Asynchronous Broadcast-Based Convex Optimization over a Network

Angelia Nedić

Abstract

We consider a distributed multi-agent network system where each agent has its own convex objective function, which can be evaluated with stochastic errors. The problem consists of minimizing the sum of the agent functions over a commonly known constraint set, but without a central coordinator and without agents sharing the explicit form of their objectives. We propose an asynchronous broadcast-based algorithm where the communications over the network are subject to random link failures. We investigate the convergence properties of the algorithm for a diminishing (random) stepsize and a constant stepsize, where each agent chooses its own stepsize independently of the other agents. Under some standard conditions on the gradient errors, we establish almost sure convergence of the method to an optimal point for diminishing stepsize. For constant stepsize, we establish some error bounds on the expected distance from the optimal point and the expected function value. We also provide numerical results.

I. INTRODUCTION

The problem of minimizing of a sum of convex functions when each component function is available (with stochastic errors) to a specific network agent is an important problem in the context of wired and wireless networks [13], [14], [24], [28], [29]. These problems require the design of optimization algorithms that are distributed and asynchronous, i.e., without a central coordinator and without synchronized actions. Furthermore, the algorithms should be obeying the local network structure in the sense that each agent can only use its own objective function and can exchange some limited information with its local neighbors only.

In this paper, we propose an *asynchronous distributed algorithm* that uses the random broadcast scheme [1] as a mechanism to distribute the operations over the network in an

asynchronous manner. We investigate the convergence properties of the algorithm for a diminishing stepsize and provide error bounds for a constant stepsize. We also provide a numerical example. An asynchronous algorithm based on a random gossip scheme has been proposed in [25] and analyzed for a scalar objective function and a diminishing stepsize. The gossip scheme used in [25] requires establishing bidirectional communication links between agent pairs and assumes reliable communications. However, wireless media is inherently broadcast and establishing bidirectional communications creates a bottleneck. The algorithm proposed in this present paper removes the need for bidirectional communications by using a different scheme, namely random broadcast, to spread the information in the network. Furthermore, unlike [25], we allow link failures to randomly occur in the agent communications. Additionally, unlike [25], we also study a constant stepsize and provide error bounds on the algorithm's performance.

Another asynchronous algorithm for optimization over networks is the Markov incremental algorithm proposed in [13], [14] and further studied in [27]. This algorithm maintains a single iterate sequence that is incrementally updated by one agent at the time. At a given time, the agent (that has just updated the iterate) passes the iterate to a randomly selected neighbor that performs the next update. The random selection of the neighbor is driven by a Markov chain. As such the Markov incremental algorithm is very different from the broadcast-based algorithm proposed in this paper, where each agent generates its own iterate sequence. Furthermore, Markov incremental algorithm is not well adapted to wireless media where the broadcasted information is available to all agents that can hear the broadcast.

Distributed consensus-based algorithms have been studied in [19], [18], [21], [26], [16], [20], which rely on deterministic consensus schemes, except for [16] where a random consensus scheme is considered. These algorithms have the following limitations in common: (i) the algorithms are synchronous as all agents in the network update at the same time; (ii) all agents in the network use the same stepsize value; (iii) the communication links are always reliable (except for [16] where random link failures are allowed); (iv) agents always communicate over bidirectional links. The broadcast-based algorithm discussed in this paper overcomes all these limitations. In particular, through the use of a random broadcast, the algorithm is totally asynchronous and allows a group of agents to update while the other agents do not. The agents neither coordinate the stepsize values nor communicate

bidirectionally. The communication links are not always reliable as they may randomly fail.

A very different distributed computational model has been proposed in [30] and also discussed in [3], [31], [6], where the network objective function is available to each agent and the aim is to distribute the processing by allowing each agent to update only a part of the decision vector. Since we are dealing with a random broadcast scheme for consensus, on a broader scale our work in this paper is related to the literature on distributed consensus and averaging [30], [31], [3], [12], [22], [32], [15], [18], [21]. Also, since we are considering (sub)gradient methods with stochastic errors, on a broader basis, our work is also related to stochastic optimization literature [10], [11], [4], [7].

The contribution of our work in this paper is mainly in three directions. First, we use broadcast model that allows for uncertainties in agent communications by accounting for possible link failures. Second, our algorithm is asynchronous and allows the agents to use uncoordinated stepsize values. Third, we study the convergence of the algorithm for a diminishing (random) stepsize and provide error bounds for a constant stepsize. The algorithm uses stochastic subgradients with a general model for the subgradient errors.

The rest of the paper is organized as follows. In Section II, we describe the problem of our interest, present our algorithm and assumptions. In Section III, among other preliminaries, we investigate the asymptotic properties of the agent disagreements. In Section IV, the convergence properties of the algorithm are studied for a (stochastic) diminishing stepsize, while in Section V we provide error bounds for a (deterministic) constant stepsize. We provide some numerical results in Section VI and conclude with a summary in Section VII.

Notation. All vectors are viewed as column vectors, and $\|x\|$ denotes the Euclidean norm of a vector x. For a matrix A, the norm $\|A\|$ is the matrix norm induced by the Euclidean vector norm (i.e., $\|A\|$ is the spectral norm given by $\|A\| = \max_{\|x\|=1} \|Ax\|$). We write x^T and A^T to denote the transpose of a vector x and a matrix A, respectively. We use x_i or $[x]_i$ to denote the i-th component of a vector x. Similarly, for a matrix A, we write $[A]_{ij}$ or A_{ij} (or a_{ij}) to indicate its (i,j)-th component. We use $\mathbf{1}$ to denote the vector in $\mathbb{R}^m$ with all entries equal to 1. The matrix A is *stochastic* when $A_{ij} \geq 0$ for all i,j and $A\mathbf{1} = \mathbf{1}$; while it is doubly stochastic if both A and A^T are stochastic. The cardinality of a set S with finitely many elements is denoted by $|S|$. We use $\mathsf{E}[X]$ to denote the expected value of a random variable X, and χ_E to denote the indicator function of a random event E.

II. PROBLEM, ALGORITHM AND ASSUMPTIONS

We consider a network of m agents that are indexed by $1, \ldots, m$; when convenient, we denote the agent set by $V = \{1, \ldots, m\}$. We assume that the agents communicate over a network with a static topology represented by an undirected graph $(V, \mathscr{E})$, where $\mathscr{E}$ is the set of undirected links $\{i, j\}$. There are no self-loops in the graph ($\{i, i\} \notin \mathscr{E}$), and we have $\{i, j\} \in \mathscr{E}$ only if agents i and j can communicate with each other. We assume a wireless communication medium where agents broadcast information over the network with possible link failures. We are interested in solving the following problem over the network:

$$
\begin{aligned}
&\text{minimize} && f(x) \triangleq \sum_{i=1}^{m} f_i(x) \\
&\text{subject to} && x \in X,
\end{aligned}
\tag{1}
$$

where X is a subset of $\mathbb{R}^n$ and each f_i is a function defined over the set X. The precise assumptions on the set X and the functions f_i will be stated in Section II-B. Problem (1) is to be solved under the following restrictions on the network information: (1) Each agent i knows only its own objective function f_i and it can compute the (sub)gradients ∇f_i with stochastic errors; and (2) Each agent communicates and exchanges some information with its local neighbors only. To solve the problem, we consider *an asynchronous distributed algorithm* based on the broadcast consensus model of [1].

As a motivating example[1] for the problem, we next discuss a distributed regression problem over wireless sensor network [27]. Suppose m sensors are deployed to sense a time invariant spatial field. Let $r_{i,k}$ be the measurement made by i^{th} sensor at time slot k. Let s_i be the location of the ith sensor. For each sensor i, let $h(s_i; x)$ be a set of candidate models for the spatial field that are selected based on a priori information and parameterized by x. Thus, for each x, the candidate $h(s_i, x)$ is a model for the measurement $r_{i,k}$. The problem in regression is to choose the best model among the set of candidate models based on the collected measurements $r_{i,k}$, i.e., to determine the value for x that best describes the spatial field. In least squares regression, when the measurements $r_{i,k}$ are corrupted by i.i.d. noise, the parameter value x^* corresponding to the best model satisfies the following relation: $x^* \in Argmin_{x \in X} \lim_{N \to \infty} \sum_{i=1}^{m} \frac{1}{N} \sum_{k=1}^{N} \left(r_{i,k} - h(s_i, x) \right)^2$. Under the i.i.d. noise assumption for the measurements $r_{i,k}$, the preceding limit exists and it is deterministic, leading to

---

[1]See [27] for another motivating example on distributed resource allocation over wireless network.
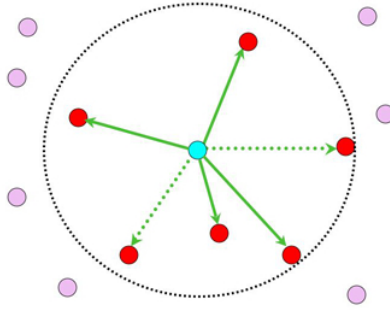
Fig. 1. The node in the center of the circle is broadcasting. Solid links point to the neighbors that can hear the broadcast. The dotted links point to the neighbors that do not hear the broadcast due to link failures. The other agents are out of range and cannot hear the broadcast.

an equivalent problem of the following form: $x^* \in Argmin_{x \in X} \sum_{i=1}^{m} \mathsf{E}\left[(R_i - h(s_i, x))^2\right]$. In linear least squares regression, the models $h(s_i, x)$, $i = 1, \ldots, m$, are linear in x and, hence, each of the functions $f_i(x) = \mathsf{E}\left[(R_i - h(s_i, x))^2\right]$ is convex in x.

A. Broadcast Optimization Algorithm

Given the agent connectivity graph $(V, \mathscr{E})$, we let $\mathcal{N}(i)$ denote the set of all neighbors of agent i, i.e. $\mathcal{N}(i) = \{j \in V \mid \{i, j\} \in \mathscr{E}\}$. We assume that each link $\{i, j\} \in \mathscr{E}$ is subject to a random failure. Each agent has its local clock that ticks at a Poisson rate of 1 independently of the clocks of the other agents. We note that the model and the analysis can be easily extended to the case when the clocks have rates different from 1. At each tick of its clock, agent i broadcasts its current iterate to its neighbors $j \in \mathcal{N}(i)$ (see Figure 1 for an illustration of the broadcast). However, due to a possible link failure, neighbor j hears the broadcasted information with probability p_{ij} and combines its own iterate with the received iterate of agent i. Then, it adjusts its iterate along the negative (sub)gradient direction ∇f_i of its objective function, which is computed with stochastic errors.

As in [1], [8], [25], we find it easier to study the algorithm in terms of a single virtual clock that ticks whenever any of the local Poisson clock ticks. Thus, the virtual clock ticks according to a Poisson process with rate m. Let Z_k be the time of the k-th tick of the global Poisson clock. We think of time as discretized according to the intervals $[Z_{k-1}, Z_k)$, i.e., this interval corresponds to timeslot k (also referred to as time k). We let x_{k-1}^i be agent i iterate at timeslot $k - 1$. We assume that only one agent broadcasts at a time, and we let I_k be the index of the agent that broadcasts at time k. Due to possible link failure, a subset

J_k of the neighbors $\mathcal{N}(I_k)$ hear the broadcast; thus, each agent $i \in J_k$ receives the iterate $x_{k-1}^{I_k}$ from agent I_k. The new iterates are generated as follows: Agent I_k and agents i that do not receive information from I_k do not update,

$$x_k^i = x_{k-1}^i \qquad \text{for } i \notin J_k. \tag{2}$$

Agents $i \in J_k$ receive the estimate $x_{k-1}^{I_k}$ and update according to the following rule:

$$v_k^i = \beta x_{k-1}^{I_k} + (1 - \beta)\, x_{k-1}^i,$$
$$x_k^i = P_X[v_k^i - \alpha_{i,k}(\nabla f_i(v_k^i) + \epsilon_k^i)], \tag{3}$$

where P_X denotes the Euclidean projection on the set X, $\nabla f_i(x)$ denotes the gradient (or subgradient) of f_i at x, and the vector ϵ_k^i is *stochastic error* in computing $\nabla f_i(v_i^k)$. The scalar $\beta \in (0, 1)$ is a *weight* and $\alpha_{i,k} > 0$ is a stepsize. The process is initialized with some initial iterates $x_0^i \in X$, $i \in V$. Each vector x_0^i is random, and it is assumed to be independent of all x_0^j for $j \neq i$ and, also, independent of all the other random variables in the process. Furthermore, it is assumed that both $\mathsf{E}[x^i(0)]$ and $\mathsf{E}[\|x^i(0)\|^2]$ are finite.

We now summarize the broadcast algorithm of Eq. (3) in phases.

Initial Phase. Each agent i has the parameter $\beta \in (0, 1)$ and an initial (random) vector $x^i \in X$. Each agent has a local clock that ticks at rate 1. All agents are initially sleeping.

Update Phase. A local clock of an agent i has ticked and the agent wakes up. Agent i broadcasts its estimate x^i and goes to sleep. Each neighbor $\ell \in \mathcal{N}(i)$ of agent i may receive x^i with probability $p_{i\ell}$. A neighbor ℓ that receives the value x^i updates its estimate, as follows:

(1) Computes $v^\ell = \beta x^i + (1 - \beta)x^\ell$ and (an approximate) gradient g of $\nabla f_i(v^i)$;

(2) Selects stepsize α_ℓ, computes $x^i = P_X[v^\ell - \alpha_i g]$, and goes to sleep.

Sleeping Phase. No updates occur in the network until another agent wakes up.

The vector g in Update Phase corresponds to an erroneous gradient $\nabla f_i(v_k^i)$ of Eq. (3). We will consider two stepsize choices: (1) A random stepsize defined by the frequency of the agent updates, $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$, where $\Gamma_k(i)$ denotes the number of updates that agent i has performed until time k, inclusively; and (2) A deterministic constant stepsize, $\alpha_{i,k} = \alpha_i > 0$ for all i and k. Observe that these stepsizes do not require any coordination of the agents.

