Regulated ADMM for Constrained Nonconvex Optimization

Yu Yang [a], Qing-Shan Jia [b], Zhanbo Xu [a], Xiaohong Guan [a,b], Costas J. Spanos [c]

[a] *School of Automation Science and Engineering, Xi'an Jiaotong University, Shaanxi, China.*

[b] *CFINS, Department of Automation, BNRist, Tsinghua University, Beijing, China.*

[c] *Electrical Engineering and Computer Science, University of California, Berkeley, CA, USA.*

Abstract

By enabling the nodes or agents to solve small-sized subproblems to achieve coordination, distributed algorithms are favored by many networked systems for efficient and scalable computation. While for convex problems, substantial distributed algorithms are available, the results for the nonconvex counterparts are extremely lacking. The generalizations of the convex results to the applications with nonconvex settings are generally hampered either due to the failure of convergence or the lack of convergence guarantee. Motivated by applications, this paper focuses on developing a distributed algorithm for a class of nonconvex problems featured by i) a nonconvex objective formed by separate and composite components regarding the decision variables of multiple interconnected agents, ii) local bounded convex constraints, and iii) coupled global linear constraints. This problem is directly originated from smart buildings and is also broad in other domains. To provide a distributed algorithm with convergence guarantee, we revise the existing powerful tool of alternating direction method of multiplier (ADMM) and proposed a regulated ADMM. Specifically, noting that the main difficulty to establish the convergence for the problem within the ADMM framework is to assume the boundness of dual updates, we propose to regulate the dual update procedure by imposing a discounted factor. This leads to the establishment of the so-called sufficiently decreasing Lyapunov function, which is critical to establish the convergence. We prove that the method converges to some approximate stationary points. We besides showcase the efficacy and performance of the method by a numeric example and the concrete application to multi-zone heating, ventilation, and air-conditioning (HVAC) control in smart buildings.

Key words: distributed nonconvex optimization, regulated ADMM, bounded Lagrangian multipliers, global convergence.

1 Introduction

By enabling the nodes or agents to solve small-sized subproblems to achieve coordination, distributed algorithms are favored by many networked systems to achieve efficient and scalable computation. While distributed algorithms for convex optimization have been studied extensively, the results for the more broad nonconvex counterparts are extremely lacking. The direct extensions of distributed algorithms for convex problems to the applications with nonconvex settings are generally hampered either due to the failure of convergence or the lack of convergence guarantee. Therefore, it is important and necessary to study distributed algorithms that can handle the nonconvex counterparts with convergence guarantee. This paper focuses on developing a distributed algorithm for a class of nonconvex problems given by

$$\min_{\mathbf{x}=\{\mathbf{x}_i\}_{i=1}^N} F(\mathbf{x}) = g(\mathbf{x}) + \sum_{i=1}^N f_i(\mathbf{x}_i) \qquad (\mathbf{P})$$

$$\text{s.t.} \ \sum_{i=1}^N \mathbf{A}_i\mathbf{x}_i = \mathbf{b}. \qquad (1a)$$

$$\mathbf{x}_i \in \mathbf{X}_i, \ i = 1, 2, \cdots, N. \qquad (1b)$$

where $i = 1, 2, \cdots, N$ denotes the computing nodes or agents. Variable $\mathbf{x}_i \in \mathbf{R}^{n_i}$ represents the decision variables of agent i and $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^N \in \mathbf{R}^n$ with $n = \sum_{i=1}^N n_i$

⋆ This work is supported by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. This work is also supported in part by National Natural Science Foundation of China (62192752, 62192750, 62125304, 62073182), 111 International Collaboration Project (BP2018006), and Tsinghua University Initiative Scientific Research Program.

Yu Yang is the corresponding author.

Email addresses: yangyu21@xjtu.edu.cn (Yu Yang), jiaqs@tsinghua.edu.cn (Qing-Shan Jia), zhanbo.xu@xjtu.edu.cn (Zhanbo Xu), xhguan@xjtu.edu.cn (Xiaohong Guan), spanos@berkeley.edu (Costas J. Spanos).

denotes the stacked decision variables for all the agents. Function $f_i : \mathbf{R}^{n_i} \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$ denote the separate and composite objective components, which are continuously differentiable but possibly nonconvex. The agents are expected to cooperatively optimize their decision variables so as to achieve the optimal overall system performance measured by the objective $F(\mathbf{x}) = g(\mathbf{x}) + \sum_{i=1}^N f_i(\mathbf{x}_i)$ while respecting their local decision boundaries indicated by the bounded convex constraints $\mathbf{X}_i$ as well as the global coupled linear constraints (1a) encoded by $\mathbf{A}_i \in \mathbf{R}^{m\times n_i}$ and $b \in \mathbf{R}^m$. For notation, we define $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_N) \in \mathbf{R}^{m\times n}$, and therefore the coupled linear constraints take the form of $\mathbf{Ax} = \mathbf{b}$.

Problem ($\mathbf{P}$) is directly originated from smart buildings where smart devices are empowered to make local decisions while accounting for the interactions or the shared resource limits with the other devices in the proximity (see, for examples [1, 2]). Many other applications also fit into this formulation, including but not limited to smart sensing [3], electric vehicle charging management [4, 5], power system control [6], wireless communication control [7]. When the number of nodes is large, centralized methods usually suffer bottlenecks from the heavy computation, storing and communication. Also, centralized methods may disrupt privacy as the complete information for all agents (i.e., private local objectives) are required. In this regard, distributed algorithms are usually preferred for privacy, computing efficiency, small data storage, and scaling properties.

When the objective functions f_i are convex and $g = 0$, various distributed methods are available. Dual decomposition is one typical option which can nicely leverage the separable structure of problem to enable distributed computation [8, 9]. Dual decomposition is a primal-dual algorithm based on the Lagrangian relaxation technique. The effectiveness of dual decomposition relies on the strong duality stating that the optimal primal value coincides with the optimal dual value. This is generally hold by convex problems. When the problem is nonconvex, the convergence of the method and the quality of solution generally can not be guaranteed due to the lack of strong duality. For convex scenarios, another major category of distributed algorithms is developed along the augmented Lagrangian (AL) framework which penalizes the coupled constraints by a quadratic term in addition to the dual variables with Larangian relaxation [10]. AL technique was proposed in the late 1960s by Hestenes [11] and Powell [12] for constrained optimization. Since then, the line of work till the late 1990s was mostly consolidated in a monograph by Bertsekas [10]. Due to the emergence of large-scale networked control systems driving by the Internet and Communication technologies (ICT) as well as the massive heavy computing tasks arising from Artificial Intelligence (AI) fields, AL technique was brought to front and renewed by Boyd [13] to enable efficient distributed computation. Compared with Lagrangian relaxation, AL relaxation is assumed to enhance the convergence properties of distributed computation but at the cost of disrupting the separable property probably hold by the problem due to the quadratic penalty terms [10]. To enable distributed computation, alternating optimization technique was first employed to the two-block case ($N = 2$ with problem ($\mathbf{P}$)), leading to the well-known alternating direction method of multiplier (ADMM) [13]. The main idea of alternating optimization is to empower the agents to update their decision components in a sequential manner. The convergence of two-block ADMM was comprehensively reviewed in [13]. Mainly due to the superior convergence and easy implementation (i.e., fixed step-size) properties, the extensions of ADMM to more general settings have attracted extensive interest from the research community. The extensions mainly lie in three folds which include multi-blocks, parallel computation, and faster convergence. For example, the convergence of the direct extension of ADMM to multi-block scenarios was examined in [14], and established under some extra conditions like strong convexity in [15, 16]. To enable parallel computation, *Jacobian-type* ADMM was proposes to enable parallel primal updates [16], which is opposed to the *Gauss-Seidel* ADMM that employs sequential primal updates. As argued in [16], *Gauss-Seidel* ADMM generally provides stronger convergence property over *Jacobian-type* ADMM, but underperforms the latter in scaling properties due to the sequential paradigms. Regarding the convergence rate, [17] proposed an accelerated Distributed Augmented Lagrangian (ADAL) method for problem ($\mathbf{P}$) with $g = 0$ in convex settings.

The above results are all for convex problems. Nevertheless, massive applications arising from the engineering systems and machine learning domains require to handle the type of problem ($\mathbf{P}$) with possibly nonconvex objectives. The non-convexity may originate from the intrinsic complex system metrics or the penalty imposed on the constraints. When the objective functions f_i and g lack convexity, the existing distributed methods generally can not be directly utilized either due to the failure of convergence or the lack of convergence guarantee. To our best knowledge, [18] is the scarce work that proposed a distributed method with local convergence guarantee for nonconvex ($\mathbf{P}$) where $g = 0$. The notion of local convergence is that the convergence is established by assuming the iterations start with a point sufficiently close to some local optima. By investigating the literature, we observe that there lack distributed methods for nonconvex ($\mathbf{P}$) where g is not null and the convergence guarantee is assured.

To fill the gap, we focus on developing a distributed algorithm for ($\mathbf{P}$). Our main contributions are

- We revise the classic ADMM framework and propose a regulated ADMM for problem ($\mathbf{P}$). The method takes the *Gauss Seidel* scheme and favors parallel computation.
- We establish the global convergence of the method towards the approximate stationary points.

- We showcase the performance of the distributed method with a numeric example and a concrete application arising from smart buildings.

The reminder of this paper is structured as follows. In Section II, we survey the existing distributed constrained nonconvex optimization. In Section III, we present the regulated ADMM. In Section IV, we study the convergence of the method. In Section V, we showcase the method with a numerical example and the smart building application. In Section VI, we conclude this paper.

2 Literature

As a powerful tool, augmented Lagrangian (AL) framework has dominated the line of distributed constrained optimization. Based on AL and by blending decomposition technique (i.e., alternating or Jacobian decomposition) with primal-dual update scheme, ADMM and its variations have been studied extensively for distributed convex optimization. Substantial solid theoretical results (see, for examples [5, 15–17, 26]) and successful applications (see, for examples [4, 25, 27–29]) are available. Mainly due to the desirable performance observed with ADMM in convex settings, the extensions to the nonconvex scenarios begin to raise interest. The existing results can be properly differentiated by `problem structures`, `main assumptions`, `update scheme` (i.e., *Jacobian-type* or *Gauss Seidel*) and `convergence guarantee` as reported in Table 1. In the sequel, we discuss the results by the categories represented by the problem structures.

As discussed, the first category (Row 1) resembles this paper most in problem structure except for $g = 0$ [18]. An accelerated distributed augmented Lagrangian (ADAL) method is proposed to handle the class of problems with nonconvex but continuously differentiable objectives f_i. This method follows the classic ADMM framework which is composed of a *Jacobian-type* primal update and a dual ascent update procedure. The exception is that a interpolation procedure is imposed on the primal update regarding the current and preceding updates, which reads as $\mathbf{A}_i\mathbf{x}_i^{k+1} = \mathbf{A}_i\mathbf{x}_i^k + \mathbf{T}\left(\mathbf{A}_i\hat{\mathbf{x}}_i^k - \mathbf{A}_i\mathbf{x}_i^k\right)$ (k the iteration and $\mathbf{T}$ is a weighted matrix). To our understanding, this can be interpreted as a means to slow down the primal update to enhance the convergence in nonconvex settings. By assuming the existence of stationary points satisfying the strong second-order optimality condition, this paper establishes the local convergence of the method towards local optima.

