GLOBAL COMPLEXITY BOUND OF A PROXIMAL ADMM FOR LINEARLY-CONSTRAINED NONSEPARABLE NONCONVEX COMPOSITE PROGRAMMING*

WEIWEI KONG† AND RENATO D.C. MONTEIRO‡

Abstract. This paper proposes and analyzes a dampened proximal alternating direction method of multipliers (DP.ADMM) for solving linearly-constrained nonconvex optimization problems where the smooth part of the objective function is nonseparable. Each iteration of DP.ADMM consists of: (ii) a sequence of partial proximal augmented Lagrangian (AL) updates, (ii) an under-relaxed Lagrange multiplier update, and (iii) a novel test to check whether the penalty parameter of the AL function should be updated. Under a basic Slater point condition and some requirements on the dampening factor and under-relaxation parameter, it is shown that DP.ADMM obtains a first-order stationary point of the constrained problem in $\mathcal{O}(\varepsilon^{-3})$ iterations for a given numerical tolerance $\varepsilon > 0$. One of the main novelties of the paper is that convergence of the method is obtained without requiring any rank assumptions on the constraint matrices.

Key words. proximal ADMM, nonseparable, nonconvex composite optimization, iteration complexity, under-relaxed update, augmented Lagrangian function

AMS subject classifications. 65K10, 90C25, 90C26, 90C30, 90C60

1. Introduction.

Consider the following composite optimization problem:

$$\min_{x\in\mathbb{R}^n} \left\{\phi(x) := f(x) + h(x) : Ax = d\right\}, \tag{1.1}$$

where h is a closed convex function, f is a (possibly) nonconvex differentiable function on the domain of h, the gradient of f is Lipschitz continuous, A is a linear operator, d is a vector in the image of A (denoted as $\mathrm{Im}(A)$), and the following B-block structure is assumed:

$$\begin{gathered} n = n_1 + \ldots + n_B, \quad x = (x_1, \ldots, x_B) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_B} \\ h(x) = \sum_{t=1}^{B} h_t(x_t), \quad Ax = \sum_{t=1}^{B} A_t x_t, \end{gathered} \tag{1.2}$$

where $\{A_t\}_{t=1}^B$ is another set of linear operators and $\{h_t\}_{t=1}^B$ is another set of proper closed convex functions with compact domains.

Due to the block structure in (1.2), a popular algorithm for obtaining stationary solutions of (1.1) is the proximal alternating direction method of multipliers (ADMM) wherein a sequence of smaller augmented Lagrangian type subproblems is solved over $x_1, ..., x_B$ sequentially or in parallel. However, the main drawbacks of existing ADMM-type methods include: (i) strong assumptions about the structure of h; (ii) iteration

*Funding: The first author has been supported by (i) the US Department of Energy (DOE) and UT-Battelle, LLC, under contract DE-AC05-00OR22725, (ii) the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration, and (iii) the IDEaS-TRIAD Fellowship (NSF Grant CCF-1740776). The second author was partially supported by ONR Grant N00014-18-1-2077 and AFOSR Grant FA9550-22-1-0088.

Versions: v0.1 (October 24, 2021), v0.2 (December 9, 2021), v1.0 (June 14, 2022)

†Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37830. wwkong92@gmail.com

‡School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205. monteiro@isye.gatech.edu

complexity bounds that scale poorly with the numerical tolerance; (iii) small stepsize parameters; or (iv) strong rank assumptions about the last block A_B, such as $\mathrm{Im}(A_B) \supseteq \{d\} \cup \mathrm{Im}(A_1) \cup \ldots \mathrm{Im}(A_{B-1})$. Of the above drawbacks, the *last block condition* in (iv) is especially limiting. For example, consider the popular multiblock distributed finite-sum problem

$$\min_{(x_1,\ldots,x_B)\in\mathbb{R}^n\times\cdots\times\mathbb{R}^n} \left\{ \sum_{t=1}^{B} (f_t + h_t)(x_t) : x_t - x_B = 0, \quad t = 1,\ldots,B-1 \right\} \tag{1.3}$$

where f_i is continuously differentiable, h_t is closed convex, and ∇f_t is Lipschitz continuous for $t = 1, ..., B$. It is easy to see[1] that (1.3) is a special case of (1.1) where $n_t = n$ for $t = 1, \ldots, B$, we have $A_s = e_s \otimes I \in \mathbb{R}^{n(B-1)\times n}$ for $s = 1, \ldots, B-1$, we have $A_B = -\mathbf{1} \otimes I \in \mathbb{R}^{n(B-1)\times n}$, and $d = 0$. Moreover, it is straightforward to show that for $s = 1, \ldots, B-1$ we have $\mathrm{Im}(A_s) \cap \mathrm{Im}(A_B) = 0$ but $\mathrm{Im}(A_s)\backslash\{0\} \neq \emptyset$, which implies that $\mathrm{Im}(A_s) \not\subseteq \mathrm{Im}(A_B)$.

Our goal in this paper is to develop and analyze the complexity of a proximal ADMM that removes all the drawbacks mentioned above. For a given $\theta \in (0, 1)$, its k^{th} iteration is based on the *dampened* augmented Lagrangian (AL) function given by

$$\mathcal{L}^{\theta}_{c_k}(x; p) := \phi(x) + (1-\theta)\langle p, Ax - d\rangle + \frac{c_k}{2}\|Ax - d\|^2, \tag{1.4}$$

where $c_k > 0$ is the *penalty parameter*. Specifically, it consists of the following updates: given $x^{k-1} = (x_1^{k-1}, \ldots, x_B^{k-1})$, p^{k-1} c_k, χ, and λ, sequentially ($t = 1, \ldots, B$) compute the t^{th} block of x^k as

$$x_t^k = \underset{u_t\in\mathbb{R}^{n_t}}{\operatorname{argmin}} \left\{ \lambda\mathcal{L}^{\theta}_{c_k}(\ldots, x_{t-1}^k, u_t, x_{t+1}^{k-1}, \ldots; p^{k-1}) + \frac{1}{2}\|u_t - x_t^{k-1}\|^2 \right\}, \tag{1.5}$$

and then update

$$p^k = (1-\theta)p^{k-1} + \chi c_k \left(Ax^k - d\right), \tag{1.6}$$

where $\chi \in (0, 1)$ is a suitably chosen under-relaxation parameter.

Contributions. For proper choices of the stepsize λ and a non-decreasing sequence of penalty parameters $\{c_k\}_{k\geq1}$, it is shown that if the Slater-like condition[2]

$$\exists \mathring{z} \in \mathrm{int}\,(\mathrm{dom}\, h) \text{ such that } A\mathring{z} = d, \tag{1.7}$$

holds, then DP.ADMM has the following features:

- ▷ for any tolerance pair $(\rho, \eta) \in \mathbb{R}^2_{++}$, it obtains a pair $(\bar{z}, \bar{q})$ satisfying

  $$\mathrm{dist}\,(0, \nabla f(\bar{z}) + A^*\bar{q} + \partial h(\bar{z})) \leq \rho, \quad \|A\bar{z} - d\| \leq \eta \tag{1.8}$$

  in $\mathcal{O}(\max\{\rho^{-3}, \eta^{-3}\})$ iterations;
- ▷ it introduces a novel approach for updating the penalty parameter c_k, instead of assuming that $c_k = c_1$ for every $k \geq 1$ and that c_1 is sufficiently large (such as in [3, 13, 14, 26, 28, 29] in Table 1.2);

[1] Here, $e_1, \ldots, e_n$ is the standard basis for $\mathbb{R}^{B-1}$, I_n is the n-by-n identity matrix, $\mathbf{1} \in \mathbb{R}^{B-1}$ is a vector of ones, and $\otimes$ is the Kronecker product of two matrices.

[2] Here, $\mathrm{int}\, S$ denotes the interior of a set S, $\mathrm{dom}\,\psi$ denotes the domain of a function ψ, and A^* is the adjoint of linear operator A.

- ▷ it does not have any of the drawbacks mentioned in the sentences preceding equation (1.3).

Related Works. Since ADMM-type methods where f is convex have been well-studied in the literature (see, for example, [1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 23, 24, 25]), we make no further mention of them here. Instead, we discuss ADMM-type methods where f is nonconvex.

Letting δ_S denote the indicator function of a convex set S (see Subsection 1.1), Table 1.1 presents a list of common assumptions found in the literature. Table 1.2

$\mathcal{R}_0$	$\mathrm{Im}(A_B) \supseteq \{d\} \cup \mathrm{Im}(A_1) \cup \ldots \cup \mathrm{Im}(A_{B-1})$.
$\mathcal{R}_1$	A_B has full column rank or, equivalently, the rows of A_B are linearly independent.
$\mathcal{S}$	The Slater-like assumption (1.7) holds.
$\mathcal{KL}$	The classic AL function, i.e. (1.4) with $\theta = 0$, has the KL property.[3]
$\mathcal{P}$	$h_i \equiv \delta_P$ for $i \in \{1, \ldots, B\}$, where P is a polyhedral set.
$\mathcal{F}$	A point $x^0 \in \mathrm{dom}\, h$ satisfying $Ax^0 = d$ is available as an input.

Table 1.1

Common nonconvex ADMM assumptions and regularity conditions. It is well-known that condition $\mathcal{R}_1$ implies condition $\mathcal{R}_0$.

presents a comparison between our proposed DP.ADMM and other ADMM-type methods for nonconvex and nonseparable problems, under a common tolerance ε given by $\varepsilon := \min\{\rho, \eta\}$.

Algorithm	θ	χ	Complexity	Assumptions	Adaptive c
ADMM [28]	0	1	None	$\mathcal{R}_0, \mathcal{KL}$	No
LPADMM [29]	0	$(0, \infty)$	None	$\mathcal{P}, \mathcal{S}$	No
PADMM-m [14]	0	1	$\mathcal{O}(\varepsilon^{-6})$	$\mathcal{F}$	No
SDD-ADMM [26]	$(0, 1]$	$[-\frac{\theta}{4}, 0)$	$\mathcal{O}(\varepsilon^{-4})$	$\mathcal{F}$	No
DP.ADMM	$(0, 1]$	$(0, \pi_\theta]$	$\mathcal{O}(\varepsilon^{-3})$	$\mathcal{S}$	Yes

Table 1.2

Comparison of existing ADMM-type methods with DP.ADMM for finding ε-stationary points with $\varepsilon := \min\{\rho, \eta\}$ and $\pi_\theta = \theta^2/[2B(2-\theta)(1-\theta)]$ if $\theta \in (0, 1)$ and $\pi_\theta = 1$ if $\theta = 1$. The algorithms in [26, 28] are non-proximal ADMMs, and the last column indicates whether the method has a way to adaptively choose the penalty parameter c to ensure convergence.

We now make five remarks about the results in papers [14, 26] compared to the ones in this paper (which were developed independently of [26]). First, both of the complexity bounds in [14, 26] require that a feasible point be readily available, while the initial point for DP.ADMM can be any point in $\mathrm{dom}\, h$. Second, the $\mathcal{O}(\varepsilon^{-6})$ complexity bound established in [14] is for an ADMM-type method applied to a penalized *reformulation* of (1.1), while DP.ADMM is applied to (1.1) directly. Third, the method in [26] considers a small stepsize (proportional to η^2) *linearized* proximal gradient update while DP.ADMM considers a large stepsize (proportional to the inverse of the weak-convexity constant of f) proximal point update as in (1.5). Fourth, paper [26] establishes an improved $\mathcal{O}(\varepsilon^{-3})$ complexity bound for SDD-ADMM only under the additional strong assumption that $\mathcal{R}_1$ in Table 1.1 holds and $\partial h(x)$ is compact for every x in the sublevel set of ϕ. Finally, it is worth emphasizing that among

[3] See [3, 13] for a definition.

the papers that establish an iteration complexity for ADMM, paper [26] and this one are the only ones that do not assume condition $\mathcal{R}_0$ or $\mathcal{R}_1$. Moreover, between these two papers, **only this examines the case of** $\chi > 0$.

To close, we discuss some related ADMM papers which assume the objective function ϕ in (1.1) is separable and has the same block structure as in (1.2), i.e., $\phi(x) = \sum_{t=1}^{B}(f_t + h_t)(x_t)$ for closed (possibly) convex functions $h_t : \mathbb{R}^n \mapsto (-\infty, \infty]$ and continuously differentiable functions $f_t : \operatorname{dom} h_t \mapsto \mathbb{R}$. All of their results restrictively assume that condition $\mathcal{R}_0$ or $\mathcal{R}_1$ in Table 1.1 holds and, as a consequence, some of them obtain an $\mathcal{O}(\varepsilon^{-2})$ iteration complexity[4]. Papers [3, 12, 27] present proximal ADMMs under the assumption that $B = 2$, $f_1 \equiv 0$, and $h_2 \equiv 0$. Papers [19,20] present linearized ADMMs that tackle a multi-block ($B \geq 2$) case of the above problem, in which $h_B \equiv 0$, and $f_1 \equiv \cdots \equiv f_{B-1} \equiv 0$. Finally, paper [13] presents a proximal ADMM for tackling the multiblock ($B \geq 2$) case of this problem in which assumption $\mathcal{KL}$ in Table 1.1 holds, $f_1 \equiv 0$, and $h_2 \equiv \cdots \equiv h_B \equiv 0$.

