Asynchronous Stochastic Convex Optimization over Random Networks: Error Bounds

B. Touri
IESE Dept.
University of Illinois
Urbana, IL 61801
touri1@illinois.edu

A. Nedić
IESE Dept.
University of Illinois
Urbana, IL 61801
angelia@illinois.edu

S. Sundhar Ram
ECE Dept.
University of Illinois
Urbana, IL 61801
ssrini5@illinois.edu

***Abstract*— We consider a distributed multi-agent network system where the goal is to minimize the sum of convex functions, each of which is known (with stochastic errors) to a specific network agent. We are interested in asynchronous algorithms for solving the problem over a connected network where the communications among the agent are random. At each time, a random set of agents communicate and update their information. When updating, an agent uses the (sub)gradient of its individual objective function and its own stepsize value. The algorithm is completely asynchronous as it neither requires the coordination of agent actions nor the coordination of the stepsize values. We investigate the asymptotic error bounds of the algorithm with a constant stepsize for strongly convex and just convex functions. Our error bounds capture the effects of agent stepsize choices and the structure of the agent connectivity graph. The error bound scales at best as m in the number m of agents when the agent objective functions are strongly convex.**

***Index Terms*— convex optimization, networked system, stochastic algorithms, asynchronous algorithms, random consensus.**

I. Introduction

An important problem in the context of wired and wireless networks is the problem of minimizing of a sum of functions where each component function is available (with stochastic errors) to a specific network agent [14], [23], [27], [28]. Such a problem requires the design of optimization algorithms that are *distributed and local*. The algorithms are to be distributed in the sense that they have to execute their actions without a central coordinator and/or access to a central information. The algorithms are to be local in the sense that each agent can only use its local objective function and can exchange some limited information with its immediate neighbors.

In this paper, we propose an asynchronous distributed algorithm for optimization over random networks arising from random communications among the agents. At each iteration of the algorithm, a random subset of agents is active, whereby each agent in the set performs a consensus step followed by a gradient step. In the consensus step, an agent computes a "weighted" average of its estimate and the estimates received from its neighbors. In the gradient step, an agent updates the weighted average based on the gradient of its local objective function computed with a stochastic error. We are interested in the case when the agents use constant but uncoordinated stepsize values. We investigate the asymptotic bounds on the system error resulting from the stepsize choices, network structure and the stochastic gradient errors. In particular, we provide such error bounds for the iterates of the algorithm when the agent objective functions are strongly convex, and for the averaged iterates when the functions are just convex.

The algorithm in this paper is closely related to the asynchronous gossip algorithm proposed in [24], where the convergence of the gossip-based algorithm is investigated for convex and nonconvex (scalar) objective functions. In contrast with [24], this present paper develops a more general algorithm with the focus on establishing the *error bounds for approximate solutions in the case of convex functions*.

The work in this paper is also closely related to the work in [17], where a distributed subgradient method is considered over a network with a random dynamic network connectivity structure. There, the subgradient evaluations are exact, the stepsize values have to be coordinated among the agents, and the matrices used by the agents in the consensus step are assumed to be doubly stochastic almost surely. The last two requirements are somewhat restrictive since both require some coordination of the agents. Unlike [17], we consider the method with stochastic errors, uncoordinated stepsize values, and we relax the doubly stochasticity requirement on the matrices used in the consensus step. However, our work is limited to random communications occurring over a network with a static underlying connectivity graph, which is less general than the dynamic connectivity graph used in [17].

On a broader basis, the algorithm in this paper is related to the distributed (deterministic) consensus-based optimization algorithm proposed in [21], [22] and further studied in [15], [17], [20], [25], [27]. That algorithm is requires the agents to update simultaneously and to coordinate their stepsize choices, which is in contrast with the algorithm discussed in this paper.

A different distributed model has been proposed in [31] and also studied in [3], [7], [32], where the complete objective function information is available to each agent, with the aim of distributing the processing by allowing an agent to update only a part of the decision vector. Related to the algorithm of this paper is also the literature on incremental algorithms [6], [12]–[14], [16], [18], [19], [23], [26], [29], where the network agents sequentially update a single iterate sequence and only

This research is supported by NSF Career Award CMMI-0742538.

one agent updates at any given time in a cyclic or a random order. While being local, the incremental algorithms differ fundamentally from the algorithm studied in this paper (where all agents maintain and update their own iterate sequence). In addition, since we are interested in the effect of stochastic errors, on a broader scale our work is also related to stochastic (sub)gradient methods [4], [9]–[11].

The novelty of this work is mainly in three aspects. First, we establish the *error bounds on the performance of the asynchronous distributed algorithms where the agents use uncoordinated stepsize*. The error bounds show favorable scaling with the size m of the network. Second, our model is general enough to account for *random failures of communication links*. Second, we are dealing with the general case where the agents compute their *(sub)gradients with stochastic errors*.

Our development combines the ideas used to study the basic random gossip and broadcast algorithms [1], [8] with the tools that are generally used to study the convergence of the stochastic (sub)gradient methods.

The rest of the paper is organized in the following manner. In Section II, we describe the problem of our interest, present our algorithm and assumptions. In Section III, we show some basic relation for later use, while in Section IV we provide a relation for the the disagreement of the agent iterates. We establish the error bounds for the algorithm in Section V, and we conclude with a discussion in Section VI.

Notation. All vectors are column vectors. For a vector x, $[x]_i$ denotes its ith component and $\|x\|$ denotes its Euclidean norm. The vector with all entries equal to 1 is denoted by $\mathbb{1}$. We use $\Pi_X[x]$ to denote the Euclidean projection of a vector x on a set X. For a matrix W, $[W]_{ij}$ denotes its ijth entry and $\|W\|$ denotes the matrix norm induced by the Euclidean vector norm. A matrix is termed *stochastic* if all its entries are nonnegative and the sum of the entries in each row is equal to 1. A matrix W is doubly stochastic if both W and W^T are stochastic. The cardinality of a set U with finitely many elements is denoted by $|U|$. We use $\mathsf{E}[Y]$ to denote the expectation of a random variable Y. We denote by χ_E the indicator function of a random event E.

II. Problem, algorithm and assumptions

We consider a network of m agents that are indexed by $1,\ldots,m$, and we let $V=\{1,\ldots,m\}$. The network has a static topology that is represented by the bidirectional graph $(V,\mathcal{E})$, where $\mathcal{E}$ is the set of links in the network. We have $\{i,j\}\in\mathcal{E}$ if agents i and j can communicate with each other. We assume that $\{i,i\}\in\mathcal{E}$, which models the fact that each agent has access to its own information state. The network objective is to solve the following optimization problem:

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

where $X\subseteq\mathbb{R}^n$, and $f_i:\mathcal{D}\to\mathbb{R}$ for all i, where $\mathcal{D}$ is some open set that contains the set X. The function f_i is only known to agent i that can compute the gradients $\nabla f_i(x)$ with stochastic errors[1]. The goal is to solve problem (1) using an algorithm that confirms with the distributed nature of the problem information and local connectivity structure of the agents in the network. Our interest is in an algorithm that does not require any coordination of the agents' actions.

A. Asynchronous Optimization Algorithm

We consider a generic class of random algorithms where the agent communications are randomized. We let W_k be the matrix describing the information exchange in the network at time k. The matrix W_k has nonnegative entries and its sparsity structure is compliant with the network connectivity graph $(V,\mathcal{E})$, i.e., the entry $[W_k]_{ij}$ can be positive only if $\{i,j\}\in\mathcal{E}$. Furthermore, $[W_k]_{ij}>0$ if and only if agent j communicates with agent i at time k, and otherwise $[W_k]_{ij}=0$. Thus, we have for each $i,j\in V$ with $i\neq j$,

$$\begin{aligned} &[W_k]_{ij}>0 \quad \text{iff } (j,i) \text{ is active at time } k \text{ and } \{i,j\}\in\mathcal{E}, \\ &[W_k]_{ij}=0 \quad \text{otherwise,} \end{aligned}$$

and for each $i\in V$,

$$[W_k]_{ii}>0.$$

We now describe the algorithm. First, each agent i performs a consensus-like step to combine its current estimate $x_{i,k-1}\in\mathbb{R}^n$ with the estimates $x_{j,k-1}\in\mathbb{R}^n$ received from some of its neighbors, as follows:

$$v_{i,k}=\sum_{j=1}^{m}[W_k]_{ij}\,x_{j,k-1}. \tag{2}$$

The matrix W_k captures the weights used by the agents as well as the communication pattern at time k. Define the sets

$$U_k=\{j\in V\mid [W_k]_{j\ell}>0\,\text{for some } \ell\neq j\} \qquad \text{for each } k.$$

The set U_k is the set of agents j that receive estimates $x_{\ell,k-1}$ from their neighbors ℓ at time k. Basically, the set U_k is the index set of agents that update at time k. The new iterates are defined as follows: for each $i\in V$,

$$x_{i,k}=\Pi_X\left[v_{i,k}-\alpha_i\left(\nabla f(v_{i,k})+\epsilon_{i,k}\right)\chi_{\{i\in U_k\}}\right], \tag{3}$$

where $\alpha_i>0$ is a *constant stepsize* of agent i. The vector $\nabla f_i(x)$ is the gradient[2] of f_i at x, and $\epsilon_{i,k}$ is the stochastic error of computing $\nabla f(x)$ at $x=v_{i,k}$. The initial iterates $x_{i,0}\in X$ are random and *independent of* $\{W_k\}$.

