An Adaptive Proximal ADMM for Nonconvex Linearly-Constrained Composite Programs *

Leandro Farias Maia[†] David H. Gutman[‡] Renato D.C. Monteiro[§]
Gilson N. Silva[¶]

July 12, 2024

Abstract

This paper develops an adaptive Proximal Alternating Direction Method of Multipliers (P-ADMM) for solving linearly-constrained, weakly convex, composite optimization problems. This method is adaptive to all problem parameters, including smoothness and weak convexity constants. It is assumed that the smooth component of the objective is weakly convex and possibly nonseparable, while the non-smooth component is convex and block-separable. The proposed method is tolerant to the inexact solution of its block proximal subproblem so it does not require that the non-smooth component has easily computable block proximal maps. Each iteration of our adaptive P-ADMM consists of two steps: (1) the sequential solution of each block proximal subproblem, and (2) adaptive tests to decide whether to perform a full Lagrange multiplier and/or penalty parameter update(s). Without any rank assumptions on the constraint matrices, it is shown that the adaptive P-ADMM obtains an approximate first-order stationary point of the constrained problem in a number of iterations that matches the state-of-the-art complexity for the class of P-ADMMs. The two proof-of-concept numerical experiments that conclude the paper suggest our adaptive P-ADMM enjoys significant computational benefits.

Keywords: proximal ADMM, nonseparable, nonconvex composite optimization, iteration complexity, augmented Lagrangian function

1 Introduction

This paper develops an adaptive Proximal Alternating Direction Method of Multipliers, called A-ADMM, for solving the linearly-constrained, smooth, weakly convex, composite optimization problem

$$\phi^* = \min_{y\in\mathbb{R}^n} \left\{ \phi(y) := f(y) + \sum_{t=1}^{B} h_t(y_t) \quad : \quad \sum_{t=1}^{B} A_t y_t = b \right\}, \tag{1}$$

where $n = n_1 + \ldots + n_B$, $y = (y_1, \ldots, y_B) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_B}$, $b \in \mathbb{R}^l$, $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued differentiable function which is m-weakly convex, and $h_t : \mathbb{R}^{n_t} \to (-\infty, \infty]$ is a proper, closed, convex function which is M_h-Lipschitz continuous on its compact domain, for every $t \in \{1, \ldots, B\}$. To ease notation,

*$\mathbf{Funding}$: The first and second authors are supported by the NSF grant #2410328. The third author has been partially supported by AFOSR Grant FA9550-22-1-0088. The last author has been supported by CNPq Grants 401864/2022-7 and 306593/2022-0

[†]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, 77843. farias-maia@gmail.com

[‡]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, 77843. dhgutman@gmail.com

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

[¶]Department of Mathematics, Federal University of Piauí, Teresina, PI, 64049-550. gilson.silva@ufpi.edu.br

let $A(\cdot) := \sum_{t=1}^{B} A_t(\cdot)$ and $h(\cdot) := \sum_{t=1}^{B} h_t(\cdot)$. The goal in this paper is to find a (ρ,η)-stationary solution of (1), i.e., a quadruple $(\bar{x},\bar{p},\bar{v},\bar{\varepsilon}) \in (\operatorname{dom} h) \times \mathbb{R}^l \times \operatorname{Im}(A^*) \times \mathbb{R}_+$ satisfying

$$\bar{v} \in \nabla f(\bar{x}) + \partial_{\bar{\varepsilon}} h(\bar{x}) + A^*\bar{p}, \quad \sqrt{\|\bar{v}\|^2 + \bar{\varepsilon}} \le \rho, \quad \|A\bar{x} - b\| \le \eta, \tag{2}$$

where $(\rho,\eta) \in \mathbb{R}^2_{++}$ is a given tolerance pair.

A popular primal-dual algorithmic framework for solving problem (1) that takes advantage of its block structure is the Proximal Alternating Direction Method of Multipliers (P-ADMM), which is based on the dampened augmented Lagrangian function,

$$\mathcal{L}_c^\theta(y;p) := \phi(y) + (1-\theta)\langle p, Ay - b\rangle + \frac{c}{2}\|Ay-b\|^2, \tag{3}$$

where $\theta \in (0,1)$ is a dampening parameter and $c > 0$ is a penalty parameter. Given $(\tilde{y}^{k-1}, \tilde{q}^{k-1}, c_{k-1})$, P-ADMM finds the next triple $(\tilde{y}^k, \tilde{q}^k, c_k)$ as follows. Starting from $\tilde{y}^{k-1}$, it first performs a few cyclic block updates via an inexact proximal point method, with a fixed B-tuple $(\lambda_1, \cdots, \lambda_B)$ of block prox stepsizes, applied to the augmented Lagrangian function $\mathcal{L}^\theta_{c_{k-1}}(\cdot\,; \tilde{q}^{k-1})$ to obtain $\tilde{y}_k$. Next, it performs a standard dual update to obtain $\tilde{q}^k$. Finally, it possibly increases the penalty parameter. More formally, letting $z^0 = \tilde{y}^{k-1}$ and $|\mathcal{I}_k|$ be a positive integer, it recursively computes the sequence $\{z^i\}_{i=1}^{|\mathcal{I}_k|}$ as follows: given z^{i-1}, it computes z^i by sequentially solving the subproblems

$$z_t^i \approx \operatorname{argmin}_{u_t \in \mathbb{R}^{n_t}} \left\{ \lambda_t \mathcal{L}^\theta_{c_{k-1}}(z^i_{<t}, u_t, z^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|u_t - z_t^{i-1}\|^2 \right\}, \quad t = 1,\ldots,B. \tag{4}$$

It then sets $\tilde{y}^k = z^{|\mathcal{I}_k|}$, performs a Lagrange multiplier update according to

$$\tilde{q}^k = (1-\theta)\tilde{q}^{k-1} + \chi c_k \left(A\tilde{y}^k - b\right), \tag{5}$$

where χ is a positive under-relaxation parameter, and chooses a scalar $c_k \ge c_{k-1}$ as the next penalty parameter. The rationale for denoting the number of cyclic updates, prior to a dual update, as $|\mathcal{I}_k|$ will become clear in Section 3.2.

In the recent publication [25], the authors proposed a version of P-ADMM for solving (1), which assumes that $|\mathcal{I}_k| = 1$, $\lambda_1 = \cdots = \lambda_B$, and $(\chi,\theta) \in (0,1]^2$ satisfy

$$2\chi B(2-\theta)(1-\theta) \le \theta^2. \tag{6}$$

One of the main contributions of [25] is that its convergence guarantees do not require *the last block condition*, $\operatorname{Im}(A_B) \supseteq \{b\} \cup \operatorname{Im}(A_1) \cup \ldots \cup \operatorname{Im}(A_{B-1})$ and $h_B \equiv 0$, that hinders many instances of P-ADMM, see [7, 16, 44, 49]. However, the main drawbacks of the P-ADMM of [25] include: (i) the strong assumption (6) on (χ,θ); (ii) subproblem (4) must be solved exactly; (iii) the stepsize parameter λ is conservative and requires the knowledge of f's weak convexity parameter; (iv) it (conservatively) updates the Lagrange multiplier after each primal update cycle (i.e. $|\mathcal{I}_k| = 1$); (v) its iteration-complexity has a high dependence on the number of blocks B, namely, $\mathcal{O}(B^8)$. Paper [25] also presents computational results comparing its P-ADMM with a more practical variant where (θ,χ), instead of satisfying (6), it is set to $(0,1)$. Intriguingly, this $(\theta,\chi) = (0,1)$ regime substantially outperforms the theoretical regime of (6) in the provided computational experiments. No convergence analysis for the $(\theta,\chi) = (0,1)$ regime is forwarded in [25]. Thus, [25] leaves open the tantalizing question of whether the convergence of P-ADMM with $(\theta,\chi) = (0,1)$ can be theoretically secured.

Contributions: This work partially addresses the convergence analysis issue raised above by studying a *completely parameter-free* P-ADMM, with $(\theta,\chi) = (0,1)$ and $|\mathcal{I}_k|$ adaptively chosen, called A-ADMM. Rather than making the conservative determination that $|\mathcal{I}_k| = 1$, the studied adaptive method ensures the dual updates are committed as frequently as possible. It is shown that A-ADMM finds a (ρ,η)-stationary solution in $\mathcal{O}(B\max\{\rho^{-3},\eta^{-3}\})$ iterations. A-ADMM also exhibits the following additional features:

- Similar to the P-ADMM of [25], its complexity is established without assuming that the *last block condition* holds;

- Compared to the $\mathcal{O}(B^8 \max\{\rho^{-3}, \eta^{-3}\})$ iteration-complexity of the P-ADMM of [25], the one for A-ADMM vastly *improves the dependence on* B;
- A-ADMM uses an adaptive scheme that aggressively computes *variable block prox stepsizes*, instead of fixed ones that require knowledge of the weak convexity parameters $m_1, \ldots, m_B$ (e.g., the choice $\lambda_1 = \ldots = \lambda_B \in (0, 1/(2m))$ where $m := \max\{m_1, \ldots, m_B\}$ made by the P-ADMM of [25]). In contrast to the P-ADMM of [25], A-ADMM may generate various λ_t's which are larger than $1/m_t$ (as observed in our computational results), and hence which do not guarantee convexity of (4).
- A-ADMM is also adaptive to Lipschitz parameters;
- In contrast to the P-ADMM in [25], A-ADMM allows the block proximal subproblems (4) to be either exactly or *inexactly* solved.

Related Works: ADMM methods with $B = 1$ are well-known to be equivalent to augmented Lagrangian methods. Several references have studied augmented Lagrangian and proximal augmented Lagrangian methods in the convex (see e.g., [1, 2, 29, 30, 31, 32, 38, 39, 46]) and nonconvex (see e.g. [4, 5, 17, 21, 24, 26, 27, 33, 43, 47, 48, 50]) settings. Moreover, ADMMs and proximal ADMMs in the convex setting have also been broadly studied in the literature (see e.g. [4, 6, 8, 9, 10, 11, 12, 13, 14, 15, 37, 40, 41]). So from now on, we just discuss P-ADMM variants where f is nonconvex and $B > 1$.

A discussion of the existent literature on non-convex P-ADMM is best framed by dividing it into two different corpora: those papers that assume the last block condition and those that do not. Under the *last block condition*, the iteration-complexity established is $\mathcal{O}(\varepsilon^{-2})$, where $\varepsilon := \min\{\rho, \eta\}$. Specifically, [7, 16, 44, 45] introduce P-ADMM approaches assuming $B = 2$, while [22, 23, 34, 35] present (possibly linearized) P-ADMMs assuming $B \geq 2$. A first step towards removing the last block condition was made by [23] which proposes an ADMM-type method applied to a penalty reformulation of (1) that artificially satisfies the last block condition. This method possesses an $\mathcal{O}(\varepsilon^{-6})$ iteration-complexity bound.

On the other hand, development of ADMM-type methods directly applicable to (1) is considerably more challenging and only a few works addressing this topic have surfaced. In addition to [23], earlier contributions to this topic were obtained in [20, 43, 49]. More specifically, [20, 49] develop a novel small stepsize ADMM-type method without establishing its complexity. Finally, [43] considers an interesting but unorthodox negative stepsize for its Lagrange multiplier update, that sets it outside the ADMM paradigm, and thus justifies its qualified moniker, "scaled dual descent ADMM".

1.1 Organization

In this subsection, we outline this article's structure. This section's lone remaining subsection, Subsection 1.2, briefly lays out the basic definitions and notation used throughout. Section 2 introduces a notion of an inexact solution of A-ADMM's foundational block proximal subproblem (4) along with efficient subroutines designed to find said solutions. Section 3 presents this article's centerpiece algorithm, A-ADMM, and its main subprocedure and "static" version, S-ADMM, along with the main theorems governing their iteration-complexity (Theorem 3.3 and 3.1). Section 4 undertakes the long proof of the S-ADMM complexity theorem and presents all of the supporting technical lemmas. Section 5 presents proof-of-concept numerical experiments that display the superb efficiency of A-ADMM for two different problem classes. Section 6 gives some concluding remarks that suggest further research directions.

1.2 Notation and Basic Definitions

This subsection lists the elementary notation deployed throughout the paper. We let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote the set of non-negative and positive real numbers, respectively. We shall assume that the n-dimensional Euclidean space, $\mathbb{R}^n$, is equipped with an inner product, $\langle \cdot, \cdot \rangle$. The associated induced norm will be written as $\| \cdot \|$. Given $\nu \in \mathbb{R}^n$, we let

$$\max(\nu) := \max_{1 \leq t \leq n} \{\nu_t\} \text{ and } \min(\nu) = \min_{1 \leq t \leq n} \{\nu_t\}. \tag{7}$$

When the Euclidean space of interest has the block structure $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_B}$, we will often consider, for $x = (x_1, \ldots, x_B) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_B}$, the aggregated quantities

$$x_{<t} := (x_1, \ldots, x_{t-1}), \quad x_{>t} := (x_{t+1}, \ldots, x_B), \quad x_{\le t} := (x_{<t}, x_t), \quad x_{\ge t} := (x_t, x_{>t}). \tag{8}$$

For a given closed, convex set $Z \subset \mathbb{R}^n$, we let ∂Z designate its boundary. The distance of a point $z \in \mathbb{R}^n$ to Z, measured in terms of $\|\cdot\|$, is denoted $\operatorname{dist}(z, Z)$. The indicator function of Z, δ_Z, is defined by $\delta_Z(z) = 0$ if $z \in Z$, and $\delta_Z(z) = +\infty$ otherwise.

The set of points where a function $h : \mathbb{R}^n \to (-\infty, \infty]$ is finite-valued, $\operatorname{dom} h := \{x \in \mathbb{R}^n : h(x) < +\infty\}$, is called its domain. We say that h is called proper if it is finite-valued at some point $x \in \mathbb{R}^n$, i.e. if $\operatorname{dom} h \neq \emptyset$. We call $g \in \mathbb{R}^n$ an ϵ-subgradient of a proper function $h : \mathbb{R}^n \to (-\infty, \infty]$ at $x \in \operatorname{dom} h$, with $\epsilon \ge 0$, if

$$h(z) \ge h(x) + \langle g, z - x \rangle - \varepsilon$$

holds for all $z \in \mathbb{R}^n$. The set of all ϵ-subgradients of h at x, $\partial_\epsilon h(\cdot)$, is called the ϵ-subdifferential of h. When $\epsilon = 0$, the ϵ-subdifferential recovers the classical subdifferential, $\partial h(\cdot) := \partial_0 h(\cdot)$. If ψ is a real-valued function that is differentiable at $\bar{z} \in \mathbb{R}^n$, then its affine approximation $\ell_\psi(\cdot, \bar{z})$ at $\bar{z}$ is the function defined, for arbitrary $z \in \mathbb{R}^n$, by the rule $z \mapsto \psi(\bar{z}) + \langle \nabla \psi(\bar{z}), z - \bar{z} \rangle$. The smallest positive singular value of a nonzero linear operator $Q : \mathbb{R}^n \to \mathbb{R}^l$ is denoted ν_Q^+ and its operator norm is $\|Q\| := \sup\{\|Q(w)\| : \|w\| = 1\}$.

2 Methods and Concepts for the Inexact Solution of (4)

This section introduces a notion of an inexact stationary point for the block proximal subproblem (4) along with two different efficient methods for discovering such points. These methods permit the application of our main algorithm, A-ADMM, even when (4) is not exactly solvable. In Subsection 2.1, we introduce our inexact solution concept, Definition 2.1, for (4). Subsection 2.2 introduces a general method for finding said inexact solutions to (4), while Subsection 2.3 presents accelerated schemes for solving strongly convex or convex versions of (4). It also discusses how these schemes are used to tentatively solve weakly convex versions of (4).

2.1 An Inexact Solution Concept for (4)

This subsection introduces our notion (Definition 2.1) of an inexact solution of the block proximal subproblem (4). To cleanly frame this solution concept, observe that (4) can be cast in the form

$$\psi^* = \min\{\psi(z) := \psi_s(z) + \psi_n(z) : z \in \mathbb{R}^n\}, \tag{9}$$

where

$$\psi_s(\cdot) = \lambda_t \hat{\mathcal{L}}_c(y^i_{<t}, \cdot, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|\cdot - \tilde{y}_t^{k-1}\|^2, \quad \psi_n(\cdot) = \lambda_t h_t(\cdot), \tag{10}$$

and $\hat{\mathcal{L}}_c(\cdot\,; \tilde{q}^{k-1})$ is the smooth part of (3) with $(\theta, \chi) = (0, 1)$, defined as

$$\hat{\mathcal{L}}_c(y; \tilde{q}^{k-1}) := f(y) + \left\langle \tilde{q}^{k-1}, Ay - b \right\rangle + \frac{c}{2} \left\| Ay - b \right\|^2. \tag{11}$$

The assumptions introduced in Section 3 regarding problem (1) will ensure that the above functional pair (ψ_s, ψ_n) satisfies the following conditions:

(B1) $\psi_n : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper, closed, convex function with compact domain;

(B2) function $\psi_s : \mathbb{R}^n \to \mathbb{R}$ is differentiable and there exists $M > 0$ such that

$$\|\nabla \psi_s(z) - \nabla \psi_s(\tilde{z})\| \le M \|z - \tilde{z}\| \quad \forall\, z, \tilde{z} \in \mathbb{R}^n.$$

Hence, we assume that conditions (B1) and (B2) are valid throughout our discussion in this preliminary section.

With the technical setup now established, we are able to forward our inexact stationary point concept for (9), and hence for (4).

Definition 2.1 *For a given $z^0 \in \operatorname{dom}\psi_n$ and parameter triple $(\tau_1, \tau_2, \vartheta) \in \mathbb{R}^3_+$, a triple $(\bar z, \bar r, \bar\epsilon) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ satisfying*

$$\bar r \in \nabla\psi_s(\bar z) + \partial_{\bar\varepsilon}\psi_n(\bar z) \tag{12}$$

$$\|\bar r\|^2 + 2\bar\varepsilon \le \tau_1\|z^0 - \bar z\|^2 + \tau_2[\psi(z^0) - \psi(\bar z)] + \vartheta^2 \tag{13}$$

is called a $(\tau_1, \tau_2, \vartheta; z^0)$-stationary solution of (9). If $\vartheta = 0$, then $(\bar z, \bar r, \bar\epsilon)$ is simply referred to as a $(\tau_1, \tau_2; z^0)$-stationary solution of (9).

