SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

October 2022, Vol. 65 202202:1–202202:15
https://doi.org/10.1007/s11432-020-3275-3

Distributed quantized mirror descent for strongly convex optimization over time-varying directed graph

Menghui XIONG, Baoyong ZHANG*, Deming YUAN & Shengyuan XU

School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China

Received 2 October 2020/Revised 22 December 2020/Accepted 6 May 2021/Published online 26 September 2022

Abstract This paper investigates a distributed strongly convex constrained optimization problem in the non-Euclidean sense, where the bit rate of the considered communication between nodes is assumed to be limited, and the communication topology is represented by a time-varying directed graph. By considering the limitation of communication capacity, the quantization technique is applied in the process of exchanging information over the network. Then a distributed quantized mirror descent (DQMD) algorithm, which uses the Bregman divergence and time-varying quantizers, is developed for the strongly convex optimization under a convex constraint set. The convergence of the developed algorithm is also analyzed. It is shown that the nodes' state errors are bounded by some terms related to the quantization resolutions, and the sublinear upper-bounds can be guaranteed by choosing appropriate quantization resolutions. Finally, a distributed ridge regression is provided as an example to verify the validity of the proposed method.

Keywords distributed convex optimization, strongly convex optimization, mirror descent, quantized communication, Bregman divergence

Citation Xiong M H, Zhang B Y, Yuan D M, et al. Distributed quantized mirror descent for strongly convex optimization over time-varying directed graph. Sci China Inf Sci, 2022, 65(10): 202202, https://doi.org/10.1007/s11432-020-3275-3

1 Introduction

Recently, the distributed convex optimization over a multi-agent network has been extensively studied due to its extensive applications in many fields including sensor networks, power systems, smart grids, machine learning, and so on [1–6]. In a multi-agent network, each node has a local convex objective function, and the major purpose of distributed convex optimization is to minimize the sum of these local objective convex functions by local computation and communication. The approach of distributed convex optimization can be used to solve large-scale linear matrix equations, for example, in [7]. It should be pointed out that the creative information renovation idea in [8] was used in [7] to obtain the novel distributed algorithm with a faster convergence rate. Up to now, lots of effective algorithms have been proposed for distributed convex optimization, many of which are built on the mirror descent method (see [9–24] and the references therein). The mirror descent method is a generalization of the projected subgradient method to the non-Euclidean setting, where the Bregman divergence is used in the projection. It is noted that the Bregman divergence is a more general distance-like measure that includes many widely-used distance measures as special cases.

In most of the published literature on distributed optimization algorithms, the communication between nodes is generally supposed to be perfect. In reality, however, the capacity of communication channels may be limited, and thus the information to be transmitted generally needs to be quantized before it is sent. In this case, the following two questions arise naturally: (i) How do the errors caused by the information quantization affect distributed optimization algorithms? (ii) What conditions can make the distributed optimization algorithms based on quantized communication guarantee convergence? Regarding these questions, the distributed convex optimization under quantized communication has received increasing

* Corresponding author (email: baoyongzhang@njust.edu.cn)

attention, and a few of algorithms have been developed (see [25–31]). Specifically, the distributed dual averaging algorithm with quantized communication was considered in [25]. The distributed quantized subgradient algorithm was studied in [26–28]. The distributed consensus algorithm based on quantized communication was investigated in [29, 30]. On the basis of the inexact proximal-gradient method, two distributed optimization algorithms with a quantization refinement were developed in [31].

It is worth noting that the aforementioned algorithms are developed for solving the distributed convex optimization problems in the Euclidean sense. For the non-Euclidean case, however, the quantized communication has not been fully considered in the setting of distributed convex optimization algorithms. Therefore, it is of much interest to design and analyze the distributed quantized mirror descent (DQMD) algorithm for the distributed optimization problem subject to a convex constraint set. This is the motivation of our work.

Motivated by [15, 25], under the consideration of quantized communication over a time-varying multi-agent network, we are concerned with the problem of distributed strongly convex constrained optimization in the non-Euclidean sense. A DQMD algorithm is developed to solve the investigated problem, and the convergence of the developed algorithm is analyzed. In particular, the contributions are stated as follows.

(1) A DQMD algorithm is designed based on two kinds of time-varying quantizers (i.e., the deterministic quantizer and the probabilistic quantizer).

(2) The effects of the used time-varying quantizers on the convergence performance of the proposed DQMD algorithm are discussed. It is shown that the convergence can be guaranteed by choosing the quantization resolutions properly.

The contents are arranged below: Section 2 describes the investigated problem and the algorithm. The convergence results of the algorithm are given in Section 3. Section 4 provides a discussion for a special case. Section 5 provides an illustrative example and the paper is concluded in Section 6.

Notation. $\mathbb{R}^n$ denotes the n-dimensional Euclidean space. $\mathbb{R}^{N\times N}$ and $\mathbb{R}$ stand for the set of $N\times N$ real matrices and the set of real numbers, respectively. $[W(t)]_{ij}$ and $[u]_i$ represent the (i,j)-th entry of the matrix $W(t)$ and the i-th element of the vector u, respectively. $\nabla\varphi$ represents the gradient of a differentiable function $\varphi(\cdot)$. $\|\cdot\|_2$ stands for the Euclidean norm of vectors. $\|\cdot\|_\infty$ represents the infinite norm of vectors. $\langle a,b\rangle$ denotes the inner product of vectors. $\|u\|_* = \max_{\|v\|=1}\langle u,v\rangle$ is called the dual norm with respect to $\|\cdot\|$. $[N]$ represents the set $\{1,2,\ldots,N\}$. $\mathbb{E}\{v\}$ represents the mathematical expectation of the stochastic variable v. $\mathrm{col}_n\{v\}$ is an n-dimensional column vector, where its elements are all the constant v. $|\cdot|$ is the symbol of absolute value. For the convex function $f_i(\cdot):\mathbb{R}^n\to\mathbb{R}$, its subgradient at $u\in\mathbb{R}^n$ is denoted as $d_i(u)\in\mathbb{R}^n$, which satisfies $f_i(v)\geqslant f_i(u)+\langle d_i(u),v-u\rangle$ for all $v\in\mathbb{R}^n$.

2 Problem and algorithm

2.1 Problem formulation

In this paper, we are concerned with a distributed strongly convex constrained optimization over a time-varying multi-agent network with N nodes. A time-varying directed graph $\mathcal{F}(t)=(\vartheta,\xi(t),W(t))$ is introduced to represent the communication topology between nodes at time t, where $\vartheta=[N]$ is the node set, and $\xi(t)\in\vartheta\times\vartheta$ and $W(t)\in\mathbb{R}^{N\times N}$ are the edge set and weighted matrix of nodes at time t, respectively. Moreover, at time t, we use $\mathcal{N}_i^t=\{j:(i,j)\in\xi(t)\}$ to denote the set of the neighbors of node i, where $(i,j)\in\xi(t)$ represents that the information can be transmitted from nodes j to i.

The considered optimization problem is described as follows:

$$\min\ f(u)=\sum_{i=1}^{N}f_i(u)\quad \text{s.t. } u\in U, \tag{1}$$

where $u\in\mathbb{R}^n$ is the variable; $f_i(\cdot):\mathbb{R}^n\to\mathbb{R}$ is the objective function, which is only known by node i; $U\subseteq\mathbb{R}^n$ is a convex constraint set known by all the nodes in the time-varying network. Moreover, one assumes that an optimal solution for problem (1) exists. The optimal point and value are denoted by $u^*=\arg\min_{u\in U}f(u)$ and $f(u^*)=\min_{u\in U}f(u)$, respectively.

2.2 Assumptions and lemmas

For the time-varying directed graph $\mathcal{F}(t)=(\vartheta,\xi(t),W(t))$, the following standard assumption is adopted.

Assumption 1 ([15,32]). The time-varying directed graph $\mathcal{F}(t) = (\vartheta, \xi(t), W(t))$ satisfies the following three conditions for all $i \in \vartheta$ and $t \geqslant 1$.

(i) There exists a scalar $0 < \epsilon < 1$ such that if $j \in \mathcal{N}_i^t$, one has $[W(t)]_{ij} \geqslant \epsilon$; otherwise, $[W(t)]_{ij} = 0$;

(ii) The weighted matrix $W(t)$ is doubly stochastic, i.e., $\sum_{i=1}^N [W(t)]_{ij} = 1$ and $\sum_{j=1}^N [W(t)]_{ij} = 1$;

(iii) There exists a positive integer H such that the directed graph $(\vartheta, \bigcup_{t=pH+1}^{(p+1)H} \xi(t))$ is strongly connected for any $p \geqslant 0$.

