2022 IEEE 61st Conference on Decision and Control (CDC)
December 6-9, 2022. Cancún, Mexico

Quantized Distributed Nonconvex Optimization with Linear Convergence

Lei Xu, Xinlei Yi, Jiayue Sun, Yang Shi, Karl Henrik Johansson and Tao Yang

***Abstract*— This paper considers distributed nonconvex optimization for minimizing the average of local cost functions, by using local information exchange over undirected communication networks. Since the communication channels often have limited bandwidth or capacity, we first introduce a quantization rule and an encoder/decoder scheme to reduce the transmission bits. By integrating them with a distributed algorithm, we then propose a distributed quantized nonconvex optimization algorithm. Assuming the global cost function satisfies the Polyak–Łojasiewicz condition, which does not require the global cost function to be convex and the global minimizer is not necessarily unique, we show that the proposed algorithm linearly converges to a global optimal point. Moreover, a low data rate is shown to be sufficient to ensure linear convergence when the algorithm parameters are properly chosen. The theoretical results are illustrated by numerical simulation examples.**

I. Introduction

Distributed optimization, which can be traced back to [1], [2], has received a growing and renewed interest over the last decade due to its wide applications in resource allocation, machine learning, and sensor networks, just to name a few. Various distributed optimization algorithms have been developed, see, e.g., [3], [4]. The basic convergence results of distributed optimization algorithms usually guarantee sublinear convergence to the optimal point for the case where the local cost functions are convex, see, e.g., [5]–[8]. When the local cost functions are strongly convex, linear convergence results are established [9]–[13].

Distributed optimization algorithms require the agents to communicate with each other through communication networks. Since the communication channels often have limited bandwidth or capacity, distributed optimization algorithms with quantized communications have been developed. For the convex case, the authors of [14], [15] proposed a quantized distributed incremental and subgradient algorithm, respectively. These algorithms sublinearly converge to a neighborhood around the optimal point. The authors of [16] developed a quantized distributed accelerated gradient algorithm and established the linear convergence to a neighborhood around the optimal point.

Recently, focusing on the strongly convex case, few studies proposed quantized distributed algorithms which converge to the exact optimal point. For example, the authors of [17] designed a quantized distributed algorithm by integrating the distributed subgradient algorithm and the uniform quantization, while the authors of [18], [19] developed a quantized distributed gradient algorithm by using the random quantizer and the sign of relative state, respectively. However, these algorithms only have sublinear convergence rates. The authors of [20], [21] proposed quantized distributed algorithms by equipping the distributed gradient tracking algorithm with uniform quantizers, and established linear convergence to the exact global optimal point for undirected and directed graphs, respectively.

Note that the aforementioned distributed algorithms which linearly converge to the exact optimal point only focused on the case where local cost functions are strongly convex. However, in many applications, such as optimal traffic flow problems, operating wind farm problems, and resource allocation problems, the cost functions are usually nonconvex, see, e.g., [22]–[24]. This motivates us to consider the nonconvex case. The main contributions of this work are summarized as follows. First, we introduce a quantization rule and an encoder/decoder scheme to reduce the transmission bits. Second, by integrating them with a distributed algorithm, we propose a quantized distributed algorithm for solving nonconvex optimization over an undirected connected network. Third, assuming that the global cost function satisfies the Polyak–Łojasiewicz condition, which does not require the cost function to be convex and the global minimizer is not necessarily unique, we show that the proposed algorithm linearly converges to a global optimal point for the case where the quantization level is larger than a certain threshold. Last but not least, if the communication channels allow only a low data rate, we show that the proposed algorithm also linearly converges to a global optimal point provided that the algorithm parameters are properly chosen. TABLE I summarizes the comparison between this paper and the related studies.

The remainder of the paper is organized as follows. Section II presents the problem formulation and motivation. Section III introduces a quantization rule and an encoder/decoder scheme. Section IV proposes a quantized distributed algorithm with finite data rates. Section V presents numerical simulation examples. Finally, concluding remarks are offered in Section VI. Due to the space limitation, all the proofs are omitted here, but can be found in [25].

This work was supported by the National Natural Science Foundation of China 62133003, 61991403 & 61991400, and the Knut and Alice Wallenberg Foundation, the Swedish Foundation for Strategic Research, the Swedish Research Council.

L. Xu, J. Sun and T. Yang are with the State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, 110819, Shenyang, China (e-mail: 2010345@stu.neu.edu.cn; sunjiayue@ise.neu.edu.cn; yangtao@mail.neu.edu.cn).

X. Yi and K. H. Johansson are with Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, and they are also affiliated with Digital Futures, 10044, Stockholm, Sweden (e-mail: {xinleiy, kallej}@kth.se).

Y. Shi is with the Department of Mechanical Engineering, University of Victoria, Victoria, BC V8W 2Y2, Canada (e-mail: yshi@uvic.ca).

