Distributed Constrained Optimization With Delayed Subgradient Information Over Time-Varying Network Under Adaptive Quantization

Liu, Jie; Yu, Zhan; Ho, Daniel W. C.

Published in:
IEEE Transactions on Neural Networks and Learning Systems

Published: 01/01/2024

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1109/TNNLS.2022.3172450

Publication details:
Liu, J., Yu, Z., & Ho, D. W. C. (2024). Distributed Constrained Optimization With Delayed Subgradient Information Over Time-Varying Network Under Adaptive Quantization. IEEE Transactions on Neural Networks and Learning Systems, 35(1), 143-156. https://doi.org/10.1109/TNNLS.2022.3172450

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.
© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Liu, J., Yu, Z., & Ho, D. W. C. (2022). Distributed Constrained Optimization With Delayed Subgradient Information Over Time-Varying Network Under Adaptive Quantization. IEEE Transactions on Neural Networks and Learning Systems. https://doi.org/10.1109/TNNLS.2022.3172450.
Distributed Constrained Optimization with Delayed Subgradient Information over Time-Varying Network under Adaptive Quantization

Jie Liu, Zhan Yu, and Daniel W. C. Ho  Fellow, IEEE

Abstract—In this paper, we consider a distributed constrained optimization problem with delayed subgradient information over the time-varying communication network, where each agent can only communicate with its neighbors and the communication channel has a limited data rate. We propose an adaptive quantization method to address this problem. A mirror descent algorithm with delayed subgradient information is established based on the theory of Bregman divergence. With a non-Euclidean Bregman projection-based scheme, the proposed method essentially generalizes many previous classical Euclidean projection-based distributed algorithms. Through the proposed adaptive quantization method, the optimal value without any quantization error can be obtained. Furthermore, comprehensive analysis on the convergence of the algorithm is carried out and our results show that the optimal convergence rate $O(1/\sqrt{T})$ can be obtained under appropriate conditions. Finally, numerical examples are presented to demonstrate the effectiveness of our results.

Index Terms—Distributed Optimization, Mirror Descent Algorithm, Adaptive Quantization, Delayed Subgradient Information.

I. INTRODUCTION

Recently, the distributed optimization algorithms for the network system have been studied widely (see in [1]-[11]). In the distributed optimization problem, there is no central coordination between different agents. Each agent knows about its local function and can only communicate with its neighboring agents in the network. The objective function is composed of sum of local functions. Additionally, these agents, by sending updated information to their neighboring agents, cooperatively minimize the objective function. These distributed methods are critical in many engineering problems, such as localization in sensor networks [12], smart grid optimization [13], aggregative games [14], resource allocation [15], decentralized estimation [16] and distributed control problems [17].

The purpose of distributed optimization algorithms is to solve optimization problems through distributed process in which the agents cooperatively minimize the objective function via information communication. The information communication is often carried out between an agent and its neighbours. Different kinds of distributed optimization algorithms have been proposed in recent years. A subgradient method to solve not necessarily smooth function is proposed in [1]. The distributed stochastic gradient push algorithm is constructed over time-varying directed graphs for distributed optimization in [2]. Distributed dual averaging algorithm over time-varying communication network is investigated in [5]. Distributed gradient algorithm for constrained optimization is proposed in [7]. A collaborative neurodynamic approach to distributed constrained optimization is studied in [8].

Mirror descent algorithms have been studied extensively in recent years. Compared with other distributed subgradient projection methods, mirror descent methods use customized Bregman divergence rather than Euclidean distance, which is a general class of distance measuring functions and can be viewed as non-Euclidean projection method. Euclidean distance and Kullback-Leibler divergence are two well-known types of Bregman divergence. The mirror descent has been shown to be an efficient tool for optimization over large scale distributed networked system [18], [19]. Recently, some mirror descent algorithms are developed based on stochastic sub-gradient [3], delayed gradient [6], and randomized gradient-free [11] for distributed optimization. Also the mirror descent algorithms in [6], [9], [10], [11] are established over time-varying network for distributed optimization.

In the distributed optimization (see [1], [2], [3], [5], [10], [11], [20], [21], [22], [23]), each agent needs to update parameter and calculate (sub)gradient based on local parameter in parallel. Then each agent receives current (sub)gradient information. However, the asynchronous process of updating parameter and calculating (sub)gradient will cause time delay. Under that circumstance, each agent receives outdated (sub)gradient information. Therefore, it is significant to study the distributed optimization algorithms with the presence of time delay. The asynchrony of two processes updating parameters and calculating subgradient is very common in real life, such as master worker architectures for distributed computation [24] and other similar model [25], [26]. Many distributed optimization methods based on delayed subgradient information has been proposed, such as mirror descent [6], gradient-based algorithm [24], dual averaging algorithm [27] and so on.

High requirement of data rates and low transmission delay in information transmission play important roles to design distributed algorithms over distributed networked systems. However, those requirements place strict constraints on communication channel [28]. Quantization method is widely used...
to satisfy communication constraints in different kinds of networked systems, such as networked control systems [7], [29], wireless sensor network [30], neural network [31] and 5G systems [28], [32]. In the decoding scheme of 5G discussed in [28], [32], the transmitted information is usually quantized to satisfy the constraints of communication channels before sending. Thus, information loss will usually be caused by quantizer. To design appropriate quantization method to satisfy communication constraints and also maintain the required accuracy of data in large-scale network system is a challenging research problem.

Further, previous works have proposed many distributed optimization algorithms with different kinds of quantization methods and analyze the quantization’s effect to the convergence of algorithm (see [4], [20], [33]). However, only a sub-optimal value can be obtained from static quantization methods in [4], [20] due to quantization error. Also, [33] proposes an adaptive quantization algorithm with Euclidean projection method, which not only satisfies communication constraints but also maintains required accuracy of data. Therefore, to design adaptive quantization over mirror descent algorithm to satisfy communication constraints is highly desirable in the large-scale distributed networked system.

The contribution of this paper is summarized as follows:

(i) Firstly, the distributed mirror descent algorithm with adaptive quantization is proposed to address limited communication channel. The traditional uniform quantizer uses the static quantization parameters mid-value and interval size while the proposed adaptive quantizer changes the quantization parameters mid-value and interval size at each iteration. Existing works such as [4], [33], [34], [35] design quantizers to address the limited communication channel in distributed optimization algorithm with Euclidean projection method. However, these quantization schemes can not be directly established in the non-Euclidean projection based methods. In this paper, we overcome this difficulty and establish the proposed distributed non-Euclidean quantization method by employing some new techniques on handling mirror descent structure. The appropriate adaptive quantizer is designed to realize the quantization in distributed optimization with non-Euclidean projection method. The proposed adaptive quantizer helps to asymptotically alleviate the quantized error but only uses a finite number of bits for quantization.

(ii) We analyze the convergence of the mirror descent algorithm under adaptive quantization and also derive some sufficient conditions on stepsize and quantization parameter for the convergence of the proposed algorithm. The convergence rates are comprehensively investigated by considering different stepsizes and quantization parameters. Compared with [33], the convergence rate of distributed optimization algorithm in [33] is $O(\ln T/\sqrt{T})$ while the convergence rate in this paper is $O(1/\sqrt{T})$. Our algorithm’s convergence rate is faster than that in [33]. Also, the communication network in [33] is static while we consider a class of time-varying communication network, which is more realistic. Compared with the algorithm in [6], an adaptive quantization method has been designed for mirror descent algorithm to address limited communication capacity and the convergence rate is still $O(1/\sqrt{T})$, which is the same with that in [6]. Furthermore, the assumptions in this paper are much easier to be satisfied. In addition, the objective function’s subgradient has upper bound while the objective function’s gradient in [6] should satisfy Lipschitz continuous.

(iii) The third contribution is that the asynchronous operation of optimization algorithms with the presence of time delay has been considered. After careful analysis, we conclude that the convergence is guaranteed under appropriate conditions with any time delay. Finally, the algorithm can asymptotically converge to an optimal solution without quantization error. In this paper, we significantly improve our previous works [9], [10], [11] on distributed mirror descent methods in several aspects. To the best of our knowledge, this is the first work to propose the adaptive quantization method to address limited communication channel in mirror descent algorithm and simultaneously take delayed subgradient information into consideration in the study of distributed mirror descent method.

The rest of this paper is organized as follows:

Section II introduces some notations and definitions. Section III defines the problem and propose some assumptions. Section IV proposes the mirror descent algorithm with adaptive quantization. Section V analyzes the convergence of the algorithm and discusses how to select stepsize and quantization parameter. Then we show the convergence rate of different stepsize and quantization parameter. Section VI provides numerical simulations to verify theoretical results and Section VII concludes this paper.

II. NOTATION AND DEFINITION

A. Notation

We first introduce some notations. For a vector $x \in \mathbb{R}^n$, $||x||$ and $||x||_\infty$ are Euclidean norm and infinity norm, respectively. The jth entry of vector $x \in \mathbb{R}^n$ is $[x]_j$ and the i,jth row, jth column of matrix $P \in \mathbb{R}^{n\times n}$ is $[P]_{ij}$. For the $a,b,c \in \mathbb{R}^n$, we use $a \in [b,c]$ or $a \leq c$ to denote $[a]_j \in [b]_j, [c]_j$ with $j = 1, \ldots, n$. The vector $1 \in \mathbb{R}^{n\times 1}$ whose each entry is $1$. For a given compact convex set $A$, we use $\text{Proj}_A(a)$ to denote the projection of $a$ to set $A$, with

\[ \text{Proj}_A(a) = \arg \min_{b \in A} ||a - b||. \]

For a given non-smooth convex function $h(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, we use $\partial h(a) = \{ g \in \mathbb{R}^n | h(b) \geq h(a) + g^T(b-a) \}$ to be the set of the subgradient of $h(x)$ at $a$. The set $\partial h(a)$ is nonempty since $h(x)$ is convex function.

Let $h(x) : A \rightarrow B$ be a $\sigma_B$ strongly convex function if and only if $h(b) \geq h(a) + \langle \nabla h(a), b-a \rangle + \frac{\sigma_B}{2} ||b-a||^2$ for any $a, b \in A$. $L(g)$ denotes a Lipschitz constant of the function $g$ if and only if $||g(a) - g(b)|| \leq L||a - b||$ for any $a, b \in \text{dom} g$.

