Distributed Constrained Optimization over Noisy Networks

Kunal Srivastava, Angelia Nedić, Dušan M. Stipanović

Abstract— **In this paper we deal with two problems which are of great interest in the field of distributed decision making and control. The first problem we tackle is the problem of achieving consensus on a vector of local decision variables in a network of computational agents when the decision variables of each node are constrained to lie in a subset of the Euclidean space. Such constraints arise out of consideration of local characteristics of each node. We assume that the constraint sets for the local variables are private information for each node. We provide a distributed algorithm for the case when there is communication noise present in the network. We show that we can achieve almost sure convergence under certain assumptions. The second problem we discuss is the problem of distributed constrained optimization when the constraint sets are distributed over the agents. Furthermore our model incorporates the presence of noisy communication links and the presence of stochastic errors in the evaluation of subgradients of the local objective function. We establish sufficient conditions and provide an analysis guaranteeing the convergence of the algorithm to the optimal set with probability one.**

I. Introduction

There has been a sustained effort in the research community over the years to develop algorithms for distributed decision making and control. Recently two strands of this broad problem has garnered special attention. These are the problem of reaching consensus in a network of computational agents [4], [14], [15] and the problem of fair allocation of resources in a network. The algorithms for reaching consensus have proven useful in a wide variety of contexts from formation control [4], distributed parameter estimation [12], [5], load balancing [6], to synchronization of kuramoto oscillators [1]. The problem of fair allocation of resources has been thoroughly studied in the area of microeconomic [7]. Recent interest in the resource allocation problem has arisen in the context of utility maximization in communication networks [11]. 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 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 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 [8]. In this paper, we consider the case when there is noise present in the communication channel. In this case we extend the algorithm proposed in [8] by introducing a step size sequence that attenuates the communication noise. Next we consider a distributed constrained optimization problem. A distributed optimization algorithm for such a problem has been proposed in [8], 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 [8], by combining the presence of local constraint sets and noisy communication along with the presence of stochastic errors in the evaluation of subgradients. In this case we need to introduce two step size sequences to damp out communication noise and subgradient errors. 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. Consensus over noisy links in the lack of constraint sets has been studied in [13], [3] and [5] among others. In [10], 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.

The rest of the paper is organized as follows. In section II we discuss the problem of constrained consensus with noisy communication and present our algorithm for this problem. We describe the network model used in the paper and state various assumptions on the noise process. We provide an analysis demonstrating the convergent behavior of the algorithm almost surely. In section III we introduce the problem of distributed optimization with noisy communication and subgradient error. We provide an analysis proving almost sure convergence of our algorithm to the optimal set. Finally in section IV we provide the concluding remarks.

Notation. The j^{th} component of a vector x is denoted by x_j. For an $m \times m$ real valued matrix $A \in \mathbb{R}^{m\times n}$, we denote the element in the i^{th} row and j^{th} column of A by A_{ij}. We write I_r for the $r \times r$ dimensional identity matrix. Given a convex set $X \subset \mathbb{R}^n$, the projection operator $P_X[\cdot]$ maps a vector $x \in \mathbb{R}^n$ to the closest point in X under the Euclidean norm. Given a connected graph $G = (V, E)$, a path $p = (i_1, \dots, i_m)$ is a sequence of visits to the nodes of the graph G such that all the nodes are visited at least once and $(i_{k-1}, i_k) \in E$ for all k. Given any graph G, and a function $F : V \times V \to \mathbb{R}$, we use $\sum_G F(i,j)$ to denote the sum where the function $F(i,j)$ is evaluated for all (i,j) in the edge set of graph G. We assume that, when the graph G has bidirectional links, the sum $\sum_G F(i,j)$ is evaluated at every edge only once.

II. Constrained Consensus

We consider a setup where we are give a set of m agents, which can be viewed as the node set $V = \{1, \ldots, m\}$. We use terms node and agent interchangeably. At each time k, the communication pattern among the agents is represented by a time varying directed graph $G(k)$. At each instant, each node receives information from a subset of nodes and, also, broadcasts its information to a subset of nodes. The subset of nodes from which a node receives information at any instant are termed as the neighboring nodes of the node at that instant. Such information exchanges are characterized by the edge set $E(k)$ of the graph. For the scope of this paper we deal with synchronous algorithms. This implies that each agents local clocks are synchronized and time proceeds in discrete steps $k = 0, 1, \ldots$. Let $x_i(k) \in \mathbb{R}^n$ denote the local variable at node i at instance k. The local variables $x_i(k)$ are constrained to lie in convex sets $X_i \subseteq \mathbb{R}^n$ for all $i = 1, \ldots, m$. Each local constraint set is private and local information to node i. The objective of the network is to achieve consensus among the agents on their local decision variables i.e., $x_1 = \ldots = x_m$ only through information exchanges between neighboring agents. Clearly, since the local variables are constrained to lie in local constraint sets, the above problem is feasible only when the set $X = \cap_{j=1}^m X_j$ is nonempty.

Alternatively, the problem can be casted as a quest for a distributed algorithm for the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad \textstyle\sum_{i=1}^m \|x - P_{X_i}[x]\|^2 \\ &\text{subject to} \quad x \in \mathbb{R}^n. \end{aligned}$$

A solution to the above problem is given by any vector x which lies in the constraint set X. Clearly, in this case the objective function value is zero, which is also the optimal value. An algorithm is said to solve the problem if it generates agent estimates $x_i(k)$ that converge to a common value x^*, and x^* satisfies the constraint $x^* \in X$. A distributed algorithm for this problem was proposed in [8]. In the algorithm agent i local variables $x_i(k)$ evolves as follows:

$$x_i(k+1) = P_{X_i}\left[\sum_{j=1}^m a_{ij}(k+1)x_j(k)\right], \qquad (1)$$

where $a_{ij}(k+1)$ denotes the weight assigned by node i to the estimate coming from node j. A crucial assumption needed in the analysis in [8] was the requirement that if agent i receives data from agent j then $a_{ij}(k) \geq \eta > 0$. 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 $a_{ij}(k) \geq \eta$ since we need to asymptotically damp the impact of noise. To this effect we propose the following distributed algorithm for the problem:

$$x_i(k+1) = P_{X_i}\Bigg[x_i(k) - \alpha(k+1) \sum_{j \in N_i(k)} r_{ij}(k+1)[x_i(k) - (x_j(k) + \xi_{ij}(k+1))]\Bigg].$$

