Quantized Primal-Dual Algorithms for Network Optimization With Linear Convergence

Ziqin Chen, Shu Liang, Li Li, Member, IEEE, and Shuming Cheng

Abstract—This note investigates a network optimization problem in which a group of agents cooperate to minimize a global function under the practical constraint of finite-bandwidth communication. We propose an adaptive encoding–decoding scheme to handle the quantization communication between agents. Based on this scheme, we develop a continuous-time quantized distributed primal-dual algorithm for the network optimization problem. Our algorithm achieves linear convergence to an exact optimal solution. Furthermore, we obtain the relationship between the communication bandwidth and the convergence rate. Finally, we use a distributed logistic regression problem to illustrate the effectiveness of our methods.

Index Terms—Distributed convex optimization, linear convergence rate, primal-dual algorithm, quantized communication.

I. INTRODUCTION

Network optimization problems have been intensely studied in various fields, such as large-scale machine learning [1] and economic dispatch in power systems [2]. Typically, they are formulated as minimizing the sum of local functions, each of which is assessed by a local agent in the network [3], [4], [5], [6]. Most distributed optimization algorithms do not consider communication bandwidth, which are always limited by practical factors, such as sensor battery power and computing resource. Regarding the problem of finite-bandwidth constraints, an effective method is to use quantization schemes that are capable of compressing messages exchanged among agents.

Quantization in the context of network optimization was first studied for distributed consensus problems [7], [8], [9], in which zoom-in-zoom-out quantization techniques were proposed. Later, the encoding–decoding schemes [10], [11], [12], the unbiased [13], [14], [15], and biased compression methods [16] were studied for distributed algorithms. These works achieve linear convergence rates under strong convexity assumptions. The quantized algorithms given in [17] and [18] do not require strong convexity, whereas they only achieved sublinear rates $O(\ln k/\sqrt{k})$ and $O(1/k)$. Note that, the authors in [10], [11], [12], [13], [14], [15], [17], and [18] focus on discrete-time quantized optimization algorithms.

Continuous-time optimization algorithms are also widely studied as considered in cyber-physical systems, while their quantization faces further challenges. Specifically, stabilizing the sampling and quantization error during the continuous evolution time rather than the discrete sampling instances is one of the major differences between quantized continuous-time algorithms and quantized discrete-time ones. This continuous-time error term can affect the convergence and accuracy of the algorithm, even though it vanishes at sampling instances. Eliminating the effect of this error requires trajectory prediction that depends on global information about cost functions and the communication graph. In addition, the Lyapunov function of the original continuous-time dynamics may increase due to this error term, and the monotonically decreasing Lyapunov function-based convergence analysis becomes inapplicable.

In this work, we propose a continuous-time quantized distributed primal-dual (QDPD) algorithm with a novel encoding–decoding scheme. It achieves a linear convergence rate for the network convex optimization with finite bandwidth under metric subregularity assumption. Furthermore, the relationship between the convergence rate and the required bandwidth is quantitatively analyzed. The main contributions can be summarized as follows.

1) We propose a continuous-time QDPD algorithm to solve a non-strongly convex optimization problem and achieve a linear convergence rate. We introduce an adaptive encoding–decoding scheme to overcome finite-bandwidth constraints and guarantee the accuracy of the algorithm. To the best of authors’ knowledge, achieving linear and exact convergence of continuous-time quantized algorithms for the nonstrongly convex optimization with finite-bandwidth communication has not been reported in the literature.

2) In the algorithm design, our method stabilizes the sampling and quantization error during all the evolution time, which is stronger than [10], [11], [12], [13], [14], [15], [17], [18] that focus on time instances only. We achieve the results by predicting the trajectory of dynamics and its performance in advance and also provide distributed estimation for key parameters. In the convergence analysis, we employ an envelop function in addition to the Lyapunov function to establish the linear convergence. The result is stronger than the linear convergence result in [19] without quantization.

3) The relationship between communication bandwidths and convergence rates is clearly revealed. In particular, given the communication bandwidth, we can find a lower bound for the convergence rate of our algorithms. It gives insights into transmission bandwidth configurations by different convergence rate requirements.

The rest of this note is organized as follows. Section II formulates the distributed convex optimization problem. Section III introduces the adaptive encoding–decoding scheme and the QDPD algorithm with convergence analysis. Section IV gives an example to validate our results. Section V concludes this work.
Notations: For a scalar $a \in \mathbb{R}$, $\lfloor a \rfloor$ is the smallest integer not smaller than $a$. For vectors $x_1, \ldots, x_N$, denote the stacked column vector as $x = [x_i]_{i \in \{1, \ldots, N\}}$. For $x \in \mathbb{R}^n$, $\|x\|$ represents its 2-norm. For a set $C \subseteq \mathbb{R}^n$ and a vector $x \in \mathbb{R}^n$, denote $d(x, C) = \inf_{z \in C} \|y - x\|$. For a continuously differentiable function $f(x)$, denote by $\nabla f(x)$ its gradient vector. For a map $F : \mathbb{R}^n \to \mathbb{R}^m$, $\text{gph} F = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m | y = F(x)\}$.

II. PROBLEM FORMULATION

Consider a network optimization problem in which $N$ agents cooperate to optimize a global cost function as

$$\min_{x \in \mathbb{R}^n} f(x), \quad f(x) = \sum_{i=1}^{N} f_i(x).$$

Each local cost function $f_i(x) : \mathbb{R}^n \to \mathbb{R}$ is processed by the agent $i$ in the network and satisfies the following assumption.

Assumption 1: Each local cost function $f_i(x)$ is differentiable and convex. Moreover, $\nabla f_i(x)$ is $m_{f_i}$-Lipschitz continuous for some constants $m_{f_i} > 0$.

The communication network is described by an undirected connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a node set with the cardinality $|\mathcal{V}| = N$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is an edge set. Let $A = [a_{ij}]_{N \times N}$ be the adjacency matrix of $\mathcal{G}$ with $a_{ij} = 1$ if $(i, j) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise. For agent $i$, $N_i = \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$ and $|N_i| = d_i$. The Laplacian matrix is $L_0 = D - A$ with $D = \text{diag}(d_1, \ldots, d_N)$, and its eigenvalues are arranged in ascending order as $0 = \sigma_1 < \sigma_2 \leq \cdots \leq \sigma_N$.

