2025 23rd European Control Conference (ECC)
June 24-27, 2025. Thessaloniki, Greece

Convergence Theory of Flexible ALADIN for Distributed Optimization

Xu Du*, Xiaohua Zhou, Shijie Zhu and Apostolos I. Rikos

***Abstract*— The Augmented Lagrangian Alternating Direction Inexact Newton (ALADIN) method is a cutting-edge distributed optimization algorithm known for its superior numerical performance. It relies on each agent transmitting information to a central coordinator for data exchange. However, in practical network optimization and federated learning, unreliable information transmission often leads to packet loss, posing challenges for the convergence analysis of ALADIN. To address this issue, this paper proposes Flexible ALADIN, a random polling variant of ALADIN, and presents a rigorous convergence analysis, including global convergence for convex problems and local convergence for non-convex problems.**

I. Introduction

In recent years, distributed optimization has gained significant attention, driven by advancements in machine learning [29], model predictive control [18] and optimal power flow [11]. These applications, modeled through distributed optimization, can be broadly categorized into two main types: a) *distributed resource allocation optimization* [18, Section 2], shown as Problem (1),

$$\begin{aligned} \min_{x_i\in\mathbb{R}^{n_i}} \quad & \sum_{i=1}^N f_i(x_i) \\ \text{s.t.} \quad & \sum_{i=1}^N A_i x_i = b; \end{aligned} \tag{1}$$

b) *distributed consensus optimization* [3, Chapter 7] shown as Problem (2),

$$\begin{aligned} \min_{x_i,y\in\mathbb{R}^{n}} \quad & \sum_{i=1}^N f_i(x_i) \\ \text{s.t.} \quad & x_i = y. \end{aligned} \tag{2}$$

Here $f_i:\mathbb{R}^{n_i}\to\mathbb{R}$ for Problem (1) and $f_i:\mathbb{R}^{n}\to\mathbb{R}$ for Problem (2). In the first type, we minimize the sum of separable objective functions with linear coupling relations [3], [17]. Here the local decision variables x_is are linearly coupled with the given matrices $A_i\in\mathbb{R}^{m\times n_i}$s and the vector $b\in\mathbb{R}^m$. Unlike the first type, the second type (2) introduces a global variable $y\in\mathbb{R}^n$, requiring each agent's local variable $x_i\in\mathbb{R}^n$ to reach consensus with it.

To solve the problems in (1) and (2), distributed optimization algorithms distribute data across multiple agents in the network. Two primary approaches are commonly employed: (a) primal decomposition and (b) dual decomposition [22]. This paper focuses on a class of algorithms derived from dual decomposition, specifically the Augmented Lagrangian Alternating Direction Inexact Newton (ALADIN) method [17], [9]. Here, Typical ALADIN (T-ALADIN) [17], [18] focuses on solving Problem (1), while Consensus ALADIN (C-ALADIN) [8], [10], [9] focuses on solving Problem (2). In detail, T-ALADIN consists of the following updates [18],

$$\left\{\begin{array}{l} x_i^+ = \operatorname*{argmin}_{x_i} f_i(x_i) + \lambda^\top A_i x_i + \frac{1}{2}\|x_i - y_i\|^2_{B_i}; \\ (\Delta y, \lambda^+) \\ = \left\{\begin{array}{cc} \operatorname*{argmin}_{\Delta y_i\in\mathbb{R}^{|x_i|},\forall i} & \sum_{i=1}^N \left(\frac{1}{2}\Delta y_i^\top B_i \Delta y_i + g_i^\top \Delta y_i\right) \\ \text{s.t.} & \sum_{i=1}^N A_i\left(x_i^+ + \Delta y_i\right) = b \mid \lambda \end{array}\right\}; \\ y_i = x_i^+ + \Delta y_i. \end{array}\right. \tag{3}$$

In the first step, each agent updates its local variable x_i. Notably, the first step of T-ALADIN (3) can also be formulated as

$$\operatorname*{argmin}_{x_i} f_i(x_i) + \lambda^\top A_i x_i + \frac{\rho}{2}\|x_i - y_i\|^2,$$

where $\rho > 0$, simplifying the setting of the proximal term [17]. The second step updates the dual variable λ by solving a coupled quadratic programming (QP) problem that relates to global variables y_is, where λ corresponds to the affine-coupled constraints in (1). The term $B_i \approx \nabla^2 f_i(x_i^+) \succ 0$ represents a local Hessian approximation of f_i in (1), while $g_i = B_i(y_i - x_i^+) - A_i^\top\lambda$ provides the (sub)gradient of f_i. Similarly, the update process of C-ALADIN [9] is summarized as follows,

$$\left\{\begin{array}{l} x_i^+ = \arg\min_{x_i} f_i(x_i) + \lambda_i^\top x_i + \frac{1}{2}\|x_i - y\|^2_{B_i}; \\ (y, \Delta y, \lambda^+) \\ = \left\{\begin{array}{cc} \operatorname*{argmin}_{\Delta y_i\in\mathbb{R}^{|x_i|},\forall i} & \sum_{i=1}^N \left(\frac{1}{2}\Delta y_i^\top B_i \Delta y_i + g_i^\top \Delta y_i\right) \\ \text{s.t.} & x_i^+ + \Delta y_i = y \mid \lambda_i \end{array}\right\}. \end{array}\right. \tag{4}$$

The main difference between T-ALADIN and C-ALADIN is that C-ALADIN coordinates the information of all agents by solving a *consensus QP*. Further details can be found in [8]. Notably, the statements of T-ALADIN for solving (1) also hold for C-ALADIN when solving (2).