Here $\alpha(k+1) > 0$ is a step size, $r_{ij}(k+1)$ is a weighting parameter, $\xi_{ij}(k+1)$ is a random variable denoting the additive noise in communication, and $N_i(k)$ denotes the set of agents communicating with agent i at instance k. Let us define $r_{ii}(k+1) = -\sum_{j \in N_i(k)} r_{ij}(k+1)$, and $\xi_i(k+1) = \sum_{j \in N_i(k)} r_{ij}(k+1)\xi_{ij}(k+1)$. Then, the algorithm can be re-written as

$$x_i(k+1) = P_{X_i}\Bigg[x_i(k) + \alpha(k+1)\sum_{j=1}^m r_{ij}(k+1)x_j(k) + \alpha(k+1)\xi_i(k+1)\Bigg]. \qquad (2)$$

The matrix $R(k)$, where $R_{ij}(k) = r_{ij}(k)$, is thus a weighted graph Laplacian matrix and it satisfies $\sum_{j=1}^n r_{ij}(k) = 0$ for all i. We impose the following assumptions on the graph $G(t)$ and the weight sequence.

A. Model Assumptions

Assumption 1: 1) (*Bi-directional communication*) We assume that the communication is bi-directional; i.e if at any instant k, $r_{ij}(k) > 0$ then $r_{ji}(k) > 0$.
2) (*Symmetric weights*) The neighboring agents use symmetric weights, i.e., $r_{ij}(k) = r_{ji}(k)$.
3) (*Connectedness*) We assume that the graph $G(t)$ is connected at every instance, though it is free to be time varying.
4) We assume that if $r_{ij}(k) \neq 0$ at any instant, then it satisfies $\eta \leq r_{ij}(k) \leq \eta'$.

Let us assume that all the associated random processes are adapted to the filtration $\mathcal{F}_k$. We impose the following assumptions on the spatio-temporal noise process.

Assumption 2: 1) The process $\{\xi_{ij}(k)\}$ is a martingale difference sequence, i.e., $\mathbb{E}[\xi_{ij}(k+1)|\mathcal{F}_k] = 0$ for all i, j and $k \in \mathbb{N}$.
2) At any fixed instance k, the noise on link $e = (i,j)$ is independent of the noise on link $e' = (i', j')$ for $e \neq e'$.
3) The noise process is uniformly bounded in the mean square sense, i.e., there is a deterministic scalar $\mu_i > 0$ such that $\mathbb{E}[\|\xi_i(k+1)\|^2 \,|\mathcal{F}_k] \leq \mu_i^2$ for all $k \in \mathbb{N}$.

From assumption 2.3, it follows that $\mathbb{E}[\|\xi_i(k+1)\| \,|\mathcal{F}_k] \leq \sqrt{\mathbb{E}[\|\xi_i(k+1)\|^2 \,|\mathcal{F}_k]} \leq \mu_i$ for all $k \in \mathbb{N}$ and all i. Let us also define $\mu^2 = \sum_{j=1}^m \mu_j^2$. Furthermore we impose the following assumption on the step size sequence $\{\alpha(k)\}$.

Assumption 3: 1) $\alpha(k) > 0$, $\sum_{k=1}^\infty \alpha(k) = \infty$, and $\sum_{k=1}^\infty \alpha^2(k) < \infty$.

B. Preliminary Results

In this section we provide various results which will be useful in proving our main results. The following lemma from [8] gives some relations regarding the projection operator on a convex set.

Lemma 1: [8] Let X be a nonempty closed convex set in $\mathbb{R}^n$. Then, we have for any $x \in \mathbb{R}^n$,

1) $(P_X[x]-x)'(x-y) \leq -\|P_X[x]-x\|^2$ for all $y \in X$.
2) $\|P_X[x]-y\|^2 \leq \|x-y\|^2 - \|P_X[x]-x\|^2$ for $y \in X$.
3) $\|P_X[x]-P_X[y]\| \leq \|x-y\|$ for any $y \in \mathbb{R}^n$.

The following known theorem, which is a generalization of the supermartingale convergence theorem will be instrumental in proving our results.

Theorem 1: ([9], 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 $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ for $t \geq 0$. Suppose that:

1) The random variables Y_t, X_t, Z_t and $g(t)$ are nonnegative, and are adapted to the filtration $\mathcal{F}_t$.
2) For each t, we have almost surely
$$\mathbb{E}[Y_{t+1}|\mathcal{F}_t] \leq (1+g(t))Y_t - X_t + Z_t. \quad (3)$$
3) There holds $\sum_{t=0}^{\infty} Z_t < \infty$ and $\sum_{t=0}^{\infty} g(t) < \infty$ almost surely.

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

Now, we establish certain relations for algorithm (2). Let

$$\begin{aligned} v_i(k+1) &= x_i(k) + \alpha(k+1)\textstyle\sum_{j=1}^{m} r_{ij}(k+1)x_j(k) \\ &\quad +\alpha(k+1)\xi_i(k+1). \end{aligned}$$

Then, we have

$$x_i(k+1) = v_i(k+1) + e_i(k+1),$$

where the error term is given by

$$e_i(k+1) := P_{X_i}[v_i(k+1)] - v_i(k+1) = x_i(k+1) - v_i(k+1).$$

Lemma 2: When the set $X = \cap_{i=1}^{m} X_i$ is nonempty, we surely have for all $i \in V$, $k \geq 0$, and for all $x \in X$,

$$\|x_i(k+1)-x\|^2 \leq \|v_i(k+1)-x\|^2 - \|e_i(k+1)\|^2$$

Proof: By the definition of $x_i(k+1)$ we have

$$\|x_i(k+1)-x\|^2 = \|P_{X_i}[v_i(k+1)]-x\|^2.$$

Now, applying Lemma 1, we get along every sample path

$$\begin{aligned} \|P_{X_i}[v_i(k+1)]-x\|^2 &\leq \|v_i(k+1)-x\|^2 - \\ &\quad \|P_{X_i}[v_i(k+1)]-v_i(k+1)\|^2 \\ &= \|v_i(k+1)-x\|^2 - \|e_i(k+1)\|^2. \end{aligned}$$

