IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 21 November 2022, accepted 25 December 2022, date of publication 29 December 2022, date of current version 3 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233233

THEORY

Gradient-Push Algorithm for Distributed Optimization With Event-Triggered Communications

JIMYEONG KIM AND WOOCHEOL CHOI
Department of Mathematics, Sungkyunkwan University, Suwon 16419, South Korea

Corresponding author: Woocheol Choi (choiwc@skku.edu)

This work was supported by the National Research Foundation of Korea (NRF) Grants through the Korea Government under Grant NRF-2016R1A5A1008055 and Grant NRF-2021R1F1A1059671.

ABSTRACT Decentralized optimization problems consist of multiple agents connected by a network. The agents have each local cost function, and the goal is to minimize the sum of the functions cooperatively. It requires the agents to communicate with each other, and reducing the cost for communication is desired for a communication-limited environment. Recently, the decentralized gradient descent algorithms involving event-triggered communication have been proposed when the network of the agents is an undirected graph. On the other hand, the network of agents is often directed graph for realistic scenarios whose communication resources are limited. In this work, we first propose a gradient-push algorithm involving event-triggered communication on a directed network. Each agent sends its current states to its neighbors only when the differences between the latest sent states and the current states are larger than thresholds. The convergence of the algorithm is established under suitable decays and summability conditions on a stepsize and triggering thresholds. Numerical experiments are presented to support the effectiveness and the convergence results of the algorithm. More precisely, the numerical results reveals that the proposed algorithm may reduce the communication cost significantly compared to the gradient-push algorithm not involving the event-triggered communication.

INDEX TERMS Decentralized optimization, directed graph, event-triggered communication, gradient push algorithm.

I. INTRODUCTION

In recent years, distributed optimization techniques over a multi-agent network have attracted considerable attention since they play an essential role in engineering problems in distributed control [1], [2], signal processing [3], [4], and machine learning problems [5], [6], [7]. In distributed optimization, multiple agents have their own local cost function and try to find a minimizer of the sum of those local cost functions in a collaborative way that each agent only uses the information from its neighboring agents where the neighborhood structure is depicted as a graph, often undirected or directed.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaojie Su.

There has been a significant interest in the consensus-based distributed gradient method. One fundamental work is [9], which developed the distributed gradient descent on an undirected graph. This algorithm consists of a local gradient step and consensus step based on communication between neighboring agents. The convergence property of the algorithm has been studied in the works [8], [9], [10], [11]. There are also various distributed algorithms containing the distributed dual averaging method [12], consensus-based dual decomposition [13], [14], and the alternating direction method of multipliers (ADMM) based algorithms [15], [16]. These algorithms work with a doubly-stochastic matrix associated with the undirected graph.

The gradient-push algorithm was introduced in [17] to solve the distributed optimization for a directed graph, which utilizes push-sum algorithms [18], [19]. The communication of this algorithm is represented by a column stochastic matrix, which requires each agent to know its out-degree at each time, without having the information of the number of agents. This algorithm has influenced a significant impact on later works. The work [20] studied the algorithm with gradient having a noise. The time-varying distributed optimization was also considered [21] using the gradient-push algorithm. Recently, stochastic gradient-push algorithm was designed for large scale deep learning problem [22]. This work was also extended in [23] further to quantized communication settings. We also refer to [24] and [25] where the authors studied the asynchronous version of the gradient-push algorithm. An important issue for the gradient-push algorithm in practical applications is the resistance to resilient attacks in the network. Recently, the work [26] applied the push-sum method to the distributed estimation problem under sensor attacks. The work [27] designed a decentralized robust subgradient push algorithm for detection and isolation of malicious nodes in the network for optimization.

Regardless of the types of graphs, these distributed algorithms require each agent to communicate with their neighbors at every iteration, which leads to overhead in restricted environments. Power consumption by communication may become more significant than that by computation of control inputs or optimization algorithms [28]. Recently, the event-triggering approach has appeared as a promising paradigm to reduce the communication load in distributed systems. In the distributed detection problem over sensor network [29], [30], each sensor censors its local data and sends the updated data to the fusion center only when the data is informative. For distributed control problems, agents send their coordinate information only when a triggering condition is satisfied [31], [32].

For the distributed optimization problems, recent works [33], [34], [35], [36], [37] developed distributed optimization algorithms with event-triggered communication to overcome the communication overhead of distributed systems. Lu and Li [34] designed the distributed gradient descent with event-triggered communication for the distributed optimization on the whole space, and it was further studied in Li and Mu [38] to establish a convergence rate. For the distributed optimization on a bounded domain, Kajiyama et al. [33] designed the projected distributed gradient descent with event-triggered communication. Liu et al. [39] extended the work to the case with constant step-size. Cao and Basar [40] studied the online distributed problem using the distributed event-triggered gradient method. Xiong et al. [41] considered the distribted strochastic mirror descent with event-triggered communication. The distributed estimation problem was studied by He et al. [42] utilizing the event-triggered communication. In these algorithms, each agent sends its state only when the difference between the current state and the latest sent state is larger than a threshold, therefore reducing possible unnecessary network utilization.

The consensus-based distributed optimization algorithms with event-triggering communication mentioned above have been proposed for the undirected graph. However these methods cannot be applied to the situation where the network of the agents is a directed graph. We remark that when the agents have different ranges of communication, the network of agents is usually given as a directed graph. A realistic example is the multi-robot localization problem whose communication resource is limited.

In this work, we are interested in developing a distributed optimization on a directed graph involving the even-triggered communication. Precisely we propose the gradient-push algorithm incorporating the event-triggered communication. In the proposed algorithm, each agent sends its current states only when the difference between the latest sent states and the current states is larger than a triggering threshold. We prove that the algorithm solves the distributed optimization under suitable decays and summability conditions on the stepsize and the triggering thresholds. The numerical experiments are given for the proposed algorithm, supporting the theoretical results.

The proposed algorithm can be seen as a perturbed version of the gradient-push algorithm [17]. We remark that each agent j in the gradient-push algorithm communicates two variables $x_j(t)$ and $y_j(t)$, and the proposed algorithm considers the event-triggered communications both for $x_j(t)$ and $y_j(t)$. Therefore, extending the convergence result of [17] to the proposed algorithm is non-trivial. For the convergence analysis, we carefully investigate the impact of the perturbations due to the event-triggered communications, and successfully obtain the convergence results of the proposed algorithm. The numerical result shows that the proposed algorithm may reduce the communication cost significantly compared to the gradient-push algorithm.

The rest of the paper is organized as follows. In Section 2, we state the problem and introduce the algorithm with its convergence results. Section 3 is devoted to providing a consensus estimate, which is essentially used in Section 4 to prove the convergence results. In section 5, we present the numerical results of the proposed algorithm.

Before ending this section, we state several notations used in this paper. For a matrix $A \in \mathbb{R}^{n\times m}$, a_{ij} or $[A]_{ij}$ denotes the (i,j)th entry of A. For a vector $x \in \mathbb{R}^d$, $\|x\| = \sqrt{x^T x}$ denotes the standard Euclidean norm. In addition, for $X \in \mathbb{R}^{m\times d}$ given by $X = [x_1; x_2; \cdots; x_m]^T$ with row vector $x_k \in \mathbb{R}^d$, we define the mixed norm $\|X\|_1$ by $\|X\|_1 = \sum_{k=1}^m \|x_k\|$ and the maximum norm $\|X\|_\infty = \max_{1\le k\le m} \|x_k\|$. Also we use $\bar{x}$ to denote $\bar{x} = \frac{1}{m}\sum_{k=1}^m x_k$. For a function $f : \mathbb{R}^d \to \mathbb{R}$, we denote by $\nabla f(x)$ the gradient $(\partial_1 f(x), \cdots, \partial_d f(x)) \in \mathbb{R}^d$.

II. PROBLEM, ALGORITHM, AND MAIN RESULTS

A. PROBLEM STATEMENT

We consider the distributed optimization problem, which consists of m agents connected by a network that collaboratively minimize a global cost function given by the sum of local private cost functions. Formally, the problem is described by

$$\min_{x\in\mathbb{R}^d} f(x) = \sum_{i=1}^{m} f_i(x), \tag{1}$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ is a local convex cost function only known to agent $i \in \mathcal{V} = \{1, 2, \cdots m\}$. We let f^* be the optimal value of problem (1) and denote by X^* the set of optimal solutions, i.e.,

$$X^* = \{x \in \mathbb{R}^d \ : \ f(x) = f^*\},$$

which is assumed to be nonempty. We make the following standard assumption on the local cost functions.

Assumption II.1. *For each $i \in \{1, \cdots, m\}$, there exists $D_i > 0$ such that*

$$\|\nabla f_i(x)\| \le D_i \quad \forall x \in \mathbb{R}^d. \tag{2}$$

We set $D = \max_{1\le i\le m} D_i$.

This assumption is commonly used in the literatures [17], [20], [23], [43] for the convergence analysis of gradient push type algorithms. Removing this assumption is an open issue even for the original version of the gradient push algorithm [17]. We hope to address this issue in the future.

The communication pattern among agents in (1) at each time $t \in \mathbb{N} \cup \{0\}$ is characterized by a directed graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$, where each node in $\mathcal{V}$ represents each agent, and each directed edge $(i, j) \in \mathcal{E}(t)$ means that i can send messages to j. In this work, we consider a sequence of graphs $\{\mathcal{G}(t)\}_{t\in\mathbb{N}}$ satisfying the following assumption.

Assumption II.2. *The sequence of graph $\{\mathcal{G}(t)\}_{t\in\mathbb{N}}$ is uniformly strongly connected, i.e., there exists a value $B \in \mathbb{N}$ such that the graph with edge set $\cup_{i=kB}^{(k+1)B-1}\mathcal{E}(i)$ is strongly connected for any $k \ge 0$.*

We define in-neighbors and out-neighbors of node i, respectively, as $N_i^{\text{in}}(t) = \{j|(j, i) \in \mathcal{E}(t)\} \cup \{i\}$ and $N_i^{\text{out}}(t) = \{j|(i, j) \in \mathcal{E}(t)\} \cup \{i\}$. Also the out-degree of node i is defined as $d_i^{\text{out}}(t) = |N_i^{\text{out}}(t)|$. Define the mixing matrix $A(t)$ such that $[A(t)]_{ij} = a_{ij}(t)$, where

$$a_{ij}(t) = \begin{cases} 1/d_j^{\text{out}}(t), & \text{if } i \in N_j^{\text{out}}(t), \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Here $a_{ij}(t)$ is a weight that agent i uses when it receives the state information of agent j. The mixing matrix $A(t) \in \mathbb{R}^{m\times m}$ is column stochastic and we recall some useful properties of this matrix from [17, Corollary 2].

Lemma II.3 ([17], *Corollary* 2)**.** *Suppose that the graph sequence $\{\mathcal{G}(t)\}$ is uniformly strongly connected. Then, the following statements are valid.*

1) *For each integer $s \ge 0$, there is a stochastic vector $\phi(s)$ such that for all i, j and $t \ge s$*

$$|[A(t : s)]_{ij} - \phi_i(t)| \le C_0\lambda^{t-s} \tag{4}$$

for some values $C_0 \ge 1$ and $\lambda \in (0, 1)$ depending on the graph sequence.

2) *The following inequality holds.*

$$Q := \inf_{t=0,1,\cdots} \min_{1\le i\le m} [A(t : 0)\mathbf{1}]_i \ge 1/n^{nB}. \tag{5}$$

Here we denote by $A(t : s)$ the matrix given as

$$A(t : s) = A(t)A(t-s)\cdots A(s) \quad \textit{for all } t \ge s \ge 0.$$

In Algorithm 1, each agent i maintains the current states $x_i(t) \in \mathbb{R}^d$ and $y_i(t) \in \mathbb{R}$ and the latest sent states $\hat{x}_i(t) \in \mathbb{R}^d$ and $\hat{y}_i(t) \in \mathbb{R}$ at time t. In a non-time-varying graph, the agents send their current states to neighbors simultaneously only when the differences between the current states and the latest sent states are larger than certain thresholds, which are called trigger times. For each time t, we denote by $\hat{x}_j(t) \in \mathbb{R}^d$ and $\hat{y}_j(t) \in \mathbb{R}$ the latest sent states that agent j sent to its neighbors at the latest trigger time $\kappa_j^x(t), \kappa_j^y(t)$ up to time t. Then we have

$$\hat{x}_j(s) = \hat{x}_j(\kappa_j^x(t)) = x_j(\kappa_j^x(t)), \quad \text{for all } \kappa_j^x(t) \le s \le t$$

and

$$\hat{y}_j(s) = \hat{y}_j(\kappa_j^y(t)) = y_j(\kappa_j^y(t)), \quad \text{for all } \kappa_j^y(t) \le s \le t.$$

We use the latest sent states to update $x_i(t+1)$ and $y_i(t+1)$ by (8), (9), (10), (11) in Algorithm 1 with $a_{ij}(t) = a_{ij}(0)$ for all $t \ge 0$, where the value $a_{ij}(t)$ is designed by the agent j as in (3) depending on the edge information at j. Each agent i sends the states $x_i(t + 1)$ and $y_i(t + 1)$ to its neighbors respectively if

$$\|x_i(t + 1) - \hat{x}_i(t)\| > \tau(t) \tag{6}$$

and

$$|y_i(t + 1) - \hat{y}_i(t)| > \zeta(t), \tag{7}$$

where $\tau(t), \zeta(t) > 0$ are the thresholds.