For the two function $F(t)$ and $G(t)$, $F(t) = O(G(t))$ means if there are positive $T > 0$ and $C > 0$ such that $F(t) \leq C G(t)$ when $t > T$.

B. Uniform Quantization

The uniform quantizer in $\mathbb{R}$ with mid-value $z \in \mathbb{R}$, quantization interval size $d \in \mathbb{R}$ and a fixed number of bits $K + 1$ is defined as

\[ Q(z, d, x) = \begin{cases} 
  z - d, & x - z \in (-\infty, -d) \\
  z + \frac{2j - K}{K}d, & x - z + d \in [\frac{2j - K}{K}d, \frac{2(j+1)d}{K}) \\
  z + d, & x - z \in [d, +\infty) 
\end{cases} \]

(1)
where \( j = 0, 1, \ldots, K - 1 \).

For the quantizer in \( \mathbb{R}^n \), we can also define a quantization function \( \hat{Q}(z,d,x) : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \) as follows:
\[
[\hat{Q}(z,d,x)]_i = Q(z)_i, |d|_i, |x|_i.
\]

where \( i = 1, \ldots, n \), \( [\hat{Q}(z,d,x)]_i \), \( |d|_i \) and \( |x|_i \) are the \( i \)-th element of vector \( \hat{Q}(z,d,x) \), \( z \), \( d \) and \( x \), respectively. For the uniform quantizer \( Q(z,d,x) \), \( z \in \mathbb{R}^n \) and \( d \in \mathbb{R}^n \) denote the mid value vector and quantization interval size vector, respectively. When the vector \( x \in \mathbb{R}^n \) falls inside the quantization interval \([z - d, z + d]\) \( j \), the quantization error is bounded by
\[
|x - \hat{Q}(z,d,x)| \leq \frac{2\sqrt{n}}{K} ||d||_\infty \leq \frac{2\sqrt{n}}{K} ||d||.
\]

\[\text{III. DISTRIBUTED OPTIMIZATION OVER TIME VARYING NETWORK}\]

In this paper, we consider a distributed optimization problem defined over time-varying communication network with \( N \) nodes. The function \( f_j(x) : \mathcal{X} \to \mathbb{R} \) with \( j = 1, 2, \ldots, N \) are non-smooth convex functions and \( \mathcal{X} \subset \mathbb{R}^n \) is non-empty, convex and compact. The objective function is
\[
\min_{x \in \mathcal{X}} f(x) = \sum_{j=1}^{N} f_j(x).
\]

The optimal value set \( \mathcal{X}^* = \arg \min_{x \in \mathcal{X}} f(x) \) of problem (4) is not empty. There is no central coordination between the agents and each agent \( j \) only knows its local function \( f_j(x) \). The direct graph \( \mathcal{G}(t) = \{V, E(t), P(t)\} \) denotes time-varying communication network topology, where \( V = \{1, 2, \ldots, N\} \), \( E(t) = \{(j,i) | \text{agent } j \text{ and } i \text{ are connected}, i, j \in V \} \) and \( P(t) \in \mathbb{R}^{N \times N} \) is the correspond weigh matrix at the time \( t \).

We define \( N_i(t) = \{j \in V | (i,j) \in E(t)\} \) as agent \( i \)'s neighbor set at time \( t \). The node \( i \) sends information to node \( j \) at time \( t \) if and only if \( j \in N_i(t) \). Agent \( i \) and \( j \) are not connected at time \( t \) if and only if \( [P(t)]_{ij} = 0 \).

Assumption 1: The time-varying network's corresponding communication matrix \( P(t) \) is doubly stochastic at each time \( t \), i.e. \( \sum_{i=1}^{N} [P(t)]_{ij} = 1 \) and \( \sum_{j=1}^{N} [P(t)]_{ij} = 1 \) for all \( t \) and \( i, j \in V \).

Assumption 2: The time-varying network \( (V, E(t), P(t)) \) is \( B \) connectivity. There exists a positive integer \( B \) such that the graph \( (V, (j+1)B, B \{E(t)\}) \) is strongly connected for any \( c \geq 0 \). There is a \( \theta \in (0,1) \) such that \( [P(t)]_{ij} \geq \theta \) if \( (i,j) \in E(t) \) and \( [P(t)]_{ij} \geq \theta \) for all \( i \in V \).

Assumptions 1 and 2 are widely used in distributed optimization over time-varying communication network. In this paper, we define the transition matrix \( P(t,s) = \prod_{s=0}^{t-1} P(s-i) \) for \( t \geq s \) and \( P(s,s+1) = I_n \) for any \( s \geq 0 \). The following Lemma 1 is critical in the analysis of time-varying communication network.

Lemma 1: [1] If Assumptions 1 and 2 are satisfied, we have
\[
[\bar{P}(m,n)]_{ij} \left( 1 - \frac{1}{N} \right)^m \leq \omega^m \left( 1 - \frac{1}{\gamma N} \right)^n,
\]
for all \( i, j \in V \) and \( m, n \) satisfying \( m \geq n \geq 1 \), where \( \omega = (1 - \frac{\theta}{\gamma N})^{-2} \) and \( \gamma = (1 - \frac{\theta}{\gamma N})^{-\frac{1}{2}} \).

In this paper, we will develop a mirror descent algorithm to solve the distributed optimization problem. We consider a continuously differentiable distance generating function \( \phi(x) : \mathbb{R}^n \to \mathbb{R} \), which is \( \sigma \) strongly convex over \( \mathcal{X} \) with respect to \( ||\cdot|| \), and the corresponding Bregman divergence \( V(x) \), \( \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \) as
\[
V(x) = ||x - \phi(x)||^2.
\]

The following separate convexity assumption on Bregman divergence is standard in the study of distributed mirror descent algorithms (e.g. [9, 10]).

Assumption 3: The Bregman divergence \( V(x) \) satisfies the separate convexity. For any vector \( a \in \mathbb{R}^n \) and a sequence vectors \( \{a\}_{j=1}^{N} \), we have
\[
V(x) = \sum_{j=1}^{N} w_j V(a,b),
\]
where \( \sum_{j=1}^{N} w_j = 1 \).

The following boundedness assumption is always used for (sub)gradient and Bregman divergence (e.g. [10]).

Assumption 4: The norm of \( g_i(x) \) \( \|g_i(x)\| \leq G \), \( i = 1, 2, \ldots, N \) for each time \( t \) and \( \|x - y\| \leq D_\phi \), where \( D_\phi \) is the bound of \( \sup_{x,y \in \mathcal{X}} \|x - y\| \) obtained.

We will use the following assumption on the distance generating function \( \phi(x) \).

Assumption 5: The distance generating function \( \phi(x) \) is Lipschitz continuity with constant \( L_\phi \)
\[
||\nabla \phi(a) - \nabla \phi(b)|| \leq L_\phi \|a - b\|,
\]
for any \( a, b \in \mathcal{X} \), where \( \mathcal{X} = \{x \in \mathbb{R}^n | \text{there exist } y \in \mathcal{X} \text{ such that } -\frac{\Gamma(2N+\sqrt{n})}{\sqrt{n}} \leq x - y \leq \frac{\Gamma(2N+\sqrt{n})}{\sqrt{n}} \text{ for } i = 1, 2, \ldots, n\} \).

Remark 1: After quantization, some values may not belong to \( \mathcal{X} \), which is the reason to assume distance generating function \( \phi(x) \) is Lipschitz continuity with constant \( L_\phi \) on \( \mathcal{X} \) rather than \( \mathcal{X} \).

Assumption 5 will be used in the proof of Lemma 10. The set \( \mathcal{X} \) defined in Assumption 5 is bounded and closed due to \( \mathcal{X} \) is non-empty and compact. Thus distance generating function \( \phi(x) \) is easy to satisfy Lipschitz continuity with constant \( L_\phi \) on \( \mathcal{X} \). In fact, there are many common distance generating function \( \phi(x) \) satisfy Assumption 5, such as Euclidean distance \( \frac{1}{2} ||x||^2 \) and exponential \( \sum_{i=1}^{n} e^{x_i} \), Shannon entropy \( \sum_{i=1}^{n} (x_i \log(x_i) - x_i) \) if any element \( x \in \mathcal{X} \) satisfies \( x_i > 0 \) for all \( i = 1, 2, \ldots, n \) and so on.

IV. MIRROR DESCENT ALGORITHM WITH DELAYED SUBGRADIENT INFORMATION UNDER ADAPTIVE QUANTIZATION

In this paper, we design a mirror descent algorithm (proposed in Algorithm 1) with delayed subgradient information under adaptive quantization.

Some notations in Algorithm 1 need to be introduced. \( \tau \) is the time delay, \( x_i(t) \) and \( z_i(t) \) are the state and quantization mid value of agent \( i \) at time \( t \), respectively; \( d(t) \) is the quantization interval size for all agents at time \( t \); The quantizer \( \hat{Q} \) is defined in equation (2). Also, \( y_i(t) \) is the weighted sum of quantized
information received from agent $i$’s all neighbors. The non-increasing positive sequences $\{\alpha(t)\}$ and $\{\beta(t)\}$ are defined as the sequences of stepsize and quantization parameter, respectively, where $0 < \alpha(t), \beta(t) < 1$ for any $t$ and $\alpha(t), \beta(t)$ are decreasing to 0. Note that $\{\alpha(t)\}$ and $\{\beta(t)\}$ are pre-determined and need to satisfy some conditions, which will be discussed in detail in Section V.

Algorithm 1 Mirror Descent Algorithm with Delayed Subgradient Information under Adaptive Quantization

Initialize: $x_i(0) \in X$, $z_i(0) = x_i(0)$, $\hat{y}_i(-\tau)$, $\hat{y}_i(-\tau + 1)$, $\cdots$, $\hat{y}_i(-1) \in X$ with $i = 1, \cdots, N$, stepsize sequence $\{\alpha(t)\}$ and quantization parameter sequence $\{\beta(t)\}$.