T-ALADIN and C-ALADIN exhibit excellent numerical performance, ensuring global convergence for convex prob-

Xu Du and Apostolos I. Rikos are with the Artificial Intelligence Thrust of the Information Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. Apostolos I. Rikos is also affiliated with the Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China. E-mails: duxu@hnas.ac.cn; apostolosr@hkust-gz.edu.cn.

Xiaohua Zhou is with ShanghaiTech University. E-mail: zhouxh3@shanghaitech.edu.cn.

Shijie Zhu is with China Telecom Corporation Ltd. Shanghai Branch. E-mail: zhushj@alumni.shanghaitech.edu.cn.

The work of X.D. and A.I.R. was supported by the Guangzhou-HKUST(GZ) Joint Funding Scheme (Grant No. 2025A03J3960).

lems. Additionally, they provide local convergence guarantees for non-convex problems when Linear Independence Constraint Qualification (LICQ) and Second-Order Sufficiency Conditions (SOSC) are satisfied, as shown in [18], [9]. Importantly, [10] establishes the global convergence theory of C-ALADIN for non-convex consensus problems, see (2). However, the ALADIN framework faces significant challenges due to its reliance on data synchronization at each iteration, which limits its scalability. Specifically, in the context of distributed optimization, to the best of our knowledge, no existing work has provided a rigorous convergence analysis of ALADIN under packet loss during network transmission. Furthermore, in federated learning, each agent is selected with a certain probability in each iteration to update its local optimization variables and upload them to the server, which can partially mitigate security risks. Notably, FedALADIN, a variant of C-ALADIN, represents the first attempt to integrate C-ALADIN with federated learning and enhance security—demonstrating competitive performance against classical federated learning algorithms. However, its theoretical foundations remain insufficiently explored. To overcome these synchronization-related challenges, this paper introduces random polling variants of ALADIN-type algorithms, providing a more scalable and resilient alternative.

A. Related Work

Since ALADIN is built upon ADMM (Alternating Direction Method of Multipliers) and SQP (Sequential Quadratic Programming) [17], [9], we primarily review the related work on these two types of algorithms within the context of asynchronous optimization.

ADMM with asynchronous update structure has been proposed by many authors. Interestingly, there are two main names for related work, including *Flexible (Randomized) ADMM* [16], [25], [26], [29], [24], [14], [5] and *Asynchronous ADMM* [28], [6], [21], [27], [15], [23], [1], to name a few. In the aforementioned literature, to the best of our knowledge, [27] and [28] were the first to propose the Consensus ADMM for solving (2) with the asynchronous structure. Importantly, [16] provided the convergence analysis of Flexible ADMM in the non-convex cases, which was further developed by [6], [15], [26]. Additionally, [29] applied Flexible ADMM in the federated learning while optimal control scenario has also been covered by [21].

Compared with ADMM, SQP is rarely studied in asynchronous contexts. We found that [20] provided a local convergence analysis with the random polling variant of SQP. Interestingly, [19] proposed the first asynchronous version of T-ALADIN, however, it is applicable only to *tree-structure* problems.

B. Contributions

In this paper, we propose random polling variants for both T-ALADIN and C-ALADIN in Section II to address packet loss in unstable networks, namely Flexible Typical ALADIN (FT-ALADIN) and Flexible Consensus ALADIN (FC-ALADIN), respectively. Importantly, we provide the convergence theory of FC-ALADIN in Section III. The related convergence theory papers for ALADIN type algorithms are listed in Table I.

TABLE I

CONVERGENCE ANALYSIS OF ALADIN

Attribute	Smooth	Non-smooth
Convex	[18], [8],[10], **this paper**	[18], [8],[10], **this paper**
Non-convex	[17], [13], [7], [19], [8],[10] **this paper**	[10], **this paper**

Notation: In this paper, $(\cdot)^-$ denotes the previous value while $(\cdot)^+$ represents the current value. For ease of expression, $(\cdot)^k$ indicates the value of $(\cdot)$ at the k-th iteration for the given algorithms.

II. FLEXIBLE ALADIN

In this section, we propose Flexible ALADIN for distributed optimization problems. In details, FT-ALADIN is proposed for solving (1), as shown in Algorithm 1, and FC-ALADIN for solving (2), as shown in Algorithm 2.