2022 IEEE 61st Conference on Decision and Control (CDC) | 978-1-6654-6761-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/CDC51059.2022.9992989

TABLE I

COMPARISON OF DIFFERENT QUANTIZED DISTRIBUTED ALGORITHMS

References	Exact solution	Linear convergence	Nonconvex
[14], [15], [26]	✗	✗	✗
[17]–[19]	✔	✗	✗
[16], [27]	✗	✔	✗
[20], [21], [28]	✔	✔	✗
this paper	✔	✔	✔

Notation: Let $\mathbf{1}_n$ (or $\mathbf{0}_n$) be the $n\times 1$ vector with all ones (or zeros), and $\mathbf{I}_n$ be the n-dimensional identity matrix. $\|\cdot\|$ is the Euclidean vector norm or spectral matrix norm. For a column vector $x=(x_1,\cdots,x_m)$, $\|x\|_\infty := \max_{1\le i\le m}|x_i|$. For a positive semidefinite matrix $\mathcal{C}$, $\mathcal{Q}_{\mathcal{C}}(x) := x^T\mathcal{C}x$, $\rho(\mathcal{C})$ and $\underline{\rho}(\mathcal{C})$ are the spectral radius and the minimum positive eigenvalue of matrix $\mathcal{C}$, respectively. The minimum integer greater than or equal to c is denoted by $\lceil c\rceil$. Let $\mathrm{diag}[a_1,\cdots,a_n]$ denote a diagonal matrix with the i-th diagonal element being a_i. Given any differentiable function f, ∇f is the gradient of f. $A\otimes B$ represents the Kronecker product of matrices A and B.

II. PROBLEM FORMULATION AND MOTIVATION

Consider a group of n agents distributed over an undirected graph $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{A})$, where $\mathcal{V}=\{1,2,\dots,n\}$ and $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of edges. $(i,j)\in\mathcal{E}$ indicates that the agents i and j can communicate with each other, and $\mathcal{A}=[a_{ij}]\in\mathbb{R}^{n\times n}$ is the adjacency matrix, where $a_{ij}>0$ if $(j,i)\in\mathcal{E}$, otherwise $a_{ij}=0$. Let $\mathcal{N}_i=\{j\in\mathcal{V}: a_{ij}>0\}$ and $d_i=\sum_{j=1}^n a_{ij}$ denote the neighbor set and weighted degree of agent i, respectively. The degree matrix is defined as $\mathcal{D}=\mathrm{diag}[d_1,\dots,d_n]$. The graph Laplacian matrix is $L:=[L_{ij}]=\mathcal{D}-\mathcal{A}$. A path from agent i_1 to agent i_k is a sequence of agents $\{i_1,\cdots,i_k\}$ such that $(i_j,i_{j+1})\in\mathcal{E}$ for $j=1,\cdots,k-1$. An undirected graph is connected if there exists a path between any pair of distinct agents.

Assume that each agent has a private local cost function $f_i:\mathbb{R}^m\to\mathbb{R}$. The objective is to find an optimizer $x^\star$ to minimize the following optimization problem

$$\min_{x\in\mathbb{R}^m} f(x) := \frac{1}{n}\sum_{i=1}^n f_i(x). \tag{1}$$

Throughout this paper, we make the following assumptions.

Assumption 1: The undirected graph $\mathcal{G}$ is connected.

Assumption 2: Each local cost function $f_i(x)$ is smooth with constant $L_f>0$, i.e.,

$$\|\nabla f_i(x)-\nabla f_i(y)\|\le L_f\|x-y\|,\ \forall x,y\in\mathbb{R}^m. \tag{2}$$

Assumption 3: The optimal set $\mathbb{X}^\star=\mathrm{argmin}_{x\in\mathbb{R}^m} f(x)$ is nonempty and $f^\star=\min_{x\in\mathbb{R}^m} f(x)>-\infty$.

Assumption 4: The global cost function $f(x)$ satisfies the Polyak–Łojasiewicz condition with constant $\nu>0$, i.e.,

$$\frac{1}{2}\|\nabla f(x)\|^2\ge\nu(f(x)-f^\star),\ \forall x\in\mathbb{R}^m. \tag{3}$$

Remark 1: Assumptions 1–3 are common in the literature, see, e.g., [3], [4]. Note that the convexity of the local cost functions is not assumed. Assumption 4 does not imply the convexity of the global cost function. However, it implies its invexity [29], i.e., all stationary points are global optimal points.

The following result is used in this paper.

Lemma 1: (see [30, Lemma 2]) Let L be the Laplacian matrix of an undirected and connected graph $\mathcal{G}$ with n agents and $K_n=\mathbf{I}_n-\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$. Then L and K_n are positive semi-definite, $L\le\rho(L)\mathbf{I}_n$, $\rho(K_n)=1$,

$$K_nL=LK_n=L, \tag{4a}$$

$$0<\underline{\rho}(L)K_n\le L\le\rho(L)K_n. \tag{4b}$$

Moreover, there exists an orthogonal matrix $[\ r\quad R\ ]\in\mathbb{R}^{n\times n}$ with $r=\frac{1}{\sqrt{n}}\mathbf{1}_n$ and $R\in\mathbb{R}^{n\times(n-1)}$ such that

$$PL=LP=K_n, \tag{5a}$$

$$\frac{1}{\rho(L)}I_n\le P\le\frac{1}{\underline{\rho}(L)}I_n, \tag{5b}$$

where $\Lambda_1=\mathrm{diag}([\lambda_2,\cdots,\lambda_\mathrm{n}])$ with $0<\lambda_2\le\cdots\le\lambda_n$ beging the nonzero eigenvalues of the Laplacian matrix L, and

$$P=\begin{bmatrix} r & R\end{bmatrix}\begin{bmatrix}\lambda_n^{-1} & 0\\ 0 & \Lambda_1^{-1}\end{bmatrix}\begin{bmatrix} r^T\\ R^T\end{bmatrix}.$$

