z \in \Re^n$. For a scalar $\alpha \in \Re$ and a function h, we define the sublevel set

$$L_h(\alpha) := \{z \in \Re^n : h(z) \leq \alpha\}.$$

If $\psi : \Re^n \to \Re$ is differentiable at $\bar z \in \Re^n$, then its affine approximation $\ell_\psi(\cdot; \bar z)$ at $\bar z$ is defined as

$$\ell_\psi(z; \bar z) := \psi(\bar z) + \langle \nabla\psi(\bar z), z - \bar z\rangle \quad \forall z \in \Re^n. \tag{9}$$

2. Problem of interest and background materials.

This section contains two subsections. The first one presents our main problem of interest and the assumptions made on it. It also presents a procedure to refine an approximate solution (z_k, v_k) of (5) to a pair $(\hat z_k, \hat v_k)$ which together with an approximate Lagrange multiplier $\hat p_k$ is such that the triple $(\hat z, \hat v, \hat p) := (\hat z_k, \hat v_k, \hat p_k)$ satisfies the inclusion in (3). The second subsection reviews an accelerated composite gradient variant which will be used to approximately solve the subproblems generated by the θ-IPAAL method.

2.1. Problem of interest and refinement procedure .

This subsection formally states our problem of interest as well as the main assumptions and the concept of approximate stationary point to it. It also contains a refinement procedure which will be used in the θ-IPAAL method.

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

(A1) A is a nonzero linear operator and the feasible set $\mathcal{F} := \{z \in \mathrm{dom}\, h : Az = b\} \neq \emptyset$;

(A2) h is a proper convex lower semi-continuous function;

(A3) f is nonconvex and differentiable on $\mathrm{dom}\, h$, and there exist $L \geq m > 0$ such that for every $z, z' \in \mathrm{dom}\, h$,

$$\|\nabla f(z') - \nabla f(z)\| \leq L\|z' - z\|, \tag{10}$$

$$f(z') - \ell_f(z'; z) \geq -\frac{m}{2}\|z' - z\|^2; \tag{11}$$

(A4) there exists $\bar c \geq 0$ such that $\phi^*_{\bar c} := \inf_{z\in\Re^n} \phi_{\bar c}(z) > -\infty$, where ϕ_c is defined as

$$\phi_c(\cdot) := \phi(\cdot) + \frac{c}{2}\|A\cdot - b\|^2 \quad \forall c \geq 0. \tag{12}$$

Some comments are in order. First, it can be easily shown that the condition in (10) implies that $L\|z' - z\|^2/2 \geq f(z') - \ell_f(z'; z) \geq -L\|z' - z\|^2/2$ for every $z, z' \in \mathrm{dom}\, h$, and hence that (11) holds with $m = L$. However our analysis covers the case in which (11) holds with a scalar $m < L$ in order to obtain better iteration-complexity bounds. Second, (11) implies that the function $f(\cdot) + m\|\cdot\|^2/2$ is convex on $\mathrm{dom}\, h$. Moreover, since f is nonconvex on $\mathrm{dom}\, h$, we have that the smallest m satisfying (11) is positive. Third, **(A3)** implies that $\mathrm{dom}\, h \subseteq \mathrm{dom}\, f$, and hence that $\mathrm{dom}\, h = \mathrm{dom}\, \phi$. Fourth, **(A4)** is used to obtain our iteration-complexity bounds. It trivially holds if ϕ is bounded below, which is always the case when $\mathrm{dom}\, h$ is bounded and f is lower semi-continuous on $\mathrm{cl}\,(\mathrm{dom}\, h)$. Fifth, it is easy to see that (11) implies that $\mathcal{L}^\theta_c(\cdot, p)$ is strongly convex for every $\lambda < 1/m$. Finally, for the sake of future reference, we note that the definitions of $\mathcal{F}$ and $\phi^*_{\bar c}$ immediately imply that

$$\phi^*_{\bar c} \leq \phi^* = \inf_{z\in\mathcal{F}} \phi(z). \tag{13}$$

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

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

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

Definition 1. *Given a tolerance pair $(\hat{\rho}, \hat{\eta}) \in \Re_{++} \times \Re_{++}$, a triple $(\hat{z}, \hat{v}, \hat{p}) \in \Re^n \times \Re^n \times \Re^l$ is said to be a $(\hat{\rho}, \hat{\eta})$-approximate stationary point of* (1) *if it satisfies* (3).

Subsection 3.2 formally describes the θ-IPAAL method for finding a $(\hat{\rho}, \hat{\eta})$-approximate stationary point of (1). Each one of its outer iteration generates a triple (z_k, v_k, p_k) satisfying the criterion (32) and then updates the multiplier according to (34). The method computes the latter triple by applying the ACG variant of Subsection 2.2 to the subproblem started from z_{k-1} and stops when (32) is satisfied.

In what follows, we discuss how the quadruple $(z_k, v_k, p_k, \varepsilon_k)$ obtained above can be refined to a suitable triple $(\hat{z}_k, \hat{v}_k, \hat{p}_k)$ satisfying the inclusion $\hat{v}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^* \hat{p}_k$, which will later be shown in Subsections 3.1 and 3.2 to possess the property that $\|\hat{v}_k\|$ converges to zero with a well-established convergence rate while the feasibility gap $\|A\hat{z}_k - b\|$ is shown to be either $\mathcal{O}(1/\sqrt{c})$ (see Theorem 4) or $\mathcal{O}(1/c)$ (see Theorem 5). Hence, if c is chosen sufficiently large, it follows that $(\hat{z}, \hat{v}, \hat{p}) = (\hat{z}_k, \hat{v}_k, \hat{p}_k)$ will eventually satisfy (3), or equivalently, be a $(\hat{\rho}, \hat{\eta})$-approximate solution of (1).

First, we mention that θ-IPAAL assumes that $\lambda < 1/m$. In view of (11), this implies that (5) is a strongly convex subproblem, and hence has a unique optimal solution. Consider first the case in which z_k is the exact solution of (5). Indeed, in this case, it follows from (6) and the optimality condition of (5) that $0 \in \nabla f(z_k) + \partial h(z_k) + A^* p_k + (z_k - z_{k-1})/\lambda$, and hence that $(\hat{z}_k, \hat{v}_k, \hat{p}_k) := (z_k, (z_{k-1} - z_k)/\lambda, p_k)$ satisfies the above inclusion. Assume now that z_k is an approximate solution of (5) in the sense that there exists a residual pair (v_k, ε_k) which together with z_k satisfies the approximation criterion (32) below. In this case, it is shown in Proposition 2 below that the following refinement procedure with inputs $(g, h) = (\mathcal{L}_c^\theta(\cdot; p_{k-1}) - h, h)$ and $(\lambda, z^-, z, v) = (\lambda, z_{k-1}, z_k, v_k)$ obtains a pair $(\hat{z}_k, \hat{v}_k)$ which together with $\hat{p}_k$ as in (33) satisfies the above inclusion and such that $(\|\hat{v}_k\|, \|\hat{z}_k - z_k\|)$ is conveniently bounded by $\|v_k + z_{k-1} - z_k\|$ and ε_k. Since ε_k is bounded by $\|v_k + z_{k-1} - z_k\|$ in view of the inequality in (32), and it follows from the second inequality in (66) combined with (51) that $\|v_k + z_{k-1} - z_k\|$ approaches zero, we will then be able to conclude that $\|\hat{v}_k\|$, as well as $\|\hat{z}_k - z_k\|$, approaches zero.

Refinement Procedure

Input: A pair of functions (g, h) such that $h \in \overline{\text{Conv}}\,(\Re^n)$, and g is differentiable on $\text{dom}\, h$ and its gradient is M-Lipschitz continuous, and a quadruple $(\lambda, z^-, z, v) \in \Re_{++} \times \Re^n \times \text{dom}\, h \times \Re^n$.

(1) set

$$g_\lambda := \lambda g + \frac{1}{2}\|\cdot - z^-\|^2 - \langle v, \cdot \rangle, \quad h_\lambda := \lambda h; \tag{15}$$

(2) compute

$$\hat{z} := \text{argmin}_u \left\{ \langle \nabla g_\lambda(z), u - z \rangle + \frac{\lambda M + 1}{2}\|u - z\|^2 + h_\lambda(u) \right\}, \tag{16}$$

$$\hat{v} := \frac{1}{\lambda}\left[(v + z^- - z) + (\lambda M + 1)(z - \hat{z})\right] + \nabla g(\hat{z}) - \nabla g(z), \tag{17}$$

$$\Delta := (g_\lambda + h_\lambda)(z) - (g_\lambda + h_\lambda)(\hat{z}); \tag{18}$$

Output: $(\hat{z}, \hat{v}, \Delta) \in \text{dom}\, h \times \Re^n \times \Re_{++}$.

We now state a proposition summarizing some important properties of the above procedure.

Proposition 2. *Under the assumptions stated at the beginning of the above refinement procedure, the following statements about its output $(\hat{z}, \hat{v}, \Delta)$ hold:*

a) $\Delta \geq 0$ *and*

$$\hat{v} \in \nabla g(\hat{z}) + \partial h(\hat{z}), \quad \lambda\|\hat{v}\| \leq \|v + z^- - z\| + 2\sqrt{2(\lambda M + 1)\Delta}, \qquad \|\hat{z} - z\| \leq \sqrt{2(\lambda M + 1)^{-1}\Delta}; \tag{19}$$

b) if there exists $\varepsilon \geq 0$ *such that the input* (λ, z^-, z, v) *satisfies*

$$v \in \partial_\varepsilon \left(\lambda(g + h) + \frac{1}{2}\|\cdot - z^-\|^2\right)(z), \tag{20}$$

then $\Delta \leq \varepsilon$.

Proof: The proof of this proposition can be found, for instance, in [22, Proposition 2.1 and Lemma 2.3]. ■

The above proposition shows that: 1) the pair $(\hat{z}, \hat{v})$, computed as in (16) and (17), satisfies an inclusion closely related to (3), in view of the definition $\hat{p}$ given in Step 3 of the θ-IPAAL method of Subsection 3.1; 2) the quantities $\lambda\|\hat{v}\|$ and $\|\hat{z} - z\|$ have upper bounds expressed in terms of λM and the two quantities: $\|v + z^- - z\|$ and $\sqrt{\Delta}$. Finally, the ACG method of Subsection 2.2 will be invoked in Step 1 of the θ-IPAAL method to compute a triple (ε, z, v) which together with a previously computed pair (λ, z^-) satisfy the inclusion in (20) as well as some extra conditions bounding, in particular, the scalar ε in terms of the quantity $\|v + z^- - z\|$. In view of the latter discussion and Proposition 2 b), the scalar Δ will also be controlled by $\|v + z^- - z\|$. The latter result will be essential to establish our iteration-complexity bounds.

2.2. An ACG variant. This subsection reviews the ACG variant that will be invoked by the θ-IPAAL method for solving the subproblems (5) which arise during its implementation. It also discusses the iteration-complexity for finding a certain type of approximate solutions.

With the above goal in mind, we consider the composite optimization problem (for the purpose of this subsection only)

$$\min\{\psi(x) := \psi_s(x) + \psi_n(x) : x \in \Re^n\} \tag{21}$$

where the following conditions are assumed to hold:

(C1) $\psi_n : \Re^n \to (-\infty, +\infty]$ is a proper, closed and μ-strongly convex function with $\mu \geq 0$;

(C2) ψ_s is a convex differentiable function on $\mathrm{dom}\,\psi_n$ and there exists $M_s > 0$ satisfying $\psi_s(u) - \ell_{\psi_s}(u; x) \leq M_s\|u - x\|^2/2$ for every $x, u \in \mathrm{dom}\,\psi_n$ where $\ell_{\psi_s}(\cdot\,;\cdot)$ is defined in (9).

We now state the aforementioned ACG variant for solving (21). We remark that other ACG variants such as the ones in [1, 16, 34, 35, 40] could also have been used in the development of the θ-IPAAL method.

ACG Method

---

(0) Let function and parameter pairs (ψ_s, ψ_n) and (M_s, μ) satisfying assumptions **(C1)** and **(C2)** and initial point $x_0 \in \mathrm{dom}\,\psi_n$ be given, and set $y_0 = x_0$, $A_0 = 0$, $\Gamma_0 \equiv 0$ and $j = 0$;

(1) compute

$$A_{j+1} = A_j + \frac{\mu A_j + 1 + \sqrt{(\mu A_j + 1)^2 + 4M_s(\mu A_j + 1)A_j}}{2M_s},$$

$$\tilde{x}_j = \frac{A_j}{A_{j+1}}x_j + \frac{A_{j+1} - A_j}{A_{j+1}}y_j, \quad \Gamma_{j+1} = \frac{A_j}{A_{j+1}}\Gamma_j + \frac{A_{j+1} - A_j}{A_{j+1}}\ell_{\psi_s}(\cdot\,;\tilde{x}_j),$$

$$y_{j+1} = \mathrm{argmin}_y \left\{\Gamma_{j+1}(y) + \psi_n(y) + \frac{1}{2A_{j+1}}\|y - y_0\|^2\right\},$$

$$x_{j+1} = \frac{A_j}{A_{j+1}}x_j + \frac{A_{j+1} - A_j}{A_{j+1}}y_{j+1};$$

(2) compute

$$u_{j+1} = \frac{y_0 - y_{j+1}}{A_{j+1}},$$

$$\eta_{j+1} = \psi(x_{j+1}) - \Gamma_{j+1}(y_{j+1}) - \psi_n(y_{j+1}) - \langle u_{j+1}, x_{j+1} - y_{j+1}\rangle;$$

(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 the θ-IPAAL method in Subsection 3.1. Third, the ACG method in which $\mu = 0$ is a special case of a slightly more general one studied by Tseng in [40] (see Algorithm 3 of [40]). 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

$$A_j \geq \frac{1}{M_s} \max\left\{\frac{j^2}{4}, \left(1+\sqrt{\frac{\mu}{4M_s}}\right)^{2(j-1)}\right\}. \tag{22}$$

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

PROPOSITION 3. *Let $\{(A_j, x_j, u_j, \eta_j)\}$ be the sequence generated by the ACG method applied to (21), where (ψ_s, ψ_n) is a given pair of functions satisfying* **(C1)** *and* **(C2)** *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 $\sigma > 0$, the ACG method obtains a triple $(x, u, \eta) = (x_j, u_j, \eta_j)$ satisfying

$$u \in \partial_\eta(\psi_s + \psi_n)(x) \quad \|u\|^2 + 2\eta \leq \sigma^2 \|x_0 - x + u\|^2 \tag{23}$$

in at most

$$\left\lceil 1 + \sqrt{\frac{M_s}{\mu}} \log_1^+ \left((1+\sigma^{-1})\sqrt{2M_s}\right)\right\rceil \tag{24}$$

iterations.

