Decentralized Optimization via RC-ALADIN with Efficient Quantized Communication

Xu Du, Karl H. Johansson, and Apostolos I. Rikos*

Abstract— **In this paper, we investigate the problem of decentralized consensus optimization over directed graphs with limited communication bandwidth. We introduce a novel decentralized optimization algorithm that combines the Reduced Consensus Augmented Lagrangian Alternating Direction Inexact Newton (RC-ALADIN) method with a finite time quantized coordination protocol, enabling quantized information exchange among nodes. Assuming the nodes' local objective functions are μ-strongly convex and simply smooth, we establish global convergence at a linear rate to a neighborhood of the optimal solution, with the neighborhood size determined by the quantization level. Additionally, we show that the same convergence result also holds for the case where the local objective functions are convex and L-smooth. Numerical experiments demonstrate that our proposed algorithm compares favorably against algorithms in the current literature while exhibiting communication efficient operation.**

I. Introduction

Decentralized optimization has garnered significant attention in recent years, driven by advances in areas such as control systems [1], machine learning [2], and power grids [3]. This growing interest is largely due to the increasing need to address optimization problems that involve vast amounts of data and heterogeneous objective functions.

Decentralized optimization distributes data across multiple network nodes, typically using two main approaches: (i) primal decomposition and (ii) dual decomposition. Both methods locally optimize the objective functions. Primal decomposition methods focus on sharing primal information among nodes, such as local optimal solutions or first- and second-order objective function data (e.g., [4], [5]). In contrast, dual decomposition methods update both the primal and the dual variables associated with coupling constraints (e.g., [6], [7]). In general, dual decomposition methods achieve faster convergence and higher accuracy than primal decomposition approaches [8, Section I]. In this paper we focus on dual decomposition methods.

Existing Literature. Dual decomposition-based optimization methods are built upon several core frameworks: Dual Decomposition [9], the Alternating Direction Method of Multipliers (ADMM) [6], and the Augmented Lagrangian Alternating Direction Inexact Newton (ALADIN) method [7]. While ADMM has undergone significant developments in both convergence theory and practical applications [6], [10], [11], ALADIN improves convergence performance—achieving faster rates and ensuring convergence for both convex and nonconvex problems—by incorporating sequential quadratic programming techniques [7], [12]. Subsequent developments have introduced specialized variants of ALADIN designed to tackle diverse computational complexity challenges [1], [12]–[15]. However, two key limitations of these approaches are (i) their reliance on centralized coordination mechanisms (which limits scalability in distributed systems), and (ii) the requirement of nodes transmitting real-valued messages requiring a significant amount of bandwidth (which creates a scalability bottleneck). To address the first challenge, the works [8], [16], [17] investigated its decentralization for both resource allocation and consensus problems. Furthermore, recent works [1], [18] have extended ALADIN to decentralized settings, however focusing only on resource allocation problems. Note that, despite some existing studies on variants of consensus ALADIN, such as [12], [19], a decentralized version of this framework has yet to be explored and addressed in the literature. Meanwhile, to address the second challenge of inefficient communication due to transmission of real-valued messages, the works in [20], [21] investigated decentralized Consensus ADMM with quantized communication on undirected graphs. Although the aforementioned works have advanced ADMM development, they have only addressed some of the identified bottlenecks. The aforementioned literature has certain limitations, such as assuming a centralized coordinator, requiring nodes to exchange real-valued messages, or being restricted to undirected graphs. To the best of our knowledge, [22] is the only study to explore decentralized Consensus ADMM on directed graphs while enabling nodes to communicate in a resource efficient manner by exchanging quantized valued messages. Note that, although Consensus ALADIN exhibits superior convergence performance compared to Consensus ADMM [12], its decentralized variant which however exhibits efficient communication with quantized value on directed graphs remains a problem that is unexplored in the literature.

Main Contributions. Motivated by the aforementioned challenges, we introduce a novel decentralized optimization

*Corresponding author.

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: michaelxudu@hkust-gz.edu.cn; apostolosr@hkust-gz.edu.cn.

Karl H. Johansson is with the Division of Decision and Control Systems, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. He is also affiliated with Digital Futures, SE-100 44 Stockholm, Sweden. E-mail: kallej@kth.se.

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

algorithm leveraging Consensus ALADIN to tackle these issues. Note that to the authors knowledge, it is the first that Consensus ALADIN to address (i) fully decentralized algorithm operation, (ii) communication over directed graphs, and (iii) quantized communication among nodes. Inspired by [22] and [23], we note that adopting a two-layer algorithmic structure—where the optimization steps are decoupled from the decentralized averaging procedure—can achieve faster convergence compared to approaches that interleave optimization and communication in a single layer. Our contributions are the following.

A. We present a novel two-layer decentralized optimization algorithm, termed Quantized Decentralized Reduced Consensus ALADIN (QuDRC-ALADIN) (see Algorithm 1), designed to operate over directed communication graphs. The inner layer (Algorithm 2, inspired by [22]) enables efficient communication among nodes via the exchange of quantized messages. The outer layer (i.e., the remaining steps of Algorithm 1) is responsible for updating the primal and dual variables of the original optimization problem.

B. We prove that our algorithm converges to a neighborhood of the optimal solution (where the neighborhood depends on the quantization level) with global linear convergence rate for the case where the local cost function of each node is μ-strongly convex, closed, proper, and simply smooth (see Theorem 1). Additionally, we show that our algorithm exhibits global linear convergence rate also for the case where the local cost function of each node is convex, closed, proper, and L-smooth (see Corollary 1).