Distributed optimization algorithms require the agents to communicate with each other through a communication network. Since the communication channels usually have a limited capacity, quantized distributed algorithms have been proposed to save the bandwidth and reduce the communication cost, see, e.g., [14]–[21], [26], [27]. These algorithms require the local cost functions to be convex. However, in many applications, the cost functions are usually nonconvex. This motivates us to develop a quantized distributed algorithm for the nonconvex case.

III. QUANTIZATION RULE AND ENCODER/DECODER SCHEME

In this section, we introduce a quantization rule and an encode/decode scheme.

To begin with, let us consider a uniform quantizer $q[a]$ with $2\mathcal{K}+1$ quantization levels, i.e.,

$$q[a]=\begin{cases} j, & \dfrac{2j-1}{2}<a\le\dfrac{2j+1}{2},\ j=0,\cdots,\mathcal{K},\\ \mathcal{K}, & \dfrac{2\mathcal{K}+1}{2}>a,\\ -q[a], & a\le-\dfrac{1}{2}.\end{cases} \tag{6}$$

For this $2\mathcal{K}+1$-level quantizer, the communication channel is required to be capable of transmitting $\lceil\log_2(2\mathcal{K})\rceil$ bits. Next for a vector $h=[h_1,h_2,\dots,h_m]\in\mathbb{R}^m$, we define $Q[h]=(q[h_1],\cdots,q[h_m])$. The quantizer $Q[h]$ is not saturated if

$\|h\|_\infty \leq \mathcal{K} + \frac{1}{2}$. In this case, the quantization error is bounded, i.e.,

$$\|h - Q[h]\|_\infty \leq \frac{1}{2}. \tag{7}$$

Next, we introduce an encoder/decoder pair for each agent to quantize its state and to estimate its neighbors' states.

Encoder

Agent $j \in \mathcal{V}$ recursively generates m-dimensional quantized output $z_j(k)$ and internal state $b_j(k)$ from the exact state $x_j(k)$ as follows for any $k \geq 1$:

$$z_j(k) = Q\left[\frac{1}{s(k-1)}(x_j(k) - b_j(k-1))\right], \tag{8a}$$

$$b_j(k) = s(k-1)z_j(k) + b_j(k-1), \tag{8b}$$

where the initial value $b_j(0) = 0$, $s(k) = s(0)\mu^k > 0$, which is a decreasing sequence used to adaptively adjust the encoder, and $\mu \in (0, 1)$ is a positive constant.

The agent $j \in \mathcal{V}$ at time k sends its quantized output $z_j(k)$ to its neighboring agent $i \in \mathcal{N}_j$. The following decoder scheme is used to recover agent j's state.

Decoder

When agent $i \in \mathcal{N}_j$ receives the quantized data $z_j(k)$ from agent j, a decoder recursively generates an estimate $\hat{x}_{ij}(k)$ for $x_j(k)$ by the following rule for any $k \geq 1$:

$$\hat{x}_{ij}(k) = s(k-1)z_j(k) + \hat{x}_{ij}(k-1), \tag{9}$$

where the initial value $\hat{x}_{ij}(0) = 0$.

Remark 2: Note that $b_j(k)$ is a predictor, $s(k)$ is used to adjust the prediction error $x_j(k) - b_j(k-1)$, and the initial value $s(0)$ requires to be large enough to guarantee that the quantizer is not saturated, which implies the quantization error is bounded. The positive constant $\mu \in (0, 1)$ ensures that the agent gradually improves the accuracy of the estimate for its neighbors' state. Moreover, since the initial value $b_j(0) = \hat{x}_{ij}(0) = 0$, we obtain that $b_j(k) = \hat{x}_{ij}(k)$.

IV. A Quantized Distributed Optimization Algorithm

Based on the uniform quantizer and the encode/decode scheme, we propose Algorithm 1 with quantized communication for solving distributed nonconvex optimization.

Algorithm 1 Quantized Distributed Proportional Integral Algorithm

For each agent $i \in \mathcal{V}$.
Initialization :
$x_i(0) \in \mathbb{R}^m$, $\sum_{j=1}^n u_j(0) = \mathbf{0}_m$.
Update Rule :

$$x_i(k+1) = x_i(k) - \xi \sum_{j=1}^n L_{ij}\hat{x}_{ij}(k) - \varphi u_i(k) - \sigma \nabla f_i(x_i(k)), \tag{10a}$$

$$u_i(k+1) = u_i(k) + \varphi \sum_{j=1}^n L_{ij}\hat{x}_{ij}(k), \tag{10b}$$

where $\sigma > 0$ is the fixed step-size, ξ and φ are gain parameters, and $\hat{x}_{ij}(k)$ is produced by the decoder (9).

