Distributed Asynchronous Constrained Stochastic Optimization

Kunal Srivastava and Angelia Nedić

Abstract

In this paper we study two problems which often occur in various applications arising in wireless sensor networks. These are the problem of reaching an agreement on the value of local variables in a network of computational agents and the problem of cooperative solution to a convex optimization problem, where the objective function is the aggregate sum of local convex objective functions. We incorporate the presence of a random communication graph between the agents in our model as a more realistic abstraction of the gossip and broadcast communication protocols of a wireless network. An added ingredient is the presence of local constraint sets to which the local variables of each agent is constrained. Our model allows for the objective functions to be nondifferentiable and accommodates the presence of noisy communication links and subgradient errors. For the consensus problem we provide a diminishing step size algorithm which guarantees asymptotic convergence. The distributed optimization algorithm uses two diminishing step size sequences to account for communication noise and subgradient errors. We establish conditions on these step sizes under which we can achieve the dual task of reaching consensus and convergence to the optimal set with probability one. In both cases we consider the constant step size behavior of the algorithm and establish asymptotic error bounds.

I. Introduction

There has been a sustained effort in the research community over the years to develop algorithms for distributed decision making and control. The main driver for these problems are the more application specific problems arising in wireless and sensor networks, transmission control protocols for the internet, distributed machine learning, multi-vehicle coordination and more recently social networks. The two main mathematical abstractions which have been employed to address these problems are the problem

of reaching consensus on the decision variables [1]–[5] in a network of computational agents and the problem of cooperative solution to distributed optimization problems [6]–[11]. The algorithms for reaching consensus have proven useful in a wide variety of contexts from formation control [3], distributed parameter estimation [12], [13], load balancing [14], to synchronization of Kuramoto oscillators [15]. The problem of distributed optimization, where the objective is to minimize a sum of convex functions appears widely in the context of wireless and sensor networks [16]–[18]. A more recent application area for distributed optimization is the problem of distributed machine learning. In many machine learning applications it is highly desirable to come up with distributed schemes to solve an optimization problem as the ubiquity of large and distributed data sets makes it impractical to solve the problem in a centralized fashion [19], [20]. In many cases it is not possible to store the massive amount of data at the node, which makes algorithms that rely on multiple iterations over the data sets infeasible. This feature of the problem makes stochastic gradient descent algorithms attractive for online learning problems, since these algorithms typically require a single pass over the data. A related problem to distributed optimization is the problem of fair allocation of resources. This has been thoroughly studied in the area of microeconomics [21]. Recent interest in the resource allocation problem has arisen in the context of utility maximization in communication networks [22]–[24]. One of the most important characteristics of the network utility maximization problem is the fact that the objective function to be minimized has a separable form. Under this structure various primal or dual decomposition methods can be applied to make the problem amenable to a distributed solution.

In this paper we deal with both the consensus problem and the problem of distributed optimization when the objective function has an additive, but not necessarily a separable structure. Most of the work on the consensus problem deals with the unconstrained case when the variables on which the nodes need to agree are free to lie in the Euclidean space. We deal with the case when the variables which are local to nodes are also constrained to lie in closed convex sets. The constraint set for each local variable is private information to the node. The objective is to design an algorithm which is adapted to the time varying random nature of the underlying communication graph between nodes and guarantees asymptotic consensus on the local variables while maintaining the feasibility of each variable with respect to its constraint set. A distributed algorithm for this problem was proposed in [11]. However, the analysis there was restricted to the case of a deterministic network, with noiseless communication links. Unlike [11], in this paper we consider the case when the communication graph is random and the communication links are noisy. In this case we extend the algorithm proposed in [11] by introducing a step size sequence that attenuates the communication noise. A new feature of the algorithm is that the step sizes of the agents are

asynchronous and are a function of their local clocks. We establish convergence of the algorithm (with probability one) for such uncoordinated but diminishing step sizes, and we provide error bounds for the case when the step sizes are constant. Then, we consider a distributed constrained stochastic optimization problem. A distributed optimization algorithm for the case when the objective functions are deterministic functions has been proposed in [11], but its convergence analysis was limited to two special cases: when the local constraint sets are identical and when the network is fully connected (requiring the nodes to use uniform weights). This present paper considers a more general problem than [11], by fully studying the presence of local constraint sets and noisy communication along with the presence of stochastic errors in the evaluation of subgradients. Once again we consider a random communication network. In this case we need to introduce two step size sequences to damp out both communication noise and subgradient errors arising from considering the stochastic optimization problem. Our distributed optimization algorithm is asynchronous in nature and, for each agent, both the step sizes are functions of the agent's local clock. We prove that if the step size damping the subgradient error decays fast enough when compared to the step size attenuating the communication noise, then the algorithm converges to a common point in the optimal set with probability one. We also provide an error bound for the case when the agent step sizes are uncoordinated but constant.

Our model of the random communication network is general enough to include both the gossip communication protocol of [25] and the broadcast protocol of [26]. We also consider the case when the algorithm is employed by using constant step sizes. As expected, in this case it is not possible to achieve convergence to the optimal set with probability one. Instead, we derive asymptotic error bounds on the iterates of the algorithm under some additional assumptions.

Consensus over noisy links in the lack of constraint sets has been studied in [27], [28] and [13] among others. In [7], the authors studied the distributed optimization problem in the presence of subgradient errors. However, the paper assumes a common constraint set and the absence of communication noise. A related asynchronous scheme for the distributed optimization problem is the Markov incremental algorithm proposed in [8]. According to this algorithm the index of agent updating the common decision vector is decided by the evolution of a Markov chain. A limitation of this algorithm is its serial nature. In [9], the authors consider a random network model for communication among agents, but the optimization problem is unconstrained and the proposed approach is synchronous in nature in the sense that the step sizes of the different agents are coordinated. Another relevant paper which considers the unconstrained random network model for consensus is [29].

The rest of the paper is organized as follows. In Section II we setup the problem, state our algorithms

for both the problem of constrained consensus and distributed optimization, and introduce our main assumptions. In Section III we state some results regarding asymptotic properties of some deterministic and random sequences which will be useful in deriving our result on asymptotic convergence and the asymptotic error bounds. In Section IV we tackle the problem of constrained consensus. We prove the asymptotic convergence result of our algorithm. We also consider the case of constant step size and establish an asymptotic error bound for averaged iterates. Next, in Section V we provide an asymptotic convergence result for our asynchronous distributed optimization algorithm, and derive an asymptotic error bound for the case of constant step sizes. Finally in Section VI we provide the conclusion, and in the Appendix, we prove one auxiliary result stated in Section III.

Notation and Terminology. The jth component of a vector x is denoted by x_j. In many instances when there are m vectors present, we denote the ith vector as x_i. For an $m \times n$ real valued matrix $A \in \mathbb{R}^{m \times n}$, we denote the element in the ith row and j^{th} column of A by A_{ij} or $[A]_{ij}$. We use $\| \cdot \|$ to denote the Euclidean norm, and $P_X[x]$ to denote the projection of a vector x on the set X.

We write I_r for the $r \times r$-dimensional identity matrix and $\mathbf{1}_m$ for the m-dimensional vector with each component equal to 1. We use $D(a_i)$ to denote a diagonal matrix with diagonal entries given by $\{a_1, \ldots, a_\ell\}$. The size of the diagonal matrix is thus given by the number of values the index i takes. The null space of a matrix A is denoted by $\mathcal{N}(A)$. An $m \times m$ matrix W is stochastic if $W_{ij} \geq 0$ for all i, j, and $W\mathbf{1}_m = \mathbf{1}_m$.

Given a directed graph $G = (V, E)$, the link $(i, j) \in E$ is to be interpreted as the incoming edge from j to i. For a bidirectional graph G, we have $(i, j) \in E$ if and only if $(j, i) \in E$. We will sometimes denote the edge set of a graph G as $\mathcal{E}(G)$. Given any graph $G = (V, E)$ and a function $F : V \times V \to \mathbb{R}$, we use $\sum_E F(i, j)$ to denote the sum where the function $F(i, j)$ is evaluated for all $(i, j) \in E$. When the graph G has bidirectional links, the sum $\sum_E F(i, j)$ is assumed to be evaluated by taking every edge only once. We use the terms "agent" and "node" interchangeably. We say that agent j is a *neighbor* of agent i if $(i, j) \in E$, and we denote the set of all neighbors of agent i by N_i. A graph $G = (V, E)$ is *r-regular* if $|N_i| = r$ for each node i. The Laplacian of a graph G is a matrix L such that $L_{ij} = -1$ if $(i, j) \in E$, $L_{ii} = |N_i|$ and $L_{ij} = 0$ for all $(i, j) \notin E$. For a bidirectional graph, the matrix L is symmetric, positive semidefinite, and satisfies $L\mathbf{1} = 0$ and $\mathbf{1}'L = 0$. If the graph G is connected then $\{c\mathbf{1} : c \in \mathbb{R}\}$ is the unique null space of the matrix L. Given m vectors in $\mathbb{R}^n$, $\{x_1, \ldots, x_m\}$, the consensus subspace is the subspace of the mn-dimensional product space, and it is defined as:

$$\mathcal{C} = \{\mathbf{z} \in \mathbb{R}^{mn} : \mathbf{z} = \mathbf{1}_m \otimes z, \ z \in \mathbb{R}^n\},$$

which is the subspace of m-copies of the same n-dimensional vector. Thus, the vectors $x_1, \ldots, x_m$ are in consensus if the concatenated vector $\mathbf{x} = (x_1', \ldots, x_m')'$ lies in the consensus subspace, i.e., $\mathbf{x} \in \mathcal{C}$.

Given a finite set of scalars $\{\alpha_i\}_{i \in \mathcal{I}}$, we let $\bar{\alpha} = \max_i \{\alpha_i\}$ and $\underline{\alpha} = \min_i \{\alpha_i\}$. Furthermore, we let $\Delta_\alpha = \bar{\alpha} - \underline{\alpha}$. We use $\chi_{\{p\}}$ to denote the boolean indicator function which takes the value 1 when the statement p is $true$, and 0 when p is $false$. Given a convex not necessarily differentiable function $f(x)$, a vector d is a subgradient of f at x if the following relation holds for every z in the domain of f:

$$d'(z - x) \le f(z) - f(x).$$

The set of all subgradients of a function f at a point x is the subdifferential set, denoted by $\partial f(x)$.

II. NETWORK MODEL, OUTLINE OF ALGORITHMS AND ASSUMPTIONS

In this section we introduce the main problems we consider. We provide our algorithms for these problems and discuss the assumptions we use in our convergence analysis. We consider a setup with a set of m agents, which can be viewed as the node set $V = \{1, \ldots, m\}$. We further assume that the possible communication links among agents are given by a bidirectional graph $G = (V, E)$. The communication graph at time slot k is represented by the random graph $\mathcal{W}(k) = (V, \mathcal{E}(k))$, with $\mathcal{E}(k) \subset E$. Clearly the random graph $\mathcal{W}(k)$ is a subgraph of the graph G. Note that $\mathcal{W}(k)$ is not required to be a bidirectional graph. The use of random communication graphs enables us to include various communication protocols which are prevalent in the wireless network literature such as gossip based protocols [25] and broadcast based protocol [26].

In the current work we are mainly concerned with the problem of *constrained consensus* and *distributed optimization* in the presence of various uncertainties including noisy communication links, random communication graphs, and stochastic errors in the evaluation of subgradients of the objective function. We will model the communication events (when any two agents i and j communicate) as occurring at the ticks of a global Poisson clock T_p. It has been shown [25], [26], that this abstraction fits the framework of asynchronous computation in both the gossip and broadcast framework. Let us denote the local decision variable associated with node i as $x_{i,k}$. Each local decision variable $x_{i,k}$ is restricted to lie in a local constraint set X_i.

A. Constrained consensus

The constrained consensus problem is to achieve asymptotic consensus on the local decision variables through information exchange with the neighboring nodes in the presence of the constraint sets. Alternatively, the problem can be cast as a quest for a distributed algorithm for the following optimization

problem:

$$\text{minimize} \quad \sum_{i=1}^{m} \|x - P_{X_i}[x]\|^2$$

$$\text{subject to} \quad x \in \mathbb{R}^n.$$

When the intersection $\cap_{i=1}^{m} X_i$ is nonempty, a solution to the above problem is given by any vector x which lies in the intersection. Clearly, in this case the objective function value is zero, which is also the optimal value. A distributed algorithm for this problem was proposed in [11]. In the algorithm agent i's local variable $x_{i,k}$ evolves as follows:

$$x_{i,k+1} = P_{X_i} \left[\sum_{j=1}^{m} r_{ij,k+1} x_{j,k} \right],$$

where $r_{ij,k+1}$ denotes the weight assigned by node i to the estimate coming from node j. A crucial assumption needed in the analysis in [11] was the requirement that if agent i receives data from agent j then $r_{ij,k} \geq \eta > 0$, uniformly in k. We are interested in the case when the communication links are noisy and, hence, node i has access to a noise corrupted value of its neighbor's local estimate. In this case it is detrimental to impose the requirement that $r_{ij,k} \geq \eta$ since we need to asymptotically damp the impact of the noise. We formulate our algorithm for the general case when, at the kth tick of the global Poisson clock T_p, the communication graph $\mathcal{W}(k) = (V, \mathcal{E}(k))$ is random, i.e., the edge set $\mathcal{E}(k)$ is random. A similar formulation to ours was carried out in [30], but without local set constraints. We will assume that the noise is additive in nature. With this in consideration, let us define the random variables $v_{i,k+1}$ for all $i \in V$, as follows:

$$v_{i,k+1} = x_{i,k} + \alpha_{i,k+1} \sum_{j=1}^{m} a_{ij,k+1}[x_{j,k} + \xi_{ij,k+1} - x_{i,k}], \tag{1}$$

where $\xi_{ij,k+1}$ is the additive noise on the link (i,j). The weights $a_{ij,k+1} > 0$ encode the random network and the relative weights node i imposes on the communicated variables from its neighboring nodes at instance $k+1$. If a node i doesn't receive any information from node j at instance $k+1$, then $a_{ij,k+1} = 0$. Clearly $a_{ij,k+1} = 0$ for all k, if the edges $(i,j) \notin \mathcal{E}(\mathcal{W}(k+1))$. We assign $a_{ii,k} = 0$, uniformly in k. If an agent i doesn't receive any information from its neighbors, then $v_{i,k+1} = x_{i,k}$. We define the matrix $W_k \in \mathbb{R}^{m \times m}$ such that $[W_k]_{ij} = -a_{ij,k}$ for $i \neq j$ and $[W_k]_{ii} = \sum_{j \neq i} a_{ij,k}$. Then, W_k has the property

$W_k \mathbf{1}_m = 0$ for all k. Under the new notation relation (1) can be written as

$$v_{i,k+1} = x_{i,k} - \alpha_{i,k+1} \sum_{j=1}^{m} [W_{k+1}]_{ij} x_{j,k}$$

$$- \alpha_{i,k+1} \sum_{j \neq i} [W_{k+1}]_{ij} \xi_{ij,k+1} \quad \text{for all } i \in V. \tag{2}$$

Define the set $U_k = \{ j \in V \mid [W_k]_{j\ell} < 0 \text{ for some } \ell \neq j \}$ and $\Gamma_{i,k} = \sum_{t=1}^{k} \chi_{\{i \in U_t\}}$. Essentially the set U_k denotes the set of agents updating their iterates at instance k, and $\Gamma_{i,k}$ denotes the number of times agent i has updated its value until time k. The asynchronous behavior of the algorithm becomes apparent in the way the agents use the step sizes $\alpha_{i,k}$. We start with the assumption that all the agents know the functional form of the step sizes $\alpha_{i,k}$, for example $\alpha_k = \frac{1}{k^\theta}$ with a parameter θ. For an agent i updating at time $k+1$, the step size is given by $\alpha_{i,k+1} = \alpha(\Gamma_{i,k+1})$. Let us denote the total noise experienced by agent i at time $k+1$ in Eq. (2) by $\xi_{i,k+1}$. Then, our algorithm for the constrained consensus problem is given as

$$v_{i,k+1} = x_{i,k} - \alpha_{i,k+1} \sum_{j=1}^{m} [W_{k+1}]_{ij} x_{j,k} - \alpha_{i,k+1} \xi_{i,k+1}$$

$$\xi_{i,k+1} = \sum_{j \neq i} [W_{k+1}]_{ij} \xi_{ij,k+1} \tag{3}$$

$$x_{i,k+1} = P_{X_i}[v_{i,k+1}],$$

where $P_{X_i}[\cdot]$ is the projection on the local constraint set X_i. The local variables always satisfy $x_{i,k} \in X_i$, and if at any instant the node i doesn't receive any information from its neighbors ($[W_{k+1}]_{ij} = 0$ for all $j \neq i$), then $x_{i,k+1} = x_{i,k}$. The initial points $x_{i,0} \in X_i$ are assumed to be random with finite expectations $\mathbb{E}[\|x_{i,0}\|^2]$ and independent of the other random variables involved in the communications.