II. Notation and Preliminaries

Notation. We use the symbols $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, and $\mathbb{N}$ to represent the sets of real, rational, integer, and natural numbers, respectively. Matrices are indicated by capital letters (e.g., A), and vectors are represented by lowercase letters (e.g., a). The transpose of matrix $A \in \mathbb{R}^{n\times n}$ and vector $a \in \mathbb{R}^n$ are represented as $A^\top$ and $a^\top$, respectively. For a real number $a \in \mathbb{R}$, $\lfloor a \rfloor$ and $\lceil a \rceil$ denote the greatest integer less than or equal to a and the least integer greater than or equal to a, respectively. For the real vector $a \in \mathbb{R}^n$, $\lfloor a \rfloor \in \mathbb{R}^n$ and $\lceil a \rceil \in \mathbb{R}^n$ denote the element-wise operation. Furthermore, $\mathbf{1}$ represents the vector of all ones and I denotes the identity matrix with appropriate dimensions. In addition, $\|a\|$ denotes the Euclidean norm of the vector a. The value of a variable x of node i at iteration k is denoted as $x_i^{[k]}$. The updated value is denoted as $(\cdot)^+$. Furthermore, $|\mathcal{S}|$ denotes the cardinality of a countable set $\mathcal{S}$ (e.g. $|\mathcal{V}| = N$ as we can see below). The notation $a|b$ denotes $b \in \mathbb{R}^n$ as the dual variable of constraint $a \in \mathbb{R}^n$.

Graph Theory. The communication network is captured by a directed graph (or digraph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The set of agents (nodes) is denoted as $\mathcal{V} = \{1, \cdots, N\}$ (where $|\mathcal{V}| \geq 2$). The set of edges is denoted as $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}\cup\{(i,i) \mid i \in \mathcal{V}\}$ (each node has a virtual self-edge). A directed edge from node i to node j is denoted by $e_{ji} \doteq (j,i) \in \mathcal{E}$. The subset of nodes that can directly transmit information to node i is called the set of in-neighbors of i and is denoted $\mathcal{N}_i^- = \{j \in \mathcal{V} \mid (i,j) \in \mathcal{E}\}$. The subset of nodes that can directly receive information from node i is called the set of out-neighbors of i and is denoted $\mathcal{N}_i^+ = \{l \in \mathcal{V} \mid (l,i) \in \mathcal{E}\}$. The cardinality of $\mathcal{N}_i^-$ represented as $\mathcal{D}_i^- = |\mathcal{N}_i^-|$, is called *in-degree* of node i. The cardinality of $\mathcal{N}_i^+$ represented as $\mathcal{D}_i^+ = |\mathcal{N}_i^+|$, is called *out-degree* of node i. The diameter D of digraph $\mathcal{G}$ is the longest shortest path between $i, j \in \mathcal{V}$. A digraph $\mathcal{G}$ is strongly connected if there exists a directed path from every node i to node j that $i, j \in \mathcal{V}$.

Quantization. In digital communication networks, quantization serves to reduce bandwidth requirements and improve communication efficiency. By using a finite number of bits, quantization enables the application of error-correcting codes (e.g., Reed-Solomon, LDPC) to significantly enhance the signal's resilience to interference during transmission [24]. Three primary types of quantizers have been extensively studied: asymmetric, uniform, and logarithmic (see [25]). In this paper we utilize asymmetric mid-rise quantizers with an infinite range (although our findings also hold for other types of quantizers as well). An asymmetric mid-rise quantizer is defined as

$$q_\Delta^a(b) = \left\lfloor \frac{b}{\Delta} \right\rfloor, \tag{1}$$

where $b \in \mathbb{R}^n$ is the value to be quantized, and $\Delta \in \mathbb{Q}$ denotes the quantization level, and the superscript a denotes the asymmetric type.

III. Problem Formulation

Consider a communication network represented by a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ comprising $N = |\mathcal{V}|$ nodes. We assume that the communication channels between nodes in our network $\mathcal{G}$ have limited bandwidth. Each node i is associated with a scalar local cost function $f_i(x) : \mathbb{R}^n \mapsto \mathbb{R}$, known exclusively to that node. Our objective in this paper is to develop a decentralized algorithm that enables nodes to collaboratively solve the following optimization problem

$$\min_{z\in\mathbb{R}^n} \quad \sum_{i=1}^{N} f_i(z), \quad i \in \{1, \cdots, N\}, \tag{2}$$

where z is the global optimization variable. In order to solve problem (2) we introduce a local variable x_i for each node $i \in \mathcal{V}$ (following the approach outlined in [6, Chapter 7.1]). Thus, (2) is reformulated as

$$\begin{aligned} \min_{x_i,\ i=1,\dots,N} \quad & \sum_{i=1}^{N} f_i(x_i) \\ \text{s.t.} \quad & x_i = z,\ \forall i \in \{1, \cdots, N\}, \quad x_i, z \in \mathbb{R}^n. \end{aligned} \tag{3}$$

Problem (3) is known as the consensus optimization problem, as the constraint enforces equality among all local variables. In order to solve (3) via the RC-ALADIN strategy (see [12]) while guaranteeing efficient communication among nodes, we have that (3) is reformulated as

$$\begin{aligned}\min_{x_i,\ i=1,\ldots,N} \quad & \sum_{i=1}^{N} f_i(x_i) \\ \text{s.t.} \quad & x_i = z,\ \forall i \in \{1, \cdots, N\}, \quad x_i, z \in \mathbb{R}^n, \\ & \text{nodes communicate with quantized values.}\end{aligned} \tag{4}$$

The primary contribution of this paper is the development of a decentralized algorithm that enables nodes to solve the problem in (4). In our paper, a decentralized approach refers to a distributed approach without the presence of a central coordinator. More specifically, in our proposed decentralized algorithm nodes coordinate solely through communication with their immediate neighbors (i.e., without the presence of a central coordinator).