1: for $t = 0, 1, 2, \cdots$
2: Compute $d(t) = \frac{\beta(t)}{\alpha(t)}$;
3: for $i = 1: t \leq N: i + 1$
4: Agent $i$ receives quantized values $\hat{Q}(z_i(t), d(t), x_j(t))$ from all neighbors $j \in N_i(t)$;
5: Update $y_i(t) = \sum_{j=1}^{N} [P(t)]_{ij} \hat{Q}(z_i(t), d(t), x_j(t))$;
6: Compute $\hat{y}_i(t) = \text{Proj}_{X}(y_i(t))$;
7: Compute the subgradient $g_i(t - \tau) \in \partial f_i(\hat{y}_i(t - \tau))$;
8: Update $z_i(t + 1)$ and $x_i(t + 1)$ as follows:

$$z_i(t + 1) = \arg \min_{x \in X} \left\{ g_i(t - \tau), x + \frac{V_\phi(x, \hat{y}_i(t))}{\alpha(t + 1) (1 - \beta(t + 1))} \right\}$$

$$x_i(t + 1) = \arg \min_{x \in X} \left\{ g_i(t - \tau), x + \frac{V_\phi(x, \hat{y}_i(t))}{\alpha(t + 1)} \right\}$$

9: end for
10: end for

Some steps in Algorithm 1 are explained here. Step 2 calculates the quantization interval for each agent’s quantization function. Step 4 calculates the projection of weighted sum of quantization information from agent $i$’s neighboring agents. Step 5 calculates the projection of weighted sum of quantization information in Step 4 onto the domain $X$. Step 8 updates the quantization mid-value $z_i(t + 1)$ and state $x_i(t + 1)$ of agent $i$.

The following key inequalities Lemmas 2 and 3 are provided to show that the state $x_i(t)$ falls into the quantization interval $[z_i(t) - d(t), z_i(t) + d(t)]$ at each iteration, which play a crucial role in the derivation of our main results.

Lemma 2: [36] For the Bregman divergence $V_\phi(x, z)$ and $x^+_1 = \arg \min_{x \in X} \{ \alpha(g_1, x) + V_\phi(x, y) \}$, $x^+_2 = \arg \min_{x \in X} \{ \alpha(g_2, x) + V_\phi(x, y) \}$, where for any $\alpha > 0$, $g_1, g_2 \in \mathbb{R}^n$ and $y \in X$. Then we have $\|x^+_2 - x^+_1\| \leq \frac{\alpha}{\sigma_\phi} \|g_2 - g_1\|$, where $\sigma_\phi$ is the modulus of strong convexity of distance generating function $\phi$.

Lemma 3: The $x_i(t)$ generated by Algorithm 1 falls into the quantization intervals $[z_i(t) - d(t), z_i(t) + d(t)]$ for $i = 1, 2, \cdots, N$ at each iteration.

Proof: From Lemma 2, for any $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, n$, we know that $\|x_i(t) - z_i(t)\| \leq \frac{\|g(t - 1 - \tau)\| |\alpha(t)\| \beta(t)}{\sigma_\phi} = [d(t)]_j$.

Hence, for any $j = 1, 2, \cdots, n$, we have $\|x_i(t) - z_i(t)\| \leq \|x_i(t) - z_i(t)\| \leq |d(t)|_j$.

which is equivalent to $z_i(t) - d(t) \leq x_i(t) \leq z_i(t) + d(t)$.

Therefore, $x_i(t)$ generated by Algorithm 1 will fall into quantization intervals $[z_i(t) - d(t), z_i(t) + d(t)]$ for $i = 1, 2, \cdots, N$ at each iteration.

From Lemma 3 and inequality (3), we know that the quantization error satisfies

$$\|x_i(t) - \hat{Q}(z_i(t), d(t), x_j(t))\| \leq \frac{2\sqrt{N}}{\beta(t)} \|d(t)\|.$$

The quantization interval $d(t + 1)$ in Algorithm 1 satisfies

$$\|d(t)\| = \frac{\sqrt{2\sqrt{N}}}{{\alpha(t)} \beta(t)}.$$

Both stepsize $\alpha(t)$ and quantization parameter $\beta(t)$ are to be designed to decrease to 0 and then the quantization error $\|x_i(t) - \hat{Q}(z_i(t), d(t), x_j(t))\|$ will decrease to 0.

Remark 2: In this paper, a new adaptive quantization method is designed through updating uniform quantizer’s mid-value and quantization interval at each iteration. Thus, how to find an appropriate quantization mid value $z_i(t)$ and the corresponding quantization interval size $d(t)$ to make transferred information $x_i(t)$ fall into quantization interval $[z_i(t) - d(t), z_i(t) + d(t)]$ is the key to construct adaptive quantizer over mirror descent algorithm. Lemma 2 helps us to overcome the difficulty by finding $z_i(t)$ and $d(t)$.

Moreover, Lemma 3 shows agent $i$’s transferred information $x_i(t)$ falls into agent $i$’s quantization interval $[z_i(t) - d(t), z_i(t) + d(t)]$ at each iteration. The techniques in Lemmas 2 and 3 indicate the essential differences as compared with existing works such as [4], [33], [34], [35].

V. MAIN RESULTS

In this section, we will analyze the convergence of distributed mirror descent algorithm with delayed subgradient information under adaptive quantization and discuss the convergence rate of different stepsize $\alpha(t)$ and quantization parameter $\beta(t)$.

A. Convergence Analysis

The Steps 5, 6 and 8 of Algorithm 1 are shown as follows:

$$y_i(t) = \sum_{j=1}^{N} [P(t)]_{ij} \hat{Q}(z_j(t), d(t), x_j(t)),$$

$$\hat{y}_i(t) = \text{Proj}_{X}(y_i(t)),$$

$$x_i(t + 1) = \arg \min_{x \in X} \left\{ g_i(t - \tau), x + \frac{V_\phi(x, \hat{y}_i(t))}{\alpha(t + 1)} \right\}.$$

where $P(t)$ is time-varying double stochastic communication matrix at time $t$, $\hat{y}_i(t)$ is projection of $y_i(t)$ onto $X$, $V_\phi$ is Bregman divergence and quantizer $Q$ is defined as equation (2).

We will use the upper bound of quantization error to analyze the convergence rate of Algorithm 1. In order to present clearly, we use $e_j(t)$ and $p_i(t)$ to denote quantization error and projection error, respectively with $i, j = 1, 2, \cdots, N$,

$$e_j(t) = \hat{Q}(z_j(t), d(t), x_j(t)) - x_j(t),$$

$$p_i(t) = \hat{y}_i(t) - y_i(t).$$

The equivalent forms of (12), (13) and (14) are

$$\hat{y}_i(t) = \sum_{j=1}^{N} [P(t)]_{ij} (x_j(t) + e_j(t)) + p_i(t),$$

$$x_i(t + 1) = \arg \min_{x \in X} \left\{ g_i(t - \tau), x + \frac{V_\phi(x, \hat{y}_i(t))}{\alpha(t + 1)} \right\}.$$
From inequality (10) and equations (11), (15), hence we know that

$$||e_1(t)|| = ||Q(z_j(t), d(t), x_j(t)) - x_j(t)|| \leq \frac{2G\alpha}{K\sigma_\phi} \alpha(t)\beta(t),$$  \hspace{1cm} (19)

for \( j = 1, 2, \ldots, N \). We define that

$$E(t) = \frac{2G\alpha}{K\sigma_\phi} \alpha(t)\beta(t)$$  \hspace{1cm} (20)

and from (19) and (20), then we have

$$||e_j(t)|| \leq E(t).$$  \hspace{1cm} (21)

Similarly, the upper bound of projection error \( ||p_i(t)|| \) should be obtained in order to analyze Algorithm 1’s convergence. We have the following Lemma 4 to obtain the projection error’s upper bound.

**Lemma 4:** The Euclidean norm of projection error \( p_i(t) \) satisfies

$$||p_i(t)|| \leq 2NE(t),$$  \hspace{1cm} (22)

where \( E(t) \) is defined as (20).

**Proof:** From equation (12), (15), (16) and triangle inequality , we have

$$||p_i(t)|| = \left||\tilde{y}_i(t) - y_i(t)\right||$$

$$\leq \left||y_i(t) - \sum_{j=1}^{N} [P(t)_{ij} x_j(t)] + NE(t) \right||$$

$$\leq 2NE(t),$$

where inequality (23) is obtained from projection theorem, respectively.

We define Bregman projection error as

$$\varepsilon_i(t) = x_i(t+1) - \tilde{y}_i(t)$$  \hspace{1cm} (24)

and need to obtain the Bregman projection error’s upper bound. We have the following Lemma 5 about the upper bound of \( ||\varepsilon_i(t)|| \).

**Lemma 5:** Bregman projection error \( \varepsilon_i(t) \) with \( i = 1, 2, \ldots, N \) satisfies

$$||\varepsilon_i(t)|| \leq \frac{Ga(t)}{\sigma_\phi}.$$  \hspace{1cm} (25)

**Proof:** The first order optimality of \( x_i(t+1) \) implies, for \( \forall x \in X \), we have

$$\langle \alpha(t+1)g_i(t, -\tau) + \nabla \phi(x_i(t+1)) - \nabla \phi(\tilde{y}_i(t)), x - x_i(t+1) \rangle \geq 0.$$  

Substitute \( \tilde{y}_i(t) \in X \) into above inequality, and we have

$$\langle \alpha(t+1)g_i(t, -\tau) + \nabla \phi(x_i(t+1)) - \nabla \phi(\tilde{y}_i(t)),$$

$$\tilde{y}_i(t) - x_i(t+1) \rangle \geq 0.$$  \hspace{1cm} (26)

Rearrange the terms in (25) and we have

$$\langle \alpha(t+1)g_i(t, -\tau), \tilde{y}_i(t) - x_i(t+1) \rangle$$

$$\geq \langle \nabla \phi(x_i(t+1)) - \nabla \phi(\tilde{y}_i(t)), x_i(t+1) - \tilde{y}_i(t) \rangle - \sigma_\phi \|x_i(t+1) - \tilde{y}_i(t)\|^2,$$  \hspace{1cm} (27)

where the second inequality is derived from \( \sigma_\phi \) strongly convex of \( \phi(x) \). From Cauchy inequality, we have

$$\alpha(t+1)||g_i(t, -\tau)|| ||x_i(t+1) - \tilde{y}_i(t)||$$

$$\geq \langle \alpha(t+1)g_i(t, -\tau), \tilde{y}_i(t) - x_i(t+1) \rangle.$$  \hspace{1cm} (28)

Combining (26), (27) and subgradient’s upper bound (6), we have

$$\sigma_\phi \|x_i(t+1) - \tilde{y}_i(t)\|^2 \leq Ga(t+1) \|x_i(t+1) - \tilde{y}_i(t)\|.$$  \hspace{1cm} (29)

Therefore, the upper bound of Bregman projection error \( ||\varepsilon_i(t)|| \) is obtained from equation (24) and inequality (28).

$$||\varepsilon_i(t)|| \leq \frac{Ga(t+1)}{\sigma_\phi} \leq \frac{Ga(t)}{\sigma_\phi}.$$  \hspace{1cm} (30)

The proof is completed.