Algorithm 1 Flexible Typical ALADIN (FT-ALADIN)

```
Initialization: Initial the global dual variable $\lambda$ and the local primal variables $y_i$s. Set $B_i \succ 0$.
Repeat:
  1) Agents update:
     For $i \in \mathcal{C}^+$, do:
     a) Update the local variable $x_i$:
          $x_i^+ = \operatorname*{argmin}_{x_i} f_i(x_i) + \lambda^\top A_i x_i + \frac{1}{2}\|x_i - y_i\|_{B_i}^2.$
     b) Evaluate the new Hessian[1] $B_i \succ 0$ and (sub)gradient:
          $g_i^+ = \rho(y_i - x_i^+) - A_i^\top \lambda.$
     For $i \notin \mathcal{C}^+$, set $x_i^+ = x_i, B_i^+ = B_i, g_i^+ = g_i$.
  2) Coordination:
     a) Evaluate the dual gradient and dual Hessian:
          $\begin{cases} R = \sum_{i=1}^{N} A_i \left(x_i^+ - \left(B_i^+\right)^{-1} g_i^+\right) - b, \\ M = \sum_{i=1}^{N} A_i \left(B_i^+\right)^{-1} A_i^\top. \end{cases}$   (5)
     b) Evaluate the dual variable $\lambda$ and update the primal variables $y_i$s:
          $\begin{cases} \lambda^+ = M^{-1} R, \\ y_i^+ = x_i^+ - \left(B_i^+\right)^{-1} \left(g_i^+ + A_i^\top \lambda^+\right). \end{cases}$   (6)
```

Here, we assume the agent i is randomly chosen with probability p at iteration k for an active set $\mathcal{C}^k \subseteq \{1, \cdots, N\}$, such that

$$1 \geq \mathbb{P}\left(i \in \mathcal{C}^k\right) = p > 0. \tag{7}$$

[1] Notably, depending on the applications, B_i can be approximated with various methods, i.e. BFGS [12], [8] update or Gauss-Newton Hessian approximation [7].

If $p = 1$ at every iteration, Algorithm 1 reduces to T-ALADIN [17], and Algorithm 2 reduces to C-ALADIN [8]. Notice that, for the following two algorithms, each agent is updated at least once during the total K iterations, such that $i \in \bigcup_{k=1}^{K} \mathcal{C}^k$. There are two main steps in Algorithm 1 and 2: the parallelizable steps from the agent side and the coordination steps from the master side. For clarity in the following algorithmic structure, we define the active set at each iteration as $\mathcal{C}^+$, eliminating the need to explicitly reference the iteration index k.

```
Algorithm 2 Flexible Consensus ALADIN (FC-ALADIN)
Initialization: Initial the global variable $z$, the dual variables $\lambda_i$s. Set $B_i \succ 0$.
Repeat:
  1) Agents update:
     For $i \in \mathcal{C}^+$, do:
     a) Update the local variable $x_i$s:
        $x_i^+ = \arg\min_{x_i} f_i(x_i) + \lambda_i^\top x_i + \frac{1}{2}\|x_i - y\|_{B_i}^2.$   (8)
     b) Evaluate the Hessian and the (sub)gradient:
        $\begin{cases} B_i^+ \approx \nabla^2 f_i(x_i^+) \succ 0, \\ g_i^+ = B_i(y - x_i^+) - \lambda_i. \end{cases}$   (9)
     For $i \notin \mathcal{C}^+$, set $x_i^+ = x_i, B_i^+ = B_i, g_i^+ = g_i$.
  2) Coordination:
     a) Update the global variable $y$:
        $y^+ = \left(\sum_{i=1}^{N} B_i^+\right)^{-1} \left(\sum_{i=1}^{N} \left(B_i^+ x_i^+ - g_i^+\right)\right).$   (10)
     b) Evaluate the local dual variables:
        $\lambda_i^+ = B_i(x_i^+ - y^+) - g_i^+.$   (11)
```

In Algorithm 1, the closed-form expressions for the second step of T-ALADIN (3) are given by (5) and (6) (see [18, Section 3.4]). Similarly, in Algorithm 2, (10) and (11) provide the closed-form expressions for the second step of C-ALADIN (4) (see [9]). Notably, *Stochastic SQP* [20] can be viewed as a special case of FC-ALADIN by omitting Equation (8) in Algorithm 2, which is equivalent to setting the coefficient of the proximal term in (8) to infinity (see `https://www.uiam.sk/~oravec/apvv_sk_cn/slides/aladin.pdf`, page 39). In contrast, retaining (8) allows the sub-problem update to support the Consensus QP in collaborative optimization, thereby improving numerical performance. A detailed numerical comparison in [8] evaluates FedALADIN, a variant of FC-ALADIN, against two Consensus ADMM variants, demonstrating the superior numerical stability of FC-ALADIN.

III. Convergence Theory of FC-ALADIN

In this section, we present the convergence theory of Algorithm 2 for both smooth and non-smooth cases. In details, Section III-A establishes the global convergence of FC-ALADIN for convex problems, Section III-B provides a local convergence theory for non-convex cases, Section III-C presents a global convergence analysis for the inexact version of FC-ALADIN. In this section, the probability operator is represented as (12),

$$\mathbb{E}[\cdot] = p(\cdot)^+ + (1-p)(\cdot). \tag{12}$$

Note that, the convergence analysis of Algorithm 1 is similar to that of Algorithm 2 and will be presented in an extended version of this paper.

