Asynchronous Gossip-Based Random Projection Algorithms for Fully Distributed Problems

Soomin Lee
Electrical and Computer Engineering
University of Illinois
Urbana, IL 61801, USA
Email: lee203@illinois.edu

Angelia Nedić
Industrial and Enterprise Systems Engineering
University of Illinois
Urbana, IL 61801, USA
Email: angelia@illinois.edu

***Abstract*—We consider fully distributed constrained convex optimization problems over a network, where each network agent has a distinct objective and constraint set. We discuss a gossip-based random projection algorithm (GRP) with uncoordinated diminishing stepsizes. We prove that, when the problem has a solution, the iterates of all network agents converge to the same optimal point with probability 1.**

I. Introduction

A number of important problems that arise in wired and wireless networks [1]–[4] can be formulated as a convex constrained minimization problem, where the objective is a sum of convex functions and the constraint is the intersection of convex sets. The goal of the agents in the network is to cooperatively solve the optimization problem subject to the following restrictions: 1) a component objective function and constraint set is only known to a specific network agent, 2) there is no central coordinator that works with global information or synchronizes actions on the network, and 3) the agents usually have a limited memory and computational power. These restrictions motivate the design of distributed, asynchronous and computationally simple algorithms whereby each agent exchanges local information only with its immediate neighbors in an asynchronous manner and performs only a couple of updates per iteration.

In this paper, we propose and analyze an asynchronous fully distributed algorithm that uses the gossip communication protocol [5] and random projections. Random projection based-algorithms have been proposed in [6] for distributed problems with a synchronous update rule and in [7] for centralized problems. Asynchronous algorithms based on a gossip scheme have been proposed and analyzed for a scalar objective function and a diminishing stepsize [8], and a vector objective function and a constant stepsize [9]. An asynchronous broadcast-based algorithm has also been proposed in [10]. The gradient-projection algorithms proposed in the papers [8]–[10] assume that the agents share a common constraint set and the projection is performed on the whole constraint set at each iteration.

In our algorithm, to efficiently handle the projections in gradient-projection step at each iteration, we randomly select a component of the local constraint set and perform projection on that single component. For asynchrony, each agent uses a diminishing stepsize independent of the other agents. Our main interest is in establishing the convergence of the method, which is done by handling the three random factors, namely, gossip communication protocol, stepsizes and projections.

II. Problem Formulation and Algorithm

We consider an optimization problem whose objective and constraint are distributed among m agents over a network. Let an undirected graph $G = (V, E)$ represent the topology of the network, with the vertex set $V = \{1, \ldots, m\}$ and edge set $E \subseteq V \times V$. Let $\mathcal{N}(i)$ be the set of neighbors of agent i. i.e., $\mathcal{N}(i) = \{j \in V \mid \{i, j\} \in E\}$. The goal of the agents is to cooperatively solve the following optimization problem:

$$\min f(x) \triangleq \sum_{i=1}^{m} f_i(x) \qquad \text{s.t. } x \in \mathcal{X} \triangleq \bigcap_{i=1}^{m} \mathcal{X}_i \tag{1}$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ is a convex function, representing the local objective of agent i, and $\mathcal{X}_i \subseteq \mathbb{R}^d$ is a closed convex set, representing the local constraint set of agent i. The function f_i and the set $\mathcal{X}_i$ are known to agent i only. Each set $\mathcal{X}_i$ is defined as the intersection of a collection of simple convex sets. That is, $\mathcal{X}_i = \bigcap_{j \in I_i} \mathcal{X}_i^j$, where I_i is a (possibly infinite) set of indices. We propose a new distributed optimization algorithm for problem (1) that is based on the gossip communication protocol and random projections. We refer to our algorithm as *Gossip-based Random Projection* (GRP).

GRP uses an asynchronous time model as in [5]. Each agent has a local clock that ticks at a Poisson rate of 1. Consider a single virtual clock that ticks whenever any of the local Poisson clock ticks. Thus, the ticks of the virtual clock is a Poisson random process with rate m. Let Z_k be the absolute time of the kth tick of the virtual clock. The time is discretized according to the intervals $[Z_{k-1}, Z_k)$ and this time slot corresponds to our discrete time k. Let I_k denote the index of the agent that wakes up at time k and J_k denote the index of agent I_k's neighbor that is selected for communication. We assume only one agent wakes up at a time.