B. Assumptions

Here, we provide our basic assumptions on the underlying agent connectivity graph $(V, \mathscr{E})$, the set X, and the agent objective functions f_i and their (sub)gradient errors ϵ_k^i. We also state some relations that are immediate consequences of our assumptions.

Assumption 1: The underlying communication graph $(V, \mathscr{E})$ is connected. The link failure process is i.i.d., whereby at any time the link $\{i, j\} \in \mathscr{E}$ is functioning with probability $p_{ij} > 0$ independently of the other links in the network.

Assumption 1 is crucial as it ensures that through the broadcast strategy the information of each agent reaches every other agent frequently enough. This frequent "mixing" of the agent information is necessary for the convergence of the agents' estimates x_k^i to a common vector. However, to ensure that the common vector solves problem (1), some assumptions are needed for the set X and the functions f_i. We use the following.

Assumption 2: The set $X \subseteq \mathbb{R}^n$ is nonempty closed and convex. Each function f_i is defined and convex over an open set containing the set X.

By Assumption 2, each f_i is continuous over the set X ([2], Proposition 1.4.6).

We do not assume differentiability of the functions f_i. At points where the gradient does not exist, we use a subgradient, which we denote also by $\nabla f_i(x)$ to keep the notation simple. A vector $\nabla g(x)$ is a *subgradient* of a function g at a point x in the domain of g (denoted by $\mathrm{dom}\, g$) if the following relation holds

$$\nabla g(x)^T(y - x) \leq g(y) - g(x) \qquad \text{for all } y \in \mathrm{dom}\, g. \tag{4}$$

By Assumption 2, for each i, a subgradient of $f_i(x)$ exists at each point $x \in X$ (see [2], Proposition 4.4.2). We make the following assumption on the subgradients.

Assumption 3: The subgradients of each function f_i are uniformly bounded over the set X, i.e., $\|\nabla f_i(x)\| \leq C$ for some $C > 0$, and for all $x \in X$ and $i \in V$.

Assumption 3 is satisfied for example when X is compact, or when each f_i is a polyhedral function (i.e., defined as a pointwise maximum of a finite collection of affine functions). The bounded subgradient assumption allows us to relate the subgradients at a given point $v \in X$ with function values at a different point $y \in X$. In particular, by subgradient property (4), it can be seen that for all $i \in V$,

$$\nabla f_i(v)^T(v - x) \geq f_i(y) - f_i(x) - C\|v - y\| \qquad \text{for any } x, y, v \in X. \tag{5}$$

We now discuss the random errors ϵ_k^i in computing the subgradients $\nabla f_i(x)^T$ at points $x = v_k^i$. When dealing with these errors, we often make use of Hölder's inequality

$$\mathsf{E}\left[\sum_{i=1}^{K} |x_i^T y_i|\right] \leq \sqrt{\sum_{i=1}^{K} \mathsf{E}[\|x_i\|^2]} \sqrt{\sum_{i=1}^{K} \mathsf{E}[\|y_i\|^2]}, \tag{6}$$

which holds for any (finite) collections $\{x_i\}_{i=1}^K$ and $\{y_i\}_{i=1}^K$ of random vectors with finite second moments ([5], page 242).

We let $\mathcal{F}_k$ be the σ-algebra generated by the entire history of the algorithm up to time k,

$$\mathcal{F}_k = \{(x_0^i, \ i \in V); I_\ell, J_\ell, (\epsilon_\ell^j, j \in J_\ell); \ 1 \leq \ell \leq k\} \qquad \text{for all } k \geq 1,$$

with $\mathcal{F}_0 = \{x_0^i, \ i \in V\}$. We use the following assumptions on the stochastic errors.

Assumption 4: There is a scalar ν such that $\mathsf{E}[\|\epsilon_k^i\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k] \leq \nu^2$ with probability 1 for all $i \in J_k$ and $k \geq 1$.

Assumption 4 is a nonsmooth analog of the linear growth condition that is used in stochastic gradient methods for differentiable functions with Lipschitz gradients [4], [7].

When X and each f_i are convex, every vector v_k^i is a convex combination of $x_k^j \in X$ (see Eq. (3)), implying that $v_k^i \in X$ for all k and i. When in addition Assumptions 3 and 4 hold, by using $\mathsf{E}[\|x\|] \leq \sqrt{\mathsf{E}[\|x\|^2]}$ we can see that

$$\mathsf{E}\left[\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \leq (C + \nu)^2 \qquad \text{for } i \in J_k \text{ and } k \geq 1. \tag{7}$$

We also use the following result of Robbins and Siegmund ([23], Chapter 2.2, Lemma 11).

Lemma 1: Let $(\Omega, F, \mathcal{P})$ be a probability space and $F_0 \subset F_1 \subset \ldots$ be a sequence of σ-subfields of F. Let $\{d_k\}$, $\{v_k\}$ and $\{w_k\}$ be F_k-measurable scalar random variables. Let $\{d_k\}$ be bounded below uniformly, and let $\{v_k\}$ and $\{w_k\}$ be nonnegative with $\sum_{k=1}^{\infty} w_k < \infty$. Also, let the following relation hold with probability 1,

$$\mathsf{E}[d_{k+1} \mid F_k] \leq (1 + q_k)d_k - v_k + w_k \qquad \text{for all } k \geq 1,$$

where $q_k \geq 0$ are deterministic scalars such that $\sum_{k=1}^{\infty} q_k < \infty$. Then, with probability 1, the sequence $\{d_k\}$ converges to some random variable and $\sum_{k=1}^{\infty} v_k < \infty$.

III. PRELIMINARIES

We discuss an alternative description of the algorithm that we extensively use later on. We also study some limiting properties of the agents' disagreements. The proofs of the results are provided in Appendix.

A. Alternative Representation of the Algorithm

We at first introduce some matrices to represent the broadcast-based algorithm in a more convenient form. We then discuss few properties of these matrices that are instrumental in the subsequent development. Define the matrix W_k for each k, as follows:

$$\text{for } i \in J_k : \qquad [W_k]_{i,i} = 1 - \beta, \qquad [W_k]_{i,I_k} = \beta, \qquad [W_k]_{i,j} = 0 \text{ else,}$$
$$\text{for } i \notin J_k : \qquad [W_k]_{i,i} = 1, \qquad [W_k]_{i,j} = 0 \text{ else.} \tag{8}$$

Now, we can write the method (2)–(3) equivalently as follows: for all $k \geq 1$ and $i \in V$,

$$v_k^i = \sum_{j=1}^{m} [W_k]_{i,j} \, x_{k-1}^j,$$

$$x_k^i = v_k^i + \left(P_X[v_k^i - \alpha_{i,k} \left(\nabla f_i(v_k^i) + \epsilon_k^i \right)] - v_k^i \right) \chi_{\{i \in J_k\}}, \tag{9}$$

where $\chi_{\{i \in J_k\}}$ is the indicator function of the event $\{i \in J_k\}$. Thus, when i updates at time k, we have $\chi_{\{i \in J_k\}} = 1$ and x_k^i is defined by the projection-based iterate adjustment. If agent i does not update at time k, we have $\chi_{\{i \in J_k\}} = 0$ and $x_k^i = v_k^i = x_{k-1}^i$.

Due to the construction and Assumption 1, the sequence $\{W_k\}$ is i.i.d. For each k, the random matrix W_k is stochastic (but not doubly stochastic), while the expected matrix $\mathsf{E}[W_k]$ is doubly stochastic. This and some other properties of the matrices W_k are stated below.

Lemma 2: The weight matrices W_k of (8) are stochastic. Moreover, we have:

(a) Under Assumption 1, the expected weight matrix $\bar{W} = \mathsf{E}[W_k]$ is doubly stochastic; in particular, it is given by $\bar{W} = I - \frac{\beta}{m} L_\Pi$, where L_Π is the weighted Laplacian of the weighted graph $(V, \mathscr{E}, \Pi)$ with Π being the probability matrix associated with the reliabilities of the links in $\mathscr{E}$, i.e., $[L_\Pi]_{ii} = \sum_{j \in \mathcal{N}(i)} p_{ij}$ for all $i \in V$, while $[L_\Pi]_{ij} = -p_{ij}$ for $j \in \mathcal{N}(i)$ and $[L_\Pi]_{ij} = 0$ otherwise.

(b) The matrices $D_k = W_k - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k$ are i.i.d. and $\lambda \triangleq \lambda_1(\mathsf{E}[D_k^T D_k]) < 1$, where $\lambda_1(\mathsf{E}[D_k^T D_k])$ denotes the largest eigenvalue of the matrix $\mathsf{E}[D_k^T D_k]$.

The following two relations are consequences of the stochasticity of W_k and the doubly stochasticity of $\mathsf{E}[W_k]$, combined with the convexity of the norm and squared norm.

$$\sum_{i=1}^{m} \mathsf{E}\left[\|v_k^i - x\|^2 \mid \mathcal{F}_{k-1}\right] \leq \sum_{j=1}^{m} \|x_{k-1}^j - x\|^2 \qquad \text{for all } x \in \mathbb{R}^n \text{ and } k \geq 1, \tag{10}$$

$$\sum_{i=1}^{m} \mathsf{E}\left[\|v_k^i - x\| \mid \mathcal{F}_{k-1}\right] \leq \sum_{j=1}^{m} \|x_{k-1}^j - x\| \qquad \text{for all } x \in \mathbb{R}^n \text{ and } k \geq 1. \tag{11}$$

B. Agent Disagreements

In this section, we provide some auxiliary results for the disagreement among the agents on values x_k^i. Our first result establishes some long term estimates for the stepsize $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$, based on a characterization of the number $\Gamma_k(i)$. This characterization is important as it allows us to remove some technical difficulties when dealing with cross-dependencies between the randomness in the stepsize and the randomness of the iterative processes $\{x_k^i\}$.

Define $E_{i,k} = \{i \in J_k\}$, which is the event that agent i updates at time k. The sequence $\{I_k\}$ is i.i.d. with uniform distribution over the set V. The link failure process is also i.i.d. Thus, for each i, the events $E_{i,k}$ are independent across time and have the same (time invariant) probability distribution. Let γ_i denote the probability of event $E_{i,k}$, and note that

$$\gamma_i = \frac{1}{m} \sum_{j \in \mathcal{N}(i)} p_{ij} \qquad \text{for all } i \in V, \tag{12}$$

where $p_{ij} > 0$ is the probability that link $\{i, j\}$ is functioning.

The following lemma gives long term estimates for the stepsize $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$.

Lemma 3: Let the graph $(V, \mathscr{E})$ have no isolated nodes and let $p_{\min} = \min_{\{i,j\} \in \mathscr{E}} p_{ij}$. Let $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$ for all k and i. Also, let q be a scalar such that $0 < q < 1/2$. Then, there exists a large enough $\tilde{k} = \tilde{k}(q, m)$ such that with probability 1 for all $k \geq \tilde{k}$ and $i \in V$,

$$\alpha_{i,k} \leq \frac{2}{k \gamma_i}, \qquad \alpha_{i,k}^2 \leq \frac{4m^2}{k^2 p_{\min}^2}, \qquad \left| \alpha_{i,k} - \frac{1}{k \gamma_i} \right| \leq \frac{2}{k^{\frac{3}{2} - q} p_{\min}^2}.$$

We now provide a lemma showing, among other relations, that the agent disagreement on the vectors x_k^i converges to zero with probability 1 when the stepsize is defined by the relative frequency of the updates, i.e., $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$.

Lemma 4: Let Assumptions 1–4 hold. Let $\{x_k^i\}$, $i \in V$, be the iterate sequences generated by algorithm (9), and define $\bar{y}_k = \frac{1}{m} \sum_{j=1}^m x_k^j$ for all $k \geq 0$. Then, the following holds:

(a) When $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$, we have $\sum_{k=1}^{\infty} \frac{1}{k} \|x_{k-1}^i - \bar{y}_{k-1}\| < \infty$ and $\lim_{k \to \infty} \|x_k^i - \bar{y}_k\| = 0$ with probability 1 for each $i = 1, \ldots, m$.

(b) When $\alpha_{i,k} = \alpha_i > 0$, we have $\limsup_{k \to \infty} \sum_{i=1}^m \mathsf{E}[\|x_k^i - \bar{y}_k\|] \leq \alpha_{\max} \frac{\sqrt{m d_{\max}}}{1 - \sqrt{\lambda}} (C + \nu)$, where $\alpha_{\max} = \max_{1 \leq j \leq m} \alpha_j$, $d_{\max} = \max_{1 \leq i \leq m} |\mathsf{E}[J(i)]|$ and $J(i)$ is the random set of agents that update given that agent i broadcasts. The scalar $\lambda < 1$ is as given in Lemma 2, while C and ν are as in Assumptions 3 and 4, respectively.

The expectation in the expression for $d_{\max}$ in part (b) of Lemma 4 is taken with respect to the link reliability probabilities. The given bound captures the dependence of the deviations

of x_k^i from their current average $\bar{y}_k$ in terms of the maximum stepsize and the communication graph topology. The impact of the topology of the communication graph $(V, \mathscr{E})$ is captured by the maximum expected number $d_{\max}$ of the agents updating at any time (under the link failure probabilities) and the spectral radius λ of the matrix $\mathsf{E}\big[D_k^T D_k\big]$.

IV. CONVERGENCE RESULTS

Here, we investigate the convergence of the method for the stepsize $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$. We start by establishing a special relation that is valid for this stepsize.