We now make some remarks about Definition 2.1. First, A-ADMM uses only $(\tau_1, \tau_2; z^0)$-stationary points, i.e., it always assumes that $\vartheta = 0$. Second, if $(\tau_1, \tau_2, \vartheta) = (0, 0, 0)$, then (13) implies that $(\bar r, \bar\varepsilon) = (0, 0)$ and (12) then implies that $\bar z$ is an exact stationary point of (9). Thus, if the triple $(\bar z, \bar r, \bar\varepsilon)$ is a $(\tau_1, \tau_2, \vartheta; z^0)$-stationary solution of (9), then $\bar z$ can be viewed as an approximate stationary solution of (9) where the residual pair $(\bar r, \bar\varepsilon)$ is bounded according to (13) (instead of being zero as in the exact case). Third, if $\bar z$ is an exact stationary point of (9), then the triple $(\bar z, 0, 0)$ is a $(\tau_1, 0, \vartheta; z^0)$-stationary point of (9) for any $(\tau_1, \vartheta) \in \mathbb{R}^2_+$. Fourth, if $\bar z$ is a stationary point of (9) such that $\psi(z^0) - \psi(\bar z) \ge 0$, then the triple $(\bar z, 0, 0)$ is a $(\tau_1, \tau_2, \vartheta; z^0)$-stationary point of (9) for any $(\tau_1, \tau_2, \vartheta) \in \mathbb{R}^3_+$. Hence, if z^0 is a stationary point of (9), then the triple $(z^0, 0, 0)$ is a $(\tau_1, \tau_2, \vartheta; z^0)$-stationary point of (9) for any $(\tau_1, \tau_2, \vartheta) \in \mathbb{R}^3_+$.

2.2 Composite Gradient Method

This subsection describes a variant of the composite gradient method, referred to as S-CGM, and its iteration-complexity, for finding a $(\tau_1, \tau_2, \vartheta; z^0)$-stationary point of (9) regardless of whether or not the objective function ψ is convex. In fact, this method finds somewhat higher quality inexact stationary points where $\tau_1 = 0$. Specifically, it obtains a pair $(\bar z, \bar v)$ satisfying

$$\bar v \in \nabla\psi_s(\bar z) + \partial\psi_n(\bar z) \quad \text{and} \quad \|\bar v\|^2 \le \sigma[\psi(\tilde z^0) - \psi(\bar z)] + \vartheta^2, \tag{14}$$

which implies that it is $(0, \sigma, \vartheta; z^0)$-stationary solution of (9). The S-CGM method is detailed below.

Algorithm 1 S-CGM

Input: $\tilde z^0 \in \operatorname{dom}\psi_n$, $\sigma > 0$, $\vartheta \ge 0$, $M > 0$.
Output: $(\bar z, \bar v)$ satisfying (14)
1: **for** $j = 1, 2, \ldots$ **do**
2: Compute $(\tilde z^j, \tilde v^j)$ such that:

$$\tilde z^j \in \operatorname{argmin}_{z\in\mathbb{R}^n}\left\{\ell_{\psi_s}(z, \tilde z^{j-1}) + \psi_n(z) + \frac{M}{2}\|z - \tilde z^{j-1}\|^2\right\} \tag{15}$$

$$\tilde v^j := M(\tilde z^{j-1} - \tilde z^j) + \nabla\psi_s(\tilde z^j) - \nabla\psi_s(\tilde z^{j-1}). \tag{16}$$

3: **if** $\|\tilde v^j\|^2 \le \sigma[\psi(\tilde z^0) - \psi(\tilde z^j)] + \vartheta^2$ **then** $(\bar z, \bar v) = (\tilde z^j, \tilde v^j)$

The following result, whose proof is given in Appendix B, describes the iteration-complexity of S-CGM.

Proposition 2.2 *S-CGM stops in at most*

$$\bar K := \begin{cases} 1 & \text{if } \vartheta^2 + \|\tilde v^1\|^2 = 0 \\ 1 + \left\lceil \dfrac{8M + \sigma}{\sigma} \ln\left(\dfrac{16M[\psi(\tilde z^0) - \psi^*]}{\vartheta^2 + \|\tilde v^1\|^2}\right)\right\rceil & \text{if } \vartheta^2 + \|\tilde v^1\|^2 > 0 \end{cases} \tag{17}$$

iterations and its output $(\bar z, \bar v)$ satisfies (14).

We now make some comments about S-CGM. First, if $\sigma = 0$, then it is well-known that the iteration-complexity of S-CGM is $\mathcal{O}(M\vartheta^{-2})$ (see for example [3, Theorem 10.15]). On the other hand, if σ is positive, then the above result shows that the iteration-complexity of S-CGM is $\mathcal{O}(M\ln(\vartheta^{-2}))$.

2.3 Accelerated Composite Gradient Methods in the Convex Setting

This subsection briefly surveys accelerated composite gradient methods from [18] and [42] for finding $(\tau_1, 0; z^0)$-stationary solutions of (9) when ψ is μ-strongly convex for some $\mu \geq 0$ (with the convention that ψ is convex for $\mu = 0$). Assuming the conditions (B1) and (B2) hold, then [18] shows that an accelerated composite gradient (ACG) method variant finds a $(\tau_1, 0; z^0)$-stationary solution of (9), namely a triple $(\bar{z}, \bar{v}, \bar{\varepsilon})$ satisfying

$$\bar{v} \in \nabla\psi_s(\bar{z}) + \partial_{\bar{\varepsilon}}\psi_n(\bar{z}) \quad \text{and} \quad \|\bar{v}\|^2 + 2\bar{\varepsilon} \leq \tau_1\|z^0 - \bar{z}\|^2,$$

in at most

$$\mathcal{O}\left(\left\lceil \min\left\{\sqrt{M(M+1)\lceil\tau_1^{-1}\rceil},\ 1 + \left(1 + \sqrt{M/\mu}\right)\max\{\log\left[M(M+1)\lceil\tau_1^{-1}\rceil\right], 1\}\right\}\right\rceil\right)$$

iterations. Moreover, [42] describes an ACG variant that, in at most

$$\mathcal{O}\left(\sqrt{M/\mu}\max\{\log(M), 1\}\right)$$

iterations, either finds a $(\tau_1, 0\,; z^0)$-stationary solution of (9) with $\bar{\varepsilon} = 0$, i.e., of the form $(\bar{z}, \bar{v}, 0)$, or concludes that ψ_s is not μ-strongly convex. This method is an adaptive ACG scheme (referred to as Adap-ACG) that searches for a stepsize λ_t until a $(\tau_1, 0\,; z^0)$-stationary solution of (9), with (ψ_s, ψ_n) given by (10), is found. This scheme repeatedly calls Adapt-ACG with λ_t halved each time. Under the assumption that ψ_s is m_t-weakly convex for every $t \in \{1, \ldots, B\}$, this scheme stops after a finite number of calls to Adap-ACG because ψ_s given by (10) is strongly convex for any $\lambda_t < 1/m_t$.

3 An Adaptive P-ADMM

This section presents this paper's main algorithm, A-ADMM, for solving (1). Subsection 3.1 details a few mild technical assumptions imposed on (1). Subsection 3.2 develops a "static" version of A-ADMM, called S-ADMM, that maintains a constant penalty parameter throughout its execution. This "static" version is the main subprocedure for A-ADMM. That subsection also presents the main complexity theorem (Theorem 3.1) for S-ADMM, but the proof is deferred to Section 4. Finally, Subsection 3.3 describes A-ADMM and establishes its iteration-complexity (Theorem 3.3)

3.1 Problem Assumptions

This subsection describes a series of mild assumptions on this paper's main problem of interest (1). We shall assume that $f, h : \mathbb{R}^n \to (-\infty, \infty]$, $A : \mathbb{R}^n \to \mathbb{R}^l$, and $b \in \mathbb{R}^l$, satisfy the following conditions:

(A1) for every $t \in \{1, \ldots, B\}$, $h_t : \mathbb{R}^{n_t} \to (-\infty, \infty]$ is proper, closed, and convex with $\mathcal{H}_t := \operatorname{dom} h_t$ compact;

(A2) A is a nonzero linear operator and $\mathcal{F} := \{x \in \mathcal{H} : Ax = b\} \neq \emptyset$, where $\mathcal{H} := \mathcal{H}_1 \times \cdots \times \mathcal{H}_B$;

(A3) f is block m-weakly convex for $m = (m_1, \ldots, m_B) \in \mathbb{R}^B_{++}$, that is, for every $t \in \{1, \ldots, B\}$,

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

(A4) f is differentiable on $\mathcal{H}$ and, for every $t \in \{1, \ldots, B-1\}$, there exists $L_t \geq 0$ such that

$$\|\nabla_{x_t} f(x_{<t}, x_t, \tilde{x}_{>t}) - \nabla_{x_t} f(x_{<t}, x_t, x_{>t})\| \leq L_t\|\tilde{x}_{>t} - x_{>t}\| \quad \forall\, x, \tilde{x} \in \mathcal{H}; \tag{18}$$

(A5) for some $M_h \geq 0$, $h(\cdot)$ is M_h-Lipschitz continuous on $\mathcal{H}$;

(A6) there exists $\bar{z} \in \mathcal{F}$ such that $\bar{d} := \operatorname{dist}(\bar{z}, \partial\mathcal{H}) > 0$.

Note that since $\mathcal{H}$ is compact, it follows from (A1) and (A2) that the scalars

$$D_h := \sup_{z \in \mathcal{H}} \|z - \bar{z}\|, \quad \nabla_f := \sup_{u \in \mathcal{H}} \|\nabla f(u)\|, \quad \underline{\phi} := \inf_{u \in \mathcal{H}} \phi(u), \quad \overline{\phi} := \sup_{u \in \mathcal{H}} \phi(u) \tag{19}$$

are bounded. Furthermore, throughout this paper, we let

$$\|A\|_{\dagger}^2 := \sum_{t=1}^{B} \|A_t\|^2, \tag{20}$$

3.2 S-ADMM: A "Static" Version of A-ADMM

This subsection presents S-ADMM, a static version of our A-ADMM, and its main complexity result (Theorem 3.1). The qualifier "static" is attached because this variant keeps the penalty parameter constant throughout its course. The proof of Theorem 3.1 is the focus of Section 4. We start by elaborating S-ADMM.

Algorithm 2 S-ADMM

Universal Input: $C > 0$, $\rho > 0$, $\alpha > 0$, $\sigma_1 \le 1/8$, $\sigma_2 > 0$ and $\sigma_1 + \sigma_2/2 \ge 0$
Input: $(x, p, \gamma, c) \in \mathcal{H} \times \mathbb{R}^l \times \mathbb{R}^B_{++} \times \mathbb{R}_{++}$
Output: $(x^+, p^+, \gamma^+, v^+, \varepsilon^+)$

1: $y^0 = \tilde{y}^0 = x$, $\tilde{q}^0 = p$, $\tilde{T}_0 = 0$, $\lambda^0 = \gamma$, $k = 1$
2: **for** $i \leftarrow 1, 2, \ldots$ **do**
3:     **for** $t \leftarrow 1, 2, \ldots, B$ **do**
4:         $\lambda_t = \lambda_t^{i-1}$
5:         set (ψ_s, ψ_n) as in (10) and

$$\tau_1 = \left(\sigma_1 + \frac{\sigma_2}{2}\right) \frac{\lambda_t}{1 + 2\lambda_t}, \quad \tau_2 = \frac{\sigma_2 \lambda_t}{1 + 2\lambda_t}, \quad z^0 = y_t^{i-1},$$

        and find a $(\tau_1, \tau_2; z^0)$-stationary solution $(y_t, r_t, \varepsilon_t)$ of (9) (see Definition 2.1)
6:         **if** (λ_t, y_t, r_t) does not satisfy

$$(1 + \sigma_2)(\Delta \mathcal{L}_c)_t^i(y_t) \ge \frac{1}{4\lambda_t} \|y_t - y_t^{i-1}\|^2 + \frac{c}{4} \|A_t(y_t - y_t^{i-1})\|^2$$

        then
7:             $\lambda_t = \lambda_t/2$ and go to line 5.
8:         **else**
9:             $(\lambda_t^i, y_t^i, r_t^i, \varepsilon_t^i) = (\lambda_t, y_t, r_t, \varepsilon_t)$

10:     **for** $t \leftarrow 1, 2, \ldots, B$ **do**
11:         $(\Delta f)_t^i \leftarrow \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^i) - \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^{i-1})$
12:         $v_t^i \leftarrow (\Delta f)_t^i + \frac{r_t^i}{\lambda_t^i} + cA_t^* \sum_{s=t+1}^{B} A_s(y_s^i - y_s^{i-1}) - \frac{1}{\lambda_t^i}(y_t^i - y_t^{i-1})$
13:     $v^i = (v_1^i, \ldots, v_B^i)$, $y^i = (y_1^i, \ldots, y_B^i)$, $\lambda^i = (\lambda_1^i, \ldots, \lambda_B^i)$, $\delta_i := (\varepsilon_1^i/\lambda_1^i) + \ldots + (\varepsilon_B^i/\lambda_B^i)$
14:     $T_i = \mathcal{L}_c(\tilde{y}^{k-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(y^i; \tilde{q}^{k-1}) + \tilde{T}_{k-1}$

15:     **if** $\|v^i\|^2 + \delta_i \le C^2$ and $i \ge (k\alpha T_i)/\rho^2$ **then**
16:         $(\tilde{y}^k, \tilde{v}^k, \tilde{\varepsilon}_k) = (y^i, v^i, \delta_i)$, and $\tilde{q}^k = \tilde{q}^{k-1} + c(A\tilde{y}^k - b)$
17:         $\tilde{T}_k = T_i$, $\tilde{\lambda}^k = \lambda^i$, $i_k^+ = i$
18:         **if** $\|\tilde{v}^k\|^2 + \tilde{\varepsilon}_k \le \rho^2$ **then**
19:             $(x^+, p^+, \gamma^+, v^+, \varepsilon^+) = (\tilde{y}^k, \tilde{q}^k, \tilde{\lambda}^k, \tilde{v}^k, \tilde{\varepsilon}_k)$
20:             **return** $(x^+, p^+, \gamma^+, v^+, \varepsilon^+)$
21:         **else**
22:             $k \leftarrow k + 1$

Let us now clarify several of the steps of S-ADMM. The index i is an iteration count for S-ADMM, whose iterations are referred to as S-ADMM iterations throughout the paper. The t-th pass of the loop consisting of lines 3 to 9 approximately solves the t-th augmented Lagrangian subproblem (4) with a tentative prox stepsize λ_t and then tests whether the descent condition of line 6 holds. If it holds, then λ_t is accepted as the prox stepsize for the t-th block in iteration i (line 9); otherwise, λ_t is halved and the above steps are repeated with the updated λ_t (line 5). The index k counts the number of times that S-ADMM updates the Lagrange multiplier. The set of S-ADMM iterations after the $(k-1)$-th update and up to and including the k-th update, denoted $\mathcal{I}_k$, is called the k-th epoch. We enumerate each $\mathcal{I}_k$ as

$$\mathcal{I}_k := \{i_k^-, \ldots, i_k^+\}, \text{ where } i_k^- := i_{k-1}^+ + 1. \tag{21}$$

The beginning and the end of an epoch are determined by lines 15 to 22. If the test in line 15 is passed at iteration i, then the Lagrange multiplier is updated. Provided S-ADMM does not terminate due to line 18, a new epoch begins. Line 18 tests whether the residual vector pair $(\tilde{v}^k, \tilde{\varepsilon}_k)$, computed in line 16, satisfies the first inequality in (2) that defines a (ρ, η)-stationary solution. If so, then S-ADMM terminates (see line 19); otherwise, k is updated to $k+1$ and a new epoch commences. It is shown in Lemma 4.10 below that every epoch $\mathcal{I}_k$ terminates, and hence that the last index i_k^+ in (21) is well-defined.

We now make several remarks about S-ADMM. First, it is shown in Lemma 4.4 below that if the stepsize λ_t in (4) satisfies $\lambda_t < 1/m_t$, then the test in line 6 is satisfied. Since λ_t is halved every time the test fails, this remark ensures that the loop in lines 5 to 9 ends. Second, it is shown in Lemma 4.5 that the iterate $y^i = (y_1^i, \ldots, y_B^i)$, the residual vector $v^i = (v_1^i, \ldots, v_B^i)$ and $\delta_i := (\varepsilon_1^i/\lambda_1^i) + \ldots + (\varepsilon_B^i/\lambda_B^i)$ computed in lines 10 to 13 satisfy the stationary inclusion $v^i \in \nabla f(y^i) + \partial_{\delta_i} h(y^i) + \mathrm{Im}(A^*)$. Hence, upon termination of S-ADMM, its output satisfies the first two conditions in (2) but not necessarily the last one. Third, if the (fixed) penalty parameter c is sufficiently large, then Theorem 3.1(c) below shows that the final output of S-ADMM also satisfies the last condition in (2), and hence is a (ρ, η)-stationary solution of (1).

It is not difficult to check that S-ADMM fits within the P-ADMM framework described in the introduction. First, fixing $k \geq 0$, the cyclic, inexact solution of the block proximal subproblem (4) is manifested in S-ADMM in lines 3 to 9. The damped augmented Lagrangian in this setting takes $(\theta, \chi) = (0, 1)$. The sequence $\{z^j\}_{j=1}^{|\mathcal{I}_k|}$ is related to the sequence $\{y^j\}_{j=i_k^-}^{i_k^+}$ by the rule $z^j = y^{i_k^- + j - 1}$ for $1 \leq j \leq i_k^+ - i_k^- + 1$. By extension, $|\mathcal{I}_k| = i_k^+ - i_k^- + 1$ in accordance with (21). Second, the dual update of P-ADMM, (5), manifests in S-ADMM in line 16.

