52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Gossip-based Random Projection Algorithm for Distributed Optimization: Error Bound*

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

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

Abstract— **We consider a fully distributed constrained convex optimization problem over a multi-agent network. We discuss an asynchronous gossip-based random projection (GRP) algorithm that solves the distributed problem using only local communication and computation. We analyze its error bound for a constant stepsize and provide simulation results on a distributed robust model predictive control problem.**

I. Introduction

A number of important problems that arise in various application domains, including distributed control, large-scale machine learning, wired and wireless networks [1]–[4] can be formulated as a distributed convex constrained minimization problem over a multi-agent network. The problem is usually defined as a sum of convex objective functions over an intersection of convex constraint sets. The goal of the agents is to solve the problem in a distributed way, with each agent handling a component of the objective and constraint. This is useful either when the problem data are naturally collected in a distributed way or when the data are too large to be conveniently processed by a single agent.

Common to these distributed optimization problems are the following operational restrictions: 1) a component objective function and constraint set is only known to a specific network agent (the problem is fully distributed); 2) there is no central coordinator that synchronizes actions on the network or works with global information; 3) the agents usually have a limited memory, computational power and energy; and 4) communication overhead is significant due to the expensive start-up cost and network latencies. These restrictions motivate the design of distributed, asynchronous, computationally simple and local-communication based algorithms.

The focus of this paper is the analysis of an efficient distributed algorithm whereby each agent exchanges local information only with its immediate neighbors in an asynchronous manner, and only a pair of agents locally perform updates at each iteration. We discuss a gradient descent algorithm with *random projections*, which uses a gossip scheme as a communication protocol.

Random projection-based algorithms have been proposed in [5] (see also its extended version [6]) for distributed problems with a synchronous update rule, and in [7] for centralized problems. Synchronous algorithms are often inefficient as they create bottlenecks and waste CPU cycles, while centralized approaches are inapplicable in situations where a central coordinator does not exist. 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]–[12] assume that the agents share a common constraint set and the projection is performed on the whole constraint set at each iteration. To accommodate the situations where the agents have local constraint sets, the distributed gradient methods with distributed projections on local constraint sets have been considered in [13], [14] (see also [15]). However, even the projection on the entire (local) constraint set often overburdens agents, such as wireless sensors, as it requires intensive computations. Furthermore, in some situations, the constraint set can be revealed only component-wise in time, and the whole set is not available in advance, which makes the existing distributed methods inadequate. Our proposed algorithm is intended to accommodate such situations.

In our algorithm, we efficiently handle the projection at each iteration by performing a projection step on the local constraint set that is randomly selected (by nature or by an agent itself). For asynchrony, each agent uses a constant stepsize. Our main goals are to establish the error bound for a constant stepsize and to provide simulation results for the algorithm.

To the best of our knowledge, there is no previous work on asynchronous distributed optimization algorithms that utilize random projections. Finding probabilistic feasible solutions through random sampling of constraints for optimization problems with uncertain constraints have been proposed in [16], [17]. Also, the related work is the (centralized) random projection method proposed by Polyak [18] for a class of convex feasibility problems, and recently considered in [19].

The rest of this paper is organized as follows. In Section II, we formally describe the problem of interest, propose our gossip-based random projection algorithm, and state assumptions on the problem and the network. In Section III, we state two lemmas regarding random projection errors and agent disagreements. In Section IV, we concisely provide

*The support for this work has been provided by the NSF under grants CMMI 07-42538 and CCF 11-11342.

our main results. We present the simulation results on a distributed model predictive control application in Section V and conclude with a summary in Section VI. Appendix contains the proofs of the propositions stated in Section IV.

Notation. A vector is viewed as a column. We write x' to denote the transpose of a vector x. The scalar product of two vectors x and y is $\langle x, y\rangle$. We use $\mathbf{1}$ to denote a vector whose entries are 1 and $\|x\|$ to denote the standard Euclidean norm. We write $\mathrm{dist}(x, \mathcal{X})$ for the distance of a vector x from a closed convex set $\mathcal{X}$, *i.e.*, $\mathrm{dist}(x, \mathcal{X}) = \min_{v\in\mathcal{X}} \|v - x\|$. We use $\Pi_{\mathcal{X}}[x]$ for the projection of a vector x on the set $\mathcal{X}$, *i.e.*, $\Pi_{\mathcal{X}}[x] = \arg\min_{v\in\mathcal{X}} \|v - x\|^2$. We use $\mathsf{E}[Z]$ to denote the expectation of a random variable Z. We often abbreviate *independent and identically distributed* and *with probability 1* as *i.i.d.* and *w.p.1*, respectively.

