

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

4 WEIWEI KONG[†] AND RENATO D.C. MONTEIRO[‡]

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

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

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

18 **1. Introduction.** Consider the following composite optimization problem:

19 (1.1)
$$\min_{x \in \mathbb{R}^n} \{ \phi(x) := f(x) + h(x) : Ax = d \},$$

20 where h is a closed convex function, f is a (possibly) nonconvex differentiable function
21 on the domain of h , the gradient of f is Lipschitz continuous, A is a linear operator,
22 $d \in \mathbb{R}^\ell$ is a vector in the image of A (denoted as $\text{Im}(A)$), and the following B -block
23 structure is assumed:

24 (1.2)
$$n = n_1 + \dots + n_B, \quad x = (x_1, \dots, x_B) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_B}$$

$$h(x) = \sum_{t=1}^B h_t(x_t), \quad Ax = \sum_{t=1}^B A_t x_t,$$
25

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

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

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

Versions: v0.1 (Oct. 24, 2021), v0.2 (Dec. 9, 2021), v1.0 (Jun. 14, 2022), v2.0 (Jan. 3, 2023)

[†]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

33 complexity bounds that scale poorly with the numerical tolerance; (iii) small stepsize
 34 parameters; or (iv) a strong rank assumption about the last block A_B that implies
 35 $\text{Im}(A_B) \supseteq \{d\} \cup \text{Im}(A_1) \cup \dots \cup \text{Im}(A_{B-1})$ which we refer to as the *last block condition*.

36 Of the above drawbacks, (iv) is especially limiting. To illustrate this, we give a
 37 few applications where the last block condition, and hence (iv), does not hold:

38 \triangleright *Rank-deficient Quadratic Programming (RDQP)*. It is shown in [4] that the
 39 (non-proximal) ADMM diverges on the following three-block convex RDQP:

$$40 \quad \min_{x_1, x_2, x_3, x_4} \frac{1}{2} x_1^2$$

$$41 \quad \text{s.t.} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} x_3 + \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} x_4 = 0.$$

43 \triangleright *Distributed Finite-Sum Optimization (DFSO)*. Given a positive integer B ,
 44 consider:

$$45 \quad (1.3) \quad \min_{x_i \in \mathbb{R}^n} \left\{ \sum_{t=1}^B (f_t + h_t)(x_t) : x_t - x_B = 0, \quad t = 1, \dots, B-1 \right\}$$

46 where f_i is continuously differentiable, h_t is closed convex, and ∇f_t is Lip-
 47 schitz continuous for $t = 1, \dots, B$. It is easy to see¹ that (1.3) is a special
 48 case of (1.1) where we have $A_s = e_s \otimes I \in \mathbb{R}^{n(B-1) \times n}$ for $s = 1, \dots, B-1$,
 49 we have $A_B = -\mathbf{1} \otimes I \in \mathbb{R}^{n(B-1) \times n}$, and $d = 0$. Moreover, it is straightfor-
 50 ward to show that for $s = 1, \dots, B-1$ we have $\text{Im}(A_s) \cap \text{Im}(A_B) = 0$ but
 51 $\text{Im}(A_s) \setminus \{0\} \neq \emptyset$, which implies that $\text{Im}(A_s) \not\subseteq \text{Im}(A_B)$.

52 \triangleright *Decentralized AC Optimal Power Control (DAC-OPF)*. The convex version
 53 was first considered in [27] for the rectangular coordinate formulation, and
 54 the problem itself is considered one of the most important ones in power
 55 systems decision-making. The nonconvex version of DAC-OPF is a variant
 56 where h_t is the indicator of a convex region given by a finite number of com-
 57 plicated quadratic constraints and f_t is a nonconvex quadratic cost function.
 58 A discussion of the limitations induced by assuming any rank condition which
 59 implies the last block condition is given in [29].

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

$$63 \quad (1.4) \quad \mathcal{L}_{c_k}^\theta(x; p) := \phi(x) + (1 - \theta) \langle p, Ax - d \rangle + \frac{c_k}{2} \|Ax - d\|^2,$$

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

$$68 \quad (1.5) \quad x_t^k = \underset{u_t \in \mathbb{R}^{n_t}}{\text{argmin}} \left\{ \lambda \mathcal{L}_{c_k}^\theta(\dots, x_{t-1}^k, u_t, x_{t+1}^{k-1}, \dots; p^{k-1}) + \frac{1}{2} \|u_t - x_t^{k-1}\|^2 \right\},$$

70 and then update

$$71 \quad (1.6) \quad p^k = (1 - \theta)p^{k-1} + \chi c_k (Ax^k - d),$$

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

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

73 *Contributions.* For proper choices of the stepsize λ and a non-decreasing sequence of
 74 penalty parameters $\{c_k\}_{k \geq 1}$, it is shown that if the Slater-like condition²

75 (1.7)
$$\exists z_{\dagger} \in \text{int}(\text{dom } h) \text{ such that } Az_{\dagger} = d,$$

76 holds, then DP.ADMM has the following features:

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

78 (1.8)
$$\text{dist}(0, \nabla f(\bar{z}) + A^* \bar{q} + \partial h(\bar{z})) \leq \rho, \quad \|A\bar{z} - d\| \leq \eta$$

79 in $\mathcal{O}(\max\{\rho^{-3}, \eta^{-3}\})$ iterations;

80 \triangleright it introduces a novel approach for updating the penalty parameter c_k , instead
 81 of assuming that $c_k = c_1$ for every $k \geq 1$ and that c_1 is sufficiently large (such
 82 as in [3, 14, 15, 28, 31, 32]);

83 \triangleright it does not have any of the drawbacks mentioned in the sentences preceding
 84 equation (1.3).

85 *Related Works.* Since ADMM-type methods where f is convex have been well-studied
 86 in the literature (see, for example, [1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 23, 24, 25]), we make no
 87 further mention of them here. Instead, we discuss below ADMM-type methods where
 88 f is nonconvex.

89 Letting δ_S denote the indicator function of a convex set S (see Subsection 1.1),
 90 we first present a list of common assumptions in Table 1.1.

\mathcal{Q}	$f(z) = \sum_{t=1}^B f_t(z_t)$ for subfunctions $f_t : \text{dom } h_t \mapsto \mathbb{R}$.
\mathcal{R}_0	$\text{Im}(A_B) \supseteq \{d\} \cup \text{Im}(A_1) \cup \dots \cup \text{Im}(A_{B-1})$.
\mathcal{S}	The Slater-like assumption (1.7) holds.
\mathcal{P}	$h_i \equiv \delta_P$ for $i \in \{1, \dots, B\}$, where P is a polyhedral set.
\mathcal{F}	A point $x^0 \in \text{dom } h$ satisfying $Ax^0 = d$ is available as an input.

TABLE 1.1

Common nonconvex ADMM assumptions and regularity conditions.

91 Earlier developments on ADMM for solving nonconvex instances of (1.1) all as-
 92 sume that \mathcal{R}_0 hold, and the ones dealing with complexity establish an $\mathcal{O}(\varepsilon^{-2})$ iteration
 93 complexity, where $\varepsilon := \min\{\rho, \eta\}$. More specifically, [3, 13, 30, 31] present proximal
 94 ADMMs under the assumption $B = 2$, $h_B \equiv 0$, and assumption \mathcal{Q} holds for [3, 13, 30].
 95 Papers [14, 15, 20, 21] present (possibly linearized) ADMMs under the assumption that
 96 $B \geq 2$, $h_B \equiv 0$, and assumption \mathcal{Q} holds for [14, 20, 21].

97 We next discuss papers that do not assume the restrictive condition \mathcal{R}_0 in Ta-
 98 ble 1.1, and are based on ADMM approaches directly applicable to (1.1) or some
 99 reformulation of it. An early paper in this direction is [15], which establishes an
 100 $\mathcal{O}(\varepsilon^{-6})$ iteration-complexity bound for an ADMM-type method applied to a penalty
 101 reformulation of (1.1) that artificially satisfies \mathcal{R}_0 . On the other hand, development
 102 of ADMM-type methods directly applicable to (1.1) is considerably more challenging
 103 and only a few works have recently surfaced (see Table 1.2 below).

104 We now discuss some advantages of DP.ADMM compared to the other two pa-
 105 pers in Table 1.2. First, the method in [28] considers a small stepsize (proportional

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

Algorithm	θ	χ	Complexity	Assumptions	Adaptive c
LPADMM [32]	0	$(0, \infty)$	None	\mathcal{P}, \mathcal{S}	No
SDD-ADMM [28]	$(0, 1]$	$[-\frac{\theta}{4}, 0)$	$\mathcal{O}(\varepsilon^{-4})$	\mathcal{F}	No
DP.ADMM	$(0, 1]$	$(0, \pi_\theta]$	$\mathcal{O}(\varepsilon^{-3})$	\mathcal{S}	Yes

TABLE 1.2

Comparison of existing ADMM-type methods with DP.ADMM for finding ε -stationary points with $\varepsilon := \min\{\rho, \eta\}$ and $\pi_\theta = \theta^2/[2B(2-\theta)(1-\theta)]$ if $\theta \in (0, 1)$ and $\pi_\theta = 1$ if $\theta = 1$.

106 to η^2) linearized proximal gradient update while DP.ADMM considers a large step-
107 size (proportional to the inverse of the weak-convexity constant of f) proximal point
108 update as in (1.5). Second, the method in [28] requires a feasible initial point, i.e., a
109 point $z_0 \in \text{dom } h$ satisfying $Az_0 = d$, while DP.ADMM only requires that the initial
110 point be in $\text{dom } h$. Third, the methods in [28, 32] both require certain hyperparam-
111 eters (the penalty parameter in [28] and an interpolation parameter in [32]) to be
112 chosen in a range that is hard to compute, while DP.ADMM only requires its main
113 hyperparameter pair (χ, θ) to satisfy a simple inequality (see (2.6)). Moreover, [28]
114 does not specify an easily implementable rule for updating its method’s penalty pa-
115 rameter, while DP.ADMM does. Fourth, convergence of the method in [32] requires
116 h being the indicator of a polyhedral set, whereas DP.ADMM applies to any closed
117 convex function h . Fifth, in contrast to [28] and this work, [32] does not give a com-
118 plexity bound for its proposed method. Finally, [28] considers an unusual negative
119 stepsize for its Lagrange multiplier update — which justifies its moniker “scaled dual
120 descent ADMM” — whereas DP.ADMM considers a positive stepsize.

121 *Organization.* Subsection 1.1 presents some basic definitions and notation. Section 2
122 presents the proposed DP.ADMM in two subsections. The first one precisely describes
123 the problem of interest, while the second one states the static and dynamic DP.ADMM
124 variants and their iteration complexities. Section 3 and 4 present the main properties
125 of the static and dynamic DP.ADMM, respectively. Section 5 presents some prelim-
126 inary numerical experiments. Section 6 gives some concluding remarks. Finally, the
127 end of the paper contains several appendices.

128 **1.1. Notation and Basic Definitions.** Let \mathbb{R}_+ denote the set of nonnegative
129 real numbers, and let \mathbb{R}_{++} denote the set of positive real numbers. Let \mathbb{R}_n denote the
130 n -dimensional Hilbert space with inner product and associated norm denoted by $\langle \cdot, \cdot \rangle$
131 and $\|\cdot\|$, respectively. The direct sum (or Cartesian product) of a set of sets $\{S_i\}_{i=1}^n$
132 is denoted by $\prod_{i=1}^n S_i$.

133 The smallest positive singular value of a nonzero linear operator $Q : \mathbb{R}^n \rightarrow \mathbb{R}^l$ is
134 denoted by σ_Q^+ . For a given closed convex set $X \subset \mathbb{R}^n$, its boundary is denoted by
135 ∂X and the distance of a point $x \in \mathbb{R}^n$ to X is denoted by $\text{dist}_X(x)$. The indicator
136 function of X at a point $x \in \mathbb{R}^n$ is denoted by $\delta_X(x)$ which has value 0 if $x \in X$
137 and $+\infty$ otherwise. For every $z > 0$ and positive integer b , we denote $\log_b^+(z) :=$
138 $\max\{1, \lceil \log_b(z) \rceil\}$.