The algorithm is motivated by the discrete-time proportional-integral control strategy [10], [11], [31], [32]. More specifically, in Algorithm 1, the term $-\nabla f_i(x_i(k))$ ensures that each agent follows its local gradient descent, and the term $\sum_{j=1}^n L_{ij}\hat{x}_j(k)$ ensures that consensus is achieved. However, if the update rule only contains these two terms, the agents' states would not converge since the local gradients are not the same in general. Thus, to correct the error, the additional feedback term $u_i(k)$ is introduced.

Define $e_j(k) = x_j(k) - b_j(k)$. Then, the update rule (10) can be rewritten as

$$x_i(k+1) = x_i(k) - \xi \sum_{j=1}^n L_{ij}x_j(k) - \varphi u_i(k) - \sigma \nabla f_i(x_i(k)) + \xi \sum_{j=1}^n L_{ij}e_j(k), \tag{11a}$$

$$u_i(k+1) = u_i(k) + \varphi \sum_{j=1}^n L_{ij}x_j(k) - \varphi \sum_{j=1}^n L_{ij}e_j(k), \quad \forall x_i(0) \in \mathbb{R}^m, \sum_{j=1}^n u_j(0) = \mathbf{0}_m, i \in \mathcal{V}. \tag{11b}$$

Denote $\boldsymbol{x}(k) = [x_1^T(k), \dots, x_n^T(k)]^T$, $\boldsymbol{u}(k) = [u_1^T(k), \dots, u_n^T(k)]^T$, $F(\boldsymbol{x}) = \sum_{i=1}^n f_i(x_i)$, and $\boldsymbol{L} = L \otimes \mathbf{I}_m$, $\boldsymbol{e}(k) = [e_1^T(k), \dots, e_n^T(k)]^T$, $\boldsymbol{b}(k) = [b_1^T(k), \dots, b_n^T(k)]^T$, $\bar{x}(k) = \frac{1}{n}(\mathbf{1}_n^T \otimes \mathbf{I}_m)\boldsymbol{x}(k)$, $\bar{\boldsymbol{x}}(k) = \mathbf{1}_n \otimes \bar{x}(k)$. Then, the equations (11) can be rewritten in a compact form:

$$\boldsymbol{x}(k+1) = \boldsymbol{x}(k) - \xi \boldsymbol{L}\boldsymbol{x}(k) - \varphi \boldsymbol{u}(k) - \sigma \nabla F(\boldsymbol{x}(k)) + \xi \boldsymbol{L}\boldsymbol{e}(k), \tag{12a}$$

$$\boldsymbol{u}(k+1) = \boldsymbol{u}(k) + \varphi \boldsymbol{L}\boldsymbol{x}(k) - \varphi \boldsymbol{L}\boldsymbol{e}(k), \quad \forall \boldsymbol{x}(0) \in \mathbb{R}^{nm}, \ \sum_{j=1}^n u_j(0) = \mathbf{0}_m. \tag{12b}$$

Next, we investigate the property of Algorithm 1. Before stating the main convergence result, we denote

$$
\begin{aligned}
W(k) &= V(k) + n(f(\bar{x}(k)) - f^\star),\\
V(k) &= \mathcal{Q}_{\boldsymbol{K}}(\boldsymbol{x}(k)) + \mathcal{Q}_{\frac{\varphi+\xi}{\varphi}\boldsymbol{P}}(\boldsymbol{u}(k) + \frac{\sigma}{\varphi}\boldsymbol{g}(k))\\
&\quad + 2\boldsymbol{x}^T(k)\boldsymbol{K}\boldsymbol{P}(\boldsymbol{u}(k) + \frac{\sigma}{\varphi}\boldsymbol{g}(k)),
\end{aligned}
$$

where $\boldsymbol{P} = P \otimes \mathbf{I}_m$, $\boldsymbol{K} = K_n \otimes \mathbf{I}_m$ and $\boldsymbol{g}(k) = \nabla F(\boldsymbol{x}(k))$.