Let $x_i(k)$ denote the estimate sequence of the decision variable x of agent i at time k. GRP updates according to the following rule. Each agent starts with some initial vector $x_i(0)$, which can be randomly selected for each agent $i \in V$.

For $k \geq 1$, agents other than I_k and J_k do not update:

$$x_i(k) = x_i(k-1) \quad \text{for all } i \notin \{I_k, J_k\}. \tag{2}$$

Agents I_k and J_k calculate the average of their estimates, adjust the average by using their local gradient information and projecting onto a randomly selected component of their local constraint sets, i.e., for $i \in \{I_k, J_k\}$:

$$\begin{aligned} v_i(k) &= (x_{I_k}(k-1) + x_{J_k}(k-1))/2, \\ x_i(k) &= \Pi_{\mathcal{X}_i^{\Omega_i(k)}}\left[v_i(k) - \alpha_i(k)\nabla f_i(v_i(k))\right], \end{aligned} \tag{3}$$

where $\alpha_i(k) = \frac{1}{\Gamma_i(k)}$ is a stepsize with $\Gamma_i(k)$ denoting the number of updates that agent i has performed until time k, and $\{\Omega_i(k)\}_{k\geq 1}$ is a sequence of random variables drawn from the set I_i. The key difference between the work in [8], [9] and this paper is the random projection. Instead of projecting onto the whole constraint set $\mathcal{X}_i$, a component $\mathcal{X}_i^{\Omega_i(k)}$ is selected (or revealed by nature) and the projection is made on that single component. This simple projection is important due to the following reasons: 1) agents on the network usually have a limited memory and computational power, 2) the whole $\mathcal{X}_i$ may not be known in advance but its components are revealed in time, and 3) $\mathcal{X}_i$ may have too many components which makes the projection on this set prohibitive.

For an alternative representation of GRP we define the matrix $W(k)$ as follows:

$$W(k) = I - \frac{1}{2}(e_{I_k} - e_{J_k})(e_{I_k} - e_{J_k})^T \quad \text{for } k \geq 1,$$

where I is the m-dimensional identity matrix, $e_i \in \mathbb{R}^m$ is a vector whose ith entry is equal to 1 and all the other entries are equal to 0. Each $W(k)$ is doubly stochastic by construction, implying that $\mathsf{E}[W(k)]$ is also doubly stochastic. Using $W(k)$, the algorithm (2)–(3) can be equivalently represented as

$$v_i(k) = \sum_{j=1}^{m}[W(k)]_{ij}x_j(k-1), \tag{4a}$$

$$p_i(k) = \Pi_{\mathcal{X}_i^{\Omega_i(k)}}[v_i(k) - \alpha_i(k)\nabla f(v_i(k))] - v_i(k), \tag{4b}$$

$$x_i(k) = v_i(k) + p_i(k)\chi_{\{i\in\{I_k,J_k\}\}}, \tag{4c}$$

where $\chi_{\mathscr{E}}$ is the characteristic function of an event $\mathscr{E}$.

We next discuss our assumptions, the first of which ensures that the information of each agent influences every other agent.

Assumption 1: The underlying graph $G = (V, E)$ is connected. Furthermore, the neighbor selection process is *iid*, whereby at any time agent i is chosen by its neighbor $j \in \mathcal{N}(i)$ with probability $p_{ij} > 0$ ($p_{ij} = 0$ if $j \notin \mathcal{N}(i)$) independently of the other agents in the network.

We use the following assumption for the functions f_i and the sets $\mathcal{X}_i^j$.

Assumption 2: Let the following conditions hold:

(a) The sets $\mathcal{X}_i^j$, $j \in I_i$ are closed and convex for every $i \in V$.
(b) Each function $f_i : \mathbb{R}^d \to \mathbb{R}$ is convex over $\mathbb{R}^d$.
(c) The functions f_i, $i \in V$, are differentiable and have *Lipschitz gradients* with a constant L over $\mathbb{R}^d$,
$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^d.$$
(d) The gradients $\nabla f_i(x)$, $i \in V$ are bounded over the set $\mathcal{X}$, i.e., there exists a constant G_f such that
$$\|\nabla f_i(x)\| \leq G_f \quad \text{for all } x \in \mathcal{X} \text{ and all } i \in V.$$

When each f_i has Lipschitz gradients with a constant L_i, Assumption 2(c) is satisfied with $L = \max_{i\in V} L_i$. Assumption 2(d) is satisfied, for example, when $\mathcal{X}$ is compact.