Organization. Subsection 1.1 presents some basic definitions and notation. Section 2 presents the proposed DP.ADMM in two subsections. The first one precisely describes the problem of interest, while the second one states the DP.ADMM and its iteration complexity. Section 3 presents the main properties of the DP.ADMM. Section 4 gives the proof of two important results, namely, Propositions 2.1 and 2.2. Section 5 gives some concluding remarks. Finally, the end of the paper contains several appendices.

1.1. Notation and Basic Definitions. Let $\mathbb{R}_+$ denote the set of nonnegative real numbers, and let $\mathbb{R}_{++}$ denote the set of positive real numbers. Let $\mathbb{R}_n$ denote the n-dimensional Hilbert space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. The direct sum (or Cartesian product) of a set of sets $\{S_i\}_{i=1}^n$ is denoted by $\prod_{i=1}^n S_i$.

The smallest positive singular value of a nonzero linear operator $Q : \mathbb{R}^n \to \mathbb{R}^l$ is denoted by σ_Q^+. For a given closed convex set $X \subset \mathbb{R}^n$, its boundary is denoted by ∂X and the distance of a point $x \in \mathbb{R}^n$ to X is denoted by $\operatorname{dist}_X(x)$. The indicator function of X at a point $x \in \mathbb{R}^n$ is denoted by $\delta_X(x)$ which has value 0 if $x \in X$ and $+\infty$ otherwise. For every $z > 0$ and positive integer b, we denote $\log_b^+(z) := \max\{1, \lceil \log_b(z) \rceil\}$.

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

$$\partial_\varepsilon h(z) := \{u \in \mathbb{R}^n : h(z') \geq h(z) + \langle u, z' - z \rangle - \varepsilon, \quad \forall z' \in \mathbb{R}^n\} \tag{1.9}$$

for every $z \in \mathbb{R}^n$. The classic subdifferential, denoted by $\partial h(\cdot)$, corresponds to $\partial_0 h(\cdot)$. The normal cone of a closed convex set C at $z \in C$, denoted by $N_C(z)$, is defined as

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

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

$$\ell_\psi(z; \bar{z}) := \psi(\bar{z}) + \langle \nabla \psi(\bar{z}), z - \bar{z} \rangle \quad \forall z \in \mathbb{R}^n. \tag{1.10}$$

[4] This complexity is also established in [14] for the non-separable setting of (1.1) under the assumption that $\mathcal{R}_1$ holds and $h_B \equiv 0$.

If $z=(x,y)$ then $f(x,y)$ is equivalent to $f(z)=f((x,y))$.

Iterates of a scalar quantity have their iteration number appear as a subscript, e.g., c_ℓ, while non-scalar quantities have this number appear as a superscript, e.g., v^k, and $\hat{p}^\ell$. For variables with multiple blocks, the block number appears as a subscript, e.g., x_t^k and v_t^k.

2. Alternating Direction Method of Multipliers. This section contains two subsections. The first one precisely describes the problem of interest and its underlying assumptions, while the second one presents the DP.ADMM and its corresponding iteration complexity.

Throughout this section, and subsequent ones, we let $\{\mathcal{H}_t\}_{t=1}^B \subseteq \mathbb{R}^{n_t}$ be compact convex sets and denote the aggregated quantities

$$\mathcal{H} := \prod_{t=1}^B \mathcal{H}_t, \quad x_{<t} := (x_1,\ldots,x_{t-1}), \tag{2.1}$$
$$x_{>t} := (x_{t+1},\ldots,x_B), \quad x_{\le t} := (x_{<t}, x_t), \quad x_{\ge t} := (x_t, x_{>t}),$$

for every $x=(x_1,\ldots,x_B)\in\mathcal{H}$.

2.1. Problem of Interest. This subsection presents the problem of interest and the assumptions underlying it.

Our problem of interest is finding approximate stationary points of (1.1) under the following assumptions on $(\phi, h_1,\ldots,h_B)$ and (A,d):

(A1) $h_t \in \overline{\text{Conv}}\ \mathcal{H}_t$ for every $1\le t\le B$;

(A2) $A\not\equiv 0$ and $\mathcal{F} := \{x\in\mathcal{H} : Ax = d\} \neq \emptyset$.

as well as the following assumptions on (f,h):

(A3) h is K_h-Lipschitz continuous on $\mathcal{H}$ for some $K_h\ge 0$;

(A4) f is continuously differentiable on $\mathcal{H}$ and, for every $1\le t\le B$, there exists $(m_t, M_t)\in\mathbb{R}^2_{++}$ such that

$$\|\nabla_{x_t} f(x_{\le t}, \tilde{x}_{>t}) - \nabla_{x_t} f(x)\| \le M_t \|\tilde{x}_{>t} - x_{>t}\|, \tag{2.2}$$
$$-\frac{m_t}{2}\|\tilde{x}_t - x_t\|^2 \le f(x_{<t}, \tilde{x}_t, x_{>t}) - f(x) - \langle \nabla_{x_t} f(x), \tilde{x}_t - x_t\rangle, \tag{2.3}$$

for every $x,\tilde{x}\in\mathcal{H}$;

(A5) there exists $\mathring{z}\in\mathcal{F}$ such that $d_\circ := \text{dist}_{\partial\mathcal{H}}(\mathring{z}) > 0$.

We now give a few remarks about the above assumptions. First, it is well known that (2.2) implies (2.3) with $m_t = M_{t-1}$. However, we show that better iterations complexities can be derived when scalars $\{m_t\}_{t=1}^B$ satisfying $m_t < M_{t-1}$ are available. Second, condition (2.3) implies that $f(x_{<t},\cdot,x_{>t}) + m_t\|\cdot\|^2/2$ is convex on x_t for any $x\in\mathcal{H}$. Third, since $\mathcal{H}$ is compact by (A1), the image of any continuous $\mathbb{R}^n$-valued function is bounded. In particular, this implies that the following scalars are bounded:

$$D_x := \sup_{x,x'\in\mathcal{H}} \|x-x'\|, \quad G_f := \sup_{x\in\mathcal{H}} \|\nabla f(x)\|, \quad \phi_* := \inf_{x\in\mathcal{H}}, \quad \overline{\phi} := \sup_{x\in\mathcal{H}} \phi(x). \tag{2.4}$$

We now briefly discuss the notion of an approximate stationary point of (1.1) in (1.8). It is well-known that the first-order necessary condition for a point $\bar{z}\in\text{dom}\, h$ to be a local minimum of (1.1) is that there exists $\bar{q}\in$ such that

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

Hence, the requirements in (1.8) can be viewed as a direct relaxation of the above conditions. For ease of future reference, we explicitly label the problem of obtaining (1.8) below.

Problem $\mathcal{LCCO}$: Given $(\rho, \eta) \in \mathbb{R}^2_{++}$, find a pair $(\bar{z}, \bar{q})$ satisfying (1.8).

It is worth mentioning that $(\bar{z}, \bar{q})$ is a solution of Problem $\mathcal{LCCO}$ if and only if there exists a residual $\bar{v} \in \mathbb{R}^n$ such that

$$\bar{v} \in \nabla f(\bar{z}) + A^*\bar{q} + \partial h(\bar{z}), \quad \|\bar{v}\| \le \rho, \quad \|A\bar{z} - d\| \le \eta, \tag{2.5}$$

and that this type of condition has been previously considered in the authors' previous works [15,16,17,18,22]. In the next subsection, we present a method (Algorithm 2.1) that computes such a residual in order to verify whether an incumbent solution $(\bar{z}, \bar{q})$ solves Problem $\mathcal{LCCO}$.

2.2. DP.ADMM. We present DP.ADMM in two parts. The first part presents a static version of DP.ADMM which either (i) stops with a solution of Problem $\mathcal{LCCO}$ or (ii) signals that its penalty parameter is too small. The second part presents the (dynamic) DP.ADMM that repeatedly invokes the static version on an increasing sequence of penalty parameters.

Both versions of DP.ADMM make use of the following condition on (χ, θ):

$$2\chi B(2-\theta)(1-\theta) \le \theta^2, \quad (\chi, \theta) \in (0, 1]^2. \tag{2.6}$$

For ease of reference and discussion, the pseudocode for the static DP.ADMM is given in Algorithm 2.1 below. In the special case of $(\theta, \chi) = (0, 1)$, its Steps 1 and 3 reduce to the classic proximal ADMM iteration

$$\begin{aligned} x_t^k &= \operatorname*{argmin}_{u^t \in \mathbb{R}^{n_t}} \left\{ \lambda \mathcal{L}_c^0(x_{<t}^k, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2}\|u_t - x_t^{k-1}\|^2 \right\}, \\ p^k &= p^{k-1} + c\left(Ax^k - d\right), \end{aligned}$$

for $1 \le t \le B$ and a fixed penalty parameter $c \ge 1$. Consequently, the novelty of the method lies in the careful choice of (θ, χ) and the special termination condition in its Step 2b.

The next result presents some technical properties of Algorithm 2.1. Its proof is given in Section 4, and it makes use of the following scalars:

$$\begin{aligned} M &:= \max_{1 \le t \le B} M_t, \quad m := \min_{1 \le t \le B} m_t, \quad \mathcal{N}_A := 8B^2 \sum_{t=1}^{B} \|A_t\|^2, \quad \Delta_\phi := \overline{\phi} - \phi_*, \\ \kappa_0 &:= \frac{2B^2(M + 2m)}{\sqrt{3m}}, \quad \kappa_1 := \left(K_h + G_f + B^2\left[M + 2m\right]D_x\right)D_x, \\ \kappa_2 &:= (\chi + \theta - \chi\theta)d_\circ\sigma_A^+, \quad \kappa_3 := \frac{\chi}{\theta}\sup_{x \in \mathcal{H}} \|Ax - d\|, \\ \kappa_4 &:= (1-\theta) + (1-\theta)(1-\chi)d_\circ\sigma_A^+, \quad \kappa_5 := \frac{12}{\chi}\left(1 + \frac{2\chi\kappa_1}{\kappa_2}\right), \end{aligned} \tag{2.7}$$

where $(G_f, D_x, \overline{\phi}, \phi_*)$, K_h, and (m_t, M_t) are as in (2.4), (A3), and (A4), respectively.

Algorithm 2.1 Static DP.ADMM

`Input`: $x^0 \in \mathcal{H}$, $p^0 \in A(\mathbb{R}^n)$, $c > 0$
`Require`: $\{m_t\} \subseteq \mathbb{R}_{++}$, $(\rho, \eta) \in (0,1]^2$, (χ, θ) as in (2.6)

$\lambda \leftarrow 1/(2\min_t m_t)$
for $k \leftarrow 1, 2, \ldots$ **do**
  `STEP 1` (prox update):
  for $t \leftarrow 1, 2, \ldots, B$ **do**
    $x_t^k \leftarrow \operatorname{argmin}_{u_t \in \mathbb{R}^{n_t}} \left\{ \lambda \mathcal{L}_c^\theta(x_{<t}^k, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2}\|u_t - x_t^{k-1}\|^2 \right\}$
  $q^k \leftarrow (1-\theta)p^{k-1} + c(Ax^k - d)$
  `STEP 2a` (successful termination check):
  for $t \leftarrow 1, 2, \ldots, B$ **do**
    $\delta_t^k \leftarrow \nabla_{x_t} f(x_t^k) - \nabla_{x_t} f(x_{\le t}^k, x_{>t}^{k-1})$
    $v_t^k \leftarrow \delta_t^k + cA_t^* \sum_{s=t+1}^B A_s(x_s^k - x_s^{k-1}) - \frac{1}{\lambda}(x_t^k - x_t^{k-1})$
  if $\|v^k\| \le \rho$ **and** $\|Ax^k - d\| \le \eta$ **then**
    return (x^k, q^k, v^k)
  `STEP 2b` (unsuccessful termination check):
  if $k \equiv 0 \bmod 3$ and $k \ge 9$ **then**
    $\mathcal{S}_k^{(v)} \leftarrow \frac{2}{k+1}\sum_{i=k/2}^k \|v^i\|$
    $\mathcal{S}_k^{(f)} \leftarrow \frac{2}{k+1}\sum_{i=k/2}^k \|Ax^i - d\|$
    if $\frac{1}{\rho} \cdot \mathcal{S}_k^{(v)} + \frac{1}{\eta}\sqrt{\frac{c^3}{k}} \cdot \mathcal{S}_k^{(f)} \le 1$ **then**
      return (x^k, q^k, v^k)
  `STEP 3` (multiplier update):
  $p^k \leftarrow (1-\theta)p^{k-1} + \chi c(Ax^k - d)$