However, in a time-varying graph, even though a time t is not a trigger time in the sense of (6) and (7), the agent j has to send its current states to new neighbors if the neighbors $N_j^{out}(t)$ is changed. Furthermore, the value $a_{ij}(t)$ should be transmitted to agent $i \in N_j^{out}(t) \cup N_j^{out}(t-1)$ by the agent j. Covering these cases and trying to reduce the communication as much as possible, we impose the following additional rule for transmission in the time-varying graph case:

- If the $N_j^{out}(t)$ is not changed at time t, then agent j does not send $a_{ij}(t)$ to its neighbor $i \in N_j^{out}(t)$.
- If the $N_j^{out}(t)$ is changed at time t, then we follow the below rules.
  - The agent j sends the updated value $a_{ij}(t)$ to its neighbors $k \in N_j^{out}(t)$ and inform the agent $k \in N_j^{out}(t-1) \setminus N_j^{out}(t)$ of that the weight is updated as $a_{kj}(t) = 0$.
  - The agent j sends its latest sent states $\hat{x}_j(t)$ and $\hat{y}_j(t)$ to new neighbors $i \in N_j^{out}(t) \setminus N_j^{out}(t-1)$ which have not received the states since the latest triggering time.

Algorithm 1 Distributed Event-Triggered Gradient-Push Algorithm on Directed Graph

Require: Initialize $x_i(0)$ arbitraily and $y_i(0) = \hat{y}_i(0) = 1$ for all $i \in \{1, \cdots, m\}$. Set $\hat{x}_i(0) = x_i(0)$

for $t = 0, 1, \cdots,$ **do**
  if $t = 0$, or $\kappa_i^x(t)$ is updated **then**
    Send $\hat{x}_i(t)$ to its negihbors simultaneously.
  else
    Send information using the transmission rules
  end if
  if $t = 0$, or $\kappa_i^y(t)$ is updated **then**
    Send $\hat{y}_i(t)$ to its negihbors simultaneously.
  else
    Send information using the transmission rules
  end if
  Compute the new action as

$$\widehat{w}_i(t+1) = \sum_{j=1}^{m} a_{ij}(t)\hat{x}_j(t), \tag{8}$$

$$y_i(t+1) = \sum_{j=1}^{m} a_{ij}(t)\hat{y}_j(t), \tag{9}$$

$$\widehat{z}_i(t+1) = \frac{\widehat{w}_i(t+1)}{y_i(t+1)}, \tag{10}$$

$$x_i(t+1) = \widehat{w}_i(t+1) - \alpha(t+1)\nabla f_i(\widehat{z}_i(t+1)). \tag{11}$$

  if $\|x_i(t+1) - \hat{x}_i(t)\| \geq \tau(t+1)$ **then**
    Set $\hat{x}_i(t+1) = x_i(t+1)$ and update $\kappa_i^x(t+1)$.
  else
    Set $\hat{x}_i(t+1) = \hat{x}_i(t)$ and do not send
  end if
  if $|y_i(t+1) - \hat{y}_i(t)| \geq \zeta(t+1)$ **then**
    Set $\hat{y}_i(t+1) = y_i(t+1)$ and update $\kappa_i^y(t+1)$
  else
    Set $\hat{y}_i(t+1) = \hat{y}_i(t)$ and do noet send
  end if
end for

For the convergene analysis of Algorithm 1, we consider the following assumptions on the stepsize and the thresholds for the trigger conditions.

Assumption II.4. *The sequence of stepsize $\{\alpha(t)\}_{t\in\mathbb{N}}$ is monotonically non-increasing and satisfies*

$$\sum_{t=1}^{\infty} \alpha(t) = \infty, \quad \sum_{t=1}^{\infty} \alpha(t)^2 < \infty.$$

Assumption II.5. *The sequence of event-triggering thresholds $\{\tau(t)\}_{t\in\mathbb{N}}$ is monotonically non-increasing for $t \geq 1$ and we set $\tau(0) = 0$. In addition, the sequence satisfies*

$$\sum_{t=0}^{\infty} \tau(t) < \infty.$$

Assumption II.6. *The sequence of event-triggering thresholds $\{\zeta(t)\}_{t\in\mathbb{N}}$ is monotonically non-increasing for $t \geq 0$ and we set $\zeta(t) = 0$. In addition, the sequence satisfies*

$$\sum_{t=0}^{\infty} t^{3/2}\zeta(t) < \infty, \quad \sum_{t=0}^{\infty} \zeta(t) < 1.$$

This assumption on the triggering thresholds include the case that thresholds have exponential decays. This exponential decay assumption is commonly used in the literature [31], [34], [35]. However, we mention that the above assumptions imposed for the convergence analysis may not be optimal and it will be interesting to weaken these assumptions.

Note that $\sum_{t=0}^{\infty} t^{3/2}\zeta(t) < \infty$ implies that there exists a finite M such that $\sum_{t=0}^{\infty} \zeta(t) = M$. If we set a new sequence $\{\tilde{\zeta}(t)\}_{t\in\mathbb{N}}$ by $\tilde{\zeta}(t) = \zeta(t)/(M+1)$, then it satisfy $\sum_{t=0}^{\infty} \tilde{\zeta}(t) < 1$. Hence if we have a sequence $\{\zeta(t)\}_{t\in\mathbb{N}}$ satisfying $\sum_{t=0}^{\infty} t^{3/2}\zeta(t) < \infty$, then we may divide the sequence by a positive constant to satisfy Assumption II.6. One example of the sequence that satisfies Assumption II.6 is $\zeta(t) = \frac{1}{3t^3}$ for $t \geq 1$.

B. MAIN RESULTS

Our first result establishes the convergence of $\hat{z}_i(t)$ to the optimal solutions for an arbitrary stepsize $\alpha(t)$ satisfying Assumption II.4, and event-triggering thresholds $\tau(t)$ and $\zeta(t)$ satisfying Assumption II.5 and II.6.

Theorem II.7. *Suppose that Assumptions II.1,II.2,II.4. II.5 and II.6 hold. Then the sequence $\{\hat{z}_i(t)\}_{t\in\mathbb{N}}$ for $1 \leq i \leq n$ of the Algorithm 1 satisfies the following property*:

$$\lim_{t\to\infty} \hat{z}_i(t) = x^* \text{ for all } i \text{ and for some } x^* \in X^*.$$

Next we consider the Algorithm 1 with specific stepsize $\alpha(t) = 1/\sqrt{t}$. This stepsize does not satisfy Assumption II.4, but we may obtain an explicit convergence rate as in the following result. Before stating the result, we give some notations which are used throughout the paper. First we define the summations of the event-triggering thresholds $\tau(t)$ and $\zeta(t)$ as well as their squares: For $T \geq 0$,

$$E_\tau(T) = \sum_{t=0}^{T} \tau(t),\ E_{\tau,2}(T) = \sum_{t=0}^{T} \tau(t)^2,\ E_\tau = \sum_{t=0}^{\infty} \tau(t) \tag{12}$$

and

$$F_\zeta(T) = \sum_{t=0}^{T} \zeta(t), \quad F_\zeta = \sum_{t=0}^{\infty} \zeta(t), \quad F_{\zeta_{3/2}} = \sum_{t=0}^{\infty} t^{3/2}\zeta(t). \tag{13}$$

These quantities naturally appear in the convergence analysis of the algorithm. It is worth mentioning that these quantities are finite by Assumptions II.5 and II.6. It will turn out that the asymptotic behaivor of $y(t)$ as $t \to \infty$ is related with the vector ϕ in (4) and the constant m_ζ defined by

$$m_\zeta = \mathbf{1}_m^T y(0) + \sum_{s=1}^{\infty} \mathbf{1}_m^T \theta(s) = m + \sum_{s=1}^{\infty} \mathbf{1}_m^T \theta(s), \quad (14)$$

where $\theta(s) = \hat{y}(s) - y(s)$. For notational simplicity, we also define the following ratio

$$B_\zeta = \frac{m_\zeta}{m}. \quad (15)$$

These constants are well-defined if $\{\zeta(s)\}_{s\geq 0}$ is summable since we have the inequality $|\mathbf{1}_m^T\theta(s)| \leq m\zeta(s)$ from the triggering condition. Related to the asymptotic behavior of the states $y_i(t)$, we also define the values $\beta(t)$ and $K(t)$ by

$$\beta(t) = m\Big(\Big(F_\zeta - F_\zeta(t)\Big) + C_0\lambda^t + C_0\lambda^{t/2}F_\zeta(t) + \frac{\zeta([t/2]+1)}{1-\lambda}\Big)$$

and

$$K(t) = \frac{\beta(t)}{m_\zeta},$$

where $[a]$ denotes the largest integer not larger than $a \in \mathbb{R}$. For the convergence analysis, we also need to define the lower bound of all $y_i(t)$:

$$\delta := \min_{1\leq i\leq m} \inf_{t\in\mathbb{N}} y_i(t) > 0, \quad (16)$$

whose positivity is proved in Lemma III.2 under the Assumption II.6.

Theorem II.8. *Suppose that Assumption II.1, II.2, II.5 and II.6 hold. Let* $\alpha(t) = \frac{1}{\sqrt{t}}$ *for* $t \geq 1$. *Define* $H(-1) = 1$ *and* $H(t) := \prod_{k=0}^{t}(1+\tau(k))$. *Moreover, suppose that every node i maintains the variable* $\tilde{z}_i(t) \in \mathbb{R}^d$ *initialized at time* $t = 0$ *with* $\tilde{z}_i(0) \in \mathbb{R}^d$ *and updated by*

$$\tilde{z}_i(t+1) = \frac{\frac{\alpha(t+1)}{H(t)}\hat{z}_i(t+1) + S(t)\tilde{z}_i(t)}{S(t+1)},$$

where $S(0) = 0$ *and* $S(t) = \sum_{k=0}^{t-1}\frac{\alpha(k+1)}{H(k)}$ *for* $t \geq 1$. *Then we have for each* $T \geq 0$ *and* $i = 1, \cdots, m$, *the following estimate*

$$f(\tilde{z}_i(T+1)) - f(x^*) \leq \frac{me^{E_\tau}}{2\sqrt{T+1}}J_1(T) + \frac{3mDe^{E_\tau}}{\delta\sqrt{T+1}}J_2(T) + \frac{3mDe^{E_\tau}}{\delta\sqrt{T+1}}J_3(T),$$

where

$$J_1(T) = \frac{\|\bar{x}(0) - x\|^2}{B_\zeta} + \Big[2D^2\Big(1+\ln(T+1)\Big) + 2E_{\tau,2}(T) + E_\tau(T)\Big]B_\zeta$$

$$J_2(T) = \left(\frac{C_0}{(1-\lambda)}\right)\|x(0)\|_1 + \frac{4mC_0E_\tau(T)}{(1-\lambda)} + \left(\frac{C_0mD}{(1-\lambda)}\right)(1+\ln(T))$$

$$J_3(T) = \sum_{t=0}^{T} K(t)\alpha(t+1)\Big[\|x(0)\|_1 + \sum_{s=0}^{t-1}(\alpha(s+1)D + \tau(s))\Big],$$

and $x^* \in X^*$.

We mention that $J_3(t)$ is proved to be uniformly bounded for $t \geq 1$ in Lemma IV.4 under the assumption of the above theorem. Hence Theorem II.8 implies that $f(\tilde{z}_i(t))$ converges to $f(x^*)$ at the rate of $O(\log(t)/\sqrt{t})$.

III. PROPERTIES OF THE SEQUENCE $\{y_i(t)\}_{t\in\mathbb{N}}$ AND DISAGREEMENT IN AGENT ESTIMATES

A. PROPERTIES OF THE SEQUENCE $\{y_i(t)\}_{t\in\mathbb{N}}$

A convergence property of $\{y_i(t)\}_{t\in\mathbb{N}}$ and their positive uniform lower bound are key points in proving our main results. Let us first look at the case without event-triggering ($\zeta(t) = 0$), which means all in-neighbors of agent i share the $y_i(t)$ with this agent for every time step. In this case, since $y(t) = \hat{y}(t)$ for all $t \in \mathbb{N}$, it holds that

$$y(t) = A(t-1:0)\mathbf{1}_m \quad (17)$$

by (9) in Algorithm 1. Hence we can directly show that $y(t)$ converges to $\phi(t)$ and has a uniform lower bound Q using Lemma II.3. In the event-triggered case, $y(t)$ can be written as

$$y(t) = A(t-1:0)\mathbf{1}_m + \sum_{s=1}^{t-1} A(t-1:s)\theta(s). \quad (18)$$

Therefore the convergence and uniform lower boundedness property may not hold due to the additional term

$$\sum_{s=1}^{t-1} A(t-1:s)\theta(s).$$

The following lemmas shows that $y(t)$ has a positive uniform lower bound δ and converges to $m_\zeta\phi(t)$ instead of $\phi(t)$ under Assumption II.6.

Lemma III.1. *Suppose that Assumption II.6 holds. Then we have*

$$m_\zeta \geq (1 - F_\zeta)m \quad (19)$$

and the following estimate holds:

$$\Big\|y(t+1) - m_\zeta\phi(t)\Big\|_\infty \leq \beta(t) \quad \forall\, t \in \mathbb{N}. \quad (20)$$

In addition, we have

$$\lim_{t\to\infty} t^{3/2}\beta(t) = 0.$$

Proof: Observe that $|\theta_i(s)| = |\hat{y}_i(s) - y_i(s)| \leq \zeta(s)$ for $s \geq 1$ by the event-triggering condition, and so

$$\Big|\sum_{s=1}^{\infty} 1_m^T\theta(s)\Big| \leq m\sum_{s=1}^{\infty}\zeta(s) = mF_\zeta.$$

Using this in (14), we get

$$m_\zeta = m + \sum_{s=1}^{\infty} 1_m^T \theta(s) \geq (1 - F_\zeta)m,$$

which proves (19).