■

Let us define $\mathbf{x}(k) = (x_1'(k), \ldots, x_m'(k))'$ as the joint state vector and, similarly, $\mathbf{e}(k+1) = (e_1'(k+1), \ldots, e_m'(k+1))'$, and $\xi(k+1) = (\xi_1'(k+1), \ldots, \xi_m'(k+1))$. Define the $mn \times mn$ matrix $\mathbf{R}(k+1) = R(k+1) \otimes I_n$. The joint state space representation of algorithm (2) can be given as:

$$\begin{aligned} \mathbf{x}(k+1) &= [I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]\mathbf{x}(k) \\ &\quad + \alpha(k+1)\xi(k+1) + \mathbf{e}(k+1). \end{aligned}$$

Before we state the next lemma for notational convenience let us define $H(k+1) = 1 + \alpha^2(k+1)\sigma_m^2(R(k+1))$. Here $\sigma_m^2(R(k+1))$ denotes the square of the maximum singular value of the matrix $R(k+1)$. For any vector $z \in X$ let us denote $\mathbf{z} = z \otimes 1_m$, where 1_m is the m-dimensional vector of ones. Also, let us denote the conditional expectation operator $\mathbb{E}_k[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_k]$. Since we assume that the communication graph is connected at every instant (cf. Assumption 1.3), there exists a spanning tree $S(k)$ such that the edge (i,j) belongs to the tree if and only if $r_{ij}(k+1) > \eta$.

Lemma 3: Let Assumptions 1 and 2 hold. Also, assume that the set $X = \cap_{j=1}^{m} X_j$ is nonempty. Then, the following relation holds for any $z \in X$:

$$\begin{aligned} \mathbb{E}_k[\|\mathbf{x}(k+1)-\mathbf{z}\|^2] &\leq H(k+1)\|\mathbf{x}(k)-\mathbf{z}\|^2 + \alpha^2(k+1)\mu^2 \\ &\quad - 2\eta\alpha(k+1)\sum_{S(k)}\|x_i(k)-x_j(k)\|^2. \end{aligned}$$

Proof: Note that from Lemma 2 we have

$$\begin{aligned} \mathbb{E}_k[\|\mathbf{x}(k+1)-\mathbf{z}\|^2] &\leq \mathbb{E}_k[\|\mathbf{v}(k+1)-\mathbf{z}\|^2] \\ &\quad - \mathbb{E}_k[\|\mathbf{e}(k+1)\|^2], \quad (4) \end{aligned}$$

where

$$\mathbf{v}(k+1) = [I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]\mathbf{x}(k) + \alpha(k+1)\xi(k+1).$$

Hence,

$$\begin{aligned} \mathbb{E}_k[\|\mathbf{v}(k+1)-\mathbf{z}\|^2] &= \|[I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]\mathbf{x}(k) - \mathbf{z}\|^2 \\ &\quad + \alpha^2(k+1)\mathbb{E}_k[\|\xi(k+1)\|^2] + 2\alpha(k+1) \\ &\quad \times ([I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]\mathbf{x}(k) - \mathbf{z})'\mathbb{E}_k[\|\xi(k+1)\|]. \end{aligned}$$

By our Assumption 2.1 $\mathbb{E}_k[\xi(k+1)] = 0$, so that

$$\begin{aligned} \mathbb{E}_k[\|\mathbf{v}(k+1)-\mathbf{z}\|^2] &= \|[I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]\mathbf{x}(k) - \mathbf{z}\|^2 \\ &\quad + \alpha^2(k+1)\mathbb{E}_k[\|\xi(k+1)\|^2]. \end{aligned}$$

Since $\mathbf{R}(k+1)$ is symmetric and has zero-row sums, we have

$$\begin{aligned} &\|[I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]\mathbf{x}(k) - \mathbf{z}\|^2 = \\ &\quad [\mathbf{x}(k)-\mathbf{z}]'[I_{mn} + \alpha(k+1)\mathbf{R}(k+1)]' \\ &\quad [I_{mn} + \alpha(k+1)\mathbf{R}(k+1)][\mathbf{x}(k)-\mathbf{z}] \\ &= [\mathbf{x}(k)-\mathbf{z}]'[I_{mn} + \alpha^2(k+1)\mathbf{R}'(k+1)\mathbf{R}(k+1)][\mathbf{x}(k)-\mathbf{z}] \\ &\quad + 2\alpha(k+1)[\mathbf{x}(k)-\mathbf{z}]'\mathbf{R}(k+1)[\mathbf{x}(k)-\mathbf{z}]. \quad (5) \end{aligned}$$

We also have

$$\begin{aligned} &[\mathbf{x}(k)-\mathbf{z}]'\mathbf{R}(k+1)'\mathbf{R}(k+1)[\mathbf{x}(k)-\mathbf{z}] \leq \\ &\quad \sigma_m^2(\mathbf{R}(k+1))\|\mathbf{x}(k)-\mathbf{z}\|^2, \quad (6) \end{aligned}$$

where $\sigma_m(\mathbf{R}(k+1))$ is the maximum singular value of the matrix $\mathbf{R}(k+1)$. Since $\mathbf{R}(k+1) = R(k+1) \otimes I_n$, we have $\sigma_m(\mathbf{R}(k+1)) = \sigma_m(R(k+1))$. Hence,

$$[\mathbf{x}(k) - \mathbf{z}]'[I_{mn} + \alpha^2(k+1)\mathbf{R}'(k+1)\mathbf{R}(k+1)][\mathbf{x}(k) - \mathbf{z}] \leq H(k+1)\|\mathbf{x}(k) - \mathbf{z}\|^2,$$

with $H(k+1) = 1 + \alpha^2(k+1)\sigma_m^2(R(k+1))$.