Based on Assumption 1, a lemma describing the convergence of the transition matrix $W(t,s) = W(t)W(t-1)\cdots W(s)$ is provided, where $W(t,s)$ records the weight history from round s to t, and $t \geqslant s \geqslant 1$.

Lemma 1 ([15,33]). Under Assumption 1, the following inequality is obtained for $i, j \in \vartheta$ and $t \geqslant s \geqslant 1$,

$$|[W(t,s)]_{ij} - 1/N| \leqslant \theta_1 \theta_2^{t-s+1},$$

where $\theta_1 = (1 - \frac{\epsilon}{4N^2})^{-2}$, $\theta_2 = (1 - \frac{\epsilon}{4N^2})^{1/H}$.

Now, we introduce the definitions of strongly convexity and the Bregman divergence as follows.

Definition 1 ([34]). The function $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is δ_φ-strongly convex w.r.t. $\|\cdot\|_2$, if there exists a constant $\delta_\varphi > 0$, such that the following inequality holds for any $u_1, u_2 \in \mathbb{R}^n$,

$$\varphi(u_1) \geqslant \varphi(u_2) + \langle \nabla\varphi(u_2), u_1 - u_2 \rangle + \frac{\delta_\varphi}{2}\|u_1 - u_2\|_2^2.$$

Definition 2 ([35]). For $u_1, u_2 \in \mathbb{R}^n$, the Bregman divergence associated with the function $\varphi(\cdot) : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$B_\varphi(u_1, u_2) = \varphi(u_1) - \varphi(u_2) - \langle \nabla\varphi(u_2), u_1 - u_2 \rangle.$$

Remark 1. The Bregman divergence is a difference between the associated function $\varphi(\cdot)$ and its linear approximation. It should be noted that different Bregman divergence can be obtained by choosing different associated function $\varphi(\cdot)$. Theoretically, the Bregman divergences are a class of distance-measuring functions having more general properties. Two common examples for the possible choices of the Bregman divergence are given below. (i) The Euclidean distance $B_\varphi(u,v) = \frac{1}{2}\|u - v\|_2^2$ generated by $\varphi(u) = \frac{1}{2}\|u\|_2^2$ on $\mathbb{R}^n$; (ii) the Kullback-Leibler divergence $B_\varphi(u,v) = \sum_{i=1}^n [u]_i \ln \frac{[u]_i}{[v]_i}$ generated by $\varphi(u) = \sum_{i=1}^n [u]_i \ln [u]_i$ on $\Delta_n = \{u \in \mathbb{R}^n | [u]_i \geqslant 0, \sum_{i=1}^n [u]_i = 1\}$. Moreover, in the mirror descent algorithm, the adaptive updates can be generated by appropriately choosing the Bregman divergence, such that the geometry of the underlying constraint set can be better reflected. For instance, the Bregman divergence can be chosen as the Euclidean distance in the case of a box constraint, while under the unit simplex constraint, we should choose the Kullback-Leibler divergence as the Bregman divergence.

Remark 2. As in [35,36], two properties of the Bregman divergence commonly used in the convergence analysis of the mirror descent method are given below. For $x, y, z \in U$,

(i) $\langle \nabla\varphi(x) - \nabla\varphi(y), z - y \rangle = B_\varphi(z, y) - B_\varphi(z, x) + B_\varphi(y, x)$,

(ii) $B_\varphi(x, y) \geqslant B_\varphi(x, \Pi_{U,\varphi}(y)) + B_\varphi(\Pi_{U,\varphi}(y), y)$,

where $\Pi_{U,\varphi}(y) = \arg\min_{x \in U} B_\varphi(x, y)$.

In what follows, five assumptions are formulated, which are commonly used in the study of the mirror descent algorithm.

Assumption 2 ([19, 21]). The function $\varphi(\cdot)$ adopted in this paper is δ_φ-strongly convex w.r.t. $\|\cdot\|_2$ on the set U, and it has L_φ-Lipschitz gradient on the set U, i.e., $\|\nabla\varphi(u_1) - \nabla\varphi(u_2)\|_* \leqslant L_\varphi \|u_1 - u_2\|_2$ holds for any $u_1, u_2 \in U$.

Assumption 3 ([15]). The function f_i ($i \in [N]$) is δ_f-strongly convex w.r.t. the Bregman divergence, i.e., there exists a constant $\delta_f > 0$ such that the following condition holds for any $u_1, u_2 \in U$,

$$f_i(u_1) \geqslant f_i(u_2) + \langle d_i(u_2), u_1 - u_2 \rangle + \delta_f B_\varphi(u_1, u_2).$$

Assumption 4 ([19]). The Bregman divergence $B_\varphi(u, v)$ is convex in v for every fixed u, i.e., the following inequality holds for $\sum_{i=1}^N a_i = 1$, $a_i \geqslant 0$ and $u, v_i \in U$,

$$B_\varphi\left(u, \sum_{i=1}^N a_i v_i\right) \leqslant \sum_{i=1}^N a_i B_\varphi(u, v_i).$$

Assumption 5 ([15]). There exists a constant $D \geqslant 0$ such that $\|d_i(u)\|_2 \leqslant D$ ($i \in [N]$) holds for any $u \in U$.

Assumption 6 ([16]). There exists a constant $M \geqslant 0$ such that $\|u^* - u\|_2 \leqslant M$ holds for any $u \in U$.

2.3 Quantizers

In this paper, the effects of information quantization are also considered. Two kinds of time-varying quantizers are adopted. One is deterministic, while the other is probabilistic.

For a given vector $v = v_1\ v_2\ \cdots\ v_n{}^{\mathrm{T}} \in \mathbb{R}^n$. Then, at time t, the deterministic time-varying quantizer $Q_d^t(\cdot)$ is defined below:

$$Q_d^t(v) = \left[q_d^t(v_1)\ q_d^t(v_2)\ \cdots\ q_d^t(v_n)\right]^{\mathrm{T}} \tag{2}$$

with

$$q_d^t(v_i) = \lfloor v_i \rfloor_d^t, \tag{3}$$

where $\lfloor v_i \rfloor_d^t$ represents the operation of rounding down v_i to the nearest integer multiples of $1/\Delta_d^t$ at time t (Δ_d^t is the quantization resolution at time t). It can be seen that $q_d^t(\cdot)$ satisfies the following condition at time t:

$$v_i - q_d^t(v_i) \leqslant \frac{1}{\Delta_d^t}. \tag{4}$$

At time t, the probabilistic time-varying quantizer $Q_p^t(\cdot)$ is defined as

$$Q_p^t(v) = \left[q_p^t(v_1)\ q_p^t(v_2)\ \cdots\ q_p^t(v_n)\right]^{\mathrm{T}} \tag{5}$$

with

$$q_p^t(v_i) = \begin{cases} \lfloor v_i \rfloor_p^t & \text{with probability } (\lceil v_i \rceil_p^t - v_i)\Delta_p^t, \\ \lceil v_i \rceil_p^t & \text{with probability } (v_i - \lfloor v_i \rfloor_p^t)\Delta_p^t, \end{cases} \tag{6}$$

where $\lfloor v_i \rfloor_p^t$ and $\lceil v_i \rceil_p^t$ are the operation of rounding down and up v_i to the nearest integer multiples of $1/\Delta_p^t$ at time t (Δ_p^t is the quantization resolution at time t), respectively. Moreover, at time t, two properties of $q_p^t(\cdot)$ are given below [25]:

$$\mathbb{E}\{q_p^t(v_i) - v_i\} = 0, \quad \mathbb{E}\{(q_p^t(v_i) - v_i)^2\} \leqslant \frac{1}{4(\Delta_p^t)^2}. \tag{7}$$

2.4 Distributed quantized mirror descent algorithm

Now, we propose the distributed quantized mirror descent (DQMD) algorithm. Specifically, for $i \in [N]$, $t \geqslant 1$, node i updates its state through the following scheme:

$$\nabla\varphi(w_i^t) = \nabla\varphi(u_i^t) - \gamma_t d_i^t, \tag{8}$$

$$v_i^t = \arg\min_{u\in U} B_\varphi(u, w_i^t), \tag{9}$$

$$p_i^t = \begin{cases} Q^t(v_i^t), & \text{if } Q^t(v_i^t) \in U, \\ v_i^t, & \text{if } Q^t(v_i^t) \notin U, \end{cases} \tag{10}$$

$$u_i^{t+1} = \sum_{j=1}^N [W(t)]_{ij} p_j^t, \tag{11}$$

where u_i^t is the state of node i at time t, and the initial value is set as u_i^1; $\{\gamma_t\}_{t\geqslant 1}$ represents a non-increasing sequence of positive step sizes; $d_i^t = d_i(u_i^t)$ is a subgradient of the function f_i at u_i^t; $B_\varphi(u, w_i^t)$ is the Bregman divergence between u and w_i^t based on the function $\varphi(\cdot)$; $[W(t)]_{ij}$ is the connection weight between nodes i and j at time t, which satisfies Assumption 1; $Q^t(\cdot)$ represents the quantizer, i.e., the deterministic time-varying quantizer $Q_d^t(\cdot)$ and the probabilistic time-varying quantizer $Q_p^t(\cdot)$, respectively.