PROPOSITION 2.1. *Let $(\kappa_i, \Delta_\phi, \mathcal{N}_A)$ and D_x be as in (2.7) and (2.4), respectively, and let $(\underline{c}, \chi) \in \mathbb{R}_{++}^2$ and $p^0 \in A(\mathbb{R}^n)$ be given. Moreover, define*

$$\begin{aligned}
&\tilde{\kappa}_0 := 2\left[\Delta_\phi^{1/2} + \frac{10}{\chi\sqrt{\underline{c}}}\left(1 + \frac{2\chi\kappa_1}{\kappa_2}\right)\right] \quad \tilde{\kappa}_1 := \frac{6}{\chi}\left[\sqrt{\mathcal{N}_A} + \frac{\kappa_0}{\sqrt{\underline{c}}}\right],\\
&\tau_1(c, p^0) := \left(\frac{2\kappa_4}{\kappa_2}\right)\frac{\|p^0\|^2}{c} + \frac{\kappa_4}{\kappa_2}\|p^0\| + (2\kappa_3^2 + \kappa_3)c,\\
&\tau_2(c, p^0) := \frac{4\chi D_x}{\kappa_2}\left(\left[\kappa_0 + \sqrt{\mathcal{N}_A c}\right]\left[\Delta_\phi^{1/2} + \frac{6\kappa_3\sqrt{c}}{\chi}\right] + \tilde{\kappa}_1\|p^0\|\right),
\end{aligned} \tag{2.8}$$

$$T(\rho, \eta \,|\, c, p^0) := 48\left[1 + \frac{2\tilde{\kappa}_0^2(\kappa_0^2 + \mathcal{N}_A c)}{\rho^2} + \frac{\kappa_5^2 c}{\eta^2} + \tau_1(c, p^0) + \tau_2^2(c, p^0)\right]. \tag{2.9}$$

Then, for any $c \ge \underline{c}$, the following statements hold about Algorithm 2.1 when it is given input (x^0, p^0, c):

(a) *it terminates in at most $T(\rho, \eta \,|\, c, p^0)$ iterations;*
(b) *if it terminates successfully in Step 2a, then the first two components of its output triple $(\bar{z}, \bar{q}, \bar{v})$ solve Problem $\mathcal{LCCO}$;*
(c) *if (c, p^0) satisfies $T(\rho, \eta \,|\, c, p^0) \le c^3$ then it must terminate successfully.*

We now make a few important observations about the above result. First, part (a) states that Algorithm 2.1 stops in a finite number of iterations. Second, denoting

$\varepsilon = \min\{\rho, \eta\}$, it is straightforward to verify that if $c \geq 1$, then

$$T(\rho, \eta \,|\, c, p^0) = \Theta\left(c^2 + \frac{c}{\varepsilon^2} + \|p^0\| + \|p^0\|^2\right).$$

Consequently, if $\|p^0\| + \|p^0\|^2$ is on the same order of magnitude as the other terms in the above bound, then there always exists a threshold value $\hat{c} > 0$ such that $T(\rho, \eta \,|\, c, p^0) \leq c^3$ for every $c \geq \hat{c}$. In view of part (c) and this previous observation, it follows that Algorithm 2.1 terminates successfully if its input c is sufficiently large and $\|p^0\|$ is not too large.

The above observations motivate us to develop the dynamic version of Algorithm 2.1, whose pseudocode is given in Algorithm 2.2. Specifically, Algorithm 2.2 repeatedly calls Algorithm 2.1 on an increasing sequence of penalty parameters until the final call terminates successfully.

Algorithm 2.2 DP.ADMM

`Input`: $\bar{z}^0 \in \mathcal{H}$, $\underline{c} > 0$
`Require`: $\{m_t\} \subseteq \mathbb{R}_{++}$, $(\rho, \eta) \in (0, 1]^2$, (χ, θ) as in (2.6)

$(\bar{q}^0, c_1) \leftarrow (0, \underline{c})$
for $\ell \leftarrow 1, 2, \ldots$ **do**
  call Algorithm 2.1 with inputs $(c, p^0, x^0) = (c_\ell, \bar{q}^{\ell-1}, \bar{z}^{\ell-1})$ and parameters $\{m_t\}$, (ρ, η), and (χ, θ) to obtain an output triple $(\bar{z}^\ell, \bar{q}^\ell, \bar{v}^\ell)$
  if $\|\bar{v}^\ell\| \leq \rho$ **and** $\|A\bar{z}^\ell - d\| \leq \eta$ **then**
    return $(\bar{z}^\ell, \bar{q}^\ell)$
  $c_{\ell+1} \leftarrow 2c_\ell$

In the results below, we give a uniform bound on $\bar{q}^\ell / c_\ell$, use this bound to determine the threshold value $\hat{c}$ mentioned two paragraphs above, and present a few other useful facts. For the ease of presentation, the proof of this result is given in Section 4, and it makes use of the following tolerance-independent constants:

$$\xi_0 := \frac{128\chi^2 D_x^2 \Delta_\phi}{\kappa_2^2}, \quad \xi_2 := \frac{64\chi^2 D_x^2}{\kappa_2^2}\left[\frac{72\mathcal{N}_A \kappa_3^2}{\chi^2} + \tilde{\kappa}_2^2 \kappa_3^2\right],$$
$$\xi_1 := \frac{128\chi^2 D_x^2}{\kappa_2^2}\left[\mathcal{N}_A \Delta_\phi + \frac{72\kappa_0^2 \kappa_3^2}{\chi^2}\right] + \frac{8\kappa_4\kappa_3^2 + 2\kappa_4\kappa_3}{\kappa_2} + 2\kappa_3^2 + \kappa_3, \tag{2.10}$$

where all other named constants are as in (2.7) and (2.8).

Proposition 2.2. *Let $(\kappa_i, \mathcal{N}_A)$, $\tilde{\kappa}_i$, and ξ_i be given by* (2.7), (2.8), *and* (2.10), *respectively, and define*

$$\mathcal{T}_\ell(\rho, \eta) := 48\left[1 + \xi_0 + \xi_1 c_\ell + \xi_2 c_\ell^2 + \frac{2\tilde{\kappa}_0^2(\kappa_0^2 + \mathcal{N}_A c_\ell)}{\rho^2} + \frac{\kappa_5^2 c_\ell}{\eta^2}\right], \tag{2.11}$$

for every $\ell \geq 1$. Then, the following statements hold about the ℓ^{th} iteration of Algorithm 2.2:

(a) $\|\bar{q}^\ell\| \leq 2\kappa_3 c_\ell$;

(b) the ℓ^{th} call to Algorithm 2.1 terminates in at most

$$T(\rho, \eta \,|\, c_\ell, \bar{q}^{\ell-1}) \leq \mathcal{T}_\ell(\rho, \eta) \leq \left[\max\{1, c_\ell^2\} + \frac{\max\{1, c_\ell\}}{\min\{\rho^2, \eta^2\}}\right]\mathcal{T}_1(1, 1) \tag{2.12}$$

iterations of Algorithm 2.1, where $T(\cdot, \cdot \,|\, \cdot, \cdot)$ is as in (2.9);

(c) if the ℓ^{th} *penalty parameter* $c_\ell > 0$ *satisfies*

$$c_\ell \geq \hat{c}(\rho,\eta) := \frac{\sqrt{2\mathcal{T}_1(1,1)}}{\varepsilon}, \tag{2.13}$$

then the ℓ^{th} *call to Algorithm 2.1 terminates successfully.*

The next result gives the complexity of Algorithm 2.2 in terms of the total number of iterations of Algorithm 2.1 across all of its calls.

THEOREM 2.3. *Let* $\mathcal{T}_\ell(\cdot,\cdot)$ *be as in* (2.11)*, and define*

$$\varepsilon := \min\{\rho,\eta\}, \quad E_0 := 32\max\{4, \underline{c}^2\}, \quad E_1 := 2\log_2^+(1/\underline{c}).$$

Then, Algorithm 2.2 stops and outputs a pair that solves Problem $\mathcal{LCCO}$ *in at most*

$$\mathcal{T}_1(1,1)\cdot\left\lceil \frac{E_0+E_1}{\varepsilon^2} + \frac{E_0}{\varepsilon^3}\right\rceil \tag{2.14}$$

iterations of Algorithm 2.1.

Proof. Let $\hat{c}(\cdot,\cdot)$ be as in (2.13), and define the scalars

$$\underline{\ell} := \log_2^+(1/\underline{c}), \quad \hat{\ell} := \log_2^+[\hat{c}(\rho,\eta)/\underline{c}], \quad \varepsilon = \min\{\rho,\eta\}, \quad \hat{c} := \hat{c}(\rho,\eta).$$

It follows from Proposition 2.2(c) and the penalty parameter update in Algorithm 2.2 that the number of calls of Algorithm 2.2 is at most $\hat{\ell}$. Hence, it follows from Proposition 2.1(a), Proposition 2.2, and the previous observation that Algorithm 2.2 stops and outputs a pair that solves Problem $\mathcal{LCCO}$ in at most $\sum_{\ell=1}^{\hat{\ell}}\mathcal{T}_\ell(\rho,\eta)$ iterations of Algorithm 2.1. To bound this sum, we bound the following subsums: $\sum_{\ell=1}^{\underline{\ell}-1}\mathcal{T}_\ell(\rho,\eta)$ and $\sum_{\ell=\underline{\ell}}^{\hat{\ell}}\mathcal{T}_\ell(\rho,\eta)$. For the first sum, let $1 \leq \ell < \underline{\ell}$. Since $c_\ell < 1$ (from the definition of $\underline{\ell}$) and $\varepsilon \leq 1$, it follows from Proposition 2.2(b) that

$$\sum_{\ell=1}^{\underline{\ell}-1}\mathcal{T}_\ell(\rho,\eta) \leq \sum_{\ell=1}^{\underline{\ell}-1}\frac{2\mathcal{T}_1(1,1)}{\varepsilon^2} = \frac{2\underline{\ell}\mathcal{T}_1(1,1)}{\varepsilon^2} = \frac{\mathcal{T}_1(1,1)\cdot E_1}{\varepsilon^2}. \tag{2.15}$$

For the second sum, let $\ell \geq \underline{\ell}$. Similarly, since $c_\ell \geq 1$ and $\varepsilon \leq 1$ (from the definition of $\underline{\ell}$), it follows from Proposition 2.2(b) that

$$\mathcal{T}_\ell(\rho,\eta) \leq \left(c_\ell^2 + \frac{c_\ell}{\varepsilon^2}\right)\mathcal{T}_1(1,1). \tag{2.16}$$

On the other hand, using the fact that $\log_2\hat{c} \geq 1$, we have

$$\begin{aligned}\hat{\ell} - \underline{\ell} &= \log_2^+[\hat{c}/\underline{c}] - \log_2^+[1/\underline{c}] \leq \max\{1, \log_2[\hat{c}/\underline{c}] - \log_2[1/\underline{c}] + 1\}\\ &= 1 + \max\{0, \log_2\hat{c}\} = 1 + \log_2\hat{c}\end{aligned} \tag{2.17}$$

Using (2.16), (2.17), the fact that $c_\ell = c_{\underline{\ell}}2^{\ell-\hat{\ell}}$, the bounds $\log_2\hat{c} \geq 1$ and $c_{\underline{\ell}} \leq \max\{2, \underline{c}\}$ (see the update rule for c_ℓ and the fact that $\bar{\ell}$ is the first index where c_ℓ is greater than or equal to 1), and the relation $\sum_{i=0}^{k} b^i \leq b^{k+1}$ for $b \geq 2$, it follows that

$$\frac{\sum_{\ell=\underline{\ell}}^{\hat{\ell}}\mathcal{T}_\ell(\rho,\eta)}{\mathcal{T}_1(1,1)} \leq \sum_{\ell=\underline{\ell}}^{\hat{\ell}}\left(c_\ell^2 + \frac{c_\ell}{\varepsilon^2}\right) = \sum_{i=0}^{\hat{\ell}-\underline{\ell}}\left(2^{2i}c_{\underline{\ell}}^2 + \frac{2^i c_{\underline{\ell}}}{\varepsilon^2}\right)$$

$$\le 2^{2(\hat{\ell}-\underline{\ell})+1} c_{\underline{\ell}}^2 + \frac{2^{\hat{\ell}-\underline{\ell}+1} c_{\underline{\ell}}}{\varepsilon^2} \le 4\left(2^{2[1+\log_2 \hat{c}]} c_{\underline{\ell}}^2 + \frac{2^{1+\log_2 \hat{c}} c_{\underline{\ell}}}{\varepsilon^2}\right)$$

$$\le 16\left(c_{\underline{\ell}}^2 \hat{c}^2 + \frac{c_{\underline{\ell}}\hat{c}}{\varepsilon^2}\right) \le 16\left(\hat{c}^2 + \frac{\hat{c}}{\varepsilon^2}\right)\max\{4, \underline{c}^2\}. \tag{2.18}$$

Moreover, using Proposition 2.2(c) and the relation $\mathcal{T}_1(1,1) \ge \sqrt{\mathcal{T}_1(1,1)}$, we have

$$\begin{aligned} 16\max\{4, c_1^2\}\cdot\left(\hat{c}^2 + \frac{\hat{c}}{\varepsilon^2}\right) &\le 32\mathcal{T}_1(1,1)\cdot\max\{4,\underline{c}^2\}\cdot\left(\varepsilon^{-2}+\varepsilon^{-3}\right) \\ &\le \mathcal{T}_1(1,1)\cdot E_0\cdot\left(\varepsilon^{-2}+\varepsilon^{-3}\right). \end{aligned} \tag{2.19}$$

The conclusion now follows from (2.15), (2.18), and (2.19). ∎

Notice that the bound in (2.14) is $\mathcal{O}(\varepsilon^{-3})$ in terms of the tolerances only. Hence, if $\mathcal{T}_1(1,1)$, $1/\underline{c}$, and $\underline{c}$ are $\mathcal{O}(1)$ with respect to ε then the overall complexity of Algorithm 2.2 is also $\mathcal{O}(\varepsilon^{-3})$, as claimed in Section 1.