IV. Preliminaries of RC-ALADIN

The augmented Lagrangian of Problem (4) is given by

$$\mathscr{L}(x, z, \lambda) = \sum_{i=1}^{N} f_i(x_i) + \lambda_i^\top (x_i - z) + \rho \|x_i - z\|^2, \tag{5}$$

where, $\rho > 0$ is a penalty parameter, $\lambda = [\lambda_1^\top, \cdots, \lambda_N^\top]^\top$ denotes the dual variables and $x = [x_1^\top, \cdots, x_N^\top]^\top$ collects the local primal variables. Focusing on (5), RC-ALADIN was proposed in [12] to solve the consensus optimization problem in (3). Details of RC-ALADIN are the following,

$$x_i^+ = \arg\min_{x_i} f_i(x_i) + \lambda_i^\top x_i + \frac{\rho}{2}\|x_i - z\|^2,\ \forall i \in \mathcal{V}, \tag{6a}$$

$$g_i = \rho\left(z - x_i^+\right) - \lambda_i,\ \forall i \in \mathcal{V}, \tag{6b}$$

$$z^+ = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^+ - \frac{g_i}{\rho}\right), \tag{6c}$$

$$\lambda_i^+ = \rho\left(x_i^+ - z^+\right) - g_i,\ \forall i \in \mathcal{V}. \tag{6d}$$

In equation (6a), we minimize the augmented Lagrangian with respect to each x_i. In equation (6b), we evaluate the (sub)gradient of each f_i at x_i^+. Note that (6b) always exists as long as (6a) is solvable. Equations (6c) and (6d) are the closed form expression of the following centralized reduced consensus QP problem

$$\begin{aligned}\min_{\Delta x_i \in \mathbb{R}^n,\ i=1,\ldots,N} \quad & \sum_{i=1}^{N} \frac{\rho}{2}\Delta x_i^\top \Delta x_i + g_i^\top \Delta x_i \\ \text{s.t.} \quad & x_i^+ + \Delta x_i = z | \lambda_i,\ \forall i \in \{1, \cdots, N\}.\end{aligned} \tag{7}$$

The reduced consensus QP in (7) is used from nodes to coordinate without depending on second-order information of their local cost functions f_i, $\forall i \in \mathcal{V}$. Our previous work, RC-ALADIN in (6), provides global convergence guarantees for (3) when the local functions of each node are convex. Additionally, it offers globally linear convergence guarantees for (3) when the local functions of each node are μ-strongly convex (see [26, Theorem 2, Theorem 7]).

Note that, RC-ALADIN in (6) is implemented by nodes in a distributed manner, but also considers the presence of a centralized coordinator that can exchange messages with every node in the network. More specifically, focusing on (6a) – (6d), the updates (6a), (6b), (6d) are performed locally from each node i, and the update (6c) is performed in a centralized fashion from the centralized coordinator. Motivated by this limitation, in the next section we will present a fully decentralized algorithm. Our proposed algorithm enables nodes to (i) collaboratively solve problem (4) by communicating exclusively with their immediate neighbors (eliminating the need for a central coordinator), and (ii) exhibit efficient communication within the network.

V. Decentralized RC-ALADIN with Efficient Communication

In this section, we introduce a novel decentralized algorithm designed to address the problem in (4). Before introducing the proposed decentralized algorithm, we first make the following assumptions that are important for our subsequent development.

Assumption 1. *The communication network is modeled as a strongly connected digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Also, every node i knows the diameter of the network D, and a common quantization level Δ.*

Assumption 2. *The local cost function f_i of each node $i \in \mathcal{V}$ is closed, proper, simply smooth and μ-strongly convex. Specifically, for each local cost function f_i, for every $x_\alpha, x_\beta \in \mathbb{R}^n$, there exist a* strong-convexity *constant $\mu_i > 0$ such that*

$$f_i(x_\alpha) + \nabla f_i(x_\alpha)^\top (x_\beta - x_\alpha) + \frac{\mu_i}{2}\|x_\beta - x_\alpha\|^2 \leq f_i(x_\beta). \tag{8}$$

Here, simply smooth refers to a smooth function f_i where L_i cannot be explicitly estimated as in inequality (9) *in Assumption 3 below. In contrast, L-smooth indicates that f_i is smooth with an explicitly estimable L_i as shown in* (9) *below.*

Assumption 3. *The local cost function f_i of each node $i \in \mathcal{V}$ is closed, proper, L-smooth and convex. Specifically, for each local cost function f_i, for every $x_\alpha, x_\beta \in \mathbb{R}^n$, there exist a* Lipschitz-continuity *constant $L_i > 0$ (see [8, equation (51)]) such that*

$$\begin{aligned}&\|\nabla f_i(x_\alpha) - \nabla f_i(x_\beta)\|^2 \\ &\leq L_i \left(\nabla f_i(x_\alpha) - \nabla f_i(x_\beta)\right)^\top (x_\alpha - x_\beta).\end{aligned} \tag{9}$$