Under a connected graph, problem (1) is equivalent to

$$\min_{x \in \mathbb{R}^n} f(x), \quad f(x) = \sum_{i=1}^{N} f_i(x)$$

subject to $L_0 x = 0_{Nn}$

where $x = [x_i]_{i \in \mathcal{V}}$ and $L_0 = L_0 \otimes I_n$. Define a dual variable $\lambda = [\lambda_i]_{i \in \mathcal{V}}$. The dual problem of (2) is described as

$$\max_{\lambda \in \mathbb{R}^n} q(\lambda), \quad q(\lambda) = \min_{x \in \mathbb{R}^n} \left\{ f(x) + \lambda^T L_0 x \right\}$$

and an augmented Lagrangian function of (2) is

$$L(\lambda, x) = f(x) + \lambda^T L_0 x + \frac{1}{2} x^T L_0^2 x.$$ 

By KKT conditions, $(x^*, \lambda^*)$ is a pair of optimal primal-dual solutions if and only if

$$0_{Nn} = -\nabla f(x^*) + L_0 x^* - L_0 \lambda^*$$

Define a gradient map $F(z)$ with $z = [x; \lambda] \in \mathbb{R}^{2Nn}$ as

$$F(z) = \begin{bmatrix} \nabla f(x) + L_0 x + L_0 \lambda \\ -L_0 x \end{bmatrix} \in \mathbb{R}^{2Nn}. $$

We impose an assumption on $F(z)$ in the following.

Assumption 2: $F(z)$ in (5) is $\kappa$-metrically subregular at point $(z^*, 0_{2Nn}) \in \text{gph} F$ for some constants $\kappa > 0$, that is, there exists an open set $C \supset Z^*$ such that

$$\|F(z)\| \geq \kappa^{-1} d(z, Z^*) \quad \forall z \in C$$

where $z^*$ is the projection of a point $z$ to the closed and convex $Z^* = \{z | F(z) = 0_{2Nn}\}$.

According to [19], Assumption 2 can be satisfied when $f(x)$ is quadratic but not necessarily strongly convex, and when $f(x)$ is twice differentiable and has an unique optimal solution $x^*$. A few nonstrongly convex functions, such as linear–quadratic continuous functions and logistic regression functions, satisfy Assumption 2.

In addition, the following assumption is needed in our design to handle finite-bandwidth constraints.

Assumption 3: The optimal solution set of each cost function $f_i(x)$, defined as $X_i = \arg\min_{x \in \mathbb{R}^n} f_i(x)$, is compact, nonempty, and bounded. That is, $\|\hat{x}_i\| \leq M_i$ for any $\hat{x}_i \in X_i$ and some constant $M_i > 0$.

Assumption 3 has been used in [5] to guarantee $\|x^*\| \leq M_1$ for some $M_1 > 0$, where $x^*$ is any optimal solution to (2). In the following lemma (whose proof can be found in Appendix A), we calculate an upper bound of $\|x^*\|$, which is used to design an adaptive length of quantization ranges in the encoding–decoding scheme to prevent quantizer saturation.

Lemma 1: Under Assumptions 1–3, the upper bound of the norm of an optimal variable $z^*$ satisfies

$$\|z^*\| \leq M_1 + M_2$$

where $M_1 \geq \|x^*\|$ and $M_2 = \frac{\sqrt{2\lambda(0)}}{\sqrt{N_n}} + \frac{\sqrt{N_n}M_0}{\sigma_2} + M_1$ with $M_0 = \max_{i \in \mathcal{V}} \{m_{f_i} (\frac{M_i}{\sqrt{\sigma_2}} + M_i)\}$ and the smallest nonzero eigenvalue of the Laplacian matrix $L_0$ is $\sigma_2$.

III. MAIN RESULT

In this section, we design a continuous-time QDPD algorithm with an encoding–decoding scheme in Section III-A. Then, we provide distributed estimation methods for key parameters $\kappa$ and $M_1$ in Section III-B. Next, we give the convergence analysis in Section III-E. Finally, we present the relationship between convergence rates and communication bandwidth in Section III-F.

A. QDPD Algorithm Design

The QDPD algorithm can be decomposed into the following two steps.

1) Quantized Communication Step: In the quantized communication step, both primal and dual variables are quantized and transmitted over the network graph $\mathcal{G}$. At time instance $k$, the quantized state of a scalar $s \in [a(k), b(k)]$ is written as

$$Q^L_{a(k), b(k)}(s) = \arg \min_{0 \leq \hat{s} \leq L} \left\{ \left| a(k) + \frac{b(k) - a(k)}{L} \hat{s} - s \right| \right\}$$

where $s$ is encoded into the set $\{0, \ldots, L\}$ and requires $\log_2(L)$ bits to transmit if no transmission exists at zero level.

i) Encoder: We use periodic sampling with the period $T$ to sample continuous-time inputs of an encoder $z_j(t) = [z_j(t); \lambda_j(t)] \in \mathbb{R}^{2n}$, $t \geq 0$ as the discrete variables $z_j(kT) = [z_{j,p}(kT)]_{p=1,2,\ldots,2n}, k \in \mathbb{N}$ for any $j \in \mathcal{V}$. The outputs of the encoder $S_j(z_j(kT)) = [S_{j,p}(z_j(kT))]_{p=1,2,\ldots,2n} \in \mathbb{R}^{2n}$ satisfy

$$S_j(z_j(kT)) = \left[ Q_{a(j), b(j)}^L(z_{j,p}(kT)) \right]_{p=1,2,\ldots,2n}$$

where the quantization range $P^z_{j}(k) = [P^z_{j,p}(k)]_{p=1,2,\ldots,2n} = [a_j(k), b_j(k)]$ with $a_j(k), b_j(k) \in \mathbb{R}^{2n}$ will be explicitly given in the following section. Following that, each agent $j$ encodes $S_j(z_j(kT))$ into $2n \log_2(L)$ bits that are sent to its neighbors.

ii) Decoder: If the agent $i$ receives quantized information $S_j(z_j(kT))$, it estimates neighbor’s states $z_j(kT)$ through

$$q^0_j(kT) = a_j(k) + S_j(z_j(kT)) b_j(k) - a_j(k) \quad L.$$
with $S_j(z_j(kT)) = \text{diag}(S_{j,1}(z_{j,1}(kT)); \cdots; S_{j,2n}(z_{j,2n}(kT)))$. By using zero-order hold methods, the agent $i$ recovers the continuous-time signals from the discrete-time signals via

$$q_i^j(t) = q_i^j(kT), \quad kT \leq t < (k + 1)T, \quad j \in \mathcal{N}_i \cup \{i\}. \quad (8)$$

2) **State Update Step:** Following the quantized communication step, each agent updates its real-time state via

$$\begin{align*}
\dot{x}_i(t) &= \alpha \left[ -\nabla f_i(x_i(t)) - \sum_{j \in \mathcal{N}_i} \left( q_i^j(t) - q_j^i(t) + q_j^i(t) \right) \right] \\
\dot{\lambda}_i(t) &= \alpha \sum_{j \in \mathcal{N}_i} (q_i^j(t) - q_j^i(t))
\end{align*} \quad (9)$$

where $\alpha$ is a positive constant.

