Global Complexity Bound of a Proximal ADMM for Linearly-Constrained Nonseperable Nonconvex Composite Programming*

Weiwei Kong† and Renato D.C. Monteiro‡

December 8, 2021

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: (i) 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 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.

1 Introduction

This paper presents a dampened proximal alternating direction method of multipliers (DP.ADMM) for finding approximate stationary points of the Γ-block nonseparable nonconvex composite optimization problem

$$\min_{(x^1,\ldots,x^\Gamma)\in\mathbb{R}^{n_1}\times\cdots\times\mathbb{R}^{n_\Gamma}} \left\{\phi(x^1,\ldots,x^\Gamma) := f(x^1,\ldots,x^\Gamma) + \sum_{t=1}^{\Gamma} h_t(x^t) : \sum_{t=1}^{\Gamma} A_t x^t = d\right\}, \tag{1}$$

where $\{A_t\}_{t=1}^{\Gamma}$ is a set of linear operators, $\{h_t\}_{t=1}^{\Gamma}$ is a set of proper closed convex functions with compact domains, and f is a (possibly) nonconvex differentiable function on the domain of $(x^1,\ldots,x^\Gamma) \mapsto \sum_{t=1}^{\Gamma} h_t(x^t)$ with a Lipschitz continuous gradient. The main idea of the method is to apply a fully proximal ADMM-type method to a sequence of penalty subproblems

$$\min_{(x^1,\ldots,x^\Gamma)\in\mathbb{R}^{n_1}\times\cdots\times\mathbb{R}^{n_\Gamma}} \left\{\phi(x^1,\ldots,x^\Gamma) + \frac{c_\ell}{2}\left\|\sum_{t=1}^{\Gamma} A_t x^t - d\right\|^2\right\} \tag{2}$$

*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).

†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

where $\{c_\ell\}_{\ell\geq 1}$ is a strictly increasing sequence of positive scalars.

The proximal ADMM iteration scheme for the $\ell^{\rm th}$ subproblem is based on the dampened augmented Lagrangian (AL) function

$$\mathcal{L}^\theta_{c_\ell}(x^1,\ldots,x^\Gamma;p) := \phi(x^1,\ldots,x^\Gamma) + (1-\theta)\left\langle p, \sum_{t=1}^\Gamma A_t x^t - d\right\rangle + \frac{c_\ell}{2}\left\|\sum_{t=1}^\Gamma A_t x^t - d\right\|^2, \tag{3}$$

for $\theta\in(0,1)$, and it consists of the following updates at the $k^{\rm th}$ iteration: given $(x^1_{k-1},\ldots,x^\Gamma_{k-1},p_{k-1})$, c_ℓ, and λ, compute

$$x^t_k = \underset{u\in\mathbb{R}^{n_t}}{\operatorname{argmin}}\left\{\lambda\mathcal{L}^\theta_{c_\ell}(x^1_k,\ldots,x^{t-1}_k,u,x^{t+1}_{k-1},\ldots,x^\Gamma_{k-1};p_{k-1}) + \frac{1}{2}\|u-x^t_{k-1}\|^2\right\}, \tag{4}$$

$$p_k = (1-\theta)p_{k-1} + \chi c_\ell\left(\sum_{t=1}^\Gamma A_t x^t_k - d\right), \tag{5}$$

for every $1\leq t\leq\Gamma$, where $\chi\in(0,1)$ is a suitably chosen under-relaxation parameter. Moreover, the DP.ADMM introduces a novel approach for updating c_ℓ in subproblem (2) in order to ensure fast convergence.

Under a suitable choice of the stepsize λ and the following Slater-like assumption[1]:

$$\exists(x^1_\iota,\ldots,x^\Gamma_\iota)\in\operatorname{int}\left(\prod_{t=1}^\Gamma\operatorname{dom}h_t\right)\ \text{such that}\ \sum_{t=1}^\Gamma A_t x^t_\iota = d, \tag{6}$$

it is shown that for any tolerance pair $(\rho,\eta)\in\mathbb{R}^2_{++}$, the DP.ADMM obtains a tuple $(\bar{x}^1,\ldots,\bar{x}^\Gamma,\bar{p})$ satisfying

$$\operatorname{dist}\left(0,\nabla f(\bar{x}^1,\ldots,\bar{x}^\Gamma) + \prod_{t=1}^\Gamma\{A_t^*\bar{p}\} + \prod_{t=1}^\Gamma\partial h_t(\bar{x}^t)\right)\leq\rho,\quad \left\|\sum_{t=1}^\Gamma A_t\bar{x}^t - d\right\|\leq\eta. \tag{7}$$

in $\mathcal{O}(\rho^{-2}\eta^{-1})$ iterations (across all penalty subproblems).

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, 22, 23]), we make no further mention of them here. Instead, we discuss ADMM-type methods where f is strictly nonconvex.

Denoting $\operatorname{Im}(P)$ to be the image of a linear operator P and δ_S to be the indicator function of a closed convex set (see Subsection 1.1), Table 1 presents a list of common assumptions found the literature.

Using Table 1's notation, Table 2 presents a comparison between our proposed DP.ADMM and other relevant nonconvex ADMM-type methods, under a common tolerance $\varepsilon=\min\{\rho,\eta\}$.

Before ending this review, we make some additional remarks about results in papers [14, 24] compared to the results in this paper. First, both complexity bounds in [14, 24] require that that a feasible point is readily available, while DP.ADMM does not. Second, the $\mathcal{O}(\varepsilon^{-6})$ complexity bound established in [14] is for an ADMM-type method applied to a reformulation of (1), while DP.ADMM is applied to (1) directly. Third, the method in [24] considers a small stepsize (proportional to η^2)

[1]Here, $\operatorname{int}S$ denotes the interior of a set S, $\prod_{i=1}^n S_i$ denotes the Cartesian product of the collection of sets $\{S_i\}_{i=1}^n$, and $\operatorname{dom}\psi$ denotes the domain of a function ψ.

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

[3]Contains a proximal update for the first block and a non-proximal update in the second block.

[4]Contains a linearized proximal update instead of a fully proximal one.

$\mathcal{R}_0$	$\mathrm{Im}(A) \supseteq \{d\} \cup \mathrm{Im}(B)$.
$\mathcal{R}_1$	A has full row rank.
$\mathcal{R}_2$	A has full column rank.
$\mathcal{E}$	Problem (1) has a solution.
$\mathcal{S}$	The Slater-like assumption (6) holds.
$\mathcal{KL}$	The classical AL function, i.e. (3) with $\theta = 0$, has the KL property.[2]
$\mathcal{M}$	f is separable, i.e. $f(x^1, \ldots, x^\Gamma) = \sum_{t=1}^{\Gamma} f_t(x^t)$.
$\mathcal{H}$	$h_i \equiv 0$ or $h_i \equiv \delta_P$ for $i \in \{1, \ldots, \Gamma\}$, where P is a polyhedral set.
$\mathcal{F}$	A point $(x_0^1, \ldots, x_0^\Gamma) \in \prod_{t=1}^{\Gamma} \mathrm{dom}\, h_t$ satisfying $\sum_{t=1}^{\Gamma} A_t x_0^t = d$ is available.

Table 1: Common nonconvex ADMM assumptions and regularity conditions.

Algorithm	Iteration Complexity	Fully Proximal?	Assumptions
ADMM [25]	None	✗	$\mathcal{R}_0$,$\mathcal{R}_1$,$\mathcal{E}$,$\mathcal{M}$
ADMM [26]	None	✗	$\mathcal{R}_0$,$\mathcal{KL}$
IAPADMM [3]	None	✗[3]	$\mathcal{R}_0$,$\mathcal{M}$,$\mathcal{H}$,$\mathcal{KL}$
IAPADMM [13]	None	✓	$\mathcal{R}_0$,$\mathcal{R}_2$,$\mathcal{M}$,$\mathcal{H}$,$\mathcal{KL}$
LPADMM [27]	None	✓	$\mathcal{H}$,$\mathcal{S}$
PADMM-m [14]	$\mathcal{O}(\varepsilon^{-6})$	✓	$\mathcal{R}_1$,$\mathcal{F}$
SDD-ADMM [24]	$\mathcal{O}(\varepsilon^{-4})$	✗[4]	$\mathcal{F}$
DP.ADMM	$\mathcal{O}(\varepsilon^{-3})$	✓	$\mathcal{S}$

Table 2: Comparison of existing nonconvex ADMM-type methods with the DP.ADMM.

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 (4). Finally, it worth mentioning that [24] establishes an improved $\mathcal{O}(\varepsilon^{-3})$ complexity bound for SDD-ADMM under the additional strong assumption that $\mathcal{R}_1$ in Table 1 holds and $\prod_{t=1}^{\Gamma} \partial h_t(x)$ is compact for every x in the sublevel set of ϕ.

Contributions. Our contributions in this paper are twofold. First, we improve the current state-of-the-art nonconvex ADMM complexity bound from $\mathcal{O}(\varepsilon^{-4})$ to $\mathcal{O}(\varepsilon^{-3})$. Second, we are the first to establish the (non-asymptotic) convergence of a nonconvex proximal ADMM-type method without relying on any restrictive assumptions on the structure of the objective function ($\mathcal{M}$, $\mathcal{KL}$, and $\mathcal{H}$ in Table 1), the properties of the constraint matrices ($\mathcal{R}_0$, $\mathcal{R}_1$, and $\mathcal{R}_2$ in Table 1), or whether or not the proposed method starts from a feasible point ($\mathcal{F}$ in Table 1).

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 convergence analysis of the DP.ADMM in four subsections. The first one establishes some basic technical results, the second one presents bounds on some key algorithmic residuals, the third one bounds the Lagrange multipliers generated by the DP.ADMM, and the fourth one gives the proof of an important result, namely, Proposition 2.1. Section 4 gives some concluding remarks. Finally, the end of the paper contains several appendices.

1.1 Notation and Basic Definitions

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

Let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the set of nonnegative and positive real numbers, respectively, and 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 $\mathrm{dist}_X(x)$. 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.

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

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

for every $z\in\mathbb{R}^n$. The classical 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

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

For conciseness, we denote Δ_k^s to be the difference of iterates for the variable s at iteration k, i.e. $\Delta_k^s=s_k-s_{k-1}$. To reduce bracketing, if $z=(x,y)$ then $f(x,y)$ is equivalent to $f(z)=f((x,y))$.