A. Global Convergence of Exact FC-ALADIN for Strongly Convex Cases

In this subsection, we assume the objectives f_is are smooth and strongly convex. To simplify the proof, we further assume that $B_i \in \mathbb{S}_{++}^n$s are proper, symmetric, and strongly positive definite constant matrices, as similarly required in [18, Section 4.2]. Before we provide the convergence theory, we first introduce the following energy function

$$\mathcal{L}(y, \lambda) = \sum_{i=1}^{N} \left(\|y - y^*\|_{B_i}^2 + \|\lambda_i - \lambda_i^*\|_{B_i^{-1}}^2 \right), \tag{13}$$

where y^* and λ^* denote the optimal solution of (2).

Theorem 1 *Let the local objectives f_is in Problem* (2) *be closed, proper, smooth, and strongly convex. Let $B_i \in \mathbb{S}_{++}^n$ be proper, symmetric, and strictly positive definite constant matrices. Define $x_i^* = y^*$ and λ_i^* as the optimal primal and dual solutions of* (2). *Given an initial point (y^1, λ^1), there always exists a $\delta > 0$ such that FC-ALADIN ensures the following contraction property,*

$$\mathbb{E}\left[\mathcal{L}(y^k, \lambda^k)\right] \leq \alpha^{k-1} \mathcal{L}\left(y^1, \lambda^1\right), \tag{14}$$

where $\alpha = \left(\frac{p}{1+\delta} + (1-p)\right) < 1$.

Proof: See Appendix I. ■

B. Local Convergence of Exact FC-ALADIN for Smooth Non-convex Cases

To simplify the convergence proof, we replace Equation (8) with (15) for the update of the local variable x_i, see [9] and [17],

$$x_i^+ = \arg\min_{x_i} f_i(x_i) + \lambda_i^\top x_i + \frac{\rho}{2}\|x_i - y\|^2. \tag{15}$$

The following statement is an extension of [8, Appendix H]. In this subsection, let γ be an upper bound of the Hessian approximation error, such that

$$\gamma \geq \left\|B_i - \nabla^2 f_i(x_i^+)\right\|. \tag{16}$$

Moreover, we define σ such that $\|B_i + \rho I\| > \sigma > 0$.

We establish the local convergence theory of FC-ALADIN for smooth non-convex cases by demonstrating Theorem 2.

Theorem 2 *Let the local objectives f_is of Problem* (2) *be closed, proper, twice continuously differentiable, potentially non-convex. Let the initial point (x^1, y^1, λ^1) be in a neighborhood of the optimal solution (x^*, y^*, λ^*). Let* (17)

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{\sigma}\sum_{i=1}^{N}\left(\rho\|y^k - y^*\| + \|\lambda_i^k - \lambda_i^*\|\right)\right] \\ \geq &\mathbb{E}\left[\sum_{i=1}^{N}\|x_i^{k+1} - y^*\|\right] \end{aligned} \tag{17}$$

be satisfied for all the iterations with Algorithm 2, then

$$\mathbb{E}\left[\frac{\rho N}{\sigma}\|y^k - y^*\| + \frac{1}{\sigma}\sum_{i=1}^{N}\left\|\lambda_i^k - \lambda_i^*\right\|\right] \tag{18}$$

converges linearly with rate $\frac{(\rho+1)\gamma}{\sigma} < 1$.

Proof: See Appendix II. ■

C. Convergence of Inexact FC-ALADIN for Strictly Convex Cases

In some applications, the exact update of (8) may not be desirable. In this subsection, we assume that each agent updates x_i at each iteration based solely on the closed-form expressions approximated from (8). We propose Inexact FC-ALADIN, as described by Equations (19)-(22), to address cases where (8) can not be updated precisely,

$$x_i^+ = y - B_i^{-1}\left(\lambda_i + \partial f_i(y)\right), \forall i \in \mathcal{C}^+, \tag{19}$$

$$g_i^+ = \partial f_i(x_i^+), \forall i \in \mathcal{C}^+, \tag{20}$$

$$y^+ = \left(\sum_{i=1}^{N} B_i\right)^{-1}\left(\sum_{i=1}^{N}\left(B_i\mathbb{E}[x_i^+] - \mathbb{E}[g_i^+]\right)\right), \tag{21}$$

$$\lambda_i^+ = B_i(x_i^+ - y^+) - \partial f_i(x_i^+). \tag{22}$$

Here ∂f_i denotes the (sub)gradient of f_i. Notice that, from the KKT (Karush-Kuhn-Tucker) conditions of the consensus QP in FC-ALADIN, see (10) and (11), (23) is satisfied for all iterations,

$$\sum_{i=1}^{N}\lambda_i^+ = 0, \ \sum_{i=1}^{N}\lambda_i = 0. \tag{23}$$

Moreover, for Inexact FC-ALADIN, we assume

$$\begin{cases} \left\|\sum_{i=1}^{N}\partial f_i(\cdot)\right\| \leq G < \infty, \\ 0 \preceq \Psi_{\text{min}} I \preceq \left(\sum_{i=1}^{N} B_i\right)^{-1} \preceq \Psi_{\text{max}} I \preceq \infty, \end{cases} \tag{24}$$

where $0 < \Psi_{\text{min}} < \Psi_{\text{max}} < \infty$, and define

$$\begin{cases} \varphi_1 = \dfrac{pG\sum_{k=1}^{K}\Psi_{\text{max}}^k\sum_{i=1}^{N}\left\|y^K - x_i^{K+1}\right\|}{2\sum_{k=1}^{K}\Psi_{\text{min}}^k}, \\ \varphi_2 = \dfrac{(1-p)G\sum_{k=1}^{K}\Psi_{\text{max}}^k\sum_{i=1}^{N}\left\|y^{K-1} - x_i^{K}\right\|}{2\sum_{k=1}^{K}\Psi_{\text{min}}^k}. \end{cases} \tag{25}$$

The global convergence of Inexact FC-ALADIN, see Equation (19)-(22), is established by demonstrating the following theorem for strictly convex cases.