The subsequent three categories (Row 2, 3, 4) differ from the first mainly in the presence of a last block encoded by $\mathbf{B}$. Noted that [23] can be viewed as a special case with $\mathbf{B} = \mathbf{I}$, where $\mathbf{I}$ are identity matrices of suitable sizes. The last block is exceptional due to the unconstrained and Lipschitz differentiable (Lipschitz continuous gradient) features, which are critical to bound the dual updates to establish the convergence (see the references therein). While the first category employs *Jacobian-type* scheme for primal update, the subsequent three categories fit into the *Gauss-Seidel* paradigms using alternating optimization [19, 23, 24]. Particularly, [19, 23] build a general framework to establish the convergence of *Gauss-Seidel* ADMM towards local optima or stationary points in nonconvex settings, which comprises two key steps, i.e., one is to identify the so-called sufficiently decreasing Lyapunov function, and the other one is to identify the lower boundness property of the Lyapunov function. The sufficiently decreasing property of Lyapunov function states that [19]

$$\begin{aligned}\mathbf{T}(\mathbf{x}^{k+1},&\boldsymbol{\lambda}^{k+1}) - \mathbf{T}(\mathbf{x}^k, \boldsymbol{\lambda}^k)\\ &\leq -a_{\mathbf{x}}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - a_{\boldsymbol{\lambda}}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2\end{aligned}$$

where $\mathbf{T}(\cdot,\cdot)$ is a general Lyapunov function, $\mathbf{x}$ and $\boldsymbol{\lambda}$ are primal and dual variables, $a_{\mathbf{x}}$ and $a_{\boldsymbol{\lambda}}$ are positive coefficients.

Generally, the AL function is a readily available option as exerted in most existing work (see [19, 23] and the references therein). However, this generally depends on the following two necessary conditions on the last block encoded by $\mathbf{B}$ to bound the dual updates $\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$ by the primal updates $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2$ [19, 23], i.e.,

a) $\mathbf{B}$ has full column rank and $\mathrm{Im}(\mathbf{A}) \subseteq \mathrm{Im}(\mathbf{B})$ ($\mathrm{Im}(\cdot)$ represents the image of a matrix).
b) The last block is unconstrained and with Lipschitz differentiable objective.

Noted that the third category originated from [23] is a special case of $\mathbf{B} = \mathbf{I}$ which certainly satisfies the necessary **condition** a).

Following the line of work, the fourth category (Row 4) studies the extension of ADMM to non-linearly constrained nonconvex problems [24, 25]. Since it is difficult (if not impossible) to directly handle the non-linear coupled constraints by the AL framework, [24] proposes to first convert the non-linearly constrained problems to linearly constrained problems by introducing duplicated copies of decision variables for interconnected agents. This leads to the linearly constrained nonconvex problems with non-linear local constraints as presented. The interpretation is that each agent holds an augmented local decision variable $\mathbf{x}_i$ composed of its local components and the copies for its neighbors. To drive the consistence of duplicated copies, a global copy $\bar{\mathbf{x}}$ regarding the decision variables of all agents is managed. [24] argues that the direction extension of ADMM to the reformulated problem is problematic for the two necessary conditions **condition** a) and b) can not be satisfied simultaneously. To bypass the challenge, [24] proposes to introduce a block of slack variables working as the last block. To force the slack block finally to *zero*, this paper adopts a two-level solution method where the inner-level exerts the classic ADMM to solve a relaxed problem associated with a penalty on the slack variables , and the outer-level gradually forces the slack variables to *zero*.

As can be perceived from the studies, it is difficult (if not impossible) to establish the convergence of clas-

Table 1
Distributed constrained nonconvex optimization

Problem structures	Main assumptions	Methods	Types	Convergence	Papers
$\min_{\{\mathbf{x}_i\}_{i=1}^N} \sum_{i=1}^N f_i(\mathbf{x}_i)$ s.t. $\sum_{i=1}^N \mathbf{A}_i\mathbf{x}_i = \mathbf{b}$. $\mathbf{x}_i \in \mathbf{X}_i,\ i = 1, 2, \cdots, N$.	f_i continuously differentiable. Strong second-order optimality condition.	ADAL	Jacobian	Local convergence. Local optima.	[18]
$\min_{\{\mathbf{x}_i\}_{i=0}^p, y} g(\mathbf{x}) + \sum_{i=0}^p f_i(\mathbf{x}_i) + h(\mathbf{y})$ s.t. $\sum_{i=0}^p \mathbf{A}_i\mathbf{x}_i + \mathbf{B}\mathbf{y} = 0$.	g and h Lipschitz continuous gradient. f_i weakly convex; $\mathrm{Im}(\mathbf{A}) \subseteq \mathrm{Im}(\mathbf{B})$.	ADMM	Gauss-Seidel	Global convergence. Stationary points.	[19, 20] [21, 22]
$\min_{\{\mathbf{x}_k\}_{k=0}^K} \sum_{k=1}^k g_k(\mathbf{x}_k) + h(\mathbf{x}_0)$ s.t. $\mathbf{x}_k = \mathbf{x}_0$. $\mathbf{x}_0 \in \mathbf{X}$.	g Lipschitz continuous gradient. h convex.	Flexible ADMM	Gauss-Seidel	Global convergence. Stationary points.	[23]
$\min_{\{\mathbf{x}_k\}_{k=0}^K} \sum_{k=1}^N g_x(\mathbf{x}_k) + \ell(\mathbf{x}_0)$ s.t. $\sum_{k=1}^K \mathbf{A}_k\mathbf{x}_k = \mathbf{x}_0$. $\mathbf{x}_k \in \mathbf{X}_k,\ k = 1, \cdots, N$.	ℓ Lipschitz continuous gradient. g nonconvex but smooth or convex but non-smooth.	Flexible ADMM	Gauss-Seidel	Global convergence. Stationary points.	[23]
$\min_{\{\mathbf{x}_i\}_{i=1}^N, \bar{\mathbf{x}}} \sum_{i=1}^N f_i(\mathbf{x}_i)$ s.t. $\sum_{i=1}^N \mathbf{A}_i\mathbf{x}_i + \mathbf{B}\bar{\mathbf{x}} = 0$. $\mathbf{x}_i \in \mathbf{X}_i, h_i(\mathbf{x}_i) = 0, i = 1, \cdots, N$. $\bar{\mathbf{x}} \in \bar{\mathbf{X}}$.	f_i continuously differentiable. h_i non-linear (possibly nonconvex). $\mathbf{B}$ full column rank. $\mathbf{X}_i$ possibly nonconvex.	ALM + ADMM	Gauss-Seidel	Global convergence. Stationary points.	[24, 25]
$\min_{\{\mathbf{x}_i\}_{i=1}^N} g(\mathbf{x}) + \sum_{i=1}^N f_i(\mathbf{x}_i)$ s.t. $\sum_{i=1}^N \mathbf{A}_i\mathbf{x}_i = \mathbf{b}$. $\mathbf{x}_i \in \mathbf{X}_i,\ i = 1, 2, \cdots, N$.	f_i and g Lipschitz continuous gradient.	Regulated ADMM	Jacobian	Global convergence. Approximate stationary points.	This paper

Note: the set $\mathbf{X}_i$ and $\bar{\mathbf{X}}$ are bounded convex sets.

sic ADMM for problem (**P**) due to the lack of a well-behaved last block satisfying **condition** a) and b). [18] provides a solution with local convergence guarantee but can not handle the probable composite objective components g. Though the idea of introducing slack variables proposed in [24] can provide a solution with global convergence guarantee but at the cost of heavy iteration complexity caused by the two-level structure. Despite these limitations, what we can learn from the studies is that the behaviors of dual variables is important to draw the convergence of ADMM for nonconvex problems. To overcome the challenges, we thus propose to regulate the dual update procedure to bound the behaviors of the dual variables manually by imposing a discounting factor. This leads to a regulated ADMM to be discussed. For the regulated ADMM, we are able to draw a sufficiently decreasing and lower bounded Lyapunov function which can guide the convergence of the method towards approximate stationary points.

3 Regulated ADMM

3.1 Notations

Throughout the paper, we will visit the following notations. We refer to the bold alphabets $\mathbf{x}, \mathbf{y}, \mathbf{a}, \mathbf{b}, \mathbf{c}$ as vectors, and the bold alphabets $\mathbf{A}, \mathbf{A}_i, \mathbf{Q}, \mathbf{M}$ as matrices. We use $\mathbf{I}_n$ or $\mathbf{I}$ to denote identity matrices of $n \times n$ or suitable sizes. We use $\mathbf{R}^n$ to represent the n-dimensional real space. We use $\{\mathbf{x}_i\}_{i=1}^N = (\mathbf{x}_1^\top, \mathbf{x}_2^\top, \cdots, \mathbf{x}_N^\top)^\top$ to stack the sub-vector $\mathbf{x}_i \in \mathbf{R}^{n_i}$. We have $\|\mathbf{x}\|^2 = \sum_{i=1}^N x_i^2$ denote the Euclidean norm of vector $\mathbf{x} = \{x_i\}_{i=1}^n \in \mathbf{R}^n$ without specification. We have $\langle \mathbf{x}, \mathbf{y} \rangle$ denote the dot product of vector $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$. We besides have $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^\top \mathbf{A}\mathbf{x}$. We use $\mathrm{diag}(\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_N)$ to denote the diagonal matrix formed by the sub-matrices $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_N$. We use $:=$ to denote the definitions. We have the normal cone to a convex set $\mathbf{X} \subseteq \mathbf{R}^n$ at $\mathbf{x}^*$ defined as $N_{\mathbf{X}}(\mathbf{x}^*) := \{\nu \in \mathbf{R}^n | \langle \nu, \mathbf{x} - \mathbf{x}^* \rangle \leq 0, \forall \mathbf{x} \in \mathbf{X}\}$. For $g : \mathbf{R}^n \to \mathbf{R}$, we denote $\nabla_i g(\mathbf{x}) = \partial g(\mathbf{x})/\partial x_i$ where $\mathbf{x} = \{x_i\}_{i=1}^n$.

3.2 Algorithm

As discussed, the proposed regulated ADMM fits into the AL framework. We form the AL for problem (**P**) as

$$\mathbb{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = g(\mathbf{x}) + f(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{A}\mathbf{x} - \mathbf{b} \rangle + \rho/2 \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$$

where $\boldsymbol{\lambda} \in \mathbf{R}^m$ is the Lagrangian multiplier, ρ is the penalty parameter.

Resembling most AL methods, regulated ADMM is a primal-dual method. As shown in **Algorithm** 1, the main procedures are composed of **Primal update** and **Dual update**. To handle the composite objective components in a distributed manner, we choose to linearize the term at each iteration k by $\langle \nabla g(\mathbf{x}^k), \mathbf{x} - \mathbf{x}^k \rangle$. To favor parallel computation and scaling properties, the regulated ADMM admits the *Jacobian* scheme in **Primal update** which means at each iteration the agents update their local decision variables in parallel by assuming the preceding information for the interconnected agents. Particularly, to enhance convergence, the **Primal update** optimizes the proxy of the linearized AL function plus a proximal term $\|\mathbf{x}_i - \mathbf{x}_i^{k+1}\|^2$ at each iteration (Step 3). This has been used in many *Jacobian-type* ADMM in convex settings (see [16] and the references therein) or where the linearization technique is exerted to handle the composite objective components (see, for examples [30]). Nevertheless, different from the classic ADMM, we have modified the **Dual update** in regulated ADMM by imposing a discounting factor $(1 - \tau)$ ($\tau \in (0, 1)$), which is aimed to bound the dual update in the iterations (Step 4). The idea behind is to update the dual variables in terms of the residual in a discounted manner so as to bound the dual updates. This can be perceived that

$$\begin{aligned} \boldsymbol{\lambda}^{k+1} &= (1-\tau)\boldsymbol{\lambda}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}) \\ &= (1-\tau)^2 \boldsymbol{\lambda}^{k-1} + (1-\tau)\rho(\mathbf{A}\mathbf{x}^k - \mathbf{b}) \\ &\qquad + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}) \\ &\cdots \\ &= (1-\tau)^{k+1}\boldsymbol{\lambda}^0 + \sum_{\ell=0}^{k} (1-\tau)^{k-\ell} \rho(\mathbf{A}\mathbf{x}^{\ell+1} - \mathbf{b}). \end{aligned} \tag{2}$$

This differs from the classic ADMM where we have the Lagrangian multipliers is the running sum of the residual in the meaning that

$$\begin{aligned} \boldsymbol{\lambda}^{k+1} &= \boldsymbol{\lambda}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}) \\ &= \boldsymbol{\lambda}^{k-1} + \rho(\mathbf{A}\mathbf{x}^k - \mathbf{b}) + \rho(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}) \\ &\cdots \\ &= \boldsymbol{\lambda}^0 + \sum_{\ell=0}^{k} \rho(\mathbf{A}\mathbf{x}^{\ell+1} - \mathbf{b}). \end{aligned}$$

In this aspect, the classic ADMM can be viewed as a special case of the regulated ADMM with $\tau = 0$.