The combination of Steps 1 and 2 brings the QDPD algorithm, and the key parameters are designed as follows.

i) The sampling period $T$ is any positive constant satisfying

$$e^{\alpha \sigma_N T - 1} \left( e^{\frac{\rho c_1}{2} T} - 1 \right) \rho \leq c_1 < 1 \quad (10)$$

where $c = \sqrt{\frac{3\pi^2}{2}}, \quad \eta = \sqrt{\frac{\beta}{\kappa (m_f + \alpha \sigma_N T)}}, \quad m_f = \sum_{i=1}^{N} m_i, \quad c_1$, and $\beta$ are chosen as any positive constant smaller than 1, and

$$\rho = \frac{\alpha (6\sigma_N + 2m_f)}{\epsilon \eta} \sqrt{\frac{12\sigma_N + 33}{\epsilon \eta (1 - \beta) - 4\sigma_N \beta}}.$$ 

Inequality (10) can be simplified as

$$T \leq \sqrt{\frac{2c_1 (2\sigma_N)}{\alpha^2 \sigma_N \eta}}. \quad (11)$$

Since $\lim_{T \to 0} (e^{\alpha \sigma_N T} - 1) / (e^{\frac{\rho c_1}{2} T} - 1) \rho = 0$, (10) holds if the sampling period $T$ is sufficiently small. In practice, the sampling period $T \ll 1$, and thus, $(e^{\alpha \sigma_N T} - 1) / (e^{\frac{\rho c_1}{2} T} - 1) \approx \alpha^2 \sigma_N \eta T^2 / 2$, which yields (11).

ii) The adaptive length of quantization ranges satisfies

$$l(k) = l(0) e^{\frac{\alpha c_1}{2} T k} \quad (12)$$

where

$$l(0) = \frac{c_2 M}{\kappa \sigma_N} \sqrt{\frac{6n (1 - \beta - 8\beta)}{3Nn (4\sigma_N + 11)}} \quad (13)$$

with $c_2 = 1 - c_1, \quad M = M_1 + M_2 + M_3$, and $M_3 \geq \| z(0) \|$. iii) The number of quantization levels $L$ could be any positive constant satisfying

$$L \geq \max \left\{ \frac{2M_3}{l(0)}, \frac{\sqrt{2Nn}}{c_2} e^{\frac{\alpha c_1}{2} T k} \right\}. \quad (14)$$

iv) The dynamic quantization range is determined by

$$P_j^i(0) = \left[ -\frac{L(0)}{2} 1_{2n}, \frac{L(0)}{2} 1_{2n} \right]$$

$$P_j^i(k+1) = \left[ q_i^j(kT) - \frac{L(k+1)}{2} 1_{2n}, q_i^j(kT) + \frac{L(k+1)}{2} 1_{2n} \right].$$

**Remark 1:** The novel encoding–decoding scheme relies on the design of the adaptive length of quantization ranges. On the one hand, a linearly decaying $l(k)$ is significant to guarantee the convergence of the quantization error and the QDPD algorithm. On the other hand, $l(k)$ cannot converge “too fast” to avoid the saturation of the quantizer, whose design depends on the sampling and quantization error $e_{z_i} = z_i - q_i^j$ with $z_i = [x_i; \lambda_i]$. This error is related to the updated trajectory of the algorithm. To predict this trajectory, the global parameters $\kappa$ and $M_1$ are needed. As shown in our design, these constants are used to determine $T$ in (10) and $l(k)$ in (12). Hence, in the following section, we provide distributed estimation methods for global parameters $\kappa$ and $M_1$ in specific situations.

B. **Distributed Estimations for Key Parameters**

This section provides distributed estimation methods for the parameters $\kappa$ and $M_1$ in the encoding–decoding scheme.

1) **$\kappa$ Estimation:** If $f(x)$ in (2) has a unique solution and is twice differentiable, then $1/\kappa$ equals the lower bound of the smallest nonzero singular value of the following matrix according to [19, Corollary 2]:

$$M(x^*) = \begin{bmatrix} H(x^*) + L_G & \mathbf{L}_G \\ \mathbf{L}_G & 0 \end{bmatrix}.$$ 

where $H(x^*) = \text{diag}(c_1(x^*), \ldots, c_N(x^*))$ is the Hessian matrix of $f(x^*)$. Here, we consider the case $n = 1$ since the results for $n > 1$ are the same. Since $L_G$ is symmetrical, the matrix of $M(x^*)^T M(x^*)$ is similar to the following matrix:

$$\begin{bmatrix} c_i^2(x^*) + \sigma_i^2 & \cdots & 0 & \bar{e}_1(x^*) \sigma_i & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \ddots \end{bmatrix}$$

where $\bar{e}_i(x^*) = e_i(x^*) + \sigma_i$, and the eigenvalue of $L_G$ is $\sigma_i$. Denote by $\lambda_i(x^*)$, $i = 1, \ldots, 2N$ the singular values of $M(x^*)$. By the similarity, the singular values $\lambda_i(x^*)$ satisfy

$$\lambda_i^2(x^*) = \lambda_i^2(x^*) - (\bar{e}_i(x^*) + \sigma_i)^2 - 2\sigma_i^2 + \sigma_i^4 = 0.$$

By Assumption 1, we get $c_i(x^*) \leq m_j < m_f$ with $m_f = \sum_{i=1}^{N} m_f$. Then, the nonzero singular values $\lambda_i(x^*)$ satisfy

$$\lambda_i^2(x^*) \geq \frac{\sigma_i^2}{(\max_{i \in V} c_i(x^*))^2 + 2\sigma_i^2} \geq \frac{16 (1 - \cos \left( \frac{\pi}{N} \right))^4}{(m_f + 2 (1 - \cos \left( \frac{\pi}{N} \right))^2 + 2 (1 - \cos \left( \frac{\pi}{N} \right))^2).$$

Consequently, $\kappa$ can be chosen as

$$\kappa \geq \sqrt{\frac{16 (1 - \cos \left( \frac{\pi}{N} \right))^4 + 2 (1 - \cos \left( \frac{\pi}{N} \right))^2}{16 (1 - \cos \left( \frac{\pi}{N} \right))^4}}. \quad (14)$$

Since $m_f$ and $N$ can be estimated using distributed average methods, the constant $\kappa$ can also be calculated in a distributed manner by virtue of (14).