Lemma 5: Let Assumptions 1–4 hold. Let $\{x_k^i\}$, $i \in V$, be the iterate sequences generated by algorithm (9) with $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$. Then, for any $q \in (0, 1/2)$ there is a sufficiently large $\tilde{k} = \tilde{k}(q, m)$, such that we have with probability 1, for all $x \in X$, $k \geq \tilde{k}$ and $i \in J_k$,

$$\mathsf{E}\big[\|x_k^i - x\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq (1 + a_k)\|v_k^i - x\|^2 - \frac{2}{k\gamma_i}\left(f_i(\bar{y}_{k-1}) - f_i(x)\right)$$

$$+ \frac{2mC}{k\, p_{\min}} \|v_k^i - \bar{y}_{k-1}\| + \frac{2m}{k\, p_{\min}} \|\mathsf{E}\big[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k\big]\|\|v_k^i - x\| + \left(\frac{4m^2}{k^2\, p_{\min}^2} + a_k\right)(C + \nu)^2.$$

where $p_{\min} = \min_{\{i,j\} \in \mathscr{E}} p_{ij}$, $a_k = \frac{2}{k^{\frac{3}{2}-q} p_{\min}^2}$, $\bar{y}_k = \frac{1}{m}\sum_{j=1}^m x_k^j$, and γ_i is the probability that agent i updates.

Proof: Let $x \in X$ be arbitrary. From the definition of the method in (9) and the nonexpansive property of the projection operation, we have for any $k \geq 1$ and $i \in J_k$,

$$\|x_k^i - x\|^2 \leq \|v_k^i - x\|^2 - 2\alpha_{i,k}(\nabla f_i(v_k^i) + \epsilon_k^i)^T(v_k^i - x) + \alpha_{i,k}^2\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2.$$

Writing $\alpha_{i,k} = (\alpha_{i,k} - \frac{1}{k\gamma_i}) + \frac{1}{k\gamma_i}$ and using $\left|\alpha_{i,k} - \frac{1}{k\gamma_i}\right| \leq \frac{2}{k^{\frac{3}{2}-q} p_{\min}^2}$ (see Lemma 3), for $\tilde{k}$ sufficiently large, we obtain with probability 1 for all $k \geq \tilde{k}$ and $i \in J_k$,

$$\begin{aligned}
\|x_k^i - x\|^2 &\leq \|v_k^i - x\|^2 - \frac{2}{k\gamma_i}\left(\nabla f_i(v_k^i) + \epsilon_k^i\right)^T(v_k^i - x) \\
&\quad + \frac{4}{k^{\frac{3}{2}-q} p_{\min}^2}|(\nabla f_i(v_k^i) + \epsilon_k^i)^T(v_k^i - x)| + \alpha_{i,k}^2\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2.
\end{aligned}$$

Letting $a_k = \frac{2}{k^{\frac{3}{2}-q} p_{\min}^2}$ and using $2|(\nabla f_i(v_k^i) + \epsilon_k^i)^T(v_k^i - x)| \leq \|\nabla f_i(v_k^i) + \epsilon_k^i\|^2 + \|v_k^i - x\|^2$, we obtain

$$\begin{aligned}
\|x_k^i - x\|^2 &\leq (1 + a_k)\|v_k^i - x\|^2 - \frac{2}{k\gamma_i}\left(\nabla f_i(v_k^i) + \epsilon_k^i\right)^T(v_k^i - x) \\
&\quad + (\alpha_{i,k}^2 + a_k)\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2.
\end{aligned} \tag{13}$$

Since $x_k^j \in X$ for all j and k, by convexity of X it follows that $\bar{y}_{k-1} \in X$ for all k. By the subgradient boundedness (Assumption 3) relation (5) holds, so that we have

$$\nabla f_i(v_k^i)^T(v_k^i - x) \geq f_i(\bar{y}_{k-1}) - f_i(x) - C\|v_k^i - \bar{y}_{k-1}\|.$$

By substituting the preceding estimate in (13), and using $\alpha_{i,k}^2 \leq \frac{4m^2}{k^2 p_{\min}^2}$ (Lemma 3), we obtain with probability 1 for all $k \geq \tilde{k}$ and $i \in J_k$,

$$\|x_k^i - x\|^2 \leq (1 + a_k)\|v_k^i - x\|^2 - \frac{2}{k\gamma_i}\left(f_i(\bar{y}_{k-1}) - f_i(x)\right)$$
$$+ \frac{2C}{k\gamma_i}\|v_k^i - \bar{y}_{k-1}\| + \frac{2}{k\gamma_i}(\epsilon_k^i)^T(v_k^i - x) + \left(\frac{4m^2}{k^2 p_{\min}^2} + a_k\right)\left\|\nabla f(v_k^i) + \epsilon_k^i\right\|^2.$$

Taking the conditional expectation on $\mathcal{F}_{k-1}$, I_k and J_k jointly, and using the boundedness of the subgradient norm and the noise ϵ_k^i (Eq. (7)), we obtain for $k \geq \tilde{k}$ and $i \in J_k$,

$$\mathsf{E}\left[\|x_k^i - x\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \leq (1 + a_k)\|v_k^i - x\|^2 - \frac{2}{k\gamma_i}\left(f_i(\bar{y}_{k-1}) - f_i(x)\right)$$
$$+ \frac{2C}{k\gamma_i}\|v_k^i - \bar{y}_{k-1}\| + \frac{2}{k\gamma_i}\mathsf{E}\left[(\epsilon_k^i)^T(v_k^i - x) \mid \mathcal{F}_{k-1}, I_k, J_k\right] + \left(\frac{4m^2}{k^2 p_{\min}^2} + a_k\right)(C + \nu)^2. \quad (14)$$

We note that v_k^i is completely determined given $\mathcal{F}_{k-1}$, I_k, and J_k, so that

$$\mathsf{E}\left[(\epsilon_k^i)^T(v_k^i - x) \mid \mathcal{F}_{k-1}, I_k, J_k\right] \leq \|\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k]\| \, \|v_k^i - x\|. \quad (15)$$

Substituting estimate (15) in relation (14) and using $\gamma_i \geq \frac{p_{\min}}{m}$ (cf. (12) and $|\mathcal{N}(i)| \geq 1$), we have for $k \geq \tilde{k}$ and $i \in J_k$,

$$\mathsf{E}\left[\|x_k^i - x\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\right] \leq (1 + a_k)\|v_k^i - x\|^2 - \frac{2}{k\gamma_i}\left(f_i(\bar{y}_{k-1}) - f_i(x)\right)$$
$$+ \frac{2C}{k\gamma_i}\|v_k^i - \bar{y}_{k-1}\| + \frac{2m}{k\,p_{\min}}\|\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k]\|\|v_k^i - x\| + \left(\frac{4m^2}{k^2 p_{\min}^2} + a_k\right)(C + \nu)^2.$$

The desired relation follows by using $\gamma_i \geq \frac{p_{\min}}{m}$ for γ_i in the coefficient of $\|v_k^i - \bar{y}_{k-1}\|$. ∎

We now show the convergence of the algorithm.

Proposition 1: Let Assumptions 1–4 hold. Assume that the subgradient errors ϵ_k^i are such that $\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k] = 0$ for all $k \geq 1$ and $i \in J_k$ with probability 1. Let the sequences $\{x_k^i\}$, $i \in V$, be generated by method (9) with stepsize $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$. Assume that problem (1) has a non-empty optimal set X^*. Then, the sequences $\{x_k^i\}$, $i \in V$, converge to the same random point in X^* with probability 1.

Proof: Let $x^* \in X^*$ be arbitrary. Using $x = x^*$ in the relation of Lemma 5 and the fact $\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k] = 0$, we obtain with probability 1 for any $q \in (0, 1/2)$, and for all $k \geq \tilde{k}(q, m)$ and $i \in J_k$,

$$
\begin{aligned}
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq & \; (1 + a_k)\|v_k^i - x^*\|^2 - \frac{2}{k\gamma_i}\left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right) \\
& + \frac{2mC}{k\, p_{\min}}\|v_k^i - \bar{y}_{k-1}\| + \left(\frac{4m^2}{k^2\, p_{\min}^2} + a_k\right)(C + \nu)^2,
\end{aligned}
$$

with $a_k = \frac{2}{k^{\frac{3}{2}-q} p_{\min}^2}$. Recall that when $i \notin J_k$, we have $x_k^i = v_k^i$, and by combining these two cases with the fact that agent i updates with probability γ_i, we obtain with probability 1 for any $q \in (0, 1/2)$, and for all $k \geq \tilde{k}(q, m)$ and $i \in V$,

$$
\begin{aligned}
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \leq & \; (1 + a_k)\mathsf{E}\big[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] - \frac{2}{k}\left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right) \\
& + \frac{2mC}{k\, p_{\min}}\mathsf{E}\big[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\big] + \gamma_i\left(\frac{4m^2}{k^2\, p_{\min}^2} + a_k\right)(C + \nu)^2.
\end{aligned}
$$

We next sum the preceding relations over all $i \in V$. Then, using $f = \sum_{i=1}^m f_i$ and relations (10) and (11) (with $x = x^*$ and $x = \bar{y}_{k-1}$, respectively), we have with probability 1 for any $q \in (0, 1/2)$, any $x^* \in X^*$, and for all $k \geq \tilde{k}(q, m)$,

$$
\begin{aligned}
\mathsf{E}\left[\sum_{i=1}^m \|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \leq & \; (1 + a_k)\sum_{j=1}^m \|x_{k-1}^j - x^*\|^2 - \frac{2}{k}\left(f(\bar{y}_{k-1}) - f(x^*)\right) \\
+ \frac{2mC}{k\, p_{\min}}\sum_{j=1}^m \|x_{k-1}^j - \bar{y}_{k-1}\| & + \sum_{i=1}^m \gamma_i\left(\frac{4m^2}{k^2\, p_{\min}^2} + a_k\right)(C + \nu)^2. \quad (16)
\end{aligned}
$$

Now, since $0 < q < \frac{1}{2}$ and $a_k = \frac{2}{k^{\frac{3}{2}-q} p_{\min}^2}$, we have $\sum_{k=1}^\infty a_k < \infty$ and $\sum_{k=1}^\infty \left(\frac{4m^4}{k^2\, p_{\min}^2} + a_k\right) < \infty$. By Lemma 4, we have $\sum_{k=1}^\infty \frac{1}{k}\sum_{j=1}^m \|x_{k-1}^j - \bar{y}_{k-1}\| < \infty$ with probability 1, implying $\sum_{k=1}^\infty \frac{2mC}{k\, p_{\min}}\sum_{j=1}^m \|x_{k-1}^j - \bar{y}_{k-1}\| < \infty$ with probability 1. Furthermore, $f(\bar{y}_{k-1}) - f(x^*) \geq 0$ for all k since $\bar{y}_{k-1} \in X$ for all k and x^* is an optimal solution. Therefore, relation (16) satisfies the conditions of Lemma 1 for $k \geq \tilde{k}$. Hence, with probability 1, the sequence $\{\sum_{i=1}^m \|x_k^i - x^*\|^2\}$ converges for any $x^* \in X^*$, and $\sum_{k=\tilde{k}}^\infty \frac{1}{k}\left(f(\bar{y}_{k-1}) - f(x^*)\right) < \infty$. The latter relation implies

$$
\liminf_{k \to \infty} f(\bar{y}_{k-1}) = f(x^*) \qquad \text{with probability 1.} \quad (17)
$$

The relation $\lim_{k \to \infty} \|x_k^i - \bar{y}_k\| = 0$ for all $i \in V$ with probability 1 (Lemma 4), the convergence of the sequence $\{\sum_{i=1}^m \|x_k^i - x^*\|^2\}$ for any $x^* \in X^*$ with probability 1, and relation (17) imply by continuity of f that $\{\bar{y}_k\}$ converges to a random point in the set X^*

with probability 1. Since $\lim_{k \to \infty} \|x_k^i - \bar{y}_k\| = 0$ with probability 1 for all $i \in V$, it follows that each sequence $\{x_k^i\}$ converges to the same random point in X^* with probability 1. ∎

Our next result deals with the convergence of the method when the constraint set X is bounded. In this case, the zero-mean condition on the subgradient errors of Proposition 1 can be replaced with a more general condition, as seen in the following proposition.

Proposition 2: Let Assumptions 1, 2 and 4 hold, and let the constraint set X be compact. Assume that the subgradient errors ϵ_k^i are such that with probability 1 for some deterministic scalar sequence $\{\mu_k\}$, we have $\|\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k]\| \le \mu_k$ for all k and $i \in J_k$, and $\sum_{k=1}^{\infty} \frac{\mu_k}{k} < \infty$. Then, the optimal set X^* of problem (1) is nonempty, and the sequences $\{x_k^i\}$, $i \in V$, converge to the same random point in the set X^* with probability 1.

Proof: By Assumption 2, the function f is continuous over X (see the discussion following Assumption 2). Since the set X is compact, the optimal set X^* is nonempty. Furthermore, the subgradients of each f_i are bounded over X, and let $C > 0$ be such that $\|\nabla f_i(x)\| \le C$ for all $x \in X$ and $i \in V$.