In Assumption 1, strong connectivity ensures that information can propagate between all nodes in the network and guarantees the convergence of Algorithm 2 (since strongly connected digraph implies there is a path from every node to every other node in the network). Knowing the diameter of the network is useful for each node to determine whether Algorithm 2 has converged, allowing it to proceed to step 3 of Algorithm 1. Note here that the network diameter can can be computed by employing distributed algorithms [27]. For open undirected networks, the network diameter D need not be known a priori [28]. The quantization level is important for quantizing the messages as described in (2) (thus ensuring efficient communication during the execution of Algorithm 2). Nodes can compute a common quantization level Δ in finite time via a max-consensus operation. Assumption 2 ensures strong convexity and simple smoothness, allowing us to establish a global linear convergence rate for Algorithm 1 while ensuring that the global cost function in (4) has a unique minimum [29, Theorem 13.27], [8]. Assumption 3 guarantees that the Lipschitz continuity of gradients in (9) ensures the existence of a global optimal solution x^* for (4) and enables nodes to compute it, which is a standard requirement in first-order distributed optimization frameworks (see, e.g., [29]). Many practical optimization problems satisfy Assumptions 2 and 3 [30], [31]. However, when these two assumptions are violated in specific applications, the proposed Algorithm 1 admits only global convergence, without a guaranteed convergence rate [12], [26].

A. Algorithm Development

In this section, we present our proposed decentralized algorithm, detailed below as Algorithm 1.

Algorithm 1 QuDRC-ALADIN: Quantized Decentralized Reduced Consensus ALADIN

Input. Strongly connected digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, parameter ρ, network diameter D, quantization level Δ, for each node $i \in \mathcal{V}$. Each node $i \in \mathcal{V}$ has a local cost function f_i. Assumptions 1 and 2 hold.

Initialization. Randomly chosen dual variable $\hat{\lambda}_i \in \mathbb{R}^n$, and global variable estimation $\hat{z}_i \in \mathbb{R}^n$, for each node $i \in \mathcal{V}$.

Iteration. Each node $i \in \mathcal{V}$ repeats:

1) Optimize x_i as
$$x_i^+ = \arg\min_{x_i} f_i(x_i) + \hat{\lambda}_i^\top x_i + \frac{\rho}{2}\|x_i - \hat{z}_i\|^2. \quad (10)$$
2) Evaluate the gradient g_i of f_i as
$$g_i = \rho\left(\hat{z}_i - x_i^+\right) - \hat{\lambda}_i. \quad (11)$$
3) Calculate the global variable estimation as
$$\hat{z}_i^+ = \text{Algorithm 2}\left(x_i^+ - \frac{g_i}{\rho}, D, \Delta\right). \quad (12)$$
4) Update the dual variable $\hat{\lambda}_i$ as
$$\hat{\lambda}_i^+ = \rho\left(x_i^+ - \hat{z}_i^+\right) - g_i. \quad (13)$$

Output. Each node i calculates x_i^* that solves problem (4).

The intuition of Algorithm 1 is organized into two main phases: local optimization and coordination among nodes. In the first step, each node $i \in \mathcal{V}$ performs a local optimization to determine the optimal value of its variable x_i by solving its corresponding augmented objective function, (see (10)). Following this, in the second step each node i evaluates the (sub)gradient of its local function f_i at the locally optimized solution x_i^+. This (sub)gradient evaluation serves as preparation for the aggregation process in the subsequent step (see (11)). In the third step, all nodes collaborate to update their estimates of the global variable $\hat{z}_i^+$ through the quantized, decentralized operation of Algorithm 2 (see (12)). Finally, in the fourth step each node updates its dual variable $\hat{\lambda}_i^+$ which encodes sensitivity information related to the constraints of problem (4). The updated dual variables are then utilized in the next iteration's local optimization phase (see (13)). Overall, Algorithm 1 alternates between performing local optimization (step 1) and solving problem (7), iterating until convergence is achieved and the optimal solution is obtained.

Algorithm 2 FQAC: Finite-time Quantized Average Consensus

Input. $y_i = x_i^+ - \frac{g_i}{\rho}, D, \Delta$.

Initialization. Each node $i \in \mathcal{V}$:

1) Assigns probability
$$p_{li} = \begin{cases} \dfrac{1}{1+\mathcal{D}_i^+}, & \text{if } l \in \mathcal{N}_i^+ \cup \{i\}, \\ 0, & \text{if } l \notin \mathcal{N}_i^+ \cup \{i\}, \end{cases} \quad (14)$$
to each out-neighbor of node i.
2) Sets $\xi_i = 2$, $\chi_i = 2q_\Delta^a(y_i)$ (see (1)).