Proof: The inclusion in a) follows immediately from [20, Proposition 8(c)]. The proof of b) follows from the first statement in [20, Lemma 9] and by noting that if j is larger than or equal to the number in (24), then $A_j \geq 2(1+\sigma^{-1})^2$, in view of (22). It should be noted that the parameter σ in (23) corresponds to $\sqrt{\sigma}$ in [20, Proposition 8 and Lemma 9]. ∎

As we have already mentioned earlier, the ACG method of this subsection will be used to approximately solve the subproblems (5) during the course of the θ-IPAAL method. The results of this subsection will then be invoked with ψ chosen as the objective function of (5) and condition (23) on (x, u, η) will then be used to declare $z_k = x$ as an acceptable approximate solution of (5).

3. The θ-IPAAL method and main complexity results.

This section contains two subsections. The first one states the static θ-IPAAL method as well as its iteration-complexity bounds. The second subsection presents a dynamic variant of the static θ-IPAAL method and its iteration-complexity analysis.

3.1. The static θ-IPAAL method and its iteration-complexity.

This subsection presents the static θ-IPAAL method and its iteration-complexity bounds.

The static θ-IPAAL method assumes that the parameter $\theta \in (0, 1]$ is specified by the user and then starts by computing the following two intermediate parameters τ_θ and σ_θ uniquely determined by θ:

$$\tau_\theta := \begin{cases} \frac{\theta}{16-17\theta}, & \text{if } \theta \leq \frac{16}{19}, \\ \frac{1}{2}, & \text{otherwise,} \end{cases} \tag{25}$$

which is used to define the prox-stepsize λ that appears in the prox θ-AL subproblems (5), and σ_θ defined as the the only positive root of the second-order equation

$$\left(\frac{3}{4} + \frac{2(1-\theta)(3\tau_\theta + 1)}{\theta\tau_\theta}\right)\sigma^2 + \left(\frac{8-7\theta}{2\theta}\right)\sigma - \frac{1}{8} = 0. \tag{26}$$

For the sake of future reference, we note that $\sigma_\theta \le 1/2$.

We now state the θ-IPAAL Method.

Static θ-IPAAL Method

(0) Let parameter $\theta \in (0,1]$, initial point $z_0 \in \operatorname{dom} h$, scalar $\bar{c}$ satisfying **(A4)**, tolerance $\hat{\rho} > 0$, and penalty parameter $c \in \Re$ satisfying

$$c > 2\bar{c} \tag{27}$$

be given, and set $p_0 = 0$, $k = 1$ and

$$L_c := L + c\|A\|^2, \qquad \lambda = \frac{\tau_\theta}{m}, \qquad \sigma := \min\left\{\frac{1}{\sqrt{\lambda L_c + 1}}, \sigma_\theta\right\}; \tag{28}$$

(1) let

$$g_k := f + (1-\theta)\left\langle p_{k-1}\,, A\cdot - b\right\rangle + \frac{c}{2}\|A\cdot - b\|^2 \tag{29}$$

and apply the ACG method with inputs

$$\psi_s := \lambda g_k + \frac{\tau_\theta}{2}\|\cdot - z_{k-1}\|^2, \quad \psi_n := \lambda h + \frac{1-\tau_\theta}{2}\|\cdot - z_{k-1}\|^2, \tag{30}$$

$$(M_s, \mu) = (\lambda L_c + \tau_\theta, 1 - \tau_\theta), \quad x_0 = z_{k-1} \tag{31}$$

to obtain a triple $(z_k, v_k, \varepsilon_k) \in \Re^n \times \Re^n \times \Re_+$ satisfying

$$v_k \in \partial_{\varepsilon_k}\left(\lambda \mathcal{L}_c^\theta(\cdot; p_{k-1}) + \frac{1}{2}\|\cdot - z_{k-1}\|^2\right)(z_k), \quad \|v_k\|^2 + 2\varepsilon_k \le \sigma^2\|v_k + z_{k-1} - z_k\|^2; \tag{32}$$

(2) compute a pair $(\hat{z}_k, \hat{v}_k)$ via the refinement procedure with input $(g,h) = (g_k, h)$ and $(\lambda, z^-, z, v) = (\lambda, z_{k-1}, z_k, v_k)$ and set

$$\hat{p}_k = (1-\theta)p_{k-1} + c\,(A\hat{z}_k - b)\,; \tag{33}$$

(3) if $\|\hat{v}_k\| \le \hat{\rho}$, then output $(\hat{z}, \hat{v}, \hat{p}) := (\hat{z}_k, \hat{v}_k, \hat{p}_k)$ and **stop**; otherwise, set

$$p_k = (1-\theta)p_{k-1} + c\,(Az_k - b) \tag{34}$$

and $k \leftarrow k+1$ and return to (1);

We now make a few remarks about the static θ-IPAAL method. First, it makes 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. Second, upon termination the method finds a pair satisfying the inclusion and the first inequality in (3) but not necessarily the second inequality in (3). However, it is shown in Theorems 4 and 5 below that the feasibility gap $\|A\hat{z}_k - b\|$ of $\hat{z}_k$ is either $\mathcal{O}(1/\sqrt{c})$ or $\mathcal{O}(1/c)$ depending on the assumptions made on problem (1). Hence, the static θ-IPAAL method should not be viewed as an algorithm for solving (1) but rather an intermediary step in this direction. Indeed, its dynamic version described in Subsection 3.2, which: a) doubles c whenever the aforementioned feasibility gap at the end of the static θ-IPAAL method is not small enough, and; b) repeats the static θ-IPAAL method with the updated c, is guaranteed to obtain an approximate solution of (1) in the sense of Definition 1. Finally, only the inequality in (32) needs to be checked for terminating the call to the ACG method in step 1 since the inclusion is always satisfied by any ACG iterate (see Proposition 3 (a), (4), (29), and (30)).

Our goal now is to state an iteration-complexity result for the static θ-IPAAL method under the assumptions introduced in Subsection 2.1. Its bounds are described in terms of the quantity

$$R_0 = R(z_0; \lambda, \bar{c}) := \inf \left\{ \lambda[(f+h)(z) - \phi^*_{\bar{c}}] + \|z - z_0\|^2 : z \in \mathcal{F} \right\} \tag{35}$$

where $\mathcal{F}$ and $\phi^*_{\bar{c}}$ are as **(A1)** and **(A4)**, respectively. Note that if (1) has an optimal solution, then we easily see by using (13) that

$$0 \le R_0 \le \lambda[\phi^* - \phi^*_{\bar{c}}] + d_0^2, \tag{36}$$

where ϕ^* is as in (1) and

$$d_0 := \inf\{\|z^* - z_0\| : z^* \text{ is an optimal solution of (1)}\}.$$

Hence, under the assumption above, R_0 is majorized by a quantity expressed in terms of d_0^2 and the functional gap $\phi^* - \phi^*_{\bar{c}} \ge 0$.

We are now ready to state the aforementioned iteration-complexity result for the static θ-IPAAL method. Its proof will be given in Subsection 4.1.

THEOREM 4. *Let τ_θ and R_0 be as in (25) and (35), respectively, and consider κ_θ as*

$$\kappa_\theta := 1 + \frac{16(1-\theta)}{\theta \tau_\theta}. \tag{37}$$

Then, the following statements about the static θ-IPAAL method hold:

a) its number of outer iterations is bounded by $\left\lceil 108\kappa_\theta R_0/(\lambda^2\hat{\rho}^2)\right\rceil$;

b) the number of ACG iterations performed at each outer iteration is bounded by

$$\left\lceil 1 + \sqrt{\Theta_c}\log_1^+\left(\frac{2\Theta_c}{\sigma_\theta}\right)\right\rceil \tag{38}$$

where $\Theta_c := 2\lambda L_c + 1$ and L_c is as in (28);

c) its output $(\hat{z}, \hat{v}, \hat{p})$ satisfies

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}) + A^*\hat{p}, \qquad \|\hat{v}\| \le \hat{\rho}, \tag{39}$$

$$\|A\hat{z} - b\| \le \frac{4}{\lambda}\sqrt{\frac{\kappa_\theta R_0}{c}}. \tag{40}$$

We will now state an alternative iteration-complexity result for the static θ-IPAAL method which holds under the same assumptions stated in Subsection 2.1 plus a new set of mild conditions which include the existence of a Slater point for (1). The main extra consequence of this result is that the feasibility gap of $\hat{z}$ is now shown to be $\mathcal{O}(1/c)$ instead of $\mathcal{O}(1/\sqrt{c})$. The extra conditions assumed on problem (1) are as follows:

(B1) there exists $\bar{z} \in \operatorname{int}(\operatorname{dom} h)$ such that $A\bar{z} = b$;

(B2) for any $\alpha \in \Re$,

$$D(\alpha) := \sup\{\|x - x'\| : x, x' \in L_{\phi_{\bar{c}}}(\alpha)\} < \infty,$$

where $\bar{c}$ is as in (A4); hence, the sublevel set $L_{\phi_{\bar{c}}}(\alpha)$ is bounded for any $\alpha \in \Re$;

(B3) for any $\alpha \in \Re$,

$$S(\alpha) := \inf\left\{ s \in \Re_+ : \partial h(z) \subseteq \bar{B}(0; s) + N_{\operatorname{dom} h}(z), \ \forall z \in L_{\phi_{\bar{c}}}(\alpha) \right\} < \infty.$$

We now make a few remarks about conditions **(B1)**–**(B3)**. First, **(B1)** is the so called Slater condition for (1). Second, **(B2)** trivially holds if $\operatorname{dom} h$ is bounded. Third, the latter condition plus conditions **(B1)** and **(B3)** are exactly the ones assumed in [27] where the complexity stated for θ-IPAAL method in Theorem 4 was established for the penalty scheme studied there.

We are now ready to state the aforementioned alternative iteration-complexity for the static θ-IPAAL method. Its proof will be given only in Subsection 4.2.

THEOREM 5. *In addition to* **(A1)**–**(A4)**, *assume that conditions* **(B1)**–**(B3)** *hold. Let* R_0, κ_θ, *and* $\bar{z}$ *be as in* (35), (37), *and* **(B1)**, *respectively. Define* $N_0 = N(z_0;\theta,\lambda,\bar{c},\bar{z})$ *as*

$$N_0 := \frac{1}{\sigma^+(A)}\left(1+\frac{\max\{\sqrt{3\kappa_\theta R_0}, D(\alpha)\}}{\mathrm{dist}_{\partial(\mathrm{dom}\,h)}(\bar{z})}\right)\left[\|\nabla f(z_0)\| + LD(\alpha) + S(\alpha) + \left(L+\frac{14}{\lambda\sqrt{\theta}}\right)\sqrt{3\kappa_\theta R_0}\right], \tag{41}$$

where $D(\alpha)$ *and* $S(\alpha)$ *are as in* **(B2)** *and* **(B3)** *with* $\alpha := \kappa_\theta + \max\{\phi_{\bar{c}}(z_0), \phi_{\bar{c}}(\bar{z})\}$. *Then, the output* $(\hat{z},\hat{v},\hat{p})$ *of the static* θ*-IPAAL method satisfies the relations in* (39) *and*

$$\|A\hat{z}-b\| \le \frac{2N_0}{c}. \tag{42}$$

We now make a few remarks about Theorems 4 and 5. First, given a tolerance $\hat{\eta}$, define

$$c(\hat{\eta}) := \max\left\{2\bar{c}, \frac{16\kappa_\theta R_0}{\lambda^2\hat{\eta}^2}\right\}.$$

Clearly, in view of (39) and (40), it follows that for any $c > c(\hat{\eta})$, the output $(\hat{z},\hat{v},\hat{p})$ of the static θ-IPAAL method is a $(\hat{\rho},\hat{\eta})$-approximate stationary point of (1) (see Definition 1). Second, under assumptions **(A1)–(A4)** and **(B1)–(B3)**, Theorem 5 implies that $\|A\hat{z}-b\| \le \hat{\eta}$ for any c satisfying

$$c > \max\left\{2\bar{c}, \frac{2N_0}{\hat{\eta}}\right\}, \tag{43}$$

and hence the output $(\hat{z},\hat{v},\hat{p})$ of the static θ-IPAAL method is a $(\hat{\rho},\hat{\eta})$-approximate stationary point of (1). Third, since either $c(\hat{\eta})$ or the right-hand side of (43) is difficult to compute due to its dependence on κ_θ or N_0, it is usually not possible to have at our immediate disposal a scalar c as in the previous two remarks, and hence to find a $(\hat{\rho},\hat{\eta})$-approximate stationary point of (1) by just solving a single penalized subproblem. The next subsection presents a dynamic version of the static θ-IPAAL method which dynamically updates the penalty parameter c and whose overall ACG iteration-complexity is similar to either one the complexities obtained above with c chosen as $c(\hat{\eta})$ or the right-hand side of (43).