Theorem 3 *Let the local objectives f_is of Problem* (2) *be closed, proper, strictly convex. Let the inequalities of* (24) *be satisfied. The Inexact FC-ALADIN, see Equation* (19)-(22)*, is guaranteed to converge if*

$$\begin{cases} \dfrac{\sum_{k=1}^{K}(\Psi_{max}^k)^2 G^2}{\sum_{k=1}^{K}\Psi_{min}^k} & \to 0, \\ \varphi_1 + \varphi_2 & \to 0. \end{cases} \tag{26}$$

Proof: See Appendix III. ■

Note that, if f_is are smooth, then the subgradients ∂f_is are replaced by the gradients ∇f_is in Inexact FC-ALADIN. The convergence analysis in this case is identical to that presented in Appendix III and is not repeated here. Moreover, for non-convex cases, if the local objectives f_is of Problem (2) are semi-convex [2, Definition 10 and Equation (18)], FC-ALADIN can still achieve global convergence with a *bi-level globalization strategy* [10].

IV. Conclusion

This paper proposes random polling variants of T-ALADIN and C-ALADIN, termed FT-ALADIN and FC-ALADIN, respectively, to address packet loss in unstable networks within the ALADIN framework. Additionally, we present a convergence analysis of FC-ALADIN under various scenarios, establishing theoretical guarantees for extending ALADIN to broader applications. Future research will explore diverse use cases to evaluate the numerical performance of the proposed algorithms.

Appendix I
Proof of Theorem 1

For $p = 1$ in Equation (7), FC-ALADIN (see Algorithm 2) simplifies to C-ALADIN (see [9]). In this case, for strongly convex problems (2), there always exists a $\delta > 0$ such that the following inequality holds (see [9]),

$$\mathcal{L}(y^+, \lambda^+) \leq \frac{1}{1+\delta}\mathcal{L}(y, \lambda). \tag{27}$$

In this proof, we adopt a slightly modified notation: $(\hat{y}^+, \hat{\lambda}^+)$ denotes the primal and dual solution generated by Algorithm 2 for a given (y, λ) from the previous iteration. The energy function (13), corresponding to $(\hat{y}^+, \hat{\lambda}^+)$, can be represented as follows,

$$\begin{aligned} &\mathbb{E}\left[\mathcal{L}(\hat{y}^+, \hat{\lambda}^+)\middle|(y, \lambda)\right] \\ =&p\mathcal{L}(y^+, \lambda^+) + (1-p)\mathcal{L}(y, \lambda) \leq \alpha\mathcal{L}(y, \lambda). \end{aligned} \tag{28}$$

From (28), with iteration index k, the following inequality holds,

$$\begin{aligned} &\mathbb{E}\left[\mathcal{L}(\hat{y}^k, \hat{\lambda}^k)\middle|(\hat{y}^{k-1}, \hat{\lambda}^{k-1})\right] \\ \leq&\alpha\ \mathbb{E}\left[\mathcal{L}(\hat{y}^{k-1}, \hat{\lambda}^{k-1})\middle|(\hat{y}^{k-2}, \hat{\lambda}^{k-2})\right] \\ &\vdots \\ \leq&\ \alpha^{k-2}\ \mathbb{E}\left[\mathcal{L}(\hat{y}^2, \hat{\lambda}^2)\middle|(y^1, \lambda^1)\right] \\ \leq&\ \alpha^{k-1}\ \mathcal{L}(y^1, \lambda^1). \end{aligned} \tag{29}$$

By defining $\mathbb{E}\left[\mathcal{L}(y^k, \lambda^k)\right] = \mathbb{E}[\mathcal{L}(\hat{y}^k, \hat{\lambda}^k)|(\hat{y}^{k-1}, \hat{\lambda}^{k-1})]$ for FC-ALADIN, Theorem 1 is then proved.

Appendix II
Proof of Theorem 2

For any agent $i \in \mathcal{C}^+$, if the optimal point of (15) is attained at a certain iteration, we have

$$\begin{cases} \nabla f_i(x_i^+) + \lambda_i + \rho\left(x_i^+ - y\right) = 0, \\ \nabla f_i(y^*) + \lambda_i^* = 0. \end{cases} \tag{30}$$

From Equation (30),

$$\frac{\rho}{\sigma}\|y - y^*\| + \frac{1}{\sigma}\|\lambda_i - \lambda_i^*\| \geq \|x_i^+ - y^*\|, \tag{31}$$

can be obtained with $\sigma < \|B_i + \rho I\|$, see [8, Appendix E]. For $i \notin \mathcal{C}^+$, Inequality (31) does not need to be satisfied. However, if Inequality (17) satisfies, the local convergence of Algorithm 2 can be then established.