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 $\{X_t\}_{t=1}^\Gamma\subseteq\mathbb{R}^{n_t}$ be compact convex sets and denote the aggregated quantities

$$\begin{gathered}X:=\prod_{t=1}^{\Gamma}X_t,\quad h(x):=\sum_{t=1}^{\Gamma}h_t(x^t),\quad A=(A_1,\ldots,A_\Gamma),\quad n=n_1+\cdots+n_\Gamma,\\ x^{<t}:=(x^1,\ldots,x^{t-1}),\quad x^{>t}:=(x^{t+1},\ldots,x^\Gamma),\quad x^{\leq t}:=(x^{<t},x^t),\quad x^{\geq t}:=(x^t,x^{>t}),\end{gathered} \tag{10}$$

for every $x=(x^1,\ldots,x^\Gamma)\in X$.

2.1 Problem of Interest

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

Our problem of interest is finding approximate stationary points of (1) under the following assumptions on $(\phi,h_1,\ldots,h_\Gamma)$ and (A,d):

(A1) $h_t \in \overline{\text{Conv}}\ X_t$ for every $1 \le t \le \Gamma$;

(A2) $A \not\equiv 0$ for $1 \le t \le \Gamma$ and $\mathcal{F} := \{x \in X : Ax = d\} \neq \emptyset$.

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

(A3) h is K_h-Lipschitz continuous on X for some $K_h \ge 0$;

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

$$\|\nabla_{x_t} f(x^{\le t}, \tilde{x}^{>t}) - \nabla_{x^t} f(x)\| \le M_t \|\tilde{x}^{\le t} - x^{>t}\|, \tag{11}$$

$$-\frac{m_t}{2}\|\tilde{x}^t - x^t\|^2 \le f(x^{<t}, \tilde{x}^t, x^{>t}) - f(x) - \left\langle \nabla_{x^t} f(x), \tilde{x}^t - x^t \right\rangle, \tag{12}$$

for every $x, \tilde{x} \in X$;

(A5) there exists $x_\iota \in \mathcal{F}$ such that $d_\iota := \text{dist}_{\partial X}(x_\iota) > 0$.

We now give a few remarks about the above assumptions. First, it is well known that (11) implies (12) with $m_t = M_{t-1}$. However, we show that better iterations complexities can be derived when scalars $\{m_t\}_{t=1}^{\Gamma}$ satisfying $m_t < M_{t-1}$ are available. Second, condition (12) implies that $f(x^{<t}, \cdot, x^{>t}) + m_t\|\cdot\|^2/2$ is convex on X_t for any $x \in X$. Third, since X 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:

$$D_x := \sup_{x, x' \in X} \|x - x'\|, \quad G_f := \sup_{x \in X} \|\nabla f(x)\|, \quad \phi_* := \inf_{x \in X}, \quad \overline{\phi} := \sup_{x \in X} \phi(x). \tag{13}$$

We now briefly discuss the notion of an approximate stationary point of (1) in (7). It is well-known that the first-order necessary condition for a point $\bar{x} = (\bar{x}^1, \ldots, \bar{x}^T)$ to be a local minimum of (1) is that there exists a multiplier $\bar{p} \in$ such that

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

Hence, the requirements in (7) can be viewed as a direct relaxation of the above conditions. For ease of future reference, we explicitly label the problem of obtaining (7) in the problem below.

Problem $\mathcal{LCCO}$: Given $(\rho, \eta) \in \mathbb{R}^2_{++}$, find a pair $(\bar{x}, \bar{p})$ satisfying (7).

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

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

and this type of condition has been previously considered in the authors' previous works [15, 16, 17, 19, 21]. In the next subsection, we present a subroutine (Algorithm 2.1) that computes such a residual in order to verify whether or not an incumbent solution $(\bar{x}, \bar{p})$ solves Problem $\mathcal{LCCO}$.

2.2 DP.ADMM

This subsection presents the DP.ADMM and its corresponding iteration complexity.

We present the DP.ADMM in two parts. The first part presents a static version of the DP.ADMM that either (i) terminates with a solution to Problem $\mathcal{LCCO}$ or (ii) signals that its penalty parameter is too small (see Proposition 2.1). The second part presents the actual DP.ADMM that repeatedly invokes the static version on an increasing sequence of penalty parameters. Both algorithms make use of the following condition on (χ, θ):

$$\chi \le \frac{\theta^2}{2\Gamma(1-\theta)(2-\theta)}, \quad \theta \ge \frac{1}{2}, \quad (\chi, \theta) \in [0,1)^2. \tag{14}$$

We now present a static version of our proposed DP.ADMM in Algorithm 2.1. For the special case of $(\theta, \chi) = (0, 1)$, its steps 1 and 3 reduce to the classic proximal ADMM iteration

$$\begin{aligned} x_k^t &= \underset{u \in \mathbb{R}^{n_t}}{\operatorname{argmin}} \left\{ \lambda \mathcal{L}_{c_\ell}^0 (x_k^{<t}, u, x_{k-1}^{>t}; p_{k-1}) + \frac{1}{2} \|u - x_{k-1}^t\|^2 \right\}, \\ p_k &= p_{k-1} + c_\ell \left(A x_k - d \right), \end{aligned}$$

for $1 \le t \le \Gamma$ 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 step 2b.

Algorithm 2.1: Static DP.ADMM

Require: $(\rho, \eta) \in [0,1]^2$, $x_0 \in X$, $p_0 \in A(\mathbb{R}^n)$, $\lambda \in (0, \frac{1}{\min_t m_t})$, $c > 0$, (χ, θ) satisfying (14)

for $k \leftarrow 1, 2, \ldots$ **do**

- `STEP 1` (prox update):
- **for** $t \leftarrow 1, 2, \ldots, \Gamma$ **do**
  - $x_k^t \leftarrow \operatorname{argmin}_{u \in \mathbb{R}^{n_t}} \left\{ \lambda \mathcal{L}_c^\theta (x_k^{<t}, u, x_{k-1}^{>t}; p_{k-1}) + \frac{1}{2}\|u - x_{k-1}^t\|^2 \right\}$
- `STEP 2a` (successful termination):
- $\tilde{p}_k \leftarrow (1-\theta) p_{k-1} + c(Ax_k - b)$
- **for** $t \leftarrow 1, 2, \ldots, \Gamma$ **do**
  - $v_k^t \leftarrow \nabla_{x^t} f(x_k^t) - \nabla_{x^t} f(x_k^{\le t}, x_{k-1}^{>t}) + c \sum_{s=t+1}^{\Gamma} A_t^* A_s \Delta_k^{x^s} - \frac{1}{\lambda} \Delta_k^{x^t}$
- **if** $\|v_k\| \le \rho$ **and** $\|Ax_k - d\| \le \eta$ **then**
  - **return** $(x_k, \tilde{p}_k, v_k)$
- `STEP 2b` (unsuccessful termination):
- **if** $k \ge 2$ **and** $\|v_k\| + \sqrt{\frac{c^3}{k-1}} \|Ax_k - d\| \le \rho$ **then**
  - **return** $(x_k, \tilde{p}_k, v_k)$
- `STEP 3` (multiplier update):
- $p_k \leftarrow (1-\theta) p_{k-1} + \chi c (Ax_k - d)$

The main properties of the static DP.ADMM are as follows. To keep these statements concise, we make use of the penalty parameter-independent scalars

$$M := \max_{1\le t\le \Gamma} M_t, \quad \alpha_\lambda := \min_{1\le t\le\Gamma}\left\{1-\frac{\lambda m_t}{2}\right\}, \quad B_\chi := \overline{\phi} - \phi_* + \frac{1}{\chi^2 c}\left(7\zeta + \|p_0\|^2\right)$$
$$\kappa_1 := \left(K_h + G_f + \Gamma^2\left[M+\frac{1}{\lambda}\right]D_z\right)D_z, \quad \kappa_2 := \frac{\chi\kappa_1 + 2\zeta c_1^{-1} + \sqrt{2\zeta}(1-\chi)(1-\theta)d_\iota\sigma_A^+}{(\chi+\theta-\chi\theta)d_\iota\sigma_A^+}, \tag{15}$$
$$\kappa_3 := \kappa_2 + \frac{4\chi D_x}{(\chi+\theta-\chi\theta)d_\iota\sigma_A^+}, \quad \kappa_4 := \Gamma\left(\sqrt{B_\chi} + \kappa_3\sqrt{\frac{1-\theta}{\chi}}\right),$$

and the penalty-dependent scalars

$$\beta_\lambda(c) := \frac{\lambda\Gamma^2}{\alpha_\lambda}\left(M+\frac{1}{\lambda}\right)^2 + 4c\sum_{t=1}^{\Gamma}\|A_t\|^2, \quad \tau_1(c) := \|p_0\| + \chi c\left(\frac{\sup_{x\in X}\|Ax-d\|}{\theta}\right),$$
$$\tau_2(c) := \Gamma^2\beta_\lambda(c)\left[B_\chi + \frac{1-\theta}{\chi c}\tau_1(c)\right], \quad \tau_3(c) := \frac{1}{2}\left[\kappa_4\sqrt{\frac{\beta_\lambda(c)}{\min\{1,c\}}} + \frac{2\sqrt{c}}{\chi}(\zeta+\kappa_3)\right] \tag{16}$$

for any value of $c > 0$ and a given $\zeta \ge 0$.

Proposition 2.1. *For any $c > 0$ and $(\rho,\eta) \in \mathbb{R}^2_+$, define*

$$T(c,\rho) := 1 + \tau_2(c) + \frac{\tau_3^2(c)}{\rho^2}, \quad \overline{c}(\eta) := \frac{\sqrt{T(1,1)}}{\eta}. \tag{17}$$

Then, the following statements hold about Algorithm 2.1:

(a) if (x_0,p_0) satisfies
$$\|p_1\|^2 = \|(1-\theta)p_0 + \chi c(Ax_0 - b)\|^2 \le 2\zeta, \tag{18}$$
then it terminates in at most $\lceil T(c,\rho)\rceil$ iterations;

(b) if it terminates in step 2a, then the first two components of its output triple $(x_k,\tilde{p}_k,v_k)$ solve Problem $\mathcal{LCCO}$;

(c) if it terminates in step 2b, then the output triple $(x_k,\tilde{p}_k,v_k)$ satisfies

$$v_k \in \nabla f(x_k) + \partial h(x_k) + A^*\tilde{p}_k, \quad \|v_k\| \le \rho, \quad \|Ax_k - d\| \le \frac{\sqrt{T(1,1)}}{\min\{c^2,c\}}; \tag{19}$$

(d) if its penalty parameter satisfies $c \ge \overline{c}(\eta)$ and (18) holds, then it terminates in step 2a.