Next we prove (20). For each $s \geq 0$, by definition we have

$$y(s+1) = A(s)\hat{y}(s) = A(s)(y(s) + \theta(s)),$$

where we have set $\theta(0) = 0$. Using this iteratively gives the following formula

$$\begin{aligned} &y(t+1) \\ &= A(t:0)y(0) + \sum_{s=1}^{t} A(t:s)\theta(s) \\ &= \phi(t)\Big[\mathbf{1}_m^T y(0) + \sum_{s=1}^{t} \mathbf{1}_m^T \theta(s)\Big] \\ &\quad + (A(t:0) - \phi(t)\, 1_m^T)y(0) + \sum_{s=1}^{t}\Big[(A(t:s) - \phi(t)\, \mathbf{1}_m^T)\theta(s)\Big]. \end{aligned}$$

Since $|\theta_j(s)| \leq \zeta(s)$, we find

$$\sum_{s=t+1}^{\infty} |\theta_j(s)| \leq \sum_{s=t+1}^{\infty} \zeta(s) = F_\zeta - F_\zeta(t).$$

Using the above inequality, we obtain

$$\begin{aligned} &\Big| m_\zeta - 1_m^T y(0) - \sum_{s=0}^{t} 1_m^T \theta(s) \Big| \\ &\qquad = \Big| \sum_{s=t+1}^{\infty} 1_m^T \theta(s) \Big| \leq \Big(F_\zeta - F_\zeta(t)\Big)m. \end{aligned} \tag{21}$$

Hence we have

$$\begin{aligned} &\Big\| y(t+1) - m_\zeta \phi(t) \Big\|_\infty \\ &\quad \leq \Big(F_\zeta - F_\zeta(t)\Big)m + \Big\| (A(t:0) - \phi(t)\, 1_m^T)y(0) \Big\|_\infty \\ &\qquad + \Big\| \sum_{s=1}^{t}\Big[(A(t:s) - \phi(t)\, 1_m^T)\theta(s)\Big] \Big\|_\infty. \end{aligned} \tag{22}$$

Now we estimate the second and third terms in the right hand side of (22). Using (4) we have

$$\Big\| (A(t:0) - \phi(t)\, 1_m^T)y(0) \Big\|_\infty \leq mC_0\lambda^t \tag{23}$$

and

$$\begin{aligned} &\Big\| \sum_{s=1}^{t}\Big[(A(t:s) - \phi(t)\, 1_m^T)\theta(s)\Big] \Big\|_\infty \\ &\quad \leq mC_0 \sum_{s=1}^{t} \lambda^{t-s}\zeta(s) \\ &\quad \leq mC_0 \Big(\sum_{s=1}^{[t/2]} \lambda^{t-s}\zeta(s) + \sum_{s=[t/2]+1}^{t} \lambda^{t-s}\zeta(s) \Big) \\ &\quad \leq mC_0 \Big(\lambda^{t/2} F_\zeta(t) + \zeta([t/2]+1) \sum_{s=[t/2]+1}^{t} \lambda^{t-s} \Big) \\ &\quad \leq mC_0 \Big(\lambda^{t/2} F_\zeta(t) + \frac{\zeta([t/2]+1)}{1-\lambda} \Big). \end{aligned} \tag{24}$$

Putting the estimates (23) and (24) in (22), we get

$$\begin{aligned} &\Big\| y(t+1) - \phi(t)m_\zeta \Big\|_\infty \\ &\quad \leq m\Big(\Big(F_\zeta - F_\zeta(t)\Big) \\ &\qquad + C_0\lambda^t + C_0\lambda^{t/2}F_\zeta(t) + \frac{\zeta([t/2]+1)}{1-\lambda} \Big). \end{aligned}$$

This proves the second assertion of the lemma.

Now we shall show that $\lim_{t\to\infty} t^{3/2}\beta(t) = 0$. Since $\lambda \in (0,1)$, it suffices to show that

$$\lim_{t\to\infty} t^{3/2}(F_\zeta - F_\zeta(t) + \zeta([t/2])) = 0.$$

This fact follows directly from the fact that $\sum_{t=0}^{\infty} t^{3/2}\zeta(t) < \infty$ and the following inequality

$$t^{3/2}(F_\zeta - F_\zeta(t)) = t^{3/2} \sum_{s=t+1}^{\infty} \zeta(s) \leq \sum_{s=t+1}^{\infty} s^{3/2}\zeta(s).$$

The proof is done. □

Lemma III.2. *Suppose that Assumptions II.2 and II.6 hold. Then the value $\delta \in \mathbb{R}$ defined in* (16) *is positive.*

Proof: Note that from (4) and (5), we have

$$\begin{aligned} m\phi_i(t) &= \sum_{j=1}^{m} [A(t:0)]_{ij} + \sum_{j=1}^{m} \Big(\phi_i(t) - [A(t:0)]_{ij} \Big) \\ &\geq Q - mC_0\lambda^t. \end{aligned}$$

Using the above inequality and Lemma III.1, we deduce for each $1 \leq i \leq m$ the following estimate

$$\begin{aligned} y_i(t+1) &\geq m_\zeta \phi_i(t) - \beta(t) \\ &\geq (m_\zeta/m)Q - m_\zeta C_0\lambda^t - \beta(t). \end{aligned}$$

Since $\beta(t)$ converges to zero as t goes to infinity and $\lambda \in (0,1)$, there exists a time $T \in \mathbb{N}$ and a constant $\tilde{\delta} > 0$ such that for any $t \geq T$,

$$y_i(t+1) \geq \tilde{\delta}. \tag{25}$$

Note that by Assumption II.2, each matrix $A(t)$ has no zero row. This fact, together with the definition of $\hat{y}$ and (9), for any $t \in \mathbb{N}$ we have

$$\min_{1 \leq i \leq m} y_i(t) > 0. \tag{26}$$

Therefore, combining (25) with (26), we conclude that δ defined in (16) satisfies

$$\delta \geq \min_{1 \leq i \leq m} \{y_i(0), y_i(1), \cdots, y_i(T), \tilde{\delta}\} > 0.$$

The proof is done. □

B. DISAGREEMENT IN AGENT ESTIMATES

In this subsection, we derive a bound of the disagreement in agent estimates $\{\hat{z}_i(t)\}_{i=1}^m$ that will be used in the proofs of the main theorems. In the case without event-triggering ($\tau(t) = \zeta(t) = 0$), the paper [17] proved that $\|\hat{z}_i(t+1)-\bar{x}(t)\|$ converges to zero for the stepsize satisfying Assumption II.4 as t goes to infinity. For the event-triggered case, the following proposition shows that the values $\{\hat{z}_i(t)\}_{i=1}^m$ approach $B_\zeta \bar{x}(t)$ instead of $\bar{x}(t)$ as t goes to infinity due to the effect of the threshold $\zeta(t)$ for the triggering condition of $\{y_i(t)\}_{i=1}^m$.

Proposition III.3. *Suppose that Assumptions II.1, II.2 and II.6 hold. Then for any $t \geq 1$ we have*

$$\begin{aligned}&\|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\| \\ &\quad\leq \frac{1}{\delta}\Big(C_0\lambda^t + K(t)\Big)\|x(0)\|_1 \\ &\quad\quad+\frac{m}{\delta}\sum_{s=0}^{t-1}\Big[C_0\lambda^{t-s-1} + K(t)\Big]\Big(\alpha(s+1)D + \tau(s)\Big) \\ &\quad\quad+\frac{d_i(t)\tau(t)}{\delta},\end{aligned}$$

and for $t = 0$ we have

$$\|\hat{z}_i(1) - \bar{x}(0)\| \leq \frac{2C_0}{\delta}\|x(0)\|,$$

where the constant $\delta > 0$ satisfies $y_i(t) > \delta$ for all $t > 0$.

To prove Proposition III.3, we consider a variable $w_i(t+1) \in \mathbb{R}^d$ which is a companion to the variable $\hat{w}_i(t+1) \in \mathbb{R}^n$ defined as

$$w_i(t+1) = \sum_{j=1}^m a_{ij}(t)\, x_j(t), \tag{27}$$

and their difference

$$e_i(t+1) = \hat{w}_i(t+1) - w_i(t+1). \tag{28}$$

Then we may rewrite the gradient step (II-A) as

$$x_i(t+1) = w_i(t+1) - \alpha(t+1)\nabla f_i(\hat{z}_i(t+1)) + e_i(t+1). \tag{29}$$

Summing up (29) for $1 \leq i \leq m$ and using that $A(t)$ is column-stochastic, we have

$$\begin{aligned}\bar{x}(t+1) = \bar{x}(t) - \frac{\alpha(t+1)}{m}\sum_{i=1}^m \nabla f_i(\hat{z}_i(t+1)) \\ +\frac{1}{m}\sum_{i=1}^m e_i(t+1).\end{aligned} \tag{30}$$

Now we find a bound of $e_i(t+1)$ which is the difference between $w_i(t+1)$ and $\hat{w}_i(t+1)$ associated to the event-triggering $\tau(t)$.

Lemma III.4. *Suppose that Assumption II.2 hold. The quantity $e_i(t+1)$ defined in (28) satisfies*

$$\|e_i(t+1)\| \leq d_i(t)\,\tau(t), \tag{31}$$

where $d_i(t) = \sum_{j=1}^m a_{ij}(t)$. In addition, we have

$$\sum_{i=1}^m \|e_i(t+1)\| \leq m\tau(t). \tag{32}$$

Proof: By using the triggering condition, we have

$$\begin{aligned}\|e_i(t+1)\| &\leq \|\hat{w}_i(t+1) - w_i(t+1)\| \\ &\leq \Big\|\sum_{j=1}^m a_{ij}(t)(\hat{x}_j(t) - x_j(t))\Big\| \\ &\leq \sum_{j=1}^m a_{ij}(t)\|\hat{x}_j(t) - x_j(t)\| \leq d_i(t)\,\tau(t),\end{aligned}$$

which proves (31). Summing this over $1 \leq i \leq m$ and using that $A(t)$ is column stochastic, we find

$$\begin{aligned}\sum_{i=1}^m \|e_i(t+1)\| &\leq \sum_{i=1}^m\sum_{j=1}^m \Big(a_{ij}(t)\,\tau(t)\Big) \\ &= \sum_{j=1}^m\Big(\sum_{i=1}^m a_{ij}(t)\Big)\tau(t) \;= m\tau(t).\end{aligned}$$

The proof is finished. ☐

Now we are ready to prove Proposition III.3.

Proof: [Proof of Proposition III.3] We regard $x_k(t)$ as a row vector in $\mathbb{R}^{1\times d}$ and define the variables $x(t) \in \mathbb{R}^{m\times d}$, $\nabla f(\hat{z}(t)) \in \mathbb{R}^{m\times d}$, and $e(t) \in \mathbb{R}^{m\times d}$ as

$$x(t)=\begin{pmatrix} x_1(t) \\ \vdots \\ x_m(t)\end{pmatrix},\ \nabla f(\hat{z}(t))=\begin{pmatrix}\nabla f_1(\hat{z}_1(t)) \\ \vdots \\ \nabla f_m(\hat{z}_m(t))\end{pmatrix},\ e(t) =\begin{pmatrix} e_1(t) \\ \vdots \\ e_m(t)\end{pmatrix}.$$

Note that by (28) and (10), we have

$$\hat{z}_i(t+1) = \frac{w_i(t+1) + e_i(t+1)}{y_i(t+1)}.$$

Also we see from definition (15) that

$$B_\zeta \bar{x}(t) = \frac{m\bar{x}(t)}{m_\zeta} = \frac{\mathbf{1}_m^T x(t)}{m_\zeta}.$$

Using these formulas and (27) we have

$$\begin{aligned}&\hat{z}_i(t+1) - B_\zeta\bar{x}(t) \\ &= \frac{w_i(t+1) + e_i(t+1)}{y_i(t+1)} - \frac{\mathbf{1}_m^T x(t)}{m_\zeta} \\ &= \frac{1}{y_i(t+1)}\Big([A(t)x(t)]_i - y_i(t+1)\frac{\mathbf{1}_m^T x(t)}{m_\zeta}\Big) + \frac{e_i(t+1)}{y_i(t+1)}.\end{aligned} \tag{33}$$

To estimate the first term on the right hand side of the last equality, we rewrite (29) as

$$x(t+1) = A(t)x(t) - \alpha(t+1)\nabla f(\hat{z}(t+1)) + e(t+1).$$

Using this formula recursively, for $t \geq 1$ we have

$$A(t)\,x(t) = A(t:0)x(0) - \sum_{s=0}^{t-1} A(t:s+1)\varepsilon(s), \tag{34}$$

where we have let

$$\varepsilon(s) = \alpha(s+1)\nabla f(\hat{z}(s+1)) - e(s+1).$$

Using Assumption II.1 and (32) we have the following bound

$$\|\varepsilon(s)\|_1 \leq m\Big(\alpha(s+1)D + \tau(s)\Big). \tag{35}$$

Since $A(t)$ is column stochastic we have $1_m^T A(t) = 1_m^T$, and combine this with (34) to have

$$\mathbf{1}_m^T x(t) = \mathbf{1}_m^T x(0) - \sum_{s=0}^{t-1} \mathbf{1}_m^T \varepsilon(s). \tag{36}$$

