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 $O(\varepsilon^{-3})$ 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. 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:

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

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 $\text{Im}(A)$), and the following $B$-block structure is assumed:

$$
n = n_1 + \ldots + n_B, \quad x = (x_1, \ldots, x_B) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_B}
$$

$$
(1.2) \quad h(x) = \sum_{i=1}^{B} h_i(x_i), \quad Ax = \sum_{i=1}^{B} A_ix_i,
$$

where $\{A_i\}_{i=1}^{B}$ is another set of linear operators and $\{h_i\}_{i=1}^{B}$ 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 $x_1, \ldots, x_B$ sequentially or in parallel. However, the main drawbacks of existing ADMM-type methods include: (i) strong assumptions about the structure of $h$; (ii) iteration...
complexity bounds that scale poorly with the numerical tolerance; (iii) small step-
size parameters; or (iv) strong rank assumptions about the last block $A_B$, such as
$\text{Im}(A_B) \supset \{d\} \cup \text{Im}(A_1) \cup \ldots \cup \text{Im}(A_{B-1})$. Of the above drawbacks, the last block con-
dition in (iv) is especially limiting. For example, consider the popular multiblock
distributed finite-sum problem

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

where $f_t$ is continuously differentiable, $h_t$ is closed convex, and $\nabla f_t$ is Lipschitz con-
tinuous for $t = 1, \ldots, B$. It is easy to see 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 $A_B = -I \otimes I \in \mathbb{R}^{n(B-1) \times n}$, and $d = 0$. Moreover, it is straightforward to show
that for $s = 1, \ldots, B - 1$ we have $\text{Im}(A_s) \cap \text{Im}(A_B) = 0$ but $\text{Im}(A_s) \setminus \{0\} \neq \emptyset$, which
implies that $\text{Im}(A_s) \not\subseteq \text{Im}(A_B)$.

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

\[
L_{c_k}(x; p) := \phi(x) + (1 - \theta) \langle p, Ax - d \rangle + \frac{c_k}{2} \|Ax - d\|^2,
\]

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

\[
x_t^k = \arg\min_{u_t \in \mathbb{R}^{n_t}} \left\{ \lambda L_{c_k}(\ldots, x_{t-1}^k, u_t, x_{t+1}^{k-1}, \ldots; p^{k-1}) + \frac{1}{2} \|u_t - x_{t-1}^k\|^2 \right\},
\]

and then update

\[
p^k = (1 - \theta) p^{k-1} + \chi c_k (Ax^k - d),
\]

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

Contributions. For proper choices of the stepsize $\lambda$ and a non-decreasing sequence of
penalty parameters $\{c_k\}_{k \geq 1}$, it is shown that if the Slater-like condition\(^2\)

\[
\exists \tilde{z} \in \text{int} (\text{dom } h) \text{ such that } A\tilde{z} = d,
\]

holds, then DP-ADMM has the following features:

\begin{itemize}
  \item for any tolerance pair $(\rho, \eta) \in \mathbb{R}_+^2$, it obtains a pair $(\tilde{z}, \tilde{q})$ satisfying
  \begin{equation}
  \text{dist}(0, \nabla f(\tilde{z}) + A^*\tilde{q} + \partial h(\tilde{z})) \leq \rho, \quad \|A\tilde{z} - d\| \leq \eta
  \end{equation}
  in $O(\max\{\rho^{-3}, \eta^{-3}\})$ iterations;
  \item 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);
\end{itemize}

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

\(^2\)Here, int $S$ denotes the interior of a set $S$, dom $\psi$ denotes the domain of a function $\psi$, and $A^*$
is the adjoint of linear operator $A$.

This manuscript is for review purposes only.
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 \( \delta_S \) 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 presents a comparison between our proposed DP.ADMM and other ADMM-type methods for nonconvex and nonseparable problems, under a common tolerance \( \varepsilon \) given by \( \varepsilon := \min\{\rho, \eta\} \).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( \theta )</th>
<th>Complexity Assumptions Adaptive c</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADMM [26]</td>
<td>0</td>
<td>( \mathcal{O}(\varepsilon^{-6}) ) ( \mathcal{R}_0, \mathcal{K}\mathcal{L} ) No</td>
</tr>
<tr>
<td>LPADMM [29]</td>
<td>( (0, \infty) )</td>
<td>( \mathcal{P}, \mathcal{S} ) No</td>
</tr>
<tr>
<td>PADMM-m [14]</td>
<td>0</td>
<td>( \mathcal{O}(\varepsilon^{-4}) ) ( \mathcal{F} ) No</td>
</tr>
<tr>
<td>SDD-ADMM [26]</td>
<td>[0, 1] ( [-\theta \eta^2, 0] )</td>
<td>( \mathcal{O}(\varepsilon^{-3}) ) ( \mathcal{F} ) No</td>
</tr>
<tr>
<td>DP.ADMM</td>
<td>[0, 1] ( (0, \pi_\theta] )</td>
<td>( \mathcal{O}(\varepsilon^{-3}) ) ( \mathcal{S} ) Yes</td>
</tr>
</tbody>
</table>

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 \( \text{dom} h \). 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 \( \eta^2 \)) 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 \( \phi \). 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 $\phi$ in (1.1) is separable and has the same block structure as in (1.2), i.e., $\phi(x) =
\sum_{i=1}^{B} (f_i + h_t)(x_i)$ for closed (possibly) convex functions $h_t : \mathbb{R}^n \mapsto (-\infty, \infty]$ and
continuously differentiable functions $f_i : \text{dom } h_t \mapsto \mathbb{R}$. 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 $O(\varepsilon^{-2})$ 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 $h_2 \equiv \cdots \equiv h_B \equiv 0$.

Organisation. 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 $Q : \mathbb{R}^n \rightarrow \mathbb{R}^l$ is
denoted by $\sigma_Q^+$. For a given closed convex set $X \subset \mathbb{R}^n$, its boundary is denoted by
$\partial X$ and the distance of a point $x \in \mathbb{R}^n$ to $X$ is denoted by $\text{dist}_X(x)$. 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 $\varepsilon > 0$ and positive integer $b$, we denote $\log_b^+(\varepsilon) :=
\max\{1, [\log_b(\varepsilon)]\}$.

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

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

for every $z \in \mathbb{R}^n$. The classical subdifferential, denoted by $\partial h(\cdot)$, corresponds to $\partial_{\varepsilon} h(\cdot)$.
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 $\psi$ is a real-valued function which is differentiable at $\bar{z} \in \mathbb{R}^n$, then its affine approxima-
tion $\ell_\psi(\cdot, \bar{z})$ at $\bar{z}$ is given by

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

---

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

This manuscript is for review purposes only.
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., \( c_k \), while non-scalar quantities have this number appear as a superscript, e.g., \( v^k \), and \( p^k \). 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 \( \{H_t\}_{t=1}^\infty \subseteq \mathbb{R}^n \) be compact convex sets and denote the aggregated quantities

\[
H := \prod_{t=1}^B H_t, \quad x_{<t} := (x_1, \ldots, x_{t-1}),
\]

\[
x_{>t} := (x_{t+1}, \ldots, x_B), \quad x_{\leq t} := (x_{<t}, x_t), \quad x_{\geq t} := (x_t, x_{>t}),
\]

for every \( x = (x_1, \ldots, x_B) \in H \).

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) \( h_t \in \text{Conv} \ H_t \) for every \( 1 \leq t \leq B \);
(A2) \( A \neq 0 \) and \( F := \{x \in H : Ax = d\} \neq \emptyset \).

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

(A3) \( h \) is \( K_h \)-Lipschitz continuous on \( H \) for some \( K_h \geq 0 \);
(A4) \( f \) is continuously differentiable on \( H \) and, for every \( 1 \leq t \leq B \), there exists \((m_t, M_t) \in \mathbb{R}^{2+} \) such that

\[
\|\nabla x_t f(x_{<t}, \tilde{x}_{>t}) - \nabla x_t f(x)\| \leq M_t \|\tilde{x}_{>t} - x_{>t}\|,
\]

\[
-\frac{m_k}{2} \|\tilde{x}_t - x_t\|^2 \leq f(x_{<t}, \tilde{x}_t, x_{>t}) - f(x) - \langle \nabla x_t f(x), \tilde{x}_t - x_t \rangle,
\]

for every \( x, \tilde{x} \in H \);
(A5) there exists \( \tilde{z} \in F \) such that \( d_\circ := \text{dist}_{\partial H}(\tilde{z}) > 0 \).