Now, consider the term $[\mathbf{x}(k)-\mathbf{z}]'\mathbf{R}(k+1)[\mathbf{x}(k)-\mathbf{z}]$ in (5). Since $\mathbf{R}(k+1)$ has row sum zero and $\mathbf{R}(k+1)$ is symmetric, we have $\mathbf{R}(k+1)\mathbf{z} = 0$ and $\mathbf{z}'\mathbf{R}(k+1) = 0$. Thus,

$$[\mathbf{x}(k) - \mathbf{z}]'\mathbf{R}(k+1)[\mathbf{x}(k) - \mathbf{z}] = \mathbf{x}'(k)\mathbf{R}(k+1)\mathbf{x}(k).$$

By definition $\mathbf{R}(k+1) = R(k+1) \otimes I_m$, where $R_{ij}(k+1) = r_{ij}(k+1)$ and $R_{ii}(k+1) = -\sum_{j=1, j\neq i}^m r_{ij}(k+1)$. Thus, we see that

$$\mathbf{x}'(k)\mathbf{R}(k+1)\mathbf{x}(k) = -\sum_{i=1}^m x_i'(k) \sum_{j=1, j\neq i}^m r_{ij}(k+1)(x_i(k) - x_j(k)).$$

Since, the matrix $R(k+1)$ is symmetric we have

$$\begin{aligned} -\sum_{i=1}^m x_i'(k) &\sum_{j=1, j\neq i}^m r_{ij}(k+1)(x_i(k) - x_j(k)) \\ &= -\sum_{i<j} r_{ij}(k+1)\|x_i(k) - x_j(k)\|^2 \\ &\leq -\eta \sum_{S(k)} \|x_i(k) - x_j(k)\|^2. \end{aligned}$$

Using the preceding and the bound $\mathbb{E}_k[\|\xi(k+1)\|^2] \leq \mu^2$ we arrive at

$$\mathbb{E}_k[\|\mathbf{v}(k+1) - z\|^2] \leq H(k+1)\|\mathbf{x}(k) - \mathbf{z}\|^2 + \alpha^2(k+1)\mu^2 - 2\alpha(k+1)\eta \sum_{S(k)} \|x_i(k) - x_j(k)\|^2$$

By substituting back in Eq. (4) we get

$$\mathbb{E}_k[\|\mathbf{x}(k+1) - \mathbf{z}\|^2] \leq H(k+1)\|\mathbf{x}(k) - \mathbf{z}\|^2 + \alpha^2(k+1)\mu^2 - 2\alpha(k+1)\eta \sum_{S(k)} \|x_i(k) - x_j(k)\|^2 - \mathbb{E}_k[\|e(k+1)\|^2].$$

Neglecting the error term we have the desired result. ■

C. Almost sure convergence

In this section we state out main convergence result for the constrained consensus problem.

Theorem 2: Assume that $X = \cap_{j=1}^m X_j$ is nonempty, and let Assumptions 1 and 2 hold. Let the step size sequence $\{\alpha(k)\}$ in algorithm (2) satisfy Assumption 3. Then, there exists a random variable x^* taking values in the set X such that almost surely $\lim_{k\to\infty} \|x_i(k) - x^*\| = 0$ for all agents i.

Proof: First, we consider the term $H(k+1) = 1 + \alpha^2(k+1)\sigma_m^2(R(k+1))$ in Lemma 3. The entries of the matrix $R(k+1)$ are uniformly bounded, implying $\sigma_m^2(R(k+1)) \leq C$ for some scalar C and all k. By our assumption on the step size, we have $\sum_{k=1}^\infty \alpha^2(k) < \infty$. Since all the terms appearing in Lemma 3 are non-negative, we can apply the result of Robbins-Siegmund (Theorem 1) to deduce that, with probability one, $\|\mathbf{x}(k) - \mathbf{z}\|^2$ converges for any $z \in X$ and

$$\sum_{k=1}^\infty \alpha(k+1) \sum_{S(k)} \|x_i(k) - x_j(k)\|^2 < \infty. \tag{7}$$

By $\sum_{k=1}^\infty \alpha(k) = \infty$, relation (7) implies that there is a subsequence such that $\lim_{k\to\infty} \sum_{S(n_k)} \|x_i(n_k) - x_j(n_k)||^2 = 0$. Now, since the number of spanning trees on a finite graph is finite, there exists a spanning tree S which appears infinitely often in the sequence $\{S(n_k)\}$. Let us pick a further subsequence such that $S(n_k^1) = S$, then we have along this subsequence $\lim_{k\to\infty} \sum_S \|x_i(n_k^1) - x_j(n_k^1)\|^2 = 0$ for all i and j. The spanning tree S is in the connected graph, so the preceding relation yields $\lim_{k\to\infty} \|x_i(n_k^1) - x_j(n_k^1)\|^2 = 0$ for all i, j.

Now since $\|\mathbf{x}(k) - \mathbf{z}\|^2$ converges almost surely for any $z \in X$, the subsequence $\{\mathbf{x}(n_k^1)\}$ is bounded almost surely. Again, we can extract a convergent subsequence $\mathbf{x}(n_k^2)$ such that $\lim_{k\to\infty} \|x_i(n_k^2) - x_i^*\| = 0$ almost surely for some random vector x_i^* for all i. Since $\lim_{k\to\infty} \|x_i(n_k^2) - x_j(n_k^2)\| = 0$ almost surely for all i, j, it follows that $x_i^* = x_j^* = x^*$ almost surely for all i, j. The sets X_i are closed, so that $x^* \in X_i$ almost surely for all i, which in turn implies that $x^* \in X$ almost surely. Therefore, $\lim_{k\to\infty} \|\mathbf{x}(n_k^2) - \mathbf{x}^*\|^2 = 0$ almost surely. But, we know that $\lim_{k\to\infty} \|\mathbf{x}(k) - \mathbf{z}\|^2$ converges almost surely for all $z \in X$. Hence, by looking at the sample paths, we can conclude that the limit of any subsequence is also the sequential limit, implying that $\lim_{k\to\infty} \|\mathbf{x}(k) - \mathbf{x}^*\| = 0$ almost surely. ■

III. Distributed Optimization

We next consider solving 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:

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

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 a compact and convex set, representing the local constraint set of agent i. Since the objective function is continuous and the set X is compact, we know that by Weirstrass theorem 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. The proposed projected subgradient algorithm is given by:

$$x_i(k+1) = P_{X_i}\left[x_i(k) - \alpha(k+1)\sum_{j\in\mathcal{N}_i(k)} r_{ij}(k+1)x_i(k)\right.$$
$$\left. -(x_j(k)+\xi_{ij}(k+1))] - \gamma(k+1)[d_i(k+1)+\epsilon_i(k+1)]\right].$$

Here, the vector $d_i(k)$ is a subgradient of the local objective function f_i and $\epsilon_i(k+1)$ is the error in the evaluation of the subgradient of $f_i(x)$ at $x = v_i(k+1)$, where $v_i(k+1)$ is given by (4). The step size $\gamma(k+1) > 0$ is used to attenuate the subgradient error. Proceeding similarly to the consensus part let us re-write this as follows.

$$x_i(k+1) = P_{X_i}\left[x_i(k) + \alpha(k+1)\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)\right.$$
$$\left. + \alpha(k+1)\xi_i(k+1) - \gamma(k+1)[d_i(k+1)+\epsilon_i(k+1)]\right]. \tag{8}$$

We denote the noisy subgradient by $\tilde{d}_i(k) = d_i(k)+\epsilon_i(k)$. Recalling the definition of $v_i(k+1)$ of Eq. (4), the algorithm can be re-written as:

$$x_i(k+1) = v_i(k+1) - \gamma(k+1)\tilde{d}_i(k+1) + e_i(k+1), \tag{9}$$

where $e_i(k+1) = P_{X_i}\left[v_i(k+1) - \gamma(k+1)\tilde{d}_i(k+1)\right] - v_i(k+1) + \gamma(k+1)\tilde{d}_i(k+1)$.