2) **$M_1$ Estimation:** According to [5, Lemma 4], $f_i(x)$ is coercive, and the corresponding level set $X_i = \{ x \in \mathbb{R}^n | f_i(x) \leq f_i(0) \}$ is bounded. Then, $M_1$ is selected as the radius of any open ball that contains this level set $X_i$ and takes the origin as its center. Similar to [20], we estimate $M_1 = \sqrt{N} \max \{ M^i \}$, $i \in V$ by distributed average methods.
C. Convergence Analysis

In this section, the convergence of the QDPD algorithm is analyzed. Under Assumption 1, the QDPD algorithm with the dynamics (9) can be written in a stacked form with $z = [x; \lambda]$

$$
\dot{z}(t) = -\alpha \left[ \begin{array}{c} 1_{Nn} \\ 0_{Nn} \end{array} \right] \nabla f(x(t)) - \alpha \left[ \begin{array}{cc} L_g & L_g' \\ -L_g & 0_{Nn} \end{array} \right] z(t) \\
+ \alpha \left[ \begin{array}{cc} L_g & L_g' \\ -L_g & 0_{Nn} \end{array} \right] (z(t) - q^*(t))
$$

(15)

where $q^*(t) = [q^+_i(t)]_{i \in V}$ is generated by

$$
\begin{align*}
q^+(t) &= q^+(kT), kT \leq t < (k + 1)T \\
q^+(k + 1)T &= q^+(kT) - \frac{L(k + 1)}{2} + S(z((k + 1)T))
\end{align*}

(16)

where $S(z(kT)) = \text{diag}(S_1(z_1(kT)); \ldots; S_N(z_N(kT)))$.

If $q^*(t)$ tends to $z^*$, then (15) at this equilibrium coincides with the KKT condition (4), which means that $z^*$ is an optimal primal-dual solution. We will prove both the convergence of $z(t)$ and $q^*(t)$ in (15) and (16).

Our main result for the convergence is given as follows.

**Theorem 1:** Under Assumptions 1–3, the QDPD algorithm with the dynamics (9) ensures that the states of all agents $x_i, i \in V$ converge to an optimal solution $x^*$ of problem (1) at a linear convergence rate $O(t^{-\gamma})$ with $\gamma = \frac{\alpha c \sigma}{\beta} > 1$, where $\alpha$ and $\eta$ are positive parameters given in (10) and (11).

Proof: Define $\tilde{z} = [\tilde{x}; \tilde{\lambda}], \varepsilon = z - q^*, \tilde{x} = x - x^*,$ and $\tilde{\lambda} = \lambda - \lambda^*$. By (4) and (15), we obtain error dynamics as

$$
\begin{align*}
\dot{\tilde{z}} &= -\alpha \left[ \begin{array}{c} 1_{Nn} \\ 0_{Nn} \end{array} \right] \left( \nabla f(x) - \nabla f(x^*) \right) - \alpha \left[ \begin{array}{cc} L_g & L_g' \\ -L_g & 0_{Nn} \end{array} \right] \tilde{z} \\
&+ \alpha \left[ \begin{array}{cc} L_g & L_g' \\ -L_g & 0_{Nn} \end{array} \right] e.
\end{align*}

(17)

Then, the convergence of our QDPD algorithm with the dynamics (9) is equivalent to the stability of (17). Employ a Lyapunov candidate function as

$$V(\tilde{z}) = 4\sigma_N V_1(\tilde{z}) + V_2(\tilde{z})
$$

(18)

where

$$
\begin{align*}
V_1(\tilde{z}) &= \frac{1}{2} \| \tilde{z} \|^2 \\
V_2(\tilde{z}) &= \tilde{f}(\tilde{x}) + \frac{3}{2} \tilde{x}^T L_g \tilde{x} + \tilde{\lambda}^T L_g \tilde{x} + \tilde{x}^T L_g \tilde{x}
\end{align*}

$$

(19)

with $\tilde{f}(\tilde{x}) = f(\tilde{x}) + \frac{3}{2} \tilde{x}^T L_g \tilde{x} + \tilde{\lambda}(L_g \tilde{\lambda})$.

First, we examine the positive definiteness of $V(\tilde{z})$. It follows from (4) that

$$
\tilde{x}^T \nabla f(x^*) = -\tilde{x}^T L_g \tilde{\lambda}
$$

and

$$
\begin{align*}
V_2(\tilde{z}) &= \tilde{f}(\tilde{x}) - \tilde{x}^T \nabla f(x^*) + \frac{1}{2} \tilde{x}^T L_g \tilde{x} + \tilde{\lambda}^T L_g \tilde{x}
\end{align*}

(19)

By the convexity of $f(x)$, we have $\tilde{f}(\tilde{x}) - \tilde{x}^T \nabla f(x^*) \geq 0$. Since $\frac{3}{2} \tilde{x}^T L_g \tilde{x} \geq 0$, it follows that $V_2(\tilde{z}) \geq -\frac{3\sigma_N}{2} (\| \tilde{z} \|^2 + \| \tilde{\lambda} \|^2) \geq -\frac{3\sigma_N}{2} \| \tilde{z} \|^2$. Therefore

$$
V(\tilde{z}) \geq \frac{3\sigma_N}{2} \| \tilde{z} \|^2 \geq 0.
$$

(20)

Next, we show that $V(\tilde{z})$ is an ISS Lyapunov candidate function of the dynamic (17) that satisfies

$$
V \leq \alpha \left[ -\eta V - \frac{(1 - \beta)\eta}{\beta} V + \left( 2\sigma_N + \frac{11}{2} \sigma_N^2 \right) \| e(t) \|^2 \right].
$$

(21)

The proof of (21) is given in Appendix B. According to (21), the Lyapunov function $V(\tilde{z})$ converges to zero only when the norm of estimation errors $\| e(t) \|$ is vanishing. Hence, the remaining is to show the convergences of both $\| e(t) \|$ and $V(\tilde{z})$. To this end, we use a function-enveloped method with defining the following two linearly decaying functions:

$$
a(t) = \frac{\beta M^2}{2\eta_2 e^{-\alpha t} T} \\
b(t) = \frac{M}{\kappa \sigma_N} \frac{3\eta(1 - \beta) - 4\beta}{3\eta(4\sigma_N + 11)} e^{-\frac{T}{T(t)} t}
$$

(22)

where the parameter definitions are the same as those in (10)–(12). Then, the proof of linear convergence of $\| e(t) \|$ and $V(\tilde{z})$ is equivalent to prove that $\| e(t) \|$ and $V(\tilde{z})$ are bounded by

$$
\begin{align*}
V(\tilde{z}(t)) &\leq a(t), 0 \leq t \leq kT, k \in \mathbb{N} \\
\| e(kT) \| &\leq c_2 e^{-c_1 \kappa T} b(kT) \\
\| e(t) \| &\leq b(t), kT \leq t < (k + 1)T, k \in \mathbb{N}
\end{align*}

(22)

where $c_2 = 1 - c_1 < 1$ is defined in (12) and the proof of (22) is given in Appendix C.