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 \( f(x_{<t}, x_{>t}) + m_t \|\cdot\|^2/2 \) is convex on \( x_t \) for any \( x \in H \). Third, since \( H \) 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:

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

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 \( \tilde{z} \in \text{dom} \ h \) to be a local minimum of (1.1) is that there exists \( \tilde{q} \in \mathbb{R}^n \) such that

\[
0 \in \nabla f(\tilde{z}) + A^* \tilde{q} + \partial h(\tilde{z}), \quad A\tilde{z} = d.
\]
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 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 LCCO if and only if there exists a residual $\bar{v} \in \mathbb{R}^n$ such that

\begin{equation}
\bar{v} \in \nabla f(\bar{z}) + A^* \bar{q} + \partial h(\bar{z}), \quad \|\bar{v}\| \leq \rho, \quad \|A\bar{z} - d\| \leq \eta,
\end{equation}

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 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 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 $(\chi, \theta)$:

\begin{equation}
2 \chi B(2 - \theta)(1 - \theta) \leq \theta^2, \quad (\chi, \theta) \in (0, 1]^2.
\end{equation}

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{align*}
x^k_t & = \arg\min_{u \in \mathbb{R}^t} \left\{ \chi \mathcal{L}_0^k(x^k_{<t}, u_t, x^k_{>t}; p^{k-1}) + \frac{1}{2} \|u_t - x^k_{t-1}\|^2 \right\}, \\
p^k & = p^{k-1} + c \left( Ax^k - d \right),
\end{align*}

for $1 \leq t \leq B$ and a fixed penalty parameter $c \geq 1$. Consequently, the novelty of the method lies in the careful choice of $(\theta, \chi)$ 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{equation}
M := \max_{1 \leq t \leq B} M_t, \quad m := \min_{1 \leq t \leq B} m_t, \quad N_A := 8B^2 \sum_{t=1}^{B} \|A_t\|^2, \quad \Delta_0 := \bar{\varphi} - \varphi,*,
\end{equation}

\begin{align*}
\kappa_0 & := \frac{2B^2(2 + M + 2m)}{\sqrt{3m}}, \quad \kappa_1 := (K_h + G_f + B^2 [M + 2m] D_x) D_x, \\
\kappa_2 & := (\chi + \theta - \chi \theta) d_0 \sigma_A^+, \quad \kappa_3 := \frac{1}{\chi} \sup_{x \in H} \|Ax - d\|, \\
\kappa_4 & := (1 - \theta) + (1 - \theta)(1 - \chi) d_0 \sigma_A^+, \quad \kappa_5 := \frac{12}{\chi} \left( 1 + \frac{2\chi \kappa_1}{\kappa_2} \right),
\end{align*}

where $(G_f, D_x, \bar{\varphi}, \varphi,)$, $K_h$, 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: $\{m_t\} \subseteq \mathbb{R}_{++}$, $(\rho, \eta) \in (0, 1)^2$, $(\chi, \theta)$ as in (2.6)

1: $\lambda \leftarrow 1/(2 \min_t m_t)$
2: for $k \leftarrow 1, 2, \ldots$
   STEP 1 (prox update):
   3: for $t \leftarrow 1, 2, \ldots, B$
      4: $x^k_t \leftarrow \arg\min_{u_t \in \mathbb{R}^n} \{\lambda L^0_c(x^k_t, u_t, x_{>t}^{-1}; p^{k-1}) + \frac{\lambda}{2} \|u_t - x^k_t\|^2\}$
      5: $q^k \leftarrow (1 - \theta)p^{k-1} + c(Ax^k - d)$
      6: return $(x^k, q^k, v^k)$

STEP 2a (successful termination check):
7: for $t \leftarrow 1, 2, \ldots, B$
   8: $\delta^k_t \leftarrow \nabla f(x^k_t) - \nabla f(x^k_{<t}, x_{>t}^{-1})$
   9: $v^k_t \leftarrow \delta^k_t + cA^*_t \sum_{s=t+1}^B A_s(x^k_s - x^k_{<s}) - \frac{1}{\lambda}(x^k_t - x^k_{<1})$
10: if $\|v^k_t\| \leq \rho$ and $\|Ax^k - d\| \leq \eta$ then
   11: return $(x^k, q^k, v^k)$

STEP 2b (unsuccessful termination check):
12: if $k \equiv 0$ mod 3 and $k \geq 9$ then
13: $s^k_{(v)} \leftarrow \frac{2}{k/2} \sum_{i=k/2}^k \|v^i\|$ 
14: if $\frac{1}{\rho} s^k_{(v)} + \frac{1}{\eta} \sqrt{\frac{\kappa_0}{\chi}} s^k_{(f)} \leq 1$ then
15: return $(x^k, q^k, v^k)$

STEP 3 (multiplier update):
16: $p^k \leftarrow (1 - \theta)p^{k-1} + \chi c(Ax^k - d)$

Proposition 2.1. Let $(\kappa_3, \Delta_0, \mathcal{N}_A)$ and $D_x$ be as in (2.7) and (2.4), respectively,
and let $(\xi, \chi) \in \mathbb{R}^2_{++}$ and $p^0 \in A(\mathbb{R}^n)$ be given. Moreover, define

$$\bar{\kappa}_0 := 2 \left[ \Delta_0^{1/2} + \frac{10}{\chi} \sqrt{\frac{2}{\kappa_3}} \left( 1 + \frac{2\chi \kappa_1}{\kappa_2} \right) \right],$$

$$\bar{\kappa}_1 := \frac{6}{\chi} \sqrt{\mathcal{N}_A + \frac{\kappa_0}{\kappa_3}},$$

$$\tau_1(c, p^0) := \left( c \kappa_1 \right) \frac{\|p^0\|^2}{c} + \frac{\kappa_1}{\kappa_2} \|p^0\| + (2\kappa_3^2 + \kappa_3)c,$$

$$\tau_2(c, p^0) := 4\chi \frac{D_x}{\kappa_2} \left[ \left( \kappa_0 + \sqrt{\mathcal{N}_A} \right) \left( \Delta_0^{1/2} + \frac{6\kappa_3 \sqrt{\gamma}}{c} \right) + \bar{\kappa}_1 \|p^0\| \right],$$

$$T(p, \eta | c, p^0) := 48 \left[ 1 + \frac{2\bar{\kappa}_0 (\kappa_3 + \mathcal{N} c)}{\rho^2} \right] + \frac{\kappa_2^2 c}{\eta^2} + \tau_1(c, p^0) + \tau_2(c, p^0).$$

Then, for any $c \geq \xi$, the following statements hold about Algorithm 2.1 when it is

(a) it terminates in at most $T(p, \eta | c, p^0)$ 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 LCCO;

(c) if $(c, p^0)$ satisfies $T(p, \eta | c, p^0) \leq c^3$ 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
\[ T(\rho, \eta | c, p^0) = \Theta \left( c^2 + \frac{c}{\varepsilon^2} + \|p^0\| + \|p^0\|^2 \right). \]

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:** \( z^0 \in \mathcal{H}, \varepsilon > 0 \)

**Require:** \( \{m_t\} \subseteq \mathbb{R}_{++}, (\rho, \eta) \in (0,1]^2, (\chi, \theta) \) as in (2.6)

1. \((\hat{q}^0, c_1) \leftarrow (0, \varepsilon)\)
2. for \( \ell \leftarrow 1,2, \ldots \) do
3. call Algorithm 2.1 with inputs \((c, p^0, x^0) = (c_\ell, \hat{q}^{\ell-1}, \hat{z}^{\ell-1})\) and parameters \( \{m_t\}, (\rho, \eta) \), and \((\chi, \theta)\) to obtain an output triple \((\bar{z}^\ell, \hat{q}^\ell, \tilde{v}^\ell)\)
4. if \( \|\tilde{v}^\ell\| \leq \rho \) and \( \|A\bar{z}^\ell - d\| \leq \eta \) then
5. return \((\bar{z}^\ell, \hat{q}^\ell)\)
6. \(c_{\ell+1} \leftarrow 2c_\ell\)

In the results below, we give a uniform bound on \( \hat{q}^\ell/c_\ell \), 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:

\[
\begin{align*}
\xi_0 &:= \frac{128\chi^2 D_2^2 \Delta_\phi}{\kappa_3^2}, \quad \xi_2 := \frac{64\chi^2 D_2^2}{\kappa_3^2} \left[ \frac{72N_A\kappa_3^2}{\chi^2} + \kappa_2^2\kappa_3^2 \right], \\
\xi_1 &:= \frac{128\chi^2 D_2^2}{\kappa_2^2} \left[ N_A\Delta_\phi + \frac{72\kappa_2^2\kappa_3^2}{\chi^2} + \frac{8\kappa_4\kappa_3^2 + 2\kappa_4\kappa_3}{\kappa_2} + 2\kappa_3^2 + \kappa_3, \right]
\end{align*}
\]