A. Assumptions

Let $\partial f_i(x)$ denote the set of subgradients of the objective function $f_i(x)$. We impose the following assumptions on the subgradients and the constraint sets.

Assumption 4:
1) The subgradient errors $\epsilon_i(k)$ when conditioned on the point of evaluation of the subgradient $d_i(k)$ are mean zero, i.e., $\mathbb{E}[\epsilon_i(k)|v_i(k)] = 0$ for all i and $k > 0$.
2) The subgradient errors further satisfy the bound $\mathbb{E}[\|\epsilon_i(k)\|^2 |v_i(k)] \le c$ for all i and $k > 0$.
3) The local constraint sets X_i are compact and convex.
4) The intersection set X has a nonempty interior, i.e., there exists a point $\bar{z} \in X$ such that the ball $B_\delta := \{x \in X : \|x - \bar{z}\| \le \delta\} \subset X$.

In addition to the above assumptions on the model, we use the following assumptions on the step sizes $\alpha(k)$ and $\gamma(k)$.

Assumption 5:
1) (non-summability) The step sizes satisfy $\sum_{k=1}^{\infty}\alpha(k) = \infty$ and $\sum_{k=1}^{\infty}\gamma(k) = \infty$.
2) (square summability) $\sum_{k=1}^{\infty}\alpha^2(k) < \infty$ and $\sum_{k=1}^{\infty}\gamma^2(k) < \infty$.
3) $\sum_{k=1}^{\infty}\alpha(k)\gamma(k) < \infty$ and $\sum_{k=1}^{\infty}\frac{\gamma^2(k)}{\alpha(k)} < \infty$.
4) $\sum_{k=1}^{\infty}\min(\alpha(k),\gamma(k)) = \infty$.

The assumptions 5.1 and 5.2 are standard in the stochastic approximation literature. The square summability is needed to damp out the noise terms. In addition to these conditions, our analysis relies on assumptions 5.3 and 5.4 on the cross terms involving the step sizes. To verify that the set of step sizes satisfying these conditions is non empty, we can assume that the step sizes are of the form $\alpha(k) = \frac{1}{k^{\theta_1}}$ and $\gamma(k) = \frac{1}{k^{\theta_2}}$. Then conditions 5.1 and 5.2 imply that $1/2 < \theta_1, \theta_2 \le 1$. It is clear that in this case $\sum_{k=1}^{\infty}\alpha(k)\gamma(k) < \infty$. The condition $\sum_{k=1}^{\infty}\frac{\gamma^2(k)}{\alpha(k)} < \infty$ implies that $\theta_2 > \frac{1+\theta_1}{2}$. Also, since when $\theta_2 > \theta_1$, we have $\min(\alpha(k),\gamma(k)) = \gamma(k)$; Assumption 5.4 holds by Assumption 5.1.

B. Origin of subgradient errors

The consideration of subgradient error in our model enables us to include cases where the function $f_i(x)$ is given as $\mathbb{E}[g_i(x, Z_i)]$. Here Z_i is a random variable and the expectation is taken with respect to the unknown distribution of the random variable. Thus the function $f_i(x)$ is not known to agent i. The agent i however has access to samples of Z_i. Let $\nabla_x f(x)$ denote a subgradient of the objective function $f(x)$ then under some broad assumptions we have $\nabla_x f_i(x) = \mathbb{E}[\nabla_x g_i(x, Z_i)]$. Hence, we can write $\nabla_x f_i(x) = \nabla_x g_i(x,\omega) + [\mathbb{E}[\nabla_x g_i(x,\omega)] - \nabla_x g_i(x,\omega)]$. Denoting $\epsilon_i := \mathbb{E}[\nabla_x g_i(x,\omega)] - \nabla_x g_i(x,\omega)$, we can see that it is a martingale difference sequence. Thus, we can see that the stochastic optimization problem fits in our framework.

C. Preliminary Results

The subgradient sets of each f_i are bounded over the sets X_i, i.e., there is a scalar $L > 0$ such that for all i, $\|d\| \le L$ for all $d \in \partial f_i(x)$ and all $x \in X_i$.

In this section we provide some results which will be useful in deriving our main result. The first result provides a way to bound an error term of the form $\|x_i(k) - P_{X_i}[x_i(k)]\|$. The bound is established by using some of the techniques in [2] for the alternating projection method.

Lemma 4: Let Assumptions 4.3 and 4.4 hold, and let $x_i \in X_i$ be variables belonging to local constraint sets X_i. Then, we have the following bound:

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

where B is a uniform upper bound on the norms of the vectors in the sets X_i and δ is the radius from Assumption 4.4.

Proof: Let i 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 4.4. Then, we can write

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

From definition of λ_i, it is clear that $\left\|x_i - P_{X_j}[x_i]\right\| \le \lambda_i$ 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

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

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

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

■

Let us define $s(k) = \frac{1}{m}\sum_{j=1}^{m} P_X[x_j(k)]$, which belongs to the set X since X is convex. The following lemma is crucial in proving our convergence result in the next section.

Lemma 5: Let Assumptions 1, 2 and 4 hold. Then, for the algorithm proposed in Eq. (8) the following relation holds for any z^* in the optimal set X^*,

$$\begin{aligned}\mathbb{E}_k[\|\mathbf{x}(k) - \mathbf{z}^*\|^2] &\le H(k+1)\|\mathbf{x}(k) - \mathbf{z}^*\|^2 + \alpha^2(k+1)\mu^2 \\ &- \alpha(k+1)\eta \sum_{S(k)} \|x_i(k) - x_j(k)\|^2 + m(c + L^2)\gamma^2(k+1) \\ &+ \frac{K^2\gamma^2(k+1)}{\eta\alpha(k+1)} - 2\gamma(k+1)[f(s(k)) - f(z^*)] \\ &+ 2L\gamma(k+1)\alpha(k+1)\sum_{i=1}^{m}\left\|\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)\right\|,\end{aligned}$$

where L is a uniform bound on subgradient norms of f_i over the sets X_i, $K = L(m-1)(\frac{mB+\delta}{\delta})$ and $f(x) = \sum_{i=1}^{m} f_i(x)$.