The next assumption is crucial in our convergence analysis. It makes use of the distance of a vector x from a closed convex set $\mathcal{X}$, which is defined by $\mathrm{dist}(x, \mathcal{X}) \triangleq \min_{v\in\mathcal{X}} \|x - v\|$.

Assumption 3: There exists a constant $c > 0$ such that for all $i \in V$,

$$\mathrm{dist}^2(x, \mathcal{X}) \leq c\mathsf{E}\left[\mathrm{dist}^2(x, \mathcal{X}_i^{\Omega_i(k)})\right] \text{ for all } x \in \mathbb{R}^d. \tag{5}$$

Assumption 3 is satisfied, for example, if each $\mathcal{X}_i^{\Omega_i(k)}$ is an affine set or the feasible set $\mathcal{X}$ has a nonempty interior.

III. Preliminaries

In this section, we state some definitions and results from the literature, which will be used later on.

We define the history of the algorithm as follows. Let $\mathcal{F}_k$ be the σ-algebra generated by the entire history of the algorithm up to time k inclusively, i.e., for all $k \geq 1$,

$$\mathcal{F}_k = \{x_i(0); i \in V\} \cup \{I_\ell, J_\ell, \Omega_i(\ell); i \in \{I_k, J_k\}, 1 \leq \ell \leq k\},$$

and $\mathcal{F}_0 = \{x_i(0); i \in V\}$.

Next, we state a projection theorem (see [11] for its proof).

Lemma 1: Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a nonempty closed convex set. The mapping $\Pi_{\mathcal{X}} : \mathbb{R}^d \to \mathcal{X}$ is continuous and nonexpansive:

(a) $\|\Pi_{\mathcal{X}}[x] - \Pi_{\mathcal{X}}[y]\| \leq \|x - y\|$ for all $x, y \in \mathbb{R}^d$.
(b) $\|\Pi_{\mathcal{X}}[x] - y\|^2 \leq \|x - y\|^2 - \|\Pi_{\mathcal{X}}[x] - x\|^2$ for all $x \in \mathbb{R}^d$ and for all $y \in \mathcal{X}$.

In the analysis of our algorithm, we also make use of the following convergence result due to Robbins and Siegmund (see [12, Lemma 10-11, p. 49-50]).

Lemma 2: Let $\{v_k\}$, $\{u_k\}$, $\{a_k\}$ and $\{b_k\}$ be sequences of non-negative random variables such that

$$\mathsf{E}[v_{k+1}|\mathcal{F}_k] \leq (1+a_k)v_k - u_k + b_k \quad \text{for all } k \geq 0 \quad w.p.1,$$

where $\mathcal{F}_k$ denotes the collection $v_0, \ldots, v_k$, $u_0, \ldots, u_k$, $a_0, \ldots, a_k$ and $b_0, \ldots, b_k$. Also, let $\sum_{k=0}^{\infty} a_k < \infty$ and $\sum_{k=0}^{\infty} b_k < 0$ a.s. Then, we have $\lim_{k\to\infty} v_k = v$ for a random variable $v \geq 0$ *w.p.1*, and $\sum_{k=0}^{\infty} u_k < \infty$ *w.p.1*.

In the above, notation *w.p.1* stands for *with probability 1*.

IV. Basic Relations

In this section, we provide lemmas that will help us with establishing the convergence of GRP. For the convergence analysis, we need to show that the agent's estimates $x_i(k)$ eventually arrive at a consensus and the consensus point lies in the optimal set. For the latter part, we use Lemma 2 by letting $v_k = \sum_{i=1}^{m} \|x_i(k) - x^*\|^2$ for some optimal point x^*.

We start by examining a long term behavior of the stepsizes in Lemma 3. Then, we state Lemma 4 which provides a relation similar to that of Lemma 2. The relation is derived from the nonexpansive projection property (Lemma 1) and Assumption 2. Lemma 4 also helps us view the algorithm's asymptotic error in two parts. One part of the error is standard in any gradient descent algorithms and the other part is due to the use of random projections on component sets. In Lemma 5, we characterize this projection error by showing that the intermediate iterates $\{v_i(k)\}$ approach the feasible set $\mathcal{X}$. Finally, in Lemma 6, we quantify the agent's disagreements in the estimates $v_i(k)$. Due to the lack of space, we do not provide the proofs of Lemmas 4 and 6.