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

**Proposition 2.2.** Let \((\kappa_1, N_A), \bar{\kappa}_i, \) and \( \xi_i \) be given by (2.7), (2.8), and (2.10), respectively, and define

\[
T_\ell(\rho, \eta) : = 48 \left[ \frac{1}{1 + \xi_0 + \xi_1 c_\ell + \xi_2 c^2_\ell + \frac{2\kappa_3^2(\kappa_0^2 + N_A c_\ell)}{\rho^2} + \frac{\kappa_2^2 c_\ell}{\eta^2}} \right],
\]

for every \( \ell \geq 1 \). Then, the following statements hold about the \( \ell \)th iteration of Algorithm 2.2:

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

(b) the \( \ell \)th call to Algorithm 2.1 terminates in at most

\[
(2.12) \quad T(\rho, \eta | c_\ell, \hat{q}^{\ell-1}) \leq T_\ell(\rho, \eta) \leq \left[ \max\{1, c_\ell^2\} + \frac{\max\{1, c_\ell\}}{\min\{\rho^2, \eta^2\}} \right] T_1(1,1)
\]

iterations of Algorithm 2.1, where \( T(\cdot, \cdot | \cdot, \cdot) \) is as in (2.9);
(c) if the \( \ell \)th penalty parameter \( c_\ell > 0 \) satisfies
\[
  c_\ell \geq \hat{c} (\rho, \eta) := \frac{\sqrt{2T_1(1,1)}}{\varepsilon}, \tag{2.13}
\]

Then the \( \ell \)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 \( T_\ell( \cdot, \cdot ) \) be as in (2.11), and define \( \varepsilon := \min \{ \rho, \eta \} \), \( E_0 := 32 \max \{ 4, \varepsilon^2 \} \), \( E_1 := 2 \log_2 (1/\varepsilon) \).

Then, Algorithm 2.2 stops and outputs a pair that solves Problem \( \text{LCCO} \) in at most
\[
  T_1(1,1) \cdot \left[ \frac{E_0 + E_1}{\varepsilon^2} + E_0 \right] \tag{2.14}
\]
iterations of Algorithm 2.1.

**Proof.** Let \( \hat{c}(\cdot, \cdot) \) be as in (2.13), and define the scalars
\[
  \ell := \log_2 \left( \frac{1}{\varepsilon} \right), \quad \hat{\ell} := \log_2 \left[ \hat{c}(\rho, \eta) / \varepsilon \right], \quad \varepsilon := \min \{ \rho, \eta \}, \quad \hat{\varepsilon} := \hat{c}(\rho, \eta).
\]

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 \( \text{LCCO} \) in at most \( \sum_{\ell=1}^{\hat{\ell}} T_\ell(\rho, \eta) \) iterations of Algorithm 2.1.

To bound this sum, we bound the following subsums: \( \sum_{\ell=1}^{\hat{\ell}} T_\ell(\rho, \eta) \) and \( \sum_{\ell=1}^{\hat{\ell}} T_\ell(\rho, \eta) \). For the first sum, let \( 1 \leq \ell < \hat{\ell} \). Since \( c_\ell < 1 \) (from the definition of \( \ell \)) and \( \varepsilon \leq 1 \), it follows from Proposition 2.2(b) that
\[
  \sum_{\ell=1}^{\hat{\ell}-1} T_\ell(\rho, \eta) \leq \sum_{\ell=1}^{\hat{\ell}-1} 2T_1(1,1) \frac{2^{\ell-1}}{\varepsilon^2} = \frac{2^{\hat{\ell}}T_1(1,1)}{\varepsilon^2} = T_1(1,1) \cdot E_1. \tag{2.15}
\]

For the second sum, let \( \ell \geq \hat{\ell} \). Similarly, since \( c_\ell \geq 1 \) and \( \varepsilon \leq 1 \) (from the definition of \( \hat{\ell} \)), it follows from Proposition 2.2(b) that
\[
  T_\ell(\rho, \eta) \leq \left( \frac{\hat{c}_\ell + c_\ell}{\varepsilon} \right) T_1(1,1). \tag{2.16}
\]

On the other hand, using the fact that \( \log_2 \hat{c} \geq 1 \), we have
\[
  \hat{\ell} - \ell = \log_2 \left[ \hat{c} / \varepsilon \right] - \log_2 \left[ \frac{1}{\varepsilon} \right] \leq \max \{ 1, \log_2 \left[ \hat{c} / \varepsilon \right] - \log_2 \left[ 1 / \varepsilon \right] + 1 \}
\]
\[
= 1 + \max \{ 0, \log_2 \hat{c} \} = 1 + \log_2 \hat{c} \tag{2.17}
\]

Using (2.16), (2.17), the fact that \( c_\ell = c_\ell 2^{\ell-\hat{\ell}} \), the bounds \( \log_2 \hat{c} \geq 1 \) and \( c_\ell \leq \max \{ 2, \varepsilon \} \) (see the update rule for \( c_\ell \) and the fact that \( \hat{\ell} \) is the first index where \( c_\ell \) 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
\[
  \sum_{\ell=1}^{\hat{\ell}} T_\ell(\rho, \eta) / T_1(1,1) \leq \sum_{\ell=1}^{\hat{\ell}} \left( \frac{\hat{c}_\ell^2 + c_\ell^2}{\varepsilon^2} \right) = \sum_{\ell=1}^{\hat{\ell}} \left( \frac{\hat{c}_\ell^2 + c_\ell^2}{\varepsilon^2} \right)
\]
\[ \leq 2^{iD+1}c^2 + \frac{2^iD+1}{\varepsilon^2} \leq 4 \left( \frac{2^{1+\log_2 D}c^2 + 2^{1+\log_2 D}c^2}{\varepsilon^2} \right) \]

(2.18) \[ \leq 16 \left( c^2 + \frac{c^2}{\varepsilon^2} \right) \leq 16 \left( \hat{c}^2 + \frac{\hat{c}^2}{\varepsilon^2} \right) \max\{4, \varepsilon^2\}. \]

Moreover, using Proposition 2.2(c) and the relation \( T_1(1,1) \geq \sqrt{T_1(1,1)} \), we have

\[ 16 \max\{4, c_i^2\} \cdot \left( \hat{c}^2 + \frac{\hat{c}^2}{\varepsilon^2} \right) \leq 32T_1(1,1) \cdot \max\{4, c_i^2\} \cdot (\varepsilon^{-2} + \varepsilon^{-3}) \leq T_1(1,1) \cdot E_0 \cdot (\varepsilon^{-2} + \varepsilon^{-3}) \).

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

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 \( \hat{c} \in (0, c] \) 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 \( k \geq 3 \). Moreover, for every \( i \geq 1 \) and \( (\chi, \theta) \in \mathbb{R}_+^2 \) satisfying (2.6), we make use of the following useful constants and shorthand notation

\[ a_\theta = \theta(1 - \theta), \quad b_\theta := (2 - \theta)(1 - \theta), \quad \gamma_\theta := \frac{(1 - 2B\chi b_\theta) - (1 - \theta)^2}{2\chi}, \]

(3.1) \[ f^i := Ax^i - d, \quad Q_i := \sum_{t=1}^B \sum_{s=t+1}^B \|A_t^*A_s\Delta x^i_s\|, \]

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

\[ S_{j,k}^{(p)} := \sum_{i=j}^k p^i_j, \quad S_{j,k}^{(v)} := \sum_{i=j}^k v^i_j, \quad S_{j,k}^{(f)} := \sum_{i=j}^k f^i_j, \]

(3.2) for every \( 1 \leq j \leq k \). We also denote \( \Delta y^j \) to be the difference of iterates for the variable \( y \) at iteration \( i \), i.e.,

\[ \Delta y^j \equiv y^i - y^{i-1}. \]

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 \( i \leq k \):

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

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

\[ \|v^i\| \leq B^2 (M + 2m) \|\Delta x^i\| + cQ_i, \]