Algorithm 1 Regulated ADMM for problem (**P**)

1: **Initialize:** $\mathbf{x}^0$, $\boldsymbol{\lambda}^0$ and ρ, $\tau \in [0, 1)$, and set $k \to 0$.
2: **Repeat:**
3: **Primal update**:

$$\mathbf{x}_i^{k+1} = \arg\min_{\mathbf{x}_i \in \mathbf{X}_i} \left\{ \begin{array}{l} \langle \nabla g_i(\mathbf{x}^k), \mathbf{x}_i - \mathbf{x}_i^k \rangle \\ + f_i(\mathbf{x}_i) + \langle \boldsymbol{\lambda}^k, \mathbf{A}_i \mathbf{x}_i^k \rangle \\ + \rho/2 \|\mathbf{A}_i \mathbf{x}_i^k + \sum_{j \neq i} \mathbf{A}_j \mathbf{x}_j^k - \mathbf{b}\|^2 \\ + \beta/2 \|\mathbf{x}_i - \mathbf{x}_i^k\|_{\mathbf{B}_i}^2 \end{array} \right\} \tag{3}$$

4: **Dual update**:

$$\boldsymbol{\lambda}^{k+1} = (1-\tau)\boldsymbol{\lambda}^k + \rho\left(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}\right) \tag{4}$$

5: Until convergence.

4 Convergence Analysis

Before we establish the convergence for **Algorithm** 1, we first clarify the main assumptions.

4.1 Main assumptions

(A1) Function $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$ have Lipschitz continuous gradient with modulus L_f and L_g over the compact set $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_N$, i.e., [21]

$$\begin{aligned} &|f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| \\ &\qquad \leq L_f/2 \, \|\mathbf{y} - \mathbf{x}\|^2, \forall \mathbf{x}, \mathbf{y} \in \mathbf{X}. \\ &\text{or } \ \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L_f \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbf{X}. \\ &|g(\mathbf{y}) - g(\mathbf{x}) - \langle \nabla g(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| \\ &\qquad \leq L_g/2 \, \|\mathbf{y} - \mathbf{x}\|^2, \forall \mathbf{x}, \mathbf{y} \in \mathbf{X}. \\ &\text{or } \ \|\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})\| \leq L_g \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbf{X}. \end{aligned}$$

(A2) Function $f : \mathbf{R}^n \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$ are lower bounded over the compact set $\mathbf{X} = \mathbf{X}_1 \times \mathbf{X}_2 \times \cdots \times \mathbf{X}_N$, i.e.,

$$\begin{aligned} f(\mathbf{x}) &> -\infty, \quad \forall \mathbf{x} \in \mathbf{X}. \\ g(\mathbf{x}) &> -\infty, \quad \forall \mathbf{x} \in \mathbf{X}. \end{aligned}$$

4.2 Main results

As discussed in the literature, one critical step to establish convergence for distributed AL method in nonconvex settings is to identify the so-called sufficiently decreasing Lyapunov function. To achieve the objective, we first draw the following two propositions.

Proposition 1 *For the sequences* $\{\mathbf{x}^k\}_{k \in \mathbf{K}}$ *and*

$\{\boldsymbol{\lambda}^k\}_{k\in\mathbf{K}}$ *generated by **Algorithm** 1, we have*

$$\begin{aligned}
&\frac{1-2\tau^2}{2\rho}\left\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\right\|^2+\frac{1}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|_{\mathbf{Q}}^2\\
&\quad+\frac{L_g}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+\frac{1}{2}\|\mathbf{w}^k\|_{\mathbf{Q}}^2\\
&\le\frac{1-2\tau^2}{2\rho}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\right\|^2+\frac{1}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|_{\mathbf{Q}}^2\\
&\quad+\frac{L_g}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2+\rho_F\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|^2\\
&\quad-\tau(1+\tau)/\rho\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2.
\end{aligned}$$

where we have $\mathbf{K}:=\{1,2,\cdots,K\}$ *and*

$$\begin{aligned}
\mathbf{w}^k&:=(\mathbf{x}^{k+1}-\mathbf{x}^k)-(\mathbf{x}^k-\mathbf{x}^{k-1})\\
G_{\mathbf{A}}&:=diag\left(\mathbf{A}_1^\top\mathbf{A}_1,\cdots,\mathbf{A}_N^\top\mathbf{A}_N\right)\\
G_{\mathbf{B}}&:=diag\left(\mathbf{B}_1^\top\mathbf{B}_1,\cdots,\mathbf{B}_N^\top\mathbf{B}_N\right)\\
\mathbf{Q}&:=\rho G_{\mathbf{A}}+\beta G_{\mathbf{B}}-\rho\mathbf{A}^\top\mathbf{A}\\
\rho_F&:=L_f+L_g.
\end{aligned}$$

Proof of Prop. 1: **Prop.** 1 is established based on the first-order optimality condition of the subproblems (3) and the Lipschitz continuous gradient properties of f and g.

We first establish the following equality and notation.

$$\begin{aligned}
&\mathbf{A}_i\mathbf{x}_i^{k+1}+\textstyle\sum_{j\neq i}\mathbf{A}_j\mathbf{x}_j^k-\mathbf{b} &(5)\\
&\quad=\mathbf{A}\mathbf{x}^k-\mathbf{b}+\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k).\\
&\quad=\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}+\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1})+\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k).\\
&\hat{\boldsymbol{\lambda}}^k:=\boldsymbol{\lambda}^k+\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}). &(6)
\end{aligned}$$

For $\mathbf{x}_i$-update (3), the first-order optimality condition states that there exists $\nu_i^{k+1}\in N_{\mathbf{X}_i}(\mathbf{x}_i^{k+1})$ that

$$\begin{aligned}
0=\;&\nabla f_i(\mathbf{x}_i^{k+1})+\nabla g_i(\mathbf{x}^k)+\mathbf{A}_i^\top\boldsymbol{\lambda}^k\\
&+\rho\mathbf{A}_i^\top(\mathbf{A}_i\mathbf{x}_i^{k+1}+\textstyle\sum_{j\neq i}\mathbf{A}_j\mathbf{x}_j^k-\mathbf{b})\\
&+\beta\mathbf{B}_i^\top\mathbf{B}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)+\nu_i^{k+1}\\
=\;&\nabla f_i(\mathbf{x}_i^{k+1})+\nabla g_i(\mathbf{x}^k)+\mathbf{A}_i^\top\left(\boldsymbol{\lambda}^k+\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b})\right)\\
&+\rho\mathbf{A}_i^\top\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1})+\rho\mathbf{A}_i^\top\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\\
&+\beta\mathbf{B}_i^\top\mathbf{B}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)+\nu_i^{k+1}\qquad\text{by (5)}\\
=\;&\nabla f_i(\mathbf{x}_i^{k+1})+\nabla g_i(\mathbf{x}^k)+\mathbf{A}_i^\top\hat{\boldsymbol{\lambda}}^k+\rho\mathbf{A}_i^\top\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1})\\
&+\rho\mathbf{A}_i^\top\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\\
&+\beta\mathbf{B}_i^\top\mathbf{B}_i\left(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k\right)+\nu_i^{k+1}\qquad\text{by (6).}
\end{aligned}$$

Multiplying by $(\mathbf{x}_i^{k+1}-\mathbf{x}_i)$ in both sides, we have

$$\begin{aligned}
&\langle\nabla f_i(\mathbf{x}_i^{k+1}),\mathbf{x}_i^{k+1}-\mathbf{x}_i\rangle+\langle\nabla_i g(\mathbf{x}^k),\mathbf{x}_i^{k+1}-\mathbf{x}_i\rangle\\
&\quad+\langle\hat{\boldsymbol{\lambda}}^k,\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i)\rangle\\
&\quad+\rho\langle\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1}),\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i)\rangle\\
&\quad+\rho\langle\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\mathbf{A}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i)\rangle\\
&\quad+\beta\langle\mathbf{B}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k),\mathbf{B}_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i)\rangle\\
&=-\langle\nu_i^{k+1},\mathbf{x}_i^{k+1}-\mathbf{x}_i\rangle\le 0,\quad\forall\mathbf{x}_i\in\mathbf{X}_i.
\end{aligned}\tag{7}$$

Summing up (7) over $i\in\mathbf{N}$, we have $\forall\mathbf{x}_i\in\mathbf{X}_i$,

$$\begin{aligned}
&\langle\nabla f(\mathbf{x}^{k+1}),\mathbf{x}^{k+1}-\mathbf{x}\rangle+\langle\nabla g(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}\rangle\\
&\quad+\langle\hat{\boldsymbol{\lambda}}^k,\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x})\rangle+(\mathbf{x}^{k+1}-\mathbf{x})\rho\mathbf{A}^\top\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1})\\
&\quad+\textstyle\sum_i(\mathbf{x}_i^{k+1}-\mathbf{x}_i)^\top(\rho\mathbf{A}_i^\top\mathbf{A}_i+\beta\mathbf{B}_i^\top\mathbf{B}_i)(\mathbf{x}_i^{k+1}-\mathbf{x}_i^k)\le 0.
\end{aligned}$$

Plugging in $\mathbf{Q}:=\rho G_{\mathbf{A}}+\beta G_{\mathbf{B}}-\rho\mathbf{A}^\top\mathbf{A}$, we have

$$\begin{aligned}
&\langle\nabla f(\mathbf{x}^{k+1}),\mathbf{x}^{k+1}-\mathbf{x}\rangle+\langle\nabla g(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}\rangle\\
&\quad+\langle\hat{\boldsymbol{\lambda}}^k,\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x})\rangle\\
&\quad+(\mathbf{x}^{k+1}-\mathbf{x})^\top\mathbf{Q}(\mathbf{x}^{k+1}-\mathbf{x}^k)\le 0,\quad\forall\mathbf{x}\in\mathbf{X}.
\end{aligned}\tag{8}$$

By induction, we have for iteration $k-1$ that

$$\begin{aligned}
&\left\langle\nabla f(\mathbf{x}^k),\mathbf{x}^k-\mathbf{x}\right\rangle+\langle\nabla g(\mathbf{x}^{k-1}),\mathbf{x}^k-\mathbf{x}\rangle\\
&\quad+\left\langle\hat{\boldsymbol{\lambda}}^{k-1},\mathbf{A}(\mathbf{x}^k-\mathbf{x})\right\rangle\\
&\quad+(\mathbf{x}^k-\mathbf{x})^\top\mathbf{Q}(\mathbf{x}^k-\mathbf{x}^{k-1})\le 0,\quad\forall\mathbf{x}\in\mathbf{X}.
\end{aligned}\tag{9}$$