The proof of linear convergence of $\| e(t) \|$ and $V(\tilde{z})$ is given in Appendix B.
D. Bandwidth Analysis

We analyze the relationship between the communication bandwidth and the convergence rate of the QDPD algorithm.

**Theorem 2:** Under Assumptions 1–3, the following properties hold.

i) The QDPD algorithm linearly converges to an optimal solution of problem (1) under any positive bandwidth.

ii) The relationship between the communication bandwidth $B$ and the convergence rate $\gamma$ satisfies

$$B \leq C_0 \ln \gamma + C_1$$

where the constants $C_0$ and $C_1$ are selected as

$$C_0 = (\log_2 e ) \left( 1 + \frac{\sqrt{6} + \sqrt{5}\sigma_N}{\eta} \right) + \frac{\sqrt{24 + 8\sqrt{5}\sigma_N + 2\eta}}{\eta \ln \rho_0}$$

$$C_1 = \left( \frac{1}{2} \log_2(2Nn) - \log_2 c_2 \right) \frac{2c_2}{\rho \rho_0 \eta} \quad \forall \rho_0 > 1.$$ 

**Proof:**

i) Assume that $\gamma$ in (10) is a fixed constant. For any $T > 0$, there is a sufficiently small constant $\alpha > 0$ such that (10) holds. The quantization level $L$ is estimated via (13) as

$$\lim_{\alpha \rightarrow 0} \left( \sqrt{2Nn}/c_2 \right) e^{-\alpha \sigma_N T} = \sqrt{2Nn}/c_2$$

which means that transmitted information can be represented by $\log_2 \left( \max \left( \left\lfloor \frac{2\alpha \sigma_N T}{c_2} \right\rfloor \right) \right)$ bits at each communication instant. Because $T$ can be any positive constant, the upper bound of the minimum bandwidth $B = \log_2 \left( \max \left( \left\lfloor \frac{2\alpha \sigma_N T}{c_2} \right\rfloor \right) \right)/T$ can be any positive constant as well. It implies that the QDPD algorithm with the dynamics (9) achieves linear convergence for any positive bandwidth.

ii) We compute an upper bound of the minimum bandwidth $B_\alpha$ for a fixed convergence rate $\gamma_\alpha = e^{-2\alpha}$. Let

$$T_\alpha = \frac{1}{\sqrt{24 + 8\sqrt{5}\sigma_N + 2\eta}} \ln \gamma_\alpha + \frac{2c_2}{\rho \rho_0 \eta} \quad \forall \rho_0 > 1$$

which satisfies

$$T_\alpha \leq \frac{2\eta \ln \rho_0}{\sqrt{24 + 8\sqrt{5}\sigma_N + 2\eta}} \ln \gamma_\alpha + \frac{2c_2}{\rho \rho_0 \eta} \quad \forall \rho_0 > 1$$

Furthermore, $e^{\alpha} - 1 \leq e^{\alpha} - 1 \leq e^{\alpha}$ indicate that

$$e^{\alpha \sigma_N T_\alpha} - 1 \left( e^{\alpha \sigma_N T_\alpha} - 1 \right) /\alpha \leq \frac{3T_\alpha}{2} e^{\alpha \sigma_N T_\alpha + \gamma_\alpha T_\alpha}.$$  

Invoking (25) to (26) yields $e^{\alpha \sigma_N T_\alpha + \gamma_\alpha T_\alpha} < \rho_0$, which ensures $T_\alpha$ satisfying (10). By (13), we compute $B_\alpha$ via

$$B_\alpha = \frac{\log_2(L_\alpha)}{T_\alpha} = \frac{\log_2 \left( \frac{\sqrt{2Nn} e^{\alpha \sigma_N T_\alpha}}{c_2} \right)}{T_\alpha}.$$ 

Substituting (24) into (27) yields $B_\alpha = C_0 \ln \gamma_\alpha + C_1$.

Property i) provides a necessary and sufficient condition regarding the required bandwidth for the linear convergence of the QDPD algorithm. Specifically, Property i) provides its sufficiency. The necessity is deduced directly by Shannon’s rate-distortion theory, which claims that if a quantized distributed algorithm achieves linear convergence with the rate $\gamma$, the communication bandwidth $B > \gamma \log_2 e > 0$. In addition, under a fixed bandwidth $B_\alpha$, Property ii) shows that the QDPD algorithm always guarantees the linear convergence rate $\gamma_\alpha$.

**Remark 2:** The bandwidth $B_\alpha$ in (27) may need a few bits in practice, though it may be conservative. In fact, the value of $B_\alpha$ provides an upper bound of the minimum bits at each communication step. It partially addresses the minimum bandwidth problem in the quantized optimization problem.

IV. EXAMPLE

In this section, a distributed logistic regression problem is used to demonstrate our results. It is to learn a parameter $x$ as a minimizer to problem (1) with local cost functions $f_i$ as

$$f_i(x):= \frac{1}{N} \ln \left( 1 + \exp \left( -b_i a_i^T x \right) \right), \quad i \in \mathcal{V}$$

where $x \in \mathbb{R}^2$, $N = 12$, and $\{a_i, b_i\}_{i=1}^N \in \mathbb{R}^2 \times \{\pm 1\}$ are data samples. The entries of $a_i$ are generated with uniform distributions on the interval $[0,1]$. Here, Assumption 1 can be verified since the local cost functions are twice differentiable and their Hessian matrix is semipositive definite. Also, Assumption 2 holds with $\kappa = 0.5$ because the convex global function $f(x) = \sum_{i=1}^N f_i(x)$ has a unique optimal solution. The network graph is described as a ring, and the information transmission among agents is sampled and quantized. The following parameters are used in the encoding–decoding scheme:

i) the sampling period $T = 0.05s$;

ii) the dynamic length of quantization ranges $l(k) = 0.8e^{-0.1k}$;

iii) the number of quantization levels $L + 1 = 124$.