Proof: By definition we have $x_i(k+1) = P_{X_i}[v_i(k+1) - \gamma(k+1)\tilde{d}_i(k+1)]$. From Lemma 1.3 and the definition of $e_i(k+1)$ in (9), we see that for any $z^* \in X^* \subseteq X$,

$$\begin{aligned}&\|x_i(k+1) - z^*\|^2 \\ &\le \left\|v_i(k+1) - z^* - \gamma(k+1)\tilde{d}_i(k+1)\right\|^2 \\ &= \|v_i(k+1) - z^*\|^2 + \gamma^2(k+1)\left\|\tilde{d}_i(k+1)\right\|^2 \\ &-2\gamma(k+1)\tilde{d}_i'(k+1)(v_i(k+1) - z^*).\end{aligned}$$

Taking conditional expectation, we obtain for any $z^* \in X^*$,

$$\begin{aligned}\mathbb{E}_k[\|x_i(k+1) - z^*\|^2] &\le \mathbb{E}_k[\|v_i(k+1) - z^*\|^2] \\ &+ 2\gamma^2(k+1)\mathbb{E}_k[\|d_i(k+1)\|^2 + \|\epsilon_i(k+1)\|^2] \\ &- 2\gamma(k+1)\mathbb{E}_k[d_i'(k+1)(v_i(k+1) - z^*)] \\ &- 2\gamma(k+1)\mathbb{E}_k[\epsilon_i'(k+1)(v_i(k+1) - z^*)].\end{aligned}$$

Since, $\mathbb{E}_k[\epsilon_i'(k+1)(v_i(k+1) - z^*)] = \mathbb{E}_k[(v_i(k+1) - z^*)'\mathbb{E}[\epsilon_i(k+1)|v_i(k+1)]]$ and $\mathbb{E}[\epsilon_i(k+1)|v_i(k+1)] = 0$ and $d_i(k+1)$ is a subgradient of f_i at $v_i(k+1)$, we have

$$\begin{aligned}&\mathbb{E}_k[\|x_i(k+1) - z^*\|^2] \le \mathbb{E}_k[\|v_i(k+1) - z^*\|^2] \\ &+2\gamma^2(k+1)(c+L^2) - 2\gamma(k+1)\mathbb{E}_k[f_i(v_i(k+1)) - f_i(z^*)],\end{aligned}$$

where we have used $\|d_i(k)\| \le L$ and $\mathbb{E}_k[\|\epsilon_i(k+1)\|^2] \le c$. Now, since f_i is a convex function, by Jensen's inequality $-\mathbb{E}_k[f_i(v_i(k+1))] \le -f(\mathbb{E}_k[v_i(k+1)])$. By the definition of $v_i(k+1)$ in (4) and $\mathbb{E}_k[\xi_{ij}(k)] = 0$ of Assumption 2.1, we have $\mathbb{E}_k[v_i(k+1)] = x_i(k) + \alpha(k+1)\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)$. Letting $w_i(k+1) := x_i(k) + \alpha(k+1)\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)$, we obtain

$$\begin{aligned}\mathbb{E}_k[\|x_i(k+1) - z^*\|^2] &\le \mathbb{E}_k[\|v_i(k+1) - z^*\|^2] \\ &+ 2\gamma^2(k+1)(c + L^2) \\ &- 2\gamma(k+1)[f_i(w_i(k+1)) - f_i(z^*)].\end{aligned}$$

Summing over all i and using vector notation yield

$$\begin{aligned}\mathbb{E}_k[\|\mathbf{x}(k+1) - \mathbf{z}^*\|^2] &\le \mathbb{E}_k[\|\mathbf{v}(k+1) - z^*\|^2] \\ &+ 2m\gamma^2(k+1)(c + L^2) \\ &- 2\gamma(k+1)\sum_{i=1}^{m}[f_i(w_i(k+1)) - f_i(z^*)].\end{aligned}$$

As seen in the proof of Lemma 3, we have

$$\begin{aligned}\mathbb{E}_k[\|\mathbf{v}(k+1) - \mathbf{z}^*\|^2] &\le H(k+1)\|\mathbf{x}(k) - \mathbf{z}^*\|^2 \\ &+ \alpha^2(k+1)\mu^2 - 2\alpha(k+1)\eta\sum_{S(k)}\|x_i(k) - x_j(k)\|^2 .\end{aligned}$$

Let $s(k) \in X$ be as defined earlier. Adding and subtracting $f_i(s(k))$, and using $f(x) = \sum_{i=1}^{m} f(x)$ and the above two relations, we obtain

$$\begin{aligned}\mathbb{E}_k[\|\mathbf{x}(k+1) - \mathbf{z}^*\|^2] &\le H(k+1)\|\mathbf{x}(k) - \mathbf{z}^*\|^2 \\ &+ m\gamma^2(k+1)(c + L^2) - 2\gamma(k+1)[f(s(k)) - f(z^*)] \\ &- 2\gamma(k+1)\sum_{i=1}^{m}[f_i(w_i(k+1)) - f_i(s(k))] + \alpha^2(k+1)\mu^2 \\ &- 2\alpha(k+1)\eta\sum_{S(k)}\|x_i(k) - x_j(k)\|^2 .\end{aligned}$$

By the convexity and subgradient boundedness of each f_i, $|f_i(w_i(k+1)) - f_i(s(k))| \le L\|w_i(k+1) - s(k)\|$, implying

$$\begin{aligned}\mathbb{E}_k[\|\mathbf{x}(k+1) - \mathbf{z}^*\|^2] &\le H(k+1)\|\mathbf{x}(k) - \mathbf{z}^*\|^2 \\ &+ m\gamma^2(k+1)(c + L^2) - 2\gamma(k+1)[f(s(k)) - f(z^*)] \\ &+ 2\gamma(k+1)L\sum_{i=1}^{m}\|w_i(k+1) - s(k)\| + \alpha^2(k+1)\mu^2 \\ &- 2\alpha(k+1)\eta\sum_{S(k)}\|x_i(k) - x_j(k)\|^2 . \quad (10)\end{aligned}$$