Combining (34) and (36) yields

$$A(t)x(t) = \phi(t)\mathbf{1}_m^T x(t) + \Big(A(t:0) - \phi(t)\mathbf{1}_m^T\Big)x(0) - \sum_{s=0}^{t-1}\Big(A(t:s+1) - \phi(t)\mathbf{1}_m^T\Big)\varepsilon(s), \tag{37}$$

where $\phi(t)$ is the stochastic vector satisfying (4). By Lemma III.1, for $y(t) := (y_1(t), \cdots, y_m(t))^T \in \mathbb{R}^{m\times 1}$ we have

$$y(t+1) = m_\zeta \phi(t) + r(t),$$

where $r(t)$ satisfies $\|r(t)\|_\infty \leq \beta(t)$. Combining this with (37), we obtain

$$\begin{aligned}
&[A(t)x(t)]_i - y_i(t+1)\frac{1_m^T x(t)}{m_\zeta} \\
&= \phi_i(t)\, 1_m^T x(t) + [(A(t:0) - \phi(t)1_m^T)x(0)]_i \\
&\quad - \sum_{s=0}^{t-1}\Big[(A(t:s+1) - \phi(t)\, 1_m^T)\varepsilon(s)\Big]_i \\
&\quad - [m_\zeta \phi_i(t) + r_i(t)]\frac{1_m^T x(t)}{m_\zeta} \\
&= [(A(t:0) - \phi(t)\mathbf{1}_m^T)x(0)]_i \\
&\quad - \sum_{s=0}^{t-1}\Big[(A(t:s+1) - \phi(t)\, \mathbf{1}_m^T)\varepsilon(s)\Big]_i - r_i(t)\,\frac{1_m^T x(t)}{m_\zeta}.
\end{aligned}$$

By applying (4) here, we deduce

$$\begin{aligned}
&\left\|[A(t)x(t)]_i - y_i(t+1)\frac{1_m^T x(t)}{m}\right\| \\
&\leq C_0\lambda^t \|x(0)\|_1 + \sum_{s=0}^{t-1} C_0\lambda^{t-s-1}\|\varepsilon(s)\|_1 + K(t)\|1_m^T x(t)\|,
\end{aligned} \tag{38}$$

where $K(t) = \beta(t)/m_\zeta$. From (36) we find the following estimate

$$\|1_m^T x(t)\| \leq \|x(0)\|_1 + \sum_{s=0}^{t-1}\|\varepsilon(s)\|_1.$$

Combining this with (38) and using (35), we obtain

$$\begin{aligned}
&\left\|[A(t)x(t)]_i - y_i(t+1)\frac{1_m^T x(t)}{m}\right\| \\
&\leq C_0\lambda^t\|x(0)\|_1 + \sum_{s=0}^{t-1} C_0\lambda^{t-s-1}\|\varepsilon(s)\|_1 \\
&\quad + K(t)\Big(\|x(0)\|_1 + \sum_{s=0}^{t-1}\|\varepsilon(s)\|_1\Big) \\
&\leq \Big(C_0\lambda^t + K(t)\Big)\|x(0)\|_1 \\
&\quad + m\sum_{s=0}^{t-1}\Big[C_0\lambda^{t-s-1} + K(t)\Big]\Big(\alpha(s+1)D + \tau(s)\Big).
\end{aligned} \tag{39}$$

By applying Lemma III.2, (31) and the above inequality to the norm of (33), we obtain

$$\begin{aligned}
&\|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\| \\
&\leq \frac{1}{\delta}\Big(C_0\lambda^t + K(t)\Big)\|x(0)\|_1 \\
&\quad + \frac{m}{\delta}\sum_{s=0}^{t-1}\Big[C_0\lambda^{t-s-1} + K(t)\Big]\Big(\alpha(s+1)D + \tau(s)\Big) \\
&\quad + \frac{d_i(t)\tau(t)}{\delta},
\end{aligned}$$

It remains to estimate the case $t = 0$. By the algorithm, we have

$$\hat{z}_i(1) - \bar{x}(0) = \frac{\hat{w}_i(1)}{y_i(1)} - \bar{x}(0) = \frac{\sum_{j=1}^m a_{ij}(0)x_j(0)}{\sum_{j=1}^m a_{ij}(0)} - \bar{x}(0).$$

Using this we find

$$\begin{aligned}
\|z_i(1) - \bar{x}(0)\| &\leq \frac{1}{\delta}\|x(0)\|_1 + \frac{1}{m}\|x(0)\|_1 \\
&\leq 2\|x(0)\|_1 \leq \frac{2C_0}{\delta}\|x(0)\|.
\end{aligned}$$

The proof is finished. □

By utilizing Proposition III.3, we analyze the relation between $\hat{z}_i(t)$ and $B_\zeta \bar{x}(t)$ under the assumptions on $\{\alpha(t)\}_{t\in\mathbb{N}}$, $\{\tau(t)\}_{t\in\mathbb{N}}$ and $\{\zeta(t)\}_{t\in\mathbb{N}}$ of the main theorems. To do this, we first recall a useful lemma from [8].

Lemma III.5 ([8], *Lemma* 3.1). *If* $\lim_{k\to\infty}\gamma_k = \gamma$ *and* $0 < \beta < 1$, *then*

$$\lim_{k\to\infty}\sum_{l=0}^{k}\beta^{k-l}\gamma_l = \frac{\gamma}{1-\beta}$$

Corollary III.6. *Suppose that Assumptions II.1, II.2, II.5 and II.6 hold. Also, assume that the stepsize* $\alpha(t)$ *satisfies Assumption II.4 or* $\alpha(t) = 1/\sqrt{t}$. *Then we have*

$$\lim_{t\to\infty}\|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\| = 0 \text{ for all } i.$$

Proof: We recall from Proposition III.3 the following inequality

$$\begin{aligned}
&\|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\| \\
&\leq \frac{1}{\delta}\Big(C_0\lambda^t + K(t)\Big)\|x(0)\|_1
\end{aligned}$$

$$+\frac{m}{\delta}\sum_{s=0}^{t-1}\Big[C_0\lambda^{t-s-1}+K(t)\Big]\Big(\alpha(s+1)D+\tau(s)\Big)$$
$$+\frac{d_i(t)\tau(t)}{\delta}, \tag{40}$$

We notice that $\lim_{t\to\infty} t^{3/2}K(t)=0$ by Lemma III.1. From this and the boundedness of $\alpha(s)$ and $\tau(s)$, it easily follows that

$$\lim_{t\to\infty}\frac{1}{\delta}K(t)\|x(0)\|_1+m\sum_{s=0}^{t-1}K(t)\Big(\alpha(s+1)D+\tau(s)\Big)=0.$$

In addition, by Assumptions II.4, II.5 and II.6, we know that $\lim_{s\to\infty}\alpha(s+1)=0$, $\lim_{s\to\infty}\tau(s)=0$ and $\lim_{s\to\infty}\zeta(s)=0$. Using this fact with Lemma III.5 in the right hand side of (40), we deduce

$$\lim_{t\to\infty}\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|=0,$$

which completes the proof. □

Corollary III.7. *Suppose that Assumptions II.1, II.2 II.5 and II.6 hold. Let $\alpha(t)=\frac{1}{\sqrt{t}}$. Then we have*

$$\sum_{t=0}^{T}\alpha(t+1)\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|$$
$$\le\frac{C_0}{\delta(1-\lambda)}\|x(0)\|_1+\frac{4mC_0E_\tau(T)}{\delta(1-\lambda)}+\frac{C_0mD}{\delta(1-\lambda)}(1+\ln(T))$$
$$+\frac{1}{\delta}\sum_{t=0}^{T}K(t)\alpha(t+1)\Big[\|x(0)\|_1+\sum_{s=0}^{t-1}(\alpha(s+1)D+\tau(s))\Big].$$

Proof: By Proposition III.3, we have

$$\delta\sum_{t=0}^{T}\alpha(t+1)\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|\le$$
$$\|x(0)\|_1\sum_{t=0}^{T}(C_0\lambda^t+K(t))\alpha(t+1)$$
$$+m\sum_{t=0}^{T}\alpha(t+1)\sum_{s=0}^{t-1}(C_0\lambda^{t-s-1}+K(t))(\alpha(s+1)D+\tau(s))$$
$$+\sum_{t=0}^{T}\alpha(t+1)(d_i(t)\,\tau(t)). \tag{41}$$

The terms involving $K(t)$ are fit to the inequality of the corollary. Let us estimate each summation not involving $K(t)$ in the right hand side. Using that $\alpha(t)\le 1$, the first term is bounded with

$$\sum_{t=0}^{T}\alpha(t+1)\lambda^t\le\sum_{t=0}^{T}\lambda^t\le\frac{1}{1-\lambda}. \tag{42}$$

The last term is bounded using

$$\sum_{t=0}^{T}\alpha(t+1)(d_i(t)\tau(t))\le m\sum_{t=0}^{T}\tau(t)=mE_\tau(T). \tag{43}$$

We estimate the second term using

$$\begin{aligned}\sum_{t=0}^{T}\alpha(t+1)\sum_{s=0}^{t-1}\lambda^{t-s-1}\alpha(s+1)&=\sum_{t=1}^{T+1}\frac{1}{\sqrt{t}}\sum_{s=1}^{t}\lambda^{t-s}\frac{1}{\sqrt{s}}\\&\le\sum_{t=1}^{T+1}\sum_{s=1}^{t}\lambda^{t-s}\frac{1}{s}\\&=\sum_{s=1}^{T+1}\frac{1}{s}\sum_{t=s}^{T+1}\lambda^{t-s}\\&\le\frac{1+\ln(T+1)}{1-\lambda}.\end{aligned} \tag{44}$$

In order to estimate the third term, we estimate

$$\begin{aligned}&\sum_{s=0}^{t-1}\lambda^{t-s-1}\tau(s)\\&=\sum_{s=0}^{[(t-1)/2]}\lambda^{t-s-1}\tau(s)+\sum_{s=[(t-1)/2]+1}^{t-1}\lambda^{t-s-1}\tau(s)\\&\le\lambda^{(t-1)/2}\sum_{s=0}^{[(t-1)/2]}\tau(s)+\tau([t/2])\sum_{s=[(t-1)/2]+1}^{t-1}\lambda^{t-s-1}\\&\le\lambda^{(t-1)/2}E_\tau(T)+\frac{\tau([t/2])}{1-\lambda}.\end{aligned}$$

Using this we derive

$$\begin{aligned}&\sum_{t=0}^{T}\alpha(t+1)\Big[\sum_{s=0}^{t-1}\lambda^{t-s-1}\tau(s)\Big]\\&\le E_\tau(T)\sum_{t=0}^{T}\frac{\lambda^{(t-1)/2}}{\sqrt{t+1}}+\frac{1}{1-\lambda}\sum_{t=0}^{T}\frac{\tau([t/2])}{\sqrt{t+1}}\\&\le\frac{E_\tau(T)}{1-\sqrt{\lambda}}+\frac{E_\tau(T)}{1-\lambda}<\frac{3E_\tau(T)}{1-\lambda}.\end{aligned} \tag{45}$$

Putting the above estimates (42)-(45) in (41), we obtain

$$\sum_{t=0}^{T}\alpha(t+1)\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|$$
$$\le\frac{C_0}{\delta(1-\lambda)}\|x(0)\|_1+\frac{4mC_0E_\tau(T)}{\delta(1-\lambda)}+\frac{C_0mD}{\delta(1-\lambda)}(1+\ln(T))$$
$$+\frac{1}{\delta}\sum_{t=0}^{T}K(t)\alpha(t+1)\Big[\|x(0)\|_1$$
$$+\sum_{s=0}^{t-1}(\alpha(s+1)D+\tau(s))\Big].$$

which finishes the proof. □

IV. CONVERGENCE ESTIMATES

In this section we prove our main results, namely Theorems II.7 and II.8. In Section 3, we obtained the bound of the disagreement in agent estimates. Especially, Corollary III.6 and III.7 investigate the difference between the variable $\hat{z}_i(t)$ in the Algorithm 1 and $B_\zeta\bar{x}(t)$ in (30). Based upon these results, Theorem II.7 and II.8 can be proved by comparing the cost values computed at the points $B_\zeta \bar{x}(t)$ and x^*.