3.2. The θ-IPAAL method and its iteration-complexity.

This subsection presents a dynamic version of the θ-IPAAL method to obtain a $(\hat{\rho},\hat{\eta})$-approximate stationary point to (1). The method basically consists of applying the static θ-IPAAL method and repeatedly doubling the penalty parameter until the second inequality in (3) is satisfied. The main iteration-complexity result of this scheme is also presented and proved in this subsection.

We start by stating the dynamic version of the θ-IPAAL method, which will be simply referred to the θ-IPAAL method for shortness.

θ-IPAAL method

---

(0) Let an initial point $z_0 \in \mathrm{dom}\,h$, a scalar $\theta \in (0,1]$, and a pair of tolerances $(\hat{\rho},\hat{\eta}) \in \Re_{++}\times\Re_{++}$ be given, choose an initial penalty parameter $c_1 > 2\bar{c}$ and set $c = c_1$;

(1) execute the static θ-IPAAL method with input $(z_0,\theta,c,\hat{\rho})$, and let $(\hat{z},\hat{v},\hat{p})$ be the output.

(2) if $\|A\hat{z}-b\| \le \hat{\eta}$, stop and output $(\hat{z},\hat{v},\hat{p})$; otherwise, set $c \leftarrow 2c$ and return to step 1.

The next result establishes the overall ACG iteration-complexity of the θ-IPAAL method for obtaining a $(\hat{\rho},\hat{\eta})$-approximate stationary point of (1).

THEOREM 6. *Assume that conditions* **(A1)**–**(A4)** *of Subsection 2.1 hold. Then, the following statements about the* θ*-IPAAL method hold:*

a) it obtains an approximate stationary point $(\hat{z},\hat{v},\hat{p})$ *of problem* (1) *in the sense of Definition 1 in at most*

$$\mathcal{O}\left(\left\lceil\frac{R_0}{\lambda^2\hat{\rho}^2}\right\rceil\left[\sqrt{\lambda\max\left\{c_1,\frac{R_0}{\lambda^2\hat{\eta}^2}\right\}}\|A\| + \sqrt{\lambda L+1}\log_1^+\left(\frac{R_0}{c_1\lambda^2\hat{\eta}^2}\right)\right]\log_1^+(T_{\hat{\eta}})\right), \tag{44}$$

ACG iterations, where R_0 *is as in* (35), *and*

$$T_{\hat{\eta}} := \lambda\left(L+\max\left\{c_1,\frac{R_0}{\lambda^2\hat{\eta}^2}\right\}\|A\|^2\right)+1; \tag{45}$$

b) if, in addition, the conditions **(B1)**–**(B3)** *of Subsection 3.1 hold, then its overall ACG iteration-complexity is*

$$\mathcal{O}\left(\left\lceil \frac{R_0}{\lambda^2\hat{\rho}^2}\right\rceil \left[\sqrt{\lambda \max\left\{c_1, \frac{N_0}{\hat{\eta}}\right\}}\|A\| + \sqrt{\lambda L+1}\log_1^+\left(\frac{N_0}{c_1\hat{\eta}}\right)\right]\log_1^+\left(\tilde{T}_{\hat{\eta}}\right)\right), \tag{46}$$

where

$$\tilde{T}_{\hat{\eta}} := \lambda\left(L + \max\left\{c_1, \frac{N_0}{\hat{\eta}}\right\}\|A\|^2\right) + 1$$

and N_0 *is as in* (41).

Proof:

a) First note that the l-th loop of the θ-IPAAL method invokes the static θ-IPAAL method with penalty parameter $c = c_l$ where $c_l := 2^{l-1}c_1 > 2\bar{c}$ for all $l \geq 1$. Hence, in view of the stopping criterion in step 2 of the θ-IPAAL method and Theorem 4 c), we conclude that the θ-IPAAL method performs at most $\bar{l}$ iterations where

$$\bar{l} := \min\left\{l : c_l \geq \frac{16\kappa_\theta R_0}{\lambda^2\hat{\eta}^2}\right\} \tag{47}$$

and its output $(\hat{z}, \hat{v}, \hat{p})$ is an approximate stationary point of problem (1). Moreover, it follows from statements a) and b) of Theorem 4 that the total number of ACG iterations is bounded by

$$\left(\sum_{l=1}^{\bar{l}}\left[1 + \sqrt{\Theta_{c_l}}\log_1^+\left(\frac{2\Theta_{c_l}}{\sigma_\theta}\right)\right]\right)\left\lceil\frac{108\kappa_\theta R_0}{\lambda^2\hat{\rho}^2}\right\rceil, \tag{48}$$

where $\Theta_{c_l} = 2\lambda L_{c_l} + 1$. In view of the above definition of c_l and (47), we have

$$c_l \leq \max\left\{c_1, \frac{32\kappa_\theta R_0}{\lambda^2\hat{\eta}^2}\right\}, \qquad \forall l = 1, \ldots, \bar{l}. \tag{49}$$

Hence, (45), (49), and the definitions of Θ_{c_l} and L_{c_l} (see (28)) imply that

$$\Theta_{c_l} = \left[2\lambda\left(L + c_l\|A\|^2\right) + 1\right] \leq \left[2\lambda\left(L + \max\left\{c_1, \frac{32\kappa_\theta R_0}{\lambda^2\hat{\eta}^2}\right\}\|A\|^2\right) + 1\right] = \mathcal{O}\left(T_{\hat{\eta}}\right). \tag{50}$$

It also follows from the definitions of c_l, Θ_{c_l}, and L_{c_l} that

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

From the above inequalities, (49) and definition of $\bar{l}$ in (47), we have

$$\begin{aligned}\sum_{l=1}^{\bar{l}}\sqrt{\Theta_{c_l}} &\leq 8\sqrt{2\lambda}\|A\|\sqrt{\max\left\{c_1, \frac{32\kappa_\theta R_0}{\lambda^2\hat{\eta}^2}\right\}} + \sqrt{2\lambda L+1}\log_1^+\left(\frac{64\kappa_\theta R_0}{c_1\lambda^2\hat{\eta}^2}\right)\\ &= \mathcal{O}\left(\|A\|\sqrt{\lambda\max\left\{c_1, \frac{R_0}{\lambda^2\hat{\eta}^2}\right\}} + \sqrt{\lambda L+1}\log_1^+\left(\frac{R_0}{c_1\lambda^2\hat{\eta}^2}\right)\right).\end{aligned}$$

Since $\sqrt{\Theta_{c_l}}\log_1^+(t) \geq 1$ for every $t > 0$, statement a) follows from the above conclusions, (48) and (50).

b) The proof of this statement is similar to the one of a), by noting that it uses Theorem 5 instead of Theorem 4 c), and the definition of $\bar{l}$ in (47) should be modified to

$$\bar{l} := \min\left\{l : c_l \geq \frac{2N_0}{\hat{\eta}}\right\},$$

and then (49) is replaced by

$$c_l \le \max\left\{c_1, \frac{4N_0}{\hat{\eta}}\right\}, \qquad \forall l = 1 \ldots \bar{l}.$$

■

We now make two remarks about Theorem 6. First, it follows from its statements a) and b) that the ACG iteration-complexity of the θ-IPAAL method expressed only in terms of the tolerance pair $(\hat{\rho}, \hat{\eta})$ is $\mathcal{O}([1/(\hat{\eta}\hat{\rho}^2)]\log(1/\hat{\eta}))$ under the assumption that conditions **(A1)**–**(A4)** hold and can actually be sharpened to $\mathcal{O}([1/(\sqrt{\hat{\eta}}\hat{\rho}^2)]\log(1/\hat{\eta}))$ if conditions **(B1)**–**(B3)** also hold. Second, if the initial penalty parameter c_1 is chosen so as to satisfy $c_1 = 2\bar{c} + \Theta(L/\|A\|^2)$ and the logarithms are ignored in the complexity bounds (44) and (46), then these bounds reduce to

$$\mathcal{O}\left(\left\lceil\frac{m^2 R_0}{\hat{\rho}^2}\right\rceil\left[\|A\|\sqrt{\max\left\{\frac{\bar{c}}{m}, \frac{mR_0}{\hat{\eta}^2}\right\}} + \sqrt{\frac{L}{m}+1}\right]\right)$$

and

$$\mathcal{O}\left(\left\lceil\frac{m^2 R_0}{\hat{\rho}^2}\right\rceil\left[\|A\|\sqrt{\frac{1}{m}\max\left\{\bar{c}, \frac{N_0}{\hat{\eta}}\right\}} + \sqrt{\frac{L}{m}+1}\right]\right),$$

respectively, in view of the fact that by (28) the stepsize λ satisfies $\lambda = \Theta(1/m)$. Third, when (1) has an optimal solution, then the above bounds can be majorized by ones involving the upper bound on R_0 described in (36).

4. Proofs of Theorems 4 and 5.

This section contains the proofs of Theorems 4 and 5. It is divided into two subsections. The proof of Theorem 4 is given in the first subsection while the one of Theorem 5 is given in the second subsection.

4.1. Proof of Theorem 4.

This subsection is devoted to the proof of Theorem 4.

We start by stating a result which describes some relations which are frequently used in our analysis.

LEMMA 7. *Let $\{(z_k, v_k, p_k)\}$ be generated by the static θ-IPAAL method and, for every $k \ge 1$, define*

$$\Delta p_k := p_k - p_{k-1}, \quad \Delta z_k := z_k - z_{k-1}, \quad r_k := \Delta z_k - v_k. \tag{51}$$

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

$$\Delta p_{k+1} = (1-\theta)\Delta p_k + cA\Delta z_{k+1}, \tag{52}$$

$$\|\Delta z_k\| \le (1+\sigma)\|r_k\|, \quad \|r_k\| \le \frac{1}{1-\sigma}\|\Delta z_k\|, \tag{53}$$

where σ is as in (28).

Proof: The relation in (52) follows immediately from (34) and the first two relations in (51). Now, in view of the last two relations in (51), it is easy to see that the inequality in (32) implies that $\|v_k\| \le \sigma\|r_k\|$. It follows from the latter inequality, the last relation in (51), and the triangle inequality that

$$\|\Delta z_k\| \le \|\Delta z_k - v_k\| + \|v_k\| \le (1+\sigma)\|r_k\|, \qquad \|r_k\| \le \|\Delta z_k\| + \|v_k\| \le \|\Delta z_k\| + \sigma\|r_k\|.$$

The above inequalities clearly imply that the ones in (53) hold. ■

It follows from the inequalities in (53) that the sequence of displacements $\{\Delta z_k\}$ goes to zero if and only if the residual sequence $\{r_k\}$ goes to zero.

The next lemmas describe how the sequence $\{(z_k, p_k)\}$ generated by the static θ-IPAAL method affects the value of $\mathcal{L}_c^\theta(z, p)$ defined in (4).

LEMMA 8. *Let $\{(z_k, v_k, p_k)\}$ be generated by the static θ-IPAAL method and let $\{\Delta p_k\}$, $\{\Delta z_k\}$, and $\{r_k\}$ be as in (51). Then, for every $k \ge 1$, the following relations hold:*

$$\mathcal{L}_c^\theta(z_k, p_k) - \mathcal{L}_c^\theta(z_k, p_{k-1}) = \frac{(1-\theta)(2-\theta)}{2c}\|\Delta p_k\|^2 + \frac{(1-\theta)\theta}{2c}\left(\|p_k\|^2 - \|p_{k-1}\|^2\right), \tag{54}$$

$$\mathcal{L}_c^\theta(z_k, p_k) - \mathcal{L}_c^\theta(z_{k-1}, p_{k-1}) \le -\frac{1-\sigma^2}{2\lambda}\|r_k\|^2 + \frac{(1-\theta)(2-\theta)}{2c}\|\Delta p_k\|^2 + \frac{(1-\theta)\theta}{2c}\left(\|p_k\|^2 - \|p_{k-1}\|^2\right). \tag{55}$$

Proof: In view of the definition of $\mathcal{L}_c^\theta$ in (4), relation (34) and the first relation in (51), we obtain

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

which immediately implies (54) upon using the identity $p_k=\Delta p_k+p_{k-1}$. Now, it follows from (32), (51), and the Cauchy-Schwarz inequality, that

$$\lambda\mathcal{L}_c^\theta(z_k,p_{k-1})-\lambda\mathcal{L}_c^\theta(z_{k-1},p_{k-1}) \le -\frac{1}{2}\|z_k-z_{k-1}\|^2+\langle v_k, z_k-z_{k-1}\rangle+\varepsilon_k$$
$$=-\frac{1}{2}\|\Delta z_k-v_k\|^2+\left(\frac{\|v_k\|^2}{2}+\varepsilon_k\right) \le -\frac{1-\sigma^2}{2}\|\Delta z_k-v_k\|^2 = -\frac{1-\sigma^2}{2}\|r_k\|^2.$$

The inequality in (55) follows immediately by adding the previous inequality and (54). ∎

The next result shows that the term $\|\Delta p_k\|^2$ in (55) is majorized by $\|r_k\|^2$ and some summable terms. Its proof is postponed to Section 5.