Before stating the main result of this subsection, we define some constants that are used to express its complexity bounds. Given a triple $(\bar{Q}, \bar{\gamma}, \underline{c}) \in \mathbb{R}_{++} \times \mathbb{R}_{++}^B \times \mathbb{R}_{++}$, $\sigma_1 \in (0, 1/8]$, $\sigma_2 > 0$, let

$$\begin{gathered} L := (L_1, \ldots, L_t), \quad \bar{\Gamma} := \overline{\phi} - \underline{\phi} + \frac{(\bar{Q} + 3\kappa_p)^2}{2\underline{c}}, \quad \kappa_p := \frac{2D_h(M_h + C + \nabla_f) + C^2}{\bar{d}\nu_A^+} \\ \bar{\chi} := 2(1 + \sigma_2)\left[1 + 12\|L\|^2 \max(\bar{\gamma}) + 12(4\max(m) + \max(\bar{\gamma}^{-1})) + 2\sigma_1\right] + \sigma_2, \end{gathered} \tag{22}$$

where L_t, M_h and $\bar{d}$ are as in (A4), (A5) and (A6), respectively, C is part of the universal input for S-ADMM, $\max(m)$, $\max(\bar{\gamma}^{-1})$ are as in (7), $(D_h, \nabla_f, \underline{\phi}, \overline{\phi})$ is as in (19), and ν_A^+ is the smallest positive singular value of the nonzero linear operator A.

Furthermore, define

$$\begin{aligned} &\Lambda_0 := 8\kappa_p^2(1 + \sigma_2)B\|A\|_\dagger^2 \left[\frac{(1 + \sigma_2)B\|A\|_\dagger^2}{2\alpha} + 2\bar{\Gamma}\right], \quad \Lambda_1 := \left(\frac{\bar{\chi}}{\alpha} + 1\right)\left(1 + \frac{\bar{\chi}\bar{\Gamma}}{C^2}\right) \\ &\Lambda_2 := (1 + \sigma_2)B\|A\|_\dagger^2 \left[\left(\frac{\bar{\chi}}{\alpha} + 1\right)\frac{1}{C^2} + \left(1 + \frac{\bar{\chi}\bar{\Gamma}}{C^2}\right)\frac{1}{\alpha}\right], \quad \Lambda_3 := \frac{(1 + \sigma_2)^2 B^2 \|A\|_\dagger^4 \bar{\Gamma}}{\alpha C^2} \\ &\Lambda_4 := 8\kappa_p^2(1 + \sigma_2)B\|A\|_\dagger^2 \left(\frac{\bar{\chi}}{\alpha} + \frac{3}{2}\right) + \frac{8\alpha\kappa_p^2}{\underline{c}}\left(\frac{\bar{\chi}^2}{2\alpha^2} + \frac{3\bar{\chi}}{2\alpha} + 1\right) + 2\alpha\bar{\Gamma}\left(\frac{\bar{\chi}}{\alpha} + 1\right), \end{aligned} \tag{23}$$

where $\|A\|_\dagger^2$ is as in (20), $(\bar{\chi}, \bar{\Gamma}, \kappa_p)$ is as in (22) and $\alpha \in \mathbb{R}_{++}$.

The remainder of this subsection presents and discusses one of the main results of this paper, the iteration-complexity of S-ADMM. Its proof is the sole focus of Section 4.

Theorem 3.1 (S-ADMM Complexity) *Let the tolerance pair $(\rho,\eta)\in\mathbb{R}^2_{++}$, and the triple $(\bar{Q},\bar{\gamma},\underline{c})\in\mathbb{R}_{++}\times\mathbb{R}^B_{++}\times\mathbb{R}_{++}$ be given. Assume that the input $(x,p,\gamma,c)\in\mathcal{H}\times A(\mathbb{R}^n)\times\mathbb{R}^B_{++}\times\mathbb{R}_{++}$ of S-ADMM satisfies*

$$c\|Ax-b\|\le\bar{Q},\quad \|p\|\le\kappa_p,\quad c\ge\underline{c},\tag{24}$$

$$\min\left\{\frac{1}{2m},\bar{\gamma}\right\}\le\gamma\le\bar{\gamma},\tag{25}$$

where κ_p and $\max(m)$ are as in (22) and (7), respectively. Then, the S-ADMM method with input (x,p,γ,c) satisfies the following statements:

(a) the total number of iterations performed by S-ADMM is bounded by

$$\Lambda(c,\rho):=\Lambda_0\frac{c}{\rho^2}+\Lambda_1+\Lambda_2c+\Lambda_3c^2+\Lambda_4\frac{1}{\rho^2},\tag{26}$$

where Λ_0, Λ_1, Λ_2, Λ_3, and Λ_4 are as in (23);

(b) it outputs $(x^+,p^+,\gamma^+,v^+,\varepsilon^+)$ satisfying the inclusion

$$v^+\in\nabla f(x^+)+\partial_{\varepsilon^+}h(x^+)+A^*p^+\tag{27}$$

and the following bounds

$$c\|Ax^+-b\|\le 2\kappa_p,\quad \|p^+\|\le\kappa_p,\quad \|v^+\|^2+\varepsilon^+\le\rho^2,\tag{28}$$

$$\min\left\{\frac{1}{2m_t},\bar{\gamma}_t\right\}\le\gamma_t^+\le\bar{\gamma}_t\quad\forall t\in\{1,\ldots,B\};\tag{29}$$

(c) if $c\ge 2\kappa_p/\eta$, then the output triple (x^+,p^+,v^+) of S-ADMM is a (ρ,η)-stationary solution of (1) as in (2).

We now make several comments about Theorem 3.1. First, this result will be used in the next subsection to analyze the iteration-complexity of this paper's main algorithm, A-ADMM, which repeatedly invokes S-ADMM using a warm-start scheme, i.e. if c is the penalty parameter and (x^+,p^+,γ^+) is the output of the current S-ADMM call, then the input of the next S-ADMM call is $(x^+,p^+,\gamma^+,2c)$.

Second, in the above setting, the quantities $(\bar{Q},\bar{\gamma},\underline{c})$ should be viewed as uniform (upper or lower) bounds on the quantities $(c\|Ax-b\|,\gamma,c)$ associated with the input (x,p,γ,c) of all S-ADMM calls. Moreover, the iteration-complexity bounds of Theorem 3.1 are expressed in terms of $(\bar{Q},\bar{\gamma},\underline{c})$ instead of its counterpart $(c\|Ax-b\|,\gamma,c)$.

Third, it is shown in Lemma 4.5 and Lemma 4.11 below that every quadruple $(\tilde{y}^k,\tilde{q}^k,\tilde{v}^k,\tilde{\varepsilon}_k)$ satisfies the stationary inclusion $\tilde{v}^k\in\nabla f(\tilde{y}^k)+\partial h_{\tilde{\varepsilon}_k}(\tilde{y}^k)+A^*\tilde{q}^k$ and the feasibility condition $c\|A\tilde{y}^k-b\|\le 2\kappa_p$. The assertion in Theorem 3.1(b) that S-ADMM outputs a quadruple $(x^+,p^+,v^+,\varepsilon^+)$ satisfying the inclusion (27) and the feasibility bound (28) follow from this fact and lines 11 and 12 of S-ADMM. Moreover, S-ADMM stops whenever it finds a quadruple $(x^+,p^+,v^+,\varepsilon^+)=(\tilde{y}^k,\tilde{q}^k,\tilde{v}^k,\tilde{\varepsilon}_k)$ satisfying $\|\tilde{v}^k\|^2+\tilde{\varepsilon}_k\le\rho^2$, but not necessarily the condition that $\|A\tilde{y}^k-b\|\le\eta$. However, if c is sufficiently large, i.e., it satisfies the condition in Theorem 3.1(c), then the latter condition would also hold because $c\|A\tilde{y}^k-b\|\le 2\kappa_p$ for every k, in which case $(\tilde{y}^k,\tilde{q}^k,\tilde{v}^k,\tilde{\varepsilon}_k)$ would be the desired (ρ,η)-stationary solution of (1).

Fourth, we now formally argue how Theorem 3.1 implies that a single S-ADMM call finds a (ρ,η)-stationary solution of (1) in at most $\mathcal{O}(\eta^{-1}\rho^{-2}+\rho^{-2}+\eta^{-1}+\eta^{-2})$ iterations (assuming that all other quantities in (23) are $\mathcal{O}(1)$) under the following assumptions: i) the initial iterate $x\in\mathcal{H}$ satisfies $\|Ax-b\|=\mathcal{O}(\sqrt{\eta})$; ii) the constant κ_p as in (22), which is generally not computable, and the weak convexity parameters m as in assumption (A3), are known. Indeed, choose the quantities $\bar{Q},\underline{c},c,p,\bar{\gamma}$ and γ in Theorem 3.1 as

$$\bar{Q}=\frac{2\kappa_p\|Ax-b\|}{\eta},\quad \underline{c}=c=\frac{2\kappa_p}{\eta},\quad p=0,\quad \bar{\gamma}=\gamma=\frac{1}{2m}\tag{30}$$

and observe that $\bar{\Gamma}$ as in (22) satisfies $\bar{\Gamma}=\mathcal{O}(1+\|Ax-b\|^2/\eta)=\mathcal{O}(1)$, and hence all the $\Lambda_i's$ as in (23) are also $\mathcal{O}(1)$. Since the quantities in (30) satisfy the assumptions of Theorem 3.1, it follows from its statement (c), the latter observation, and the fact that $c = \mathcal{O}(\eta^{-1})$, that the iteration-complexity of S-ADMM is $\mathcal{O}(\eta^{-1}\rho^{-2} + \rho^{-2} + \eta^{-1} + \eta^{-2})$, and hence $\mathcal{O}(\epsilon^{-3})$ where $\epsilon := \min\{\rho, \eta\}$.

The next subsection shows that A-ADMM, up to a logarithmic term, has the same complexity as in the fourth remark above, but now without assuming the knowledge of κ_p nor that x^+ is near feasible. To obtain such complexity, A-ADMM invokes S-ADMM iteratively and doubles the penalty parameter until a (ρ, η)-stationary solution of (1) is obtained.

3.3 A-ADMM: Description & Complexity

This subsection describes this paper's focal algorithm, A-ADMM, and establishes its accompanying main complexity result (Theorem 3.3). A-ADMM repeatedly calls S-ADMM as a subroutine, each time with the penalty parameter double the one used in the previous call. A-ADMM is formally stated below.

Algorithm 3 A-ADMM

Universal Input: $C > 0$, pair $(\rho, \eta) \in \mathbb{R}^2_{++}$, $\alpha > 0$, $\sigma_1 \leq 1/8$, $\sigma_2 > 0$ and $\sigma_1 + \sigma_2/2 \geq 0$
Input: $(x^0, p^0, \gamma^0, c_0) \in \mathcal{H} \times A(\mathbb{R}^n) \times \mathbb{R}^B_{++} \times \mathbb{R}_{++}$
Output: $(\hat{x}, \hat{p}, \hat{\gamma}, \hat{v}, \hat{\varepsilon})$

1: **for** $\ell \leftarrow 1, 2, \ldots$ **do**
2: $\quad (x^\ell, p^\ell, \gamma^\ell, v^\ell, \varepsilon_\ell) = \text{S-ADMM}(x^{\ell-1}, p^{\ell-1}, \gamma^{\ell-1}, c_{\ell-1})$
3: $\quad$ **if** $\|Ax^\ell - b\| \leq \eta$ **then**
4: $\quad\quad (\hat{x}, \hat{p}, \hat{\gamma}, \hat{v}, \hat{\varepsilon}) = (x^\ell, p^\ell, \gamma^\ell, v^\ell, \varepsilon_\ell)$
5: $\quad\quad$ **return** $(\hat{x}, \hat{p}, \hat{\gamma}, \hat{v}, \hat{\varepsilon})$
6: $\quad$ **else**
7: $\quad\quad c_\ell = 2c_{\ell-1}$

We now make some remarks about A-ADMM. First, the initial penalty parameter c_0 can be chosen to be any positive scalar. Second, the initial Lagrange multiplier p^0 is required to be in $A(\mathbb{R}^n)$, e.g., it can be set to zero. Third, it uses a "warm-start" strategy for calling S-ADMM, i.e., after the first call to S-ADMM, the input of any S-ADMM call is the output of the previous S-ADMM call. Fourth, Lemma 3.2 below and Theorem 3.1(b) imply that each S-ADMM call in line 2 of A-ADMM generates a triple (x^ℓ, p^ℓ, v^ℓ) satisfying the first two conditions in (2). Finally, A-ADMM stops if the test in line 3 is satisfied; if the test fails, then S-ADMM is called again with c_ℓ set to $2c_{\ell-1} = 2^\ell c_0$.

The next result guarantees that all of the hypotheses used in the iteration-complexity theorem for S-ADMM (Theorem 3.1) hold each time A-ADMM calls S-ADMM.

Lemma 3.2 *Assume that $p^0 = 0$ and define*

$$\bar{Q} := \max\{4\kappa_p, c_0\|Ax^0 - b\|\}, \quad \bar{\gamma} = \gamma^0, \quad \underline{c} = c_0 \tag{31}$$

where κ_p is as in (22). If A-ADMM performs the ℓ-th iteration, then

$$\begin{gathered} c_{\ell-1}\|Ax^{\ell-1} - b\| \leq \bar{Q}, \quad \|p^{\ell-1}\| \leq \kappa_p, \quad \|v^\ell\|^2 + \varepsilon_\ell \leq \rho^2, \\ \min\left\{\frac{1}{2\max(m)}, \bar{\gamma}_t\right\} \leq \gamma_t^{\ell-1} \leq \bar{\gamma}_t, \ \forall t \in \{1, \ldots, B\}, \end{gathered} \tag{32}$$

where $\max(m)$ is as in (7). Moreover, the number of S-ADMM iterations performed in the ℓ-th call in line 2 is bounded by $\Lambda(c_{\ell-1}, \rho)$.

Proof: First, note that all inequalities in (32) hold trivially with $\ell = 1$, except the third one. This is due to the assumption $p^0 = 0$ and the definitions of κ_p and $\bar{Q}$ as in (22) and (31), respectively. Second, it follows from the logic of A-ADMM that $c_\ell/2 = c_{\ell-1} \geq c_0 = \underline{c}$. This observation and Theorem 3.1(b) with $(x, p, \gamma, c) = (x^{\ell-1}, p^{\ell-1}, \gamma^{\ell-1}, c_{\ell-1})$ imply that if all inequalities in (32), except the third one, hold for ℓ, then all the inequalities in (32) hold for $\ell + 1$. These two observations and a simple induction argument show that the first conclusion of the lemma holds. The second one follows from (32) and Theorem 3.1(a) with $(x, p, \gamma, c) = (x^{\ell-1}, p^{\ell-1}, \gamma^{\ell-1}, c_{\ell-1})$. ∎

Theorem 3.3 (A-ADMM Complexity) *The following statements about A-ADMM hold:*

(a) it obtains a (ρ, η)-stationary solution of (1) in no more than $\log_2 \left[\bar{Q}/(\underline{c}\eta)\right] + 2$ calls to S-ADMM;

(b) the total number of S-ADMM iterations performed by it is no more than

$$\left(\frac{\Lambda_0}{\rho^2} + \Lambda_2\right)\frac{4\bar{Q}}{\eta} + \left(\Lambda_1 + \frac{\Lambda_4}{\rho^2}\right)\left[\log_2\left(\frac{\bar{Q}}{\underline{c}\eta}\right) + 2\right] + 16\Lambda_3\frac{\bar{Q}^2}{\eta^2},$$

where $(\bar{Q}, \underline{c})$ and $(\Lambda_0, \Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4)$ are the constants that appears in (31) and (23), respectively.

Proof: (a) Assume for the sake of contradiction that A-ADMM generates an iteration index $\hat{\ell}$ such that $\hat{\ell} - 2 > \log_2 \left[\bar{Q}/(\underline{c}\eta)\right]$, and hence satisfying

$$c_{\hat{\ell}-2} = \underline{c}2^{\hat{\ell}-2} > \underline{c}2^{\log_2[\bar{Q}/(\underline{c}\eta)]} = \frac{\bar{Q}}{\eta}.$$

Combining Lemma 3.2, with $\ell = \hat{\ell} - 2$, and Theorem 3.1(a) and (c), with $(x, p, \gamma, c) = (x^{\hat{\ell}-2}, p^{\hat{\ell}-2}, \gamma^{\hat{\ell}-2}, c_{\hat{\ell}-2})$, we conclude that the output $(x^{\hat{\ell}-1}, p^{\hat{\ell}-1}, v^{\hat{\ell}-1}, \gamma^{\hat{\ell}-1})$ of the call to S-ADMM in the $(\hat{\ell}-1)$-th iteration of A-ADMM is a (ρ, η) stationary solution of (1), and hence satisfies $\|Ax^{\hat{\ell}-1} - b\| \leq \eta$. This yields the desired contradiction since the method should have terminated in the $(\hat{\ell}-1)$-th iteration in view of the stopping criterion in its line 3.

