Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-023-02218-z

A Proximal Augmented Lagrangian Method for Linearly Constrained Nonconvex Composite Optimization Problems

Jefferson G. Melo[1] · Renato D. C. Monteiro[2] · Hairong Wang[2]

Received: 20 February 2022 / Accepted: 1 April 2023

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. For any given (possibly infeasible) initial point and tolerance $\rho > 0$, it is shown that IPAAL generates an approximate stationary solution in $\mathcal{O}(\rho^{-3}\log(\rho^{-1}))$ ACG iterations, which can be improved to $\mathcal{O}(\rho^{-2.5}\log(\rho^{-1}))$ if it is further assumed that a certain Slater condition holds.

Keywords Inexact proximal augmented Lagrangian methods · Linearly constrained smooth nonconvex composite programs · Accelerated first-order methods · Iteration complexity

---

Communicated by Florian Jarre.

This work was posted on optimization online (see https://optimization-online.org/2020/04/7772/) in April 29, 2020 and posted on Arxiv (see http://arxiv.org/abs/2006.08048) in June 14, 2020. Both of these releases appear with a different title, namely "Iteration-complexity of an Inexact Proximal Accelerated Augmented Lagrangian Method for Solving Linearly Constrained Smooth Nonconvex Composite Optimization Problems".

✉ Jefferson G. Melo
jefferson@ufg.br

Renato D. C. Monteiro
monteiro@isye.gatech.edu

Hairong Wang
hairongwhr@gatech.edu

[1] Instituto de Matemática e Estatística, Universidade Federal de Goiás, Goiânia, GO, Brazil

[2] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

Published online: 26 April 2023

Mathematics Subject Classification 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

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

where $b \in \mathbb{R}^l$, $A : \mathbb{R}^n \mapsto \mathbb{R}^l$ is a linear operator, $h : \mathbb{R}^n \to (-\infty, \infty]$ is a closed proper convex function, and, for some $0 < m_f \le L_f$, f is a real-valued differentiable nonconvex function which is m_f-weakly convex, i.e., $f + m_f\|\cdot\|^2/2$ is convex, and whose gradient is L_f-Lipschitz continuous.

The method, referred to as θ-IPAAL, depends on a given perturbation parameter $\theta > 0$ which in turn determines the θ-augmented Lagrangian (θ-AL) function $\mathcal{L}_c^\theta(z, q)$ defined as

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

where $c > 0$ is a penalty parameter. Note that when $\theta = 0$, $\mathcal{L}_c^\theta(\cdot, \cdot)$ reduces to the classical quadratic augmented Lagrangian (AL) function which has been thoroughly studied in the literature (see, for example, [2, 4, 23, 27, 35]). Moreover, when $\theta = 1$, $\mathcal{L}_c^\theta(\cdot, \cdot)$ does not depend on q and reduces to the quadratic penalty function frequently used by penalty methods for solving (1). For a given tolerance pair $(\hat\rho, \hat\eta) \in \mathbb{R}^2_{++}$, the goal of θ-IPAAL is to find a triple $(\hat z, \hat q, \hat v)$ satisfying

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

Before discussing θ-IPAAL, we first outline its static version, referred to as the static θ-IPAAL, which keeps c always constant. Indeed, for a fixed stepsize $\lambda_\theta > 0$ depending on θ, the static θ-IPAAL repeatedly performs the following iteration for any $k \ge 1$: given $(z_{k-1}, q_{k-1}) \in \mathrm{dom} h \times \mathbb{R}^l$, it computes (z_k, q_k) as

$$\begin{aligned} z_k &\approx \mathrm{argmin}_z \left\{\lambda_\theta \mathcal{L}_c^\theta(z, q_{k-1}) + \frac{1}{2}\|z - z_{k-1}\|^2\right\}, \\ q_k &= (1-\theta)q_{k-1} + c(Az_k - b), \end{aligned} \tag{4}$$

where z_k should be understood as a suitable approximate solution of the prox subproblem (4). We now briefly describe how z_k is computed without elaborating on the inexactness criterion used to solve (4). First note that since f is m_f-weakly convex, the objective function of (4) is strongly convex whenever $\lambda_\theta < 1/m_f$. The static θ-IPAAL sets $\lambda_\theta = \tau_\theta/m_f$ for some $\tau_\theta \in (0, 1)$ such that $\tau_\theta = \mathcal{O}(\theta)$ and then approximately solves the corresponding subproblem (4) by a strongly convex version of an accelerated composite gradient (ACG) method (see, for example, [3, 19, 30, 33]) to obtain z_k. It is shown that each pair (z_k, q_k) obtained in the above manner can always be refined to a triple $(\hat{z}, \hat{q}, \hat{v}) = (\hat{z}_k, \hat{q}_k, \hat{v}_k)$ satisfying the inclusion in (3).

The static θ-IPAAL is then stopped whenever $\hat{v} = \hat{v}_k$ satisfies the first inequality in (3). Moreover, it is shown that the static θ-IPAAL satisfies the following properties: (i) it stops in $\mathcal{O}(\theta^{-5/2}\hat{\rho}^{-2}c^{1/2}\log(c))$ ACG iterations[1]; (ii) every refined iterate $\hat{z}_k$ satisfies $\|A\hat{z}_k - b\| = \mathcal{O}(\theta^{-1}c^{-1/2})$; and (iii) if h in (1) satisfies some mild assumptions such as the Slater condition int $(\mathrm{dom} h) \cap \{z : Az = b\} \neq \emptyset$, then every $\hat{z}_k$ satisfies $\|A\hat{z}_k - b\| = \mathcal{O}(\theta^{-5/2}c^{-1})$.

Observe that the above property (ii) and/or (iii) guarantees that $\hat{z} = \hat{z}_k$ is a near feasible point, i.e., satisfies the second inequality in (3), only when c is sufficiently large. The θ-IPAAL method on the other hand adaptively increases c so as to also obtain the desired near feasibility, and hence a triple $(\hat{z}, \hat{q}, \hat{v})$ satisfying (3). 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 $(\hat{z}, \hat{q}, \hat{v})$ satisfying the inclusion and the first inequality in (3); and (b) doubles c whenever the second inequality in (3) is violated, until it obtains a triple $(\hat{z}, \hat{q}, \hat{v})$ satisfying all the conditions in (3). It is then shown that the ACG iteration complexity of this adaptive variant in terms of the tolerance pair $(\hat{\rho}, \hat{\eta})$ and the parameter θ only is $\mathcal{O}(\theta^{-7/2}\hat{\eta}^{-1}\hat{\rho}^{-2}\log(\hat{\eta}^{-1}))$. Moreover, if some mild additional assumptions hold, then the ACG iteration complexity of θ-IPAAL improves to $\mathcal{O}(\theta^{-15/4}\hat{\eta}^{-1/2}\hat{\rho}^{-2}\log(\hat{\eta}^{-1}))$. It is worth emphasizing that all the results mentioned above are derived without assuming that the initial point $z_0 \in \mathrm{dom} h$ is feasible, i.e., it satisfies $Az_0 = b$. Moreover, the iteration complexities which are mentioned here refer to the effort of obtaining an approximate stationary point as in (3). Note that even though these complexities are described as bounds on the number of (possibly ACG) iterations, they are also bounds on the total number of h-resolvent computations and/or gradient evaluations of f.

Related Works. Iteration-complexity analysis of penalty- and/or AL-type methods for solving convex versions of (1) was considered in [1, 2, 22, 23, 27, 28, 32, 34, 38]. Inexact proximal point methods for solving convex-concave saddle point problems and monotone variational inequalities that use accelerated gradient algorithms to solve their prox subproblems were previously considered in [6, 11, 12, 15, 31].

We will now focus on methods for solving the nonconvex CO problem (1). Iteration complexities for proximal quadratic penalty (PQP) methods were developed in [16, 17, 21, 26]. More specifically, the first bound, namely, $\mathcal{O}(\hat{\eta}^{-1}\hat{\rho}^{-2})$, was obtained in [16], which was then improved to $\mathcal{O}(\hat{\eta}^{-1/2}\hat{\rho}^{-2}\log(1/\hat{\eta}))$ in [26] under the assumption that h is Lipschitz continuous on its domain and that the aforementioned Slater condition holds. AL methods for solving (1) with function h identically zero were studied in [13, 37].

[1] Since each ACG iteration of IPAAL requires $\mathcal{O}(1)$ resolvent evaluations of h and/or gradient evaluations of f, the ACG iteration complexity also bounds the number of h-resolvent computations (i.e., evaluations of $(I + \eta\partial h)^{-1}$ for $\eta > 0$) and gradient evaluations of f performed by θ-IPAAL.

This paragraph discusses AL methods for solving (1) with nontrivial composite function h. Paper [10][2] studies an unaccelerated proximal AL method based on the θ-AL function (2) with θ in $(0, 1]$ and establishes an $\mathcal{O}(1/(\hat{\eta}^4+\hat{\rho}^4))$ iteration complexity to compute an approximate stationary point for (1). Each iteration of this method exactly solves a prox subproblem such as (4) except that f in (2) 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 suitably chosen matrix B.

Paper [36] studies the iteration complexity of a non-proximal AL-type method for solving a nonlinearly constrained version of (1) which uses the AG method of [8] to obtain approximate stationary points of $\mathcal{L}_{c_k}(\cdot; q_{k-1})$ for increasing values of c_k. It is worth mentioning that [36] assumes a strong condition relating the feasibility of an iterate to its stationarity and shows that it holds in the setting of problem (1) with h being the indicator function of a bounded polyhedron or a 2-norm ball. After the first release of our paper, [24] proposed a non-proximal AL method related to [36] where each AL subproblem is solved by an inexact proximal inner accelerated method.

We now describe other papers that are tangentially related to this paper.

Paper [5] considers a primal-dual proximal point scheme and analyzes its complexity under strong conditions on the initial point. Paper [14] considers a penalty-ADMM method that solves an equivalent reformulation of (1). Paper [39] presents a primal-dual first-order algorithm for solving (1) with h being the indicator function of a Euclidean box. Additional discussion of how θ-IPAAL compares with [39] is given in Sect. 6.

Contributions. We now compare the iteration complexity of θ-IPAAL to the ones obtained for others methods for solving (1) under the same assumptions made in this paper (and hence without assuming that h is the indicator function of a Euclidean box as in [39] or making the strong assumption as in [36]). An $\widetilde{\mathcal{O}}(\hat{\eta}^{-1/2}\hat{\rho}^{-2})$ ACG iteration complexity had already been established in [26] for a prox quadratic penalty-type method and in [25] for the aforementioned hybrid inexact proximal type method. Paper [25] also establishes a similar ACG iteration complexity for an inexact proximal point method applied to (1) (i.e., to the function defined as $\phi(z)$ if z is feasible and $+\infty$ otherwise where ϕ is as in (1)), where each (strongly convex) prox subproblem is either solved by a penalty-type method or an AL-type method (and hence its adjective hybrid). Our work though is the first one to establish the aforementioned iteration-complexity bound for a prox AL method directly applied to (1). Furthermore, it has the following features: i) it updates the Lagrange multiplier after every ACG call to solve a composite unconstrained subproblem; ii) its initial point can be infeasible; and iii) it is a generalization of the prox AL method in Section 3 of [35] to nonconvex setting of (1). The prox AL method of [10] also satisfies properties i)–iii), but its iteration complexity is worse, i.e., $\mathcal{O}(1/(\hat{\eta}^4 + \hat{\rho}^4))$, and is established under the restrictive assumption that the initial point z_0 is feasible, i.e., satisfies $Az_0 = b$ and $z_0 \in \text{dom}h$.

Organization of the Paper. Section 2 is divided into three subsections. Section 2.1 provides some notation and basic definitions. Section 2.2 reviews the ACG variant

[2] This method generates prox subproblems of the form $\text{argmin}_{x\in X}\{\lambda h(x)+[c/2]\|Ax-b\|^2/2+[1/2]\|x-x_0\|^2_{B^*B}\}$, which are assumed to be exactly solvable for any x_0, c, λ and matrix B.

that is used to approximately solve the prox subproblems of the θ-IPAAL method. Section 2.3 states our main problem of interest and our assumptions. Section 3 contains two subsections. The first one states the static θ-IPAAL method and its main iteration-complexity results. The second subsection introduces an adaptive variant of the static θ-IPAAL method and establishes its iteration-complexity bounds. Section 4 contains the proofs of Theorems 3.1. This section contains two subsections. The first one is devoted to the proof of the first two statements of Theorem 3.1, while the second subsection contains the proof of the last statement of Theorem 3.1. Section 5 presents the proof of an auxiliary technical result. Section 6 presents some concluding remarks. The appendix contains some basic auxiliary results.

2 Notation, Basic Definitions, and Background Material

This section contains three subsections. The first one describes some basic notation and definitions used in our presentation. The second one reviews an accelerated composite gradient (ACG) variant which will be used as a subroutine by the main algorithm of this paper. The third subsection presents our problem of interest and our assumptions.