139 The domain of a function $h : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is the set $\text{dom } h := \{x \in \mathbb{R}^n :$
140 $h(x) < +\infty\}$. Moreover, h is said to be proper if $\text{dom } h \neq \emptyset$. The set of all lower
141 semi-continuous proper convex functions defined in \mathbb{R}^n is denoted by $\overline{\text{Conv}} \mathbb{R}^n$. The
142 set of functions in $\overline{\text{Conv}} \mathbb{R}^n$ which have domain $Z \subseteq \mathbb{R}^n$ is denoted by $\overline{\text{Conv}} Z$. The
143 ε -subdifferential of a proper function $h : \mathbb{R}^n \rightarrow (-\infty, \infty]$ is defined by

$$144 \quad (1.9) \quad \partial_\varepsilon h(z) := \{u \in \mathbb{R}^n : h(z') \geq h(z) + \langle u, z' - z \rangle - \varepsilon, \quad \forall z' \in \mathbb{R}^n\}$$

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

$$147 \quad N_C(z) := \{\xi \in \mathbb{R}^n : \langle \xi, u - z \rangle \leq \varepsilon, \quad \forall u \in C\}.$$

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

$$150 \quad (1.10) \quad \ell_\psi(z; \bar{z}) := \psi(\bar{z}) + \langle \nabla \psi(\bar{z}), z - \bar{z} \rangle \quad \forall z \in \mathbb{R}^n.$$

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

152 Iterates of a scalar quantity have their iteration number appear as a subscript, e.g.,
 153 c_ℓ , while non-scalar quantities have this number appear as a superscript, e.g., v^k , and
 154 \hat{p}^ℓ . For variables with multiple blocks, the block number appears as a subscript, e.g.,
 155 x_t^k and v_t^k . Finally, we define the following norm for any quantity $u = (u_1, \dots, u_B)$
 156 following a block structure as in (1.2):

$$157 \quad (1.11) \quad \|u\|_{\dagger} = \|(u_1, \dots, u_B)\|_{\dagger} := \sum_{t=1}^B \|u_t\|.$$

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

162 **2.1. Problem of Interest.** This subsection presents the problem of interest and
 163 the assumptions underlying it.

164 Denote the aggregated quantities

$$165 \quad (2.1) \quad \begin{aligned} x_{<t} &:= (x_1, \dots, x_{t-1}), & x_{>t} &:= (x_{t+1}, \dots, x_B), \\ x_{\leq t} &:= (x_{<t}, x_t), & x_{\geq t} &:= (x_t, x_{>t}), \end{aligned}$$

167 for every $x = (x_1, \dots, x_B) \in \mathcal{H}$. Our problem of interest is finding approximate
 168 stationary points of (1.1) under the following assumptions:

- 169 (A1) for every $t = 1, \dots, B$, we have $h_t \in \overline{\text{Conv}} \mathbb{R}^{n_t}$ and $\mathcal{H}_t := \text{dom } h_t$ is compact;
- 170 (A2) $A \neq 0$ and $\mathcal{F} := \{x \in \mathcal{H} : Ax = d\} \neq \emptyset$ where $\mathcal{H} := \mathcal{H}_1 \times \dots \times \mathcal{H}_B$;
- 171 (A3) h in (1.2) is K_h -Lipschitz continuous on \mathcal{H} for some $K_h \geq 0$;
- 172 (A4) for every $t = 1, \dots, B$, there exists $m_t \geq 0$ such that

$$173 \quad (2.2) \quad f(x_{<t}, \cdot, x_{>t}) + \delta_{\mathcal{H}_t}(\cdot) + \frac{m_t}{2} \|\cdot\|^2 \text{ is convex for all } x \in \mathcal{H};$$

- 175 (A5) f is differentiable on \mathcal{H} and, for every $t = 1, \dots, B - 1$, there exists $M_t \geq 0$
 176 such that

$$177 \quad (2.3) \quad \|\nabla_{x_t} f(x_{\leq t}, \tilde{x}_{>t}) - \nabla_{x_t} f(x_{\leq t}, x_{>t})\| \leq M_t \|\tilde{x}_{>t} - x_{>t}\| \quad \forall x, \tilde{x} \in \mathcal{H};$$

- 179 (A6) there exists $z_{\dagger} \in \mathcal{F}$ such that $d_{\dagger} := \text{dist}_{\partial \mathcal{H}}(z_{\dagger}) > 0$.

180 We now give a few remarks about the above assumptions. First, in view of the
 181 fact that \mathcal{H} is compact, the following scalars are bounded:

$$182 \quad (2.4) \quad \begin{aligned} D_{\dagger} &:= \sup_{z \in \mathcal{H}} \|z - z_{\dagger}\|, & G_f &:= \sup_{x \in \mathcal{H}} \|\nabla f(x)\|, \\ \underline{\phi} &:= \inf_{x \in \mathcal{H}} \phi(x), & \bar{\phi} &:= \sup_{x \in \mathcal{H}} \phi(x). \end{aligned}$$

183 Second, if f is a separable function, i.e., it is of the form $f(z) = f_1(z_1) + \dots + f_B(z_B)$,
 184 then each M_t can be chosen to be zero. Third, any function h given by (1.2) such that
 185 each h_t for $t = 1, \dots, B$ has the form $h_t = \tilde{h}_t + \delta_{Z_t}$, where \tilde{h}_t is a finite everywhere
 186 Lipschitz continuous convex function and Z_t is a compact convex set, clearly satisfies
 187 condition (A3) for some K_h .

188 For a given tolerance pair (ρ, η) , we define a (ρ, η) -stationary pair of (1.1) as being
 189 a pair $(\bar{z}, \bar{q}) \in \mathcal{H} \times \mathbb{R}^\ell$ satisfying (1.8). It is well known that the first-order necessary
 190 condition for a point $z \in \mathcal{H}$ to be a local minimum of (1.1) is that there exists $q \in \mathbb{R}^\ell$
 191 such that the stationary conditions

$$192 \quad 0 \in \nabla f(z) + A^*q + \partial h(z), \quad Az = d$$

194 hold. Hence, the requirements in (1.8) can be viewed as a direct relaxation of the
 195 above stationary conditions. For ease of future reference, we consider the following
 196 problem.

197 **Problem $\mathcal{S}_{\rho, \eta}$** : Find a (ρ, η) -stationary pair (\bar{z}, \bar{q}) satisfying (1.8).

198 We now make three remarks about Problem $\mathcal{S}_{\rho, \eta}$. First, (\bar{z}, \bar{q}) is a solution of
 199 Problem $\mathcal{S}_{\rho, \eta}$ if and only if there exists a residual $\bar{v} \in \mathbb{R}^n$ such that

$$200 \quad (2.5) \quad \bar{v} \in \nabla f(\bar{z}) + A^*\bar{q} + \partial h(\bar{z}), \quad \|\bar{v}\| \leq \rho, \quad \|A\bar{z} - d\| \leq \eta.$$

201 Second, condition (2.5) has been considered in many previous works (e.g., see [16,
 202 17, 18, 19, 22]). Third, in the case where $\|\cdot\| = \|\cdot\|_2$ and $\rho = \eta$, the stationarity
 203 condition in (1.8) implies the stationarity condition of the papers [15, 28] in Table 1.2.
 204 Specifically, [15, Definition 3.6] and [28, Definition 3.3] consider a pair $(z, q) \in \mathcal{H} \times \mathbb{R}^\ell$
 205 to be an ε -stationary pair if it satisfies

$$206 \quad \text{dist}(0, \nabla_{z_t} f(z_1, \dots, z_B) + A_t^*q + \partial h_t(z_t)) \leq \varepsilon, \quad \|Az - d\| \leq \varepsilon,$$

207 for every $t = 1, \dots, B$.

208 In the following subsection, we present a method (Algorithm 2.1) that computes
 209 a triple $(\bar{z}, \bar{q}, \bar{v})$ satisfying (2.5), and hence which guarantees that (\bar{z}, \bar{q}) is a solution
 210 of Problem $\mathcal{S}_{\rho, \eta}$.

211 **2.2. DP-ADMM.** We present DP-ADMM in two parts. The first part presents
 212 a static version of DP-ADMM which either (i) stops with a solution of Problem $\mathcal{S}_{\rho, \eta}$
 213 or (ii) signals that its penalty parameter is too small. The second part presents the
 214 (dynamic) DP-ADMM that repeatedly invokes the static version on an increasing
 215 sequence of penalty parameters.

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

$$217 \quad (2.6) \quad 2\chi B(2 - \theta)(1 - \theta) \leq \theta^2, \quad (\chi, \theta) \in (0, 1]^2.$$

218 For ease of reference and discussion, the pseudocode for the static DP-ADMM is given
 219 in Algorithm 2.1 below. Notice that the classic proximal ADMM iteration

$$220 \quad x_t^k = \underset{u^t \in \mathbb{R}^{n_t}}{\text{argmin}} \left\{ \lambda \mathcal{L}_c^0(x_{<t}^k, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|u_t - x_t^{k-1}\|^2 \right\}, \quad t = 1, \dots, B,$$

$$221 \quad p^k = p^{k-1} + c(Ax^k - d),$$

Algorithm 2.1 Static DP-ADMM

 Input: $x^0 \in \mathcal{H}$, $p^0 \in A(\mathbb{R}^n)$, $\lambda \in (0, 1/(2m)]$, $c > 0$;

 Require: m as in (2.7), $(\rho, \eta) \in \mathbb{R}_{++}^2$, (χ, θ) as in (2.6)

```

1: for  $k \leftarrow 1, 2, \dots$  do
   STEP 1 (prox update):
2:   for  $t \leftarrow 1, 2, \dots, B$  do
3:      $x_t^k \leftarrow \operatorname{argmin}_{u_t \in \mathbb{R}^{n_t}} \{ \lambda \mathcal{L}_c^\theta(x_{<t}^k, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|u_t - x_t^{k-1}\|^2 \}$ 
4:      $q^k \leftarrow (1 - \theta)p^{k-1} + c(Ax^k - d)$ 
   STEP 2a (successful termination check):
5:   for  $t \leftarrow 1, 2, \dots, B$  do
6:      $\delta_t^k \leftarrow \nabla_{x_t} f(x_{\leq t}^k, x_{>t}^k) - \nabla_{x_t} f(x_{\leq t}^k, x_{>t}^{k-1})$ 
7:      $v_t^k \leftarrow \delta_t^k + cA_t^* \sum_{s=t+1}^B A_s(x_s^k - x_s^{k-1}) - \frac{1}{\lambda}(x_t^k - x_t^{k-1})$ 
8:   if  $\|v^k\| \leq \rho$  and  $\|Ax^k - d\| \leq \eta$  then
9:     return  $(x^k, p^k, q^k, v^k)$ 
   STEP 2b (unsuccessful termination check):
10:  if  $k \equiv 0 \pmod{2}$  and  $k \geq 3$  then
11:     $\mathcal{S}_k^{(v)} \leftarrow \frac{2}{k+2} \sum_{i=k/2}^k \|v^i\|$ 
12:     $\mathcal{S}_k^{(f)} \leftarrow \frac{2}{k+2} \sum_{i=k/2}^k \|Ax^i - d\|$ 
13:    if  $\frac{1}{\rho} \cdot \mathcal{S}_k^{(v)} + \frac{1}{\eta} \sqrt{\frac{c^3}{k}} \cdot \mathcal{S}_k^{(f)} \leq 1$  then
14:      return  $(x^k, p^k, q^k, v^k)$ 
   STEP 3 (multiplier update):
15:    $p^k \leftarrow (1 - \theta)p^{k-1} + \chi c(Ax^k - d)$ 

```

222

 223 corresponds to the case of $(\chi, \theta) = (1, 0)$, where $c \geq 1$ is a fixed penalty parameter.

224 The next result describes the iteration complexity and some useful technical prop-
 225 erties of Algorithm 2.1. Its proof is given in Section 3.3, and it uses three sets of scalars.