Lemma 9. *Let $\{\Delta p_k\}$, $\{\Delta z_k\}$, and $\{r_k\}$ be as in (51). Then, for every $k\ge 2$, we have*

$$\frac{(2-\theta)}{2c}\|\Delta p_k\|^2 \le \frac{2}{\lambda\theta}\left[\frac{2\tau_\theta(1+\sigma)^2}{\tau_\theta+1}+2\sigma(1+\sigma)+\frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right]\|r_k\|^2+\frac{(1-\theta)^2}{\theta c}\left[\|\Delta p_{k-1}\|^2-\|\Delta p_k\|^2\right]$$
$$+\frac{1}{\lambda\theta}\left[\sigma(1+\sigma)+\frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right]\left[\|r_{k-1}\|^2-\|r_k\|^2\right]+\frac{1}{\lambda\theta}\left[\|\Delta z_{k-1}\|^2-\|\Delta z_k\|^2\right].$$

Next result is a first step in order to show that the residual sequence $\{r_k\}$ is controlled by a certain decreasing sequence associated to $\{\mathcal{L}_c^\theta(z_k,p_k)\}$. This fact is crucial to establish the iteration-complexity bounds for the θ-IPAAL method.

Lemma 10. *Define*

$$C_\theta := \frac{1-\sigma^2}{2}-\frac{2(1-\theta)}{\theta}\left[\frac{2\tau_\theta(1+\sigma)^2}{\tau_\theta+1}+2\sigma(1+\sigma)+\frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right] \tag{56}$$

and

$$\tilde{C}_\theta := \frac{1-\theta}{\theta}\left[\sigma(1+\sigma)+\frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right]. \tag{57}$$

Then, $C_\theta\ge 1/8$, $\tilde{C}_\theta\ge 0$ and, for every $k\ge 2$, we have

$$\mathcal{L}_c^\theta(z_k,p_k)-\mathcal{L}_c^\theta(z_{k-1},p_{k-1}) \le -\frac{C_\theta}{\lambda}\|r_k\|^2+\frac{(1-\theta)\theta}{2c}\left(\|p_k\|^2-\|p_{k-1}\|^2\right)$$
$$+\frac{(1-\theta)^3}{\theta c}\left[\|\Delta p_{k-1}\|^2-\|\Delta p_k\|^2\right]+\frac{\tilde{C}_\theta}{\lambda}\left[\|r_{k-1}\|^2-\|r_k\|^2\right]+\frac{1-\theta}{\theta\lambda}\left[\|\Delta z_{k-1}\|^2-\|\Delta z_k\|^2\right].$$

Proof: The proof that $C_\theta\ge 1/8$ is simple but tedious, and hence it is given in Appendix B. It is immediate to see that $\tilde{C}_\theta\ge 0$ in view of $\theta,\sigma,\tau_\theta\in(0,1]$. Now the last statement of the lemma follows immediately by combining the inequality in Lemma 9 with (55) and the definitions of C_θ and $\tilde{C}_\theta$. ∎

The next result summarizes some useful relations about $\{\mathcal{L}_c^\theta(z_k,p_k)\}$.

Lemma 11. *Let C_θ and $\tilde{C}_\theta$ be as in (56) and (57), and let η_k be defined by*

$$\eta_k := \frac{(1-\theta)^3}{\theta c}\|\Delta p_k\|^2+\frac{\tilde{C}_\theta}{\lambda}\|r_k\|^2+\frac{1-\theta}{\theta\lambda}\|\Delta z_k\|^2-\frac{(1-\theta)\theta}{2c}\|p_k\|^2-\phi_{\bar{c}}^*, \qquad \forall k\ge 1. \tag{58}$$

Then, the following statements hold:

a) for every $k \geq 2$,

$$\mathcal{L}_c^\theta(z_k,p_k)+\eta_k+\frac{C_\theta}{\lambda}\|r_k\|^2 \leq \mathcal{L}_c^\theta(z_{k-1},p_{k-1})+\eta_{k-1}, \tag{59}$$

$$\mathcal{L}_c^\theta(z_k,p_k)+\eta_k+\frac{C_\theta}{\lambda}\sum_{i=2}^{k}\|r_i\|^2 \leq \mathcal{L}_c^\theta(z_1,p_1)+\eta_1. \tag{60}$$

b) for every $k \geq 1$,

$$\frac{c}{4}\|Az_k-b\|^2+\frac{(1-\theta)\theta}{4c}\|p_k\|^2+\phi_{\bar c}(z_k)-\phi_{\bar c}^* \leq \mathcal{L}_c^\theta(z_k,p_k)+\eta_k; \tag{61}$$

Proof: a) The inequality in (59) follows immediately from Lemma 10 and the definition of η_k given in (58). The inequality in (60) follows immediately from (59).

b) From the definitions of $\mathcal{L}_c^\theta$ and η_k given in (4) and (58), respectively, and the fact that $\tilde C_\theta \geq 0$ (see Lemma 10), we obtain, for every $k\geq 1$,

$$\mathcal{L}_c^\theta(z_k,p_k)+\eta_k+\phi_{\bar c}^* \geq \mathcal{L}_c^\theta(z_k,p_k)+\frac{(1-\theta)^3}{\theta c}\|\Delta p_k\|^2+\frac{1-\theta}{\theta\lambda}\|\Delta z_k\|^2-\frac{(1-\theta)\theta}{2c}\|p_k\|^2$$

$$\geq (f+h)(z_k)+\frac{c}{2}\|Az_k-b\|^2+(1-\theta)\left\langle p_k\,,Az_k-b-\frac{\theta p_k}{c}\right\rangle+\frac{(1-\theta)\theta}{2c}\|p_k\|^2+\frac{(1-\theta)^3}{\theta c}\|\Delta p_k\|^2.$$

Since, for every $k\geq 1$, (34) implies that $(1-\theta)\Delta p_k = c(Az_k-b)-\theta p_k$, we conclude from the above relations, definition of $\phi_{\bar c}$ in (12), and the fact that $\langle u,v\rangle \geq -(1/2)(\|u\|^2/2+2\|v\|^2)$ for all $u,v\in\Re^n$, that

$$\mathcal{L}_c^\theta(z_k,p_k)+\eta_k+\phi_{\bar c}^* \geq \phi_{\bar c}(z_k)+\frac{c-\bar c}{2}\|Az_k-b\|^2+\frac{1-\theta}{c}\left[\langle p_k\,,(1-\theta)\Delta p_k\rangle+\frac{\theta}{2}\|p_k\|^2+\frac{(1-\theta)^2}{\theta}\|\Delta p_k\|^2\right]$$
$$\geq \phi_{\bar c}(z_k)+\frac{c}{4}\|Az_k-b\|^2+\frac{1-\theta}{2c}\left[-\frac{\theta}{2}\|p_k\|^2-\frac{2(1-\theta)^2}{\theta}\|\Delta p_k\|^2+\theta\|p_k\|^2+\frac{2(1-\theta)^2}{\theta}\|\Delta p_k\|^2\right],$$

where in the last inequality we also used $c-\bar c > c-c/2 = c/2$, in view of (27). The statement in b) easily follows. ∎

We now make a few comments about Lemma 11. Its first statement shows that the θ-AL function $\mathcal{L}_c^\theta(z_k,p_k)$ plus the scalar η_k defined in (58) works as a merit descent function for the static θ-IPAAL method. Moreover, its second statement shows that both the feasibility gap $\|Az_k-b\|$ and the magnitude of the Lagrange multiplier $\|p_k\|$ is well-controlled by this merit function.

At a first sight, it is natural to think that the left hand side of (60) grows linearly with c but the following proposition shows that it is actually uniformly bounded with respect to c. This result plays an important role in showing that the θ-IPAAL method finds a $(\hat\rho,\hat\eta)$-approximate stationary point of (1) as in Definition 1, regardless of whether the initial point $z_0\in \operatorname{dom} h$ is feasible or infeasible.

Proposition 12. *Let $c>2\bar c$, let (z_1,p_1) be generated by the static θ-IPAAL method, and consider R_0, κ_θ, r_1, and η_1 as in (35), (37), (51) and (58), respectively. Then, the following inequalities hold:*

$$c\|Az_1-b\|^2+\frac{1}{\lambda}\|r_1\|^2 \leq \frac{4R_0}{\lambda}; \tag{62}$$

$$\mathcal{L}_c^\theta(z_1,p_1)+\eta_1 \leq \frac{\kappa_\theta R_0}{\lambda}. \tag{63}$$

Proof: The definitions of $\mathcal{L}_c^\theta$ and $\mathcal{F}$ in (4) and **(A1)**, respectively, imply that $\mathcal{L}_c^\theta(z,p_0)=(f+h)(z)$ for all $z\in\mathcal{F}$. Hence, from (32) and (51) with $k=1$, and Lemma 23 with $\tilde\phi=\lambda\mathcal{L}_c^\theta(\cdot;p_0)$ and $s=1$, we obtain, for every $z\in\mathcal{F}$,

$$\lambda\mathcal{L}_c^\theta(z_1,p_0)+\frac{1-2\sigma^2}{2}\|r_1\|^2 \leq \lambda\mathcal{L}_c^\theta(z,p_0)+\|z-z_0\|^2 = \lambda\left[(f+h)(z)-\phi_{\bar c}^*\right]+\|z-z_0\|^2+\lambda\phi_{\bar c}^*.$$

The above inequality combined with the definition of R_0 in (35) and the fact that $\sigma \le 1/2$ imply that

$$\frac{1}{4\lambda}\|r_1\|^2 + \mathcal{L}_c^\theta(z_1,p_0) - \phi_{\bar{c}}^* \le \frac{R_0}{\lambda}. \tag{64}$$

Now note that the definition of $\phi_{\bar{c}}^*$ in **(A4)** implies that $-(\bar{c}/2)\|Az_1-b\|^2 \le (f+h)(z_1) - \phi_{\bar{c}}^*$. Since $p_0 = 0$, it follows from the latter inequality and the definition of $\mathcal{L}_c^\theta$ in (4) that

$$\frac{c-\bar{c}}{2}\|Az_1-b\|^2 \le (f+h)(z_1) + \frac{c}{2}\|Az_1-b\|^2 - \phi_{\bar{c}}^* = \mathcal{L}_c^\theta(z_1,p_0) - \phi_{\bar{c}}^*.$$

Hence, (62) follows by combining the above inequality with (64) and by using that $c - \bar{c} > c - c/2 = c/2$, in view of the fact that $c > 2\bar{c}$.

Now, we proceed to prove (63). Since $p_0 = 0$, (34) and (51), both with $k=1$, imply that

$$\Delta p_1 = c(Az_1 - b). \tag{65}$$

Also note that (57) and the fact that $\tau_\theta, \sigma \le 1/2$ yield $\tilde{C}_\theta \le (1-\theta)/(\theta\tau_\theta)$. Hence, combining Lemma 8, (58), the first inequality in (53), all three with $k=1$, (65), and the fact that $\sigma,\theta \in (0,1]$, we obtain

$$\mathcal{L}_c^\theta(z_1,p_1) + \eta_1 - \mathcal{L}_c^\theta(z_1,p_0) + \phi_{\bar{c}}^* = \frac{(1-\theta)(2-\theta)}{2c}\|\Delta p_1\|^2 + \frac{(1-\theta)^3}{\theta c}\|\Delta p_1\|^2 + \frac{\tilde{C}_\theta}{\lambda}\|r_1\|^2 + \frac{1-\theta}{\theta\lambda}\|\Delta z_1\|^2$$

$$\le \frac{1-\theta}{\theta}\left\{\frac{2}{c}\|\Delta p_1\|^2 + \frac{\|r_1\|^2}{\lambda\tau_\theta} + \frac{(1+\sigma)^2\|r_1\|^2}{\lambda}\right\} \le \frac{1-\theta}{\theta}\left\{2c\|Az_1-b\|^2 + \frac{4\|r_1\|^2}{\tau_\theta\lambda}\right\}$$

which combined with (62), (64), and the fact that $\tau_\theta < 1$ imply that

$$\mathcal{L}_c^\theta(z_1,p_1) + \eta_1 \le \frac{R_0}{\lambda} + \frac{16(1-\theta)R_0}{\lambda\theta\tau_\theta}.$$

Inequality (63) follows immediately from the latter one and the definition of κ_θ given in (37). ∎

The following result presents some estimates regarding the sequence $\{(z_k,r_k,p_k)\}$ which are useful to establish iteration-complexity bounds for the θ-IPAAL method.

PROPOSITION 13. *Let $\{(z_k,v_k,p_k)\}$ be generated by the static θ-IPAAL method and consider $\{r_k\}$ as in (51). Then, the following inequalities hold*

$$\|Az_k-b\|^2 \le \frac{4\kappa_\theta R_0}{\lambda c}, \qquad \sum_{j=1}^{k}\|r_j\|^2 \le 12\kappa_\theta R_0, \qquad (1-\theta)\|p_k\|^2 \le \frac{4c\kappa_\theta R_0}{\lambda\theta}, \tag{66}$$

where κ_θ is as in (37).

Proof: The first and third inequalities in (66) follow immediately by combining (61) and (63). Now, note that (59) implies that $\mathcal{L}_c^\theta(z_k,p_k)+\eta_k \ge 0$ for every $k \ge 1$. Hence, since, in view of (56), $C_\theta \ge 1/8$, we obtain by combining (60) and (63) that

$$\sum_{j=2}^{k}\frac{1}{8\lambda}\|r_j\|^2 \le \mathcal{L}_c^\theta(z_1,p_1) + \eta_1 - \left(\mathcal{L}_c^\theta(z_k,p_k)+\eta_k\right) \le \kappa_\theta.$$

We also obtain from the inequality in (62) and the definition of κ_θ in (37) that

$$\|r_1\|^2 \le 4\lambda\kappa_\theta.$$

The above inequalities easily imply the second inequality in (66). ∎

The next proposition shows some useful relations on the sequence $\{(\hat{z}_k,\hat{v}_k,\hat{p}_k)\}$. It shows that $(\hat{z}_k,\hat{v}_k,\hat{p}_k)$ satisfies the inclusion in (3) and that $\{\|\hat{v}_k\|\}$ and $\{\|\hat{z}_k - z_k\|\}$ are controlled by the residual sequence $\{\|r_k\|\}$ which has a subsequence converging to zero, in view of the second inequality in (66). Moreover, it also presents some useful bounds on the feasibility of $\{\hat{z}_k\}$ and the boundedness of the auxiliary sequence $\{\hat{p}_k\}$ associated to the Lagrange multipliers.