Lemma IV.1. *Suppose Assumptions II.1, II.2 hold. Then for any $t \geq 0$ and $x \in \mathbb{R}^d$ we have*

$$\begin{aligned}\sum_{i=1}^m \big(f_i(B_\zeta \bar{x}(t)) - f_i(x)\big) &\leq \frac{m}{2\alpha(t+1)B_\zeta}(\|B_\zeta \bar{x}(t) - x\|^2 - \|B_\zeta \bar{x}(t+1) - x\|^2) \\ &+ \frac{B_\zeta m}{2\alpha(t+1)}\big(2\alpha(t+1)^2 D^2 + 2\tau(t)^2\big) \\ &+ \frac{m}{\alpha(t+1)}\|B_\zeta \bar{x}(t) - x\|\tau(t) \\ &+ 2D\sum_{i=1}^m \|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\|. \end{aligned} \tag{46}$$

Proof: By convexity, we have

$$\begin{aligned} f_i(\hat{z}_i(t+1)) &\leq f_i(x) + (\hat{z}_i(t+1) - x)\nabla f_i(\hat{z}_i(t+1)) \\ &= f_i(x) + (B_\zeta \bar{x}(t) - x)\nabla f_i(\hat{z}_i(t+1)) \\ &\quad + (\hat{z}_i(t+1) - B_\zeta \bar{x}(t))\nabla f_i(\hat{z}_i(t+1)) \\ &= f_i(x) + \frac{1}{\alpha(t+1)}(B_\zeta \bar{x}(t) - x)(w_i(t+1) - x_i(t+1) \\ &\quad + e_i(t+1)) + (\hat{z}_i(t+1) - B_\zeta \bar{x}(t))\nabla f_i(\hat{z}_i(t+1)), \end{aligned}$$

where (27) is used in the last equality. Summing up the above inequality from $i = 1$ to $i = m$, we find that

$$\begin{aligned}\sum_{i=1}^m f_i(\hat{z}_i(t+1)) - f_i(x) &\leq \underbrace{\frac{m}{\alpha(t+1)}(B_\zeta \bar{x}(t) - x)(\bar{x}(t) - \bar{x}(t+1))}_{I} \\ &+ \underbrace{\frac{1}{\alpha(t+1)}(B_\zeta \bar{x}(t) - x)\sum_{i=1}^m e_i(t+1)}_{II} \\ &+ \underbrace{\sum_{i=1}^m (\hat{z}_i(t+1) - B_\zeta \bar{x}(t))\nabla f_i(\hat{z}_i(t+1))}_{III}. \end{aligned}$$

Now we estimate each term in the right hand side. First using the equality $\langle a, b\rangle = \frac{1}{2}(\|a\|^2 + \|b\|^2 - \|a-b\|^2)$ for $a, b \in \mathbb{R}^d$, we have

$$\begin{aligned} I = \frac{m}{2\alpha(t+1)B_\zeta}(&\|B_\zeta \bar{x}(t) - x\|^2 - \|B_\zeta \bar{x}(t+1) - x\|^2 \\ &+ \|B_\zeta \bar{x}(t+1) - B_\zeta \bar{x}(t)\|^2). \end{aligned}$$

Using (30) along with (31) and (2), we estimate the right-most term as

$$\begin{aligned} &\|\bar{x}(t+1) - \bar{x}(t)\|^2 \\ &\leq 2\left\|\frac{\alpha(t+1)}{m}\sum_{i=1}^m \nabla f_i(\hat{z}_i(t+1))\right\|^2 + 2\left\|\frac{1}{m}\sum_{i=1}^m e_i(t+1)\right\|^2 \\ &\leq 2\alpha(t+1)^2 D^2 + 2\tau(t)^2. \end{aligned}$$

We apply (31) again to estimate

$$II \leq \frac{m}{\alpha(t+1)}\|B_\zeta \bar{x}(t) - x\|\tau(t),$$

and use (2) to deduce

$$III \leq D\sum_{i=1}^m \|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\|.$$

Combining the above estimates on I, II and III, we have

$$\begin{aligned}\sum_{i=1}^m \big(f_i(\hat{z}_i(t+1)) - f_i(x)\big) &\leq \frac{m}{2\alpha(t+1)B_\zeta}(\|B_\zeta \bar{x}(t) - x\|^2 - \|B_\zeta \bar{x}(t+1) - x\|^2) \\ &+ \frac{mB_\zeta}{2\alpha(t+1)}\big(2\alpha(t+1)^2 D^2 + 2\tau(t)^2\big) \\ &+ \frac{m}{\alpha(t+1)}\|B_\zeta \bar{x}(t) - x\|\tau(t) \\ &+ D\sum_{i=1}^m \|\hat{z}_i(t+1) - B_\zeta \bar{x}(t)\|. \end{aligned}$$

Finally we observe that (2) gives us the estimate

$$\sum_{i=1}^m \Big(f_i(B_\zeta \bar{x}(t)) - f_i(\hat{z}_i(t+1))\Big) \leq D\sum_{i=1}^m \|B_\zeta \bar{x}(t) - \hat{z}_i(t+1)\|.$$

Summing up the above two inequalities, we obtain the desired estimate. □

A. PROOF OF THEOREM II.7

We recall the following lemma for proving Theorem II.7.

Lemma IV.2. *([17] Lemma 7). Consider a minimization problem*

$$\min_{x \in \mathbb{R}^d} f(x),$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is a continuous function. Assume that the solution X^* of the problem is nonempty. Let $\{x(t)\}_{t\in\mathbb{N}}$ be a sequence such that for all $x \in X^*$ and for all $t \geq 0$,

$$\begin{aligned} &\|x(t+1) - x\|^2 \\ &\quad \leq (1 + b(t))\|x(t) - x\|^2 - a(t)(f(x(t)) - f(x)) + c(t) \end{aligned}$$

where $b(t) \geq 0$, $a(t) \geq 0$ and $c(t) \geq 0$ for all $t \geq 0$ with $\sum_{t=0}^\infty b(t) < \infty$, $\sum_{t=0}^\infty a(t) = \infty$ and $\sum_{t=0}^\infty c(t) < \infty$. Then the sequence $\{x(t)\}_{t\in\mathbb{N}}$ converges to some solution $x^* \in X^*$

By manipulating the estimate in Lemma IV.1, we obtain the following estimate which is suitable for applying Lemma IV.2.

Corollary IV.3. *Suppose Assumptions II.1 and II.2 hold. Then we have*

$$\begin{aligned} \|B_\zeta \bar{x}(t+1) - x\|^2 &\leq (1 + \tau(t))\|B_\zeta \bar{x}(t) - x\|^2 \\ &- \frac{2\alpha(t+1)B_\zeta}{m}(f(B_\zeta \bar{x}(t)) - f(x)) + c(t) + d(t), \end{aligned}$$

where

$$c(t) = \Big[2\alpha(t+1)^2 D^2 + 2\tau(t)^2 + \tau(t)\Big]B_\zeta,$$

and

$$d(t) = \frac{4\alpha(t+1)B_\zeta D}{m}\sum_{i=1}^m \|\hat{z}_i(t+1) - \bar{x}(t)\|.$$

Proof: We use Young's inequality to find

$$\|B_\zeta \bar{x}(t) - x\|\tau(t) \leq \frac{\|B_\zeta \bar{x}(t) - x\|^2\,\tau(t)}{2B_\zeta} + \frac{\tau(t)B_\zeta}{2}.$$

Applying this to (46), we get

$$\begin{aligned}\sum_{i=1}^m &\big(f_i(B_\zeta\bar{x}(t)) - f_i(x)\big)\\ &\leq \frac{m}{2\alpha(t+1)B_\zeta}(1+\tau(t))\|B_\zeta\bar{x}(t) - x\|^2\\ &\quad - \frac{m}{2\alpha(t+1)B_\zeta}\|B_\zeta\bar{x}(t+1) - x\|^2\\ &\quad + \frac{mB_\zeta}{2\alpha(t+1)}\big(2\alpha(t+1)^2D^2 + 2\tau(t)^2\big) + \frac{B_\zeta m\tau(t)}{2\alpha(t+1)}\\ &\quad + 2D\sum_{i=1}^m \|\hat{z}_i(t+1) - B_\zeta\bar{x}(t)\|. \end{aligned} \tag{47}$$

Dividing both sides by $\frac{m}{2\alpha(t+1)B_\zeta}$, it follows that

$$\begin{aligned}\frac{2\alpha(t+1)B_\zeta}{m}&\sum_{i=1}^m \big(f_i(B_\zeta\bar{x}(t)) - f_i(x)\big)\\ &\leq (1+\tau(t))\|B_\zeta\bar{x}(t) - x\|^2 - \|B_\zeta\bar{x}(t+1) - x\|^2\\ &\quad + B_\zeta^2\Big[2\alpha(t+1)^2D^2 + 2\tau(t)^2 + \tau(t)\Big]\\ &\quad + \frac{4\alpha(t+1)DB_\zeta}{m}\sum_{i=1}^m \|\hat{z}_i(t+1) - B_\zeta\bar{x}(t)\|. \end{aligned} \tag{48}$$

Rearranging this we obtain the desired estimate. □

Now we are ready to prove Theorem II.7. *Proof:* [Proof of Theorem II.7] By Lemmas IV.2 and Corollary IV.3 it is enough to prove $\sum_{t=0}^\infty (c(t) + d(t)) < \infty$, where

$$c(t) = 2\alpha(t+1)^2D^2 + 2\tau(t)^2 + \tau(t),$$

and

$$d(t) = \frac{4\alpha(t+1)DB_\zeta}{m}\sum_{i=1}^m \|\hat{z}_i(t+1) - B_\zeta\bar{x}(t)\|.$$

It follows that $\sum_{t=0}^\infty c(t) < \infty$ by Assumptions II.4 and II.5.

Next we will show that $\sum_{t=0}^\infty d(t) < \infty$. By Proposition III.3, it suffices to show that

$$\sum_{t=0}^\infty \alpha(t+1)\Big(\lambda^t + \tau(t) + \sum_{s=0}^{t-1}\lambda^{t-s-1}\alpha(s+1) + \sum_{s=0}^{t-1}\lambda^{t-s-1}\tau(s)\Big) < \infty, \tag{49}$$

and

$$\sum_{t=0}^\infty \alpha(t+1)\Big[K(t) + \sum_{s=0}^{t-1}K(t)\Big(\alpha(s+1)D + \tau(s)\Big)\Big] < \infty.$$

The latter one is proved in Lemma IV.4 below. We proceed to prove (49). Using the Cauchy-Schwarz inequality, we have

$$\sum_{t=0}^\infty \alpha(t+1)\lambda^t \leq \frac{1}{2}\sum_{t=0}^\infty \alpha(t+1)^2 + \frac{1}{2}\sum_{t=0}^\infty \lambda^{2t} < \infty.$$

By rearranging and using the decreasing property of $\alpha(t)$ in Assumption II.4, we find

$$\begin{aligned}\sum_{t=0}^\infty \alpha(t+1)&\sum_{s=0}^{t-1}\lambda^{t-s-1}\alpha(s+1)\\ &= \sum_{s=0}^\infty\sum_{t=s+1}^\infty \lambda^{t-s-1}\alpha(t+1)\alpha(s+1)\\ &\leq \sum_{s=0}^\infty\Big(\sum_{t=s+1}^\infty \lambda^{t-s-1}\Big)\alpha(s+1)^2\\ &= \frac{1}{1-\lambda}\sum_{s=0}^\infty \alpha(s+1)^2 < \infty. \end{aligned}$$

Similarly, due to Assumption II.5, the last term is bounded as

$$\begin{aligned}\sum_{t=0}^\infty \alpha(t+1)\sum_{s=0}^{t-1}\lambda^{t-s-1}\tau(s) &= \sum_{s=0}^\infty\sum_{s<t}\lambda^{t-s-1}\alpha(t+1)\tau(s)\\ &\leq \sum_{s=0}^\infty\sum_{s<t}\lambda^{t-s-1}\alpha(s+1)\tau(s)\\ &= \frac{1}{1-\lambda}\sum_{s=0}^\infty \alpha(s+1)\tau(s) < \infty. \end{aligned}$$

Gathering the above estimates, we find that $\sum_{t=0}^\infty (c(t) + d(t)) < \infty$. Hence by Lemma IV.2, the sequence $\{B_\zeta\bar{x}(t)\}$ converges to some solution $x^* \in X^*$. Finally, we apply Corollary III.6 to conclude that each sequence $\{\hat{z}_i(t)\}$, $i = 1, \cdots, n$, converges to the same solution x^*. The proof is done. □

Lemma IV.4. *Suppose that Assumption II.5 and Assumption II.6 hold. Then for the stepsize $\{\alpha(t)\}_{t\geq 0}$ satisfying Assumption II.4 or $\alpha(t) = 1/\sqrt{t}$, we have*

$$\sum_{t=0}^\infty \alpha(t+1)\left[K(t) + \sum_{s=0}^{t-1}K(t)\left(\alpha(s+1) + \tau(s)\right)\right] < \infty.$$

Proof: From Lemma III.1, we know that $\lim_{t\to\infty} t^{3/2}K(t) = 0$. Using this fact the summability of $\tau(s)$, it easily follows that

$$\sum_{t=0}^\infty \alpha(t+1)\Big[K(t) + \Big(\sum_{s=0}^{t-1}\tau(s)\Big)K(t)\Big] < \infty.$$

Next, for $\alpha(t)$ satisfying Assumption II.4, we observe that

$$\sum_{t=0}^{\infty}\alpha(t+1)K(t)\sum_{s=0}^{t-1}\alpha(s+1)\leq \sum_{t=0}^{\infty}K(t)\sum_{s=0}^{t-1}\alpha(s+1)^2<\infty.$$

For $\alpha(t) = 1/\sqrt{t}$, by using that $\sum_{s=0}^{t-1}1/\sqrt{s+1} \leq 2\sqrt{t}$, we deduce

$$\sum_{t=0}^{\infty}\alpha(t+1)K(t)\sum_{s=0}^{t-1}\alpha(s+1)\leq 2\sum_{t=0}^{\infty}K(t)<\infty.$$

The proof is done. □

B. PROOF OF THEOREM II.8

We now turn to the proof of Theorem II.8. we recall that

$$H(-1)=1 \quad \text{and} \quad H(t)=\prod_{k=0}^{t}(1+\tau(k)) \quad \text{for} \quad t\geq 0$$

and

$$S(0)=0 \quad \text{and} \quad S(t)=\sum_{s=0}^{t-1}\frac{\alpha(s+1)}{H(s)} \quad \text{for} \quad t\geq 1.$$

First we find the boundedness of $H(t)$ and $S(t)$.

Lemma IV.5. *Let $\alpha(t) = \frac{1}{\sqrt{t}}$ and Assumptions II.5 and II.6 hold. Then we have*

$$\sup_{t\geq 0}H(t)<e^{E_\tau} \quad \text{and} \quad S(t)\geq e^{-E_\tau}\sqrt{t}, \tag{50}$$

where $E_\tau = \sum_{t=0}^{\infty}\tau(t) < \infty$.