 226 The first set is independent of (c, p^0) and is given by

$$\begin{aligned}
 M &:= \max_{1 \leq t \leq B} M_t, & m &:= \max_{1 \leq t \leq B} m_t, & \Delta_\phi &:= \bar{\phi} - \underline{\phi}, & \kappa_0 &:= \frac{2B^2(\lambda M + 1)}{\sqrt{\lambda}}, \\
 (2.7) \quad \kappa_1 &:= \frac{\chi \|A\| D_\dagger}{\theta}, & \kappa_2 &:= \frac{1}{\theta} \left[1 + \frac{2\chi D_\dagger (K_h + G_f)}{\theta d_\dagger \sigma_A^+} \right] + 1, \\
 \kappa_3 &:= \frac{108\kappa_2^2}{\chi^2}, & \kappa_4 &:= \frac{\theta d_\dagger \sigma_A^+}{\chi D_\dagger}, & \kappa_5 &:= 8(B-1) \|A\|_\dagger^2, & \kappa_6 &:= 3 + \frac{8\kappa_0^2 \Delta_\phi}{\kappa_4^2}.
 \end{aligned}$$

228

 229 where $(G_f, D_\dagger, \bar{\phi}, \underline{\phi})$, K_h , and (m_t, M_t) are as in (2.4), (A3), and (A4). The second
 230 set is dependent on a given lower bound \underline{c} on c and is given by

$$(2.8) \quad \tilde{\kappa}_{\underline{c}}^{(0)} := 2 \left(\sqrt{\Delta_\phi} + \frac{5\kappa_2}{\chi \sqrt{\underline{c}}} \right), \quad \tilde{\kappa}_{\underline{c}}^{(1)} := 3\kappa_5 [\tilde{\kappa}_{\underline{c}}^{(0)}]^2, \quad \tilde{\kappa}_{\underline{c}}^{(2)} := 3\kappa_0^2 [\tilde{\kappa}_{\underline{c}}^{(0)}]^2.$$

 231
 232

233 The third set is dependent on a given upper bound \mathcal{R} on $\|p^0\|/c$ and is given by

$$\begin{aligned}
234 \quad (2.9) \quad \xi_{\mathcal{R}}^{(0)} &:= \frac{8}{\kappa_4^2} \left[\frac{9\kappa_0^2(\mathcal{R} + \kappa_1)^2}{\chi^2} + \kappa_5 \Delta_\phi \right] + (1 - \theta)(\mathcal{R} + \kappa_1), \\
\xi_{\mathcal{R}}^{(1)} &:= \frac{72\kappa_5(\mathcal{R} + \kappa_1)^2}{\chi^2 \kappa_4^2}.
\end{aligned}$$

235
236
237 **PROPOSITION 2.1.** *Let $\mathcal{R} \geq 0$ and $\underline{c} > 0$ be given, and assume that the pair (c, p^0)*
238 *given to Algorithm 2.1 satisfies*

$$239 \quad (2.10) \quad \|p_0\| \leq c\mathcal{R}, \quad c \geq \underline{c}.$$

240 *Then, the following statements hold about the call to Algorithm 2.1:*

241 *(a) it terminates in a number of iterations bounded by*

$$(2.11)$$

$$242 \quad \mathcal{T}_c(\rho, \eta | \underline{c}, \mathcal{R}) := 48 \left(\left\{ \kappa_6 + \frac{\tilde{\kappa}_{\underline{c}}^{(1)}}{\rho^2} \right\} + \left\{ \xi_{\mathcal{R}}^{(0)} + \frac{\kappa_3}{\eta^2} + \frac{\tilde{\kappa}_{\underline{c}}^{(2)}}{\rho^2} \right\} c + \xi_{\mathcal{R}}^{(1)} c^2 \right),$$

244 *where (κ_3, κ_6) , $(\tilde{\kappa}_{\underline{c}}^{(1)}, \tilde{\kappa}_{\underline{c}}^{(2)})$, and $(\xi_{\mathcal{R}}^{(0)}, \xi_{\mathcal{R}}^{(1)})$ are as in (2.7), (2.8), and (2.9),*
245 *respectively;*

246 *(b) if it terminates successfully in Step 2a, then the first and third components of*
247 *its output quadruple $(\bar{z}, \bar{p}, \bar{q}, \bar{v})$ solve Problem $\mathcal{S}_{\rho, \eta}$;*

248 *(c) if c satisfies*

$$249 \quad (2.12) \quad c \geq \hat{c}(\rho, \eta | \underline{c}, \mathcal{R}) := \frac{1}{\underline{c}^2} \left[\mathcal{T}_{\underline{c}}(1, 1 | \underline{c}, \mathcal{R}) + \frac{\sqrt{\underline{c}^3 \cdot \mathcal{T}_{\underline{c}}(1, 1 | \underline{c}, \mathcal{R})}}{\min\{\rho, \eta\}} \right],$$

250 *where $\mathcal{T}_c(\rho, \eta | \underline{c}, \mathcal{R})$ is as in (a), then it must terminate successfully.*

251 We now make some remarks about Proposition 2.1. First, statement (c) implies
252 that Algorithm 2.1 terminates successfully if its penalty parameter c is sufficiently
253 large, i.e., $c = \Omega(\varepsilon^{-1})$ where $\varepsilon := \min\{\rho, \eta\}$. Moreover, if a penalty parameter c
254 satisfying (2.12) and the condition that $c = \mathcal{O}(\varepsilon^{-1})$ is known, then it follows from
255 Proposition 2.1(a) that the iteration complexity of Algorithm 2.1 for finding a solution
256 of Problem $\mathcal{S}_{\rho, \eta}$ is $\mathcal{O}(\varepsilon^{-3})$.

257 Since a penalty parameter c as in the above paragraph is nearly impossible to
258 compute, we next present an adaptive method, namely, Algorithm 2.2 below, which
259 adaptively increases the penalty parameter c , and whose overall number of iterations
260 is also $\mathcal{O}(\varepsilon^{-3})$.

261 Some comments about Algorithm 2.2 are in order. First, it employs a “warm-
262 start” type strategy for calling Algorithm 2.1 at each iteration ℓ . Specifically, the
263 input of the ℓ^{th} to Algorithm 2.1 is the pair $(\bar{z}^{\ell-1}, \bar{p}^{\ell-1})$ output by the previous call
264 to Algorithm 2.1. Second, the initial penalty parameter c_1 can be chosen to be any
265 positive scalar, in contrast to many of the methods listed in Section 1 where this
266 parameter must be chosen sufficiently large. Third, the initial point \bar{z}^0 only needs to
267 be in the domain of h and need not be feasible or near feasible. Finally, while the
268 initial Lagrange multiplier \bar{p}^0 is chosen to be zero, the analysis in this paper can be
269 carried out for any $\bar{p}^0 \in A(\mathbb{R}^n)$, at the cost of more complicated complexity bounds.

270 The next result, whose proof is given in Section 4, gives the complexity of Algo-
271 rithm 2.2 in terms of the total number of iterations of Algorithm 2.1 across all of its
272 calls.

Algorithm 2.2 DP.ADMM

 Input: $\bar{z}^0 \in \mathcal{H}$, $\lambda \in (0, 1/(2m)]$, $c_1 > 0$

 Require: m as in (2.7), $(\rho, \eta) \in (0, 1)^2$, (χ, θ) as in (2.6)

- 1: $\bar{p}^0 \leftarrow 0$
 - 2: **for** $\ell \leftarrow 1, 2, \dots$ **do**
 - 3: **call** Algorithm 2.1 with inputs $(x^0, p^0, \lambda, c) = (\bar{z}^{\ell-1}, \bar{p}^{\ell-1}, \lambda, c_\ell)$ and parameters m , (ρ, η) , and (χ, θ) to obtain an output quadruple $(\bar{z}^\ell, \bar{p}^\ell, \bar{q}^\ell, \bar{v}^\ell)$
 - 4: **if** $\|\bar{v}^\ell\| \leq \rho$ **and** $\|A\bar{z}^\ell - d\| \leq \eta$ **then**
 - 5: **return** $(\bar{z}^\ell, \bar{q}^\ell)$
 - 6: $c_{\ell+1} \leftarrow 2c_\ell$
-

 273 THEOREM 2.2. *Define the scalars*

274 (2.13) $T_1 := \mathcal{T}_{c_1}(1, 1 | c_1, 2\kappa_1), \quad \varepsilon := \min\{\rho, \eta\},$

 275 where κ_1 and $\mathcal{T}_c(\cdot, \cdot | \cdot, \cdot)$ are as in (2.7) and (2.11), respectively. Then, Algorithm 2.2
 276 stops and outputs a pair that solves Problem $\mathcal{S}_{\rho, \eta}$ in a number of iterations of Algo-
 277 rithm 2.1 bounded by

278 (2.14) $T_1 \left(2E_0^2 + \frac{E_0 + 2E_1^2}{\varepsilon^2} + \frac{E_1}{\varepsilon^3} \right)$

 279 where
 280 where

281 (2.15) $E_0 := 2 \left(1 + \frac{T_1^2}{c_1^3} \right), \quad E_1 := 2 \sqrt{\frac{T_1}{c_1^3}}.$

 282 Since $T_1 = \mathcal{O}(c_1^{-1})$ in view of (2.11) and (2.13), it follows from (2.14) and (2.15)
 283 that if $c_1^{-1} = \mathcal{O}(1)$, then the overall complexity of Algorithm 2.2 is $\mathcal{O}(\varepsilon^{-3})$.

 284 **3. Analysis of Algorithm 2.1.** This section presents the main properties of
 285 Algorithm 2.1, and it contains three subsections. More specifically, the first (resp.,
 286 second) subsection establishes some key bounds on the ergodic means of the sequences
 287 $\{\|v^k\|\}_{k \geq 0}$ and $\{\|Ax^k - d\|\}_{k \geq 0}$ (resp., the sequence $\{\|p_k\|\}_{k \geq 0}$). The third one proves
 288 Proposition 2.1.

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

293 (3.1)
$$\begin{aligned} a_\theta &= \theta(1 - \theta), & b_\theta &:= (2 - \theta)(1 - \theta), \\ \gamma_\theta &:= \frac{(1 - 2B\chi b_\theta) - (1 - \theta)^2}{2\chi}, & f^i &:= Ax^i - d, \end{aligned}$$

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

295 (3.2)
$$S_{j,k}^{(p)} := \frac{\sum_{i=j}^k \|p^i\|}{k - j + 1}, \quad S_{j,k}^{(v)} := \frac{\sum_{i=j}^k \|v^i\|}{k - j + 1}, \quad S_{j,k}^{(f)} := \frac{\sum_{i=j}^k \|f^i\|}{k - j + 1}.$$

 297 for every $j = 1, \dots, k$. Notice that $\gamma_\theta \geq \theta/\chi$ in view of (2.6). We also denote Δy^i to
 298 be the difference of iterates for any variable y at iteration i , i.e.,

299 (3.3)
$$\Delta y^i \equiv y^i - y^{i-1}.$$

300 **3.1. Properties of the Key Residuals.** This subsection presents bounds on
 301 the residuals $\{\|v^i\|\}_{i=2}^k$ and $\{\|f^i\|\}_{i=2}^k$ generated by Algorithm 2.1. These bounds will
 302 be particularly helpful for proving Proposition 2.1 in Subsection 3.3.

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

304 LEMMA 3.1. For $i = 1, \dots, k$,

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

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

$$307 \quad (3.4) \quad \|v^i\| \leq B \left(M + \frac{1}{\lambda} \right) \|\Delta x^i\|_{\dagger} + c \|A\|_{\dagger} \sum_{t=2}^B \|A_t \Delta x_t^i\|,$$

308 where $\|\cdot\|_{\dagger}$ is as in (1.11).

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

311 (b) We first prove the required inclusion. The optimality of x_t^k in Step 1 of
 312 Algorithm 2.1, and assumption (A4), imply that

$$313 \quad 0 \in \partial \left[\mathcal{L}_c^\theta(x_{\leq t}^i, \cdot, x_{> t}^{i-1}; p^{i-1}) + \frac{1}{2\lambda} \|\cdot - x_k^{i-1}\|^2 \right] (x^i)$$

$$314 \quad = \nabla_{x_t} f(x_{\leq t}^i, x_{> t}^{i-1}) + A_t^* [(1 - \theta)p^{i-1} + c[A(x_{\leq t}^i, x_{> t}^{i-1}) - d]] + \partial h_t(x_t^i) + \frac{1}{\lambda} \Delta x_t^i$$

$$315 \quad = \nabla_{x_t} f(x_{\leq t}^i, x_{> t}^{i-1}) + A_t^* \left(q^i - c \sum_{s=t+1}^B A_s \Delta x_s^i \right) + \partial h_t(x_t^i) + \frac{1}{\lambda} \Delta x_t^i$$

$$316 \quad = \nabla_{x_t} f(x^i) + A_t^* q^i + \partial h_t(x_t^i) - v_t^i.$$

318 for every $1 \leq t \leq B$. Hence, the inclusion holds. To show the inequality, let $1 \leq t \leq B$
 319 be fixed and use the triangle inequality, the definition of v_t^i , and assumption (A5) to
 320 obtain

$$321 \quad \|v_t^i\| \leq \|\nabla_{x_t} f(x_{\leq t}^i, x_{> t}^i) - \nabla_{x_t} f(x_{\leq t}^i, x_{> t}^{i-1})\| + c \sum_{s=t+1}^B \|A_t^* A_s \Delta x_s^i\| + \frac{1}{\lambda} \|\Delta x_t^i\|$$

$$322 \quad \leq M_t \|x_{> t}^i - x_{> t}^{i-1}\| + c \|A_t\| \sum_{s=t+1}^B \|A_s \Delta x_s^i\| + \frac{1}{\lambda} \|\Delta x_t^i\|$$

$$323 \quad \leq \left(M + \frac{1}{\lambda} \right) \sum_{s=t}^B \|\Delta x_s^i\| + c \|A_t\| \sum_{t=2}^B \|A_t \Delta x_t^i\|.$$

325 Summing the above bound from $t = 1$ to B , and using the definition of M in (2.7)
 326 and the triangle inequality, we conclude that

$$327 \quad \|v^i\| \leq \sum_{t=1}^B \|v_t^i\| \leq \left(M + \frac{1}{\lambda} \right) \sum_{t=1}^B \sum_{s=t}^B \|\Delta x_s^i\| + c \|A\|_{\dagger} \sum_{t=2}^B \|A_t \Delta x_t^i\|$$

$$328 \quad \leq B \left(M + \frac{1}{\lambda} \right) \|\Delta x^i\|_{\dagger} + c \|A\|_{\dagger} \sum_{t=2}^B \|A_t \Delta x_t^i\|. \quad \square$$

330 Notice that part (b) of the above result implies that $(\bar{x}, \bar{v}, \bar{p}) = (x^i, v^i, q^i)$ satisfies
 331 the inclusion in (2.5). Hence, if $\|v^i\|$ and $\|f^i\|$ are sufficiently small at some iteration

332 i , then Algorithm 2.1 clearly returns a solution of Problem $\mathcal{S}_{\rho,\eta}$ at iteration i , i.e.,
 333 Proposition 2.1(b) holds. However, to understand when Algorithm 2.1 terminates, we
 334 will need to develop more refined bounds on $\|v_i\|$ and $\|f_i\|$.