Assigning $\mathbf{x}:=\mathbf{x}^k$ and $\mathbf{x}:=\mathbf{x}^{k+1}$ with (8) and (9), we have

$$\begin{aligned}
&\langle\nabla f(\mathbf{x}^{k+1}),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle+\langle\nabla g(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&\quad+\langle\hat{\boldsymbol{\lambda}}^k,\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\rangle\\
&\quad+(\mathbf{x}^{k+1}-\mathbf{x}^k)^\top\mathbf{Q}(\mathbf{x}^{k+1}-\mathbf{x}^k)\le 0. &(10)\\
&\langle\nabla f(\mathbf{x}^k),\mathbf{x}^k-\mathbf{x}^{k+1}\rangle+\langle\nabla g(\mathbf{x}^{k-1}),\mathbf{x}^k-\mathbf{x}^{k+1}\rangle\\
&\quad+\langle\hat{\boldsymbol{\lambda}}^{k-1},\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1})\rangle\\
&\quad+(\mathbf{x}^k-\mathbf{x}^{k+1})^\top\mathbf{Q}(\mathbf{x}^k-\mathbf{x}^{k-1})\le 0. &(11)
\end{aligned}$$

Summing up (10) and (11) and plugging in $\mathbf{w}^k:=(\mathbf{x}^{k+1}-\mathbf{x}^k)-(\mathbf{x}^k-\mathbf{x}^{k-1})$, we have

$$\begin{aligned}
&\langle\nabla f(\mathbf{x}^{k+1})-\partial f(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&\quad+\langle\nabla g(\mathbf{x}^k)-\nabla g(\mathbf{x}^{k-1}),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&\quad+\langle\hat{\boldsymbol{\lambda}}^k-\hat{\boldsymbol{\lambda}}^{k-1},\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\rangle\\
&\quad+(\mathbf{x}^{k+1}-\mathbf{x}^k)^\top\mathbf{Q}\mathbf{w}^k\le 0.
\end{aligned}\tag{12}$$

Based on the Lipschitz continuous gradient property of $f(\mathbf{x})$ over the compact set $\mathbf{X}$, we have

$$\langle\nabla f(\mathbf{x}^{k+1})-\nabla f(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\ge -L_f\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2.\tag{13}$$

We also have

$$\begin{aligned}
&\langle\nabla g(\mathbf{x}^k)-\nabla g(\mathbf{x}^{k-1}),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&=\langle\frac{\nabla g(\mathbf{x}^k)-\nabla g(\mathbf{x}^{k-1})}{\sqrt{L_g}},\sqrt{L_g}(\mathbf{x}^{k+1}-\mathbf{x}^k)\rangle\\
&\ge-\frac{1}{2L_g}\|\nabla g(\mathbf{x}^k)-\nabla g(\mathbf{x}^{k-1})\|^2-\frac{L_g}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2\\
&\ge-\frac{L_g}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2-\frac{L_g}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2
\end{aligned}$$

where the last equality is based on the Lipschitz continuous gradient property of g.

Besides, we have

$$\begin{aligned}
&\langle\hat{\boldsymbol{\lambda}}^k-\hat{\boldsymbol{\lambda}}^{k-1},\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\rangle\\
&=\left\langle\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k+\tau(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}),\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\right\rangle\\
&=\Big\langle\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k+\tau(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}),\\
&\qquad\frac{\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k}{\rho}-\frac{(1-\tau)}{\rho}(\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1})\Big\rangle\\
&=\frac{\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2}{\rho}-\frac{(1-2\tau)}{\rho}\langle\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k,\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\rangle\\
&\quad-\frac{\tau(1-\tau)}{\rho}\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\|^2\\
&\geq\frac{\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2}{\rho}-\frac{1-2\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2\\
&\quad-\frac{1-2\tau}{2\rho}\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\|^2-\frac{\tau(1-\tau)}{\rho}\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\|^2\\
&=\frac{1-2\tau^2}{2\rho}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2-\frac{1-2\tau^2}{2\rho}\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\|^2\\
&\quad+\tau(\tau+1)/\rho\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2
\end{aligned}\tag{14}$$

where the inequality is based on $\langle\mathbf{a},\mathbf{b}\rangle\leq\frac{1}{2}(\|\mathbf{a}\|^2+\|\mathbf{b}\|^2)$.

Based on the inequality $\mathbf{b}^\top\mathbf{M}(\mathbf{b}-\mathbf{c})=\frac{1}{2}(\|\mathbf{b}-\mathbf{c}\|^2_{\mathbf{M}}+\|\mathbf{b}\|^2_{\mathbf{M}}-\|\mathbf{c}\|^2_{\mathbf{M}})$, and by setting $\mathbf{M}=\mathbf{Q}$, $\mathbf{b}=\mathbf{x}^{k+1}-\mathbf{x}^k$, and $\mathbf{c}=\mathbf{x}^k-\mathbf{x}^{k-1}$, we have

$$\begin{aligned}(\mathbf{x}^{k+1}-\mathbf{x}^k)^\top\mathbf{Q}\mathbf{w}^k=&\frac{1}{2}(\|\mathbf{w}^k\|^2_{\mathbf{Q}}+\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2_{\mathbf{Q}}\\
&-\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2_{\mathbf{Q}}).\end{aligned}\tag{15}$$

Plugging (13), (14), (15) into (12), we have

$$\begin{aligned}
&\frac{1-2\tau^2}{2\rho}\left\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\right\|^2+\frac{1}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2_{\mathbf{Q}}\\
&\quad+\frac{L_g}{2}\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2+\frac{1}{2}\|\mathbf{w}^k\|^2_{\mathbf{Q}}\\
&\leq\frac{1-2\tau^2}{2\rho}\left\|\boldsymbol{\lambda}^k-\boldsymbol{\lambda}^{k-1}\right\|^2+\frac{1}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2_{\mathbf{Q}}\\
&\quad+\frac{L_g}{2}\|\mathbf{x}^k-\mathbf{x}^{k-1}\|^2+(L_g+L_f)\left\|\mathbf{x}^{k+1}-\mathbf{x}^k\right\|^2\\
&\quad-\tau(1+\tau)/\rho\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2.
\end{aligned}$$

We therefore complete the proof.

Let $\mathbb{L}^+_\rho(\mathbf{x},\boldsymbol{\lambda}):=\mathbb{L}_\rho(\mathbf{x},\boldsymbol{\lambda})-\frac{\tau}{2\rho}\|\boldsymbol{\lambda}\|^2$ be the regularized AL function. The subsequent proposition is regarding the change of the regularized AL function.

Proposition 2 *For the sequences* $\{\mathbf{x}^k\}_{k\in\mathbf{K}}$ *and* $\{\boldsymbol{\lambda}^k\}_{k\in\mathbf{K}}$ *generated by* ***Algorithm*** *1, we have*

$$\begin{aligned}
&\mathbb{L}^+_\rho(\mathbf{x}^{k+1},\boldsymbol{\lambda}^{k+1})-\mathbb{L}^+_\rho(\mathbf{x}^k,\boldsymbol{\lambda}^k)\\
&\leq-\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2_{\mathbf{Q}}+\frac{\rho_F}{2}\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2\\
&\quad-\frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\|^2+\frac{2-\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1}-\boldsymbol{\lambda}^k\|^2.
\end{aligned}$$

Proof of Prop. 2: Before starting the proof, we first establish the following inequalities to be used. Based on the Lipschitz continuous gradient property of $f(\mathbf{x})$ over $\mathbf{x}\in\mathbf{X}$ (see (A1)), we have

$$\begin{aligned}
&|f(\mathbf{x}^k)-f(\mathbf{x}^{k+1})-\langle\nabla f(\mathbf{x}^{k+1}),\mathbf{x}^k-\mathbf{x}^{k+1}\rangle|\\
&\qquad\leq L_f/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2.\\
&\Rightarrow f(\mathbf{x}^{k+1})-f(\mathbf{x}^k)\leq\langle\nabla f(\mathbf{x}^{k+1}),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&\qquad+L_f/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2.
\end{aligned}\tag{16}$$

Similarly, for $g(\mathbf{x})$ with Lipschitz continuous gradient over $\mathbf{x}\in\mathbf{X}$ (see (A1)), we have

$$\begin{aligned}
&|g(\mathbf{x}^{k+1})-g(\mathbf{x}^k)-\langle\nabla g(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle|\\
&\qquad\leq L_f/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2.\\
&\Rightarrow g(\mathbf{x}^{k+1})-g(\mathbf{x}^k)\leq\langle\nabla g(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&\qquad+L_g/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2.
\end{aligned}\tag{17}$$

Besides, we have

$$\begin{aligned}
&\frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}\|^2-\frac{\rho}{2}\|\mathbf{A}\mathbf{x}^k-\mathbf{b}\|^2\\
=&\ \frac{\rho}{2}\left\langle\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{A}\mathbf{x}^{k+1}+\mathbf{A}\mathbf{x}^k-2\mathbf{b}\right\rangle \qquad (18)\\
=&\ \frac{\rho}{2}\left\langle\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\mathbf{A}(\mathbf{x}^k-\mathbf{x}^{k+1})+2(\mathbf{A}\mathbf{x}^k-\mathbf{b})\right\rangle\\
=&-\frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\|^2+\left\langle\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b})\right\rangle.
\end{aligned}$$

We next quantify the decrease of $\mathbb{L}_\rho(\mathbf{x},\boldsymbol{\lambda})$ with respect to (w.r.t.) primal update. We have

$$\begin{aligned}
&\mathbb{L}_\rho(\mathbf{x}^{k+1},\boldsymbol{\lambda}^k)-\mathbb{L}_\rho(\mathbf{x}^k,\boldsymbol{\lambda}^k)\\
&=f(\mathbf{x}^{k+1})-f(\mathbf{x}^k)+g(\mathbf{x}^{k+1})-g(\mathbf{x}^k)+\left\langle\boldsymbol{\lambda}^k,\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\right\rangle\\
&\quad+\frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b}\|^2-\frac{\rho}{2}\|\mathbf{A}\mathbf{x}^k-\mathbf{b}\|^2\\
&\leq\langle\nabla f(\mathbf{x}^{k+1})+\nabla g(\mathbf{x}^k),\mathbf{x}^{k+1}-\mathbf{x}^k\rangle+\rho_F/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2\\
&\quad+\left\langle\boldsymbol{\lambda}^k,\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\right\rangle-\frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\|^2\\
&\quad+\left\langle\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k),\rho(\mathbf{A}\mathbf{x}^{k+1}-\mathbf{b})\right\rangle\ \text{by (16), (17), (18)}\\
&=\langle\nabla f(\mathbf{x}^{k+1})+\nabla g(\mathbf{x}^k)+\mathbf{A}^\top\hat{\boldsymbol{\lambda}}^k,\mathbf{x}^{k+1}-\mathbf{x}^k\rangle\\
&\quad+\rho_F/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2-\rho/2\|\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\|^2\ \text{by (6)}\\
&\leq-\|\mathbf{x}^{k+1}-\mathbf{x}^k\|^2_{\mathbf{Q}}+\rho_F/2\|\mathbf{x}^k-\mathbf{x}^{k+1}\|^2\\
&\quad-\rho/2\ \|\mathbf{A}(\mathbf{x}^{k+1}-\mathbf{x}^k)\|^2\ \text{by (10).}
\end{aligned}\tag{19}$$

We next quantify the change of $\mathbb{L}_\rho(\mathbf{x}, \boldsymbol{\lambda})$ w.r.t. dual update. We have

$$
\begin{aligned}
&\mathbb{L}_\rho(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}) - \mathbb{L}_\rho(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k}) \\
=& \left\langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \mathbf{A}\mathbf{x}^{k+1} - \mathbf{b} \right\rangle \\
=& \left\langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \frac{\boldsymbol{\lambda}^{k+1} - (1-\tau)\boldsymbol{\lambda}^k}{\rho} \right\rangle \\
=& \left\langle \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k, \frac{1-\tau}{\rho}(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) + \frac{\tau}{\rho}\boldsymbol{\lambda}^{k+1} \right\rangle \\
=& \frac{(1-\tau)}{\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{\tau}{2\rho}\Big(\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \\
&+ \|\boldsymbol{\lambda}^{k+1}\|^2 - \|\boldsymbol{\lambda}^k\|^2\Big) \\
=& \frac{2-\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 - \frac{\tau}{2\rho}\|\boldsymbol{\lambda}^k\|^2.
\end{aligned}
\tag{20}
$$

Combining (19) and (20), we have

$$
\begin{aligned}
&\mathbb{L}_\rho(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}) - \frac{\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 - \left(\mathbb{L}_\rho(\mathbf{x}^{k}, \boldsymbol{\lambda}^{k}) - \frac{\tau}{2\rho}\|\boldsymbol{\lambda}^k\|^2\right) \\
\le& -\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{\mathbf{Q}}^2 + \frac{\rho_F}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \\
&- \frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k)\|^2 + \frac{2-\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2.
\end{aligned}
\tag{21}
$$

We therefore close the proof.