Proof: By applying the inequality $1+x \leq e^x$ for $x \geq 0$ we estimate $H(t)$ as

$$\sup_{t\geq 0}H(t)<\exp\Big(\sum_{t=0}^{\infty}\tau(t)\Big)=e^{E_\tau}.$$

Using this inequality, we deduce the following estimate

$$S(t)=\sum_{s=0}^{t-1}\frac{\alpha(s+1)}{H(s)}\geq e^{-E_\tau}\sum_{s=1}^{t}\frac{1}{\sqrt{s}}\geq e^{-E_\tau}\sqrt{t} \quad \forall\, t\geq 1.$$

The proof is done. □

To prove Theorem II.8, we first modify Lemma IV.1 which states the boundedness of $\sum_{i=1}^m\big(f_i(\bar{x}(t))-f_i(x)\big)$, by replacing $\bar{x}(t)$ to $\Big(\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}B_\zeta\bar{x}(t)\Big)/S(T+1)$.

Lemma IV.6. *Suppose that all the conditions are same as in Theorem II.8. Then we have*

$$f\left(\frac{\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}B_\zeta\bar{x}(t)}{S(T+1)}\right)-f(x) \leq \frac{me^{E_\tau}}{2\sqrt{T+1}}J_1(T)+\frac{2mDe^{E_\tau}}{\delta\sqrt{T+1}}J_2(T)+\frac{2mDe^{E_\tau}}{\delta\sqrt{T+1}}J_3(T)$$

for any $T\in\mathbb{N}$, where $J_1(T)$, $J_2(T)$, and $J_3(T)$ are defined in Theorem II.8

Proof: We recall from (51) the following inequality

$$\begin{aligned}\sum_{i=1}^m&\big(f_i(B_\zeta\bar{x}(t))-f_i(x)\big)\\ &\leq \frac{m}{2\alpha(t+1)B_\zeta}(1+\tau(t))\|B_\zeta\bar{x}(t)-x\|^2\\ &\quad-\frac{m}{2\alpha(t+1)B_\zeta}\|B_\zeta\bar{x}(t+1)-x\|^2\\ &\quad+\frac{mB_\zeta}{2\alpha(t+1)}\big(2\alpha(t+1)^2D^2+2\tau(t)^2\big)\\ &\quad+\frac{B_\zeta m\tau(t)}{2\alpha(t+1)}+2D\sum_{i=1}^m\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|. \end{aligned}\tag{51}$$

Dividing both sides of (51) by $\frac{mH(t)}{2\alpha(t+1)}$, we get

$$\begin{aligned}\frac{2\alpha(t+1)}{mH(t)}&\sum_{i=1}^m\big(f_i(B_\zeta\bar{x}(t))-f_i(x)\big)\\ &\leq \frac{\|B_\zeta\bar{x}(t)-x\|^2}{B_\zeta H(t-1)}-\frac{\|B_\zeta\bar{x}(t+1)-x\|^2}{B_\zeta H(t)}\\ &\quad+\frac{B_\zeta}{H(t)}\Big(2\alpha(t+1)^2D^2+2\tau(t)^2+\tau(t)\Big)\\ &\quad+\frac{4\alpha(t+1)D}{mH(t)}\sum_{i=1}^m\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|. \end{aligned}\tag{52}$$

Summing this from $t=0$ to $t=T$ we obtain

$$\begin{aligned}\sum_{t=0}^{T}&\left[\frac{2\alpha(t+1)}{mH(t)}\sum_{i=1}^m\big(f_i(B_\zeta\bar{x}(t))-f_i(x)\big)\right]\\ &\leq \frac{\|\bar{x}(0)-x\|^2}{B_\zeta H(-1)}-\frac{\|\bar{x}(T+1)-x\|^2}{B_\zeta H(T)}\\ &\quad+\sum_{t=0}^{T}\frac{B_\zeta}{H(t)}\Big(2\alpha(t+1)^2D^2+2\tau(t)^2+\tau(t)\Big)\\ &\quad+\sum_{t=0}^{T}\frac{4\alpha(t+1)D}{mH(t)}\sum_{i=1}^m\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|. \end{aligned}\tag{53}$$

This, together with the fact that $H(-1) = 1$ and $H(t) \geq 1$, gives

$$\begin{aligned}\sum_{t=0}^{T}&\left[\frac{2\alpha(t+1)}{mH(t)}\sum_{i=1}^m\big(f_i(B_\zeta\bar{x}(t))-f_i(x)\big)\right]\\ &\leq \frac{\|\bar{x}(0)-x\|^2}{B_\zeta}+\sum_{t=0}^{T}\Big(2\alpha(t+1)^2D^2+2\tau(t)^2+\tau(t)\Big)B_\zeta\\ &\quad+\sum_{t=0}^{T}\left(\frac{4\alpha(t+1)D}{m}\sum_{i=1}^m\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|\right). \end{aligned}\tag{54}$$

We find

$$\sum_{t=0}^{T}2\alpha(t+1)^2D^2=2D^2\sum_{t=1}^{T+1}\frac{1}{t}\leq 2D^2\Big(1+\ln(T+1)\Big), \tag{55}$$

and by definition (13) we have

$$\sum_{t=0}^{T}\Big(2\tau(t)^2+\tau(t)\Big)=2E_{\tau,2}(T)+E_\tau(T). \tag{56}$$

Finally we estimate the last term of the right hand side of (54) using Corollary III.7 as follows:

$$\begin{aligned}&\sum_{t=0}^{T}\Big(\frac{4\alpha(t+1)D}{m}\sum_{i=1}^{m}\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|\Big)\\&\le 4D\Big(\frac{C_0}{\delta(1-\lambda)}\|x(0)\|_1+\frac{4mC_0E_\tau(T)}{\delta(1-\lambda)}+\frac{C_0mD}{\delta(1-\lambda)}(1+\ln(T))\Big)\\&+\frac{4D}{\delta}\sum_{t=0}^{T}K(t)\alpha(t+1)\Big[\|x(0)\|_1+\sum_{s=0}^{t-1}(\alpha(s+1)D+\tau(s))\Big].\end{aligned}$$

Putting this estimate, (55) and (56) in (54), we achieve the following estimate

$$\begin{aligned}&\sum_{t=0}^{T}\Big[\frac{2\alpha(t+1)}{mH(t)}\sum_{i=1}^{m}\Big(f_i(B_\zeta\bar{x}(t))-f_i(x)\Big)\Big]\\&\le J_1(T)+\frac{4D}{\delta}J_2(T)+\frac{4D}{\delta}J_3(T).\end{aligned}$$

Now we set $S(T)=\sum_{t=0}^{T-1}\frac{\alpha(t+1)}{H(t)}$ for $T\in\mathbb{N}$ and divide the both sides by $\frac{2S(T+1)}{m}$. Then we apply the convexity of f_i in the left hand side and use the lower bound $S(T+1)\ge e^{-E_\tau}\sqrt{T+1}$ to the right hand side, which leads to

$$\begin{aligned}&f\Big(\frac{\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}B_\zeta\bar{x}(t)}{S(T+1)}\Big)-f(x)\\&\le\frac{me^{E_\tau}}{2\sqrt{T+1}}J_1(T)+\frac{2mDe^{E_\tau}}{\delta\sqrt{T+1}}J_2(T)+\frac{2mDe^{E_\tau}}{\delta\sqrt{T+1}}J_3(T).\end{aligned}$$

The proof is finished. □

Now we are ready to give the proof of Theorem II.8.

Proof: [Proof of Theorem II.8] Using the definition of $\tilde{z}_i$ and Assumption 2.1, we find

$$\begin{aligned}&f(\tilde{z}_i(T+1))-f\Big(\frac{\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}B_\zeta\bar{x}(t)}{S(T+1)}\Big)\\&=f\Big(\frac{\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}\hat{z}_i(t+1)}{S(T+1)}\Big)-f\Big(\frac{\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}B_\zeta\bar{x}(t)}{S(T+1)}\Big)\\&\le\frac{mD}{S(T+1)}\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}\|\hat{z}_i(t+1)-B_\zeta\bar{x}(t)\|.\end{aligned}$$

Then by Corollary III.7 with the fact that $H(t)\ge 1$ and (50), we have

$$\begin{aligned}&f(\tilde{z}_i(T+1))-f\Big(\frac{\sum_{t=0}^{T}\frac{\alpha(t+1)}{H(t)}B_\zeta\bar{x}(t)}{S(T+1)}\Big)\\&\le\frac{mDe^{E_\tau}}{\delta\sqrt{T+1}}J_2(T)+\frac{mDe^{E_\tau}}{\delta\sqrt{T+1}}J_3(T).\end{aligned}$$

Combining this inequality with Lemma IV.6, we obtain

$$\begin{aligned}&f(\tilde{z}_i(T+1))-f(x^*)\\&\le\frac{me^{E_\tau}}{2\sqrt{T+1}}J_1(T)+\frac{3mDe^{E_\tau}}{\delta\sqrt{T+1}}J_2(T)+\frac{3mDe^{E_\tau}}{\delta\sqrt{T+1}}J_3(T).\end{aligned}$$

which is the desired estimate. Moreover, we see that the right hand side is bounded by $O(\log(T+1)/\sqrt{T+1})$ using Lemma IV.4. □

V. SIMULATIONS

In this section, we present simulation results of the proposed event-triggered gradient-push method to demonstrate that the theoretical results can be realized in practice.

Example 1 (Least square solution): We consider the decentralized least squares problem:

$$\min_{x\in\mathbb{R}^d}\sum_{i=1}^{m}f_i(x)\quad\text{with}\quad f_i(x)=\|q_i-p_i^Tx\|^2,$$

where, each agent i in $\mathcal{V}=\{1,\cdots,m\}$ is given the local cost function f_i. The variable $p_i\in\mathbb{R}^{d\times p}$ is the input data and the variable $q_i\in\mathbb{R}^p$ is the output data. This type of problem arises in various areas containing machine learning and signal processing. The data are generated according to the linear regression model $q_i=p_i^T\tilde{x}+\varepsilon_i$ where $\tilde{x}\in\mathbb{R}^d$ is the true weight vector and $\varepsilon_i\in\mathbb{R}^p$ is the noise. We generate $\tilde{x}$ and p_i in the way that the value of each component is randomly chosen in [0, 1] with uniform distribution. In addition, the component values of the noise $\varepsilon_i\in\mathbb{R}^p$ are jointly Gaussian with zero mean and variance 1. The initial points $x_i(0)$ are independent random variables, generated by a standard Gaussian distribution. In this simulation, we set the problem dimensions and the number of agents as $d=5$, $p=1$, and $m=50$. We use connected directed graph where every node has four out-neighbors.

Test 1: Here we fix $\alpha(t)=1/t^{0.52}$ which satisfies Assumption II.4 and $\zeta(t)=1/(3t^3)$ which satisfies Assumption II.6, and consider various choices of $\tau(t)$. We measure the relative distance between the variable $z_i(t)$ and the optimal point x^* the value

$$R_d(t)=\frac{\sum_{i=1}^{m}\|z_i(t)-x^*\|}{\sum_{i=1}^{m}\|z_i(0)-x^*\|}, \tag{57}$$

We set the termination time k_f as the first time $k\in\mathbb{N}$ when $R_d(k)<10^{-2}$. And we let N_x and N_y be the average of total number of triggers for all agents until the termination time associated with $\tau(t)$ and $\zeta(t)$, respectively. Table 1 indicates the average of those values depending on $\tau(t)$ and $\zeta(t)$ in 100 trials.

We first look at the effect of $\zeta(t)$, the threshold for variables $y_i(t)$. Table 1 shows that an existence of the threshold ($\zeta\neq 0$) does not bring a big difference in the termination time if we compare the cases $\zeta(t)=0$ and $\zeta(t)=1/(3t^3)$ with same $\tau(t)$, but there is a big improvement in the number of triggers for $y_i(t)$. Next we discuss the values N_x and k_f of Table 1 in terms of $\tau(t)$, the threshold for variables $x_i(t)$. As in Table 1, some cases give us similar or worse results compared to the cases $\tau(t)=0$. For $\tau(t)=1/t^{1.1}$, the number of triggers is decreased by more than 70%, and the

TABLE 1. **The number of triggers and termination time depending on $\tau(t)$ and $\zeta(t)$. Here the case $\tau(t) = 0$ and $\zeta(t) = 0$ corresponds to the gradient-push algorithm [17] not involving the event-triggered communications.**

$\tau(t)$ $\zeta(t)$	0 0	0 $1/(3t^3)$	$1/t^{1.1}$ 0	$1/t^{1.1}$ $1/(3t^3)$	$1/t^{1.3}$ 0	$1/t^{1.3}$ $1/(3t^3)$	$1/t^{1.5}$ 0	$1/t^{1.5}$ $1/(3t^3)$	$1/t^{1.7}$ 0	$1/t^{1.7}$ $1/(3t^3)$
N_x	11425	11305	3125	3152	3299	3288	8860	8767	15537	11177
N_y	11425	26	72705	32	15696	27	11644	26	15572	26
k_f	11425	11305	72705	72757	15696	15572	11644	11514	15572	11207