Note that, for a sufficiently small $0 < \gamma < 1$ in (16), $\frac{(\rho+1)\gamma}{\sigma} < 1$ is guaranteed. See [18, Equation (24)], the following inequalities are satisfied,

$$\begin{cases} \mathbb{E}\left[N\|y^+ - y^*\|\right] \le \gamma \mathbb{E}\left[\sum_{i=1}^N \|x_i^+ - y^*\|\right], \\ \mathbb{E}\left[\sum_{i=1}^N \|\lambda_i^+ - \lambda_i^*\|\right] \le \gamma \mathbb{E}\left[\sum_{i=1}^N \|x_i^+ - y^*\|\right]. \end{cases} \tag{32}$$

With Equation (32) and (17), (33) is then derived,

$$\begin{aligned} &\mathbb{E}\left[\frac{\rho N}{\sigma}\|y^+ - y^*\| + \frac{1}{\sigma}\sum_{i=1}^N \left\|\lambda_i^+ - \lambda_i^*\right\|\right] \\ \le &\frac{(\rho+1)\gamma}{\sigma}\mathbb{E}\left[\frac{\rho N}{\sigma}\|y - y^*\| + \frac{1}{\sigma}\sum_{i=1}^N \|\lambda_i - \lambda_i^*\|\right]. \end{aligned} \tag{33}$$

Let the initial point (x^1, y^1, λ^1) be in a neighborhood of the optimal solution (x^*, y^*, λ^*), Equation (34) is guaranteed,

$$\begin{aligned} &\mathbb{E}\left[\frac{\rho N}{\sigma}\|y^k - y^*\| + \frac{1}{\sigma}\sum_{i=1}^N \left\|\lambda_i^k - \lambda_i^*\right\|\right] \\ \le &\left(\frac{(\rho+1)\gamma}{\sigma}\right)^{k-1}\left(\frac{\rho N}{\sigma}\|y^1 - y^*\| + \frac{1}{\sigma}\sum_{i=1}^N \left\|\lambda_i^1 - \lambda_i^*\right\|\right). \end{aligned} \tag{34}$$

Theorem 2 is then proved.

Appendix III
Proof of Theorem 3

The proof starts from the update of the global variable y:

$$\begin{aligned} &\|y^+ - y^*\|^2 \\ \overset{(21)}{=} &\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\left(\sum_{i=1}^N \left(B_i\mathbb{E}[x_i^+] - \mathbb{E}[g_i^+]\right)\right) - y^*\right\|^2 \\ \overset{(12)}{=} &\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\left(\sum_{i=1}^N B_i\left(px_i^+ + (1-p)x_i\right)\right)\right. \\ &\left.-\left(\sum_{i=1}^N B_i\right)^{-1}\sum_{i=1}^N \left(pg_i^+ + (1-p)g_i^-\right) - y^*\right\|^2. \end{aligned} \tag{35}$$

By expending the convexity of $\|\cdot\|^2$ [4, Equation 3.1, A.1], (36) is obtained

$$\begin{aligned} (35) \le p &\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\left(\sum_{i=1}^N \left(B_ix_i^+ - g_i^+\right)\right) - y^*\right\|^2 \\ +(1-p) &\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\left(\sum_{i=1}^N \left(B_ix_i - g_i^-\right)\right) - y^*\right\|^2. \end{aligned} \tag{36}$$

By plugging (19) and (20) into (36), (37) is derived,

$$\begin{aligned} (35) \le &p\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\right. \\ &\left.\left(\sum_{i=1}^N \left(B_iy - (\lambda_i + \partial f_i(y)) - \partial f_i(x_i^+)\right)\right) - y^*\right\|^2 \\ &+(1-p)\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\right. \\ &\left.\left(\sum_{i=1}^N \left(B_iy^- - \left(\lambda_i^- + \partial f_i(y^-)\right) - \partial f_i(x_i)\right)\right) - y^*\right\|^2. \end{aligned} \tag{37}$$

Taking (23) into account, (38) is guaranteed,

$$\begin{aligned} (35) \le &p\left\|y - y^* - \left(\sum_{i=1}^N B_i\right)^{-1}\left(\sum_{i=1}^N \left(\partial f_i(y) + \partial f_i(x_i^+)\right)\right)\right\|^2 \\ &+(1-p)\left\|y^- - y^*\right. \\ &\left.-\left(\sum_{i=1}^N B_i\right)^{-1}\left(\sum_{i=1}^N \left(\partial f_i(y^-) + \partial f_i(x_i)\right)\right)\right\|^2 \\ = &p\left\|y - y^* - \left(\sum_{i=1}^N B_i\right)^{-1}\mathcal{F}\right\|^2 \\ &+(1-p)\left\|y^- - y^* - \left(\sum_{i=1}^N B_i\right)^{-1}\mathcal{F}^-\right\|^2, \end{aligned} \tag{38}$$

where $\mathcal{F} = \sum_{i=1}^N \left(\partial f_i(y) + \partial f_i(x_i^+)\right)$ and $\mathcal{F}^- = \sum_{i=1}^N \left(\partial f_i(y^-) + \partial f_i(x_i)\right)$.

