Performance Evaluation of Multi-Agent Distributed Collaborative Optimization under Random Communication Topologies

Ion Matei and John S. Baras

***Abstract*— We investigate collaborative optimization in a multi-agent setting, when the agents execute in a distributed manner using local information, while the communication topology used to exchange messages and information is modeled by a graph-valued random process, independent of other time instances. Specifically, we study the performance of the consensus-based multi-agent subgradient method, for the case of a constant stepsize, as measured by two metrics: rate of convergence and guaranteed region of convergence, evaluated via their expected values. Under a strong convexity type of assumption, we provide upper bounds on the performance metrics, which explicitly depend on the probability distribution of the random graph and on the agents' estimates of the optimal solution. This provides a guide for tuning the parameters of the communication protocol such that good performance of the multi-agent subgradient method is ensured.**

I. INTRODUCTION

Multi-agent distributed optimization problems appear naturally in many distributed processing problems (such as network resource allocation), where the optimization cost is a convex function which is not necessarily separable. A distributed subgradient method for multi-agent optimization of a sum of convex functions was proposed in [7], where each agent has only local knowledge of the optimization cost, i.e. knows only one term of the sum. The agents exchange information according to a communication topology, modeled as an undirected, time varying graph, which defines the communication neighborhoods of the agents. The agents maintain *estimates* of the optimal decision vector, which are updated in two stages. The first stage consists of a consensus step among the estimate of an agent and its neighbors. In the second stage, the result of the consensus step is updated in the direction of a subgradient of the local knowledge of the optimization cost. Another multi-agent subgradient method was proposed in [5], where the communication topology is assumed time invariant and where the order of the two stages mentioned above is inverted.

In this note we investigate the collaborative optimization problem in a multi-agent setting, when the agents execute in a distributed manner using local information, while the communication topology used to exchange messages and information is modeled by a graph-valued random process, independent of other time instances. Specifically, we study the performance of the consensus-based multi-agent subgradient method proposed in [7], for the case of a constant stepsize, as measured by two metrics: rate of convergence and guaranteed region of convergence, evaluated via their expected values. Random graphs are suitable models for networks that change with time due to link failures, packet drops, node failures, etc. An analysis of the multi-agent subgradient method under random communication topology is addressed in [8]. The authors assume that the consensus weights are lower bounded by some positive scalar and give upper bounds on the performance metrics as functions of this scalar and other parameters of the problems. More precisely, the authors give upper bounds on the distance between the cost function and the optimal solution (in expectation), where the cost is expressed as a function of the (weighted) time average of the optimal decision vector's estimate. *In this paper, our main goal is the provide upper bounds on the performance metrics, which explicitly depend on the probability distribution of the random graph*. We first derive an upper bound on how close the cost function, evaluated at the estimate, gets to the optimal solution. Next, under a strong convexity type of assumption, we focus on the squared distance between the estimate of the optimal decision and some minimizer. We provide an upper bound for this metric, which will give us the rate of convergence of the estimate to a guaranteed neighborhood of the optimum. The performance metrics are evaluated via their expected values. The explicit dependence on the graph's probability distribution allows us to determine the optimal probability distributions that would ensure the best guaranteed upper bounds on the metrics. This idea has relevance especially in the wireless networks case, where the communication topology has a random nature with a probability distribution (partially) determined by the communication protocol parameters. As example of possible applications of our results, in [10] we address two simple scenarios where the goal is to tune the communication protocol parameters such that the performance of the multi-agent subgradient method is improved. In the first scenario we consider that the agents use a randomize scheme for enabling packet transmissions, where the agents decide to act like a transmitter or a receiver with some probability. This probability will play the role of the protocol parameter. In the second scenario, we assume that the transmissions happen according to a pre-established order (TDMA based protocol) but they are affected by interferences. In this case, the power allocated for transmissions is going to play the

This material is based upon work supported by the US Air Force Office of Scientific Research MURI award FA9550-09-1-0538

Ion Matei and John S. Baras are with the Department of Electrical Engineering and Institute for Systems Research, University of Maryland, College Park, USA, imatei@umd.edu, baras@umd.edu

role of protocol parameters. Both scenarios are applied on a small world type of communication topology.