B. Distributed optimization

We now discuss the problem of distributed optimization. We consider the distributed optimization problem when the objective function is a sum of m local convex objective functions corresponding to m agents. The objective of the agents is to cooperatively solve the following constrained optimization problem:

$$\text{minimize} \quad \sum_{i=1}^{m} f_i(x)$$

$$\text{subject to} \quad x \in X = \bigcap_{i=1}^{m} X_i,$$

where each $f_i : \mathbb{R}^n \to \mathbb{R}$ is a convex function, representing the local objective function of agent i, and each set $X_i \subseteq \mathbb{R}^n$ is compact and convex, representing the local constraint set of agent i. Since the objective function is continuous and the set X is compact, by Weierstrass theorem it follows that the optimal set is nonempty. Let us denote the optimal set by X^*. We assume that the local constraint set X_i and the objective function f_i are known to agent i only. In our formulation we are not restricted to the deterministic optimization. In many applications arising in sensor networks and distributed machine learning [31], [32] the local objective functions can take the form of the following stochastic optimization:

$$f_i(x) = \mathbb{E}_\omega[g_i(x, \omega)] + \Omega(x),$$

where ω is a random variable and the expectation is taken with respect to the distribution of ω. The term $\Omega(x)$ is a regularization term that is often included to improve the generalization ability [33]. Recently a lot of interest in signal processing has been generated towards the use of the l_1-norm as the regularization term. In many cases it has been shown that such a regularization yields sparse solutions. Since the l_1-penalty is nondifferentiable, our algorithm which doesn't require the objective function to be differentiable is suitable for this problem. It is well known that the stochastic optimization problems of the form above can be dealt with by using first-order stochastic gradient descent methods [34]–[38]. Such algorithms are also known as stochastic approximation algorithms. Our algorithm can be classified as a stochastic approximation based method.

We propose the following update mechanism for each agent:

$$v_{i,k+1} = x_{i,k} - \alpha_{i,k+1} \sum_{j=1}^{m} [W_{k+1}]_{ij} x_{j,k} - \alpha_{i,k+1} \xi_{i,k+1}$$

$$\xi_{i,k+1} = \sum_{j \neq i} [W_{k+1}]_{ij} \xi_{ij,k+1} \tag{4}$$

$$x_{i,k+1} = P_{X_i} \left[v_{i,k+1} - \gamma_{i,k+1} \tilde{d}_{i,k+1} \chi_{\{i \in U_{k+1}\}} \right].$$

The algorithm is an adjustment of the consensus algorithm in (3), where the agent i update of $x_{i,k+1}$ is modified to account for the local objective function f_i. The vector $\tilde{d}_{i,k+1}$ is a noisy subgradient of $f_i(x)$ at $x = v_{i,k+1}$, given by $\tilde{d}_{i,k+1} = d_{i,k+1} + \epsilon_{i,k+1}$ for all k, where the vector $d_{i,k+1}$ is a subgradient of $f_i(x)$ at $x = v_{i,k+1}$ and $\epsilon_{i,k}$ is the error associated with the evaluation of the subgradient $d_{i,k+1}$. A unique feature of our algorithm is the ability to handle both communication noise and subgradient error. The presence of two sources of noise makes it imperative to use two step size sequences $\alpha_{i,k}$ and $\gamma_{i,k}$ corresponding to the communication noise and subgradient error respectively. These stepsizes are functions of the number of times agent i updates its variable, as given later (see (10)). Analogous to

the consensus algorithm, here we also assume that the initial points $x_{i,0} \in X_i$ are random with finite expectations $\mathbb{E}[\|x_{i,0}\|^2]$ and independent of the other random variables involved in the process.

C. Assumptions and implications

In this section we introduce various assumptions which we use to prove convergence of our algorithms for both constrained consensus and the distributed optimization problem. We group the assumptions into three categories; namely, network assumptions, constraint sets and objective functions assumptions, and the assumptions on communication noise and subgradient errors. Typically in our results we only require a subset of the following assumptions to hold. To prevent imposing blanket assumptions for our results, we explicitly refer the assumptions required in each of our result.

Network Assumptions. Our first assumption lists all the conditions on the underlying communication graph that will be used, not necessarily all at once. Recall that the graph $G = (V, E)$ has the edge set E consisting of all possible communication links between nodes, and that $\mathcal{W}(k) = (V, \mathcal{E}(k))$ is the random communication graph at time k. Also, recall that the matrices W_k are defined by

$$[W_k]_{ij} = -a_{ij,k} \quad \text{for } i \neq j \text{ and } [W_k]_{ii} = \sum_{j \neq i} a_{ij,k}, \tag{5}$$

where $a_{ij,k}$ are the relative weights. We use the following conditions on the graph G and the weight matrices.

Assumption 1: Suppose that the following hold:

a) The graph $G = (V, E)$ is bidirectional, connected and without self-loops i.e., $(i, i) \notin E$ for all i.

b) For all edges $(i, j) \in \mathcal{E}(k)$, we have $0 < a_{ij,k} \leq \eta$ and $a_{ij,k} = 0$ otherwise.

c) The random matrix sequence $\{W_k\}$ is independent identically distributed (*i.i.d.*). The expected matrix $\bar{W} = \mathbb{E}[W_k]$ is positive semidefinite and the vector $\mathbf{1}_m$ is the unique (up to scaling) nonzero vector satisfying $\bar{W}\mathbf{1}_m = 0$.

d) The probability p_i of update at any instance for each node i is positive, i.e., $p_i > 0$ for all i.

By the definition of the weight matrices W_k in (5), Assumption 1-b implies that

$$-\eta \leq [W_k]_{ij} < 0 \quad \text{for } (i, j) \in \mathcal{E}(k),$$

$$[W_k]_{ij} = 0 \quad \text{for } (i, j) \notin \mathcal{E}(k). \tag{6}$$

Note that Assumption 1-c does not require that the random communication graph $\mathcal{W}(k)$ is connected at any instance. However, the expected graph $\bar{\mathcal{W}} = (V, \mathbb{E}[\mathcal{E}(k)])$ is connected in view of the assumption on the vector $\mathbf{1}_m$. As shown later in Lemmas 1 and 2, the expected weight matrix arising in gossip

and broadcast communication protocol satisfy the positive semidefiniteness condition of Assumption 1-c. Furthermore, under the $i.i.d$ condition of Assumption 1-c, the random variables $\chi_{\{i \in U_k\}}$ are also $i.i.d$ for any $i \in V$. Thus, the stationary probability p_i of update at any instance by agent i is well defined, and we have $p_i = \mathbb{E}\left[\chi_{\{i \in U_k\}}\right]$.

Constraints sets and objective functions. We here discuss the assumptions we make on the sets X_i and the functions f_i.

Assumption 2: Let the following conditions hold:

a) The constraint sets $X_i \subseteq \mathbb{R}^n$ are closed and convex. Their intersection set $X = \cap_{i=1}^m X_i$ is nonempty.

b) The constraint sets X_i are compact.

c) The intersection set X has a nonempty interior, i.e., there are $\bar{z} \in X$ and $\delta > 0$ such that $\{x : \|x - \bar{z}\| \leq \delta\} \subset X$.

d) The functions $f_i(x)$ are defined and convex on $\mathbb{R}^n$.

e) We assume that the subgradients of each function f_i are uniformly bounded i.e., there is a scalar $C > 0$ such that $\|d\| \leq C$ for all $d \in \partial f_i(x)$, all $x \in \mathbb{R}^n$, and all i.

When the sets X_i are compact, we will use the notation

$$C_{X_i} = \max_{x,y \in X_i} \|x - y\| \qquad \text{for all } i \in V.$$

Under Assumption 2-d, it is known [39] that each function $f_i(x)$ is differentiable almost everywhere. Wherever the functions are not differentiable, the subdifferential set ∂f_i is nonempty. Note that even when the sets X_i are compact, we still assume the subgradient boundedness for each f_i since the optimization algorithm in (4) uses the subgradients at points $v_{i,k+1}$, which may lie outside the set X_i.

Communication noise and subgradient error. We now state the assumptions we use for the link noise and the stochastic subgradient errors. For this, let us denote the σ-algebra generated by the entire history of the algorithm up to time k, as follows:

$$F_k = \{x_{i,0}, W_\ell, \xi_{ij,\ell}, \epsilon_{i,\ell}; 1 \leq \ell \leq k, \ i, j \in V\}.$$

Assumption 3: Let the following hold:

a) The noise is zero mean, $\mathbb{E}\left[\xi_{ij,k+1} \mid F_k, W_{k+1}\right] = 0$ for any link $(i, j) \in \mathcal{E}(k+1)$.

b) For any realization of the random graph $\mathcal{W}(k+1)$, the noise on link (i, j) is independent of the noise on link (i', j'), where $(i, j) \neq (i', j')$.

c) There is a uniform bound on the noise variance, i.e., there is $\mu > 0$ such that for all $(i, j) \in \mathcal{E}(k+1)$ and $k \geq 0$,

$$\mathbb{E}\left[\|\xi_{ij,k+1}\|^2 \mid F_k, W_{k+1}\right] \leq \mu^2.$$

d) The error $\epsilon_{i,k+1}$ associated with the subgradient $d_{i,k+1}$ of the function $f_i(x)$ at $x = v_{i,k+1}$ is such

that for $k \geq 0$,

$$\mathbb{E}\left[\epsilon_{i,k+1} \mid F_k, W_{k+1}, v_{i,k+1}\right] = 0 \quad \text{for all } i \in U_{k+1}.$$

e) The subgradient errors satisfy the following: for all $k \geq 0$,

$$\mathbb{E}\left[\|\epsilon_{i,k+1}\|^2 \mid F_k, W_{k+1}, v_{i,k+1}\right] \leq \nu^2 \quad \text{for all } i \in U_{k+1}.$$

For algorithm (4), we have $\xi_{i,k+1} = \sum_{j \neq i}[W_{k+1}]_{ij}\xi_{ij,k+1}$. Thus, as a consequence of Assumptions 3-a and 3-b on the noise process and relation (6), which holds under Assumption 1-b on the weights, we have

$$\mathbb{E}\left[\|\xi_{i,k+1}\|^2 \mid F_k, W_{k+1}\right]$$
$$= \sum_{j \neq i}\mathbb{E}\left[[W_{k+1}^2]_{ij}\|\xi_{ij,k+1}\|^2 \mid F_k, W_{k+1}\right] \leq \bar{N}\eta^2\mu^2, \tag{7}$$

where $\bar{N} = \max_i\{|N_i|\}$ and N_i is the set of neighbors of agent i in the graph $G = (V, E)$.

As a consequence of the subgradient norm and subgradient error boundedness (Assumptions 2-e and 3-e), it can be seen that for all $i \in U_{k+1}$ and $k \geq 0$,

$$\mathbb{E}\left[\|d_{i,k+1} + \epsilon_{i,k+1}\|^2 \mid F_k, W_{k+1}, v_{i,k+1}\right] \leq (C + \nu)^2. \tag{8}$$

Here we have used Hölder's inequality, which states that for random vectors $\mathbf{x}$ and $\mathbf{y}$ there holds $\mathbb{E}\left[|\mathbf{x}'\mathbf{y}|\right] \leq \sqrt{\mathbb{E}[\|\mathbf{x}\|^2]}\sqrt{\mathbb{E}[\|\mathbf{y}\|^2]}$. We now show that the two most widely studied communication protocols in the consensus literature satisfy our assumptions on the random matrix W_k.

D. Gossip-based communication protocol

In [25], a widely used model for asynchronous gossip based communication protocol is provided, which can be represented as follows. At each tick of the global Poisson clock T_p, an agent represented by the random variable I_k, wakes up with uniform probability. Then, the agent I_k selects a neighboring agent J_k with a stationary probability $P_{I_kJ_k}$. Let us denote by W_k^g the random matrix W_k corresponding to this case. Then, the matrix W_k^g takes the following form:

$$W_k^g = \frac{1}{2}(e_{I_k} - e_{J_k})(e_{I_k} - e_{J_k})' \quad \text{with probability } \frac{1}{m}P_{I_kJ_k},$$

where e_j is a vector with j^{th} component equal to 1 and the other components equal to 1. The probability of update for any agent i at any of the clock ticks is given by $p_i = \frac{1}{m}\left[1 + \sum_{j \in N_i} P_{ji}\right]$. For the special case when the probability of selecting a neighboring agent is uniform, we have $p_i = \frac{1}{m}\left[1 + \sum_{j \in N_i}\frac{1}{|N_j|}\right]$.

Note that, in this case, when the underlying graph $G = (V, E)$ is regular, the probability of update for all agents is the same and is equal to $p = \frac{2}{m}$. We also have the following properties for the random matrix W_k^g.

Lemma 1: [25] The random matrix W_k^g has the following properties:

a) $\bar{W}_g = \mathbb{E}[W_k^g]$ is a symmetric positive semidefinite matrix such that $\bar{W}_g \mathbf{1} = 0$ and $\mathbf{1}' \bar{W}_g = 0$.

b) $R_g = \mathbb{E}[(W_k^g)' W_k^g] = \bar{W}_g$.

Proof: We can explicitly compute the expected value of the random matrix $W(k)$ as follows:

$$\bar{W}_g = \mathbb{E}[W_k^g] = \frac{1}{2m} \sum_{(i,j) \in E} P_{ij}(e_i - e_j)(e_i - e_j)',$$

which can be represented as $\bar{W}_g = \frac{1}{2m}(D - (P + P'))$, where D is a diagonal matrix with entries $D_i = \sum_{j \neq i}[P_{ij} + P_{ji}]$. Now, it can be verified that $\bar{W}_g \mathbf{1} = 0$ and $\mathbf{1}' \bar{W}_g = 0$. Moreover since we have

$$0 \leq (W_k^g)' W_k^g = [W_k^g]^2 = W_k^g,$$

upon taking the expectation, we obtain the desired result. ∎

Observe that our notation is slightly different from that of [25], where the convergence properties of the recursion $x(k+1) = \tilde{W}_k x(k)$ is considered with $\tilde{W}_k = I - W_k^g$.

E. Broadcast-based communication protocol

The consensus algorithm based on a broadcast based communication protocol was discussed in [26]. According to the broadcast protocol once a node i wakes up at the tick of its local clock, it broadcasts its local variable, which is successfully received by its neighbors. The random matrix W_{k+1} in this case is labeled W_k^b. Formally, representing the random agent whose clock ticks in slot k by I_k, the matrix W_k^b takes the following form:

$$W_k^b = \beta \sum_{j \in N_{I_k}} (e_j e_j' - e_j e_{I_k}') \qquad \text{with probability } \frac{1}{m}.$$

The probability of update for any agent i is given as $p_i = \frac{|N_i|}{m}$. If the graph G is r-regular then the probability of update for all agents is uniformly equal to $p = \frac{r}{m}$. Furthermore, the following result holds.

Lemma 2: [40] Under the assumption that the graph G is bidirectional and connected, the random matrix W_k^b has the following properties:

a) $\bar{W}_b = \mathbb{E}[W_k^b]$ is a symmetric positive semidefinite matrix such that $\bar{W}_b \mathbf{1} = 0$ and $\mathbf{1}' \bar{W}_b = 0$.

b) $R_b = \mathbb{E}[(W_k^b)' W_k^b] = 2\beta^2 \bar{W}_b$.

Proof: The expected matrix can be computed as

$$\bar{W}_b = \mathbb{E}[W_k^b] = \frac{\beta}{m} \sum_{i \in V} \sum_{j \in N_i} (e_j e_j' - e_j e_i') = \frac{\beta}{m} L,$$

where L is the Laplacian of the graph G. By the definition of the graph Laplacian matrix L, and the assumptions that the graph G is bidirectional and connected (Assumption 1-a), we have that the matrix $\bar{W}_b$ is a positive semidefinite matrix which satisfies $\bar{W}_b \mathbf{1} = 0$ and $\mathbf{1}' \bar{W}_b = 0$. It can also be seen that $R_b = \mathbb{E}[(W_k^b)' W_k^b] = 2\beta^2 \bar{W}_b$. ∎

Recently in [40], a variant of the broadcast algorithm is presented including the possibility of collisions between simultaneous transmissions. In this case it can be seen that, if the graph is r-regular then, for the matrix $\bar{W}_{bc} = \mathbb{E}[W_k]$, we have $\bar{W}_{bc} = \beta p (1-p)^r L$, which is a positive semidefinite matrix with $\bar{W}_{bc} \mathbf{1} = 0$ and $\mathbf{1}' \bar{W}_{bc} = 0$.