Iteration. For time steps $t = 1, 2, \cdots$ each node $i \in \mathcal{V}$ does:

1) **If** $t \bmod (D) = 1$, sets $M_i = \left\lceil \frac{\chi_i}{\xi_i} \right\rceil$ and $m_i = \left\lfloor \frac{\chi_i}{\xi_i} \right\rfloor$.
2) Broadcasts M_i, m_i to each out-neighbor $l \in \mathcal{N}_i^+$ and receives M_j, m_j from each in-neighbor $j \in \mathcal{N}_i^-$. Then, sets $M_i = \max_{j \in \mathcal{N}_i^- \cup \{i\}} M_j$, $m_i = \min_{j \in \mathcal{N}_i^- \cup \{i\}} m_j$.
3) Sets $\tau_i = \xi_i$.
4) **While** $\tau_i > 1$ **do**
   a) $c_i = \left\lfloor \frac{\chi_i}{\xi_i} \right\rfloor$.
   b) Sets $\chi_i = \chi_i - c_i$, $\xi_i = \xi_i - 1$, $\tau_i = \tau - 1$.
   c) Transmits c_i to randomly chosen out-neighbor $l \in \mathcal{N}_i^+ \cup \{i\}$ with probability p_{li}.
   d) Receives c_i from $j \in \mathcal{N}_i^-$ and updates
$$\chi_i^{[t+1]} = \chi_i^{[t]} + \sum_{j \in \mathcal{N}_i^-} w_{ij}^{[t]} c_j^{[t]}, \quad (15a)$$
$$\xi_i^{[t+1]} = \xi_i^{[t]} + \sum_{j \in \mathcal{N}_i^-} w_{ij}^{[t]}. \quad (15b)$$
   Here $w_{ij}^{[t]} = 1$ if node i receives $c_j^{[t]}$ from node j at step t. Otherwise $w_{ij}^{[t]} = 0$ and node i does not receive information from node j.
5) **if** t mod $(D) = 0$ and $\|M_i - m_i\|_\infty \leq 1$, set $\hat{z}_i^+ = m_i\Delta$, and stop the operation of the algorithm.

Output. $\hat{z}_i^+$.

Algorithm 2 follows a structure similar to [32, Algorithm 1], and consists of three main operations: quantization, averaging, and a stopping criterion. During initialization each node i quantizes its local information $y_i = x_i^+ - \frac{g_i}{\rho}$ into a quantized value χ_i. Then, it splits χ_i into ξ_i pieces (the value of some pieces might be greater than others by one). It retains the piece with the smallest value to itself and transmits the rest $\xi_i - 1$ pieces to randomly chosen out-neighbors $l \in \mathcal{N}_i^+$ or to itself. Then, it receives the pieces c_j transmitted from each in-neighbor $j \in \mathcal{N}_i^-$ and updates χ_i and ξ_i as in (15). The algorithm also performs max- and min-consensus operations every D time steps. If the results of the max-consensus M_i and min-consensus m_i have a difference less or equal to one, then each node i scales its solution according to the quantization level to compute $\hat{z}_i^+$. At this point, Algorithm 2 terminates, and each node i transitions to step 4 of Algorithm 1. Note that Algorithm 2 is guaranteed to converge in finite time, ensuring $\|M_i - m_i\|_\infty \leq 1$ (see Step 5). The convergence time depends on the network diameter D, as established in [33, Theorem 1].

Comparison with Previous Works. Note that the key contribution of this work is the development of a communication-efficient fully decentralized algorithm for solving problem (4). More specifically, while our previous work RC-ALADIN in (6) (also see [12]) is distributed and considers the existence of a server node able to communicate with every node in the network, our main contribution in this paper lies in achieving full decentralization (i.e., nodes coordinate without the presence of a server node). This is achieved by replacing the centralized update in (6c) by a decentralized operation through the introduction of local copies $\hat{z}_i$ of z. Moreover, by implementing Algorithm 2, Algorithm 1 enhances its communication-efficiency while ensuring convergence precision (see for example [22], [33], [34]). This characteristic is not present in our previous work RC-ALADIN in [12] where nodes are operating in a resource inefficient manner by exchanging real-valued messages that require a significant amount of bandwidth. Additionally, in our previous work [22], the convexity (without simple smoothness) of f_i for all i is sufficient for establishing convergence, demonstrating a global *sub-linear* convergence rate. Furthermore, the algorithms in [33], [34] require L-smoothness *and* μ-strong convexity to establish the global *linear* convergence rate. In contrast, in this paper, *either* simple smoothness with μ-strong convexity *or* L-smoothness with simple convexity of f_i for all i is required to establish the global linear convergence rate of Algorithm 1. Finally, note that our previous works [12], [26] require the same assumptions as this paper (i.e., *either* simple smoothness with μ-strong convexity *or* L-smoothness with simple convexity of f_i for all i). However, as we mentioned above they do not exhibit communication efficient operation among nodes. Details of this will be provided in the next subsection.

B. Convergence Analysis

In this section, we provide the convergence analysis of Algorithm 1. First, we introduce the two lemmas that are important for our analysis. Then, we prove our main result via a theorem.

Lemma 1. *The update of the global variable estimation $\hat{z}_i$ of each node $i \in \mathcal{V}$ is given by Algorithm 1 in* (12)*. According to the constraints of problem* (4)*, the following equation is satisfied*

$$\begin{cases} \hat{z}_i^+ = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \Delta \left\lfloor \frac{y_i}{\Delta} \right\rfloor + \kappa_i, \ \|\kappa_i\|_\infty \leq \Delta, \\ \left\| z^+ - \hat{z}_i^+ \right\|_\infty \leq 2\Delta, \\ \left\| \hat{\lambda}_i^+ - \lambda_i^+ \right\|_\infty = \left\| \rho(z^+ - \hat{z}_i^+) \right\|_\infty \leq 2\rho\Delta, \end{cases} \tag{16}$$

where $y_i = x_i^+ - \frac{g_i}{\rho}$.