Remark 3. Different from the distributed mirror descent algorithm in [15,16], the quantization step (10) is applied in the DQMD algorithm, that means in the consensus step (11), the information to be exchanged comes from the information released by the quantization step (10). Moreover, one assumes that the nodes can store the real-valued data, and one point we need to make clear is that the quantization process is carried out in the convex constraint set U, that is, the information released by the quantization step (10) still belongs to the convex constraint set U.

Remark 4. In the process of utilizing the DQMD method (8)–(11) to solve problem (1), the main challenge lies in the proof of the basic convergence result (i.e., Lemma 3), that is the quantization error caused by the quantizer makes it difficult to deal with the corresponding Bregman divergence. For this challenge, our solution is to make an assumption (i.e., Assumption 2) on the distance-measuring function $\varphi(\cdot)$, and then use the generalized Cauchy-Schwarz inequality to expand and shrink accordingly.

3 Convergence analysis

Before presenting the main results of the DQMD algorithm, we first give two lemmas. Lemma 2 provides a bound on the differences among the states of all the nodes in the time-varying multi-agent network. Lemma 3 provides a basic convergence result for the algorithm.

Lemma 2. Let Assumptions 1, 3 and 5 hold.

(a) The following condition holds for the case with the deterministic time-varying quantizer $Q_d^t(\cdot)$:

$$\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\|u_i^t-u_j^t\|_2 \leqslant N\left(1+\frac{2\theta_1\theta_2}{1-\theta_2}\right)\sum_{j=1}^{N}\|u_j^1\|_2+\frac{2\theta_1\theta_2N^2D}{\delta_\varphi(1-\theta_2)}\sum_{t=1}^{\mathrm{T}}\gamma_t+\frac{2\theta_1\theta_2N^2\sqrt{n}}{1-\theta_2}\sum_{t=1}^{\mathrm{T}}\frac{1}{\Delta_d^t}. \tag{12}$$

(b) The following condition holds for the case with the probabilistic time-varying quantizer $Q_p^t(\cdot)$:

$$\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\mathbb{E}\{\|u_i^t-u_j^t\|_2\} \leqslant N\left(1+\frac{2\theta_1\theta_2}{1-\theta_2}\right)\sum_{j=1}^{N}\|u_j^1\|_2+\frac{2\theta_1\theta_2N^2D}{\delta_\varphi(1-\theta_2)}\sum_{t=1}^{\mathrm{T}}\gamma_t+\frac{\theta_1\theta_2N^2\sqrt{n}}{1-\theta_2}\sum_{t=1}^{\mathrm{T}}\frac{1}{\Delta_p^t}. \tag{13}$$

Proof. For convenience, we define the quantization error e_i^t, the Bregman projection error r_i^t and the average state $\bar{u}^t$ of all the nodes as

$$e_i^t = p_i^t - v_i^t, \tag{14}$$
$$r_i^t = v_i^t - u_i^t, \tag{15}$$
$$\bar{u}^t = \frac{1}{N}\sum_{i=1}^{N}u_i^t. \tag{16}$$

We first compute the upper bound of r_i^t. By utilizing the first-order optimality condition for the update formula (9) and the property $\nabla B_\varphi(u_1,u_2)=\nabla\varphi(u_1)-\nabla\varphi(u_2)$, the following inequality can be obtained:

$$\langle\nabla\varphi(v_i^t)-\nabla\varphi(w_i^t), v_i^t-u\rangle \leqslant 0, \quad \forall u\in U. \tag{17}$$

Since $u_i^t\in U$, one can gain the following inequality by setting $u=u_i^t$ in the inequality (17):

$$\langle\nabla\varphi(v_i^t)-\nabla\varphi(w_i^t), v_i^t-u_i^t\rangle \leqslant 0. \tag{18}$$

By combining the update formula (8), the inequality (18) and δ_φ-strongly convexity of the function $\varphi(\cdot)$, we can obtain that

$$\gamma_t\langle d_i^t, u_i^t-v_i^t\rangle \geqslant \langle\nabla\varphi(v_i^t)-\nabla\varphi(u_i^t), v_i^t-u_i^t\rangle \geqslant \delta_\varphi\|v_i^t-u_i^t\|_2^2, \tag{19}$$

which yields the following bound of r_i^t by employing Assumption 5 and Hölder's inequality:

$$\|r_i^t\|_2 \leqslant \frac{D}{\delta_\varphi}\gamma_t. \tag{20}$$

Next, it is obtained based on Assumption 1 and (14) and (15) that, for $t \geqslant 2$,

$$\begin{aligned}\bar{u}^t &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}[W(t-1)]_{ij}p_j^{t-1}\\ &= \frac{1}{N}\sum_{j=1}^{N}\left(\sum_{i=1}^{N}[W(t-1)]_{ij}\right)p_j^{t-1}\\ &= \frac{1}{N}\sum_{i=1}^{N}(u_i^{t-1}+r_i^{t-1}+e_i^{t-1})\\ &= \bar{u}^{t-1}+\frac{1}{N}\sum_{i=1}^{N}r_i^{t-1}+\frac{1}{N}\sum_{i=1}^{N}e_i^{t-1},\end{aligned} \tag{21}$$

which can induce an iterative formula as follows:

$$\bar{u}^t = \bar{u}^1 + \frac{1}{N}\sum_{k=1}^{t-1}\sum_{i=1}^{N}r_i^k + \frac{1}{N}\sum_{k=1}^{t-1}\sum_{i=1}^{N}e_i^k. \tag{22}$$

Similarly, calculating the recursive relation of u_i^t gives the following iteration for $t \geqslant 2$:

$$u_i^t = \sum_{j=1}^{N}[W(t-1,1)]_{ij}u_j^1 + \sum_{k=1}^{t-1}\sum_{j=1}^{N}[W(t-1,k)]_{ij}r_j^k + \sum_{k=1}^{t-1}\sum_{j=1}^{N}[W(t-1,k)]_{ij}e_j^k. \tag{23}$$

Based on (22) and (23), we get the following inequality for $t \geqslant 2$:

$$\begin{aligned}\|\bar{u}^t - u_i^t\|_2 \leqslant &\sum_{j=1}^{N}\left|[W(t-1,1)]_{ij}-\frac{1}{N}\right|\cdot\|u_j^1\|_2 + \sum_{k=1}^{t-1}\sum_{j=1}^{N}\left|[W(t-1,k)]_{ij}-\frac{1}{N}\right|\cdot\|r_j^k\|_2\\ &+\sum_{k=1}^{t-1}\sum_{j=1}^{N}\left|[W(t-1,k)]_{ij}-\frac{1}{N}\right|\cdot\|e_j^k\|_2.\end{aligned} \tag{24}$$

Applying Lemma 1 to the inequality (24), and then using the inequality (20), the following inequality can be derived under $t \geqslant 2$:

$$\|\bar{u}^t - u_i^t\|_2 \leqslant \theta_1\theta_2^{t-1}\sum_{j=1}^{N}\|u_j^1\|_2 + \frac{\theta_1 ND}{\delta_\varphi}\sum_{k=1}^{t-1}\theta_2^{t-k}\gamma_k + \theta_1\sum_{k=1}^{t-1}\theta_2^{t-k}\rho_k, \tag{25}$$

where $\rho_k = \sum_{j=1}^{N}\|e_j^k\|_2$. Summing $\|u_i^t - u_j^t\|_2$ over $t = 1, 2, \ldots, T$ yields

$$\begin{aligned}\sum_{t=1}^{\mathrm{T}}\|u_i^t - u_j^t\|_2 &= \|u_i^1 - u_j^1\|_2 + \sum_{t=2}^{\mathrm{T}}\|u_i^t - u_j^t\|_2\\ &\leqslant \sum_{j=1}^{N}\|u_j^1\|_2 + \sum_{t=2}^{\mathrm{T}}(\|\bar{u}^t - u_i^t\|_2 + \|\bar{u}^t - u_j^t\|_2).\end{aligned} \tag{26}$$