III. PRELIMINARY RESULTS

In this section we provide various results which will be useful in proving our main results. The following is a well-known non-expansiveness property of the projection operator on a convex closed set X:

$$\|P_X[x] - P_X[y]\| \le \|x - y\| \qquad \text{for all } x, y, \tag{9}$$

see for example [39], Proposition 2.2.1, page 88.

The second method of Lyapunov has been instrumental in the stability analysis of dynamic systems. The Foster-Lyapunov theorem [41] gives an analogous stability result for systems with stochastic dynamics. However, the Foster-Lyapunov technique is not sufficient for our case. We need the following result of Robbins and Siegmund, which is a generalization of the supermartingale convergence theorem.

Lemma 3: ([36], page 50) Let $\{X_t\}$, $\{Y_t\}$, $\{Z_t\}$ and $\{g(t)\}$ be sequences of random variables and let $\mathcal{F}_t$, $t = 0, 1, 2, \ldots$, be a filtration such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t \ge 0$. Suppose that:

a) The random variables Y_t, X_t, Z_t and $g(t)$ are nonnegative, and are adapted to the filtration $\mathcal{F}_t$.

b) For each t, we have almost surely

$$\mathbb{E}[Y_{t+1} \mid \mathcal{F}_t] \le (1 + g(t))Y_t - X_t + Z_t.$$

c) There holds $\sum_{t=0}^{\infty} Z_t < \infty$ and $\sum_{t=0}^{\infty} g(t) < \infty$ almost surely.

Then, almost surely, $\sum_{t=0}^{\infty} X_t < \infty$ and the sequence $\{Y_t\}$ converges to a nonnegative random variable Y.

The following result, as proven in [30], is helpful in establishing asymptotic error bounds associated with constant step size algorithms.

Lemma 4: ([30]) Let $\beta \in (0,1)$, and let $\{d_k\}$ and $\{u_k\}$ be scalar sequences such that

$$d_k \leq \beta d_{k-1} + u_{k-1} \quad \text{for all} \quad k \geq 1.$$

Then,

$$\limsup_{k \to \infty} d_k \leq \frac{1}{1-\beta} \limsup_{k \to \infty} u_k.$$

In the current work we restrict ourself to step sizes of the form $\frac{1}{k^\theta}$, where $\frac{1}{2} < \theta \leq 1$. Thus, in our algorithms (3) and (4), we will use

$$\alpha_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_1}}, \qquad \gamma_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_2}}, \tag{10}$$

where $\frac{1}{2} < \theta_1, \theta_2 \leq 1$. In the following lemma, we give some results for these step sizes. The proof of the lemma is in Appendix.

Lemma 5: Under the $i.i.d$ assumption on the random network $\mathcal{W}(k)$ and $p_i > 0$ (Assumptions 1-c and 1-d), for any step size of the form $\frac{1}{\Gamma_{i,k}^\theta}$ with $\frac{1}{2} < \theta \leq 1$, the following hold with probability one: for all $i \in V$,

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\frac{1}{\Gamma_{i,k}^{2\theta}} \,|\, F_k\right] < \infty,$$

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\left|\frac{1}{\Gamma_{i,k}^\theta} - \frac{1}{p_i^\theta k^\theta}\right| \,|\, F_k\right] < \infty.$$

We are now ready to discuss our main results regarding the asymptotic behavior of the consensus and optimization algorithms given by Eqs. (3) and (4), respectively. We investigate the algorithms under diminishing and constant step sizes.

IV. CONSTRAINED CONSENSUS

In this section we present our analysis of the constrained consensus algorithm (3) under both diminishing and constant step sizes. We adopt the classical Lyapunov kind of analysis for stochastic systems, which is as follows. First, we derive a descent-type relation involving a particular choice of a Lyapunov function in Lemma 6. Then, using the descent relation and a standard supermartingale convergence result, in Theorem 1 we show the almost sure convergence of the local variables $x_{i,k}$ as generated by the consensus algorithm (3) to a common random point in the intersection of the local constraint sets. Following this, we consider the constant step size case in Theorem 2. It is well known that in this case, an almost sure convergence of the algorithm cannot be guaranteed, but error bounds may be provided for expected value of a Lyapunov function. We derive asymptotic error bounds based on a metric which penalizes the expected deviations from the consensus subspace.

We now proceed with some basic relations which will be valuable in deriving both the almost sure convergence and the error bound for a constant step size. Let us introduce the notation for the joint state vector as $\mathbf{x}_k = (x'_{1,k}, \ldots, x'_{m,k})'$, where $x_{i,k}$ is the local variable at node i constrained to the set X_i. Correspondingly, we denote the vector $\mathbf{v}_k = (v'_{1,k}, \ldots, v'_{m,k})'$, where $v_{i,k}$ is given by Eq. (3). For a fixed vector $z \in \mathbb{R}^n$, we denote $\mathbf{z} \in \mathbb{R}^{mn}$ as the vector consisting of m copies of the vector z, i.e., $\mathbf{z} = \mathbf{1}_m \otimes z$. By definition any such vector $\mathbf{z}$ belongs to the consensus subspace $\mathcal{C} \subset \mathbb{R}^{mn}$. To get further insight in the problem, let us define the product constraint set $\tilde{X} = X_1 \times \cdots \times X_m$. Then, clearly $\tilde{X} \subset \mathbb{R}^{mn}$. According to this definition, the consensus algorithm (3) implies that $\mathbf{x}_k \in \tilde{X}$ for all k. The constrained consensus problem can be equivalently thought of as the task of finding a point in the intersection set $\tilde{X} \cap \mathcal{C}$.

Recalling that $D(a_i)$ is a diagonal matrix with diagonal elements a_i, we introduce the diagonal matrix $D(p_i^{\theta_1})$, where $\theta_1 \in (1/2, 1]$ and the matrix $\mathbf{Q} = D(p_i^{\theta_1}) \otimes I_n$. Also, we use the diagonal matrix $D(\alpha_{i,k})$ to define $\bar{\Lambda}_k = D(\alpha_{i,k}) \otimes I_n$. The matrix $\mathbf{W}_k$ is defined as $\mathbf{W}_k = W_k \otimes I_n$. Finally, we define the net noise vector by $\xi'_k = (\xi'_{1,k}, \ldots, \xi'_{m,k})'$. According to this notation, we can represent the consensus algorithm (3) in an equivalent form:

$$\mathbf{v}_{k+1} = \left[I - \bar{\Lambda}_{k+1} \mathbf{W}_{k+1} \right] \mathbf{x}_k - \bar{\Lambda}_{k+1} \xi_{k+1}$$

$$\xi_{k+1} = \mathbf{W}_{k+1} \xi_{k+1} \qquad (11)$$

$$\mathbf{x}_{k+1} = P_{\tilde{X}} \left[\mathbf{v}_{k+1} \right].$$

Then, we have the following result.

Lemma 6: Consider a Lyapunov function of the form $V_k = (\mathbf{x}_k - \mathbf{z}^*)' \mathbf{Q} (\mathbf{x}_k - \mathbf{z}^*)$, where $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$ for any $z^* \in X$. Then, for the consensus algorithm (3) under Assumption 2-a on the constraint sets and the zero mean assumption on the communication noise (Assumption 3-a), we have for all $k \geq 0$,

$$\mathbb{E}[V_{k+1} \mid F_k] \leq V_k$$

$$+ (\mathbf{x}_k - \mathbf{z}^*)' \mathbb{E} \left[\mathbf{W}'_{k+1} \bar{\Lambda}^2_{k+1} \mathbf{Q} \mathbf{W}_{k+1} \mid F_k \right] (\mathbf{x}_k - \mathbf{z}^*)$$

$$- 2(\mathbf{x}_k - \mathbf{z}^*)' \mathbb{E} \left[\mathbf{Q} \bar{\Lambda}_{k+1} \mathbf{W}_{k+1} \mid F_k \right] (\mathbf{x}_k - \mathbf{z}^*)$$

$$+ \mathbb{E} \left[\xi'_{k+1} \bar{\Lambda}^2_{k+1} \mathbf{Q} \xi_{k+1} \mid F_k \right].$$

Proof: By the non-expansiveness property of the projection operator (cf. (9)), we obtain

$$p_i^{\theta_1} \left\| x_{i,k+1} - z^* \right\|^2 = p_i^{\theta_1} \left\| P_{X_i}[v_{i,k+1}] - P_{X_i}[z^*] \right\|^2$$

$$\leq p_i^{\theta_1} \left\| v_{i,k+1} - z^* \right\|^2.$$

On summing the relations above over index $i \in V$ and using $\mathbf{Q} = D(p_i^{\theta_1})$, we get

$$V_{k+1} = (\mathbf{x}_{k+1} - \mathbf{z}^*)'\mathbf{Q}(\mathbf{x}_{k+1} - \mathbf{z}^*)$$

$$\leq (\mathbf{v}_{k+1} - \mathbf{z}^*)'\mathbf{Q}(\mathbf{v}_{k+1} - \mathbf{z}^*).$$

Taking the expectation conditional on the past F_k, we obtain

$$\mathbb{E}[V_{k+1} \mid F_k] \leq \mathbb{E}[(\mathbf{v}_{k+1} - \mathbf{z}^*)'\mathbf{Q}(\mathbf{v}_{k+1} - \mathbf{z}^*) \mid F_k]. \tag{12}$$

We have $\mathbf{W}_{k+1}\mathbf{z}^* = 0$ which holds by $\mathbf{W}_{k+1} = W_{k+1} \otimes I_n$ and the definition of W_k in Eq. (5). Using this and the expression for $\mathbf{v}_{k+1}$ as given in relation (11), we obtain

$$\mathbb{E}[(\mathbf{v}_{k+1} - \mathbf{z}^*)'\mathbf{Q}(\mathbf{v}_{k+1} - \mathbf{z}^*) \mid F_k]$$

$$= V_k + (\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{W}'_{k+1}\bar{\Lambda}^2_{k+1}\mathbf{Q}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)$$

$$- 2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{Q}\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)$$

$$+ \mathbb{E}\left[\xi'_{k+1}\bar{\Lambda}^2_{k+1}\mathbf{Q}\xi_{k+1} \mid F_k\right]$$

$$- 2\mathbb{E}\left[\left([I - \bar{\Lambda}_{k+1}\mathbf{W}_{k+1}]\mathbf{x}_k - \mathbf{z}^*\right)'\mathbf{Q}\bar{\Lambda}_{k+1}\xi_{k+1} \mid F_k\right]. \tag{13}$$

Using the iterated expectation rule and $\mathbb{E}[\xi_{k+1} \mid F_k, W_{k+1}] = 0$ (Assumption 3-a), we have

$$\mathbb{E}\left[\left([I - \bar{\Lambda}_{k+1}\mathbf{W}_{k+1}]\mathbf{x}_k - \mathbf{z}^*\right)'\mathbf{Q}\bar{\Lambda}_{k+1}\xi_{k+1} \mid F_k\right] = 0. \tag{14}$$

The desired relation follows by combining Eqs. (12)-(14). ∎

A. Almost sure convergence

The following theorem brings together our various assumptions and the preliminary results to provide conditions guaranteeing almost sure convergence of the consensus algorithm.

Theorem 1: Let Assumption 1 on the network hold. Also, let Assumptions 2-a and 2-b on the constraint sets, and Assumptions 3-a, 3-b, and 3-c on the noise process hold. Let the step size $\alpha_{i,k}$ be of the form $\alpha_{i,k} = \frac{1}{\Gamma_{i,k}^{\theta_1}}$, where $\frac{1}{2} < \theta_1 \leq 1$. Then, the iterations of the consensus algorithm (3) satisfy with probability one for all i and some random vector $z^* \in X$:

$$\lim_{k \to \infty} \|x_{i,k} - z^*\| = 0.$$

Proof: The starting point of our analysis for almost sure convergence is the relation derived in Lemma 6. Let us focus on the term $-2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{Q}\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)$, where $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$ for an arbitrary $z^* \in X$. According to our notation, we have $\mathbf{Q}\bar{\Lambda}_{k+1} = D(p_i^{\theta_1}\alpha_{i,k+1}) \otimes I_n$, where

$D(p_i^{\theta_1} \alpha_{i,k+1})$ is the diagonal matrix with diagonal entries given by $p_i^{\theta_1} \alpha_{i,k+1}$, $i = 1, \ldots, n$. Thus, we can write

$$p_i^{\theta_1} \alpha_{i,k+1} = \frac{p_i^{\theta_1}}{\Gamma_{i,k+1}^{\theta_1}}$$
$$= \frac{p_i^{\theta_1}}{p_i^{\theta_1}(k+1)^{\theta_1}} + p_i^{\theta_1} \left(\frac{1}{\Gamma_{i,k+1}^{\theta_1}} - \frac{1}{p_i^{\theta_1}(k+1)^{\theta_1}} \right).$$

Let us denote $H_{k,\theta_1} = D\left(\frac{1}{\Gamma_{i,k}^{\theta_1}} - \frac{1}{p_i^{\theta_1} k^{\theta_1}} \right)$ and $\bar{H}_{k,\theta_1} = H_{k,\theta_1} \otimes I_n$. Then, we have

$$\mathbf{Q}\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} = \frac{1}{(k+1)^{\theta_1}}\mathbf{W}_{k+1} + \mathbf{Q}\bar{H}_{k+1,\theta_1}\mathbf{W}_{k+1}.$$

Since $\bar{W}\mathbf{1}_m = 0$ (Assumption 1-c), it follows that

$$- 2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}[\mathbf{Q}\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} \,|\, F_k](\mathbf{x}_k - \mathbf{z}^*)$$
$$= -\frac{2}{(k+1)^{\theta_1}}\,\mathbf{x}_k'\bar{\mathbf{W}}\mathbf{x}_k$$
$$- 2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{Q}\bar{H}_{k+1,\theta_1}\mathbf{W}_{k+1} \,|\, F_k \right](\mathbf{x}_k - \mathbf{z}^*). \tag{15}$$

Now, using $p_i^{\theta_1} \leq 1$, $|[W_{k+1}]_{ij}| \leq \eta$ for $i \neq j$ and $|[W_{k+1}]_{ii}| \leq |N_i|\eta$ (cf. Eq. (5) as implied by Assumption 1-b), the boundedness of constraint sets (Assumption 2-b), we can upper bound the last term in Eq. (15) as follows:

$$- 2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{Q}\bar{H}_{k+1,\theta_1}\mathbf{W}_{k+1} \,|\, F_k \right](\mathbf{x}_k - \mathbf{z}^*)$$
$$\leq 2\sum_{i,j} |x_{i,k} - z^*|\,|x_{j,k} - z^*|\mathbb{E}\left[|[W_{k+1}]_{ij}[H_{k+1,\theta_1}]_i| \,|\, F_k \right]$$
$$\leq 2\eta\bar{N} \sum_{i,j} C_{X_i} C_{X_j} \mathbb{E}\left[|[H_{k+1,\theta_1}]_i| \,|\, F_k \right], \tag{16}$$

where C_{X_i} is the diameter of the set X_i (i.e., $C_{X_i} = \max_{x,y \in X_i} \|x - y\|$) and $\bar{N}$ is the maximal node degree in the graph G (i.e., $\bar{N} = \max_i |N_i|$). Similarly we can derive the following bound

$$(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{W}_{k+1}'\bar{\Lambda}_{k+1}^2\mathbf{Q}\mathbf{W}_{k+1} \,|\, F_k \right](\mathbf{x}_k - \mathbf{z}^*)$$
$$\leq \bar{N}^2\eta^2 \sum_{j,j'} C_{X_j} C_{X_{j'}} \sum_{i \in V} \mathbb{E}\left[\alpha_{i,k+1}^2 \,|\, F_k \right]. \tag{17}$$

Under zero mean, independent across the links, and bounded second moment assumptions on the link noise (Assumptions 3-a, 3-b, and 3-c), the term involving communication noise in Lemma 6 can be bound

as follows:

$$\mathbb{E}\left[\xi'_{k+1}\bar{\Lambda}^2_{k+1}\mathbf{Q}\xi_{k+1} \,|\, F_k\right] = \sum_{i \in V} \mathbb{E}\left[p_i^{\theta_1}\alpha^2_{i,k+1}\|\xi_{i,k+1}\|^2 \,|\, F_k\right]$$

$$= \sum_{i \in V} \mathbb{E}\left[p_i^{\theta_1}\alpha^2_{i,k+1}\mathbb{E}[\|\xi_{i,k+1}\|^2 \,|\, F_k, W_{k+1}] \,|\, F_k\right]$$

$$\leq \bar{N}\eta^2\mu^2 \sum_{i \in V} \mathbb{E}\left[\alpha^2_{i,k+1} \,|\, F_k\right]. \tag{18}$$

Here, we have used the inequality in Eq. (7).