Note that, due to the convexity of f_is, the first part of Equation (38) is upper bounded according to (24), such that

$$\begin{aligned} &p\|y - y^*\|^2 + p\left\|\left(\sum_{i=1}^N B_i\right)^{-1}\mathcal{F}\right\|^2 \\ &-2p\left(y - y^*\right)^\top\left(\sum_{i=1}^N B_i\right)^{-1}\mathcal{F} \\ \overset{(24)}{\le} &p\left(\|y - y^*\|^2 + 4\Psi_{\text{max}}^2G^2 + 2\Psi_{\text{max}}G\sum_{i=1}^N \left\|y - x_i^+\right\|\right. \\ &\left.-2\Psi_{\text{min}}\left(\sum_{i=1}^N f_i(y) + \sum_{i=1}^N f_i(x_i^+) - 2\sum_{i=1}^N f_i(y^*)\right)\right). \end{aligned} \tag{39}$$

For the same reason, the second part of Equation (37) is upper bounded by (40),

$$\begin{aligned} &(1-p)\left(\left\|y^- - y^*\right\|^2 + 4\Psi_{\text{max}}^2G^2 + 2\Psi_{\text{max}}G\sum_{i=1}^N \left\|y^- - x_i\right\|\right. \\ &\left.-2\Psi_{\text{min}}\left(\sum_{i=1}^N f_i(y^-) + \sum_{i=1}^N f_i(x_i) - 2\sum_{i=1}^N f_i(y^*)\right)\right). \end{aligned} \tag{40}$$

By combining (39) and (40), (41) is then derived,

$$
\begin{aligned}
&\left\|y^+ - y^*\right\|^2 \\
&\le p\left\|y - y^*\right\|^2 + (1-p)\left\|y^- - y^*\right\|^2 + 4\Psi_{\max}^2 G^2 \\
&\quad + 2\Psi_{\max} G\left(p\sum_{i=1}^{N}\left\|y - x_i^+\right\| + (1-p)\sum_{i=1}^{N}\left\|y^- - x_i\right\|\right) \\
&\quad - 2p\Psi_{\min}\left(\sum_{i=1}^{N} f_i(y) + \sum_{i=1}^{N} f_i(x_i^+) - 2\sum_{i=1}^{N} f_i(y^*)\right) \\
&\quad - 2(1-p)\Psi_{\min}\left(\sum_{i=1}^{N} f_i(y^-) + \sum_{i=1}^{N} f_i(x_i) - 2\sum_{i=1}^{N} f_i(y^*)\right).
\end{aligned} \tag{41}
$$

This indicates that,

$$
\begin{aligned}
&2p\Psi_{\min}\left(\sum_{i=1}^{N} f_i(y) + \sum_{i=1}^{N} f_i(x_i^+) - 2\sum_{i=1}^{N} f_i(y^*)\right) \\
&\quad + 2(1-p)\Psi_{\min}\left(\sum_{i=1}^{N} f_i(y^-) + \sum_{i=1}^{N} f_i(x_i) - 2\sum_{i=1}^{N} f_i(y^*)\right) \\
&\le p\left\|y - y^*\right\|^2 + (1-p)\left\|y^- - y^*\right\|^2 - \left\|y^+ - y^*\right\|^2 + 4\Psi_{\max}^2 G^2 \\
&\quad + 2\Psi_{\max} G\left(p\sum_{i=1}^{N}\left\|y - x_i^+\right\| + (1-p)\sum_{i=1}^{N}\left\|y^- - x_i\right\|\right).
\end{aligned} \tag{42}
$$

By summing up (42) over the iteration index k, (43) is obtained,

$$
\begin{aligned}
&\sum_{i=1}^{N} f_i(y^{\text{best}}) - \sum_{i=1}^{N} f_i(y^*) \\
&\le \frac{1}{4\sum_{k=1}^{K}\Psi_{\min}^k}\left((1-p)\|y^1 - y^*\|^2 + \|y^2 - y^*\|^2\right. \\
&\qquad \left. + (p-1)\left\|y^{K-1} - y^*\right\|^2 - \|y^K - y^*\|^2\right) \\
&\qquad + \frac{G^2\sum_{k=1}^{K}(\Psi_{\max}^k)^2}{\sum_{k=1}^{K}\Psi_{\min}^k} + \varphi_1 + \varphi_2,
\end{aligned} \tag{43}
$$

where $\sum_{i=1}^{N} f_i(y^{\text{best}})$ denotes the minimum value that the recursion can achieve during the K iterations. If Equation (25) satisfies, Inexact FC-ALADIN converges. This completes the proof.

References

[1] N. Bastianello, R. Carli, L. Schenato, and M. Todescato. Asynchronous distributed optimization over lossy networks via relaxed admm: Stability and linear convergence. *IEEE Transactions on Automatic Control*, 66(6):2620–2635, 2020.