335 To begin, we present some relations between the perturbed augmented Lagrangian
 336 $\mathcal{L}_c^\theta(\cdot; \cdot)$ and the iterates $\{(x^i, p^i)\}_{i=1}^k$. For conciseness, its proof is given in Appendix A.

337 LEMMA 3.2. For $i = 1, \dots, k$,

- 338 (a) $\mathcal{L}_c^\theta(x^i; p^i) - \mathcal{L}_c^\theta(x^i; p^{i-1}) = b_\theta \|\Delta p^i\|^2 / (2\chi c) + a_\theta (\|p^i\|^2 - \|p^{i-1}\|^2) / (2\chi c)$;
 339 (b) $\mathcal{L}_c^\theta(x^i; p^{i-1}) - \mathcal{L}_c^\theta(x^{i-1}; p^{i-1}) \leq -\|\Delta x^i\|^2 / (2\lambda) - c \sum_{t=1}^B \|A_t \Delta x_t^i\|^2 / 2$;
 340 (c) if $i \geq 2$, it holds that

$$341 \quad (3.5) \quad \frac{b_\theta}{2\chi c} \|\Delta p^i\|^2 - \frac{c}{4} \sum_{t=1}^B \|A_t \Delta x_t^i\|^2 \leq \frac{\gamma_\theta}{4B\chi c} (\|\Delta p^{i-1}\|^2 - \|\Delta p^i\|^2).$$

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

344 LEMMA 3.3. For $j = 1, \dots, k$,

$$345 \quad (3.6) \quad \sum_{i=j+1}^k \|v^i\|^2 \leq (\kappa_0^2 + \kappa_5 c) [\Psi_j(c) - \Psi_k(c)],$$

346 where (κ_0, κ_5) is as in (2.7), and denoting $(a_\theta, \gamma_\theta)$ and as in (3.1), we have

$$348 \quad (3.7) \quad \Psi_i(c) := \mathcal{L}_c^\theta(x^i; p^i) - \frac{a_\theta}{2\chi c} \|p^i\|^2 + \frac{\gamma_\theta}{4B\chi c} \|\Delta p^i\|^2 \quad \forall i \geq 1.$$

350 *Proof.* Using the inequality $\|z\|_1^2 \leq n\|z\|_2^2$ for $z \in \mathbb{R}^n$ and (3.4), we first have that

$$351 \quad \sum_{i=j+1}^k \|v^i\|^2 \stackrel{(3.4)}{\leq} \sum_{i=j+1}^k \left[B \left(M + \frac{1}{\lambda} \right) \|\Delta x^i\|_\dagger + c \|A\|_\dagger \sum_{t=2}^B \|A_t \Delta x_t^i\| \right]^2$$

$$352 \quad \leq \sum_{i=j+1}^k 2B^2 \left(M + \frac{1}{\lambda} \right)^2 \|\Delta x^i\|_\dagger^2 + c^2 \|A\|_\dagger^2 \left(\sum_{t=2}^B \|A_t \Delta x_t^i\| \right)^2$$

$$353 \quad \leq \sum_{i=j+1}^k 2B^4 \left(M + \frac{1}{\lambda} \right)^2 \|\Delta x^i\|^2 + 2(B-1)c^2 \|A\|_\dagger^2 \sum_{t=2}^B \|A_t \Delta x_t^i\|^2$$

$$354 \quad (3.8) \quad \leq (\kappa_0^2 + \kappa_5 c) \sum_{i=j+1}^k \left[\frac{1}{2\lambda} \|\Delta x^i\|^2 + \frac{c}{4} \sum_{t=2}^B \|A_t \Delta x_t^i\|^2 \right].$$

356 Combining Lemma 3.2(a)–(c), the definition of Ψ_θ^i , and the bound $(a+b)^2 \leq 2a^2 + 2b^2$
 357 for $a, b \in \mathbb{R}_+$, we also have that

$$358 \quad \frac{1}{2\lambda} \|\Delta x^i\|^2 + \frac{c}{4} \sum_{t=2}^B \|A_t \Delta x_t^i\|^2$$

$$359 \quad \stackrel{\text{L.3.2(a)-(b)}}{\leq} \mathcal{L}_c^\theta(x^{j-1}; p^{j-1}) - \mathcal{L}_c^\theta(x^j; p^j) + \frac{a_\theta}{2\chi c} \Delta_{p,j}^{(2)} + \frac{b_\theta}{2\chi c} \|\Delta p^i\|^2 - \frac{c}{4} \sum_{t=1}^B \|A_t \Delta x_t^i\|^2$$

$$360 \quad \stackrel{\text{L.3.2(c)}}{\leq} \mathcal{L}_c^\theta(x^{j-1}; p^{j-1}) - \mathcal{L}_c^\theta(x^j; p^j) + \frac{a_\theta}{2\chi c} \Delta_{p,j}^{(2)} + \frac{\gamma_\theta}{4B\chi c} (\|\Delta p^{i-1}\|^2 - \|\Delta p^i\|^2)$$

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

where $\Delta_{p,j}^{(2)} := \|p^j\|^2 - \|p^{j-1}\|^2$. Consequently, summing the above inequality from $i = j + 1$ to k , and combining the resulting inequality with (3.8), yields the desired bound. \square

We now bound the quantity on the right-hand-side of (3.6)

- LEMMA 3.4. For any $j \geq 1$ and $k \geq 1$,
- (a) $\mathcal{L}_c^\theta(x^j; p^j) \leq \phi(x^j) + 3(\|p^j\|^2 + \|p^{j-1}\|^2)/(\chi^2 c)$;
 - (b) $\mathcal{L}_c^\theta(x^k; p^k) \geq \phi(x^k) - \|p^k\|^2/(2c)$;
 - (c) it holds that

$$(3.9) \quad \Psi_j(c) - \Psi_k(c) \leq \Delta_\phi + 4 \left(\frac{\|p^j\|^2 + \|p^{j-1}\|^2 + \|p^k\|^2}{\chi^2 c} \right),$$

where $\Psi_i(\cdot)$ and Δ_ϕ are as in (3.6) and (2.7), respectively.

Proof. (a)–(b) See Appendix A.

(c) Using parts (a)–(b), the fact that $a_\theta \in (0, 1)$ 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 $\gamma_\theta \leq 1/(2\chi)$, it holds that

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

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

PROPOSITION 3.5. For $j = 1, \dots, k - 1$,

$$(3.10) \quad S_{j+1,k}^{(f)} \leq \frac{\|p^j\| + 2S_{j+1,k}^{(p)}}{\chi c},$$

$$(3.11) \quad S_{j+1,k}^{(v)} \leq 2\sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k - j}} \left(\Delta_\phi^{1/2} + \frac{\|p^j\| + \|p^{j-1}\| + \|p^k\|}{\chi\sqrt{c}} \right),$$

where $(\kappa_0, \kappa_5, \Delta_\phi)$ is as in (2.7).

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

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

393 which is (3.10). On the other hand, to show (3.11), we use the definition of $S_{j+1,k}^{(v)}$,
 394 the fact that $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \in \mathbb{R}_+$, Lemma 3.3, and Lemma 3.4(c), to
 395 conclude that

$$\begin{aligned}
 396 \quad S_{j+1,k}^{(v)} &= \frac{\sum_{i=j+1}^k \|v^i\|}{k-j} \leq \left(\frac{\sum_{i=j+1}^k \|v^i\|^2}{k-j} \right)^{1/2} \\
 397 \quad &\stackrel{\text{L.3.3}}{\leq} \left(\frac{[\kappa_0^2 + \kappa_5 c][\Psi_j(c) - \Psi_k(c)]}{k-j} \right)^{1/2} \\
 398 \quad &\stackrel{\text{L.3.4(c)}}{\leq} \sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}} \left[\Delta_\phi + 4 \left(\frac{\|p^j\|^2 + \|p^{j-1}\|^2 + \|p^k\|^2}{\chi^2 c} \right) \right]^{1/2} \\
 399 \quad (3.12) \quad &\leq 2 \sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}} \left(\Delta_\phi^{1/2} + \frac{\|p^j\| + \|p^{j-1}\| + \|p^k\|}{\chi \sqrt{c}} \right). \quad \square
 \end{aligned}$$

401 Observe that both residuals $S_{j+1,k}^{(v)}$ and $S_{j+1,k}^{(f)}$ depend on the size of the Lagrange
 402 multipliers p^j , p^{j-1} , and p^k . If all the multipliers generated by Algorithm 2.1 could be
 403 shown to be bounded independent of c then it would be easy to see that (3.10)–(3.11)
 404 with $j = 1$ and some $c = \Theta(\eta^{-1})$ would imply the existence of some $k = O(\eta^{-1}\rho^{-2})$
 405 such that $[S_{2,k}^{(v)}/\rho] + [S_{2,k}^{(f)}/\eta] \leq 1$. Consequently, Algorithm 2.1 would find a solution
 406 of Problem $\mathcal{S}_{\rho,\eta}$ in $O(\eta^{-1}\rho^{-2})$ iterations.

407 Unfortunately, we do not know how to bound $\{\|p_i\|\}$ independent of c , so we
 408 will instead show the existence of $1 \leq j \leq k$ such that (i) indices j and $k - j$ are
 409 $\Theta(\eta^{-1}\rho^{-2})$ and (ii) the three multipliers p^j , p^{j-1} , and p^k are bounded. This fact and
 410 Proposition 3.5 suffice to show that the last (hypothetical) conclusion in the previous
 411 paragraph actually holds.

412 **3.2. Bounding the Lagrange Multipliers.** This subsection generalizes the
 413 analysis in [19]. More specifically, Proposition 3.8 shows that if k is sufficiently large
 414 relative to an index j , the penalty parameter c , and $\|p^0\|$, then $S_{j+1,k}^{(p)} = \mathcal{O}(1)$.

415 The proof of the first result can be found in [26, Lemma B.3] using the variable
 416 substitution $(q, q^-, \chi) = (q^i, [1 - \theta]p^{i-1}, c)$ and step 4 of Algorithm 2.1.

417 **LEMMA 3.6.** *For every $i \geq 1$ and $r \in \partial h(z^i) + A^*q^i$, it holds that*

$$418 \quad \|q^i\| \leq \max \left\{ (1 - \theta)\|p^{i-1}\|, \frac{2D_\dagger(K_h + \|r\|)}{d_\dagger\sigma_A^+} \right\}.$$

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

420 **LEMMA 3.7.** *For every $1 \leq j \leq k$,*
 421 *(a) $p^j = \chi q^j + (1 - \chi)(1 - \theta)p^{j-1}$;*
 422 *(b) $\|p^j\| \leq \|p^0\| + \kappa_1 c$;*
 423 *(c) it holds that*

$$424 \quad \frac{(1 - \theta)\|p^k\|}{k - j} + \theta S_{j+1,k}^{(p)} \leq \frac{(1 - \theta)\|p^j\|}{k - j} + \frac{2\chi D_\dagger [K_h + G_f + S_{j+1,k}^{(v)}]}{d_\dagger\sigma_A^+},$$

425 where K_h , d_\dagger , and (D_\dagger, G_f) are as in (A3), (A6), and (2.4), respectively.