We now estimate the term with $\|w_i(k+1) - s(k)\|$. Since $w_i(k+1) := x_i(k) + \alpha(k+1)\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)$, it follows that

$$\begin{aligned}&2L\gamma(k+1)\sum_{i=1}^{m}\|w_i(k+1) - s(k)\| \\ &\le 2L\gamma(k+1)\sum_{i=1}^{m}\|x_i(k) - s(k)\| \\ &+ 2L\gamma(k+1)\alpha(k+1)\sum_{i=1}^{m}\left\|\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)\right\|. \quad (11)\end{aligned}$$

Next, we focus on the term $\sum_{i=1}^{m} \|x_i(k) - s(k)\|$. Substituting for $s(k) = \frac{1}{m}\sum_{j=1}^{m} P_X[x_j(k)]$, adding and subtracting the term $P_X[x_i(k)]$ inside the norm and using the convexity of norm, we have

$$\sum_{i=1}^{m} \|x_i(k) - s(k)\| = \sum_{i=1}^{m} \left\| x_i(k) - \frac{1}{m}\sum_{j=1}^{m} P_X[x_j(k)] \right\| \leq \sum_{i=1}^{m} \|x_i(k) - P_X[x_i(k)]\| + \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m} \|P_X[x_i(k)] - P_X[x_j(k)]\|.$$

By Lemma 4 we have that

$$\|x_i(k) - P_X[x_i(k)]\| \leq \frac{B}{\delta}\sum_{j=1}^{m} \|x_i(k) - x_j(k)\|,$$

and by the non-expansiveness property of projection $\|P_X[x_i(k)] - P_X[x_j(k)]\| \leq \|x_i(k) - x_j(k)\|$. Hence,

$$\sum_{i=1}^{m} \|x_i(k) - s(k)\| \leq \frac{mB+\delta}{m\delta}\sum_{i=1}^{m}\sum_{j=1}^{m} \|x_i(k) - x_j(k)\| = 2\frac{mB+\delta}{m\delta}\sum_{i<j} \|x_i(k) - x_j(k)\|. \quad (12)$$

For any two nodes i and j, there is a path from node i to node j in the spanning tree $S(k)$. By suppressing the dependence on time, we represent such a path by $p_{ij} = (i_1, \ldots, i_\ell)$, where $i_1 = i$, $i_\ell = j$ and $\{i_\tau, i_{\tau+1}\}$ is edge in the spanning tree for $1 \leq \tau \leq \ell - 1$. Therefore, we have $\|x_i(k) - x_j(k)\| \leq \sum_{\tau=1}^{\ell-1} \|x_{i_{\tau+1}(k)}(k) - x_{i_\tau(k)}(k)\|$. Let all the (undirected) edges in the tree $S(k)$ be enumerated by $e_s = \{i_s, i_{s+1}\}$ for $1 \leq s \leq m-1$. By summing over i and j with $i < j$ we get $\sum_{i<j} \|x_i(k) - x_j(k)\| \leq \sum_{e_s \in S(k)} \kappa_i \left\|x_{i_s}(k) - x_{i_{s+1}}(k)\right\|$, where κ_i denotes the number of times the edge e_s appears in the collection of the paths $\{p_{ij}, i < j\}$. A simple upper bound on the number κ_i is given by $\binom{m}{2}$. Thus,

$$\sum_{i<j} \|x_i(k) - x_j(k)\| \leq \binom{m}{2} \sum_{S(k)} \|x_i(k) - x_j(k)\|,$$

where the summation on the right hand side is over all edges $\{i, j\}$ in $S(k)$. Using the preceding estimate in relation (12), we obtain

$$\sum_{i=1}^{m} \|x_i(k) - s(k)\| \leq \frac{mB+\delta}{\delta}(m-1)\sum_{S(k)} \|x_i(k) - x_j(k)\|$$

Multiplying the preceding relation by $2\gamma(k+1)L$ and letting $K = L(m-1)\frac{mB+\delta}{\delta}$, we have

$$2\gamma(k+1)L\sum_{i=1}^{m} \|x_i(k) - s(k)\| \leq 2\gamma(k+1)K\sum_{S(k)} \|x_i(k) - x_j(k)\|. \quad (13)$$

By using the identity $2ab \leq a^2 + b^2$, we obtain

$$2\gamma(k+1)K\sum_{S(k)} \|x_i(k) - x_j(k)\| = \frac{2\gamma(k+1)K}{\sqrt{\eta}\sqrt{\alpha(k+1)}}\sum_{S(k)} \sqrt{\eta\alpha(k+1)}\,\|x_i(k) - x_j(k)\| \leq \frac{\gamma^2(k+1)K^2}{\eta\alpha(k+1)} + \eta\alpha(k+1)\sum_{S(k)} \|x_i(k) - x_j(k)\|^2.$$

Substituting the preceding inequality in relation (13), and then combining the resulting relation with Eq. (11), we obtain the following estimate:

$$2\gamma(k+1)L\sum_{i=1}^{m} \|w_i(k+1) - s(k)\| \leq \frac{\gamma^2(k+1)K^2}{\eta\alpha(k+1)} + \eta\alpha(k+1)\sum_{S(k)} \|x_i(k) - x_j(k)\|^2 + 2L\gamma(k+1)\alpha(k+1)\sum_{i=1}^{m}\left\|\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)\right\|$$

Substituting the preceding relation in Eq. (10), we arrive at the desired result. ■

D. Almost sure convergence

In this section we prove our main convergence result for the projected subgradient algorithm.

Theorem 3: Let Assumptions 1, 2 and 4 hold. Then if the step size sequences $\{\alpha(k)\}$ and $\{\gamma(k)\}$ in the algorithm (8) are chosen to satisfy the conditions in Assumption 5, then almost surely the iterate sequences $\{x_i(k)\}$ converge to a common (random) point in the optimal set X^*.