A. Random Stepsizes

Define $E_i(k) = \{i \in \{I_k, J_k\}\}$, which is the event that agent i updates at time k. Let γ_i be the probability of the event $E_i(k)$, then

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

where $p_{ij} > 0$ is the probability that agent i is chosen by its neighbor j to communicate. In the next lemma, we establish long term estimates for the stepsize $\alpha_i(k) = \frac{1}{\Gamma_i(k)}$ in terms of γ_i (see [10] for the proof).

Lemma 3: Let $\alpha_i(k) = 1/\Gamma_i(k)$ for all $k \geq 1$ and $i \in V$. Let $p_{\min} = \min_{\{i,j\} \in E} p_{ij}$. Also, let q be some constant such that $0 < q < \frac{1}{2}$. Then, there exists a large enough $\tilde{k}$ (which depends on q and m) such that with probability 1 for all $k \geq \tilde{k}$ and $i \in V$,

$$\text{(a)} \ \alpha_i(k) \leq \frac{2}{k\gamma_i}, \quad \text{(b)} \ \alpha_i^2(k) \leq \frac{4m^2}{k^2(1+p_{\min})^2},$$
$$\text{(c)} \ \left|\alpha_i(k) - \frac{1}{k\gamma_i}\right| \leq \frac{2}{k^{\frac{3}{2}-q}(1+p_{\min})^2}.$$

This lemma states that the stepsizes $\alpha_i(k)$ behaves like a deterministic stepsize $1/k$ in the long run. It enables us to handle some technical difficulties due to the cross dependencies of the random stepsizes and the other randomness in GRP.

B. Iterate Relation

The following lemma provides a relation among the iterates obtained after one step of the algorithm (4a)-(4c) and a point in the feasible set $\mathcal{X}$.

Lemma 4: Let Assumptions 1-3 hold. Let $\{x_i(k)\}$ be the iterates generated by the algorithm (4a)-(4c). Then, for any $q \in (0, 1/2)$ there is a sufficiently large $\hat{k}$, such that we have with probability 1, for all $x \in \mathcal{X}$, $k \geq \hat{k}$ and $i \in V$,

$$\begin{aligned} \mathsf{E}\big[\|x_i(k) - x\|^2 \mid \mathcal{F}_{k-1}\big] &\leq (1+A(k))\mathsf{E}[\|v_i(k) - x\|^2 \mid \mathcal{F}_{k-1}] \\ &\quad + B(k)G_f^2 - \frac{2}{k}\mathsf{E}[f_i(z_i(k)) - f_i(x) \mid \mathcal{F}_{k-1}] \\ &\quad - C(k)\mathsf{E}[\text{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}], \qquad (6) \end{aligned}$$

where

$$A(k) = (1+4L^2)\bar{\gamma}a_k + \frac{(8+24c)L^2}{k^2\underline{\gamma}},$$

$$B(k) = 4\bar{\gamma}a_k + \frac{8+16c}{k^2\underline{\gamma}}, \quad C(k) = \frac{\underline{\gamma}}{2c} - \frac{\bar{\gamma}a_k}{c} - \frac{2L}{k},$$

$$a_k = \frac{2}{k^{\frac{3}{2}-q}(1+p_{\min})^2}, \ \bar{\gamma} = \max_i \gamma_i \text{ and } \underline{\gamma} = \min_i \gamma_i.$$

Later in Proposition 1, we let x be a point in the optimal set. Hence, $\|x_i(k) - x\|^2$ measures the distance from the current state and a point in the optimal set. The first two terms in the upper bound of (6) are common to any gradient algorithms. The remaining terms are due to the random projections and will be handled later (in Proposition 1).

C. Projection Estimate

In the next lemma, we show that the distance between the estimates $\{v_i(k)\}$, $i \in V$, and the feasible set $\mathcal{X}$ goes to zero as $k \to \infty$.