As discussed, we require to identify a sufficiently decreasing Lyapunov function to establish the convergence. In the literature, the AL function is generally exerted when the dual updates $\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$ can be bounded by the primal updates $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2$ (see, for examples [19–22]). As discussed, this is to provide the sufficiently decreasing property of Lyapunov function. This can be further interpreted that dual updates will lead to the increase of AL function by $1/\rho\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$ (set $\tau = 0$ with (20)). However, this is not the case for problem ($\mathbf{P}$) due to the lack of a well-behaved last block.

For the regulated ADMM, it turns out that we face the same challenge if the AL function or regularized AL function are exerted as the Lyapunov function for dual update will lead to the increase of (regularized) AL function by $\frac{2-\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$ (see (21)). However, another suitable Lyapunov function is possible due to the introduction discounting factor $1-\tau$ which can help bypass the challenge to bound the dual updates $\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$. Particularly, based on **Prop.** 1, we note that the term $\frac{1-2\tau^2}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{\mathbf{Q}}^2$ is descending by $\frac{\tau(1+\tau)}{\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2$ and ascending by $\rho_F\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2$, which is exactly opposite to the descending and ascending properties of the regularized AL function $\mathbb{L}_\rho^+(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1})$ as stated in **Prop.** 2. We therefore build the following Lyapunov function.

$$
\begin{aligned}
&\mathbf{T}_c(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}; \mathbf{x}^k, \boldsymbol{\lambda}^k) = \mathbb{L}_\rho^+(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}) \\
&+ c\Bigg(\frac{1-2\tau^2}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{1}{2}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{\mathbf{Q}}^2 \\
&\quad + \frac{L_g}{2}\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\Bigg)
\end{aligned}
\tag{22}
$$

where c is a constant parameter to be determined.

Let $\mathbf{T}_c^{k+1} := \mathbf{T}_c(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}; \mathbf{x}^k, \boldsymbol{\lambda}^k)$, we have the following proposition stating the sufficiently decreasing property of the Lyapunov function.

Proposition 3 *For the sequences $\{\mathbf{x}^k\}_{k\in\mathbf{K}}$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbf{K}}$ generated by **Algorithm** 1, we have*

$$
\mathbf{T}_c^{k+1} - \mathbf{T}_c^k \le -a_{\mathbf{x}}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - a_{\boldsymbol{\lambda}}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \frac{c}{2}\|\mathbf{w}^k\|^2
$$

where we have $\rho_F = L_f + L_g$ and

$$
\begin{aligned}
a_{\mathbf{x}} &:= \frac{2\rho G_{\mathbf{A}} + 2\beta G_{\mathbf{B}} - \rho\mathbf{A}^\top\mathbf{A} - (2c+1)\rho_F\mathbb{I}_N}{2} \\
a_{\boldsymbol{\lambda}} &:= \frac{2c\tau(1+\tau) - (2-\tau)}{2\rho}.
\end{aligned}
$$

Proof of Prop. 3: Based on **Prop.** 1, **Prop.** 2, we have

$$
\begin{aligned}
&\mathbf{T}_c^{k+1} - \mathbf{T}_c^k = -\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{\mathbf{Q}}^2 + \frac{\rho_f}{2}\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2 \\
&- \frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^{k+1} - \mathbf{x}^k)\|^2 + \frac{2-\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \\
&+ c\Big(\rho_f\left\|\mathbf{x}^{k+1} - \mathbf{x}^k\right\|^2 - \tau(1+\tau)/\rho\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \\
&- 1/2\|\mathbf{w}^k\|_{\mathbf{Q}}^2\Big) \\
&\le -a_{\mathbf{x}}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 - a_{\boldsymbol{\lambda}}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 - \frac{c}{2}\|\mathbf{w}^k\|_{\mathbf{Q}}^2
\end{aligned}
$$

which closes the proof.

Remark 1 ***Prop.** 3 implies that we would have the sufficiently decreasing property hold by $\mathbf{T}_c^k$ if $a_{\mathbf{x}} > 0$, $a_{\boldsymbol{\lambda}} > 0$, $c \ge 0$ and $\mathbf{Q} \ge 0$. This is can be achieved by setting the tuples (τ, ρ, β, $\mathbf{B}_i$, c) properly for **Algorithm** 1, which will be discussed shortly.*

As discussed, the other key step to draw the convergence is to examine the lower boundness property of the Lyapunov function. To this end, we first provide the lower boundness property of the Lagrangian multipliers below.

Proposition 4 *Let $\Delta := \max\{\|\mathbf{A}\mathbf{x} - \mathbf{b}\| : \mathbf{x} \in \mathbf{X}\}$, and **Algorithm** 1 starts with any given initial dual variable $\boldsymbol{\lambda}^0$, we have $\|\boldsymbol{\lambda}^k\|$ is upper bounded by*

$$
\begin{aligned}
&\|\boldsymbol{\lambda}^k\| \le \|\boldsymbol{\lambda}^0\| + \tau^{-1}\rho\Delta \\
\text{or}\ &\|\boldsymbol{\lambda}^k\|^2 \le 2\|\boldsymbol{\lambda}^0\|^0 + 2\tau^{-2}\rho^2\Delta^2.
\end{aligned}
\tag{23}
$$

Proof of Prop. 4: We define the residual of coupled constraints at iteration k as $\Delta^k := \mathbf{A}\mathbf{x}^k - \mathbf{b}$. According to (2), we have

$$\boldsymbol{\lambda}^k = (1-\tau)^{k+1}\boldsymbol{\lambda}^0 + \sum_{\ell=0}^{k}(1-\tau)^{k-\ell}\rho(\mathbf{A}\mathbf{x}^{\ell+1} - \mathbf{b})$$

We therefore have

$$\begin{aligned}\|\boldsymbol{\lambda}^k\| &= \|(1-\tau)^{k+1}\boldsymbol{\lambda}^0 + \sum_{\ell=0}^{k}\rho(1-\tau)^{k-\ell}\Delta^{\ell+1}\| \\ &\leq \|(1-\tau)^{k+1}\boldsymbol{\lambda}^0\| + \sum_{\ell=0}^{k}\|\rho\Delta(1-\tau)^{k-\ell}\| \\ &= \|(1-\tau)^{k+1}\boldsymbol{\lambda}^0\| + \rho\Delta\frac{1-(1-\tau)^k}{\tau} \\ &\leq \|\boldsymbol{\lambda}^0\| + \tau^{-1}\rho\Delta\end{aligned}$$

where the last inequality holds because of $\tau \in (0,1)$.

Further, we have $\|\boldsymbol{\lambda}^k\|^2 \leq 2\|\boldsymbol{\lambda}^0\|2 + 2\tau^{-2}\rho^2\Delta^2$, we therefore complete the proof.

Based on **Prop.** 4, we can draw the lower boundness property of Lyapunov function below.

Proposition 5 *For the sequences $\{\mathbf{x}^k\}_{k\in\mathbf{K}}$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbf{K}}$ generated by **Algorithm** 1, we have*

$$\mathbf{T}_c^{k+1} > -\infty. \tag{24}$$

Proof of Prop. 5: Recall the definition of $\mathbf{T}_c^{k+1} := \mathbf{T}_c(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}; \mathbf{x}^k, \boldsymbol{\lambda}^k)$ in (22), we have $-\frac{\tau}{2\rho}\|\boldsymbol{\lambda}^{k+1}\|^2$ is lower bounded since $\|\boldsymbol{\lambda}^{k+1}\|^2$ is upper bounded (see **Prop.** 4). We next prove $\mathbb{L}_\rho(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}) = f(\mathbf{x}^{k+1}) + \langle\boldsymbol{\lambda}^{k+1}, \mathbf{A}\mathbf{x}^{k+1} - b\rangle + \rho/2\left\|\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}\right\|^2$ is lower bounded. Note that we have $f(\mathbf{x}^{k+1}) > -\infty$ over the compact set $\mathbf{X}$ (see (A2)) and the last two terms are all positive, it suffices to prove the lower boundness of the second term on the right. We have

$$\begin{aligned}&\langle\boldsymbol{\lambda}^{k+1}, \mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}\rangle = \langle\boldsymbol{\lambda}^{k+1}, \frac{\boldsymbol{\lambda}^{k+1} - (1-\tau)\boldsymbol{\lambda}^k}{\rho}\rangle \\ &= \langle\boldsymbol{\lambda}^{k+1}, \frac{1-\tau}{\rho}(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) + \frac{\tau}{\rho}\boldsymbol{\lambda}^{k+1}\rangle \\ &= \frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 + \frac{1-\tau}{\rho}\langle\boldsymbol{\lambda}^{k+1}, \boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\rangle \\ &= \frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 + \frac{1-\tau}{2\rho}\big(\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \|\boldsymbol{\lambda}^{k+1}\|^2 - \|\boldsymbol{\lambda}^k\|^2\big)\end{aligned} \tag{25}$$

Since we have $\|\boldsymbol{\lambda}^k\|^2$ is upper bounded (see **Prop.** 4), we therefore conclude that $\mathbb{L}_\rho(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1})$ is lower bounded. Note that the other terms of $\mathbf{T}_c^{k+1}$ are all non-negative, we therefore complete the proof.

To present the main results regarding the convergence of **Algorithm** 1, we first give the following definition on **Approximate stationary solution**.

Definition 1 ***(Approximate stationary solution)*** *For any given ϵ, we say a tuple $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is an ϵ-stationary solution of problem $(\mathbf{P})$, if we have*

$$dist\big(\nabla f(\mathbf{x}^*) + \nabla g(\mathbf{x}^*) + \mathbf{A}^\top\boldsymbol{\lambda}^* + N_{\mathbf{X}}(\mathbf{x}^*), \mathbf{0}\big) + \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2 \leq \epsilon.$$

In terms of the convergence of **Algorithm** 1 for problem $(\mathbf{P})$, we have the main results below.