By combining the inequalities (25) and (26), one has

$$\begin{aligned}\sum_{t=1}^{\mathrm{T}}\|u_i^t - u_j^t\|_2 &\leqslant \left(1+2\theta_1\sum_{t=2}^{\mathrm{T}}\theta_2^{t-1}\right)\sum_{j=1}^{N}\|u_j^1\|_2 + \frac{2\theta_1 ND}{\delta_\varphi}\sum_{t=2}^{\mathrm{T}}\sum_{k=1}^{t-1}\theta_2^{t-k}\gamma_k + 2\theta_1\sum_{t=2}^{\mathrm{T}}\sum_{k=1}^{t-1}\theta_2^{t-k}\rho_k\\ &\leqslant \left(1+\frac{2\theta_1\theta_2}{1-\theta_2}\right)\sum_{j=1}^{N}\|u_j^1\|_2 + \frac{2\theta_1\theta_2 ND}{\delta_\varphi(1-\theta_2)}\sum_{t=1}^{\mathrm{T}}\gamma_t + \frac{2\theta_1\theta_2}{1-\theta_2}\sum_{t=1}^{\mathrm{T}}\rho_t,\end{aligned} \tag{27}$$

where the last inequality derived based on the following three inequalities:

$$\sum_{t=2}^{\mathrm{T}}\theta_2^{t-1} \leqslant \frac{\theta_2}{1-\theta_2},$$
$$\sum_{t=2}^{\mathrm{T}}\sum_{k=1}^{t-1}\theta_2^{t-k}\gamma_k \leqslant \sum_{t=1}^{T-1}\gamma_t\left(\sum_{l=1}^{T-1}\theta_2^l\right) \leqslant \frac{\theta_2}{1-\theta_2}\sum_{t=1}^{T-1}\gamma_t \leqslant \frac{\theta_2}{1-\theta_2}\sum_{t=1}^{T}\gamma_t,$$
$$\sum_{t=2}^{\mathrm{T}}\sum_{k=1}^{t-1}\theta_2^{t-k}\rho_k \leqslant \sum_{t=1}^{T}\rho_t\left(\sum_{l=1}^{T-1}\theta_2^l\right) \leqslant \frac{\theta_2}{1-\theta_2}\sum_{t=1}^{T-1}\rho_t \leqslant \frac{\theta_2}{1-\theta_2}\sum_{t=1}^{T}\rho_t.$$

The subsequent proofs of parts (a) and (b) are different, which will be discussed separately.

(a) In the case with quantizer $Q_d^t(\cdot)$, it follows from the property (4) that

$$\rho_t = \sum_{j=1}^{N}\|e_j^t\|_2 = \sum_{j=1}^{N}\sqrt{\sum_{i=1}^{n}[e_j^t]_i^2} \leqslant \frac{N\sqrt{n}}{\Delta_d^t}. \tag{28}$$

Then, substituting the inequality (28) into the inequality (27) yields the inequality (12).

(b) In the case with quantizer $Q_p^t(\cdot)$, it follows from the property (7) that

$$\mathbb{E}\{\rho_t\} = \sum_{j=1}^{N}\mathbb{E}\{\|e_j^t\|_2\} \leqslant \sum_{j=1}^{N}\sqrt{\sum_{i=1}^{n}\mathbb{E}\{[e_j^t]_i^2\}} \leqslant \frac{N\sqrt{n}}{2\Delta_p^t}. \tag{29}$$

Then the inequality (13) can be derived by taking expectation of the inequality (27), and using the inequality (29).

Lemma 3. Let Assumptions 1–6 hold.

(a) The following condition holds for the case with the deterministic time-varying quantizer $Q_d^t(\cdot)$:

$$\begin{aligned}\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\langle d_i^t, u_i^t-u^*\rangle \leqslant &\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\frac{B_\varphi(u^*,u_i^t)-B_\varphi(u^*,u_i^{t+1})}{\gamma_t} + \frac{ND^2}{2\delta_\varphi}\sum_{t=1}^{\mathrm{T}}\gamma_t\\ &+\left(L_\varphi-\frac{\delta_\varphi}{2}\right)Nn\sum_{t=1}^{\mathrm{T}}\frac{1}{(\Delta_d^t)^2}\frac{1}{\gamma_t}+ML_\varphi N\sqrt{n}\sum_{t=1}^{\mathrm{T}}\frac{1}{\Delta_d^t}\frac{1}{\gamma_t}.\end{aligned} \tag{30}$$

(b) The following condition holds for the case with the probabilistic time-varying quantizer $Q_p^t(\cdot)$:

$$\begin{aligned}\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\mathbb{E}\{\langle d_i^t, u_i^t-u^*\rangle\} \leqslant &\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\frac{\mathbb{E}\{B_\varphi(u^*,u_i^t)\}-\mathbb{E}\{B_\varphi(u^*,u_i^{t+1})\}}{\gamma_t}\\ &+\frac{ND^2}{2\delta_\varphi}\sum_{t=1}^{\mathrm{T}}\gamma_t+\frac{ML_\varphi N\sqrt{n}}{2}\sum_{t=1}^{\mathrm{T}}\frac{1}{\Delta_p^t}\frac{1}{\gamma_t}.\end{aligned} \tag{31}$$

Proof. Based on the updating formula (8) and recalling Remark 2, we have

$$\begin{aligned}\langle d_i^t, u_i^t-u^*\rangle &= \frac{1}{\gamma_t}\langle\nabla\varphi(w_i^t)-\nabla\varphi(u_i^t), u^*-u_i^t\rangle\\ &=\frac{1}{\gamma_t}(B_\varphi(u^*,u_i^t)-B_\varphi(u^*,w_i^t)+B_\varphi(u_i^t,w_i^t))\\ &\leqslant\frac{1}{\gamma_t}(B_\varphi(u^*,u_i^t)-B_\varphi(u^*,v_i^t)+B_\varphi(u_i^t,w_i^t)-B_\varphi(v_i^t,w_i^t)).\end{aligned} \tag{32}$$

We scale $B_\varphi(u_i^t,w_i^t)-B_\varphi(v_i^t,w_i^t)$ as

$$B_\varphi(u_i^t,w_i^t)-B_\varphi(v_i^t,w_i^t)=\varphi(u_i^t)-\varphi(v_i^t)-\langle\nabla\varphi(w_i^t),u_i^t-v_i^t\rangle$$

$$
\begin{aligned}
&\leqslant \langle \nabla\varphi(u_i^t) - \nabla\varphi(w_i^t), u_i^t - v_i^t\rangle - \frac{\delta_\varphi}{2}\|u_i^t - v_i^t\|_2^2 \\
&= \langle \gamma_t d_i^t, u_i^t - v_i^t\rangle - \frac{\delta_\varphi}{2}\|u_i^t - v_i^t\|_2^2 \\
&\leqslant \gamma_t D\|u_i^t - v_i^t\|_2 - \frac{\delta_\varphi}{2}\|u_i^t - v_i^t\|_2^2 \\
&\leqslant \frac{(\gamma_t D)^2}{2\delta_\varphi},
\end{aligned} \tag{33}
$$

where the first equation is based on Definition 2; the second inequality is derived from δ_φ-strongly convexity of the function $\varphi(\cdot)$, i.e., Definition 1; the third equation follows from the updating formula (8); the fourth inequality is based on Hölder's inequality and Assumption 5; the last inequality follows from the inequality $a_1 x - a_2 x^2 \leqslant a_1^2/4a_2$ based on $a_2 > 0$ for all $x \in \mathbb{R}$.

Based on Assumptions 1 and 4, one has

$$
\begin{aligned}
\sum_{i=1}^N B_\varphi(u^*, u_i^{t+1}) &= \sum_{i=1}^N B_\varphi\left(u^*, \sum_{j=1}^N [W(t)]_{ij} p_j^t\right) \\
&\leqslant \sum_{j=1}^N \sum_{i=1}^N [W(t)]_{ij} B_\varphi(u^*, p_j^t) \\
&= \sum_{i=1}^N B_\varphi(u^*, v_i^t + e_i^t).
\end{aligned} \tag{34}
$$

Moreover, by Definition 2, one has

$$
B_\varphi(u^*, v_i^t + e_i^t) = B_\varphi(u^*, v_i^t) - B_\varphi(v_i^t + e_i^t, v_i^t) + \langle \nabla\varphi(v_i^t) - \nabla\varphi(v_i^t + e_i^t), u^* - v_i^t - e_i^t\rangle. \tag{35}
$$

Because of the different types of the quantizers, the following proof is divided into two parts.