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

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

$$\begin{aligned}
 430 \quad \|p^j\| &\leq (1 - \theta)\|p^{j-1}\| + \chi c \|f^j\| \leq (1 - \theta)^j \|p^0\| + \chi c \sum_{i=0}^{j-1} (1 - \theta)^i \|f^i\| \\
 431 \quad &\leq \|p^0\| + \chi c \|A\| \sup_{z \in \mathcal{H}} \|z - z_\dagger\| \sum_{i=0}^{\infty} (1 - \theta)^i \\
 432 \quad &= \|p^0\| + \frac{\chi c \|A\| D_\dagger}{\theta} = \|p^0\| + \kappa_1 c. \\
 433
 \end{aligned}$$

434 (c) Let $i \geq 1$ be fixed, define

$$435 \quad d_{\chi, \theta} := (1 - \theta)(1 - \chi),$$

436 and recall that Lemma 3.1(b) implies $v^i - \nabla f(x^i) \in \partial h(x^i) + A^* q^i$. Using Lemma 3.6
 437 with $r = v^i - \nabla f(x^i)$, the definition of G_f in (2.4), and part (a), we first have that

$$\begin{aligned}
 438 \quad \|p^i\| &\stackrel{(a)}{=} \|\chi q^i + d_{\chi, \theta} \cdot p^{i-1}\| \leq \chi \|q^i\| + d_{\chi, \theta} \|p^{i-1}\| \\
 439 \quad &\stackrel{\text{L.3.6}}{\leq} d_{\chi, \theta} \|p^{i-1}\| + \chi \max \left\{ (1 - \theta) \|p^{i-1}\|, \frac{2D_\dagger(K_h + \|v^i - \nabla f(x^i)\|)}{d_\dagger \sigma_A^+} \right\} \\
 440 \quad &\leq (1 - \theta)(1 - \chi) \|p^{i-1}\| + \chi \left[(1 - \theta) \|p^{i-1}\| + \frac{2D_\dagger(K_h + \|v^i - \nabla f(x^i)\|)}{d_\dagger \sigma_A^+} \right] \\
 441 \quad &\leq (1 - \theta) \|p^{i-1}\| + \frac{2\chi D_\dagger(K_h + \|\nabla f(x^i)\| + \|v^i\|)}{d_\dagger \sigma_A^+} \\
 442 \quad &\leq (1 - \theta) \|p^{i-1}\| + \frac{2\chi D_\dagger(K_h + G_f + \|v^i\|)}{d_\dagger \sigma_A^+}. \\
 443
 \end{aligned}$$

444 Summing the above inequality from $i = j + 1$ to k and dividing by $k - j$ yields the
 445 desired conclusion. \square

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

447 PROPOSITION 3.8. *Let $\mathcal{R} \geq 0$ and $\underline{c} > 0$ be given and suppose c and p^0 satisfy*
 448 *(2.10). Then, for any positive integers j and k such that $k - j \geq \kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2$,*
 449 *we have*

$$450 \quad S_{j+1, k}^{(p)} \leq \kappa_2,$$

451 where (κ_2, κ_6) and $(\xi_{\mathcal{R}}^{(0)}, \xi_{\mathcal{R}}^{(1)})$ are as in (2.7) and (2.9), respectively.

452 *Proof.* Using (2.10), (3.11), Lemma 3.7(b), and the relation $\sqrt{a} + \sqrt{b} \leq \sqrt{2(a + b)}$
 453 for $a, b \in \mathbb{R}_+$, we first have that

$$\begin{aligned}
 454 \quad S_{j+1, k}^{(v)} &\leq 2\sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k - j}} \left(\Delta_\phi^{1/2} + \frac{\|p^j\| + \|p^{j-1}\| + \|p^k\|}{\chi \sqrt{c}} \right) \\
 455 \quad &\leq \sqrt{\frac{4(\kappa_0^2 + \kappa_5 c)}{k - j}} \left(\Delta_\phi^{1/2} + \frac{3[\|p^0\| + \kappa_1 c]}{\chi \sqrt{c}} \right)
 \end{aligned}$$

$$\begin{aligned}
456 \quad & \leq \sqrt{\frac{4(\kappa_0^2 + \kappa_5 c)}{k-j}} \left(\Delta_\phi^{1/2} + \frac{3[\mathcal{R} + \kappa_1]\sqrt{c}}{\chi} \right) \\
457 \quad & \leq \sqrt{\frac{8(\kappa_0^2 + \kappa_5 c)}{k-j}} \left(\Delta_\phi + \frac{9[\mathcal{R} + \kappa_1]^2 c}{\chi^2} \right) \leq \kappa_4 \sqrt{\frac{\xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2}{k-j}}. \\
458 \quad &
\end{aligned}$$

459 Using the above bound, Lemma 3.7(b)–(c), our assumed bound on $k - j$, and the
460 definition of κ_2 , we conclude that

$$\begin{aligned}
461 \quad S_{j+1,k}^{(p)} & \leq \frac{2\chi D_\dagger(K_h + G_f)}{\theta d_\dagger \sigma_A^+} + \frac{(1-\theta)\|p^j\|}{\theta(k-j)} + \frac{S_{j+1,k}^{(v)}}{\kappa_4} \\
462 \quad & \leq \frac{2\chi D_\dagger(K_h + G_f)}{\theta d_\dagger \sigma_A^+} + \frac{(1-\theta)(\|p^0\| + \kappa_1 c)}{\theta(k-j)} + \sqrt{\frac{\kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2}{k-j}} \\
463 \quad & \leq \frac{2\chi D_\dagger(K_h + G_f)}{\theta d_\dagger \sigma_A^+} + \frac{(1-\theta)(\mathcal{R} + \kappa_1)c}{\theta(k-j)} + \sqrt{\frac{\kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2}{k-j}} \\
464 \quad & \leq \frac{2\chi D_\dagger(K_h + G_f)}{\theta d_\dagger \sigma_A^+} + \frac{\xi_{\mathcal{R}}^{(0)} c}{\theta(k-j)} + \sqrt{\frac{\kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2}{k-j}} \\
465 \quad & \leq \frac{1}{\theta} \left[1 + \frac{2\chi D_\dagger(K_h + G_f)}{\theta d_\dagger \sigma_A^+} \right] + 1 = \kappa_2. \quad \square \\
466 \quad &
\end{aligned}$$

467 We end this subsection by discussing some implications of the above results.
468 Suppose ζ is an integer satisfying $\zeta \geq \kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2 = \Theta(c^2)$. It then follows from
469 Proposition 3.8 that $S_{2,\zeta}^{(p)} = \mathcal{O}(1)$ and $S_{2\zeta,3\zeta}^{(p)} = \mathcal{O}(1)$. Since the minimum of a set of
470 scalars minorizes its average, there exist indices $j_0 \in \{2, \dots, \zeta\}$ and $k_0 \in \{2\zeta, \dots, 3\zeta\}$
471 such that $\|p^{j_0}\| = \mathcal{O}(1)$ and $\|p^{k_0}\| = \mathcal{O}(1)$. Using the fact that $k_0 - j_0 \geq \zeta$, the
472 above bounds, and (3.10)–(3.11) with $(j, k) = (j_0, k_0)$, it is reasonable to expect that
473 $S_{j_0+1,k_0}^{(f)} = \mathcal{O}(1/c)$ and $S_{j_0+1,k_0}^{(v)} = \mathcal{O}(\sqrt{c/\zeta})$. In the next section, we give the exact
474 steps of this argument and use the resulting bounds to prove Proposition 2.1.

475 **3.3. Proof of Proposition 2.1.** Before presenting the proof of Proposition 2.1,
476 we first give two technical results. The first one refines the bounds in Proposition 3.5
477 using Proposition 3.8, while the second one gives an important implication of (2.12).

478 **LEMMA 3.9.** *Let $\mathcal{R} \geq 0$ and $\underline{c} > 0$ be given and suppose (c, p^0) satisfies (2.10) for
479 some $\mathcal{R} \geq \iota$ and $\underline{c} > 0$. For any integer ζ such that $\zeta \geq \kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2$, there exist
480 $j \in \{3, \dots, \zeta\}$ and $k \in \{2\zeta + 1, \dots, 3\zeta\}$ satisfying*

$$481 \quad (3.13) \quad S_{j+1,k}^{(v)} \leq \tilde{\kappa}_{\underline{c}}^{(0)} \sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}}, \quad S_{j+1,k}^{(f)} \leq \frac{6\kappa_2}{\chi c}, \\
482 \quad$$

483 where $(\kappa_0, \kappa_2, \kappa_5)$ and $\tilde{\kappa}_0$ is as in (2.7) and (2.8), respectively.

484 *Proof.* Suppose $\zeta \in \mathbb{N}$ satisfies $\zeta \geq \kappa_6 + \xi_{\mathcal{R}}^{(0)} c + \xi_{\mathcal{R}}^{(1)} c^2$. Using Proposition 3.8 with
485 $(j, k) = (1, \zeta)$ it holds that there exists $3 \leq j \leq \zeta$ such that

$$486 \quad \|p^{j-1}\| + \|p^j\| \leq \frac{\sum_{i=3}^{\zeta} (\|p^{i-1}\| + \|p^i\|)}{\zeta - 2} \leq \frac{2 \sum_{i=2}^{\zeta} \|p^i\|}{\zeta - 2}$$

$$\begin{aligned}
487 \quad (3.14) \quad &= \frac{2(\zeta - 1)S_{2,\zeta}^{(p)}}{\zeta - 2} \leq 4S_{2,\zeta}^{(p)} \leq 4\kappa_2. \\
488
\end{aligned}$$

489 On the other hand, using Proposition 3.8 with $(j, k) = (2\zeta, 3\zeta)$ it holds that there
490 exists $k \in \{2\zeta + 1, \dots, 3\zeta\}$ such that

$$491 \quad (3.15) \quad \|p^k\| \leq \frac{\sum_{i=2\zeta+1}^{3\zeta} \|p^i\|}{\zeta} = S_{2\zeta+1, 3\zeta} \leq \kappa_2.$$

492 Combining (3.14), (3.15), and Proposition 3.5, it follows that

$$\begin{aligned}
493 \quad S_{j+1,k}^{(v)} &\leq 2\sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}} \left(\Delta_\phi^{1/2} + \frac{\|p^{j_0}\| + \|p^{j_0-1}\| + \|p^{k_0}\|}{\chi\sqrt{c}} \right) \\
494 \quad &\stackrel{(3.14)-(3.15)}{\leq} 2\sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}} \left(\Delta_\phi^{1/2} + \frac{5\kappa_2}{\chi\sqrt{c}} \right) \\
495 \quad &\leq 2\sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}} \left(\Delta_\phi^{1/2} + \frac{5\kappa_2}{\chi\sqrt{c}} \right) = \tilde{\kappa}_{\underline{c}}^{(0)} \sqrt{\frac{\kappa_0^2 + \kappa_5 c}{k-j}}, \\
496
\end{aligned}$$

497 which is the first bound in (3.13). To show the other bound in (3.13), we use (3.14)
498 and Proposition 3.8 to conclude that

$$499 \quad S_{j+1,k}^{(f)} \leq \frac{\|p^j\| + 2S_{j+1,k}^{(p)}}{\chi c} \leq \frac{6\kappa_2}{\chi c}. \quad \square$$

500 We now state a technical result which will be used in the proof of Proposi-
501 tion 2.1(c).

502 LEMMA 3.10. For any $\mathcal{R} \geq 0$ and $c \geq \underline{c} > 0$, the following statements hold:

503 (a) the quantity $\mathcal{T}_c(\cdot, \cdot | \cdot, \cdot)$ defined in (2.11) satisfies

$$504 \quad \mathcal{T}_c(\rho, \eta | \underline{c}, \mathcal{R}) \leq \left[\left(\frac{c}{\underline{c}} \right)^2 + \frac{c}{\underline{c} \cdot \min\{\rho^2, \eta^2\}} \right] \mathcal{T}_{\underline{c}}(1, 1 | \underline{c}, \mathcal{R});$$

505 (b) if c satisfies (2.12), then $\mathcal{T}_c(\rho, \eta | \underline{c}, \mathcal{R}) \leq c^3$.