The above proposition shows that Algorithm 2.1 obtains a solution of Problem $\mathcal{LCCO}$ in $\mathcal{O}(c\rho^{-2})$ iterations if $\|p_1\|$ is $\mathcal{O}(1)$ and the penalty parameter $c > 0$ is sufficiently large. Hence, to solve Problem $\mathcal{LCCO}$, one could invoke Algorithm 2.1 on an increasing sequence of penalty parameters and properly initialized (x_0,p_0). This the idea of the DP.ADMM, and we give its full description in Algorithm 2.2.

Note that Algorithm 2.2 does not specify how to obtain the point $\hat{x}_\ell$ given as the starting point of the ℓ^{th} call of Algorithm 2.1. In Appendix B, we show how an accelerated convex solver, started from any point in X, obtains this point in $\mathcal{O}(\sqrt{c_\ell})$ solver iterations, which is significantly less work than what performed by the ℓ^{th} call of Algorithm 2.1. Another (possibly more difficult) method is to find $x_0 \in \mathcal{F}$ and set $\hat{x}_\ell = x_0$, for every $\ell \ge 1$, and $\zeta = \|(1-\theta)p_0\|^2/2$.

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.

Algorithm 2.2: DP.ADMM

```
Require: ζ ≥ 0, (ρ, η) ∈ [0,1]^2, p_0 ∈ A(ℝ^n), λ ∈ (0, 1/(min_t m_t)), c_1 > 0, (χ, θ) satisfying (14)
for ℓ ← 1, 2, ... do
    find x̂_ℓ ∈ X satisfying ‖(1 − θ)p_0 + χc_ℓ(Ax̂_ℓ − d)‖^2 ≤ 2ζ
    call Algorithm 2.1 with (ρ, η, p_0, λ, χ, θ) as already initialized and (c, x_0) = (c_ℓ, x̂_ℓ) to
      obtain an output triple (x̄_ℓ, p̄_ℓ, v̄_ℓ)
    if ‖v̄_ℓ‖ ≤ ρ and ‖Ax̄_ℓ − d‖ ≤ η then
        return (x̄_ℓ, p̄_ℓ)
    else
        c_{ℓ+1} ← 2c_ℓ
```

Theorem 2.2. *Algorithm 2.2 stops and outputs a pair $(\bar{x}, \bar{p})$ that solves Problem $\mathcal{LCCO}$ in at most*

$$S(\eta,\rho) := \sum_{\ell=1}^{\bar{\ell}(\eta)} \lceil T(c_\ell,\rho) \rceil \le \frac{4T(1,1)}{\rho^2} \max\left\{\frac{1}{2}, \frac{1}{c_1}, \frac{\sqrt{T(1,1)}}{\eta}\right\} \tag{20}$$

iterations of Algorithm 2.1, where $\bar{\ell}(\eta) := \max\{1, \lceil \log_2[\bar{c}(\eta)/c_1]) \rceil\}$, $T(\cdot,\cdot)$ *is as in (17), and* $\bar{c}(\cdot)$ *is as in (17). Moreover, if it holds that*

$$\max\left\{\frac{1}{\chi}, \frac{1}{\lambda}, \frac{1}{\alpha_\lambda}, c_1, \zeta, \|p_0\|\right\} = \mathcal{O}(1), \tag{21}$$

then $S(\eta,\rho) = \mathcal{O}(\rho^{-2}\eta^{-1})$.

Proof. It follows from Lemma 2.1(b) and the penalty parameter update in Algorithm 2.2 that the number of calls of Algorithm 2.2 is $\bar{\ell}(\eta)$. Hence, it follows from Lemma 2.1(a)–(b) and the previous observation that Algorithm 2.2 stops and outputs a pair that solves Problem $\mathcal{LCCO}$ in at most $S(\eta,\rho) = \sum_{\ell=1}^{\bar{\ell}(\eta)} \lceil T(c_\ell,\rho) \rceil$ iterations of Algorithm 2.1. Moreover, using the fact that $\lceil T(c,\rho) \rceil \le 2\max\{1/c, c\} T(1,1)\rho^{-2}$, for every $(c,\rho) \in \mathbb{R}^2_+$, and the identity $c_\ell = c_1 2^{\ell-1}$, for every $\ell \ge 1$, we have that

$$\begin{aligned}\sum_{\ell=1}^{\bar{\ell}(\eta)} \lceil T(c_\ell,\rho) \rceil &\le \frac{2T(1,1)}{\rho^2} \max\left\{\frac{1}{c_1}\sum_{\ell=1}^{\bar{\ell}(\eta)} 2^{-(\ell-1)}, c_1 \sum_{\ell=1}^{\bar{\ell}(\eta)} 2^{\ell-1}\right\} \le \frac{2T(1,1)}{\rho^2} \max\left\{1, \frac{2}{c_1}, 2\bar{c}(\eta)\right\} \\ &= \frac{4T(1,1)}{\rho^2} \max\left\{\frac{1}{2}, \frac{1}{c_1}, \frac{\sqrt{T(1,1)}}{\eta}\right\}.\end{aligned}$$

The last conclusion follows from the above bound and the fact that (21) implies $T(1,1) = \mathcal{O}(1)$. □

3 Convergence Analysis

This section establishes the main properties of Algorithm 2.1 and contains four subsections. The first three subsections establish (in order) the following key properties[5] for every $k \ge 1$ and $\zeta = \mathcal{O}(1)$:

[5] See Lemma 3.1 for item 1, Lemma 3.6 for item 2, Lemmas 3.8 and 3.9 for item 3, and Lemma 3.11 for item 4.

1. The triple $(v_k, x_k, \tilde{p}_k)$ satisfies the inclusion of (19), and the feasibility $\|Ax_k - d\|$ is on the order
$$\|Ax_k - d\| = \Theta\left(\frac{\|p_k\| + \|p_{k-1}\|}{c}\right).$$

2. If (18) holds, then the average of the residuals $\{\|v_i\|\}_{i=1}^k$ is on the order of
$$\frac{1}{k}\sum_{i=1}^{k}\|v_k\| = \mathcal{O}\left(\frac{\sqrt{c} + \|p_k\|}{\sqrt{k}}\right).$$

3. The multiplier size $\|p_k\|$ is $\mathcal{O}(c)$, and if (18) holds, then there exists $k_0 = \Omega(c)$ such that for any $k \geq k_0$ the average of the multipliers sizes $\{\|p_i\|\}_{i=1}^k$ is $\mathcal{O}(1)$.

4. If (18) holds, then for every $k \geq 2$, the minimum of mixed residuals generated in step 2b of Algorithm 2.1 is on the order of
$$\min_{2\leq j\leq k}\left\{\|v_j\| + \sqrt{\frac{c^3}{j-1}}\|Ax_j - d\|\right\} = \mathcal{O}\left(\sqrt{\frac{c}{k}}\right).$$

The final subsection then uses the above properties to prove Proposition 2.1.

Throughout this section, we let $\{(v_i, x_i, p_i, \tilde{p}_i)\}_{i=1}^k$ denote the iterates generated by Algorithm 2.1 up to and including the k^{th} iteration for some $k \geq 2$. Moreover, for every $i \geq 1$ and $(\chi, \theta) \in \mathbb{R}^2_{++}$, we make use of the following useful constants and shorthand notation

$$\begin{gathered} a_\theta = \theta(1-\theta), \quad b_\theta := (2-\theta)(1-\theta), \quad \alpha_{\chi,\theta} := \frac{(1 - 2\Gamma\chi b_\theta) - (1-\theta)^2}{2\chi}, \\ f_i := Ax_i - d, \quad \mathcal{A}_i := \sum_{t=1}^{\Gamma}\sum_{s=t+1}^{\Gamma}\|A_t^* A_s \Delta_i^{x^s}\|, \end{gathered} \tag{22}$$

the aggregated quantities in (10), and the averaged quantities

$$S_{p,j} := \frac{\sum_{i=2}^j \|p_i\|}{j-1}, \quad S_{f,j} := \frac{\sum_{i=2}^j \|f_i\|}{j-1}, \quad S_{v,j} := \frac{\sum_{i=2}^j \|v_i\|}{j-1}. \tag{23}$$

3.1 Basic Technical Results

This section presents several technical results that will be repeatedly used in later analyses.

The first result presents some key relationships between the iterates generated by the method.

Lemma 3.1. *The following statements hold for every $i \leq k$:*

(a) $f_i = [p_i - (1-\theta)p_{i-1}]/(\chi c)$;

(b) if $i \geq 2$, then $\chi c(f_i - f_{i-1}) = \Delta_i^p - (1-\theta)\Delta_{i-1}^p$;

(c) $v_i \in \nabla f(x_i) + A^*\tilde{p}_i + \partial h(x_i)$ *and*
$$\|v_i\| \leq \Gamma^2\left(M + \frac{1}{\lambda}\right)\|\Delta_i^x\| + c\mathcal{A}_i.$$

Proof. (a) This is immediate from step 3 of the DP.ADMM and the definition of f_i in (22).

(b) This follows immediately from applying part (a) at indices i and $i-1$.