(b) It follows from Lemma 3.2 that the number of S-ADMM iterations is bounded by $\Lambda(c_{\tilde{\ell}-1}, \rho)$, where $\tilde{\ell}$ is the last iteration performed by it. Then, we conclude from the previous item (a) that $\tilde{\ell} \leq \log_2[\bar{Q}/(\underline{c}\eta)] + 2$. As a consequence, we can deduce that

$$2^{\tilde{\ell}} \leq \frac{4\bar{Q}}{\underline{c}\eta}. \tag{33}$$

Hence, these conclusions imply that the overall number of iterations performed by S-ADMM is bounded by

$$\begin{aligned}
\sum_{\ell=1}^{\tilde{\ell}} \Lambda(c_{\ell-1}, \rho) &\overset{(26)}{=} \sum_{\ell=1}^{\tilde{\ell}} \left(\Lambda_0 \frac{c_{\ell-1}}{\rho^2} + \Lambda_1 + \Lambda_2 c_{\ell-1} + \Lambda_3 c_{\ell-1}^2 + \Lambda_4 \frac{1}{\rho^2}\right) \\
&= \Lambda_1 \tilde{\ell} + \frac{\underline{c}\Lambda_0}{\rho^2} \sum_{\ell=1}^{\tilde{\ell}} 2^{\ell-1} + \frac{\Lambda_4}{\rho^2}\tilde{\ell} + \Lambda_2 \underline{c} \sum_{\ell=1}^{\tilde{\ell}} 2^{\ell-1} + \Lambda_3 \underline{c}^2 \sum_{\ell=1}^{\tilde{\ell}} 2^{2(\ell-1)} \\
&= \left(\Lambda_1 + \frac{\Lambda_4}{\rho^2}\right)\tilde{\ell} + \left(\frac{\underline{c}\Lambda_0}{\rho^2} + \Lambda_2 \underline{c}\right)(2^{\tilde{\ell}} - 1) + \frac{\Lambda_3 \underline{c}^2 (4^{\tilde{\ell}} - 1)}{3} \\
&\overset{(33)}{<} \left(\Lambda_1 + \frac{\Lambda_4}{\rho^2}\right)\left[\log_2\left(\frac{\bar{Q}}{\underline{c}\eta}\right) + 2\right] + \left(\frac{\Lambda_0}{\rho^2} + \Lambda_2\right)\frac{4\bar{Q}}{\eta} + 16\Lambda_3 \frac{\bar{Q}^2}{\eta^2}.
\end{aligned}$$

■

We now make some comments about Theorem 3.3. First, it follows from Theorem 3.3(a) that A-ADMM ends with a (ρ, η)-stationary solution of (1) by calling S-ADMM no more than $\mathcal{O}(\log_2(\eta^{-1}))$ times. Second, Theorem 3.3(b) implies that the complexity of A-ADMM, in terms of the tolerances only, is

$$\mathcal{O}(\rho^{-2} + \rho^{-2}\eta^{-1} + \eta^{-2}).$$

Third, if $\epsilon := \min\{\rho, \eta\}$ then the iteration-complexity of A-ADMM is $\mathcal{O}(\epsilon^{-3})$.

4 The Proof of S-ADMM's Complexity Theorem (Theorem 3.1)

The goal of this section is proving the theorem governing the iteration-complexity of S-ADMM (Theorem 3.1). Our approach is comprised of three main steps. First, Subsection 4.1 bounds the number of epochs, $\mathcal{I}_1, \mathcal{I}_2, \ldots$. Second, Subsection 4.2 derives a uniform bound on the size of each of the epochs. Third and finally, Subsection 4.3 synthesizes the results of the previous two steps to prove Theorem 3.1.

4.1 Step 1: Bounding the Number of Epochs

The goal of this subsection is to bound the total number of epochs $\mathcal{I}_1, \mathcal{I}_2, \ldots$ generated by S-ADMM. Before presenting the main result of this subsection, we state and prove some preliminary technical results. The first one summarizes some straightforward facts about S-ADMM.

Lemma 4.1 *The following statements about an epoch $\mathcal{I}_k$ generated by S-ADMM hold:*

(a) if $\mathcal{I}_k$ ends, then $\tilde{v}^k = v^{i_k^+}$, $\tilde{y}^k = y^{i_k^+}$, $\tilde{\varepsilon}_k = \delta_{i_k^+}$ and $\tilde{T}_k = T_{i_k^+}$; moreover,

$$\|\tilde{v}^k\|^2 + \tilde{\varepsilon}_k \le C^2 \quad \text{and} \quad i_k^+ \ge \frac{k\alpha}{\rho^2}\tilde{T}_k; \tag{34}$$

(b) if $i < j$ are indices lying in $\mathcal{I}_k \cup \{i_{k-1}^+\}$, then $T_j - T_i = \mathcal{L}_c(y^i; \tilde{q}^{k-1}) - \mathcal{L}_c(y^j; \tilde{q}^{k-1})$.

Proof: (a) Recalling from (21) that i_k^+ is the last index of $\mathcal{I}_k$, the identities $\tilde{v}^k = v^{i_k^+}$, $\tilde{y}^k = y^{i_k^+}$, $\tilde{\varepsilon}_k = \delta_{i_k^+}$ and $\tilde{T}_k = T_{i_k^+}$ immediately follow from lines 16 and 17 of S-ADMM. Thus, the inequalities in (34) readily result from the tests in line 15 with $i = i_k^+$, and the aforementioned identities.

(b) We first note that if $i = i_{k-1}^+$ then the previous item of this lemma, applied with $k-1$ replacing k, implies that $\tilde{T}_{k-1} = T_{i_{k-1}^+}$ and $\tilde{y}^{k-1} = y^{i_{k-1}^+}$. Hence, the conclusion follows from the definition of T_i in line 14 of S-ADMM with $i = j$. Now, if $i < j$ are indices lying in $\mathcal{I}_k$, then the statement is also a direct consequence of the definition of T_i and T_j as in line 14 of S-ADMM. ∎

The following result describes the properties of the iterate $(y_t, r_t, \varepsilon_t)$ obtained in line 5 of S-ADMM using the data of subproblem (9).

Lemma 4.2 *Every triple $(y_t, r_t, \varepsilon_t)$ generated in line 5 of S-ADMM satisfies*

$$r_t \in \nabla\left[\lambda_t \hat{\mathcal{L}}_c(y^i_{<t}, \cdot, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|\cdot - y_t^{i-1}\|^2\right](y_t) + \partial_{\varepsilon_t}(\lambda_t h_t)(y_t), \tag{35}$$

$$(\lambda_t^{-1} + 2)\|r_t\|^2 + 2\varepsilon_t \le \sigma_1\|y_t - y_t^{i-1}\|^2 + \lambda_t\sigma_2(\Delta\mathcal{L}_c)^i_t(y_t), \tag{36}$$

where $\hat{\mathcal{L}}_c(\cdot\,; \tilde{q}^{k-1})$ is as in (11) and

$$(\Delta\mathcal{L}_c)^i_t(\cdot) := \mathcal{L}_c(y^i_{<t}, y^{i-1}_t, y^{i-1}_{>t}; \tilde{q}^{k-1}) - \mathcal{L}_c(y^i_{<t}, \cdot, y^{i-1}_{>t}; \tilde{q}^{k-1}). \tag{37}$$

Proof: We first prove (35). It follows from line 5 of S-ADMM that $(y_t, r_t, \varepsilon_t)$ is a $(\tau_1, \tau_2, 0; z^0)$-stationary solution of (9) with (ψ_s, ψ_n) given by (10). Hence, in view of Definition 2.1 and the definition of (τ_1, τ_2, z^0) in line 5 of S-ADMM, we conclude that (35) holds and

$$\begin{aligned}\|r_t\|^2 + 2\varepsilon_t &\overset{(13)}{\le} \left(\sigma_1 + \frac{\sigma_2}{2}\right)\frac{\lambda_t}{1+2\lambda_t}\|y_t - y_t^{i-1}\|^2 + \frac{\sigma_2\lambda_t}{1+2\lambda_t}[(\psi_s+\psi_n)(y_t^{i-1}) - (\psi_s+\psi_n)(y_t)] \\ &= \frac{\sigma_1\lambda_t}{1+2\lambda_t}\|y_t - y_t^{i-1}\|^2 + \frac{\sigma_2\lambda_t}{1+2\lambda_t}\lambda_t(\Delta\mathcal{L}_c)^i_t(y_t),\end{aligned}$$

where the equality is due to the fact that (10) and (37) imply that

$$(\psi_s+\psi_n)(y_t^{i-1}) - (\psi_s+\psi_n)(y_t) = \lambda_t(\Delta\mathcal{L}_c)^i_t(y_t) - \frac{1}{2}\|y_t - y_t^{i-1}\|^2.$$

∎

The next result shows that the loop consisting of lines 5 to 9 of S-ADMM ends and describes a property of its output.

Lemma 4.3 *Let $i \in \mathcal{I}_k$, $t \in \{1,\ldots,B\}$, and the triple (λ_t, y_t, r_t) be obtained from lines 4 and 5 of* S-ADMM*. If λ_t lies in the interval $(0, 1/m_t)$, then (λ_t, y_t, r_t) satisfies*

$$(1+\sigma_2)(\Delta\mathcal{L}_c)_t^i(y_t) \geq \frac{1}{4\lambda_t}\|y_t - y_t^{i-1}\|^2 + \frac{c}{4}\|A_t(y_t - y_t^{i-1})\|^2 + \frac{\|r_t\|^2}{\lambda_t^2}, \tag{38}$$

where $\sigma_2 > 0$ is part of the input for S-ADMM *and $(\Delta\mathcal{L}_c)_t^i(y_t)$ is as in (37).*

Proof: Let $i \in \mathcal{I}_k$, $t \in \{1,\ldots,B\}$, and $\lambda_t \in (0, 1/m_t)$ be given. We first claim that, for every $u_t \in \mathcal{H}_t$, the quadruple $(\lambda_t, y_t, r_t, \varepsilon_t)$ obtained in lines 4 and 5 of S-ADMM satisfies

$$\begin{aligned}\mathcal{L}_c(y^i_{<t}, u_t, y^{i-1}_{>t}; \tilde{q}^{k-1}) - \mathcal{L}_c(y^i_{<t}, y_t, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2\lambda_t}\|u_t - y_t^{i-1}\|^2 \\ \geq \frac{1}{2\lambda_t}\|y_t - y_t^{i-1}\|^2 + \frac{c}{4}\|A_t(u_t - y_t)\|^2 - \frac{1}{\lambda_t}\left(\|r_t\|\|y_t - u_t\| + 2\varepsilon_t\right).\end{aligned} \tag{39}$$

Since $\lambda_t \in (0, 1/m_t)$, the matrix $B_t := (1 - \lambda_t m_t)I + \lambda_t c A_t^* A_t$ satisfies $B_t \succ 0$, and hence defines the norm $\|\cdot\|_{B_t}$ whose square satisfies

$$\|\cdot\|_{B_t}^2 := \langle\, \cdot\,,\, B_t(\cdot)\,\rangle \geq \lambda_t c\|A_t(\cdot)\|^2. \tag{40}$$

Moreover, using assumption (A3), we can easily see that the function

$$\lambda_t \mathcal{L}_c(y^i_{<t}, \cdot, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|\cdot - y_t^{i-1}\|^2 - \langle r_t, u_t\rangle$$

is 1-strongly convex with respect to this norm. Using that the triple $(y_t, r_t, \varepsilon_t)$ satisfies

$$r_t \in \partial_{\varepsilon_t}\left(\lambda_t \hat{\mathcal{L}}_c(y^i_{<t}, \cdot, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|\cdot - y_t^{i-1}\|^2 + \lambda_t h_t(\cdot)\right)(y_t),$$

due to (35), we can apply Lemma A.4 with $(\xi, \tau, Q, \hat{\eta}) = (1, 1, B_t, \varepsilon_t)$, to conclude that

$$\begin{aligned}\lambda_t \mathcal{L}_c(y^i_{<t}, u_t, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|u_t - y_t^{i-1}\|^2 - \frac{1}{4}\|u_t - y_t\|^2_{B_t} \\ \geq \lambda_t \mathcal{L}_c(y^i_{<t}, y_t, y^{i-1}_{>t}; \tilde{q}^{k-1}) + \frac{1}{2}\|y_t - y_t^{i-1}\|^2 - 2\varepsilon_t + \langle r_t, u_t - y_t\rangle.\end{aligned}$$

The preliminary claim (39) follows by dividing the previous inequality by λ_t, and using both (40) and the Cauchy-Schwarz inequality.

We now prove (38). Using (39) with $u_t = y_t^{i-1}$ and the definition of $(\Delta\mathcal{L}_c)_t^i(y_t)$ (see (37)), we have

$$(\Delta\mathcal{L}_c)_t^i(y_t) \geq \frac{1}{2\lambda_t}\|y_t - y_t^{i-1}\|^2 + \frac{c}{4}\|A_t(y_t^{i-1} - y_t)\|^2 - \frac{1}{\lambda_t}(\|r_t\|\|y_t - y_t^{i-1}\| + 2\varepsilon_t). \tag{41}$$

Using the inequality $2ab \leq a^2 + b^2$ with $(a,b) = (2\|r_t\|, (1/2)\|y_t - y_t^{i-1}\|)$ and the condition on the error (r_t, ε_t) as in (36), we conclude that

$$\begin{aligned}&\frac{1}{\lambda_t}\left(\frac{1}{\lambda_t}\|r_t\|^2 + \|r_t\|\|y_t - y_t^{i-1}\| + 2\varepsilon_t\right) \\ &\leq \frac{1}{\lambda_t}\left(\frac{1}{\lambda_t}\|r_t\|^2 + 2\|r_t\|^2 + \frac{1}{8}\|y_t - y_t^{i-1}\|^2 + 2\varepsilon_t\right) \\ &= \frac{1}{\lambda_t}\left[(\lambda_t^{-1} + 2)\|r_t\|^2 + \frac{1}{8}\|y_t - y_t^{i-1}\|^2 + 2\varepsilon_t\right] \\ &= \frac{1}{\lambda_t}\left[(\lambda_t^{-1} + 2)\|r_t\|^2 + 2\varepsilon_t\right] + \frac{1}{8\lambda_t}\|y_t - y_t^{i-1}\|^2 \\ &\overset{(36)}{\leq} \frac{\sigma_1 + (1/8)}{\lambda_t}\|y_t - y_t^{i-1}\|^2 + \sigma_2(\Delta\mathcal{L}_c)_t^i(y_t) \\ &\leq \frac{1}{4\lambda_t}\|y_t - y_t^{i-1}\|^2 + \sigma_2(\Delta\mathcal{L}_c)_t^i(y_t),\end{aligned}$$

where the last inequality is due to the assumption $\sigma_1 \leq 1/8$. The conclusion now follows by combining the previous inequality with (41). ∎

To shorten the formulas in this subsection, for each $i \in \mathbb{N}$ and $t \in \{1, \ldots, B\}$, we define the quantity

$$\Delta y_t^i := y_t^i - y_t^{i-1}. \tag{42}$$

Lemma 4.4 *Let $i \in \mathcal{I}_k$, $t \in \{1, \ldots, B\}$, and the triple (λ_t, y_t, r_t) be obtained from lines 4 and 5 of S-ADMM. Then, the loop consisting of lines 5 to 9 terminates with a triple $(\lambda_t^i, y_t^i, r_t^i) \in \mathbb{R}_{++}^B \times \mathcal{H} \times \mathcal{H}$ satisfying*

$$(1+\sigma_2)(\Delta\mathcal{L}_c)_t^i(y_t^i) \geq \frac{1}{4\lambda_t^i}\|\Delta y_t^i\|^2 + \frac{c}{4}\|A_t \Delta y_t^i\|^2 + \frac{\|r_t^i\|^2}{(\lambda_t^i)^2} \tag{43}$$

and

$$\min\left\{\frac{1}{2m_t}, \gamma_t\right\} \leq \lambda_t^i \leq \gamma_t, \tag{44}$$

where $(c, \gamma, \sigma_2) \in \mathbb{R}_{++} \times \mathbb{R}_{++}^B \times \mathbb{R}_{++}$ is part of the input for S-ADMM, Δy_t^i is as in (42), and $(\Delta\mathcal{L}_c)_t^i(\cdot)$ is as in (37).

Proof: First, note that if (λ_t, y_t, r_t), as in line 5 of S-ADMM, is such that $\lambda_t \in (0, 1/m_t)$, then it follows from Lemma 4.3 that the loop consisting of lines 5 to 9 terminates with $(\lambda_t^i, y_t^i, r_t^i) = (\lambda_t, y_t, r_t)$. The above observation, in conjunction with the facts that the loop is initialized with $\lambda_t = \lambda_t^{i-1}$ and λ_t is halved every time a loop iteration fails to satisfy (38), then imply that

$$\min\left\{\frac{1}{2m_t}, \lambda_t^{i-1}\right\} \leq \lambda_t^i \leq \lambda_t^{i-1}.$$

This conclusion together with the fact that $\lambda_t^0 = \gamma_t$ can now be easily seen to imply the conclusion of the lemma. ∎

The next result shows that v^i, computed in line 13 of S-ADMM, is a stationary residual for the pair (y^i, q^i) and also establishes an upper bound for it.

Lemma 4.5 *Let $\mathcal{I}_k$ be an epoch generated by S-ADMM and define*

$$q^i := \tilde{q}^{k-1} + c(Ay^i - b) \quad \forall i \in \mathcal{I}_k. \tag{45}$$

Then, for every $i \in \mathcal{I}_k$, we have

$$v^i \in \nabla f(y^i) + \partial_{\delta_i} h(y^i) + A^* q^i \tag{46}$$

and

$$\|v^i\|^2 + \delta_i \leq (\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2)(\Delta\mathcal{L}_c)^i \tag{47}$$

where $(c, \gamma, \sigma_2) \in \mathbb{R}_{++} \times \mathbb{R}_{++}^B \times \mathbb{R}_{++}$ is part of the input for S-ADMM, (v^i, δ_i) is as in line 13 of S-ADMM,

$$(\Delta\mathcal{L}_c)^i := \mathcal{L}_c(y^{i-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(y^i; \tilde{q}^{k-1}) \tag{48}$$

and

$$\chi(\gamma) := 2(1+\sigma_2)\left[1 + 12\|L\|^2 \max(\gamma) + 12(2\max(m) + \max(\gamma^{-1})) + 2\sigma_1\right] + \sigma_2. \tag{49}$$

Proof: We first prove that (46) holds. For each $t \in \{1, \ldots, B\}$, let $(\lambda_t^i, y_t^i, r_t^i, \varepsilon_t^i)$ be the quadruple obtained in line 9 of S-ADMM. Using (35) with $(\lambda_t, y_t, r_t, \varepsilon_t) = (\lambda_t^i, y_t^i, r_t^i, \varepsilon_t^i)$, we have

$$\begin{aligned}
\frac{r_t^i}{\lambda_t^i} &\overset{(35)}{\in} \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^{i-1}) + A_t^*\left[\tilde{q}^{k-1} + c[A(y_{<t}^i, y_t^i, y_{>t}^{i-1}) - b]\right] + \frac{1}{\lambda_t^i}\Delta y_t^i + \partial_{(\varepsilon_t^i/\lambda_t^i)} h_t(y_t^i) \\
&\overset{(45)}{=} \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^{i-1}) + A_t^*\left(q^i - c\sum_{s=t+1}^{B} A_s \Delta y_s^i\right) + \frac{1}{\lambda_t^i}\Delta y_t^i + \partial_{(\varepsilon_t^i/\lambda_t^i)} h_t(y_t^i),
\end{aligned}$$

where in the inclusion above we used the well-known ε-subdifferential property that $\partial_{\varepsilon_t}\lambda_t h(y_t) = \lambda_t \partial_{\varepsilon_t/\lambda_t} h(y_t)$ (e.g., see [19, Proposition 1.3.1]). By rearranging the above inclusion and using the definition of $(\Delta f)_t^i$, as in line 11 of S-ADMM, we have

$$(\Delta f)_t^i + \frac{r_t^i}{\lambda_t^i} + cA_t^* \sum_{s=t+1}^{B} A_s \Delta y_s^i - \frac{1}{\lambda_t^i}\Delta y_t^i \in \nabla_{y_t} f(y^i) + \partial_{(\varepsilon_t^i/\lambda_t^i)} h_t(y_t^i) + A_t^* q^i.$$

Noting that the left-hand side of the above inclusion equals v_t^i (see line 12 of S-ADMM) and using the well-known fact that $\delta_i = (\varepsilon_1^i/\lambda_1^i) + \ldots + (\varepsilon_B^i/\lambda_B^i)$ implies that

$$\partial_{\delta_i} h(y^i) \supset \partial_{(\varepsilon_1^i/\lambda_1^i)} h_1(y_1^i) \times \ldots \times \partial_{(\varepsilon_B^i/\lambda_B^i)} h_B(y_B^i),$$

we conclude that inclusion (46) holds.