II. Problem Setup, Algorithm and Assumptions

A. Problem Formulation

We consider an optimization problem where the objective function and constraint sets 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 $V = \{1, \dots, m\}$ and the edge set $E \subseteq V \times V$. Let $\mathcal{N}(i)$ be the set of the 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.

We assume that problem (1) is feasible, and we denote its optimal value by f^*. Moreover, we assume each set $\mathcal{X}_i$ is defined as the intersection of a collection of simple convex sets. That is, $\mathcal{X}_i$ can be represented as $\mathcal{X}_i = \bigcap_{j\in I_i} \mathcal{X}_i^j$, where the superscript j is used to identify a component set and I_i is a (possibly infinite) set of indices. In some applications, $\mathcal{X}_i$ may not be explicitly given in advance due to online constraints or uncertainty. For example, consider the case when $\mathcal{X}_i$ is given by

$$\mathcal{X}_i = \{x \in \mathbb{R}^d \mid \langle a + \xi, x\rangle \le b\},$$

where $a \in \mathbb{R}^d$, $b \in \mathbb{R}$ are deterministic and $\xi \in \mathbb{R}^d$ is a Gaussian random noise. In such a case, a projection-based distributed algorithm cannot be directly applied to solve problem (1) since $|I_i|$ is infinite and the projection of a point on the uncertain set $\mathcal{X}_i$ is impossible. However, a component $\mathcal{X}_i^j$ can be realized from a random selection of ξ and the projection onto the realized component is always possible. Our algorithm is based on such random projections.

B. Asynchronous Gossip-based Random Projection

We discuss a distributed optimization algorithm for problem (1) that is based on random projections and the gossip communication protocol. Gossip algorithms robustly achieve consensus through sparse communications in the network. That is, only one edge $\{i, j\}$ in the network is randomly selected for communication at each iteration and agents i and j simply average their values. From now on, we refer to our algorithm as *Gossip-based Random Projection* (GRP).

GRP uses an asynchronous time model as in [20]. Each agent has a local clock that ticks at a Poisson rate of 1. The setting can be visualized as having 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 that only one agent wakes up at a time.

The distribution by which J_k is selected is characterized by a nonnegative stochastic $m \times m$ matrix $[\Pi]_{ij} = \pi_{ij}$ that conforms with the graph topology $G = (V, E)$, i.e., $\pi_{ij} > 0$ only if $\{i, j\} \in E$. At iteration k, agent I_k wakes up and contacts one of its neighbors J_k with probability $\pi_{I_k J_k}$.

Let $x_i(k)$ denote the estimate of agent i at time k. GRP updates these estimates according to the following rule. Each agent starts with some initial vector $x_i(0)$, which can be randomly selected. For $k \ge 1$, 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, and adjust the average by using their local gradient information and by projecting onto a randomly selected component of their local constraint sets, i.e., for $i \in \{I_k, J_k\}$:

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

where α_i is a constant stepsize of agent i, and $\Omega_i(k)$ is a random variable drawn from the set I_i. The key difference between the work in [11], [13], [14] and this paper is the random projection step. Instead of projecting on the whole constraint set $\mathcal{X}_i$, a component set $\mathcal{X}_i^{\Omega_i(k)}$ is selected (or revealed by nature) and the projection is made on that set, which reduces the required computations per iteration.

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

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

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 other entries are equal to 0. Each $W(k)$ conforms with the network topology G (i.e., if $\{i, j\} \notin E$ and $i \ne j$, then $[W(k)]_{ij} = 0$). Furthermore, each $W(k)$ is doubly stochastic by construction, implying that $\mathsf{E}[W(k)]$ is also doubly stochastic. Using

$W(k)$, algorithm (2)–(3) can be equivalently represented as

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

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

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

where $\chi_{\mathscr{E}}$ is the characteristic function of an event $\mathscr{E}$, that is, $\chi_{\mathscr{E}} = 1$ if $\mathscr{E}$ happens, and $\chi_{\mathscr{E}} = 0$ otherwise.