Now substituting the bounds from Eqs. (15), (16), (17) and (18) in the relation of Lemma 6, we obtain

$$\mathbb{E}\left[V_{k+1} \,|\, F_k\right] \leq V_k - \frac{2}{(k+1)^{\theta_1}}\mathbf{x}'_k\bar{\mathbf{W}}\mathbf{x}_k$$

$$+ 2\eta\bar{N}\sum_{i,j} C_{X_i}C_{X_j}\mathbb{E}\left[|[H_{k+1,\theta_1}]_i| \,|\, F_k\right]$$

$$+ \left(\bar{N}\eta^2\mu^2 + \bar{N}^2\eta^2\sum_{j,j'} C_{X_j}C_{X_{j'}}\right)\sum_{i \in V}\mathbb{E}\left[\alpha^2_{i,k+1} \,|\, F_k\right].$$

Note that from Lemma 5, we have the result that for the choice of step size $\alpha_{i,k+1} = \frac{1}{\Gamma^{\theta_1}_{i,k+1}}$ the last two terms are summable with probability one. Thus, we can apply the supermartingale convergence result of Lemma 3 to deduce that, with probability one, both the sequence $V_k = (\mathbf{x}_k - \mathbf{z}^*)'\mathbf{Q}(\mathbf{x}_k - \mathbf{z}^*)$ converges for any $z^* \in X$ and the following holds:

$$\sum_{k=0}^{\infty}\frac{1}{(k+1)^{\theta_1}}\mathbf{x}'_k\bar{\mathbf{W}}\mathbf{x}_k < \infty. \tag{19}$$

Since $\theta_1 \in (1/2, 1]$, we have $\sum_{k=0}^{\infty}\frac{1}{(k+1)^{\theta_1}} = \infty$. Moreover, since $\bar{\mathbf{W}}$ is positive semidefinite (Assumption 1-c), from Eq. (19) it follows that with probability one $\liminf_{k \to \infty}\mathbf{x}'_k\bar{\mathbf{W}}\mathbf{x}_k = 0$. Thus, with probability one, there exists a subsequence such that

$$\lim_{\ell \to \infty}\mathbf{x}'_{k_\ell}\bar{\mathbf{W}}\mathbf{x}_{k_\ell} = 0.$$

Now, under Assumption 1-c, the consensus subspace $\mathcal{C}$ is the null space of the matrix $\bar{\mathbf{W}}$, hence for all agents i and j we have with probability one

$$\lim_{\ell \to \infty}\|x_{i,k_\ell} - x_{j,k_\ell}\| = 0. \tag{20}$$

Since the constraint sets are compact (Assumption 2-b) and $\{x_{i,k}\} \subset X_i$ for all i, without loss of generality, we may assume that $\lim_{\ell \to \infty}\|x_{i,k_\ell} - \tilde{z}_i\| = 0$ with probability one for some random vector $\tilde{z}_i \in X_i$ and for all i. In view of (20), it follows that $\tilde{z}_1 = \ldots = \tilde{z}_m = \tilde{z}^*$. Hence, with probability one,

$\lim_{\ell \to \infty} Vk_\ell = \lim_{\ell \to \infty} \sum_i p_i^{\theta_1} \|x_{i,k_\ell} - \tilde{z}^*\|^2 = 0$. But, V_k converges almost surely for any $\tilde{z}^* \in X$, so the sub-sequential limit is also the sequential limit, implying that almost surely

$$\lim_{k \to \infty} V_k = \lim_{k \to \infty} \sum_{i=1}^{m} p_i^{\theta_1} \|x_{i,k} - \tilde{z}^*\|^2 = 0.$$

Since $p_i > 0$, we have $\lim_{k \to \infty} \|x_{i,k} - \tilde{z}^*\| = 0$ almost surely for all i and a random point $\tilde{z}^* \in X$. ∎

B. Constant step size error bound

In this section we focus on a constant step size algorithm, where $\alpha_{i,k} = \alpha_i$ for all i and $k > 0$. Using constant step sizes does not give us almost sure convergence to the consensus subspace, but in this case we can provide an asymptotic error bound on the iterations of the consensus algorithm.

Theorem 2: Let Assumption 1 on the network hold. Also, let Assumptions 2-a and 2-b on the constraint set hold together with Assumptions 3-a, 3-b and 3-c on the link noise. Then, for the iterates generated by consensus algorithm (3), we have the following asymptotic bound:

$$\limsup_{T \to \infty} \mathbb{E}\left[(\hat{\mathbf{x}}_T - \mathbf{z}^*)' \bar{\mathbf{W}} (\hat{\mathbf{x}}_T - \mathbf{z}^*) \right] \leq \frac{\eta \Delta_\alpha \bar{N}}{\underline{\alpha}} \sum_E C_{X_i} C_{X_j}$$

$$+ \frac{\bar{\alpha}^2}{2\underline{\alpha}} \left[m\bar{N}\eta^2\mu^2 + \|\mathbf{R}\| \sum_{i=1}^{m} C_{X_i}^2 \right],$$

where $\hat{\mathbf{x}}_T = \frac{1}{T} \sum_{k=0}^{T-1} \mathbf{x}_k$, $\bar{\mathbf{W}} = \mathbb{E}[\mathbf{W}_k]$, $\mathbf{R} = \mathbb{E}\left[\mathbf{W}_{k+1}' \mathbf{W}_{k+1} \right]$, $\|\mathbf{R}\|$ is the matrix norm of $\mathbf{R}$ induced by the Euclidean vector norm, and $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$ for any point $z^* \in X$.

Proof: Using a line of analysis similar to that of Lemma 6, it can be seen that for the constant step sizes α_i, the following relation holds with probability one for all $k \geq 0$:

$$\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k \right] \leq \|\mathbf{x}_k - \mathbf{z}^*\|^2$$

$$+ (\mathbf{x}_k - \mathbf{z}^*)' \mathbb{E}\left[\mathbf{W}_{k+1}' \bar{\Lambda}^2 \mathbf{W}_{k+1} \mid F_k \right] (\mathbf{x}_k - \mathbf{z}^*)$$

$$- 2(\mathbf{x}_k - \mathbf{z}^*)' \mathbb{E}\left[\bar{\Lambda} \mathbf{W}_{k+1} \mid F_k \right] (\mathbf{x}_k - \mathbf{z}^*)$$

$$+ \mathbb{E}\left[\xi_{k+1}' \bar{\Lambda}^2 \xi_{k+1} \mid F_k \right],$$

where $\bar{\Lambda} = D(\alpha_i) \otimes I_n$ and $D(\alpha_i)$ is diagonal matrix with diagonal entries α_i. Similar to the derivation of relation (18), we can show that the following bound is valid for all $k \geq 0$,

$$\mathbb{E}\left[\xi_{k+1}' \bar{\Lambda}^2 \xi_{k+1} \mid F_k \right] \leq \bar{N}\eta^2\mu^2 \sum_{i \in V} \alpha_i^2 \leq m\bar{N}\eta^2\mu^2\bar{\alpha}^2,$$

with $\bar{\alpha} = \max_{i \in V} \alpha_i$. From the preceding two relations, we obtain

$$
\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k\right] \leq \|\mathbf{x}_k - \mathbf{z}^*\|^2
$$
$$
+ \bar{\alpha}^2(\mathbf{x}_k - \mathbf{z}^*)'\mathbf{R}(\mathbf{x}_k - \mathbf{z}^*)
$$
$$
- 2(\mathbf{x}_k - \mathbf{z}^*)'\bar{\Lambda}\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*) + m\bar{N}\eta^2\mu^2\bar{\alpha}^2, \tag{21}
$$

where $\mathbf{R} = \mathbb{E}\left[\mathbf{W}'_{k+1}\mathbf{W}_{k+1}\right]$ and $\bar{\mathbf{W}} = \mathbb{E}\left[\mathbf{W}_{k+1}\right]$. Using $\bar{\Lambda} = D(\alpha_i) \otimes I_n$, we can derive the following estimate:

$$
-2(\mathbf{x}_k - \mathbf{z}^*)'\bar{\Lambda}\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*) \leq -2\underline{\alpha}(\mathbf{x}_k - \mathbf{z}^*)'\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*)
$$
$$
+ 2\eta\Delta_\alpha\bar{N}\sum_E C_{X_i}C_{X_j}, \tag{22}
$$

where $\underline{\alpha} = \min_{i \in V} \alpha_i$ and δ_α is the difference between the maximum and the minimum values of the step sizes α_i. By substituting (22) in (21) and taking total expectation, we obtain

$$
2\underline{\alpha}\mathbb{E}\left[(\mathbf{x}_k - \mathbf{z}^*)'\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*)\right] \leq \mathbb{E}\left[\|\mathbf{x}_k - \mathbf{z}^*)\|^2\right]
$$
$$
- \mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2\right] + 2\eta\Delta_\alpha\bar{N}\sum_E C_{X_i}C_{X_j}
$$
$$
+ m\bar{N}\eta^2\mu^2\bar{\alpha}^2 + \bar{\alpha}^2\mathbb{E}\left[(\mathbf{x}_k - \mathbf{z}^*)'\mathbf{R}(\mathbf{x}_k - \mathbf{z}^*)\right]. \tag{23}
$$

Using the compactness condition (Assumption 2-b), we can see that $(\mathbf{x}_k - \mathbf{z}^*)'\mathbf{R}(\mathbf{x}_k - \mathbf{z}^*) \leq \|\mathbf{R}\|\sum_{i=1}^m C_{X_i}^2$, implying that

$$
\mathbb{E}[(\mathbf{x}_k - \mathbf{z}^*)'\mathbf{R}(\mathbf{x}_k - \mathbf{z}^*)] \leq \|\mathbf{R}\|\sum_{i=1}^m C_{X_i}^2.
$$

Using this bound and summing relations (23) from $k = 0$ to $k = T - 1$ for some $T > 0$, we have

$$
\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{E}\left[(\mathbf{x}_k - \mathbf{z}^*)'\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*)\right]
$$
$$
\leq \frac{1}{2\underline{\alpha}T}\left[\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{z}^*)\|^2\right] - \mathbb{E}\left[\|\mathbf{x}_T - \mathbf{z}^*\|^2\right]\right]
$$
$$
+ \frac{\eta\Delta_\alpha\bar{N}}{\underline{\alpha}}\sum_E C_{X_i}C_{X_j} + \frac{\bar{\alpha}^2}{2\underline{\alpha}}\left[m\bar{N}\eta^2\mu^2 + \|\mathbf{R}\|\sum_{i=1}^m C_{X_i}^2\right].
$$

Letting $\hat{\mathbf{x}}_T = \frac{1}{T}\sum_{k=0}^{T-1}\mathbf{x}_k$, and using Jensen's inequality and the convexity of the function $g(\mathbf{y}) = \mathbf{y}'\bar{\mathbf{W}}\mathbf{y}$, we obtain

$$\mathbb{E}\left[(\hat{\mathbf{x}}_T - \mathbf{z}^*)'\bar{\mathbf{W}}(\hat{\mathbf{x}}_T - \mathbf{z}^*)\right]$$
$$\leq \frac{1}{2\underline{\alpha}T}\left[\mathbb{E}\left[\|\mathbf{x}_0 - \mathbf{z}^*)\|^2\right] - \mathbb{E}\left[\|\mathbf{x}_T - \mathbf{z}^*\|^2\right]\right]$$
$$+ \frac{\eta\Delta_\alpha\bar{N}}{\underline{\alpha}}\sum_E C_{X_i}C_{X_j} + \frac{\bar{\alpha}^2}{2\underline{\alpha}}\left[m\bar{N}\eta^2\mu^2 + \|\mathbf{R}\|\sum_{i=1}^m C_{X_i}^2\right].$$

The result follows by taking the limsup as $T \to \infty$ and noting that, by the compactness of the sets X_i (Assumption 2-b), the limsup of the first term on the right-hand side converges to 0, as $T \to \infty$. ∎

In Theorem 2, the error term $\frac{\eta\Delta_\alpha\bar{N}}{\underline{\alpha}}\sum_E C_{X_i}C_{X_j}$ is due to misaligned step sizes. In other words, this term would be zero if all step sizes α_i take the same value ($\Delta_\alpha = 0$). The error term $\frac{\bar{\alpha}^2}{2\underline{\alpha}}m\bar{N}\eta^2\mu^2$ is due to the noise in the communication links, which would not be present if the links are perfect ($\mu = 0$). Note that this term is of the order $\frac{\bar{\alpha}^2}{2\underline{\alpha}}$ and, hence, diminishes with diminishing step sizes. The error term $\frac{\bar{\alpha}^2}{2\underline{\alpha}}\|\mathbf{R}\|\sum_{i=1}^m C_{X_i}^2$ can also be controlled by controlling the step sizes. In the case of gossip and broadcast protocols, we have $\mathbf{R}_g = \bar{\mathbf{W}}_g$ and $\mathbf{R}_b = 2\beta^2\bar{\mathbf{W}}_b$, thus the last error term can be written in terms of the expected matrices $\bar{\mathbf{W}}_g$ and $\bar{\mathbf{W}}_b$, respectively. Here, $\bar{\mathbf{W}}_g$ and $\bar{\mathbf{W}}_b$ are the expected weighted communication matricess for the gossip and broadcast protocols, as discussed in Sections II-D and II-E.

V. DISTRIBUTED OPTIMIZATION

In this section we discuss the asymptotic properties of the optimization algorithm given in Eq. (4) using both diminishing and constant step sizes. As mentioned earlier, a novel aspect of the proposed algorithm is the use of two step size sequences to damp communication noise and subgradient errors. As in the problem of constrained consensus, for the case of diminishing step sizes we can show almost sure convergence of the iterates to the optimal solution set $X^* \subseteq X$. An interesting feature which emerges out of our analysis is the requirement that, for almost sure convergence, the step size corresponding to the subgradient error needs to decay to zero at a faster rate than the step size sequence corresponding to the communication noise.

To study the convergence properties of the proposed algorithm, once again we use tools from Lyapunov analysis. First we start with establishing a bound as given in Lemma 7, which is crucial for the later development. To derive this bound we need to impose that the intersection of the constraint sets have a nonempty interior as stated in Assumption 2-c. This was also required to prove convergence of the alternating projection method to find the intersection of convex sets in [42]. It was also used in [11] to establish convergence of a distributed algorithm and the rate of convergence guarantees.

We then proceed to derive a descent-type inequality for the iterates of the algorithm, which is done in two steps. At first, we derive an auxiliary bound in Lemma 8 which holds for diminishing step sizes $\alpha_{i,k}$ and $\gamma_{i,k}$ and, at second, in Lemma 9 we refine the result for a specific choice of the form of the step sizes to get the desired descent-type inequality. Then, we proceed by showing almost sure convergence of the optimization algorithm in Theorem 3 and Theorem 4. These theorems differ in their set of assumptions on the network and the step sizes employed by the agents. Finally, we establish an asymptotic error bound for the algorithm in Theorem 5, when both the step sizes corresponding to the subgradient step and the alignment task are constant. In this case, under strong convexity assumption on the objective functions $f_i(x)$, we obtain a contractive property which enables us to derive the result.

A. Preliminary results

In this section we provide several results which we use when deriving our main result. The first result provides a way to bound an error term of the form $\|x_{i,k} - P_X[x_{i,k}]\|$. The bound is established by using some of the techniques in [42].

Lemma 7: Let Assumptions 2-a, 2-b and 2-c on the constraint sets hold, and let $x_i \in X_i$ for all $i \in V$. Then, we have the following bound:

$$\|x_i - P_X[x_i]\| \leq \frac{B}{\delta} \sum_{j=1}^{m} \|x_i - x_j\| \qquad \text{for all } i,$$

where $B = \max_i\{C_{X_i}\}$ is a uniform upper bound on the diameters of the sets X_i and δ is the radius from the interior point Assumption 2-c.