Proof. See [35, Lemma 1]. ∎

Lemma 2. *For the distributed consensus optimization problem presented in* (4)*, Algorithm 1 establishes a relationship between the local primal update x_i^+, the local dual variables $\hat{\lambda}_i$ and $\hat{\lambda}_i^+$, and the global primal variable approximations $\hat{z}_i$ and $\hat{z}_i^+$. This relationship is*

$$x_i^+ = \frac{\hat{\lambda}_i^+ - \hat{\lambda}_i}{2\rho} + \frac{\hat{z}_i^+ + \hat{z}_i}{2}. \tag{17}$$

Proof. From (13) we have

$$\begin{aligned} \hat{\lambda}_i^+ &= \rho\left(x_i^+ - \hat{z}_i^+\right) - g_i \\ &\overset{(6b)}{=} \rho\left(x_i^+ - \hat{z}_i^+\right) - \left(\rho\left(\hat{z}_i - x_i^+\right) - \lambda_i\right) \\ &= 2\rho x_i^+ - \rho(\hat{z}_i^+ + \hat{z}_i) + \hat{\lambda}_i. \end{aligned} \tag{18}$$

equation (17) is then derived from (18). ∎

Moreover, from the KKT (Karush-Kuhn-Tucker) system of problem (7) and (16), the following formulas can be obtained,

$$\sum_{i=1}^{N} \lambda_i^+ = 0, \text{ and } \left\| \sum_{i=1}^{N} \hat{\lambda}_i^+ \right\|_\infty \leq 2\rho N \Delta. \tag{19}$$

In (19) the first equality arises from the KKT stationarity condition for the reduced QP problem in (7). The second inequality arises from substituting the third inequality of (16) into (13) and then summing over all nodes. For establishing the global convergence of Algorithm 1, we introduce the following Lyapunov function,

$$\mathcal{L}(z, \lambda) = \frac{1}{\rho}\sum_{i=1}^{N} \|\lambda_i - \lambda_i^*\|^2 + \rho N \|z - z^*\|^2, \tag{20}$$

where $\lambda = [\lambda_1^\top, \lambda_2^\top, \cdots, \lambda_N^\top]^\top$, and (z^*, λ^*) denotes the optimal solution pair of problem (4). Following the analysis in [17], we define the finite positive scalars $0 < M_z < \infty$ for simplifying our later analysis, such that $\|z - z^*\| \leq M_z$ for every node $i \in \mathcal{V}$. Note that the proof of the theorem below is available in [36]. It will also be available in an extended version of our paper.

Theorem 1. *Let us consider a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node $i \in \mathcal{V}$ has a local cost function f_i, and Assumptions 1 and 2 hold. Each node $i \in \mathcal{V}$ in the network executes Algorithm 1 for solving the consensus optimization problem in* (4) *in a decentralized fashion. Given parameter $\rho > 0$, during the operation of Algorithm 1 there always exists a $\delta > 0$ such that*

$$\delta \mathcal{L}\left(\hat{z}^+, \hat{\lambda}^+\right) \leq 4\sum_{i=1}^{N} \mu_i \left\| x_i^+ - z^* \right\|^2, \tag{21}$$

where $\hat{z} = \hat{z}_i, \forall i \in \mathcal{V}$. From (21)*, we have that during the operation of Algorithm 1 the following inequality is satisfied*

$$\mathcal{L}\left(\hat{z}^{+}, \hat{\lambda}^{+}\right) \leq \frac{1}{1+\delta} \mathcal{L}\left(\hat{z}, \hat{\lambda}\right) + \frac{4}{1+\delta} \mathcal{O}(N\Delta), \qquad (22)$$

where $\mathcal{O}(N\Delta) = 6\rho M_z N\Delta$, Δ is the utilized quantization level, and $\|z - z^*\| \leq M_z < \infty$.

Remark 1. *Focusing on* (22) *of Theorem 1, the term $\frac{4}{1+\delta}\mathcal{O}(N\Delta)$ represents the quantization error introduced from Algorithm 2. As we will see later in Section VI, this error causes nodes to converge to a Δ-dependent neighborhood of the optimal solution. While decentralized approaches to progressively refine Δ can enhance solution precision [34], this typically incurs higher communication overhead in terms of bits per message compromising our algorithm's communication efficiency. In contrast, employing quantizers with base-shifting capabilities [37] allows for maintaining high communication efficiency while still enabling nodes to approximate the optimal solution with greater precision. This latter strategy however, results in a trade-off as it may reduce the convergence speed of Algorithm 2.*

Relaxing Strong Convexity. While Theorem 1 relies on μ-strong convexity (see Assumption 2), it is important to note that Algorithm 1 can achieve a symmetric global linear convergence rate when the local cost function f_i of each node $i \in \mathcal{V}$ is convex (and not strongly convex) provided it remains closed, proper, and L-smooth (i.e., satisfies Assumption 3). We present this in the following corollary. Note that the proof of the corollary is available in [36] and will also be available in an extended version of our paper.