PROPOSITION 14. *Consider the sequences* $\{(z_k, v_k, \varepsilon_k)\}$ *and* $\{(\hat{z}_k, \hat{v}_k, \hat{p}_k)\}$ *be generated by the static* θ*-IPAAL method and let* $\{r_k\}$ *be as in* (51)*. Then, the following relations hold, for every* $k \geq 1$*:*

$$\hat{v}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^*\hat{p}_k, \qquad \|\hat{v}_k\| \leq \frac{\left(1 + 2\sigma\sqrt{\lambda L_c + 1}\right)\|r_k\|}{\lambda}, \tag{67}$$

$$\|\hat{z}_k - z_k\| \leq \frac{\sigma\|r_k\|}{\sqrt{\lambda L_c + 1}}, \qquad \|A\hat{z}_k - b\| \leq 4\sqrt{\frac{\kappa_\theta R_0}{\lambda c}}, \qquad \|\hat{p}_k\| \leq 6\sqrt{\frac{c\kappa_\theta R_0}{\theta\lambda}}, \tag{68}$$

where L_c *and* κ_θ *are as in* (28) *and* (37)*, respectively.*

Proof: First note that Step 2 of the θ-IPAAL method computes $(\hat{z}_k, \hat{v}_k)$ by the refinement procedure with input $(g, h) = (g_k, h)$ and $(\lambda, z^-, z, v) = (\lambda, z_{k-1}, z_k, v_k)$ where g_k as defined in (29) has gradient L_c-Lipschitz continuous with L_c as in (28). Hence, the inclusion in (67) follows from the inclusion in Proposition 2 a) and the definitions of g_k and $\hat{p}_k$ given in (29) and (33), respectively. Now note that the scalar Δ output by the refinement procedure satisfies $\Delta \leq \varepsilon_k$, in view of the inclusion in (32), the fact that $g + h = \mathcal{L}_c^\theta(\cdot; p_{k-1})$, and Proposition 2 b). Hence, the inequalities in Proposition 2 a) imply that, for every $k \geq 1$,

$$\lambda\|\hat{v}_k\| \leq \|v_k + z_k - z_{k-1}\| + 2\sqrt{2(\lambda L_c + 1)]\varepsilon_k}, \quad \|\hat{z}_k - z_k\| \leq \sqrt{2(\lambda L_c + 1)^{-1}\varepsilon_k}$$

which combined with the inequality in (32) and the definition of r_k in (51) proves the inequality in (67) and the first inequality in (68).

Using the triangle inequality for norms, the first inequality in (68), and the first and second inequalities in (66), we obtain

$$\|A\hat{z}_k - b\| \leq \|Az_k - b\| + \|A\|\|\hat{z}_k - z_k\| \leq 2\sqrt{\frac{\kappa_\theta R_0}{\lambda c}} + \frac{2\sigma\|A\|\sqrt{3\kappa_\theta R_0}}{\sqrt{\lambda L_c + 1}} \leq 2\sqrt{\frac{\kappa_\theta R_0}{\lambda c}} + \sqrt{\frac{3\kappa_\theta R_0}{\lambda c}},$$

where the last inequality is due to $\sqrt{\lambda L_c + 1} \geq \|A\|\sqrt{\lambda c}$ and the fact that $\sigma \leq \sigma_\theta \leq 1/2$. Hence, the second inequality in (68) follows. Now, from the definition of $\hat{p}_k$ given in (33), the last inequality in (66), the second inequality in (68), and the Cauchy-Schwarz inequality, we obtain

$$\|\hat{p}_k\| \leq (1 - \theta)\|p_{k-1}\| + c\|A\hat{z}_k - b\| \leq 2\sqrt{\frac{c\kappa_\theta R_0}{\theta\lambda}} + 4c\sqrt{\frac{\kappa_\theta R_0}{\lambda c}} = 6\sqrt{\frac{c\kappa_\theta R_0}{\theta\lambda}}.$$

■

The following inequality, which follows by combining the second inequality in (66), the inequality in (67) and the definition of σ in (28), will be frequently used:

$$\|\hat{v}_k\| \leq \frac{3\|r_k\|}{\lambda} \leq \frac{6\sqrt{3\kappa_\theta R_0}}{\lambda} \qquad \forall k \geq 1. \tag{69}$$

We are now ready to prove Theorem 4.

Proof of Theorem 4: a) It follows from the second inequality in (66) that, for every $k \geq 1$,

$$k \min\{\|r_j\|^2, j = 1, \ldots, k\} \leq 12\kappa_\theta R_0.$$

Consider the index $j \in \{1, \ldots, k\}$ that achieves the above minimum. It follows from the above inequality and the first inequality in (69) with $k = j$ that

$$\|\hat{v}_j\| \leq \frac{3\|r_j\|}{\lambda} \leq \frac{6\sqrt{3\kappa_\theta R_0}}{\lambda\sqrt{k}},$$

which combined with the stopping criterion of the static θ-IPAAL method proves the desired result.

b) First note that the static θ-IPAAL method invokes the ACG method with ψ_s and ψ_n as in (30) and then $\nabla\psi_s$ has Lipschitz constant M_s and ψ_n is μ-strongly convex, where M_s and μ are as in (31). Hence, it follows from Proposition 3 and the definition of σ given in (28) that the static θ-IPAAL method performs at most

$$\left\lceil 1+\sqrt{\frac{\lambda L_c+\tau_\theta}{1-\tau_\theta}}\log_1^+\left(\left(1+\max\left\{\sqrt{\lambda L_c+1},\frac{1}{\sigma_\theta}\right\}\right)\sqrt{2(\lambda L_c+\tau_\theta)}\right)\right\rceil$$

ACG iterations at each outer iteration. Hence, the statement in b) follows in view of the definition of Θ_c and the fact that $\tau_\theta \le 1/2$.

c) First note that a) ensures that the static θ-IPAAL method has finite termination. Hence, in view of its stopping criterion and the inclusion in (67), we immediately obtain the relations in (39). Moreover, the output of the θ-IPAAL method combined with the second inequality in (68) imply that (40) holds, concluding the proof.

4.2. Proof of Theorem 5. This subsection is devoted to the proof of Theorem 5.

We start by specifying some bounds on the sequence $\{z_k\}$ and $\{\hat z_k\}$ generated by the θ-IPAAL method.

LEMMA 15. *In addition to* **(A1)**–**(A4)***, assume that conditions* **(B1)**–**(B2)** *hold and let $\bar z$ be as in* **(B1)**. *Then, the sequences $\{z_k\}$ and $\{\hat z_k\}$ generated by the θ-IPAAL method satisfy*

$$\max\{\|z_k-z_0\|,\|z_k-\bar z\|\}\le D(\alpha),\qquad \max\{\|\hat z_k-z_0\|,\|\hat z_k-\bar z\|\}\le D(\alpha)+\sqrt{3\kappa_\theta R_0},\qquad \forall k\ge 1, \tag{70}$$

where α is as in Theorem 5 and $D(\alpha)$ is as in **(B2)**.

Proof: Since $\alpha=\kappa_\theta R_0/\lambda+\max\{\phi_{\bar c}(z_0),\phi_{\bar c}(\bar z)\}$, Lemma 11 b) combined with (63) and the definition of $\phi^*_{\bar c}$ given in **(A4)** imply that

$$\phi_{\bar c}(z_k)\le \frac{\kappa_\theta R_0}{\lambda}+\phi^*_{\bar c}\le \frac{\kappa_\theta R_0}{\lambda}+\phi_{\bar c}(z_0)\le \alpha,\qquad \forall k\ge 1, \tag{71}$$

which means that $\{z_k\}$ is in the α-sublevel set $L_{\phi_{\bar c}}(\alpha)$. Since $\kappa_\theta>0$, the definition of α immediately yields $\phi_{\bar c}(z_0)\le\alpha$ and $\phi_{\bar c}(\bar z)\le\alpha$, or equivalently $z_0,\bar z\in L_{\phi_{\bar c}}(\alpha)$. Hence, the first inequality in (70) holds in view of assumption **(B2)**. Now, using the first inequality in (68), the second inequality in (66), and the fact that $\sigma\le\sigma_\theta\le 1/2$ (see (28)), we obtain

$$\|\hat z_k-z_k\|\le\frac{\sigma\|r_k\|}{\sqrt{\lambda L_c+1}}\le\sqrt{3\kappa_\theta R_0},\quad \forall k\ge 1.$$

Hence, in view of the first inequality in (70) and the triangle inequality, we conclude that, for every $k\ge 1$,

$$\max\{\|\hat z_k-z_0\|,\|\hat z_k-\bar z\|\}\le\max\{\|z_k-z_0\|,\|z_k-\bar z\|\}+\|\hat z_k-z_k\|\le D(\alpha)+\sqrt{3\kappa_\theta R_0},$$

which proves the last inequality in (70). ∎

Next we state a technical result which will be used in the proof of the subsequently lemma. Its proof can be found, for instance, in [27, Lemma 1].

LEMMA 16. *Assume that X is a convex set and $\bar x\in \operatorname{int}(X)$, and let ∂X denote the boundary of X. Then,* $\operatorname{dist}_{\partial X}(\bar x)>0$ *and*

$$\|\xi\|\le\frac{\langle\xi,x-\bar x\rangle}{\operatorname{dist}_{\partial X}(\bar x)}\quad \forall x\in X,\ \forall\xi\in N_X(x).$$

The following result shows that the component of the inclusion in (67) lying in $\partial h(\hat z_k)$ is bounded. It is worth noting that its proof strongly relies on the bound for $\|\hat p_k\|$ derived in Proposition 14.

LEMMA 17. *In addition to* **(A1)**–**(A4)***, assume that conditions* **(B1)**–**(B3)** *hold and let $M_\theta(\alpha)$ be defined by*

$$M_\theta(\alpha):=\frac{2\max\left\{D(\alpha),\sqrt{3\kappa_\theta R_0}\right\}}{\operatorname{dist}_{\partial(\operatorname{dom}h)}(\bar z)}\left[LD(\alpha)+\left(L+\frac{25}{\sqrt{\theta}\lambda}\right)\sqrt{3\kappa_\theta R_0}+\|\nabla f(z_0)\|+S(\alpha)\right]. \tag{72}$$

Let $\{(\hat z_k,\hat v_k,\hat p_k)\}$ be generated by the static θ-IPAAL method and consider the sequence $\{\hat\xi_k\}$ given by

$$\hat\xi_k:=\hat v_k-\nabla f(\hat z_k)-A^*\hat p_k,\qquad\forall k\ge 1. \tag{73}$$

Then, $\hat\xi_k\in\partial h(\hat z_k)$ and $\|\hat\xi_k\|\le M_\theta(\alpha)+S(\alpha)$, for every $k\ge 1$.

Proof: The first statement of the lemma immediately follows from the inclusion in (67) and the definition of $\hat{\xi}_k$ given in (73). Now note that by using the Cauchy-Schwarz inequality and the last two inequalities in (68), we obtain

$$\|\langle \hat{p}_k, A\hat{z}_k - b\rangle\| \le \|\hat{p}_k\|\|A\hat{z}_k - b\| \le 24\sqrt{\frac{c\kappa_\theta R_0}{\lambda\theta}}\sqrt{\frac{\kappa_\theta R_0}{\lambda c}} \le \frac{24\kappa_\theta R_0}{\lambda\sqrt{\theta}}, \qquad \forall k \ge 1, \tag{74}$$

where the last inequality is due to the fact that $c - \bar{c} \ge c/2$ (in view of (27)). On the other hand, since $\hat{\xi}_k \in \partial h(\hat{z}_k)$ for every $k \ge 1$, assumption **(B3)** yields

$$\hat{\xi}_k = \hat{\xi}^s_k + \hat{\xi}^N_k, \quad \|\hat{\xi}^s_k\| \le S(\alpha), \quad \hat{\xi}^N_k \in N_{\operatorname{dom} h}(\hat{z}_k). \tag{75}$$

where α is as in Theorem 5. Let $\bar{z}$ be as in (**B1**) and note that $A\bar{z} = b$. Hence, it follows from Lemma 16 with $x = \hat{z}_k$, $\bar{x} = \bar{z}$ and $X = \operatorname{dom} h$, the Cauchy-Schwarz inequality, the triangle inequality, (74) and (75), the Lipshichtz continuity of ∇f and (69) that, for every $k \ge 1$,

$$\begin{aligned}
\operatorname{dist}_{\partial(\operatorname{dom} h)}(\bar{z})\|\hat{\xi}^N_k\| &\le \langle \hat{\xi}^N_k, \hat{z}_k - \bar{z}\rangle = \langle \hat{\xi}_k - \hat{\xi}^s_k, \hat{z}_k - \bar{z}\rangle = \langle \hat{v}_k - \nabla f(\hat{z}_k) - \hat{\xi}^s_k, \hat{z}_k - \bar{z}\rangle - \langle \hat{p}_k, A\hat{z}_k - b\rangle \\
&\le \left(\|\nabla f(\hat{z}_k) - \nabla f(z_0)\| + \|\nabla f(z_0)\| + \|\hat{v}_k\| + \|\hat{\xi}^s_k\|\right)\|\bar{z} - \hat{z}_k\| + \frac{24\kappa_\theta R_0}{\lambda\sqrt{\theta}} \\
&\le \left(L\|\hat{z}_k - z_0\| + \|\nabla f(z_0)\| + \frac{6\sqrt{3\kappa_\theta R_0}}{\lambda} + S(\alpha)\right)\|\bar{z} - \hat{z}_k\| + \frac{24\kappa_\theta R_0}{\lambda\sqrt{\theta}},
\end{aligned}$$

which combined with the last inequality in (70) and the fact that $\theta < 1$, imply that

$$\begin{aligned}
\operatorname{dist}_{\partial(\operatorname{dom} h)}(\bar{z})\|\hat{\xi}^N_k\| &\le \left[LD(\alpha) + \left(L + \frac{6}{\lambda}\right)\sqrt{3\kappa_\theta R_0} + \|\nabla f(z_0)\| + S(\alpha)\right]\left(D(\alpha) + \sqrt{3\kappa_\theta R_0}\right) + \frac{24\kappa_\theta R_0}{\lambda\sqrt{\theta}} \\
&\le 2\left[LD(\alpha) + \left(L + \frac{14}{\sqrt{\theta}\lambda}\right)\sqrt{3\kappa_\theta R_0} + \|\nabla f(z_0)\| + S(\alpha)\right]\max\left\{D(\alpha), \sqrt{3\kappa_\theta R_0}\right\}
\end{aligned} \tag{76}$$

The above inequalities and (72) imply that

$$\|\hat{\xi}^N_k\| \le M_\theta(\alpha), \qquad \forall k \ge 1.$$

Hence, (75) and the triangle inequality imply that

$$\|\hat{\xi}_k\| \le \|\hat{\xi}^N_k\| + \|\hat{\xi}^s_k\| \le M_\theta(\alpha) + S(\alpha), \qquad \forall k \ge 1,$$

proving the last statement of the lemma. ∎

In the following, we state a basic result that will be used in the proof of the next proposition. Its proof can be found, for instance, in [13, Lemma 1.4].