Let $x^* \in X^*$ be arbitrary. Taking $x = x^*$ in Lemma 5 and using $\|\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k]\| \le \mu_k$, we have for any $q \in (0, 1/2)$, with probability 1, for all $k \ge \tilde{k}(q, m)$ and $i \in J_k$,

$$
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \le (1 + a_k)\|v_{k-1}^i - x^*\|^2 - \frac{2}{k\gamma_i}\left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right)
$$
$$
+ \frac{2mC}{k\,p_{\min}}\|v_k^i - \bar{y}_{k-1}\| + \frac{2m}{k\,p_{\min}}\mu_k\|v_k^i - x^*\| + \left(\frac{4m^2}{k^2\,p_{\min}^2} + a_k\right)(C + \nu)^2,
$$

where $a_k = \frac{2}{k^{\frac{3}{2}-q}\,p_{\min}^2}$ and $\bar{y}_k = \frac{1}{m}\sum_{j=1}^m x_k^j$. In view of $v_k^i = \sum_{j=1}^m [W_k]_{i,j}\,x_{k-1}^j$ with $x_k^j \in X$ for all j and k, and the stochasticity of the matrix W_k, we have $v_k^i \in X$ for all i and k, implying $\|v_k^i - x^*\| \le \max_{x,y \in X}\|x - y\|$. Letting $C_X = \max_{x,y \in X}\|x - y\|$, we thus obtain with probability 1 for any $x \in X^*$,

$$
\|v_k^i - x^*\| \le C_X \qquad \text{for any } k \ge 0 \text{ and } i \in V.
$$

Using the preceding two relations, and noting that $x_k^i = v_k^i$ when $i \notin J_k$ and that agent i updates with probability γ_i, we can see that for any $q \in (0, 1/2)$, with probability 1, for all $k \ge \tilde{k}(q, m)$ and $i \in V$,

$$
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \le (1 + a_k)\mathsf{E}\big[\|v_{k-1}^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] - \frac{2}{k}\left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right)
$$
$$
+ \frac{2mC}{k\,p_{\min}}\mathsf{E}\big[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\big] + \frac{2m}{k\,p_{\min}}\mu_k C_X + \gamma_i\left(\frac{4m^2}{k^2\,p_{\min}^2} + a_k\right)(C + \nu)^2.
$$

By summing the preceding relations over $i \in V$, and using $\sum_{i=1}^{m} \mathsf{E}[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}] \leq \sum_{j=1}^{m} \|x_{k-1}^j - x^*\|^2$ and $\sum_{i=1}^{m} \mathsf{E}[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}] \leq \sum_{j=1}^{m} \|x_{k-1}^j - \bar{y}_{k-1}\|$ (see (10)–(11)), we have with probability 1 for any $q \in (0, 1/2)$, $x^* \in X^*$, and $k \geq \tilde{k}(q, m)$,

$$\mathsf{E}\left[\sum_{i=1}^{m} \|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\right] \leq (1 + a_k) \sum_{j=1}^{m} \|x_{k-1}^j - x^*\|^2 - \frac{2}{k} \left(f(\bar{y}_{k-1}) - f(x^*)\right)$$

$$+ \frac{2mC}{k\, p_{\min}} \sum_{j=1}^{m} \|x_{k-1}^j - \bar{y}_{k-1}\| + \frac{2m^2}{k\, p_{\min}} \mu_k C_X + \sum_{i=1}^{m} \gamma_i \left(\frac{4m^2}{k^2\, p_{\min}^2} + a_k\right)(C + \nu)^2.$$

The result now follows by Lemma 1, relations $\sum_{k=1}^{\infty} \frac{1}{k}\mu_k < \infty$ and $\sum_{k=1}^{\infty} \frac{2mC}{k\, p_{\min}} \sum_{j=1}^{m} \|x_{k-1}^j - \bar{y}_{k-1}\| < \infty$, and the line of analysis similar to the proof of Proposition 1. ∎

V. ERROR BOUNDS

We now focus on a constant stepsize $\alpha_{i,k} = \alpha_i$, and we establish some limiting error bounds assuming that the set X is compact. We consider the case when each f_i is strongly convex and the case the functions are just convex. We make use of the following result for a scalar sequence, which is proved in Appendix.

Lemma 6: Let $\{d_k\}$ and $\{u_k\}$ be scalar sequences such that $d_k \leq cd_{k-1} + u_{k-1}$ for all $k \geq 1$ and some scalar $c \in (0, 1)$. Then, $\limsup_{k \to \infty} d_k \leq \frac{1}{1-c} \limsup_{k \to \infty} u_k$.

Consider now the case when each function is strongly convex. In particular, let f_i be strongly convex over X with a constant $\sigma_i > 0$:

$$(\nabla f_i(x) - \nabla f_i(y))^T (x - y) \geq \sigma_i \|x - y\|^2 \qquad \text{for all } x, y \in X.$$

In this case, the function f is strongly convex with constant $\sigma = \sum_{i=1}^{m} \sigma_i$ and, therefore, problem (1) has a unique solution. We establish an error bound on the expected distances between the optimal solution and the agent iterates in the following proposition.

Proposition 3: Let Assumptions 1, 2 and 4 hold. Let the set X be bounded. Assume that the subgradient errors ϵ_k^i are such that $\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k] = 0$ with probability 1 for all $k \geq 1$ and $i \in J_k$. Let each function f_i be strongly convex over the set X with a constant σ_i. Let sequences $\{x_k^i\}$, $i \in V$, be generated by method (9) with the stepsize $\alpha_{i,k} = \alpha_i > 0$ such that $2\alpha_i \sigma_i < 1$ for all $i \in V$. Then, we have

$$\limsup_{k \to \infty} \sum_{i=1}^{m} \mathsf{E}\left[\|x_k^i - x^*\|^2\right] \leq \frac{\varepsilon}{1-c} + \frac{2C}{1-c}\, \alpha_{\max}^2 \frac{\sqrt{md_{\max}}}{1 - \sqrt{\lambda}}\, (C + \nu),$$

where x^* is the optimal solution of problem (1), $c = 1 - 2\gamma_{\min} \min_j \alpha_j \sigma_j$, $\gamma_{\min} = \min_j \gamma_j$, $\alpha_{\max} = \max_i \alpha_i$, $d_{\max}$ is the maximum expected number of agents that update at any time as given in Lemma 4(b), and

$$\varepsilon = 2m\Delta_{\alpha,\sigma} \left(1 - \gamma_{\min}\right) C_X^2 + 2m\Delta_{\alpha,\gamma} CC_X + m\alpha_{\max} \left(\max_\ell \alpha_\ell \gamma_\ell\right) (C + \nu)^2.$$

Here, $\Delta_{\alpha,\sigma} = \max_\ell \alpha_\ell \sigma_\ell - \min_j \alpha_j \sigma_j$, $\Delta_{\alpha,\gamma} = \max_\ell \alpha_\ell \gamma_\ell - \min_j \alpha_j \gamma_j$, $C_X = \max_{x,y} \|x - y\|$, and C and ν are the bounds on the subgradient and subgradient error norms, respectively.

Proof: From the definition of the method in (9), the nonexpansive property of the projection operation, and the zero mean assumption on the errors, we obtain for the optimal point x^*, and any k and $i \in J_k$,

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq \|v_k^i - x^*\|^2 - 2\alpha_i \nabla f_i(v_k^i)^T(v_k^i - x^*) + \alpha_i^2 \|\nabla f_i(v_k^i) + \epsilon_k^i\|^2.$$
(18)

Adding and subtracting $\nabla f_i(x^*)^T(v^i - x^*)$, and using the strong convexity of f_i, we have

$$\begin{aligned}
\nabla f_i(v_k^i)^T(v_k^i - x^*) &= \left(\nabla f_i(v_k^i) - \nabla f_i(x^*)\right)^T (v_k^i - x^*) + \nabla f_i(x^*)^T(v_k^i - x^*) \\
&\geq \sigma_i \|v_k^i - x^*\|^2 + \nabla f_i(x^*)^T(v_k^i - x^*).
\end{aligned}$$

We further have $\nabla f_i(x^*)^T(v_k^i - x^*) = \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + \nabla f_i(x^*)^T(v_k^i - \bar{y}_{k-1})$, with $\bar{y}_{k-1} = \frac{1}{m} \sum_{j=1}^m x_{k-1}^j$. Thus, we obtain

$$\nabla f_i(x^*)^T(v_k^i - x^*) \geq \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) - C \|v_k^i - \bar{y}_{k-1}\|,$$

where the inequality follows from $\|\nabla f_i(x^*)\| \leq C$ for all i. The preceding two relations imply for all $k \geq 1$ and all $i \in J_k$,

$$\nabla f_i(v_k^i)^T(v_k^i - x^*) \geq \sigma_i \|v_k^i - x^*\|^2 + \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) - C \|v_k^i - \bar{y}_{k-1}\|.$$

Using the preceding estimate and the inequality

$$\mathsf{E}\big[\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq (C + \nu)^2$$

(see Eq. (7)), from relation (18), we obtain for all $k \geq 1$ and $i \in J_k$,

$$\begin{aligned}
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] &\leq (1 - 2\alpha_i \sigma_i)\|v_k^i - x^*\|^2 - 2\alpha_i \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) \\
&\quad + 2\alpha_i C \|v_k^i - \bar{y}_{k-1}\| + \alpha_i^2 (C + \nu)^2.
\end{aligned}$$

When $i \notin J_k$, we have $x_k^i = v_k^i$ by the definition of the method. Hence, for $i \notin J_k$,

$$\|x_k^i - x^*\|^2 = \|v_k^i - x^*\|^2 = (1 - 2\alpha_i \sigma_i)\|v_k^i - x^*\|^2 + 2\alpha_i \sigma_i \|x_{k-1}^i - x^*\|^2,$$

where we also use the fact $v_k^i = x_{k-1}^i$ when agent i does not update. By combining the preceding two relations, and using the indicator function $\chi_{\{i \in J_k\}}$ of the event that agent i updates at time k, we can see that for any any $k \geq 1$ and $i \in V$,

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq (1 - 2\alpha_i\sigma_i)\|v_k^i - x^*\|^2 + 2\alpha_i\sigma_i\|x_{k-1}^i - x^*\|^2\left(1 - \chi_{\{i \in J_k\}}\right)$$
$$- 2\alpha_i\nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*)\chi_{\{i \in J_k\}} + 2\alpha_i C\|v_k^i - \bar{y}_{k-1}\| + \alpha_i^2(C + \nu)^2\chi_{\{i \in J_k\}},$$

where we have also used $\|v_k^i - \bar{y}_{k-1}\|\chi_{\{i \in J_k\}} \leq \|v_k^i - \bar{y}_{k-1}\|$. Taking the expectation conditioned on the past $\mathcal{F}_{k-1}$ and using the fact that agent i updates with probability γ_i independently of the past, we obtain with probability 1 for all $k \geq 1$ and $i \in V$,

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \leq (1 - 2\alpha_i\sigma_i)\mathsf{E}\big[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] + 2\alpha_i\sigma_i\|x_{k-1}^i - x^*\|^2\left(1 - \gamma_{\min}\right)$$
$$-2\alpha_i\gamma_i\nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + 2\alpha_i C\mathsf{E}\big[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\big] + \alpha_i^2\gamma_i(C + \nu)^2,$$

where $\gamma_{\min} = \min_i \gamma_i$. Let $\Delta_{\alpha,\sigma} = \max_\ell \alpha_\ell\sigma_\ell - \min_j \alpha_j\sigma_j$ and $\Delta_{\alpha,\gamma} = \max_\ell \alpha_\ell\gamma_\ell - \min_j \alpha_j\gamma_j$. Note that by $\|\nabla f_i(x^*)\| \leq C$, the fact $x_k^i, \bar{y}_k \in X$ for all i and k, and the compactness of X, we can see that

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \leq (1 - 2\min_s \alpha_s\sigma_s)\mathsf{E}\big[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big]$$
$$+ 2\left(\min_j \alpha_j\sigma_j\right)(1 - \gamma_{\min})\|x_{k-1}^i - x^*\|^2 + 2\Delta_{\alpha,\sigma}(1 - \gamma_{\min})C_X^2$$
$$- 2\left(\min_j \alpha_j\gamma_j\right)\nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + 2\Delta_{\alpha,\gamma}CC_X$$
$$+ 2\alpha_{\max}C\mathsf{E}\big[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\big] + \alpha_i^2\gamma_i(C + \nu)^2,$$

where $C_X = \max_{x,y \in X}\|x - y\|$ and $\alpha_{\max} = \max_\ell \alpha_\ell$.

We sum the preceding inequalities over all $i \in V$, and use $\sum_{i=1}^m \mathsf{E}[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}] \leq \sum_{j=1}^m \|x_{k-1}^j - x^*\|^2$ and $\sum_{i=1}^m \mathsf{E}[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}] \leq \sum_{j=1}^m \|x_{k-1}^j - \bar{y}_{k-1}\|$ (see (10)–(11)), to obtain with probability 1 for all $k \geq 1$,

$$\sum_{i=1}^m \mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \leq (1 - 2\min_s \alpha_s\sigma_s)\sum_{j=1}^m \|x_{k-1}^j - x^*\|^2$$
$$+ 2\left(\min_j \alpha_j\sigma_j\right)(1 - \gamma_{\min})\sum_{i=1}^m \|x_{k-1}^i - x^*\|^2 + \varepsilon$$
$$- 2\left(\min_j \alpha_j\gamma_j\right)\sum_{i=1}^m \nabla f_i(x^*)^T(\bar{y}_{k-1} - x^*) + 2\alpha_{\max}C\sum_{j=1}^m \|x_{k-1}^j - \bar{y}_{k-1}\|,$$

where $\varepsilon = 2m\Delta_{\alpha,\sigma}(1 - \gamma_{\min})C_X^2 + 2m\Delta_{\alpha,\gamma}CC_X + m\alpha_{\max}\max\alpha_\ell\gamma_\ell(C + \nu)^2$. The last term in ε is obtained by using $\sum_{i=1}^m \alpha_i^2\gamma_i \leq m\alpha_{\max}\max\alpha_\ell\gamma_\ell$. Observe that $\sum_{i=1}^m \nabla f_i(x^*)^T(\bar{y}_{k-1} -$

$x^*) \geq 0$ since $\bar{y}_{k-1} \in X$ and x^* is the optimal point of $\sum_{i=1}^m f_i(x)$ over $x \in X$. Thus, by taking the total expectation, we obtain for all $k \geq 1$,

$$\sum_{i=1}^m \mathsf{E}\big[\|x_k^i - x^*\|^2\big] \leq c \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - x^*\|^2\big] + \varepsilon + 2\alpha_{\max} C \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big], \quad (19)$$

where $c = 1 - 2\gamma_{\min} \min_j \alpha_j \sigma_j$. We have $c \in (0,1)$ since $\gamma_{\min} \leq 1$ and $2\alpha_i \sigma_i < 1$ for all i (by our assumption). Thus, relation (19) satisfies Lemma 6, implying

$$\limsup_{k \to \infty} \sum_{i=1}^m \mathsf{E}\big[\|x_k^i - x^*\|^2\big] \leq \frac{\varepsilon}{1-c} + \frac{2\alpha_{\max} C}{1-c} \limsup_{k \to \infty} \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big].$$

The desired relation is obtained by using Lemma 4(b). $\blacksquare$

Proposition 3 provides an asymptotic error bound for the sum of the expected distances between the optimal solution x^* and the iterates of asynchronous broadcast-based algorithm in the presence of random link failures. The bound captures the effects of the network connectivity topology and the link failure probabilities (through $\gamma_{\min}$, λ and $d_{\max}$), as well as the objective function properties. Note that the error ε grows as m and the error term $\frac{2C}{1-c} \alpha_{\max}^2 \frac{\sqrt{m d_{\max}}}{1-\sqrt{\lambda}} (C+\nu)$ grows as $\sqrt{m}$ in the number m of agents, provided that $\gamma_{\min}$ and λ do not depend on m. In this case, the limiting error grows as m in the number m of agents. This will happen, for example, if the graph $(V, \mathscr{E})$ is full and the links are perfectly reliable (i.e., $p_{ij} = 1$ for all $\{i,j\} \in \mathscr{E}$), which will imply that $\gamma_{\min} = \frac{m-1}{m}$ and $d_{\max} = m-1$, while $\lambda \leq 1 - \beta^2$ where $\beta \in (0,1)$ are weights as given in (3).