Corollary 1. *Let us consider a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Each node $i \in \mathcal{V}$ has a local cost function f_i, and Assumptions 1 and 3 hold. Each node $i \in \mathcal{V}$ executes Algorithm 1 for solving the consensus optimization problem in* (4) *in a decentralized manner. During the operation of Algorithm 1, there always exists a $\delta > 0$ such that*

$$\delta \mathcal{L}(\hat{z}^{+}, \hat{\lambda}^{+}) \leq 4 \sum_{i=1}^{N} \frac{1}{L_i} \|g_i - g_i^*\|^2 . \qquad (23)$$

As a result, we have that during the operation of Algorithm 1 the inequality (22) *in Theorem 1 is satisfied for every node.*

VI. Numerical Simulation

In this section, we present numerical simulations to demonstrate the operation of Algorithm 1 and to highlight the improvements it offers over existing distributed optimization algorithms.

We focus on a random digraph consisting of 20 nodes. Each node i is associated with a function $f_i(z) = \frac{1}{2} z^\top P_i z + p_i^\top z$ where $z \in \mathbb{R}^n$, $P_i \in \mathbb{S}^n_{++}$, and $p_i \in \mathbb{R}^n$ for each node $i \in \mathcal{V}$, with $n = 20$. Furthermore, we have $\rho = 1$ and that Assumptions 1 and 2 hold. For each node i, $P_i \succ 0$ was initialized as the square of a randomly generated symmetric matrix A_i, which guarantees that it is positive definite. Additionally, q_i is set as the negative of the product of the transpose of A_i and a randomly generated vector b_i (i.e., it represents a linear term). For further details, please refer to [22, Section VI]. We execute Algorithm 1 and show how the nodes' decision variables convergence to the optimal solution for $\Delta = 10^{-3}, 10^{-4}$, and 10^{-5}, respectively. We plot the error $\sum_{i=1}^{N} \left\| x_i^{[k]} - x_i^* \right\|$ where $x_i^* = z^*, \forall i \in \mathcal{V}$ represents the optimal solution of problem (4).

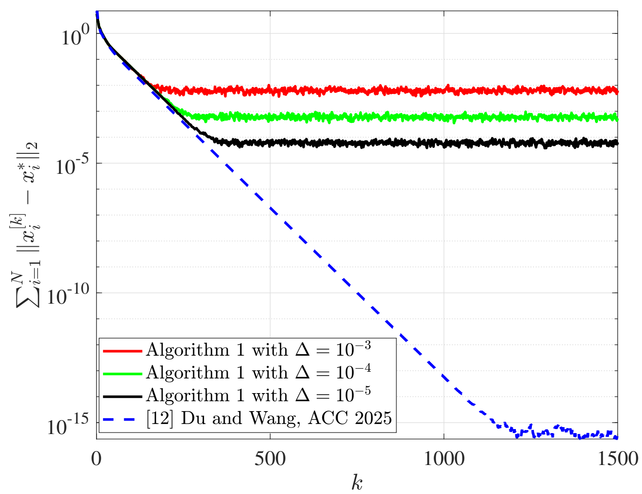

Fig. 1. Comparison of Algorithm 1 with RC-ALADIN [12] over a directed graph with quantization level $\Delta = 10^{-3}, 10^{-4}$ and 10^{-5}.

In Fig. 1 we can see that Algorithm 1 converges to the optimal solution, achieving an approximation precision that is directly influenced by the quantization level Δ. Specifically, a smaller Δ enables nodes to approximate the optimal solution with greater accuracy, a behavior consistent with the theoretical results presented in Theorem 1. The oscillatory behavior observed in the convergence is attributed to the nonlinearities introduced by quantized communication, which affect the values of parameters such as x_i, g_i, and $\hat{\lambda}_i$. While Algorithm 1 demonstrates comparable performance to the approach in [12] up to the neighborhood of the optimal solution, [12] relies on a distributed framework with a server node and real-valued message exchanges among nodes. Thus, the approach in [12] introduces scalability challenges, suffers from single-point-of-failure risks, limits its applicability to bandwidth-constrained environments and compromises resource efficiency. In contrast, Algorithm 1 offers a significant advantages. It provides comparable performance with algorithms in the literature that enable nodes to exchange real-valued messages (see [12]), while exhibiting efficient (quantized) communication among nodes (a larger Δ implies a reduced communication bandwidth requirement among nodes) thereby reducing data transmission. Moreover, Algorithm 1 operates in a fully decentralized manner, eliminating the need for a server node. A numerical example in [13, Section 5.1] investigates the role of the penalty parameter ρ in Algorithm 1, applied to a consistent convex problem under Assumptions 2 and 3. The example demonstrates that ρ influences the observed linear convergence rate. Nevertheless, a comprehensive theoretical characterization of this dependence for the full class of problems satisfying the same assumptions remains an open question.

VII. Conclusions and Future Directions

In this paper, we presented a novel decentralized optimization algorithm named QuDRC-ALADIN. Our algorithm enables computation of the optimal solution in a fully decentralized manner over directed communication networks, while ensuring efficient quantized communication among nodes. We analyzed our algorithm's operation and established its linear convergence to a neighborhood of the optimal solution that depends on the utilized quantization level. Finally, we presented numerical simulations validating our algorithm's performance, and highlighted its advantages compared to existing algorithms in the literature.

References

[1] G. Stomberg, A. Engelmann, M. Diehl, and T. Faulwasser, "Decentralized real-time iterations for distributed nonlinear model predictive control," *arXiv preprint arXiv:2401.14898*, 2024.

[2] L. Yuan, Z. Wang, L. Sun, P. S. Yu, and C. G. Brinton, "Decentralized federated learning: A survey and perspective," *IEEE Internet of Things Journal*, vol. 11, no. 21, pp. 34 617–34 638, 2024.