(a) Under the quantizer $Q_d^t(\cdot)$, based on (35), the generalized Cauchy-Schwarz inequality, and Assumptions 2 and 6, the following inequality can be gained:

$$
\begin{aligned}
B_\varphi(u^*, v_i^t + e_i^t) &\leqslant B_\varphi(u^*, v_i^t) - \frac{\delta_\varphi}{2}\|e_i^t\|_2^2 + \|\nabla\varphi(v_i^t) - \nabla\varphi(v_i^t + e_i^t)\|_* \cdot \|u^* - v_i^t - e_i^t\|_2 \\
&\leqslant B_\varphi(u^*, v_i^t) - \frac{\delta_\varphi}{2}\|e_i^t\|_2^2 + L_\varphi\|e_i^t\|_2(\|u^* - v_i^t\|_2 + \|e_i^t\|_2) \\
&\leqslant B_\varphi(u^*, v_i^t) + \left(L_\varphi - \frac{\delta_\varphi}{2}\right)\|e_i^t\|_2^2 + ML_\varphi\|e_i^t\|_2.
\end{aligned} \tag{36}
$$

Combining the inequalities (34) and (36) yields

$$
\begin{aligned}
\sum_{i=1}^N B_\varphi(u^*, u_i^{t+1}) &\leqslant \sum_{i=1}^N B_\varphi(u^*, v_i^t) + \left(L_\varphi - \frac{\delta_\varphi}{2}\right)\sum_{i=1}^N \|e_i^t\|_2^2 + ML_\varphi \sum_{i=1}^N \|e_i^t\|_2 \\
&= \sum_{i=1}^N B_\varphi(u^*, v_i^t) + \left(L_\varphi - \frac{\delta_\varphi}{2}\right)\sigma_t + ML_\varphi\rho_t,
\end{aligned} \tag{37}
$$

where $\sigma_t = \sum_{i=1}^N \|e_i^t\|_2^2$ and $\rho_t = \sum_{i=1}^N \|e_i^t\|_2$. By using the inequalities (32), (33) and (37), one has

$$
\sum_{i=1}^N \langle d_i^t, u_i^t - u^*\rangle \leqslant \frac{ND^2\gamma_t}{2\delta_\varphi} + \left(\left(L_\varphi - \frac{\delta_\varphi}{2}\right)\sigma_t + ML_\varphi\rho_t\right)\frac{1}{\gamma_t} + \sum_{i=1}^N \frac{B_\varphi(u^*, u_i^t) - B_\varphi(u^*, u_i^{t+1})}{\gamma_t}. \tag{38}
$$

Moreover, the following inequality can be derived by using the property (4):

$$
\sigma_t = \sum_{j=1}^N \|e_j^t\|_2^2 = \sum_{j=1}^N \sum_{i=1}^n [e_j^t]_i^2 \leqslant \frac{Nn}{(\Delta_d^t)^2}. \tag{39}
$$

Then, the inequality (30) can be obtained by using the inequalities (28), (38) and (39) and by summing $\sum_{i=1}^N \langle d_i^t, u_i^t - u^* \rangle$ over $t = 1, 2, \dots, T$.

(b) Under the quantizer $Q_d^t(\cdot)$, taking the expectation of (35), and utilizing the generalized Cauchy-Schwarz inequality, Assumptions 2 and 6, and the property (7), we have

$$\begin{aligned} \mathbb{E}\{B_\varphi(u^*, v_i^t + e_i^t)\} &= \mathbb{E}\{B_\varphi(u^*, v_i^t)\} - \mathbb{E}\{B_\varphi(v_i^t + e_i^t, v_i^t)\} + \mathbb{E}\{\langle \nabla\varphi(v_i^t) - \nabla\varphi(v_i^t + e_i^t), u^* - v_i^t \rangle\} \\ &\leqslant \mathbb{E}\{B_\varphi(u^*, v_i^t)\} - \frac{\delta_\varphi}{2}\mathbb{E}\{\|e_i^t\|_2^2\} + L_\varphi \mathbb{E}\{\|e_i^t\|_2 \cdot \|u^* - v_i^t\|_2\} \\ &\leqslant \mathbb{E}\{B_\varphi(u^*, v_i^t)\} + ML_\varphi \mathbb{E}\{\|e_i^t\|_2\}. \end{aligned} \tag{40}$$

Taking expectation on the inequality (34), and using the inequality (40), we obtain that

$$\sum_{i=1}^N \mathbb{E}\{B_\varphi(u^*, u_i^{t+1})\} \leqslant \sum_{i=1}^N \mathbb{E}\{B_\varphi(u^*, v_i^t)\} + ML_\varphi \mathbb{E}\{\rho_t\}. \tag{41}$$

By taking the expectation of the inequalities (32) and (33), and utilizing the inequality (41), we have

$$\sum_{i=1}^N \mathbb{E}\{\langle d_i^t, u_i^t - u^* \rangle\} \leqslant \frac{ND^2\gamma_t}{2\delta_\varphi} + ML_\varphi \mathbb{E}\{\rho_t\}\frac{1}{\gamma_t} + \sum_{i=1}^N \frac{\mathbb{E}\{B_\varphi(u^*, u_i^t)\} - \mathbb{E}\{B_\varphi(u^*, u_i^{t+1})\}}{\gamma_t}. \tag{42}$$

The inequality (31) can be readily obtained by summing $\sum_{i=1}^N \mathbb{E}\{\langle d_i^t, u_i^t - u^* \rangle\}$ over $t = 1, 2, \dots, T$, and utilizing the inequality (29).

Now, we define an average point of node i as follows:

$$\hat{u}_i^{\mathrm{T}} = \frac{1}{T}\sum_{t=1}^{\mathrm{T}} u_i^t, \quad \forall i \in [N].$$

On the basis of Lemmas 2 and 3, we obtain the following theorem showing the convergence of $\hat{u}_i^{\mathrm{T}}$ to u^*.

Theorem 1. Let Assumptions 1–6 hold, and $\gamma_t = \frac{1}{\delta_f t}$ for $t = 1, 2, \dots, T$. Then, for $j \in [N]$, $T \geqslant 3$, one has the following results.

(a) The following condition holds for the case with the deterministic time-varying quantizer $Q_d^t(\cdot)$:

$$\|\hat{u}_j^{\mathrm{T}} - u^*\|_2^2 \leqslant O_d \frac{\ln(T)}{T} + (E_d^1 + E_d^2 + E_d^3 + E_d^4)\frac{1}{T}, \tag{43}$$

where

$$O_d = \frac{2D^2}{\delta_f^2 \delta_\varphi^2}\left(1 + \frac{4\theta_1\theta_2 N}{1 - \theta_2}\right), \quad E_d^1 = \frac{2D}{\delta_f \delta_\varphi}\left(1 + \frac{2\theta_1\theta_2}{1 - \theta_2}\right)\sum_{j=1}^N \|u_j^1\|_2, \quad E_d^2 = \frac{2n}{\delta_\varphi}\left(L - \frac{\delta_\varphi}{2}\right)\sum_{t=1}^{\mathrm{T}} \frac{t}{(\Delta_d^t)^2},$$

$$E_d^3 = \frac{2ML_\varphi\sqrt{n}}{\delta_\varphi}\sum_{t=1}^{\mathrm{T}} \frac{t}{\Delta_d^t}, \quad E_d^4 = \frac{4\theta_1\theta_2 DN\sqrt{n}}{\delta_f \delta_\varphi(1 - \theta_2)}\sum_{t=1}^{\mathrm{T}} \frac{1}{\Delta_d^t}.$$

(b) The following condition holds for the case with the probabilistic time-varying quantizer $Q_p^t(\cdot)$:

$$\mathbb{E}\{\|\hat{u}_j^{\mathrm{T}} - u^*\|_2^2\} \leqslant O_p \frac{\ln(T)}{T} + (E_p^1 + E_p^2 + E_p^3)\frac{1}{T}, \tag{44}$$

where

$$O_p = O_d, \quad E_p^1 = E_d^1, \quad E_p^2 = \frac{ML_\varphi\sqrt{n}}{\delta_\varphi}\sum_{t=1}^{\mathrm{T}} \frac{t}{\Delta_d^t}, \quad E_p^3 = \frac{2\theta_1\theta_2 DN\sqrt{n}}{\delta_f \delta_\varphi(1 - \theta_2)}\sum_{t=1}^{\mathrm{T}} \frac{1}{\Delta_d^t}.$$

Proof. Based on Assumption 3, we have

$$\sum_{i=1}^N \langle d_i^t, u_i^t - u^* \rangle \geqslant \sum_{i=1}^N (f_i(u_i^t) - f_i(u^*) + \delta_f B_\varphi(u^*, u_i^t))$$

$$\geqslant f(u_j^t) - f(u^*) + \delta_f \sum_{i=1}^{N} B_\varphi(u^*, u_i^t) - D\sum_{i=1}^{N} \|u_i^t - u_j^t\|_2, \tag{45}$$

where $f(u_j^t) = \sum_{i=1}^{N} f_i(u_j^t)$, $f(u^*) = \sum_{i=1}^{N} f_i(u^*)$, and the last inequality is based on the following inequality:

$$\begin{aligned} f_i(u_i^t) &\geqslant f_i(u_j^t) + \langle d_i(u_j^t), u_i^t - u_j^t\rangle \\ &\geqslant f_i(u_j^t) - \|d_i(u_j^t)\|_2 \cdot \|u_i^t - u_j^t\|_2 \\ &\geqslant f_i(u_j^t) - D\|u_i^t - u_j^t\|_2. \end{aligned}$$

Then, according to the inequality (45), we obtain that

$$\sum_{t=1}^{\mathrm{T}} (f(u_j^t) - f(u^*)) \leqslant \sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N} \langle d_i^t, u_i^t - u^*\rangle - \sum_{t=1}^{\mathrm{T}} \delta_f \sum_{i=1}^{N} B_\varphi(u^*, u_i^t) + D\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N} \|u_i^t - u_j^t\|_2. \tag{46}$$

The following proof is divided into two parts.