We now prove (47). It follows from the definition of $(\Delta f)_t^i$ and v_t^i in lines 11 and 12 of S-ADMM, respectively, the triangle and Cauchy-Schwarz inequalities, and assumption (A4), that

$$\begin{aligned}
\left\| v_t^i - \frac{r_t^i}{\lambda_t^i} \right\|^2 &= \left\| \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^i) - \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^{i-1}) + cA_t^* \sum_{s=t+1}^{B} A_s \Delta y_s^i - \frac{1}{\lambda_t^i}\Delta y_t^i \right\|^2 \\
&\le 3\left(\|\nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^i) - \nabla_{y_t} f(y_{<t}^i, y_t^i, y_{>t}^{i-1})\|^2 + c^2\|A_t\|^2 \left\| \sum_{s=t+1}^{B} A_s \Delta y_s^i \right\|^2 + \frac{1}{(\lambda_t^i)^2}\|\Delta y_t^i\|^2 \right) \\
&\overset{(18)}{\le} 3\left(L_t^2 \|y_{>t}^i - y_{>t}^{i-1}\|^2 + c^2\|A_t\|^2 (B-t) \sum_{s=t+1}^{B} \|A_s \Delta y_s^i\|^2 + \frac{1}{(\lambda_t^i)^2}\|\Delta y_t^i\|^2 \right) \\
&\le 3\left(L_t^2 \sum_{s=1}^{B} \|\Delta y_s^i\|^2 + Bc^2\|A_t\|^2 \sum_{s=1}^{B} \|A_s \Delta y_s^i\|^2 + \frac{1}{(\lambda_t^i)^2}\|\Delta y_t^i\|^2 \right).
\end{aligned} \tag{50}$$

We now upper bound the first and third terms of the right-hand of (50). For the first term, we have

$$\sum_{s=1}^{B} \|\Delta y_s^i\|^2 = \sum_{s=1}^{B} \lambda_s^i \cdot \frac{1}{\lambda_s^i}\|\Delta y_s^i\|^2 \overset{(44)}{\le} \sum_{s=1}^{B} \gamma_s \cdot \frac{1}{\lambda_s^i}\|\Delta y_s^i\|^2 \overset{(7)}{\le} \max(\gamma) \cdot \sum_{s=1}^{B} \frac{1}{\lambda_s^i}\|\Delta y_s^i\|^2. \tag{51}$$

To bound the third term of (50), first note that for $t \in \{1, \ldots, B\}$,

$$\frac{1}{\lambda_t^i} \overset{(44)}{\le} \left(\min\left\{ \frac{1}{2m_t}, \gamma_t \right\} \right)^{-1} = \max\left\{ 2m_t, \frac{1}{\gamma_t} \right\} \overset{(7)}{\le} 2\max(m) + \max(\gamma^{-1}) =: \hat{m}, \tag{52}$$

which yields

$$\frac{1}{(\lambda_t^i)^2}\|\Delta y_t^i\|^2 = \frac{1}{\lambda_t^i} \cdot \frac{1}{\lambda_t^i}\|\Delta y_t^i\|^2 \le \frac{\hat{m}}{\lambda_t^i}\|\Delta y_t^i\|^2. \tag{53}$$

Combining the inequalities (50), (51), (52), and (53), and using (43), we have

$$\begin{aligned}
\left\| v_t^i - \frac{r_t^i}{\lambda_t^i} \right\|^2 &\le 3\left(L_t^2 \max(\gamma) \sum_{s=1}^{B} \frac{1}{\lambda_s^i}\|\Delta y_s^i\|^2 + Bc^2\|A_t\|^2 \sum_{s=1}^{B} \|A_s \Delta y_s^i\|^2 + \frac{\hat{m}}{\lambda_t^i}\|\Delta y_t^i\|^2 \right). \\
&\overset{(43)}{\le} 3\left[4L_t^2 \max(\gamma)(1+\sigma_2)(\Delta\mathcal{L}_c)^i + 4Bc\|A_t\|^2(1+\sigma_2)(\Delta\mathcal{L}_c)^i + 4\hat{m}(1+\sigma_2)(\Delta\mathcal{L}_c)_t^i \right] \\
&= 12(1+\sigma_2)\left[L_t^2 \max(\gamma)(\Delta\mathcal{L}_c)^i + Bc\|A_t\|^2(\Delta\mathcal{L}_c)^i + \hat{m}(\Delta\mathcal{L}_c)_t^i \right],
\end{aligned}$$

where $(\Delta\mathcal{L}_c)^i$ and $(\Delta\mathcal{L}_c)_t^i(\cdot)$ are as in (48) and (37), respectively. Using the reverse triangle inequality in the left-hand side of the previous inequality, we conclude after simple manipulations that

$$\|v_t^i\|^2 - 2\frac{\|r_t^i\|^2}{(\lambda_t^i)^2} \le 24(1+\sigma_2)\left[L_t^2 \max(\gamma)(\Delta\mathcal{L}_c)^i + Bc\|A_t\|^2(\Delta\mathcal{L}_c)^i + \hat{m}(\Delta\mathcal{L}_c)_t^i \right].$$

Now, summing up the last inequality from $t = 1$ to $t = B$ and using (20) and (48), one has

$$\begin{aligned} \|v^i\|^2 + \sum_{t=1}^{B} \frac{\varepsilon_t^i}{\lambda_t^i} &\overset{(20)}{\leq} \left(\sum_{t=1}^{B} \frac{2\|r_t^i\|^2}{(\lambda_t^i)^2} + \frac{\varepsilon_t^i}{\lambda_t^i} \right) + 24(1+\sigma_2)\left[\|L\|^2 \max(\gamma) + Bc\|A\|_\dagger^2 + \hat{m}\right](\Delta\mathcal{L}_c)^i \\ &\overset{(43)}{\leq} \sum_{t=1}^{B} \frac{\varepsilon_t^i}{\lambda_t^i} + 2(1+\sigma_2)[1 + 12(\|L\|^2 \max(\gamma) + Bc\|A\|_\dagger^2 + \hat{m})](\Delta\mathcal{L}_c)^i \\ &\leq [4\sigma_1(1+\sigma_2) + \sigma_2](\Delta\mathcal{L}_c)^i + 2(1+\sigma_2)[1 + 12(\|L\|^2 \max(\gamma) + Bc\|A\|_\dagger^2 + \hat{m})](\Delta\mathcal{L}_c)^i, \end{aligned}$$

where the last inequality above is due to (36) with $(\lambda_t^i, y_t^i, \varepsilon_t^i) = (\lambda_t, y_t, \varepsilon_t)$ and (43). The conclusion now follows from the previous inequality and the definitions of $\hat{m}$ and $\chi(\gamma)$ as in (52) and (49), respectively. ■

The following result shows that $\|v^i\|^2 = \mathcal{O}(T_i - T_{i-1})$, and hence that $\{\|v^i\|^2\}$ is summable.

Lemma 4.6 *For any iteration $i \geq 1$ of S-ADMM, we have*

$$\frac{1}{\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2} \left(\|v^i\|^2 + \delta_i\right) \leq T_i - T_{i-1}, \tag{54}$$

where (δ_i, T_i) is as in lines 13 and 14 of S-ADMM (with $T_0 := 0$ by convention), $(\gamma, c) \in \mathbb{R}_{++}^B \times \mathbb{R}_{++}$ is part of the input for S-ADMM, and $\chi(\gamma)$ is as in (49). As a consequence, $T_i \leq T_j$ for any two iterations $i < j$ of S-ADMM.

Proof: Consider an arbitrary iteration index i of S-ADMM and assume that the k-th epoch $\mathcal{I}_k$ is the one that contains this index. Then,

$$(\Delta\mathcal{L}_c)^i \overset{(47)}{\geq} \frac{1}{[\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2]} \left(\|v^i\|^2 + \delta_i\right).$$

Thus, it follows from the definition of $(\Delta\mathcal{L}_c)^i$ as in (48) and Lemma 4.1 with $(i, j) = (i-1, i)$, that the left-hand side of the previous inequality is equal to $T_i - T_{i-1}$, and this concludes the proof. ■

We now present the main result of this subsection, which establishes an upper bound on the number of epochs generated by S-ADMM.

Proposition 4.7 *The total number of epochs performed by S-ADMM is bounded by*

$$\bar{K} := \left\lceil \frac{\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2}{\alpha} \right\rceil, \tag{55}$$

where $(\gamma, c, \sigma_2) \in \mathbb{R}_{++}^B \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ are part of the input for S-ADMM, and $\chi(\gamma)$ is as in (49).

Proof: Assume for the sake of contradiction, that S-ADMM generates an epoch $\mathcal{I}_k$ such that $k \geq \bar{K} + 1$. Summing the inequality (54) from $i = 1$ to $i = i_{k-1}^+$ and using the fact that $T_0 = 0$ and $T_{i_{k-1}^+} = \tilde{T}_{k-1}$, due to Lemma 4.1(a), we have

$$\sum_{i=1}^{i_{k-1}^+} \left(\|v^i\|^2 + \delta_i\right) \overset{(54)}{\leq} [\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2][T_{i_{k-1}^+} - T_0] = [\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2]\tilde{T}_{k-1}. \tag{56}$$

Since S-ADMM did not terminate during epochs $\mathcal{I}_1, \ldots, \mathcal{I}_{k-1}$, it follows from its termination criterion in line 18 that $\|v^i\|^2 + \delta_i > \rho^2$ for every iteration $i \leq i_{k-1}^+$. This observation and (56) then imply that

$$\begin{aligned} \rho^2 < \frac{1}{i_{k-1}^+} \sum_{i=1}^{i_{k-1}^+} \left(\|v^i\|^2 + \delta_i\right) &\overset{(56)}{\leq} [\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2] \frac{\tilde{T}_{k-1}}{i_{k-1}^+} \\ &\leq [\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2] \frac{\rho^2}{(k-1)\alpha} \overset{(55)}{\leq} \frac{\bar{K}}{k-1}\rho^2, \end{aligned}$$

where the second last inequality is due to the last inequality of Lemma 4.1(a) with $k = k-1$. Since by assumption $k - 1 \geq \bar{K}$, the above inequality yields the desired contradiction. ■

4.2 Step 2: Bounding the Epoch Size

The goal of this subsection is to prove a uniform bound on sizes of the epochs $\mathcal{I}_1, \mathcal{I}_2, \ldots$. The first result of this subsection, which establishes a bound on the difference of the augmented Lagrangian function evaluated at two different pairs, will be used a few times later on.

Lemma 4.8 *Given triples $(z^j, u^j, c) \in \mathcal{H} \times \mathbb{R}^l \times \mathbb{R}_{++}$ with $j = 1, 2$. If $\bar\zeta = \max_j \|u^j\|$, then*

$$\mathcal{L}_c(z^1; u^1) - \mathcal{L}_c(z^2; u^2) \le \overline{\phi} - \underline{\phi} + \frac{(c\|Az^1 - b\| + \bar\zeta)^2}{2c} \tag{57}$$

where $(\overline{\phi}, \underline{\phi})$ is as in (19).

Proof: Using the definitions of $\mathcal{L}_c(\cdot\,;\cdot)$ (with $(\theta,\chi) = (0,1)$) and $\underline{\phi}$ as in (3) and (19), respectively, we have

$$\begin{aligned}\mathcal{L}_c(z^2; u^2) - \underline{\phi} &\overset{(19)}{\ge} \mathcal{L}_c(z^2; u^2) - (f+h)(z^2) \\ &\overset{(3)}{=} \langle u^2, Az^2 - b\rangle + \frac{c}{2}\|Az^2 - b\|^2 = \frac{1}{2}\left\|\frac{u^2}{\sqrt{c}} + \sqrt{c}(Az^2 - b)\right\|^2 - \frac{\|u^2\|^2}{2c} \ge -\frac{\bar\zeta^2}{2c},\end{aligned}$$

where the last inequality is due to the definition of $\bar\zeta$. On the other hand, using the definitions of $\mathcal{L}_c(\cdot\,;\cdot)$ (with $(\theta,\chi) = (0,1)$) and $\bar\phi$ as in (3) and (19), respectively, and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\mathcal{L}_c(z^1; u^1) - \overline{\phi} &\overset{(19)}{\le} \mathcal{L}_c(z^1; u^1) - (f+h)(z^1) \overset{(3)}{=} \langle u^1, Az^1 - b\rangle + \frac{c}{2}\|Az^1 - b\|^2 \\ &\le \|u^1\|\|Az^1 - b\| + \frac{c}{2}\|Az^1 - b\|^2 \le \bar\zeta\|Az^1 - b\| + \frac{c}{2}\|Az^1 - b\|^2,\end{aligned}$$

where the last inequality is due to the definition of $\bar\zeta$. Combining the above two relations we then conclude that (57) holds. ■

For any $k \ge 1$, let

$$\mathcal{I}_k(C) := \{i \in \mathcal{I}_k : \|v^i\|^2 + \delta_i \le C^2\}, \tag{58}$$

where $C > 0$ is part of the input for S-ADMM and (v^i, δ_i) is as in line 13 of S-ADMM. The next result shows that the Lagrange multiplier q^j remains bounded as long as $j \in \mathcal{I}_k(C)$. It also shows that if $i \in \mathcal{I}_k(C)$ and epoch $\mathcal{I}_k$ does not terminate at the i-th iteration, then it generates another index $j \in \mathcal{I}_k(C)$ such that $0 < j - i = \mathcal{O}(c)$.

Lemma 4.9 *The following statements about an epoch $\mathcal{I}_k$ generated by S-ADMM hold:*

(a) for every $j \in \mathcal{I}_k$, we have

$$T_j - \tilde T_{k-1} \le \Gamma_{k-1} \tag{59}$$

where

$$\Gamma_{k-1} := \overline{\phi} - \underline{\phi} + \frac{(c\|A\tilde y^{k-1} - b\| + \|\tilde q^{k-1}\|)^2}{2c}; \tag{60}$$

(b) if either $i = i^+_{k-1}$ or $i \in \mathcal{I}_k(C)$ and $\mathcal{I}_k$ does not terminate at the i-th iteration, then there exists $j \in \mathcal{I}_k(C)$ such that

$$0 < j - i \le \left\lceil \frac{\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2}{C^2}\Gamma_{k-1}\right\rceil, \tag{61}$$

where $\chi(\gamma)$ is as in (49) and σ_2 is part of the input for S-ADMM;

(c) if $\tilde q^{k-1} \in A(\mathbb{R}^n)$ and $\|\tilde q^{k-1}\| \le \kappa_p$, then for every $j \in \mathcal{I}_k(C)$, the pair (y^j, q^j), where q^j as in (45), satisfies

$$q^j \in A(\mathbb{R}^n), \quad \|q^j\| \le \kappa_p \quad \text{and} \quad c\|Ay^j - b\| \le 2\kappa_p. \tag{62}$$

Proof: (a) Assume that $j \in \mathcal{I}_k$. Using the definition of T_j as in line 14 of S-ADMM, and Lemma 4.8 with $(u^1, u^2) = (\tilde{q}^{k-1}, \tilde{q}^{k-1})$ and $(z^1, z^2) = (\tilde{y}^{k-1}, y^j)$, we have

$$T_j - \tilde{T}_{k-1} = \mathcal{L}_c(\tilde{y}^{k-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(y^j; \tilde{q}^{k-1}) \overset{(57)}{\leq} \overline{\phi} - \underline{\phi} + \frac{(c\|A\tilde{y}^{k-1} - b\| + \|\tilde{q}^{k-1}\|)^2}{2c} \overset{(60)}{=} \Gamma_{k-1},$$

which yields (59).

(b) Assume that $i \in \mathcal{I}_k(C) \cup \{i_{k-1}^+\}$ is an iteration for which the termination criterion of the k-th epoch is not satisfied and assume for the sake of contradiction that $j \notin \mathcal{I}_k(C)$ for every $j \in \{i+1, \ldots, i+s\}$, where

$$s := \left\lceil \frac{\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2}{C^2} \Gamma_{k-1} \right\rceil. \tag{63}$$

Since by assumption the k-th epoch does not stop at iteration i and it can only stop at some iteration j lying in $\mathcal{I}_k(C)$ (see line 15 of S-ADMM), we conclude that every $j \in \{i+1, \ldots, i+s\}$ lies in $\mathcal{I}_k$ and $\|v^j\|^2 + \delta_j > C^2$. Using this conclusion, (59), Lemma 4.6, the fact $\tilde{T}_{k-1} = T_{i_{k-1}^+}$ (see Lemma 4.1(a)), and the nondecreasing property of the sequence T_i (see Lemma 4.6), we conclude that

$$\begin{aligned}\Gamma_{k-1} &\overset{(59)}{\geq} T_{i+s} - \tilde{T}_{k-1} = T_{i+s} - T_{i_{k-1}^+} \geq T_{i+s} - T_i = \sum_{l=i+1}^{i+s} (T_l - T_{l-1}) \\ &\overset{(54)}{\geq} \frac{1}{\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2} \sum_{l=i+1}^{i+s} (\|v^l\|^2 + \delta_l) > \frac{sC^2}{\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2} \geq \Gamma_{k-1},\end{aligned}$$

where the last inequality is due to (63). Since the above inequality is not possible, statement (b) follows.