[3] D. Liang, S. Su, L. Zeng, and H.-D. Chiang, "Decentralized method for nonconvex robust static state estimation of integrated electricity-gas systems," *CSEE Journal of Power and Energy Systems*, pp. 1–13, 2024.

[4] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[5] S. Wang, F. Roosta, P. Xu, and M. W. Mahoney, "GIANT: Globally improved approximate newton method for distributed optimization," in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[7] B. Houska, J. Frasch, and M. Diehl, "An augmented Lagrangian based algorithm for distributed nonconvex optimization," *SIAM Journal on Optimization*, vol. 26, no. 2, pp. 1101–1127, 2016.

[8] Q. Ling, W. Shi, G. Wu, and A. Ribeiro, "DLM: Decentralized linearized alternating direction method of multipliers," *IEEE Transactions on Signal Processing*, vol. 63, no. 15, pp. 4051–4064, 2015.

[9] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations research*, vol. 8, no. 1, pp. 101–111, 1960.

[10] B. He and X. Yuan, "On the O(1/n) convergence rate of the douglas–rachford alternating direction method," *SIAM Journal on Numerical Analysis*, vol. 50, no. 2, pp. 700–709, 2012.

[11] 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.

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

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

[14] A. Engelmann, Y. Jiang, T. Mühlpfordt, B. Houska, and T. Faulwasser, "Toward distributed OPF using ALADIN," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 584–594, 2019.

[15] X. Du, A. Engelmann, Y. Jiang, T. Faulwasser, and B. Houska, "Distributed state estimation for AC power systems using Gauss-Newton ALADIN," in *IEEE Conference on Decision and Control*, 2019, pp. 1919–1924.

[16] A. Falsone and M. Prandini, "Augmented Lagrangian tracking for distributed optimization with equality and inequality coupling constraints," *Automatica*, vol. 157, p. 111269, 2023.

[17] W. Jiang, A. Grammenos, E. Kalyvianaki, and T. Charalambous, "An asynchronous approximate distributed alternating direction method of multipliers in digraphs," in *IEEE Conference on Decision and Control*, 2021, pp. 3406–3413.

[18] 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*, vol. 7, no. 4, pp. 1848–1858, 2020.

[19] 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.

[20] S. Zhu, M. Hong, and B. Chen, "Quantized consensus ADMM for multi-agent distributed optimization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 4134–4138.

[21] A. Elgabli, J. Park, A. S. Bedi, C. B. Issaid, M. Bennis, and V. Aggarwal, "Q-GADMM: Quantized group ADMM for communication efficient decentralized machine learning," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 164–181, 2020.

[22] A. I. Rikos, W. Jiang, T. Charalambous, and K. H. Johansson, "Asynchronous distributed optimization via ADMM with efficient communication," in *IEEE Conference on Decision and Control*, 2023, pp. 7002–7008.

[23] ——, "Distributed optimization with efficient communication, event-triggered solution enhancement, and operation stopping," *arXiv preprint arXiv:2504.16477*, 2025.

[24] J. G. Proakis and M. Salehi, *Communication Systems Engineering*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall, 2002.

[25] J. Wei, X. Yi, H. Sandberg, and K. H. Johansson, "Nonlinear consensus protocols with applications to quantized communication and actuation," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 2, pp. 598–608, 2019.

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

[27] G. Oliva, R. Setola, and C. N. Hadjicostis, "Distributed finite-time calculation of node eccentricities, graph radius and graph diameter," *Systems & Control Letters*, vol. 92, pp. 20–27, 2016.

[28] D. Deplano, N. Bastianello, M. Franceschelli, and K. H. Johansson, "Optimization and learning in open multi-agent systems," *arXiv preprint arXiv:2501.16847*, 2025.

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

[30] J. B. Rawlings, D. Q. Mayne, and M. Diehl, *Model Predictive Control: Theory, Computation, and Design, 2nd Edition*. Nob Hill Publishing, 2017.

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

[32] A. I. Rikos, C. N. Hadjicostis, and K. H. Johansson, "Non-oscillating quantized average consensus over dynamic directed topologies," *Automatica*, vol. 146, p. 110621, 2022.

[33] A. I. Rikos, A. Grammenos, E. Kalyvianaki, C. N. Hadjicostis, T. Charalambous, and K. H. Johansson, "Distributed optimization for quadratic cost functions with quantized communication and finite-time convergence," *IEEE Transactions on Control of Network Systems*, vol. 12, no. 1, pp. 930–942, 2025.

[34] A. I. Rikos, W. Jiang, T. Charalambous, and K. H. Johansson, "Distributed optimization via gradient descent with event-triggered zooming over quantized communication," in *IEEE Conference on Decision and Control*, 2023, pp. 6321–6327.

[35] ——, "Distributed optimization with gradient descent and quantized communication," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 5900–5906, 2023.

[36] X. Du, K. H. Johansson, and A. I. Rikos, "Decentralized optimization via RC-ALADIN with efficient quantized communication," *arXiv preprint arXiv:2508.06197*, 2025.

[37] A. I. Rikos, W. Jiang, T. Charalambous, and K. H. Johansson, "Distributed optimization with finite bit adaptive quantization for efficient communication and precision enhancement," in *IEEE Conference on Decision and Control*, 2024, pp. 2531–2537.