We note that our model for W_k includes the network connectivity models with random link failures, as well as random gossip and broadcast. In the gossip algorithm [8], the matrix W_k has the form $W_k=I-\frac{1}{2}(e_{I_k}-e_{J_k})(e_{I_k}-e_{J_k})^T$, where I_k and J_k are two neighboring agents that update, i.e., $U_k=\{I_k,J_k\}$. For the case of the broadcast algorithm [1], the set U_k consists of all the neighbors j of some agent I_k, i.e., $U_k=\{j\in V\mid\{I_k,j\}\in\mathcal{E}\}$.

[1]See [26] for the motivation for studying stochastic errors.

[2]If the function is not differentiable but is convex, then $\nabla f_i(x)$ denotes a subgradient, as discussed later.

B. Assumptions

For the constraint set X and the agent objective functions f_i, we use the following assumption.

Assumption 1: The set X is compact and convex. Each f_i is defined and convex over some open set containing the set X.

A convex function is continuous over the relative interior of its domain (see [2], Proposition 1.4.6). Thus, by Assumption 1, each f_i is continuous over the set X, and the sum $f = \sum_{i=1}^m f_i$ is also continuous over X. Moreover, by the compactness of X, problem (1) has an optimal solution.

Differentiability of the functions f_i is not assumed. At a point where the gradient does not exist, we use a subgradient: $\nabla g(x)$ is a *subgradient* of a function g at a point x in the domain of g (denoted by $\operatorname{dom} g$) if the following relation holds

$$\nabla g(x)^T(y-x) \le g(y) - g(x) \qquad \text{for all } y \in \operatorname{dom} g.$$

By Assumption 1, a subgradient of $f_i(x)$ exists at every point $x \in X$ for each i (see [2], Proposition 4.4.2). Under the compactness of X, the subgradients of each function f_i are bounded uniformly over X. We let a scalar C be such that

$$\|\nabla f_i(x)\| \le C \qquad \text{for all } x \in X \text{ and } i \in V. \tag{4}$$

For the underlying graph $(V,\mathcal{E})$ and the matrices W_k, we assume the following.

Assumption 2: The graph $(V,\mathcal{E})$ is connected. The matrix sequence $\{W_k\}$ is i.i.d. and each W_k has positive diagonal entries almost surely. Each matrix W_k is stochastic and the expected matrix $\bar{W} = \mathsf{E}[W_k]$ is doubly stochastic. Furthermore, $\bar{W}$ has the sparsity pattern compliant with the underlying graph, i.e., for $i,j \in V$ with $i \ne j$,

$$\bar{W}_{ij} > 0 \quad \text{if and only if} \quad \{i,j\} \in \mathcal{E}.$$

Note that the requirement that $\bar{W}_{ij} > 0$ for $\{i,j\} \in \mathcal{E}$ does not require that $\bar{W}$ is symmetric, but rather that $\bar{W}$ has a symmetric sparsity pattern.

Assumption 2 is satisfied in randomized gossip and broadcast schemes. Furthermore, in both of these schemes, when the graph $(V,\mathcal{E})$ is connected there exists $\lambda < 1$ such that:

$$\mathsf{E}\big[\|D_k z\|^2\big] \le \lambda \|z\|^2$$

for the matrices $D_k = W_k - \frac{1}{m}\mathbb{1}\mathbb{1}^T W_k$ and all $z \in \mathbb{R}^m$. Essentially, the relation states that the largest eigenvalue of the matrix $\mathsf{E}\big[D_k^T D_k\big]$ is less than 1, and λ could be taken to be equal to the largest eigenvalue. We later show that the relation holds for the matrices satisfying Assumption 2 (see Theorem 1 in Section IV).

We have several additional comments regarding the implications of Assumption 2. Under the assumption that $\{W_k\}$ is *i.i.d.*, the random index-set sequence $\{U_k\}$ is also i.i.d. Thus, for every $i \in V$, the event sequence $\{i \in U_k\}$ is i.i.d. We let γ_i denote the probability of the event $\{i \in U_k\}$, which is actually the probability that agent i updates at any given time.

Under the assumption that the matrices W_k are *stochastic*, we have the following observation. When agent i does not receive any new information at time k, i.e., $i \notin U_k$, there holds $[W_k]_{i\ell} = 0$ for all $\ell \ne i$. By the stochasticity of W_k, it follows $[W_k]_{ii} = 1$. Hence, from relations (2)–(3) we see that

$$x_{i,k} = v_{i,k}, \qquad v_{i,k} = x_{i,k-1} \qquad \text{for } i \notin U_k. \tag{5}$$

Thus, when $i \notin U_k$, agent i does not update.

Under Assumptions 1–2, the iterates $x_{i,k}$ lie in the set X. In particular, since the set X is closed and convex, the projection on X is well defined. Since $x_{i,0} \in X$ for all i and the matrices W_k are stochastic, it follows that

$$v_{i,k} \in X \text{ and } x_{i,k} \in X \text{ for all } i \in V \text{ and } k \ge 1. \tag{6}$$

We next discuss the assumptions on the gradient errors $\epsilon_{i,k}$. Let $\mathscr{F}_k$ be the σ-algebra generated by the entire history of the algorithm up to time k, i.e., for $k \ge 1$,

$$\mathscr{F}_k = \{(x_{i,0}, i \in V);\ W_t,\ (\epsilon_{j,t}, j \in U_t),\ 1 \le t \le k\},$$

with $\mathscr{F}_0 = \{x_{i,0}, i \in V\}$. We use the following assumption.

Assumption 3: With probability 1, we have:

(a) $\mathsf{E}[\epsilon_{i,k} \mid \mathscr{F}_{k-1}, W_k] = 0$ for all $i \in U_k$ and all $k \ge 1$.
(b) $\mathsf{E}\big[\|\epsilon_{i,k}\|^2 \mid \mathscr{F}_{k-1}, W_k\big] \le \nu^2$ for some deterministic scalar $\nu > 0$, and for all $i \in U_k$ and $k \ge 1$.

Assumption 3 holds, for example, when $\epsilon_{i,k}$ are zero mean, independent in time, and have bounded second moments.

III. Preliminaries

In this section, we establish some preliminary results that we use later in our error analysis.

We make use of the following lemma for a scalar sequence.

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

$$d_k \le \beta d_{k-1} + u_{k-1} \qquad \text{for all } k \ge 1.$$

Then,

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

Proof: From the relation between d_k and u_k, we can see by induction (on k) that

$$d_k \le \beta^k d_0 + \sum_{t=0}^{k-1} \beta^{k-t-1} u_t \qquad \text{for all } k \ge 1.$$

Since $\beta \in (0,1)$, it follows

$$\limsup_{k\to\infty} d_k \le \limsup_{k\to\infty} \sum_{t=0}^{k-1} \beta^{k-t-1} u_t.$$

It remains to show that

$$\limsup_{k\to\infty} \sum_{t=0}^{k} \beta^{k-t} u_t \le \frac{1}{1-\beta} \limsup_{t\to\infty} u_t. \tag{7}$$

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

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

Since $\sum_{t=K+1}^{k} \beta^{k-t} \le \frac{1}{1-\beta}$ and $\sum_{t=0}^{K} \beta^{k-t} \le \frac{\beta^{k-K}}{1-\beta}$, it follows that for all $k \ge K$,

$$\sum_{t=0}^{k} \beta^{k-t} u_t \le \left(\max_{0 \le s \le K} u_s \right) \frac{\beta^{k-K}}{1-\beta} + \frac{a}{1-\beta}.$$

Therefore, $\limsup_{k \to \infty} \sum_{t=0}^{k} \beta^{k-t} u_t \le \frac{a}{1-\beta}$, and relation (7) follows by the definition of a. ■

We also use the following lemma providing two basic relations between $v_{i,k}$ and $x_{j,k-1}$.