2.1 Basic Notation and Definitions

Let $\mathbb{R}^n$ denote the n-dimensional Euclidean space with inner product and associated norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. We use $\mathbb{R}^{l\times n}$ to denote the set of all $l\times n$ matrices. The image space of a matrix $S\in\mathbb{R}^{l\times n}$ is defined as $\mathrm{Im}(S):=\{Sx : x\in\mathbb{R}^n\}$, and $\mathcal{P}_S$ denotes the Euclidean projection onto Im (S). The smallest positive eigenvalue of $(S^*S)^{1/2}$ is denoted by $\sigma^+(S)$. If S is a symmetric and positive semidefinite matrix, the seminorm induced by S on $\mathbb{R}^n$, denoted by $\|\cdot\|_S$, is defined as $\|\cdot\|_S:=\langle S(\cdot),\cdot\rangle^{1/2}$. The distance of a point x to a closed convex set X is denoted by $\mathrm{dist}_X(x)$. The normal cone of X at a point $x\in X$, denoted by $N_X(x)$, is defined by $N_X(x)=\{v\in\mathbb{R}^n : \langle v, z-x\rangle\le 0,\ \forall z\in X\}$. For $t>0$, we define $\bar{B}(0,t):=\{z\in\mathbb{R}^n : \|z\|\le t\}$ and $\log_1^+(t):=\max\{\log t, 1\}$. The domain of a function $g:\mathbb{R}^n\to(-\infty,\infty]$ is the set $\mathrm{dom}g:=\{x\in\mathbb{R}^n : g(x)<+\infty\}$. Moreover, g is said to be proper if $g(x)<\infty$ for some $x\in\mathbb{R}^n$. The set of closed proper convex functions defined in $\mathbb{R}^n$ is denoted by $\overline{\mathrm{Conv}}\,\mathbb{R}^n$. The ε-subdifferential of a function $g\in\overline{\mathrm{Conv}}\,\mathbb{R}^n$ is defined by

$$\partial_\varepsilon g(z):=\{u\in\mathbb{R}^n : g(z')\ge g(z)+\langle u, z'-z\rangle-\varepsilon,\quad \forall z'\in\mathbb{R}^n\},\qquad \forall z\in\mathrm{dom}g. \tag{5}$$

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

$$\ell_\psi(z;\bar{z}):=\psi(\bar{z})+\langle\nabla\psi(\bar{z}), z-\bar{z}\rangle\quad\forall z\in\mathbb{R}^n. \tag{6}$$

2.2 An ACG Variant

This subsection reviews and describes the iteration complexity of an ACG variant which will later be used in the context of the method outlined in the Introduction to approximately solve the prox subproblem (4).

For the purpose of this subsection only, consider the following composite optimization problem:

$$\min\{\psi(x) := \psi^{(s)}(x) + \psi^{(n)}(x) : x \in \mathbb{R}^n\}, \tag{7}$$

where the following conditions are assumed to hold:

(A1) $\psi^{(n)} \in \overline{\text{Conv}}\,\mathbb{R}^n$;

(A2) $\psi^{(s)}$ is convex differentiable on $\text{dom}\psi^{(n)}$ and there exists $(\mu_0, M_0) \in \mathbb{R}^2$ such that $M_0 > \mu_0 > 0$ and

$$\mu_0\|u - x\|^2/2 \le \psi^{(s)}(u) - \ell_{\psi^{(s)}}(u; x) \le M_0\|u - x\|^2/2 \tag{8}$$

for every $x, u \in \text{dom}\psi^{(n)}$, where $\ell_{\psi^{(s)}}(\cdot;\cdot)$ is as in (6).

We are now ready to state the ACG variant.

ACG

---

(0) Let a pair of functions $(\psi^{(s)}, \psi^{(n)})$ satisfying **(A1)** and **(A2)** for some $(\mu_0, M_0) \in \mathbb{R}^2_+$, a scalar $\tilde{\sigma} > 0$, and an initial point $y_0 \in \text{dom}\psi_n$ be given; set $x_0 = y_0$, $A_0 = 0$, $\tau_0 = 1$, $\zeta = 1/(M_0 - \mu_0)$, and $j = 0$;

(1) compute the iterates

$$a_j = \frac{\zeta\tau_j + \sqrt{(\zeta\tau_j)^2 + 4\tau_j A_j}}{2}, \quad A_{j+1} = A_j + a_j, \quad \tilde{x}_j = \frac{A_j y_j + a_j x_j}{A_{j+1}},$$

$$\tau_{j+1} = \tau_j + \mu_0 a_j,$$

$$y_{j+1} = \text{argmin}_{y\in\mathbb{R}^n}\left\{\ell_{\psi^{(s)}}(y; \tilde{x}_j) + \psi^{(n)}(y) + \frac{M_0}{2}\|y - \tilde{x}_j\|^2\right\},$$

$$x_{j+1} = \frac{1}{\tau_{j+1}}\left[\frac{a_j}{\zeta}(y_{j+1} - \tilde{x}_j) + \mu_0 a_j y_{j+1} + \tau_j x_j\right];$$

(2) compute the quantities

$$u_{j+1} = \mu_0(y_{j+1} - x_{j+1}) + \frac{x_0 - x_{j+1}}{A_{j+1}},$$

$$\eta_{j+1} = \frac{1}{2A_{j+1}}\left(\|x_0 - y_{j+1}\|^2 - \tau_{j+1}\|x_{j+1} - y_{j+1}\|^2\right);$$

(3) if the inequality

$$\|u_{j+1}\|^2 + 2\eta_{j+1} \le \sigma_0^2 \|y_0 - y_{j+1} + u_{j+1}\|^2$$

holds, then stop and output $(y, u, \eta) := (y_{j+1}, u_{j+1}, \eta_{j+1})$; otherwise, set $j = j + 1$ and go to (1).

Some remarks about ACG follow. First, the most common way of describing an iteration of ACG is as in Step 1. Second, the auxiliary iterates $\{u_j\}$ and $\{\eta_j\}$ computed in Step 2 are used to develop a stopping criterion for ACG when it is called as a subroutine for solving the subproblems generated in Step 1 of static θ-IPAAL in Sect. 3.

Third, it can be shown (see, for example, [7, 19]) that ACG (without Steps 2 and 3) with $\mu_0 = 0$ corresponds to the well-known FISTA algorithm.

Fourth, the sequence $\{A_j\}$ has the following increasing property:

$$A_j \ge \frac{1}{M_0 - \mu_0} \max\left\{\frac{j^2}{4}, \left(1 + \sqrt{\frac{\mu_0}{4(M_0 - \mu_0)}}\right)^{2(j-1)}\right\}, \qquad \forall j \ge 1.$$

The next result, whose proof can be found, for example, in [19, Lemma 2.13], summarizes the main properties of ACG used in this paper.

Proposition 2.1 *Let $\{(y_j, u_j, \eta_j)\}_{j \ge 1}$ be the sequence generated by ACG applied to (7), where $(\psi^{(s)}, \psi^{(n)})$ is a given pair of data functions satisfying* **(A1)** *and* **(A2)**. *Then, the following statements hold:*

(a) *for every $j \ge 1$, we have $\eta_j \ge 0$ and $u_j \in \partial_{\eta_j}(\psi^{(s)} + \psi^{(n)})(y_j)$;*
(b) *for any $\sigma_0 > 0$, the ACG method outputs a triple $(y, u, \eta) \in \mathbb{R}^n \times \mathbb{R}^n \times [0, +\infty)$ satisfying*

$$u \in \partial_\eta(\psi^{(s)} + \psi^{(n)})(y) \quad \|u\|^2 + 2\eta \le \sigma_0^2 \|y_0 - y + u\|^2$$

in at most

$$\left\lceil 3\sqrt{\frac{M_0}{\mu_0}} \log_1^+ \mathcal{Q}_0 \right\rceil \tag{9}$$

iterations, where

$$\mathcal{Q}_0 := \frac{4(2\mu_0 + 3)M_0}{\sigma_0^2}. \tag{10}$$

2.3 Problem of Interest and Assumptions

This section contains notation, basic definitions, and assumptions considered in this paper.

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

(B1) A is a nonzero linear operator and the feasible set $\mathcal{F} := \{z \in \mathrm{dom} h : Az = b\} \neq \emptyset$;

(B2) $h \in \overline{\mathrm{Conv}}\, \mathbb{R}^n$ and there exists k_h such that

$$\partial h(z) \subseteq \bar{B}(0, k_h) + N_{\mathcal{H}}(z), \quad \forall z \in \mathcal{H} := \mathrm{dom} h;$$

(B3) the diameter $D_h := \sup\{\|z - z'\| : z, z' \in \mathcal{H}\}$ is finite;

(B4) f is nonconvex and differentiable on $\mathcal{H}$, and there exist $L_f \geq m_f > 0$ such that, for every $z, z' \in \mathcal{H}$,

$$\|\nabla f(z') - \nabla f(z)\| \leq L_f \|z' - z\|, \qquad f(z') - \ell_f(z'; z) \geq -\frac{m_f}{2}\|z' - z\|^2. \tag{11}$$

Some results of this paper also assume (in addition to the above assumptions) that the following Slater condition holds:

(B5) there exists $\bar{z} \in \mathrm{int}\, \mathcal{H}$ such that $A\bar{z} = b$.

Some comments about the above conditions are in order. First, **(B2)**[3] has been previously used as an assumption in [26] where (up to a logarithmic term) an $\mathcal{O}(\hat{\rho}^{-2}\hat{\eta}^{-1/2})$ iteration complexity was established for a penalty-type method. Second, it can be easily shown that the first condition in (11) implies that $|f(z') - \ell_f(z'; z)| \leq L_f\|z' - z\|^2/2$ for every $z, z' \in \mathcal{H}$, and hence that the second condition in (11) holds with $m_f = L_f$. However, our analysis considers the case in which $m_f < L_f$ since it may potentially lead to better iteration-complexity bounds. Third, the second condition in (11) is equivalent to the function $f(\cdot) + m_f\|\cdot\|^2/2$ being convex on $\mathcal{H}$. Moreover, since f is nonconvex on $\mathcal{H}$, the smallest m_f satisfying (11) is positive. Finally, **(B4)** implies that $\mathcal{H} \subseteq \mathrm{dom} f$, and hence $\mathcal{H} = \mathrm{dom}\phi$, where ϕ is as in (1).

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

$$0 \in \nabla f(\bar{z}) + \partial h(\bar{z}) + A^*\bar{q}, \quad A\bar{z} - b = 0. \tag{12}$$

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 (12).

[3] It is shown in Lemma A.2 of paper [18], which appeared after this work, that **(B2)** is equivalent to the more usual condition that h restricted to its domain $\mathcal{H}$ is k_h-Lipschitz continuous.

Definition 2.1 Given a tolerance pair $(\hat{\rho}, \hat{\eta}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$, a triple $(\hat{z}, \hat{q}, \hat{v}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ is said to be a $(\hat{\rho}, \hat{\eta})$-approximate stationary point of (1) if it satisfies (3), i.e.,

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}) + A^* \hat{q}, \qquad \|\hat{v}\| \le \hat{\rho}, \qquad \|A\hat{z} - b\| \le \hat{\eta}. \tag{13}$$

Sections 3.1 and 3.2 formally describe the static θ-IPAAL method and its adaptive version, respectively, for finding a $(\hat{\rho}, \hat{\eta})$-approximate stationary point of (1) in the sense of Definition 2.1.

3 The θ-IPAAL Method and Main Complexity Results

This section contains two subsections. The first one states the static θ-IPAAL method and its iteration-complexity bounds. The second subsection presents an adaptive variant of the static θ-IPAAL method and its iteration-complexity bounds.

3.1 The Static θ-IPAAL Method and its Iteration Complexity

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

We start by stating the static θ-IPAAL.

Static θ-IPAAL Method

---

(0) Let parameters $\theta \in (0, 1)$ and $\nu > 0$, initial point $z_0 \in \mathcal{H}$, tolerance $\hat{\rho} > 0$, and penalty parameter $c > 0$ be given, and set $q_0 = 0$, $k = 1$, and

$$L_c := L_f + c\|A\|^2, \quad \tau_\theta := \min\left\{\frac{1}{2}, \frac{\theta}{88(1-\theta)}\right\}, \quad \lambda_\theta := \frac{\tau_\theta}{m_f}, \tag{14}$$

where (m_f, L_f) is as in **(B4)**; also, choose $\sigma \in (0, \tau_\theta]$ and set

$$\tilde{\sigma} := \min\left\{\frac{\nu}{\sqrt{\lambda_\theta L_c + 1}}, \sigma\right\}; \tag{15}$$

(1) let $(z_k, v_k, \varepsilon_k) = (y, u, \eta)$ denote the output of the ACG described in Sect. 2.2 with inputs

$$(\psi^{(s)}, \psi^{(n)}) = (\psi_k^{(s)}, \lambda_\theta h), \quad (M_0, \mu_0) = (\lambda_\theta L_c + 1, 1/2), \quad x_0 = z_{k-1}, \quad \sigma_0 = \tilde{\sigma}, \tag{16}$$

where

$$\psi_k^{(s)} := \lambda_\theta \left(\mathcal{L}_c^\theta(\cdot, q_{k-1}) - h\right) + \frac{1}{2}\|\cdot - z_{k-1}\|^2, \tag{17}$$

and set

$$q_k = (1-\theta)q_{k-1} + c(Az_k - b); \tag{18}$$

(2) compute $(\hat{z}_k, \hat{q}_k, \hat{v}_k)$ as

$$\hat{z}_k := \operatorname{argmin}_u \left\{ \langle \lambda_\theta \left[\nabla f(z_k) + A^* q_k\right] - r_k, u \rangle + \lambda_\theta h(u) + \frac{\lambda_\theta L_c + 1}{2}\|u - z_k\|^2 \right\}, \tag{19}$$

$$\hat{q}_k := (1-\theta)q_{k-1} + c\left(A\hat{z}_k - b\right),$$

$$\hat{v}_k := \frac{1}{\lambda_\theta}\left[(\lambda_\theta L_c + 1)(z_k - \hat{z}_k) + r_k\right] + \nabla f(\hat{z}_k) - \nabla f(z_k) + cA^*A(\hat{z}_k - z_k),$$

where r_k is given by

$$r_k := v_k + z_{k-1} - z_k. \tag{20}$$

(3) if $\|\hat{v}_k\| \le \hat{\rho}$, then output $(\hat{z}, \hat{q}, \hat{v}) := (\hat{z}_k, \hat{q}_k, \hat{v}_k)$ and **stop**; otherwise, set $k \leftarrow k+1$ and return to Step 1.