The result of Proposition 3 requires that each node selects a stepsize α_i so that $2\alpha_i \sigma_i < 1$, which can be done since each node knows its strong convexity constant σ_i. Then, since $\gamma_i \in (0,1)$ for all i, we have $c = 1 - 2\gamma_{\min} \min_i \alpha_i \sigma_i < 1$. Thus, the relation $c \in (0,1)$ can be ensured globally over the network without any coordination among the agents.

To get additional insights into the result of Proposition 3, suppose that $\Delta_{\alpha,\sigma} = \max_\ell \alpha_\ell \sigma_\ell - \min_j \alpha_j \sigma_j \approx 0$ (or $\gamma_{\min} \approx 1$) and $\Delta_{\alpha,\gamma} = \max_\ell \alpha_\ell \gamma_\ell - \min_j \alpha_j \gamma_j \approx 0$. Substituting the value for c, we obtain the following limiting error bound:

$$\limsup_{k \to \infty} \sum_{i=1}^m \mathsf{E}\big[\|x_k^i - x^*\|^2\big] \leq \frac{m\alpha_{\max}}{2\gamma_{\min}} \frac{\max_i \alpha_i \gamma_i}{\min_j \alpha_j \sigma_j} (C+\nu)^2 + \frac{C\alpha_{\max}^2}{\gamma_{\min} \min_j \alpha_j \sigma_j} \frac{\sqrt{m d_{\max}}}{1-\sqrt{\lambda}} (C+\nu).$$

Note that the preceding error bound is proportional to $\alpha_{\max}$, implying that the error can be small when all the agents use small stepsize values α_i.

We next provide another error estimate that holds without requiring strong convexity.

Proposition 4: Let Assumptions 1, 2 and 4 hold. Let the set X be bounded and the errors ϵ_k^i be such that $\|\mathsf{E}[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k]\| \leq \mu$ with probability 1 for all $k \geq 1$, $i \in J_k$ and for a deterministic scalar $\mu > 0$. Let the sequences $\{x_k^i\}$, $i \in V$, be generated by method (9) with stepsize $\alpha_{i,k} = \alpha_i$. Then, for the averaged iterates $\hat{x}_t^i = \frac{1}{t}\sum_{t=1}^k x_{k-1}^i$ we have

$$\limsup_{t\to\infty} \mathsf{E}\big[f(\hat{x}_t^i)\big] - f^* \leq \theta + \left(mC + \frac{\alpha_{\max}C}{\min_\ell \alpha_\ell\gamma_\ell}\right)\frac{\alpha_{\max}\sqrt{md_{\max}}}{1-\sqrt{\lambda}}(C+\nu) \text{ for all } i \in V,$$

$$\theta = mCC_X\left(\frac{\max_\ell \alpha_\ell\gamma_\ell}{\min_j \alpha_j\gamma_j} - 1\right) + \frac{m\max_\ell \alpha_\ell\gamma_\ell}{2\min_j \alpha_j\gamma_j}\left(2\mu\,C_X + \alpha_{\max}(C+\nu)^2\right),$$

where f^* is the optimal value of problem (1), $C_X = \max_{x,y\in X}\|x-y\|$, $\alpha_{\max} = \max_i \alpha_i$, and $d_{\max}$ is the maximum expected number of agents updating as defined in Lemma 4(b).

Proof: Since X is compact, the optimal set X^* is nonempty. Thus, from $x_k^i = P_X[v_k^i - \alpha_i(\nabla f_i(v_k^i) + \epsilon_k^i)]$ for $i \in J_k$ (see Eq. (9)), it follows for all $k \geq 1$ and $i \in J_k$,

$$\|x_k^i - x^*\|^2 \leq \|v_k^i - x^*\|^2 - 2\alpha_i(\nabla f_i(v_k^i) + \epsilon_k^i)^T(v_k^i - x^*) + \alpha_i^2\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2.$$

By compactness of X, we have $\|\nabla f_i(x)\| \leq C$ for all $x \in X$ and all i, and some scalar C. Taking the expectation conditional on $\mathcal{F}_{k-1}$, I_k, and J_k, and using

$$\mathsf{E}\big[\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2\mathcal{F}_{k-1}, I_k, J_k\big] \leq (C+\nu)^2$$

(see Eq. (7)), we obtain with probability 1 for all $k \geq 1$ and $i \in J_k$,

$$\begin{aligned}
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq{}& \|v_k^i - x^*\|^2 - 2\alpha_i\nabla f_i(v_k^i)^T(v_k^i - x^*)\\
&- 2\alpha_i\mathsf{E}\big[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k\big]^T(v_k^i - x^*) + \alpha_i^2(C+\nu)^2.
\end{aligned}$$

From approximate subgradient relation (5) it follows

$$\nabla f_i(v_k^i)^T(v_k^i - x^*) \geq f_i(\bar{y}_{k-1}) - f_i(x^*) - C\|v_k^i - \bar{y}_{k-1}\|.$$

Furthermore, by the assumption on the mean of the error, we have

$$\mathsf{E}\big[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k\big]^T(v_k^i - x) \leq \|\mathsf{E}\big[\epsilon_k^i \mid \mathcal{F}_{k-1}, I_k, J_k\big]\|\,\|v_k^i - x\| \leq \mu\|v_k^i - x\|.$$

Combining the preceding relations, we obtain with probability 1 for all $k \geq 1$ and $i \in J_k$,

$$\begin{aligned}
\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq{}& \|v_k^i - x^*\|^2 - 2\alpha_i(f_i(\bar{y}_{k-1}) - f_i(x^*))\\
&+ 2\alpha_i C\|v_k^i - \bar{y}_{k-1}\| + \alpha_i\left(2\mu\,C_X + \alpha_i(C+\nu)^2\right),
\end{aligned}$$

where $C_X = \max_{x,y \in X} \|x - y\|$, which is finite since $v_k^i \in X$ and X is compact. Using the indicator function of the event that agent i updates at the time k, from the preceding relation we further obtain with probability 1 for all $k \geq 1$ and $i \in J_k$,

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\big] \leq \big(\|v_k^i - x^*\|^2 - 2\alpha_i(f_i(\bar{y}_{k-1}) - f_i(x^*))\big) \chi_{\{i \in J_k\}}$$
$$+ 2\alpha_i C \|v_k^i - \bar{y}_{k-1}\| + \alpha_i \big(2\mu\, C_X + \alpha_i(C + \nu)^2\big) \chi_{\{i \in J_k\}}. \tag{20}$$

We combine relation (20) with $x_k^i = v_k^i$ for $i \notin J_k$ (see (9)). Since agent i updates with probability γ_i independently of the past, we have with probability 1 for $k \geq 1$ and $i \in V$,

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \leq \mathsf{E}\big[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] - 2\alpha_i \gamma_i (f_i(\bar{y}_{k-1}) - f_i(x^*))$$
$$+ 2\alpha_i C \mathsf{E}\big[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\big] + \alpha_i \gamma_i \big(2\mu\, C_X + \alpha_i(C + \nu)^2\big). \tag{21}$$

Let $\Delta_{\alpha,\gamma} = \max_\ell \alpha_\ell \gamma_\ell - \min_j \alpha_j \gamma_j$. Using $|f_i(\bar{y}_{k-1}) - f_i(x^*)| \leq C\|\bar{y}_{k-1} - x^*\| \leq CC_X$, which holds by the subgradient boundedness and the fact $\bar{y}_k \in X$, from (21) we see that with probability 1 for all $k \geq 1$ and $i \in V$,

$$\mathsf{E}\big[\|x_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] \leq \mathsf{E}\big[\|v_k^i - x^*\|^2 \mid \mathcal{F}_{k-1}\big] - 2\left(\min_j \alpha_j \gamma_j\right)(f_i(\bar{y}_{k-1}) - f_i(x^*))$$
$$+ 2\Delta_{\alpha,\gamma} CC_X + 2\alpha_{\max} C \mathsf{E}\big[\|v_k^i - \bar{y}_{k-1}\| \mid \mathcal{F}_{k-1}\big] + \max_\ell \alpha_\ell \gamma_\ell \big(2\mu\, C_X + \alpha_{\max}(C + \nu)^2\big),$$

where $\alpha_{\max} = \max_i \alpha_i$. We next sum the preceding inequalities over all i, and note that $f = \sum_{i=1}^m f_i$ and $f^* = f(x^*)$. By doing this and using relations (10)–(11) (with $x = x^*$ and $x = \bar{y}_{k-1}$, respectively) after taking the total expectation and rearranging the terms, we obtain with probability 1 for all $k \geq 1$,

$$2 \min_j \alpha_j \gamma_j \, \mathsf{E}[f(\bar{y}_{k-1}) - f^*] \leq \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - x^*\|^2\big] - \sum_{i=1}^m \mathsf{E}\big[\|x_k^i - x^*\|^2\big] + 2m\Delta_{\alpha,\gamma} CC_X$$
$$+ 2\alpha_{\max} C \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big] + m \max_\ell \alpha_\ell \gamma_\ell \big(2\mu\, C_X + \alpha_{\max}(C + \nu)^2\big). \tag{22}$$

Next, we divide the preceding relation by $2 \min_j \alpha_j \gamma_j$. We note that by convexity and the boundedness of the subgradients of each f_i, we have

$$f(x_{k-1}^i) - f^* \leq f(\bar{y}_{k-1}) - f^* + mC\|x_{k-1}^i - \bar{y}_{k-1}\| \leq f(\bar{y}_{k-1}) - f^* + mC \sum_{i=1}^m \|x_{k-1}^i - \bar{y}_{k-1}\|.$$

Thus, we obtain from (22) for all $k \geq 1$ and $i \in V$,

$$\mathsf{E}\big[f(x_{k-1}^i) - f^*\big] \leq \frac{1}{2 \min_j \alpha_j \gamma_j} \left(\sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - x^*\|^2\big] - \sum_{i=1}^m \mathsf{E}\big[\|x_k^i - x^*\|^2\big] \right)$$
$$+ \theta + \left(mC + \frac{\alpha_{\max} C}{\min_j \alpha_j \gamma_j} \right) \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big],$$

where $\theta = \frac{m \Delta_{\alpha,\gamma} C C_X}{\min_j \alpha_j \gamma_j} + \frac{m \max_\ell \alpha_\ell \gamma_\ell}{2 \min_j \alpha_j \gamma_j} \left(2\mu\, C_X + \alpha_{\max}(C + \nu)^2 \right)$. Using $\Delta_{\alpha,\gamma} = \max_\ell \alpha_\ell \gamma_\ell - \min_j \alpha_j \gamma_j$, the error θ can be written as:

$$\theta = m C C_X \left(\frac{\max_\ell \alpha_\ell \gamma_\ell}{\min_j \alpha_j \gamma_j} - 1 \right) + \frac{m \max_\ell \alpha_\ell \gamma_\ell}{2 \min_j \alpha_j \gamma_j} \left(2\mu\, C_X + \alpha_{\max}(C + \nu)^2 \right).$$

By summing the preceding relations over k from $k = 1$ to $k = t$ for some $t \geq 1$, and then averaging with respect to t, we obtain for any $t \geq 1$ and $i \in V$,

$$\frac{1}{t} \sum_{k=1}^t \mathsf{E}\big[f(x_{k-1}^i) - f^*\big] \leq \frac{1}{2t \min_j \alpha_j \gamma_j} \sum_{j=1}^m \mathsf{E}\big[\|x_0^j - x^*\|^2\big]$$
$$+ \theta + \left(mC + \frac{\alpha_{\max} C}{\min_\ell \alpha_\ell \gamma_\ell} \right) \frac{1}{t} \sum_{k=1}^t \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big].$$

Letting $t \to \infty$ and using the relation

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{k=1}^t \left(\sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big] \right) \leq \limsup_{k \to \infty} \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big],$$

we have for any $i \in V$,

$$\limsup_{t \to \infty} \frac{1}{t} \sum_{k=1}^t \mathsf{E}\big[f(x_{k-1}^i) - f^*\big] \leq \theta + \left(mC + \frac{\alpha_{\max} C}{\min_\ell \alpha_\ell \gamma_\ell} \right) \limsup_{k \to \infty} \sum_{j=1}^m \mathsf{E}\big[\|x_{k-1}^j - \bar{y}_{k-1}\|\big].$$