506 *Proof.* (a) This statement follows immediately from the definition of $\mathcal{T}_c(\cdot, \cdot | \cdot, \cdot)$
507 and the fact that for any $c \geq \bar{c}$ any nonnegative scalars α, β , and γ , we have

$$508 \quad \alpha + \beta c \leq (\alpha + \beta \underline{c}) \left(\frac{c}{\underline{c}} \right), \quad \alpha + \beta c + \gamma c^2 \leq (\alpha + \beta \underline{c} + \gamma \underline{c}^2) \left(\frac{c}{\underline{c}} \right)^2.$$

509 (b) Define $\hat{c} := \hat{c}(\rho, \eta | \underline{c}, \mathcal{R})$, $\varepsilon := \min\{\rho, \eta\}$, and $T := \mathcal{T}_{\underline{c}}(1, 1 | \underline{c}, \mathcal{R})$, and assume
510 that c satisfies (2.12), or equivalently, $c \geq \hat{c}$. To show the conclusion of (b), it suffices
511 to show that

$$512 \quad (3.16) \quad \left[\left(\frac{c}{\underline{c}} \right)^2 + \frac{c}{\underline{c} \cdot \varepsilon^2} \right] T \leq c^3.$$

513 in view of part (a). It is easy to see that the above inequality is satisfied by any c
514 such that

$$515 \quad c \geq \pi_\varepsilon := \frac{T/\underline{c}^2 + \sqrt{T^2/\underline{c}^4 + 4T/(\varepsilon^2 \underline{c})}}{2}.$$

516 Since the definition of \hat{c} in (2.12) and the relation $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \in \mathbb{R}_+$
517 imply that $\hat{c} \geq \pi_\varepsilon$, the conclusion of (b) follows from the assumption that $c \geq \hat{c}$ and
518 the previous observation. \square

519 We now remark on Lemma 3.9. For any integer $\zeta \geq \kappa_6 + \xi_{\mathcal{R}}^{(0)}c + \xi_{\mathcal{R}}^{(1)}c^2$, it follows
520 that there exist $i_1, i_2 \leq 3\zeta$ such that $\|v_{i_1}\| = \mathcal{O}(\sqrt{c/\zeta})$ and $\|f_{i_2}\| = \mathcal{O}(1/c)$. Hence,
521 for some $c = \Theta(\eta^{-1})$ and some $\zeta \geq \Omega(\rho^{-2}\eta^{-1})$, we can guarantee that $\|v_{i_1}\| \leq \rho$
522 and $\|f_{i_2}\| \leq \eta$. Clearly, if $i_1 = i_2$ then this argument shows that a solution of
523 Problem $\mathcal{S}_{\rho, \eta}$ can be found in $\mathcal{O}(\rho^{-2}\eta^{-1})$ iterations of Algorithm 2.1. In the proof (of
524 Proposition 2.1) below, we give a more involved argument that guarantees that the
525 above i_1 and i_2 can be chosen so that $i_1 = i_2$.

526 *Proof of Proposition 2.1.* (a) Let $(\rho, \eta) \in \mathbb{R}_{++}^2$, $p^0 \in A(\mathbb{R}^n)$, and $c > 0$ be given,
527 and define

$$528 \quad T := \mathcal{T}_c(\rho, \eta | \underline{c}, \mathcal{R}), \quad r_j := \frac{\mathcal{S}_j^{(v)}}{\rho} + \frac{\mathcal{S}_j^{(f)}}{\eta} \sqrt{\frac{c^3}{j}} \quad \forall j \geq 1,$$

529 where $\mathcal{S}_j^{(v)}$ and $\mathcal{S}_j^{(f)}$ are as in Step 2b of Algorithm 2.1 and $\mathcal{T}_c(\cdot, \cdot | \cdot, \cdot)$ is as in (2.11).
530 For the sake of contradiction, suppose that Algorithm 2.1 has not terminated by the
531 end of iteration $k = T$. Since Algorithm 2.1 (see its Step 2b) terminates unsuccessfully
532 at iteration k exactly when $r_k \leq 1$, we will obtain the desired contradiction by showing
533 that there exists $k \leq T$ such that $r_k \leq 1$.

534 First, consider an arbitrary pair of integers j and k such that $1 \leq j \leq k \leq T$
535 and assume without loss of generality that k is even. Then, combining (3.18), the
536 relations $\mathcal{S}_{k/2, k}^{(v)} = \mathcal{S}_k^{(v)}$ and $\mathcal{S}_{k/2, k}^{(f)} = \mathcal{S}_k^{(f)}$, we easily see that

$$537 \quad r_k = \frac{\mathcal{S}_{k/2, k}^{(v)}}{\rho} + \frac{c^{3/2}\mathcal{S}_{k/2, k}^{(f)}}{\eta\sqrt{k}} = \frac{k-j+1}{k-k/2+1} \left[\frac{\mathcal{S}_{j, k}^{(v)}}{\rho} + \frac{c^{3/2}\mathcal{S}_{j, k}^{(f)}}{\eta\sqrt{k}} \right]$$

$$538 \quad (3.17) \quad \leq \frac{k+2}{k/2+1} \left[\frac{\mathcal{S}_{j, k}^{(v)}}{\rho} + \frac{c^{3/2}\mathcal{S}_{j, k}^{(f)}}{\eta\sqrt{k}} \right] = 2 \left[\frac{\mathcal{S}_{j, k}^{(v)}}{\rho} + \frac{c^{3/2}\mathcal{S}_{j, k}^{(f)}}{\eta\sqrt{k}} \right],$$

$$539$$

540 We now show that there exists suitable j and k so that the last expression is bounded
541 by 1 and hence that our desired contradiction follows. Note first that the definition
542 of $T = \mathcal{T}_c(\rho, \eta)$ in (2.11) implies that $\zeta := T/3$ satisfies the assumption of Lemma 3.9.
543 Hence, the conclusion of this lemma implies the existence of $j \in \{3, \dots, T/3\}$ and
544 $k \in \{2T/3 + 1, \dots, T\}$ such that

$$545 \quad \frac{\mathcal{S}_{j, k}^{(v)}}{\rho} + \frac{c^{3/2}\mathcal{S}_{j, k}^{(f)}}{\eta\sqrt{k}} \leq \frac{\tilde{\kappa}_{\underline{c}}^{(0)}\sqrt{\kappa_0^2 + \kappa_5 c}}{\rho\sqrt{k-j}} + \frac{6\kappa_2\sqrt{c}}{\chi\eta\sqrt{k}} \leq \frac{\tilde{\kappa}_{\underline{c}}^{(0)}\sqrt{\kappa_0^2 + \kappa_5 c}}{\rho\sqrt{T/3}} + \frac{6\kappa_2\sqrt{c}}{\chi\eta\sqrt{T/3}}$$

$$546 \quad (3.18) \quad = \sqrt{\frac{\tilde{\kappa}_1 + \tilde{\kappa}_2 c}{\rho^2 T}} + \sqrt{\frac{\kappa_3 c}{\eta^2 T}} \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$547$$

548 where the last inequality follows from the definition of T . Combining (3.17) and (3.18)
549 we conclude that $r_k \leq 1$, which yields our desired contradiction.

550 (b) This follows immediately from the stopping condition in Step 2a of Algo-
551 rithm 2.1 and Lemma 3.1(b).

552 (c) Let (T, r_k) be as in part (a) and assume that c satisfies (2.12). Assume, for
553 contradiction, that Algorithm 2.1 does not terminate successfully. Then, by part (a),

554 the algorithm terminates in an iteration $k \leq T$ such that $r_k \leq 1$. Using the fact that
 555 r_k itself is an average of scalars, there exists $k/2 \leq i \leq k$ such that

$$556 \quad \frac{\|v^i\|}{\rho} + \frac{c^{3/2}\|f^i\|}{\eta\sqrt{k}} \leq \frac{S_{k/2,k}^{(v)}}{\rho} + \frac{c^{3/2}S_{k/2,k}^{(f)}}{\eta\sqrt{k}} \leq 1.$$

557 Hence, it holds that $\|v^i\| \leq \rho$ and $\|f^i\| \leq \eta\sqrt{k}c^{-3/2} \leq \eta\sqrt{T}c^{-3/2}$ where the last
 558 inequality is due to the fact that $k \leq T$. Moreover, the assumption that c satisfies
 559 (2.12) together with Lemma 3.10(b) then imply that $T \leq c^3$ and, hence, that $\|f^i\| \leq$
 560 η . Consequently, this means that the algorithm actually terminates successfully at
 561 iteration $i \leq k$. We have thus established the desired contradiction and, hence, that
 562 part (c) holds. \square

563 **4. Analysis of Algorithm 2.2.** This section presents the main properties of
 564 Algorithm 2.2, including the proof of Theorem 2.2.

565 We first start with two crucial technical results.

566 **PROPOSITION 4.1.** *The following statements hold about the ℓ^{th} iteration of Algo-*
 567 *rithm 2.2:*

- 568 (a) $\|\bar{p}^{\ell-1}\|/c_\ell \leq 2\kappa_1$, where κ_1 is as in (2.7);
 569 (b) its call to Algorithm 2.1 terminates in $\mathcal{T}_{c_\ell}(\rho, \eta | c_1, 2\kappa_1)$ iterations and, if the
 570 ℓ^{th} penalty parameter $c_\ell > 0$ satisfies

$$571 \quad (4.1) \quad c_\ell \geq \hat{c}(\rho, \eta | c_1, 2\kappa_1),$$

572 then this call terminates successfully, where κ_1 , $\mathcal{T}_c(\cdot, \cdot | \cdot, \cdot)$, and $\hat{c}(\cdot, \cdot | \cdot, \cdot)$ are
 573 as in (2.7), (2.11), and (2.12), respectively.

574 *Proof.* (a) We proceed by induction. Since $\bar{p}^0 = 0$, the case of $\ell = 1$ is immediate.
 575 Suppose the statement holds for some iteration ℓ and, hence, that $\|\bar{p}^{\ell-1}\| \leq 2\kappa_1 c_\ell$.
 576 Then, it follows from Lemma 3.7(b) with $(p^0, c) = (\bar{p}^{\ell-1}, c_\ell)$ and the relation $c_{\ell+1} =$
 577 $2c_\ell$ that

$$578 \quad \|\bar{p}^\ell\| \leq \|\bar{p}^{\ell-1}\| + \kappa_1 c_\ell \leq 2\kappa_1 c_\ell + \kappa_1 c_\ell = 3\kappa_1 c_\ell = \frac{3\kappa_1}{2} c_{\ell+1} < 2\kappa_1 c_{\ell+1}.$$

579 (b) This follows from part (a), the fact that $\{c_\ell\}_{\ell \geq 1}$ is an increasing sequence,
 580 and Proposition 2.1 with $(c, \underline{c}, \mathcal{R}) = (c_\ell, c_1, 2\kappa_1)$. \square

581 We are now ready to give the proof of Theorem 2.2.

582 *Proof of Theorem 2.2.* Define the scalars

$$583 \quad \hat{c} := \hat{c}(\rho, \eta | c_1, 2\kappa_1), \quad \hat{\ell} := \lceil \log_2^+(\hat{c}/c_1) \rceil, \quad \mathcal{T}_{c_\ell} := \mathcal{T}_{c_\ell}(\rho, \eta | c_1, 2\kappa_1),$$

585 where $\hat{c}(\cdot, \cdot | \cdot, \cdot)$ is as in (2.12). Proposition 4.1(b) and the update rule for c_ℓ imply
 586 that Algorithm 2.2 performs at most $\hat{\ell}$ iterations, and terminates with a pair that
 587 solves Problem $\mathcal{S}_{\rho, \eta}$. Moreover, the total number of iterations of Algorithm 2.1 (per-
 588 formed by all of Algorithm 2.2's calls to it) is bounded by $\sum_{\ell=1}^{\hat{\ell}} \mathcal{T}_{c_\ell}$. Now, using
 589 Lemma 3.10(a) with $\underline{c} = c_1$, it follows that

$$590 \quad (4.2) \quad \frac{\sum_{\ell=1}^{\hat{\ell}} \mathcal{T}_{c_\ell}}{T_1} \leq \frac{\sum_{\ell=1}^{\hat{\ell}} c_\ell^2}{c_1^2} + \frac{\sum_{\ell=1}^{\hat{\ell}} c_\ell}{c_1 \varepsilon^2} = \sum_{\ell=1}^{\hat{\ell}} 2^{2(\ell-1)} + \frac{\sum_{\ell=1}^{\hat{\ell}} 2^{(\ell-1)}}{\varepsilon^2} \leq 4^{\hat{\ell}} + \frac{2^{\hat{\ell}}}{\varepsilon^2},$$

591 where (T_1, ε) are as in (2.13). We now derive suitable bounds for $4^{\hat{\ell}}$ and $2^{\hat{\ell}}$. Using
 592 the definitions of \hat{c} and $\hat{\ell}$, and the definition of (E_0, E_1) in (2.15), we first have that

$$593 \quad 2^{\hat{\ell}} \leq \max \left\{ 2, 2^{(1+\log_2 \hat{c}/c_1)} \right\} \leq 2 \max \left\{ 1, \frac{\hat{c}}{c_1} \right\} = 2 \max \left\{ 1, \frac{1}{c_1^3} \left(T_1 + \frac{\sqrt{c_1^3 T_1}}{\varepsilon} \right) \right\}$$

$$594 \quad (4.3) \quad \leq 2 \left(1 + \frac{T_1}{c_1^3} + \frac{1}{\varepsilon} \sqrt{\frac{T_1}{c_1^3}} \right) = E_0 + \frac{E_1}{\varepsilon}.$$
 595

596 Combining the above inequality above with the bound $(a+b)^2 \leq 2a^2+2b^2$ for $a, b \in \mathbb{R}$,
 597 it is also easy to see that

$$598 \quad (4.4) \quad 4^{\hat{\ell}} \leq (2^{\hat{\ell}})^2 \leq 2E_0^2 + \frac{2E_1^2}{\varepsilon^2}.$$
 599

600 The conclusion now follows by applying (4.4) and (4.3) to (4.2). □

601 **5. Numerical Experiments.** This section examines the performance of the
 602 proposed DP-ADMM (Algorithm 2.2) for finding stationary points of a nonconvex
 603 three-block distributed quadratic programming problem. Specifically, given a radius
 604 $\gamma > 0$ and a dimension $n \in \mathbb{N}$, it considers the three-block problem

$$605 \quad \min_{(x_1, x_2, x_3) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n} - \sum_{i=1}^2 \left[\frac{\alpha_i}{2} \|x_i\|^2 + \langle x_i, \beta_i \rangle \right]$$

$$606 \quad \text{s.t. } \|x\|_\infty \leq \gamma,$$

$$607 \quad x_1 - x_3 = 0,$$

$$608 \quad x_2 - x_3 = 0,$$

610 where $\{\alpha_i\}_{i=1}^2 \subseteq [0, 1]$, $\{\beta_i\}_{i=1}^2 \subseteq [0, 1]^n$, and the entries of these quantities are
 611 sampled from the uniform distribution on $[0, 1]$. It is clear that the above problem is
 612 an instance of (1.1) if we take h_i to be the indicator of the set $\{x \in \mathbb{R}^n : \|x\|_\infty \leq \gamma\}$
 613 for $i = 1, \dots, 3$. At the end of this section, we give some elucidating remarks.