From here onward, we will shorten $\mathsf{E}[W(k)] = \bar{W}$ since the matrices $W(k)$ are identically distributed. Let λ denote to the second largest eigenvalue of $\bar{W}$. If Assumption 1 holds, the incidence graph associated with the positive entries in the matrix $\bar{W}$ is connected and with a self-loop at each node. Hence, we have $\lambda < 1$.

C. Assumptions

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 *i.i.d.*, whereby at any time agent i is chosen by its neighbor $j \in \mathcal{N}(i)$ with probability $\pi_{ji} > 0$ ($\pi_{ji} = 0$ 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 strongly convex with a constant $\sigma_i > 0$.

(c) Each function f_i is differentiable and has *Lipschitz gradients* with a constant L_i over $\mathbb{R}^d$,
$$\|\nabla f_i(x) - \nabla f_i(y)\| \le L_i \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^d.$$

(d) The set $\mathcal{X}$ is compact.

By part (d) of the assumption, there exists a constant G_f such that

$$\|\nabla f_i(x)\| \le G_f \quad \text{for all } x \in \mathcal{X} \text{ and all } i \in V.$$

The next assumption states set regularity, which is crucial in our convergence analysis.

Assumption 3: There exists a constant $c > 0$ such that for all $i \in V$ and $x \in \mathbb{R}^d$,

$$\operatorname{dist}^2(x, \mathcal{X}) \le c\, \mathsf{E}\left[\operatorname{dist}^2(x, \mathcal{X}_i^{\Omega_i(k)}) | \Omega_\ell(t), t \in [1, k), \ell \in V\right].$$

Assumption 3 is satisfied, for example, if each set $\mathcal{X}_i^j$ is an affine set, or the constraint set $\mathcal{X}$ has a nonempty interior.

III. Preliminaries

Under the assumption that each f_i is strongly convex, the objective function f is also strongly convex and, consequently, problem (1) has a unique solution, denoted by x^*. The GRP method with constant stepizes α_i is not able to generate sequences $\{x_i(k)\}$ converging to the optimal point x^*. However, it can generate sequences that are producing approximate solutions to problem (1). We provide error bounds for the GRP method, which are the estimates for such approximate solutions that characterize the sub-optimality and near-feasibility of the solutions.

In this section, we state some results which we use to establish the error bounds. In Lemma 1, we give a basic iterate relation for the method, while in Lemma 2, we provide an asymptotic error due to the random projections. In Lemma 3, we provide an upper bound on the asymptotic disagreement among the agents. We do not provide the proofs of these lemmas due to the lack of space (see our extended version [21] for their proofs).

A. Iterate Relation

Lemma 1: Let Assumptions 2-3 hold. Let the stepsizes in method (4a)-(4c) be such that $\alpha_i > 0$ and $8\alpha_i^2 L_i^2 - \frac{1}{2c} \le 0$ for all i, where c is the constant from Assumption 3. Then, for the iterates of the GRP method and the solution x^* of problem (1) we have *w.p.1* for all $k \ge 1$ and $i \in \{I_k, J_k\}$,

$$\mathsf{E}[\|x_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k] \le \rho_i \|v_i(k) - x^*\|^2 \\ - 2\alpha_i \langle \nabla f_i(x^*), z_i(k) - x^* \rangle + 8(1+c)\alpha_i^2 G_f^2,$$

where $\rho_i = 1 - \sigma_i\alpha_i + 8(1+c)\alpha_i^2 L_i^2$ for all i.

B. Projection Error Estimate

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

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

where $\pi_{ji} > 0$ is the probability that agent i is chosen by its neighbor j to communicate.

In the following lemma we provide an asymptotic upper bound for the distance between the iterates $x_i(k)$ of the method and the set $\mathcal{X}$.

Lemma 2: Let Assumptions 2-3 hold. Assume that the stepsizes α_i in the GRP method are such that $q_i = 1 - \alpha_i\sigma_i + 8\alpha_i^2 L_i^2 \in (0, 1)$ for all i, and let $q_{pe} = \max_i 1 + \gamma_i(q_i - 1)$. Then, we have $q_{pe} \in (0, 1)$ and

$$\limsup_{k \to \infty} \sum_{i=1}^{m} \mathsf{E}[\operatorname{dist}^2(x_i(k), \mathcal{X})] \le C_{pe} m \bar{\gamma} \bar{\alpha}^2 G_f^2,$$

where $C_{pe} = \frac{8(1+c)}{1-q_{pe}}$, $\bar{\gamma} = \max_i \gamma_i$, and $\bar{\alpha} = \max_i \alpha_i$.