Proposition 1: Suppose that Assumptions 1–4 hold. Let each agent $i \in \mathcal{V}$ run Algorithm 1, and the algorithm parameters are given as follows:

$$
\begin{aligned}
&\xi \in [\frac{5}{\underline{\rho}(L)}\varphi, \kappa_1\varphi],\ \varphi \in [\sigma\kappa_2, \sigma\kappa_3],\\
&\sigma \in (0, \min\{\frac{\varepsilon}{\gamma_1}, \frac{\varepsilon}{\gamma_2}, \frac{2}{\nu}, \frac{1}{4L_f}\}),
\end{aligned}
$$

where the parameters $\gamma_1 > \eta_1$, $\gamma_2 > \eta_2$, $\varepsilon \in (0, \min\{\frac{\kappa_2}{2} - 2 - 3L_f^2\kappa_1^2 - \frac{L_f^2}{2}, \kappa_2 - 1 - \frac{3L_f^2+8}{\underline{\rho}(L)}\})$, $\kappa_1 > \frac{5}{\underline{\rho}(L)}$, $\kappa_2 > \max\{6L_f^2(\kappa_1+1)^2\kappa_1^2\rho(L), 4+6L_f^2\kappa_1^2+L_f^2, 6L_f^2(\kappa_1+1)^2, 1+\frac{3L_f^2+8}{\underline{\rho}(L)}\}$ and $\kappa_3 > \kappa_2$ with

$$
\begin{aligned}
\eta_1 &:= \kappa_3^2\rho(L) + \frac{2}{\underline{\rho}(L)} + 2\kappa_3^2\rho(L)\\
&\quad + 3\kappa_3^2L_f^2(\frac{\kappa_1+1}{\kappa_2^2} + \frac{3}{2}\rho(L)),\\
\eta_2 &:= 4\kappa_1^2\kappa_3^2\rho^2(L) + 2(\kappa_3^2(\kappa_1+1)\rho(L) + 1 + \kappa_3^2)\\
&\quad + 3\kappa_1^2L_f^2((\kappa_1+1)\rho(L) + \frac{3}{2}\kappa_3^2\rho^2(L)).
\end{aligned}
$$

Then, for any

$$
\mathcal{K} \geq \epsilon_1\sqrt{\frac{\epsilon_2 nm}{4\mu^2(\mu^2-\epsilon_3)}} + \frac{(1+2\xi d)}{2\mu} - \frac{1}{2}, \qquad (13)
$$

with $\mu \in (\sqrt{\epsilon_3}, 1)$ and

$$
\begin{aligned}
\epsilon_1 &:= \max\{\xi^2\rho^2(L), \frac{\varphi^3\rho(L)}{\varphi+\xi}, \xi\varphi\rho^2(L)\},\\
\epsilon_2 &:= \xi\rho(L) + 2\varphi\rho(L) + 4\xi^2\rho^2(L) + 2(\varphi(\xi+\varphi)\rho(L)\\
&\quad + \sigma^2 + \varphi^2 + 2\varphi),\\
\epsilon_3 &:= 1 - \frac{\epsilon_4}{\epsilon_5},\ \epsilon_4 := \min\{\epsilon_6, \epsilon_7, \frac{\sigma}{2}\nu\},\\
\epsilon_5 &:= \max\{\frac{\xi\underline{\rho}(L)+\varphi}{\xi\underline{\rho}(L)},\ 1 + \frac{2\xi}{\varphi}\},\\
\epsilon_6 &:= \varphi - \frac{8\sigma}{\underline{\rho}(L)} - \frac{6\sigma^2\varphi^2L_f^2(\xi+\varphi)^2}{\varphi^5} - \frac{3\sigma L_f^2}{\underline{\rho}(L)}\\
&\quad - (\varphi^2\rho(L) + \frac{2\sigma^2}{\underline{\rho}(L)} + 2\varphi^2\rho(L)\\
&\quad + 3\varphi^2L_f^2(\frac{\sigma^2(\xi+\varphi)}{\varphi^3} + \frac{3}{2}\rho(L))),\\
\epsilon_7 &:= \epsilon_8 - \frac{\sigma}{2}L_f^2,\\
\epsilon_8 &:= \xi\underline{\rho}(L) - \frac{9\varphi}{2} - \sigma - \frac{6\sigma^2\xi^2L_f^2(\xi+\varphi)^2}{\varphi^5}\rho(L) - \frac{3\sigma L_f^2\xi^2}{\varphi^2}\\
&\quad - (4\xi^2\rho^2(L) + 2(\varphi(\xi+\varphi)\rho(L) + \sigma^2 + \varphi^2)\\
&\quad + 3\xi^2L_f^2(\frac{\sigma^2(\xi+\varphi)}{\varphi^3}\rho(L) + \frac{3}{2}\rho^2(L)))\\
&\quad + 3\xi^2L_f^2(\frac{\sigma^2(\xi+\varphi)}{\varphi^3}\rho(L) + \frac{2\sigma^2(\xi+\varphi)^2}{\varphi^5}\rho(L) + \rho^2(L)\\
&\quad + \frac{\sigma}{\varphi^2} + \frac{1}{2}\rho^2(L)),
\end{aligned}
$$

the quantizer (8a) is never saturated provided that

$$
s(0) \geq \max\left\{\frac{C_x + \varphi C_u + \sigma C_g}{\mathcal{K} + \frac{1}{2}}, \sqrt{\frac{4\mu^2(\mu^2-\epsilon_3)W(0)}{\epsilon_2 nm}}\right\}, \qquad (14)
$$

where $C_x \geq \|\boldsymbol{x}(0)\|_\infty,\ C_u \geq \|\boldsymbol{u}(0)\|_\infty,\ C_g \geq \|\boldsymbol{g}(0)\|_\infty$.