(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 $x^k_i$ in Step 1 of Algorithm 2.1, assumption (A4), and the fact that $\lambda = 1/(2m)$ imply that

\[ 0 \in \partial \left[ \mathcal{L}^\theta_i(x^i_{<t}, \cdot, x^i_{>t}; p^{i-1}) + \frac{1}{2\lambda} \| -x^i_k \|^2 \right](x^i) \]

\[ = \nabla x_i f(x^i_{<t}, x^i_{>t}) + A^*_i \left[ (1-\theta)p^{i-1} + c\{A(x^i_{<t}, x^i_{>t}) - d\} \right] + \partial h_i(x^i) + \frac{1}{\lambda} \Delta x^i \]

\[ = \nabla x_i f(x^i_{<t}, x^i_{>t}) + A^*_i \left( q^i + c \sum_{s=t+1}^B A_s \Delta x^i_s \right) + \partial h_i(x^i) + \frac{1}{\lambda} \Delta x^i \]

\[ = \nabla x_i f(x^i) + A^*_i q^i + \partial h_i(x^i) - v^i. \]

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

\[ \| v^i \| \leq \| \nabla x_i f(x^i) - \nabla x_i f(x^i_{<t}, x^i_{>t}) \| + c \sum_{s=t+1}^B \| A^*_i A_s \Delta x^i_s \| + \frac{1}{\lambda} \| \Delta x^i \| \]

\[ \leq M_t \| x^i_{>t} - x^i_{<t} \| + c \sum_{s=t+1}^B \| A^*_i A_s \Delta x^i_s \| + 2m \| \Delta x^i \| \]

\[ \leq \sum_{s=t}^B (M_t + 2m) \| \Delta x^i_s \| + c \sum_{s=t+1}^B \| A^*_i A_s \Delta x^i_s \|. \]

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

\[ \| v^i \| \leq \sum_{t=1}^B \| v^i \| \leq \sum_{t=1}^B \sum_{s=t}^B (M_t + 2m) \| \Delta x^i_s \| + c \sum_{t=1}^B \sum_{s=t+1}^B \| A^*_i A_s \Delta x^i_s \| \]

\[ \leq (M + 2m) \sum_{t=1}^B \| \Delta x^i_s \| + c M \leq B^2 (M + 2m) \| \Delta x^i \| + c M. \]

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 $LCCO$ 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 $L^\theta_c(\cdot; \cdot)$ and the iterates $\{(x^i, p^i)\}_{i=1}^k$. For brevity, its proof is given in Appendix A.

Lemma 3.2. For every $i \leq k$, it holds that:

(a) $L^\theta_c(x^i; p^i) - L^\theta_c(x^i_{<t}; p^i_{<t}) = b_0 \| \Delta p^i \|^2 / (2\chi c) + a_0 \| p^i - p^{i-1} \|^2 / (2\chi c)$;

(b) $L^\theta_c(x^i; p^i_{<t}) - L^\theta_c(x^i_{<t}; p^i_{<t}) \leq -3m \| \Delta x^i \|^2 / 2 - c \sum_{t=1}^B \| A_t \Delta x^i_t \|^2 / 2$;

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

\[ \frac{b_0}{2\chi c} \| \Delta p^i \|^2 - \frac{c}{4} \sum_{t=1}^B \| A_t \Delta x^i_t \|^2 \leq \frac{\gamma_0}{4B\chi c} \left( \| \Delta p^i \|^2 - \| \Delta p^{i-1} \|^2 \right); \]
(d) $L^0_i(x^i;p^i) \leq \phi(x^{i-1}) + 3(\|p^i\|^2 + \|p^{i-1}\|^2)/(\chi^2 c)$;

(e) $L^0_i(x^i;p^i) \geq \phi(x^i) - \|p^i\|^2/(2c)$.

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

\[\Psi_i(c) := L^0_i(x^i;p^i) - \frac{a_0}{2\chi c} \|p^i\|^2 + \frac{\gamma_0}{4B\chi c} \|\Delta p^i\|^2 \quad \forall i \geq 1.\]

Then, for $1 \leq j \leq k$, it holds that

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

\[\leq \Delta_\phi + 4 \left( \frac{\|p^j\|^2 + \|p^{j-1}\|^2 + \|p^k\|^2}{2\chi c} \right).\]

**Proof.** Using the fact that $\|z\|^2 \leq n\|z\|^2$ for every $z \in \mathbb{R}^n$, the definition of $Q_i$ in (3.1), and the fact that $\|Mx\| \leq \|M\||x|_{\kappa}$ for any matrix $M$, we first have

\[cQ_i^2 \leq B^2 c \sum_{t=1}^{B} \sum_{s=t+1}^{B} \|A^*A_s \Delta x^i_s\|^2 \leq B^2 c \sum_{t=1}^{B} \|A_t\|^2 \sum_{s=t+1}^{B} \|A_s \Delta x^i_s\|^2 \]

\[\leq \left( B^2 \sum_{t=1}^{B} \|A_t\|^2 \right) \left( c \sum_{s=1}^{B} \|A_s \Delta x^i_s\|^2 \right) \]

\[= \left( 4B^2 \sum_{t=1}^{B} \|A_t\|^2 \right) \left( \frac{c}{4} \sum_{s=1}^{B} \|A_s \Delta x^i_s\|^2 \right).\]

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

\[\left[ \frac{B^2(M + 2m)\|\Delta x^i\| + cQ_i}{\kappa_0 + \sqrt{\mathcal{N}_A c}} \right]^2 \leq \frac{2B^4(M + 2m)^2\|\Delta x^i\|^2 + 2c^2Q_i^2}{\kappa_0^2 + \mathcal{N}_A c} \]

\[\leq \frac{3m}{2} \|\Delta x^i\|^2 + \frac{cQ_i^2}{4B^2 \sum_{t=1}^{B} \|A_t\|^2} \leq \frac{3m}{2} \|\Delta x^i\|^2 + \frac{c}{4} \sum_{t=1}^{B} \|A_t \Delta x^i_t\|^2 \]

\[\leq L^0_i(x^{i-1};p^{i-1}) - L^0_i(x^i;p^i) + \frac{a_0}{2\chi c} (\|p^j\|^2 - \|p^{j-1}\|^2) + \frac{b_0}{2\chi c} \|\Delta p^i\|^2 - \frac{\gamma_0}{4B\chi c} (\|\Delta p^{i-1}\|^2 - \|\Delta p^i\|^2) \]

\[= \Psi_i-1(c) - \Psi_i(c).\]

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 $(a+b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}_+$, and the bound $\gamma_0 \leq 1/(2\chi)$ to obtain

\[\Psi_j(c) - \Psi_k(c) \]
\[= \mathcal{L}_0^b(x^j; p^j) - \mathcal{L}_0^b(x^k; p^k) + \frac{a_0\|p^k\|^2 - \|p^j\|^2}{2\chi c} + \frac{\gamma_0(\|\Delta p^j\|^2 - \|\Delta p^k\|^2)}{4B\chi c} \]

\[\leq \mathcal{L}_0^b(x^j; p^j) - \mathcal{L}_0^b(x^k; p^k) + \frac{a_0\|p^j\|^2}{2\chi c} + \frac{\gamma_0\|\Delta p^j\|^2}{4B\chi c} \]

\[\leq \mathcal{L}_0^b(x^j; p^j) - \mathcal{L}_0^b(x^k; p^k) + \frac{\|p^j\|^2}{2\chi c} + \frac{\|p^j\|^2}{4B\chi^2 c} + \frac{\|p^j\|^2}{2c} \]

\[\leq \left[ \phi(x^{j-1}) - \phi(x^k) + \frac{3(\|p^j\|^2 + \|p^j\|^2)}{\chi^2 c} + \frac{\|p^k\|^2}{2c} \right] \]

\[\leq \frac{\|p^k\|^2}{2\chi c} + \frac{\|p^j\|^2}{2\chi c} + \frac{\|p^j\|^2}{4B\chi^2 c} \leq \Delta_\phi + 4 \left( \frac{\|p^j\|^2 + \|p^j\|^2 + \|p^k\|^2}{\chi^2 c} \right). \]

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

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

\[ S_{j+1,k}^{(f)} \leq \frac{\|p^j\| + 2S_{j+1,k}^{(p)}}{\chi c}, \]

\[ S_{j+1,k}^{(v)} \leq \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A}c)}{\sqrt{k-j}} \left( \Delta_{\phi}^{1/2} + \frac{\|p^j\| + \|p^j\| + \|p^k\|}{\chi c} \right). \]