Lemma 2: Let $\{W_k\}$ be a sequence of stochastic and i.i.d. matrices with $\mathsf{E}[W_k]$ being doubly stochastic. Then, for the vectors $v_{i,k}$ of (2), we have for any $z \in \mathbb{R}^n$ and all $k \ge 1$,

$$\sum_{i=1}^{m} \mathsf{E}[\|v_{i,k} - z\| \mid \mathscr{F}_{k-1}] \le \sum_{j=1}^{m} \|x_{j,k-1} - z\|, \qquad (8)$$

$$\sum_{i=1}^{m} \mathsf{E}\left[\|v_{i,k} - z\|^2 \mid \mathscr{F}_{k-1}\right] \le \sum_{j=1}^{m} \|x_{j,k-1} - z\|^2. \qquad (9)$$

Proof: The convexity of the norm and the stochasticity of W_k yield $\|v_{i,k} - z\| \le \sum_{j=1}^{m} [W_k]_{ij} \|x_{j,k-1} - z\|$ for any $z \in \mathbb{R}^n$, $i \in V$ and $k \ge 1$. By taking the conditional expectation on the past history $\mathscr{F}_{k-1}$, using the independency of W_k, and by summing over all i, we have

$$\sum_{i=1}^{m} \mathsf{E}[\|v_{i,k} - z\| \mid \mathscr{F}_{k-1}] \le \sum_{i=1}^{m} \sum_{j=1}^{m} \mathsf{E}[W_k]_{ij} \|x_{j,k-1} - z\|.$$

Exchanging the order of summation and using the doubly stochasticity of $\mathsf{E}[W_k]$, we obtain relation (8). The proof of relation (9) follows the same line of argument starting with the convexity of the squared Euclidean norm. ■

We next provide an estimate for the expected value $\mathsf{E}\left[\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 \mid \mathscr{F}_{k-1}, W_k\right]$.

Lemma 3: Under Assumption 1, Assumption 3, and the stochasticity of W_k, we have for all i and $k \ge 1$,

$$\mathsf{E}\left[\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 \mid \mathscr{F}_{k-1}, W_k\right] \le (C + \nu)^2.$$

Proof: Expanding the term $\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2$, we have

$$\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 = \|\nabla f_i(v_{i,k})\|^2 + 2\nabla f_i(v_{i,k})^T \epsilon_{i,k} + \|\epsilon_{i,k}\|^2.$$

When X is convex and W_k's are stochastic, we have $v_{i,k} \in X$ for all i and $k \ge 1$, see (6). Under Assumption 1, the (sub)gradients of each f_i are uniformly bounded over the set X by some scalar C (see (4)). Hence, $\|\nabla f_i(v_{i,k})\| \le C$. Using this, the fact that $v_{i,k}$ is completely determined given $\mathscr{F}_{k-1}$ and W_k, and using Assumption 3 on $\epsilon_{i,k}$, from the preceding equality we obtain

$$\mathsf{E}\left[\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 \mid \mathscr{F}_{k-1}, W_k\right] \le C^2 + \nu^2 \le (C + \nu)^2.$$

■

We now provide a basic iterate relation for the method. This relation relies on the preceding lemma and the nonexpansive property of the projection on a closed convex set $X \subset \mathbb{R}^n$:

$$\|\Pi_X[x] - z\| \le \|x - z\| \qquad \text{for any } x \in \mathbb{R}^n \text{ and } z \in X. \quad (10)$$

The basic iterate relation is given in the following.

Lemma 4: Under Assumption 1, Assumption 3, and the stochasticity of W_k, we have for all $z \in X$, $k \ge 1$, and $i \in U_k$,

$$\mathsf{E}\left[\|x_{i,k} - z\|^2 \mid \mathscr{F}_{k-1}, W_k\right] \le \|v_{i,k} - z\|^2 + \alpha_i^2 (C + \nu)^2 - 2\alpha_i \nabla f_i(v_{i,k})^T (v_{i,k} - z).$$

Proof: For $i \in U_k$, from relation (3) and the nonexpansive property of the projection operation in (10), we obtain

$$\begin{aligned} \|x_{i,k} - z\|^2 &\le \|v_{i,k} - \alpha_i \left(\nabla f_i(v_{i,k}) + \epsilon_{i,k}\right) - z\|^2 \\ &= \|v_{i,k} - z\|^2 + \alpha_i^2 \|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 \\ &\quad -2\alpha_i (\nabla f_i(v_{i,k}) + \epsilon_{i,k})^T (v_{i,k} - z). \end{aligned}$$

We now take the conditional expectation on the past $\mathscr{F}_{k-1}$ and W_k. Under the assumptions on the errors, we have $\mathsf{E}\left[\epsilon_{i,k}^T (v_{i,k} - z) \mid \mathscr{F}_{k-1}, W_k\right] = 0$. Using this and Lemma 3, we obtain the desired relation. ■

In our subsequent analysis, we invoke Hölder's inequality

$$\mathsf{E}\left[\sum_{i=1}^{K} |y_i^T z_i|\right] \le \sqrt{\sum_{i=1}^{K} \mathsf{E}[\|y_i\|^2]} \sqrt{\sum_{i=1}^{K} \mathsf{E}[\|z_i\|^2]}, \qquad (11)$$

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

IV. Disagreement Estimate

The disagreement estimates as a function of time are important in our development of the error bounds for the method. We start the development by showing that under our assumptions, there exists $\lambda < 1$ such that $\mathsf{E}\left[\|D_k z\|^2\right] \le \lambda \|z\|^2$ for the matrices $D_k = W_k - \frac{1}{m}\mathbb{1}\mathbb{1}^T W_k$ and all $z \in \mathbb{R}$. This relation is instrumental for the establishment of the estimates in this and the following section. The relation is justified by the following.

Theorem 1: Let W be a random stochastic matrix such that $\mathsf{E}[W]$ is doubly stochastic. Assume that the diagonal elements of W are positive almost surely. Let $(V, \mathcal{E})$ be an undirected connected graph such that for all $i, j \in V$ with $i \ne j$,

$$\mathsf{E}[W_{ij}] > 0 \quad \text{if } \{i, j\} \in \mathcal{E}.$$

Then, we have for a scalar $\lambda < 1$,

$$\mathsf{E}\left[\|Dz\|^2\right] \le \lambda \|z\|^2 \qquad \text{for all } z \in \mathbb{R}^m,$$

where $D = W - \frac{1}{m}\mathbf{1}\mathbf{1}^T W$.

Proof: Consider the vector $D\mathbb{1}$ and note that from the stochasticity of W we have $D\mathbb{1} = \mathbb{1} - \frac{1}{m}\mathbf{1}\mathbf{1}^T \mathbb{1} = 0$. Thus,

$$Dz = 0 \qquad \text{for all } z = c\mathbb{1} \text{ for some } c \in \mathbb{R}.$$

Therefore, it suffices to show that there exists $\lambda < 1$ such that $\mathsf{E}\left[\|Dz\|^2\right] \le \lambda \|z\|^2$ for all $z \in \mathbb{R}^m$ with $z^T \mathbb{1} = 0$.