The bound shows the asymptotic distance in terms of the number of agents, the maximum stepsize, the properties of the objective function and the agent-update probability γ_i.

C. Disagreement Estimate

We provide an estimate for the disagreement among the agents in the following lemma.

Lemma 3: Let Assumptions 1–3 hold. Let the stepsizes in method (4a)-(4c) be such that $\alpha_i > 0$, $8\alpha_i^2 L_i^2 - \frac{1}{2c} \le 0$ and $q_i = 1 - \sigma_i\alpha_i + 8\alpha_i^2 L_i^2 \in (0, 1)$ for all i. Let

$\bar{x}(k) = \frac{1}{m}\sum_{i=1}^{m} x_i(k)$ for all k. Then, for the iterates $\{x_i(k)\}$ generated by method (4a)-(4c), we have

$$\limsup_{k\to\infty} \sum_{i=1}^{m} \mathsf{E}[\|x_i(k) - \bar{x}(k)\|^2] \leq C_{de} \frac{m\bar{\alpha}^2 G_f^2}{(1-\sqrt{\lambda})^2},$$

where $C_{de} = 4\left(\frac{8\bar{\gamma}(1+\bar{\alpha}^2\bar{L}^2)(1+c)}{1-q_{pe}} + 1\right)$, $q_{pe} = \max_i 1 + \gamma_i(q_i - 1)$, $\bar{\gamma} = \max_i \gamma_i$, $\bar{\alpha} = \max_i \alpha_i$, and $\bar{L} = \max L_i$.

The bound in Lemma 3 captures the variance of the estimates $x_i(k)$ in terms of the number of agents, the maximum stepsize and the spectral gap $1-\sqrt{\lambda}$ of the matrix $\bar{W}$ induced by the communication graph.

IV. Error Bound

We provide a limiting error bound of algorithm (4a)-(4c) in the following proposition. The bound shows that the convergence of the algorithm is controlled by three factors. The first term shows the relavance of the spectral gap $1-\sqrt{\lambda}$ (or connectivity) of the matrix $\bar{W}$ and the convergence of the algorithm. The second term is shows an error due to a nondiminishing stepsize. The last term is about a network balance. We need $\Delta_{\gamma\alpha} = \max_i \gamma_i\alpha_i - \min_i \gamma_i\alpha_i \approx 0$ to set the term nearly equal to zero. This means that an agent that is selected less frequently (the one with smaller γ_i) has to perform more aggressive updates by choosing a larger stepsize.

Proposition 1: Let Assumptions 1–3 hold. Let the stepsizes in method (4a)-(4c) be such that $\alpha_i > 0$, $8\alpha_i^2 L_i^2 - \frac{1}{2c} \leq 0$ and $\rho_i = 1 - \sigma_i\alpha_i + 8(1+c)\alpha_i^2 L_i^2 \in (0,1)$ for all i. Then, for the sequences $\{x_i(k)\}$, $i \in V$ generated by the algorithm, we have

$$\limsup_{k\to\infty} \frac{1}{m}\sum_{i=1}^{m} \mathsf{E}[\|x_i(k) - x^*\|^2] \leq \left(\frac{C_1}{1-\sqrt{\lambda}} + C_2\right)\bar{\gamma}\bar{\alpha}^2 G_f^2 + C_3 G_f C_x,$$

where $C_x = \max_{x,y\in\mathcal{X}} \|x - y\|$, $C_1 = \frac{2\sqrt{C_{de}}}{1-q}$, $C_2 = \frac{8(1+c)}{1-q}$ and $C_3 = \frac{2\Delta_{\gamma\alpha}}{1-q}$ with $\Delta_{\gamma\alpha} = \max_i \gamma_i\alpha_i - \min_i \gamma_i\alpha_i$, $q = \max_i 1 + \gamma_i(\rho_i - 1)$, $\bar{\gamma} = \max_i \gamma_i$, and $\bar{\alpha} = \max_i \alpha_i$.