614 Before presenting the results, we first describe the algorithms tested. The first
 615 set of algorithms, labeled DP1–DP2, are modifications of Algorithm 2.2. Specifically,
 616 both DP1 and DP2 replace the original definition of $\mathcal{S}_k^{(f)}$ (resp. $\mathcal{S}_k^{(f)}$) in Step 2b
 617 of Algorithm 2.1 with $2 \sum_{i=1}^k \|v^i\|/[k+2]$ (resp. $2 \sum_{i=1}^k \|Ax^i - d\|/[k+2]$) and
 618 choose $(\lambda, c_1) = (1/2, 1)$. Moreover, DP1 chooses $(\theta, \chi) = (0, 1)$ while DP2 chooses
 619 $(\theta, \chi) = (1/2, 1/18)$ which satisfies (2.6) at equality. The second set of algorithms,
 620 labeled SDD1–SDD3, are instances of the SDD-ADMM of [28] for different values
 621 of the penalty parameter ρ . Specifically, all of these instances uses the parameters
 622 $(\omega, \theta, \tau) = (4, 2, 1)$, following the same choice as in [28, Section 5.1], and select the fol-
 623 lowing curvature constants: $(M_h, K_h, J_h, L_h) = (4\gamma, 1, 1, 0)$. Moreover, SDD1–SDD3
 624 respectively choose the penalty parameter ρ to be 0.1, 1.0, and 10.0, and termination
 625 of the method occurs when the norm of the stationary residual ξ^k and feasibility are
 626 both less than a given numerical tolerance.

627 The results of our experiment are now given in Tables 5.1–5.2, which present
 628 both iteration counts and runtimes for either varying choices of γ (Table 5.1) or n
 629 (Table 5.2). We now describe a few more details about these experiments and tables.
 630 First, the starting point for all methods is the zero vector and the numerical tolerances
 631 (e.g., ρ and η in DP1–DP2) for each method were set to be 10^{-9} . Second, the bolded

632 text in the tables highlight the method that performed the best in terms of iteration
633 count. Third, we imposed an iteration limit of 100,000 and marked the runs which
634 did not terminate by this limit with a ‘-’ symbol. Fourth, the experiments were
635 implemented and executed in Matlab R2021b on a Windows 64-bit desktop machine
636 with 12GB of RAM and two Intel(R) Xeon(R) Gold 6240 processors, and the code is
637 readily available online³.

γ	Iteration Count					Runtime (ms)				
	DP1	DP2	SDD1	SDD2	SDD3	DP1	DP2	SDD1	SDD2	SDD3
10^0	21	29	363	135	528	1.8	1.9	38.2	13.4	50.4
10^1	76	83	427	223	976	4.0	4.9	41.3	22.4	88.1
10^2	151	156	497	309	1394	7.9	7.7	45.2	28.3	121.7
10^3	228	232	569	399	1855	10.8	10.8	51.2	34.3	159.3
10^4	306	308	647	489	2316	15.5	17.6	58.9	42.9	223.1
10^5	385	385	-	581	2778	17.9	18.5	-	48.0	241.5

TABLE 5.1
Results with $n = 10$ and different values of γ

n	Iteration Count					Runtime (ms)				
	DP1	DP2	SDD1	SDD2	SDD3	DP1	DP2	SDD1	SDD2	SDD3
10	151	156	497	309	1394	7.8	7.5	65.8	29.0	121.8
40	55	60	-	-	3117	3.7	3.5	-	-	319.0
160	139	144	-	388	1836	8.5	8.2	-	42.0	202.7
640	53	54	-	349	16243	4.0	3.9	-	40.4	1901.5
2560	58	59	-	458	8464	7.1	6.7	-	77.4	1553.7
10240	108	110	-	1058	4334	44.4	40.3	-	623.5	2790.6

TABLE 5.2
Results with $\gamma = 100$ and different values of n

638 From the results in Tables 5.1–5.2, we see that DP1 performed the best in terms
639 of iteration count and DP2 had iteration counts that were close to DP1. On the other
640 hand, SDD2 outperformed its other SDD-ADMM variant on all problems except one.
641 Finally, notice that the DP-ADMM variants scaled better against the dimension n
642 compared to the SDD-ADMM variants.

643 To close this section, we give some elucidating remarks. First, we excluded the
644 algorithm in [15] due to its poor iteration complexity bound and the fact that it is an
645 algorithm applied to a reformulation of (1.1) rather than to (1.1) directly. Second,
646 we had to choose different values of the penalty parameter ρ for the SDD-ADMM
647 variants because the analysis in [28] did not present a practical way of adaptively
648 updating ρ (note that the “adaptive” method in [28, Algorithm 3.2] is not practical
649 because it requires an estimate of $\sup_{x \in \mathcal{H}} \phi(x) - \inf_{x \in \mathcal{H}} \phi$ for (1.1)).

650 **6. Concluding Remarks.** The analysis of this paper also applies to instances
651 of (1.1) where f is not necessarily differentiable on \mathcal{H} as in our condition (A5), but
652 instead satisfies a more relaxed version of (A5), namely: for every $x \in \mathcal{H}$, the function
653 $f(x_{<t}, \cdot, x_{>t})$ has a Fréchet subgradient at x_t , denoted by $\nabla_{x_t} f(x_{<t}, x_{>t})$, and (2.3)
654 is satisfied for every $t = 1, \dots, B - 1$. Hence, our analysis immediately applies to

³See https://github.com/wwkong/nc_opt/tree/master/tests/papers/dp_admm.

655 the case where $f(z)$ is of the form $\sum_{t=1}^B f_t(z_t)$ in which, for every $t = 1, \dots, B$, the
656 function $f_t(\cdot) + m_t \|\cdot\|^2/2 + \delta_{\mathcal{H}_t}(\cdot)$ is convex and has a subgradient everywhere in \mathcal{H}_t .

657 We now discuss some possible extensions of our analysis in this paper. First,
658 our analysis was done under the assumption that \mathcal{H} is bounded (see (A3)), but
659 it is straightforward to see that it is still valid under the weaker assumption that
660 $\sup_{k \geq 1} \|x^k - z_\dagger\| \leq D_\dagger$ for some $D_\dagger > 0$ where z_\dagger is as in (A6). It would be interest-
661 ing to extend the analysis in this paper to the case where \mathcal{H} is unbounded, possibly
662 by assuming conditions on the sublevel sets of ϕ which guarantee that the aforemen-
663 tioned bound holds. Second, the convergence of Algorithm 2.2 is established under
664 the assumption that exact solutions to the subproblems in Step 1 of Algorithm 2.1
665 are easy to obtain. We believe that convergence can also be established when only
666 inexact solutions, e.g.,

$$667 \quad (6.1) \quad x_t^k \approx \operatorname{argmin}_{u_t \in \mathbb{R}^{n_t}} \left\{ \lambda \mathcal{L}_c^\theta(x_{<t}^k, u_t, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|u_t - x_t^{k-1}\|^2 \right\}$$

668 are available. For example, one could consider applying an accelerated composite
669 gradient (ACG) method to the problem associated with (6.1) so that x_t^k satisfies

$$670 \quad \exists r_t^k \quad \text{s.t.} \quad \begin{cases} r_t^k \in \partial \left(\lambda \mathcal{L}_c^\theta(x_{<t}^k, \cdot, x_{>t}^{k-1}; p^{k-1}) + \frac{1}{2} \|\cdot - x_t^{k-1}\|^2 \right) (x_t^k), \\ \|r_t^k\|^2 \leq \sigma^2 \|x_t^{k-1} - x_t^k\|^2, \end{cases}$$

671 for some $\sigma \in (0, 1)$.

672 Appendix A. Proof of Lemma 3.2 and Lemma 3.4(a)–(b).

673 Before giving the proofs, we present some auxiliary results. To avoid repetition,
674 we assume the reader is already familiar with (3.1)–(3.3).

675 The proof of the first result can be found in [19, Lemma B.2].

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

$$678 \quad (A.1) \quad \|a - (1 - \theta)b\|^2 - \zeta \|a\|^2 \geq \left[\frac{(1 - \zeta) - (1 - \theta)^2}{2} \right] (\|a\|^2 - \|b\|^2).$$

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

680 LEMMA A.2. *For every $i \geq 1$, index $t = 1, \dots, B$, and $u_t \in \mathcal{H}_t$, it holds that*

$$681 \quad \begin{aligned} & \lambda [\mathcal{L}_c^\theta(x_{<t}^i, u_t, x_{>t}^{i-1}; p^{i-1}) - \mathcal{L}_c^\theta(x_{<t}^i, x_t^i, x_{>t}^{i-1}; p^{i-1})] + \frac{1}{2} \|u_t - x_t^{i-1}\|^2 \\ 682 & \geq \frac{1}{2} \|\Delta x_t^i\|^2 + \left(\frac{1 - \lambda m_t}{2} \right) \|u_t - x_t^i\|^2 + \frac{\lambda c}{2} \|A_t(u_t - x_t^i)\|^2. \end{aligned}$$

684 *Proof.* Let $i \geq 1$, $t = 1, \dots, B$, and $u_t \in \mathcal{H}_t$ be fixed, and define $\mu := 1 - \lambda m_t$
685 and $\|\cdot\|_\alpha^2 := \langle \cdot, (\mu I + \lambda c A_t^* A_t)(\cdot) \rangle$. Since the prox stepsize λ is chosen in $(0, 1/(2m))$
686 and $m \geq m_t$ in view of (2.7), it follows that $\mu \geq 1/2$. Using the optimality of x_t^i ,
687 assumption (A4), and the fact that $\lambda \mathcal{L}_c^\theta(x_{<t}^i, \cdot, x_{>t}^{i-1}; p^{i-1}) + \|\cdot - x_t^{i-1}\|^2/2$ is 1-strongly
688 convex with respect to $\|\cdot\|_\alpha^2$, it follows that

$$689 \quad \begin{aligned} & \lambda \mathcal{L}_c^\theta(x_{<t}^i, x_t^i, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2} \|\Delta x_t^i\|^2 \\ 690 & \leq \lambda \mathcal{L}_c^\theta(x_{<t}^i, u_t, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2} \|u_t - x_t^{i-1}\|^2 - \frac{1}{2} \|u_t - x_t^i\|_\alpha^2 \\ 691 & = \lambda \mathcal{L}_c^\theta(x_{<t}^i, u_t, x_{>t}^{i-1}; p^{i-1}) + \frac{1}{2} \|u_t - x_t^{i-1}\|^2 - \frac{\mu}{2} \|u_t - x_t^i\|^2 - \frac{\lambda c}{2} \|A_t(u_t - x_t^i)\|^2. \quad \square \end{aligned}$$