Theorem 1 *For **Algorithm** 1 with the tuples (τ, ρ, β, $\mathbf{B}_i$, c) satisfying*

$$\begin{aligned}&2\rho G_{\mathbf{A}} + 2\beta G_{\mathbf{B}} - \rho\mathbf{A}^\top\mathbf{A} \geq (2c+1)\rho_f \\ &\mathbf{Q} := \rho G_{\mathbf{A}} + \beta G_{\mathbf{B}} - \rho\mathbf{A}^\top\mathbf{A} \geq 0 \\ &2c\tau(1+\tau) \geq (2-\tau), \quad c \geq 0\end{aligned}$$

(a) The generated sequence $\{\mathbf{x}^k\}_{k\in\mathbf{K}}$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbf{K}}$ are bounded and convergent, i.e.,

$$\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k \to 0, \quad \mathbf{x}^{k+1} - \mathbf{x}^k \to 0.$$

(b) The limit tuples $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ are $\tau^2\rho^{-2}\|\boldsymbol{\lambda}^*\|^2$ -stationary solution of problem $(\mathbf{P})$.

Proof of Theorem 1: (a) Recall **Prop.** 3, we have

$$\begin{aligned}&\sum_{k=1}^{K}\big(\mathbf{T}_c^k - \mathbf{T}_c^{k+1}\big) \geq a_{\mathbf{x}}\sum_{k=1}^{K}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\ &\qquad + a_{\boldsymbol{\lambda}}\sum_{k=1}^{K}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{c}{2}\sum_{k=1}^{K}\|\mathbf{w}^k\|^2 \\ &\Rightarrow \mathbf{T}_c^0 - \lim_{K\to\infty}\mathbf{T}_c^{k+1} \geq a_{\mathbf{x}}\sum_{k=1}^{\infty}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 \\ &\qquad + a_{\boldsymbol{\lambda}}\sum_{k=1}^{\infty}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{c}{2}\sum_{k=1}^{\infty}\|\mathbf{w}^k\|^2 \\ &\Rightarrow \infty \geq a_{\mathbf{x}}\sum_{k=1}^{\infty}\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 + a_{\boldsymbol{\lambda}}\sum_{k=1}^{\infty}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \frac{c}{2}\sum_{k=1}^{\infty}\|\mathbf{w}^k\|^2.\end{aligned}$$

We therefore conclude

$$\begin{aligned}&\|\mathbf{x}^{k+1} - \mathbf{x}^k\| \to 0, \quad \|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\| \to 0, \\ &\mathbf{w}^k := \|(\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k) - (\boldsymbol{\lambda}^k - \boldsymbol{\lambda}^{k-1})\| \to 0.\end{aligned}$$

(b) According to (a), we have the sequences $\{\mathbf{x}^k\}_{k\in\mathbf{K}}$ and $\{\boldsymbol{\lambda}^k\}_{k\in\mathbf{K}}$ converge to some limit point $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$, i.e., $\mathbf{x}^{k+1} \to \mathbf{x}^*$, $\boldsymbol{\lambda}^{k+1} \to \boldsymbol{\lambda}^*$ and $\mathbf{x}^{k+1} \to \mathbf{x}^k$ and $\boldsymbol{\lambda}^{k+1} \to \boldsymbol{\lambda}^k$.

Recall the first-order optimality condition (8), we therefore have

$$\begin{aligned}&\nabla f(\mathbf{x}^*) + \nabla g(\mathbf{x}^*) + \mathbf{A}^\top\boldsymbol{\lambda}^* + N_{\mathbf{X}}(\mathbf{x}^*) \in 0. \\ &\Rightarrow \text{dist}\Big(\nabla f(\mathbf{x}^*) + \nabla g(\mathbf{x}^*) + \mathbf{A}^\top\boldsymbol{\lambda}^* + N_{\mathbf{X}}(\mathbf{x}^*), 0\Big) = 0.\end{aligned}$$

Based on the dual update procedure (4), we have

$$\mathbf{A}\mathbf{x}^* - \mathbf{b} = \tau\rho^{-1}\boldsymbol{\lambda}^*. \tag{26}$$

We thus have

$$\operatorname{dist}\big(\nabla f(\mathbf{x}^*) + \nabla g(\mathbf{x}^*) + \mathbf{A}^\top \boldsymbol{\lambda}^* + N_{\mathbf{X}}(\mathbf{x}^*), 0\big) + \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\| \le \tau^2 \rho^{-2} \|\boldsymbol{\lambda}^*\|^2. \tag{27}$$

We therefore complete the proof.

From **Theorem** 1, we note that if the convergent $\boldsymbol{\lambda}^*$ does not depend on τ and ρ, we could decrease τ or increase ρ to achieve any sub-optimality. If that is not the case, we given the following corollary to show we can still achieve any sub-optimality by properly setting the initial point and parameters.

Corollary 1 *For any given $\epsilon > 0$ and $\tau \in (0,1)$, if **Algorithm** 1 starts with $\boldsymbol{\lambda}^0 = 0$ and $\mathbf{A}\mathbf{x}^0 = \mathbf{b}$, and the penalty parameter ρ is selected that*

$$\rho \ge \ \epsilon^{-1}\tau\big(4 + c(1-2\tau^2) + c/2\big)d_F + cL_g/2\|\mathbf{x}^0\|^2 + \epsilon^{-1}\tau c L_g/2\|\mathbf{x}^0\|^2 + \epsilon^{-1}\tau c \rho_F/4\ d_{\mathbf{x}},$$

the limit tuples $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ are ϵ-stationary solution of problem $(\mathbf{P})$. where we have $d_F = \max_{\mathbf{x}\in\mathbf{X}} f(\mathbf{x}) + g(\mathbf{x})$, $d_{\mathbf{x}} = \max_{\mathbf{x},\mathbf{y}\in\mathbf{X}} \|\mathbf{x}-\mathbf{y}\|^2$, and we assume $f(\mathbf{x}) \ge 0$ and $g(\mathbf{x}) \ge 0$ without losing any generality.

Proof of Corollary 1: The proof is structured into two parts which include i) Prove $\|\boldsymbol{\lambda}^*\|^2 \le \rho\tau^{-1}\mathbf{T}_c^0$, and ii) Prove ii) $\mathbf{T}_c^0 \le \big(4+c(1-2\tau^2)+c/2\big)d_F + cL_g/2\|\mathbf{x}^0\|^2 + c\rho_F/4\ d_{\mathbf{x}}$.

i) Prove $\|\boldsymbol{\lambda}^*\|^2 \le \rho\tau^{-1}\mathbf{T}_c^0$: Based on the sufficiently decreasing property of $\mathbf{T}_c^{k+1}$ (see **Prop.** 3), we have

$$\mathbf{T}_c^{k+1} \le \mathbf{T}_c^0 \tag{28}$$

Recalling the definition of the Lyapunov function in (22) and invoking (25), we have

$$\begin{aligned}
&\mathbf{T}_c^{k+1} = f(\mathbf{x}^{k+1}) + g(\mathbf{x}^{k+1}) + \frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 \\
&+ \frac{1-\tau}{2\rho}\big(\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 + \|\boldsymbol{\lambda}^{k+1}\|^2 - \|\boldsymbol{\lambda}^k\|^2\big) \\
&+ \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}\|^2 + c\Big(\frac{1-2\tau^2}{2\rho}\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|^2 \\
&+ 1/2\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{\mathbf{Q}}^2 + L_g/2\|\mathbf{x}^k - \mathbf{x}^{k-1}\|^2\Big)
\end{aligned} \tag{29}$$

We have $f \ge 0$ and $g \ge 0$ over $\mathbf{X}$. By combing (28) and (29), we have (the other terms are all non-negative)

$$\frac{1-\tau}{2\rho}\big(\|\boldsymbol{\lambda}^{k+1}\|^2 - \|\boldsymbol{\lambda}^k\|^2\big) + \frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 \le \mathbf{T}_c^0 \tag{30}$$

We next prove $\frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 \le \mathbf{T}_c^0$ by induction. For $k = 0$, we can properly pick the initial point to satisfy the inequality. For iteration k, we assume $\frac{\tau}{\rho}\|\boldsymbol{\lambda}^k\|^2 \le \mathbf{T}_c^0$. We consider two possible cases for iteration $k+1$, i.e., if $\|\boldsymbol{\lambda}^{k+1}\|^2 \le \|\boldsymbol{\lambda}^k\|^2$, we straightforwardly have $\frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 \le \frac{\tau}{\rho}\|\boldsymbol{\lambda}^k\|^2 \le \mathbf{T}_c^0$, and else if $\|\boldsymbol{\lambda}^{k+1}\|^2 \ge \|\boldsymbol{\lambda}^k\|^2$, we also have $\frac{\tau}{\rho}\|\boldsymbol{\lambda}^{k+1}\|^2 \le \mathbf{T}_c^0$ by (30). We therefore conclude $\|\boldsymbol{\lambda}^*\|^2 \le \rho\tau^{-1}\mathbf{T}_c^0$.

ii) Prove $\mathbf{T}_c^0 \le \big(4 + c(1-2\tau^2) + c/2\big)d_F + cL_g/2\|\mathbf{x}^0\|^2 + c\rho_F/4\ d_{\mathbf{x}}$: Invoking **Prop.** 2 and set $k = 0$, we have

$$\begin{aligned}
&\mathbb{L}_\rho(\mathbf{x}^1, \boldsymbol{\lambda}^1) - \frac{\tau}{2\rho}\|\boldsymbol{\lambda}^1\|^2 \le \mathbb{L}_\rho(\mathbf{x}^0, \boldsymbol{\lambda}^0) - \frac{\tau}{2\rho}\|\boldsymbol{\lambda}^0\|^2 \\
&- \|\mathbf{x}^1 - \mathbf{x}^0\|_{\mathbf{Q}}^2 + \frac{\rho_F}{2}\|\mathbf{x}^1 - \mathbf{x}^0\|^2 - \frac{\rho}{2}\|\mathbf{A}(\mathbf{x}^1 - \mathbf{x}^0)\|^2 \\
&+ \frac{2-\tau}{2\rho}\|\boldsymbol{\lambda}^1 - \boldsymbol{\lambda}^0\|^2.
\end{aligned}$$

By invoking (25), $\boldsymbol{\lambda}^0 = 0$, $\mathbf{A}\mathbf{x}^0 = b$, $\mathbf{Q} := \rho G_{\mathbf{A}} + \beta G_{\mathbf{B}} - \rho\mathbf{A}^\top\mathbf{A}$ and set $\boldsymbol{\lambda}^{-1} = 0$, we have

$$\begin{aligned}
&\frac{\rho}{2}\|\mathbf{A}\mathbf{x}^1 - \mathbf{b}\|^2 + \frac{2\mathbf{Q} + \rho\mathbf{A}^\top\mathbf{A} - \rho_f\mathbf{I}_N}{2}\|\mathbf{x}^1 - \mathbf{x}^0\|^2 \\
&\le f(\mathbf{x}^0) + g(\mathbf{x}^0) - f(\mathbf{x}^1) - g(\mathbf{x}^1)
\end{aligned}$$

Since we have $f(\mathbf{x}) \ge 0$ and $f(\mathbf{x}) \ge 0$ over the $\mathbf{X}$, we have (the term $\frac{\rho\mathbf{A}^\top\mathbf{A}}{2}\|\mathbf{x}^1 - \mathbf{x}^0\|^2$ is non-negative)

$$\frac{\rho}{2}\|\mathbf{A}\mathbf{x}^1 - \mathbf{b}\|^2 \le d_F. \tag{31}$$

$$\begin{aligned}
&\frac{2\mathbf{Q} - \rho_F\mathbf{I}_N}{2}\|\mathbf{x}^1 - \mathbf{x}^0\|^2 \le d_F \\
\Rightarrow\ &\|\mathbf{x}^1 - \mathbf{x}^0\|_{\mathbf{Q}}^2 \le d_F + \rho_F/2 d_{\mathbf{x}}.
\end{aligned} \tag{32}$$

where the last inequality is by $d_{\mathbf{x}} := \max_{\mathbf{x},\mathbf{y}} \|\mathbf{x} - \mathbf{y}\|^2$.