(c) The inclusion in (62) follows from the definition of q^j in (45) and the assumption that $\tilde{q}^{k-1} \in A(\mathbb{R}^n)$. We now prove the first inequality in (62). Relations (45) and (46) imply that the triple $(z, q, r) = (y^j, q^j, v^j - \nabla f(y^j))$ satisfies the conditions in (79) of Lemma A.3 with $(\chi, \delta) = (c, \delta_j)$, and hence the conclusion of this lemma implies that

$$\|q^j\| \leq \max\left\{ \|\tilde{q}^{k-1}\|, \frac{2D_h(M_h + \|v^j - \nabla f(y^j)\|) + \delta_j}{\bar{d}\nu_A^+} \right\},$$

where M_h, $\bar{d} > 0$ and D_h are as in assumptions (A5), (A6) and (19), respectively, and ν_A^+ denotes the smallest positive singular value of A. This inequality together with the triangle inequality, the facts that $\|v^j\| \leq C$, $\delta_j \leq C^2$, and $\|\nabla f(y^j)\| \leq \nabla_f$, due to (58) and (19), respectively, and the assumption that $\|\tilde{q}^{k-1}\| \leq \kappa_p$, imply that

$$\|q^j\| \overset{(19)}{\leq} \max\left\{ \kappa_p, \frac{2D_h(M_h + C + \nabla_f) + C^2}{\bar{d}\nu_A^+} \right\} \overset{(22)}{=} \kappa_p,$$

and hence the first inequality in (62) holds. Moreover, using the definition of q^j in (45), the first inequality in (62), and the assumption that $\|\tilde{q}^{k-1}\| \leq \kappa_p$, we have

$$c\|Ay^j - b\| \overset{(45)}{=} \|q^j - \tilde{q}^{k-1}\| \leq \|q^j\| + \|\tilde{q}^{k-1}\| \overset{(62)}{\leq} 2\kappa_p,$$

and hence the second inequality in (62) also holds. ∎

The next result shows that every epoch of S-ADMM ends.

Lemma 4.10 *Every epoch $\mathcal{I}_k$ generated by S-ADMM terminates and the following relations hold:*

$$\mathcal{L}_c(\tilde{y}^k; \tilde{q}^{k-1}) \leq \mathcal{L}_c(y^i; \tilde{q}^{k-1}) \quad \forall i \in \mathcal{I}_k, \tag{64}$$

$$\mathcal{L}_c(\tilde{y}^k; \tilde{q}^{k-1}) + \frac{\|\tilde{q}^k - \tilde{q}^{k-1}\|^2}{c} = \mathcal{L}_c(\tilde{y}^k; \tilde{q}^k). \tag{65}$$

Proof: We first show that every epoch $\mathcal{I}_k$ generated by S-ADMM terminates. Assume for the sake of contradiction that $\mathcal{I}_k$ is infinite. Lemma 4.9(b) then implies that $\mathcal{I}_k(C)$ is also infinite. This in turn implies that $i < (\alpha k T_i)/\rho^2$ for every $i \in \mathcal{I}_k(C)$ since otherwise the k-epoch would terminate at the first iteration $i \in \mathcal{I}_k(C)$ that violates this condition (see line 15 of S-ADMM). This conclusion together with (59) then imply that any $i \in \mathcal{I}_k(C)$ satisfies

$$i < \frac{\alpha k}{\rho^2} T_i \overset{(59)}{\leq} \frac{\alpha k}{\rho^2}\left(\tilde{T}_{k-1} + \Gamma_{k-1}\right),$$

and hence that $\mathcal{I}_k$ is finite as the right-hand side of the above inequality is independent of $i \in \mathcal{I}_k$. Since this contradicts our previous assumption that $\mathcal{I}_k$ is infinite, we conclude that $\mathcal{I}_k$ terminates.

We now prove (64). First, observe that, by Lemma 4.1(a), we have $\tilde{y}^k = y^{i_k^+}$. Hence, it follows from Lemma 4.1(b) with $i = i_k^+$ and $j = i$ that $\mathcal{L}_c(y^i; \tilde{q}^{k-1}) - \mathcal{L}_c(\tilde{y}^k; \tilde{q}^{k-1}) = \tilde{T}_k - T_i$. Inequality (64) now follows from the last conclusion of Lemma 4.6. Finally, the identity in (65) follows from the definition of the augmented Lagrangian function in (3) (with $(\theta, \chi) = (0, 1)$) and the fact that $\tilde{q}^k = \tilde{q}^{k-1} + c(A\tilde{y}^k - b)$, due to line 16 of S-ADMM. ∎

The next result provides some preliminary bounds on the Lagrangian multiplier sequence $\{\tilde{q}^k\}$, the feasibility residual sequence $\{\|A\tilde{y}^k - b\|\}$, and the sequence $\{T_i\}$.

Lemma 4.11 *If the final iterate of the $(k-1)$-th epoch satisfies the conditions that $\tilde{q}^{k-1} \in A(\mathbb{R}^n)$ and $\|\tilde{q}^{k-1}\| \leq \kappa_p$, then the following statements hold:*

(a) the pair $(\tilde{y}^k, \tilde{q}^k)$ satisfies

$$\tilde{q}^k \in A(\mathbb{R}^n), \quad \|\tilde{q}^k\| \leq \kappa_p \quad \text{and} \quad c\|A\tilde{y}^k - b\| \leq 2\kappa_p; \tag{66}$$

(b) for every $i \in \mathcal{I}_k(C)$, there holds

$$T_i \leq \tilde{T}_{k-1} + \frac{4\kappa_p^2}{c} + \Delta_k, \tag{67}$$

where

$$\Delta_k := \mathcal{L}_c(\tilde{y}^{k-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(\tilde{y}^k; \tilde{q}^k); \tag{68}$$

(c) there holds

$$\tilde{T}_k - \tilde{T}_{k-1} \leq \frac{4\kappa_p^2}{c} + \Delta_k, \tag{69}$$

where $\tilde{T}_k$ is as in Lemma 4.1(a).

Proof: Assume that the conditions $\tilde{q}^{k-1} \in A(\mathbb{R}^n)$ and $\|\tilde{q}^{k-1}\| \leq \kappa_p$ hold.

(a) The conclusion that the pair $(\tilde{y}^k, \tilde{q}^k)$ satisfies (66) follows from (62) with $j = i_k^+$ and the fact that $(\tilde{y}^k, \tilde{q}^k) = (y^{i_k^+}, q^{i_k^+})$.

(b) Using the definition of T_i as in line 14 of S-ADMM, inequality (64), and the identity in (65), we have

$$\begin{aligned} T_i - \tilde{T}_{k-1} &= \mathcal{L}_c(\tilde{y}^{k-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(y^i; \tilde{q}^{k-1}) \overset{(64)}{\leq} \mathcal{L}_c(\tilde{y}^{k-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(\tilde{y}^k; \tilde{q}^{k-1}) \\ &\overset{(65)}{=} \mathcal{L}_c(\tilde{y}^{k-1}; \tilde{q}^{k-1}) - \mathcal{L}_c(\tilde{y}^k; \tilde{q}^k) + \frac{\|\tilde{q}^k - \tilde{q}^{k-1}\|^2}{c}. \end{aligned}$$

Inequality (67) now follows from the above identity, the definition of Δ_k in (68), the triangle inequality for norms, and the first inequality in (66).

(c) Inequality (69) follows from (67) with $i = i_k^+$ and the fact that $\tilde{T}_k = T_{i_k^+}$ (see Lemma 4.1(a)). ∎

The main purpose of the next result is to provide a preliminary bound on the cardinality of the k-th epoch.

Lemma 4.12 *Assume that the initial Lagrange multiplier satisfies $p \in A(\mathbb{R}^n)$ and $\|p\| \le \kappa_p$, and define*

$$\Gamma(x;c) := \overline{\phi} - \underline{\phi} + \frac{(\|Ax-b\| + 3\kappa_p)^2}{2c}, \tag{70}$$

where (x,c) is part of the input for S-ADMM*. Then, the following statements about any epoch $\mathcal{I}_k$ generated* S-ADMM *hold:*

(a) $\Gamma_{k-1} \le \Gamma(x;c)$, *where Γ_{k-1} is as in* (60)*;*

(b) there hold

$$\sum_{l=1}^{k} \Delta_l \le \Gamma(x;c), \quad \sum_{l=1}^{k} l\Delta_l \le k\Gamma(x;c) \tag{71}$$

where Δ_l is as in (68)*; moreover,*

$$\tilde{T}_k \le \frac{4k\kappa_c^2}{c} + \Gamma(x;c); \tag{72}$$

(c) its length satisfies

$$|\mathcal{I}_k| \le \frac{\alpha}{\rho^2}\left(\frac{8k\kappa_p^2}{c} + \Gamma(x;c) + k\Delta_k\right) + 1 + \frac{(\chi(\gamma) + (1+\sigma_2)cB\|A\|_\dagger^2)\Gamma(x;c)}{C^2}, \tag{73}$$

where $\chi(\gamma)$ is as in (49) *and σ_2 is part of the input for* S-ADMM*.*

Proof: (a) Recalling the definitions of Γ_{k-1} and $\Gamma(x;c)$ in (60) and (70), respectively, the inequality $\Gamma_{k-1} \le \Gamma(x;c)$ for $k=1$ follows from the facts that $\tilde{y}^0 = x$ and $\tilde{q}^0 = p$ (see line 1 of S-ADMM) and the assumption $\|p\| \le \kappa_p$, and for $k \ge 2$ follows from both inequalities in (66).

(b) We first claim that $\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) - \mathcal{L}_c(\tilde{y}^k;\tilde{q}^k) \le \Gamma(x;c)$ for every $l \in \{1,\ldots,k\}$. Using the fact that (66) implies that $\|\tilde{q}^l\| \le \kappa_p$ and $c\|A\tilde{y}^l - b\| \le 2\kappa_p$ for every $l \in \{1,\ldots,k\}$, Lemma 4.8 with $(u^1,u^2) = (\tilde{q}^{l-1},\tilde{q}^k)$, $(z^1,z^2) = (\tilde{y}^{l-1},\tilde{y}^k)$, the assumption that $\|p\| \le \kappa_p$, we conclude that

$$\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) - \mathcal{L}_c(\tilde{y}^k;\tilde{q}^k) \le \overline{\phi} - \underline{\phi} + \frac{(c\|A\tilde{y}^{l-1} - b\| + \kappa_p)^2}{2c} \le \Gamma(x;c)$$

for every $l \in \{1,\ldots,k\}$, where the last inequality above follows by repeating the same argument as in statement (a). Thus, we conclude that the claim holds. We now prove the first inequality in (71). Using the definition of Δ_l in (68), we have

$$\sum_{l=1}^{k} \Delta_l = \sum_{l=1}^{k} \left(\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) - \mathcal{L}_c(\tilde{y}^l;\tilde{q}^l)\right) = \mathcal{L}_c(\tilde{y}^0;\tilde{q}^0) - \mathcal{L}_c(\tilde{y}^k;\tilde{q}^k) \le \Gamma(x;c),$$

where the last inequality above follows from the previous claim with $l=1$ and the fact that $(\tilde{y}^0,\tilde{q}^0) = (x,p)$ (see line 1 of S-ADMM). We now prove the second inequality in (71). Using the identity in (68) and simple algebraic manipulations, we have

$$\begin{aligned}
\sum_{l=1}^{k} l\Delta_l &\overset{(68)}{=} \sum_{l=1}^{k} l[\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) - \mathcal{L}_c(\tilde{y}^l;\tilde{q}^l)] \\
&= \sum_{l=1}^{k} \mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) + \sum_{l=1}^{k}[(l-1)\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) - l\mathcal{L}_c(\tilde{y}^l;\tilde{q}^l)] \\
&= \sum_{l=1}^{k}(\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1})) - k\mathcal{L}_c(\tilde{y}^k;\tilde{q}^k) = \sum_{l=1}^{k}\left(\mathcal{L}_c(\tilde{y}^{l-1};\tilde{q}^{l-1}) - \mathcal{L}_c(\tilde{y}^k;\tilde{q}^k)\right) \le k\Gamma(x;c),
\end{aligned}$$

where the last inequality above follows from the previous claim.

Inequality (72) follows by summing inequality in (69) from $k = 1$ to $k = k$, using the fact that $(\tilde{y}^0, \tilde{q}^0, \tilde{T}_0) = (x, p, 0)$ (see line 1 of S-ADMM), and the first inequality in (71).

(c) We now prove (73). First, we recall from Lemma 4.10 that $\mathcal{I}_k$ is finite. Let i be the largest index in $\mathcal{I}_k(C) \cup \{i^+_{k-1}\}$ satisfying $i < i^+_k$ (with the convention that $i^+_0 = 0$). Then, since i and i^+_k are consecutive indices in $\mathcal{I}_k(C)$, it follows from Lemma 4.9(b) and the fact that $\Gamma_{k-1} \le \Gamma(x; c)$ from statement (a), that

$$|\mathcal{I}_k| = i^+_k - i^+_{k-1} = (i - i^+_{k-1}) + (i^+_k - i) \overset{(61)}{\le} i - i^+_{k-1} + 1 + \frac{(\chi(\gamma) + (1+\sigma_2)cB\|A\|^2_\dagger)\Gamma(x; c)}{C^2}. \tag{74}$$

If $i = i^+_{k-1}$ then (74) implies that (73) holds. Suppose now that $i \in \mathcal{I}_k(C)$. Using the facts that $i \in \mathcal{I}_k(C) \setminus \{i^+_k\}$ and i^+_k is the only index in $\mathcal{I}_k(C)$ satisfying the second condition in line 15 of S-ADMM, we then conclude that

$$\begin{aligned}
\frac{\rho^2 i}{\alpha} &< kT_i \overset{(67)}{\le} k\left(\tilde{T}_{k-1} + \frac{4\kappa_p^2}{c} + \Delta_k\right) = (k-1)\tilde{T}_{k-1} + \tilde{T}_{k-1} + \frac{4k\kappa_p^2}{c} + k\Delta_k \\
&\overset{(34)}{\le} \frac{\rho^2 i^+_{k-1}}{\alpha} + \tilde{T}_{k-1} + \frac{4k\kappa_p^2}{c} + k\Delta_k \\
&\overset{(72)}{\le} \frac{\rho^2 i^+_{k-1}}{\alpha} + \left[\frac{4(k-1)\kappa_p^2}{c} + \Gamma(x; c)\right] + \frac{4k\kappa_p^2}{c} + k\Delta_k \\
&\le \frac{\rho^2 i^+_{k-1}}{\alpha} + \frac{8k\kappa_p^2}{c} + \Gamma(x; c) + k\Delta_k,
\end{aligned}$$

where the second last inequality is due to the second inequality in (34) with $k = k - 1$, and the third inequality is due to (72) with $k = k - 1$. Inequality (73) follows by combining (74) and the last inequality. ■

4.3 Step 3: Proof of Theorem 3.1

In this subsection, we marshal all of the results of the previous subsections to finally prove S-ADMM's main iteration-complexity result (Theorem 3.1).

We first show that the assumptions (24) and (25) implies that

$$\chi(\gamma) \le \bar{\chi}, \quad \Gamma(x; c) \le \bar{\Gamma}, \tag{75}$$

where $\chi(\gamma)$ and $(\bar{\chi}, \bar{\Gamma})$ are as in (49) and (22), respectively. Indeed, the lower bound (25) on γ_t implies that $\gamma^{-1} \le \max\{2m, \bar{\gamma}^{-1}\} \le 2m + \bar{\gamma}^{-1}$. This inequality, the second inequality in (25), and the definition of $\chi(\gamma)$ in (49), then imply that $\chi(\gamma) \le \bar{\chi}$. The second inequality in (75) follows immediately from the first and third inequality in (24) combined with (22) and (70).

(a) By Proposition 4.7, S-ADMM performs a finite number K of epochs that is bounded by $\bar{K}$ as in (55). Since $|\mathcal{I}_k|$ is the number of iterations performed during the k-epoch, it suffices to show that $\sum_{k=1}^K |\mathcal{I}_k| \le \Lambda(c, \rho)$ where $\Lambda(c, \rho)$ is as in (26). Using the bound in (73), second inequality in (75), and the second inequality in (71), we have

$$\begin{aligned}
\sum_{k=1}^K |\mathcal{I}_k| &\overset{(73)}{\le} \sum_{k=1}^K \left[\frac{8k\alpha\kappa_p^2}{\rho^2 c} + \frac{\alpha\Gamma(x; c)}{\rho^2} + \frac{\alpha k\Delta_k}{\rho^2} + 1 + \frac{(\chi(\gamma) + (1+\sigma_2)cB\|A\|^2_\dagger)\Gamma(x; c)}{C^2}\right] \\
&\overset{(71)}{\le} \frac{K(K+1)}{2} \cdot \frac{8\alpha\kappa_p^2}{c\rho^2} + K\left[1 + \left(\frac{2\alpha}{\rho^2} + \frac{(\chi(\gamma) + (1+\sigma_2)cB\|A\|^2_\dagger)}{C^2}\right)\Gamma(x; c)\right] \\
&\overset{(75)}{\le} \frac{K(K+1)}{2} \cdot \frac{8\alpha\kappa_p^2}{c\rho^2} + K\left[1 + \left(\frac{2\alpha}{\rho^2} + \frac{(\chi(\gamma) + (1+\sigma_2)cB\|A\|^2_\dagger)}{C^2}\right)\bar{\Gamma}\right].
\end{aligned}$$

After some algebraic manipulations, we can easily see that the previous inequality, the fact that $K \le \bar{K} \le [\bar{\chi} + (1+\sigma_2)cB\|A\|^2_\dagger]/\alpha + 1$ (see Proposition 4.7), and the first inequality in (75), imply that

$$\sum_{k=1}^K |\mathcal{I}_k| \le \Lambda_0 \frac{c}{\rho^2} + \Lambda_1 + \Lambda_2 c + \Lambda_3 c^2 + \Lambda_4 \frac{1}{\rho^2} = \Lambda(c, \rho),$$

where Λ_0, Λ_1, Λ_2, Λ_3, and Λ_4 are as in (23).