(c) We first prove the required inclusion. The optimality of x_k^k in step 1 of the DP.ADMM, assumption (A4), and the fact that $\lambda < 1/m$, imply that

$$
\begin{aligned}
0 &\in \partial \left[\mathcal{L}_c^\theta(x_i^{<t}, \cdot, x_{i-1}^{>t}; p_{i-1}) + \frac{1}{2\lambda}\|\cdot - x_{i-1}^k\|^2\right](x_i) \\
&= \nabla_{x^t} f(x_i^{\leq t}, x_{i-1}^{>t}) + A_t^*\left[(1-\theta)p_{i-1} + c[A(x_i^{\leq t}, x_{i-1}^{>t}) - d]\right] + \partial h_t(x_i^t) + \frac{1}{\lambda}\Delta_i^{x^t} \\
&= \nabla_{x^t} f(x_i^{\leq t}, x_{i-1}^{>t}) + A_t^*\left(\tilde{p}_i + c\sum_{s=t+1}^{\Gamma} A x_{i-1}^s\right) + \partial h_t(x_i^t) - \frac{1}{\lambda}\Delta_i^{x^t} \\
&= \nabla_{x^t} f(x_i) + A_t^*\tilde{p}_i + \partial h_t(x_i^t) - v_i^t.
\end{aligned}
$$

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

$$
\begin{aligned}
\|v_i^t\| &\leq \|\nabla_{x^t} f(x_i^t) - \nabla_{x^t} f(x_i^{\leq t}, x_{i-1}^{>t})\| + c\sum_{s=t+1}^{\Gamma}\|A_t^* A_s \Delta_i^{x_s}\| + \frac{1}{\lambda}\|\Delta_k^{x^t}\| \\
&\leq M_t\|x_i^{>t} - x_{i-1}^{>t}\| + c\sum_{s=t+1}^{\Gamma}\|A_t^* A_s \Delta_i^{x^s}\| + \frac{1}{\lambda}\|\Delta_k^{x^t}\| \\
&\leq \sum_{s=t}^{\Gamma}\left(M_t + \frac{1}{\lambda}\right)\|\Delta_i^{x^s}\| + c\sum_{s=t+1}^{\Gamma}\|A_t^* A_s \Delta_i^{x^s}\|.
\end{aligned}
$$

Using the above bound, the definition of M in 22, and the triangle inequality, we conclude that

$$
\begin{aligned}
\|v_i\| &\leq \sum_{t=1}^{\Gamma}\|v_i^t\| \leq \sum_{t=1}^{\Gamma}\sum_{s=t}^{\Gamma}\left(M_t + \frac{1}{\lambda}\right)\|\Delta_i^{x^s}\| + c\sum_{s=t}^{\Gamma}\sum_{s=t+1}^{\Gamma}\|A_t^* A_s \Delta_i^{x^s}\| \\
&\leq \left(M + \frac{1}{\lambda}\right)\sum_{t=1}^{\Gamma}\sum_{s=t}^{\Gamma}\|\Delta_i^{x^s}\| + c\mathcal{A}_i \leq \Gamma^2\left(M + \frac{1}{\lambda}\right)\|\Delta_i^x\| + c\mathcal{A}_i.
\end{aligned}
$$

□

The next result presents some general bounds given by the updates in (4).

Lemma 3.2. *Then for every $i \leq k$, $1 \leq t \leq \Gamma$, and $u \in X_t$, it holds that*

$$
\begin{aligned}
&\lambda\left[\mathcal{L}_c^\theta(x_i^{<t}, u, x_{i-1}^{>t}; p_{i-1}) - \mathcal{L}_c^\theta(x_i^{\leq t}, x_{i-1}^{>t}; p_{i-1})\right] + \frac{1}{2}\|u - x_{i-1}^t\|^2 \\
&\geq \frac{1}{2}\|\Delta_i^{x^t}\|^2 + \left(\frac{1-\lambda m_t}{2}\right)\|u - x_i^t\|^2 + \frac{\lambda c}{2}\|A_t(u - x_i^t)\|^2.
\end{aligned}
$$

Proof. Let $i \leq k$, $1 \leq t \leq \Gamma$, and $u \in X_t$ 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_i^t and the fact that $\lambda\mathcal{L}_c^\theta(x_i^{<t}, \cdot, x_{i-1}^{>t}; p_{i-1}) + \|\cdot\|^2/2$ is μ-strongly convex with respect to $\|\cdot\|_\alpha^2$, it holds that

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

or equivalently,

$$
\begin{aligned}
&\lambda \mathcal{L}_c^\theta(x_i^{\leq t}, x_{i-1}^{>t}; p_{i-1}) + \frac{1}{2}\|\Delta_i^{x^t}\|^2 \\
&\leq \lambda \mathcal{L}_c^\theta(x_i^{<t}, u, x_{i-1}^{>t}; p_{i-1}) + \frac{1}{2}\|u - x_{i-1}^t\|^2 - \frac{1}{2}\|u - x_i^t\|_\alpha^2 \\
&= \lambda \mathcal{L}_c^\theta(x_i^{<t}, u, x_{i-1}^{>t}; p_{i-1}) + \frac{1}{2}\|u - x_{i-1}^t\|^2 - \frac{\mu}{2}\|u^t - x_i\|^2 - \frac{\lambda c}{2}\|A_t(u - x_i)\|^2,
\end{aligned}
$$

which clearly implies the desired bound. □

The result below specializes to the previous result, and another one from Appendix A, to obtain more refined bounds.

Lemma 3.3. *The following statements hold for every $i \leq k$:*

(a) $\mathcal{L}_c^\theta(x_i; p_i) - \mathcal{L}_c^\theta(x_i; p_{i-1}) = b_\theta \|\Delta_i^p\|^2/(2\chi c) + a_\theta(\|p_i\|^2 - \|p_{i-1}\|^2)/(2\chi c)$;

(b) $\mathcal{L}_c^\theta(x_i; p_{i-1}) - \mathcal{L}_c^\theta(x_{i-1}; p_{i-1}) \leq -\alpha_\lambda \|\Delta_i^x\|^2/\lambda - c\sum_{t=1}^{\Gamma} \|A_t \Delta_i^{x^t}\|^2/2$;

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

$$
\frac{b_\theta}{2\chi c}\|\Delta_i^p\|^2 - \frac{c}{4}\sum_{t=1}^{\Gamma} \|A_t \Delta_i^{x^t}\|^2 \leq \frac{\alpha_{\chi,\theta}}{4\Gamma\chi c}\left(\|\Delta_{i-1}^p\|^2 - \|\Delta_i^p\|^2\right). \tag{24}
$$

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

$$
\begin{aligned}
\mathcal{L}_c^\theta(x_i; p_i) - \mathcal{L}_c^\theta(x_i; p_{i-1}) = (1-\theta)\langle \Delta_i^p, f_i\rangle &= \left(\frac{1-\theta}{\chi c}\right)\|\Delta_i^p\|^2 + \frac{a_\theta}{\chi c}\langle \Delta_i^p, p_{i-1}\rangle \\
&= \left(\frac{1-\theta}{\chi c}\right)\|\Delta_i^p\|^2 + \frac{a_\theta}{\chi c}\left(\langle p_i, p_{i-1}\rangle - \|p_{i-1}\|^2\right) \\
&= \left(\frac{1-\theta}{\chi c}\right)\|\Delta_i^p\|^2 + \frac{a_\theta}{\chi c}\left(\frac{1}{2}\|p_i\|^2 - \frac{1}{2}\|\Delta_i^p\|^2 - \frac{1}{2}\|p_{i-1}\|^2\right) \\
&= \frac{b_\theta}{2\chi c}\|\Delta_i^p\|^2 + \frac{a_\theta}{2\chi c}\left(\|p_i\|^2 - \|p_{i-1}\|^2\right),
\end{aligned}
$$

which clearly implies the desired bound.

(b) Using part (a) and Lemma 3.2 for every $1 \leq t \leq \Gamma$ and $u = x_i^t$, it holds that

$$
\begin{aligned}
\alpha_\lambda \|\Delta_i^x\|^2 + \frac{\lambda c}{2}\sum_{t=1}^{\Gamma}\|A_t \Delta_i^{x^t}\|^2 &\leq \sum_{i=1}^{t}\left(1 - \frac{\lambda m_t}{2}\right)\|\Delta_i^{x^t}\|^2 + \frac{\lambda c}{2}\sum_{t=1}^{\Gamma}\|A_t\Delta_i^{x^t}\|^2 \\
&\leq \lambda\left[\mathcal{L}_c^\theta(z_{i-1}; p_{i-1}) - \mathcal{L}_c^\theta(z_i; p_{i-1})\right],
\end{aligned}
$$

which clearly implies the desired bound.

(c) We first use (14) and Lemma A.2 with $(a, b, \tau) = (\Delta_i^p, \Delta_{i-1}^p, 2\Gamma\chi b_\theta)$ to observe that

$$
\|\Delta_k^p - (1-\theta)\Delta_{i-1}^p\|^2 \geq \Gamma\chi b_\theta \|\Delta_i^p\|^2 + \chi\alpha_{\chi,\theta}\left(\|\Delta_i^p\|^2 - \|\Delta_{i-1}^p\|^2\right). \tag{25}
$$

Using (25), Lemma 3.1(b), and the inequality $(\sum_{i=1}^n a_i)^2 \le n\sum_{i=1}^n a_i^2$ for every $\{a_i\}_{i=1}^n \subseteq \mathbb{R}$, we then have that

$$\begin{aligned}
\frac{c}{4}\sum_{t=1}^{\Gamma}\|A_t\Delta_i^{x^t}\|^2 \ge \frac{c}{4\Gamma}\|A\Delta_i^x\|^2 &= \frac{\|\Delta_i^p - (1-\theta)\Delta_{i-1}^p\|^2}{4\Gamma\chi^2 c} \\
&\ge \frac{1}{4\Gamma\chi c}\left[2\Gamma b_\theta\|\Delta_k^p\|^2 + \alpha_{\chi,\theta}\left(\|\Delta_k^p\|^2 - \|\Delta_{k-1}^p\|^2\right)\right] \\
&= \frac{b_\theta}{2\chi c}\|\Delta_k^p\|^2 + \frac{\alpha_{\chi,\theta}}{4\Gamma\chi c}\left(\|\Delta_k^p\|^2 - \|\Delta_{k-1}^p\|^2\right).
\end{aligned}$$

Rearranging terms, yields (24). □

3.2 Key Residual Bounds

This subsection presents several residual bounds that will be important in characterizing how the stationarity and feasibility of the iterates $\{(x_i, p_i)\}_{i=1}^k$ progress during the execution of Algorithm 2.1.

We first show a unified bound on the sum of the residuals $\|\Delta_i^x\|^2$ and $c\sum_{t=1}^{\Gamma}\|A_t\Delta_i^{x^t}\|^2$.