LEMMA 18. *Let $S \in \Re^{l\times n}$ be a non-zero matrix and let $\sigma^+(S)$ denote the smallest positive eigenvalue of $(S^*S)^{1/2}$. Then, for every $u \in \Re^l$, there holds*

$$\|\mathcal{P}_S(u)\| \le \frac{1}{\sigma^+(S)}\|S^*u\|.$$

The next result shows that the sequences of multipliers $\{p_k\}$ and $\{\hat{p}_k\}$ are bounded by a quantity which does not depend on the penalty parameter c. This fact in turn is easily seen to imply an $\mathcal{O}(1/c)$ bound on $\|A\hat{z}_k - b\|$.

PROPOSITION 19. *Under the assumptions of Theorem 5, the following inequalities hold*

$$\|\hat{p}_k\| \le N_0, \qquad \|p_k\| \le N_0 \qquad \forall k \ge 1, \tag{77}$$

where N_0 is as in (41).

Proof: It follows from (73), the triangle inequality for norms, the Lipschitz continuity of ∇f, the relation in (69), and the last statement of Lemma 17, that

$$\begin{aligned} \|A^*\hat{p}_k\| &= \|\hat{v}_k - \nabla f(\hat{z}_k) - \hat{\xi}_k\| \le \|\nabla f(\hat{z}_k) - \nabla f(z_0)\| + \|\nabla f(z_0)\| + \|\hat{v}_k\| + \|\hat{\xi}_k\| \\ &\le L\|\hat{z}_k - z_0\| + \|\nabla f(z_0)\| + \frac{6\sqrt{3\kappa_\theta R_0}}{\lambda} + M_\theta(\alpha) + S(\alpha). \end{aligned} \tag{78}$$

On the other hand, since $p_0 = 0$, (33) and (34) imply that $p_k, \hat{p}_k \in \operatorname{Im} A$, for every $k \ge 1$. Hence, it follows from Lemma 18 with $S = A$ that

$$\|\hat{p}_k\| \le \frac{1}{\sigma^+(A)}\|A^*\hat{p}_k\|,$$

which combined with (78) and the last inequality in (70) imply that

$$\begin{aligned} \|\hat{p}_k\| &\le \frac{1}{\sigma^+(A)}\left[LD(\alpha) + L\sqrt{3\kappa_\theta R_0} + \|\nabla f(z_0)\| + \frac{6\sqrt{3\kappa_\theta R_0}}{\lambda} + M_\theta(\alpha) + S(\alpha)\right], \\ &= \frac{1}{\sigma^+(A)}\left[LD(\alpha) + \left(L + \frac{6}{\lambda}\right)\sqrt{3\kappa_\theta R_0} + \|\nabla f(z_0)\| + M_\theta(\alpha) + S(\alpha)\right] =: \hat{N}_0, \qquad \forall k \ge 1. \end{aligned} \tag{79}$$

Hence, the first inequality in (77) follows, in view of the fact that $\hat{N}_0 \le N_0$ (see the definition of N_0 in (41)). Now, subtracting (33) from (34), using the triangle inequality for norms, the first inequality in (68), the second inequality in (66), and the definitions of L_c and σ given in (28), we obtain

$$\begin{aligned} \|p_k\| &\le \|\hat{p}_k\| + c\|A(z_k - \hat{z}_k)\| \le \|\hat{p}_k\| + \frac{\sigma c\|A\|\|r_k\|}{\sqrt{\lambda L_c + 1}} \\ &\le \hat{N}_0 + \frac{c\|A\|\|r_k\|}{\lambda(L + c\|A\|^2) + 1} \le \hat{N}_0 + \frac{2\sqrt{3\kappa_\theta R_0}}{\lambda\|A\|}. \end{aligned}$$

Hence, the second inequality in (77) follows by using the definitions of N_0 and $\hat{N}_0$ given in (41) and (79), respectively, and that $\sigma^+(A) \le \|A\|$. ■

Now we are ready to proof of Theorem 5.

Proof of Theorem 5: The first statement follows from the first one in Theorem 4 c). Now, using (33), Proposition 19, the triangle inequality for norms and the fact that $\theta \in (0, 1]$, we obtain

$$c\|A\hat{z}_k - b\| \le \|\hat{p}_k\| + (1-\theta)\|p_{k-1}\| \le 2N_0,$$

where N_0 is as in (41). Hence, (42) immediately follows.

5. Proof of Lemma 9.

The main goal of this section is to prove Lemma 9.

Before giving its proof, we state and prove two technical results. The first one essentially describes equivalent but useful ways of expressing the inclusion in (32). The second one establishes an intermediate bound used in the proof of Lemma 9.

LEMMA 20. *Let $\{(z_k, p_k, v_k, \varepsilon_k)\}$ be generated by the static θ-IPAAL method, let Δz_k, Δp_k and r_k be as in (51), and define $\Delta v_{k+1} := v_{k+1} - v_k$ and*

$$a_k(\cdot) := \frac{1}{2}\|\cdot - (z_k + v_{k+1})\|^2 - \frac{1}{2}\|\cdot - (z_{k-1} + v_k)\|^2 + \lambda(1-\theta)\left\langle \Delta p_k \,, A\cdot - b\right\rangle. \tag{80}$$

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

a) a_k is an affine function whose gradient ∇a_k is given by

$$\nabla a_k = -\Delta z_k - \Delta v_{k+1} + \lambda(1-\theta)A^*\Delta p_k; \tag{81}$$

b) the following inclusions hold

$$0 \in \partial_{\varepsilon_k}\left[\lambda \mathcal{L}_c^\theta(\cdot, p_{k-1}) + \frac{1}{2}\|\cdot - (z_{k-1} + v_k)\|^2\right](z_k),$$

$$-\nabla a_k \in \partial_{\varepsilon_{k+1}}\left[\lambda \mathcal{L}_c^\theta(\cdot, p_{k-1}) + \frac{1}{2}\|\cdot - (z_{k-1} + v_k)\|^2\right](z_{k+1}).$$

Proof: a) This statement follows trivially from the definition of a_k given in (80) and the relations in (51).

b) In view of (8), it is immediate to see that the inclusion in (32) is equivalent to the first one in b). From the definitions of $\mathcal{L}_c^\theta$ and a_k given in (4) and (80), respectively, we see that

$$\lambda \mathcal{L}_c^\theta(\cdot, p_k) + \frac{1}{2}\|\cdot - (z_k + v_{k+1})\|^2 = \lambda \mathcal{L}_c^\theta(\cdot, p_{k-1}) + \frac{1}{2}\|\cdot - (z_{k-1} + v_k)\|^2 + a_k(\cdot).$$

It follows from this relation and the first inclusion of b) with $k = k+1$ that

$$0 \in \partial_{\varepsilon_{k+1}} \left[\lambda \mathcal{L}_c^\theta(\cdot, p_{k-1}) + \frac{1}{2}\|\cdot - (z_{k-1} + v_k)\|^2 + a_k(\cdot)\right](z_{k+1}).$$

Since a_k is an affine function, the latter inclusion easily implies that the second one of b) holds. ■

LEMMA 21. *Let $\{(z_k, p_k, v_k, \varepsilon_k)\}$ be generated by the static θ-IPAAL method and consider Δz_k, Δp_k and r_k as in (51). Then, for every $k \geq 1$, the following inequality holds*

$$\Theta_k := \frac{c\lambda}{1+\tau_\theta}\|A\Delta z_{k+1}\|^2 + \lambda(1-\theta)\langle \Delta p_k, A\Delta z_{k+1}\rangle \leq \left[\frac{2\tau_\theta(1+\sigma)^2}{\tau_\theta+1} + 2\sigma(1+\sigma) + \frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right]\|r_{k+1}\|^2$$

$$+ \left[\frac{\sigma(1+\sigma)}{2} + \frac{(\tau_\theta+1)\sigma^2}{2\tau_\theta}\right]\left[\|r_k\|^2 - \|r_{k+1}\|^2\right] + \frac{1}{2}\left[\|\Delta z_k\|^2 - \|\Delta z_{k+1}\|^2\right]. \tag{82}$$

Proof: For every k, let us consider the function

$$\psi_k := \lambda \mathcal{L}_c^\theta(\cdot, p_{k-1}) + \frac{1}{2}\|\cdot - (z_{k-1} + v_k)\|^2. \tag{83}$$

Note that, in view of **(A2)**, (11), (83), and the definition of $\mathcal{L}_c^\theta$ in (4), we obtain $\psi_k - (1/2)\|\cdot\|_Q^2$ is convex where $Q := (1-\lambda m)I + c\lambda A^*A \in S_{++}^n$. Hence, in view of Lemma 20 b), using Lemma 22 twice, first with $\psi = \psi_k$, $\xi = 1$, $(y, v, \eta) = (z_k, 0, \varepsilon_k)$, and $\tau = \tau_\theta$, and second with the same ψ, ξ, and τ but $(y, v, \eta) = (z_{k+1}, -\nabla a_k, \varepsilon_{k+1})$ where a_k is as in (80), we obtain

$$\psi_k(u) \geq \psi_k(z_k) + \frac{1}{2(1+\tau_\theta)}\|u - z_k\|_Q^2 - (1+\tau_\theta^{-1})\varepsilon_k, \tag{84}$$

$$\psi_k(u') \geq \psi_k(z_{k+1}) + \langle -\nabla a_k, u' - z_{k+1}\rangle + \frac{1}{2(1+\tau_\theta)}\|u' - z_{k+1}\|_Q^2 - (1+\tau_\theta^{-1})\varepsilon_{k+1}, \tag{85}$$

for all $u, u' \in \Re^n$. Adding both inequalities with $u = z_{k+1}$ and $u' = z_k$, and using the relation in (81) and the definition of Q given above, we conclude that

$$(1+\tau_\theta^{-1})(\varepsilon_k + \varepsilon_{k+1}) \geq \langle \nabla a_k, \Delta z_{k+1}\rangle + \frac{\|\Delta z_{k+1}\|_Q^2}{1+\tau_\theta}$$

$$= -\langle \Delta z_k, \Delta z_{k+1}\rangle - \langle \Delta v_{k+1}, \Delta z_{k+1}\rangle + \lambda(1-\theta)\langle \Delta p_k, A\Delta z_{k+1}\rangle + \frac{1-\lambda m}{1+\tau_\theta}\|\Delta z_{k+1}\|^2 + \frac{c\lambda}{1+\tau_\theta}\|A\Delta z_{k+1}\|^2.$$

Rewriting this inequality and using the definition of Θ_k in (82), we obtain

$$\Theta_k = \frac{c\lambda}{1+\tau_\theta}\|A\Delta z_{k+1}\|^2 + \lambda(1-\theta)\langle \Delta p_k, A\Delta z_{k+1}\rangle$$

$$\leq \langle \Delta z_k, \Delta z_{k+1}\rangle - \frac{1-\lambda m}{\tau_\theta+1}\|\Delta z_{k+1}\|^2 + \langle \Delta v_{k+1}, \Delta z_{k+1}\rangle + \frac{\tau_\theta+1}{\tau_\theta}(\varepsilon_k + \varepsilon_{k+1}). \tag{86}$$

We now proceed to estimate the right-hand side of the last inequality. Using that $\langle u, \tilde{u}\rangle \leq (\|u\|^2 + \|\tilde{u}\|^2)/2$

for every $u, \tilde{u} \in \Re^n$, (53) and that $\tau_\theta = \lambda m$, we have

$$\begin{aligned}
\langle \Delta z_k, \Delta z_{k+1} \rangle - \frac{1-\lambda m}{\tau_\theta + 1}\|\Delta z_{k+1}\|^2 &\le \frac{\|\Delta z_k\|^2}{2} + \frac{\|\Delta z_{k+1}\|^2}{2} - \frac{1-\tau_\theta}{\tau_\theta+1}\|\Delta z_{k+1}\|^2 \\
&= \frac{\|\Delta z_k\|^2 - \|\Delta z_{k+1}\|^2}{2} + \frac{2\tau_\theta}{\tau_\theta+1}\|\Delta z_{k+1}\|^2 \\
&\le \frac{\|\Delta z_k\|^2 - \|\Delta z_{k+1}\|^2}{2} + \frac{2\tau_\theta(1+\sigma)^2}{\tau_\theta+1}\|r_{k+1}\|^2.
\end{aligned}$$