Proof: The proof starts with the relation of Lemma 1. Define $\bar{z}(k) = \frac{1}{m}\sum_{i=1}^{m} z_i(k)$, so that $\bar{z}(k) \in \mathcal{X}$. We have $\langle \nabla f_i(x^*), z_i(k) - x^*\rangle = \langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle + \langle \nabla f_i(x^*), z_i(k) - \bar{z}(k)\rangle$, which in view of compactness of $\mathcal{X}$ implies

$$\langle \nabla f_i(x^*), z_i(k) - x^*\rangle \geq \langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle - G_f\|z_i(k) - \bar{z}(k)\|.$$

Substituting the above relation in the relation of Lemma 1, we have for all $k \geq 1$ *w.p.1*,

$$\begin{aligned}&\mathsf{E}[\|x_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}, I_k, J_k] \\ &\leq \rho_i\|v_i(k) - x^*\|^2 - 2\alpha_i\langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle \\ &\quad + 2\alpha_i G_f\|z_i(k) - \bar{z}(k)\| + 8(1+c)\alpha_i^2 G_f^2.\end{aligned}$$

Taking the expectation with respect to $\mathcal{F}_{k-1}$ and using the fact that the preceding inequality holds with probability γ_i, and $x_i(k) = v_i(k)$ with probability $1-\gamma_i$, we obtain *w.p.1* for all $k \geq 1$ and $i \in V$,

$$\begin{aligned}&\mathsf{E}[\|x_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] \\ &\leq (1 + \gamma_i(\rho_i - 1))\,\mathsf{E}[\|v_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] \\ &- 2\gamma_i\alpha_i\mathsf{E}[\langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle \mid \mathcal{F}_{k-1}] \\ &+ 2\gamma_i\alpha_i G_f\mathsf{E}[\|z_i(k) - \bar{z}(k)\| \mid \mathcal{F}_{k-1}] + 8(1+c)\gamma_i\alpha_i^2 G_f^2.\end{aligned}$$

Let $\underline{\alpha} = \min_i \alpha_i$, $\underline{\gamma} = \min_i \gamma_i$, $\bar{\alpha} = \max_i \alpha_i$ and $\bar{\gamma} = \max_i \gamma_i$. By adding and subtracting $2\gamma_i\alpha_i\mathsf{E}[\langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle \mid \mathcal{F}_{k-1}]$, we find that

$$\begin{aligned}&\mathsf{E}[\|x_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] \\ &\leq (1 + \gamma_i(\rho_i - 1))\,\mathsf{E}[\|v_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] \\ &- 2\underline{\gamma}\underline{\alpha}\mathsf{E}[\langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle \mid \mathcal{F}_{k-1}] \\ &+ 2\Delta_{\gamma\alpha}\|\nabla f_i(x^*)\|\|\bar{z}(k) - x^*\| \\ &+ 2\bar{\gamma}\bar{\alpha}G_f\mathsf{E}[\|z_i(k) - \bar{z}(k)\| \mid \mathcal{F}_{k-1}] + 8(1+c)\bar{\gamma}\bar{\alpha}^2 G_f^2,\end{aligned}$$

where $\Delta_{\gamma\alpha} = \max_i \gamma_i\alpha_i - \min_i \gamma_i\alpha_i$. By the compactness of $\mathcal{X}$, we have $\|\nabla f_i(x^*)\| \leq G_f$ and $\|\bar{z}(k) - x^*\| \leq C_x$, where $C_x = \max_{x,y\in\mathcal{X}} \|x - y\|$. Summing the above relations over $i = 1, \ldots, m$, and using $\sum_{i=1}^{m}\langle \nabla f_i(x^*), \bar{z}(k) - x^*\rangle \geq f(\bar{z}(k)) - f(x^*) \geq 0$, we have

$$\begin{aligned}&\sum_{i=1}^{m} \mathsf{E}[\|x_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] \\ &\leq q\sum_{i=1}^{m} \mathsf{E}[\|v_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] + 2\Delta_{\gamma\alpha} m G_f C_x \\ &+ 2\bar{\gamma}\bar{\alpha}G_f\sum_{i=1}^{m} \mathsf{E}[\|z_i(k) - \bar{z}(k)\| \mid \mathcal{F}_{k-1}] + 8(1+c)m\bar{\gamma}\bar{\alpha}^2 G_f^2,\end{aligned}$$

where $q = \max_i 1 + \gamma_i(\rho_i - 1)$. Using the definition of $v_i(k)$ in (4a), the convexity of the norm square function and the doubly stochasticity of the weights, we have

$$\begin{aligned}\sum_{i=1}^{m} \|v_i(k) - x^*\|^2 &\leq \sum_{i=1}^{m}\sum_{j=1}^{m} [W(k)]_{ij}\|x_j(k-1) - x^*\|^2 \\ &= \sum_{j=1}^{m} \|x_j(k-1) - x^*\|^2.\end{aligned}$$