Lemma 5: Let Assumptions 1-3 hold. Then, with probability 1, we have

$$\sum_{k=1}^{\infty} \mathsf{E}\left[\text{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}\right] < \infty \qquad \text{for all } i \in V.$$

Proof: In Lemma 4, let $x = z_i(k) \triangleq \Pi_{\mathcal{X}}[v_i(k)]$. Then, for any $k \geq \hat{k}$ and $i \in V$, we obtain

$$\begin{aligned} &\mathsf{E}\big[\|x_i(k) - \Pi_{\mathcal{X}}[v_i(k)]\|^2 \mid \mathcal{F}_{k-1}\big] \\ &\quad \leq (1 + A(k))\,\mathsf{E}[\text{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}] \\ &\qquad + B(k)G_f^2 - C(k)\mathsf{E}[\text{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}]. \quad (7) \end{aligned}$$

By using the definition of $v_i(k)$ (as a convex combination of $x_j(k-1)$ in (4a)) and the convexity of the distance function $x \mapsto \text{dist}^2(x, \mathcal{X})$ (see [11, p. 88]), we find that

$$\mathsf{E}\big[\text{dist}^2(v_i(k), \mathcal{X}) | \mathcal{F}_{k-1}\big] \leq \sum_{j=1}^{m} \bar{W}_{ij} \text{dist}^2(x_j(k-1), \mathcal{X}),$$

where $\bar{W} = \mathsf{E}[W(k)]$. Also, by the definition of the projection, we have

$$\text{dist}(x_i(k), \mathcal{X}) = \|x_i(k) - \Pi_{\mathcal{X}}[x_i(k)]\| \leq \|x_i(k) - \Pi_{\mathcal{X}}[v_i(k)]\|.$$

Upon substituting these estimates in relation (7), we have for all $k \geq \hat{k}$ and $i \in V$ with probability 1,

$$\begin{aligned} &\mathsf{E}\big[\text{dist}^2(x_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}\big] \\ &\quad \leq (1 + A(k)) \sum_{j=1}^{m} \bar{W}_{ij} \text{dist}^2(x_j(k-1), \mathcal{X}) \\ &\qquad + B(k)G_f^2 - C(k)\mathsf{E}[\text{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}]. \end{aligned}$$

By summing over all i and using the fact that $\bar{W}$ has column sums equal to 1, we arrive at the following relation: with probability 1 for all $k \geq \hat{k}$ and $i \in V$,

$$\begin{aligned} \sum_{i=1}^{m} \mathsf{E}\Big[\text{dist}^2(x_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}\Big] &\leq (1+A(k)) \sum_{i=1}^{m} \text{dist}^2(x_i(k-1), \mathcal{X}) \\ &\quad + B(k)mG_f^2 - C(k) \sum_{i=1}^{m} \mathsf{E}\left[\text{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}\right]. \end{aligned}$$

Since a_k is a decreasing sequence, from the definition of $C(k)$ in Lemma 4, there exists a k' such that for $k \geq k'$ the coefficient of the last term on the right hand side is positive. Therefore, for $k \geq \max(\hat{k}, k')$, all the conditions of Lemma 2 are satisfied. By applying the Lemma (to a time-delayed process from $\max(\hat{k}, k')$ onward) we conclude that

$$\sum_{k=1}^{\infty} \mathsf{E}\left[\mathrm{dist}^2(v_i(k), \mathcal{X}) \mid \mathcal{F}_{k-1}\right] < \infty \text{ for all } i \in V,\ w.p.1.$$

■

D. Disagreement Estimate

The following lemma shows that the agent disagreement on the intermediate vectors $v_i(k)$ and the error $x_i(k) - v_i(k)$ converges to zero with probability 1.

Lemma 6: Let Assumptions 1-3 hold. Let the iterates $\{v_i(k)\}$ be generated by method (4a)-(4c). Then, we have for all $i \in V$, with probability 1,

(a) $\sum_{k=1}^{\infty} \frac{1}{k}\mathsf{E}[\|v_i(k) - \bar{v}(k)\| \mid \mathcal{F}_{k-1}] < \infty,$

(b) $\sum_{k=1}^{\infty} \mathsf{E}[\|e_i(k)\|^2 \mid \mathcal{F}_{k-1}] < \infty \quad$ for all $i \in V,$

where $\bar{v}(k) = \frac{1}{m}\sum_{i=1}^{m} v_i(k)$ and $e_i(k) = x_i(k) - v_i(k)$.