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

\[ S_{j+1,k}^{(f)} = \sum_{i=j+1}^{k} \|p^i - (1 - \theta)p^{i-1}\| \leq \sum_{i=j+1}^{k} (\|p^{i-1}\| + \|p^i\|) \leq \frac{\|p^j\| + 2S_{j+1,k}^{(p)}}{\chi c}, \]

which is (3.9). On the other hand, to show (3.10), we use the fact that \( \|a\|_1 \leq \sqrt{n}\|a\|_2 \)

for \( a \in \mathbb{R}^n \), Lemma 3.1(b), Lemma 3.3, and the fact that \( \sqrt{a} + \sqrt{b} \leq \sqrt{a + b} \)

for \( a, b \in \mathbb{R}_+ \), to obtain

\[ S_{j+1,k}^{(v)} = \frac{\sum_{i=j+1}^{k} \|v^i\|}{k-j} \leq \left( \frac{\sum_{i=j+1}^{k} \|v^i\|^2}{k-j} \right)^{1/2} \]

\[ \leq \left( \frac{\sum_{i=j+1}^{k} \left[ B^2(M + 2m)\|\Delta x^i\| + c Q_i \right]^2}{k-j} \right)^{1/2} \]

\[ \leq \frac{\kappa_0 + \sqrt{\mathcal{N}_A}c}{\sqrt{k-j}} \left[ \Delta_{\phi}^{1/2} + \frac{\|p^j\|^2 + \|p^j\|^2 + \|p^k\|^2}{\chi^2 c} \right] \]

\[ \leq \frac{2(\kappa_0 + \sqrt{\mathcal{N}_A}c)}{\sqrt{k-j}} \left[ \Delta_{\phi}^{1/2} + \frac{\|p^j\|^2 + \|p^j\|^2 + \|p^k\|^2}{\chi c} \right]. \]

Now, observe that both residuals \( S_{j+1,k}^{(v)} \) and \( S_{j+1,k}^{(f)} \) depend on the size of the Lagrange multipliers. Since both termination conditions in Algorithm 2.1 require \( \|v^j\|, \|f^j\| \), 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 \( S_{j+1,k}^{(p)} = O(1) \).
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 $\sigma_S^+\|u\| \leq \|Su\|$. The proof of the next result can be found in [21, Lemma 4.7].

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

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

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

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

**Lemma 3.7.** Let $d_o, D_x, \kappa_i$ be as in (A5), (2.4), (2.7), respectively. Then, for every $i \geq 1$,

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

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

(c) it holds that

$$\frac{1}{c}\|q^i\|^2 + d_o \sigma_A^+\|q^i\| \leq \left(\frac{1 - \theta}{c}\right) \langle q^i, p^{i-1}\rangle + 2cD_x q_i + 2\kappa_1.$$

**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

$$\|p^i\| \leq (1 - \theta)\|p^{i-1}\| + \chi c\|Ax^i - d\|$$

$$\leq (1 - \theta)^i\|p^0\| + \chi c \sum_{j=0}^{i-1} (1 - \theta)^j\|Ax^j - d\|$$

$$\leq \|p^0\| + \chi c \sup_{x \in X} \|Ax - d\| \sum_{j=0}^{\infty} (1 - \theta)^j$$

$$= \|p^0\| + \chi c \left(\sup_{x \in X} \|Ax - d\|/\theta\right) = \|p^0\| + \kappa_3 c.$$

(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

$$\frac{1}{c}\|q^i\|^2 + d_o \sigma_A^+\|q^i\| \leq \frac{1}{c}\|q^i\|^2 + d_o \|A^*q^i\|$$

$$\leq \frac{1}{c}\|q^i\|^2 + d_o \left[\|v^i - \nabla f(x^i) - A^*q^i\| + \|\nabla f(x^i)\| + \|v^i\|\right]$$

$$\leq \frac{1}{c}\|q^i\|^2 + d_o \left[\|v^i - \nabla f(x^i) - A^*q^i\| + G_f + B^2(M + 2m) D_x + cQ_i\right]$$

$$\leq \frac{1}{c}\|q^i\|^2 + d_o \|v^i - \nabla f(x^i) - A^*q^i\| + cQ_iD_x + \kappa_i - K_h D_x.$$  

(3.11) We now derive a suitable bound on $d_o \|v^i - \nabla f(x^i) - A^*q^i\|$. 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 \((\psi, z, \bar{z}) = (h, x^i, \hat{x})\) and \(r = v^i - \nabla f(x^i) - A^* q^i\), it follows that

\[d_o \|v^i - \nabla f(x^i) - A^* q^i\| = \|v^i - \nabla f(x^i) - A^* q^i\| \text{dist}_{\partial H}(\hat{x})\]

\[\leq \left[\text{dist}_{\partial H}(\hat{x}) + \|x^i - \hat{x}\|\right] K_h + \left<v^i - \nabla f(x^i) - A^* q^i, x^i - \hat{x}\right>\]

\[(3.12) \leq 2K_h D_x + \left<v^i - \nabla f(x^i) - A^* q^i, x^i - \hat{x}\right>\].

On the other hand, Lemma 3.1(b), the Cauchy-Schwarz inequality, the definition of \(\kappa_1\), and the fact that \(Ax^i - d = [q^i - (1 - \theta)p^{i-1}] / c\) imply that

\[
\left<v^i - \nabla f(x^i) - A^* q^i, x^i - \hat{x}\right> \\
\leq (\|v^i\| + \|\nabla f(x^i)\|) \|x^i - \hat{x}\| - \left<q^i, Ax^i - d\right> \\
\leq \left[B^2 (M + 2m) D_x + cQ_i + G_j\right] D_x - \left<q^i, Ax^i - d\right> \\
= \kappa_1 - K_h D_x + cQ_i D_x + \left(1 - \frac{\theta}{c}\right) \left<q^i, p^{i-1}\right> - \frac{1}{c} \|q^i\|^2. \tag{3.13}
\]

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 \(e_{\sum_{i=j+1}^k Q_i}\).

**Lemma 3.8.** Let \(d_o, D_x, \kappa_i, \text{ and } \tau_i(\cdot, \cdot)\) be as in (A5), (2.4), (2.7), and (2.8), respectively, and define

\[d_\theta := \frac{2(1 - \theta)^2}{1 + \sqrt{1 + 4(1 - \theta)^2}} \quad e_\theta := (1 - \theta)(1 - \chi).
\]

Then, it holds that:

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

\[\kappa_2 \|p^i\| \leq 4\chi(1 + cQ_i D_x) + e_\theta d_o \sigma_{\chi}^+(\|p^{i-1}\| - \|p^i\|) + \frac{d_\theta(\|p^{i-1}\| - \|p^i\|)}{c};
\]

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

\[\frac{\epsilon \sum_{i=j+1}^k Q_i}{k - j} \leq \left[\frac{\kappa_2}{4\chi D_x}\right] \left[\frac{\tau_2(c, p^0)}{\sqrt{k - j}}\right].
\]

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

\[\nu_i(c) := \kappa_1 + cQ_i D_x, \quad g_\theta := \frac{1 + \sqrt{1 + 4(1 - \theta)^2}}{2(1 - \theta)},
\]

\[\Delta_{p,i}^{(1)} := \|p^i\| - \|p^{i-1}\|, \quad \Delta_{p,i}^{(2)} := \|p^i\|^2 - \|p^{i-1}\|^2.
\]

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

\[\frac{1}{c} \|p^i\|^2 + d_o \sigma_{\chi}^+ \|p^i\| = \frac{1}{c} \|\chi q^i + e_\theta p^{i-1}\| + d_o \sigma_{\chi}^+ \|\chi q^i + e_\theta p^{i-1}\|.
\]

This manuscript is for review purposes only.
\[ \leq 2 \chi \left[ \frac{1}{c} \|q_i\|^2 + d_\sigma \sigma_\Delta^+ \|q_i\|^2 \right] + \frac{2c^2}{c} \|p_i\| \|p_i\| + c_\theta d_\sigma \sigma_\Delta^+ \|p_i\| \|p_i\| \]

\[ \leq 2 \chi \left[ \frac{1 - \theta}{c} \langle q_i, p_i \rangle + 2 \nu_i(c) \right] + \frac{2c^2}{c} \|p_i\| \|p_i\| + e_\theta d_\sigma \sigma_\Delta^+ \|p_i\| \|p_i\| \]

\[ = 2 \left[ \frac{1 - \theta}{c} \langle p_i, p_i \rangle + 2 \nu_i(c) \right] + \frac{2c^2}{c} \|p_i\| \|p_i\| + e_\theta d_\sigma \sigma_\Delta^+ \|p_i\| \|p_i\| \]

\[ \leq \frac{2(1 - \theta)}{c} \langle p_i, p_i \rangle + \frac{2c_\theta (c_\theta - 1)}{c} \|p_i\|^2 + e_\theta d_\sigma \sigma_\Delta^+ \|p_i\| \|p_i\| + 4 \chi \nu_i(c) \]