**Remark 3:** It is important to note that (25) is a crucial part for proving the convergence of Algorithm 1. The Step 6 in Algorithm 1 ensure that the \( \tilde{y}_i(t) \) always belongs to \( X \), which is a necessary condition in (25).

In order to derive the expression of \( x_i(t) \) from equation (17), we define that \( c_i(t) = \sum_{s=1}^{N} [P(t)_{ij} x_j(s-1) + \varepsilon_j(s-1)] \) and have

$$x_i(t) = \sum_{s=1}^{t} \sum_{i=1}^{N} [P(t-s)_{ij} x_j(s-1) + \varepsilon_j(s-1)]$$

$$+ \sum_{j=1}^{N} [P(t-1, 0)_{ij} x_j(0)].$$

The average state of all nodes at time \( t \) is \( \bar{x}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t) \) and we have

$$\bar{x}(t) = \frac{1}{N} \sum_{s=1}^{t} \sum_{i=1}^{N} (c_j(s-1) + \varepsilon_j(s-1)) + \frac{1}{N} \sum_{j=1}^{N} x_j(0).$$

For each agent \( i \in V \), at iteration \( t \), we consider the following variable

$$\tilde{x}_i(T) = \frac{1}{T} \sum_{i=1}^{T} x_i(t).$$

We will analyze the property of \( f(\tilde{x}_i(T)) - f(x^*) \) and derive its upper bound. From the convexity of \( f(x) \), we have

$$f(\tilde{x}_i(T)) - f(x^*) \leq \frac{1}{T} \sum_{t=1}^{T} f(x_i(t)) - f(x^*).$$  \hspace{1cm} (29)

Thus, we need to derive the upper bound of \( \frac{1}{T} \sum_{t=1}^{T} f(x_i(t)) - f(x^*) \). \hspace{1cm} (30)

The inequality in (30) is obtained from definition of subgradient \( g_i(t) = \partial f_i(\tilde{y}_i(t)) \) for \( 1 \leq i \leq N \), from definition of subgradient \( g_i(x_i(t-1)) \) in \( \partial f_i(x_i(t-1)) \) and \( ||g_i(x_i(t-1))|| \leq G \), we have

$$\frac{1}{N} \sum_{i=1}^{N} f_i(\tilde{y}_i(t))$$

$$\geq f(x_i(t)) - \frac{G}{N} \sum_{i=1}^{N} ||\tilde{y}_i(t) - x_i(t)||,$$  \hspace{1cm} (31)

where the inequality (31) is obtained from Cauchy inequality. From equation (17) and triangle inequality, we have

$$||\tilde{y}_i(t) - x_i(t)||$$

$$\leq \sum_{j=1}^{N} [P(t)_{ij} x_j(t-\tau) - x_i(t-\tau)]$$

$$+ 2NE(t-\tau),$$

where inequality (32) is obtained from inequality (21) and (22). Therefore, from inequalities (31) and (32), we have

$$\frac{1}{N} \sum_{i=1}^{N} f_i(\tilde{y}_i(t))$$

$$\geq f(x_i(t)) - \frac{G}{N} \sum_{i=1}^{N} ||\tilde{y}_i(t) - x_i(t)||$$

$$\geq f(x_i(t)) - 3NGE(t-\tau)$$

$$- \frac{G}{N} \sum_{i=1}^{N} ||x_i(t) - x_i(t-\tau)||.$$  \hspace{1cm} (33)

Substitute (33) into (30) and we have

$$\frac{1}{N} \sum_{i=1}^{N} [g_i(t) - \tilde{y}_i(t) - x^*]$$

$$\geq f(x_i(t)) - f(x^*) - 3NGE(t-\tau)$$

$$- \frac{G}{N} \sum_{i=1}^{N} ||x_i(t) - x_i(t-\tau)||.$$  \hspace{1cm} (34)
After summing up inequality (34) from \( t = \tau + 1 \) to \( T + \tau \) and dividing both side by \( T \), we have

\[
\frac{1}{NT} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} (g_i(t, \tau), \tilde{y}_i(t, \tau) - x^*)
\]
\[
\geq \frac{1}{T} \sum_{t=\tau+1}^{T+\tau} f(x(t, \tau)) - f(x^*) - \frac{3NG}{T} \sum_{t=\tau+1}^{T+\tau} E(t) - \frac{G}{TN} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} ||x_j(t, \tau) - x_i(t, \tau)||. \tag{35}
\]

Rearrange terms in inequality (35) and we can derive the upper bound of the following terms in (35),

\[
\frac{1}{T} \sum_{t=\tau+1}^{T+\tau} f(x(t, \tau)) - f(x^*) \leq \frac{1}{NT} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} (g_i(t, \tau), \tilde{y}_i(t, \tau) - x^*)
\]
\[
+ \frac{G}{TN} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} ||x_j(t, \tau) - x_i(t, \tau)||
\]
\[
+ \frac{3NG}{T} \sum_{t=\tau+1}^{T+\tau} E(t). \tag{36}
\]

Substitute inequality (36) into inequality (29) and we have

\[
f(\tilde{x}(T)) - f(x^*) \leq \frac{1}{NT} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} (g_i(t, \tau), \tilde{y}_i(t, \tau) - x^*)
\]
\[
+ \frac{1}{NT} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} (g_i(t, \tau), \tilde{y}_i(t, \tau) - \tilde{y}_i(t))
\]
\[
+ \frac{G}{TN} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} ||x_j(t, \tau) - x_i(t, \tau)||
\]
\[
+ \frac{3NG}{T} \sum_{t=\tau+1}^{T+\tau} E(t). \tag{37}
\]

In order to obtain the upper bound of \( f(\tilde{x}(T)) - f(x^*) \), we need Lemmas 6-8 to estimate the upper bound of the first, second and third term in (37), respectively. The proofs of Lemmas 6, 7, 8 are shown in Appendices A, B, C, respectively.

**Lemma 6:** For the first term in (37), we have

\[
\frac{1}{NT} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} (g_i(t, \tau), \tilde{y}_i(t, \tau) - x^*) \leq B_1. \tag{38}
\]

where \( B_1 = \sqrt{2D_o} (2N+1) L_o \sum_{t=\tau+1}^{T+\tau} E(t) \alpha(t) + \frac{G^2}{\sigma_o} \sum_{t=\tau+1}^{T+\tau} \alpha(t) + \frac{D_o}{T \alpha(T+\tau)} + L_o (8N^2 + 2) \frac{1}{T} \sum_{t=\tau+1}^{T+\tau} E(t) \alpha(t). \)

**Lemma 7:** For the second term in (37), we have

\[
\frac{1}{NT} \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} (g_i(t, \tau), \tilde{y}_i(t, \tau) - \tilde{y}_i(t)) \leq B_2. \tag{39}
\]

where \( B_2 = \frac{(2B+1) \gamma^2}{T \sigma_o^2} \sum_{t=0}^{T+\tau} \alpha(t) + \frac{6(2B+1) N G \gamma}{T} \sum_{t=0}^{T+\tau} E(t) + \frac{2N^2 \omega A}{T \gamma}. \)

**Lemma 8:** We define that \( A = \sum_{j=1}^{N} ||x_j(0)||, \quad B = (2N + N^2 \omega) \gamma \) and have

\[
\sum_{t=1}^{T} \sum_{N} ||x_i(t) - \tilde{x}(t)|| \leq \frac{N \omega}{1-\gamma} A + B \sum_{t=0}^{T} \frac{G \alpha(t)}{\sigma_o} + 3NE(t)
\]
\[
\sum_{t=1}^{T} \sum_{N} ||x_i(t) - x_j(t)|| \leq \frac{2N \omega}{1-\gamma} A + 2B \sum_{t=0}^{T} \frac{G \alpha(t)}{\sigma_o} + 3NE(t)
\]

From Lemma 8, for the third term in (37), we define that

\[
B_3 = \frac{2G A \omega}{T \gamma(1-\gamma)} + \frac{6BG}{T} \sum_{t=0}^{T} E(t) + \frac{2BG^2}{\sigma_o T} \sum_{t=0}^{T} \alpha(t) \quad \text{and have}
\]
\[
\left| \sum_{t=\tau+1}^{T+\tau} \sum_{i=1}^{N} ||x_j(t, \tau) - x_i(t, \tau)|| \right| \leq B_3. \tag{40}
\]

With the above Lemmas 6-8, we have established the upper bound of the first, second and third terms in (37) and then we have the following theorem.

**Theorem 1:** Suppose that Assumptions 1-5 holds and let the sequence \((x_t(t))_{t \geq 1}\) for all \( t \) be generated by Algorithm 1. Let \( f(x^*) \) be the optimal value of problem (4). Then for any \( t \) we have

\[
f(\tilde{x}(T)) - f(x^*) \leq C(T) + D(T), \tag{41}
\]

where \( C(T) = \left( \frac{2B+1 \gamma^2}{T \sigma_o} + \frac{G^2}{\sigma_o} \sum_{t=0}^{T+\tau} \alpha(t) + \frac{2N^2 \omega A}{T \gamma} \right) + \frac{D_o}{T} \alpha(T+\tau), \quad D(T) = \frac{L_o (8N^2 + 2) \gamma}{T} \sum_{t=0}^{T+\tau} E(t) \alpha(t) + (6B + 3NG \gamma) \frac{1}{T} \sum_{t=0}^{T+\tau} E(t) \alpha(t), \quad A = \sum_{j=1}^{N} ||x_j(0)|| \quad \text{and} \quad B = (2N + N^2 \omega) \gamma. \)

**Proof:** Substitute inequalities (38), (39) and (40) into inequality (37), and then we have

\[
f(\tilde{x}(T)) - f(x^*) \leq B_1 + B_2 + B_3 + \frac{3NG}{T} \sum_{t=\tau+1}^{T+\tau} E(t). \tag{42}
\]

From the expressions of \( B_1, B_2, B_3 \) in Lemmas 6, 7 and 8, inequality (41) can be obtained by rearranging the terms. The proof is completed.