The estimate now follows by using Lemma 4(b) and convexity of the function f. ∎

The error bound of Proposition 4 scales as $m\sqrt{m}$ in the number m of agents, when $d_{\max}$ and λ do not depend on m. As a consequence of Proposition 4, in view of $f(x_k^i) - f^* \geq 0$ (since $x_k^i \in X$ for all k and i) and Fatou's lemma, it can be seen that for all $i \in V$,

$$\mathsf{E}\Big[\liminf_{k \to \infty} f(x_k^i) - f^* \Big] \leq \theta + \left(mC + \frac{\alpha_{\max} C}{\min_j \alpha_j \gamma_j} \right) \frac{\alpha_{\max} \sqrt{m d_{\max}}}{1 - \sqrt{\lambda}} (C + \nu),$$

with $\theta = m C C_X \left(\frac{\max_\ell \alpha_\ell \gamma_\ell}{\min_j \alpha_j \gamma_j} - 1 \right) + \frac{m \max_\ell \alpha_\ell \gamma_\ell}{2 \min_j \alpha_j \gamma_j} \left(2\mu\, C_X + \alpha_{\max}(C + \nu)^2 \right)$. When the mean of the errors ϵ_k^i are zero (corresponding to $\mu = 0$) and the ratio $\frac{\max_\ell \alpha_\ell \gamma_\ell}{\min_j \alpha_j \gamma_j}$ is close to 1, the result of Proposition 4 reduces to

$$\limsup_{t \to \infty} \mathsf{E}\big[f(\hat{x}_t^i)\big] - f^* \leq \frac{m}{2} \alpha_{\max}(C + \nu)^2 + \left(mC + \frac{\alpha_{\max} C}{\min_j \alpha_j \gamma_j} \right) \frac{\alpha_{\max} \sqrt{m d_{\max}}}{1 - \sqrt{\lambda}} (C + \nu).$$

This bound scales by order $5/2$ less than the scaling of the bound for the distributed (deterministic) consensus-based subgradient algorithm of [26], which scales at best as m^4. Specifically, in [26] it is shown that for consensus-based distributed algorithm (with zero mean errors ϵ_k^i) the corresponding error bound on the difference in function values at the time-averages of the agent iterates and the optimal value is given by (Theorem 5.4 in [26]): $m\alpha(C+\nu)^2\left(\frac{9}{2}+\frac{2mb\beta}{1-\beta}\right)$. Here, α is the constant stepsize common to all agents, and b and β are some positive constants such that $\frac{b\beta}{1-\beta}$ is of the order m^2. Thus, random broadcast algorithm has much better scaling of the asymptotic bound with the size m of the network.

VI. NUMERICAL RESULTS

We implement the broadcast-based algorithm for solving the following problem

$$\min_{x\in X}\sum_{i=1}^{m}f_i(x)\qquad\text{where }f_i(x)=\mathbb{E}[(R_i-s_i^Tx)^2]\text{ and }X\subseteq\mathbb{R}^2,$$

where $s_i\in\mathbb{R}^2$ and R_i is a random noise. The set X is chosen to be the unit ball in $\mathbb{R}^2$ centered at the origin $(0,0)$, while the vectors s_i are chosen randomly in X with the uniform distribution. In the experiments, we have $\mathbb{E}[R_i]=0.5$ and $\mathbb{E}[R_i^2]=1$ for all agents i.

We implement the broadcast-based algorithm (3) over a network with 50 agents ($m=50$) for two network topologies, namely, a cycle and a wheel. In the cycle network, the agents are connected to form a single undirected cycle, while in the wheel network, agent 1 is connected to all the other agents i by undirected link $\{1,i\}$. In the experiments, all the links are reliable, viz $p_{ij}=1$ for all the links.

We evaluate the algorithm performance by carrying out 25 Monte-Carlo runs, each with 10,000 iterations, for both the diminishing and constant stepsize rules. The diminishing stepsize for agent i is based on the number $\Gamma_k(i)$ of times the agent updates by the given time k. For each agent i, the constant stepsize value α_i is generated randomly with the uniform distribution on $[0.001, 0.01]$, which is denoted by U(0.001, 0.01). Table **??** provides the parameter values for the algorithm and the network topologies.

The algorithm performance is numerically evaluated by plotting the averaged values (across the Monte-Carlo runs) of the network function objective $f(x)=\sum_{i=1}^{m}f_i(x)$ along the iterates $\{x_k^i\}$ for randomly selected agents i. The results for the cycle network are shown for 7 randomly (and uniformly) selected agents out of 50. The results for the wheel network

TABLE I

ALGORITHM AND NETWORK PARAMETERS

Parameter	Symbol	Cycle	Wheel
Mixing weight	β	0.35	0.35
Maximum expected degree	$d_{\max}$	2	49
Diminishing stepsize	α_i	$\frac{1}{\Gamma_k(i)}$	$\frac{1}{\Gamma_k(i)}$
Constant stepsize	α_i	U(0.001, 0.01)	U(0.001, 0.01)

are given for the central agent (the one that is connected with all other agents) and for another 6 agents randomly (uniformly) selected out of the remaining 49 (peripheral) agents.

The results for diminishing stepsize are shown in Figure 2, where the network function objective $f(x) = \sum_{i=1}^{m} f_i(x)$ is plotted along the iterate sequences $\{x_k^i\}$ for randomly selected agents i, as described above. The results are in compliance with the convergence established in Proposition 1.
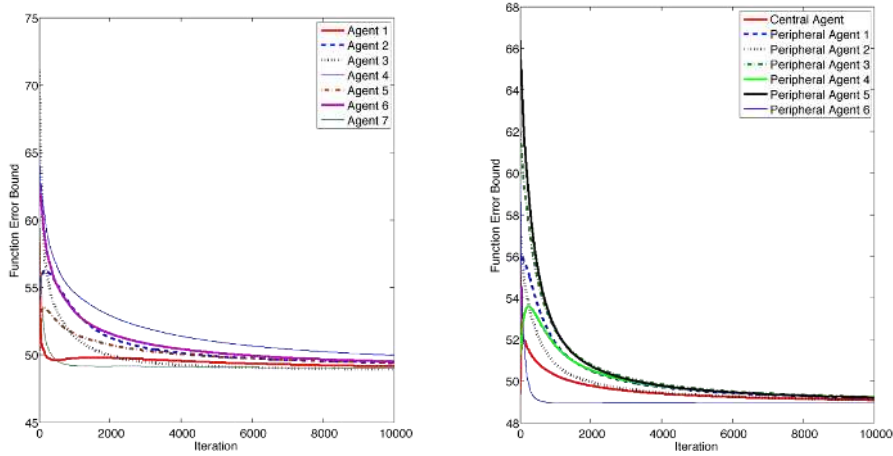

Fig. 2.   Network objective $f(x)$ evaluated along $\{x_k^i\}$. The plot to the left is for the cycle network, while the plot to the right is for the wheel network.

The asymptotic results for constant stepsize are shown in Figures 3 and 4, where the network objective $f(x) = \sum_{i=1}^{m} f_i(x)$ is plotted for the cycle and the wheel network, respectively. The function $f(x)$ is evaluated along the averaged iterate sequences $\{\hat{x}_k^i\}$ for randomly selected agents i, as discussed earlier. The plots also compare the numerically observed "asymptotic error" with the theoretical error bound predicted by Proposition 4. As expected, Figures 3 and 4 indicate reaching an "asymptotic error" level. However, the experimentally observed error level is by far less than the error bound of Proposition 4. This

is no surprise since the error bound is for the worse case scenario that does not account for any special properties that the functions f_i may have aside from convexity.
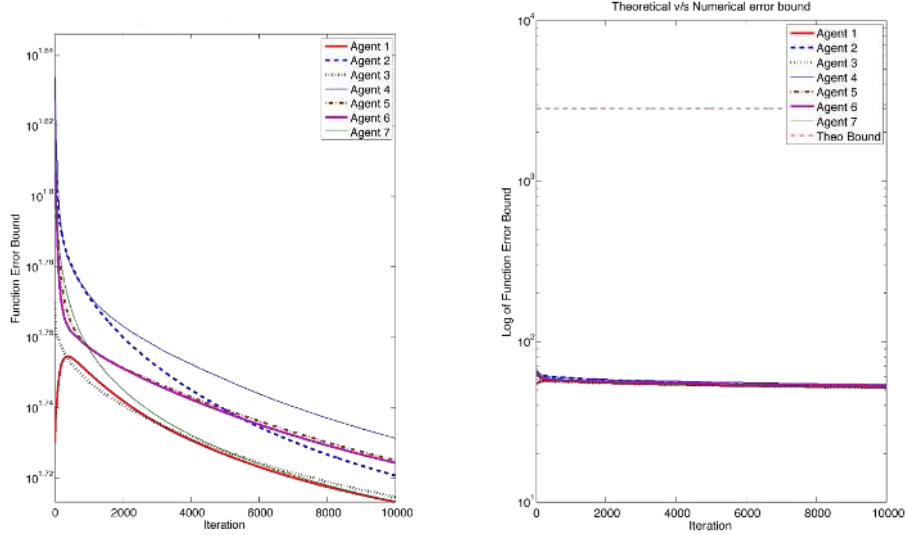

Fig. 3. Network objective $f(x)$ evaluated along $\{\hat{x}_k^i\}$ for the cycle network. The plot to the left shows the function values for different agents, while the plot to the right compares these values with the theoretical bound (on the log-scale).
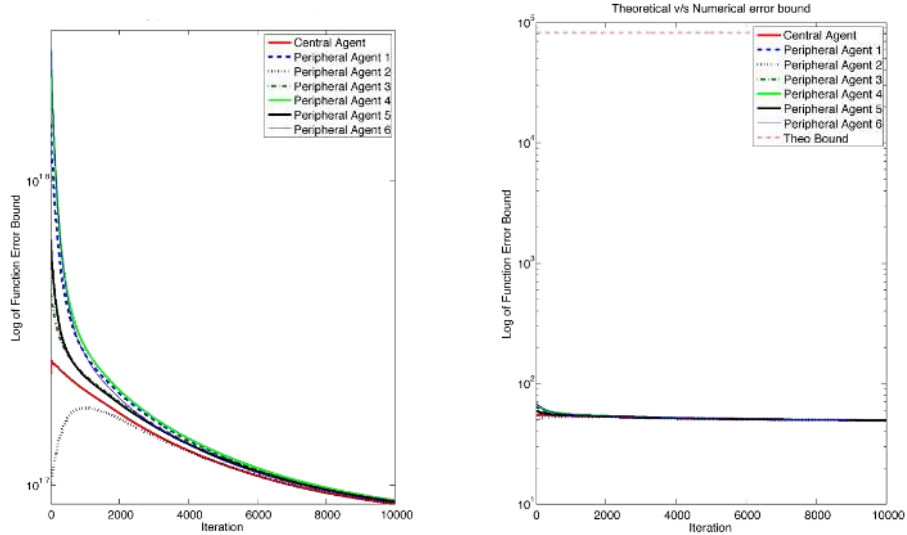

Fig. 4. Network objective $f(x)$ evaluated along $\{\hat{x}_k^i\}$ for the wheel network. The plot to the left shows the function values, while the plot to the right compares these values with the theoretical bound (on the log-scale).

VII. CONCLUSIONS

We have considered a convex problem of minimizing the sum of agents' objective functions over a common constraint set X, and proposed a broadcast-based optimization algorithm suitable for solving the problem over a wireless network. We discussed the general

case when agents evaluations of the (sub)gradients of their objective functions are erroneous and studied the properties of the algorithm for a random diminishing stepsize and a constant deterministic stepsize. We showed convergence with probability 1 to an optimal solution for the diminishing stepsize, and established two error bounds for a constant stepsize. We have also provided a numerical example.

APPENDIX

Proof: **Lemma 2**. Each matrix W_k is stochastic by the construction. Part (a) follows from the structure of the weighted Laplacian L_Π, the weight rule used in the update (see (3)), and from each agent i broadcasting with probability $\frac{1}{m}$. We next show the result in part (b). For the matrix $D_k = W_k - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k$ we have

$$D_k^T D_k = \left(W_k^T - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k \right) \left(W_k - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k \right) = W_k^T W_k - \frac{1}{m} W_k^T \mathbf{1}\mathbf{1}^T W_k, \quad (23)$$

where we use $\left(\frac{1}{m} \mathbf{1}\mathbf{1}^T \right)^2 = \frac{1}{m} \mathbf{1}\mathbf{1}^T$. Consider the quadratic form $z^T \mathsf{E}\left[D_k^T D_k \right] z$ for $z \in \mathbb{R}^m$. When $z = c\mathbf{1}$ for some $c \in \mathbb{R}$, from the stochasticity of W_k it is easily seen that

$$\mathbf{1}^T \mathsf{E}\left[D_k^T D_k \right] \mathbf{1} = 0. \quad (24)$$

We next estimate the maximum of $z^T \mathsf{E}\left[D_k^T D_k \right] z$ on the subspace of the vectors orthogonal to $\mathbf{1}$. For any vector z such that $z^T \mathbf{1} = 0$, using (23) we have

$$z^T \mathsf{E}\left[D_k^T D_k \right] z = \mathsf{E}\left[z W_k^T W_k z \right] - \frac{1}{m} \mathsf{E}\left[z^T W_k^T \mathbf{1}\mathbf{1}^T W_k z \right] \leq \|z\|^2 - \frac{1}{m} \mathsf{E}\left[z^T W_k^T \mathbf{1}\mathbf{1}^T W_k z \right], (25)$$

where the inequality follows by $\|W_k z\|^2 \leq \|z\|^2$, which holds by the stochasticity of W_k.