(3.16) \[ e_{\theta \in (0,1)} \leq \frac{(1 - \theta) g_\theta}{c} \|p_i\|^2 \|p_i\| + \frac{1 - \theta}{c} \|p_i\|^2 \|p_i\| + e_\theta d_\sigma \sigma_\Delta^+ \|p_i\| \|p_i\| + 4 \chi \nu_i(c). \]

Subtracting \( e_\theta d_\sigma \sigma_\Delta^+ \|p_i\| + d_\theta \|p_i\|^2 + (1 - \theta) g_\theta/c \|p_i\|^2 \) from both sides and using the relations \( \kappa_2 = (1 - e_\theta) d_\sigma \sigma_\Delta^+ c, \) \( d_\theta = (1 - \theta)/g_\theta, \) and \( (1 - \theta) g_\theta + (1 - \theta) = 0, \) we conclude that

\[ 4 \chi (\kappa_1 + cQ_i Dz) - e_\theta d_\sigma \sigma_\Delta^+ \|p_i\| - \frac{d_\theta \Delta_p(1)}{c} \]

\[ \geq (1 - e_\theta) d_\sigma \sigma_\Delta^+ \|p_i\| + \frac{\|p_i\|^2}{c} [1 - d_\theta - (1 - \theta) g_\theta] \]

\[ = \kappa_2 \|p_i\|^2 - \frac{\|p_i\|^2}{g_\theta \cdot c} \left[(1 - \theta) g_\theta + (1 - \theta)\right] = \kappa_2 \|p_i\|^2 \]

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

\[ \sum_{i=j+1}^{k} \frac{cQ_i}{k-j} \leq \frac{(\sum_{i=j+1}^{k} c^2 Q_i^2)^{1/2}}{\sqrt{k-j}} \]

\[ \leq \frac{\kappa_0 + \sqrt{NAC}}{\sqrt{k-j}} \left[ \frac{\Delta_{1/2}^1 + 6 \|p_0\|^2 + 6 \kappa_3 \sqrt{c}}{\chi \sqrt{c}} \right] \]

\[ \leq \frac{\kappa_0 + \sqrt{NAC}}{\sqrt{k-j}} \left[ \frac{\Delta_{1/2}^1 + 6 \kappa_3 \sqrt{c}}{\chi \sqrt{c}} \right] + \frac{6 \|p_0\|^2}{\chi \sqrt{c} - j} \left[ \sqrt{N_A} + \kappa_0 \right] \]

\[ = \kappa_2 \left[ \frac{\kappa_0}{4 \chi Dz} \right] \left[ \frac{\tau_2(c, p_0)}{\sqrt{k-j}} \right]. \]

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

**Proposition 3.9.** Let \( \kappa_i \) and \( \tau_i \) be as in (2.7) and (2.8), respectively. Then, for every \( 1 \leq j < k, \) it holds that

(3.17) \[ S_{j+1,k}^{(p)} \leq \frac{4 \chi \kappa_1}{\kappa_2} + \frac{\tau_1(c, p_0)}{k-j} + \frac{\tau_2(c, p_0)}{\sqrt{k-j}}. \]

Moreover, if \( k \geq j + \tau_1(c, p_0) + \tau_2^2(c, p_0), \) then \( S_{j+1,k}^{(p)} \leq 2 + 4 \chi \kappa_1/\kappa_2. \)
Proof. Let \( \Delta^{(1)} \), \( \Delta^{(2)} \), \( d_9 \), and \( e_9 \) 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 \( d_9 \) is smaller than the first term in \( \kappa_4 \), it follows that

\[
\kappa_2 \sum_{i=j+1}^{k} \|p_i\| \leq \frac{d_9}{c} (\|p_i\|^2 - \|p_k\|^2) + e_9 d_2 \sigma^2 \|p_i\| - \|p_k\| + 4 \chi \sum_{i=j+1}^{k} \nu_i(c)
\]

\[
\leq \kappa_4 \left( \frac{\|p_i\|^2}{c} + \|p_i\| \right) + 4 \chi \sum_{i=j+1}^{k} \nu_i(c)
\]

\[
(3.18)
\]

\[
\leq \kappa_4 \left[ \frac{2\|p_i\|^2}{c} + \|p_i\| + (2\kappa_3 + \kappa_3) c \right] + 4 \chi \sum_{i=j+1}^{k} \nu_i(c).
\]

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

\[
S_{j+1,k}^{(p)} \leq \frac{\kappa_4}{\kappa_2(k-j)} \left[ \frac{2\|p_i\|^2}{c} + \|p_i\| + (2\kappa_3 + \kappa_3) c \right] + \frac{4 \chi \sum_{i=j+1}^{k} \nu_i(c)}{\kappa_2(k-j)}
\]

\[
= \frac{4 \chi \kappa_1}{\kappa_2} + \frac{\tau_1(c,p_i)}{k-j} + \frac{4 \chi D_x \sum_{i=j+1}^{k} c Q_i}{\kappa_2(k-j)}
\]

\[
\leq \frac{4 \chi \kappa_1}{\kappa_2} + \frac{\tau_1(c,p_i)}{k-j} + \frac{\tau_2(c,p_i)}{\sqrt{k-j}},
\]

which is exactly (3.17). The last statement of the proposition follows immediately from the fact that \( k \geq \tau_1(c,p_i) + \tau_2(c,p_i) \) implies \( k-j \geq \tau_1(c,p_i) + \sqrt{k-j} \). \( \Box \)

We end this subsection by discussing some implications of the above results. Suppose \( \zeta \) is an integer satisfying \( \zeta \geq 1 + \tau_1(c,p_i) + \tau_2(c,p_i) = \Omega(c^2 + \|p_i\|^2) \). It then follows from Proposition 3.9 that \( S_{2,2}^{(p)} = O(1) \) and \( S_{3,3}^{(p)} = O(1) \). Since the minimum of a set of scalars minorizes the average of these scalars, there exists indices \( j_0 \in \{2,\ldots,\zeta\} \) and \( k_0 \in \{\zeta,\ldots,3\zeta\} \) such that \( \|p_j\| = O(1) \) and \( \|p_\zeta\| = O(1) \). Using the fact that \( k_0 - j_0 \geq \zeta \), the above bounds, and (3.9)–(3.10), it is reasonable to expect \( S_{j_0+1,k_0}^{(f)} = O(1/c) \) and \( S_{j_0+1,k_0}^{(v)} = O(\zeta_0/c) \). 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 \( (\kappa_i,N_\Delta) \) and \( (\tilde{\kappa}_i,\tau_i) \) be as in (2.7) and (2.8), respectively, and suppose \( \zeta \in \mathbb{N} \) satisfies \( \zeta \geq 1 + \tau_1(c,p_i) + \tau_2(c,p_i) \). Then, there exists \( j_0 \in \{3,\ldots,\zeta\} \) and \( k_0 \in \{\zeta,\ldots,3\zeta\} \) such that

\[
(4.1)
\]

\[
S_{j_0,k_0}^{(v)} \leq \frac{\tilde{\kappa}_0(\kappa_0 + \sqrt{N_\Delta} \zeta)}{\sqrt{\zeta}} \leq \frac{\kappa_3}{c}.
\]

**Proof.** Suppose \( \zeta \in \mathbb{N} \) satisfies \( \zeta \geq 1 + \tau_1(c,p_i) + \tau_2(c,p_i) \). Using Proposition 3.9 with \( (j,k) = (1,\zeta) \) it holds that there exists \( 3 \leq j_0 \leq k \) such that

\[
\|p_i^{j_0-1}\| + \|p_i^k\| \leq \sum_{i=3}^{\zeta} \|p_i^{j_0-1}\| + \|p_i^k\| \leq \frac{2}{\zeta-2} \sum_{i=2}^{\zeta} \|p_i^k\| \leq \frac{2}{\zeta-2} \sum_{i=2}^{\zeta} \|p_i^k\|.
\]
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

\begin{equation}
\|p^{k_0}\| \leq \frac{\sum_{i=2\zeta+1}^{3\zeta} \|p^i\|}{\zeta} = S_{2\zeta+1,3\zeta} \leq 2 + \frac{4\chi k_1}{k_2}.
\end{equation}