We now make a few remarks about the static θ-IPAAL. First, it makes two types of iterations, namely the outer iterations indexed by k and the ACG iterations performed during its calls to the ACG method in Step 1. Second, the triple $(z_k, v_k, \varepsilon_k)$ computed in Step 1 can be regarded as an approximate stationary solution of subproblem (4) in a sense that will be described in Proposition 3.1(d) and the remarks following the proof of this proposition. Third, Step 2 uses (z_k, q_k, v_k) computed in Step 1 to construct a triple $(\hat{z}_k, \hat{q}_k, \hat{v}_k)$ satisfying $\hat{v}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^*\hat{q}_k$ and some other technical properties described in Lemma 4.1. Fourth, the static θ-IPAAL stops when the kth residual $\hat{v}_k$ is small as described in Step 3. Fifth, the method terminates, it is shown in Theorem 3.1 that, if c is sufficiently large, then the feasibility residual $\|A\hat{z}_k - b\|$ is also small, and hence that $\hat{z}_k$ is an approximate solution in the sense of Definition 2.1. Sixth, using (14) and (15), it can be easily seen that

$$\tilde{\sigma} \le \sigma \le \tau_\theta \le 1/2, \quad \max\{\sigma, \tau_\theta\} = \mathcal{O}(\theta). \tag{21}$$

Finally, the second paragraph in the Concluding Remarks discusses an equivalent way of describing the static θ-IPAAL method in terms of the classical Lagrangian, i.e., the perturbed Lagrangian $\mathcal{L}_c^\theta(\cdot,\cdot)$ in (2) with $\theta = 0$.

The result below describes some properties of the ACG call in Step 1 of the static θ-IPAAL.

Proposition 3.1 *For every $k \ge 1$, if we define*

$$\psi_k(\cdot) := \psi_k^{(s)}(\cdot) + \lambda_\theta h(\cdot) = \lambda_\theta \mathcal{L}_c^\theta(\cdot, q_{k-1}) + \|\cdot - z_{k-1}\|^2/2, \tag{22}$$

where $\psi_k^{(s)}(\cdot)$ is as in (17), then the following statements hold:

(a) *the function* $\psi_k^{(s)}(\cdot) - \lambda_\theta c\|A(\cdot)\|^2/2$*, and hence* $\psi_k(\cdot) - \lambda_\theta c\|A(\cdot)\|^2/2$*, is* $1/2$*-strongly convex; as a consequence both* $\psi_k^{(s)}(\cdot)$ *and* $\psi_k(\cdot)$ *are* $1/2$*-strongly convex;*

(b) $\psi^{(s)}(\cdot) = \psi_k^{(s)}(\cdot)$ *satisfies the second inequality in* (8) *with* $M_0 = \lambda_\theta L_c + 1$*, where* (λ_θ, L_c) *is as in* (14)*;*

(c) *the ACG method invoked in Step 1 of the static* θ*-IPAAL method finds its output* $(z_k, v_k, \varepsilon_k)$ *in at most*

$$I(c) := \left\lceil 6\left(\frac{\sqrt{L_f} + \sqrt{\tau_\theta c}\|A\|}{\sqrt{m_f}}\right)\log(\mathcal{Q}(c))\right\rceil \tag{23}$$

ACG iterations, where (τ_θ, L_c) *is as in* (14) *and* $\mathcal{Q}(c)$ *is given by*

$$\mathcal{Q}(c) := 64\left[\frac{L_f + \tau_\theta c\|A\|^2}{m_f}\right]^2 \max\left\{\nu^{-2}, \sigma^{-2}\right\}; \tag{24}$$

(d) $(z_k, v_k, \varepsilon_k)$ *satisfies*

$$v_k \in \partial_{\varepsilon_k}\left(\lambda_\theta \mathcal{L}_c^\theta(\cdot, q_{k-1}) + \frac{1}{2}\|\cdot - z_{k-1}\|^2\right)(z_k), \quad \|v_k\|^2 + 2\varepsilon_k \le \tilde{\sigma}^2\|r_k\|^2, \tag{25}$$

where $(\sigma, \tilde{\sigma})$ *is as in* (15) *and* r_k *is as in* (20).

Proof (a) To prove this statement, it suffices to show that $\psi_k^{(s)}(\cdot) - \lambda_\theta c\|A(\cdot)\|^2/2$ is $(1/2)$-strongly convex as this claim trivially implies all the other three claims in a). Indeed, first observe that assumption **(B4)** and the fact that $\tau_\theta \le 1/2$ (see (21)) imply that $\lambda_\theta f(\cdot) + \|\cdot\|^2/2$ is strongly convex with modulus $1 - \lambda_\theta m_f = 1 - \tau_\theta \ge 1/2$. The aforementioned claim now follows by noticing that $\psi_k^{(s)}(\cdot) - \lambda_\theta c\|A(\cdot)\|^2/2$ is equal to $\lambda_\theta f(\cdot) + \|\cdot\|^2/2$ plus a suitably defined affine function.

(b) The fact that ∇f is L_f-continuous in $\operatorname{dom} h$ (see **(B5)**) and the definitions of $\mathcal{L}_c^\theta$ and $\psi_k^{(s)}$ in (2) and (17), respectively, imply that $\nabla\psi_k^{(s)}$ is $(\lambda_\theta L_c + 1)$-Lipschitz continuous in $\operatorname{dom} h$, and hence that b) holds.

(c) & (d) First note that (a) and (b) imply that $(\psi^{(s)}, \psi^{(n)})$ and (M_0, μ_0) in (16) satisfy the assumptions of Proposition 2.1. Now, from the definitions of $\mathcal{Q}_0$ in (10), $\tilde{\sigma}$ in (15), and (μ_0, M_0, σ_0) in (16), we have

$$\begin{aligned} \mathcal{Q}_0 &= \frac{4(2\mu_0 + 3)M_0}{\sigma_0^2} = \frac{16M_0}{\tilde{\sigma}^2} \le 16M_0(\lambda_\theta L_c + 1)\max\left\{\nu^{-2}, \sigma^{-2}\right\} \\ &= 16M_0^2 \max\left\{\nu^{-2}, \sigma^{-2}\right\}. \end{aligned}$$

On the other hand, using the definitions of (λ_θ, L_c) in (14) and M_0 in (16), we have

$$M_0 = \lambda_\theta L_c + 1 = \frac{\tau_\theta(L_f + c\|A\|^2)}{m_f} + 1 \le \frac{(\tau_\theta + 1)L_f + c\tau_\theta\|A\|^2)}{m_f} \le 2\left(\frac{L_f + \tau_\theta c\|A\|^2}{m_f}\right)$$

where the last inequality is due to the fact that $m_f \le L_f$ and $\tau_\theta < 1$ (see **(B4)** and (14)). It follows from the above two inequalities and the definition of $\mathcal{Q}(c)$ in (24) that $Q_0 \le \mathcal{Q}(c)$. Hence, statement (c) follows from (9), the last displayed inequality above, and the fact that $\mu_0 = 1/2$. Statement (d) follows immediately from Proposition 2.1(b).

□

We now give some comments about (25). First, observe that if $(v_k, \varepsilon_k) = (0, 0)$ then (25) implies that z_k is an optimal solution of (4). Second, the quantity $\|v_k\|^2 + 2\varepsilon_k$ in (25) can be viewed as the size of the residual pair (v_k, ε_k), which in turn measures the inexactness of z_k as an approximate solution of (4). Finally, the inequality in (25) requires the size of (v_k, ε_k) to be small relative to $\|r_k\|^2$, and hence it is a criterion that does not involve the tolerances $\hat{\rho}$ and $\hat{\eta}$.

The following result, whose proof will be given in Sects. 4.1 and 4.2, presents some iteration-complexity bounds for the static θ-IPAAL method, under the assumptions introduced in Sect. 2.3. Its statement and several proofs below make use of the quantities

$$\phi_* = \inf_{z\in\mathcal{F}} \phi(z), \qquad \underline{\phi} := \inf_{z\in\mathbb{R}^n} \phi(z), \tag{26}$$

where $\mathcal{F}$ is as in **(B1)**. Clearly, $\phi_* \ge \underline{\phi}$ and they can be easily seen to be finite in view of **(B1)–(B4)**.

Theorem 3.1 *Assume that conditions* **(B1)–(B4)** *hold and define*

$$\bar{R}_\theta := \frac{88(1+2\nu)^2}{\theta\tau_\theta}\left(\phi_* - \underline{\phi} + m_f D_h^2\right), \tag{27}$$

$$\hat{R}_\theta := \frac{4}{\sigma^+(A)\operatorname{dist}_{\partial\mathcal{H}}(\bar{z})}\left[\left(L_f D_h + \|\nabla f(z_0)\| + K_h + \frac{(1+\nu)\sqrt{m_f\bar{R}_\theta}}{\sqrt{2\tau_\theta}}\right)D_h + \frac{3\bar{R}_\theta}{\sqrt{2\theta}}\right], \tag{28}$$

where τ_θ is as in (14), $(\phi_*, \underline{\phi})$ is as in (26), and $(K_h, \mathcal{H})$, D_h, (m_f, L_f), and $\bar{z}$ are as in **(B2)**, **(B3)**, **(B4)**, *and* **(B5)**, *respectively. Then, the following statements about the static θ-IPAAL method hold:*

(a) *its number of outer iterations is bounded by*

$$O_\theta(\hat{\rho}) := \left\lceil 1 + \frac{m_f\bar{R}_\theta}{2\tau_\theta\hat{\rho}^2}\right\rceil; \tag{29}$$

(b) *its output* $(\hat{z}, \hat{q}, \hat{v})$ *satisfies*

$$\hat{v} \in \nabla f(\hat{z}) + \partial h(\hat{z}) + A^*\hat{q}, \qquad \|\hat{v}\| \le \hat{\rho}, \qquad \|A\hat{z} - b\| \le \sqrt{\frac{\bar{R}_\theta}{c}}; \tag{30}$$

as a consequence, if

$$c \ge \bar{c}_\theta(\hat{\eta}) := \frac{\bar{R}_\theta}{\hat{\eta}^2}, \tag{31}$$

for some tolerance $\hat{\eta} > 0$*, then* $(\hat{z}, \hat{q}, \hat{v})$ *is a* $(\hat{\rho}, \hat{\eta})$*-approximate stationary solution of* (1)*;*

(c) *if in addition, the Slater condition* **(B5)** *holds, then* $\hat{z}$ *also satisfies*

$$\|A\hat{z} - b\| \le \frac{\hat{R}_\theta}{c}; \tag{32}$$

as a consequence, if

$$c \ge \hat{c}_\theta(\hat{\eta}) := \min\left\{\bar{c}_\theta(\hat{\eta}), \frac{\hat{R}_\theta}{\hat{\eta}}\right\}, \tag{33}$$

then $(\hat{z}, \hat{q}, \hat{v})$ *is a* $(\hat{\rho}, \hat{\eta})$*-approximate stationary solution of* (1).

We now make a few remarks about Theorem 3.1. First, it follows from Proposition 3.1(b) and Theorem 3.1 that, if c is sufficiently large, then the static θ-IPAAL successfully terminates with a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution of (1) in $O_\theta(\hat{\rho})I(c)$ ACG iterations, where $I(c)$ and $O_\theta(\hat{\rho})$ are as in (23) and (29), respectively. Second, the smaller c is, the smaller the latter complexity is. Hence, if we could only make a single choice of c so as to guarantee successful termination of the static θ-IPAAL and (**B5**) does not hold (resp., (**B5**) holds), then a safe one would be $c = \bar{c}_\theta(\hat{\eta})$ given in (31) (resp. $c = \hat{c}_\theta(\hat{\eta})$ given in (33)), in which case its total ACG complexity in terms of the tolerances $(\hat{\rho}, \hat{\eta})$ would be $\mathcal{O}(\hat{\rho}^{-2}\hat{\eta}^{-1})$ (resp., $\mathcal{O}(\hat{\rho}^{-2}\hat{\eta}^{-1/2})$ in view of (23) and (29). Third, since the quantities $\bar{c}_\theta(\hat{\eta})$ and $\hat{c}_\theta(\hat{\eta})$ are difficult to compute, it is usually impossible to have at our disposal a scalar c as in statements c) or d) of Theorem 3.1, and hence it is not clear how to obtain a $(\hat{\rho}, \hat{\eta})$-approximate stationary point of (1) by just solving a single penalized subproblem.

The next subsection presents an adaptive version of the static θ-IPAAL which adaptively updates the penalty parameter c and whose overall ACG iteration complexity is essentially the same as that of the static θ-IPAAL method with $c = \bar{c}_\theta(\hat{\eta})$ (or with $c = \hat{c}_\theta(\hat{\eta})$ if **(B5)** holds).

3.2 The θ-IPAAL Method and its Iteration Complexity

This subsection presents an adaptive version of θ-IPAAL to obtain a $(\hat{\rho}, \hat{\eta})$-approximate stationary point to (1). This method basically consists of applying the static θ-IPAAL and repeatedly doubling the penalty parameter until the second inequality in (13) is satisfied. The main iteration-complexity result of this scheme is also presented in this subsection.

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

θ-IPAAL Method

---

(0) Let an initial point $z_0 \in \mathcal{H}$, scalars $\theta \in (0, 1)$ and $\nu > 0$, and a pair of tolerances $(\hat{\rho}, \hat{\eta}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ be given, choose an initial penalty parameter $c_1 > 0$ and set $c = c_1$;
(1) execute the static θ-IPAAL with input $(z_0, \theta, \nu, c, \hat{\rho})$, and let $(\hat{z}, \hat{v}, \hat{q})$ be its output;
(2) if $\|A\hat{z} - b\| \leq \hat{\eta}$, stop and output $(\hat{z}, \hat{v}, \hat{q})$; otherwise, set $c \leftarrow 2c$ and return to Step 1.