(a) For the quantizer $Q_d^t(\cdot)$, the following inequality can be gained by combining with the inequalities (30) and (46):

$$\begin{aligned} \sum_{t=1}^{\mathrm{T}} (f(u_j^t) - f(u^*)) &\leqslant \sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N} \frac{B_\varphi(u^*, u_i^t) - B_\varphi(u^*, u_i^{t+1})}{\gamma_t} - \sum_{t=1}^{\mathrm{T}} \delta_f \sum_{i=1}^{N} B_\varphi(u^*, u_i^t) + \frac{ND^2}{2\delta_\varphi}\sum_{t=1}^{\mathrm{T}} \gamma_t \\ &\quad + \left(L_\varphi - \frac{\delta_\varphi}{2}\right) Nn \sum_{t=1}^{\mathrm{T}} \frac{1}{(\Delta_d^t)^2}\frac{1}{\gamma_t} + ML_\varphi N\sqrt{n}\sum_{t=1}^{\mathrm{T}} \frac{1}{\Delta_d^t}\frac{1}{\gamma_t} + D\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N} \|u_i^t - u_j^t\|_2 \\ &\leqslant \frac{ND^2}{\delta_\varphi \delta_f}\ln(T) + \left(L_\varphi - \frac{\delta_\varphi}{2}\right) Nn\delta_f \sum_{t=1}^{\mathrm{T}} \frac{t}{(\Delta_d^t)^2} + ML_\varphi N\sqrt{n}\delta_f \sum_{t=1}^{\mathrm{T}} \frac{t}{\Delta_d^t} \\ &\quad + D\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N} \|u_i^t - u_j^t\|_2, \end{aligned} \tag{47}$$

where the last inequality is derived based on the fact that $\gamma_t = \frac{1}{\delta_f t}$ and the following two inequalities:

$$\begin{aligned} &\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N} \frac{B_\varphi(u^*, u_i^t) - B_\varphi(u^*, u_i^{t+1})}{\gamma_t} - \sum_{t=1}^{\mathrm{T}} \delta_f \sum_{i=1}^{N} B_\varphi(u^*, u_i^t) \\ &= \left(\frac{1}{\gamma_1} - \delta_f\right)\sum_{i=1}^{N} B_\varphi(u^*, u_i^1) + \sum_{t=2}^{\mathrm{T}} \left(\frac{1}{\gamma_t} - \frac{1}{\gamma_{t-1}} - \delta_f\right)\sum_{i=1}^{N} B_\varphi(u^*, u_i^t) - \frac{1}{\gamma_T}\sum_{i=1}^{N} B_\varphi(u^*, u_i^{T+1}) \\ &= -\frac{1}{\gamma_T}\sum_{i=1}^{N} B_\varphi(u^*, u_i^{T+1}) \leqslant 0, \end{aligned} \tag{48}$$

$$\sum_{t=1}^{\mathrm{T}} \gamma_t = \frac{1}{\delta_f}\sum_{t=1}^{\mathrm{T}} \frac{1}{t} \leqslant \frac{1}{\delta_f}\left(1 + \int_1^{\mathrm{T}} \frac{1}{t}\mathrm{d}t\right) \leqslant \frac{2}{\delta_f}\cdot \ln(T), \quad \forall T \geqslant 3. \tag{49}$$

Hence, based on the convexity of the function $f(\cdot)$ and the first-order optimality condition, one has that

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{\mathrm{T}} (f(u_j^t) - f(u^*)) &\geqslant f(\hat{u}_j^{\mathrm{T}}) - f(u^*) \\ &\geqslant \langle d(u^*), \hat{u}_j^{\mathrm{T}} - u^*\rangle + N\delta_f B_\varphi(\hat{u}_j^{\mathrm{T}}, u^*) \\ &\geqslant N\delta_f B_\varphi(\hat{u}_j^{\mathrm{T}}, u^*) \\ &\geqslant \frac{N\delta_f\delta_\varphi}{2}\|\hat{u}_j^{\mathrm{T}} - u^*\|_2^2. \end{aligned} \tag{50}$$

Combining the inequalities (12), (47) and (50) yields the inequality (43).

(b) For the quantizer $Q_p^t(\cdot)$, taking the expectation of the inequalities (46) and (48), and combining the inequalities (31) and (49), one has

$$\sum_{t=1}^{\mathrm{T}}\mathbb{E}\{f(u_j^t)-f(u^*)\}\leqslant\frac{ND^2}{\delta_\varphi\delta_f}\ln(T)+\frac{ML_\varphi N\sqrt{n}\delta_f}{2}\sum_{t=1}^{\mathrm{T}}\frac{t}{\Delta_p^t}+D\sum_{t=1}^{\mathrm{T}}\sum_{i=1}^{N}\mathbb{E}\{\|u_i^t-u_j^t\|_2\}. \tag{51}$$

Then the inequality (44) can be readily derived by taking expectation of the inequality (50), and combining with the inequalities (13) and (51).

Remark 5. Based on the results in Theorem 1, we can summarize the rules of choosing the quantization resolutions (Δ_d^t and Δ_p^t) as follows. Δ_d^t and Δ_p^t should be chosen so that $(E_d^1+E_d^2+E_d^3+E_d^4)/T$ and $(E_p^1+E_p^2+E_p^3)/T$ converge to zero or a fixed constant as time goes to infinity, such that the convergence of the algorithm can be guaranteed. Under these circumstances, the convergence rate of the proposed DQMD algorithm is at least $\mathcal{O}(\ln(T)/T)$.

Based on the results in Theorem 1, one has two following corollaries.

Corollary 1. Let the conditions of Theorem 1 hold, and the quantization resolution is chosen as $\Delta_d^t=\Delta_p^t=\alpha t$ (α is a positive constant). Then for $T\geqslant 3$, one has the following results.

(a) The following condition holds for the case with the deterministic time-varying quantizer $Q_d^t(\cdot)$:

$$\|\hat{u}_j^{\mathrm{T}}-u^*\|_2^2\leqslant C_d^1\frac{\ln(T)}{T}+C_d^2\frac{1}{T}+C_d^3,$$

where

$$C_d^1=\frac{2D^2}{\delta_f^2\delta_\varphi^2}\left(1+\frac{4\theta_1\theta_2N}{1-\theta_2}\right)+\frac{4n}{\delta_\varphi\alpha^2}\left(L-\frac{\delta_\varphi}{2}\right)+\frac{8\theta_1\theta_2DN\sqrt{n}}{\delta_f\delta_\varphi\alpha(1-\theta_2)},$$

$$C_d^2=\frac{2D}{\delta_f\delta_\varphi}\left(1+\frac{2\theta_1\theta_2}{1-\theta_2}\right)\sum_{j=1}^{N}\|u_j^1\|_2,\quad C_d^3=\frac{2ML_\varphi\sqrt{n}}{\delta_\varphi\alpha}.$$

(b) The following condition holds for the case with the probabilistic time-varying quantizer $Q_p^t(\cdot)$:

$$\mathbb{E}\{\|\hat{u}_j^{\mathrm{T}}-u^*\|_2^2\}\leqslant C_p^1\frac{\ln(T)}{T}+C_p^2\frac{1}{T}+C_p^3,$$

where

$$C_p^1=\frac{2D^2}{\delta_f^2\delta_\varphi^2}\left(1+\frac{4\theta_1\theta_2N}{1-\theta_2}\right)+\frac{4\theta_1\theta_2DN\sqrt{n}}{\delta_f\delta_\varphi\alpha(1-\theta_2)},\quad C_p^2=C_d^2,\quad C_p^3=\frac{ML_\varphi\sqrt{n}}{\delta_\varphi\alpha}.$$

Corollary 2. Let the conditions of Theorem 1 hold, and the quantization resolution is chosen as $\Delta_d^t=\Delta_p^t=\alpha t^{(3/2)}$ (α is a positive constant). Then for $T\geqslant 3$, one has the following results.