V. Convergence of GRP Algorithm

In this section, we assert the convergence of the method (4a)-(4c) using the lemmas established in Section IV. Note that Lemma 5 allows us to infer that $v_i(k)$ approaches the feasible set $\mathcal{X}$, while Lemma 6 allows us to claim that any two sequences $\{v_i(k)\}$ and $\{v_j(k)\}$ have the same limit points with probability 1. To claim the convergence of the iterates to an optimal solution, it remains to relate the limit points of $\{v_i(k)\}$ to the optimal solutions of problem (1). This last piece is provided by the iterate relation of Lemma 4, supported by the convergence result in Lemma 2.

From here onward, we use the following notation regarding the optimal value and optimal solutions of problem (1):

$$f^* = \min_{x \in \mathcal{X}} f(x), \qquad \mathcal{X}^* = \{x \in \mathcal{X} \mid f(x) = f^*\}.$$

We have the following convergence result.

Proposition 1: Let Assumptions 1-3 hold. Also, assume that the problem (1) has a nonempty optimal set $\mathcal{X}^*$. Let $\{x_i(k)\}$ be the iterates generated by algorithm (4a)-(4c). Then, the sequences $\{x_i(k)\}$, for $i \in V$, converge to some random point $x^\star$ in the optimal set $\mathcal{X}^*$ with probability 1, *i.e.*,

$$\lim_{k\to\infty} x_i(k) = x^\star \quad \text{for all } i \in V.$$

Proof: Since a_k is a decreasing sequence, from the definition of $C(k)$ in Lemma 4, there exists a k' such that $C(k) > 0$ for all $k \geq k'$ in the last term on the right hand side in (6). We let $\bar{k} \triangleq \max(\hat{k}, k')$ and consider $k \geq \bar{k}$.

We use the definition of the iterate $x_i(k)$ in (4a)-(4c) and Lemma 4. By using the definition of $v_i(k)$ (as a convex combination of $x_j(k-1)$ in (4a)) and the convexity of the norm function, we find that

$$\mathsf{E}\left[\|v_i(k) - x\|^2 \mid \mathcal{F}_{k-1}\right] \leq \sum_{j=1}^{m} \bar{W}_{ij}\|x_j(k-1) - x\|^2,$$

with $\bar{W} = \mathsf{E}[W(k)]$. By summing relations in (6) over $i = 1, \ldots, m$, and using the anpreceding relation with the doubly stochasticity of $\bar{W}$, for any $x \in \mathcal{X}$, $k \geq \bar{k}$ and $i \in V$, we have with probability 1,

$$\sum_{i=1}^{m} \mathsf{E}\left[\|x_i(k) - x\|^2 \mid \mathcal{F}_{k-1}\right] \leq (1{+}A(k)) \sum_{i=1}^{m} \|x_i(k{-}1) - x\|^2 + B(k)mG_f^2 - \frac{2}{k}\sum_{i=1}^{m} \mathsf{E}[f_i(z_i(k)) - f_i(x) \mid \mathcal{F}_{k-1}]. \tag{8}$$

Next, we estimate the last term of (8). Recall that $f(x) = \sum_{i=1}^{m} f_i(x)$. Let $\bar{z}(k) \triangleq \frac{1}{m}\sum_{\ell=1}^{m} z_\ell(k)$. Using $\bar{z}(k)$ and f, we can rewrite the term $f_i(z_i(k)) - f_i(x)$ as follows.

$$\sum_{i=1}^{m}(f_i(z_i(k)) - f_i(\check{x})) = \sum_{i=1}^{m}(f_i(z_i(k)) - f_i(\bar{z}(k))) + (f(\bar{z}(k)) - f(\check{x})). \tag{9}$$

Furthermore, using the convexity of each function f_i, we obtain

$$\sum_{i=1}^{m}(f_i(z_i(k)) - f_i(\bar{z}(k))) \geq \sum_{i=1}^{m}\langle \nabla f_i(\bar{z}(k)), z_i(k) - \bar{z}(k)\rangle \geq -\sum_{i=1}^{m}\|\nabla f_i(\bar{z}(k))\|\,\|z_i(k) - \bar{z}(k)\|.$$