Using Cauchy-Schwarz and triangle inequalities, the inequality in (32) combined with the definitions of Δz_k and r_k in (51), and (53), we obtain

$$\begin{aligned}
\langle \Delta v_{k+1}, \Delta z_{k+1} \rangle &\le (\|v_{k+1}\| + \|v_k\|)\|\Delta z_{k+1}\| \le \sigma(\|r_{k+1}\| + \|r_k\|)\|\Delta z_{k+1}\| \\
&\le \sigma(1+\sigma)(\|r_{k+1}\| + \|r_k\|)\|r_{k+1}\| \le \frac{\sigma(1+\sigma)\|r_k\|^2}{2} + \frac{3\sigma(1+\sigma)\|r_{k+1}\|^2}{2} \\
&= \frac{\sigma(1+\sigma)}{2}\left[\|r_k\|^2 - \|r_{k+1}\|^2\right] + 2\sigma(1+\sigma)\|r_{k+1}\|^2,
\end{aligned}$$

where the last inequality is due to $ab \le (a^2+b^2)/2$, $\forall a, b \in \Re$. The inequality in (32) combined with the definitions of Δz_k and r_k in (51), also yields

$$\varepsilon_k + \varepsilon_{k+1} \le \frac{\sigma^2\|r_k\|^2}{2} + \frac{\sigma^2\|r_{k+1}\|^2}{2} = \sigma^2\|r_{k+1}\|^2 + \frac{\sigma^2}{2}\left[\|r_k\|^2 - \|r_{k+1}\|^2\right].$$

Hence, in view of (86) and the above inequalities, we obtain

$$\begin{aligned}
\Theta_k \le{}& \frac{1}{2}\left[\|\Delta z_k\|^2 - \|\Delta z_{k+1}\|^2\right] + \frac{2\tau_\theta(1+\sigma)^2}{\tau_\theta+1}\|r_{k+1}\|^2 + \frac{\sigma(1+\sigma)}{2}\left[\|r_k\|^2 - \|r_{k+1}\|^2\right] + 2\sigma(1+\sigma)\|r_{k+1}\|^2 \\
&+ \frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\|r_{k+1}\|^2 + \frac{(\tau_\theta+1)\sigma^2}{2\tau_\theta}\left[\|r_k\|^2 - \|r_{k+1}\|^2\right],
\end{aligned}$$

which after simple algebraic manipulations proves the desired inequality. ∎

We are now ready to prove Lemma 9.

Proof of Lemma 9: In view of (52) and the definition of Θ_k in (82), we obtain

$$\begin{aligned}
\Theta_k &= \frac{c\lambda}{1+\tau_\theta}\|A\Delta z_{k+1}\|^2 + \lambda(1-\theta)\langle \Delta p_k, A\Delta z_{k+1}\rangle \\
&= \frac{\lambda}{c}\left(\frac{1}{1+\tau_\theta}\|\Delta p_{k+1} - (1-\theta)\Delta p_k\|^2 + (1-\theta)\langle \Delta p_k, \Delta p_{k+1} - (1-\theta)\Delta p_k\rangle\right) \\
&= \frac{\lambda}{c(1+\tau_\theta)}\left(\|\Delta p_{k+1}\|^2 - \tau_\theta(1-\theta)^2\|\Delta p_k\|^2 + (\tau_\theta - 1)(1-\theta)\langle \Delta p_k, \Delta p_{k+1}\rangle\right).
\end{aligned}$$

Hence, using Cauchy-Schwarz inequality and the facts that $\tau_\theta \in (0,1)$ and $ab \ge -(a^2+b^2)/2$ for all $a, b \in \Re$, we obtain

$$\begin{aligned}
\Theta_k \ge{}& \frac{\lambda}{c(1+\tau_\theta)}\left(\|\Delta p_{k+1}\|^2 - \tau_\theta(1-\theta)^2\|\Delta p_k\|^2 - \frac{(1-\tau_\theta)}{2}\|\Delta p_{k+1}\|^2 - \frac{(1-\tau_\theta)(1-\theta)^2}{2}\|\Delta p_k\|^2\right) \\
={}& \frac{\lambda}{c}\left(\frac{1}{2}\|\Delta p_{k+1}\|^2 - \frac{(1-\theta)^2}{2}\|\Delta p_k\|^2\right) = \frac{\lambda}{c}\left(\frac{\theta(2-\theta)}{2}\|\Delta p_{k+1}\|^2 + \frac{(1-\theta)^2}{2}\left[\|\Delta p_{k+1}\|^2 - \|\Delta p_k\|^2\right]\right).
\end{aligned}$$

The conclusion of the lemma now follows by combining the above inequality with (82).

6. Numerical experiments. This section presents computational results to illustrate the performance of the θ-IPAAL method for different values of the parameter θ. The computational results are limited only to a class of linearly constrained quadratic matrix (LCQM) problems but they should provide a good indication of how promising the θ-IPAAL is as the parameter θ decreases towards zero.

In order to describe the LCQM problems considered here, let $l, n \in \mathcal{N}$, $\alpha_1, \alpha_2 \in \Re_{++}$, $b, d \in \Re^l$, matrices $\{A_i\}_{i=1}^l, \{B_j\}_{j=1}^n, \{C_i\}_{i=1}^l \subseteq \Re^{n\times n}$, and a positive diagonal matrix $D \in \Re^{n\times n}$ be given and define the linear operators $\mathcal{A} : S^n_+ \mapsto \Re^l$, $\mathcal{B} : S^n_+ \mapsto \Re^n$, and $\mathcal{C} : S^n_+ \mapsto \Re^l$ by

$$[\mathcal{A}(z)]_i = \langle A_i, z\rangle_F\,, \quad [\mathcal{B}(z)]_j = \langle B_j, z\rangle_F\,, \quad [\mathcal{C}(z)]_i = \langle C_i, z\rangle_F\,, \qquad \forall i = 1,\dots,l,\ \forall j = 1,\dots,n.$$

The LCQM problem is

$$\begin{aligned} \min_z \quad & \frac{\alpha_1}{2}\|\mathcal{C}(z) - d\|^2 - \frac{\alpha_2}{2}\|D\mathcal{B}(z)\|^2 \\ \text{s.t.} \quad & \mathcal{A}(z) = b, \quad z \in P_n, \end{aligned} \tag{87}$$

where $P_n = \{z \in S^n_+ : \operatorname{tr} z = 1\}$ denotes the n–dimensional spectraplex. Note that (87) can be put in the setting of (1) by considering

$$f(z) = \frac{\alpha_1}{2}\|\mathcal{C}(z) - d\|^2 - \frac{\alpha_2}{2}\|D\mathcal{B}(z)\|^2, \quad h(z) = \delta_{P_n}(z).$$

In the numerical experiments, the entries of A_i, B_j, C_i, b, and d (resp., D) were generated by sampling from the uniform distribution $\mathcal{U}[0,1]$ (resp., $\mathcal{U}[1,1000]$) with 5.0% and 1.0% of the entries of the matrices A_i, B_j, and C_i being nonzero when $(l,n) = (5,20)$ and $(l,n) = (25,100)$, respectively. The scalars $\alpha_1, \alpha_2 \in \Re_{++}$ are selected such that the pair of lower and upper curvatures (m, L) satisfies $L = \lambda_{\max}(\nabla^2 f)$ and $-m = \lambda_{\min}(\nabla^2 f)$. In particular, the inequalities (10) and (11) are satisfied.

The numerical experiments were performed using MATLAB 2019b and a MacOS 64-bit machine with an Intel Core i5 processor and 8 GB of memory.

This section reports computational results for the "theoretical" version of the θ-IPAAL method studied in Subsection 3.2 as well as of a more aggressive version whose motivation is as follows. Note that the parameters τ_θ and σ_θ defined in (25) and (26), respectively, and hence the prox stepsize λ and the parameter σ in (28) become too small as θ approaches 0 (see the two columns under the "theoretical version" part of Table 1). Since the smallness of σ and λ directly affects the number of ACG and outer iterations performed by the method, respectively, it is natural to consider its more aggressive variant (referred to here as the constant version of the θ-IPAAL method) which simply sets these parameters to $\lambda = 0.5/m$ and $\sigma^2 = 0.5$. Even though our theoretical results derived in Sections 3 and 4 do not apply to this constant version, the computational results reported in Tables 2 and 3 show that it performs considerably better than its theoretical counterpart.

The implementation of the theoretical version of the θ-IPAAL method essentially follows its description in Subsection 3.2 with the exception that σ is set to be σ_θ instead of the value in (28). It is straightforward to see that this choice of σ would still be covered by our analysis if we modify the formula in (28) for σ to

$$\sigma = \min\left\{\frac{\tau}{\sqrt{\lambda L_c + 1}}, \sigma_\theta\right\}$$

where $\tau > 0$ is a fixed constant. Choosing $\tau > 0$ large enough, it clearly follows that $\sigma = \sigma_\theta$, and hence that the latter choice for σ is still under the scope of our analysis.

We now discuss implementation details which are common to both versions. The initial multiplier p_0 is set to be zero and the initial point $z_0 \in S^n_+$ is randomly generated, namely, $z_0 = \nu\nu^\top$ where $\nu := \tilde{\nu}/\|\tilde{\nu}\|$, $\tilde{\nu} \sim \mathcal{U}^n[0,1]$ with 10.0% of the entries being nonzero. For a given tolerance pair $(\hat{\rho}, \hat{\eta}) \in \Re^2_{++}$, the θ-IPAAL method stops when it obtains a point $(\hat{z}, \hat{v}, \hat{p})$ satisfying

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}) + A^*\hat{p}, \quad \frac{\|\hat{v}\|}{\|\nabla f(z_0)\| + 1} \le \hat{\rho}, \quad \frac{\|A\hat{z} - b\|}{\|Az_0 - b\| + 1} \le \hat{\eta}. \tag{88}$$

The penalty parameter is chosen as $c = c_1 := 10^{-5}L/(\|A\|^2 + 1)$ and is updated according to $c \leftarrow 5c$, instead of $c \leftarrow 2c$ as in Step 2 of the θ-IPAAL method. It can be easily seen that such a choice does not affect the overall iteration complexity of the θ-IPAAL method. Both versions use the following warm start strategy: instead of starting the call to the static θ-IPAAL method in step 1 of the θ-IPALL method from the iterate-Lagrangian multiplier pair $(z_0, p_0) = (z_0, 0)$ where z_0 is as in step 0 of the θ-IPALL method, it initializes from $(z_0, p_0) = (\hat{z}, \hat{p})$ where $(\hat{z}, \hat{p})$ is the pair obtained in the previous call to the static θ-IPALL method.

θ	Theoretical version		Constant version	
	τ_θ	σ^2	τ_θ	σ^2
1	0.5	3.75e-02	0.5	0.5
0.5	0.067	5.44e-04	0.5	0.5
0.1	0.0070	8.08e-06	0.5	0.5
0	*	*	0.5	0.5

Table 1: The values of τ_θ and σ_θ.

Tables 2 and 3 illustrate the performance of the θ-IPAAL method with different values of the parameter θ for solving some instances of (87). Table 2 displays the results for instances with $(l, n) = (5, 20)$ whereas table 3 considers instances with $(l, n) = (25, 100)$. In these two tables, "ACG iter" denotes the total number of ACG iterations, "outer iter" denotes the total number of outer iterations, "cycle" is the number of cycles, i.e., the number of times that the penalty parameter is updated (hence, every outer iteration within a cycle uses the same penalty parameter c), and runtime is in seconds.

(L, m)	θ	Theoretical version				Constant version			
		ACG iter	Outer iter	Cycle	Runtime	ACG iter	Outer iter	Cycle	Runtime
$(10^4, 1)$	1	25704	16	13	37.91	6606	16	13	8.71
	0.5	7404	27	12	12.37	2639	15	12	4.14
	0.1	5188	129	11	7.44	1323	14	11	2.34
	0	*	*	*	*	756	13	10	1.44
$(10^5, 1)$	1	93443	13	13	145.78	25697	14	13	45.01
	0.5	24337	13	12	33.75	10092	13	12	15.78
	0.1	7662	28	11	11.92	4057	12	11	6.45
	0	*	*	*	*	2226	11	10	3.52
$(10^6, 1)$	1	328146	13	13	458.28	94579	13	13	179.83
	0.5	89737	12	12	149.06	40578	12	12	75.18
	0.1	19568	12	11	32.65	17491	11	11	32.39
	0	*	*	*	*	8005	10	10	13.23
$(10^7, 10)$	1	327119	13	13	517.19	94613	13	13	143.58
	0.5	89983	12	12	142.12	40719	12	12	69.29
	0.1	19791	12	11	29.80	17977	11	11	25.74
	0	*	*	*	*	7942	10	10	11.34
$(10^7, 10^2)$	1	93835	13	13	150.35	25791	14	13	38.33
	0.5	24160	13	12	32.84	10113	13	12	13.98
	0.1	7548	28	11	10.22	4189	12	11	6.36
	0	*	*	*	*	2226	11	10	3.91
$(10^7, 10^3)$	1	26061	16	13	44.04	6552	16	13	10.24
	0.5	7424	27	12	11.28	2639	15	12	4.43
	0.1	5208	129	11	7.20	1323	14	11	1.97
	0	*	*	*	*	756	13	10	1.43

Table 2: Performance of the θ-IPAAL method with $(l, n) = (5, 20), \hat{\rho} = \hat{\eta} = 10^{-4}$.