3. Analysis of Algorithm 2.1.

This section contains two subsections. The first one establishes some key bounds on its main residuals, while the second one gives a bound on its generated Lagrange multipliers.

Throughout this section, we let $\bar{c} \in (0, c]$ and let $\{(v^i, x^i, p^i, q^i)\}_{i=1}^k$ denote the iterates generated by Algorithm 2.1 up to and including the k^{th} iteration for some $k \ge 3$. Moreover, for every $i \ge 1$ and $(\chi, \theta) \in \mathbb{R}^2_{++}$ satisfying (2.6), we make use of the following useful constants and shorthand notation

$$\begin{gathered} a_\theta = \theta(1-\theta), \quad b_\theta := (2-\theta)(1-\theta), \quad \gamma_\theta := \frac{(1-2B\chi b_\theta)-(1-\theta)^2}{2\chi}, \\ f^i := Ax^i - d, \quad \mathcal{Q}_i := \sum_{t=1}^{B}\sum_{s=t+1}^{B}\|A_t^* A_s \Delta x_s^i\|, \end{gathered} \tag{3.1}$$

the aggregated quantities in (2.1), and the averaged quantities

$$S_{j,k}^{(p)} := \frac{\sum_{i=j}^k \|p^i\|}{k-j+1}, \quad S_{j,k}^{(v)} := \frac{\sum_{i=j}^k \|v^i\|}{k-j+1}, \quad S_{j,k}^{(f)} := \frac{\sum_{i=j}^k \|f^i\|}{k-j+1}. \tag{3.2}$$

for every $1 \le j \le k$. We also denote Δy^i to be the difference of iterates for the variable y at iteration i, i.e.,

$$\Delta y^i \equiv y^i - y^{i-1}. \tag{3.3}$$

3.1. Properties of the Key Residuals.

This subsection presents bounds on the residuals $\{\|v^i\|\}_{i=2}^k$ and $\{\|f^i\|\}_{i=2}^k$ generated by Algorithm 2.1. These bounds will be particularly helpful for proving Proposition 2.1 in Section 4.

The first result presents some key properties about the generated iterates.

LEMMA 3.1. *The following statements hold for every $i \le k$:*

(a) $f^i = \left[p^i - (1-\theta)p^{i-1}\right]/(\chi c)$;

(b) $v^i \in \nabla f(x^i) + A^* q^i + \partial h(x^i)$ *and*

$$\|v^i\| \le B^2\,(M + 2m)\,\|\Delta x^i\| + c\mathcal{Q}_i. \tag{3.4}$$

Proof. (a) This is immediate from step 3 of Algorithm 2.1 and the definition of f^i in (3.1).

(b) We first prove the required inclusion. The optimality of x_t^k in Step 1 of Algorithm 2.1, assumption (A4), and the fact that $\lambda = 1/(2m)$, imply that

$$
\begin{aligned}
0 &\in \partial \left[\mathcal{L}_c^\theta(x^i_{<t}, \cdot, x^{i-1}_{>t}; p^{i-1}) + \frac{1}{2\lambda}\| \cdot - x_k^{i-1}\|^2\right](x^i) \\
&= \nabla_{x_t} f(x^i_{\le t}, x^{i-1}_{>t}) + A_t^*\left[(1-\theta)p^{i-1} + c[A(x^i_{\le t}, x^{i-1}_{>t}) - d]\right] + \partial h_t(x_t^i) + \frac{1}{\lambda}\Delta x_t^i \\
&= \nabla_{x_t} f(x^i_{\le t}, x^{i-1}_{>t}) + A_t^*\left(q^i + c\sum_{s=t+1}^{B} A_s \Delta x_s^i\right) + \partial h_t(x_t^i) + \frac{1}{\lambda}\Delta x_t^i \\
&= \nabla_{x_t} f(x^i) + A_t^* q^i + \partial h_t(x_t^i) - v_t^i.
\end{aligned}
$$

for every $1 \le t \le B$. Hence, the inclusion holds. To show the inequality, let $1 \le t \le B$ be fixed and use the triangle inequality, the definition of v_t^i, and assumption (A4) to obtain

$$
\begin{aligned}
\|v_t^i\| &\le \|\nabla_{x_t} f(x_t^i) - \nabla_{x_t} f(x^i_{\le t}, x^{i-1}_{>t})\| + c\sum_{s=t+1}^{B}\|A_t^* A_s \Delta x_s^i\| + \frac{1}{\lambda}\|\Delta x_t^k\| \\
&\le M_t\|x^i_{>t} - x^{i-1}_{>t}\| + c\sum_{s=t+1}^{B}\|A_t^* A_s \Delta x_s^i\| + 2m\|\Delta x_t^k\| \\
&\le \sum_{s=t}^{B}(M_t + 2m)\|\Delta x_s^i\| + c\sum_{s=t+1}^{B}\|A_t^* A_s \Delta x_s^i\|.
\end{aligned}
$$

Using the above bound, the definition of M in 3.1, the fact that $\lambda = 1/(2m)$, and the triangle inequality, we conclude that

$$
\begin{aligned}
\|v^i\| &\le \sum_{t=1}^{B}\|v_t^i\| \le \sum_{t=1}^{B}\sum_{s=t}^{B}(M_t + 2m)\|\Delta x_s^i\| + c\sum_{t=1}^{B}\sum_{s=t+1}^{B}\|A_t^* A_s \Delta x_s^i\| \\
&\le (M + 2m)\sum_{t=1}^{B}\sum_{s=t}^{B}\|\Delta x_s^i\| + c\mathcal{Q}_i \le B^2(M+2m)\|\Delta x^i\| + c\mathcal{Q}_i.
\end{aligned}
$$ ∎

Notice that part (c) of the above result implies that $(\bar{x}, \bar{v}, \bar{p}) = (x^i, v^i, q^i)$ satisfies the inclusion in (2.5). Hence, if $\|v^i\|$ and $\|f^i\|$ are sufficiently small at some iteration i, then Algorithm 2.1 clearly returns a solution to Problem $\mathcal{LCCO}$ at iteration i, i.e., Proposition 2.1(b) holds. However, to understand when Algorithm 2.1 terminates, we will need to develop more refined bounds on $\|v_i\|$ and $\|f_i\|$.

To begin, we present some relations between the perturbed augmented Lagrangian $\mathcal{L}_c^\theta(\cdot;\cdot)$ and the iterates $\{(x^i, p^i)\}_{i=1}^k$. For brevity, its proof is given in Appendix A.

LEMMA 3.2. *For every $i \le k$, it holds that:*

(a) $\mathcal{L}_c^\theta(x^i; p^i) - \mathcal{L}_c^\theta(x^i; p^{i-1}) = b_\theta\|\Delta p^i\|^2/(2\chi c) + a_\theta\left(\|p^i\|^2 - \|p^{i-1}\|^2\right)/(2\chi c)$;

(b) $\mathcal{L}_c^\theta(x^i; p^{i-1}) - \mathcal{L}_c^\theta(x^{i-1}; p^{i-1}) \le -3m\|\Delta x^i\|^2/2 - c\sum_{t=1}^B\|A_t \Delta x_t^i\|^2/2$;

(c) if $i \ge 2$, it holds that

$$
\frac{b_\theta}{2\chi c}\|\Delta p^i\|^2 - \frac{c}{4}\sum_{t=1}^{B}\|A_t \Delta x_t^i\|^2 \le \frac{\gamma_\theta}{4B\chi c}\left(\|\Delta p^{i-1}\|^2 - \|\Delta p^i\|^2\right); \tag{3.5}
$$

(d) $\mathcal{L}_c^\theta(x^i;p^i) \le \phi(x^{i-1}) + 3(\|p^i\|^2 + \|p^{i-1}\|^2)/(\chi^2 c)$*;*
(e) $\mathcal{L}_c^\theta(x^i;p^i) \ge \phi(x^i) - \|p^i\|^2/(2c)$*.*

The next result uses the above relations to establish a bound on the quantities in the right-hand-side of (3.4).

LEMMA 3.3. *Let* $(\kappa_0, \Delta_\phi, \mathcal{N}_A)$ *be as in* (2.7)*, and define the scalars*

$$\Psi_i(c) := \mathcal{L}_c^\theta(x^i;p^i) - \frac{a_\theta}{2\chi c}\|p^i\|^2 + \frac{\gamma_\theta}{4B\chi c}\|\Delta p^i\|^2 \quad \forall i \ge 1. \tag{3.6}$$

Then, for $1 \le j \le k$*, it holds that*

$$\sum_{i=j+1}^{k}\left[\frac{B^2(M+2m)\|\Delta x^i\| + c\mathcal{Q}_i}{\kappa_0 + \sqrt{\mathcal{N}_A c}}\right]^2 \le \Psi_j(c) - \Psi_k(c)$$

$$\le \Delta_\phi + 4\left(\frac{\|p^j\|^2 + \|p^{j-1}\|^2 + \|p^k\|^2}{\chi^2 c}\right). \tag{3.7}$$

Proof. Using the fact that $\|z\|_1^2 \le n\|z\|_2^2$ for every $z \in \mathbb{R}^n$, the definition of $\mathcal{Q}_i$ in (3.1), and the fact that $\|Mx\| \le \|M\|\|x\|$ for any matrix M, we first have

$$c\mathcal{Q}_i^2 \le B^2 c \sum_{t=1}^{B}\sum_{s=t+1}^{B}\|A_t^* A_s \Delta x_s^i\|^2 \le B^2 c \sum_{t=1}^{B}\|A_t\|^2 \sum_{s=t+1}^{B}\|A_s\Delta x_s^i\|^2$$

$$\le \left(B^2\sum_{t=1}^{B}\|A_t\|^2\right)\left(c\sum_{s=1}^{B}\|A_s\Delta x_s^i\|^2\right)$$

$$= \left(4B^2\sum_{t=1}^{B}\|A_t\|^2\right)\left(\frac{c}{4}\sum_{s=1}^{B}\|A_s\Delta x_s^i\|^2\right). \tag{3.8}$$

Combining (3.8), Lemma 3.2(a)–(b), the definition of Ψ_θ^i, and the bound $(a+b)^2 \le 2a^2 + 2b^2$ for $a, b \in \mathbb{R}_+$, it follows that

$$\left[\frac{B^2(M+2m)\|\Delta x^i\| + c\mathcal{Q}_i}{\kappa_0 + \sqrt{\mathcal{N}_A c}}\right]^2 \le \frac{2B^4(M+2m)^2\|\Delta x^i\|^2 + 2c^2\mathcal{Q}_i^2}{\kappa_0^2 + \mathcal{N}_A c}$$

$$\le \frac{3m}{2}\|\Delta x_t^i\|^2 + \frac{c\mathcal{Q}_i^2}{4B^2\sum_{t=1}^{B}\|A_t\|^2} \overset{(3.8)}{\le} \frac{3m}{2}\|\Delta x_t^i\|^2 + \frac{c}{4}\sum_{t=1}^{B}\|A_t\Delta x_t^i\|^2$$

$$\le \mathcal{L}_c^\theta(x^{i-1};p^{i-1}) - \mathcal{L}_c^\theta(x^i;p^i) +$$

$$\qquad \frac{a_\theta}{2\chi c}(\|p^i\|^2 - \|p^{i-1}\|^2) + \frac{b_\theta}{2\chi c}\|\Delta p^i\|^2 - \frac{c}{4}\sum_{t=1}^{B}\|A_t\Delta x_t^i\|^2$$

$$\le \mathcal{L}_c^\theta(x^{i-1};p^{i-1}) - \mathcal{L}_c^\theta(x^i;p^i) + \frac{a_\theta}{2\chi c}(\|p^i\|^2 - \|p^{i-1}\|^2) + \frac{\gamma_\theta}{4B\chi c}\left(\|\Delta p^{i-1}\|^2 - \|\Delta p^i\|^2\right)$$

$$= \Psi_{i-1}(c) - \Psi_i(c).$$

Consequently, summing the above inequality from $i = j+1$ to k yields the leftmost bound. To prove the rightmost bound, we use Lemma 3.2(d)–(e), the inclusions $a_\theta \in (0,1)$ and $(\chi,\theta) \in (0,1)^2$, the relation $(a+b)^2 \le 2a^2 + 2b^2$ for $a, b \in \mathbb{R}_+$, and the bound $\gamma_\theta \le 1/(2\chi)$ to obtain

$$\Psi_j(c) - \Psi_k(c)$$

$$
\begin{aligned}
&= \left[\mathcal{L}_c^\theta(x^j;p^j) - \mathcal{L}_c^\theta(x^k;p^k)\right] + \frac{a_\theta(\|p^k\|^2 - \|p^j\|^2)}{2\chi c} + \frac{\gamma_\theta(\|\Delta p^j\|^2 - \|\Delta p^k\|^2)}{4B\chi c} \\
&\le \left[\mathcal{L}_c^\theta(x^j;p^j) - \mathcal{L}_c^\theta(x^k;p^k)\right] + \frac{a_\theta\|p^k\|^2}{2\chi c} + \frac{\gamma_\theta\|\Delta p^j\|^2}{4B\chi c} \\
&\le \left[\mathcal{L}_c^\theta(x^j;p^j) - \mathcal{L}_c^\theta(x^k;p^k)\right] + \frac{\|p^k\|^2}{2\chi c} + \frac{\|p^{j-1}\|^2 + \|p^j\|^2}{4B\chi^2 c} \\
&\le \left[\phi(x^{j-1}) - \phi(x^k) + \frac{3(\|p^j\|^2 + \|p^{j-1}\|^2)}{\chi^2 c} + \frac{\|p^k\|^2}{2c}\right] + \\
&\qquad \frac{\|p^k\|^2}{2\chi c} + \frac{\|p^{j-1}\|^2 + \|p^j\|^2}{4B\chi^2 c} \le \Delta_\phi + 4\left(\frac{\|p^j\|^2 + \|p^{j-1}\|^2 + \|p^k\|^2}{\chi^2 c}\right). \qquad \square
\end{aligned}
$$

The next result presents bounds on $S^{(f)}_{j+1,k}$ and $S^{(v)}_{j+1,k}$.