Further, based on the dual update, we have

$$\frac{1}{2\rho}\|\boldsymbol{\lambda}^1\|^2 = \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^1 - \mathbf{b}\|^2 \le d_F \tag{33}$$

Further, we bound $\mathbf{T}_c^0$ and we have

$$\begin{aligned}
\mathbf{T}_c^0 &= f(\mathbf{x}^1) + g(\mathbf{x}^1) + \frac{2 + c(1-2\tau^2)}{2\rho}\|\boldsymbol{\lambda}^1\|^2 + \frac{\rho}{2}\|\mathbf{A}\mathbf{x}^1 - \mathbf{b}\|^2 \\
&\quad + \frac{c}{2}\|\mathbf{x}^1 - \mathbf{x}^0\|_{\mathbf{Q}}^2 + \frac{cL_g}{2}\|\mathbf{x}^0\|^2 \ \text{ by (29) and } \boldsymbol{\lambda}^0 = 0 \\
&\le\ d_F + (2 + c(1-2\tau^2))d_F + d_F \\
&\qquad + \frac{c}{2}d_F + \frac{c\rho_F}{4}d_{\mathbf{x}} + \frac{cL_g}{2}\|\mathbf{x}^0\|^2 \ \text{ by (31), (32), (33)} \\
&=\ \big(4 + c(1-2\tau^2) + c/2\big)d_F + cL_g/2\|\mathbf{x}^0\|^2 + c\rho_F/4\ d_{\mathbf{x}}
\end{aligned}$$

Based on **Theorem** 1 and i), ii), we therefore have

$$\operatorname{dist}\big(\nabla f(\mathbf{x}^*) + \nabla g(\mathbf{x}^*) + \mathbf{A}^\top\boldsymbol{\lambda}^* + N_{\mathbf{X}}(\mathbf{x}^*), 0\big) + \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|^2 \le \epsilon,$$

which thus close the proof.

5 Numeric Experiments

5.1 A numeric example

We first consider a numerical example with $N = 2$ agents given by

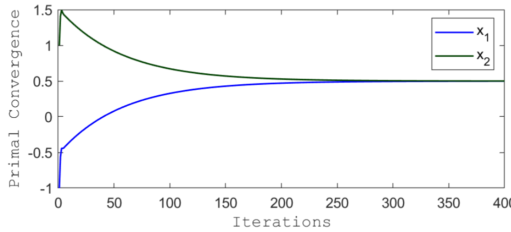
$$\begin{aligned} \min\ & 0.1x_1^3 + 0.1x_2^3 + 0.1x_1x_2 \\ \text{s.t. } & x_1 + x_2 = 1 \\ & -1 \leq x_1 \leq 1 \\ & -1 \leq x_2 \leq 1 \end{aligned} \tag{34}$$

For this example, we have $f_1(x_1) = 0.1x_1^3$, $f_2(x_2) = 0.1x_2^3$, and $g(x_1, x_2) = 0.1x_1x_2$. The Lipschitz continuous gradient modulus for f and g are $L_f = 0.6$ and $L_g = 0.2$. Besides, we have $\mathbf{A}_1 = 1$, $\mathbf{A}_2 = 1$, $\mathbf{A} = (1\ 1)$. The stationary point of the problem is $x_1^\star = 0.5, x_2^\star = 0.5$. We apply the proposed regulated ADMM to solve this problem in a distributed manner. The configurations of **Algorithm** 1 are $\mathbf{B}_1 = \mathbf{B}_2 = 1$, $\tau = 0.1$, $\rho = 5$, $\beta = 6$, $x_1^0 = 0$, $x_2^0 = 0$ and $\lambda^0 = 0$. Before starting the algorithm, we first examine the convergence conditions as stated in **Theorem** 1. For this example, we have $G_{\mathbf{A}} = G_{\mathbf{B}} = \mathbf{I}_2$ and $\mathbf{Q} = [6, -5; -5, 6] > 0$. We select $c = 8.7$. We therefore have $a_{\mathbf{x}} = [2.98, -2.5; -2.5, 2.98] > 0,$ and $a_\lambda = 0.0014 > 0$. This justifies the sufficiently decreasing property of the Lyapunov function.

Next we evaluate the numeric convergence of regulated ADMM for this example. We run **Algorithm** 1 sufficiently long ($K = 400$ iterations). We observe the method converges to $x_1^* = 0.4981$ and $x_2^* = 0.4997$. The relative (coupled) `constraints residual` measured by $\|\mathbf{Ax} - \mathbf{b}\|/\mathbf{b}$ is about 0.22%. We display the evolution of the primal variables x_1^k and x_2^k as well as the Lyapunov function $\mathbf{T}_c^k$ over the iterations in Fig. 1. We note that x_1^k and x_2^k are convergent and approximately approaching the stationary points $x_1^\star = 0.5$ and $x_2^\star = 0.5$. Besides, the `Lyapunov function` decreases w.r.t. the iterations and finally stabilizes. This is consistent with our theoretical analysis in Section III. We further compare the proposed regulated ADMM (`Reg-ADMM`) with the proximal ADMM (`Prox-ADMM`) established for convex problems [16] (the composite objective components are also handled by the linearization technique). For fair comparison, we set the same ρ and β for the two algorithms. We run both algorithms suitably long ($K = 400$ iterations). We finally obtain the reports in Table 2. Note that both algorithms yield desirable stationary solutions for this example, i.e., the `sub-optimality` measured by $\|\mathbf{x}^\star - \mathbf{x}^*\|/\|\mathbf{x}^\star\|$ are about 0.16% (`Prox-ADMM`) versus 0.27% (`Reg-ADMM`). However, only the proposed `Reg-ADMM` provides theoretical convergence guarantee but at the cost of a minor `constraints residual` 0.22%.

Table 2
Prox-ADMM vs. Reg-ADMM (N: No, Y: Yes)

Method	x_1^*	x_2^*	Sub-optimality	Constraints residual	Convergence guarantee
Prox-ADMM	0.4992	0.5008	0.16%	0	N
Reg-ADMM	0.4981	0.4997	0.27%	0.22%	Y



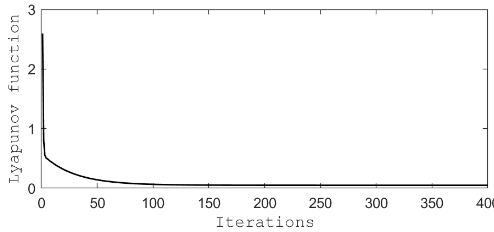


Fig. 1. (a) The evolution of primal variables x_1^k and x_2^k. (b) The evolution of the Lyapunov function $\mathbf{T}_c^k$.

5.2 Application: multi-zone HVAC control

To showcase the performance of regulated ADMM, this section presents the application to multi-zone heating, ventilation, and air conditioning (HVAC) control arising from smart buildings. The goal is to optimize the HVAC operation to provide the comfortable temperature with minimal electricity bill. Due to the thermal capacity of buildings, the evolution of indoor temperature is a slow process affected both by the dynamic indoor occupancy (thermal loads) and the HVAC operation (cooling loads). The general solution is to design a model predictive controller for optimizing HVAC operation (i.e., zone mass flow, zone temperature trajectories) to minimize the overall electricity cost while respecting the comfortable temperature ranges based on the predicted information (i.e., indoor occupancy, outdoor temperature, electricity price, etc.). The general problem formulation is presented as below.

$$\begin{aligned} \min_{\mathbf{m}^z, \mathbf{T}} & \sum_t c_t \{ c_p(1-d_r) \sum_i m_t^{zi}(T_t^o - T^c) \\ & + c_p \eta d_r \sum_i m_t^{zi}(T_t^i - T^c) + \kappa_f \big(\textstyle\sum_i m_t^{zi} \big)^2 \} \Delta_t \end{aligned} \tag{P1}$$

$$\begin{aligned} \text{s.t. } \ T_{t+1}^i = & A_{ii}T_t^i + \textstyle\sum_{j \in N_i} A_{ij} T_t^j \\ & + C_{ii} m_t^{zi}(T_t^i - T^c) + D_t^{ii}, \ \forall i, t. \end{aligned} \tag{35a}$$

$$T_{\min}^i \leq T_t^i \leq T_{\max}^i, \quad \forall i, t. \tag{35b}$$

$$m_{\min}^{zi} \leq m_t^{zi} \leq m_{\max}^{zi}, \ \forall i, t. \tag{35c}$$

$$\textstyle\sum_i m_t^{zi} \leq \overline{m}, \ \forall t. \tag{35d}$$

where i and t are zone and time indices. $\mathbf{T} = \{T_t^i\}_{i,t}$ and $\mathbf{m}^z = \{m_t^{zi}\}_{i,t}$ are zone temperature and zone mass flow rates, which are decision variables. The other notations are constant parameters. The main task is to optimize the zone mass flow rate to provide the temperature trajectories within the comfortable ranges $[T_{\min}^i, T_{\max}^i]$ with the minimal electricity cost measured by the objective.

The problem is subject to the constraints covering zone thermal dynamics (35a), comfortable temperature margins (35b), zone mass flow rate limits (35c), and total zone mass flow rate limits (35d).

For the multi-zone HVAC control, centralized strategies are generally not suitable due to the computation and communication overheads, and distributed methods have been regarded as desirable solutions. However, the non-convexity makes it challenging to develop a distributed mechanism which can enable zone-level computation while still achieving the coordination among the zones to minimize the overall cost. This section demonstrates that the proposed method can work as an effective distributed solution. Before we show the results, we first restate the problem in the standard format:

$$\begin{aligned}
&\min_{\mathbf{m}^z, \mathbf{T}} \sum_t c_t \big\{ c_p(1-d_r) \sum_i m_t^{zi}(T_t^o - T^c) && (\mathbf{P2})\\
&+ c_p \eta d_r \sum_i m_t^{zi}(T_t^{ii} - T^c) + \kappa_f \big(\textstyle\sum_i m_t^{zi} \big)^2 \big\} \Delta_t \\
&+ M \textstyle\sum_i \sum_t \big(T_{t+1}^{ii} - A_{ii} T_t^{ii} - \sum_{j \in N_i} A_{ij} T_t^{ij} \\
&\quad - C_{ii} m_t^{zi}(T_t^{ii} - T^c) - D_t^{ii} \big)^2 \\
&\text{s.t. } \; T_t^{ij} = \overline{T}_t^j, \quad \forall i, j, t. && (36a)\\
&\quad T_{\min}^i \leq T_t^{ii} \leq T_{\max}^i, \quad \forall i, t. && (36b)\\
&\quad T_{\min}^i \leq \overline{T}_t^j \leq T_{\max}^i, \quad \forall i, t. && (36c)\\
&\quad m_{\min}^{zi} \leq m_t^{zi} \leq m_{\max}^{zi}, \quad \forall i, t. && (36d)\\
&\quad \textstyle\sum_i m_t^{zi} \leq \overline{m}, \; \forall t. && (36e)
\end{aligned}$$

where we have augmented the decision component for each zone to involve the copy of temperature for its neighboring zones, i.e., $\mathbf{T}^i := \{T_t^{ij}\}_{j \in \mathbf{N}_i, t}$. Besides, to drive the consistence of zone temperature, we introduce a block of consensus variable $\overline{\mathbf{T}} = \{\overline{T}_t^j\}_{j,t}$. Considering the challenging to handle the hard non-linear constraints (35a), we employ the penalty method and penalize the violations of constraints with quadratic terms. In this regard, problem (**P2**) fits into the template (**P**). Particularly, we have $N+1$ computing agents, where agents 1 to N correspond to the zones with the augmented decision variable $\mathbf{x}_i = (\{T_t^{ij}\}_{j \in N_i, t}, \{m_t^{zi}\}_t)$, and agent 0 control the consensus decision variable $\overline{\mathbf{T}} := \{\overline{T}_j\}_{j \in N}$. Constraints (36a) and (36e) represents the coupled linear constraints which can be expressed in the compact form $\mathbf{A}\mathbf{x} = \mathbf{b}$ if necessary. The other constraints comprise the local bounded convex constraints for the agents.