The initial states of all agents are set as $x_i(0) = [-7 + i; 12 - 2i], \quad i \in \{1, \ldots, 12\}$.

Fig. 1 demonstrates the convergence of the QDPD algorithm. Fig. 2 shows that our QDPD algorithm can achieve a similar tracking performance to that of the PD algorithm even with finite-bandwidth constraints. Fig. 3 implies that for any given convergence rate $\gamma$, the actual bandwidth $B$ is less than the upper bound of the minimum bandwidth $C_0 \ln \gamma + C_1$.
Define \( \vec{x}^* = Px^*, \tilde{\lambda} = PX^* \), and \( \nabla \tilde{f}(\vec{x}^*) = P \nabla f(\vec{x}^*) \). Substituting them into (29), we obtain that

\[
\tilde{\lambda}_i = \vec{x}_i^* - \frac{1}{\sigma_i} \nabla \tilde{f}_i(\vec{x}_i^*), \quad i = 2, \ldots, N.
\]

Further define \( \tilde{\lambda}_1 = [\tilde{\lambda}_2; \ldots; \tilde{\lambda}_N] \in \mathbb{R}^{N-1} \) and norm it as

\[
\|\tilde{\lambda}_i\| = \sqrt{\sum_{i=2}^{N} \left( \frac{1}{\sigma_i} - 1/\sigma \right)^2} \leq \frac{\|\nabla f(\vec{x}^*)\|}{\sigma_2} + \|\vec{x}^*\|.
\]

Next, we compute the upper bound of \( \|\nabla f(\vec{x}^*)\| \). Assumption 1 illustrates that \( \|\nabla f_i(\vec{x}_i^*) - \nabla f_i(\vec{x}_i)\| \leq m_{f_i} \|\vec{x}^* - \vec{x}_i\| \). Recalling \( \nabla f_i(\vec{x}_i) = 0_{n_i} \), \( \|\vec{x}^* - \vec{x}_i\| \leq M_1 \) and \( \|\vec{x}_i\| \leq M_i \) yields

\[
\|\nabla f(\vec{x}^*)\| \leq \sqrt{N} \max_{i \in \mathcal{V}} \left\{ 2m_{f_i} \left( \frac{M_i}{\sqrt{N}} + M_i \right) \right\} = \sqrt{N} M_0.
\]

We combine \( \|\vec{x}^*\| \leq M_1 \) with (31) to obtain that

\[
\|\lambda^*\| = \|\tilde{\lambda}\| \leq \frac{\|\tilde{\lambda}\|}{\sigma_2} \leq \frac{\|\lambda^\ast\|}{\sigma_2} \leq \frac{\sqrt{N} M_0}{\|\lambda^\ast\|} + M_1.
\]

Due to \( L_G P^T = PA \) and \( L_G P^T \tilde{x} = 0_N \) with \( P_1 = 1_{N}^T/\sqrt{N} \), there is \( \tilde{\lambda}_1 = 1_{N}^T \lambda^*/\sqrt{N} \). Further from (9), we have \( 1_{N}^T \lambda^* = 0 \) such that \( 1_{N}^T \lambda^*(t) = 1_{N}^T \lambda^*(0) \) for any \( t \geq 0 \). Hence, \( \tilde{\lambda}_1 = 1_{N}^T \lambda^*(0)/\sqrt{N} \). If \( n \neq 1 \), the upper bound of \( \|\lambda^*\| \) is

\[
\|\lambda^*\| \leq \frac{1_{N}^T \lambda^*(0)}{\sqrt{N}} + \frac{\sqrt{N} M_0}{\|\lambda^\ast\|} + M_1 = M_2.
\]

which further yields that \( \|z^*\| \leq \|\vec{x}^*\| + \|\lambda^\ast\| \leq M_1 + M_2 \).

\[P \text{ of the ISS Lyapunov function (21)}\]

Regarding (20) provides \( V(\vec{x}) \geq (3\sigma_N/2) \|\vec{x}\|^2 \), we now establish the upper bound of \( V(\vec{x}) \). From (19), we obtain that

\[
V_2(\vec{x}) = \tilde{f}^T(\vec{x}) - \tilde{\lambda}^T \nabla \tilde{f}(\vec{x}^*) + \frac{1}{2} \tilde{\lambda}^T L_G \tilde{\lambda} + \tilde{\lambda}^T L_G \tilde{\lambda}
\]

\[
\leq \frac{m_{f_i}}{2} \|\vec{x}\|^2 + \frac{\sigma_N}{2} \|\vec{x}\|^2 + \frac{\sigma_N}{2} \left( \|\vec{x}\|^2 + \|\tilde{\lambda}\|^2 \right)
\]

\[
\leq \frac{m_{f_i} + 2\sigma_N}{2} \|\tilde{\lambda}\|^2
\]

where \( \sigma = \sigma N/(m_f + \sigma N) \) and Assumption 1 are used. Following (18) and (20) yields

\[
\frac{3\sigma_N}{2} \|\tilde{\lambda}\|^2 \leq V(\vec{x}) \leq \frac{m_{f_i} + 6\sigma_N}{2} \|\tilde{\lambda}\|^2.
\]

Based on (33), we finish the proof of (21) by calculating the first-order derivatives of \( V_1 \) and \( V_2 \) along with (17)

\[
\dot{V}_1(\vec{x}) = \alpha \left[ -\tilde{f}^T(\vec{x}) - L_G \tilde{\lambda} + L_G (e_x + e_s) - \tilde{\lambda}^T L_G e_s \right]
\]

\[
\leq \alpha \left[ -\sigma_N \|L_G \tilde{\lambda}\|_{\tilde{\lambda}} + \tilde{\lambda}^T L_G (e_x + e_s) - \tilde{\lambda}^T L_G e_s \right]
\]

where \( \tilde{\lambda}^T \nabla \tilde{f}(\vec{x}) \geq 0 \) is used. According to the Young inequality, \( ab \leq \lambda^\ast + 2\lambda^T \vec{x} \) for any \( a, b \in \mathbb{R}, c > 0 \). Thus, for any \( \epsilon_i \in (0, 1) \)

\[
\tilde{\lambda}^T L_G (e_x + e_s) \leq \|L_G \tilde{\lambda}\| \|e\| \leq \frac{\epsilon_i}{2\sigma_N} \|L_G \tilde{\lambda}\| \|e\|^2 + \frac{\sigma_N}{2\epsilon_1} \|e\|^2
\]
which further leads to
\[ V_1(\hat{\tilde{z}}) \leq \alpha \left[ -\frac{(1-\varepsilon_1)||L_0\tilde{\omega}||^2}{\sigma_N} + \frac{\sigma_N}{2\varepsilon_1} \| e \|^2 - \tilde{\omega}^T L_0 e_\omega \right]. \] (34)