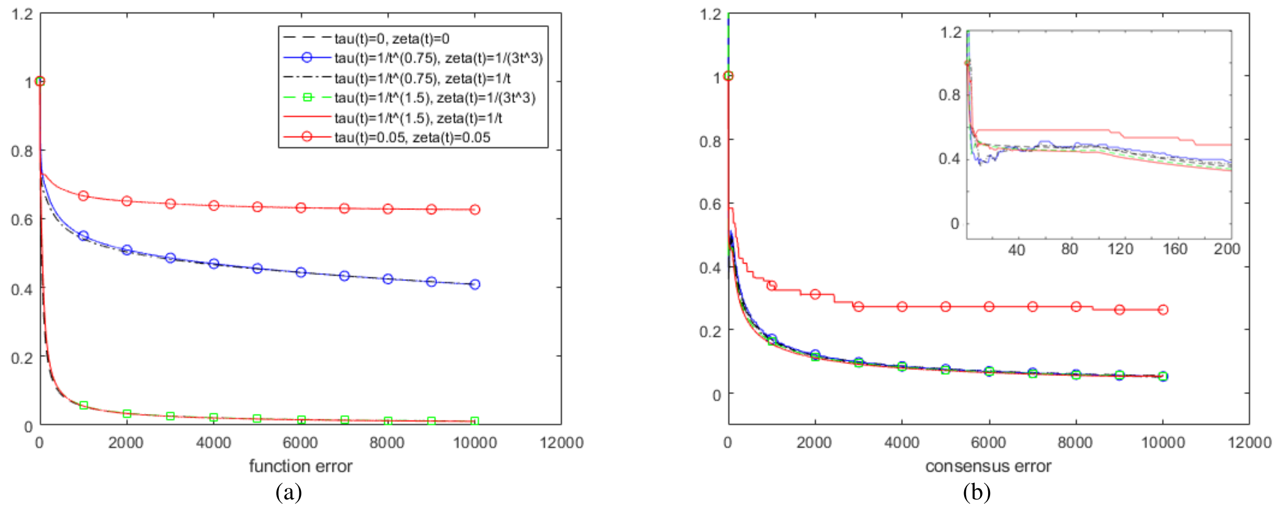


FIGURE 1. **(a) The values of $R_f(t)$ for Test 2 with different choices of $\tau(t)$ and $\zeta(t)$. (b) The values of $R_c(t)$ for Test 2 with different choices of $\tau(t)$ and $\zeta(t)$. The choice of $\tau(t)$ and $\zeta(t)$ for each graph of (b) corresponds to that of (a) with the same color and style.**

TABLE 2. **The exact values of $R_f(t)$ in Figure 1a with different choices of $\tau(t)$ and $\zeta(t)$ for every 1250 steps. The numbers in bold are the numbers of steps t. Here the case $\tau(t) = 0$ and $\zeta(t) = 0$ corresponds to the gradient-push algorithm [17] not involving the event-triggered communications.**

$\tau(t)$	$\zeta(t)$	**0**	**1250**	**2500**	**3750**	**5000**	**6250**	**7500**	**8750**	**10000**
0	0	1	0.0468	0.0295	0.0224	0.0185	0.0158	0.0140	0.0126	0.0115
$1/t^{0.75}$	$1/(3t^3)$	1	0.5367	0.04966	0.4729	0.4554	0.4414	0.4291	0.4186	0.4089
$1/t^{0.75}$	$1/t$	1	0.5288	0.4916	0.4700	0.4540	0.4409	0.4293	0.4190	0.4096
$1/t^{1.5}$	$1/(3t^3)$	1	0.0482	0.0298	0.0225	0.0184	0.0158	0.0139	0.0125	0.0114
$1/t^{1.5}$	$1/t$	1	0.0472	0.0288	0.0217	0.0177	0.0151	0.0133	0.0119	0.0108
0.05	0.05	1	0.6611	0.6470	0.6394	0.6345	0.6314	0.6292	0.6276	0.6264

termination time increased by more than 530%. For $\tau(t) = 1/t^{1.7}$, both the termination time and the number of triggers increased by almost 36% when $\zeta(t) = 0$ and remain similar when $\zeta(t) = 1/(3t^3)$ compared to the cases $\tau(t) = 0$. For $\tau(t) = 1/t^{1.5}$, the termination time is almost same, the number of triggers decreased by more than 20%. These results show that the proposed gradient-push with event-triggered communication with proper $\tau(t)$ and $\zeta(t)$ can diminish the number of communications to achieve the convergence compared to the gradient-push algorithm without triggering. The threshold functions $\tau(t)$ and $\zeta(t)$ should be chosen carefully considering the characteristics of the given optimization problem.

Test 2: Here we fix $\alpha(t) = 1/\sqrt{t}$ and take several choices of $\tau(t)$ and $\zeta(t)$. We measure the relative cost error and the consensus error given by

$$R_f(t) = \frac{\sum_{i=1}^m (f(\tilde{z}_i(t)) - f^*)}{\sum_{i=1}^m (f(\tilde{z}_i(0)) - f^*)},$$
$$R_c(t) = \frac{\max_{i,j\in V} \|z_i(t) - z_j(t)\|}{\max_{i,j\in V} \|z_i(0) - z_j(t)\|}.$$

For $\tau(t)$, we consider two cases where $\tau(t) = 1/t^{1.5}$ satisfying Assumption II.5 and $\tau(t) = 1/t^{0.75}$ not satisfying Assumption II.5. And for $\zeta(t)$, we consider also consider two cases where $\zeta(t) = 1/(3t^3)$ satisfying Assumption II.6 and $\zeta(t) = 1/t$ not satisfying Assumption II.6. Additionally,

TABLE 3. The exact values of $R_d(t)$ in Figure 1b with different choices of $\tau(t)$ and $\zeta(t)$ for every 1250 steps. The numbers in bold are the numbers of steps t. The case $\tau(t) = 0$ and $\zeta(t) = 0$ corresponds to the gradient-push algorithm [17] without the event-triggered communications.

$\tau(t)$	$\zeta(t)$	**0**	**1250**	**2500**	**3750**	**5000**	**6250**	**7500**	**8750**	**10000**
0	0	1	0.2314	0.1657	0.1361	0.1183	0.1061	0.0970	0.0899	0.0842
$1/t^{0.75}$	$1/(3t^3)$	1	0.2216	0.1584	0.1302	0.1130	0.1001	0.0917	0.0849	0.0802
$1/t^{0.75}$	$1/t$	1	0.2152	0.1518	0.1239	0.1074	0.0961	0.0886	0.0821	0.0766
$1/t^{1.5}$	$1/(3t^3)$	1	0.2333	0.1672	0.1373	0.1193	0.1070	0.0978	0.0907	0.0849
$1/t^{1.5}$	$1/t$	1	0.2254	0.1611	0.1322	0.1149	0.1030	0.0942	0.0873	0.0818
0.05	0.05	1	0.2903	0.2903	0.2510	0.2494	0.2494	0.2494	0.2494	0.2494

we test two constant cases $\tau(t), \zeta(t) = 0$ and $\tau(t), \zeta(t) = 0.05$. Figure 1a depicts the graph of the values of $R_f(t)$ versus the iteration time. The result shows that the cost error decreases to zero when $\sum_{t=0}^{\infty} \tau(t) < \infty$ while it does not converge to zero when $\sum_{t=0}^{\infty} \tau(t) = \infty$ regardless of the choice of $\zeta(t)$. This supports the convergence result of Theorem II.8. Figure 1b illustrates the consensus error $R_c(t)$. The result shows that the consensus error decreases to zero for any choices $\tau(t)$ and $\zeta(t)$ except the case $\tau(t), \zeta(t) = 0.05$. This numerical result supports the theoretical result obtained in Corollary III.6.

Example 2 (Network localization): We consider the network localization problem where N free agents that only have estimates of their own positions and M anchor agents that have the information of their own exact positions in a global coordinate system. The goal of this problem is that each free agent achieves its own position in the global coordinate system only by communicating with its nearby neighbors and using the anchor agent's information. The communication pattern among agents is depicted by a directed (N+M)-node graph $\mathbb{G} = (\mathcal{V}, \mathcal{E})$, where each node in $\mathcal{V}$ represents each agent, and each edge $\{i.j\} \in \mathcal{E}$ means agent i can send its information to j depending on its own sensor power. To formulate this problem, we let $x_i = (x_{i1}, x_{i2}) \in \mathbb{R}^2$ be the position of the agent i in $\mathbb{G}$. Without loss of generality, the point x_i is free agent $i \in \{1, \cdots, N\}$ and the point x_j is anchor agent for $j \in \{N+1, \cdots, N+M\}$. For each $i \in \{1, \cdots, N\}$, agent i has the set of neighboring agents, denoted by N_i, and may find the barycentric coodinates p_{ij} with respect to the neighboring agents $j \in N_i$ using their relative coordinates (coordniates with center x_i).To determin p_{ij}, each agent may solve the following problem

$$\min_{\{p_{ij}\}_{j \in N_i}} \sum_{j \in N_i} p_{ij}^2 \tag{58}$$

subject to

$$\sum_{j \in N_i} p_{ij} x_j = x_i$$
$$\sum_{j \in N_i} p_{ij} = 1. \tag{59}$$

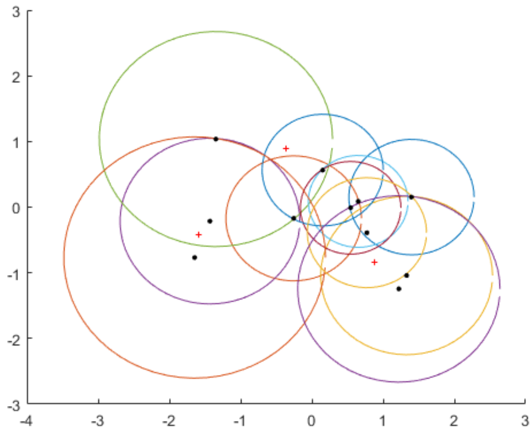


FIGURE 2. Each dot represents free agents and Each plus sigh represents known anchor agents.Each circle displays the sensor power of its corresponding free agent.

We also define p_{ij} for $j \notin N_i$ by

$$p_{ii} = -1 \quad \text{and} \quad p_{ij} = 0 \quad \text{for } j \notin N_i \cup \{i\}. \tag{60}$$

Then we have the following relation between the free agents and the anchor agents.

$$\sum_{j=1}^{N} p_{ij} x_j = q_i, \text{ for all } i \in \{1, \cdots, N\}, \tag{61}$$

where

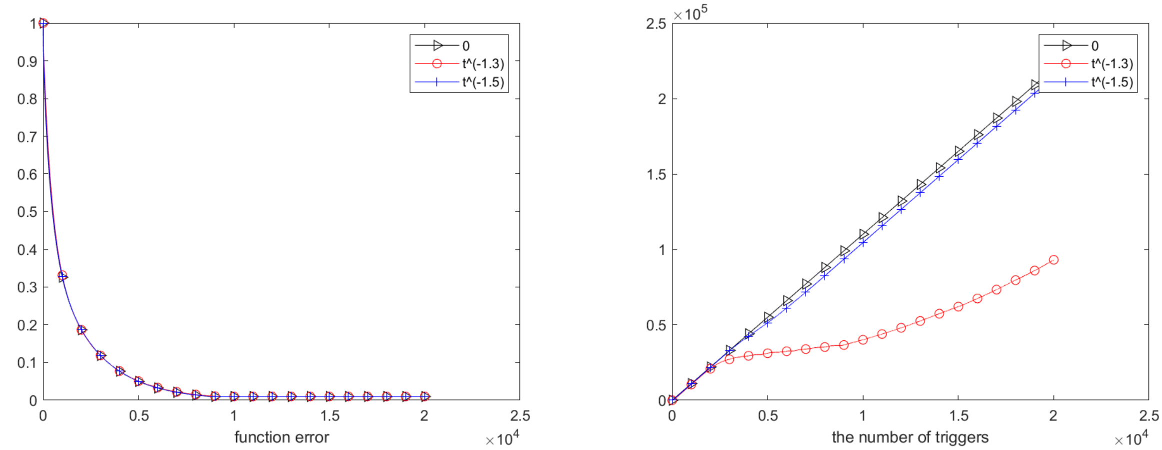
$$q_i = (q_{i1}, q_{i2}) = \sum_{j=N+1}^{N+M} a_{ij} x_j \in \mathbb{R}^2. \tag{62}$$

Let $\overline{P} = P \otimes I_2$ and $x = (x_{11}, x_{12}, x_{21}, \cdots, x_{N1}, x_{N2})^T$. Then, $x \in \mathbb{R}^{2d}$ is the solution of the following decentralized problem:

$$\min_{s \in \mathbb{R}^{2d}} \sum_{i=1}^{N} f_i(x) := \left(|q_{i1} - \overline{P}_{2i-1} s|^2 + |q_{i2} - \overline{P}_{2i}^T s|^2 \right),$$

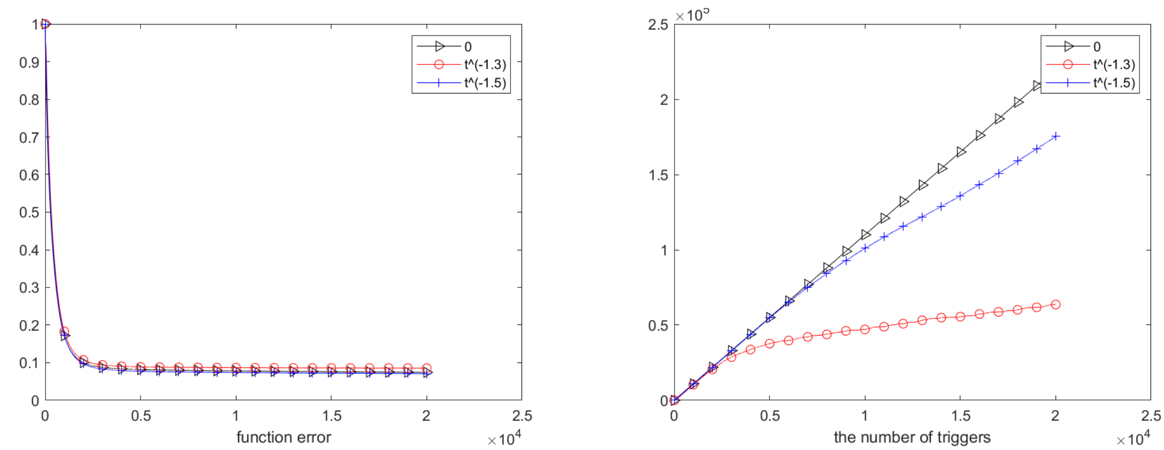
where $\overline{P}_i$ is the ith row of the $\overline{P}$. We remark that agent i has informations of $\overline{P}_{2i-1}, \overline{P}_{2i}, q_{i1}, q_{i2}$.