We consider a case study with $N = 10$ zones and the predicted horizon is set as $T = 48$ time slots (one day with a sampling interval of 30 mins). We set the lower and upper comfortable temperature bounds as $T_{\min}^i = 24°$C and $T_{\max}^i = 26°$C. The specifications for HVAC system can refer to [1, 2]. We apply the proposed regulated ADMM to solve this problem in a distributed manner. The algorithm configurations are $\rho = 2.0$, $\tau = 0.1$, $\beta = 3.0$, $\mathbf{B}_i = \mathbf{I}$ (suitable sizes), and $c = 8.7$. We first examine the numeric convergence of the algorithm measured by the `Lyapunov function` and the norm of (coupled) `constraints residual`. We run the algorithm suitably long when both the residual and Lyapunov function do not change apparently ($K = 200$ iterations for this example). We visualize the `Lyapunov function` and `constraints residual` in Fig. 2. Note that the `Lyapunov function` declines rapidly along the iterations, which is consistent with our theoretical analysis. Besides, the `constraints residual` is almost strictly decreasing toward *zero* along the iterations. We find the overall norm of the `constraints residual` at the end of iterations is about 0.38, which is quite small considering the problem scale $T \cdot N = 480$. This justifies the convergence property of regulated ADMM.

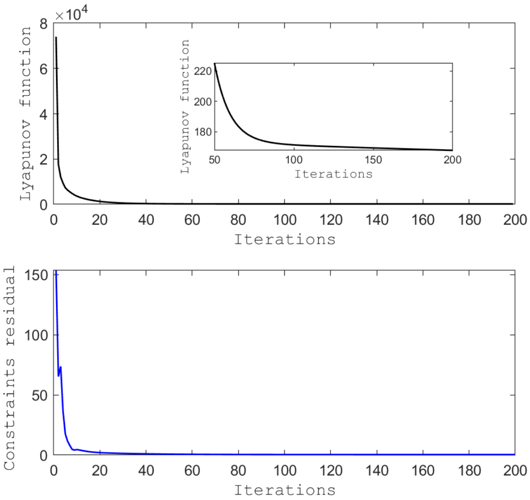


Fig. 2. (a) The evolution of Lyapunov function. (b) The evolution of the norm of constraints residual.

Table 3
Centralized vs. Reg-ADMM (Y: Yes)

Method	Electricity cost (s$)	Human comfort	Constraints residual	Computing time
Centralized	160.54	Y	0.38	50 min
Reg-ADMM	153.12	Y	0	$\geq$ 10h

We next evaluate the solution quality measured by HVAC electricity cost and human comfort. We randomly pick 3 zones (zone 1, zone 3, and zone 7) and display the predicted `zone occupancy` (inputs), the zone mass flow rates (`zone MFR`, control variables) , and the zone temperature (`zone temp.`, control variables) over the $T = 48$ time slots in Fig. 3. We note that the variation of `zone MFR` is almost consistent with the `zone occupancy`. This is reasonable as the `zone occupancy` determines the thermal loads which need to be balanced by the supplied cooling air. We see that the `zone temp.` are all maintained within the comfortable range $[24, 26]°$C. This infers the satisfaction of `human comfort`. To further evaluate the solution quality and computation efficiency, we compare the proposed regulated ADMM (`Reg-ADMM`) with centralized method (`Centralized`). Specifically, the centralized method solves the problem directly with the `fmincon` solver embodied in MATLAB without considering the running time. We compare the two methods in three folds, i.e., `electricity cost`, the norm of `constraints residual`, and `computation time` as reported in Table 3. We see that `electricity cost` under the `Reg-ADMM` is about 160.20 (s$) versus 153.12 (s$) yield by `Centralized`. This infers the sub-optimality of `Reg-ADMM` in terms of the objective is about 5.0%. Particularly, we observe a marginal `constraints residual` (0.38) for `Reg-ADMM` but not with the `Centralized`. This is consistent with our theoretical analysis and caused by the discounted factor $1-\tau$. However, we see that the `Reg-ADMM` is obviously advantageous over the `Centralized` in computation efficiency. The average `computing time` for each zone is about 50 min with `Reg-ADMM` (parallel computation) while the `Centralized` takes more than 10 h. Note that we have picked $T=48$ time slots (a whole day) as the predicted horizon, the computing time could be largely sharpened in practice with a much smaller prediction horizon, say $T=10$ time slots (5h). This is to our expectations as `Reg-ADMM` empowers the agents to solve small subproblems in parallel instead of relying on a central agent solving the overall heavy problem as with the `Centralized`.

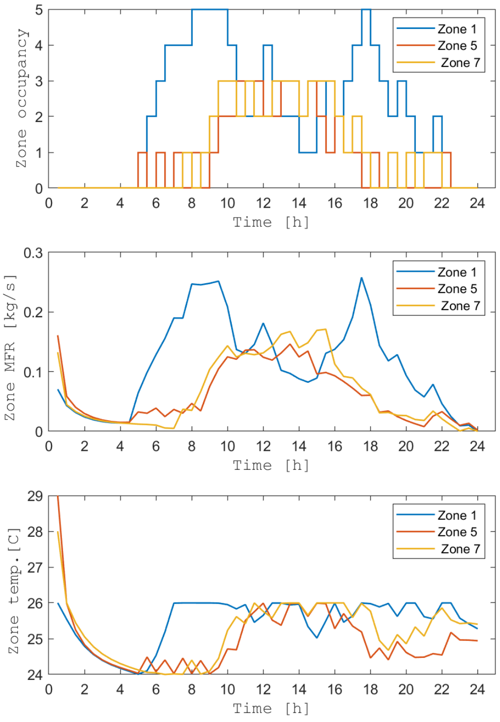


Fig. 3. (a) zone occupancy. (b) zone mass flow rate (zone MFR). (c) zone temperature (zone temp.).

6 Conclusion

This paper focused on developing a distributed algorithm for a class of structured nonconvex problems with convergence guarantee. The problems are featured by i) a possibly nonconvex objective composed of both separate and composite components, ii) local bounded convex constraints, and iii) global coupled linear constraints. This class of problems are broad in application but lack distributed solutions with convergence guarantee. We employed the powerful alternating direction method of multiplier (ADMM) tool for constrained optimization but faced the challenges to establish the convergence. Noting that the underlying obstacle is to assume the boundness of dual updates, we revised the classic ADMM and proposed to regulate the dual update procedure. This leads to a `regulated ADMM` with the convergence guarantee towards approximate stationary points of the problem. We demonstrated the convergence and solution quality of the distributed method by a numeric example and a concrete application to the multi-zone heating, ventilation, and air-condition (HVAC) control arising from smart buildings.

References

[1] Y. Yang, G. Hu, and C. J. Spanos, "Hvac energy cost optimization for a multizone building via a decentralized approach," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 4, pp. 1950–1960, 2020.

[2] Y. Yang, S. Srinivasan, G. Hu, and C. J. Spanos, "Distributed control of multizone hvac systems considering indoor air quality," *IEEE Transactions on Control Systems Technology*, 2021.

[3] J. A. Ansere, G. Han, L. Liu, Y. Peng, and M. Kamal, "Optimal resource allocation in energy-efficient internet-of-things networks with imperfect csi," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5401–5411, 2020.

[4] L. Zhang, V. Kekatos, and G. B. Giannakis, "Scalable electric vehicle charging protocols," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1451–1462, 2016.

[5] A. Falsone, I. Notarnicola, G. Notarstefano, and M. Prandini, "Tracking-admm for distributed constraint-coupled optimization," *Automatica*, vol. 117, p. 108962, 2020.

[6] M. K. Arpanahi, M. H. Golshan, and P. Siano, "A comprehensive and efficient decentralized framework for coordinated multiperiod economic dispatch of transmission and distribution systems," *IEEE Systems Journal*, 2020.

[7] S. Hashempour, A. A. Suratgar, and A. Afshar, "Distributed nonconvex optimization for energy efficiency in mobile ad hoc networks," *IEEE Systems Journal*, 2021.

[8] I. Necoara and V. Nedelcu, "On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems," *Automatica*, vol. 55, pp. 209–216, 2015.

[9] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization with coupling constraints," *Automatica*, vol. 84, pp. 149–158, 2017.

[10] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.

[11] M. R. Hestenes, "Multiplier and gradient methods," *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.

[12] M. J. Powell, "A method for nonlinear constraints in minimization problems," *Optimization*, pp. 283–298, 1969.

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

[14] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of admm for multi-block convex minimization problems is not necessarily convergent," *Mathematical Programming*, vol. 155, no. 1-2, pp. 57–79, 2016.

[15] T.-Y. Lin, S.-Q. Ma, and S.-Z. Zhang, "On the sublinear convergence rate of multi-block admm," *Journal of the Operations Research Society of China*, vol. 3, no. 3, pp. 251–274, 2015.

[16] W. Deng, M.-J. Lai, Z. Peng, and W. Yin, "Parallel multi-block admm with o (1/k) convergence," *Journal of Scientific Computing*, vol. 71, no. 2, pp. 712–736, 2017.

[17] N. Chatzipanagiotis, D. Dentcheva, and M. M. Zavlanos, "An augmented lagrangian method for distributed optimization," *Mathematical Programming*, vol. 152, no. 1, pp. 405–434, 2015.

[18] N. Chatzipanagiotis and M. M. Zavlanos, "On the convergence of a distributed augmented lagrangian method for nonconvex optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4405–4420, 2017.

[19] Y. Wang, W. Yin, and J. Zeng, "Global convergence of admm in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.

[20] L. Yang, T. K. Pong, and X. Chen, "Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 74–110, 2017.

[21] K. Guo, D. Han, and T.-T. Wu, "Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints," *International Journal of Computer Mathematics*, vol. 94, no. 8, pp. 1653–1669, 2017.

[22] G. Li and T. K. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2434–2460, 2015.

[23] 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*, vol. 26, no. 1, pp. 337–364, 2016.

[24] K. Sun and X. A. Sun, "A two-level distributed algorithm for general constrained non-convex optimization with global convergence," *arXiv preprint arXiv:1902.07654*, 2019.

[25] K. Sun and X. A. Sun, "A two-level admm algorithm for ac opf with convergence guarantees," *IEEE Transactions on Power Systems*, 2021.

[26] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the admm in decentralized consensus optimization," *IEEE Transactions on Signal Processing*, vol. 62, no. 7, pp. 1750–1761, 2014.

[27] T. Erseghe, "Distributed optimal power flow using admm," *IEEE transactions on power systems*, vol. 29, no. 5, pp. 2370–2380, 2014.

[28] T. Baroche, P. Pinson, R. L. G. Latimier, and H. B. Ahmed, "Exogenous cost allocation in peer-to-peer electricity markets," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 2553–2564, 2019.

[29] G. Taylor, R. Burmeister, Z. Xu, B. Singh, A. Patel, and T. Goldstein, "Training neural networks without gradients: A scalable admm approach," in *International conference on machine learning*, pp. 2722–2731, PMLR, 2016.

[30] X. Li, G. Feng, and L. Xie, "Distributed proximal algorithms for multiagent optimization with coupled inequality constraints," *IEEE Transactions on Automatic Control*, vol. 66, no. 3, pp. 1223–1230, 2020.