Lemma 3.4. *If $k \ge 2$, then it holds that*

$$\frac{\alpha_\lambda}{\lambda}\sum_{i=2}^{k}\|\Delta_i^x\|^2 + \frac{c}{4}\sum_{i=2}^{k}\sum_{t=1}^{\Gamma}\|A_t\Delta_i^{x^t}\|^2 \le \Psi_1^\theta - \Psi_k^\theta$$

where

$$\Psi_i^\theta := \mathcal{L}_c^\theta(z_i; p_i) - \frac{a_\theta}{2\chi c}\|p_i\|^2 + \frac{\alpha_{\chi,\theta}}{4\Gamma\chi c}\|\Delta_i^p\|^2 \quad \forall i \ge 1. \tag{26}$$

Proof. Define the function It follows from Lemma 3.3(b)–(c) and the definition of Ψ_i^θ above that

$$\begin{aligned}
&\frac{\alpha_\lambda}{\lambda}\|\Delta_i^{x^t}\|^2 + \frac{c}{4}\sum_{t=1}^{\Gamma}\|A_t\Delta_i^{x^t}\|^2 \\
&\le \mathcal{L}_c^\theta(x_{i-1}; p_{i-1}) - \mathcal{L}_c^\theta(x_i; p_i) + \frac{a_\theta}{2\chi c}(\|p_i\|^2 - \|p_{i-1}\|^2) + \frac{b_\theta}{2\chi c}\|\Delta_i^p\|^2 - \frac{c}{4}\sum_{t=1}^{\Gamma}\|A_t\Delta_i^{x^t}\|^2 \\
&\le \mathcal{L}_c^\theta(x_{i-1}; p_{i-1}) - \mathcal{L}_c^\theta(x_i; p_i) + \frac{a_\theta}{2\chi c}(\|p_i\|^2 - \|p_{i-1}\|^2) + \frac{\alpha_{\chi,\theta}}{4\Gamma\chi c}\left(\|\Delta_{i-1}^p\|^2 - \|\Delta_i^p\|^2\right) \\
&= \Psi_{i-1}^\theta - \Psi_i^\theta.
\end{aligned}$$

and summing the above inequality from $i = 2$ to k yields the desired bound. □

We next bound some quantities in the potential function Ψ_j^θ from the previous result, under the assumption that (18) holds.

Lemma 3.5. *If (18) holds for some $\zeta \ge 0$, then:*

(a) $\|p_1\|^2 \le 2\zeta$;

(b) $\mathcal{L}_c^\theta(x_1; p_1) \le \phi(x_0) + (13\zeta + \|p_0\|^2)/(2\chi^2 c)$;

(c) $\mathcal{L}_c^\theta(x_k; p_k) \ge \phi(x_k) - (1-\theta)^2\|p_k\|^2/(2c)$.

Proof. (a) This is immediate from (18) and step 3 of the DP.ADMM.

(b) Using (18), part (a), Lemma 3.3(a)–(b) at $i = 1$, and the fact that $a_\theta \in (0,1)$, $b_\theta \in (0,2)$, and $\theta \in (0,1)$, we have that

$$\begin{aligned}
\mathcal{L}_c^\theta(x_1;p_1) &= \mathcal{L}_c^\theta(x_1;p_0) + \frac{b_\theta}{2\chi c}\|\Delta_1^p\|^2 + \frac{a_\theta}{2\chi c}\left[\|p_1\|^2 - \|p_0\|^2\right] \\
&\le \mathcal{L}_c^\theta(x_0;p_0) + \frac{b_\theta}{2\chi c}\|\Delta_1^p\|^2 + \frac{a_\theta}{2\chi c}\|p_1\|^2 \le \mathcal{L}_c^\theta(x_0;p_0) + \frac{3}{\chi c}\|p_1\|^2 + \frac{1}{\chi c}\|p_0\|^2 \\
&= \phi(x_0) + (1-\theta)\left\langle p_0, Ax_0 - d\right\rangle + \frac{c}{2}\|Ax_0 - d\|^2 + \frac{3}{\chi c}\|p_1\|^2 + \frac{1}{\chi c}\|p_0\|^2 \\
&\le \phi(x_0) + \frac{1}{\chi^2 c}\left[\langle (1-\theta)p_0, \chi c(Ax_0 - d)\rangle + \frac{1}{2}\|\chi c(Ax_0-d)\|^2 + \frac{1}{2}\|p_0\|^2 + 6\|p_1\|^2\right] \\
&= \phi(x_0) + \frac{1}{\chi^2 c}\left[\frac{1}{2}\|(1-\theta)p_0 + \chi c(Ax - d)\|^2 + \frac{\theta}{2}\|p_0\|^2 + 6\|p_1\|^2\right] \\
&\le \phi(x_0) + \frac{1}{2\chi^2 c}\left(13\zeta + \|p_0\|^2\right).
\end{aligned}$$

(c) It holds that

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

□

We now present some important technical bounds that will be needed in later analyses.

Lemma 3.6. *Let Ψ_i^θ, B_χ, and $\beta_\lambda(\cdot)$ be as in (26), (15), and (16), respectively. If (18) holds for some $\zeta \ge 0$, then*

$$\begin{gathered}
\Psi_1^\theta - \Psi_k^\theta \le B_\chi + \frac{1-\theta}{\chi c}\|p_k\|^2, \\
\max\left\{\frac{c}{k-1}\sum_{i=2}^k \mathcal{A}_i, \frac{1}{\sqrt{2}}S_{v,k}\right\} \le \Gamma\sqrt{\frac{\beta_\lambda(c)\left[\Psi_1^\theta - \Psi_k^\theta\right]}{k-1}}.
\end{gathered} \tag{27}$$

Proof. Using Lemma 3.5(a)–(c), the bound $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$ for $a, b \in \mathbb{R}^n$, the fact that $c_1 \le c$, and the bounds $\alpha_{\chi,\theta} \le 1/(2\chi)$ and $(\chi,\theta) \in (0,1)^2$, we first have that

$$\begin{aligned}
\chi\left(\Psi_1^\theta - \Psi_k^\theta\right) &= \chi\left[\mathcal{L}_c^\theta(x_1;p_1) - \mathcal{L}_c^\theta(x_k;p_k)\right] + \frac{a_\theta}{2c}\left(\|p_k\|^2 - \|p_1\|^2\right) + \frac{\alpha_{\chi,\theta}}{4\Gamma c}\left(\|\Delta_1^p\|^2 - \|\Delta_k^p\|^2\right) \\
&\le \chi\left[\phi(x_0) + \frac{1}{2\chi^2 c}\left(13\zeta + \|p_0\|^2\right) - \phi(x_k) + \frac{(1-\theta)^2}{2c}\|p_k\|^2\right] + \frac{1}{8\Gamma\chi c}\|\Delta_1^p\|^2 + \frac{a_\theta}{2c}\|p_k\|^2 \\
&\le \chi\left[\overline{\phi} + \frac{1}{2\chi^2 c}\left(13\zeta + \|p_0\|^2\right) - \phi_* + \frac{1-\theta}{2c}\|p_k\|^2\right] + \frac{1}{4\Gamma\chi c}\left(\|p_1\|^2 + \|p_0\|^2\right) + \frac{1-\theta}{2c}\|p_k\|^2 \\
&\le \chi\left[\overline{\phi} - \phi_* + \frac{1}{\chi^2 c}\left(7\zeta + \|p_0\|^2\right)\right] + \frac{1-\theta}{c}\|p_k\|^2 \\
&\le \chi B_\chi + \frac{1-\theta}{c}\|p_k\|^2
\end{aligned}$$

which is the first desired bound. To show the other bounds, we use the relation $\|z\|_1 \leq \sqrt{n}\|z\|_2$ for every $z \in \mathbb{R}^n$, the definition of $\beta_\lambda(\cdot)$, and Lemma 3.4 to obtain

$$\begin{aligned}\frac{c}{k-1}\sum_{i=2}^{k}\mathcal{A}_i &\leq \left(\frac{c^2\Gamma^2}{k-1}\sum_{i=2}^{k}\sum_{t=1}^{\Gamma}\sum_{s=t+1}^{\Gamma}\|A_t^*A_s\Delta_i^{x^s}\|^2\right)^{1/2}\\ &\leq \left(\frac{4c\Gamma^2\sum_{t=1}^{\Gamma}\|A_t\|^2}{k-1}\cdot\frac{c}{4}\sum_{i=2}^{k}\sum_{s=1}^{\Gamma}\|A_s\Delta_i^{x^s}\|^2\right)^{1/2}\\ &\leq \Gamma\left[\frac{2c^2\sum_{t=1}^{\Gamma}\|A_t\|^2\left(\Psi_1^\theta-\Psi_k^\theta\right)}{k-1}\right]^{1/2} \leq \Gamma\left[\frac{\beta_\lambda(c)\left(\Psi_1^\theta-\Psi_k^\theta\right)}{k-1}\right]^{1/2},\end{aligned}$$

which implies the desired bound on the average of $\mathcal{A}_i$. Similarly, we use the relations $\|a\|_1 \leq \sqrt{n}\|a\|_2$ every $a \in \mathbb{R}^n$, the definition of $\beta_\lambda(\cdot)$, and Lemma 3.4 to conclude that

$$\begin{aligned}\frac{1}{\sqrt{2}}S_{v,k} &= \frac{\sum_{i=2}^{k}\|v_i\|}{\sqrt{2}(k-1)} \leq \left[\frac{1}{2(k-1)}\sum_{i=2}^{k}\|v_i\|^2\right]^{1/2}\\ &\leq \frac{1}{(k-1)^{1/2}}\left[\sum_{i=2}^{k}\Gamma^4\left(M+\frac{1}{\lambda}\right)^2\|\Delta_i^x\|^2 + c^2\sum_{i=2}^{k}\mathcal{A}_i^2\right]^{1/2}\\ &\leq \frac{1}{(k-1)^{1/2}}\left[\Gamma^4\left(M+\frac{1}{\lambda}\right)^2\sum_{i=2}^{k}\|\Delta_i^x\|^2 + \Gamma^2c^2\sum_{i=2}^{k}\sum_{t=1}^{\Gamma}\sum_{s=t+1}^{\Gamma}\|A_t^*A_s\Delta_i^{x^s}\|^2\right]^{1/2}\\ &\leq \frac{\Gamma}{(k-1)^{1/2}}\left[\Gamma^2\left(M+\frac{1}{\lambda}\right)^2\sum_{i=2}^{k}\|\Delta_i^x\|^2 + \left(4c\sum_{t=1}^{\Gamma}\|A_t\|^2\right)\frac{c}{4}\sum_{i=2}^{k}\sum_{s=1}^{\Gamma}\|A_s\Delta_i^{x^s}\|^2\right]^{1/2}\\ &\leq \Gamma\left[\frac{\beta_\lambda(c)}{k-1}\right]^{1/2}\left[\frac{\alpha_\lambda}{\lambda}\sum_{i=2}^{k}\|\Delta_i^x\|^2 + \frac{c}{4}\sum_{i=2}^{k}\sum_{s=1}^{\Gamma}\|A_s\Delta_i^{x^s}\|^2\right]^{1/2}\\ &\leq \Gamma\left[\frac{\beta_\lambda(c)\left(\Psi_1^\theta-\Psi_k^\theta\right)}{k-1}\right]^{1/2}, \qquad (28)\end{aligned}$$

which implies the desired bound on $S_{v,k}$.