The main idea of proofing Theorem 1 is shown as follows: Firstly, we use the convexity of global objective function \( f(x) \), definition of local objective function \( f_i(x) \)'s subgradient \( g_i(t) \) of each agent \( i \) at time \( t \) and its upper bound from Assumption 5 to derive (29), (30) and (31). Secondly, we use the definition of quantization error \( e_i(t) \), projection error \( p_i(t) \) of agent \( i \) at time \( t \) from (15), (16) and their corresponding upper bound from (21), (22) to derive (37). Finally, we derive the upper bound of first term, second term and third term of (37) through Lemma 6, Lemma 7 and Lemma 8, respectively.

From (41), we know that \( C(T) \) and \( D(T) \) defined in Theorem 1 show the convergence of Algorithm 1. Time delay \( \tau \) is related to the convergence because both \( C(T) \) and \( D(T) \) involve time delay \( \tau \). The convergence of Algorithm 1 under subgradient information without delay and adaptive quantizer can be obtained by considering \( \tau = 0 \) in (41). In addition, the term \( D(T) \) in (41) shows the quantization effect related to the convergence of Algorithm 1. When \( E(t) \) is set to 0, the remaining \( C(T) \) shows the convergence of Algorithm 1 under delayed subgradient information without quantization. In fact, the proposed mirror descent algorithm method provides a general setting which covers those cases without both delay subgradient and quantization.

**Remark 4:** We get the upper bound of \( f(\tilde{x}(T)) - f(x^*) \) in Theorem 1. From inequality (41), we know that the properties of stepsize \( \alpha(t) \) and quantization parameter \( \beta(t) \) decide the convergence of Algorithm 1. Thus, the appropriate conditions of \( \alpha(t) \) and \( \beta(t) \) need to be discussed. Next, we will discuss how to design stepsize and quantization parameter.

**B. How to Design Stepsize and Quantization Parameter**

In this subsection, we will show some conditions of stepsize \( \alpha(t) \) and quantization parameter \( \beta(t) \) for Algorithm 1's convergence.
Theorem 2: When the stepsize $\alpha(t)$ and quantization parameter $\beta(t)$ satisfy the following conditions (42), (43) and (44),
\[
\lim_{T \to \infty} T\alpha(T + \tau) = 0, \quad (42)
\]
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T+\tau} \alpha(t) = 0, \quad (43)
\]
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T+\tau} \beta(t) = 0, \quad (44)
\]
then the convergence of the Algorithm 1 is guaranteed.

Proof: Since $E(t) = \frac{4G_0}{K\sigma}(\alpha(t)\beta(t))$ from equation (20) and we have
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} E(t) = \frac{2G_0}{K\sigma} \frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t)\beta(t),
\]
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} E^2(t) = \frac{4G^2_0}{K^2\sigma^2} \frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t)^2\beta^2(t),
\]
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} E(t) = \frac{2G_0}{K\sigma} \frac{1}{T} \sum_{t=0}^{T+\tau} \beta(t).
\]

From Theorem 2, we know that the convergence of the Algorithm 1 is guaranteed if the conditions (52), (53) and (54) are satisfied. This fact indicates that the time delay does not affect the convergence of Algorithm 1.

Corollary 1: The equivalent forms of the conditions (42), (43) and (44) in Theorem 2 are
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \alpha(t) = 0, \quad (52)
\]
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \alpha(t) = 0, \quad (53)
\]
\[
\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \beta(t) = 0, \quad (54)
\]

Remark 5: Corollary 1 shows that the convergence of Algorithm 1 is guaranteed if the conditions without time delay (52), (53) and (54) are satisfied. This fact indicates that the time delay does not affect the convergence of Algorithm 1.

Furthermore, different stepsize $\alpha(t)$ and quantization parameter $\beta(t)$ satisfying conditions (52), (53) and (54) will affect the convergence rate of Algorithm 1. We have the following Corollary 2.

Corollary 2: When stepsize $\alpha(t)$ and quantization parameter $\beta(t)$ satisfy the condition (42), (43) and (44), the convergence rate is determined by
\[
\max \left\{ \frac{1}{T} \sum_{t=1}^{T+\tau} \beta(t), \frac{1}{T} \sum_{t=1}^{T+\tau} \alpha(t), \frac{1}{T\alpha(T + \tau)} \right\}.
\]

Proof: When $T$ is sufficiently large, we have
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t)\beta(t) \leq \frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t), \quad \frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t)^2\beta^2(t) \leq \frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t).
\]

From condition (43), we know that
\[
\lim_{T \to \infty} \alpha(t) = 0.
\]
Therefore, when $T$ is sufficiently large, we have
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t) = \frac{1}{T} \sum_{t=0}^{T+\tau} \beta(t) = 0.
\]

From Theorem 2, we know that the convergence rate of Algorithm 1 is determined by
\[
\max \left\{ \frac{1}{T} \sum_{t=1}^{T+\tau} \beta(t), \frac{1}{T} \sum_{t=1}^{T+\tau} \alpha(t), \frac{1}{T\alpha(T + \tau)} \right\}
\]

Corollary 2 analyze the convergence rate of Algorithm 1 under limited communication capacity. Next we will discuss the convergence rate of Algorithm 1 with stepsize $\alpha(t) = \frac{1}{(T+1)^{\rho}}$, quantization parameter $\beta(t) = \frac{1}{(T+1)^{\rho}}$, where $0 < \rho_1, \rho_2 < 1$.

Corollary 3: For the stepsize $\alpha(t) = \frac{1}{(T+1)^{\rho}}$, quantization parameter $\beta(t) = \frac{1}{(T+1)^{\rho}}$, where $0 < \rho_1, \rho_2 < 1$, the convergence rate of Algorithm 1 is $O(1/T^\rho)$, where $\rho = \min\{\rho_1, \rho_2, 1 - \rho_1\}$.

Proof: From Corollary 2, we need to analyze three terms $\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t)$, $\frac{1}{T} \sum_{t=0}^{T+\tau} \beta(t)$ and $\frac{1}{T\alpha(T + \tau)}$ in (55). For the term $\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t)$ and $T > \tau + 2$, we have
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t) \leq \frac{1}{T} \left( 1 + \int_{1}^{T+\tau+1} \frac{1}{1 + z^{\rho}} dx \right)
\]
\[
\leq \frac{2^{1-\rho_1} - 1}{1 - \rho_1} - \frac{1}{T - 1}. \quad (56)
\]
For the term $\frac{1}{T} \sum_{t=0}^{T+\tau} \beta(t)$ and $T > \tau + 2$, similar to (56), we have
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} \beta(t) \leq \frac{2^{1-\rho_2} - 1}{1 - \rho_2} - \frac{1}{T - 1}. \quad (57)
\]
For the term $\frac{1}{T\alpha(T + \tau)}$ and $T > \tau + 1$, we have
\[
\frac{1}{T\alpha(T + \tau)} \leq \frac{T^{1-\rho_1}}{2^{\rho_1} - 1}. \quad (58)
\]
We define that \( \rho = \min\{\rho_1, \rho_2, 1 - \rho_1\} \). Thus the convergence rate is \( O(1/T^\rho) \). The proof is completed. ■

The following table shows the convergence rate of different stepsize \( \alpha(t) \) and quantization parameter \( \beta(t) \).

| \( \alpha(t) \) | \( \beta(t) \) | \( 1/T^0.1 \) | \( 1/T^0.5 \) | \( 1/T^0.9 \) |
|----------------|----------------|--------------|--------------|--------------|
| \( 1/T^0.5 \)  | \( 1/T^0.1 \) | \( 1/T^0.25 \) | \( 1/T^0.5 \) | \( 1/T^0.25 \) |
| \( 1/T^0.75 \) | \( 1/T^0.1 \) | \( 1/T^0.25 \) | \( 1/T^0.75 \) | \( 1/T^0.25 \) |

Next we will discuss how to select appropriate stepsize \( \alpha(t) \) and quantization parameter \( \beta(t) \) to obtain the optimal convergence rate.

**Corollary 4:** The optimal convergence rate is \( O(1/\sqrt{T}) \) as we select \( \alpha(t) = \frac{1}{(t+1)^\rho} \) and \( \beta(t) = \frac{1}{(t+1)^{2\rho}} \), where \( \rho_1 = \frac{1}{2} \) and \( \frac{1}{2} \leq \rho_2 < 1 \).

**Proof:** From Corollary 3, we know that the optimal convergence rate is \( O(1/T^\rho) \) and 
\[
\rho = \min\{\rho_1, 1 - \rho_1, \rho_2\} \leq \frac{1}{2},
\]
where \( \rho_1 = \frac{1}{2} \) and then we have \( \min\{\rho_1, 1 - \rho_1, \rho_2\} = \frac{1}{2} \). The optimal convergence rate is \( O(1/\sqrt{T}) \). The proof is completed. ■

**Remark 6:** The proposed adaptive quantization method is designed from uniform quantization method by updating uniform quantizer’s mid-value and interval size at each iteration. As shown in (1), the bitsize of proposed quantizer is \( K + 1 \) in this paper and it can be adjusted to satisfy communication capacity constraints. However, as shown in (3) and (48), small bit size of uniform quantizer would cause large information loss, thus it slows down the convergence rate of algorithm. When bitsize \( K + 1 \) converges to infinity, it implies from (20) that quantization error \( E(t) = 0 \). Then, it implies \( D(T) = 0 \) and remaining \( C(T) \) is the convergence rate of Algorithm 1, where both \( C(T) \) and \( D(T) \) are shown in Theorem 1.

**Remark 7:** The static quantization over subgradient descent algorithm is proposed in [4] and that over dual averaging algorithm is developed in [20]. Compared with static quantization method in [4] and [20], the proposed adaptive quantizer helps to asymptotically alleviate the quantized error to obtain optimal value. However, due to quantization error, only suboptimal value can be obtained in [4], [20]. Moreover, the proposed adaptive quantization will not slow down the convergence rate of distributed optimization algorithm if appropriate parameters are selected according to Corollary 4.

For a special case, when the communication between each agent is perfect, i.e. \( E(t) = 0 \), we have Corollary 5.