Proposition 3.4. *Let $(\kappa_0, \Delta_\phi, \mathcal{N}_A)$ be as in (2.7). Then, for every $1 \le j < k$, it holds that*

$$
S^{(f)}_{j+1,k} \le \frac{\|p^j\| + 2S^{(p)}_{j+1,k}}{\chi c}, \tag{3.9}
$$

$$
S^{(v)}_{j+1,k} \le \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{k-j}}\left(\Delta_\phi^{1/2} + \frac{\|p^j\| + \|p^{j-1}\| + \|p^k\|}{\chi\sqrt{c}}\right). \tag{3.10}
$$

Proof. Using Lemma 3.1(a), the fact that $\theta \in (0,1)$, and the triangle inequality, it holds that

$$
S^{(f)}_{j+1,k} = \frac{\sum_{i=j+1}^k \|p^i - (1-\theta)p^{i-1}\|}{\chi c(k-j)} \le \frac{\sum_{i=j+1}^k (\|p^{i-1}\| + \|p^i\|)}{\chi c(k-j)} \le \frac{\|p^j\| + 2S^{(p)}_{j+1,k}}{\chi c},
$$

which is (3.9). On the other hand, to show (3.10), we use the fact that $\|a\|_1 \le \sqrt{n}\|a\|_2$ for $a \in \mathbb{R}^n$, Lemma 3.1(b), Lemma 3.3, and the fact that $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ for $a, b \in \mathbb{R}_+$, to obtain

$$
\begin{aligned}
S^{(v)}_{j+1,k} = \frac{\sum_{i=j+1}^k \|v^i\|}{k-j} &\le \left(\frac{\sum_{i=j+1}^k \|v^i\|^2}{k-j}\right)^{1/2} \\
&\le \left(\frac{\sum_{i=j+1}^k \left[B^2(M+2m)\|\Delta x^i\| + c\mathcal{Q}_i\right]^2}{k-j}\right)^{1/2} \\
&\le \frac{\kappa_0 + \sqrt{\mathcal{N}_A c}}{\sqrt{k-j}}\left[\Delta_\phi + 4\left(\frac{\|p^j\|^2 + \|p^{j-1}\|^2 + \|p^k\|^2}{\chi^2 c}\right)\right]^{1/2} \\
&\le \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{k-j}}\left(\Delta_\phi^{1/2} + \frac{\|p^j\| + \|p^{j-1}\| + \|p^k\|}{\chi\sqrt{c}}\right). \qquad \square
\end{aligned}
$$

Now, observe that both residuals $S^{(v)}_{j+1,k}$ and $S^{(f)}_{j+1,k}$ depend on the size of the Lagrange multipliers. Since both termination conditions in Algorithm 2.1 require $\|v^i\|$, $\|f^i\|$, or some combination of the two to be sufficiently small, our goal for the next subsection is to bound the size of generated multipliers.

3.2. Bounding the Lagrange Multipliers. This subsection generalizes the analysis in [18]. More specifically, Proposition 3.9 shows that if k is sufficiently large relative to an index j, the penalty parameter c, and $\|p^0\|$, then $S^{(p)}_{j+1,k} = \mathcal{O}(1)$.

The first result, whose proof can be found in [12, Lemma 1.2], presents a relation on elements in the image of a linear operator.

LEMMA 3.5. *For any $S \in \mathbb{R}^{m\times n}$ and $u \in S(\mathbb{R}^{m\times n})$, we have $\sigma_S^+\|u\| \le \|Su\|$.*

The proof of the next result can be found in [21, Lemma 4.7].

LEMMA 3.6. *Suppose $\psi \in \overline{\mathrm{Conv}}\ \mathbb{R}^n$ is K_ψ-Lipschitz continuous. Then, for every $z, \bar z \in \mathrm{dom}\,\psi$ and $r \in \partial\psi(z)$, it holds that*

$$\|r\|\mathrm{dist}_{\partial(\mathrm{dom}\,\psi)}(\bar z) \le \left[\mathrm{dist}_{\partial(\mathrm{dom}\,\psi)}(\bar z) + \|z-\bar z\|\right] K_\psi + \langle r, z-\bar z\rangle,$$

where $\partial(\mathrm{dom}\,\psi)$ denotes the boundary of $\mathrm{dom}\,\psi$.

The following result presents some fundamental properties about p^{i-1}, p^i, and q^i.

LEMMA 3.7. *Let $d_\circ$, D_x, κ_i be as in (A5), (2.4), (2.7), respectively. Then, for every $i \ge 1$,*

(a) $p^i = \chi q^i + (1-\chi)(1-\theta)p^{i-1}$;

(b) $\|p^i\| \le \|p^0\| + \kappa_3 c$;

(c) it holds that

$$\frac{1}{c}\|q^i\|^2 + d_\circ\sigma_A^+\|q^i\| \le \left(\frac{1-\theta}{c}\right)\left\langle q^i, p^{i-1}\right\rangle + 2cD_x\mathcal{Q}_i + 2\kappa_1.$$

Proof. (a) This is an immediate consequence of the updates for p^i and q^i in Algorithm 2.1.

(b) In view of Step 3 of Algorithm 2.1, the fact that $\theta \in (0,1)$, and the triangle inequality, it holds that

$$\begin{aligned}
\|p^i\| &\le (1-\theta)\|p^{i-1}\| + \chi c\|Ax^i - d\| \\
&\le (1-\theta)^i\|p^0\| + \chi c\sum_{j=0}^{i-1}(1-\theta)^j\|Ax^j - d\| \\
&\le \|p^0\| + \chi c\cdot\sup_{x\in\mathcal{F}}\|Ax-d\|\sum_{j=0}^{\infty}(1-\theta)^j \\
&= \|p^0\| + \chi c\left(\frac{\sup_{x\in\mathcal{H}}\|Ax-d\|}{\theta}\right) = \|p^0\| + \kappa_3 c.
\end{aligned}$$

(c) Using Lemma 3.5 with $(S,u) = (A, q^i)$, Lemma 3.1(b), the fact that $q^i \in A(\mathbb{R}^n)$, and the triangle inequality, we first have that

$$\begin{aligned}
\frac{1}{c}\|q^i\|^2 + d_\circ\sigma_A^+\|q^i\| &\le \frac{1}{c}\|q^i\|^2 + d_\circ\|A^*q^i\| \\
&\le \frac{1}{c}\|q^k\|^2 + d_\circ\left[\|v^i - \nabla f(x^i) - A^*q^i\| + \|\nabla f(x^i)\| + \|v^i\|\right] \\
&\le \frac{1}{c}\|q^i\|^2 + d_\circ\left[\|v^i - \nabla f(x^i) - A^*q^i\| + G_f + B^2(M+2m)D_x + c\mathcal{Q}_i\right] \\
&\le \frac{1}{c}\|q^i\|^2 + d_\circ\|v^i - \nabla f(x^i) - A^*q^i\| + c\mathcal{Q}_iD_x + \kappa_1 - K_hD_x. \qquad (3.11)
\end{aligned}$$

We now derive a suitable bound on $d_\circ\|v^i - \nabla f(x^i) - A^*q^i\|$. First, observe that Lemma 3.1(b) implies that $v^i - \nabla f(x^i) - A^*q^i \in \partial h(x^i)$. Using the definition of D_x in

(2.4) and Lemma 3.6 with $(\psi, z, \bar{z}) = (h, x^i, \mathring{x})$ and $r = v^i - \nabla f(x^i) - A^* q^i$, it follows that

$$\begin{aligned} d_\circ \|v^i - \nabla f(x^i) - A^* q^i\| &= \|v^i - \nabla f(x^i) - A^* q^i\| \mathrm{dist}_{\partial\mathcal{H}}(\mathring{x}) \\ &\le \left[\mathrm{dist}_{\partial\mathcal{H}}(\mathring{x}) + \|x^i - \mathring{x}\|\right] K_h + \left\langle v^i - \nabla f(x^i) - A^* q^i, x^i - \mathring{x}\right\rangle \\ &\le 2K_h D_x + \left\langle v^i - \nabla f(x^i) - A^* q^i, x^i - \mathring{x}\right\rangle. \end{aligned} \tag{3.12}$$

On the other hand, Lemma 3.1(b), the Cauchy-Schwarz inequality, the definition of κ_1, and the fact that $Ax^i - d = [q^i - (1-\theta)p^{i-1}]/c$ imply that

$$\begin{aligned} &\left\langle v^i - \nabla f(x^i) - A^* q^i, x^i - \mathring{x}\right\rangle \\ &\le \left(\|v^i\| + \|\nabla f(x^i)\|\right) \|x^i - \mathring{x}\| - \left\langle q^i, Ax^i - d\right\rangle \\ &\le \left[B^2 (M + 2m) D_x + c\mathcal{Q}_i + G_f\right] D_x - \left\langle q^i, Ax^i - d\right\rangle \\ &= \kappa_1 - K_h D_x + c\mathcal{Q}_i D_x + \left(\frac{1-\theta}{c}\right) \left\langle q^i, p^{i-1}\right\rangle - \frac{1}{c}\|q^i\|^2. \end{aligned} \tag{3.13}$$

The conclusion now follow from combining (3.11), (3.12), and (3.13). ∎

The next result presents two important technical bounds. One of them shows that $\|p^i\|$ is bounded by a *nearly* telescopic quantity, while the other gives a bound on $c\sum_{i=j+1}^k \mathcal{Q}_i$.

LEMMA 3.8. *Let $d_\circ$, D_x, κ_i, and $\tau_i(\cdot,\cdot)$ be as in (A5),* (2.4)*,* (2.7)*, and* (2.8)*, respectively, and define*

$$d_\theta := \frac{2(1-\theta)^2}{1 + \sqrt{1 + 4(1-\theta)^2}}, \quad e_\theta := (1-\theta)(1-\chi). \tag{3.14}$$

Then, it holds that:

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

$$\kappa_2 \|p^i\| \le 4\chi(\kappa_1 + c\mathcal{Q}_i D_x) + e_\theta d_\circ \sigma_A^+ (\|p^{i-1}\| - \|p^i\|) + \frac{d_\theta(\|p^{i-1}\|^2 - \|p^i\|^2)}{c};$$

(b) for every $1 \le j < k$, we have

$$\frac{c\sum_{i=j+1}^k \mathcal{Q}_i}{k-j} \le \left[\frac{\kappa_2}{4\chi D_x}\right]\left[\frac{\tau_2(c, p^0)}{\sqrt{k-j}}\right].$$

Proof. (a) Let $i \le k$ be arbitrary, suppose $\theta \in (0,1)$, and define

$$\begin{aligned} \nu_i(c) &:= \kappa_1 + c\mathcal{Q}_i D_x, \quad g_\theta := \frac{1 + \sqrt{1 + 4(1-\theta)^2}}{2(1-\theta)}, \\ \Delta_{p,i}^{(1)} &:= \|p^i\| - \|p^{i-1}\|, \quad \Delta_{p,i}^{(2)} := \|p^i\|^2 - \|p^{i-1}\|^2. \end{aligned} \tag{3.15}$$

Using Lemma 3.7(a) *thrice*, Lemma 3.7(c), the relations $e_0 \in (0,1)$, $\theta \in (0,1)$, and $\chi \le \chi^2 \in (1,0)$, and the bounds $2ab \le g_\theta a^2 + b^2/g_\theta$ and $(a+b)^2 \le 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}_+$, we first have that

$$\frac{1}{c}\|p^i\|^2 + d_\circ \sigma_A^+ \|p^i\| = \frac{1}{c}\|\chi q^i + e_\theta p^{i-1}\|^2 + d_\circ \sigma_A^+ \|\chi q^i + e_\theta p^{i-1}\|$$

$$
\begin{aligned}
&\le 2\chi\left[\frac{1}{c}\|q^i\|^2 + d_\circ\sigma_A^+\|q^i\|\right] + \frac{2e_\theta^2}{c}\|p^{i-1}\|^2 + e_\theta d_\circ \sigma_A^+\|p^{i-1}\| \\
&\le 2\chi\left[\frac{1-\theta}{c}\left\langle q^i, p^{i-1}\right\rangle + 2\nu_i(c)\right] + \frac{2e_\theta^2}{c}\|p^{i-1}\|^2 + e_\theta d_\circ \sigma_A^+\|p^{i-1}\| \\
&= 2\left[\frac{1-\theta}{c}\left\langle p^i - e_\theta p^{i-1}, p^{i-1}\right\rangle + 2\chi\nu_i(c)\right] + \frac{2e_\theta^2}{c}\|p^{i-1}\|^2 + e_\theta d_\circ \sigma_A^+\|p^{i-1}\| \\
&\le \frac{2(1-\theta)}{c}\left\langle p^i, p^{i-1}\right\rangle + \frac{2e_\theta(e_\theta-1)}{c}\|p^{i-1}\|^2 + e_\theta d_\circ \sigma_A^+\|p^{i-1}\| + 4\chi\nu_i(c) \\
&\overset{e_\theta\in(0,1)}{\le} \frac{(1-\theta)g_\theta}{c}\|p^i\|^2 + \frac{1-\theta}{c\cdot g_\theta}\|p^{i-1}\|^2 + e_\theta d_\circ \sigma_A^+\|p^{i-1}\| + 4\chi\nu_i(c). \qquad (3.16)
\end{aligned}
$$