By \( \nabla f(x^z) = -L_0\alpha^*, \) the first order derivative of \( V_2 \) satisfies
\[ V_2(\hat{z}) = \alpha \left[ -||\nabla f(\tilde{x}) + L_0\tilde{x} + L_0\lambda||^2 - (L_0\tilde{x})^T L_0 e_x \right. \\
+ \left. ||L_0\tilde{x}||^2 + (\nabla f(\tilde{x}) + L_0\tilde{x} + L_0\lambda)^T L_0 e_x \right]. \]

Using again Young inequality, for any \( \varepsilon_2 \in (0, 1), \) there is
\[ V_2(\hat{z}) \leq \alpha \left[ -||\nabla f(\tilde{x}) + L_0\tilde{x} + L_0\lambda||^2 + (4\varepsilon_2)^{-1} \right] \\
\times \left[ ||L_0\tilde{x}||^2 + ||L_0\tilde{x}||^2 - (L_0\tilde{x})^T L_0 e_x \right]. \] (35)

Summing up (34) and (35), we obtain the upper bound of \( V(\hat{z}) \) for selecting \( \varepsilon_1 = \varepsilon_2 = 1/2 \) as
\[ \hat{V}(\hat{z}) \leq \alpha \left[ -2||L_0\tilde{x}||^2 + 4\sigma_N^2 ||e||^2 - 4\sigma_N \hat{\lambda}^T L_0 e_x \right. \\
- \left. \frac{1}{2}||\nabla f(\tilde{x}) + L_0\tilde{x} + L_0\lambda||^2 + \frac{\sigma_N}{2} \| e \|^2 \right] \\
+ \left[ 4\sigma_N^2 \right] \frac{1}{2} ||L_0 e_x||^2. \] (36)

By \( ||\hat{\lambda}|| \leq ||\hat{x}|| \leq \frac{24(||x||}{\sigma_N} \) in (33) and \( ||L_0 e_x|| \leq \frac{\sigma_N}{2} ||e|| \)
\[ \hat{V}(\hat{z}) \leq \alpha \left[ \frac{1}{2} ||F(\hat{z})||^2 + \left( 2\sigma_N^2 + 1 \right) ||e(t)||^2 + 4\sigma_N ||\hat{\lambda}||^2 \right] \\
\leq \alpha \left[ \frac{\beta}{2} ||F(\hat{z})||^2 - \frac{1-\beta}{2} ||F(\hat{z})||^2 + \frac{4\sigma_N^2}{3} \right] \\
+ \left[ 2\sigma_N^2 + \frac{11}{2} \sigma_N^2 \right] ||e(t)||^2 \] (37)
for any \( \beta \in (0, 1). \) Then, it follows from \( -\kappa ||F(\hat{z})|| \leq ||\hat{\lambda}|| \) and 
\( -||\hat{x}||^2 \leq (2/(m_f + 6\sigma_N)) V(\hat{z}) \) in (33) that (21) holds.

C. Proof of Three Inequalities (22)

We employ the mathematical induction method to prove three inequalities in (22). When \( k = 0, \) Assumption 3 naturally ensures (22). The remaining is to prove that (22) holds for \( k = k_1 + 1 \) provided that it holds for \( k = k_1, \forall k_1 \in N. \)

Step 1: We prove that for any \( t' \in (k_1T, k_1 + 1)T \)
\[ \text{when } ||e(t)|| < b(t), \ t \in (k_1T, t') \Rightarrow V(\tilde{z}(t)) \leq a(t), \ t \in (k_1T, t'). \] (38)

For any \( t \in (k_1T, t'), \) we define an auxiliary function
\[ V_1(t) = V(\tilde{z}(t))e^{\alpha t}. \]
Assume \( V_1(t) \geq \frac{\beta}{\sigma_N} V(z(0)), \) which further derives that \( V(\tilde{z}(t)) \geq a(t) \) by (20). Since \( ||e(t)|| < b(t) \)
\[ \left( 2\sigma_N^2 + \frac{11}{2} \sigma_N^2 \right) ||e(t)||^2 < \left( \frac{1-\beta}{\beta} \eta - \frac{4}{3} \right) a(t). \] (39)

According to (21), if \( V(\tilde{z}(t)) \geq a(t), \) then \( \dot{V} < -\alpha \eta V. \) It is equivalent to state that
\[ \text{If } V_1(t) \geq a(0) \Rightarrow V_1(t) = e^{\alpha t}(V(\tilde{z}) + \alpha \eta V(\tilde{z})) < 0. \] (40)

Hence, \( \{z(t)V(\tilde{z}(t))e^{\alpha t} \leq \frac{\beta}{\sigma_N} V(z(0)) \} \) is a positively invariant set. Combining with \( V(\tilde{z}(k_1T)) \leq a(k_1T) \) such that \( V_1(t) \leq \frac{\beta}{\sigma_N} V(z(0)) \), we conclude that (38) holds.

Step 2: We prove that for any \( t' \in (k_1T, (k_1 + 1)T) \)
\[ \text{when } V(\tilde{z}(t)) \leq a(t) \Rightarrow ||e(t)|| < b(t), \ t \in (k_1T, t'). \] (41)

Since \( \dot{\hat{z}} = \hat{z} + \hat{q} \) and \( \hat{q} = 0_{2N_k} \) from (17), we obtain that
\[ \dot{e} - \alpha L_0 e = \left[ -L_0\tilde{x} - L_0\lambda - \nabla f(\tilde{x}) \right] \] (42)
with \( \hat{L}_0 = \begin{bmatrix} L_0 & L_0 \\ L_0 & 0_{2N_k} \end{bmatrix}. \) Integrating both sides of (42) from \( k_1T \) to \( t \) for any \( t \in (k_1T, t') \) yields that
\[ e(t) = e(k_1T)e^{-\alpha \int_{k_1T}^{t} L_0 e^T d\tau} - \alpha \int_{k_1T}^{t} \int_{k_1T}^{\tau} \left[ -L_0\tilde{\omega} - L_0\lambda - \nabla f(\tilde{x}) \right] e^{- \alpha \int_{k_1T}^{\tau} L_0 e^T d\tau} d\tau. \] (43)