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

\begin{equation}
S_{j_0+1,k_0}^{(\varepsilon)} \leq \frac{2(\kappa_0 + \sqrt{N_A}c)}{\sqrt{k_0 - j_0}} \left( \frac{\Delta_\phi^{1/2} + \frac{10}{\chi \sqrt{c}} \left( 1 + \frac{2\chi k_1}{k_2} \right)}{\chi c} \right)
\end{equation}

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

\begin{equation}
S_{j_0+1,k_0}^{(f)} \leq \frac{\|p^{j_0}\| + 2S_{j_0+1,k_0}^{(p)}}{\chi c} \leq \frac{12}{\chi c} \left( 1 + \frac{2\chi k_1}{k_2} \right) = \frac{\kappa_5}{c}.
\end{equation}

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

\[ T := T(\rho,\eta | c, p^0), \quad r_j := S_j^{(\varepsilon)} + S_j^{(f)} \quad \forall j \geq 1, \]

where $S_j^{(\varepsilon)}$ and $S_j^{(f)}$ 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 $(a+b)^2 \leq 2a^2 + 2b^2$ 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{equation}
\frac{S_{j_0,k_0}^{(\varepsilon)}}{\rho} + \frac{c^3/2S_{j_0,k_0}^{(f)}}{\eta \sqrt{T/3}} \leq \frac{\kappa_0(\kappa_0 + \sqrt{N_A}c)}{\rho \sqrt{T/3}} + \frac{\kappa_5 \sqrt{c}}{\eta \sqrt{T/3}}
\end{equation}

\begin{equation}
= \sqrt{\frac{3\kappa_0^2(\kappa_0 + \sqrt{N_A}c)^2}{\rho^2 T}} + \frac{3\kappa_5^2 c}{\eta^2 T} \leq \sqrt{\frac{6\kappa_0^2(\kappa_0 + \sqrt{N_A}c)^2}{\rho^2 T}} + \frac{1}{4} \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.
\end{equation}

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

\begin{equation}
r_{k_0} = \frac{S_{k_0/2,k_0}^{(\varepsilon)}}{\rho} + \frac{c^3/2S_{k_0/2,k_0}^{(f)}}{\eta \sqrt{k_0}} \leq \frac{k_0 - j_0 + 1}{k_0 - k_0/2 + 1} \left[ \frac{S_{j_0,k_0}^{(\varepsilon)}}{\rho} + \frac{c^3/2S_{j_0,k_0}^{(f)}}{\eta \sqrt{T/3}} \right] \]

This manuscript is for review purposes only.
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

\[
\left\| v^i \right\| + \frac{c^{3/2} \left\| f^i \right\|}{\eta \sqrt{j}} \leq \frac{S_{j/2,j}^{(v)}}{\rho} + \frac{c^{3/2} S_{j/2,j}^{(f)}}{\eta \sqrt{j}} \leq 1.
\]

Hence, it holds that \( \left\| v^i \right\| \leq \rho \) and, from our initial bound on \( T \), we have \( \left\| f^i \right\| \leq \eta \sqrt{Tc^{-3/2}} \leq \eta \sqrt{T}c^{-3/2} \leq \eta \). 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 \((\bar{q}^0, c) = (\bar{q}^{\ell - 1}, c_\ell)\) and the relation \( c_\ell = 2c_{\ell - 1} \) that

\[
\left\| q^\ell \right\| \leq \left\| q^{\ell - 1} \right\| + \kappa_3 c_\ell = \kappa_3(2c_{\ell - 1} + c_\ell) = 2\kappa_3 c_\ell.
\]

(b) The fact that the iteration count is bounded by \( T(\rho, \eta \mid c_\ell, \bar{q}^{\ell - 1}) \) 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 \( T_\ell(\rho, \eta) \) 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 \( \tau_1(c_\ell, \bar{q}^{\ell - 1}) + \tau_2(c_\ell, \bar{q}^{\ell - 1}) \leq \bar{\xi} c_\ell + \xi_2 c_\ell^2 \). Using part (a) and the definition of \( \tau_1(\cdot, \cdot) \), we first have that

\[
\tau_1(c_\ell, \bar{q}^{\ell - 1}) \leq \left( \frac{2\kappa_4}{\kappa_2} \right) \left( \frac{4\kappa_3^2 c_\ell^2}{c_\ell} + \frac{2\kappa_4 \kappa_3 c_\ell}{\kappa_2} + (2\kappa_3^2 + \kappa_3) c_\ell \right)
\]

\[
= \left( \frac{8\kappa_4 \kappa_3^2 + 2\kappa_4 \kappa_3}{\kappa_2} + \frac{2\kappa_3^2 + \kappa_3}{\kappa_2} \right) c_\ell.
\]

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

\[
\tau_2^2(c_\ell, \bar{q}^{\ell - 1}) \leq \frac{16\chi_2^2 D_\ell^2}{\kappa_2^2} \left( \left[ \kappa_0 + \sqrt{N_A} c_\ell \right] \left[ \Delta_\phi^{1/2} + \frac{6\kappa_3 \sqrt{c_\ell \chi}}{\chi} \right] + 2\kappa_2 \kappa_3 c_\ell \right)^2
\]

\[
\leq \frac{16\chi_2^2 D_\ell^2}{\kappa_2^2} \left( 2 \left[ \kappa_0 + \sqrt{N_A} c_\ell \right]^2 \left[ \Delta_\phi^{1/2} + \frac{6\kappa_3 \sqrt{c_\ell \chi}}{\chi} \right]^2 + 4\kappa_2^2 \kappa_3^2 c_\ell^2 \right)
\]

\[
\leq \frac{16\chi_2^2 D_\ell^2}{\kappa_2^2} \left( 8 \left[ \kappa_0^2 + N_A c_\ell \right] \left[ \Delta_\phi + \frac{36\kappa_3^2 c_\ell}{\chi^2} \right] + 4\kappa_2^2 \kappa_3^2 c_\ell^2 \right)
\]
Combining (4.5), (4.6), and the definitions of \( \xi_0, \xi_1, \) and \( \xi_2 \), yields the desired bound on \( \tau_1(c_t, q^{t-1}) + \tau_2(c_t, q^{t-1}) \).

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

\[
T(\rho, \eta | c_t, q^{t-1}) = \left( c_t^2 + \frac{c_t}{\varepsilon^2} \right) \mathcal{T}_t(1, 1) = \left( c_t^2 + \frac{c_t}{\varepsilon^2} \right) \left( \frac{\varepsilon^2 c_t^2}{2} \right) \leq \left( c_t^2 + \frac{c_t}{\varepsilon^2} \right) \left( \frac{\varepsilon^2 c_t^2}{2} \right)
\]

\[
= \frac{1}{2} (c_t^2 + c_t) \leq \frac{1}{2} (c_t^3 + c_t) \leq c_t^3.
\]

In view of Proposition 2.1(c) with \( (c, p^0) = (c_t, q^{t-1}) \), it follows that the \( \ell \)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 \( c_t \geq \hat{c} \) then the \( \ell \)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 available. For example, one could consider applying an accelerated composite gradient (ACG) method to the problem associated with (5.1) so that only inexact solutions, i.e.,

\[
(x_t^k) \approx \arg \min_{u_t \in \mathbb{R}^n} \left\{ \lambda \mathcal{L}_c^0(x_{<t}, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|u_t - x_t^{k-1}\|^2 \right\}
\]

are easy to obtain. We believe that convergence can be also be established for when inexact solutions, i.e.,

\[
(x_t^k) \approx \arg \min_{u_t \in \mathbb{R}^n} \left\{ \lambda \mathcal{L}_c^0(x_{<t}, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|u_t - x_t^{k-1}\|^2 \right\}
\]

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

\[
\exists (r_t^k, \varepsilon_t^k) \quad \text{s.t.} \quad \left\{ \begin{array}{l}
 r_t^k \in \partial \varepsilon_t^k \left( \lambda \mathcal{L}_c^0(x_{<t}, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|x_t^{k-1}\|^2 \right) (x_t^k), \\
 \|r_t^k\| + 2\varepsilon_t^k \leq \sigma \left\| r_t^k + x_t^{k-1} - x_t^{k-2} \right\|^2,
\end{array} \right.
\]