Notations: Let X be a subset of $\mathbb{R}^n$ and let y be a point in $\mathbb{R}^n$. Using an abuse of notation, by $\|y-X\|$ we understand the distance from the point y to the set X, i.e. $\|y-X\| \triangleq \min_{x\in X}\|y-x\|$, where $\|\cdot\|$ is the standard two norm.

Let $f:\mathbb{R}^n\to\mathbb{R}$ be a convex function. We denote by $\partial f(x)$ the subdifferential of f at x, i.e. the set of all subgradients of f at x:

$$\partial f(x) = \{d\in\mathbb{R}^n | f(y)\geq f(x)+d'(y-x),\ \forall y\in\mathbb{R}^n\}. \quad (1)$$

Let $\varepsilon\geq 0$ be a nonnegative real number. We denote by $\partial_\varepsilon f(x)$ the ε-subdifferential of f at x, i.e. the set of all ε-subgradients of f at x:

$$\partial f(x) = \{d\in\mathbb{R}^n | f(y)\geq f(x)+d'(y-x)-\varepsilon,\ \forall y\in\mathbb{R}^n\}. \quad (2)$$

We will denote by LEM and SLEM the largest and second largest eigenvalue of a stochastic matrix, respectively. We will use MASM for multi-agent subgradient method and pmf for probability mass function.

Paper structure: Section II contains the problem formulation. More precisely presents in details the communication and optimization models assumed in this note. In Section III, we introduce a set of preliminary results, which mainly consist in providing upper bounds for a number a quantities of interest. By combining the preliminary results, in Section IV we give upper bounds for the expected values of two performance metrics: the distance between the cost function evaluated at the estimate and the optimal solution and the (squared) distance between the estimate and some minimizer.

Due to space limitation some of the proofs of our results are omitted. The missing proofs can be found in reference [10].

II. Problem formulation

A. Communication model

Consider a network of N agents, indexed by $i=1,\dots,N$. The communication topology is time varying and is modeled by a random graph $\mathbf{G}(k)$, whose edges correspond to communication links among agents. Given a positive integer M, the graph $\mathbf{G}(k)$ takes values in a finite set $\mathcal{G}=\{G_1,G_2,\dots,G_M\}$. The underlying random process of $\mathbf{G}(k)$ is assumed i.i.d. with probability distribution $Pr(\mathbf{G}(k)=G_i)=p_i,\ \forall k\geq 0$, where $\sum_{i=1}^M p_i=1$. Throughout this note, we will consider only bidirectional communication topologies, i.e. $\mathbf{G}(k)$ is undirected.

Assumption 2.1: (Connectivity assumption) The graph resulting from the union of graphs in $\mathcal{G}$, i.e. $\bigcup_{i=1}^N G_i$, is connected.

Let G be a graph of order N and let $A\in\mathbb{R}^{N\times N}$ be a row stochastic matrix, with positive diagonal entries. We say that the matrix A, *corresponds* to the graph G or the graph G is *induced* by A if any non-zero entry (i,j) of A implies a link from j to i in G and vice-versa. Consider a matrix product of stochastic matrices of length m, $\prod_{i=1}^m A_i$. We say that the graph induced by the aforementioned matrix product, is given by the union of graphs corresponding to the matrices A_i, $i=1,\dots,m$, i.e. $\bigcup_{i=1}^m G_i$, where G_i is induced by A_i.