To do so, let us define

$$V(z) = \frac{1}{m}\sum_{i=1}^{m}(z_i - \bar{z})^2 \qquad \text{for any } z \in \mathbb{R}^m.$$

Let z be arbitrary and let $y = Dz$. Note that by the definition of D, we have $\bar{y} = 0$. Therefore $\|Dz\|^2 = \|y - \bar{y}\mathbb{1}\|^2 = mV(y)$. It can be seen that (see Theorem 4 in [30]),

$$\mathsf{E}[V(y)] \leq V(z) - \frac{1}{m}\sum_{i<j} H_{ij}(z_i - z_j)^2,$$

where $H = \mathsf{E}\left[W^T W\right]$. Thus,

$$\mathsf{E}\left[\|Dz\|^2\right] \leq mV(z) - \sum_{i<j} H_{ij}(z_i - z_j)^2. \qquad (12)$$

Note that for any $i, j \in V$ such that $\mathsf{E}[W_{ij}] > 0$ or $\mathsf{E}[W_{ji}] > 0$, we have

$$H_{ij} = \mathsf{E}\left[\sum_{\ell=1}^{m} W_{\ell i} W_{\ell j}\right] \geq \mathsf{E}[W_{ii}W_{ij}] + \mathsf{E}[W_{ji}W_{jj}].$$

Since $W_{ii} > 0$ and $W_{jj} > 0$ almost surely, it follows $H_{ij} > 0$ whenever $\mathsf{E}[W_{ij}] > 0$ or $\mathsf{E}[W_{ji}] > 0$. Since $\mathsf{E}[W_{ij}] > 0$ whenever $\{i,j\} \in \mathcal{E}$, we see that $H_{ij} > 0$ whenever $\{i,j\} \in \mathcal{E}$. Thus, let $h = \frac{1}{m}\min_{\{i,j\}\in\mathcal{E}} H_{ij} > 0$. From (12) we have

$$\mathsf{E}\left[\|Dz\|^2\right] \leq mV(z) - mh\sum_{\{i,j\}\in\mathcal{E}}(z_i - z_j)^2. \qquad (13)$$

Since the properties assumed in the theorem are invariant under permutation of indices, let us assume that the entries of z are sorted in nondecreasing order, i.e., $z_1 \leq z_2 \leq \cdots \leq z_m$ (otherwise, we can permute coordinate indices of z, and the rows and columns of W accordingly). Then, we have $(z_i - z_j)^2 \geq \sum_{\ell=i}^{j-1}(z_{\ell+1} - z_\ell)^2$ and hence,

$$\sum_{\{i,j\}\in\mathcal{E}}(z_i - z_j)^2 \geq \sum_{\{i,j\}\in\mathcal{E}}\sum_{\ell=i}^{j-1}(z_{\ell+1} - z_\ell)^2. \qquad (14)$$

Now, since the graph $(V, \mathcal{E})$ is connected, for any $\ell = 1, 2, \ldots, m-1$, there exists $i \leq \ell$ and $j \geq \ell + 1$ such that $\{i,j\} \in \mathcal{E}$, for otherwise, the two vertex sets $\{1, \cdots, \ell\}$ and $\{\ell+1, \cdots, m\}$ will be disconnected. Hence, by relation (14) we have $\sum_{\{i,j\}\in\mathcal{E}}(z_i - z_j)^2 \geq \sum_{\ell=1}^{m-1}(z_{\ell+1} - z_\ell)^2$. By convexity of the squared-norm, we obtain

$$\begin{aligned}\sum_{\ell=1}^{m-1}(z_{\ell+1} - z_\ell)^2 &\geq \frac{1}{m-1}(z_m - z_1)^2 \\ &\geq \frac{1}{m(m-1)}\sum_{i=1}^{m}(z_i - \bar{z})^2.\end{aligned}$$

Therefore, by combining (13), (14) and the preceding relation and using the definition of $V(z)$, we see

$$\begin{aligned}\mathsf{E}\left[\|Dz\|^2\right] &\leq mV(z) - \frac{h}{m-1}\sum_{i=1}^{m}(z_i - \bar{z})^2 \\ &= \left(1 - \frac{h}{m-1}\right)\|z - \bar{z}\|^2.\end{aligned}$$

It follows that for all $z \in \mathbb{R}^m$ with $z^T\mathbb{1} = 0$ we have

$$\mathsf{E}\left[\|Dz\|^2\right] \leq \left(1 - \frac{h}{m-1}\right)\|z\|^2.$$

Hence, the desired relation holds with $\lambda = 1 - \frac{h}{m-1} < 1$. ■

When Assumption 2 holds, each W_k satisfies the conditions of Theorem 1 so that, in view of i.i.d. property of $\{W_k\}$, we have for $\lambda < 1$,

$$\mathsf{E}\left[\|D_k z\|^2\right] \leq \lambda\|z\|^2 \qquad \text{for all } z \in \mathbb{R}^m. \qquad (15)$$

We next measure the agent disagreement as the dispersion of the iterates $x_{i,k}$ around their average at any given time. For this, we define

$$\bar{y}_k = \frac{1}{m}\sum_{j=1}^{m} x_{j,k} \qquad \text{for all } k \geq 0.$$

We quantify the agent disagreement by $\sum_{i=1}^{m}\mathsf{E}[\|x_{i,k} - \bar{y}_k\|]$, for which we have the following estimate.

Proposition 1: Let Assumptions 1–3 hold. Let $\{x_{i,k}\}$, $i \in V$, be the sequences generated by algorithm (2)–(3). Then, we have for all $k \geq 0$,

$$\sum_{i=1}^{m}\mathsf{E}[\|x_{i,k} - \bar{y}_k\|] \leq \sqrt{m}\,\lambda^{\frac{k}{2}}\sqrt{\sum_{i=1}^{m}\mathsf{E}[\|x_{i,0} - \bar{y}_0\|^2]} + \varepsilon_{\text{net}},$$

where ε_{net} is the error given by

$$\varepsilon_{\text{net}} = \frac{\sqrt{m}}{1 - \sqrt{\lambda}}\sqrt{N}\alpha_{\max}(C + \nu),$$

$\alpha_{\max} = \max_{1\leq j\leq m}\alpha_j$, and N is the maximum number of nodes updating at any time, i.e., $N = \max_k |U_k|$, and $\lambda < 1$ is the scalar as in (15).

Proof: To prove the estimate, we will consider components of the vectors $x_{i,k}$. Define for each $\ell = 1, \ldots, n$, the vectors

$$z_k^\ell = ([x_{1,k}]_\ell, \ldots, [x_{m,k}]_\ell)^T \quad \text{for } k \geq 0. \qquad (16)$$

From the definition of the method, we have

$$z_k^\ell = W_k\, z_{k-1}^\ell + \zeta_k^\ell \qquad \text{for } \ell = 1, \ldots, n,\ k \geq 1, \qquad (17)$$

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

$$[\zeta_k^\ell]_i = \begin{cases}[\Pi_X[v_{i,k} - \alpha_i(\nabla f_i(v_{i,k}) + \epsilon_{i,k})] - v_{i,k}]_\ell & \text{if } i \in U_k, \\ 0 & \text{otherwise.}\end{cases} \qquad (18)$$

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

$$[\bar{y}_k]_\ell = \frac{1}{m}\mathbb{1}^T z_k^\ell. \qquad (19)$$

By (17) and (19), we have $[\bar{y}_k]_\ell = \frac{1}{m}\left(\mathbf{1}^T W_k z_{k-1}^\ell + \mathbb{1}^T\zeta_k^\ell\right)$ for all ℓ and $k \geq 1$, implying

$$\begin{aligned}z_k^\ell - [\bar{y}_k]_\ell\mathbb{1} &= W_k z_{k-1}^\ell + \zeta_k^\ell - \frac{1}{m}\mathbb{1}\mathbb{1}^T(W_k z_{k-1}^\ell + \zeta_k^\ell) \\ &= D_k\, z_{k-1}^\ell + M\zeta_k^\ell,\end{aligned}$$

where $D_k = W_k - \frac{1}{m}\mathbb{1}\mathbb{1}^T W_k$, $M = I - \frac{1}{m}\mathbb{1}\mathbb{1}^T$ and I denotes the identity matrix. Since the matrices W_k are stochastic ($W_k\mathbb{1} = \mathbb{1}$), it follows $D_k\mathbb{1} = 0$, implying that $D_k[\bar{y}_k]_\ell\mathbb{1} = 0$. Hence, for all ℓ and k,

$$z_k^\ell - [\bar{y}_k]_\ell\mathbb{1} = D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbb{1}) + M\zeta_k^\ell.$$

From the preceding relation we have for all ℓ and $k \geq 1$,

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

By summing these relations over $\ell = 1,\ldots,n$, and then taking the expectation and using Hölder's inequality (11), we obtain

$$\sum_{\ell=1}^n \mathsf{E}\Big[\|z_k^\ell - [\bar{y}_k]_\ell\mathbb{1}\|^2\Big] \leq \left(\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbb{1})\|^2\big]} + \sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|M\zeta_k^\ell\|^2\big]}\right)^2. \tag{20}$$

We estimate the term $\mathsf{E}\big[\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbb{1})\|^2\big]$ by using the iterated expectation rule, conditioning on $\mathscr{F}_{k-1}$, and using the fact the matrix W_k is independent of $\mathscr{F}_{k-1}$. This yields for all ℓ and $k \geq 1$,

$$\mathsf{E}\Big[\big\|D_k(z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbb{1})\big\|^2\Big] \leq \lambda\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbb{1}\|^2, \tag{21}$$

where $\lambda = \|\mathsf{E}\big[D_k^T D_k\big]\|^2$ with $\lambda < 1$ (see (15)).