In our experiment, we set $N = 11$ and $M = 3$. We design the network of agents so that each free agent has a sensor



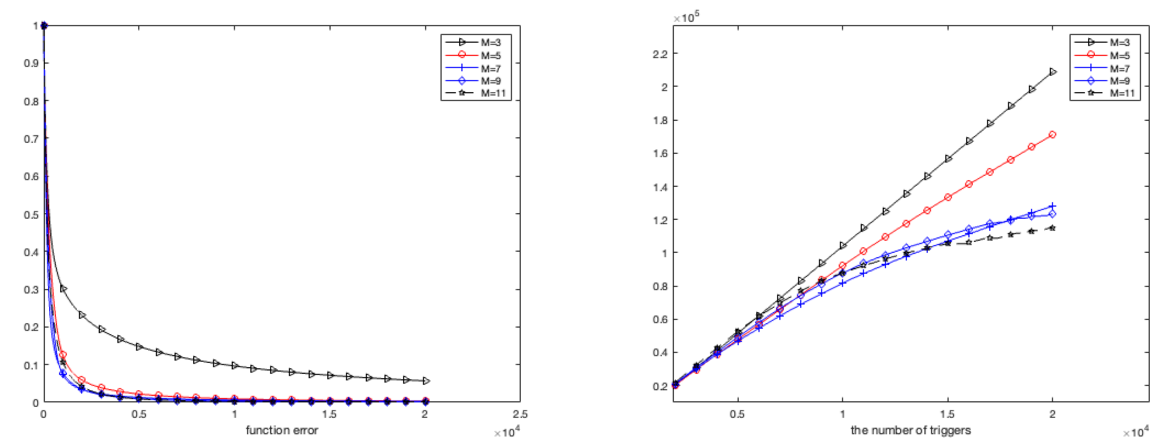
(a) (b)

FIGURE 3. (a) The values of $R_d(t)$ with different choices of $\tau(t)$ for fixed $\alpha(t) = 1/t^{(0.7)}$. (b) The values of the number of triggers with different choices of $\tau(t)$ for fixed $\alpha(t) = 1/t^{(0.7)}$.



(a) (b)

FIGURE 4. (a) The values of $R_d(t)$ with different choices of $\tau(t)$ for fixed $\alpha(t) = 1/\sqrt{t}$. (b) The values of the number of triggers with different choices of $\tau(t)$ for fixed $\alpha(t) = 1/\sqrt{t}$.



(a) (b)

FIGURE 5. We fixed $\alpha(t) = 1/t^{0.7}$, $\tau(t) = 1/t^{1.5}$ and $\zeta(t) = 0$. (a) The values of $R_d(t)$ with different choices of the number of anchor agents M. (b) The values of the number of triggers with different choices of the number of anchor agents M.

power to send its information to at least 4 free agents (See Figure 2). We choose the stepsize $\alpha(t) = 1/t^{(0.7)}$, which satisfies Assumption II.4 and $\alpha(t) = 1/\sqrt{t}$. For each stepsize, we consider various choices of $\tau(t)$ but fix $\zeta(t) = 0$. We measure $R_d(t)$ (see Figures 3a, 4a) and the number of triggers with respect to the choice of $\tau(t)$ (see Figures 3b, 4b). Comparing the cases $\tau(t) = 0$ and $\tau(t) = 1/t^{1.3}$, we find that the number of triggering times of the case $\tau(t) = 1/t^{1.3}$ is much smaller than that of the case $\tau(t) = 0$ while they have similar decay in function errors. From this result, we see that one may reduces the communication cost for resource aware scenarios.

Lastly, we look at the effect of the number of anchor agents on the performance of the algorithm. Precisely, we compare the graphs of $R_d(t)$ (see (57)) and the number of triggers for various choices of M with fixing $N = 11$ (see Figures 5a, 5b). As in the previous test, we have set $\alpha(t) = 1/t^{0.7}$, $\tau(t) = 1/t^{1.5}$ and $\zeta(t) = 0$. We perform the test for $M \in \{3, 5, 7, 9, 11\}$. In figure 3b, we observe that for the case $M = 3$ the number of triggering times is not significantly smaller than that of the time-triggered case $\tau(t) = 0$. However, if the number M is larger of equal to 7, then the number of triggers becomes notably smaller comparing to the case $M = 3$. Also, the error value $R_d(t)$ decreases faster if the number of anchor agents is larger (see figure 5a). These results imply that if there are more information of anchors that agents can access, the event-triggering strategy of gradient-push algorithm becomes more effective by reducing the power consumption for communication.

VI. CONCLUSION

In this work, we considered the gradient-push algorithm with event-triggered communication for distributed optimization problems whose agents are connected by directed graphs. We showed that by the algorithm each agent's state converges to a common minimizer under a diminishing and summability condition on the stepsize and the triggering function. Numerical simulations have been conducted to support the convergence results. It would be interesting further about how to choose an optimal triggering function for reducing the communication burden, reflecting various elements such as the connectivity of graphs and the number of agents.

REFERENCES

[1] F. Bullo, J. Cortes, and S. Martinez, *Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms*. Princeton, NJ, USA: Princeton Univ. Press, 2009.

[2] Y. Cao, W. Yu, W. Ren, and G. Chen, "An overview of recent progress in the study of distributed multi-agent coordination," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 427–438, Feb. 2013.

[3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 2508–2530, Dec. 2006.

[4] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3816–3827, Jul. 2010.

[5] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, 2018.

[6] P. A. Forero, A. Cano, and G. B. Giannakis, "Consensus-based distributed support vector machines," *J. Mach. Learn. Res.*, vol. 11, pp. 1663–1707, Jan. 2010.

[7] H. Raja and W. U. Bajwa, "Cloud K-SVD: A collaborative dictionary learning algorithm for big, distributed data," *IEEE Trans. Signal Process.*, vol. 64, no. 1, pp. 173–188, Jan. 2016.

[8] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *J. Optim. Theory Appl.*, vol. 147, no. 3, pp. 516–545, Dec. 2010.

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

[10] I.-A. Chen, "Fast distributed first-order methods," M.S. thesis, Massachusetts Inst. Technol., Cambridge, MA, USA, 2012.

[11] K. Yuan, Q. Ling, and W. Yin, "On the convergence of decentralized gradient descent," *SIAM J. Optim.*, vol. 26, no. 3, pp. 1835–1854, 2016.

[12] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: Convergence analysis and network scaling," *IEEE Trans. Autom. Control*, vol. 57, no. 3, pp. 592–606, Mar. 2012.

[13] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization with coupling constraints," *Automatica*, vol. 84, pp. 149–158, Oct. 2017.

[14] A. Simonetto and H. Jamali-Rad, "Primal recovery from consensus-based dual decomposition for distributed convex optimization," *J. Optim. Theory Appl.*, vol. 168, no. 1, pp. 172–197, Jan. 2016.

[15] M. Maros and J. Jalden, "On the Q-linear convergence of distributed generalized ADMM under non-strongly convex function components," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 5, no. 3, pp. 442–453, Sep. 2019.

[16] W. Shi, Q. Ling, K. Yuan, G. Wu, and W. Yin, "On the linear convergence of the ADMM in decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1750–1761, Apr. 2014.

[17] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[18] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th Annu. IEEE Symp. Found. Comput. Sci.*, Washington, DC, USA, Oct. 2003, pp. 482–491.

[19] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proc. IEEE 51st IEEE Conf. Decis. Control (CDC)*, Maui, HI, USA, Dec. 2012, pp. 5453–5458.

[20] A. Nedić and A. Olshevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3936–3947, Dec. 2016.

[21] M. Akbari, B. Gharesifard, and T. Linder, "Distributed online convex optimization on time-varying directed graphs," *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 3, pp. 417–428, Sep. 2017.

[22] M. Assran, N. Loizou, N. Ballas, and M. Rabbat, "Stochastic gradient push for distributed deep learning," in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds. Long Beach, CA, USA, Jun. 2019, pp. 344–353.

[23] H. Taheri, A. Mokhtari, H. Hassani, and R. Pedarsani, "Quantized decentralized stochastic learning over directed graphs," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 9324–9333.

[24] M. S. Assran and M. G. Rabbat, "Asynchronous gradient push," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 168–183, Jan. 2021.

[25] J. Zhang and K. You, "AsySPA: An exact asynchronous algorithm for convex optimization over digraphs," *IEEE Trans. Autom. Control*, vol. 65, no. 6, pp. 2494–2509, Jun. 2020.

[26] M. Meng, X. Li, and G. Xiao, "Distributed estimation under sensor attacks: Linear and nonlinear measurement models," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 7, pp. 156–165, 2021.

[27] N. Ravi and A. Scaglione, "Detection and isolation of adversaries in decentralized optimization for non-strongly convex objectives," *IFAC-PapersOnLine*, vol. 52, no. 20, pp. 381–386, 2019.

[28] V. Shnayder, M. Hempstead, B.-R. Chen, G. W. Allen, and M. Welsh, "Simulating the power consumption of large-scale sensor network applications," in *Proc. 2nd Int. Conf. Embedded Networked Sensor Syst.*, 2004, pp. 188–200.

[29] S. Appadwedula, V. V. Veeravalli, and D. L. Jones, "Decentralized detection with censoring sensors," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1362–1373, Apr. 2008.

[30] P. Addesso, S. Marano, and V. Matta, "Sequential sampling in sensor networks for detection with censoring nodes," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5497–5505, Nov. 2007.
[31] M. Mazo and P. Tabuada, "Decentralized event-triggered control over wireless sensor/actuator networks," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2456–2461, Oct. 2011.
[32] D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson, "Distributed event-triggered control for multi-agent systems," *IEEE Trans. Autom. Control*, vol. 57, no. 5, pp. 1291–1297, May 2012.
[33] Y. Kajiyama, N. Hayashi, and S. Takai, "Distributed subgradient method with edge-based event-triggered communication," *IEEE Trans. Autom. Control*, vol. 63, no. 7, pp. 2248–2255, Jul. 2018.
[34] Q. Lü and H. Li, "Event-triggered discrete-time distributed consensus optimization over time-varying graphs," *Complexity*, vol. 2017, pp. 1–12, May 2017.
[35] M. Meinel, M. Ulbrich, and S. Albrecht, "A class of distributed optimization methods with event-triggered communication," *Comput. Optim. Appl.*, vol. 57, no. 3, pp. 517–553, Oct. 2013.
[36] T. Wu, K. Yuan, Q. Ling, W. Yin, and A. H. Sayed, "Decentralized consensus optimization with asynchrony and delays," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 4, no. 2, pp. 293–307, Jun. 2018.
[37] M. Zhong and C. G. Cassandras, "Asynchronous distributed optimization with event-driven communication," *IEEE Trans. Autom. Control*, vol. 55, no. 12, pp. 2735–2750, Dec. 2010.
[38] R. Li and X. Mu, "Distributed event-triggered subgradient method for convex optimization with general step-size," *IEEE Access*, vol. 8, pp. 14253–14264, 2020.
[39] C. Liu, H. Li, Y. Shi, and D. Xu, "Distributed event-triggered gradient method for constrained convex minimization," *IEEE Trans. Autom. Control*, vol. 65, no. 2, pp. 778–785, Feb. 2020.
[40] X. Cao and T. Basar, "Decentralized online convex optimization with event-triggered communications," *IEEE Trans. Signal Process.*, vol. 69, pp. 284–299, 2021.
[41] M. Xiong, B. Zhang, D. W. C. Ho, D. Yuan, and S. Xu, "Event-triggered distributed stochastic mirror descent for convex optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 4, 2022, doi: 10.1109/TNNLS.2021.3137010.
[42] X. He, Y. Xing, J. Wu, and K. H. Johansson, "Event-triggered distributed estimation with decaying communication rate," *SIAM J. Control Optim.*, vol. 60, no. 2, pp. 992–1017, Apr. 2022.
[43] C. Wang, S. Xu, D. Yuan, B. Zhang, and Z. Zhang, "Push-sum distributed online optimization with bandit feedback," *IEEE Trans. Cybern.*, vol. 52, no. 4, pp. 2263–2273, Apr. 2022.

JIMYEONG KIM received the B.S. degree in mathematics from Sungkyunkwan University, Suwon, South Korea, in 2018, where he is currently pursuing the Ph.D. degree in mathematics. His research interests include optimization, control theory, and reinforcement learning.

WOOCHEOL CHOI received the B.Sc. and Ph.D. degrees in mathematics from Seoul National University, Seoul, South Korea, in 2009 and 2015, respectively. He is currently an Assistant Professor with the Department of Mathematics, Sungkyunkwan University, Suwon, South Korea. His research interests include distributed control, optimization, optimal control, asymptotic analysis, and numerical analysis of PDEs.