Proof: Let $i \in V$ be arbitrary. Define $\lambda_i = \sum_{j=1}^{m} \|x_i - P_{X_j}[x_i]\|$ and the variable s_i as follows:

$$s_i = \frac{\lambda_i}{\lambda_i + \delta} \bar{z} + \frac{\delta}{\lambda_i + \delta} x_i,$$

where $\bar{z}$ is the interior point of the set X from Assumption 2-c. Then, we can write

$$s_i = \frac{\lambda_i}{\lambda_i + \delta} \left[\bar{z} + \frac{\delta}{\lambda_i} \left(x_i - P_{X_j}[x_i] \right) \right] + \frac{\delta}{\lambda_i + \delta} P_{X_j}[x_i].$$

From definition of λ_i, it is clear that $\left\| x_i - P_{X_j}[x_i] \right\| \leq \lambda_i$ for any j, implying by the interior point assumption that the vector $\bar{z} + \frac{\delta}{\lambda_i} \left(x_i - P_{X_j}[x_i] \right)$ lies in the set X and hence, in set X_j for any j. Since the vector s_i is a convex combination of two vectors in the set X_j, by the convexity assumption on the set X_j, we have that $s_i \in X_j$ for any j. Therefore, we have $s_i \in X$. Now, we can see that

$$\|x_i - s_i\| \leq \frac{\lambda_i}{\lambda_i + \delta} \|x_i - \bar{z}\| \leq \frac{\|x_i - \bar{z}\|}{\delta} \sum_{j=1}^{m} \left\| x_i - P_{X_j}[x_i] \right\|.$$

By our assumption the sets X_i are compact (Assumption 2-b), so $\|x_i - \bar{z}\| \leq B$ for $B > 0$. Since $x_j \in X_j$, by the properties of the projection operator it follows $\left\|x_i - P_{X_j}[x_i]\right\| \leq \|x_i - x_j\|$. Thus,

$$\|x_i - P_X[x_i]\| \leq \|x_i - s_i\| \leq \frac{B}{\delta} \sum_{j=1}^m \|x_i - x_j\|.$$

$\blacksquare$

We now provide a basic iterate relation for algorithm (4), which we use later on to establish convergence and error bounds.

Lemma 8: Let Assumption 1 on the network hold. Also, let Assumption 2 on the constraint sets and objective functions hold as well as Assumption 3 on the link noise and subgradient errors. Further, let the step sizes be such that $\lim_{k\to\infty} \alpha_{i,k} = 0$ for all i. Then, for the iterates generated by algorithm (4) the following inequality holds for any $z^* \in X^*$ and for all large enough k,

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k, W_{k+1}] \leq \mathbb{E}[\|\mathbf{v}_{k+1} - \mathbf{z}^*\|^2 \mid F_k, W_{k+1}]$$

$$+ 2(C + \nu)^2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}^2 - 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}[f_i(s_k) - f_i(z^*)]$$

$$+ 2CK \sum_{i \in U_{k+1}} \gamma_{i,k+1} \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\|,$$

where $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$, $s_k = \frac{1}{m} \sum_{i=1}^m P_X[x_{i,k}]$ and $K = \frac{mB+\delta}{m\delta}$.

Proof: By definition we have $x_{i,k+1} = P_{X_i}[v_{i,k+1} - \gamma_{i,k+1}\tilde{d}_{i,k+1}\chi_{\{i \in U_{k+1}\}}]$. We see that for any $z^* \in X^* \subseteq X$ and all $i \in V$,

$$\|x_{i,k+1} - z^*\|^2$$

$$= \left\| P_{X_i}[v_{i,k+1} - \gamma_{i,k+1}\tilde{d}_{i,k+1}\chi_{i \in U_{k+1}}] - P_{X_i}[z^*] \right\|^2$$

$$\leq \left\| v_{i,k+1} - z^* - \gamma_{i,k+1}\tilde{d}_{i,k+1}\chi_{i \in U_{k+1}} \right\|^2$$

$$= \|v_i(k+1) - z^*\|^2$$

$$+ \left(\gamma_{i,k+1}^2 \|\tilde{d}_{i,k+1}\|^2 - 2\gamma_{i,k+1}\tilde{d}'_{i,k+1}(v_{i,k+1} - z^*) \right)\chi_{\{i \in U_{k+1}\}}.$$

Taking conditional expectation with respect to the past information F_k and the matrix W_{k+1}, we obtain for any $z^* \in X^*$,

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}] \leq \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$
$$+ 2\gamma_{i,k+1}^2 \mathbb{E}\left[\|\tilde{d}_{i,k+1}\|^2 \mid F_k, W_{k+1}\right] \chi_{\{i \in U_{k+1}\}}$$
$$- 2\gamma_{i,k+1} \mathbb{E}[d_{i,k+1}'(v_{i,k+1} - z^*) \mid F_k, W_{k+1}] \chi_{\{i \in U_{k+1}\}}$$
$$- 2\gamma_{i,k+1} \mathbb{E}[\epsilon_{i,k+1}'(v_{i,k+1} - z^*) \mid F_k, W_{k+1}] \chi_{\{i \in U_{k+1}\}}. \tag{24}$$

By Assumption 3-d we have $\mathbb{E}[\epsilon_{i,k+1} \mid F_k, W_{k+1}, v_{i,k+1}] = 0$, thus by the iterated expectation, we get

$$\mathbb{E}[\epsilon_{i,k+1}'(v_{i,k+1} - z^*) \mid F_k, W_{k+1}] = 0.$$

Also, using the fact that $d_{i,k+1}$ is a subgradient of f_i at $v_{i,k+1}$ and the bound on the subgradient error of Eq. (8) (implied by Assumptions 2-e and 3-e), we have

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}] \leq \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$
$$+ 2\gamma_{i,k+1}^2 (C + \nu)^2 \chi_{\{i \in U_{k+1}\}}$$
$$- 2\gamma_{i,k+1} \mathbb{E}[f_i(v_{i,k+1}) - f_i(z^*) \mid F_k, W_{k+1}] \chi_{\{i \in U_{k+1}\}}.$$

Since f_i is a convex function, by Jensen's inequality we have $-\mathbb{E}[f_i(v_{i,k+1}) \mid F_k, W_{k+1}] \leq -f(\mathbb{E}[v_{i,k+1} \mid F_k, W_{k+1}])$. By the definition of $v_{i,k+1}$ in Eq. (4) and $\mathbb{E}[\xi_{ij,k} \mid F_k, W_{k+1}] = 0$ (Assumption 3-a), we have for $i \in U_{k+1}$,

$$\mathbb{E}[v_{i,k+1} \mid F_k, W_{k+1}] = x_{i,k} - \alpha_{i,k+1} \sum_{j=1}^m [W_{k+1}]_{ij} x_{j,k}.$$

Letting $y_{i,k} = x_{i,k} - \alpha_{i,k+1} \sum_{j=1}^m [W_{k+1}]_{ij} x_{j,k}$, we obtain

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$
$$\leq \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$
$$+ 2\gamma_{i,k+1}^2 (C + \nu)^2 \chi_{\{i \in U_{k+1}\}}$$
$$- 2\gamma_{i,k+1} [f_i(y_{i,k}) - f_i(z^*)] \chi_{\{i \in U_{k+1}\}}. \tag{25}$$

Summing over all i in Eq. (25) and using vector notation yield

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k, W_{k+1}] \leq \mathbb{E}[\|\mathbf{v}_{k+1} - \mathbf{z}^*\|^2 \mid F_k, W_{k+1}]$$
$$+ 2(C + \nu)^2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}^2 - 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1} [f_i(y_{i,k}) - f_i(z^*)]. \tag{26}$$

Define $s_k = \frac{1}{m} \sum_{i=1}^{m} P_X[x_{i,k}]$. Upon adding and subtracting the term $\sum_{i \in U_{k+1}} f_i(s_k)$ in Eq. (26) we get

$$
\begin{aligned}
\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k, W_{k+1}] &\leq \mathbb{E}[\|\mathbf{v}_{k+1} - \mathbf{z}^*\|^2 \mid F_k, W_{k+1}] \\
&+ 2(C+\nu)^2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}^2 \\
&- 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1} [f_i(s_k) - f_i(z^*)] \\
&- 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1} [f_i(y_{i,k}) - f_i(s_k)].
\end{aligned}
\tag{27}
$$

Let us now focus on the last term in Eq. (27), which we can bound as follows:

$$
\begin{aligned}
- 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1} [f_i(y_{i,k}) - f_i(s_k)] \\
\leq 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1} |f_i(y_{i,k}) - f_i(s_k)|.
\end{aligned}
$$

By the convexity and subgradient boundedness of each $f_i(x)$, we have

$$
|f_i(y_{i,k}) - f_i(s_k)| \leq C \|y_{i,k} - s_k\|.
\tag{28}
$$

We now derive an upper bound on the term $\|y_{i,k} - s_k\|$. Since $\alpha_{i,k} \to 0$ and $y_{i,k} = x_{i,k} - \alpha_{i,k+1} \sum_{j=1}^{m} [W_{k+1}]_{ij} x_{j,k}$, the matrix $I_m - \alpha_{i,k+1} W_{k+1}$ is a stochastic matrix for all k large enough. Therefore, it follows by the convexity of the norm function and the fact that $0 \leq 1 - \alpha_{i,k+1} [W_{k+1}]_{ij} \leq 1$ for all i, j and all k large enough

$$
\|y_{i,k} - s_k\| \leq \sum_{j=1}^{m} \|x_{j,k} - s_k\|.
\tag{29}
$$

Now, consider $\|x_{i,k} - s_k\|$. Since $s_k = \frac{1}{m} \sum_{j=1}^{m} P_X[x_{j,k}]$, by adding and subtracting the term $P_X[x_{i,k}]$ inside the norm and using the convexity of norm function, we have

$$
\begin{aligned}
\|x_{i,k} - s_k\| &= \left\| x_{i,k} - \frac{1}{m} \sum_{j=1}^{m} P_X[x_{j,k}] \right\| \\
&\leq \|x_{i,k} - P_X[x_{i,k}]\| + \frac{1}{m} \sum_{j=1}^{m} \|P_X[x_{i,k}] - P_X[x_{j,k}]\|.
\end{aligned}
$$

By Lemma 7, $\|x_{i,k} - P_X[x_{i,k}]\| \leq \frac{B}{\delta} \sum_{j=1}^{m} \|x_{i,k} - x_{j,k}\|$, and by the non-expansiveness property of projection, $\|P_X[x_{i,k}] - P_X[x_{j,k}]\| \leq \|x_{i,k} - x_{j,k}\|$ (cf. Eq. (9)). Hence, for all i,

$$
\|x_{i,k} - s_k\| \leq \frac{mB + \delta}{m\delta} \sum_{j=1}^{m} \|x_{i,k} - x_{j,k}\|.
\tag{30}
$$

Combining relations in Eqs. (28), (29) and (30), we obtain for k large enough,

$$|f_i(y_{i,k}) - f_i(s_k)| \le C \frac{mB + \delta}{m\delta} \sum_{j=1}^{m} \sum_{r=1}^{m} \|x_{j,k} - x_{r,k}\|.$$

Therefore, for all k large enough, we have

$$-2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}[f_i(y_{i,k+1}) - f_i(s_k)]$$

$$\le 2C \frac{mB + \delta}{m\delta} \sum_{i \in U_{k+1}} \gamma_{i,k+1} \sum_{j=1}^{m} \sum_{r=1}^{m} \|x_{j,k} - x_{r,k}\|.$$

Letting $K = \frac{mB+\delta}{m\delta}$ and substituting the preceding relation in Eq. (27), we obtain the desired result. ∎

Lemma 8 is true for any step size sequence $\gamma_{i,k}$. We next give a corollary of the lemma for the special case when $\alpha_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_1}}$ and $\gamma_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_2}}$ with $\frac{1}{2} < \theta_1, \theta_2 \le 1$. Let us define $\mathbf{P} = P \otimes I_n$, where P is defined by $P = D\left(\frac{1}{p_i^{\theta_1}}\right)$.

Lemma 9: In addition to the assumptions in Lemma 8, let the step sizes be given by $\alpha_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_1}}$ and $\gamma_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_2}}$ with $\frac{1}{2} < \theta_1, \theta_2 \le 1$. Then, the following relation holds for all large enough k, any $z^* \in X^*$, and $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$,

$$\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k\right] \le \|\mathbf{x}_k - \mathbf{z}^*\|^2 - \frac{2\mathbf{x}_k' \mathbf{P}\bar{\mathbf{W}}\mathbf{x}_k}{(k+1)^{\theta_1}}$$

$$+ \left(\bar{N}\eta^2\mu^2 + \bar{N}^2\eta^2 \sum_E C_{X_j} C_{X_{j'}}\right) \sum_{i \in V} \mathbb{E}\left[\alpha_{i,k+1}^2 \mid F_k\right]$$

$$+ 2\eta\bar{N} \sum_E C_{X_i} C_{X_j} \mathbb{E}\left[|[H_{k+1,\theta_1}]_i| \mid F_k\right]$$

$$+ 2(C + \nu)^2 \mathbb{E}\left[\sum_{i \in U_{k+1}} \gamma_{i,k+1}^2 \mid F_k\right]$$

$$- \frac{2}{(k+1)^{\theta_2}} \sum_{i=1}^{m} p_i^{1-\theta_2}[f_i(s_k) - f_i(z^*)]$$

$$+ \frac{2CK}{(k+1)^{\theta_2}} \sum_{j=1}^{m} \sum_{r=1}^{m} \|x_{j,k} - x_{r,k}\| \sum_{i=1}^{m} p_i^{1-\theta_2}$$

$$+ 2C \left(\|s_k - z^*\| + K \sum_{j=1}^{m} \sum_{r=1}^{m} \|x_{j,k} - x_{r,k}\|\right)$$

$$\times \mathbb{E}\left[\sum_{i \in U_{k+1}} |[H_{k+1,\theta_2}]_i| \mid F_k\right].$$

Proof: Consider the result of Lemma 8, and focus on the term $\mathbb{E}\left[\|\mathbf{v}_{k+1} - \mathbf{z}^*\|^2 \mid F_k\right]$. Note that we can proceed as in Eq. (13) to get

$$
\mathbb{E}\left[\|\mathbf{v}_{k+1} - \mathbf{z}^*)\|^2 \mid F_k\right] = \|\mathbf{x}_k - \mathbf{z}^*\|^2
$$
$$
+ (\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{W}'_{k+1}\bar{\Lambda}^2_{k+1}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)
$$
$$
- 2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)
$$
$$
+ \mathbb{E}\left[\xi'_{k+1}\bar{\Lambda}^2_{k+1}\xi_{k+1} \mid F_k\right], \tag{31}
$$

where we have used the assumption that noise is zero mean. Proceeding similarly as in Eqs. (15) and (16) we can write

$$
\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} = \frac{1}{(k+1)^{\theta_1}}\mathbf{P}\mathbf{W}_{k+1} + \bar{H}_{k+1,\theta_1}\mathbf{W}_{k+1},
$$

and obtain

$$
- 2(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\bar{\Lambda}_{k+1}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)
$$
$$
\leq -\frac{2\mathbf{x}'_k\mathbf{P}\bar{\mathbf{W}}\mathbf{x}_k}{(k+1)^{\theta_1}} + 2\eta\bar{N}\sum_E C_{X_i}C_{X_j}\mathbb{E}\left[|[H_{k+1,\theta_1}]_i| \mid F_k\right], \tag{32}
$$

where we use $\mathbf{P}\bar{\mathbf{W}}\mathbf{z}^* = 0$, and $\mathbf{z}^{*'}\mathbf{P}\bar{\mathbf{W}} = 0$. Identically to Eqs. (18) and (17), we can see that

$$
(\mathbf{x}_k - \mathbf{z}^*)'\mathbb{E}\left[\mathbf{W}'_{k+1}\bar{\Lambda}^2_{k+1}\mathbf{W}_{k+1} \mid F_k\right](\mathbf{x}_k - \mathbf{z}^*)
$$
$$
+ \mathbb{E}\left[\xi'_{k+1}\bar{\Lambda}^2_{k+1}\xi_{k+1} \mid F_k\right]
$$
$$
\leq \left(\bar{N}\eta^2\mu^2 + \bar{N}^2\eta^2\sum_E C_{X_j}C_{X_{j'}}\right)\sum_{i \in V}\mathbb{E}\left[\alpha^2_{i,k+1} \mid F_k\right]. \tag{33}
$$

Combining Eqs. (31), (32) and (33), we obtain

$$
\mathbb{E}\left[\|\mathbf{v}_{k+1} - \mathbf{z}^*)\|^2 \mid F_k\right] \leq \|\mathbf{x}_k - \mathbf{z}^*\|^2
$$
$$
+ \left(\bar{N}\eta^2\mu^2 + \bar{N}^2\eta^2\sum_E C_{X_j}C_{X_{j'}}\right)\sum_{i \in V}\mathbb{E}\left[\alpha^2_{i,k+1} \mid F_k\right]
$$
$$
- \frac{2\mathbf{x}'_k\mathbf{P}\bar{\mathbf{W}}\mathbf{x}_k}{(k+1)^{\theta_1}} + 2\eta\bar{N}\sum_E C_{X_i}C_{X_j}\mathbb{E}\left[|[H_{k+1,\theta_1}]_i| \mid F_k\right]. \tag{34}
$$

We now focus on the last two terms in the relation of Lemma 8. Using the given forms of the step sizes, we can write

$$
\gamma_{i,k+1} = \frac{1}{[\Gamma_{i,k+1}]^{\theta_2}}
$$
$$
= \frac{1}{p_i^{\theta_2}(k+1)^{\theta_2}} + \left(\frac{1}{[\Gamma_{i,k+1}]^{\theta_2}} - \frac{1}{p_i^{\theta_2}(k+1)^{\theta_2}}\right).
$$