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

Theorem 3.2 *Assume that conditions* **(B1)–(B4)** *hold and let $\bar{c}_\theta(\hat{\eta})$ and $\hat{c}_\theta(\hat{\eta})$) as in (31) and (33), respectively. Then, the following statements hold:*

(a) *the θ-IPAAL method obtains a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution $(\hat{z}, \hat{q}, \hat{v})$ of the problem (1) in*

$$\mathcal{O}\left(\left\lceil \frac{\sqrt{L_f}\log\left(1+\bar{\chi}^{-1}\right) + \left((1+\bar{\chi})^{1/2}\sqrt{\tau_\theta \bar{c}_\theta(\hat{\eta})}\right)\|A\|}{\sqrt{m_f}} \log\left(\bar{\mathcal{Q}}_1(\hat{\eta})\right)\right\rceil O_\theta(\hat{\rho})\right) \tag{34}$$

ACG iterations, where $O_\theta(\hat{\rho})$ is as in (29), $\bar{c}_\theta(\hat{\eta})$ is as in (31), $\bar{\chi} := c_1/\bar{c}_\theta(\hat{\eta})$, $\bar{\mathcal{Q}}_1(\hat{\eta}) := \mathcal{Q}(c_1 + \bar{c}_\theta(\hat{\eta}))$, and the quantity $\mathcal{Q}(\cdot)$ is as in (24);

(b) *if in addition, the Slater condition* **(B5)** *holds, then θ-IPAAL obtains a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution $(\hat{z}, \hat{q}, \hat{v})$ of the problem (1) in*

$$\mathcal{O}\left(\left\lceil \frac{\sqrt{L_f}\log\left(1+\hat{\chi}^{-1}\right) + \left((1+\hat{\chi})^{1/2}\sqrt{\tau_\theta \hat{c}_\theta(\hat{\eta})}\right)\|A\|}{\sqrt{m_f}} \log\left(\hat{\mathcal{Q}}_1(\hat{\eta})\right)\right\rceil O_\theta(\hat{\rho})\right) \tag{35}$$

ACG iterations, where $\hat{c}_\theta(\hat{\eta})$ is as in (33), $\hat{\chi} := c_1/\hat{c}_\theta(\hat{\eta})$ and $\hat{\mathcal{Q}}_1(\hat{\eta}) := \mathcal{Q}(c_1 + \hat{c}_\theta(\hat{\eta}))$.

Proof (a) First note that the lth loop of the θ-IPAAL method invokes the static θ-IPAAL with penalty parameter $c = c_l$, where $c_l := 2^{l-1}c_1$ for all $l \geq 1$. Hence, it follows from the stopping criterion in Step 2 of θ-IPAAL and Theorem 3.1(b) that θ-IPAAL computes a $(\hat{\rho}, \hat{\eta})$-approximate stationary solution $(\hat{z}, \hat{q}, \hat{v})$ of (1) in at most $\bar{l}$ iterations, where $\bar{l} := \min\left\{l : c_l \geq \bar{c}_\theta(\hat{\eta})\right\}$ and $\bar{c}_\theta(\hat{\eta})$ is as (31). Hence, we have

$$c_l = c_1 2^{l-1} \leq \max\left\{c_1, 2\bar{c}_\theta(\hat{\eta})\right\} \leq 2(c_1 + \bar{c}_\theta(\hat{\eta})) \qquad \forall l = 1, \ldots, \bar{l}. \tag{36}$$

Moreover, in view of Theorem 3.1(a) and Proposition 3.1(b), we conclude that the total number of ACG iterations performed by θ-IPAAL is on the order of

$$\mathcal{O}\left(O_\theta(\hat{\rho}) \sum_{l=1}^{\bar{l}} I(c_l)\right), \tag{37}$$

where $I(c)$ is as in (23) and $O_\theta(\hat{\rho})$ is as in(29). To simplify this bound, note that the definitions of $\mathcal{Q}(\cdot)$ in (24) and $\bar{\mathcal{Q}}_1(\hat{\eta})$ in statement (a), combined with (36) and the fact that $\tau_\theta \leq 1/2$ (see (21)), imply that

$$\mathcal{Q}(c_l) = 64\left[\frac{L_f + \tau_\theta c_l \|A\|^2}{m_f}\right]^2 \max\left\{\nu^{-2}, \sigma^{-2}\right\} \leq \mathcal{Q}(2\tau_\theta[c_1 + \bar{c}_\theta(\hat{\eta})]) \leq \bar{\mathcal{Q}}_1(\hat{\eta}),$$
$$\forall l = 1, \ldots, \bar{l}.$$

Moreover, (36) and the definition of $\bar{\chi}$ in the statement of Theorem 3.2(a) imply that

$$\begin{aligned}
\sum_{l=1}^{\bar{l}}\left(\sqrt{L_f} + \sqrt{\tau_\theta c_l}\|A\|\right) &= \bar{l}\sqrt{L_f} + \sqrt{\tau_\theta c_1}\|A\| \sum_{l=1}^{\bar{l}} \sqrt{2}^{l-1} \\
&\leq \bar{l}\sqrt{L_f} + (1+\sqrt{2})\sqrt{\tau_\theta}\sqrt{2c_1}^{\bar{l}}\|A\| \\
&\leq \sqrt{L_f}\log\left(\frac{4[c_1 + \bar{c}_\theta(\hat{\eta})]}{c_1}\right) + 8\sqrt{\tau_\theta[c_1 + \bar{c}_\theta(\hat{\eta})]}\|A\| \\
&= \left(\sqrt{L_f}\log\left(4\left(1+\bar{\chi}^{-1}\right)\right) + 8\sqrt{(1+\bar{\chi})\tau_\theta \bar{c}_\theta(\hat{\eta})}\|A\|\right).
\end{aligned}$$

It then follows from the last two inequalities above and the definition of $I(c)$ in (23) that

$$\sum_{l=1}^{\bar{l}} I(c_l) = \mathcal{O}\left(\frac{1}{\sqrt{m_f}}\left(\sqrt{L_f}\log\left(1 + \bar{\chi}^{-1}\right) + \sqrt{(1+\bar{\chi})\tau_\theta \bar{c}_\theta(\hat{\eta})}\|A\|\right)\log\left(\bar{\mathcal{Q}}_1(\hat{\eta})\right)\right),$$

which combined with (37) proves statement (a).

(b) The proof of this statement follows the same steps of the proof of (a), but uses Theorem 3.1(c) and the quantities $\hat{c}_\theta(\hat{\eta})$ given in (33) and $(\hat{\chi}, \hat{\mathcal{Q}}_1(\hat{\eta}))$ defined in statement (b), instead of Theorem 3.1(b) and the quantities $\bar{c}_\theta(\hat{\eta})$ given in (31) and $(\bar{\chi}, \bar{\mathcal{Q}}_1(\hat{\eta}))$ defined in statement (a), respectively.

□

We now make some remarks about Theorem 3.2.

First, if the initial penalty parameter c_1 is chosen so as to satisfy $\max\{\bar{\chi}^{-1}, \bar{\chi}\} = \mathcal{O}(1)$ (resp., $\max\{\hat{\chi}^{-1}, \hat{\chi}\} = \mathcal{O}(1)$), then the iteration-complexity bound in (34) (resp., (35)) reduces to

$$O_\theta(\hat{\rho}) I(\bar{c}_\theta(\hat{\eta})), \qquad \left(\text{resp., } O_\theta(\hat{\rho}) I(\hat{c}_\theta(\hat{\eta}))\right), \tag{38}$$

where $I(\cdot)$ and $O_\theta(\hat{\rho})$ are as in (23) and (29), respectively, i.e., to the ACG iteration complexity of the static θ-IPAAL (see the paragraph following Theorem 3.1) with $c = \bar{c}_\theta(\hat{\eta})$ (resp., $c = \hat{c}_\theta(\hat{\eta})$). Second, bound (38) in terms of $(\hat{\rho}, \hat{\eta})$ and θ becomes

$$\mathcal{O}\left(\frac{1}{\theta^{7/2}\hat{\rho}^2\hat{\eta}}\right), \qquad \left(\text{resp., } \mathcal{O}\left(\frac{1}{\theta^{15/4}\hat{\rho}^2\hat{\eta}^{1/2}}\right)\right), \tag{39}$$

due to the definition of $\bar{c}_\theta(\hat{\eta})$ in (31) (resp., $\hat{c}_\theta(\hat{\eta})$ in (33)), and the fact that (21) implies that $\tau_\theta = \mathcal{O}(\theta)$ and hence that the quantity $\bar{R}_\theta$ defined in (27) (resp., $\hat{R}_\theta$ defined in (28)) satisfy $\bar{R}_\theta = \mathcal{O}(\theta^{-2})$ (resp., $\hat{R}_\theta = \mathcal{O}(\theta^{-5/2})$).

Third, a choice of c_1 which satisfies the condition in the first remark and does not depend on $\bar{R}_\theta$ (resp., $\hat{R}_\theta$) is $c_1 = \kappa\hat{\eta}^{-2}$ (resp., $c_1 = \kappa \min\{\hat{\eta}^{-1}, \hat{\eta}^{-2}\}$), for some constant $\kappa > 0$ that does not depend on the tolerances $\hat{\rho}$ and $\hat{\eta}$ (e.g., $\kappa = 1$).

4 Proof of Theorem 3.1

This section contains two subsections. More specifically, the first subsection contains the proofs of statements (a) and (b) of Theorem 3.1, whereas the second subsection presents the proof of its statement (c).

4.1 Proofs of Statements (a) and (b) of Theorem 3.1

The first technical result below describes some important properties about the sequence $\{(\hat{z}_k, \hat{q}_k, \hat{w}_k)\}$ computed in Step 2 of static θ-IPAAL.

Lemma 4.1 *The sequence $\{(\hat{z}_k, \hat{q}_k, \hat{v}_k)\}$ generated by the static θ-IPAAL and the sequence $\{r_k\}$ defined in (20) satisfy, for every $k \geq 1$, the following relations:*

$$\hat{v}_k \in \nabla f(\hat{z}_k) + \partial h(\hat{z}_k) + A^*\hat{q}_k, \tag{40}$$

$$\lambda_\theta \|\hat{v}_k\| \leq \left(1 + 2\tilde{\sigma}\sqrt{\lambda_\theta L_c + 1}\right)\|r_k\|, \qquad \|\hat{z}_k - z_k\| \leq \frac{\tilde{\sigma}}{\sqrt{\lambda_\theta L_c + 1}}\|r_k\|, \tag{41}$$

where L_c is as in (14).

Proof First note that the elements $(\tilde{g}, \tilde{h})$, (z, ε), and $\tilde{L}$ defined as

$$\begin{aligned}
\tilde{g} :=& \lambda_\theta[\mathcal{L}_c^\theta(\cdot, q_{k-1}) - h] \qquad -\langle v_k, \cdot - z_k\rangle + \frac{1}{2}\|\cdot - z_{k-1}\|^2, \\
\tilde{h} :=& \lambda_\theta h, \quad (z, \varepsilon) := (z_k, \varepsilon_k), \\
\tilde{L} :=& \lambda_\theta L_c + 1,
\end{aligned} \tag{42}$$

satisfy the assumptions of Lemma A.2. Indeed, the first assumption of Lemma A.2 holds due to **(B2)**, the second and third assumptions hold due to **(B4)** and by taking into account the definitions of $\mathcal{L}_c^\theta$ in (2) and $\ell_{\tilde{g}}$ in (6) with $\psi = \tilde{g}$, (11), and the fact that $\lambda_\theta = \tau_\theta / m_f$ with $\tau_\theta \in (0, 1/2]$ (see (14)). The inclusion in the assumption of Lemma A.2 holds due to the inclusion in (25) and by taking into account the definitions of $\mathcal{L}_c^\theta$ in (2) and ε-subdifferential in (5). Note also that the relations $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$ and $r_k = v_k + z_{k-1} - z_k$ together with (2) and the first relation in (42) imply that $\nabla\tilde{g}(z) = \lambda_\theta(\nabla f(z_k) + A^* q_k) - r_k$. Moreover, $(\tilde{z}, \widetilde{w})$ in (81) corresponds to $(\hat{z}_k, (\lambda_\theta L_c + 1)(z_k - \hat{z}_k))$, in view of the definition of $\hat{z}_k$ in Step 2 of the static θ-IPAAL. Thus, it follows from the conclusion of Lemma A.2 that

$$(\lambda_\theta L_c + 1)(z_k - \hat{z}_k) - [\lambda_\theta(\nabla f(z_k) + A^* q_k) - r_k] \in \partial(\lambda_\theta h)(\hat{z}_k), \tag{43}$$

$$(\lambda_\theta L_c + 1)\|(z_k - \hat{z}_k)\| \leq \sqrt{2(\lambda_\theta L_c + 1)\varepsilon_k}. \tag{44}$$

Hence, inclusion (40) follows by dividing (43) by λ_θ, adding $\nabla f(\hat{z}_k) + A^*\hat{q}_k$ to both sides of the resulting inclusion, and by noting that $A^*\hat{q}_k - A^* q_k = cA^*A(\hat{z}_k - z_k)$ in view of the identities $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$ and $\hat{q}_k = (1-\theta)q_{k-1} + c(A\hat{z}_k - b)$ given in (18) and (19), respectively. Now, the L_f-Lipschitz continuity of ∇f, the fact that $L_c = L_f + c\|A\|^2$ (see (14), the definition of $\hat{v}_k$ in (19) together with the Cauchy–Schwarz inequality imply that

$$\begin{aligned}
\lambda_\theta\|\hat{v}_k\| &\leq \|(\lambda_\theta L_c + 1)(z_k - \hat{z}_k) + r_k\| + \lambda_\theta(L_f + c\|A\|^2)\|\hat{z}_k - z_k\| \\
&\leq \|r_k\| + 2(\lambda_\theta L_c + 1)\|\hat{z}_k - z_k\|.
\end{aligned}$$

Hence, the first inequality in (41) follows from the inequalities in (25) and (44). The second inequality in (41) immediately follows from the inequalities in (25) and (44). □

Note that (40) of Lemma 4.1 implies that the triple $(\hat{z}, \hat{q}, \hat{v}) := (\hat{z}_k, \hat{q}_k, \hat{v}_k)$ satisfies the inclusion in (13).

The next result describes some relations related to the quantities

$$\Delta q_k := q_k - q_{k-1}, \quad \Delta z_k := z_k - z_{k-1}, \qquad \forall k \geq 1, \tag{45}$$

which are frequently used in our analysis.

Lemma 4.2 *Let $\{(z_k, v_k, q_k)\}$ be generated by the static θ-IPAAL method and consider $\{r_k\}$ as in (20). Then, for every $k \geq 1$, we have*

$$\Delta q_{k+1} = (1-\theta)\Delta q_k + cA\Delta z_{k+1}, \qquad (1-\tilde{\sigma})\|r_k\| \leq \|\Delta z_k\| \leq (1+\tilde{\sigma})\|r_k\|. \tag{46}$$

Moreover, $\|\Delta z_k\| \leq D_h$ and $\|r_k\| \leq 2D_h$, where D_h is as in **(B3)**.