We now estimate the term $\mathsf{E}\big[\|M\zeta_k^\ell\|^2\big]$ of (20). The matrix $M = I - \frac{1}{m}\mathbb{1}\mathbb{1}^T$ is a projection matrix (it projects on the subspace orthogonal to the vector $\mathbb{1}$), so that we have $\|M\|^2 = 1$, implying that $\|M\zeta_k^\ell\|^2 \leq \|\zeta_k^\ell\|^2$ for all k. Using this and the definition of ζ_k^ℓ in (18), we obtain

$$\|M\zeta_k^\ell\|^2 \leq \sum_{i\in U_k} \big|[\Pi_X[v_{i,k} - \alpha_i\left(\nabla f_i(v_{i,k}) + \epsilon_{i,k}\right)] - v_{i,k}]_\ell\big|^2.$$

Thus, by summing these relations over all ℓ, by using $v_{i,k} \in X$ (cf. (6)), and by using the nonexpansive property of the projection (cf. (10)), we see

$$\begin{aligned}\sum_{\ell=1}^n \|M\zeta_k^\ell\|^2 &\leq \sum_{i\in U_k} \|\Pi_X[v_{i,k} - \alpha_i\left(\nabla f_i(v_{i,k}) + \epsilon_{i,k}\right)] - v_{i,k}\|^2 \\ &\leq \sum_{i\in U_k} \alpha_i^2\,\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2.\end{aligned}$$

Taking the expectation, and using the iterated expectation rule and Lemma 3, we obtain

$$\begin{aligned}\sum_{\ell=1}^n \mathsf{E}\Big[\|M\zeta_k^\ell\|^2\Big] \quad &\leq \quad \mathsf{E}\left[\sum_{i\in U_k} \alpha_i^2 \mathsf{E}\big[\|\nabla f_i(v_{i,k}) + \epsilon_{i,k}\|^2 \mid \mathscr{F}_{k-1}\big]\right] \\ &\leq \quad N\alpha_{\max}^2(C+\nu)^2,\end{aligned} \tag{22}$$

where $N = \max_k |U_k|$ and $\alpha_{\max} = \max_i \alpha_i$.

From relations (20)–(22) we have for all $k \geq 1$,

$$\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell\mathbb{1}\|^2\big]} \leq \sqrt{\lambda}\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_{k-1}^\ell - [\bar{y}_{k-1}]_\ell\mathbb{1}\|^2\big]} + \sqrt{N}\alpha_{\max}\left(C+\nu\right).$$

Since $\lambda < 1$, it follows that for all $k \geq 1$,

$$\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell\mathbb{1}\|^2\big]} \leq \sqrt{\lambda^k}\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_0^\ell - [\bar{y}_0]_\ell\mathbb{1}\|^2\big]} + \frac{1-\sqrt{\lambda^k}}{1-\sqrt{\lambda}}\sqrt{N}\alpha_{\max}\left(C+\nu\right). \tag{23}$$

Using $z_k^\ell = ([x_{1,k}]_\ell,\ldots,[x_{m,k}]_\ell)^T$ (see (16)), we have for all $k \geq 0$,

$$\sqrt{\sum_{\ell=1}^n \mathsf{E}\big[\|z_k^\ell - [\bar{y}_k]_\ell\mathbb{1}\|^2\big]} = \sqrt{\sum_{i=1}^m \mathsf{E}[\|x_{i,k} - \bar{y}_k\|^2]}. \tag{24}$$

By Hölder's inequality (11), we also have for all $k \geq 0$,

$$\sum_{i=1}^m \mathsf{E}[\|x_{i,k} - \bar{y}_k\|] \leq \sqrt{m}\sqrt{\sum_{i=1}^m \mathsf{E}[\|x_{i,k} - \bar{y}_k\|^2]}. \tag{25}$$

Combining relation (23) (where we use $1 - \sqrt{\lambda^k} \leq 1$), equality (24) (for $k = 0$), and inequality (25), we obtain

$$\sum_{i=1}^m \mathsf{E}[\|x_{i,k} - \bar{y}_k\|] \leq \sqrt{m}\,\sqrt{\lambda^k}\sqrt{\sum_{i=1}^m \mathsf{E}[\|x_{i,0} - \bar{y}_0\|^2]} + \frac{\sqrt{m}}{1-\sqrt{\lambda}}\sqrt{N}\alpha_{\max}\left(C+\nu\right) \qquad \text{for } k \geq 1.$$

Note that this relation holds for $k = 0$ in view of (25). ■

The result of Proposition 1 captures the effects of the connectivity structure of the underlying graph $(V,\mathcal{E})$ and the information flow due to choice of matrices W_k. In particular, the random gossip algorithm we have $N = 2$, while in the random broadcast, N is equal to the maximum number of the neighbors of an agent [the maximum node degree in the graph $(V,\mathcal{E})$]. We also note that the error grows as $\sqrt{m}$ in the size m of the network for the network structures where N and λ do not depend on m.

Observe that, since $\lambda < 1$, as $k \to \infty$ the estimate of Proposition 1 yields the following asymptotic error:

$$\limsup_{k\to\infty} \sum_{i=1}^m \mathsf{E}[\|x_{i,k} - \bar{y}_k\|] \leq \frac{\sqrt{m}}{1-\sqrt{\lambda}}\sqrt{N}\alpha_{\max}\left(C+\nu\right).$$

Thus, the growth of the disagreement is of the order $\sqrt{m}$ in the size m of the network if N and λ do not depend on m.

V. Error Estimates

In this section, we provide error estimates for the method. We consider the iterates of the method for the case when the functions f_i are strongly convex in Section V-A. When the functions are not strongly convex, we look at the running averages of the iterates in Section V-B.

A. Estimates for Iterates

Assuming that the functions f_i are strongly convex over the set X, we provide asymptotic error estimates for $\sum_{i=1}^m \mathsf{E}\big[\|x_{i,k}-x^*\|^2\big]$ where x^* is an optimal solution.

Based on Proposition 1, we have the following result.

Proposition 2: Let Assumptions 1–3 hold. Assume that, for each $i \in V$, the function f_i is strongly convex over the set X with a constant σ_i. Also, let the stepsize $\alpha_i > 0$ be such that $2\alpha_i\sigma_i < 1$. Then, for the optimal solution x^* of problem (1) and the sequences $\{x_k^i\}$, $i \in V$, of method (2)–(3), we have

$$\limsup_{k\to\infty} \sum_{i=1}^m \mathsf{E}\big[\|x_{i,k}-x^*\|^2\big] \le \frac{\varepsilon + 2\alpha_{\max} C \varepsilon_{\rm net}}{1-q},$$

where

$$\varepsilon = 2m(1-\gamma_{\min})\delta_{\alpha,\sigma}C_X^2 + \sum_{i=1}^m \alpha_i^2\gamma_i(C+\nu)^2 + 2m\delta_{\alpha,\gamma}CC_X.$$

$\alpha_{\max} = \max_\ell \alpha_\ell$, $\gamma_{\min} = \min_j \gamma_j$, $q = 1 - 2\gamma_{\min}\min_\ell \alpha_\ell\sigma_\ell$, $\delta_{\alpha,\sigma} = \max_\ell \alpha_\ell\sigma_\ell - \min_j \alpha_j\sigma_j$, $\delta_{\alpha,\gamma} = \max_\ell \alpha_\ell\gamma_\ell - \min_j \alpha_j\gamma_j$, $C_X = \max_{x,y\in X}\|x-y\|$ and $\varepsilon_{\rm net}$ is as given in Proposition 1.