[2] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet. Characterizations of łojasiewicz inequalities: subgradient flows, talweg, convexity. *Transactions of the American Mathematical Society*, 362(6):3319–3363, 2010.

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

[4] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[5] T.-H. Chang. A proximal dual consensus admm method for multi-agent constrained optimization. *IEEE Transactions on Signal Processing*, 64(14):3719–3734, 2016.

[6] T.-H. Chang, M. Hong, W.-C. Liao, and X. Wang. Asynchronous distributed admm for large-scale optimization-part i: Algorithm and convergence analysis. *IEEE Transactions on Signal Processing*, 64(12):3118–3130, 2016.

[7] X. Du, A. Engelmann, Y. Jiang, T. Faulwasser, and B. Houska. Distributed state estimation for AC power systems using Gauss-Newton ALADIN. In *In Proceedings of the 58th IEEE Conference on Decision and Control*, pages 1919–1924, 2019.

[8] X. Du and J. Wang. Consensus aladin: A framework for distributed optimization and its application in federated learning. *arXiv preprint arXiv:2306.05662*, 2023.

[9] X. Du and J. Wang. Distributed consensus optimization with consensus ALADIN. In *American Control Conference*, 2025 (accepted for publication).

[10] X. Du, J. Wang, X. Zhou, and Y. Mao. A bi-level globalization strategy for non-convex consensus admm and aladin. *arXiv preprint arXiv:2309.02660*, 2023.

[11] A. Engelmann, Y. Jiang, B. Houska, and T. Faulwasser. Decomposition of nonconvex optimization via bi-level distributed aladin. *IEEE Transactions on Control of Network Systems*, 7(4):1848–1858, 2020.

[12] A. Engelmann, Y. Jiang, T. Mühlpfordt, B. Houska, and T. Faulwasser. Toward distributed OPF using ALADIN. *IEEE Transactions on Power Systems*, 34(1):584–594, 2019.

[13] A. Engelmann, T. Mühlpfordt, Y. Jiang, B. Houska, and T. Faulwasser. Distributed AC optimal power flow using ALADIN. *IFAC-PapersOnLine*, 50(1):5536 – 5541, 2017. 20th IFAC World Congress.

[14] X. Gao, Y.-Y. Xu, and S.-Z. Zhang. Randomized primal–dual proximal block coordinate updates. *Journal of the Operations Research Society of China*, 7(2):205–250, 2019.

[15] M. Hong. A distributed, asynchronous, and incremental algorithm for nonconvex optimization: An admm approach. *IEEE Transactions on Control of Network Systems*, 5(3):935–945, 2017.

[16] 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.

[17] B. Houska, J. Frasch, and M. Diehl. An augmented Lagrangian based algorithm for distributed nonconvex optimization. *SIAM Journal on Optimization*, 26(2):1101–1127, 2016.

[18] B. Houska and Y. Jiang. Distributed optimization and control with aladin. *Recent Advances in Model Predictive Control: Theory, Algorithms, and Applications*, pages 135–163, 2021.

[19] Y. Jiang, D. Kouzoupis, H. Yin, M. Diehl, and B. Houska. Decentralized optimization over tree graphs. *Journal of Optimization Theory and Applications*, 189:384–407, 2021.

[20] D. Kovalev, K. Mishchenko, and P. Richtárik. Stochastic newton and cubic newton methods with simple local linear-quadratic rates. *arXiv preprint arXiv:1912.01597*, 2019.

[21] S. E. Li, Z. Wang, Y. Zheng, Q. Sun, J. Gao, F. Ma, and K. Li. Synchronous and asynchronous parallel computation for large-scale optimal control of connected vehicles. *Transportation research part C: emerging technologies*, 121:102842, 2020.

[22] Q. Ling, W. Shi, G. Wu, and A. Ribeiro. Dlm: Decentralized linearized alternating direction method of multipliers. *IEEE Transactions on Signal Processing*, 63(15):4051–4064, 2015.

[23] A. I. Rikos, W. Jiang, T. Charalambous, and K. H. Johansson. Asynchronous distributed optimization via ADMM with efficient communication. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 7002–7008. IEEE, 2023.

[24] R. Sun, Z.-Q. Luo, and Y. Ye. On the efficiency of random permutation for admm and coordinate descent. *Mathematics of Operations Research*, 45(1):233–271, 2020.

[25] H. Wang, Y. Gao, Y. Shi, and R. Wang. Group-based alternating direction method of multipliers for distributed linear classification. *IEEE transactions on cybernetics*, 47(11):3568–3582, 2016.

[26] X. Wang, J. Yan, B. Jin, and W. Li. Distributed and parallel admm for structured nonconvex optimization problem. *IEEE transactions on cybernetics*, 51(9):4540–4552, 2019.

[27] E. Wei and A. Ozdaglar. On the o (1= k) convergence of asynchronous distributed alternating direction method of multipliers. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 551–554. IEEE, 2013.

[28] R. Zhang and J. Kwok. Asynchronous distributed admm for consensus optimization. In *International conference on machine learning*, pages 1701–1709. PMLR, 2014.

[29] S. Zhou and G. Y. Li. Federated learning via inexact admm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.