Proof The first relation in (46) follows from (45) and the fact that $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$. Now, note that the inequality in (25) implies that $\|v_k\| \leq \tilde{\sigma}\|r_k\|$, where $r_k = v_k + z_{k-1} - z_k$. Hence, the last relation in (45) together with the triangle inequality implies that

$$\begin{aligned}\|r_k\| &\leq \|\Delta z_k\| + \|v_k\| \leq \|\Delta z_k\| + \tilde{\sigma}\|r_k\|,\\ \|\Delta z_k\| &\leq \|\Delta z_k - v_k\| + \|v_k\| \leq (1+\tilde{\sigma})\|r_k\|.\end{aligned}$$

The above inequalities clearly imply that both inequalities in (46) hold. The last statement of the lemma follows from the definition of Δz_k in (45), the definition of D_h in **(B3)**, the first inequality in (46) and the fact that $\tilde{\sigma} \leq 1/2$ (see (21)).

The next result describes a preliminary estimate of the sequence $\{\mathcal{L}_c^\theta(z_k, q_k) - \mathcal{L}_c^\theta(z_{k-1}, q_{k-1})\}$ where $\mathcal{L}_c^\theta(\cdot,\cdot)$ is as in (2). □

Lemma 4.3 *Let $\{(z_k, v_k, q_k)\}$ be generated by the static θ-IPAAL method and let $\{r_k\}$ and $\{(\Delta q_k, \Delta z_k)\}$ be as in (20) and (45), respectively. Then, for every $k \geq 1$, the following relations hold:*

$$\mathcal{L}_c^\theta(z_k, q_k) - \mathcal{L}_c^\theta(z_k, q_{k-1}) = \frac{(1-\theta)(2-\theta)}{2c}\|\Delta q_k\|^2 + \frac{(1-\theta)\theta}{2c}\left(\|q_k\|^2 - \|q_{k-1}\|^2\right), \tag{47}$$

$$\begin{aligned}\mathcal{L}_c^\theta(z_k, q_k) - \mathcal{L}_c^\theta(z_{k-1}, q_{k-1}) \leq &-\frac{3m_f}{8\tau_\theta}\|r_k\|^2 + \frac{(1-\theta)(2-\theta)}{2c}\|\Delta q_k\|^2\\ &+ \frac{(1-\theta)\theta}{2c}\left(\|q_k\|^2 - \|q_{k-1}\|^2\right).\end{aligned} \tag{48}$$

Proof In view of the definition of $\mathcal{L}_c^\theta$ given in (2), the fact that $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$, and the first relation in (45), we have

$$\begin{aligned}\mathcal{L}_c^\theta(z_k, q_k) - \mathcal{L}_c^\theta(z_k, q_{k-1}) &= (1-\theta)\left\langle \Delta q_k, Az_k - b\right\rangle\\ &= (1-\theta)\left\langle \Delta q_k, \frac{q_k - (1-\theta)q_{k-1}}{c}\right\rangle\\ &= \frac{1-\theta}{c}\left[\|\Delta q_k\|^2 + \theta\left\langle \Delta q_k, q_{k-1}\right\rangle\right]\\ &= \frac{1-\theta}{c}\left[\|\Delta q_k\|^2 + \frac{\theta}{2}\left([\|\Delta q_k + q_{k-1}\|^2 - \|\Delta q_k\|^2 - \|q_{k-1}\|^2\right)\right],\end{aligned}$$

which immediately implies (47) upon using the identity $q_k = \Delta q_k + q_{k-1}$. Now, it follows from (25), the identity $r_k = v_k + z_{k-1} - z_k$, the second relation in (45), and the Cauchy–Schwarz inequality, that

$$\lambda_\theta \mathcal{L}_c^\theta(z_k, q_{k-1}) - \lambda_\theta \mathcal{L}_c^\theta(z_{k-1}, q_{k-1}) \leq -\frac{1}{2}\|z_k - z_{k-1}\|^2 + \langle v_k, z_k - z_{k-1}\rangle + \varepsilon_k$$
$$= -\frac{1}{2}\|\Delta z_k - v_k\|^2 + \left(\frac{\|v_k\|^2}{2} + \varepsilon_k\right) \leq -\frac{1-\tilde\sigma^2}{2}\|\Delta z_k - v_k\|^2 = -\frac{1-\tilde\sigma^2}{2}\|r_k\|^2,$$

and hence that (48) holds in view of (47) and the facts that $\lambda_\theta = \tau_\theta / m_f$ and $\tilde\sigma \leq 1/2$ (see (14) and (21)). □

While the previous result does not use the fact that τ_θ is given by (14), the next one uses (14) to show that the "bad" term $\|\Delta q_k\|^2$ in (48) is majorized by a multiple of $\|r_k\|^2$ and some summable terms.

Lemma 4.4 *Let $\{r_k\}$ and $\{(\Delta q_k, \Delta z_k)\}$ be as in (20) and (45), respectively, and define*

$$B_\theta := \frac{2(1-\theta)}{\theta}\left[\frac{\tau_\theta(1+\tilde\sigma)^2}{2(1+\tau_\theta)} + 2\tilde\sigma(1+\tilde\sigma) + \frac{\tilde\sigma^2(1+\tau_\theta)}{\tau_\theta}\right] \tag{49}$$
$$t_k := \frac{m_f(1-\theta)}{\tau_\theta \theta}\left[\|\Delta z_k\|^2 + \left(\tilde\sigma_\theta(1+\tilde\sigma_\theta) + \frac{\tilde\sigma^2(1+\tau_\theta)}{\tau_\theta}\right)\|r_k\|^2\right], \tag{50}$$

where τ_θ is as in (14), $\tilde\sigma$ is as in (15), and m_f is as in **(B2)**. *Then, the following statements hold:*

(a) *there holds $B_\theta \leq 1/8$;*
(b) *for every $k \geq 2$, we have*

$$\frac{(1-\theta)(2-\theta)}{c}\|\Delta q_k\|^2 + \frac{(1-\theta)^3}{\theta c}\left[\|\Delta q_k\|^2 - \|\Delta q_{k-1}\|^2\right]$$
$$\leq \frac{m_f B_\theta}{\tau_\theta}\|r_k\|^2 + t_{k-1} - t_k. \tag{51}$$

Proof (a) The definition of B_θ in (49) and the facts that $\tilde\sigma \leq \tau_\theta$ and $\tau_\theta \leq 1/2$ (see (21)) imply that

$$B_\theta \leq \frac{7(1-\theta)}{\theta}\tau_\theta(\tau_\theta + 1) \leq \frac{21(1-\theta)}{2\theta}\tau_\theta,$$

which proves the conclusion of b) in view of the fact that $\tau_\theta \leq \theta/[88(1-\theta)]$ (see (14).

(b) The proof of this statement will be presented in Sect. 5.

Lemma 4.5 *Let $\{r_k\}$, $\{(\Delta q_k, \Delta z_k)\}$, and $\{t_k\}$ be as in (20), (45), and (50), respectively. Define*

$$\eta_k := \frac{(1-\theta)^3}{\theta c}\|\Delta q_k\|^2 - \frac{(1-\theta)\theta}{2c}\|q_k\|^2 + t_k, \qquad \forall k \geq 1. \tag{52}$$

Then, for every $k \geq 2$, we have

$$\mathcal{L}_c^\theta(z_k, q_k) + \eta_k + \frac{m_f}{4\tau_\theta}\|r_k\|^2 \leq \mathcal{L}_c^\theta(z_{k-1}, q_{k-1}) + \eta_{k-1}, \tag{53}$$

$$\mathcal{L}_c^\theta(z_k, q_k) + \eta_k + \frac{m_f}{4\tau_\theta}\sum_{i=2}^{k}\|r_i\|^2 \leq \mathcal{L}_c^\theta(z_1, q_1) + \eta_1, \tag{54}$$

where τ_θ and m_f are as in (14) *and* **(B2)**, *respectively.*

Proof Inequality (53) follows by combining Lemma 4.4 with (48) and by using the definitions of $\{t_k\}$ and $\{\eta_k\}$ given in (50) and (52), respectively. Finally, (54) is obtained by summing (53) from $k = 2$ to k. □

Lemma 4.5 shows that the quantity $\mathcal{L}_c^\theta(z_k, q_k) + \eta_k$ plays the role of a (iterate dependent) potential for the static θ-IPAAL. The next result provides a lower bound on the kth potential $\mathcal{L}_c^\theta(z_k, q_k) + \eta_k$ in terms of the feasibility gap $\|Az_k - b\|$ and the quantity $\|q_k\|$.

Lemma 4.6 *Let $\{(z_k, v_k, q_k)\}$ be generated by the static θ-IPAAL method and consider $\{\eta_k\}$ as in* (52). *Then, for every $k \geq 1$, we have:*

$$\mathcal{L}_c^\theta(z_k, q_k) + \eta_k \geq \phi(z_k) + \frac{c}{2}\|Az_k - b\|^2 + \frac{(1-\theta)\theta}{4c}\|q_k\|^2. \tag{55}$$

Proof Noting that (52) and the assumption that $\theta \in (0, 1)$ imply that $t_k \geq 0$, and using the definitions of $\mathcal{L}_c^\theta$ and $\{\eta_k\}$ given in (2) and (52), respectively, we conclude that for every $k \geq 1$,

$$\begin{aligned}
&\mathcal{L}_c^\theta(z_k, q_k) + \eta_k \geq \mathcal{L}_c^\theta(z_k, q_k) - \frac{(1-\theta)\theta}{2c}\|q_k\|^2 + \frac{(1-\theta)^3}{\theta c}\|\Delta q_k\|^2 \\
&= \phi(z_k) + \frac{c}{2}\|Az_k - b\|^2 + (1-\theta)\langle q_k\,,\, Az_k - b\rangle \\
&- \frac{(1-\theta)\theta}{2c}\|q_k\|^2 + \frac{(1-\theta)^3}{\theta c}\|\Delta q_k\|^2 \\
&= \phi(z_k) + \frac{c}{2}\|Az_k - b\|^2 + \frac{(1-\theta)\theta}{2c}\|q_k\|^2 \\
&+ \frac{(1-\theta)^2}{c}\langle q_k\,,\, \Delta q_k\rangle + \frac{(1-\theta)^3}{\theta c}\|\Delta q_k\|^2,
\end{aligned}$$

where the second equality follows from the fact that $c(Az_k - b) = (1-\theta)\Delta q_k + \theta q_k$, in view of the relations $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$ and $\Delta q_k = q_k - q_{k-1}$ (see (18) and (45)). The conclusion of the lemma now follows by noting that the minimum value with respect to Δq_k of the expression in the third line of the above inequality agrees with the right side of (55). □

The next result provides an upper bound on the first potential $\mathcal{L}_c^\theta(z_1, q_1) + \eta_1$ which is independent of c.

Lemma 4.7 *The first potential* $\mathcal{L}_c^\theta(z_1, q_1)+\eta_1$ *of the static* θ*-IPAAL method is bounded by*

$$\mathcal{L}_c^\theta(z_1, q_1) + \eta_1 \le \frac{11N_\theta}{\theta} + \underline{\phi}, \tag{56}$$

where $\underline{\phi}$ *and* η_1 *are as in* (26) *and* (52)*, respectively, and* N_θ *is defined as*

$$N_\theta := \phi_* - \underline{\phi} + \frac{m_f D_h^2}{\tau_\theta}. \tag{57}$$

Proof Using (47) and (52) both with $k = 1$ and the fact that $q_0 = 0$ (and hence that $\Delta q_1 = q_1$), we have

$$\begin{aligned}&\mathcal{L}_c^\theta(z_1, q_1) + \eta_1 - \mathcal{L}_c^\theta(z_1, q_0) - t_1\\ &\quad= \left[\theta(2-\theta) + 2(1-\theta)^2\right]\frac{(1-\theta)\|q_1\|^2}{2\theta c} \le \frac{3\|q_1\|^2}{2\theta c},\end{aligned} \tag{58}$$

where the last inequality is due to the fact that $\theta \in (0, 1)$. Moreover, noting that (25) with $k = 1$ together with the fact that $r_1 = v_1 + z_0 - z_1$ (see (20)) implies that the assumptions of Lemma A.3 with $\tilde{\phi} = \lambda_\theta \mathcal{L}_c^\theta(\cdot, q_0)$ hold, the conclusion of this lemma with $s = 1$ and $z \in \mathcal{F}$, the fact $\mathcal{L}_c^\theta(z, q_0) = \phi(z)$ when $z \in \mathcal{F}$ (see (2) and **(B1)**) and $\tilde{\sigma} \le 1/2$ (see (21)), then imply that

$$\begin{aligned}\mathcal{L}_c^\theta(z_1, q_0) &\le \min_{z\in\mathcal{F}}\left\{\phi(z) + \frac{1}{\lambda_\theta}\|z - z_0\|^2\right\} \le \frac{D_h^2}{\lambda_\theta} + \min_{z\in\mathcal{F}}\phi(z),\\ &= \frac{m_f D_h^2}{\tau_\theta} + \phi_* = N_\theta + \underline{\phi}\end{aligned} \tag{59}$$

where the second inequality follows from the definition of D_h in **(B3)**, the first equality follows from the fact that $\lambda_\theta = \tau_\theta / m_f$ and the definition of ϕ_* in (26), and the last equality is due to (57). The above inequality, the definition of $\mathcal{L}_c^\theta$ in (2), the definition of $\underline{\phi}$ in (26), and the fact that $q_0 = 0$, and hence $q_1 = c(Az_1 - b)$ in view of (18) with $k = 1$, then imply that

$$N_\theta \ge \mathcal{L}_c^\theta(z_1, q_0) - \underline{\phi} = (\phi(z_1) - \underline{\phi}) + \frac{c}{2}\|Az_1 - b\|^2 \ge \frac{c}{2}\|Az_1 - b\|^2 = \frac{\|q_1\|^2}{2c}. \tag{60}$$

Combining (58), (59), and (60), and the fact that $0 < \theta < 1$, we then conclude that

$$\mathcal{L}_c^\theta(z_1, q_1) + \eta_1 - t_1 \le N_\theta + \underline{\phi} + \frac{3N_\theta}{\theta} \le \frac{4N_\theta}{\theta} + \underline{\phi}.$$

Now, using the definition of t_1 in (50) and the bounds on $\|\Delta z_k\|$ and $\|r_k\|$ in Lemma 4.2, we conclude that

$$
\begin{aligned}
t_1 &\leq \left[1 + 4\left(\tilde{\sigma}(1+\tilde{\sigma}) + \frac{(1+\tau_\theta)\tilde{\sigma}^2}{\tau_\theta}\right)\right] \frac{(1-\theta)m_f D_h^2}{\tau_\theta \theta} \\
&\leq \frac{7m_f D_h^2}{\tau_\theta \theta} \leq \frac{7N_\theta}{\theta},
\end{aligned}
$$

where the second inequality is due to the fact that $\theta \in (0, 1)$ and $\tilde{\sigma} \leq \tau_\theta \leq 1/2$ (see (21)), and the last inequality is due to the definition of N_θ in (57). The conclusion of the lemma now follows by combining the above two relations. □

The next result shows that $\|A\hat{z}_k - b\| = \mathcal{O}(1/\sqrt{c})$ and $\|\hat{q}_k\| = \mathcal{O}(\sqrt{c})$.