Proof: Referring to Lemma 5, if the step size sequences are chosen to satisfy Assumption 5.2 we have

$$\sum_{k=1}^{\infty} \left[\alpha^2(k+1)\mu^2 + m(c+L^2)\gamma^2(k+1)\right] < \infty. \quad (14)$$

By Assumption 5.3 we have $\sum_{k=1}^{\infty} \frac{K^2\gamma^2(k+1)}{\eta\alpha(k+1)} < \infty$. The term $\sum_{i=1}^{m}\left\|\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)\right\|$ is bounded since the sets X_i are compact by Assumption 4.4 and $r_{ij}(k+1) \leq \eta'$ by Assumption 1.4. Also, we have $\sum_{k=1}^{\infty} \alpha(k)\gamma(k) < \infty$ by Assumption 5.3, so we can conclude that

$$2L\sum_{k=1}^{\infty} \gamma(k+1)\alpha(k+1)\sum_{i=1}^{m}\left\|\sum_{j=1}^{m} r_{ij}(k+1)x_j(k)\right\| < \infty. \quad (15)$$

Now, since $H(k+1) = (1 + \alpha^2(k+1)\sigma_m^2(R(k+1)))$ and the elements of matrix $R(k)$ are uniformly bounded, we have $\sigma_m^2(R(k+1)) \leq C$. Hence $\sum_{k=1}^{\infty} \alpha^2(k+1)\sigma_m^2(R(k+1))) \leq C\sum_{k=1}^{\infty} \alpha^2(k+1) < \infty$. Also since z^* belongs to the optimal set X^* and $s(k) \in X$, we have $f(s(k)) - f(z^*) \geq 0$ for all k. Thus, we can apply the result of Robbins-Siegmund (Lemma 1) to conclude that $\|\mathbf{x}(k+1) - \mathbf{z}^*\|^2$ converges almost surely for any $z^* \in X^*$ and the following holds almost surely:

$$\sum_{k=1}^{\infty}\left[\eta\alpha(k+1)\sum_{S(k)}\|x_i(k)-x_j(k)\|^2 + \gamma(k+1)(f(s(k))-f(z^*))\right] < \infty.$$

Letting $\theta(k)=\min(\alpha(k),\gamma(k))$, we can conclude that

$$\sum_{k=1}^{\infty}\theta(k+1)\left[\eta\sum_{S(k)}\|x_i(k)-x_j(k)\|^2 + (f(s(k))-f(z^*))\right] < \infty,$$

with $\sum_{k=1}^{\infty}\theta(k)=\infty$ (Assumption 5.4). Therefore, there is a subsequence such that $\lim_{k\to\infty}\sum_{S(n_k)}\|x_i(n_k)-x_j(n_k)\|^2=0$ and $\lim_{k\to\infty}f(s(n_k))=f(z^*)$ almost surely.

Since the number of spanning trees on a finite graph is finite, there exists a spanning tree S which appears infinitely often in the sequence $\{S(n_k)\}$. Let us pick a further subsequence such that $S(n_k^1)=S$, then we have along this subsequence $\lim_{k\to\infty}\sum_S\left\|x_i(n_k^1)-x_j(n_k^1)\right\|^2=0$ and $\lim_{k\to\infty}f(s(n_k^1)) = f(z^*)$ almost surely. Since S is a spanning tree of a connected graph we can conclude that $\lim_{k\to\infty}\left\|x_i(n_k^1)-x_j(n_k^1)\right\|=0$ for all nodes i and j almost surely. Using the earlier proven bound in equation (12) and taking limit along the time subsequence $\{n_k^1\}$, we obtain

$$\lim_{k\to\infty}\sum_{i=1}^{m}\left\|x_i(n_k^1)-s(n_k^1)]\right\| \leq \lim_{k\to\infty}2\left(\frac{mB+\delta}{m\delta}\right)\sum_{i<j}\left\|x_i(n_k^1)-x_j(n_k^1)\right\|=0,$$

implying that we have almost surely

$$\lim_{k\to\infty}\left\|x_i(n_k^1)-s(n_k^1)\right\|=0 \qquad \text{for all nodes } i. \tag{16}$$

Since $\lim_{k\to\infty}f(s(n_k^1))=f^*$ and the function f is continuous we conclude that there exists a subsequence along which $s(n_k^1)$ converges almost surely to a (random) point x^* that lies in the set X^*. Without any loss of generality we can assume that the sequence $s(n_k^1)$ itself converges to the limit point x^* almost surely. From equation (16) we conclude that $x_i(n_k^1)$ converges to x^* for all i almost surely. However from our conclusion earlier we know that $\|\mathbf{x}(k)-\mathbf{z}^*\|^2$ converges almost surely for any $z^*\in X^*$. We can consider sample paths for which both $\|\mathbf{x}(k)-\mathbf{z}^*\|^2$ converges for any $z^*\in X^*$ and $x_i(n_k^1)$ converges to the corresponding realization $\tilde{x}^*\in X^*$ of the random point x^*. Then, by letting $z^*=x^*$, we can conclude that for each such realization, the realization of the sequence $\{x_i(k)\}$ converges to the corresponding realization $\tilde{x}^*$ of the random point $x^*\in X^*$. Hence, the sequences $\{x_i(k)\}$,$i=1,\ldots,m$, converge almost surely to a common (random) point in the set X^*. ■

IV. Conclusion

In this paper we consider distributed algorithms for the constrained consensus and constrained optimization problem in the presence of noisy communication links and stochastic subgradient errors. We provide the analysis showing almost sure convergence of the algorithms in both of these cases. In our analysis we assumed that the communication graph is connected at every instant. However, our analysis can be generalized to the case when the union of communication graphs over a finite time interval is connected.

References

[1] N. Chopra and M.W. Spong, *On exponential synchronization of kuramoto oscillators*, **54** (2009), 353–357.

[2] 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 **7** (1967), no. 6, 1211–1228.

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

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

[5] S. Kar, J.M.F. Moura, and K. Ramanan, *Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication*, (2008), (submitted).

[6] A. Kashyap, T. Basar, and R. Srikant, *Quantized consensus*, Automatica **43** (2007), 1192–1203.

[7] D.G. Luenberger, *Microeconomic theory*, Mcgraw-Hill College, 1994.

[8] A. Nedić, A. Ozdaglar, and P.A. Parrilo, *Constrained consensus and optimization in multi-agent networks*, (2009).

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

[10] S.S. Ram, A. Nedić, and V.V. Veeravalli, *Distributed stochastic subgradient projection algorithms for convex optimization*, (2008), (submitted).

[11] R. Srikant, *The mathematics of internet congestion control*, Birkhäuser, Boston, 2003.

[12] S.S. Stanković, M.S. Stanković, and D. Stipanović, *Consensus based overlapping decentralized estimator*, IEEE Trans. Autom. Control **54** (2009), 410–415.

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

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

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