Define $[H_{k+1,\theta_2}]_i = \frac{1}{[\Gamma_{i,k+1}]^{\theta_2}} - \frac{1}{p_i^{\theta_2}(k+1)^{\theta_2}}$. Then, we can bound the terms under consideration as

$$- 2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}[f_i(s_k) - f_i(z^*)]$$

$$+ 2CK \sum_{i \in U_{k+1}} \gamma_{i,k+1} \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\|$$

$$\leq -2\frac{1}{(k+1)^{\theta_2}} \sum_{i \in U_{k+1}} \frac{1}{p_i^{\theta_2}}[f_i(s_k) - f_i(z^*)]$$

$$+ 2CK \sum_{i \in U_{k+1}} \frac{1}{p_i^{\theta_2}(k+1)^{\theta_2}} \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\|$$

$$+ 2C \sum_{i \in U_{k+1}} |[H_{k+1,\theta_2}]_i| \, \|s_k - z^*\|$$

$$+ 2CK \sum_{i \in U_{k+1}} |[H_{k+1,\theta_2}]_i| \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\| \,,$$

where we have used $|f_i(s_k) - f_i(z^*)| \leq C \|s_k - z^*\|$ which follows from the bounded subgradient condition (Assumption 2-e). Now, taking conditional expectation we have

$$\mathbb{E}\Bigg[-2 \sum_{i \in U_{k+1}} \gamma_{i,k+1}[f_i(s_k) - f_i(z^*)]$$

$$+ 2CK \sum_{i \in U_{k+1}} \gamma_{i,k+1} \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\| \, | \, F_k \Bigg]$$

$$\leq -2\frac{1}{(k+1)^{\theta_2}} \sum_{i=1}^m p_i^{1-\theta_2}[f_i(s_k) - f_i(z^*)]$$

$$+ \frac{2CK}{(k+1)^{\theta_2}} \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\| \sum_{i=1}^m p_i^{1-\theta_2}$$

$$+ 2C \left(\|s_k - z^*\| + K \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\| \right)$$

$$\times \mathbb{E}\Bigg[\sum_{i \in U_{k+1}} |[H_{k+1,\theta_2}]_i| \, | \, F_k \Bigg] \,.$$

The result follows from the above relation and Eq. (34). ∎

B. Almost sure convergence

We are now ready to derive one of our main results regarding the almost sure convergence of the algorithm in Eq. (4). In this section we prove that almost sure convergence holds under two cases. In

one case we assume that the quadratic form generated by the matrix $\mathbf{P\bar{W}}$ is positive everywhere except on the consensus subspace $\mathcal{C}$, where it vanishes. In this case, we can prove asymptotic convergence for the step size $\gamma_{i,k} = \frac{1}{\Gamma_{i,k}}$, i.e., $\theta_2 = 1$. We also require that $\alpha_{i,k} = \frac{1}{\Gamma_{i,k}^{\theta_1}}$ with the condition $\theta_1 \in (1/2, 1)$. Note that this is a stricter requirement than in the constrained consensus part, where θ_1 could take the value 1. Thus, we need the step sizes $\gamma_{i,k}$ to decay at a faster rate than $\alpha_{i,k}$ for all i.

In the second case we assume that the random network has the property that the probability of update p_i is the same for all agents i, i.e., $p_1 = \ldots = p_m = p$. In this case the earlier condition reduces to positive semidefiniteness of $\mathbf{\bar{W}}$ together with the requirement that $\mathcal{N}(\mathbf{\bar{W}}) = \mathcal{C}$. However in this case we are free to chose a wider class of step sizes $\gamma_{i,k} = \frac{1}{[\Gamma_{i,k}]^{\theta_2}}$. The faster decay constraint on $\gamma_{i,k}$ in this case arises as the requirement that $\frac{1+\theta_1}{2} < \theta_2 \leq 1$. Note that as illustrated in the section on Gossip and Broadcast communication protocols the requirement of a uniform update probability translates to the requirement that the graph G is regular. In this case the update probability is $p = \frac{2}{m}$ for gossip and $p = \frac{r}{m}$ for broadcast algorithm, where r is the common degree of each node.

The following theorem considers the first case. The theorem relies on the positive semidefiniteness of the matrix $P\bar{W}$, for which we recall that $P = D\left(\frac{1}{p_i^{\theta_1}}\right)$, p_i is the probability of update at any instance by agent i, and $\bar{W} = \mathbb{E}[W_k]$.

Theorem 3: Let Assumption 1 on the network hold. Also, let Assumption 2 on the constraint sets and objective functions hold as well as Assumption 3 on the link noise and subgradient errors. Further, let the step sizes be such that $\gamma_{i,k} = \frac{1}{\Gamma_{i,k}}$ and $\alpha_{i,k} = \frac{1}{\Gamma_{i,k}^{\theta_1}}$ with $\theta_1 \in (1/2, 1)$. Assume that the matrix $P\bar{W}$ is positive semidefinite and that the vector $\mathbf{1}_m$ is unique (up to scaling) nonzero vector such that $P\bar{W}\mathbf{1}_m = 0$. Then, almost surely, the iterate sequences $\{x_{i,k}\}$ generated by optimization algorithm (4) converge to a common random point $\tilde{z}^*$ in the optimal set X^*.

Proof: We use Lemma 9 with $\theta_2 = 1$ and obtain for any $z^* \in X^*$,

$$\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k\right] \leq \|\mathbf{x}_k - \mathbf{z}^*\|^2 - \frac{2}{(k+1)^{\theta_1}}\mathbf{x}_k'\mathbf{P\bar{W}x}_k$$

$$- \frac{2}{k+1}[f(s_k) - f^*] + \frac{2mCK}{k+1}\sum_{j=1}^{m}\sum_{r=1}^{m}\|x_{j,k} - x_{r,k}\| + \varphi_k,$$

where $f(x) = \sum_{i=1}^m f_i(x)$, f^* is the optimal value of the problem, $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$, and φ_k is given by

$$
\begin{aligned}
\varphi_k = &\left(\bar{N}\eta^2\mu^2 + \bar{N}^2\eta^2 \sum_E C_{X_j} C_{X_{j'}} \right) \sum_{i \in V} \mathbb{E}\left[\alpha_{i,k+1}^2 \,|\, F_k \right] \\
&+ 2(C+\nu)^2 \mathbb{E}\left[\sum_{i \in U_{k+1}} \gamma_{i,k+1}^2 \,|\, F_k \right] \\
&+ 2\eta\bar{N} \sum_E C_{X_i} C_{X_j} \mathbb{E}\left[|[H_{k+1,\theta_1}]_i| \,|\, F_k \right] \\
&+ 2C \left(\|s_k - z^*\| + K \sum_{j=1}^m \sum_{r=1}^m \|x_{j,k} - x_{r,k}\| \right) \\
&\times \mathbb{E}\left[\sum_{i \in U_{k+1}} |[H_{k+1,\theta_2}]_i| \,|\, F_k \right].
\end{aligned}
$$

Now, for each pair $\{i, j\}$ of nodes in the graph $G = (V, E)$, we can find a path $i = s_1, \ldots, s_\ell = j$ from node i to node j, where $(s_{\tau-1}, s_\tau) \in E$ for $\tau = 2, \ldots, \ell$. Thus, each term $\|x_{i,k} - x_{j,k}\|$ can be bounded above by $\sum_{r=2}^\ell \|x_{s_r,k} - x_{s_{r-1},k}\|$. Using this we arrive at the bound

$$
\begin{aligned}
&\frac{2mCK}{(k+1)^{\theta_2}} \sum_{i=1}^m \sum_{j=1}^m \|x_{i,k} - x_{j,k}\| \\
&\leq \frac{2mCK}{k+1} \binom{m}{2} \sum_E \|x_{i,k} - x_{j,k}\|.
\end{aligned}
\tag{35}
$$

For convenience of notation, we define $K_1 = mCK\binom{m}{2}$, then for any constant $\psi > 0$ we have

$$
\begin{aligned}
&\frac{2K_1(k+1)^{\frac{\theta_1}{2}}}{\psi(k+1)} \frac{\psi}{(k+1)^{\frac{\theta_1}{2}}} \sum_E \|x_{i,k} - x_{j,k}\| \\
&\leq \frac{K_1^2}{\psi^2(k+1)^{2-\theta_1}} + \frac{\psi^2}{(k+1)^{\theta_1}} \sum_E \|x_{i,k} - x_{j,k}\|^2 \\
&= \frac{K_1^2}{\psi^2(k+1)^{2-\theta_1}} + \frac{\psi^2}{(k+1)^{\theta_1}} \mathbf{x}_k' \mathbf{L}\mathbf{x}_k,
\end{aligned}
$$

where $\mathbf{L} = L \otimes I_n$, and L is the graph Laplacian for the bidirectional graph G. Substituting this estimate in the relation for $\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \,|\, F_k \right]$, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \,|\, F_k \right] \leq &\|\mathbf{x}_k - \mathbf{z}^*\|^2 + \frac{K_1^2}{\psi^2(k+1)^{2-\theta_1}} \\
&- \frac{1}{(k+1)^{\theta_1}} \mathbf{x}_k' \left[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L} \right] \mathbf{x}_k - \frac{2}{k+1}[f(s_k) - f^*] + \varphi_k.
\end{aligned}
$$

We next discuss how to choose ψ so that the matrix $2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}$ is positive semidefinite. Let the eigenvalues of both $\mathbf{P}\bar{\mathbf{W}}$ and $\mathbf{L}$ be denoted in the increasing order $0 = \lambda_1(\mathbf{P}\bar{\mathbf{W}}) < \lambda_2(\mathbf{P}\bar{\mathbf{W}}) \leq \ldots \leq$

$\lambda_m(\mathbf{P}\bar{\mathbf{W}})$, and $0 = \lambda_1(\mathbf{L}) < \lambda_2(\mathbf{L}) \le \ldots \le \lambda_m(\mathbf{L})$. Note that according to our definition of the matrices $\mathbf{P}\bar{\mathbf{W}}$ and $\mathbf{L}$, each of the eigenvalues λ_i has multiplicity n. According to our assumption that $\mathbf{1}_m$ is the unique null vector of the matrix $P\bar{W}$, we have $\lambda_2(\mathbf{P}\bar{\mathbf{W}}) > 0$. Moreover, under the assumption that the graph $G = (V, E)$ is connected, we also have $\lambda_2(\mathbf{L}) > 0$. Since $L\mathbf{1}_m = 0$ and $P\bar{W}\mathbf{1}_m = 0$, it follows $\mathcal{C} \subset \mathcal{N}(2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L})$. Any vector can be decomposed as $\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2$, where $\mathbf{x}_1 \in \mathcal{C}$ and $\mathbf{x}_2 \perp \mathbf{x}_1$, so that

$$\mathbf{x}'[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}]\mathbf{x} = a_2^2\mathbf{x}_2'[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}]\mathbf{x}_2.$$

From the variational characterization of eigenvalues, we have for all vectors $\mathbf{x}_2 \perp \mathcal{C}$,

$$2\lambda_2(\mathbf{P}\bar{\mathbf{W}})\left\|\mathbf{x}_2\right\|^2 \le 2\mathbf{x}_2'\mathbf{P}\bar{\mathbf{W}}\mathbf{x}_2,$$

$$-\psi^2\lambda_m(\mathbf{L})\left\|\mathbf{x}_2\right\|^2 \le -\psi^2\mathbf{x}_2'\mathbf{L}\mathbf{x}_2.$$

Hence, we see that for any vector $\mathbf{x}$ and $\psi > 0$,

$$\left(2\lambda_2(\mathbf{P}\bar{\mathbf{W}}) - \psi^2\lambda_m(\mathbf{L})\right)\left\|\mathbf{x}_2\right\|^2 \le \mathbf{x}'[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}]\mathbf{x}.$$

Thus, for $\psi > 0$ such that $\psi < \sqrt{\frac{2\lambda_2(\mathbf{P}\bar{\mathbf{W}})}{\lambda_m(\mathbf{L})}}$, we see that $2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}$ is positive semidefinite and its null-space coincides with the consensus space $\mathcal{C}$, i.e., $\mathbf{x}_k'[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}]\mathbf{x}_k \ge 0$ with equality holding only for $\mathbf{x} \in \mathcal{C}$. Next, note that since $\theta_1 \in (1/2, 1)$, we have that $\sum_{k=0}^\infty \frac{K_1^2}{\psi^2(k+1)^{2-\theta_1}} < \infty$. Also, by Lemma 5 and the compactness assumption on the sets X_i, we deduce that $\sum_{k=0}^\infty \varphi_k < \infty$ with probability one. Now, we apply the supermartingale convergence result of Lemma 3 and obtain that, with probability one, both $\left\|\mathbf{x}_k - \mathbf{z}^*\right\|^2$ converges for every $z^* \in X^*$ and the following sum

$$\sum_{k=0}^\infty \left[\frac{1}{(k+1)^{\theta_1}}\mathbf{x}_k'\left[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}\right]\mathbf{x}_k + \frac{2}{k+1}[f(s_k) - f^*]\right]$$

is finite. Since $\frac{1}{k+1} < \frac{1}{(k+1)^{\theta_1}}$ for $\theta_1 < 1$, we have

$$\sum_{k=0}^\infty \frac{1}{k+1}\left[\mathbf{x}_k'\left[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}\right]\mathbf{x}(k) + [f(s_k) - f(z^*)]\right] < \infty.$$

Furthermore, since $\sum_{k=0}^\infty \frac{1}{k+1} = \infty$ it follows that, with probability one, there exists a subsequence such that $\lim_{\ell\to\infty}\mathbf{x}_{k_\ell}'\left[2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}\right]\mathbf{x}_{k_\ell} = 0$ and $\lim_{\ell\to\infty}[f(s_{k_\ell}) - f^*] = 0$. By our choice of ψ, the matrix $2\mathbf{P}\bar{\mathbf{W}} - \psi^2\mathbf{L}$ is positive semidefinite and it vanishes only on the consensus subspace $\mathcal{C}$; therefore, the sequence $\{\mathbf{x}_{k_\ell}\}$ approaches the consensus subspace $\mathcal{C}$ with probability one, i.e., $\lim_{\ell\to\infty}\|x_{i.k_\ell} - x_{j,k_\ell}\| = 0$ for every i, j with probability one. On the other hand, using Eq. (30), we obtain

$$\sum_{i=1}^m \|x_{i,k_\ell} - s_{k_\ell}\| \le \frac{mB + \delta}{\delta}\sum_{i=1}^m\sum_{j=1}^m \|x_{i,k_\ell} - x_{j,k_\ell}\|. \tag{36}$$

Thus, with probability one we have for every i,

$$\lim_{\ell \to \infty} \|x_{i,k_\ell} - s_{k_\ell}\| = 0. \tag{37}$$

Since $\lim_{\ell \to \infty} f(s_{k_\ell}) = f^*$ and the function f is continuous, and since the sets X_i are compact (Assumption 2-b), we conclude that there exists a subsequence along which s_{k_ℓ} converges almost surely to a (random) point $\tilde{z}^*$ that lies in the set X^*. Without any loss of generality, we can assume that the sequence $\{s_{k_\ell}\}$ itself converges to the limit point $\tilde{z}^*$ almost surely. By Eq. (37) it follows that x_{i,k_ℓ} converges to $\tilde{z}^*$ for all i almost surely. This and the fact $\|\mathbf{x}_k - \mathbf{z}^*\|^2$ converges almost surely for $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$ with any $z^* \in X^*$ (as established earlier), imply that the sequence $\{\mathbf{x}_k\}$ must converge to $\mathbf{1}_m \otimes \tilde{z}^*$ almost surely. Hence, the sequences $\{x_{i,k}\}$, $i = 1, \ldots, m$, converge almost surely to a common (random) point in the set X^*. ∎

We now provide a convergence result for the case when the probability of updates for each agents is the same. We have the following result.

Theorem 4: Let Assumption 1 on the network hold. Also, let Assumption 2 on the constraint sets and objective functions hold as well as Assumption 3 on the link noise and subgradient errors. Assume that $p_i = p$ for all i. Further, let the step sizes be such that $\gamma_{i,k} = \frac{1}{\Gamma_{i,k}^{\theta_2}}$ and $\alpha_{i,k} = \frac{1}{\Gamma_{i,k}^{\theta_1}}$, where $\theta_2 > \frac{1+\theta_1}{2}$ and $\theta_1 \in (1/2, 1)$. Then, almost surely, the iterate sequences $\{x_{i,k}\}$ of optimization algorithm (4) converge to a common random point in the optimal set X^*.