Lemma 4.8 *The sequence $\{(\hat{z}_k, \hat{q}_k)\}$ generated by the static θ-IPAAL satisfies, for every $k \geq 2$,*

$$
\|A\hat{z}_k - b\| \leq \sqrt{\frac{\bar{R}_\theta}{c}}, \qquad \|\hat{q}_k\| \leq 3\sqrt{\frac{c\bar{R}_\theta}{2\theta}}, \qquad \sum_{j=2}^{k} \|\hat{v}_j\|^2 \leq \frac{m_f \bar{R}_\theta}{2\tau_\theta}, \tag{61}
$$

where ν and $\bar{R}_\theta$ are as in Step 0 of the θ-IPAAL and (27), respectively.

Proof Using the fact that $\phi(z_k) \geq \underline{\phi}$ (see (26)) and the relations (54), (55), and (56), we conclude that

$$
\begin{aligned}
\frac{c}{2}\|Az_k - b\|^2 + \frac{(1-\theta)\theta}{4c}\|q_k\|^2 + \sum_{j=2}^{k} \frac{m_f}{4\tau_\theta}\|r_j\|^2 &\leq \mathcal{L}_c^\theta(z_k, q_k) + \eta_k \\
&\quad - \phi(z_k) + \sum_{j=2}^{k} \frac{m_f}{4\tau_\theta}\|r_j\|^2 \\
&\leq \mathcal{L}_c^\theta(z_1, q_1) + \eta_1 - \phi(z_k) \leq \frac{11N_\theta}{\theta} - [\phi(z_k) - \underline{\phi}] \leq \frac{\bar{R}_\theta}{8(1+2\nu)^2},
\end{aligned}
$$

where the last inequality is due to the definitions of $\bar{R}_\theta$ and N_θ given in (27) and (57), respectively. Hence, since $\nu > 0$, the following inequalities hold:

$$
\|Az_k - b\| \leq \frac{1}{2}\sqrt{\frac{\bar{R}_\theta}{c}}, \qquad \sqrt{1-\theta}\,\|q_k\| \leq \sqrt{\frac{c\bar{R}_\theta}{2\theta}}, \qquad \sum_{j=2}^{k} \|r_j\|^2 \leq \frac{\tau_\theta \bar{R}_\theta}{2m_f(1+2\nu)^2}. \tag{62}
$$

Now, using the triangle inequality for norms and the second inequality in (41), we obtain

$$\|A\hat{z}_k - b\| - \|Az_k - b\| \le \|A(\hat{z}_k - z_k)\| \le \|A\|\|\hat{z}_k - z_k\|$$
$$\le \frac{\tilde{\sigma}\|A\|}{\sqrt{\lambda_\theta L_c + 1}}\|r_k\| \le \frac{1}{2}\sqrt{\frac{m_f}{\tau_\theta c}}\|r_k\|,$$

where the last inequality is due to the $\lambda_\theta L_c = [\tau_\theta/m_f](L_f + c\|A\|^2) \ge [\tau_\theta c/m_f]\|A\|^2$ and $\tilde{\sigma} \le 1/2$ (see (14) and (21)). Hence, the first inequality in (61) follows in view of the first and third inequalities in (62), and the fact that $\nu > 0$. Since $\theta \in (0, 1)$ and $\hat{q}_k = (1-\theta)q_{k-1} + c\left(A\hat{z}_k - b\right)$ (see (19)), it follows from the triangle inequality for norms, the first inequality in (61), and the second inequality in (62) that

$$\|\hat{q}_k\| \le (1-\theta)\|q_{k-1}\| + c\|A\hat{z}_k - b\| \le \sqrt{\frac{c\bar{R}_\theta}{2\theta}} + c\sqrt{\frac{\bar{R}_\theta}{c}},$$

which clearly proves the second inequality in (61). Finally, the last inequality in (61) follows immediately from the last inequality in (62) combined with the fact that $\|\hat{v}_k\| \le \|r_k\|\,(1+2\nu)\,m_f/\tau_\theta$ in view of the first inequality in (41), the fact that $\lambda_\theta = \tau_\theta/m_f$, and the definition of $\tilde{\sigma}$ (see (14) and (15)). □

We are now ready to prove statements (a), (b) and (c) of Theorem 3.1.

Proofs of Statements (a) and (b) of Theorem 3.1 (a) The last inequality in (61) implies that

$$(k-1)\min_{j=1,\dots,k}\|\hat{v}_j\|^2 \le \frac{m_f\bar{R}_\theta}{2\tau_\theta} \qquad \forall k \ge 2,$$

and hence that the stopping criterion in step 3 of the static θ-IPAAL is satisfied at some iteration $k \le O_\theta(\hat{\rho})$ where $O_\theta(\hat{\rho})$ is as in (29).

(b) First note that the output $(\hat{z}, \hat{q}, \hat{v})$ is equal to $(\hat{z}_k, \hat{q}_k, \hat{v}_k)$ for some iteration k. Hence, the inclusion in (30) follows from the one in (40). The first inequality in (30) follows from the stopping criterion of static θ-IPAAL given in Step 3. The last inequality in (30) follows from the first inequality in (61). The last statement of (b) follows from (30), the assumption on c in (31), and Definition 2.1.

□

4.2 Proof of Statement (c) of Theorem 3.1

This subsection contains the proof of statement (c) of Theorem 3.1. Before presenting this proof, we establish the boundedness of the Lagrangian multipliers sequence $\{q_k\}$ and its associated sequence $\{\hat{q}_k\}$.

We start by recalling a technical result which will be used in the proof of the subsequent lemma. Its proof can be found, for instance, in [26, Lemma 1].

Lemma 4.9 *Assume that X is a convex set and $\bar{x} \in \mathrm{int}\,(X)$, and let ∂X denote the boundary of X. Then,* $\mathrm{dist}_{\partial X}(\bar{x}) > 0$ *and*

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

The following result shows that the component of the inclusion in (40) lying in $\partial h(\hat{z}_k)$ is bounded. It is worth noting that its proof strongly relies on the bound for $\|\hat{q}_k\|$ derived in Lemma 4.8.

Lemma 4.10 *Assume that* **(B1)–(B5)** *hold and let M_θ be defined by*

$$M_\theta := \frac{1}{\mathrm{dist}_{\partial\mathcal{H}}(\bar{z})}\left[\left(L_f D_h + \|\nabla f(z_0)\| + K_h + \frac{\sqrt{m_f \bar{R}_\theta}}{\sqrt{2\tau_\theta}}\right) D_h + \frac{3\bar{R}_\theta}{\sqrt{2\theta}}\right], \tag{63}$$

where $\bar{R}_\theta$ is as in (27), D_h is as in **(B3)***, and $\bar{z}$ is as in* **(B5)***. Let $\{(\hat{z}_k, \hat{q}_k, \hat{v}_k)\}$ be generated by the static θ-IPAAL method and consider the sequence $\{\hat{\xi}_k\}$ given by*

$$\hat{\xi}_k := \hat{v}_k - \nabla f(\hat{z}_k) - A^*\hat{q}_k, \qquad \forall k \ge 1. \tag{64}$$

Then, we have $\hat{\xi}_k \in \partial h(\hat{z}_k)$ and $\|\hat{\xi}_k\| \le K_h + M_\theta$, for every $k \ge 2$.

Proof The first statement of the lemma immediately follows from (40) and the definition of $\hat{\xi}_k$ given in (64). It follows from the Cauchy–Schwarz inequality and the first two inequalities in (61) that

$$|\langle \hat{q}_k, A\hat{z}_k - b\rangle| \le \|\hat{q}_k\|\|A\hat{z}_k - b\| \le 3\sqrt{\frac{c\bar{R}_\theta}{2\theta}}\sqrt{\frac{\bar{R}_\theta}{c}} \le \frac{3\bar{R}_\theta}{\sqrt{2\theta}}, \qquad \forall k \ge 2. \tag{65}$$

On the other hand, since $\hat{\xi}_k \in \partial h(\hat{z}_k)$ for every $k \ge 2$, it follows from **(B2)** that $\hat{\xi}_k \in \bar{B}(0, K_h) + N_{\mathcal{H}}(\hat{z}_k)$, or equivalently, there exist $\hat{\xi}_k^s$ and $\hat{\xi}_k^n$ such that

$$\hat{\xi}_k = \hat{\xi}_k^s + \hat{\xi}_k^n, \quad \hat{\xi}_k^n \in N_{\mathcal{H}}(\hat{z}_k), \quad \|\hat{\xi}_k^s\| \le K_h. \tag{66}$$

Let $\bar{z}$ be as in (**B5**) and note that $A\bar{z} = b$. Hence, it follows from Lemma 4.9 with $x = \hat{z}_k$, $\bar{x} = \bar{z}$ and $X = \mathcal{H}$, the definition of $\hat{\xi}_k$ in (64), and the first two relations in (66) that, for every $k \ge 2$,

$$\begin{aligned}\mathrm{dist}_{\partial\mathcal{H}}(\bar{z})\|\hat{\xi}_k^n\| &\le \langle \hat{\xi}_k^n, \hat{z}_k - \bar{z}\rangle = \langle \hat{\xi}_k - \hat{\xi}_k^s, \hat{z}_k - \bar{z}\rangle \\ &= \langle \hat{v}_k - \nabla f(\hat{z}_k) - \hat{\xi}_k^s, \hat{z}_k - \bar{z}\rangle - \langle A^*\hat{q}_k, \hat{z}_k - \bar{z}\rangle,\end{aligned}$$

which in view of the Cauchy–Schwarz inequality, the triangle inequality for norms, the L_f-Lipschitz continuity of ∇f, and the inequality in (66), imply that

$$\begin{aligned}\mathrm{dist}_{\partial\mathcal{H}}(\bar{z})\|\hat{\xi}_k^n\| &\le \Big(\|\nabla f(\hat{z}_k) - \nabla f(z_0)\| + \|\nabla f(z_0)\| + \|\hat{\xi}_k^s\| + \|\hat{v}_k\|\Big) \| \\ &\quad \bar{z} - \hat{z}_k\| - \langle \hat{q}_k, A\hat{z}_k - b\rangle \\ &\le \left(L_f\|z_k - z_0\| + \|\nabla f(z_0)\| + K_h + \frac{\sqrt{m_f \bar{R}_\theta}}{\sqrt{2\tau_\theta}} \right) \|\bar{z} - \hat{z}_k\| + \frac{3\bar{R}_\theta}{\sqrt{2\theta}},\end{aligned}$$

where the last inequality is due to the last inequality in (61) and (65). It then follows from the above inequality and the definitions of D_h and M_θ in **(B3)** and (63), respectively, that $\|\hat{\xi}_k^n\| \le M_\theta$, for all $k \ge 2$. From the last inequality, the first and third relations in (66), and the triangle inequality for norms, we obtain $\|\hat{\xi}_k\| \le \|\hat{\xi}_k^s\| + \|\hat{\xi}_k^n\| \le K_h + M_\theta$ for every $k \ge 2$, which proves the last statement of the lemma. □

In the following, we state a basic result that will be used in the proof of the subsequent lemma. Its proof can be found, for instance, in [9, Lemma 1.4].

Lemma 4.11 *Let $S \in \mathbb{R}^{l\times n}$ be a nonzero matrix and let $\sigma^+(S)$ denote the smallest positive eigenvalue of $(S^*S)^{1/2}$. Then, for every $u \in \mathbb{R}^l$, there holds*

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

The next result shows that the sequences of multipliers $\{q_k\}$ and $\{\hat{q}_k\}$ are bounded by a quantity which does not depend on the penalty parameter c. This fact easily implies that $\|A\hat{z}_k - b\| = \mathcal{O}(1/c)$.

Lemma 4.12 *Under the assumptions* **(B1)–(B5)***, we have*

$$\|\hat{q}_k\| \le \frac{\hat{R}_\theta}{2}, \qquad \|q_k\| \le \frac{\hat{R}_\theta}{2}, \qquad \forall k \ge 2, \tag{67}$$

where $\hat{R}_\theta$ is as in (28).