Let L_Π be the weighted Laplacian of the weighted graph $(V, \mathscr{E}, \Pi)$, where Π is the symmetric matrix with entries $\Pi_{ij} = p_{ij} > 0$ when $\{i, j\} \in \mathscr{E}$ and otherwise $p_{ij} = 0$ (recall p_{ij} is the probability that link $\{i, j\} \in \mathscr{E}$ is functioning). In this case, the weighted Laplacian L_Π is symmetric and positive semidefinite (see [17], page 6), and therefore it has m nonnegative eigenvalues $0 \leq \lambda_1(L_\Pi) \leq \lambda_2(L_\Pi) \leq \cdots \leq \lambda_m(L_\Pi)$, which are repeated according to their multiplicities. Since the graph $(V, \mathscr{E})$ is connected the smallest eigenvalue is zero with multiplicity 1 (see [17], Proposition 2.3). Furthermore, since $L_\Pi \mathbf{1} = 0$, it follows

$$0 < \lambda_2(L_\Pi) = \min_{\{z \in \mathbb{R}^m | z^T \mathbf{1} = 0, \ \|z\| = 1\}} \|L_\Pi z\|. \quad (26)$$

Let W^i be the expected matrix of weights corresponding to the event that agent i is broadcasting, where the expectation is taken with respect to the link failure events for the

links $\{i,j\} \in \mathscr{E}$ with $j \in \mathcal{N}(i)$. Using the definition of the weighted Laplacian, we can see that $\mathbf{1}^T W^i$ is given by $\mathbf{1}^T W^i = \mathbf{1}^T + \beta[L_\Pi]_i$, where $[L_\Pi]_i$ denotes the ith row of the weighted Laplacian L_Π. Since each agent broadcasts with probability $\frac{1}{m}$, we have

$$\frac{1}{m}\mathsf{E}\big[W_k^T \mathbf{1}\mathbf{1}^T W_k\big] = \frac{1}{m^2}\sum_{i=1}^{m}\left(\mathbf{1}^T + \beta[L_\Pi]_i\right)^T\left(\mathbf{1}^T + \beta[L_\Pi]_i\right).$$

Hence, for any z such that $z^T\mathbf{1} = 0$, we have

$$\frac{1}{m}z^T\mathsf{E}\big[W_k^T\mathbf{1}\mathbf{1}^T W_k\big]z = \frac{1}{m^2}\sum_{i=1}^{m}\beta^2\left([L_\Pi]_i z\right)^2.$$

Since $[L_\Pi]_i$ is the ith row of the weighted Laplacian L_Π, we have $\sum_{i=1}^{m}\left([L_\Pi]_i z\right)^2 = \|L_\Pi z\|^2$, implying $\frac{1}{m}z^T\mathsf{E}\big[W_k^T\mathbf{1}\mathbf{1}^T W_k\big]z = \frac{\beta^2}{m^2}\|L_\Pi z\|^2$. Therefore, from (25) we have

$$\max_{\{z \mid z^T\mathbf{1}=0\}} z^T\mathsf{E}\big[D_k^T D_k\big]z \le \|z\|^2 - \frac{\beta^2}{m^2}\min_{\{z \mid z^T\mathbf{1}=0\}}\|L_\Pi z\|^2 = \left(1 - \frac{\beta^2}{m^2}\lambda_2^2(L_\Pi)\right)\|z\|^2,$$

where the last inequality follows by relation (26). Hence, from the preceding relation and equality (24), we conclude that $\lambda_1(\mathsf{E}\big[D_k^T D_k\big]) \le 1 - \frac{\beta^2}{m^2}\lambda_2^2(L_\Pi) < 1$. ∎

Proof: **Lemma 3**. Note that $\Gamma_k(i) = \sum_{t=1}^{k}\chi_{E_{i,t}}$, where $E_{i,t}$ is the event that agent i updates at time t and $\chi_{E_{i,t}}$ is the indicator function of the event $E_{i,t}$. Since the events $\{\chi_{E_{i,t}}\}$ are i.i.d. with mean $\mathsf{E}\big[\chi_{E_{i,t}}\big] = \gamma_i$ for each $i \in V$, by the law of iterated logarithms (see [9], pages 476–479), we have for any $q > 0$, with probability 1,

$$\lim_{k\to\infty}\frac{|\Gamma_k(i) - k\gamma_i|}{k^{\frac{1}{2}+q}} = 0 \qquad \text{for all } i \in V.$$

Thus, for sufficiently large $\tilde{k}$ (depending on q and m), we have

$$\frac{|\Gamma_k(i) - k\gamma_i|}{k^{\frac{1}{2}+q}} \le \frac{1}{m^2} \qquad \text{for all } k \ge \tilde{k} \text{ and } i \in V, \tag{27}$$

implying that with probability 1 for all $i \in V$ and $k \ge \tilde{k}$,

$$\Gamma_k(i) \ge k\gamma_i - \frac{1}{m^2}k^{\frac{1}{2}+q} = \left(k^{\frac{1}{2}-q}\gamma_i - \frac{1}{m^2}\right)k^{\frac{1}{2}+q}.$$

For $q < 1/2$, the term $k^{\frac{1}{2}-q}\gamma_i$ tends to infinity as k increases. Thus, we can choose a larger $\tilde{k}$ (if needed) so that with probability 1 we have

$$k^{\frac{1}{2}-q}\gamma_i - \frac{1}{m^2} \ge \frac{1}{2}k^{\frac{1}{2}-q}\gamma_i \qquad \text{for all } k \ge \tilde{k} \text{ and } i \in V.$$

By combining the preceding two relations, we obtain $\Gamma_k(i) \ge \frac{1}{2}k\gamma_i$ with probability 1 for $k \ge \tilde{k}$ and $i \in V$. Therefore, for any $q \in (0, 1/2)$ we have with probability 1,

$$\frac{1}{\Gamma_k(i)} \le \frac{2}{k\gamma_i} \qquad \text{for all } k \ge \tilde{k} \text{ and } i \in V, \tag{28}$$

thus showing the first relation of the lemma in view of $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$.

Since agent i updates with probability $\gamma_i = \frac{1}{m} \sum_{j \in \mathcal{N}(i)} p_{ij}$ (cf. (12)), it follows

$$\gamma_i \geq \frac{1}{m} \left(\min_{\{i,j\} \in \mathcal{E}} p_{ij} \right) |\mathcal{N}(i)| \geq \frac{1}{m} \left(\min_{\{i,j\} \in \mathcal{E}} p_{ij} \right), \tag{29}$$

where the last inequality follows from $|\mathcal{N}(i)| \geq 1$ (since the graph $(V, \mathcal{E})$ has no isolated node). By letting $p_{\min} = \min_{\{i,j\} \in \mathcal{E}} p_{ij}$, and using $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$ and (28)–(29), we obtain with probability 1 $\alpha_{i,k}^2 \leq \frac{4m^2}{k^2 p_{\min}^2}$ for all $k \geq \tilde{k}$ and $i \in V$.

We now consider $\left| \alpha_{i,k} - \frac{1}{k\gamma_i} \right|$. We have with probability 1 for all $k \geq \tilde{k}$ and $i \in V$,

$$\left| \alpha_{i,k} - \frac{1}{k\gamma_i} \right| = \frac{1}{k\gamma_i} \frac{1}{\Gamma_k(i)} |k\gamma_i - \Gamma_k(i)| \leq \frac{2}{k^2 \gamma_i^2} |k\gamma_i - \Gamma_k(i)|,$$

where the inequality follows by (28). From relations (27) and (29), we obtain with probability 1, for all $k \geq \tilde{k}$ and $i \in V$,

$$\left| \alpha_{i,k} - \frac{1}{k\gamma_i} \right| \leq \frac{2m^2}{k^2 \, p_{\min}^2} \frac{1}{m^2} k^{\frac{1}{2}+q} = \frac{2}{k^{\frac{3}{2}-q} \, p_{\min}^2}.$$

∎

Proof: **Lemma 4**. We will consider coordinate-wise relations, by defining the vector $z_k^\ell \in \mathbb{R}^m$ as the vector with entries $[x_k^i]_\ell$, $i = 1, \ldots, m$. From the definition of the method in (9), we have

$$z_k^\ell = W_k z_{k-1}^\ell + \zeta_k^\ell \qquad \text{for } \ell = 1, \ldots, n, \text{ and all } k \geq 1, \tag{30}$$

where $\zeta_k^\ell \in \mathbb{R}^m$ is a vector with coordinates $[\zeta_k^\ell]_i$ given by

$$[\zeta_k^\ell]_i = \left[P_X[v_k^i - \alpha_{i,k} \left(\nabla f_i(v_k^i) + \epsilon_k^i \right)] - v_k^i \right]_\ell \quad \text{if } i \in J_k \quad \text{and} \quad \text{otherwise } [\zeta_k^\ell]_i = 0. \tag{31}$$

Furthermore, note that $[\bar{y}_k]_\ell$ is the average of the entries of the vector z_k^ℓ, i.e.,

$$[\bar{y}_k]_\ell = \frac{1}{m} \mathbf{1}^T z_k^\ell \qquad \text{for all } k \geq 0. \tag{32}$$

From now on, let $\ell \in \{1, \ldots, n\}$ be an arbitrary, but fixed coordinate index. By relations (30) and (32), we have $[\bar{y}_k]_\ell = \frac{1}{m} \left(\mathbf{1}^T W_k z_{k-1}^\ell + \mathbf{1}^T \zeta_k^\ell \right)$ for all k, implying

$$z_k^\ell - [\bar{y}_k]_\ell \mathbf{1} = \left(W_k - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k \right) z_{k-1}^\ell + \left(I - \frac{1}{m} \mathbf{1}\mathbf{1}^T \right) \zeta_k^\ell,$$

where I denotes the identity matrix. By Lemma 2, the matrices W_k are stochastic, so that $W_k \mathbf{1} = \mathbf{1}$. Thus $\left(W_k - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k \right) \mathbf{1} = 0$, implying $\left(W_k - \frac{1}{m} \mathbf{1}\mathbf{1}^T W_k \right) [\bar{y}_{k-1}]_\ell \mathbf{1} = 0$. Hence,

$$z_k^\ell - [\bar{y}_k]_\ell \mathbf{1} = D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}) + M\zeta_k^\ell \qquad \text{for all } k \geq 1, \tag{33}$$

where $D_k = W_k - \frac{1}{m}\mathbf{1}\mathbf{1}^T W_k$ and $M = I - \frac{1}{m}\mathbf{1}\mathbf{1}^T$. By taking the norm and then, the expectation conditioned on the past history, from the preceding relation we have for $k \geq 1$,

$$\mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| \mid \mathcal{F}_{k-1}\big] \leq \mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\| \mid \mathcal{F}_{k-1}\big] + \mathsf{E}\big[\|M\zeta_k^\ell\| \mid \mathcal{F}_{k-1}\big]. \quad (34)$$

We estimate the term $\mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| \mid \mathcal{F}_{k-1}\big]$ by using the fact the matrix W_k is independent of the past history $\mathcal{F}_{k-1}$, as follows:

$$\mathsf{E}\Big[\big\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\big\|^2 \mid \mathcal{F}_{k-1}\Big] \leq \lambda \|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\|^2, \quad (35)$$

where $\lambda = \|\mathsf{E}\big[D_k^T D_k\big]\|$ and $\lambda < 1$ (Lemma 2). Using $\mathsf{E}[\|x\|] \leq \sqrt{\mathsf{E}[\|x\|^2]}$, we obtain

$$\mathsf{E}\big[\big\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\big\| \mid \mathcal{F}_{k-1}\big] \leq \sqrt{\lambda}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\| \qquad \text{for all } k \geq 1. \quad (36)$$

We next estimate the second term in (34). The matrix $M = I - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ is a projection matrix (it projects on the subspace orthogonal to the vector $\mathbf{1}$), so that $\|M\|^2 = 1$, implying that $\|M\zeta_k^\ell\|^2 \leq \|\zeta_k^\ell\|^2$ for all k. Using this and the definition of ζ_k^ℓ in (31), we obtain

$$\|M\zeta_k^\ell\|^2 \leq \sum_{i \in J_k} \big|\big[P_X[v_k^i - \alpha_{i,k}\left(\nabla f_i(v_k^i) + \epsilon_k^i\right)] - v_k^i\big]_\ell\big|^2. \quad (37)$$

At this point the proofs for parts (a) and (b) for the (random) stepsize $\alpha_{i,k} = \frac{1}{\Gamma_k(i)}$ and the constant stepsize $\alpha_{i,k} = \alpha_i$ are different, and we consider them separately.