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

694 *Proof of Lemma 3.2.* (a) Using the definition of $\mathcal{L}_c^\theta(\cdot; \cdot)$ in (1.4) and the relation
695 in Lemma 3.1(a), we conclude that

$$\begin{aligned}
696 \quad \mathcal{L}_c^\theta(x^i; p^i) - \mathcal{L}_c^\theta(x^i; p^{i-1}) &= (1 - \theta) \langle \Delta p^i, f^i \rangle = \left(\frac{1 - \theta}{\chi c} \right) \|\Delta p^i\|^2 + \frac{a_\theta}{\chi c} \langle \Delta p^i, p^{i-1} \rangle \\
697 &= \left(\frac{1 - \theta}{\chi c} \right) \|\Delta p^i\|^2 + \frac{a_\theta}{\chi c} (\langle p^i, p^{i-1} \rangle - \|p^{i-1}\|^2) \\
698 &= \left(\frac{1 - \theta}{\chi c} \right) \|\Delta p^i\|^2 + \frac{a_\theta}{\chi c} \left(\frac{1}{2} \|p^i\|^2 - \frac{1}{2} \|\Delta p^i\|^2 - \frac{1}{2} \|p^{i-1}\|^2 \right) \\
699 \quad (\text{A.2}) \quad &= \frac{b_\theta}{2\chi c} \|\Delta p^i\|^2 + \frac{a_\theta}{2\chi c} (\|p^i\|^2 - \|p^{i-1}\|^2). \\
700
\end{aligned}$$

701 (b) Using the definition of m in (2.7) and summing the inequality of Lemma A.2
702 with $u_t = x_t^{i-1}$ from $t = 1$ to B , we have that

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

706 The conclusion now follows from dividing the above inequality by λ and using the
707 fact that $\lambda \leq 1/m$.

708 (c) Note that the definition of b_θ in (3.1) and (2.6) imply

$$709 \quad \zeta := 2B\chi b_\theta \leq \theta^2.$$

710 Hence, using the definition of γ_θ in (3.1), and Lemma A.1 with $(a, b) = (\Delta p^i, \Delta p^{i-1})$
711 it follows that

$$712 \quad (\text{A.3}) \quad \|\Delta p^i - (1 - \theta)\Delta p^{i-1}\|^2 \geq 2B\chi b_\theta \|\Delta p^i\|^2 + \chi\gamma_\theta (\|\Delta p^i\|^2 - \|\Delta p^{i-1}\|^2).$$

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

$$\begin{aligned}
715 \quad \frac{c}{4} \sum_{t=1}^B \|A_t \Delta x_t^i\|^2 &\geq \frac{c}{4B} \|A \Delta x^i\|^2 = \frac{\|\Delta p^i - (1 - \theta)\Delta p^{i-1}\|^2}{4B\chi^2 c} \\
716 &\geq \frac{1}{4B\chi c} [2Bb_\theta \|\Delta p^i\|^2 + \gamma_\theta (\|\Delta p^i\|^2 - \|\Delta p^{i-1}\|^2)] \\
717 &= \frac{b_\theta}{2\chi c} \|\Delta p^i\|^2 + \frac{\gamma_\theta}{4B\chi c} (\|\Delta p^i\|^2 - \|\Delta p^{i-1}\|^2). \quad \square \\
718
\end{aligned}$$

719 Next, we give the proof of Lemma 3.4(a)–(b).

720 *Proof of Lemma 3.4(a)–(b).* (a) Using Lemma 3.2(a), the definition of $\mathcal{L}_c^\theta(\cdot; \cdot)$ in
721 (1.4), the fact that $\theta \in (0, 1)$, and the relations $2 \langle a, b \rangle \leq \|a\|^2 + \|b\|^2$ and $\|a + b\|^2 \leq$
722 $2\|a\|^2 + 2\|b\|^2$ for $a, b \in \mathbb{R}^n$, it follows that

$$\begin{aligned}
723 \quad \mathcal{L}_c^\theta(x^j; p^j) &= \phi(x^j) + (1 - \theta) \langle p^j, f^j \rangle + \frac{c}{2} \|f^j\|^2 \\
724 &\stackrel{\text{L.3.2(a)}}{=} \frac{(1 - \theta)}{\chi c} \langle p^j, p^j - (1 - \theta)p^{j-1} \rangle + \frac{1}{2c\chi^2} \|p^j - (1 - \theta)p^{j-1}\|^2
\end{aligned}$$

$$\begin{aligned}
725 \quad & \leq \frac{(1-\theta)}{2\chi c} \|p^i\|^2 + \frac{(1-\theta)}{2\chi c} \|p^i - (1-\theta)p^{i-1}\|^2 + \frac{1}{2\chi^2 c} \|p^i - (1-\theta)p^{i-1}\|^2 \\
726 \quad & \leq \frac{1}{2\chi c} \|p^i\|^2 + \frac{1}{\chi^2 c} \|p^i - (1-\theta)p^{i-1}\|^2 \\
727 \quad & \leq \frac{1}{2\chi c} \|p^i\|^2 + \frac{2}{\chi^2 c} \|p^i\|^2 + \frac{2}{\chi^2 c} \|p^{i-1}\|^2 \leq \frac{3(\|p^i\|^2 + \|p^{i-1}\|^2)}{\chi^2 c}. \\
728 \quad &
\end{aligned}$$

729 (b) It holds that

$$\begin{aligned}
730 \quad \mathcal{L}_c^\theta(x^k; p^k) &= \phi(x^k) + (1-\theta) \langle p^k, f^k \rangle + \frac{c}{2} \|f^k\|^2 \\
731 \quad &= \phi(x^k) + \frac{1}{2} \left\| \frac{(1-\theta)p^k}{\sqrt{c}} + \sqrt{c}f^k \right\|^2 - \frac{(1-\theta)^2 \|p^k\|^2}{2c} \\
732 \quad &\geq \phi(x^k) - \frac{(1-\theta)^2 \|p^k\|^2}{2c} \geq \phi(x^k) - \frac{\|p^k\|^2}{2c}. \quad \square \\
733 \quad &
\end{aligned}$$

734 REFERENCES

- 735 [1] D. P. BERTSEKAS, *Nonlinear programming*, Taylor & Francis, 3ed ed., 2016.
- 736 [2] S. BOYD, N. PARIKH, AND E. CHU, *Distributed optimization and statistical learning via the*
- 737 *alternating direction method of multipliers*, Now Publishers Inc, 2011.
- 738 [3] M. T. CHAO, Y. ZHANG, AND J. B. JIAN, *An inertial proximal alternating direction method of*
- 739 *multipliers for nonconvex optimization*, International Journal of Computer Mathematics,
- 740 (2020), pp. 1–19.
- 741 [4] C. CHEN, B. HE, Y. YE, AND X. YUAN, *The direct extension of admm for multi-block convex*
- 742 *minimization problems is not necessarily convergent*, Mathematical Programming, 155
- 743 (2016), pp. 57–79.
- 744 [5] J. ECKSTEIN AND D. P. BERTSEKAS, *On the douglas–rachford splitting method and the proximal*
- 745 *point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992),
- 746 pp. 293–318.
- 747 [6] J. ECKSTEIN AND M. C. FERRIS, *Operator-splitting methods for monotone affine variational in-*
- 748 *equalities, with a parallel application to optimal control*, INFORMS Journal on Computing,
- 749 10 (1998), pp. 218–235.
- 750 [7] J. ECKSTEIN AND M. FUKUSHIMA, *Some reformulations and applications of the alternating*
- 751 *direction method of multipliers*, in Large scale optimization, Springer, 1994, pp. 115–134.
- 752 [8] J. ECKSTEIN AND B. F. SVAITER, *A family of projective splitting methods for the sum of two*
- 753 *maximal monotone operators*, Mathematical Programming, 111 (2008), pp. 173–199.
- 754 [9] J. ECKSTEIN AND B. F. SVAITER, *General projective splitting methods for sums of maximal*
- 755 *monotone operators*, SIAM Journal on Control and Optimization, 48 (2009), pp. 787–811.
- 756 [10] D. GABAY, *Applications of the method of multipliers to variational inequalities*, in Studies in
- 757 *mathematics and its applications*, vol. 15, Elsevier, 1983, pp. 299–331.
- 758 [11] D. GABAY AND B. MERCIER, *A dual algorithm for the solution of nonlinear variational problems*
- 759 *via finite element approximation*, Computers & mathematics with applications, 2 (1976),
- 760 pp. 17–40.
- 761 [12] R. GLOWINSKI AND A. MARROCO, *Sur l’approximation, par éléments finis d’ordre un, et la*
- 762 *résolution, par pénalisation-dualité d’une classe de problèmes de dirichlet non linéaires*,
- 763 ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et
- 764 Analyse Numérique, 9 (1975), pp. 41–76.
- 765 [13] M. L. N. GONCALVES, J. G. MELO, AND R. D. C. MONTEIRO, *Convergence rate bounds for*
- 766 *a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly*
- 767 *constrained problems*, Pacific Journal of Optimization, 15 (2019), pp. 379–398.
- 768 [14] Z. JIA, J. HUANG, AND Z. WU, *An incremental aggregated proximal ADMM for linearly con-*
- 769 *strained nonconvex optimization with application to sparse logistic regression problems*,
- 770 Journal of Computational and Applied Mathematics, 390 (2021), p. 113384.
- 771 [15] B. JIANG, T. LIN, S. MA, AND S. ZHANG, *Structured nonconvex and nonsmooth optimization:*
- 772 *algorithms and iteration complexity analysis*, Computational Optimization and Applica-
- 773 *tions*, 72 (2019), pp. 115–157.

- 774 [16] W. KONG, *Accelerated inexact first-order methods for solving nonconvex composite optimization problems*, arXiv preprint arXiv:2104.09685, (2021).
775
776 [17] W. KONG, J. G. MELO, AND R. D. C. MONTEIRO, *Complexity of a quadratic penalty accelerated inexact proximal point method for solving linearly constrained nonconvex composite programs*, SIAM Journal on Optimization, 29 (2019), pp. 2566–2593.
777
778 [18] W. KONG, J. G. MELO, AND R. D. C. MONTEIRO, *An efficient adaptive accelerated inexact proximal point method for solving linearly constrained nonconvex composite problems*, Computational Optimization and Applications, 76 (2020), pp. 305–346.
779
780 [19] W. KONG AND R. D. C. MONTEIRO, *An accelerated inexact dampened augmented Lagrangian method for linearly-constrained nonconvex composite optimization problems*, arXiv preprint arXiv:2110.11151, (2021).
781
782 [20] J. G. MELO AND R. D. C. MONTEIRO, *Iteration-complexity of a Jacobi-type non-euclidean ADMM for multi-block linearly constrained nonconvex programs*, arXiv preprint arXiv:1705.07229, (2017).
783
784 [21] J. G. MELO AND R. D. C. MONTEIRO, *Iteration-complexity of a linearized proximal multiblock ADMM class for linearly constrained nonconvex optimization problems*, Optimization Online preprint, (2017).
785
786 [22] J. G. MELO, R. D. C. MONTEIRO, AND W. KONG, *Iteration-complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical Lagrangian function and a full lagrange multiplier update*, arXiv preprint arXiv:2008.00562, (2020).
787
788 [23] R. D. C. MONTEIRO AND B. F. SVAITER, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM Journal on Optimization, 23 (2013), pp. 475–507.
789
790 [24] R. T. ROCKAFELLAR, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of operations research, 1 (1976), pp. 97–116.
791
792 [25] A. RUSZCZYŃSKI, *An augmented Lagrangian decomposition method for block diagonal linear programming problems*, Operations Research Letters, 8 (1989), pp. 287–294.
793
794 [26] A. SUJANANI AND R. D. C. MONTEIRO, *An adaptive superfast inexact proximal augmented Lagrangian method for smooth nonconvex composite optimization problems*, arXiv preprint arXiv:2207.11905, (2022).
795
796 [27] A. X. SUN, D. T. PHAN, AND S. GHOSH, *Fully decentralized ac optimal power flow algorithms*, in 2013 IEEE Power & Energy Society General Meeting, IEEE, 2013, pp. 1–5.
797
798 [28] K. SUN AND A. SUN, *Dual descent ALM and ADMM*, arXiv preprint arXiv:2109.13214, (2021).
799
800 [29] K. SUN AND X. A. SUN, *A two-level distributed algorithm for general constrained non-convex optimization with global convergence*, arXiv preprint arXiv:1902.07654, (2019).
801
802 [30] A. THEMELIS AND P. PATRINOS, *Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results*, SIAM Journal on Optimization, 30 (2020), pp. 149–181.
803
804 [31] Y. WANG, W. YIN, AND J. ZENG, *Global convergence of ADMM in nonconvex nonsmooth optimization*, Journal of Scientific Computing, 78 (2019), pp. 29–63.
805
806 [32] J. ZHANG AND Z.-Q. LUO, *A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization*, SIAM Journal on Optimization, 30 (2020), pp. 2272–2302.
807
808
809
810
811
812
813
814
815
816