Since $\bar{z}(k)$ is a convex combination of the points $z_i(k) \in \mathcal{X}$, it follows that $\bar{z}(k) \in \mathcal{X}$. This observation and Assumption 2(d), stating that the gradients $\nabla f_i(x)$ are uniformly bounded for $x \in \mathcal{X}$, yield

$$\sum_{i=1}^{m}(f_i(z_i(k)) - f_i(\bar{z}(k))) \geq -G_f\sum_{i=1}^{m}\|z_i(k) - \bar{z}(k)\|. \tag{10}$$

We next consider the term $\|z_i(k) - \bar{z}(k)\|$, for which by using $\bar{z}(k) \triangleq \frac{1}{m}\sum_{\ell=1}^{m} z_\ell(k)$ we have

$$\|z_i(k) - \bar{z}(k)\| = \left\|\frac{1}{m}\sum_{\ell=1}^{m}(z_i(k) - z_\ell(k))\right\| \leq \frac{1}{m}\sum_{\ell=1}^{m}\|z_i(k) - z_\ell(k)\| \leq \frac{1}{m}\sum_{\ell=1}^{m}\|v_i(k) - v_\ell(k)\|,$$

where the first inequality is obtained by the convexity of the norm and the last inequality follows by the nonexpansive projection property (Lemma 1(a)). Furthermore, by using $\|v_i(k) - v_\ell(k)\| \leq \|v_i(k) - \bar{v}(k)\| + \|v_\ell(k) - \bar{v}(k)\|$, we obtain for every $i \in V$,

$$\|z_i(k) - \bar{z}(k)\| \leq \|v_i(k) - \bar{v}(k)\| + \frac{1}{m}\sum_{\ell=1}^{m}\|v_\ell(k) - \bar{v}(k)\|.$$

Upon summing over $i \in V$, we find that

$$\sum_{i=1}^{m}\|z_i(k) - \bar{z}(k)\| \leq 2\sum_{i=1}^{m}\|v_i(k) - \bar{v}(k)\|. \tag{11}$$

Combining relations (11) and (10), and substituting the resulting relation in equation (9), we find that

$$\sum_{i=1}^{m}(f_i(z_i(k)) - f_i(x)) \geq -2G_f \sum_{i=1}^{m} \|v_i(k) - \bar{v}(k)\| + (f(\bar{z}(k)) - f(x)).$$

Finally, by using the preceding estimate in inequality (8) and letting $x = x^*$ for an arbitrary $x^* \in \mathcal{X}^*$, we obtain for any $x^* \in \mathcal{X}^*$ and $k \geq \bar{k}$,

$$\sum_{i=1}^{m} \mathsf{E}\left[\|x_i(k){-}x^*\|^2 | \mathcal{F}_{k-1}\right] \leq (1{+}A(k)) \sum_{i=1}^{m} \|x_i(k{-}1){-}x^*\|^2 - \frac{2}{k}\mathsf{E}[f(\bar{z}(k)) - f^* \mid \mathcal{F}_{k-1}] + B(k)mG_f^2 + \frac{4G_f}{k}\sum_{i=1}^{m}\mathsf{E}[\|v_i(k) - \bar{v}(k)\| \mid \mathcal{F}_{k-1}]. \quad (12)$$

Since $\bar{z}(k) \in \mathcal{X}$, we have $f(\bar{z}(k)) - f^* \geq 0$. Thus, in the light of Lemma 6, relation (12) satisfies all the conditions of Lemma 2. Hence, the sequence $\{\|x_i(k) - x^*\|^2\}$ is convergent with probability 1 for any $i \in V$ and $x^* \in \mathcal{X}^*$, and

$$\sum_{k=0}^{\infty} \frac{1}{k}(f(\bar{z}(k)) - f^*) < \infty \quad w.p.1.$$

The preceding relation and $\sum_{k=0}^{\infty} \frac{1}{k} = \infty$ imply that

$$\liminf_{k\to\infty}(f(\bar{z}(k)) - f^*) = 0 \quad w.p.1. \quad (13)$$

By Lemma 5, noting that $z_i(k) = \Pi_{\mathcal{X}}[v_i(k)]$, we have $\sum_{k=1}^{\infty}\sum_{i=1}^{m} \|v_i(k) - z_i(k)\|^2 < \infty$ with probability 1, implying that

$$\lim_{k\to\infty} \|v_i(k) - z_i(k)\| = 0 \quad \text{for all } i \in V \quad w.p.1. \quad (14)$$