Proposition 1 provides a sufficient condition to ensure that the quantizer causes nonsaturation. We are now ready to present the first convergence result.

Theorem 1: (high data rate). Suppose that Assumptions 1–4 hold. Let each agent $i \in \mathcal{V}$ run the Algorithm 1 with the same ξ, φ, σ, μ, $\mathcal{K}$ and $s(0)$ given in Proposition 1. Then,

$$
\|\boldsymbol{x}(k) - \bar{\boldsymbol{x}}(k)\|^2 + n(f(\bar{x}(k)) - f^\star) \leq \epsilon_9\mu^{2k},\ \forall k \geq 0, \quad (15)
$$

where $\epsilon_9 := \frac{nm\epsilon_2 s^2(0)}{4\epsilon_{10}\mu^2(\mu^2-\epsilon_3)}$, $\epsilon_{10} := \min\{\frac{\xi\underline{\rho}(L)-\varphi}{\xi\underline{\rho}(L)}, 1\}$.

Theorem 1 establishes linear convergence of the proposed algorithm provided that the quantization level is larger than a certain threshold given in (13). However, in distributed networks, the communication channels normally have a limited capacity or bandwidth, the smaller quantization level is more preferred. The following theorem establishes linear convergence result for arbitrarily low data rate, even one bit rate.

Theorem 2: (low data rate). Suppose that Assumptions 1–4 hold. Let each agent $i \in \mathcal{V}$ run the Algorithm 1 with the same ξ, φ given in Proposition 1 and $(\mu, \sigma) \in \bar{\Pi}$, where

$$
\begin{aligned}
\bar{\Pi} &:= \{(\mu, \sigma) : \sigma \in (0, \min\{\frac{\varepsilon}{\gamma_1}, \frac{\varepsilon}{\gamma_2}, \frac{2}{\nu}, \frac{1}{4L_f}\}),\\
&\qquad \mu \in (\sqrt{\epsilon_3}, 1),\ \bar{\Omega}(\mu, \sigma) \leq \mathcal{K} + \frac{1}{2}\},\\
\bar{\Omega}(\mu, \sigma) &:= \epsilon_1\sqrt{\frac{\epsilon_2 nm}{4\mu^2(\mu^2-\epsilon_3)}} + \frac{(1+2\xi d)}{2\mu}.
\end{aligned}
$$

Then, for any $\mathcal{K} \geq 1$ and $s(0)$ satisfying (14) in Proposition 1,

$$
\|\boldsymbol{x}(k) - \bar{\boldsymbol{x}}(k)\|^2 + n(f(\bar{x}(k)) - f^\star) \leq \epsilon_9\mu^{2k},\ \forall k \geq 0. \quad (16)
$$

Remark 3: For the strongly convex case, the authors of [27] proposed a quantized distributed algorithm with linear convergence. However, it does not converge to the exact global optimal point. The authors of [20], [21], [28] proposed quantized distributed algorithms which linearly converge to the global optimal point under the condition that each local cost function is strongly convex. Theorem 2 shows that our proposed algorithm linearly converges to a global optimal point provided that the global cost function satisfies the

Polyak–Łojasiewicz condition, which is a weaker condition since the local cost functions can be nonconvex.

V. Numerical examples

In this section, we demonstrate the effectiveness of the proposed quantized distributed algorithm. Consider an undirected network consisting of 100 agents and the communication graph is randomly generated as shown in Fig. 1.

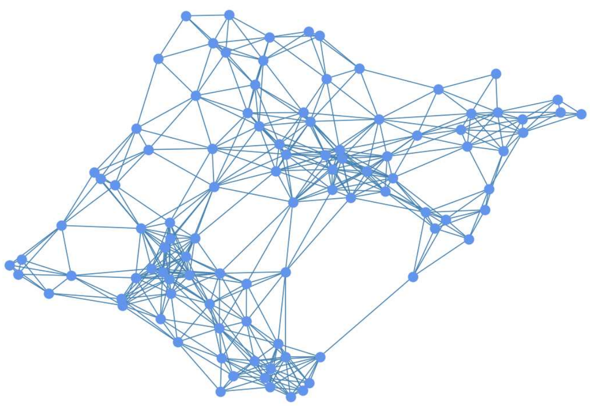

Fig. 1. A random connected network of 100 agents.

The local nonconvex cost functions are given by:

$$
\begin{aligned}
f_j(x) &= 0.2\sqrt{x^4+3} + 0.7\cos^2 x, \\
f_{10+j}(x) &= 2\sin x - 0.1(x^2+2)^{\frac{1}{3}}, \\
f_{20+j}(x) &= \frac{0.3x^2}{\sqrt{x^2+1}}, \\
f_{30+j}(x) &= -0.1\sqrt{x^4+3} - \sin x, \\
f_{40+j}(x) &= \frac{-0.2x^2}{\sqrt{x^2+1}} + 2\sin^2 x, \\
f_{50+j}(x) &= -0.1\sqrt{x^4+3} - \frac{0.1x^2}{\sqrt{x^2+1}}, \\
f_{60+j}(x) &= -\sin x - 1, \\
f_{70+j}(x) &= x^2 + 0.3\cos^2 x, \\
f_{80+j}(x) &= 2\sin^2 x + 0.2(x^2+2)^{\frac{1}{3}}, \\
f_{90+j}(x) &= -0.1(x^2+2)^{\frac{1}{3}},
\end{aligned}
$$

where $j = 1, \cdots, 10$. It is easy to check that Assumptions 1–4 are satisfied.