Before continuing, we briefly outline some approaches that use the results obtained so far. First, if we can show that $\|p_k\|^2/c = \mathcal{O}(1)$ for every $k \geq 1$, then (27), Lemma 3.1(a), and the fact that $\beta_\lambda(c) = \Theta(c)$ imply

$$\min_{2\leq i\leq k}\|v_i\|^2 \leq S_{v,k} = \mathcal{O}\left(\frac{c}{k}\right), \quad \|f_j\| = \mathcal{O}\left(\frac{1}{c}\right) \quad \forall j \geq 1.$$

On the other hand, if we show that $S_{p,k} = \mathcal{O}(1)$ for every $k \geq \underline{k} \in \mathbb{N}$, then we could use a weighted sum of $\{S_{v,i}\}_{i=2}^k$ and $\{S_{f,i}\}_{i=2}^k$ — and analogous arguments as above — to bound $\|v_i\|$ and $\|f_i\|$. Our developments in the next subsections will follow the latter approach. □

3.3 Bounding the Lagrange Multipliers

This subsection presents some specialized Lagrange multiplier bounds and generalizes the analysis in [19]. More specifically, we show that $\|p_k\| = \mathcal{O}(c)$ and $S_{p,k} = \mathcal{O}(1)$ when (18) holds and k is sufficiently large.

The first result presents some important relationships between p_{i-1}, p_i, and $\tilde{p}_i$.

Lemma 3.7. *The following statements hold for every $i \le k$:*

(a) $p_i = \chi \tilde{p}_i + (1-\chi)(1-\theta)p_{i-1}$;

(c) it holds that

$$\frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota \sigma_Q^+ \|\tilde{p}_i\| \le \left(\frac{1-\theta}{c}\right)\langle \tilde{p}_i, p_{i-1}\rangle + 2cD_z\mathcal{A}_i + 2\kappa_1$$

where κ_1 and D_z are as in (15) and (13), respectively.

Proof. (a) This is an immediate consequence of the updates for p_i and $\tilde{p}_i$ in the DP.ADMM.

(b) Let $i \le k$ be fixed and define the auxiliary residual

$$\tilde{v}_i := v_i - \nabla f(z_i) - A^*\tilde{p}_i.$$

Using the fact that $\tilde{p}_i \in A(\mathbb{R}^n)$, Lemma A.1 with $(S,u) = (Q, \tilde{p}_i)$, Lemma 3.1(c), and the triangle inequality, we have that

$$\begin{aligned}
\frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota \sigma_A^+ \|\tilde{p}_i\| &\le \frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota \|A^*\tilde{p}_i\| = \frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota\|v_i - \nabla f(z_i) - \tilde{v}_i\| \\
&\le \frac{1}{c}\|\tilde{p}_k\|^2 + d_\iota\left[\|\tilde{v}_i\| + \|\nabla f(z_i)\| + \|v_i\|\right] \\
&\le \frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota\left[\|\tilde{v}_i\| + G_f + \Gamma^2\left(M + \frac{1}{\lambda}\right)D_x + c\mathcal{A}_i\right] \\
&\le \frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota\|\tilde{v}_i\| + c\mathcal{A}_i D_x + \kappa_1 - K_h D_x. \qquad (29)
\end{aligned}$$

We now derive a suitable bound on $d_\iota\|\tilde{v}_i\|$. First, note that Lemma 3.1(c) and the definition of $\tilde{v}_i$ imply that $\tilde{v}_i \in \partial h(z_i)$. Using the definition of D_z in (13) and Lemma A.3 with $(\psi, r, z, \bar{z}) = (h, \tilde{v}_i, x_i, x_\iota)$, it now follows that

$$\begin{aligned}
d_\iota\|\tilde{v}_i\| = \|\tilde{v}_i\|\mathrm{dist}_{\partial X}(x_\iota) &\le \left[\mathrm{dist}_{\partial X}(x_\iota) + \|x_i - x_\iota\|\right]K_h + \langle \tilde{v}_i, x_i - x_\iota\rangle \\
&\le 2K_h D_x + \langle \tilde{v}_i, x_i - x_\iota\rangle. \qquad (30)
\end{aligned}$$

On the other hand, Lemma 3.1(c) and the definitions of κ_1 and $\tilde{p}_i$ imply that

$$\begin{aligned}
\langle \tilde{v}_i, x_i - x_\iota\rangle &= \langle v_i - \nabla f(z_i) - A^*\tilde{p}_i, x_i - x_\iota\rangle \\
&\le \left(\|v_i\| + \|\nabla f(z_i)\|\right)\|x_i - x_\iota\| - \langle \tilde{p}_i, Ax_i - d\rangle \\
&\le \left[\Gamma^2\left(M + \frac{1}{\lambda}\right)D_x + c\mathcal{A}_i + G_f\right]D_x - \langle \tilde{p}_i, Ax_i - d\rangle \\
&= \kappa_1 - K_h D_z + c\mathcal{A}_i D_x + \left(\frac{1-\theta}{c}\right)\langle \tilde{p}_i, p_{i-1}\rangle - \frac{1}{c}\|\tilde{p}_i\|^2. \qquad (31)
\end{aligned}$$

The conclusion now follow from combining (29), (30), and (31). ☐

The next result establishes bounds on $\|p_i\|$ and $S_{p,k}$.

Lemma 3.8. *Let κ_2 and $\tau_1(\cdot)$ be as in (15) and (16), respectively. Then:*

(a) for every $i \le k$, it holds that $\|p_i\| \le \tau_1(c)$.

(b) if (18) *holds for some* $\zeta > 0$ *and* $k \geq 2$*, then*

$$S_{p,k} \leq \kappa_2 + \left[\frac{4\chi D_x}{(\chi+\theta-\chi\theta)d_\iota\sigma_A^+}\right]\left[\frac{c}{k-1}\sum_{i=2}^{k}\mathcal{A}_i\right].$$

Proof. (a) In view of step 3 of Algorithm 2.1, the fact that $\theta \in (0,1)$, and the triangle inequality, it holds that

$$\begin{aligned}
\|p_i\| &\leq (1-\theta)\|p_{i-1}\| + \chi c\|Ax_i - d\| \leq (1-\theta)^i\|p_0\| + \chi c\sum_{j=0}^{i-1}(1-\theta)^i\|Ax_i - d\| \\
&\leq \|p_0\| + \chi c \cdot \sup_{x\in\mathcal{F}}\|Ax-d\|\sum_{j=0}^{\infty}(1-\theta)^i = \|p_0\| + \chi c\left(\frac{\sup_{x\in X}\|Ax-d\|}{\theta}\right) = \tau_1(c).
\end{aligned}$$

(b) Let $i \leq k$ be arbitrary, and define

$$\xi_k := \frac{c}{k-1}\sum_{i=2}^{k}\mathcal{A}_i, \quad \nu_i(c) := \kappa_1 + c\mathcal{A}_i D_x, \quad e_0 := (1-\theta)(1-\chi). \tag{32}$$

Using Lemma 3.7(a) thrice, and the bounds $2ab \leq a^2 + b^2$ and $(a+b)^2 \leq 2a^2 + 2b^2$ for every $a, b \in \mathbb{R}_+$, we first have that

$$\begin{aligned}
&\frac{1}{c}\|p_i\|^2 + d_\iota\sigma_A^+\|p_i\| = \frac{1}{c}\|\chi\tilde{p}_i + e_0 p_{i-1}\|^2 + d_\iota\sigma_A^+\|\chi\tilde{p}_i + e_0 p_{i-1}\| \\
&\leq 2\chi\left[\frac{1}{c}\|\tilde{p}_i\|^2 + d_\iota\sigma_A^+\|\tilde{p}_i\|\right] + \frac{2e_0^2}{c}\|p_{i-1}\|^2 + e_0 d_\iota\sigma_A^+\|p_{i-1}\| \\
&\leq 2\chi\left[\frac{1-\theta}{c}\langle\tilde{p}_i, p_{i-1}\rangle + 2\nu_i(c)\right] + \frac{2e_0^2}{c}\|p_{i-1}\|^2 + e_0 d_\iota\sigma_A^+\|p_{i-1}\| \\
&= \frac{2(1-\theta)}{c}\langle p_i, p_{i-1}\rangle + \underbrace{\frac{2e_0(e_0-1)}{c}}_{\leq 0}\|p_{i-1}\|^2 + e_0 d_\iota\sigma_A^+\|p_{i-1}\| + 4\chi\nu_i(c) \\
&\leq \frac{1-\theta}{c}\|p_i\|^2 + \frac{1-\theta}{c}\|p_{i-1}\|^2 + e_0 d_\iota\sigma_A^+\|p_{i-1}\| + 4\chi\nu_i(c).
\end{aligned} \tag{33}$$

Rearranging terms yields

$$\frac{2\theta-1}{c}\|p_i\|^2 + (1-e_0)d_\iota\sigma_A^+\|p_i\| \leq \frac{1-\theta}{c}\left[\|p_{i-1}\|^2 - \|p_i\|^2\right] + e_0 d_\iota\sigma_A^+\left[\|p_{i-1}\| - \|p_i\|\right] + 4\chi\nu_i(c)$$

Summing the above inequality from $i = 2$ to k and combining the resulting bound with Lemma 3.5(a), (14), and the definition of κ_2 in (15), we conclude that

$$\begin{aligned}
(1-e_0)d_\iota\sigma_A^+\sum_{i=2}^{k}\|p_i\| &\leq \frac{2\theta-1}{c}\sum_{i=2}^{k}\|p_i\|^2 + (1-e_0)d_\iota\sigma_A^+\sum_{i=2}^{k}\|p_i\| \\
&\leq \frac{1-\theta}{c}\left[\|p_1\|^2 - \|p_k\|^2\right] + e_0 d_\iota\sigma_A^+\left[\|p_1\| - \|p_k\|\right] + 4\chi\sum_{i=2}^{k}\nu_i(c) \\
&\leq \frac{1}{c_1}\|p_1\|^2 + e_0 d_\iota\sigma_A^+\|p_1\| + 4\chi(k-1)(D_x\xi_k + \kappa_1) \\
&\leq 2\zeta c_1^{-1} + \sqrt{2\zeta}e_0 d_\iota\sigma_A^+ + \|p_0\|^2 + 4\chi(k-1)(D_x\xi_k + \kappa_1) \\
&\leq (k-1)\left[(1-e_0)d_\iota\sigma_A^+\kappa_2 + 4\chi D_z\xi_k\right],
\end{aligned}$$

which, in view of the definitions of e_0, ξ_k, and $S_{p,k}$, imply the desired bound. □

We now show that $S_{p,k}$ is $\mathcal{O}(1)$ when (18) holds and k is sufficiently large.