(b) Since K is the last epoch generated by S-ADMM and the inclusion in (46) holds for every $i \in \mathcal{I}_K$, we conclude that $(y^{i_K^+}, q^{i_K^+}, v^{i_K^+}, \delta_{i_K^+}) = (x^+, p^+, v^+, \varepsilon^+)$ satisfies (46), and hence that (27) holds. The two first inequalities in (28) follows from Lemma 4.11(a), due to the fact that $(y^{i_K^+}, q^{i_K^+}) = (\tilde{y}^K, \tilde{q}^K)$. Finally, the last inequality in (28) follows from the fact that S-ADMM terminates on line 18 with the condition $\|v^+\|^2 + \varepsilon^+ = \|v^{i_K^+}\|^2 + \delta_{i_K^+} \le \rho^2$ satisfied. We now prove the inequalities in (29). First, observe that $\gamma^+ = \lambda^{i_K^+}$ (see lines 16 and 19 of S-ADMM). Second, using (25) and (44), we have

$$\bar{\gamma}_t \overset{(25)}{\ge} \gamma_t \overset{(44)}{\ge} \lambda_t^{i_K^+} \overset{(44)}{\ge} \min\left\{\frac{1}{2m_t}, \gamma_t\right\} \overset{(25)}{\ge} \min\left\{\frac{1}{2m_t}, \bar{\gamma}_t\right\},$$

for every $t \in \{1, \ldots, B\}$. Combining these two observations we conclude that (29) holds.

(c) Using the assumption that $c \ge 2\kappa_p/\eta$, Lemma 4.11(a) guarantees that S-ADMM output $y^{i_K^+} = x^+$ satisfying

$$\|Ax^+ - b\| \le \frac{2\kappa_p}{c} \le \eta.$$

Hence, the conclusion that $(x^+, p^+, v^+, \varepsilon^+) = (\tilde{y}^K, \tilde{q}^K, \tilde{v}^K, \tilde{\varepsilon}_K)$ satisfies (2) follows from the previous inequality, the inclusion in (27), and the last inequality in (28).

5 Numerical Experiments

This section showcases the numerical performance of A-ADMM on two linearly and box constrained, nonconvex, quadratic programming problems. Subsection 5.1 summarizes the performance of A-ADMM on a distributed variant of our experimental problem, while Subsection 5.2 focuses on a non-distributable variant. The distributed variant employs a small number of high-dimensional blocks while the non-distributable variant conversely has a large number of uni-dimensional blocks. These two proof-of-concept experiments indicate that A-ADMM may not only substantially outperform the relevant benchmarking methods in practice, but also be relatively robust to the relationship between block counts and sizes.

All experiments were implemented and executed in MATLAB 2021b and run on a macOS machine with a 1.7 GHz Quad-Core Intel processor, and 8 GB of memory.

5.1 Distributed Quadratic Programming Problem

This subsection studies the performance of A-ADMM for finding stationary points of a box-constrained, nonconvex block distributed quadratic programming problem with B blocks (DQP).

The B-block DQP is formulated as

$$\begin{aligned} \min_{(x_1,\ldots,x_B)\in\mathbb{R}^{Bn}} \quad & -\sum_{i=1}^{B-1}\left[\frac{\alpha_i}{2}\|x_i\|^2 + \langle x_i, \beta_i\rangle\right] \\ \text{s.t.} \quad & \|x\|_\infty \le \omega && (76) \\ & x_i - x_B = 0 \quad \text{for } i = 1, \ldots, B-1 && (77) \end{aligned}$$

where $\omega > 0$, $n \in \mathbb{N}$, $\{\alpha_i\}_{i=1}^{B-1} \subseteq [0,1]$, and $\{\beta_i\}_{i=1}^{B-1} \subseteq [0,1]^n$. It is not difficult to see that DQP fits the template of (1). The smooth component is taken to be

$$f(x) = -\sum_{i=1}^{B-1}\left[\frac{\alpha_i}{2}\|x_i\|^2 + \langle x_i, \beta_i\rangle\right].$$

The non-smooth function h_i is set to the indicator of the set $\{x \in \mathbb{R}^n : \|x_i\|_\infty \le \omega\}$ for $1 \le i \le B$. For $1 \le i \le B-1$, we take $A_i \in \mathbb{R}^{n\times Bn}$ to be the operator which includes $A_i \in \mathbb{R}^n$ into the i-th block of $\mathbb{R}^{Bn}$, i.e.

$$A_i = \begin{bmatrix} 0_{(i-1)n\times n} \\ I_{n\times n} \\ 0_{(B-i)n\times n} \end{bmatrix}$$

where $0_{j\times k}$ denotes a $j \times k$ zero matrix. The matrix $A_B \in \mathbb{R}^{n\times Bn}$ is defined by the action $A_B x = (-x, \ldots, -x)^\top$.

We shall now outline how we conducted our DQP experiments. The number of blocks, B, for each experiment was set to 3, while for the block-size, n, the dimensions $n = 10, 20, 100, 500$ (5000)??? were considered. For each setting of n, we ran experiments where $\omega = 10^1, 10^3, 10^5, 10^7, 10^9$. The values of $\{\alpha_i\}_{i=1}^{B-1} \subseteq [0,1]$, and $\{\beta_i\}_{i=1}^{B-1} \subseteq [0,1]^n$ were sampled uniformly at random. To generate b, we sampled $x_b \in [-\omega,\omega]^{Bn}$ uniformly at random, then set $b = Ax_b$. The initial iterate x_0 was also selected uniformly at random from $[-\omega,\omega]^{Bn}$.

For this problem, A-ADMM was applied with $c_0 = 1$, $C = 1$, $\alpha = 10^{-2}$, $p^0 = \mathbf{0}$, and each block's initial stepsize set to 10, i.e. $\gamma_i^0 = 10$ for $1 \le i \le B$. To provide an adequate benchmark for A-ADMM, we compared its performance against two instances of the method from [25] and three instances of the method from [43]. The method of [25] was deployed with two different choices of (θ,χ): $(0,1)$ and $(1/2, 1/18)$. We call these two instances DP1 and DP2, respectively. Both DP1 and DP2 set $(\lambda, c_1) = (1/2, 1)$. The method of [43] was deployed with three different settings of the penalty parameter ρ: 0.1, 1.0, and 10.0. We call the resultant instances SD1, SD2, and SD3, respectively. Moreover, all three instances make the parameter selections $(\omega,\theta,\tau) = (4,2,1)$ and $(M_h, K_k, J_h, L_h) = (4\gamma, 1, 1, 0)$ in accordance with [43, Section 5.1]. We reiterate that [25] provides no convergence guarantees for the pragmatic choice of $(\theta,\chi) = (0,1)$. All executed algorithms were run for a maximum of 500,000 iterations. Any algorithm that met this limit took at least 10 milliseconds to complete.

		Iteration						Time (ms)					
n	ω	AD	DP1	DP2	SD1	SD2	SD3	AD	DP1	DP2	SD1	SD2	SD3
10	10^1	**18**	76	83	427	223	976	**1.592**	5.402	4.291	28.192	13.881	60.184
10	10^3	**34**	228	232	569	399	1855	**2.259**	11.752	11.858	65.049	30.209	119.417
10	10^5	**50**	385	385	*	581	2778	**3.228**	13.374	13.004	*	35.368	168.964
10	10^7	**66**	541	537	*	*	3701	**4.419**	18.706	18.363	*	*	239.235
10	10^9	**81**	697	689	*	*	4625	**5.866**	24.540	24.237	*	*	323.790
20	10^1	**22**	62	68	433	298	1261	**1.538**	2.520	2.484	27.928	19.164	80.375
20	10^3	**44**	166	171	*	498	2304	**2.560**	6.484	6.153	*	31.821	152.866
20	10^5	**65**	273	275	*	700	3347	**4.213**	10.548	9.879	*	45.812	223.235
20	10^7	**84**	379	379	*	*	4383	**4.684**	13.961	14.009	*	*	290.884
20	10^9	**103**	485	483	*	*	5418	**5.629**	17.635	17.393	*	*	365.368
100	10^1	**20**	40	46	*	433	6231	2.072	**1.820**	1.831	*	28.705	420.276
100	10^3	**33**	78	77	*	695	9444	**2.132**	3.013	2.871	*	44.640	617.636
100	10^5	**45**	116	107	*	*	12664	**3.148**	4.545	4.041	*	*	898.522
100	10^7	**57**	155	137	*	*	15876	**3.736**	5.844	5.465	*	*	1051.830
100	10^9	**68**	193	167	*	*	19087	**4.841**	7.629	6.436	*	*	1267.924
5000	10^1	**25**	121	125	*	646	2257	**13.733**	26.455	27.511	*	206.646	861.279
5000	10^3	**37**	221	223	*	851	3324	**20.084**	52.456	50.828	*	282.999	1264.225
5000	10^5	**49**	321	321	*	*	4390	**27.591**	72.829	72.810	*	*	1692.375
5000	10^7	**61**	422	419	*	*	5449	**32.080**	96.010	97.450	*	*	1968.163
5000	10^9	**72**	522	517	*	*	6507	**41.682**	118.872	118.632	*	*	2440.377

Bolded values equal to the best algorithm according to iteration count or time.

** indicates the algorithm failed to find a stationary point meeting the tolerances by the 500,000th iteration.*

Table 1: Performance for all algorithms applied to the DQP problem (76), with $B = 3$, $C = 1$ and $\alpha = 10^{-2}$ for different pair of values (n,ω). The iteration and time columns record the number of iterations and time in seconds to find a $(10^{-5}, 10^{-5})$-stationary point.

Table 1, the record of the performance of all algorithms on this experimental problem, lays bare the superior performance of A-ADMM. In this table, we label A-ADMM as AD for the sake of concision. In terms of iterations, A-ADMM outperforms all other algorithms for all settings of B and m. Along the dimension of time, A-ADMM is faster than all algorithms, for all settings of n and ω, except DP1 when $n = 100$ and $\omega = 10$.

5.2 Nonconvex QP with Box Constraints

In this subsection, we evaluate the performance of A-ADMM for solving a general nonconvex quadratic problem with box constraints (QP-BC). The QP-BC problem is formulated as

$$\min_{\|x\|_\infty \le \omega} \left\{ f(x) := \frac{1}{2}\langle x, Px\rangle + \langle r, x\rangle : Ax = b \right\}. \tag{78}$$

where $P \in \mathbb{R}^{B\times B}$ is negative definite, $A \in \mathbb{R}^{m\times B}$, $r, b \in \mathbb{R}^m \times \mathbb{R}^m$ and $\omega \in \mathbb{R}_{++}$. As for the previous problem, it is not difficult to check that QP-BC fits within the template of (1). For this problem, we take our blocks to be single coordinates. Consequently, each column of A gives rise to a A_i matrix. The non-smooth components of the objective are again picked to be the indicator functions of the sets $\{x_i \in \mathbb{R} : |x_i| \le \omega\}$ for $i \in \{1, \ldots, B\}$.

We now describe how we orchestrated our QP-BC experiments. In all instances, $\omega = 1$. To generate $\tilde{r} \in \mathbb{R}^m$, $\tilde{P} \in \mathbb{R}^{B\times B}$ and $\tilde{A} \in \mathbb{R}^{m\times B}$, we started by generating a diagonal matrix $D \in \mathbb{R}^{B\times B}$ whose entries are selected uniformly at random in $[1, 1000]$. Next, we generated $\tilde{r} \in [-1, 1]^m$, $\tilde{P} \in [-1, 1]^{B\times B}$ negative definite, and $\tilde{A} \in [-1, 1]^{m\times B}$ uniformly at random. Finally, we set $P = D\tilde{P}D$, $A = \tilde{A}D$, and $r = D\tilde{r}$. The vector $b \in \mathbb{R}^m$ was set as $b = Ax_b$, where x_b is a uniformly at random selected vector in $[-1, 1]^B$. The initial starting point x^0 was chosen in this same fashion.

For this problem, three instances of A-ADMM, referred to as AD1, AD2, and AD3, were applied with $C = 1$, $\alpha = 10^{-2}$, $p^0 = \mathbf{0}$, and each block's initial stepsize set to 10, i.e. $\gamma_i^0 = 1000$ for $1 \le i \le B$. The three methods differ only in their choice of initial penalty parameter c_0: $c_0 = 10$ for AD1, $c_0 = 1$ for AD2, and $c_0 = .1$ for AD3. The benchmarking algorithms for this experiment were three instances of the method from [25], which we refer to as DP1, DP2 and DP3. Like the three instances of A-ADMM, these instances differ only in their choice of c_0: $c_0 = 10$ in DP1, $c_0 = 1$ in DP2 and $c_0 = 0.1$ in DP3. Each of these three methods were applied with $(\theta, \chi) = (0, 1)$. Yet again, we remind the reader that [25] provides no convergence guarantees for this choice of (θ, χ). To ensure timely execution of all algorithms, each algorithm terminated upon meeting a 500,000 iteration limit or the discovery of an approximate stationary triple (x^+, p^+, v^+) satisfying the relative error criterion

$$v^+ \in \nabla f(x^+) + \partial h(x^+) + A^* p^+, \quad \frac{\|v^+\|}{1 + \|\nabla f(x^0)\|} \le \rho, \quad \frac{\|Ax^+ - b\|}{1 + \|Ax^0 - b\|} \le \eta.$$

for $\rho = \eta = 10^{-5}$.

The results for this experiment, shown in Table 2, echo those for its predecessor by again displaying the computational superiority of A-ADMM. Measured by iteration count, A-ADMM performed better in 87% of the problem instances. In terms of time, A-ADMM performed better in 100% of the instances. It is worth mentioning that A-ADMM converged for all instances, while the same cannot be said for the DP1, DP2, and DP3 benchmark methods. DP1 converged for 96% instances, DP2 converged for 54% instances, and DP3 converged only for 27% instances. For $m = 1, 2, 5$, our method was at least 10 times faster in terms of iteration count and time than any DP variant. Notably, we attempted to apply multiple versions of the method from [25] with choices of (θ, χ) that theoretically should ensure convergence. None of the methods managed to find the desired point within the iteration limit, so we omitted their results from the table.

6 Concluding Remarks

We now discuss some further related research directions. First, the ability of A-ADMM to allow for the inexact solution of its block subproblem opens up many possible avenues for application. A systematic numerical study of its performance when applied to problems requiring inexact computation would be intriguing. Second, it would be interesting to develop a P-ADMM that performs Lagrange multiplier updates with $(\theta, \chi) = (0, 1)$ at every iteration, rather than just at the last iteration of each epoch. This P-ADMM would then be an instance of the class of ADMMs outlined in the Introduction with $|\mathcal{I}_k| = 1$. The P-ADMM of [25] satisfies this last property but chooses (θ, χ) in a very conservative way, namely, satisfying (6). Third, it would be interesting to consider inexact P-ADMMs that solve the prox subproblems based on the more relaxed error criterion (13) with $\vartheta > 0$ since this generalization would possibly allow us to develop an inexact

		Iteration						Time (sec)					
B	m	AD1	AD2	AD3	DP1	DP2	DP3	AD1	AD2	AD3	DP1	DP2	DP3
10	1	44	**23**	33	3554	3560	3532	0.067	0.010	**0.008**	0.296	0.286	0.275
10	2	23	**19**	37	1355	1282	1395	0.025	**0.005**	0.007	0.128	0.111	0.116
10	5	**1280**	2421	1469	*	*	*	**0.162**	0.287	0.171	*	*	*
20	1	**23**	30	30	803	296	417	0.021	0.009	**0.008**	0.148	0.052	0.075
20	2	87	**44**	89	297	2233	*	0.032	**0.014**	0.026	0.063	0.441	*
20	5	147	144	**114**	1862	6210	*	0.036	0.034	**0.027**	0.333	1.088	*
20	10	682	1105	**550**	847	*	*	0.168	0.267	**0.132**	0.152	*	*
20	15	**1286**	2753	3691	1808	*	*	**0.308**	0.656	0.879	0.329	*	*
50	1	21	**17**	66	327	1616	1385	0.022	**0.014**	0.049	0.176	0.850	0.746
50	2	219	**22**	66	377	1180	2772	0.178	**0.019**	0.055	0.243	0.760	1.704
50	5	188	**123**	226	1880	2296	*	0.147	**0.098**	0.175	1.208	1.467	*
50	10	462	**377**	1647	1456	699	*	0.352	**0.285**	1.229	0.972	0.466	*
50	20	**1082**	56530	9363	2058	*	*	**0.842**	42.686	7.063	2.173	*	*
50	25	1326	2361	2307	**1157**	*	*	**1.230**	1.913	1.835	1.243	*	*
50	30	3430	**1262**	2045	3981	*	*	2.989	**1.044**	1.654	4.412	*	*
100	1	95	**22**	182	2792	1476	9554	0.446	**0.084**	0.572	9.545	4.887	33.713
100	2	104	**32**	102	802	1531	*	0.429	**0.120**	0.361	3.154	5.996	*
100	5	449	256	**83**	4603	*	*	1.570	0.902	**0.295**	20.776	*	*
100	10	1675	37263	**427**	3050	3281	*	5.724	124.269	**1.429**	15.528	16.771	*
100	25	2388	12916	**2346**	2687	*	*	8.041	43.605	**7.885**	15.529	*	*
100	50	4596	3526	*	**3395**	*	*	16.336	**12.488**	*	23.325	*	*
100	75	7070	27964	123020	**4816**	*	*	**26.387**	104.134	459.301	38.100	*	*

Bolded values equal to the best algorithm according to iteration count or time.
** indicates the algorithm failed to find a stationary point meeting the tolerances by the 500,000th iteration.*

Table 2: Performance for all algorithms applied to the QP-BC Problem (78), with $C = 1$ and $\alpha = 10^{-2}$, for different pair of values (B, ω). The iteration and time columns record the number of iterations and time in seconds to find a stationary point satisfying the relative error condition with $(\rho, \eta) = (10^{-5}, 10^{-5})$.