Proof: We use Lemma 9 where $p_i = p$ for all i, and we obtain

$$\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k\right] \le \|\mathbf{x}_k - \mathbf{z}^*\|^2 - \frac{2p^{-\theta_1}}{(k+1)^{\theta_1}} \mathbf{x}_k' \bar{\mathbf{W}} \mathbf{x}_k$$

$$- \frac{2p^{1-\theta_2}}{(k+1)^{\theta_2}}[f(s_k) - f^*]$$

$$+ \frac{2p^{1-\theta_2}CK}{(k+1)^{\theta_2}} \sum_{j=1}^{m} \sum_{r=1}^{m} \|x_{j,k} - x_{r,k}\| + \varphi_k, \tag{38}$$

where φ_k denotes the remaining terms, $f(x) = \sum_i f_i(x)$, and f^* is the optimal value of the problem. Proceeding similarly as in the derivation of Eq. (35), we arrive at the following:

$$\frac{2p^{1-\theta_2}CK}{(k+1)^{\theta_2}} \sum_{i=1}^{m} \sum_{j=1}^{m} \|x_{i,k} - x_{j,k}\|$$

$$\le \frac{2p^{1-\theta_2}CK}{(k+1)^{\theta_2}} \binom{m}{2} \sum_{E} \|x_{i,k} - x_{j,k}\|.$$

Let us define $K_2 = p^{1-\theta_2} C K \binom{m}{2}$, so we can write

$$\frac{2K_2(k+1)^{\frac{\theta_1}{2}}}{\psi(k+1)^{\theta_2}} \frac{\psi}{(k+1)^{\frac{\theta_1}{2}}} \sum_{(i,j)\in E} \|x_{i,k} - x_{j,k}\|$$

$$\leq \frac{K_2^2}{\psi^2(k+1)^{2\theta_2-\theta_1}} + \frac{\psi^2}{(k+1)^{\theta_1}} \sum_E \|x_{i,k} - x_{j,k}\|^2$$

$$= \frac{K_2^2}{\psi^2(k+1)^{2\theta_2-\theta_1}} + \frac{\psi^2}{(k+1)^{\theta_1}} \mathbf{x}_k' \mathbf{L} \mathbf{x}_k.$$

Substituting this relation in Eq. (38) we obtain

$$\mathbb{E}\left[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k\right] \leq \|\mathbf{x}_k - \mathbf{z}^*\|^2 + \frac{K_2^2}{\psi^2(k+1)^{2\theta_2-\theta_1}}$$

$$- \frac{1}{(k+1)^{\theta_1}} \mathbf{x}_k' \left[2p^{-\theta_1}\bar{\mathbf{W}} - \psi^2\mathbf{L}\right] \mathbf{x}_k$$

$$- \frac{2p^{1-\theta_2}}{(k+1)^{\theta_2}} [f(s_k) - f^*] + \varphi_k.$$

Now, by assumption we have $\theta_2 > \frac{1+\theta_1}{2}$, implying that $\sum_{k=0}^{\infty} \frac{K_2^2}{\psi^2(k+1)^{2\theta_2-\theta_1}} < \infty$. We choose ψ such that $0 < \psi < \sqrt{\frac{2p^{-\theta_1}\lambda_2(\bar{\mathbf{W}})}{\lambda_m(\mathbf{L})}}$, so that the matrix $2p^{-\theta_1}\bar{\mathbf{W}} - \psi^2\mathbf{L}$ is positive semidefinite. The rest of the proof follows using similar arguments as in the proof of Theorem 3. ∎

C. Constant step size error bound

In this section we derive asymptotic error bounds when each agent uses constant step sizes α_i, and γ_i. The main additional assumption we need the requirement that the local agent functions $f_i(x)$ are continuously differentiable and strongly convex, i.e., for all $i \in V$,

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

Note that the convexity parameter σ_i of each agent is allowed to be different. This problem was also considered in [43] in the absence of local constraint sets X_i on the decision variables.

Theorem 5: Let Assumption 1 on the network hold. Let also Assumptions 2 and 3 hold. Further assume that each function f_i is continuously differentiable and strongly convex with a constant $\sigma_i > 0$. Also, assume that the step size α_i is such that $\alpha_i \leq \frac{1}{\eta|N_i|}$ and the step size γ_i satisfies $\gamma_i < \frac{1}{2\sigma_i}$ for all i. Then, for the iterate sequences $\{x_{i,k}\}$ of the optimization algorithm in Eq. (4), we have the following asymptotic bound:

$$\limsup_{k\to\infty} \mathbb{E}\left[\|\mathbf{x}_k - \mathbf{z}^*\|^2\right] \leq \frac{\epsilon_1}{2\underline{\gamma}\,\underline{\sigma}\,\underline{p}} + \epsilon_2,$$

where $\mathbf{z}^* = \mathbf{1}_m \otimes z^*$ for any $z^* \in X^*$, and

$$\epsilon_1 = 2(C + \nu)^2 \sum_{i \in V} p_i \gamma_i^2 + 2m \Delta_{\gamma,p} C C_X$$

$$+ 2\Delta_{\gamma,\sigma}(1 - \underline{p}) \sum_{i \in V} C_{X_i}^2$$

$$+ (1 - 2\underline{\gamma\sigma}) \left(2\eta \Delta_\alpha \bar{N} \sum_E C_{X_i} C_{X_j} \right.$$

$$\left. + m\bar{N}\eta^2\mu^2\bar{\alpha}^2 + \bar{\alpha}^2 \|\mathbf{R}\| \sum_i C_{X_i}^2 \right)$$

$$\epsilon_2 = \frac{\bar{\gamma}^2 C^2 \lambda_m(\mathbf{L})\tilde{K}^2}{\underline{\gamma\sigma}\,\underline{p}(1 - 2\underline{\gamma\sigma})\underline{\alpha}\lambda_2(\bar{\mathbf{W}})}, \qquad \tilde{K} = m\frac{mB + \delta}{2\delta}\binom{m}{2}.$$

Proof: From Eq. (24) we have the following relation

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}] \le \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$

$$+ 2\gamma_i^2 \mathbb{E}\left[\|\tilde{d}_{i,k+1}\|^2 \mid F_k, W_{k+1}\right] \chi_{\{i \in U_{k+1}\}}$$

$$- 2\gamma_i \mathbb{E}[d'_{i,k+1}(v_{i,k+1} - z^*) \mid F_k, W_{k+1}]\chi_{\{i \in U_{k+1}\}}, \tag{39}$$

where $d_{i,k} = \nabla f_i(v_{i,k})$ and $\tilde{d}_{i,k}$ is a noisy gradient. Under the assumption of strong convexity of the function f_i, we obtain

$$d'_{i,k+1}(v_{i,k+1} - z^*) = (d_{i,k+1} - \nabla f_i(z^*))'(v_{i,k+1} - z^*)$$

$$+ \nabla f_i(z^*)'(v_{i,k+1} - z^*)$$

$$\ge \sigma_i \|v_{i,k+1} - z^*\|^2 + \nabla f_i(z^*)'(v_{i,k+1} - z^*).$$

Let $i \in U_{k+1}$. Then, we add and subtract the term $-2\gamma_i \nabla f_i(z^*)'s_k$, where $s_k = \frac{1}{m}\sum_{i=1}^m P_X[x_{i,k}]$ to get

$$-2\gamma_i d'_{i,k+1}(v_{i,k+1} - z^*) \le -2\gamma_i \sigma_i \|v_{i,k+1} - z^*\|^2$$

$$- 2\gamma_i \nabla f_i(z^*)'(s_k - z^*) + 2\gamma_i \nabla f_i(z^*)'(s_k - v_{i,k+1}).$$

Then, letting $y_{i,k} = \mathbb{E}[v_{i,k+1}|F_k, W_{k+1}]$ and using the gradient boundedness, we obtain

$$-2\gamma_i \mathbb{E}[d'_{i,k+1}(v_{i,k+1} - z^*) \mid F_k, U_{k+1}]$$

$$\le -2\gamma_i \sigma_i \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$

$$- 2\gamma_i \nabla f_i(z^*)'(s_k - z^*) + 2\gamma_i C \|y_{i,k} - s_k\|.$$

Thus, plugging back in Eq. (39), we get the relation

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$

$$\leq \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}] + 2\gamma_i^2(C+\nu)^2 \chi_{\{i \in U_{k+1}\}}$$

$$- 2\gamma_i \sigma_i \mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}] \chi_{\{i \in U_{k+1}\}}$$

$$+ 2\gamma_i \left(C \|y_{i,k} - s_k\| - \nabla f_i(z^*)'(s_k - z^*) \right) \chi_{\{i \in U_{k+1}\}}.$$

Using the fact that $v_{i,k+1} = x_{i,k}$ for $i \notin U_{k+1}$, the preceding relation can equivalently be written as

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k, U_{k+1}]$$

$$\leq (1 - 2\gamma_i \sigma_i)\mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k, W_{k+1}]$$

$$+ 2\gamma_i^2(C+\nu)^2 \chi_{\{i \in U_{k+1}\}}$$

$$+ 2\gamma_i \sigma_i \|x_{i,k} - z^*\|^2 \left(1 - \chi_{\{i \in U_{k+1}\}}\right)$$

$$+ 2\gamma_i \left(C \|y_{i,k} - s_k\| - \nabla f_i(z^*)'(s_k - z^*) \right) \chi_{\{i \in U_{k+1}\}}.$$

Now, taking conditional expectation with respect to the history F_k, we obtain

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k] \leq (1 - 2\gamma_i \sigma_i)\mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k]$$

$$+ 2p_i\gamma_i^2(C+\nu)^2 + 2\gamma_i \sigma_i \|x_{i,k} - z^*\|^2 \left(1 - \underline{p}\right)$$

$$+ 2\gamma_i C \mathbb{E}\left[\|y_{i,k} - s_k\| \chi_{\{i \in U_{k+1}\}} \mid F_k \right]$$

$$- 2\gamma_i p_i \nabla f_i(z^*)'(s_k - z^*),$$

where $\underline{p} = \min_i\{p_i\}$. Letting $\Delta_{\gamma,p} = \max_i\{\gamma_i p_i\} - \min_j\{\gamma_j p_j\}$ and similarly defining $\Delta_{\gamma,\sigma}$, we have $-\gamma_i p_i \leq -\underline{\gamma p} + \Delta_{\gamma,p}$ and $\gamma_i \sigma_i \leq \underline{\gamma \sigma} + \Delta_{\gamma,\sigma}$. Therefore, it follows

$$\mathbb{E}[\|x_{i,k+1} - z^*\|^2 \mid F_k] \leq (1 - 2\underline{\gamma\sigma})\mathbb{E}[\|v_{i,k+1} - z^*\|^2 \mid F_k]$$

$$+ 2p_i\gamma_i^2(C+\nu)^2 + 2\underline{\gamma\sigma}(1 - \underline{p}) \|x_{i,k} - z^*\|^2$$

$$+ 2\bar{\gamma}C \mathbb{E}\left[\|y_{i,k} - s_k\| \chi_{\{i \in U_{k+1}\}} \mid F_k \right]$$

$$- 2\underline{\gamma p} \nabla f_i(z^*)'(s_k - z^*) + 2\Delta_{\gamma,p}CC_X + 2\Delta_{\gamma,\sigma}(1 - \underline{p})C_{X_i}^2.$$

Summing over all i we obtain

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k] \leq (1 - 2\underline{\gamma\sigma})\mathbb{E}[\|\mathbf{v}_{k+1} - \mathbf{z}^*\|^2 \mid F_k]$$

$$+ 2(C + \nu)^2 \sum_{i \in V} p_i \gamma_i^2 + 2\underline{\gamma\sigma}(1 - \underline{p}) \|\mathbf{x}_k - \mathbf{z}^*\|^2$$

$$+ 2\bar{\gamma}C\mathbb{E}\left[\sum_{i \in U_{k+1}} \|y_{i,k} - s_k\| \mid F_k\right] + 2\Delta_{\gamma,\sigma}(1 - \underline{p}) \sum_{i \in V} C_{X_i}^2$$

$$- 2\underline{\gamma p}\nabla f(z^*)'(s_k - z^*) + 2m\Delta_{\gamma,p}CC_X.$$

Using relations in (31) and (33), where we only use the bound on the term with noise, we can see that

$$\mathbb{E}\left[\|\mathbf{v}_{k+1} - \mathbf{z}^*)\|^2 \mid F_k\right] = \|\mathbf{x}_k - \mathbf{z}^*\|^2$$

$$- 2(\mathbf{x}_k - \mathbf{z}^*)'\bar{\Lambda}\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*) + \bar{\alpha}^2 \mathbf{x}_k' \mathbf{R} \mathbf{x}_k + m\bar{N}\eta^2\mu^2\bar{\alpha}^2.$$

Further, we have

$$-2(\mathbf{x}_k - \mathbf{z}^*)'\bar{\Lambda}\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*) \leq -2\underline{\alpha}(\mathbf{x}_k - \mathbf{z}^*)'\bar{\mathbf{W}}(\mathbf{x}_k - \mathbf{z}^*)$$

$$+ 2\eta\Delta_{\alpha}\bar{N} \sum_{E} C_{X_i}C_{X_j},$$

and $\bar{\alpha}^2 \mathbf{x}_k' \mathbf{R} \mathbf{x}_k \leq \bar{\alpha}^2\|\mathbf{R}\| \sum_i C_{X_i}^2$. We also have $\nabla f(z^*)'(s_k - z^*) \geq 0$ since z^* is the optimal point of the objective $f(x) = \sum_{i \in V} f_i(x)$ and $s_k \in X$ (recall that s_k is the average sum of the projections $P_X[x_{j,k}]$). Thus, by neglecting this term and using $1 - 2\underline{\gamma\sigma} < 1$, we obtain

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k] \leq (1 - 2\underline{\gamma\sigma}\,\underline{p})\mathbb{E}[\|\mathbf{x}_k - \mathbf{z}^*\|^2 \mid F_k]$$

$$+ 2(C + \nu)^2 \sum_{i \in V} p_i \gamma_i^2 + 2\bar{\gamma}C\mathbb{E}\left[\sum_{i \in U_{k+1}} \|y_{i,k} - s_k\| \mid F_k\right]$$

$$+ 2m\Delta_{\gamma,p}CC_X + 2\Delta_{\gamma,\sigma}(1 - \underline{p}) \sum_{i \in V} C_{X_i}^2$$

$$+ (1 - 2\underline{\gamma\sigma})\bar{\alpha}^2\|\mathbf{R}\| \sum_i C_{X_i}^2$$

$$+ (1 - 2\underline{\gamma\sigma})\left(2\eta\Delta_{\alpha}\bar{N} \sum_{E} C_{X_i}C_{X_j} + m\bar{N}\eta^2\mu^2\bar{\alpha}^2\right).$$

We can write the preceding relation more compactly as:

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k] \leq (1 - 2\underline{\gamma\sigma}\,\underline{p})\mathbb{E}[\|\mathbf{x}_k - \mathbf{z}^*\|^2 \mid F_k]$$

$$+ 2\bar{\gamma}C\mathbb{E}\left[\sum_{i \in U_{k+1}} \|y_{i,k} - s_k\| \mid F_k\right] + \epsilon_1, \qquad (40)$$

where ϵ_1 includes all the other terms. Note that from Eqs. (29) and (30), similar to the derivation of relation (35), we can see that

$$\sum_{i \in U_{k+1}} \|y_{i,k} - s_k\| \leq m \frac{mB + \delta}{\delta} \sum_{i=1}^{m} \sum_{j=1}^{m} \|x_{i,k} - x_{j,k}\|$$

$$\leq 2\tilde{K} \sum_{E} \|x_{i,k} - x_{j,k}\|,$$

where $\tilde{K} = m \frac{mB+\delta}{2\delta} \binom{m}{2}$. We further have

$$2\tilde{K} \sum_{E} \|x_{i,k} - x_{j,k}\| = \frac{2\tilde{K}}{\psi} \sum_{E} \psi \|x_{i,k} - x_{j,k}\|$$

$$\leq \frac{\tilde{K}^2}{\psi^2} + \psi^2 \sum_{E} \|x_{i,k} - x_{j,k}\|^2$$

$$= \frac{\tilde{K}^2}{\psi^2} + \psi^2 \mathbf{x}_k' \mathbf{L} \mathbf{x}_k,$$

where $\psi > 0$ is an arbitrary scalar to be chosen later. When substituted back in Eq. (40) this yields

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2 \mid F_k] \leq (1 - 2\underline{\gamma}\sigma\,\underline{p})\mathbb{E}[\|\mathbf{x}_k - \mathbf{z}^*\|^2 \mid F_k] + \epsilon_1$$

$$+ 2\bar{\gamma}C\frac{\tilde{K}^2}{\psi^2} - (1 - 2\underline{\gamma}\sigma)\mathbf{x}_k' \left[2\underline{\alpha}\bar{\mathbf{W}} - 2\frac{\bar{\gamma}C\psi^2}{1 - 2\underline{\gamma}\sigma}\mathbf{L} \right] \mathbf{x}_k.$$