**Corollary 5:** The different stepsize \( \alpha(t) = \frac{1}{(t+1)^{\rho}} \), with \( 0 < \rho < 1 \) will affect the convergence rate under perfect communication. The convergence rate is determined by 
\[
\max\{\frac{1}{T} \sum_{t=1}^{T+\tau} \alpha(t), \frac{1}{T_0(T+\tau)}\}.
\]
Furthermore, the optimal convergence rate is \( O(1/\sqrt{T}) \) and the corresponding stepsize is \( \alpha(t) = \frac{1}{\sqrt{T}} \).

**Proof:** When the communication between each agent is perfect, that is to say there is no quantization between communication. Then we have \( E(t) = 0 \) and substitute it into inequality (41). That is 
\[
f(\hat{z}(T)) - f(x^*) 
\leq \left\{ \frac{(2B + 1)G^2}{2\sigma_\phi} + \frac{G^2}{2\sigma_\phi} + \frac{2BG^2}{N\sigma_\phi} \right\} \frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t) 
+ \left( \frac{2N\sigma_\phi\gamma_\omega}{1 - \gamma^2} + \frac{2\sigma_\phi\gamma_\omega}{1 - \gamma} \right) \frac{1}{T} + \frac{D_\phi}{T\alpha(T+\tau)}.
\]
When \( T \) is large enough, we have 
\[
\frac{1}{T} \leq \alpha(T+\tau) T.
\]
Therefore, the convergence rate is determined by 
\[
\max\{\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t), \frac{1}{T_0(T+\tau)}\}.
\]
From inequalities (56) and (58), we know that 
\[
\frac{1}{T} \sum_{t=0}^{T+\tau} \alpha(t) \leq \frac{2\rho\alpha}{1 - \rho} \frac{1}{T_0} \leq \frac{1}{T_0}(1 - \rho_1).
\]
Therefore, the convergence rate is \( O(1/T^\rho) \), where \( \rho = \min\{\rho_1, 1 - \rho_1, \rho_2\} \). We know that 
\[
\min\{\rho_1, 1 - \rho_1, \rho_2\} = \frac{1}{2}.
\]
When \( \rho_1 = \frac{1}{2} \), we have \( \min\{\rho_1, 1 - \rho_1\} = \frac{1}{2} \). The optimal convergence rate is \( O(1/\sqrt{T}) \). The proof is completed. ■

**Remark 8:** The distributed subgradient method with adaptive quantization (DSAQ) is proposed in [33] to solve distributed optimization problems over a static communication network with timedly subgradient information. Compared with [33], the convergence rate in this paper is \( O(1/\sqrt{T}) \) while that is \( O(\ln T/\sqrt{T}) \) in [33]. Moreover, a class of time-varying communication network and delayed subgradient information are considered in this paper, which are more realistic.

**Remark 9:** Our previous works propose a distributed zeroth-order mirror descent algorithm for constrained optimization over time-varying network in [11], [22], [23] and distributed stochastic mirror descent method for strongly convex objective functions in [10]. A novel online distributed mirror descent method has been proposed for composite objective functions in [9]. In this paper, we largely improve our previous works [9], [11], [10] on distributed mirror descent methods in several aspects. In the literature of distributed mirror descent, this paper is the first work to propose the adaptive quantization method to address limited communication channel and simultaneously take delayed subgradient information into consideration in distributed mirror descent type algorithms. Moreover, the optimal convergence rate \( O(1/\sqrt{T}) \) is derived under appropriate conditions.

| VI. SIMULATIONS |

In this section, we consider the following distributed estimation problem [2]: 
\[
\min_{x \in \mathbb{R}} \frac{1}{N} \sum_{j=1}^{N} a_j \|x - b_j\|^2, \tag{59}
\]
over a sequence of time-varying sensor network, where \( a_j \in \mathbb{R} \) and \( b_j \in \mathbb{R}^n \). The size of the network is \( N = 30 \) and the dimension of the variable \( x \) is \( n = 10 \). The sequence of time-varying network satisfies \( B \) connectivity.

The domain of this system \( \mathcal{X} \) is bounded and closed with 
\[
\mathcal{X} = \{ x \in \mathbb{R}^n | 0 \leq x \leq 100, j = 1, 2, \ldots, n \}. 
\]
We will use mirror descent algorithm under delayed subgradient information $\tau$, where $\tau = 0$ means that subgradient information is timely. We choose the distance generating function $\phi_1(x) = \sum_{i=1}^n e^{x_i}$, and the corresponding Bregman divergence $V_1(x,y) = \sum_{i=1}^n (e^{x_i} - (x_i - y_i + 1)e^{y_i})$. Note that $\phi_1(x)$ satisfies Lipschitz continuity on compact set and $V_1(x,y)$ satisfies separate convexity. We use the relative error $e_1(T)$ of agent $i$ as $e_1(T) = \frac{|\hat{x}_i(T)-f(x^*)|}{f(x^*)}$ to show $f(\hat{x}_i(T))$ converges to $f(x^*)$ intuitively.

### A. Quantization Effect

In this subsection, we will show the quantization’s effect to the convergence of Algorithm 1. We consider Algorithm 1 with stepsize $\alpha(t) = \frac{1}{\sqrt{T+1}}$ with adaptive quantizer ($\beta(t) = \frac{1}{\sqrt{T+1}}$) and without quantizer under different time delay $\tau$. In the Fig. 1, we choose nodes 1 and 2 arbitrarily among the network for illustration. The line means the relative error between arbitrarily selected node $f(\hat{x}_i(T))$ and optimal value $f(x^*)$.

![Fig. 1. Different Parameters Comparisons](image)

From Fig. 1, we can find that the relative error of each agent keeps on decreasing. However, we can also find that the convergence rate under quantizer (red and yellow lines) is much slower than that under perfect communication channel (blue and green lines). The Fig. 1 reflects quantization will slow down the convergence rate.

### B. Stepsize Effect and Quantization Effect

In this subsection, we will show the convergence rate of Algorithm 1 under different stepsizes and different quantization parameters.

![Fig. 2. Different Parameters Comparisons](image)

In Fig. 2(a), quantization parameter is $\beta(t) = \frac{1}{\sqrt{T+1}}$ and stepsize is $\alpha(t) = \frac{1}{\sqrt{T+1}}$, where $\rho_1 = 0.38, 0.42, 0.46, 0.5$. The correspond convergence rate is $O(1/T^{1-\rho_1})$ under different $\rho_1$. In Fig. 2(b), stepsize is $\alpha(t) = \frac{1}{\sqrt{T+1}}$ and quantization parameter is $\beta(t) = \frac{1}{(T+1)^{\rho_2}}$, where $\rho_2 = 0.4, 0.43, 0.46, 0.5$. The correspond convergence rate is $O(1/T^{\rho_2})$ under different $\rho_2$. Thus, the results in Corollary 3 is verified in this numerical experiment.

### C. Bitsize Effect

In this subsection, we compare the convergence rate of algorithm under different bitizes.

![Fig. 3. Different Bitsizes Comparisons](image)

From Fig. 3, higher bitsize has faster convergence rate as discussed in Remark 6.

### D. Comparisons with Static Quantizer

In this subsection, we compare the proposed adaptive quantization method with static quantizer in [20].

![Fig. 4. Static Quantizer and Adaptive Quantizer](image)

From yellow and red lines in Fig. 4, we can know that only suboptimal value can be obtained under mirror descent algorithm with static quantizer. After about 10000 iterations, the relative errors of nodes 1 and 2 are no longer decreasing. From blue and black lines in Fig. 4, the relative errors of nodes 1 and 2 keep on decreasing. This numerical example demonstrates the results in Remark 7.

### E. Comparisons with Distributed Subgradient Method

In this subsection, we will compare the convergence performance of Algorithm 1 with DSAQ in [33] (see Remark 8). We will use Algorithm 1 and DSAQ to solve problem (59). The stepsize of Algorithm 1 and DSAQ is $\alpha(t) = \frac{1}{\sqrt{T+1}}$ and the quantization parameter of Algorithm 1 is $\beta(t) = \frac{1}{\sqrt{T+1}}$.

![Fig. 5. Algorithm 1 and DSAQ](image)
From Fig. 5, we know that the relative error of each agent keeps on decreasing in Algorithm 1 and DSAQ. Moreover, the convergence performance of Algorithm 1 is better than that of DSAQ, which demonstrates the conclusions in Remark 8.

VII. CONCLUSION

This paper has studied the mirror descent algorithm with delayed subgradient information under adaptive quantization. We design adaptive quantization method to solve the limited communication channel and analyze the convergence of Algorithm 1 with different stepizes and quantization parameters. The optimal convergence rate $O(n^{-1/2})$ can be obtained under appropriate conditions. This paper has improved many previous works (e.g. [6], [9], [11], [10], [33]) in different aspects. Some numerical examples have been presented to demonstrate the effectiveness of the algorithm and verify the theoretical results. For the future work, we can consider how to use adaptive method in other settings, such as dual averaging algorithm [5], online composite optimization [9]. Furthermore, we shall investigate distributed optimization over event triggering schemes to relieve the burden of network bandwidth occupation [37], [38]. We shall also consider the possibility to extend the obtained results to communication time delay [39], and input time delay setting [40], [41], [42].

VIII. APPENDIX

A. Proof of Lemma 6

Before proving Lemma 6, we need to extend some Lemmas in the text of the paper.

Lemma 9: $\{\tilde{y}_i(t)\}$ and $\{y_i(t - \tau)\}$ are sequences generated by Algorithm 1. Then we have

$$\begin{align*}
\langle g_i(t - \tau), \tilde{y}_i(t) - x^* \rangle & \leq \frac{1}{\alpha(t)} \langle V_{\phi}(x^*, \tilde{y}_i(t)) - \frac{1}{\alpha(t)} V_{\phi}(x^*, x_i(t + 1)) + \frac{G^2 \alpha(t)}{2} \rangle.
\end{align*}$$

Proof: For the first order optimality condition, for $\forall x \in X$ we have $\langle \alpha(t) \tilde{y}_i(t - \tau), x - x_i(t + 1) \rangle \geq 0$. Thus we select $x = x^*$ and we have

$$\langle \alpha(t) g_i(t - \tau), x - x_i(t + 1) \rangle \geq 0.$$

From Bregman divergence’s definition and strong convexity of $\phi$, rearrange terms in (60) and we have

$$\begin{align*}
\langle \alpha(t) g_i(t - \tau), x_i(t + 1) - x^* \rangle & \leq V_{\phi}(x^*, \tilde{y}_i(t)) - V_{\phi}(x^*, x_i(t + 1)) - \frac{\sigma_\phi}{2} \langle x_i(t + 1) - \tilde{y}_i(t) \rangle^2.
\end{align*}$$

(61)

From Cauchy inequality, we have

$$\begin{align*}
\langle \alpha(t) g_i(t - \tau), x_i(t + 1) - x^* \rangle & \geq -\frac{\alpha^2(t)}{2 \sigma_\phi} \|g_i(t - \tau)\|^2 - \frac{\sigma_\phi}{2} \langle x_i(t + 1) - \tilde{y}_i(t) \rangle^2.
\end{align*}$$

(62)

Combining inequality (61) and (62), we have $\langle \alpha(t) g_i(t - \tau), \tilde{y}_i(t) - x^* \rangle \leq \frac{G^2 \alpha(t)}{2} + \frac{1}{\alpha(t)} (V_{\phi}(x^*, \tilde{y}_i(t)) - V_{\phi}(x^*, x_i(t + 1)))$. The proof of Lemma 8 is completed.