Therefore,

$$\begin{aligned}&\sum_{i=1}^{m} \mathsf{E}[\|x_i(k) - x^*\|^2 \mid \mathcal{F}_{k-1}] \\ &\leq q\sum_{j=1}^{m} \mathsf{E}[\|x_j(k-1) - x^*\|^2 \mid \mathcal{F}_{k-1}] + 2\Delta_{\gamma\alpha} m G_f C_x \\ &+ 2\bar{\gamma}\bar{\alpha}G_f\sum_{i=1}^{m} \mathsf{E}[\|z_i(k) - \bar{z}(k)\| \mid \mathcal{F}_{k-1}] + 8(1+c)m\bar{\gamma}\bar{\alpha}^2 G_f^2.\end{aligned}$$

Since α_i are chosen such that that $\rho_i \in (0,1)$, we have $q \in (0,1)$. Therefore, we obtain

$$\limsup_{k\to\infty} \sum_{i=1}^{m} \mathsf{E}[\|x_i(k) - x^*\|^2]$$

$$\leq \frac{2\bar{\gamma}\bar{\alpha}G_f}{1-q}\limsup_{k\to\infty}\sum_{i=1}^{m}\mathsf{E}[\|z_i(k)-\bar{z}(k)\|] + \frac{2\Delta_{\gamma\alpha}mG_fC_x}{1-q} + \frac{8(1+c)m\bar{\gamma}\bar{\alpha}^2G_f^2}{1-q}. \quad (5)$$

From the projection property it follows that

$$\sum_{i=1}^{m}\mathsf{E}[\|z_i(k)-\bar{z}(k)\|] \leq \sum_{i=1}^{m}\mathsf{E}[\|z_i(k)-\Pi_{\mathcal{X}}[\bar{v}(k)]\|] \leq \sum_{i=1}^{m}\mathsf{E}[\|v_i(k)-\bar{v}(k)\|]. \quad (6)$$

Furthermore, using Hölder's inequality, we have

$$\sum_{i=1}^{m}\mathsf{E}[\|v_i(k)-\bar{v}(k)\|] \leq \sqrt{m\mathsf{E}\left[\sum_{i=1}^{m}\|v_i(k)-\bar{v}(k)\|^2\right]} \leq \sqrt{m\mathsf{E}\left[\sum_{i=1}^{m}\|x_i(k)-\bar{x}(k)\|^2\right]}, \quad (7)$$

where the last inequality follows by the convexity of the norm-squared and the definition of $v_i(k)$. The result follows from (5)–(7) and Lemma 3. ∎

V. Simulation Results

In this section, we apply our GRP algorithm to a distributed robust model predictive control (MPC) example. The purpose of this experiment is to verify the error bound obtained in Section IV and to show in how many iterations the proposed method actually arrive at almost-consensus in various distributed settings.

A linear, time-invariant, discrete-time system is given by the following state equation for $t = 1, \ldots, T$,

$$x(t) = Ax(t-1) + Bu(t), \quad (8)$$

where

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix},$$

with initial state $x(0) = [7,\ 0]'$.

The goal of the agents on the network is to find an optimal control $\mathbf{u} \triangleq [u(1), \ldots, u(T)]'$ of the system (8) over time $t = 1, \ldots, T$ with some random terminal constraints. The distributed optimization problem is given as the following:

$$\min_{\mathbf{u}}\ f(\mathbf{u}) = \sum_{i=1}^{m} f_i(\mathbf{u}) \quad \text{s.t. } \mathbf{u} \in \mathcal{X}. \quad (9)$$

Here,

$$f_i(\mathbf{u}) = \sum_{t=1}^{T}\|x(t)-z_i\|^2 + ru(t), \text{ for } i = 1, \ldots, m,$$

is the local objective of agent i and $r > 0$ is a control parameter. Hence, the agents on the network jointly find a control $\mathbf{u}$ that generates a trajectory $x(t)$, for $t = 1, \ldots, T$ such that the trajectory minimizes the deviations from the points $z_i \in \mathbb{R}^2$ together with the control effort. The information about the points z_i for $i = 1, \ldots, m$ are private and only agent i knows where the z_i is located.