Subtracting $e_0 d_\circ \sigma_A^+\|p_i\| + d_\theta\|p^i\|^2 + (1-\theta)g_\theta/c\|p^i\|^2$ from both sides and using the relations $\kappa_2 = (1-e_0)d_\circ\sigma_A^+$, $d_\theta = (1-\theta)/g_\theta$, and $(1-\theta)g_\theta^2 - g_\theta + (1-\theta) = 0$, we conclude that

$$
\begin{aligned}
&4\chi(\kappa_1 + c\mathcal{Q}_i D_x) - e_\theta d_\circ \sigma_A^+ \Delta_{p,i}^{(1)} - \frac{d_\theta \Delta_{p,i}^{(2)}}{c} \\
&\ge (1-e_0)d_\circ\sigma_A^+\|p^i\| + \frac{\|p^i\|^2}{c}\left[1 - d_\theta - (1-\theta)g_\theta\right] \\
&= \kappa_2\|p^i\| - \frac{\|p^i\|^2}{g_\theta\cdot c}\underbrace{\left[(1-\theta)g_\theta^2 - g_\theta + (1-\theta)\right]}_{=0} = \kappa_2\|p^i\|
\end{aligned}
$$

and, hence, the desired bound holds for $\theta \in (0,1)$. Taking the limit of the bound as $\theta \uparrow 1$ implies that the bound also holds for $\theta = 1$.

(b) Using the relation $\|z\|_1 \le \sqrt{d}\|z\|_2$ for any $z \in \mathbb{R}^d$, the bound $\sqrt{a+b} \le \sqrt{a}+\sqrt{b}$ for $a, b \in \mathbb{R}_+$, Lemma 3.7(b), and Lemma 3.3, it holds that

$$
\begin{aligned}
\frac{\sum_{i=j+1}^k c\mathcal{Q}_i}{k-j} &\le \frac{(\sum_{i=j+1}^k c^2\mathcal{Q}_i^2)^{1/2}}{\sqrt{k-j}} \\
&\le \frac{\kappa_0 + \sqrt{\mathcal{N}_A c}}{\sqrt{k-j}}\left[\Delta_\phi^{1/2} + 2\left(\frac{\|p^j\| + \|p^{j-1}\| + \|p^k\|}{\chi\sqrt{c}}\right)\right] \\
&\le \frac{\kappa_0 + \sqrt{\mathcal{N}_A c}}{\sqrt{k-j}}\left[\Delta_\phi^{1/2} + \frac{6(\|p^0\| + \kappa_3 c)}{\chi\sqrt{c}}\right] \\
&\le \frac{\kappa_0 + \sqrt{\mathcal{N}_A c}}{\sqrt{k-j}}\left[\Delta_\phi^{1/2} + \frac{6\kappa_3\sqrt{c}}{\chi}\right] + \frac{6\|p^0\|}{\chi\sqrt{k-j}}\left[\sqrt{\mathcal{N}_A} + \frac{\kappa_0}{\sqrt{\underline{c}}}\right] \\
&= \left[\frac{\kappa_2}{4\chi D_x}\right]\left[\frac{\tau_2(c,p^0)}{\sqrt{k-j}}\right].
\end{aligned}
$$

□

We are now ready to present the claimed bound on $S_{j+1,k}^{(p)}$.

PROPOSITION 3.9. *Let κ_i and τ_i be as in (2.7) and (2.8), respectively. Then, for every $1 \le j < k$, it holds that*

$$
S_{j+1,k}^{(p)} \le \frac{4\chi\kappa_1}{\kappa_2} + \frac{\tau_1(c,p^0)}{k-j} + \frac{\tau_2(c,p^0)}{\sqrt{k-j}}. \qquad (3.17)
$$

Moreover, if $k \ge j + \tau_1(c,p^0) + \tau_2^2(c,p^0)$, then $S_{j+1,k}^{(p)} \le 2 + 4\chi\kappa_1/\kappa_2$.

Proof. Let $\Delta_p^{(1)}$, $\Delta_p^{(2)}$, d_θ, and e_θ be as in Lemma 3.8, and let $\nu_i(c)$ be as in (3.15). Summing the bound in Lemma 3.8(a) from $i = j+1$ to k and using the resulting bound with Lemma 3.7(b) and the fact that d_θ is smaller than the first term in κ_4, it follows that

$$\begin{aligned}\kappa_2 \sum_{i=j+1}^{k} \|p^i\| &\le \frac{d_\theta}{c}(\|p^j\|^2 - \|p^k\|^2) + e_0 d_\circ \sigma_A^+(\|p^j\| - \|p^k\|) + 4\chi \sum_{i=j+1}^{k} \nu_i(c) \\ &\le \kappa_4 \left(\frac{\|p^j\|^2}{c} + \|p^j\| \right) + 4\chi \sum_{i=j+1}^{k} \nu_i(c) \\ &\le \kappa_4 \left[\frac{2\|p^0\|^2}{c} + \|p^0\| + (2\kappa_3^2 + \kappa_3)c \right] + 4\chi \sum_{i=j+1}^{k} \nu_i(c). \qquad (3.18)\end{aligned}$$

Dividing the above bound by $\kappa_2(k-j)$ and using the definitions of $S_{j+1,k}^{(p)}$ and $\nu_i(c)$ with Lemma 3.8(b), it holds that

$$\begin{aligned}S_{j+1,k}^{(p)} &\le \frac{\kappa_4}{\kappa_2(k-j)} \left[\frac{2\|p^0\|^2}{c} + \|p^0\| + (2\kappa_3^2 + \kappa_3)c \right] + \frac{4\chi \sum_{i=j+1}^{k} \nu_i(c)}{\kappa_2(k-j)} \\ &= \frac{4\chi\kappa_1}{\kappa_2} + \frac{\tau_1(c,p^0)}{k-j} + \frac{4\chi D_x \sum_{i=j+1}^{k} c\mathcal{Q}_i}{\kappa_2(k-j)} \\ &\le \frac{4\chi\kappa_1}{\kappa_2} + \frac{\tau_1(c,p^0)}{k-j} + \frac{\tau_2(c,p^0)}{\sqrt{k-j}},\end{aligned}$$

which is exactly (3.17). The last statement of the proposition follows immediately from the fact that $k \ge j + \tau_1(c,p^0) + \tau_2^2(c,p^0)$ implies $k - j \ge \tau_1(c,p^0)$ and $\sqrt{k-j} \ge \tau_2(c,p^0)$. ∎

We end this subsection by discussing some implications of the above results. Suppose ζ is an integer satisfying $\zeta \ge 1 + \tau_1(c,p^0) + \tau_2^2(c,p^0) = \Omega(c^2 + \|p^0\|^2)$. It then follows from Proposition 3.9 that $S_{2,\zeta}^{(p)} = \mathcal{O}(1)$ and $S_{2\zeta,3\zeta}^{(p)} = \mathcal{O}(1)$. Since the minimum of a set of scalars minorizes the average of these scalars, there exists indices $j_0 \in \{2,\ldots,\zeta\}$ and $k_0 \in \{2\zeta,\ldots,3\zeta\}$ such that $\|p^{j_0}\| = \mathcal{O}(1)$ and $\|p^{k_0}\| = \mathcal{O}(1)$. Using the fact that $k_0 - j_0 \ge \zeta$, the above bounds, and (3.9)–(3.10), it is reasonable to expect $S_{j_0+1,k_0}^{(f)} = \mathcal{O}(1/c)$ and $S_{j_0+1,k_0}^{(v)} = \mathcal{O}(\tau_0(c)/\sqrt{\zeta})$. In the next section, we give the exact steps of this argument and use the resulting bounds to prove Proposition 2.1.

4. Proof of Propositions 2.1 and 2.2. Before presenting the proofs, we first refine the bounds in Proposition 3.4.

LEMMA 4.1. *Let $(\kappa_i, \mathcal{N}_A)$ and $(\tilde\kappa_i, \tau_i)$ be as in (2.7) and (2.8), respectively, and suppose $\zeta \in \mathbb{N}$ satisfies $\zeta \ge 1 + \tau_1(c,p^0) + \tau_2^2(c,p^0)$. Then, there exists $j_0 \in \{3,\ldots,\zeta\}$ and $k_0 \in \{2\zeta+1,\ldots,3\zeta\}$ such that*

$$S_{j_0,k_0}^{(v)} \le \frac{\tilde\kappa_0(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{\zeta}}, \quad S_{j_0,k_0}^{(f)} \le \frac{\kappa_5}{c}. \qquad (4.1)$$

Proof. Suppose $\zeta \in \mathbb{N}$ satisfies $\zeta \ge 1 + \tau_1(c,p^0)\tau_2^2(c,p^0)$. Using Proposition 3.9 with $(j,k) = (1,\zeta)$ it holds that there exists $3 \le j_0 \le k$ such that

$$\|p^{j_0-1}\| + \|p^{j_0}\| \le \frac{\sum_{i=3}^{\zeta}(\|p^{i-1}\| + \|p^i\|)}{\zeta - 2} \le \frac{2\sum_{i=2}^{\zeta}\|p^i\|}{\zeta - 2}$$

$$= \frac{2(\zeta-1)S^{(p)}_{2,\zeta}}{\zeta-2} \le 4S^{(p)}_{2,\zeta} \le 8 + \frac{16\chi\kappa_1}{\kappa_2}. \tag{4.2}$$

On the other hand, using Proposition 3.9 with $(j,k) = (2\zeta, 3\zeta)$ it holds that there exists $k_0 \in \{2\zeta+1, \ldots, 3\zeta\}$ such that

$$\|p^{k_0}\| \le \frac{\sum_{i=2\zeta+1}^{3\zeta} \|p^i\|}{\zeta} = S_{2\zeta+1,3\zeta} \le 2 + \frac{4\chi\kappa_1}{\kappa_2}. \tag{4.3}$$

Combining (4.2), (4.3), the fact that $k_0 - j_0 \ge \zeta$, and Proposition 3.4 with $(j,k) = (j_0, k_0)$, it follows that

$$\begin{aligned} S^{(v)}_{j_0+1,k_0} &\le \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{k_0 - j_0}} \left(\Delta_\phi^{1/2} + \frac{\|p^{j_0}\| + \|p^{j_0-1}\| + \|p^{k_0}\|}{\chi\sqrt{c}} \right) \\ &\overset{(4.2)\text{–}(4.3)}{\le} \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{k_0 - j_0}} \left[\Delta_\phi^{1/2} + \frac{10}{\chi\sqrt{c}} \left(1 + \frac{2\chi\kappa_1}{\kappa_2}\right) \right] \\ &\le \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{\zeta}} \left[\Delta_\phi^{1/2} + \frac{10}{\chi\sqrt{\underline{c}}} \left(1 + \frac{2\chi\kappa_1}{\kappa_2}\right) \right] = \frac{\tilde{\kappa}_0(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\sqrt{\zeta}}, \end{aligned}$$

which is the first bound in (4.1). To show the other bound in (4.1), we use (4.2) and Proposition 3.9 with $(j,k) = (j_0, k_0)$ to conclude that

$$S^{(f)}_{j_0+1,k_0} \le \frac{\|p^{j_0}\| + 2S^{(p)}_{j_0+1,k_0}}{\chi c} \le \frac{12}{\chi c}\left(1 + \frac{2\chi\kappa_1}{\kappa_2}\right) = \frac{\kappa_5}{c}.$$ ∎

We are now ready to give the proof of Proposition 2.1.