Since the sequence $\{\|x_i(k) - x^*\|\}$ is convergent with probability 1 for any $i \in V$ and every $x^* \in \mathcal{X}^*$, in view of the relations (4a) and (14), respectively, so are the sequences $\{\|v_i(k) - x^*\|\}$ and $\{\|z_i(k) - x^*\|\}$, as well as their average sequences $\{\|\bar{v}(k) - x^*\|\}$ and $\{\|\bar{z}(k) - x^*\|\}$. Therefore, the sequences $\{\bar{v}(k)\|$ and $\{\bar{z}(k)\}$ are bounded with probability 1, and they have accumulation points. From relation (13) and the continuity of f, the sequence $\{\bar{z}(k)\}$ must have one accumulation point in $\mathcal{X}^*$ with probability 1. This and the fact that $\{\|\bar{z}(k) - x^*\|\}$ is convergent with probability 1 for every $x^* \in \mathcal{X}^*$ imply that for a random point $x^\star \in \mathcal{X}^*$,

$$\lim_{k\to\infty} \bar{z}(k) = x^\star \quad w.p.1. \quad (15)$$

Now, from $\bar{z}(k) = \frac{1}{m}\sum_{\ell=1}^{m} z_\ell(k)$ and $\bar{v}(k) = \frac{1}{m}\sum_{i=\ell}^{m} v_\ell(k)$, using relation (14) and the convexity of the norm, we obtain with probability 1,

$$\lim_{k\to\infty} \|\bar{v}(k) - \bar{z}(k)\| \leq \frac{1}{m}\sum_{\ell=1}^{m} \lim_{k\to\infty} \|v_\ell(k) - z_\ell(k)\| = 0.$$

In view of relation (15), it follows that

$$\lim_{k\to\infty} \bar{v}(k) = x^\star \quad w.p.1. \quad (16)$$

By Lemma 6(a), we have

$$\liminf_{k\to\infty} \|v_i(k) - \bar{v}(k)\| = 0 \quad \text{for all } i \in V \quad w.p.1. \quad (17)$$

The fact that $\{\|v_i(k) - x^*\|\}$ is convergent with probability 1 for all i and any $x^* \in \mathcal{X}^*$, together with (16) and (17), implies that

$$\lim_{k\to\infty} \|v_i(k) - x^\star\| = 0 \quad \text{for } i \in V \quad w.p.1. \quad (18)$$

Finally, from Lemma 6(b), we have $\lim_{k\to\infty} \|x_i(k){-}v_i(k)\| = 0$ for all $i \in V$ with probability 1, which together with the limit in (18) yields $\lim_{k\to\infty} x_i(k) = x^\star$ for all $i \in V$ with probability 1. ■

VI. Conclusions

We have considered a distributed problem of minimizing the sum of agents' objective functions over a distributed constraint set $\mathcal{X}_i$. We proposed a gossip-based random projection algorithm for solving the problem over a network. We established convergence with probability 1 to an optimal solution when uncoordinated diminishing stepsizes are used.

Acknowledgement

This work has been funded by National Science Foundation grants CMMI 07-42538 and CCF 11-11342.

References

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

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

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

[4] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks: Quantized data and random link failures," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1383 –1400, March 2010.

[5] S. Boyd, A. Ghosh, S. Member, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[6] S. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *Journal of Selected Topics in Signal Processing (submitted)*, Aug 2012.

[7] A. Nedić, "Random algorithms for convex minimization problems," *Mathematical Programming - B*, vol. 129, pp. 225–253, 2011.

[8] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Asynchronous gossip algorithms for stochastic optimization," in *48th IEEE Conference on Decision and Control*, 2009, pp. 3581–3586.

[9] S. S. Ram, A. Nedi, and V. V. Veeravalli, "Asynchronous gossip algorithm for stochastic optimization: Constant stepsize analysis*," in *Recent Advances in Optimization and its Applications in Engineering*, M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, Eds. Springer Berlin Heidelberg, 2010, pp. 51–60.

[10] A. Nedić, "Asynchronous broadcast-based convex optimization over a network," *IEEE Trans. Automat. Contr.*, vol. 56, no. 6, pp. 1337–1351, 2011.

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

[12] B. Polyak, *Intro. to optimization*. Optimization software, Inc., Publications division, New York, 1987.