The constraint set $\mathcal{X}$ is a set of control inputs that satisfies the following constraints.

$$\|u(t)\|_\infty \leq 2, \quad \text{for } t = 1, \ldots, T, \quad (10a)$$
$$x(t) = Ax(t-1) + Bu(t), \quad \text{for } t = 1, \ldots, T, \quad (10b)$$
$$x(0) = [7,\ 0]', \quad (10c)$$
$$\max_{\ell=1,2,3,4}\{(a_\ell+\delta_\ell)'x(T) - b_\ell\} \leq 0. \quad (10d)$$

The constraint (10a) states that the control inputs are constrained so that $\|u(t)\|_\infty \leq 2$, for $t = 1, \ldots, T$. The constraints (10b)-(10c) describe the system dynamics. (10d) refers to the random terminal constraints given by the linear inequalities $(a_\ell+\delta_\ell)'x(T) \leq b_\ell$ and the perturbations δ_ℓ are uniform random vectors in boxes $\|\delta_\ell\|_\infty \leq \beta_\ell$. Note that $u(t)$, for $t = 1, \ldots, T$, are the only variables here since $x(t)$, for $t = 1, \ldots, T$, are fully determined by the state equation (10b)-(10c) once $u(t)$, for $t = 1, \ldots, T$, are given.

The constraint set $\mathcal{X}$ is uncertain and not exactly known in advance since the perturbations are uniform random vectors in boxes. To apply the GRP algorithm (4a)-(4c) in solving this robust optimal control problem, at iteration k, each agent I_k and J_k draws a realization of one of the linear inequality terminal constraints, and each of them projects its current iterate on the selected constraint. Subsequently, they perform their projections onto the box constraint (10a).

Since the uncertainty exists in a box, the problem (9) has an equivalent Quadratic Programming (QP) formulation. Note that the following representations are all equivalent:

$$(a_\ell+\delta_\ell)'x(T) \leq b_\ell, \quad \forall(\delta_\ell : \|\delta_\ell\|_\infty \leq \beta_\ell) \quad (11a)$$
$$\Leftrightarrow \max_{\|\delta_\ell\|_\infty \leq \beta_\ell} \delta_\ell' x(T) \leq b_\ell - a_\ell' x(T) \quad (11b)$$
$$\Leftrightarrow a_\ell' x(T) + \beta_\ell|[x(T)]_1| + \beta_\ell|[x(T)]_2| \leq b_\ell. \quad (11c)$$

Therefore, the inequality (10d) admits an equivalent representation of (11c) by a system of linear inequalities with additional variables t_1 and t_2:

$$-t_j \leq [x(T)]_j \leq t_j, \quad \text{for } j = 1, 2, \quad (12a)$$
$$\max_{\ell=1,2,3,4}\{a_\ell' x(T) + \beta_\ell t_1 + \beta_\ell t_2 - b_\ell\} \leq 0. \quad (12b)$$

This alternative representation is only available since we are considering simple box uncertainty sets for the sake of comparison. Note that our GRP algorithm is applicable not just to box uncertainty but to more complicated perturbations such as Gaussian or other distributions.

In the experiment, we use $m = 4$ and $m = 10$ agents with $T = 10$ and $r = 0.1$. We solve the problem on three different network topologies, namely, clique, cycle and star (see Figure 1). For the agent selection probability, we use uniform distribution. That is, at each iteration, one of the m agents is uniformly selected and the selected agent uniformly picks one of its neighbors. Table I shows the second largest eigenvalue λ of $\bar{W}$ for the three network topologies when $m = 4$ and $m = 10$. When m is larger, we can see that λ is

TABLE I

NUMBER OF AGENTS AND λ

m	Clique	Cycle	star
4	0.6667	0.7500	0.8333
10	0.8889	0.9809	0.9444

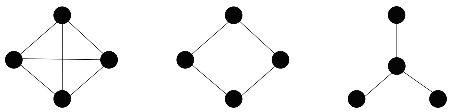

Fig. 1. Clique (left), cycle (center) and star (right) graph used for communication topology ($m = 4$)

very close to one for all of the three cases.