Lemma 10: $\{x_i(t)\}$ is generated by Algorithm 1. $e_i(t)$ and $p_i(t)$ are defined as equations (15) and (16). Then we have

$$\begin{align*}
V_{\phi}(x^*, x_j(t) + e_j(t) + p_i(t)) & \leq V_{\phi}(x^*, x_j(t)) + 2(4N^2 + 1) L_{\phi} E^2(t) + (2N + 1) L_{\phi} \sqrt{\frac{2D \phi}{\sigma_\phi}} E(t).
\end{align*}$$

(63)

Proof: Follow from mean value formula, there exists a $0 \leq \xi \leq 1$ such that $\phi(x_j(t) + e_j(t) + p_i(t)) = \phi(x_j(t)) + \langle \nabla \phi(x_j(t) + \xi e_j(t) + p_i(t)), e_j(t) + p_i(t) \rangle$. Before we use Lipschitz continuous of $\phi$ from Assumption 5, we need to show that $x_j(t) + e_j(t) + p_i(t) \in X$. From definition of quantization error, we know that $Q_j(z_j(t), d(t), x_j(t)) = x_j(t) + e_j(t)$.

Moreover, from definition of quantization function $Q_j$ (2), we know that

$$\begin{align*}
z_j(t) - d(t) \leq Q_j(z_j(t), d(t), x_j(t)) \leq z_j(t) + d(t).
\end{align*}$$

From (11), (20), (22) and $0 < \alpha(t), \beta(t) < 1$, we know that

$$\begin{align*}
\|d(t)\| & \leq \frac{G_0(t) \beta(t)}{\alpha_0(t)} \leq \frac{G\alpha(t)}{\sigma_\phi}.
\end{align*}$$

(64)

From the first term (64), from separate convexity of $V_{\phi}(x, y)$ and Lemma 10, we have

$$\begin{align*}
&\sum_{t=1}^{T+\tau} \sum_{i=1}^{N} \left\{ V_{\phi}(x, y_i(t)) - V_{\phi}(x, x_i(t + 1)) \right\} \\
&\leq \frac{1}{\alpha(t)} \sum_{t=1}^{T+\tau} \sum_{i=1}^{N} \left\{ N V_{\phi}(x^*, y_i(t)) - N V_{\phi}(x^*, x_i(t + 1)) \right\} \\
&+ \frac{N G^2}{2 \sigma_\phi} \sum_{t=1}^{T+\tau} \sum_{i=1}^{N} \langle \nabla \phi(x^*, x_i(t + 1)) \rangle.
\end{align*}$$

(65)

For the first term in (65), it is upper bounded by $\sum_{j=1}^{N} D_{\phi}(\frac{1}{\alpha(t) + 1}) + \sum_{t=1+\tau}^{T+\tau} (\frac{1}{\alpha(t) - 1})$, which is equal to $\frac{\sum_{j=1}^{N} D_{\phi}(\frac{1}{\alpha(t) + 1}) + \sum_{t=1+\tau}^{T+\tau} (\frac{1}{\alpha(t) - 1})}{\alpha(t)}$. Then substitute it into (65) and then substitute
inequality \((65)\) into \((64)\), we have
\[
\sum_{t=1}^{T+\tau} \sum_{i=1}^{N} (g_i(t-r), \tilde{y}_i(t) - x^*) 
\leq (2N^2 + N)L\phi \sqrt{\frac{2D_D}{\phi}} \sum_{t=1+\tau}^{T+\tau} \frac{E(t)}{\alpha(t)} + \frac{NG^2}{2\phi} \sum_{t=1+\tau}^{T+\tau} \alpha(t) 
+ \frac{ND\phi}{\alpha(T+\tau)} + (8N^3 + 2N)L\phi \sum_{t=1+\tau}^{T+\tau} \frac{E^2(t)}{\alpha(t)}.
\]
Therefore, we have \(\frac{1}{T+\tau} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} (g_i(t-r), \tilde{y}_i(t) - x^*) \leq B_1\). The proof is completed. 

B. Proof of Lemma 7

Proof: From Cauchy inequality and the upper bound of subgradient, we have
\[
(g_i(t-r), \tilde{y}_i(t-r) - \tilde{y}_i(t)) \leq ||g_i(t-r)|| \cdot ||\tilde{y}_i(t-r) - \tilde{y}_i(t)|| \leq G||g_i(t-r) - \tilde{y}_i(t)||. \tag{66}
\]
After summing up inequality \((66)\) from \(t = r+1\) to \(T+\tau\) and dividing both side by \(N_T\), we have
\[
\frac{1}{N_T} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} (g_i(t-r), \tilde{y}_i(t-r) - \tilde{y}_i(t)) 
\leq \frac{G}{N_T} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||\tilde{y}_i(t-r) - \tilde{y}_i(t)||. \tag{67}
\]
For \(||\tilde{y}_i(t-r) - \tilde{y}_i(t)||\), from triangle inequality, we have
\[
||\tilde{y}_i(t-r) - \tilde{y}_i(t)|| \leq \sum_{k=t-r}^{t} ||\tilde{y}_i(k) - \tilde{y}_i(k+1)||. \tag{68}
\]
For \(||\tilde{y}_i(k) - \tilde{y}_i(k+1)||\), from triangle inequality, we have \(||\tilde{y}_i(k) - \tilde{y}_i(k+1)|| \leq \sum_{k=t-r}^{t} ||x_j(k+1) - x_j(k)||\). \tag{69}

The sequence \(\{\alpha(t)\}\) and \(\{E(t)\}\) are non-increasing and substitute inequality \((69)\) into \((68)\), we have
\[
\frac{||\tilde{y}_i(t-r) - \tilde{y}_i(t)||}{\alpha(t)} \leq \frac{G}{\alpha(t)} r + 3NrE(t-r) 
+ \sum_{k=t-r}^{t-1} \sum_{j=1}^{N} ||x_j(k+1) - x_j(k)||. \tag{70}
\]
Substitute inequality \((70)\) into \((67)\) and we have
\[
G \frac{T+\tau}{N_T} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||\tilde{y}_i(t-r) - \tilde{y}_i(t)|| 
= \frac{G}{N_T} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} \sum_{k=t-r}^{t-1} \sum_{j=1}^{N} ||x_j(k+1) - x_j(k)|| 
+ \frac{G}{\alpha(T+\tau)} ||\alpha(t-r) + 3NrG \sum_{t=1+\tau}^{T+\tau} E(t-r). \tag{71}\]
For the first term of \((71)\), we have
\[
\frac{G}{N_T} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} \sum_{k=t-r}^{t-1} \sum_{j=1}^{N} ||x_j(k+1) - x_j(k)|| 
\leq \frac{G}{\alpha(T+\tau)} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||x_j(t) - x_j(t)||. \tag{72}\]
From Lemma 8, we have
\[
\sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||x_j(t) - x_j(t)|| \leq \frac{2N\omega}{1-\gamma} A + 2B \sum_{t=0}^{T} \frac{Go(t)}{\alpha} + 3NE(t) \tag{73}\]
Thus we have
\[
\sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||x_j(t) - x_j(t)|| 
\leq 2N^2\omega \frac{A}{1-\gamma} + 2BNG\frac{Go(t)}{\alpha} + 3NE(t). \tag{74}\]
Substitute inequality \((73)\) into \((72)\) and we have
\[
\sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||x_j(t) - x_j(t)|| 
\leq 2NG\frac{Go(t)}{\alpha} + 3NE(t). \tag{75}\]
Substitute inequality \((74)\) into \((71)\) and substitute \((71)\) into \((67)\), then we have \(\frac{1}{T+\tau} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} ||x_j(t-r), \tilde{y}_i(t-r) - \tilde{y}_i(t)|| \leq \frac{1}{T+\tau} \sum_{t=1+\tau}^{T+\tau} \sum_{i=1}^{N} \sum_{k=t-r}^{t-1} ||x_j(k+1) - x_j(k)|| \leq B_2\). The proof is completed. 

C. Proof of Lemma 8

Proof: We define that \(A = \sum_{j=1}^{N} ||x_j(0)||\), \(B = (2N + \frac{N^2\omega}{1-\gamma})\) and we will show the bound of \(||x_j(t) - x\||\).
\[
||x_j(t) - x\|| = ||\sum_{j=1}^{N} ||P(t)||_{ij}(x_j(t) - P(t))|| \leq 3NE(t). \tag{76}\]
Therefore, from Lemma 1, Lemma 5 and \((75)\), we have
\[
||x_j(t) - x\|| \leq \omega^{t-1} A + \sum_{s=1}^{t-1} \omega^{t-s-1} \sum_{j=1}^{N} ||c_j(s-1) + \epsilon_j(s-1)|| 
\leq \omega^{t-1} A + 2|\frac{Go(t-1)}{\alpha} + 3NE(t-1)| \tag{77}\]
Then we have
\[
||x_j(t) - x\|| \leq 2\omega^{t-1} A + 4\{|\frac{Go(t-1)}{\alpha} + 3NE(t-1)| \}
+ 2N \omega^{t-s-1} \left(\frac{Go(s-1)}{\alpha} + 3NE(s-1)\right). \tag{78}\]
Therefore, we have
\[
\sum_{j=1}^{N} ||x_j(t) - x\|| \leq \frac{2N\omega}{1-\gamma} A + 2B \sum_{t=0}^{T} \frac{Go(t)}{\alpha} + 3NE(t) \tag{79}\]
\[
\sum_{j=1}^{N} ||x_j(t) - x\|| \leq \frac{2N\omega}{1-\gamma} A + 2B \sum_{t=0}^{T} \frac{Go(t)}{\alpha} + 3NE(t). \tag{80}\]
The proof is completed.