(a)   Using the relation $v_k^i \in X$, the nonexpansive property of the projection operation, and $\alpha_{i,k}^2 \leq \frac{4m^2}{k^2}$ (holding by Lemma 3), from inequality (37) we obtain with probability 1, for all i, for large enough $\tilde{k}$ and all $k \geq \tilde{k}$,

$$\big\|M\zeta_k^\ell\big\|^2 \leq \sum_{i \in J_k} \alpha_{i,k}^2 \|\nabla f_i(v_k^i) + \epsilon_k^i\|^2 \leq \frac{4m^2}{k^2\, p_{\min}^2} \sum_{i \in J_k} \big\|\nabla f_i(v_k^i) + \epsilon_k^i\big\|^2.$$

Since the subgradients are bounded and the error norms are bounded (cf. Assumptions 3 and 4), we have $\mathsf{E}\Big[\|\nabla f_i(v_k^i) + \epsilon_k^i\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k\Big] \leq (C + \nu)^2$ for $i \in J_k$ (see Eq. (7)), implying by $|J_k| \leq m$ that

$$\mathsf{E}\Big[\big\|M\zeta_k^\ell\big\|^2 \mid \mathcal{F}_{k-1}\Big] \leq \frac{4m^3}{k^2\, p_{\min}^2} (C + \nu)^2. \quad (38)$$

By inequality $\mathsf{E}[\|x\|] \leq \sqrt{\mathsf{E}[\|x\|^2]}$, we obtain for all $k \geq \tilde{k}$ with probability 1,

$$\mathsf{E}\big[\big\|M\zeta_k^\ell\big\| \mid \mathcal{F}_{k-1}\big] \leq \frac{2m\sqrt{m}}{k\, p_{\min}} (C + \nu). \quad (39)$$

By relations (36) and (39), from inequality (34) we have for all $k \geq \tilde{k}$,

$$\mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| \mid \mathcal{F}_{k-1}\big] \leq \sqrt{\lambda}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\| + \frac{2m\sqrt{m}}{k\, p_{\min}}\,(C+\nu).$$

Therefore, with probability 1 we have for $k \geq \tilde{k}$,

$$
\begin{aligned}
\frac{1}{k}\,\mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| \mid \mathcal{F}_{k-1}\big] &\leq \frac{1}{k-1}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\| \\
&\quad -\frac{1-\sqrt{\lambda}}{k}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\| + \frac{2m\sqrt{m}}{k^2\, p_{\min}}\,(C+\nu),
\end{aligned}
$$

where $1 - \sqrt{\lambda} > 0$ since $\lambda \in (0,1)$. By Lemma 1, it follows that

$$\sum_{k=\tilde{k}} \frac{1}{k}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\| < \infty \qquad \text{with probability 1}, \tag{40}$$

implying that $\sum_{k=1}^{\infty} \frac{1}{k}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\| < \infty$ with probability 1, for any $\ell = 1,\ldots,n$. This and the definition of z_k^ℓ, being $z_k^\ell = ([x_k^1]_\ell,\ldots,[x_k^m]_\ell)^T$, implies that $\sum_{k=1}^{\infty} \frac{1}{k}\|x_{k-1}^i - \bar{y}_{k-1}\mathbf{1}\| < \infty$ for all i, with probability 1.

We next prove that $\lim_{k\to\infty}\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| = 0$ with probability 1. As a consequence of relation (40), it follows that $\liminf_{k\to\infty}\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| = 0$ with probability 1. To complete the proof, we only need to prove almost sure convergence of $\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|$ as $k \to \infty$.

By taking the square norm in (33) and then, the expectation conditioned on the past and using Hölder's inequality (6), we obtain for all k,

$$
\begin{aligned}
\mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2 \mid \mathcal{F}_{k-1}\big] &\leq \mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2 \mid \mathcal{F}_{k-1}\big] + \mathsf{E}\big[\|M\zeta_k^\ell\|^2 \mid \mathcal{F}_{k-1}\big] \\
&\quad +2\sqrt{\mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2 \mid \mathcal{F}_{k-1}\big]}\,\sqrt{\mathsf{E}\big[\|M\zeta_k^\ell\|^2 \mid \mathcal{F}_{k-1}\big]}.
\end{aligned}
$$

Combining the relations in (35) and (38) with the preceding inequality, with probability 1 we obtain for all $k \geq \tilde{k}$,

$$
\begin{aligned}
\mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2 \mid \mathcal{F}_{k-1}\big] &\leq \lambda\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\|^2 + \frac{4m^3}{k^2\, p_{\min}^2}\,(C+\nu)^2 \\
&\quad +\sqrt{\lambda}\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\|\frac{2m\sqrt{m}}{k\, p_{\min}}\,(C+\nu).
\end{aligned}
$$

Taking into account that $\lambda < 1$ and $\sum_{k=\tilde{k}}^{\infty} \frac{1}{k}\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\| < \infty$ with probability 1 (see Eq. (40)), we can apply the supermartingale convergence result of Lemma 1 to conclude that $\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|$ converges with probability 1 for any $\ell = 1,\ldots,n$. This, and the relation $z_k^\ell = ([x_k^1]_\ell,\ldots,[x_k^m]_\ell)$ imply that $\|x_k^i - \bar{y}_k\mathbf{1}\|$ converges with probability 1 for every i.

(b) Let $\alpha_{i,k} = \alpha_i > 0$. From relation (33), we have for $\ell = 1, \ldots, n$ and $k \geq 1$,

$$\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2 \leq \|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2 + \|M\zeta_k^\ell\|^2 + 2\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\| \, \|M\zeta_k^\ell\|.$$

By summing these relations over $\ell = 1, \ldots, n$, and then taking the expectation and using Hölder's inequality (6), we obtain for all $k \geq 1$,

$$\sum_{\ell=1}^n \mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2\big] \leq \left(\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2\big]} + \sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|M\zeta_k^\ell\|^2\big]} \right)^2.$$

Using the estimates in (35) and (37), with $\alpha_{i,k} = \alpha_i$, we see that

$$\sum_{\ell=1}^n \mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1})\|^2\big] \leq \lambda \sum_{\ell=1}^n \|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell \mathbf{1}\|^2,$$

$$\sum_{\ell=1}^n \mathsf{E}\big[\|M\zeta_k^\ell\|^2\big] \leq \mathsf{E}\left[\mathsf{E}\left[\sum_{j \in J_k} \alpha_i^2 \|\nabla f_i(v_k^i) + \epsilon_k^i\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k \right] \right] \leq \alpha_{\max}^2 d_{\max}(C + \nu)^2,$$

where $\alpha_{\max} = \max_i \alpha_i$, $d_{\max} = \max_{1 \leq i \leq m} |\mathsf{E}[J(i)]|$, and we use relation (7) in the last inequality. In the definition of $d_{\max}$, the set $J(i)$ is the random set of agents that update given that agent i broadcasts, and the expectation is taken with respect to the link failure probabilities. Letting $u_k = \sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2\big]}$ from the preceding three relations, we obtain for all $k \geq 1$,

$$u_k \leq \sqrt{\lambda}\, u_{k-1} + \sqrt{d_{\max}}\,(C + \nu) \leq \cdots \leq \sqrt{\lambda^k}\, u_0 + \alpha_{\max} \frac{\sqrt{d_{\max}}}{1 - \sqrt{\lambda}}\,(C + \nu).$$

Since $\lambda < 1$ by Lemma 2, it follows

$$\limsup_{k \to \infty} u_k \leq \alpha_{\max} \frac{\sqrt{d_{\max}}}{1 - \sqrt{\lambda}}\,(C + \nu). \tag{41}$$

Using $z_k^\ell = ([x_k^1]_\ell, \ldots, [x_k^m]_\ell)^T$, we see that

$$u_k = \sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell \mathbf{1}\|^2\big]} = \sqrt{\sum_{i=1}^m \mathsf{E}[\|x_k^i - \bar{y}_k\|^2]}. \tag{42}$$

Furthermore, by Hölder's inequality (6), we have

$$\sum_{i=1}^m \mathsf{E}\big[\|x_k^i - \bar{y}_k\|\big] \leq \sqrt{m} \sqrt{\sum_{i=1}^m \mathsf{E}[\|x_k^i - \bar{y}_k\|^2]}. \tag{43}$$

By (41)–(43), we obtain $\limsup_{k \to \infty} \sum_{i=1}^m \mathsf{E}[\|x_k^i - \bar{y}_k\|] \leq \alpha_{\max} \frac{\sqrt{m d_{\max}}}{1 - \sqrt{\lambda}}\,(C + \nu)$. ∎

Proof: **Lemma 6**. From the given relation for d_k and u_{k-1}, we can see by induction (on k) that $d_k \leq c^k d_0 + \sum_{t=0}^{k-1} c^{k-t-1} u_t$ for all $k \geq 1$. Since $c \in (0, 1)$, it follows that $\limsup_{k \to \infty} d_k \leq \limsup_{k \to \infty} \sum_{t=0}^{k-1} c^{k-t-1} u_t$. Thus, it remains to show that

$$\limsup_{k \to \infty} \sum_{t=0}^{k} c^{k-t} u_t \leq \frac{1}{1-c} \limsup_{t \to \infty} u_t. \tag{44}$$

Let $\gamma = \limsup_{k \to \infty} u_k$. If $\gamma = +\infty$, the relation is satisfied. Let $\epsilon > 0$ be arbitrary but small and let $M > 0$ be large. Define $a = \gamma + \epsilon$ if γ is finite and $a = -M$ if $\gamma = -\infty$. Choose index K large enough so that $u_k \leq a$ for all $k \geq K$. We then have for $k \geq K$,

$$\sum_{t=0}^{k} c^{k-t} u_t = \sum_{t=0}^{K} c^{k-t} u_t + \sum_{t=K+1}^{k} c^{k-t} u_t \leq \max_{0 \leq s \leq K} u_s \sum_{t=0}^{K} c^{k-t} + a \sum_{t=K+1}^{k} c^{k-t}.$$

Since $\sum_{t=K+1}^{k} c^{k-t} \leq \frac{1}{1-c}$ and $\sum_{t=0}^{K} c^{k-t} \leq \frac{c^{k-K}}{1-c}$, it follows that $\sum_{t=0}^{k} c^{k-t} u_t \leq \left(\max_{0 \leq s \leq K} u_s \right) \frac{c^{k-K}}{1-c} + \frac{a}{1-c}$ for all $k \geq K$. Thus, $\limsup_{k \to \infty} \sum_{t=0}^{k} c^{k-t} u_t \leq \frac{a}{1-c}$, and relation (44) follows by the definition of a. ∎

REFERENCES

[1] T. C. Aysal, M. E. Yildriz, A. D. Sarwate, and A. Scaglione, *Broadcast gossip algorithms for consensus*, IEEE Transactions on Signal processing **57** (2009), 2748–2761.

[2] D. P. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex analysis and optimization*, Athena Scientific, Belmont, MA, 2003.

[3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, Belmont, MA, 1997.

[4] _____, *Gradient convergence in gradient methods with errors*, SIAM Journal of Optimization **10** (2000), no. 3, 627–642.

[5] P. Billingsley, *Probability and measure*, 3 ed., John Wiley and Sons, 1979.

[6] V. Borkar, *Asynchronous stochastic approximations*, SIAM Journal on Control and Optimization **36** (1998), no. 3, 840–851.

[7] _____, *Stochastic approximation: A dynamical viewpoint*, Cambridge University Press, 2008.

[8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Transactions on Information Theory **52** (2006), no. 6, 2508–2530.

[9] R. Dudley, *Real analysis and probability*, Cambridge University Press, 2002.

[10] Y. Ermoliev, *Stochastic quasi-gradient methods and their application to system optimization*, Stochastics **9** (1983), no. 1, 1–36.

[11] _____, *Stochastic quazigradient methods*, Numerical Techniques for Stochastic Optimization, Springer-Verlag, N.Y., 1988, pp. 141–186.

[12] A. Jadbabaie, J. Lin, and S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Transactions on Automatic Control **48** (2003), no. 6, 998–1001.

[13] B. Johansson, *On distributed optimization in networked systems*, Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden, 2008.

[14] B. Johansson, M. Rabi, and M. Johansson, *A simple peer-to-peer algorithm for distributed optimization in sensor networks*, Proceedings of the 46th IEEE Conference on Decision and Control, 2007, pp. 4705–4710.

[15] S. Kar and J.M.F. Moura, *Distributed consensus algorithms in sensor networks: Link failures and channel noise*, IEEE Transactions on Signal Processing **57** (2009), no. 1, 355–369.

[16] I. Lobel and A. Ozdaglar, *Distributed subgradient methods over random networks*, Lab. for Information and Decision Systems, MIT, Report 2800, 2008.

[17] B. Mohar, *Some aapplications of laplace eigenvalues of graphs*, Graph Symmetry: Algebraic Methods and Applications (G. Hahn and G. Sabidussi, eds.), NATO ASI Ser. C 497, Kluwer, 1997, pp. 225–275.

[18] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, *Distributed subgradient algorithms and quantization effects*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008, pp. 4177–4184.

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

[20] ———, *Cooperative distributed multi-agent optimization*, Convex Optimization in Signal Processing and Communications (Y. Eldar and D. Palomar, eds.), Cambridge University Press, 2010, pp. 340–386.

[21] A. Nedić, A. Ozdaglar, and P. A. Parrilo, *Constrained consensus*, to appear in IEEE Transactions on Automatic Control, 2009.

[22] R. Olfati-Saber and R. M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Transactions on Automatic Control **49** (2004), no. 9, 1520–1533.

[23] B. T. Polyak, *Introduction to optimization*, Optimization Software Inc., 1987.

[24] M. G. Rabbat and R. D. Nowak, *Quantized incremental algorithms for distributed optimization*, IEEE Journal on Select Areas in Communications **23** (2005), no. 4, 798–808.

[25] S. Sundhar Ram, A. Nedić, and V.V. Veeravalli, *Asynchronous gossip algorithms for stochastic optimization*, accepted in CDC Conference, 2009.

[26] ———, *Distributed stochastic subgradient projection algorithms for convex optimization*, submitted, 2009.

[27] ———, *Incremental stochastic sub-gradient algorithms for convex optimization*, SIAM Journal on Optimization **20** (2009), no. 2, 691–717.

[28] S. Sundhar Ram, V.V. Veeravalli, and A. Nedić, *Distributed and non-autonomous power control through distributed convex optimization*, The 28th IEEE Conference on Computer Communications INFOCOM, 2009, pp. 3001–3005.

[29] ———, *Distributed and recursive estimation*, Sensor Networks: When theory meets practice (G. Ferrari, ed.), Springer, 2010.

[30] J. N. Tsitsiklis, *Problems in decentralized decision making and computation*, Ph.D. thesis, Massachusetts Institute of Technology, 1984.

[31] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control **31** (1986), no. 9, 803–812.

[32] L. Xiao, S. Boyd, and S.-J. Kim, *Distributed average consensus with least mean square deviation*, Journal of Parallel and Distributed Computing **67** (2007), no. 1, 33–46.