P-ADMM in the setting of (1) with f being non-smooth. Finally, we have assumed in this paper that $\operatorname{dom} h$ is bounded (see assumption (A1)). It would be interesting to extend its analysis to the case where $\mathcal{H}$ is unbounded.

A Technical Results for Proof of Lagrange Multipliers

This appendix provides some technical results to show that under certain conditions the Lagrangian multiplier is bound.

The next two results, used in Lemma A.3, can be found in [16, Lemma B.3] and [28, Lemma 3.10], respectively.

Lemma A.1 *Let $A : \mathbb{R}^n \to \mathbb{R}^l$ be a nonzero linear operator. Then,*

$$\nu_A^+ \|u\| \leq \|A^* u\|, \quad \forall u \in A(\mathbb{R}^n).$$

Lemma A.2 *Let h be a function as in (A5). Then, for every $\delta \geq 0$, $z \in \mathcal{H}$, and $\xi \in \partial_\delta h(z)$, we have*

$$\|\xi\| \operatorname{dist}(u, \partial\mathcal{H}) \leq [\operatorname{dist}(u, \partial\mathcal{H}) + \|z - u\|] \, M_h + \langle \xi, z - u \rangle + \delta \quad \forall u \in \mathcal{H}$$

where $\partial\mathcal{H}$ denotes the boundary of $\mathcal{H}$.

Lemma A.3 *Assume that h is a function as in assumption (A5) and $A : \mathbb{R}^n \to \mathbb{R}^l$ is a nonzero linear operator satisfying assumption (A2). If $(q^-, \chi) \in \mathbb{R}^l \times (0, \infty)$ and $(z, q, r) \in \operatorname{dom} h \times A(\mathbb{R}^n) \times \mathbb{R}^n$ satisfy*

$$r \in \partial_\delta h(z) + A^* q \quad \textit{and} \quad q = q^- + \chi(Az - b), \tag{79}$$

then we have

$$\|q\| \le \max\left\{\|q^-\|, \frac{2D_h(M_h+\|r\|)+\delta}{\bar{d}\nu_A^+}\right\}, \tag{80}$$

where M_h, $\bar{d}>0$, and D_h are as in (A5), (A6), and (19), respectively, and ν_A^+ denotes the smallest positive singular value of A.

Proof: We first claim that

$$\bar{d}\nu_A^+\|q\| \le 2D_h\left(M_h+\|r\|\right) - \langle q, Az-b\rangle + \delta \tag{81}$$

holds. The assumption on (z,q,r) implies that $r - A^*q \in \partial_\delta h(z)$. Hence, using the Cauchy-Schwarz inequality, the definitions of $\bar{d}$ and $\bar{z}$ in (A6), and Lemma A.2 with $\xi = r - A^*q$, and $u = \bar{z}$, we have:

$$\bar{d}\|r-A^*q\| - \left[\bar{d}+\|z-\bar{z}\|\right]M_h \overset{(A.2)}{\le} \langle r-A^*q, z-\bar{z}\rangle \le \|r\|\|z-\bar{z}\| - \langle q, Az-b\rangle + \delta. \tag{82}$$

Now, using the above inequality, the triangle inequality, the definition of D_h in (19), and the facts that $\bar{d} \le D_h$ and $\|z-\bar{z}\| \le D_h$, we conclude that:

$$\bar{d}\|A^*q\| + \langle q, Az-b\rangle \overset{(82)}{\le} \left[\bar{d}+\|z-\bar{z}\|\right]M_h + \|r\|\left(D_h+\bar{d}\right) \le 2D_h\left(M_h+\|r\|\right) + \delta. \tag{83}$$

Noting the assumption that $q \in A(\mathbb{R}^n)$, inequality (81) now follows from the above inequality and Lemma A.1.

We now prove (80). Relation (79) implies that $\langle q, Az-b\rangle = \|q\|^2/\chi - \langle q^-, q\rangle/\chi$, and hence that

$$\bar{d}\nu_A^+\|q\| + \frac{\|q\|^2}{\chi} \le 2D_h(M_h+\|r\|) + \frac{\langle q^-, q\rangle}{\chi} \le 2D_h(M_h+\|r\|) + \frac{\|q\|}{\chi}\|q^-\| + \delta, \tag{84}$$

where the last inequality is due to the Cauchy-Schwarz inequality. Now, letting K denote the right hand side of (80) and using (84), we conclude that

$$\left(\bar{d}\nu_A^+ + \frac{\|q\|}{\chi}\right)\|q\| \overset{(84)}{\le} \left(\frac{2D_h(M_h+\|r\|)+\delta}{K} + \frac{\|q\|}{\chi}\right)K \le \left(\bar{d}\nu_A^+ + \frac{\|q\|}{\chi}\right)K, \tag{85}$$

and hence that (80) holds. ∎

We conclude this section with a technical result of convexity which is used in the proof of Lemma 4.3. Its proof can be found in [36, Lemma A1].

Lemma A.4 *Assume that $\xi > 0$, ψ is a proper, closed, and convex function, and Q is a $n \times n$ positive definite matrix such that $\psi - (\xi/2)\|\cdot\|_Q^2$ is convex and let $(y, v, \hat{\eta}) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ be such that $y \in \partial_{\hat{\eta}}\psi(y)$. Then, for any $\tau > 0$,*

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

B Proof of Proposition 2.2

This appendix contains the proof of Proposition 2.2.

Before giving the proof of Proposition 2.2, we present a preliminary result related to S-CGM.

Lemma B.1 *For every $j \ge 1$, the vector $\tilde{v}^j \in \mathbb{R}^n$ defined in (16) satisfies*

$$\tilde{v}^j \in \nabla\psi_s(\tilde{z}^j) + \partial\psi_n(\tilde{z}^j) \quad \text{and} \quad 8M[\psi(\tilde{z}^{j-1}) - \psi(\tilde{z}^j)] \ge \|\tilde{v}^j\|^2. \tag{86}$$

Proof: The optimality condition for (15) implies that

$$0 \in \nabla\psi_s(\tilde{z}^{j-1}) + \partial\psi_n(\tilde{z}^j) + M(\tilde{z}^j - \tilde{z}^{j-1}).$$

The inclusion (86) now follows from the above inclusion and the definition of $\tilde{v}^j$ in (16). We now prove the inequality in (86). Using the definition of $\tilde{v}^j$ in (16), the triangle inequality, and assumption (B2), we have

$$\left\|\tilde{v}^j\right\| \leq M \left\|\tilde{z}^{j-1} - \tilde{z}^j\right\| + \left\|\nabla\psi_s(\tilde{z}^j) - \nabla\psi_s(\tilde{z}^{j-1})\right\| \leq 2M\|\tilde{z}^{j-1} - \tilde{z}^j\|. \tag{87}$$

On the other hand, using the fact that the objective function in (15) is M-strongly convex and that $\tilde{z}^j$ is its optimal solution, we have

$$\ell_{\psi_s}(w, \tilde{z}^{j-1}) + \psi_n(w) + \frac{M}{2}\|w - \tilde{z}^{j-1}\|^2 - \frac{M}{2}\|w - \tilde{z}^j\|^2 \geq \ell_{\psi_s}(\tilde{z}^j, \tilde{z}^{j-1}) + \psi_n(\tilde{z}^j) + \frac{M}{2}\|\tilde{z}^j - \tilde{z}^{j-1}\|^2,$$

for every $w \in \mathbb{R}^n$. Thus, using the previous inequality with $w = \tilde{z}^{j-1}$ and the facts that, for every $\tilde{x} \in \mathbb{R}^n$, we have $\ell_{\psi_s}(\tilde{x}; \tilde{x}) = \psi_s(\tilde{x})$ and $\ell_{\psi_s}(\tilde{z}^j, \tilde{z}^{j-1}) \geq \psi_s(\tilde{z}^j) - (M/2)\|\tilde{z}^j - \tilde{z}^{j-1}\|^2$, which is a consequence of assumption (B2), we obtain that

$$\psi(\tilde{z}^{j-1}) - \psi(\tilde{z}^j) \geq \frac{M}{2}\|\tilde{z}^j - \tilde{z}^{j-1}\|^2. \tag{88}$$

The conclusion now follows by combining (87) and (88). ∎

Proof of Proposition 2.2

Assume for the sake of contradiction that

$$\|\tilde{v}^j\|^2 > \sigma[\psi(\tilde{z}^0) - \psi(\tilde{z}^j)] + \vartheta^2 \quad \forall j \in \{1, \ldots, \bar{K}\}. \tag{89}$$

Inequality (86) and the previous inequality, both with $j = 1$, imply that

$$\psi(\tilde{z}^0) - \psi(\tilde{z}^1) \overset{(86)}{\geq} \frac{\|\tilde{v}^1\|^2}{8M} \overset{(89)}{>} \frac{\vartheta^2 + \|\tilde{v}^1\|^2}{16M}. \tag{90}$$

Combining inequalities (86) and (89), we have that for $j \in \{1, \ldots, \bar{K} - 1\}$,

$$8M[\psi(\tilde{z}^j) - \psi(\tilde{z}^{j+1})] \overset{(86)}{\geq} \|\tilde{v}^{j+1}\|^2 \overset{(89)}{>} \sigma[\psi(\tilde{z}^0) - \psi(\tilde{z}^{j+1})]. \tag{91}$$

Hence, for every $j \in \{1, \ldots, \bar{K} - 1\}$, we have

$$\begin{aligned} 8M[\psi(\tilde{z}^0) - \psi(\tilde{z}^{j+1})] &= 8M[\psi(\tilde{z}^0) - \psi(\tilde{z}^j)] + 8M[\psi(\tilde{z}^j) - \psi(\tilde{z}^{j+1})] \\ &\overset{(91)}{>} 8M[\psi(\tilde{z}^0) - \psi(\tilde{z}^j)] + \sigma[\psi(\tilde{z}^0) - \psi(z^{j+1})] \\ &\overset{(88)}{\geq} 8M[\psi(\tilde{z}^0) - \psi(\tilde{z}^j)] + \sigma[\psi(\tilde{z}^0) - \psi(\tilde{z}^j)] = (8M + \sigma)\,[\psi(\tilde{z}^0) - \psi(z^j)]. \end{aligned}$$

As a consequence, the previous inequality and (90) imply that

$$\begin{aligned} \psi(\tilde{z}^0) - \psi^* \geq \psi(\tilde{z}^0) - \psi(z^{\bar{K}}) &> \left(1 + \frac{\sigma}{8M}\right)^{\bar{K}-1} (\psi(\tilde{z}^0) - \psi(\tilde{z}^1)) \\ &> \left(1 + \frac{\sigma}{8M}\right)^{\bar{K}-1} \frac{\vartheta^2 + \|\tilde{v}^1\|^2}{16M}, \end{aligned}$$

and hence that

$$\ln\left(\frac{16M[\psi(\tilde{z}^0) - \psi^*]}{\vartheta^2 + \|\tilde{v}^1\|^2}\right) > (\bar{K} - 1)\ln\left(1 + \frac{\sigma}{8M}\right) > (\bar{K} - 1)\frac{\sigma/(8M)}{1 + \sigma/(8M)} = \frac{\sigma(\bar{K} - 1)}{8M + \sigma},$$

where in the last inequality above, we used the fact that $\ln(1 + x) > x/(1 + x)$ for every $x \in (-1, \infty)$. Thus,

$$\bar{K} - 1 < \frac{8M + \sigma}{\sigma} \ln\left(\frac{16M[\psi(\tilde{z}^0) - \psi^*]}{\vartheta^2 + \|\tilde{v}^1\|^2}\right).$$

Since the last inequality contradicts with the definition of $\bar{K}$ in (17), the conclusion of the proposition follows.

References

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

[2] N. Aybat and G. Iyengar. A first-order augmented Lagrangian method for compressed sensing. *SIAM J. Optim.*, 22(2):429–459, 2012. doi: 10.1137/100786721.

[3] A. Beck. *First-order methods in optimization.* SIAM, 2017.

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

[5] E. Birgin, G. Haeser, and J. M. Martínez. Safeguarded augmented Lagrangian algorithms with scaled stopping criterion for the subproblems. *Computational Optimization and Applications*, pages 1–19, 04 2024. doi: 10.1007/s10589-024-00572-w.

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

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

[8] J. Eckstein and D. P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318, 1992.

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

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

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

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

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

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

[15] R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.

[16] M. L. N. Goncalves, J. G. Melo, and R. D. C. Monteiro. Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems. *Pacific Journal of Optimization*, 15:379–398, 2019.

[17] D. Hajinezhad and M. Hong. Perturbed proximal primal–dual algorithm for nonconvex nonsmooth optimization. *Math. Program.*, 176:207–245, 2019.

[18] Y. He and R. D. C. Monteiro. Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player Nash equilibrium problems. *SIAM J. Optim.*, 25:2182–2211, 2015.

[19] J. B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis and Minimization Algorithms II. Advanced Theory and Bundle Methods.* Springer, Berlin, 1993.

[20] M. Hong, Z.-Q. Luo, and M. Razaviyayn. Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. *SIAM Journal on Optimization*, 26(1):337–364, 2016.

[21] A. Izmailov, M. Solodov, and E. Uskov. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM Journal on Optimization*, 22:1579–1606, 02 2012. doi: 10.1137/120868359.

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

[23] B. Jiang, T. Lin, S. Ma, and S. Zhang. Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis. *Computational Optimization and Applications*, 72(1):115–157, 2019.

[24] W. Kong and R. D. C. Monteiro. An accelerated inexact dampened augmented Lagrangian method for linearly-constrained nonconvex composite optimization problems. *Comput. Optim. Appl.*, 2023.

[25] W. Kong and R. D. C. Monteiro. Global complexity bound of a proximal admm for linearly constrained nonseparable nonconvex composite programming. *SIAM Journal on Optimization*, 34(1):201–224, 2024. doi: 10.1137/22M1503129.

[26] W. Kong, J. Melo, and R. D. C. Monteiro. Iteration-complexity of a proximal augmented Lagrangian method for solving nonconvex composite optimization problems with nonlinear convex constraints. *Mathematics of Operations Research*, 2023.

[27] W. Kong, J. Melo, and R. D. C. Monteiro. Iteration complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical Lagrangian function. *SIAM Journal on Optimization*, 33(1):181–210, 2023. doi: 10.1137/20M136147X. URL https://doi.org/10.1137/20M136147X.

[28] W. Kong, J. G. Melo, and R. D. C. Monteiro. Iteration complexity of an inner accelerated inexact proximal augmented Lagrangian method based on the classical Lagrangian function. *SIAM Journal on Optimization*, 33(1):181–210, 2023.

[29] G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order penalty methods for convex programming. *Math. Program.*, 138(1):115–139, Apr 2013.

[30] G. Lan and R. D. C. Monteiro. Iteration-complexity of first-order augmented Lagrangian methods for convex programming. *Math. Program.*, 155(1):511–547, Jan 2016.

[31] Y. Liu, X. Liu, and S. Ma. On the nonergodic convergence rate of an inexact augmented Lagrangian framework for composite convex programming. *Math. Oper. Res.*, 44(2):632–650, 2019.

[32] Z. Lu and Z. Zhou. Iteration-complexity of first-order augmented Lagrangian methods for convex conic programming. *SIAM journal on optimization*, 33(2):1159–1190, 2023.

[33] J. Melo, R. D. C. Monteiro, and H. Wang. Iteration-complexity of an inexact proximal accelerated augmented Lagrangian method for solving linearly constrained smooth nonconvex composite optimization problems. *Available on arXiv:2006.08048*, 2020.

[34] 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.

[35] 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.

[36] J. G. Melo, R. D. C. Monteiro, and H. Wang. A proximal augmented Lagrangian method for linearly constrained nonconvex composite optimization problems. *Journal of Optimization Theory and Applications*, pages 1–33, 2023.

[37] 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(1):475–507, 2013.

[38] I. Necoara, A. Patrascu, and F. Glineur. Complexity of first-order inexact Lagrangian and penalty methods for conic convex programming. *Optim. Methods Softw.*, pages 1–31, 2017.

[39] A. Patrascu, I. Necoara, and Q. Tran-Dinh. Adaptive inexact fast augmented Lagrangian methods for constrained convex optimization. *Optim. Lett.*, 11(3):609–626, 2017.

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

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

[42] A. Sujanani and R. D. C. Monteiro. An adaptive superfast inexact proximal augmented Lagrangian method for smooth nonconvex composite optimization problems. *Journal of Scientific Computing*, 97 (2):34, 2023.

[43] K. Sun and X. A. Sun. Dual descent augmented Lagrangian method and alternating direction method of multipliers. *SIAM Journal on Optimization*, 34(2):1679–1707, 2024.

[44] A. Themelis and P. Patrinos. Douglas–Rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181, 2020.

[45] Y. Wang, W. Yin, and J. Zeng. Global convergence of ADMM in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78(1):29–63, 2019.

[46] Y. Xu. Iteration complexity of inexact augmented Lagrangian methods for constrained convex programming. *Math. Program.*, 2019. ISSN 1436-4646. doi: 10.1007/s10107-019-01425-9. URL https://doi.org/10.1007/s10107-019-01425-9.

[47] J. Zeng, W. Yin, and D. Zhou. Moreau Envelope Augmented Lagrangian method for Nonconvex Optimization with Linear Constraints. *J. Scientific Comp.*, 91(61), April 2022.

[48] J. Zhang and Z.-Q. Luo. A proximal alternating direction method of multiplier for linearly constrained nonconvex minimization. *SIAM J. Optim.*, 30(3):2272–2302, 2020. doi: 10.1137/19M1242276.

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

[50] J. Zhang and Z.-Q. Luo. A global dual error bound and its application to the analysis of linearly constrained nonconvex optimization. *SIAM Journal on Optimization*, 32(3):2319–2346, 2022.