Consider the parameters $\mathcal{K} = 1, 10, 100$, and based on the condition (13), we set $s(0) = 10.198, 1.4569, 0.1522$, respectively. Fig. 2 illustrates the convergence of $\sum_{i=1}^n \|x_i(k) - \bar{x}(k)\|^2 + n(f(\bar{x}(k)) - f^\star)$ with respect to the number of iterations k for Algorithm 1. We set $\alpha = 0.00235$, $\beta = 0.002$, $\sigma = 0.001$, $\mu = 0.999$. Fig. 2 clearly shows that the proposed algorithm has a linear convergence rate, even the exchanged information is one bit. Moreover, the larger quantization level leads to the faster convergence. This result is reasonable since a larger quantization level implies a smaller quantization error.

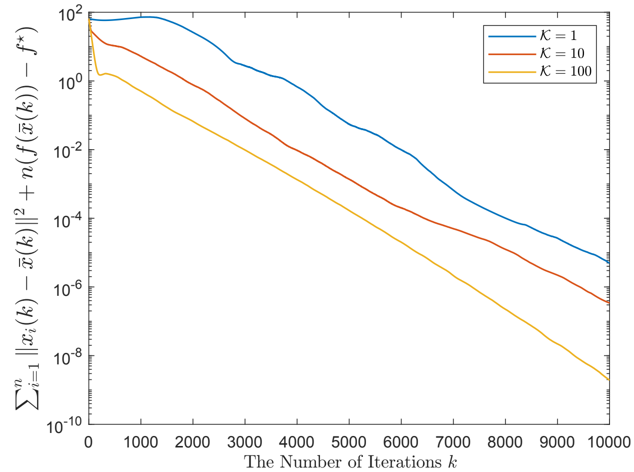

Fig. 2. Convergence of Algorithm 1 under different quantization levels.

Next, we consider the strongly convex optimization problem studied in [20], i.e.

$$
\min_{x \in \mathbb{R}^m} f(x) = \frac{1}{100}\left(\sum_{i=1}^{100} \left\|\phi_i^{\mathrm{T}} x - \psi_i\right\|^2 + \ell_2 \|x\|^2\right).
$$

Select the same parameters ϕ_i, ℓ_2 and ψ_i in [20]. We use the same quantizer (8a), and all the algorithm parameters used in the experiment are given in TABLE II.

TABLE II

Parameter Settings for Different Quantized Distributed Algorithms.

Algorithm	ξ	φ	σ	α	h	$\mathcal{K}$	μ	$s(0)$
Algorithm 1	0.235	0.2	0.1	–	–	300	0.9999	0.0026
[17]	–	–	–	–	0.1	300	0.9999	0.0026
[20]	–	–	–	0.1	0.1	300	0.9999	0.0026

Fig. 3 plots the evolution of $\sum_{i=1}^n \|x_i(k) - \bar{x}(k)\|^2 + n(f(\bar{x}(k)) - f^\star)$. It shows that our proposed algorithm has a comparable performance with distributed quantized gradient tracking algorithm in [20], and is faster than the quantized subgradient algorithm in [17]. Note that the authors of [17], [20] require the local cost functions to be strongly convex, while we only require the local cost functions to be nonconvex.

VI. Conclusions

In this paper, we introduced a quantization rule and an encoder/decoder scheme to reduce the transmitting bits. Then, by integrating them with a distributed algorithm, we proposed a quantized distributed algorithm to solve the nonconvex optimization problem. For the case where local cost functions are smooth and the global cost function satisfies the Polyak–Łojasiewicz condition, we showed that the proposed algorithm linearly converges to a global optimal point provided that the quantization level is larger than a

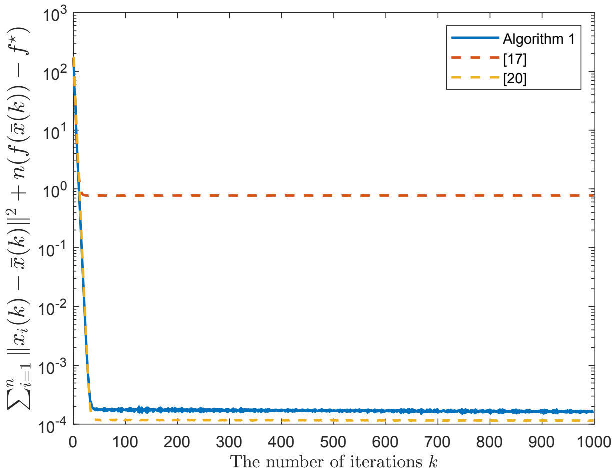

Fig. 3. Convergence under different quantized distributed algorithms.

certain threshold. Finally, we showed that, with appropriate algorithm parameters, the proposed algorithm with a low data rate, even one bit, is sufficient to ensure linear convergence. One future direction is to consider directed graphs.