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

### 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

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

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

\[
\lambda \left[ \mathcal{L}_c^0(x_{<t}^{i-1}, u_t, x_{>t}^{i-1}; p^{i-1}) - \mathcal{L}_c^0(x_{<t}^{i-1}, x_{>t}^{i-1}; p^{i-1}) \right] + \frac{1}{2} \|u_t - x_t^{i-1}\|^2 \\
\geq \frac{1}{2} \|\Delta x_t^{i}\|^2 + \left( \frac{1 - \lambda m_t}{2} \right) \|u_t - x_t^i\|^2 + \frac{\lambda c}{2} \|A_t(u_t - x_t^i)\|^2.
\]
Proof. Let \( i \leq k, 1 \leq t \leq B, \) and \( u_t \in \mathcal{F}_t \) be fixed, and define \( \mu := 1 - \lambda m_i \) and \( \| \cdot \|_a^2 := \langle \cdot, (\mu I + \lambda c A_i^* A_i) \cdot \rangle \). Using the optimality of \( x^i_t \) and the fact that \( \lambda L_c^0(x^i_{t,s}, x^i_{s+1}; p^i-1) + \| \cdot \|^2/2 \) is \( \mu \)-strongly convex with respect to \( \| \cdot \|_a^2 \), it holds that
\[
0 \in \partial \left[ \lambda L_c^0(x^i_{t<s}, x^i_{s+1}; p^i-1) + \frac{1}{2} \| \cdot - x^i_{t-1} \|^2 - \frac{\mu}{2} \| \cdot - x^i_t \|^2 \right] (x^i_t),
\]
or, equivalently,
\[
\lambda L_c^0(x_{t<s}, x^i_{s+1}; p^i-1) + \frac{1}{2} \| \Delta x^i_t \|^2
\]
\[
\leq \lambda L_c^\theta(x_{t<s}, x^i_{s+1}; p^i-1) + \frac{1}{2} \| u_t - x^i_{t-1} \|^2 - \frac{1}{2} \| u_t - x^i_t \|^2
\]
\[
= \lambda L_c^\theta(x_{t<s}, x^i_{s+1}; p^i-1) + \frac{1}{2} \| u_t - x^i_{t-1} \|^2 - \frac{\mu}{2} \| u_t - x^i_t \|^2 - \frac{\lambda c}{2} \| A_i(u_t - x^i_t) \|^2. \quad \Box
\]

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

Proof of Lemma 3.2. (a) Using the definition of \( L_c^0(\cdot; \cdot) \) and Lemma 3.1(a), we conclude that
\[
L_c^\theta(x^i; p^i) - L_c^\theta(x^i; p^i-1) = (1 - \theta) \langle \Delta p^i, f^i \rangle = \left( \frac{1 - \theta}{\chi c} \right) \| \Delta p^i \|^2 + \frac{a_\theta}{\chi c} \| \Delta p^i - p^i-1 \|^2
\]
\[
= \left( \frac{1 - \theta}{\chi c} \right) \| \Delta p^i \|^2 + \frac{a_\theta}{\chi c} \left( \frac{1}{2} \| p^i \|^2 - \| \Delta p^i \|^2 \right)
\]
\[
= \frac{b_\theta}{2 \chi c} \| \Delta p^i \|^2 + \frac{a_\theta}{\chi c} \| p^i \|^2.
\]

(b) Using the fact that \( 1 > \lambda m/2 \) and Lemma A.2 for \( 1 \leq t \leq B \) and \( u = x^i_t-1 \), we conclude that
\[
(1 - \lambda m/2) \| \Delta p^i \|^2 + \frac{\lambda c}{2} \sum_{i=1}^{B} \| A_i \Delta x^i_t \|^2 \leq \sum_{i=1}^{B} \left( 1 - \frac{\lambda m_i}{2} \right) \| \Delta x^i_t \|^2 + \frac{\lambda c}{2} \sum_{i=1}^{B} \| A_i \Delta x^i_t \|^2
\]
\[
\leq \lambda \left[ L_c^\theta(x^i_t; p^i-1) - L_c^\theta(x^i; p^i-1) \right],
\]
which, in view of the fact that \( \lambda = 1/(2m) \), implies the desired bound.

(c) We first use (2.6), the definition of \( \gamma_\theta \) in (3.1), and Lemma A.1 with \( (a, b, \zeta) = (\Delta p^i, \Delta p^i-1, 2B\chi b_\theta) \) to obtain
\[
\| \Delta p^i - (1 - \theta) \Delta p^i-1 \|^2 \geq 2B\chi b_\theta \| \Delta p^i \|^2 + \chi \gamma_\theta \left( \| \Delta p^i \|^2 - \| \Delta p^i-1 \|^2 \right).
\]

Using (A.3) at \( i \) and \( i-1 \), Lemma 3.1(a), and the relation \( \| a \|^2 \leq n\| a \|^2 \) for \( a \in \mathbb{R}^n \), we have that
\[
\frac{c}{4} \sum_{i=1}^{B} \| A_i \Delta x^i_t \|^2 \geq \frac{c}{4B} \| A \Delta x^i_t \|^2 = \frac{\| \Delta p^i - (1 - \theta) \Delta p^i-1 \|^2}{4B\chi^2 c}
\]
\[
\geq \frac{1}{4B\chi c} \left[ 2B\chi b_\theta \| \Delta p^i \|^2 + \gamma_\theta \left( \| \Delta p^i \|^2 - \| \Delta p^i-1 \|^2 \right) \right]
\]
\[
= \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).
(d) Using Lemma 3.1, we first have that
\[ 2(1 - \theta) \langle p^{i-1}, \chi f^{i-1} \rangle + \| \chi f^{i-1} \|^2 = \| (1 - \theta)p^{i-1} + \chi f^{i-1} \|^2 - (1 - \theta)\|p^{i-1}\|^2 \]
\[ \tag{A.4} \leq \|p^i\|^2 - (1 - \theta)\|p^{i-1}\|^2 \leq \|p^i\|^2 - \|p^{i-1}\|^2. \]

Now, using (A.4), parts (a)–(b), the relation \(\|\Delta p^i\|^2 \leq 2\|p^i\|^2 + 2\|p^{i-1}\|^2\), and the inclusions \(a_0 \in (0, 1), b_0 \in (0, 2), \chi \in (0, 1), \) and \(\theta \in (0, 1)\), we conclude that
\[ \mathcal{L}^\theta_c(x^i; p^i) \overset{(a)}{=} \mathcal{L}^\theta_c(x^i; p^{i-1}) + \frac{b_0\|\Delta p^i\|^2 + a_0 \|p^i\|^2}{2\chi c} \]
\[ \overset{(b)}{\leq} \mathcal{L}^\theta_c(x^i; p^{i-1}) + \frac{b_0\|\Delta p^i\|^2 + a_0 \|p^i\|^2}{2\chi c} \]
\[ \leq \mathcal{L}^\theta_c(x^i; p^{i-1}) + \frac{2\|\Delta p^i\|^2 + \|p^i\|^2}{\chi c} \]
\[ = \phi(x^{i-1}) + (1 - \theta) \langle p^{i-1}, f^{i-1} \rangle + \frac{c}{2} \|f^{i-1}\|^2 + \frac{2\|\Delta p^i\|^2 + \|p^i\|^2}{\chi c} \]
\[ \leq \phi(x^{i-1}) + \frac{2(1 - \theta)\langle p^{i-1}, \chi f^{i-1} \rangle + \|\chi f^{i-1}\|^2 + 4\|p^{i-1}\|^2 + 4\|p^i\|^2}{2\chi^2 c} \]
\[ \overset{(A.4)}{\leq} \phi(x^{i-1}) + \frac{3\|p^{i-1}\|^2 + 5\|p^i\|^2}{2\chi^2 c} \leq \phi(x^{i-1}) + \frac{3\left( \|p^{i-1}\|^2 + \|p^i\|^2 \right)}{\chi^2 c}. \]

(e) It holds that
\[ \mathcal{L}^\theta_c(x^k; p^k) = \phi(x^k) + (1 - \theta) \langle p^k, Ax^k - d \rangle + \frac{c}{2} \|Ax - d\|^2 \]
\[ = \phi(x^k) + \frac{1}{2} \left( \frac{(1 - \theta)p^k}{\sqrt{c}} + \sqrt{c}(Ax^k - d) \right)^2 - \frac{(1 - \theta)^2\|p^k\|^2}{2c} \]
\[ \geq \phi(x^k) - \frac{(1 - \theta)^2\|p^k\|^2}{2c} \geq \phi(x^k) - \frac{\|p^k\|^2}{2c}. \]

REFERENCES


This manuscript is for review purposes only.