(a) The following condition holds for the case with the deterministic time-varying quantizer $Q_d^t(\cdot)$:

$$\|\hat{u}_j^{\mathrm{T}}-u^*\|_2^2\leqslant D_d^1\frac{1}{\sqrt{T}}+D_d^2\frac{\ln(T)}{T}+D_d^3\frac{1}{T},$$

where

$$D_d^1=\frac{4ML_\varphi\sqrt{n}}{\delta_\varphi\alpha},\quad D_d^2=\frac{2D^2}{\delta_f^2\delta_\varphi^2}\left(1+\frac{4\theta_1\theta_2N}{1-\theta_2}\right),$$

$$D_d^3=\frac{2D}{\delta_f\delta_\varphi}\left(1+\frac{2\theta_1\theta_2}{1-\theta_2}\right)\sum_{j=1}^{N}\|u_j^1\|_2+\frac{4n}{\delta_\varphi\alpha^2}\left(L-\frac{\delta_\varphi}{2}\right)+\frac{12\theta_1\theta_2DN\sqrt{n}}{\delta_f\delta_\varphi\alpha(1-\theta_2)}.$$

(b) The following condition holds for the case with the probabilistic time-varying quantizer $Q_p^t(\cdot)$:

$$\mathbb{E}\{\|\hat{u}_j^{\mathrm{T}}-u^*\|_2^2\}\leqslant D_p^1\frac{1}{\sqrt{T}}+D_p^2\frac{\ln(T)}{T}+D_p^3\frac{1}{T},$$

where

$$D_p^1 = \frac{2ML_\varphi\sqrt{n}}{\delta_\varphi \alpha}, \quad D_p^2 = D_d^2, \quad D_p^3 = \frac{2D}{\delta_f\delta_\varphi}\left(1+\frac{2\theta_1\theta_2}{1-\theta_2}\right)\sum_{j=1}^{N}\|u_j^1\|_2 + \frac{6\theta_1\theta_2 DN\sqrt{n}}{\delta_f\delta_\varphi\alpha(1-\theta_2)}.$$

Remark 6. Corollary 1 shows that when $\Delta_d^t = \Delta_p^t = \alpha t$, the proposed DQMD algorithm achieves a rate of the convergence at $\mathcal{O}(\ln(T)/T)$ within the error levels C_d^3 and C_p^3, respectively. From Corollary 2, it can be seen that when $\Delta_d^t = \Delta_p^t = \alpha t^{(3/2)}$, the convergence rate of the algorithm is $\mathcal{O}(1/\sqrt{T})$. By comparing Corollaries 1 and 2, we can conclude that the quantization resolutions affect the convergence rate of the DQMD algorithm, and the larger quantization resolutions lead to better convergence rate.

4 Discussion for a special case

In this part, we consider a closed convex set $V \subseteq \mathbb{R}^n$ such that

$$\mathrm{col}_n\{0\} \in V, \quad \|x\|_\infty \leqslant R, \quad \forall x \in V. \tag{52}$$

We also consider a time-varying quantizer described as follows:

$$Q^t(v) = \left[q^t(v_1)\ q^t(v_2)\ \cdots\ q^t(v_n)\right]^{\mathrm{T}}, \quad v \in \mathbb{R}^n, \tag{53}$$

where

$$q^t(v_i) = \begin{cases} \frac{1}{\Delta^t}\lfloor \Delta^t v_i \rfloor, & \text{if } v_i \geqslant 0, \\ \frac{1}{\Delta^t}\lceil \Delta^t v_i \rceil, & \text{if } v_i < 0, \end{cases} \tag{54}$$

and $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are the operation of rounding down and up, respectively. Then, one has that

$$|v_i - q^t(v_i)| \leqslant \frac{1}{\Delta^t}. \tag{55}$$

For the quantizer adopted above, we make the following assumption.

Assumption 7. For the constant R given in (52), the inequality $\frac{1}{\Delta^t} \leqslant R$ holds.

Considering the quantizer (53), we have that $Q^t(u) \in V$ for any $u \in V$. Then, the corresponding DQMD method is given as follows:

$$\begin{cases} \nabla\varphi(w_i^t) = \nabla\varphi(u_i^t) - \gamma_t d_i^t, \\ v_i^t = \arg\min_{u\in V} B_\varphi(u, w_i^t), \\ u_i^{t+1} = \sum\limits_{j=1}^{N} [W(t)]_{ij} Q^t(v_j^t). \end{cases} \tag{56}$$

In the consensus step of the algorithm (56), we can see that the information to be exchanged between nodes is the quantized information, which is different from the algorithms (8)–(11). From this, the algorithm (56) can be viewed as a special case of the algorithms (8)–(11). For the convex optimization problem $\min f(u) = \sum_{i=1}^{N} f_i(u)$ subject to $u \in V$, the desired results based on the algorithm (56) can be obtained as the same as those in Theorem 1.

5 Simulations

In this section, we apply the DQMD algorithm developed in our paper to study the distributed ridge regression problem that is formulated as follows [24]:

$$\begin{aligned} \min \quad & f(u) = \sum_{i=1}^{N}\left(\frac{1}{2}(c_i^{\mathrm{T}}u - d_i) + \alpha\|u\|_2^2\right) \\ \text{s.t.} \quad & u \in U = \{u \in \mathbb{R}^n \mid \|u\|_2 \leqslant 1\}, \end{aligned} \tag{57}$$

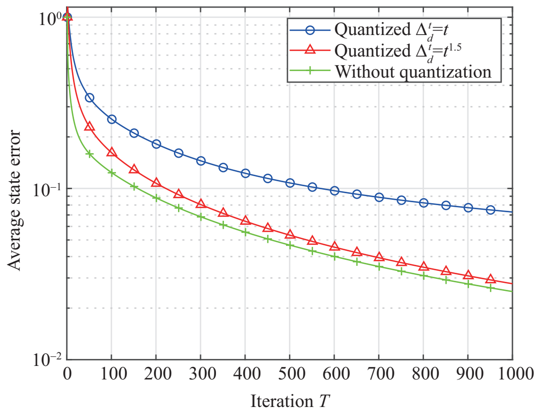

Figure 1 (Color online) The average state error with different quantization resolutions under the deterministic quantizer.

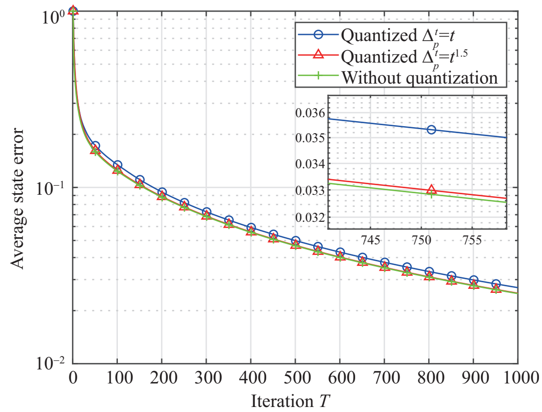

Figure 2 (Color online) The average state error with different quantization resolutions under the probabilistic quantizer.

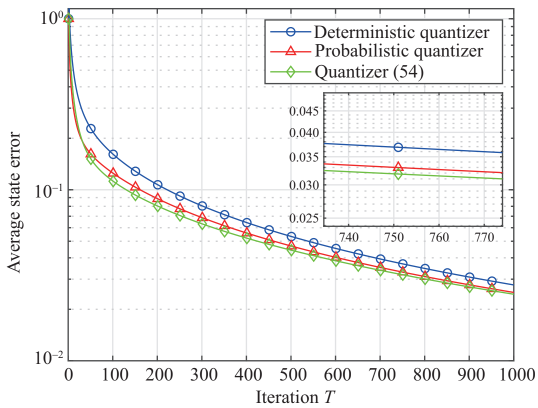

Figure 3 (Color online) The average state error under three kinds of quantizers.

where $c_i \in \mathbb{R}^n$ and $d_i \in \mathbb{R}$ are the data that can only be known by node i. c_i is generated from the unit normal distribution, and $d_i = c_i^{\mathrm{T}} x + e_i$, where $x \in \mathbb{R}^n$ satisfies $[x]_j = 2$ for $1 \leqslant j \leqslant \lfloor n/2 \rfloor$ and 0 otherwise, and e_i represents the noise which is generated from the unit normal distribution.

For the implementation of the DQMD algorithm, a time-varying ring network with 30 nodes is introduced, such that the links of nodes can be changed with the parity of iteration times. The distance-generating function adopted here is $\varphi(u) = \frac{1}{2}\|u\|_2^2$. To explore the effects of the quantizers on the convergence performance of the proposed algorithm, we define the average state error as $\frac{1}{N}\sum_{j=1}^{N} \|\hat{u}_j^{\mathrm{T}} - u^*\|_2^2$ and $\frac{1}{N}\sum_{j=1}^{N} \mathbb{E}\{\|\hat{u}_j^{\mathrm{T}} - u^*\|_2^2\}$ under the deterministic and probabilistic time-varying quantizers, respectively. Moreover, the dimension of the nodes' state is chosen as $n = 3$, and we choose $\alpha = 0.1$.