According to the condition on step sizes γ_i, we have $1 - 2\underline{\gamma}\sigma > 0$. Moreover, under the connectivity of the network, we have $\lambda_2(\bar{\mathbf{W}}) > 0$ and $\lambda_2(\mathbf{L}) > 0$. Then, it can be seen that for any ψ such that

$$0 < \psi \leq \sqrt{\frac{(1 - 2\underline{\gamma}\sigma)\underline{\alpha}\lambda_2(\bar{\mathbf{W}})}{\bar{\gamma}C\lambda_m(\mathbf{L})}},$$

the matrix $2\underline{\alpha}\bar{\mathbf{W}} - 2\frac{\bar{\gamma}C\psi^2}{1-2\underline{\gamma}\sigma}\mathbf{L}$ is positive semidefinite. Neglecting that term and taking full expectation, we obtain

$$\mathbb{E}[\|\mathbf{x}_{k+1} - \mathbf{z}^*\|^2] \leq (1 - 2\underline{\gamma}\sigma\,\underline{p})\mathbb{E}[\|\mathbf{x}_k - \mathbf{z}^*\|^2] + \epsilon_1$$

$$+ 2\bar{\gamma}C\frac{\tilde{K}^2}{\psi^2}.$$

Now we can apply the result of Lemma 4 to deduce

$$\limsup_{k \to \infty} \mathbb{E}\left[\|\mathbf{x}_k - \mathbf{z}^*\|^2 \right] \leq \frac{\epsilon_1}{2\underline{\gamma}\sigma\,\underline{p}} + \frac{\bar{\gamma}C}{\underline{\gamma}\sigma\,\underline{p}} \frac{\tilde{K}^2}{\psi^2}.$$

We can minimize the error by letting $\psi = \sqrt{\frac{(1 - 2\underline{\gamma}\sigma)\underline{\alpha}\lambda_2(\bar{\mathbf{W}})}{\bar{\gamma}C\lambda_m(\mathbf{L})}}$, which when substituted in the preceding relation yields the desired result. ∎

Note that the condition $\gamma_i < \frac{1}{2\sigma_i}$ can be easily implemented in a distributed manner as each agent knows its own objective function. The error $\epsilon_2 = \frac{\bar{\gamma}^2 C^2 \lambda_m(\mathbf{L})\tilde{K}^2}{\underline{\gamma\sigma}\,\underline{p}(1-2\underline{\gamma\sigma})\underline{\alpha}\lambda_2(\bar{\mathbf{W}})}$ emphasizes the effect of the network structure on the asymptotic error. It can be seen that networks with large eigenvalue $\lambda_2(\bar{\mathbf{W}})$ have reduced asymptotic error. To see the effect of step size on the error ϵ_2, let us assume that the step sizes $\gamma_i = \gamma$ for all i. Then, the error ϵ_2 reduces to $\epsilon_2 = \frac{\gamma C^2 \lambda_m(\mathbf{L})\tilde{K}^2}{\underline{\sigma p}(1-2\gamma\underline{\sigma})\underline{\alpha}\lambda_2(\bar{\mathbf{W}})}$, which evidently decreases with decreasing step size γ. Note that for the special case when all the step sizes α_i and γ_i are chosen to be equal across the agents, the probability of update for all agents is p, and the functions f_i have the same convexity parameter σ, the upper bound reduces to $\epsilon = \frac{m(C+\nu)^2}{\sigma}\gamma + \frac{1}{p}(\frac{1}{2\gamma\sigma}-1)(m\bar{N}\eta^2\mu^2\alpha^2 + \alpha^2\|\mathbf{R}\|\sum_i C_{X_i}^2) + \epsilon_2$. It is clear that the first term is the contribution of the subgradient error and the term involving μ^2 is the contribution of the communication noise. As noted earlier, for gossip and broadcast algorithms the norm $\|\mathbf{R}\|$ can be written in terms of $\lambda_m(\bar{\mathbf{W}})$. We can choose to minimize the total error bound by choosing the various step sizes carefully. Another interesting fact is that the convexity parameters affect the error bound inversely. This implies that the asymptotic error is smaller when the objective functions have a higher *curvature* as characterized by their strong convexity parameter.

VI. CONCLUSION

In this paper we considered the problem of reaching agreement on a set of local variables and the problem of minimizing the sum of local objective functions when the local variables are constrained to local convex constraint sets. We proposed algorithms for these problems which can be applied over a random communication network. We showed that our model of the random communication network is general enough to include the widely used gossip and broadcast based communication protocols arising in wireless networks. Further, our algorithms are robust to the presence of communication noise and errors in the evaluation of subgradients of the objective functions. This generality allows us to consider the distributed stochastic optimization problem in our framework. We established conditions under which we can guarantee almost sure convergence of our algorithms, and provided asymptotic error bounds when almost sure convergence cannot be achieved.

APPENDIX

Proof of Lemma 5: Note that according to our definition $\Gamma_{i,k} = \sum_{t=1}^{k} \chi_{E_i(t)}$, where $E_i(t) = \{i \in U(t)\}$, and $\chi_{E_i(t)}$ is the indicator function of the event that agent i updates its local variable at instance t. Since according to our assumption the random graph sequence $\{\mathcal{W}(k)\}$ is assumed to be drawn in an $i.i.d$ fashion, this implies that the events $E_i(t)$ are $i.i.d$ and $\mathbb{E}[\chi_{E_i(t)}] = \mathbb{P}\{i \in U(t)\} = p_i$ for each

$i \in V$. By the law of iterated logarithms ([44], pages 476–479), for any $q > 0$, the following holds with probability 1,

$$\limsup_{k \to \infty} \frac{|\Gamma_{i,k} - kp_i|}{k^{\frac{1}{2}+q}} = 0 \qquad \text{for all } i \in V.$$

Hence, given any constant c there exists a large enough $\tilde{k}$ with probability one such that

$$\frac{|\Gamma_{i,k} - kp_i|}{k^{\frac{1}{2}+q}} \le c \qquad \text{for all} \quad k > \tilde{k}.$$

Then it can be shown [45] that the following bound holds with probability one

$$\frac{1}{\Gamma_{i,k}} \le \frac{2}{kp_i} \qquad \text{for all} \quad k > \tilde{k}. \tag{41}$$

Now, we consider the term $\left| \frac{1}{\Gamma_{i,k}} - \frac{1}{p_i k} \right|$. Clearly, we have with probability one

$$\left| \frac{1}{\Gamma_{i,k}} - \frac{1}{p_i k} \right| = \frac{1}{\Gamma_{i,k} p_i k} |\Gamma_{i,k} - p_i k| \le \frac{2}{p_i^2 k^2} c k^{\frac{1}{2}+q} = \frac{2c}{p_i^2 k^{\frac{3}{2}-q}}.$$

Using the notation $\underline{p} = \min_i \{p_i\}$ we have

$$\left| \frac{1}{\Gamma_{i,k}} - \frac{1}{p_i k} \right| \le \frac{2c}{\underline{p}^2 k^{\frac{3}{2}-q}}.$$

The above bound was derived for the case of gossip and broadcast communication in [31] and [45] respectively. Now for our general case we need a similar bound on the difference $\left| \frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta} \right|$, where $\theta \in (1/2, 1]$. By applying the mean value theorem to the function x^θ we get

$$\left| \frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta} \right| \le \frac{\theta}{x^{1-\theta}} \left| \frac{1}{[\Gamma_{i,k}]} - \frac{1}{p_i k} \right|,$$

where x is between $\frac{1}{\Gamma_{i,k}}$, and $\frac{1}{p_i k}$. However since both $\Gamma_{i,k} \le k$, and $p_i k \le k$, we get $\frac{1}{x} \le k$. Thus using this as an upper bound and using the earlier bound, we get that the following bound holds with probability one for every $i \in V$ and $k \ge \tilde{k}$,

$$\left| \frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta} \right| \le \frac{2\theta c k^{1-\theta}}{\underline{p}^2 k^{\frac{3}{2}-q}} = \frac{2\theta c}{k^{\frac{1}{2}+\theta-q} \underline{p}^2}.$$

Then, clearly we have that with probability one

$$\mathbb{E}\left[\left| \frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta} \right| \mid F_k \right] \le \frac{2\theta c k^{1-\theta}}{\underline{p}^2 k^{\frac{3}{2}-q}} = \frac{2\theta c}{k^{\frac{1}{2}+\theta-q} \underline{p}^2},$$

for all $k \geq \tilde{k}$. Now summing up we obtain

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\left|\frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta}\right| \mid F_k\right]$$

$$= \sum_{k=1}^{\tilde{k}} \mathbb{E}\left[\left|\frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta}\right| \mid F_k\right]$$

$$+ \sum_{k=\tilde{k}+1}^{\infty} \mathbb{E}\left[\left|\frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta}\right| \mid F_k\right]$$

$$\leq \sum_{k=1}^{\tilde{k}} \mathbb{E}\left[\left|\frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta}\right| \mid F_k\right] + \sum_{k=\tilde{k}+1}^{\infty} \frac{2\theta c}{k^{\frac{1}{2}+\theta-q}\underline{p}^2}.$$

However since $\theta \in (1/2, 1]$, and $q > 0$ is any arbitrary constant, we have $\sum_{k=\tilde{k}+1}^{\infty} \frac{2\theta c}{k^{\frac{1}{2}+\theta-q}\underline{p}^2} < \infty$ with probability one. Thus, with probability one,

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\left|\frac{1}{[\Gamma_{i,k}]^\theta} - \frac{1}{p_i^\theta k^\theta}\right| \mid F_k\right] < \infty.$$

A similar argument can be carried out starting from Eq. (41) to show that with probability one,

$$\sum_{k=1}^{\infty} \mathbb{E}\left[\frac{1}{\Gamma_{i,k}^{2\theta}} \mid F_k\right] < \infty.$$

REFERENCES

[1] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, Boston, 1984.

[2] J. N. Tsitsiklis and M. Athans, "Convergence and asymptotic agreement in distributed decision problems," *IEEE Transactions on Automatic Control*, vol. 29, pp. 42–50, 1984.

[3] A. Jadbabaie, J. Lin, and S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, pp. 988–1001, 2003.

[4] L. Xiao, S. Boyd, and S. J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, pp. 33–46, 2005.

[5] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 355–369, 2009.

[6] A. Nedić and D. P. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 109–138, 2001.

[7] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[8] B. Johansson, M. Rabi, and M. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM Journal on Optimization*, vol. 20, no. 3, pp. 1157–1170, 2009.

[9] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," LIDS, MIT, Tech. Rep. 2800, 2008.

[10] D. Mosk-Aoyama, T. Roughgarden, and D. Shah, "Fully distributed algorithms for convex optimization problems," in *DISC '07: Proceedings of the 21st international symposium on Distributed Computing*. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 492–493.

[11] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, pp. 922–938, 2010.

[12] S. S. Stanković, M. S. Stanković, and D. Stipanović, "Consensus based overlapping decentralized estimator," *IEEE Transactions on Automatic Control*, vol. 54, pp. 410–415, 2009.

[13] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," 2008, preprint. [Online]. Available: http://arxiv.org/abs/0809.0009

[14] A. Kashyap, T. Basar, and R. Srikant, "Quantized consensus," *Automatica*, vol. 43, pp. 1192–1203, 2007.

[15] N. Chopra and M. W. Spong, "On exponential synchronization of Kuramoto oscillators," *IEEE Transactions on Automatic Control*, vol. 54, pp. 353–357, 2009.

[16] M. Rabbat and R. D. Nowak, "Distributed optimization in sensor networks," in *IPSN*, 2004, pp. 20–27.

[17] S. S. Ram, V. V. Veeravalli, and A. Nedić, "Distributed non-autonomous power control through distributed convex optimization," in *IEEE INFOCOM*, 2009, pp. 3001–3005.

[18] B. Johansson, "On distributed optimization in networked systems," Ph.D. dissertation, Royal Institute of Technology (KTH), Dec. 2008, tRITA-EE 2008:065.

[19] T. Alpcan and C. Bauckhage, "A distributed machine learning framework," in *Proc. of 48th IEEE Conf. on Decision and Control*, Shanghai, China, December 2009, pp. 2546–2551.

[20] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "A collaborative training algorithm for distributed learning," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1856–1871, 2009.

[21] D. G. Luenberger, *Microeconomic Theory*. Mcgraw-Hill College, 1994.

[22] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[23] S. H. Low and D. E. Lapsley, "Optimization flow control—i: basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–874, 1999.

[24] R. Srikant, *The Mathematics of Internet Congestion Control*. Boston: Birkhäuser, 2003.

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

[26] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, 2009.

[27] B. Touri and A. Nedić, "Distributed consensus over network with noisy links," in *12th International Conference on Information Fusion*, Seattle, USA, July 2009, pp. 146–154.

[28] M. Huang and J. H. Manton, "Stochastic approximation for consensus seeking: Mean square and almost sure convergence," in *Proceedings of the 46th IEEE Conference on Decision and Control*, 2007, pp. 306–311.

[29] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas of Communications*, vol. 26, pp. 634–649, 2008.

[30] B. Touri, A. Nedić, and S. S. Ram, "Asynchronous stochastic convex optimization over random networks: Error bounds," in *Information Theory and Applications Workshop*, San Diego, USA, 2010.

[31] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Asynchronous gossip algorithms for stochastic optimization," in *GameNets'09:*

Proceedings of the First ICST international conference on Game Theory for Networks. Piscataway, NJ, USA: IEEE Press, 2009, pp. 80–81.

[32] L. Xiao, "Dual averaging methods for regularized stochastic learning and online optimization," Microsoft Research, Tech. Rep. MSR-TR-2010-23, 2009.

[33] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, 2003.

[34] Y. Ermoliev, *Stochastic programming methods*. Moscow: Nauka, 1976.

[35] ——, "Stochastic quasi-gradient methods and their application to system optimization," *Stochastics*, vol. 9, no. 1, pp. 1–36, 1983.

[36] B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software, Inc., 1987.

[37] Y. Ermoliev, "Stochastic quazigradient methods," in *Numerical Techniques for Stochastic Optimization, (Eds.) Y. Ermoliev and R. J-B. Wets, Springer-Verlag, N.Y.*, 1988, pp. 141–186.

[38] H. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 2003.

[39] D. P. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA, USA: Athena Scientific, 2003.

[40] P. Frasca and F. Fagnani, "Broadcast gossip averaging algorithms: interference and asymptotical error in large networks," 2010, preprint. [Online]. Available: arXiv:1005.1292v1[math.OC]

[41] S. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. New York, NY, USA: Cambridge University Press, 2009.

[42] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *U.S.S.R Computational Mathematics and Mathematical Physics*, vol. 7, pp. 1211–1228, 1967.

[43] I. Matei and J. Baras, "Performance evaluationof the consensus-based distributed subgradient method under random communication topologies," in *Proceedings of the 19th Internation Symposium of Mathematical Theory of Networks and Systems*, 2008, pp. 1247–1254.

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

[45] A. Nedić, "Asynchronous broadcast-based convex optimization over a network," 2011, to appear in IEEE Transactions on Automatic Control.

Kunal Srivastava received his Bachelor of Technology and Master of Technology under the dual degree program from Indian Institute of Technology, Madras, (IIT-M) India in 2004. He is currently a PhD. student in the Department of Industrial and Enterprise Systems Engineering at University of Illinois, Urbana-Champaign, (UIUC) U.S.A. He received his M.S. in Mathematics from UIUC in 2009.

His general research interest is in the area of optimization, signal processing, statistical learning and control theory. His current research is focused on developing distributed algorithms for large scale batch and online optimization problems arising in sensor networks and machine learning.

Angelia Nedić received her B.S. degree from the University of Montenegro (1987) and M.S. degree from the University of Belgrade (1990), both in Mathematics. She received her Ph.D. degrees from Moscow State University (1994) in Mathematics and Mathematical Physics, and from Massachusetts Institute of Technology in Electrical Engineering and Computer Science (2002). She has been at the BAE Systems Advanced Information Technology from 2002-2006. Since 2006 she is an Assistant professor at the Department of Industrial and Enterprise Systems Engineering (ISE), University of Illinois at Urbana-Champaign.

Her general interest is in optimization including fundamental theory, models, algorithms, and applications. Her current research interest is focused on large-scale convex optimization, distributed multi-agent optimization, and duality theory with applications in decentralized optimization. She received an NSF Faculty Early Career Development (CAREER) Award in 2008 in Operations Research.