Taking the norm on both sides of (43) gives that
\[ ||e(t)|| \leq ||e(k_1T)|| ||e^\alpha L_0(t-k_1T)|| + \alpha \int_{k_1T}^{t} ||e^\alpha L_0(t-\tau)|| \] (44)
\[ \left( 2||L_0\tilde{x}|| + ||L_0\tilde{\omega}|| + ||\nabla f(\tilde{x})|| \right) d\tau. \]

Next, we analyze the upper bound of each term of (44)
\[ ||e(k_1T)|| \leq e^{\alpha \sigma_1 T} ||e(k_1T)|| \leq \alpha \sum_{p=0}^{\infty} \frac{\left[ L_0(t-k_1T) \right]^p}{p!} ||e(k_1T)|| \]
\[ \leq \frac{\left[ ||L_0(t-k_1T)||^p \right]}{p!} ||e(k_1T)|| \leq e^{\alpha \sigma_1 T} ||e(k_1T)|| \]
where \( \sigma_1 \) is the largest eigenvalue of \( \hat{L}_0. \) Following from \( V(\tilde{z}(t)) \leq a(t), \ t \in (k_1T, t') \) and \( ||\tilde{x}|| \leq ||\tilde{\lambda}|| \) in (33) that (21) holds.
Based on Steps 1–3, we finish the proof of

\[ \|e(t)\| \leq b(t), \quad t \in [k(T), (k + 1)T) \]

Since \( \|e(t)\| < b(t), \quad t \in [k(T), (k + 1)T) \), it follows (38) and (47) that (22) holds for \( k = k_1 + 1 \). By mathematical induction methods, (22) holds for any \( k \in \mathbb{N} \).

References

[1] K. Yuan, B. Ying, X. Zhao, and A. H. Sayed, "Exact diffusion for distributed optimization and learning—Part I: Algorithm development," IEEE Trans. Signal Process., vol. 67, no. 3, pp. 708–723, Feb. 2019.

[2] S. Bolognani and S. Zampieri, "A distributed control strategy for reactive power compensation in smart microgrids," IEEE Trans. Autom. Control, vol. 58, no. 11, pp. 2818–2833, Nov. 2013.

[3] S. A. Alghunaim and A. H. Sayed, "Linear convergence of primal–dual gradient methods and their performance in distributed optimization," Automatica, vol. 117, 2020, Art. no. 109003.

[4] A. Nedíc and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," IEEE Trans. Autom. Control, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[5] P. Lin, W. Ren, and J. A. Farrell, "Distributed continuous-time optimization: Nonuniform gradient gains, finite-time convergence, and convex constraint set," IEEE Trans. Autom. Control, vol. 62, no. 5, pp. 2239–2253, May 2017.

[6] Y. Zhu, W. Yu, G. Wen, G. Chen, and W. Ren, "Continuous-time distributed subgradient algorithm for convex optimization with general constraints," IEEE Trans. Autom. Control, vol. 64, no. 4, pp. 1694–1701, Apr. 2019.

[7] T. Li, M. Fu, L. Xie, and J. Zhang, "Distributed consensus with limited communication data rate," IEEE Trans. Autom. Control, vol. 56, no. 2, pp. 279–292, Feb. 2011.

[8] R. Carli, F. Bullo, and S. Zampieri, "Quantized average consensus via dynamic coding/decoding schemes," Int. J. Robust Nonlinear Control, vol. 20, no. 2, pp. 156–175, 2010.

[9] G. Baldan and S. Zampieri, "An efficient quantization algorithm for solving average-consensus problems," in Proc. IEEE Eur. Control Conf., 2009, pp. 761–766.

[10] S. Magnussson, H. Shokri-Ghadikolaei, and N. Li, "On maintaining linear convergence of distributed learning and optimization under limited communication," IEEE Trans. Signal Process., vol. 68, pp. 6101–6116, Oct. 2020.

[11] Y. Pu, M. N. Zeilinger, and C. N. Jones, "Quantization design for distributed optimization," IEEE Trans. Autom. Control, vol. 62, no. 5, pp. 2107–2120, May 2017.

[12] Y. Kajiyama, N. Hayashi, and S. Takai, "Linearization of consensus-based quantized optimization for smooth and strongly convex cost functions," IEEE Trans. Autom. Control, vol. 66, no. 3, pp. 1254–1261, Mar. 2021.

[13] X. Liu, Y. Li, R. Wang, J. Tang, and M. Yan, "Linear convergence decentralized optimization with compression," in Proc. Int. Conf. Learn. Represent., 2021.

[14] D. Kovalev, A. Koloskova, M. Jaggi, P. Richtarik, and S. Stich, "A linearly convergent algorithm for decentralized optimization: Sending less bits for free!," in Proc. Int. Conf. Artif. Intell. Statist., 2021, pp. 4087–4095.

[15] Y. Liu, Z. Li, K. Huang, and S. Pu, "A compressed gradient tracking method for decentralized optimization with linear convergence," IEEE Trans. Autom. Control, vol. 67, no. 10, pp. 5622–5629, Oct. 2022.

[16] N. Michelusi, G. Scutari, and C.-S. Lee, "Finite-bit quantization for distributed algorithms with linear convergence," IEEE Trans. Inf. Theory, vol. 68, no. 11, pp. 7524–7528, Nov. 2022.

[17] T. T. Doan, S. T. Magni, and J. Romberg, "Fast convergence rates of distributed subgradient methods with adaptive quantization," IEEE Trans. Autom. Control, vol. 66, no. 5, pp. 2191–2205, May 2021.

[18] X. Zhang, J. Liu, Z. Zhu, and E. S. Bentley, "Compressed distributed gradient descent: Communication-efficient consensus over networks," in Proc. IEEE Conf. Comput. Commun., 2019, pp. 2431–2439.

[19] S. Liang, L. Y. Wang, and G. Yin, "Exponential convergence of distributed primal–dual convex optimization algorithm without strong convexity," Automatica, vol. 105, pp. 298–306, 2019.

[20] Z. Chen and H. Ji, "Distributed quantized optimization design of continuous-time multiagent systems over switching graphs," IEEE Trans. Syst. Man, Cybern., Syst., vol. 51, no. 11, pp. 7152–7163, Nov. 2020.

[21] R. I. Boj and D.-K. Nguyen, "Improved convergence rates and trajectory convergence for primal–dual dynamical systems with vanishing damping," J. Differ. Equ., vol. 303, pp. 369–406, 2021.