Proof: Since each f_i is strongly convex, the sum $f = \sum_{i=1}^m f_i$ is also strongly convex. Thus, problem (1) has a unique optimal solution $x^* \in X$. From Lemma 4 where we let $z = x^*$, we obtain for any $k \ge 1$ and any $i \in U_k$,

$$\mathsf{E}\big[\|x_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}, W_k\big] \le \|v_{i,k}-x^*\|^2 + \alpha_i^2(C+\nu)^2 - 2\alpha_i \nabla f_i(v_{i,k})^T(v_{i,k}-x^*). \tag{26}$$

Using the strong convexity of f_i, we have

$$\begin{aligned}\nabla f_i(v_{i,k})^T(v_{i,k}-x^*) &= \left(\nabla f_i(v_{i,k}) - \nabla f_i(x^*)\right)^T(v_{i,k}-x^*) \\ &\quad + \nabla f_i(x^*)^T(v_{i,k}-x^*) \\ &\ge \sigma_i\|v_{i,k}-x^*\|^2 \\ &\quad + \nabla f_i(x^*)^T(v_{i,k}-x^*).\end{aligned}$$

Adding and subtracting $\bar{y}_{k-1} = \frac{1}{m}\sum_{j=1}^m x_{j,k}$, we further have $\nabla f_i(x^*)^T(v_{i,k}-x^*) = \nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) + \nabla f_i(x^*)^T(v_{i,k}-\bar{y}_{k-1})$, implying

$$\nabla f_i(x^*)^T(v_{i,k}-x^*) \ge \nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) - C\,\|v_{i,k}-\bar{y}_{k-1}\|. \tag{27}$$

By combining the preceding two relations with inequality (26), we obtain for any $i \in U_k$ and $k \ge 1$,

$$\begin{aligned}&\mathsf{E}\big[\|x_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}, W_k\big] \le (1-2\alpha_i\sigma_i)\|v_{i,k}-x^*\|^2 \\ &+ \alpha_i^2(C+\nu)^2 - 2\alpha_i\nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) \\ &+ 2\alpha_i C\|v_{i,k}-\bar{y}_{k-1}\|.\end{aligned}$$

Since $x_{i,k} = v_{i,k} = x_{i,k-1}$ when $i \notin U_k$ (agent i does not update), we can write

$$\begin{aligned}&\mathsf{E}\big[\|x_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}, W_k\big] \le (1-2\alpha_i\sigma_i)\|v_{i,k}-x^*\|^2 \\ &+ 2\alpha_i\sigma_i\|x_{i,k-1}-x^*\|^2(1-\chi_{\{i\in U_k\}}) \\ &+ \left(\alpha_i^2(C+\nu)^2 - 2\alpha_i\nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*)\right)\chi_{\{i\in U_k\}} \\ &+ 2\alpha_i C\|v_{i,k}-\bar{y}_{k-1}\|.\end{aligned}$$

Taking the expectation with respect to $\mathscr{F}_{k-1}$, and noting that the agent updates with the probability $\gamma_i > 0$, independently of the past, we obtain for any $i \in V$ and $k \ge 1$,

$$\begin{aligned}&\mathsf{E}\big[\|x_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}\big] \le (1-2\alpha_i\sigma_i)\mathsf{E}\big[\|v_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}\big] \\ &+ 2\alpha_i\sigma_i\|x_{i,k-1}-x^*\|^2(1-\gamma_{\min}) \\ &+ \alpha_i^2\gamma_i(C+\nu)^2 - 2\alpha_i\gamma_i\nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) \\ &+ 2\alpha_i C\mathsf{E}[\|v_{i,k}-\bar{y}_{k-1}\| \mid \mathscr{F}_{k-1}]\end{aligned}$$

with $\gamma_{\min} = \min_\ell \gamma_\ell$. Now we note that $\alpha_i\sigma_i \le \min_\ell \alpha_\ell\sigma_\ell + \delta_{\alpha,\sigma}$, with $\delta_{\alpha,\sigma} = \max_s \alpha_s\sigma_s - \min_\ell \alpha_\ell\sigma_\ell$. Similarly, $\alpha_i\gamma_i \le \min_\ell \alpha_\ell\gamma_\ell + \delta_{\alpha,\gamma}$, with $\delta_{\alpha,\gamma} = \max_s \alpha_s\gamma_s - \min_\ell \alpha_\ell\gamma_\ell$. We use these relations, the boundedness of $\|x_{i,k-1}-x^*\|$, and

$$\nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) \le C\max_{x,y\in X}\|y-x\|$$

(which follows from the subgradient boundedness and $\bar{y}_{k-1} \in X$), to obtain

$$\begin{aligned}&\mathsf{E}\big[\|x_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}\big] \\ &\le (1-2\min_\ell \alpha_\ell\sigma_\ell)\mathsf{E}\big[\|v_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}\big] \\ &+ 2(1-\gamma_{\min})\left(\min_\ell \alpha_\ell\sigma_\ell\|x_{i,k-1}-x^*\|^2 + \delta_{\alpha,\sigma}C_X^2\right) \\ &+ \alpha_i^2\gamma_i(C+\nu)^2 - 2\min_\ell \alpha_\ell\gamma_\ell\nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) \\ &+ 2\delta_{\alpha,\gamma}CC_X + 2\alpha_i C\mathsf{E}[\|v_{i,k}-\bar{y}_{k-1}\| \mid \mathscr{F}_{k-1}],\end{aligned}$$

with $C_X = \max_{x,y\in X}\|x-y\| < \infty$ by the compactness of X.

Now, by summing the preceding relations over all i, using relations (8)–(9) and $\sum_{i=1}^m \nabla f_i(x^*)^T(\bar{y}_{k-1}-x^*) \ge 0$ (which holds since $\bar{y}_{k-1} \in X$ and x^* is optimal), we obtain

$$\begin{aligned}&\sum_{i=1}^m \mathsf{E}\big[\|x_{i,k}-x^*\|^2 \mid \mathscr{F}_{k-1}\big] \le q\sum_{j=1}^m \|x_{j,k-1}-x^*\|^2 \\ &+ 2m(1-\gamma_{\min})\delta_{\alpha,\sigma}C_X^2 + \sum_{i=1}^m \alpha_i^2\gamma_i(C+\nu)^2 \\ &+ 2m\delta_{\alpha,\gamma}CC_X + 2\alpha_{\max}C\sum_{j=1}^m \|x_{j,k-1}-\bar{y}_{k-1}\|.\end{aligned}$$

where $\alpha_{\max} = \max_i \alpha_i$ and $q = 1 - 2\gamma_{\min}\min_\ell \alpha_\ell\sigma_\ell$. Since $\gamma_i \in (0,1)$ and $2\alpha_i\sigma_i \in (0,1)$ (by our assumption), it follows that $q \in (0,1)$. Using this and taking the total expectation, from the preceding relation, we have for all $k \ge 1$,

$$\begin{aligned}&\sum_{i=1}^m \mathsf{E}\big[\|x_{i,k}-x^*\|^2\big] \le q\sum_{j=1}^m \mathsf{E}\big[\|x_{j,k-1}-x^*\|^2\big] + \varepsilon \\ &+ 2\alpha_{\max}C\sum_{j=1}^m \mathsf{E}[\|x_{j,k-1}-\bar{y}_{k-1}\|],\end{aligned}$$

where

$$\varepsilon = 2m(1-\gamma_{\min})\delta_{\alpha,\sigma}C_X^2 + \sum_{i=1}^m \alpha_i^2\gamma_i(C+\nu)^2 + 2m\delta_{\alpha,\gamma}CC_X.$$

Since $q \in (0,1)$ by applying Lemma 1, we obtain

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

The desired estimate now follows by using Proposition 1. ■

Proposition 2 captures the effects of the stepsize, the connectivity structure of the graph and the information flow governed by the matrices W_k on the asymptotic error. We now take a closer look into the error estimate. We note that when $\gamma_{\min} \approx 1$, the error term $2m(1-\gamma_{\min})\delta_{\alpha,\sigma} C_X^2$ would be negligible. Thus, if all agents update with high probability, this error term will be small. This term will also be small if the difference $\delta_{\alpha,\sigma} = \max_\ell \alpha_\ell \sigma_\ell - \min_j \alpha_j \sigma_j$ is small. The error term $2m\delta_{\alpha,\gamma} C C_X$ will be small when the difference $\delta_{\alpha,\gamma} = \max_\ell \alpha_\ell \gamma_\ell - \min_j \alpha_j \gamma_j$ is small.