REFERENCES

[1] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, vol. 31, no. 9, pp. 803–812, 1986.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, USA: Prentice Hall, 1989, vol. 23.

[3] A. Nedić and J. Liu, "Distributed optimization for control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, pp. 77–103, 2018.

[4] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[5] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *IEEE Conference on Decision and Control*, 2008, pp. 4185–4190.

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

[7] M. Zhu and S. Martínez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2011.

[8] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[9] J. Lu and C. Y. Tang, "Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2348–2354, 2012.

[10] S. S. Kia, J. Cortés, and S. Martínez, "Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication," *Automatica*, vol. 55, pp. 254–264, 2015.

[11] J. Wang and N. Elia, "Control approach to distributed optimization," in *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010, pp. 557–561.

[12] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes," in *IEEE Conference on Decision and Control*, 2015, pp. 2055–2060.

[13] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

[14] M. G. Rabbat and R. D. Nowak, "Quantized incremental algorithms for distributed optimization," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 798–808, 2005.

[15] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "Distributed subgradient methods and quantization effects," in *IEEE Conference on Decision and Control*, 2008, pp. 4177–4184.

[16] Y. Pu, M. N. Zeilinger, and C. N. Jones, "Quantization design for distributed optimization," *IEEE Transactions on Automatic Control*, vol. 62, no. 5, pp. 2107–2120, 2016.

[17] P. Yi and Y. Hong, "Quantized subgradient algorithm and data-rate analysis for distributed optimization," *IEEE Transactions on Control of Network Systems*, vol. 1, no. 4, pp. 380–392, 2014.

[18] T. T. Doan, S. T. Maguluri, and J. Romberg, "Convergence rates of distributed gradient methods under random quantization: A stochastic approximation approach," *IEEE Transactions on Automatic Control*, vol. 66, no. 10, pp. 4469–4484, 2020.

[19] J. Zhang, K. You, and T. Başar, "Distributed discrete-time optimization in multiagent networks using only sign of relative state," *IEEE Transactions on Automatic Control*, vol. 64, no. 6, pp. 2352–2367, 2018.

[20] X. Ma, P. Yi, and J. Chen, "Distributed gradient tracking methods with finite data rates," *Journal of Systems Science and Complexity*, vol. 34, no. 5, pp. 1927–1952, 2021.

[21] Y. Xiong, L. Wu, K. You, and L. Xie, "Quantized distributed gradient tracking algorithm with linear convergence in directed networks," *arXiv preprint arXiv:2104.03649*, 2021.

[22] D. Hajinezhad, M. Hong, and A. Garcia, "ZONE: Zeroth-order nonconvex multiagent optimization over networks," *IEEE Transactions on Automatic Control*, vol. 64, no. 10, pp. 3995–4010, 2019.

[23] Y. Tang, J. Zhang, and N. Li, "Distributed zero-order algorithms for nonconvex multi-agent optimization," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 1, pp. 269–281, 2021.

[24] T. Chang, M. Hong, H. Wai, X. Zhang, and S. Lu, "Distributed learning in the nonconvex world: From batch data to streaming and beyond," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 26–38, 2020.

[25] L. Xu, X. Yi, J. Sun, T. Yang, Y. Shi, and K. H. Johansson, "Quantized distributed nonconvex optimization algorithms with linear convergence," *arXiv preprint arXiv:2207.08106*, 2022.

[26] S. Liu, L. Xie, and D. E. Quevedo, "Event-triggered quantized communication-based distributed convex optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 1, pp. 167–178, 2016.

[27] Y. Kajiyama, N. Hayashi, and S. Takai, "Linear convergence of consensus-based quantized optimization for smooth and strongly convex cost functions," *IEEE Transactions on Automatic Control*, vol. 66, no. 3, pp. 1254–1261, 2021.

[28] J. Lei, P. Yi, G. Shi, and B. D. Anderson, "Distributed algorithms with finite data rates that solve linear equations," *SIAM Journal on Optimization*, vol. 30, no. 2, pp. 1191–1222, 2020.

[29] F. Dinuzzo, S. O. Cheng, P. V. Gehler, and G. Pillonetto, "Learning output kernels with block coordinate descent," in *International Conference on Machine Learning*, 2011.

[30] X. Yi, S. Zhang, T. Yang, T. Chai, and K. H. Johansson, "Communication compression for decentralized nonconvex optimization," *arXiv preprint arXiv:2201.03930*, 2022.

[31] T. Yang, Y. Wan, H. Wang, and Z. Lin, "Global optimal consensus for discrete-time multi-agent systems with bounded controls," *Automatica*, vol. 97, pp. 182–185, 2018.

[32] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2013.