We evaluate the algorithm performance by carrying out 100 Monte-Carlo runs, each with 40,000 iterations for $m = 4$ and 100,000 iterations for $m = 10$. For the stepsize, $\alpha_i = 10^{-5}$ is used for $m = 4$ and $\alpha_i = 10^{-6}$ is used for $m = 10$. The optimal solution $\mathbf{u}^*$ was obtained by solving the equivalent QP problem (i.e., problem (9) with constraints (10a)-(10c) and (12a)-(12b)) using a commercial QP solver.

In Figure 2, we depict $\frac{1}{m}\sum_{i=1}^{m}\|\mathbf{u}_i(k)-\mathbf{u}^*\|^2$ over 40,000 and 100,000 iterations when $m = 4$ and $m = 10$. As the bound in Section IV shows, we can observe that the errors for both cases decrease nearly to zero. Table I lists the second largest eigenvalue λ of the matrix $\bar{W}$ for the three network topologies and the number of agents $m = 4, 10$. When $m = 4$, we can observe that the star graph converges slower than the other two regular graphs. This is because λ is relatively small, and therefore the network balance term (i.e., the last term in Proposition 1) is more dominant. When $m = 10$, however, we can see that the three graphs show almost the same performance. This is because λ is very close to one, and therefore the network topology term (i.e., the first term in Proposition 1) dominates the other two factors. This also illustrates that the gossip communication protocol is robust to any network topologies when the network size is large.

VI. CONCLUSIONS

We have considered a distributed problem of minimizing the sum of convex objective functions f_i over a distributed constraint set $\mathcal{X}_i$. We discussed an asynchronous gossip-based random projection algorithm for solving the problem over a network. We studied an asymptotic error bound of the algorithm for a constant stepsize. We have also provided simulation results for a distributed robust model predictive control problem.

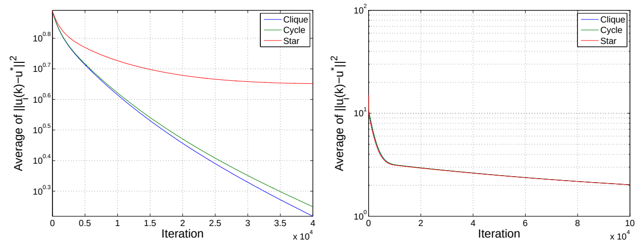

Fig. 2. Iteration vs $\frac{1}{m}\sum_{i=1}^{m}\|\mathbf{u}_i(k)-\mathbf{u}^*\|^2$ when $m = 4$ (left) and $m = 10$ (right)

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. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, pp. 221–229, April 2013.

[6] ——, "Distributed random projection algorithm for convex optimization," 2013, http://arxiv.org/abs/1211.5611.

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

[12] P. Bianchi and J. Jakubowicz, "Convergence of a multi-agent projected stochastic gradient algorithm," 2012, http://arxiv.org/abs/1107.2526.

[13] A. Nedić, A. Ozdaglar, and A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[14] K. Srivastava and A. Nedić, "Distributed asynchronous constrained stochastic optimization," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772–790, 2011.

[15] K. Srivastava, "Distributed optimization with applications to sensor networks and machine learning," Ph.D. dissertation, ISE Department, University of Illinois at Urbana-Champaign, 2011.

[16] T. Alamo, R. Tempo, and E. Camacho, "Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2545 –2559, Nov 2009.

[17] G. C. Calafiore, "Random convex programs," *SIAM J. Optimiz.*, vol. 20, no. 6, pp. 3427–3464, Dec. 2010.

[18] B. Polyak, "Random algorithms for solving convex inequalities," in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications*, ser. Studies in Computational Mathematics, Y. C. Dan Butnariu and S. Reich, Eds. Elsevier, 2001, vol. 8, pp. 409 – 422.

[19] A. Nedić, "Random projection algorithms for convex set intersection problems," in *Proc. of the 49th IEEE Conference on Decision and Control*, 2010, pp. 7655–7660.

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

[21] S. Lee and A. Nedić, "Asynchronous gossip-based random projection algorithms over networks," 2013, http://arxiv.org/abs/1304.1757.