When both $\delta_{\alpha,\sigma}$ and $\delta_{\alpha,\gamma}$ are negligible, recalling $q = 1 - 2\gamma_{\min} \min_\ell \alpha_\ell \sigma_\ell$, the estimate of Proposition 2, reduces to

$$\limsup_{k\to\infty} \sum_{i=1}^{m} \mathsf{E}\left[\|x_{i,k} - x^*\|^2\right] \leq \frac{1}{2\gamma_{\min} \min_\ell \alpha_\ell \sigma_\ell} \sum_{i=1}^{m} \alpha_i^2 \gamma_i (C+\nu)^2 + \frac{\alpha_{\max} C \varepsilon_{\text{net}}}{\gamma_{\min} \min_\ell \alpha_\ell \sigma_\ell},$$

where $\gamma_{\min} = \min_j \gamma_j$, $\alpha_{\max} = \max_\ell \alpha_\ell$, and

$$\varepsilon_{\text{net}} = \frac{\sqrt{m}}{1-\sqrt{\lambda}} \sqrt{N} \alpha_{\max} \, (C+\nu).$$

The error bound captures the effects of the agent stepsize values α_i, the probabilities γ_i of agents' updates, the strong convexity constants σ_i of the agent objective functions f_i, and the underlying network connectivity graph $(V, \mathcal{E})$. The effect of connectivity graph is seen through the maximum number N of agents that update at any given time and the parameter λ that characterizes the convergence rate of random consensus.

Observe that the error is of the order of the largest stepsize, $\alpha_{\max}$. Finally, note that the error grows linearly, as m, with the number m of agents in the network, when the network connectivity structure is such that λ and N do not depend on m. This is seen from the error term involving the summation.

Proposition 2 requires that each agent selects a stepsize α_i so that $2\alpha_i \sigma_i < 1$, which can be ensured when each agent knows the strong convexity constant σ_i of its own objective function f_i. Since $\gamma_i \in (0,1)$ for all i, the relation $0 < 1 - 2\gamma_{\min} \min_i \alpha_i \sigma_i < 1$, i.e., $q \in (0,1)$, holds globally over the network without any coordination among the agents.

B. Estimate for Averaged Iterates

Here, we provide another error estimate that does not require strong convexity. The estimate is for the network objective function values along the time-averaged iterates of each agent.

Proposition 3: Let Assumptions 1–3 hold, and let the sequences $\{x_{i,k}\}$, $i \in V$, be generated by method (2)–(3). Then, for the average vectors $z_{i,t} = \frac{1}{t}\sum_{k=0}^{t-1} x_{i,k}$ for $t \geq 1$ and $i \in V$, we have for any optimal solution x^* of problem (1) and all $i \in V$,

$$\limsup_{t\to\infty} \mathsf{E}[f(z_{i,t})] - f(x^*) \leq \frac{\varepsilon}{2\min_j \gamma_j \alpha_j} + \left(\frac{\alpha_{\max}}{\min_j \gamma_j \alpha_j} + m\right) C\, \varepsilon_{\text{net}},$$

where

$$\varepsilon = \sum_{i=1}^{m} \gamma_i \alpha_i^2 (C+\nu)^2 + 2m \left(\max_\ell \gamma_\ell \alpha_\ell - \min_j \gamma_j \alpha_j\right) C C_X,$$

$C_X = \max_{x,y\in X} \|x-y\|$ and ε_{net} is as given in Proposition 1.

Proof: Since X is compact and $f = \sum_{i=1}^m f_i$ is continuous, the optimal set X^* is nonempty. Thus, by using Lemma 4 with $z = x^*$ for any $x^* \in X^*$, we obtain for any $i \in U_k$ and $k \geq 1$,

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}, W_k\right] \leq \|v_{i,k} - x^*\|^2 + \alpha_i^2 (C+\nu)^2 - 2\alpha_i \nabla f_i(v_{i,k})^T (v_{i,k} - x^*).$$

By using the estimate

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

(see (27)), and the relation $\nabla f_i(x^*)^T (\bar{y}_{k-1} - x^*) \geq f_i(\bar{y}_{k-1}) - f_i(x^*)$, we obtain for $i \in U_k$ and any $k \geq 1$,

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}, W_k\right] \leq \|v_{i,k} - x^*\|^2 + \alpha_i^2 (C+\nu)^2 - 2\alpha_i \left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right) + 2\alpha_i C \|v_{i,k} - \bar{y}_{k-1}\|.$$

The preceding relation holds when $i \in U_k$, which happens with probability γ_i. When $i \notin U_k$, we have $x_k^i = v_{i,k}$ (see the discussion after Assumption 2) which happens with probability $1-\gamma_i$. Thus, by taking the expectation conditioned on $\mathscr{F}_{k-1}$, we obtain

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}\right] \leq \mathsf{E}\left[\|v_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}\right] + \gamma_i \alpha_i^2 (C+\nu)^2 - 2\gamma_i \alpha_i \left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right) + 2\alpha_i C \mathsf{E}[\|v_{i,k} - \bar{y}_{k-1}\| \mid \mathscr{F}_{k-1}] \,. \tag{28}$$

We now add and subtract $2(\min_j \gamma_j \alpha_j)\left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right)$ in the right hand side of (28), and use the estimate

$$|f_i(\bar{y}_{k-1}) - f_i(x^*)| \leq C \|\bar{y}_{k-1} - x^*\| \leq C C_X,$$

which holds by the subgradient boundedness and $\bar{y}_{k-1} \in X$. By doing so, we can see that

$$\mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}\right] \leq \mathsf{E}\left[\|v_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}\right] + \gamma_i \alpha_i^2 (C+\nu)^2 + 2\left(\max_\ell \gamma_\ell \alpha_\ell - \min_j \gamma_j \alpha_j\right) C C_X - 2\left(\min_j \gamma_j \alpha_j\right)\left(f_i(\bar{y}_{k-1}) - f_i(x^*)\right) + 2\alpha_{\max} C\, \mathsf{E}[\|v_{i,k} - \bar{y}_{k-1}\| \mid \mathscr{F}_{k-1}] \,.$$

By summing the preceding inequalities over i and using relations (8) and (9) with $x = \bar{y}_{k-1} \in X$ and $x = x^*$ respectively, after rearranging the terms, we obtain

$$2\left(\min_j \gamma_j\alpha_j\right)(f(\bar{y}_{k-1}) - f(x^*)) \le \sum_{j=1}^m \|x_{j,k-1} - x^*\|^2$$
$$- \sum_{i=1}^m \mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}\right]$$
$$+ \sum_{i=1}^m \gamma_i\alpha_i^2(C+\nu)^2 + 2m\left(\max_\ell \gamma_\ell\alpha_\ell - \min_j \gamma_j\alpha_j\right)CC_X$$
$$+ 2\alpha_{\max}C\sum_{j=1}^m \|x_{j,k-1} - \bar{y}_{k-1}\|,$$

where $f = \sum_{i=1}^m f_i$. By convexity of f and the boundedness of the subgradients of each f_i, we have

$$f(x_{i,k-1}) - f^* \le f(\bar{y}_{k-1}) - f^* + mC\|x_{i,k-1} - \bar{y}_{k-1}\|.$$

Substituting this in the preceding inequality and using the notation

$$\varepsilon = \sum_{i=1}^m \gamma_i\alpha_i^2(C+\nu)^2 + 2m\left(\max_\ell \gamma_\ell\alpha_\ell - \min_j \gamma_j\alpha_j\right)CC_X,$$

we obtain for all $i \in V$ and $k \ge 1$,

$$2\left(\min_j \gamma_j\alpha_j\right)(f(x_{i,k-1}) - f(x^*)) \le \sum_{j=1}^m \|x_{j,k-1} - x^*\|^2$$
$$- \sum_{i=1}^m \mathsf{E}\left[\|x_{i,k} - x^*\|^2 \mid \mathscr{F}_{k-1}\right] + \varepsilon$$
$$+ 2\left(\alpha_{\max} + m\min_j \gamma_j\alpha_j\right)C\sum_{j=1}^m \|x_{j,k-1} - \bar{y}_{k-1}\|.$$

Now, we take the total expectation and by summing the resulting relations from $k = 1$ to $k = t$ for any $t \ge 1$, we obtain for any $i \in V$,

$$2\min_j \gamma_j\alpha_j \sum_{k=1}^t \mathsf{E}[f(x_{i,k-1}) - f(x^*)] \le \sum_{j=1}^m \mathsf{E}\left[\|x_{j,0} - x^*\|^2\right]$$
$$- \sum_{i=1}^m \mathsf{E}\left[\|x_{i,t} - x^*\|^2\right] + t\varepsilon$$
$$+ 2\left(\alpha_{\max} + m\min_j \gamma_j\alpha_j\right)C\sum_{k=1}^t\sum_{j=1}^m \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|].$$

By dividing with $2(\min_j \gamma_j\alpha_j)t$ and then taking the limit superior as $t \to \infty$, we have for any $i \in V$,