Lemma 3.9. *Let κ_3 and $\tau_3(\cdot)$ be as in (15) and (16), respectively. If (18) holds for some $\zeta \geq 0$, then for every $k \geq 1 + \tau_2(c)$ it holds that $S_{p,k} \leq \kappa_3$.*

Proof. Let ξ_k be as in (32), and suppose that $k \geq 1 + \tau_2(c)$. Using Lemmas 3.6, 3.8(a), and the definition of $\tau_3(c)$ it holds that

$$\begin{aligned}\xi_k^2 &\leq \frac{\Gamma^2 \beta_\lambda(c)\left(\Psi_1^\theta - \Psi_k^\theta\right)}{k-1} \leq \frac{\Gamma^2\beta_\lambda(c)}{k-1}\left[B_\chi + \frac{1-\theta}{\chi c}\|p_k\|^2\right] \\ &\leq \frac{\Gamma^2\beta_\lambda(c)}{\tau_2(c)}\left[B_\chi + \frac{1-\theta}{\chi c}\|p_k\|^2\right] \leq \frac{\Gamma^2\beta_\lambda(c)}{\tau_2(c)}\left[B_\chi + \frac{1-\theta}{\chi c}\tau_1(c)\right] = 1.\end{aligned}$$

The conclusion now follows from the above bound, the definition of ξ_k, Lemma 3.8(b), and the definition of κ_3. $\square$

3.4 Proof of Proposition 2.1

This subsection gives the proof of Proposition 2.1 and follows the approach described at the end of Subsection 3.2.

We first show some important bounds involving $S_{v,j}$ and $S_{f,j}$.

Lemma 3.10. *Let (κ_3, κ_4) and (τ_2, τ_3) be as in (15) and (16), respectively. If (18) holds for some $\zeta \geq 0$, then for every $k \geq 1 + \tau_2(c)$ the following statements hold.*

(a) $\sum_{j=2}^k (j-1) S_{v,j} \leq (k-1)^{3/2}\kappa_4\sqrt{\beta_\lambda(c)/\min\{1,c\}}$;

(b) $\sum_{j=2}^k (j-1) S_{f,j} \leq 2\left(\zeta + \kappa_3\right)(k-1)^2/(\chi c)$;

Proof. (a) Let Ψ_i^θ be as in (26). Using Lemmas 3.6 and 3.9 and the bound $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \in \mathbb{R}_+$, it follows that

$$\begin{aligned}\frac{1}{\sqrt{2}}\sum_{j=2}^k (j-1)S_{v,j} &= \frac{1}{\sqrt{2}}\sum_{j=2}^k\sum_{i=2}^j \|v_i\| \leq \Gamma\sum_{j=2}^k \sqrt{(j-1)\beta_\lambda(c)[\Psi_1^\theta - \Psi_k^\theta]} \\ &\leq \Gamma\sum_{j=2}^k\sqrt{(j-1)\beta_\lambda(c)\left[B_\chi + \frac{1-\theta}{\chi c}\|p_j\|^2\right]} \\ &\leq \Gamma\sqrt{\beta_\lambda(c)}\left[(k-1)^{3/2}\sqrt{B_\chi} + (k-1)^{1/2}\sqrt{\frac{1-\theta}{\chi c}}\sum_{j=2}^k\|p_j\|\right] \\ &\leq \Gamma(k-1)^{3/2}\sqrt{\frac{\beta_\lambda(c)}{\min\{1,c\}}}\left(\sqrt{B_\chi} + \kappa_3\sqrt{\frac{1-\theta}{\chi}}\right) \\ &= (k-1)^{3/2}\kappa_4\sqrt{\frac{\beta_\lambda(c)}{\min\{1,c\}}}.\end{aligned}$$

(b) Let δ_j be as in part (a). Using Lemmas 3.8(b), Lemma 3.1, Lemma 3.5(a), the triangle inequality, and the fact that $\theta \in (0,1)$ and $k \geq 2$, it follows that

$$\sum_{j=2}^{k}(j-1)S_{f,j} = \sum_{j=2}^{k}\sum_{i=2}^{j}\|f_i\| = \frac{1}{\chi c}\sum_{j=2}^{k}\sum_{i=2}^{j}\|p_i-(1-\theta)p_{i-1}\|$$
$$\leq \frac{1}{\chi c}\sum_{j=2}^{k}\left(\|p_1\|+2\sum_{i=2}^{j}\|p_i\|\right) \leq \frac{2}{\chi c}(\zeta+\kappa_3)\sum_{j=2}^{k}(j-1)$$
$$\leq \frac{2}{\chi c}(\zeta+\kappa_3)(k-1)^2$$

□

The next result shows that if (18) holds, then the minimum of the residuals in step 2b of Algorithm 2.1 are on the order of $\mathcal{O}(\sqrt{\min\{1/c,c\}\cdot k^{-1}})$ for sufficiently large enough k.

Lemma 3.11. *Let (τ_2,τ_3) be as in (16). If (18) holds for some $\zeta \geq 0$, then for every $k \geq 1+\tau_2(c)$ it holds that*

$$\min_{2\leq j\leq k}\left\{\|v_j\|+\sqrt{\frac{c^3}{j-1}}\|Ax_j-d\|\right\} \leq \frac{\tau_3(c)}{\sqrt{k-1}}.$$

Hence, the number of iterations performed by Algorithm 2.1 is finite in view of its step 2b.

Proof. Define

$$r_j := \|v_j\|+\sqrt{\frac{c^3}{j-1}}\|Ax_j-d\|, \quad S_{r,j} := \frac{\sum_{i=2}^{j}\|r_i\|}{j-1}, \quad \forall j\geq 2. \tag{34}$$

Using Lemma 3.10, we have that

$$\frac{(k-1)^2}{2}\min_{2\leq j\leq k} r_j \leq \sum_{j=2}^{k}(j-1)\min_{2\leq i\leq k} S_{r,i} \leq \sum_{j=2}^{k}(j-1)S_{r,j} = \sum_{j=2}^{k}(j-1)\left(S_{v,j}+\sqrt{\frac{c^3}{k-1}}S_{f,j}\right)$$
$$\leq (k-1)^{3/2}\left[\kappa_4\sqrt{\frac{\beta_\lambda(c)}{\min\{1,c\}}}+\frac{2\sqrt{c}}{\chi}(\zeta+\kappa_3)\right] = \frac{(k-1)^{3/2}\tau_3(c)}{2},$$

which immediately implies the desired bound.

□

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

Proof of Proposition 2.1. (a) Let $T(\cdot,\cdot)$ be as in (17), and, for the sake of contradiction, suppose that the current cycle has not terminated by the beginning of iteration $T(c,\rho)$. Since $T(c,\rho) \geq 1+\tau_2(c)$ then by Lemma 3.10(c), it follows that

$$\min_{2\leq j\leq\lceil T(c,\rho)\rceil} r_j \leq \frac{\tau_3(c)}{\sqrt{\lceil T(c,\rho)\rceil-1}} \leq \frac{\tau_3(c)}{\sqrt{\tau_3^2(c)\rho^{-2}}} = \rho,$$

which contradicts our assumption that the cycle has not terminated due to step 2b. As a consequence, it must be the case that each call of Algorithm 2.1 is run for at most $T(\rho)$ iterations.

(b) This is immediate from the termination condition in step 2b and Lemma 3.1(c).

(c) In view of part (a), let $k \le T(c,\rho)$ be the first index where $r_k \le \rho$. Using step 2b of the DP.ADMM and the bounds

$$\tau_2(c) \le \max\{1/c, c\} \cdot \tau_2(1), \quad \tau_3(c) \le \max\{1/c, c\} \cdot \tau_3(1), \quad \rho \le 1,$$

it holds that $\|v_j\| \le r_k \le \rho$ and

$$\begin{aligned}
\|f_k\| &\le \sqrt{\frac{k-1}{c^3}} \left[\|v_k\| + \frac{c^{3/2}\|f_k\|}{\sqrt{k-1}} \right] = \sqrt{\frac{(k-1)r_k^2}{c^3}} \le \sqrt{\frac{[\lceil T(c,\rho)\rceil - 1]\rho^2}{c^3}} \\
&\le \sqrt{\frac{[1+\tau_2(c)+\tau_3^2(c)\rho^{-2}]\rho^2}{c^3}} \le \sqrt{\frac{[1+\tau_2(1)+\tau_3^2(1)]\cdot \max\{1/c, c\}}{c^3}} \\
&\le \frac{\sqrt{1+\tau_2(1)+\tau_3^2(1)}}{\min\{c^2, c\}} = \frac{\sqrt{T(1,1)}}{\min\{c^2, c\}}.
\end{aligned}$$

The conclusion now follows from the inclusion in Lemma 3.1(c) and the above bounds.

(d) This is an immediate consequence of parts (a) and (c), the fact that $\overline{c}(\eta) \ge 1$, the fact that step 2a precedes step 2b, and the termination criterion in step 2a. □

4 Concluding Remarks

The convergence of the DP.ADMM (Algorithm 2.2) is established under the assumption that exact solutions to the subproblems in step 2 of its static version (Algorithm 2.1) are easy obtain. We believe that convergence can be also be established for when only inexact solutions, i.e.,

$$x_k^t \approx \operatorname*{argmin}_{u \in \mathbb{R}^{n_t}} \left\{ \lambda \mathcal{L}_c^\theta(x_k^{<t}, u, x_{k-1}^{>t}; p_{k-1}) + \frac{1}{2}\|u - x_{k-1}^t\|^2 \right\} \tag{35}$$

are available. For example, one could consider applying an accelerated composite gradient (ACG) method to the problem associated with (35) so that x_k^t satisfies

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

for some $\sigma \in (0,1)$, where $\partial_\varepsilon \psi(x) := \{v \in \mathbb{R}^n : \psi(y') \ge \psi(y) + \langle v, y' - y\rangle - \varepsilon, \forall y' \in \operatorname{dom}\psi\}$. Using the techniques developed in [19], we believe that the ACG iteration complexity of the resulting inexact DP.ADMM is $\mathcal{O}(\Gamma \eta^{-1.5}\rho^{-2}\log(\eta\sigma)^{-1})$ compared to the $\mathcal{O}(\eta^{-1}\rho^{-2})$ iteration complexity of the exact DP.ADMM.