Proof It follows from the definition of $\hat{\xi}_k$ in (64) and the triangle inequality for norms that

$$\begin{aligned}\|A^*\hat{q}_k\| = \|\hat{v}_k - \nabla f(\hat{z}_k) - \hat{\xi}_k\| &\le \|\nabla f(\hat{z}_k) - \nabla f(z_0)\| + \|\nabla f(z_0)\| + \|\hat{\xi}_k\| + \|\hat{v}_k\| \\ &\le L_f D_h + \|\nabla f(z_0)\| + K_h + M_\theta + \frac{\sqrt{m_f\bar{R}_\theta}}{\sqrt{2\tau_\theta}},\end{aligned} \tag{68}$$

where the last inequality is due to the L_f-Lipschitz continuity of ∇f (see (11)), the definition of D_h in **(B3)**, the last inequality in (61), and the last statement of

Lemma 4.10. On the other hand, since $q_0 = 0$ and $b \in \operatorname{Im} A$, the relations $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$ and $\hat{q}_k = (1-\theta)q_{k-1} + c(A\hat{z}_k - b)$ given in (18) and (19), respectively, imply that $q_k, \hat{q}_k \in \operatorname{Im} A$, for every $k \geq 1$. Hence, it follows from Lemma 4.11 with $S = A$ that $\|\hat{q}_k\| \leq \|A^*\hat{q}_k\|/[\sigma^+(A)]$, which combined with (68) imply that

$$\|\hat{q}_k\| \leq \frac{1}{\sigma^+(A)}\left[L_f D_h + \|\nabla f(z_0)\| + K_h + M_\theta + \frac{\sqrt{m_f \bar{R}_\theta}}{\sqrt{2\tau_\theta}}\right], \qquad \forall k \geq 2. \tag{69}$$

The first inequality in (67) then follows from the above inequality, the definitions of $\hat{R}_\theta$ and M_θ in (28) and (63), respectively, and the facts that $D_h/\operatorname{dist}_{\partial\mathcal{H}}(\bar{z}) \geq 1$ and $\sigma^+(A)/\|A\| \leq 1$. Now, from the relations of q_k and $\hat{q}_k$ given above, we obtain $q_k - \hat{q}_k = cA(z_k - \hat{z}_k)$. Then, using the triangle inequality for norms, the second inequality in (41), the facts that $\lambda_\theta = \tau_\theta/m_f$ and $L_c = L_f + c\|A\|^2$ (see (14)), and the definition of $\tilde{\sigma}$ in (15), we have

$$\begin{aligned}\|q_k\| - \|\hat{q}_k\| &\leq c\|A(z_k - \hat{z}_k)\| \leq \frac{\tilde{\sigma}c\|A\|\|r_k\|}{\sqrt{\lambda_\theta L_c + 1}} \leq \frac{\nu m_f c\|A\|}{\tau_\theta(L_f + c\|A\|^2)}\|r_k\| \leq \frac{\nu m_f}{\tau_\theta\|A\|}\|r_k\| \\ &\leq \frac{\nu\sqrt{m_f \bar{R}_\theta}}{\|A\|\sqrt{2\tau_\theta}},\end{aligned}$$

where the last inequality is due to the last inequality in (62). Hence, the last inequality in (67) follows from the above inequality, (69), and the definitions of $\hat{R}_\theta$ in (28) and M_θ in (63), and the facts that $D_h/\operatorname{dist}_{\partial\mathcal{H}}(\bar{z}) \geq 1$ and $\sigma^+(A)/\|A\| \leq 1$. □

Now we are ready to prove statement (c) of Theorem 3.1.

Proof of Statement (c) of Theorem 3.1 From the relation $\hat{q}_k = (1-\theta)q_{k-1} + c(A\hat{z}_k - b)$ given in (19), we have $c\left(A\hat{z}_k - b\right) = \hat{q}_k - (1-\theta)q_{k-1}$, and then it follows from the triangle inequality for norms that $c\|A\hat{z}_k - b\| \leq \|\hat{q}_k\| + (1-\theta)\|q_{k-1}\|$, which immediately implies that (32) holds, in view of (67) and the fact that $\theta \in (0, 1)$. To conclude the proof, note that if $\hat{c}(\hat{\eta}) = \bar{c}(\hat{\eta})$, then the last statement of (c) follows from the last statement of (b); otherwise, the last statement of (c) follows from the first two relations in (30), inequality (32), the assumption on c in (33), and Definition 2.1. □

5 Proof of Statement (b) of Lemma 4.4

Before presenting the proof of statement (b) of Lemma 4.4, we state and prove a technical result which provides a preliminary bound on the left-hand side of (51).

Lemma 5.1 *Let $\{(z_k, q_k, \varepsilon_k)\}$ generated by the static θ-IPAAL and define*

$$s_k := \frac{(1-\theta)(2-\theta)}{c}\|\Delta q_k\|^2 + \frac{(1-\theta)^3}{\theta c}\left[\|\Delta q_k\|^2 - \|\Delta q_{k-1}\|^2\right], \tag{70}$$

where τ_θ is as in (14) and $\{\Delta q_k\}$ is as in (45). Then, for every $k \geq 2$, we have

$$\frac{\theta\lambda_\theta}{2(1-\theta)}s_k \leq \langle \Delta z_{k-1}, \Delta z_k\rangle - \frac{1}{2(1+\tau_\theta)}\|\Delta z_k\|^2 + \langle \Delta v_k, \Delta z_k\rangle$$
$$+(1+\tau_\theta^{-1})(\varepsilon_{k-1}+\varepsilon_k). \tag{71}$$

Proof Let Q denote the positive definite matrix $Q := I/2 + c\lambda_\theta A^*A$ and let ψ_k be as in (22).

Then, it follows from Proposition 3.1a) and the above definition of Q that $\psi_k(\cdot) - \|\cdot\|_Q^2/2$ is convex. Hence, it follows from the inclusion in (25) with $k = k-1$ and Lemma A.1 with $\psi(\cdot) = \psi_{k-1}(\cdot)$, $(\xi, \tau) = (1, \tau_\theta)$, $(y, v, \eta) = (z_{k-1}, v_{k-1}, \varepsilon_{k-1})$, and $u = z_k$, that

$$\langle v_{k-1}, z_k - z_{k-1}\rangle + \frac{1}{2(1+\tau_\theta)}\|z_k - z_{k-1}\|_Q^2 - (1+\tau_\theta^{-1})\varepsilon_{k-1}$$
$$\leq \psi_{k-1}(z_k) - \psi_{k-1}(z_{k-1}). \tag{72}$$

Similarly, it follows from the inclusion in (25) and Lemma A.1 with $\psi(\cdot) = \psi_k(\cdot)$, $(\xi, \tau) = (1, \tau_\theta)$, $(y, v, \eta) = (z_k, v_k, \varepsilon_k)$, and $u = z_{k-1}$, that

$$\langle v_k, z_{k-1} - z_k\rangle + \frac{1}{2(1+\tau_\theta)}\|z_{k-1} - z_k\|_Q^2 - (1+\tau_\theta^{-1})\varepsilon_k \leq \psi_k(z_{k-1}) - \psi_k(z_k). \tag{73}$$

Also, it is easy to see that the definition of $\psi_k(\cdot)$ in (22) implies that $\psi_k(\cdot) = \psi_{k-1}(\cdot) + a_{k-1}(\cdot)$ where $a_{k-1}(\cdot)$ is an affine function such that its constant gradient ∇a_{k-1} is given by

$$\nabla a_{k-1} = -\Delta z_{k-1} + \lambda_\theta(1-\theta)A^*\Delta q_{k-1}, \quad \forall k \geq 2. \tag{74}$$

Now, summing (72) and (73), and using the previous observation and the fact that $\Delta z_k = z_k - z_{k-1}$ and $\Delta v_k = v_k - v_{k-1}$, we obtain

$$-\langle \Delta v_k, \Delta z_k\rangle + \frac{1}{(1+\tau_\theta)}\|\Delta z_k\|_Q^2 - (1+\tau_\theta^{-1})(\varepsilon_{k-1}+\varepsilon_k) \leq a_{k-1}(z_{k-1}) - a_{k-1}(z_k)$$
$$= -\langle \nabla a_{k-1}, \Delta z_k\rangle = \langle \Delta z_{k-1}, \Delta z_k\rangle - \lambda_\theta(1-\theta)\langle \Delta q_{k-1}, A\Delta z_k\rangle.$$

Rearranging the above inequality and using the fact that $Q = I/2 + c\lambda_\theta A^*A$, we have

$$\langle \Delta v_k, \Delta z_k\rangle + \langle \Delta z_{k-1}, \Delta z_k\rangle - \frac{1}{2(1+\tau_\theta)}\|\Delta z_k\|^2 + (1+\tau_\theta^{-1})(\varepsilon_{k-1}+\varepsilon_k)$$
$$\geq \frac{\lambda_\theta c}{1+\tau_\theta}\|A\Delta z_k\|^2 + \lambda_\theta(1-\theta)\langle \Delta q_{k-1}, A\Delta z_k\rangle =: \tilde{s}_k. \tag{75}$$

Now, using the facts that $\tau_\theta < 1$ (see (21)) and $A\Delta z_k = [\Delta q_k - (1-\theta)\Delta q_{k-1}]/c$ in view of the first identity in (46), we have

$$\begin{aligned}\tilde{s}_k &\geq \frac{\lambda_\theta}{2c}\left[\|\Delta q_k - (1-\theta)\Delta q_{k-1}\|^2 + 2(1-\theta)\langle \Delta q_{k-1}, \Delta q_k - (1-\theta)\Delta q_{k-1}\rangle\right] \\ &= \frac{\lambda_\theta}{2c}\left\{\theta(2-\theta)\|\Delta q_k\|^2 + (1-\theta)^2\left[\|\Delta q_k\|^2 - \|\Delta q_{k-1}\|^2\right]\right\} = \frac{\theta\lambda_\theta s_k}{2(1-\theta)},\end{aligned}$$

where the last identity is due to the definition of s_k in (70). Hence, (71) follows immediately by combining (75) with the above inequality.

Now, we are ready to prove Lemma 4.4b).

Proof of Lemma 4.4b) To prove this result, we proceed to bound the right-hand side of (71) in terms of $\|r_k\|$ and $t_{k-1} - t_k$.

Indeed, using the fact that $\langle a_1, a_2\rangle \leq (\|a_1\|^2 + \|a_2\|^2)/2$, we have

$$\begin{aligned}\langle \Delta z_{k-1}, \Delta z_k\rangle - \frac{1}{2(1+\tau_\theta)}\|\Delta z_k\|^2 &\leq \frac{\|\Delta z_{k-1}\|^2}{2} + \frac{\|\Delta z_k\|^2}{2} - \frac{1}{2(1+\tau_\theta)}\|\Delta z_k\|^2 \\ &= \frac{1}{2}\left(\|\Delta z_{k-1}\|^2 - \|\Delta z_k\|^2\right) + \frac{\tau_\theta}{2(1+\tau_\theta)}\|\Delta z_k\|^2 \\ &\leq \frac{1}{2}\left(\|\Delta z_{k-1}\|^2 - \|\Delta z_k\|^2\right) + \frac{\tau_\theta(1+\tilde{\sigma})^2}{2(1+\tau_\theta)}\|r_k\|^2,\end{aligned} \tag{76}$$

where the last inequality is due to the second inequality in (46). Now, using the fact that $\Delta v_k = v_k - v_{k-1}$, the Cauchy–Schwarz inequality, the triangle inequality, and the inequality in (25), we have

$$\begin{aligned}\langle \Delta v_k, \Delta z_k\rangle &\leq (\|v_k\| + \|v_{k-1}\|)\|\Delta z_k\| \leq \tilde{\sigma}(\|r_k\| + \|r_{k-1}\|)\|\Delta z_k\| \\ &\leq \tilde{\sigma}(1+\tilde{\sigma})(\|r_k\| + \|r_{k-1}\|)\|r_k\| \leq \frac{\tilde{\sigma}(1+\tilde{\sigma})\|r_{k-1}\|^2}{2} + \frac{3\tilde{\sigma}(1+\tilde{\sigma})\|r_k\|^2}{2} \\ &= \frac{\tilde{\sigma}(1+\tilde{\sigma})}{2}\left[\|r_{k-1}\|^2 - \|r_k\|^2\right] + 2\tilde{\sigma}(1+\tilde{\sigma})\|r_k\|^2,\end{aligned} \tag{77}$$

where the last two inequalities are due to the second inequality in (46) and the fact that $\|r_{k-1}\|\|r_k\| \leq (\|r_k\|^2 + \|r_{k-1}\|^2)/2$, respectively. The inequality in (25) also yields

$$\varepsilon_{k-1} + \varepsilon_k \leq \frac{\tilde{\sigma}^2\|r_{k-1}\|^2}{2} + \frac{\tilde{\sigma}^2\|r_k\|^2}{2} = \tilde{\sigma}^2\|r_k\|^2 + \frac{\tilde{\sigma}^2}{2}\left[\|r_{k-1}\|^2 - \|r_k\|^2\right]. \tag{78}$$

Now, combining (71),(76), (77), and (78), and using the definitions of s_k, B_θ, and t_k given in (70), (49), and (50), respectively, we easily see that (51) holds. $\square$

6 Concluding Remarks

We start by making two remarks about the analysis of this paper.

First, the static θ-IPAAL method was stated using the perturbed Lagrangian (2) but it can also be stated in terms of the usual Lagrangian $\mathcal{L}_c(\cdot,\cdot) := \mathcal{L}_c^0(\cdot,\cdot)$, i.e., the special case of (2) with $\theta = 0$. Indeed, if we define the scaled multipliers $p_k = (1-\theta)q_k$ and $\hat{p}_k = (1-\theta)\hat{q}_k$ for every $k \geq 1$, then we have $\mathcal{L}_c^\theta(\cdot, q_k) = \mathcal{L}_c(\cdot, p_k)$ and the multiplier formulae $q_k = (1-\theta)q_{k-1} + c(Az_k - b)$ and $\hat{q}_k = (1-\theta)q_{k-1} + c(A\hat{z}_k - b)$ given in (18) and (19), respectively, reduce to $p_k = (1-\theta)[p_{k-1} + c(Az_k - b)]$ and $\hat{p}_k := (1-\theta)[p_{k-1} + c\left(A\hat{z}_k - b\right)]$, respectively. This observation and a close examination of the statement of the static θ-IPAAL show that this method can be stated solely in terms of the Lagrangian $\mathcal{L}_c(\cdot,\cdot)$ and the multiplier sequences $\{p_k\}$ and $\{\hat{p}_k\}$.