Proof of Proposition 2.1. (a) Let $(\rho, \eta) \in (0,1)$, $p^0 \in A(\mathbb{R}^n)$, and $c > 0$ be given, and define

$$T := T(\rho, \eta \,|\, c, p^0), \quad r_j := \frac{\mathcal{S}^{(v)}_j}{\rho} + \frac{\mathcal{S}^{(f)}_j}{\eta}\sqrt{\frac{c^3}{j}} \quad \forall j \ge 1,$$

where $\mathcal{S}^{(v)}_j$ and $\mathcal{S}^{(f)}_j$ are as in Step 2b of Algorithm 2.1. For the sake of contradiction, suppose that Algorithm 2.1 has not terminated by the end of iteration $k = T$. It then follows from the definition of T, Lemma 4.1 with $\zeta = T/3$, and the relation $(a+b)^2 \le 2a^2 + 2b^2$ for $a, b \in \mathbb{R}_+$ that there exists $j_0 \in \{3, \ldots, T/3\}$ and $k_0 \in \{2T/3+1, \ldots, T\}$ such that

$$\begin{aligned} &\frac{S^{(v)}_{j_0,k_0}}{\rho} + \frac{c^{3/2}S^{(f)}_{j_0,k_0}}{\eta\sqrt{T/3}} \le \frac{\tilde{\kappa}_0(\kappa_0 + \sqrt{\mathcal{N}_A c})}{\rho\sqrt{T/3}} + \frac{\kappa_5\sqrt{c}}{\eta\sqrt{T/3}} \\ &= \sqrt{\frac{3\tilde{\kappa}_0^2(\kappa_0 + \sqrt{\mathcal{N}_A c})^2}{\rho^2 T}} + \sqrt{\frac{3\kappa_5^2 c}{\eta^2 T}} \le \sqrt{\frac{6\tilde{\kappa}_0^2(\kappa_0^2 + \mathcal{N}_A c)}{\rho^2 T}} + \frac{1}{4} \le \frac{1}{4} + \frac{1}{4} = \frac{1}{2}. \end{aligned} \tag{4.4}$$

Now, without loss of generality, suppose k_0 is even. Combining (4.4), the relations $S^{(v)}_{k_0/2,k_0} = \mathcal{S}^{(v)}_{k_0}$, $S^{(f)}_{k_0/2,k_0} = \mathcal{S}^{(f)}_{k_0}$, and $j_0 \le T/3 < k_0/2 < k_0$, we conclude that

$$r_{k_0} = \frac{S^{(v)}_{k_0/2,k_0}}{\rho} + \frac{c^{3/2}S^{(f)}_{k_0/2,k_0}}{\eta\sqrt{k_0}} \le \frac{k_0 - j_0 + 1}{k_0 - k_0/2 + 1}\left[\frac{S^{(v)}_{j_0,k_0}}{\rho} + \frac{c^{3/2}S^{(f)}_{j_0,k_0}}{\eta\sqrt{T/3}} \right]$$

$$\leq \frac{k_0+2}{k_0/2+1}\left[\frac{S^{(v)}_{j_0,k_0}}{\rho}+\frac{c^{3/2}S^{(f)}_{j_0,k_0}}{\eta\sqrt{T/3}}\right] \leq 2\left[\frac{S^{(v)}_{j_0,k_0}}{\rho}+\frac{c^{3/2}S^{(f)}_{j_0,k_0}}{\eta\sqrt{T/3}}\right] \overset{(4.4)}{\leq} 1,$$

which, in view of Step 2b of Algorithm 2.1, implies that termination must occur at or before iteration $k_0 \leq T$. Since this contradicts our initial assumption, it must be the case that each call of Algorithm 2.1 is run for at most T iterations.

(b) This follows from the stopping condition in Step 2a and Lemma 3.1(b).

(c) Let (T, r_j) be as in part (a) and suppose that $T \leq c^3$. In view of the conclusion of part (a), let $j \leq T$ be the first even index where $r_j \leq 1$. Using the fact that r_j itself is an average of scalars, there exists $j/2 \leq i \leq j$ such that

$$\frac{\|v^i\|}{\rho}+\frac{c^{3/2}\|f^i\|}{\eta\sqrt{j}} \leq \frac{S^{(v)}_{j/2,j}}{\rho}+\frac{c^{3/2}S^{(f)}_{j/2,j}}{\eta\sqrt{j}} \leq 1.$$

Hence, it holds that $\|v^i\| \leq \rho$ and, from our initial bound on T, we have $\|f^i\| \leq \eta\sqrt{j}c^{-3/2} \leq \eta\sqrt{T}c^{-3/2} \leq \eta$. Since $i \leq j \leq T$, it follows from part (a) that Algorithm 2.1 terminates successfully in Step 2a at iteration i, which is before the first index j where it can terminate unsuccessfully. ∎

Finally, we give the proof of Proposition 2.2.

Proof of Proposition 2.2. (a) We proceed by induction. Since $\bar{q}^0 = 0$, the case of $\ell = 0$ is immediate. Suppose the statement holds for some iteration $\ell - 1$. Then, it follows from Lemma 3.7(b) with $(p^0, c) = (\bar{q}^{\ell-1}, c_\ell)$ and the relation $c_\ell = 2c_{\ell-1}$ that

$$\|\bar{q}^\ell\| \leq \|\bar{q}^{\ell-1}\| + \kappa_3 c_\ell \leq \kappa_3(2c_{\ell-1}+c_\ell) = 2\kappa_3 c_\ell.$$

(b) The fact that the iteration count is bounded by $T(\rho,\eta \,|\, c_\ell, \bar{q}^{\ell-1})$ follows immediately from Proposition 2.1(a) and how Algorithm 2.1 is called in Algorithm 2.2. We now show that the leftmost bound in (2.12) holds. Notice that the scalar $\mathcal{T}_\ell(\rho,\eta)$ is non-decreasing in terms of the variables $\max\{1, c_\ell\}$ and $1/\min\{\rho,\eta\}$ and that these variables are clearly lower bounded by 1 for $(\rho,\eta) \in (0,1)^2$. Hence, the desired bound follows from these facts and the requirement that $(\rho,\eta) \in (0,1)^2$ in Algorithm 2.2.

We next show that the rightmost bound in (2.12) holds. Notice first that (2.9) implies that it suffices to show that $\tau_1(c_\ell,\bar{q}^{\ell-1}) + \tau_2(c_\ell,\bar{q}^{\ell-1}) \leq \xi_0 + \xi_1 c_\ell + \xi_2 c_\ell^2$. Using part (a) and the definition of $\tau_1(\cdot,\cdot)$, we first have that

$$\begin{aligned}\tau_1(c_\ell,\bar{q}^{\ell-1}) &\leq \left(\frac{2\kappa_4}{\kappa_2}\right)\frac{4\kappa_3^2c_\ell^2}{c_\ell} + \frac{2\kappa_4\kappa_3c_\ell}{\kappa_2} + (2\kappa_3^2+\kappa_3)c_\ell \\ &= \left(\frac{8\kappa_4\kappa_3^2+2\kappa_4\kappa_3}{\kappa_2} + 2\kappa_3^2+\kappa_3\right)c_\ell. \end{aligned} \tag{4.5}$$

On the other hand, using part (a), the relation $(a+b)^2 \leq 2a^2+2b^2$ for $a,b \in \mathbb{R}_+$, and the definition of $\tau_2(\cdot,\cdot)$ yields

$$\begin{aligned}\tau_2^2(c_\ell,\bar{q}^{\ell-1}) &\leq \frac{16\chi^2D_x^2}{\kappa_2^2}\left(\left[\kappa_0+\sqrt{\mathcal{N}_Ac_\ell}\right]\left[\Delta_\phi^{1/2}+\frac{6\kappa_3\sqrt{c_\ell}}{\chi}\right]+2\tilde{\kappa}_2\kappa_3c_\ell\right)^2 \\ &\leq \frac{16\chi^2D_x^2}{\kappa_2^2}\left(2\left[\kappa_0+\sqrt{\mathcal{N}_Ac_\ell}\right]^2\left[\Delta_\phi^{1/2}+\frac{6\kappa_3\sqrt{c_\ell}}{\chi}\right]^2+4\tilde{\kappa}_2^2\kappa_3^2c_\ell^2\right) \\ &\leq \frac{16\chi^2D_x^2}{\kappa_2^2}\left(8\left[\kappa_0^2+\mathcal{N}_Ac_\ell\right]\left[\Delta_\phi+\frac{36\kappa_3^2c_\ell}{\chi^2}\right]+4\tilde{\kappa}_2^2\kappa_3^2c_\ell^2\right)\end{aligned}$$

$$(4.6) \qquad = \frac{64\chi^2 D_x^2}{\kappa_2^2}\left(2\kappa_0^2\Delta_\phi + 2\left[\mathcal{N}_A\Delta_\phi + \frac{72\kappa_0^2\kappa_3^2}{\chi^2}\right]c_\ell + \left[\frac{72\mathcal{N}_A\kappa_3^2}{\chi^2} + \tilde{\kappa}_2^2\kappa_3^2\right]c_\ell^2\right).$$

Combining (4.5), (4.6), and the definitions of ξ_0, ξ_1, and ξ_2 yields the desired bound on $\tau_1(c_\ell, \bar{q}^{\ell-1}) + \tau_2(c_\ell, \bar{q}^{\ell-1})$.

(c) Suppose $c_\ell \geq \hat{c} := \hat{c}(\rho, \eta)$ and let $\varepsilon = \min\{\rho, \eta\}$. Moreover, notice from the definition of $\mathcal{T}_\ell(\cdot,\cdot)$ and the requirement that $\varepsilon \in (0,1)$ in Algorithm 2.2 that $c_\ell \geq \hat{c} \geq 1$. Now, for the sake of contradiction, suppose that the ℓ^{th} call of Algorithm 2.1 terminates in Step 2b unsuccessfully. It then follows from parts (b)–(c) of this proposition, the relations $c_\ell \geq 1$ and $\varepsilon \in (0,1)$ that:

$$T(\rho, \eta \,|\, c_\ell, \bar{q}^{\ell-1}) \leq \left(c_\ell^2 + \frac{c_\ell}{\varepsilon^2}\right)\mathcal{T}_1(1,1) = \left(c_\ell^2 + \frac{c_\ell}{\varepsilon^2}\right)\left(\frac{\varepsilon^2\hat{c}^2}{2}\right) \leq \left(c_\ell^2 + \frac{c_\ell}{\varepsilon^2}\right)\left(\frac{\varepsilon^2 c_\ell^2}{2}\right)$$
$$= \frac{1}{2}\left(\varepsilon^2 c_\ell^4 + c_\ell^3\right) \leq \frac{1}{2}\left(c_\ell^3 + c_\ell^3\right) \leq c_\ell^3.$$

In view of Proposition 2.1(c) with $(c, p^0) = (c_\ell, \bar{q}^{\ell-1})$, it follows that the ℓ^{th} call of Algorithm 2.1 must have terminated successfully, which is impossible due to our initial assumption. Hence, it must be the case that if $c_\ell \geq \hat{c}$ then the ℓ^{th} call of Algorithm 2.1 terminates successfully. ∎

5. Concluding Remarks.

The convergence of Algorithm 2.2 is established under the assumption that exact solutions to the subproblems in Step 1 of Algorithm 2.1 are easy to obtain. We believe that convergence can be also be established for when only inexact solutions, i.e.,

$$(5.1) \qquad x_t^k \approx \operatorname*{argmin}_{u_t \in \mathbb{R}^{n_t}} \left\{\lambda\mathcal{L}_c^\theta(x_{<t}^k, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2}\|u_t - x_t^{k-1}\|^2\right\}$$

are available. For example, one could consider applying an accelerated composite gradient (ACG) method to the problem associated with (5.1) so that x_t^k satisfies

$$\exists (r_t^k, \varepsilon_t^k) \quad \text{s.t.} \quad \begin{cases} r_k^t \in \partial_{\varepsilon_t^k}\left(\lambda\mathcal{L}_c^\theta(x_{<t}^k, \cdot, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2}\|\cdot - x_t^{k-1}\|^2\right)(x_t^k), \\ \|r_t^k\| + 2\varepsilon_t^k \leq \sigma^2\|r_t^k + x_t^{k-1} - x^k\|^2, \end{cases}$$

for some $\sigma \in (0,1)$, where $\partial_\varepsilon\psi(x) := \{v \in \mathbb{R}^n : \psi(y') \geq \psi(y) + \langle v, y' - y\rangle - \varepsilon, \forall y' \in \operatorname{dom}\psi\}$.

Appendix A. Proof of Lemma 3.2.

Before giving the proof, we present some auxiliary results. To avoid repetition, we assume the reader is already familiar with the quantities and notation in (3.1)–(3.3).

The proof of the first result can be found in [18, Lemma B.2].

Lemma A.1. *For any $(\chi, \theta) \in [0,1]^2$ satisfying $\zeta \leq \theta^2$ and $(a,b) \in \mathbb{R}^n \times \mathbb{R}^n$, we have that*

$$(A.1) \qquad \|a - (1-\theta)b\|^2 - \zeta\|a\|^2 \geq \left[\frac{(1-\zeta) - (1-\theta)^2}{2}\right]\left(\|a\|^2 - \|b\|^2\right).$$

The next result establishes some general bounds given by the updates in (1.5).