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{k=1}^t \mathsf{E}[f(x_{i,k-1}) - f(x^*)] \le \frac{\varepsilon}{2\min_j \gamma_j\alpha_j} +$$
$$\left(\frac{\alpha_{\max}}{\min_j \gamma_j\alpha_j} + m\right)C\limsup_{t\to\infty}\frac{1}{t}\sum_{k=1}^t\sum_{j=1}^m \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|].$$

By convexity of f and $z_{i,t} = \frac{1}{t}\sum_{k=1}^t x_{i,k-1}$, we have

$$\frac{1}{t}\sum_{k=1}^t \mathsf{E}[f(x_{i,k-1}) - f(x^*)] \ge \mathsf{E}[f(z_{i,t})] - f(x^*).$$

Note that for any scalar sequence $\{u_k\}$, there holds

$$\limsup_{t\to\infty}\frac{1}{t}\sum_{k=1}^t u_k \le \limsup_{k\to\infty} u_k.$$

Therefore, by combining the preceding relations, we see that

$$\limsup_{t\to\infty} \mathsf{E}[f(z_{i,t})] - f(x^*) \le \frac{\varepsilon}{2\min_j \gamma_j\alpha_j}$$
$$+ \left(\frac{\alpha_{\max}}{\min_j \gamma_j\alpha_j} + m\right)C\limsup_{k\to\infty}\sum_{j=1}^m \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|].$$

From Proposition 1 we have

$$\limsup_{k\to\infty}\sum_{j=1}^m \mathsf{E}[\|x_{j,k-1} - \bar{y}_{k-1}\|] \le \varepsilon_{\text{net}},$$

and the desired relation follows. ■

As a direct consequence of Proposition 3 and the convexity of f, we have for $i \in V$,

$$\limsup_{t\to\infty} f(\mathsf{E}[z_{i,t}]) \le f^* + \frac{\varepsilon}{2\min_j \gamma_j\alpha_j}$$
$$+ \left(\frac{\alpha_{\max}}{\min_j \gamma_j\alpha_j} + m\right)C\,\varepsilon_{\text{net}},$$

where f^* is the optimal value of problem (1). We have another consequence of Proposition 3. In particular, since $z_{i,t}$ are averages of $x_{i,0}, \ldots, x_{i,t-1}$ for $t \ge 1$, we have for every $i \in V$,

$$\liminf_{k\to\infty} \mathsf{E}[f(x_{i,k})] \le \liminf_{t\to\infty} \mathsf{E}[f(z_{i,t})] \le \limsup_{t\to\infty} \mathsf{E}[f(z_{i,t})].$$

In view of $f(x_{i,k}) - f^* \ge 0$ (since $x_{i,k} \in X$ for all k and i), and Fatou's lemma, by Proposition 3 we have

$$\mathsf{E}\left[\liminf_{k\to\infty} f(x_{i,k})\right] \le \liminf_{k\to\infty} \mathsf{E}[f(x_{i,k})]$$
$$\le f^* + \frac{\varepsilon}{2\min_j \gamma_j\alpha_j}$$
$$+ \left(\frac{\alpha_{\max}}{\min_j \gamma_j\alpha_j} + m\right)C\,\varepsilon_{\text{net}}.$$

To get a closer insight into the behavior of the limiting error of Proposition 3, let ρ be the ratio between the largest and the smallest values of $\gamma_i\alpha_i$ for $i \in V$, i.e.,

$$\rho = \frac{\max_\ell \gamma_\ell\alpha_\ell}{\min_j \gamma_j\alpha_j}.$$

Then, the result of Proposition 3 can be written as:

$$\limsup_{t\to\infty} \mathsf{E}[f(z_{i,t})] \le f^* + \frac{m\alpha_{\max}}{2}\rho(C+\nu)^2$$
$$+ m(\rho-1)CC_X + \left(\frac{\alpha_{\max}}{\min_j \gamma_j\alpha_j} + m\right)C\varepsilon_{\text{net}},$$

where $\alpha_{\max} = \max_i \alpha_i$ and $\varepsilon_{\text{net}} = \frac{\sqrt{m}}{1-\sqrt{\lambda}}\sqrt{N}\,\alpha_{\max}(C+\nu)^2$.

We note that the error bound captures the effects of the agent stepsizes α_i, the probabilities γ_i of agents' updates [through the ratio ρ], and the network connectivity structure (through N and λ). The bound grows as $m\sqrt{m}$ in the number m of the agents, when λ and N do not depend on m. This scaling is worse than scaling as m obtained for strongly convex functions in Proposition 2. Nevertheless, the scaling of the order $m\sqrt{m}$ much is better that the scaling obtained for the distributed consensus-based subgradient algorithm of [25]. Specifically, for the consensus-based distributed algorithm of [25], it is shown that[3]

$$\limsup_{t\to\infty} \mathsf{E}[f(z_{i,t})] \le f^* + m\alpha(C+\nu)^2\left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right).$$

where α is the constant stepsize common to all agents, and θ and β are some constants related to the network structure (which is assumed to be dynamic, but deterministic) The parameter θ is of the order m^2, while the ratio $\beta/(1-\beta)$ does not depend on m, so that the error bound scales as m^4 in the size m of the network.

VI. Conclusion

We have considered asynchronous algorithm for optimization of a network objective given by the sum of objective functions of the agents in the network. We considered the situation where the agent communicate and update at random over a connected communication network. The proposed algorithm is analyzed for the case when agents use constant but uncoordinated stepsize values. We provided two asymptotic error bounds for the expected function values. The first is for the iterates of the algorithm when the agent objective functions are strongly convex, while the second is for the averages of the iterates when the functions are merely convex. The bounds are given explicitly as the function of the number of the agents in the network, the agent stepsize values and some parameters that depend on the network connectivity structure.

References

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

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

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

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

[5] P. Billingsley, *Probability and measure*, John Wiley & Sons, Inc., USA, 1995.

[6] D. Blatt, A.O. Hero, and H. Gauchman, *A convergent incremental gradient method with constant stepsize*, SIAM Journal of Optimization **18** (2007), no. 1, 29–51.

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

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

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

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

[3]See, Theorem 5.4 in [25].

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

[12] A.A. Gaivoronski, *Convergence properties of backpropogation for neural nets via theory of stochastic gradient methods. Part 1.*, Optimization Methods and Software **4** (1994), no. 2, 117–134.

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

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

[15] B. Johanssson, T. Keviczsky, M. Johansson, and K. Johansson, *Subgradient methods and consensus algorithms for solving convex optimization problems*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008, pp. 4185–4190.

[16] K.C. Kiwiel, *Convergence of approximate and incremental subgradient methods for convex optimization*, SIAM Journal on Optimization **14** (2003), no. 3, 807–840.

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

[18] A. Nedić and D.P. Bertsekas, *Incremental subgradient method for nondifferentiable optimization*, SIAM Journal of Optimization **12** (2001), 109–138.

[19] ———, *The effect of deterministic noise in sub-gradient methods*, Tech. report, Lab. for Information and Decision Systems, MIT, 2007.

[20] A. Nedić, A. Olshevsky, A. Ozdaglar, and J.N. Tsitsiklis, *Distributed subgradient methods and quantization effects*, Proceedings of the 47th IEEE Conference on Decision and Control, 2008.

[21] A. Nedić and A. Ozdaglar, *On the rate of convergence of distributed asynchronous subgradient methods for multi-agent optimization*, Proceedings of the 46th IEEE Conference on Decision and Control, 2007, pp. 4711–4716.

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

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

[24] S. Sundhar Ram, A. Nedić, and V.V. Veeravalli, *Asynchronous gossip algorithms for stochastic optimization*, Proceedings of the IEEE CDC Conference, 2009.

[25] ———, *Distributed stochastic subgradient algorithm for convex optimization*, submitted, available at http://arxiv.org/abs/0811.2595, 2009.

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

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

[28] ———, *Sensor networks: When theory meets practice*, ch. Distributed and recursive estimation, Springer, 2009.

[29] M.V. Solodov and S.K. Zavriev, *Error stability properties of generalized gradient-type algorithms*, Journal of Optimization Theory and Applications **98** (1998), no. 3, 663–680.

[30] B. Touri and A. Nedić, *On ergodicity, infinite flow and consensus in random models*, arXiv:1001.1890v2 [math.PR], 2010.

[31] J.N. Tsitsiklis and M. Athans, *Convergence and asymptotic agreement in distributed decision problems*, IEEE Transactions on Automatic Control **29** (1984), no. 1, 42–50.

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