A Technical Inequalities

The first result, whose proof can be found in [12, Lemma 1.3], presents a relationship between elements in the image of a linear operator.

Lemma A.1. *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 following result can be found in [19, Lemma B.2].

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

$$\|a - (1-\theta)b\|^2 - \tau\|a\|^2 \ge \left[\frac{(1-\tau)-(1-\theta)^2}{2}\right]\left(\|a\|^2 - \|b\|^2\right). \tag{36}$$

Finally, the proof of the next result can be found in [20, Lemma 4.7].

Lemma A.3. *Suppose* $\psi \in \overline{\text{Conv}}\ \mathbb{R}^n$ *is* K_ψ*-Lipschitz continuous. Then, for every* $z, \bar{z} \in \operatorname{dom}\psi$ *and* $r \in \partial\psi(z)$*, it holds that*

$$\|r\|\text{dist}(\bar{z}, \partial(\operatorname{dom}\psi)) \le [\text{dist}(\bar{z}, \partial(\operatorname{dom}\psi)) + \|z - \bar{z}\|]\, K_\psi + \langle r, z - \bar{z}\rangle,$$

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

B Finding $\hat{x}_\ell$ in Algorithm 2.2

Recall that the DP.ADMM (Algorithm 2.2) at the ℓ^{th} call of the static DP.ADMM (Algorithm 2.1) requires finding a point $\hat{x}_\ell$ that satisfies (18) with $(x_0, c) = (\hat{x}_\ell, c_\ell)$ for a given (p_0, χ, θ) and $\zeta \ge 0$. One efficient way is to obtain $\hat{x}_\ell$ is to apply an accelerated composite gradient (ACG) method to minimize the function

$$\Phi_\ell(x) := \frac{1}{2}\|(1-\theta)p_0 + \chi c_\ell(Ax - d)\|^2. \tag{37}$$

For this appendix, we examine how the FISTA-like ACG methods studied in [15, 18, 19] behave when minimizing $\Phi_\ell(\cdot)$. The first result gives the complexity bound for finding a near optimal point of a general convex function with Lipschitz continuous gradient. Its proof can be found, for example, in [15, Lemma 3.3.1].

Lemma B.1. *For any* $\sigma > 0$ *and starting point* $x_0 \in \mathbb{R}^n$*, the ACG methods studied in [15, 18, 19], applied to a convex function* $\Phi : \mathbb{R}^n \mapsto \mathbb{R}$ *with* L*-Lipschitz continuous gradient, obtain a triple* $(\bar{x}, \bar{v}, \bar{\varepsilon})$ *satisfying*

$$\bar{v} \in \partial_{\bar{\varepsilon}}\Phi(\bar{x}), \quad \|\bar{v}\|^2 + 2\bar{\varepsilon} \le \sigma^2\|x_0 - \bar{x} + v\|^2$$

in $\lceil 2\sqrt{2L}(1+\sigma)/\sigma \rceil$ *ACG iterations, where*

$$\partial_\varepsilon\Phi(x) := \left\{v \in \mathbb{R}^n : \Phi(x') \ge \Phi(x) + \langle v, x' - x\rangle - \varepsilon, \forall x' \in \mathbb{R}^n\right\}.$$

The next result shows that minimizing $\Phi_\ell(\cdot)$ in (37) with the above ACG methods yields $\bar{x}$ satisfying $\Phi(\bar{x}) \le \zeta$, for some $\zeta > 0$, in $\mathcal{O}(\sqrt{c})$ ACG iterations.

Proposition B.2. *The ACG methods methods studied in [15, 18, 19], when minimizing* $\Phi_\ell(\cdot)$ *with* $L = c\|A\|^2$ *and some starting point* $\bar{x}_0 \in X$*, obtain a point* $\bar{x}$ *satisfying*

$$\Phi_\ell(\bar{x}) \le D_x^2 + \frac{(1-\theta)^2}{2}\|p_0\|^2$$

in at most $\lceil 6\sqrt{2c}\|A\| \rceil$ *ACG iterations, where* D_x *is as in* (13).

Proof. Note first that $\Phi_\ell(\cdot)$ is convex and continuously differentiable with $c\|A\|^2$-Lipschitz continuous gradient. Using Lemma B.1 with $\phi = \Phi_\ell$, $L = c\|A\|^2$ and $\sigma = 1/2$, it holds that the considered ACG methods obtain $(\bar{x}, \bar{v}, \bar{\varepsilon})$ satisfying

$$\bar{v} \in \partial_{\bar{\varepsilon}}\Phi_\ell(\bar{x}), \quad \|\bar{v}\|^2 + 2\bar{\varepsilon} \le \frac{1}{4}\|\bar{x}_0 - \bar{x} + \bar{v}\|^2 \tag{38}$$

in $\lceil 6\sqrt{2c}\|A\| \rceil$ ACG iterations. Using the triangle inequality and the above inequality, we have

$$\|v\| \le \frac{1}{2}\|\bar{x}_0 - \bar{x} + \bar{v}\| \le \frac{1}{2}\|\bar{x}_0 - \bar{x}\| + \frac{1}{2}\|\bar{v}\|,$$

which implies that $\|\bar{v}\| \le \|\bar{x}_0 - \bar{x}\| \le D_x$. Using this bound, the inequality $(a+b)^2 \le 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$, and (38), it follow that for any feasible (see assumption (A2)) $x_{\mathcal{F}} \in \mathcal{F}$, we have

$$\begin{aligned}
\frac{(1-\theta)^2}{2}\|p_0\|^2 = \Phi_\ell(x_{\mathcal{F}}) &\ge \Phi_\ell(\bar{x}) + \langle \bar{v}, x_{\mathcal{F}} - \bar{x} \rangle - \varepsilon \ge \Phi_\ell(\bar{x}) - \frac{1}{2}\|\bar{v}\|^2 - \frac{1}{2}\|x_{\mathcal{F}} - \bar{x}\|^2 - \varepsilon \\
&\ge \Phi_\ell(\bar{x}) - \frac{1}{8}\|\bar{x}_0 - \bar{x} + \bar{v}\|^2 - \frac{1}{2}\|x_{\mathcal{F}} - \bar{x}\|^2 \\
&\ge \Phi_\ell(\bar{x}) - \frac{1}{4}\|\bar{x}_0 - \bar{x}\|^2 - \frac{1}{4}\|\bar{v}\|^2 - \frac{1}{2}\|x_{\mathcal{F}} - \bar{x}\|^2 \\
&\ge \Phi_\ell(\bar{x}) - D_x^2,
\end{aligned}$$

which clearly implies the desired bound.

□

Notice that the above complexity bound only depends on c, $\|A\|$, and $\|p_0\|$, as opposed to other ACG complexity bounds that typically depend on values of the objective function.

References

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

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

[3] Mian Tao Chao, Ye Zhang, and Jin Bao Jian. An inertial proximal alternating direction method of multipliers for nonconvex optimization. *International Journal of Computer Mathematics*, pages 1–19, 2020.

[4] Jonathan Eckstein and Dimitri P Bertsekas. On the douglas–rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

[5] Jonathan Eckstein and Michael C Ferris. Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS Journal on Computing*, 10(2):218–235, 1998.

[6] Jonathan Eckstein and Masao Fukushima. Some reformulations and applications of the alternating direction method of multipliers. In *Large scale optimization*, pages 115–134. Springer, 1994.

[7] Jonathan Eckstein and Benar Fux Svaiter. A family of projective splitting methods for the sum of two maximal monotone operators. *Mathematical Programming*, 111(1):173–199, 2008.

[8] Jonathan Eckstein and Benar Fux Svaiter. General projective splitting methods for sums of maximal monotone operators. *SIAM Journal on Control and Optimization*, 48(2):787–811, 2009.

[9] Daniel Gabay. Applications of the method of multipliers to variational inequalities. In *Studies in mathematics and its applications*, volume 15, pages 299–331. Elsevier, 1983.

[10] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & mathematics with applications*, 2(1):17–40, 1976.

[11] Roland Glowinski and Americo 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(R2):41–76, 1975.

[12] Max L. N. Goncalves, Jefferson G. Melo, and Renato D. C. Monteiro. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *arXiv preprint arXiv:1702.01850*, 2017.

[13] Zehui Jia, Jieru Huang, and Zhongming Wu. An incremental aggregated proximal ADMM for linearly constrained nonconvex optimization with application to sparse logistic regression problems. *Journal of Computational and Applied Mathematics*, 390:113384, 2021.

[14] Bo Jiang, Tianyi Lin, Shiqian Ma, and Shuzhong Zhang. Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis. *Computational Optimization and Applications*, 72(1):115–157, 2019.

[15] Weiwei Kong. Accelerated inexact first-order methods for solving nonconvex composite optimization problems. *arXiv preprint arXiv:2104.09685*, 2021.

[16] Weiwei Kong, Jefferson G Melo, and Renato DC Monteiro. Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs. *SIAM Journal on Optimization*, 29(4):2566–2593, 2019.

[17] Weiwei Kong, Jefferson G Melo, and Renato DC Monteiro. An efficient adaptive accelerated inexact proximal point method for solving linearly constrained nonconvex composite problems. *Computational Optimization and Applications*, 76(2):305–346, 2020.

[18] Weiwei Kong and Renato D. C. Monteiro. Accelerated inexact composite gradient methods for nonconvex spectral optimization problems. *arXiv preprint arXiv:2007.11772*, 2020.

[19] Weiwei Kong and Renato D. C. Monteiro. An accelerated inexact dampened augmented Lagrangian method for linearly-constrained nonconvex composite optimization problems. *arXiv preprint arXiv:2110.11151*, 2021.

[20] Jefferson G. Melo and Renato 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.

[21] Jefferson G Melo, Renato DC Monteiro, and Weiwei 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.

[22] Ralph Tyrrell Rockafellar. Augmented Lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research*, 1(2):97–116, 1976.

[23] Andrzej Ruszczyński. An augmented Lagrangian decomposition method for block diagonal linear programming problems. *Operations Research Letters*, 8(5):287–294, 1989.

[24] Kaizhao Sun and Andy Sun. Dual descent ALM and ADMM. *arXiv preprint arXiv:2109.13214*, 2021.

[25] Andreas Themelis and Panagiotis Patrinos. Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

[26] Yu Wang, Wotao Yin, and Jinshan Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

[27] Jiawei Zhang and Zhi-Quan Luo. A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization. *SIAM Journal on Optimization*, 30(3):2272–2302, 2020.