REFERENCES

[1] A. Nedic and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization”, IEEE Transactions on Automatic Control, vol. 54, no. 1, pp. 48-61, 2009.
[2] A. Nedic and A. Olshesky, “Stochastic gradient-push for strongly convex functions on time varying directed graphs”, IEEE Transactions on Automatic Control, vol. 61, no. 12, pp. 3936-3947, 2016.
[3] A. Nedic and S. Lee, “On stochastic subgradient mirror-descent algorithm with weighted averaging”, SIAM Journal on Optimization, vol. 24, no. 1, pp. 84-107, 2014.
[4] A. Nedic, A. Olshesky, A. Ozdaglar and J. N. Tsitsiklis, “Distributed subgradient methods and quantization effects”, Proceedings of the 47th IEEE Conference on Decision and Control, pp. 4177-4184, 2008.
[5] I. C. Duchi, A. Alekh and J. W. Martin, “Dual averaging for distributed optimization: convergence analysis and network scaling”, IEEE Transactions on Automatic Control, vol. 57, no. 3, pp. 592-606, 2011.
[6] J. Li, G. Chen, Z. Dong and Z. Wu, “Distributed mirror descent method for multi-agent optimization with delay”, Neurocomputing, vol. 177, pp. 643-650, 2016.
[7] P. Yi, Y. Hong and F. Liu, “Distributed gradient algorithm for constrained optimization with application to load sharing in power systems”, Systems and Control Letters, vol. 83, pp. 45-52, 2015.
[8] J. Liu, S. Yang and J. Wang, “A collective neurodynamic approach to distributed constrained optimization”, IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 8, pp. 1747-1758, 2017.
[9] D. Yuan, Y. Hong, D. W. C. Ho and S. Xu, “Distributed mirror descent for online composite optimization”, IEEE Transactions on Automatic Control, vol. 66, no. 2, pp. 714-729, 2021.
[10] D. Yuan, Y. Hong, D. W. C. Ho and G. Jiang, “Optimal distributed stochastic mirror descent for strongly convex optimization”, Automatica, vol. 90, pp. 196-203, 2018.
[11] Z. Yu, D. W. C. Ho and D. Yuan, “Distributed randomized gradient-free mirror descent algorithm for constrained optimization”, IEEE Transactions on Automatic Control, vol. 67, no. 2, pp. 957-964, 2022.
[12] C. Huang, Daniel W. C. Ho and J. Lu, “Partial-information-based distributed filtering in two-targets tracking sensor networks”, IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, no. 4, pp. 820-832, 2012.
[13] Z. Deng, “Distributed algorithm design for aggregative games of Euler-Lagrange systems and its application to smart grids”, IEEE Transactions on Cybernetics, DOI: 10.1109/TCYB.2021.3049462, 2021.
[14] Z. Deng and X. Nian, “Distributed generalized nash equilibrium seeking algorithm design for aggregative games over weight-balanced digraphs”, IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 3, pp. 695-706, 2018.
[15] P. Yi, Y. Hong and F. Liu, “Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and its application to economic dispatch of power systems”, Automatica, vol. 74, pp. 259-269, 2016.
[16] B. Chen, G. Hu, D. W. C. Ho and L. Lu, “A new approach to linear/nonlinear distributed fusion estimation problem”, IEEE Transactions on Automatic Control, vol. 64, no. 3, pp. 1301-1308, 2019.
[17] J. Song, D. W. C. Ho and Y. Niu, “Model-based event-triggered sliding-mode control for multi-input systems: Performance analysis and optimization”, IEEE Transactions on Cybernetics, DOI: 10.1109/TCYB.2020.3020253, 2020.
[18] V. Hovhannisyan, P. Papras and S. Zafeiriou, “MAGMA: Multilevel accelerated gradient mirror descent algorithm for large-scale convex composite minimization”, SIAM Journal on Imaging Sciences, vol. 9, no. 4, pp. 1829-1857.
[19] A. Afkanpour, A. Gygörgy, C. Szepesvari and M. Bowling, “A randomized mirror descent algorithm for large scale multiple kernel learning”, Proceedings of the 30th International Conference on Machine Learning, vol. 28, no. 1, pp. 374-382, 2013.
[20] D. Yuan, S. Xu, H. Zhao and L. Rong, “Distributed dual averaging method for multi-agent optimization with quantized communication”, Systems and Control Letters, vol. 61, no. 11, pp. 1053-1061, 2012.
[21] D. Yuan, D. W. C. Ho and S. Xu, “Stochastic strongly convex optimization via distributed epoch stochastic gradient algorithm”, IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 6, pp. 2344-2357, 2021.
[22] D. Yuan, D. W. C. Ho and S. Xu, “Zeroth-order method for distributed optimization with approximate projections”, IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 2, pp. 284-294, 2016.
[23] D. Yuan and D. W. C. Ho, “Randomized gradient-free method for multi-agent optimization over time-varying networks”, IEEE Transactions on Neural Networks and Learning Systems, vol. 26, no. 6, pp. 1342 -1347, 2015.
[24] A. Agarwal and J. C. Duchi, “Distributed delayed stochastic optimization”, Advances in Neural Information Processing Systems, 2011.
[25] J. Langford, A. Wikipedia, “Multiple similar methods for multi-agent optimization”, In Advances in Neural Information Processing Systems vol. 22, pp. 2331-2339, 2009.
[26] A. Nedic, D.P. Bertsekas and V.S. Borkar. “Distributed asynchronous stochastic gradient algorithms”, Inherently Parallel Algorithms in Data Science, Matrix Computations, and Optimization, SIAM, 2010.
[27] H. Wang, X. Liao, T. Huang and C. Li, “Cooperative distributed optimization in multiagent networks with delays”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 45, no. 2, pp. 363-369, 2015.
[28] M. Stark, L. Wang, G. Bauch and R. D. Wesel. “Decoding rate-compatible 5G-LDPC codes with coarse quantization using the information bottleneck method”, IEEE Open Journal of the Communications Society, vol. 1, pp. 646-660, 2020.
[29] L. Li, D. W. C. Ho and J. Lu, “A unified approach to practical consensus with quantized data and time delay”, IEEE Transactions on Circuits and Systems-I: Regular Papers, vol. 60, no. 10, pp. 2668-2678, 2013.
[30] J. Song, D. W. C. Ho and Y. Niu, “Event-based network consensus with retraining using outlier channel splitting”, Proceedings of the 36th International Conference on Machine Learning, pp. 7543 - 7552, 2019.
[31] T. Monses, D. Wübben and A. Dekorsy, “Information preserving quantization and decoding for satellite-aided 5G communications”, IEEE 2nd 5G World Forum, pp. 516-519, 2019.
[32] T. T. Doan, S. T. Maguluri and J. Romberg. “Fast convergence rates of distributed subgradient methods with adaptive quantization”, IEEE Transactions on Automatic Control, vol. 66, no. 5, pp. 2191-2205, 2021.
[33] 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, 2017.
[34] P. Yi and Y. Hong, “Quantized subgradient algorithm and date-rate analysis for distributed optimization”, IEEE Transactions on Control of Network Systems, vol. 1, no. 4, pp. 380-392, 2014.
[35] S. Ghadimi, G. Lan and H. Zhang, “Mini-batch stochastic approximation methods for nonconvex stochastic composite optimization”, Mathematical Programming, vol. 155, pp. 267-305, 2016.
[36] E. Tian and D. Yue, “Decentralized control of network-based interconnected systems: A state-dependent triggering method”, International Journal of Robust and Nonlinear Control, vol. 25, no. 8, pp. 1126-1144, 2015.
[37] C. Peng, D. Yue and M. Fei, “A higher energy-efficient sampling scheme for networked control systems over IEEE 802.15.4 wireless networks”, IEEE Transactions on Industrial Informatics, vol. 12, no. 5, pp. 1766-1774, 2016.
[38] L. Li, D. W. C. Ho and J. Lu, “Event-based network consensus with communication delays”, Nonlinear Dynamics, vol. 87, pp. 1847-1858, 2017.
[39] H. Chu, D. Yue, C. Dou and L. Chu, “Consensus of multiagent systems with time-varying input delay and relative state saturation constraints”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 11, pp. 6938-6944, 2021.
[40] H. Chu, D. Yue, C. Xie and L. Chu, “Consensus of multiagent systems with time-varying input delay via truncated predictor feedback”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 10, pp. 6062-6073, 2021.
[41] H. Chu, L. Gao, D. Yue and C. Dou, “Consensus of Lipschitz nonlinear multiagent systems with input delay via observer-based truncated prediction feedback”, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 5, no. 10, pp. 3784-3794, 2020.
Zhan Yu received the Ph.D. degree in applied mathematics from the City University of Hong Kong, Hong Kong, in 2021. He is currently a Postdoctoral Fellow with the School of Data Science, City University of Hong Kong. His current research interests include learning theory, optimization theory, and applied harmonic analysis.

Daniel W. C. Ho (M’89-SM’05-F’17) received the B.S., M.S., and Ph.D. degrees in mathematics from the University of Salford, Greater Manchester, U.K., in 1980, 1982, and 1986, respectively.

From 1985 to 1988, he was a Research Fellow with the Industrial Control Unit, University of Strathclyde, Glasgow, U.K. In 1989, he joined the City University of Hong Kong, Hong Kong, where he is currently a Chair Professor of applied mathematics, and an Associate Dean with the College of Science. He has over 250 publications in scientific journals. His current research interests include control and estimation theory, complex dynamical distributed networks, multi-agent systems, and stochastic systems.

Prof. Ho is a Fellow of the IEEE. He was a recipient of the Chang Jiang Chair Professor Awarded by the Ministry of Education, China, in 2012 and the ISI Highly Cited Researchers Award in Engineering by Clarivate Analytics, from 2014 to 2021. He has been on the Editorial Board of a number of journals including the IEEE Transactions on Neural Networks and Learning Systems, IET Control Theory and Its Applications, the Journal of the Franklin Institute, and the Asian Journal of Control.