Second, our analysis assumes that dom$h$ is bounded and that all the constraints are linear since these assumptions greatly simplifies its analysis. The analysis for the case where dom$h$ is unbounded but all the constraints are still linear can be found in the first version of this paper (see [29]) and is considerably more involved than the current one. Although not pursued in this paper, the authors wonder whether the techniques developed in [18] and in the paper [20] (which was released after the current work) can be used to extend θ-IPAAL to the case where some of or all the constraints in (1) are smooth nonlinear convex functions.

We now discuss how θ-IPAAL compares with the S-prox-AL method of [39], which also sequentially solves prox subproblems but updates multipliers in a manner that differs from the one in this paper. First, it is shown in [39] that S-prox-ALM has an $\mathcal{O}(\hat{\rho}^{-2})$ iteration complexity under the assumption that the function h in (1) is the indicator of a box in $\mathbb{R}^n$. Second, S-prox-AL generates a sequence of proximal subproblems as in (4) but, instead of solving them by using an ACG-type subroutine, it applies a single composite gradient step to inexactly solve a variant[4] of (4). Third, while the θ-IPAAL method only requires choosing its parameters based on the scalars m_f and L_f to guarantee convergence, the S-prox-ALM requires choosing its parameters based on the supremum of a set of Hoffman constants (see the proof of [39, Lemma 3.10] and [39, Lemma 4.8]) that is generally difficult to compute and even to adaptively estimate. It is worth noting that, after the first release of this paper, the analysis of [39] was extended to the case where h in (1) is the indicator function of a polyhedron by [40], for which all remarks above still apply.

We end this section by discussing some possible extensions of our paper. First, it would be interesting to study an adaptive version of θ-IPAAL where the prox stepizes vary from iteration to iteration and are chosen in an adaptive manner. Second, a drawback of θ-IPAAL is that the following issues arise as θ approaches zero: (1) the complexity bounds in (39) diverge to infinity which make our analysis invalid for the case where $\theta = 0$; and (2) potential deterioration of its computational performance due to the fact that the prox stepsize λ_θ defined in (14) converges to zero. Hence, it would be interesting to develop a prox AL method which does not depend on θ and whose Lagrange multiplier update formula is (18) with $\theta = 0$.

[4] Instead of inexactly minimizing the function $\lambda\mathcal{L}(\cdot; p_{k-1}) + \|\cdot - z_{k-1}\|^2/2$, the S-prox-AL exactly minimizes the linear approximation of the function $\lambda\mathcal{L}(\cdot; p_{k-1}) + \|z - \tilde{z}_{k-1}\|/2$ for a point $\tilde{z}_{k-1}$ different from z_{k-1}.

Acknowledgements The first author was partially supported by CNPq Grants 312559/2019-4 and 312223/2022-6. The second author was partially supported by ONR Grant N00014-18-1-2077 and AFOSR Grant FA9550-22-1-0088. The authors would like to thank the two anonymous referees for their insightful comments on previous drafts of this paper.

Data availability Due to the theoretical nature of this research, no additional data is necessary.

Appendices

A Three Technical Results

This section contains three technical results about the ε-subdifferential of a convex function perturbed by a prox term.

The following result is used in the proof of Lemma 4.4b).

Lemma A.1 *Assume that* $\xi > 0$, $\psi \in \overline{\text{Conv}}\,\mathbb{R}^n$ *and* $Q \in \mathcal{S}^n_{++}$ *are such that* $\psi - (\xi/2)\|\cdot\|^2_Q$ *is convex and let* $(y, v, \eta) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ *be such that* $v \in \partial_\eta \psi(y)$. *Then, for any* $\tau > 0$,

$$\psi(u) \geq \psi(y) + \langle v, u - y\rangle - (1 + \tau^{-1})\eta + \frac{(1+\tau)^{-1}\xi}{2}\|u - y\|^2_Q \quad \forall u \in \mathbb{R}^n. \tag{79}$$

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

$$\psi_v(u) \geq \psi_v(\bar{y}) + \frac{\xi}{2}\|u - \bar{y}\|^2_Q \geq \psi_v(y) - \eta + \frac{\xi}{2}\|u - \bar{y}\|^2_Q \tag{80}$$

for every $u \in \mathbb{R}^n$. The above inequalities with $u = y$ imply that $(\xi/2)\|\bar{y} - y\|^2_Q \leq \eta$. Hence, adding and subtracting the term $(\tau^{-1}\xi/2)\|\bar{y} - y\|^2_Q$ in the right-hand side of (80) and using the inequality $\tau^{-1}\|\tilde{u}\|^2 + \|u'\|^2 \geq (1+\tau)^{-1}\|\tilde{u} + u'\|^2$ with $\tilde{u} = u - \bar{y}$ and $u' = \bar{y} - y$, we obtain

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

for every $u \in \mathbb{R}^n$. Hence, (79) follows from the above conclusion and the definition of ψ_v. □

The next result is used to prove Lemma 4.1. Its proof easily follows from [11, Lemma 32].

Lemma A.2 *Assume that* $\tilde{h} \in \overline{\text{Conv}}\,\mathbb{R}^n$, $\tilde{g}$ *is a differentiable convex function on* $\text{dom}\tilde{h}$, *and* $(z, \tilde{L}) \in \text{dom}\tilde{h} \times \mathbb{R}_+$ *is such that* $\tilde{g}(u) - \ell_{\tilde{g}}(u; z) \leq \tilde{L}\|u - z\|^2/2$ *for every* $u \in \text{dom}\tilde{h}$, *and define*

$$\tilde{z} := \text{argmin}_u \left\{ \ell_{\tilde{g}}(u; z) + \tilde{h}(u) + \frac{\tilde{L}}{2}\|u - z\|^2 \right\}, \qquad \widetilde{w} := \tilde{L}(z - \tilde{z}). \tag{81}$$

If $0 \in \partial_\varepsilon(\tilde{g} + \tilde{h})(z)$ *for some* $\varepsilon \geq 0$, *then* $\widetilde{w} \in \nabla\tilde{g}(z) + \partial\tilde{h}(\tilde{z})$ *and* $\|\widetilde{w}\| \leq \sqrt{2\tilde{L}\varepsilon}$.

The following result is used in the proof of Lemma 4.7.

Lemma A.3 *Assume that* $\sigma_0 > 0$, $(z_0, z_1) \in \mathbb{R}^n \times \text{dom}\tilde{\phi}$, $(v_1, \varepsilon_1) \in \mathbb{R}^n \times \mathbb{R}_+$, *and a proper function* $\tilde{\phi} : \mathbb{R}^n \to (-\infty, \infty]$ *such that* $\tilde{\phi} + \|\cdot\|^2/2$ *is convex, satisfy*

$$v_1 \in \partial_{\varepsilon_1}\left(\tilde{\phi} + \frac{1}{2}\|\cdot - z_0\|^2\right)(z_1), \quad \|v_1\|^2 + 2\varepsilon_1 \leq \widetilde{\sigma}^2\|v_1 + z_0 - z_1\|^2. \tag{82}$$

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

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

Proof Using the inclusion in (82), the definition of ε-subdifferential in (5), and the fact that $|\langle u, \tilde{u}\rangle| \leq [s\|u\|^2 + s^{-1}\|\tilde{u}\|^2]/2$ for every $u, \tilde{u} \in \mathbb{R}^n$ and $s > 0$, we conclude that for every $z \in \mathbb{R}^n$,

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

which immediately implies the conclusion of the lemma in view of the inequality in (82). □

References

1. Aybat, N.S., Iyengar, G.: A first-order smoothed penalty method for compressed sensing. SIAM J. Optim. **21**(1), 287–313 (2011)

2. Aybat, N.S., Iyengar, G.: A first-order augmented Lagrangian method for compressed sensing. SIAM J. Optim. **22**(2), 429–459 (2012)
3. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)
4. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York (1982)
5. Boob, D., Deng, Q., Lan, G.: Stochastic first-order methods for convex and nonconvex functional constrained optimization. Math. Program. **197**, 215–279 (2023)
6. Chen, Y., Lan, G., Ouyang, Y.: Optimal primal-dual methods for a class of saddle point problems. SIAM J. Optim. **24**(4), 1779–1814 (2014)
7. Florea, M.I., Vorobyov, S.A.: An accelerated composite gradient method for large-scale composite objective problems. IEEE Trans. Signal Process. **67**(2), 444–459 (2018)
8. Ghadimi, S., Lan, G.: Accelerated gradient methods for nonconvex nonlinear and stochastic programming. Math. Program. **156**, 59–99 (2016)
9. Gonçalves, M.L.N., Melo, J.G., Monteiro, R.D.C.: 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)
10. Hajinezhad, D., Hong, M.: Perturbed proximal primal-dual algorithm for nonconvex nonsmooth optimization. Math. Program. **176**, 207–245 (2019)
11. He, Y., Monteiro, R.D.C.: 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)
12. He, Y., Monteiro, R.D.C.: An accelerated HPE-type algorithm for a class of composite convex-concave saddle-point problems. SIAM J. Optim. **26**(1), 29–56 (2016)
13. Hong, M.: Decomposing linearlyconstrained nonconvexproblems by a proximal primaldual approach: Algorithms,convergence, and applications (2016). Available on arXiv:1604.00543
14. Jiang, B., Lin, T., Ma, S., Zhang, S.: Structured nonconvex and nonsmooth optimization algorithms and iteration complexity analysis. Comput. Optim. Appl. **72**(3), 115–157 (2019)
15. Kolossoski, O., Monteiro, R.D.C.: An accelerated non-Euclidean hybrid proximal extragradient-type algorithm for convex-concave saddle-point problems. Optim. Methods Softw. **32**(6), 1244–1272 (2017)
16. Kong, W., Melo, J.G., Monteiro, R.D.C.: Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs. SIAM J. Optim. **29**(4), 2566–2593 (2019)
17. Kong, W., Melo, J.G., Monteiro, R.D.C.: An efficient adaptive accelerated inexact proximal point method for solving linearly constrained nonconvex composite problems. Comput. Optim. Appl. **76**(2), 305–346 (2019)
18. Kong, W., Melo, J.G., Monteiro, R.D.C.: Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical Lagrangian function. SIAM J. Optim. **33**(1), 181–210 (2023)
19. Kong, W., Melo, J.G., Monteiro, R.D.C.: FISTA and extensions–review and new insights. (2021) Available on arXiv:2107.01267
20. Kong, W., Melo, J.G., Monteiro, R.D.C.: Iteration complexity of a proximal augmented Lagrangian method for solving nonconvex composite optimization problems with nonlinear convex constraints. Math. Oper. Res. (2022). https://doi.org/10.1287/moor.2022.1301
21. Kong, W., Monteiro, R.D.C.: An accelerated inexact proximal point method for solving nonconvex-concave min-max problems. SIAM J. Optim. **31**(4), 2558–2585 (2021)
22. Lan, G., Monteiro, R.D.C.: Iteration-complexity of first-order penalty methods for convex programming. Math. Program. **138**(1), 115–139 (2013)
23. Lan, G., Monteiro, R.D.C.: Iteration-complexity of first-order augmented Lagrangian methods for convex programming. Math. Program. **155**(1), 511–547 (2016)
24. Li, Z., Chen, P., Liu, S., Lu, S., Xu, Y.: Rate-improved inexact augmented Lagrangian method for constrained nonconvex optimization. In: International Conference on Artificial Intelligence and Statistics, pp. 2170–2178 (2021) (Posted in arXiv:2007.01284 on July, 2)
25. Li, Z., Xu, Y.: Augmented Lagrangian based first-order methods for convex-constrained programs with weakly-convex objective. INFORMS J. Optim. **3**(4), 373–397 (2021)
26. Lin, Q., Ma, S., Xu, Y.: Complexity of an inexact proximal-point penalty method for constrained smooth non-convex optimization. Comput. Optim. Appl. **82**, 175–224 (2022)

27. Liu, Y.F., Liu, X., Ma, S.: On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. Math. Oper. Res. **44**(2), 632–650 (2019)
28. Lu, Z., Zhou, Z.: Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. (2018) Available on arXiv:1803.09941
29. Melo, J.G., Monteiro, R.D.C, Wang, H.: Iteration-complexity of an inexact proximal accelerated augmented Lagrangian method for solving linearly constrained smooth nonconvex composite optimization problems. (2020) Posted on optimization online (see https://optimization-online.org/2020/04/7772/) in April 29, 2020 and on Arxiv (see arXiv:2006.08048) in June 14, 2020
30. Monteiro, R.D.C., Ortiz, C., Svaiter, B.F.: An adaptive accelerated first-order method for convex optimization. Comput. Optim. Appl. **64**, 31–73 (2016)
31. Monteiro, R.D.C., Svaiter, B.F.: 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. Necoara, I., Patrascu, A., Glineur, F.: Complexity of first-order inexact Lagrangian and penalty methods for conic convex programming. Optim. Methods Softw. **34**(2), 305–335 (2017)
33. Nesterov, Y.E.: Introductory Lectures on Convex Optimization: A Basic Course. Kluwer Academic Publ. Springer, Boston (2004)
34. Patrascu, A., Necoara, I., Tran-Dinh, Q.: Adaptive inexact fast augmented Lagrangian methods for constrained convex optimization. Optim. Lett. **11**(3), 609–626 (2017)
35. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976)
36. Sahin, M., Eftekhari, A., Alacaoglu, A., Latorre, F., Cevher, V.: An inexact augmented Lagrangian framework for nonconvex optimization with nonlinear constraints. In: Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F.D., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32. Curran Associates Inc, Red Hook (2019)
37. Xie, Y., Wright, S.J.: Complexity of proximal augmented Lagrangian for nonconvex optimization with nonlinear equality constraints. J. Sci. Comput. **86** (38) (2021). https://doi.org/10.1007/s10915-021-01409-y
38. Xu, Y.: Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. Math. Program. **185**, 199–244 (2021)
39. Zhang, J., Luo, Z.Q.: A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization. SIAM J. Optim. **30**(3), 2272–2302 (2020)
40. Zhang, J., Luo, Z.-Q.: A global dual error bound and its application to the analysis of linearly constrained nonconvex optimization. SIAM J. Optim. **32**(3), 2319–2346 (2022) (Posted in arXiv:2006.16440 on June 30, 2020)