(L, m)	θ	Theoretical version				Constant version			
		ACG iter	Outer iter	Cycle	Runtime	ACG iter	Outer iter	Cycle	Runtime
$(10^4, 1)$	1	140390	20	15	1037.44	35221	20	15	288.64
	0.5	33794	27	14	250.59	11787	19	14	106.09
	0.1	10218	103	13	92.99	4054	19	13	34.83
	0	*	*	*	*	1753	18	12	15.26
$(10^5, 1)$	1	496375	15	15	3739.52	139589	16	15	1420.47
	0.5	125599	17	14	1223.06	55049	15	14	437.98
	0.1	27887	35	13	264.72	20826	14	13	174.20
	0	*	*	*	*	6044	13	12	54.30
$(10^6, 1)$	1	1662990	15	15	11582.70	507773	15	15	3525.21
	0.5	431788	14	14	3383.33	212601	14	14	1500.66
	0.1	90949	16	13	635.13	88880	13	13	646.18
	0	*	*	*	*	33908	12	12	224.12
$(10^7, 10)$	1	1705212	15	15	12073.87	508098	15	15	4126.89
	0.5	435710	14	14	3957.91	212679	14	14	1981.36
	0.1	90586	16	13	1012.90	88970	13	13	869.04
	0	*	*	*	*	34858	12	12	249.89
$(10^7, 10^2)$	1	493067	15	15	4793.98	139926	16	15	978.11
	0.5	125408	17	14	894.30	52874	15	14	377.65
	0.1	27941	35	13	194.83	20739	14	13	165.19
	0	*	*	*	*	5952	13	12	42.42
$(10^7, 10^3)$	1	139745	20	15	977.19	35227	20	15	246.07
	0.5	33462	27	14	236.89	11634	19	14	84.47
	0.1	10075	103	13	72.31	4041	19	13	29.63
	0	*	*	*	*	1753	18	12	13.55

Table 3: Performance of the θ-IPAAL method with $(l, n) = (25, 100), \hat{\rho} = \hat{\eta} = 10^{-4}$.

We conclude from Table 2 and 3 that the total number of ACG iterations for both versions of the θ-IPAAL method as well as the runtime decrease as θ approaches zero. The computational results reported in Table 2 and 3 show that the constant version performs considerably better than its theoretical counterpart. This phenomenon can be attributed to the sizes of the scalars σ and λ which directly affect the number of ACG and outer iterations, respectively. Finally, even though the theoretical version is undefined for $\theta = 0$, the above computational results indicate that its constant version counterpart with $\theta = 0$ is quite promising.

7. Concluding remarks. This paper has presented an inexact proximal accelerated augmented Lagrangian (IPAAL) method, based on the θ-AL function (4), for finding an approximate stationary point of the linearly constrained smooth nonconvex composite optimization problem (1) where the prox subproblems are inexactly solved by an accelerated composite gradient (ACG) scheme. It is shown that the θ-IPAAL obtains a $(\hat{\rho}, \hat{\eta})$-approximate stationary point (see Definition 1) of (1) in $\mathcal{O}([1/(\hat{\eta}\hat{\rho}^2)] \log(1/\hat{\eta}))$ ACG iterations. Moreover, it is also shown that the previous bound can be improved to $\mathcal{O}([1/(\sqrt{\hat{\eta}}\hat{\rho}^2)] \log(1/\hat{\eta}))$ under the additional mildly stronger conditions **(B1)**–**(B3)**. The above bounds are derived assuming that the initial point is neither feasible nor the domain of the composite term of the objective function is bounded.

We now make some remarks about the derived complexity bounds in light of the choice of the parameter θ. Our complexity bounds are derived under the assumption that $\theta \in (0, 1]$, and hence does not apply to the case in which $\theta = 0$, i.e., the θ-AL function reduces to the classical quadratic AL function. It turns out that as θ approaches zero, the constant involved in the $\mathcal{O}(\cdot)$ iteration-complexity bounds explodes to infinity and the stepsize λ given by (28) approaches zero. Hence, the case $\theta = 0$ is still an open case whose resolution possibly requires an insight different than the one used in this paper.

Our analysis has assumed that the pair (z_0, p_0) used as part of the input in step 1 of θ-IPAAL is always the same pair. In practice, this pair can be chosen using the following simple warm strategy, namely, set it to be the output $(\hat{z}, \hat{p})$ obtained in step 2 of the θ-IPAAL during its previous loop when c was $c/2$. An interesting topic for future research is to analyze the ACG iteration-complexity of the θ-IPAAL method endowed with this warm strategy.

Appendix A. Two Technical Results.

This section contains two technical results concerning some properties of the ε-subdifferential of a convex function perturbed by a prox-term.

The following result is used in the proof of Lemma 9.

LEMMA 22. *Assume that $\xi > 0$, $\psi \in \overline{\text{Conv}}\,(\Re^n)$ and $Q \in \mathcal{S}^n_{++}$ are such that $\psi - (\xi/2)\|\cdot\|^2_Q$ is convex and let $(y, v, \eta) \in \Re^n \times \Re^n \times \Re$ be such that $v \in \partial_\eta \psi(y)$. Then, for any $\tau > 0$,*

$$\psi(u) \geq \psi(y) + \langle v, u - y\rangle - (1 + \tau^{-1})\eta + \frac{\xi}{2(1+\tau)}\|u - y\|^2_Q \quad \forall u \in \Re^n.$$

Proof: Let $\psi_v := \psi - \langle v, \cdot\rangle$. The assumptions imply that ψ_v has a unique global minimum $\bar{y}$ and that

$$\psi_v(u) \geq \psi_v(\bar{y}) + \frac{\xi}{2}\|u - \bar{y}\|^2_Q \tag{89}$$

for every $u \in \Re^n$. Moreover, Since $v \in \partial_\eta \psi(y)$, we have $\psi_v(u) \geq \psi_v(y) - \eta$ for every $u \in \Re^n$, and hence that

$$\psi_v(\bar{y}) \geq \psi_v(y) - \eta. \tag{90}$$

The relations in (89) and (90) then imply that for every $u \in \Re^n$,

$$\begin{aligned}\psi_v(u) &\geq \psi_v(y) - \eta + \frac{\xi}{2}\|u - \bar{y}\|^2_Q = \psi_v(y) - \left(\eta + \frac{\xi}{2\tau}\|\bar{y} - y\|^2_Q\right) + \frac{\xi}{2}\left[\frac{1}{\tau}\|y - \bar{y}\|^2_Q + \|u - \bar{y}\|^2_Q\right]\\ &\geq \psi_v(y) - \eta' + \frac{\xi}{2(1+\tau)}\|u - y\|^2_Q, \qquad (91)\end{aligned}$$

where the last inequality is due to the following relations

$$\frac{1}{1+\tau}\|\tilde{u} + u'\|^2 \leq \frac{1}{\tau}\|\tilde{u}\|^2 + \|u'\|^2, \qquad \eta' := \eta + \frac{\xi}{2\tau}\|\bar{y} - y\|^2_Q.$$

Also, inequality (89) with $u = y$ and the relation in (90) imply that $(\xi/2)\|\bar{y} - y\|^2_Q \leq \eta$ and hence that $\eta' \leq (1 + \tau^{-1})\eta$. The conclusion now follows from (91), definition of ψ_v and the latter conclusion. ■

The following result is used in the proof of Proposition 12.

LEMMA 23. *Let proper function $\tilde{\phi} : \Re^n \to (-\infty, \infty]$, scalar $\sigma \in (0, 1)$ and $(z_0, z_1) \in \Re^n \times \text{dom}\,\tilde{\phi}$ be given, and assume that there exists (v_1, ε_1) such that*

$$v_1 \in \partial_{\varepsilon_1}\left(\tilde{\phi} + \frac{1}{2}\|\cdot - z_0\|^2\right)(z_1), \quad \|v_1\|^2 + 2\varepsilon_1 \leq \sigma^2\|v + z_0 - z_1\|^2. \tag{92}$$

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

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

Proof: Using the inclusion in (92), the definition of ε-subdifferential in (8), and the fact that $|\langle u, \tilde{u}\rangle| \leq [s\|u\|^2 + s^{-1}\|\tilde{u}\|^2]/2$ for every $u, \tilde{u} \in \Re^n$ and $s > 0$, we conclude that for every $z \in \Re^n$,

$$\begin{aligned}\tilde{\phi}(z) + \frac{\|z - z_0\|^2}{2} - \tilde{\phi}(z_1) &\geq \frac{\|z_1 - z_0\|^2}{2} + \langle v_1, z - z_1\rangle - \varepsilon_1\\ &= \frac{\|z_1 - z_0\|^2}{2} + \langle v_1, z_0 - z_1\rangle + \langle v_1, z - z_0\rangle - \varepsilon_1\\ &\geq \frac{\|z_1 - z_0\|^2}{2} + \langle v_1, z_0 - z_1\rangle - \varepsilon_1 - \frac{\|v_1\|^2}{2s} - \frac{s\|z - z_0\|^2}{2}\\ &\geq \frac{\|v_1 + z_0 - z_1\|^2}{2} - \frac{1}{2}(1 + s^{-1})\left[\|v_1\|^2 + 2\varepsilon_1\right] - \frac{s\|z - z_0\|^2}{2},\end{aligned}$$

which immediately implies the conclusion of the lemma in view of the inequality in (92). ■

Appendix B. Proof of the first statement of Lemma 10.

In this section we prove the first statement of Lemma 10, i.e., $C_\theta \geq 1/8$, where C_θ is as in (56).

Proof: First note that the definition of τ_θ in (25) immediately yields $\tau_\theta(16-17\theta)/\theta \leq 1$, which can be easily seen to be equivalent to

$$\frac{2(1-\theta)}{\theta}\left[\frac{2\tau_\theta}{\tau_\theta+1}\right] \leq \frac{1}{4}. \tag{93}$$

Now, in view of (56), $C_\theta \geq 1/8$ if and only if σ as in (28) satisfies

$$\frac{\sigma^2}{2} + \frac{2(1-\theta)}{\theta}\left[\frac{2\tau_\theta(1+\sigma)^2}{\tau_\theta+1} + 2\sigma(1+\sigma) + \frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right] \leq \frac{3}{8}.$$

Hence, in view of (93), in order to show that $C_\theta \geq 1/8$, it is sufficient to prove that

$$\frac{\sigma^2}{2} + \frac{(1+\sigma)^2}{4} + \frac{2(1-\theta)}{\theta}\left[2\sigma(1+\sigma) + \frac{(\tau_\theta+1)\sigma^2}{\tau_\theta}\right] \leq \frac{3}{8},$$

or equivalently

$$\left(\frac{3}{4} + \frac{2(1-\theta)\,(3\tau_\theta+1)}{\theta\tau_\theta}\right)\sigma^2 + \left(\frac{8-7\theta}{2\theta}\right)\sigma - \frac{1}{8} \leq 0.$$

Since $\theta \in (0,1]$ and $0 < \sigma \leq \sigma_\theta$ in view of (28), the above inequality holds immediately from the fact that σ_θ is the only positive solution of equation (26) associated to the above quadratic inequality. Hence, from the above conclusions, we obtain $C_\theta \geq 1/8$. ■

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 Journal on Imaging Sciences*, 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] Yair Carmon, John C. Duchi, Oliver Hinder, and Aaron Sidford. Accelerated methods for nonconvex optimization. *SIAM Journal on Optimization*, 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. Programming*, 178:503–558, 2019.

[11] S. Ghadimi and G. Lan. Accelerated gradient methods for nonconvex nonlinear and stochastic programming. *Math. Programming*, 156:59–99, 2016.

[12] S. Ghadimi, G. Lan, and H. Zhang. Generalized uniformly optimal methods for nonlinear programming. *Journal of Scientific Computing*, 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 nonsmooth optimization. *Math. Programming*, 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] M. Hong. Decomposing linearly constrained nonconvex problems by a proximal primal dual approach: algorithms, convergence, and applications. *available on arXiv:1604.00543*, 2016.

[18] 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.
[19] 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.
[20] 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 Journal on Optimization*, 29(4):2566–2593, 2019.
[21] 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.
[22] Weiwei Kong, Jefferson G. Melo, and Renato 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.
[23] G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. *Math. Programming*, 138(1):115–139, Apr 2013.
[24] G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order augmented Lagrangian methods for convex programming. *Math. Programming*, 155(1):511–547, Jan 2016.
[25] H. Li and Z. Lin. Accelerated proximal gradient methods for nonconvex programming. *Adv. Neural Inf. Process. Syst.*, 28:379–387, 2015.
[26] 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.
[27] Q. Lin, R. Ma, and Y. Xu. Inexact proximal-point penalty methods for non-convex optimization with non-convex constraints. *Available on https://arxiv.org/abs/1908.11518*, 2019.
[28] 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.
[29] Z. Lu and Z. Zhou. Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. *Available on arXiv:1803.09941*, 2018.
[30] 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.
[31] 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.
[32] R.D.C. Monteiro and B.F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM Journal on Optimization*, 22(3):914–935, 2012.
[33] 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.
[34] Y. Nesterov. Gradient methods for minimizing composite functions. *Math. Programming*, pages 1–37, 2012.
[35] Y. E. Nesterov. *Introductory lectures on convex optimization : a basic course*. Kluwer Academic Publ., 2004.
[36] Y.E. Nesterov and B.T. Polyak. Cubic regularization of newton method and its global performance. *Math. Program.*, 108(1):177–205, 2006.
[37] 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.
[38] 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.
[39] R. T. Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Math. Oper. Res.*, 1(2):97–116, 1976.
[40] P. Tseng. On accelerated proximal gradient methods for convex-concave optimization. `http://www.mit.edu/~dimitrib/PTseng/papers.html`, 2008.
[41] Yangyang Xu. Iteration complexity of inexact augmented lagrangian methods for constrained convex programming. *Math. Programming*, 2019.