In order to explore the effects of the quantization resolutions on the convergence of the proposed DQMD algorithm, two simulation results are given. Under the quantizer $Q_d^t(\cdot)$, Figure 1 provides the trajectories of the average state error versus the number of iterations T for $\Delta_d^t = t$, $\Delta_d^t = t^{1.5}$ and no-quantization case. For the quantizer $Q_p^t(\cdot)$, Figure 2 shows the trajectories of the average state error versus the number of iterations T in the case of $\Delta_p^t = t$, $\Delta_p^t = t^{1.5}$ and without quantized communication. From the above simulation results, we can see that the larger quantization resolutions (Δ_d^t and Δ_p^t) lead to better convergence performance of the proposed method under two kinds of quantizers, but the convergence performance is slightly worse than that of the distributed mirror descent algorithm without quantized communication. However, this is more or less expected, since our algorithms consider information quantization and the quantization error will deteriorate the convergence performance.

Moreover, by choosing $\Delta_d^t = \Delta_p^t = \Delta^t = t^{1.5}$, the trajectories of the average state error under three kinds of quantizers are plotted in Figure 3. From this result, we can see that in the case of the same quantization resolution, the convergence performance under the probabilistic quantizer is better than that under the deterministic quantizer, while the convergence performance under the quantizer (54) is slightly better than that under the probabilistic quantizer.

6 Conclusion

In this paper, a DQMD method has been proposed to study the strongly convex constrained optimization problem. We have analyzed the effects of the deterministic and probabilistic time-varying quantizers on the convergence rate of the DQMD algorithm, and the obtained results show that for the suitable quantization resolutions, the DQMD algorithm can converge to the optimal value of the considered problem as the iterations increase. Finally, a distributed ridge regression example has been provided to verify the theoretical results. Based on the work in this paper, our future studies will focus on the distributed mirror descent algorithm combined with other types of quantizers, the event-triggered distributed mirror descent algorithm and the online DQMD algorithm. Moreover, motivated by [37], the quantized scheme for distributed optimization with coupled inequality constraints will also be considered in our future work.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61922044, 62022042), and in part by Open Project of the Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment (Grant No. GDSC202017), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX20_0292).

References

1 Shi Q, He C, Chen H, et al. Distributed wireless sensor network localization via sequential greedy optimization algorithm. IEEE Trans Signal Process, 2010, 58: 3328–3340
2 Liu Z, Yuan Y, Guan X, et al. An approach of distributed joint optimization for cluster-based wireless sensor networks. IEEE/CAA J Autom Sin, 2015, 2: 267–273
3 Molzahn D K, Dorfler F, Sandberg H, et al. A survey of distributed optimization and control algorithms for electric power systems. IEEE Trans Smart Grid, 2017, 8: 2941–2962
4 Braun P, Grune L, Kellett C M, et al. A distributed optimization algorithm for the predictive control of smart grids. IEEE Trans Automat Contr, 2016, 61: 3898–3911
5 Yu W W, Li C J, Yu X H, et al. Economic power dispatch in smart grids: a framework for distributed optimization and consensus dynamics. Sci China Inf Sci, 2018, 61: 012204
6 Li H, Fang C, Lin Z. Accelerated first-order optimization algorithms for machine learning. Proc IEEE, 2020, 108: 2067–2082
7 Hu W, Wu A. Accelerated distributed algorithm for solving linear algebraic equations. In: Proceedings of 2020 Chinese Intelligent Systems Conference, 2021. 705: 389-398
8 Wu A G, Sun H J, Zhang Y. An SOR implicit iterative algorithm for coupled Lyapunov equations. Automatica, 2018, 97: 38–47
9 Yang T, Yi X, Wu J, et al. A survey of distributed optimization. Annu Rev Control, 2019, 47: 278–305
10 Nedic A. Distributed gradient methods for convex machine learning problems in networks: distributed optimization. IEEE Signal Process Mag, 2020, 37: 92–101
11 Yang T, Chai T Y. Research status and prospects of distributed collaborative optimization. Sci Sin Tech, 2020, 50: 1414–1425
12 Lin P, Qi H S. Distributed gradient-based sampling algorithm for least-squares in switching multi-agent networks. Sci China Inf Sci, 2020, 63: 199203
13 Beck A, Teboulle M. Mirror descent and nonlinear projected subgradient methods for convex optimization. Operations Res Lett, 2003, 31: 167–175
14 Nedić A, Lee S. On stochastic subgradient mirror-descent algorithm with weighted averaging. SIAM J Optim, 2014, 24: 84–107
15 Yuan D, Hong Y, Ho D W C, et al. Optimal distributed stochastic mirror descent for strongly convex optimization. Automatica, 2018, 90: 196–203
16 Li J, Li G, Wu Z, et al. Stochastic mirror descent method for distributed multi-agent optimization. Optim Lett, 2018, 12: 1179–1197
17 Shahrampour S, Jadbabaie A. Distributed online optimization in dynamic environments using mirror descent. IEEE Trans Automat Contr, 2018, 63: 714–725
18 Lei Y, Zhou D X. Convergence of online mirror descent. Appl Comput Harmonic Anal, 2020, 48: 343–373
19 Yuan D, Hong Y, Ho D W C, et al. Distributed mirror descent for online composite optimization. IEEE Trans Automat Contr, 2021, 66: 714–729
20 Yi X, Li X, Xie L, et al. Distributed online convex optimization with time-varying coupled inequality constraints. IEEE Trans Signal Process, 2020, 68: 731–746
21 Nedic A, Olshevsky A. Stochastic gradient-push for strongly convex functions on time-varying directed graphs. IEEE Trans Automat Contr, 2016, 61: 3936–3947
22 Sayin M O, Vanli N D, Kozat S S, et al. Stochastic subgradient algorithms for strongly convex optimization over distributed networks. IEEE Trans Netw Sci Eng, 2017, 4: 248–260
23 Yang T, Lu J, Wu D, et al. A distributed algorithm for economic dispatch over time-varying directed networks with delays. IEEE Trans Ind Electron, 2017, 64: 5095–5106
24 Yuan D, Ho D W C, Xu S. Stochastic strongly convex optimization via distributed epoch stochastic gradient algorithm. IEEE Trans Neural Netw Learn Syst, 2021, 32: 2344–2357
25 Yuan D, Xu S, Zhao H, et al. Distributed dual averaging method for multi-agent optimization with quantized communication. Syst Control Lett, 2012, 61: 1053–1061
26 Yi P, Hong Y. Quantized subgradient algorithm and data-rate analysis for distributed optimization. IEEE Trans Control Netw Syst, 2014, 1: 380–392

27 Huang C, Li H, Xia D, et al. Quantized subgradient algorithm with limited bandwidth communications for solving distributed optimization over general directed multi-agent networks. Neurocomputing, 2016, 185: 153–162
28 Li J, Chen G, Wu Z, et al. Distributed subgradient method for multi-agent optimization with quantized communication. Math Meth Appl Sci, 2017, 40: 1201–1213
29 Liu W, Deng F, Liang J, et al. Distributed average consensus in multi-agent networks with limited bandwidth and time-delays. IEEE/CAA J Autom Sin, 2014, 1: 193–203
30 Li H, Huang C, Chen G, et al. Distributed consensus optimization in multiagent networks with time-varying directed topologies and quantized communication. IEEE Trans Cybern, 2017, 47: 2044–2057
31 Pu Y, Zeilinger M N, Jones C N. Quantization design for distributed optimization. IEEE Trans Automat Contr, 2017, 62: 2107–2120
32 Ram S S, Nedić A, Veeravalli V V. Distributed stochastic subgradient projection algorithms for convex optimization. J Optim Theor Appl, 2010, 147: 516–545
33 Nedic A, Ozdaglar A. Distributed subgradient methods for multi-agent optimization. IEEE Trans Automat Contr, 2009, 54: 48–61
34 Beck A. First-Order Methods in Optimization. Philadelphia: SIAM, 2017
35 Bregman L M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comput Math Math Phys, 1967, 7: 200–217
36 Chen G, Teboulle M. Convergence analysis of a proximal-like minimization algorithm using bregman functions. SIAM J Optim, 1993, 3: 538–543
37 Li X, Feng G, Xie L. Distributed proximal algorithms for multiagent optimization with coupled inequality constraints. IEEE Trans Automat Contr, 2021, 66: 1223–1230