Lemma A.2. *For every $i \leq k$, $1 \leq t \leq B$, and $u_t \in x_t$, it holds that*

$$\lambda\left[\mathcal{L}_c^\theta(x_{<t}^i, u_t, x_{>t}^{i-1}; p^{i-1}) - \mathcal{L}_c^\theta(x_{\leq t}^i, x_{>t}^{i-1}; p^{i-1})\right] + \frac{1}{2}\|u_t - x_t^{i-1}\|^2$$
$$\geq \frac{1}{2}\|\Delta x_t^i\|^2 + \left(\frac{1-\lambda m_t}{2}\right)\|u_t - x_t^i\|^2 + \frac{\lambda c}{2}\|A_t(u_t - x_t^i)\|^2.$$

Proof. Let $i \leq k$, $1 \leq t \leq B$, and $u_t \in x_t$ be fixed, and define $\mu := 1 - \lambda m_t$ and $\|\cdot\|_\alpha^2 := \langle \cdot, (\mu I + \lambda c A_t^* A_t)(\cdot)\rangle$. Using the optimality of x_t^i and the fact that $\lambda \mathcal{L}_c^\theta(x_{<t}^i, \cdot, x_{>t}^{i-1}; p^{i-1}) + \|\cdot\|^2/2$ is μ-strongly convex with respect to $\|\cdot\|_\alpha^2$, it holds that

$$0 \in \partial \left[\lambda \mathcal{L}_c^\theta(x_{<t}^i, \cdot, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2}\|\cdot - x_t^{i-1}\|^2 - \frac{\mu}{2}\|\cdot - x_t^i\|_\alpha^2\right](x_t^i),$$

or, equivalently,

$$\begin{aligned}
&\lambda \mathcal{L}_c^\theta(x_{\leq t}^i, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2}\|\Delta x_t^i\|^2 \\
&\leq \lambda \mathcal{L}_c^\theta(x_{<t}^i, u_t, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2}\|u_t - x_t^{i-1}\|^2 - \frac{1}{2}\|u_t - x_t^i\|_\alpha^2 \\
&= \lambda \mathcal{L}_c^\theta(x_{<t}^i, u_t, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2}\|u_t - x_t^{i-1}\|^2 - \frac{\mu}{2}\|u_t - x_t^i\|^2 - \frac{\lambda c}{2}\|A_t(u_t - x_t^i)\|^2. \quad \square
\end{aligned}$$

We are now ready to give the proof of Lemma 3.2.

Proof of Lemma 3.2. (a) Using the definition of $\mathcal{L}_c^\theta(\cdot;\cdot)$ and Lemma 3.1(a), we conclude that

$$\begin{aligned}
\mathcal{L}_c^\theta(x^i; p^i) - \mathcal{L}_c^\theta(x^i; p^{i-1}) &= (1-\theta)\left\langle \Delta p^i, f^i \right\rangle = \left(\frac{1-\theta}{\chi c}\right)\|\Delta p^i\|^2 + \frac{a_\theta}{\chi c}\left\langle \Delta p^i, p^{i-1}\right\rangle \\
&= \left(\frac{1-\theta}{\chi c}\right)\|\Delta p^i\|^2 + \frac{a_\theta}{\chi c}\left(\left\langle p^i, p^{i-1}\right\rangle - \|p^{i-1}\|^2\right) \\
&= \left(\frac{1-\theta}{\chi c}\right)\|\Delta p^i\|^2 + \frac{a_\theta}{\chi c}\left(\frac{1}{2}\|p^i\|^2 - \frac{1}{2}\|\Delta p^i\|^2 - \frac{1}{2}\|p^{i-1}\|^2\right) \\
&= \frac{b_\theta}{2\chi c}\|\Delta p^i\|^2 + \frac{a_\theta}{2\chi c}\left(\|p^i\|^2 - \|p^{i-1}\|^2\right). \qquad \text{(A.2)}
\end{aligned}$$

(b) Using the fact that $1 > \lambda m/2$ and Lemma A.2 for $1 \leq t \leq B$ and $u = x_t^{i-1}$, we conclude that

$$\begin{aligned}
\left(1 - \frac{\lambda m}{2}\right)\|\Delta x^i\|^2 + \frac{\lambda c}{2}\sum_{t=1}^B \|A_t \Delta x_t^i\|^2 &\leq \sum_{i=1}^t \left(1 - \frac{\lambda m_t}{2}\right)\|\Delta x_t^i\|^2 + \frac{\lambda c}{2}\sum_{t=1}^B \|A_t \Delta x_t^i\|^2 \\
&\leq \lambda\left[\mathcal{L}_c^\theta(x^{i-1}; p^{i-1}) - \mathcal{L}_c^\theta(x^i; p^{i-1})\right],
\end{aligned}$$

which, in view of the fact that $\lambda = 1/(2m)$, implies the desired bound.

(c) We first use (2.6), the definition of γ_θ in (3.1), and Lemma A.1 with $(a, b, \zeta) = (\Delta p^i, \Delta p^{i-1}, 2B\chi b_\theta)$ to obtain

$$\|\Delta p^i - (1-\theta)\Delta p^{i-1}\|^2 \geq 2B\chi b_\theta \|\Delta p^i\|^2 + \chi\gamma_\theta\left(\|\Delta p^i\|^2 - \|\Delta p^{i-1}\|^2\right). \qquad \text{(A.3)}$$

Using (A.3) at i and $i-1$, Lemma 3.1(a), and the relation $\|a\|_1^2 \leq n\|a\|_2^2$ for $a \in \mathbb{R}^n$, we have that

$$\begin{aligned}
\frac{c}{4}\sum_{t=1}^B \|A_t \Delta x_t^i\|^2 \geq \frac{c}{4B}\|A\Delta x^i\|^2 &= \frac{\|\Delta p^i - (1-\theta)\Delta p^{i-1}\|^2}{4B\chi^2 c} \\
&\geq \frac{1}{4B\chi c}\left[2Bb_\theta\|\Delta p^i\|^2 + \gamma_\theta\left(\|\Delta p^i\|^2 - \|\Delta p^{i-1}\|^2\right)\right] \\
&= \frac{b_\theta}{2\chi c}\|\Delta p^i\|^2 + \frac{\gamma_\theta}{4B\chi c}\left(\|\Delta p^i\|^2 - \|\Delta p^{i-1}\|^2\right).
\end{aligned}$$

(d) Using Lemma 3.1, we first have that

$$
\begin{aligned}
2(1-\theta)\left\langle p^{i-1}, \chi c f^{i-1}\right\rangle + \|\chi c f^{i-1}\|^2 &= \|(1-\theta)p^{i-1} + \chi c f^{i-1}\|^2 - (1-\theta)\|p^{i-1}\|^2 \\
&= \|p^i\|^2 - (1-\theta)\|p^{i-1}\|^2 \le \|p^i\|^2 - \|p^{i-1}\|^2. \qquad \text{(A.4)}
\end{aligned}
$$

Now, using (A.4), parts (a)–(b), the relation $\|\Delta p^i\|^2 \le 2\|p^i\|^2 + 2\|p^{i-1}\|^2$, and the inclusions $a_\theta \in (0,1)$, $b_\theta \in (0,2)$, $\chi \in (0,1)$, and $\theta \in (0,1)$, we conclude that

$$
\begin{aligned}
\mathcal{L}_c^\theta(x^i; p^i) &\overset{(a)}{=} \mathcal{L}_c^\theta(x^i; p^{i-1}) + \frac{b_\theta\|\Delta p^i\|^2 + a_\theta\left[\|p^i\|^2 - \|p^{i-1}\|^2\right]}{2\chi c} \\
&\overset{(b)}{\le} \mathcal{L}_c^\theta(x^{i-1}; p^{i-1}) + \frac{b_\theta\|\Delta p^i\|^2 + a_\theta\|p^i\|^2}{2\chi c} \\
&\le \mathcal{L}_c^\theta(x^{i-1}; p^{i-1}) + \frac{2[\|\Delta p^i\|^2 + \|p^i\|^2]}{\chi c} \\
&= \phi(x^{i-1}) + (1-\theta)\left\langle p^{i-1}, f^{i-1}\right\rangle + \frac{c}{2}\|f^{i-1}\|^2 + \frac{2[\|\Delta p^i\|^2 + \|p^i\|^2]}{\chi c} \\
&\le \phi(x^{i-1}) + \frac{2(1-\theta)\left\langle p^{i-1}, \chi c f^{i-1}\right\rangle + \|\chi c f^{i-1}\|^2 + 4\|p^{i-1}\|^2 + 4\|p^i\|^2}{2\chi^2 c} \\
&\overset{(A.4)}{\le} \phi(x^{i-1}) + \frac{3\|p^{i-1}\|^2 + 5\|p^i\|^2}{2\chi^2 c} \le \phi(x^{i-1}) + \frac{3\left(\|p^{i-1}\|^2 + \|p^i\|^2\right)}{\chi^2 c}.
\end{aligned}
$$

(e) It holds that

$$
\begin{aligned}
\mathcal{L}_c^\theta(x^k; p^k) &= \phi(x^k) + (1-\theta)\left\langle p^k, Ax^k - d\right\rangle + \frac{c}{2}\|Ax - d\|^2 \\
&= \phi(x^k) + \frac{1}{2}\left\|\frac{(1-\theta)p^k}{\sqrt{c}} + \sqrt{c}(Ax^k - d)\right\|^2 - \frac{(1-\theta)^2\|p^k\|^2}{2c} \\
&\ge \phi(x^k) - \frac{(1-\theta)^2\|p^k\|^2}{2c} \ge \phi(x^k) - \frac{\|p^k\|^2}{2c}.
\end{aligned}
$$

□

REFERENCES

[1] D. P. Bertsekas, *Nonlinear programming*, Taylor & Francis, 3ed ed., 2016.

[2] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc, 2011.

[3] M. T. Chao, Y. Zhang, and J. B. Jian, *An inertial proximal alternating direction method of multipliers for nonconvex optimization*, International Journal of Computer Mathematics, (2020), pp. 1–19.

[4] J. Eckstein and D. P. Bertsekas, *On the douglas–rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318.

[5] J. Eckstein and M. C. Ferris, *Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control*, INFORMS Journal on Computing, 10 (1998), pp. 218–235.

[6] J. Eckstein and M. Fukushima, *Some reformulations and applications of the alternating direction method of multipliers*, in Large scale optimization, Springer, 1994, pp. 115–134.

[7] J. Eckstein and B. F. Svaiter, *A family of projective splitting methods for the sum of two maximal monotone operators*, Mathematical Programming, 111 (2008), pp. 173–199.

[8] J. Eckstein and B. F. Svaiter, *General projective splitting methods for sums of maximal monotone operators*, SIAM Journal on Control and Optimization, 48 (2009), pp. 787–811.

[9] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Studies in mathematics and its applications, vol. 15, Elsevier, 1983, pp. 299–331.
[10] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & mathematics with applications, 2 (1976), pp. 17–40.
[11] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires*, ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 9 (1975), pp. 41–76.
[12] M. L. N. Goncalves, J. G. Melo, and R. D. C. Monteiro, *Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems*, Pacific Journal of Optimization, 15 (2019), pp. 379–398.
[13] Z. Jia, J. Huang, and Z. Wu, *An incremental aggregated proximal ADMM for linearly constrained nonconvex optimization with application to sparse logistic regression problems*, Journal of Computational and Applied Mathematics, 390 (2021), p. 113384.
[14] B. Jiang, T. Lin, S. Ma, and S. Zhang, *Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis*, Computational Optimization and Applications, 72 (2019), pp. 115–157.
[15] W. Kong, *Accelerated inexact first-order methods for solving nonconvex composite optimization problems*, arXiv preprint arXiv:2104.09685, (2021).
[16] W. Kong, J. G. Melo, and R. D. C. Monteiro, *Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs*, SIAM Journal on Optimization, 29 (2019), pp. 2566–2593.
[17] W. Kong, J. G. Melo, and R. D. C. Monteiro, *An efficient adaptive accelerated inexact proximal point method for solving linearly constrained nonconvex composite problems*, Computational Optimization and Applications, 76 (2020), pp. 305–346.
[18] W. Kong and R. D. C. Monteiro, *An accelerated inexact dampened augmented Lagrangian method for linearly-constrained nonconvex composite optimization problems*, arXiv preprint arXiv:2110.11151, (2021).
[19] J. G. Melo and R. D. C. Monteiro, *Iteration-complexity of a Jacobi-type non-euclidean ADMM for multi-block linearly constrained nonconvex programs*, arXiv preprint arXiv:1705.07229, (2017).
[20] J. G. Melo and R. D. C. Monteiro, *Iteration-complexity of a linearized proximal multiblock ADMM class for linearly constrained nonconvex optimization problems*, Optimization Online preprint, (2017).
[21] J. G. Melo and R. D. C. Monteiro, *Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical lagrangian function and a full Lagrange multiplier update*, Available on arXiv:2008.00562, (2020).
[22] J. G. Melo, R. D. C. Monteiro, and W. Kong, *Iteration-complexity of an inner accelerated inexact proximal augmented lagrangian method based on the classical lagrangian function and a full lagrange multiplier update*, arXiv preprint arXiv:2008.00562, (2020).
[23] R. D. C. Monteiro and B. F. Svaiter, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM Journal on Optimization, 23 (2013), pp. 475–507.
[24] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of operations research, 1 (1976), pp. 97–116.
[25] A. Ruszczyński, *An augmented Lagrangian decomposition method for block diagonal linear programming problems*, Operations Research Letters, 8 (1989), pp. 287–294.
[26] K. Sun and A. Sun, *Dual descent ALM and ADMM*, arXiv preprint arXiv:2109.13214, (2021).
[27] A. Themelis and P. Patrinos, *Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results*, SIAM Journal on Optimization, 30 (2020), pp. 149–181.
[28] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, Journal of Scientific Computing, 78 (2019), pp. 29–63.
[29] J. Zhang and Z.-Q. Luo, *A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization*, SIAM Journal on Optimization, 30 (2020), pp. 2272–2302.