B. Optimization model

The task of the N agents consists in minimizing a convex function $f:\mathbb{R}^n\to\mathbb{R}$. The function f is expressed as a sum of N functions, i.e.

$$f(x)=\sum_{i=1}^N f_i(x), \quad (3)$$

where $f_i:\mathbb{R}^n\to\mathbb{R}$ are convex. Formally expressed, the agents want to cooperatively solve the following optimization problem

$$\min_{x\in\mathbb{R}^n}\sum_{i=1}^N f_i(x). \quad (4)$$

The fundamental assumption is that each agent i, has access only to the function f_i. Let f^* denote the optimal value of f and let X^* denote the set of optimizers of f, i.e. $X^*=\{x\in\mathbb{R}^n | f(x)=f^*\}$. Let $x_i(k)\in\mathbb{R}^n$ designate the *estimate of the optimal decision vector* of (4), maintained by agent i, at time k. The agents exchange estimates among themselves according to the communication topology described by the random graph $\mathbf{G}(k)$.

As proposed in [7], the agents update their estimates using a modified incremental subgradient method. Compared to the standard subgradient method, the local estimate $x_i(k)$ is replaced by a convex combination of $x_i(k)$ with the estimates received from the neighbors:

$$x_i(k+1)=\sum_{j=1}^N a_{ij}(k)x_j(k)-\alpha(k)d_i(k), \quad (5)$$

where $a_{ij}(k)$ is the $(i,j)^{th}$ entry of a stochastic random matrix $\mathbf{A}(k)$ which corresponds to the communication graph $\mathbf{G}(k)$. The matrices $\mathbf{A}(k)$ form an i.i.d random process taking values in a finite set of symmetric stochastic matrices with positive diagonal entries $\mathcal{A}=\{A_i\}_{i=1}^M$, where A_i is a stochastic matrix corresponding to the graph $G_i\in\mathcal{G}$, for $i=1,\dots,M$. The probability distribution of $\mathbf{A}(k)$ is inherited from $\mathbf{G}(k)$, i.e. $Pr(\mathbf{A}(k)=A_i)=Pr(\mathbf{G}(k)=G_i)=p_i$. The real valued scalar $\alpha(k)$ is the stepsize, while the vector $d_i(k)\in\mathbb{R}^n$ is a subgradient of f_i at $x_i(k)$, i.e. $d_i(k)\in\partial f_i(x_i(k))$.

Assumption 2.2: (Subgradient Boundedness and Constant Stepsize) The subgradients of function f_i at any point are bounded, i.e. there exists a scalar φ such that

$$\|d\|\leq\varphi, \forall d\in\partial f_i(x),\ \forall x\in\mathbb{R}^n,\quad i=1,\dots,N,$$

and the stepsize $\alpha(k)$ is assumed constant and known by all agents, i.e. $\alpha(k)=\alpha,\ \forall k\geq 0$.

Assumption 2.3: (Existence of an Optimal Solution) The optimal solution set X^* is nonempty.

III. Preliminary Results

In this section we lay the ground for our main results in Section IV. The preliminary results introduced here revolve around the idea of providing upper-bounds on a number of quantities of interest. The first quantity is represented by the distance between an estimate of the optimal decision and the average of all estimates. The second quantity is described by the distance between the average of the estimates and the set of optimizers.

We introduce the *average* vector of estimates of the optimal decision vector, denoted by $\bar{x}(k)$ and defined by

$$\bar{x}(k) \triangleq \frac{1}{N}\sum_{i=1}^{N} x_i(k). \tag{6}$$

The dynamic equation for the average vector can be derived from (5) and takes the form

$$\bar{x}(k+1) = \bar{x}(k) - \frac{\alpha(k)}{N} h(k), \tag{7}$$

where $h(k) = \sum_{i=1}^{N} d_i(k)$.

We introduce also the *deviation* of the local estimates $x_i(k)$ from the average estimate $\bar{x}(k)$, which is denoted by $z_i(k)$ and defined by

$$z_i(k) \triangleq x_i(k) - \bar{x}(k),\ i = 1,\dots,N. \tag{8}$$

Let us define the *aggregate* vectors of estimates, average estimates, deviations and subgradients, respectively:

$$\mathbf{x}(k)' \triangleq [x_1(k)', x_2(k)', \dots, x_N(k)'] \in \mathbb{R}^{Nn},$$
$$\bar{\mathbf{x}}(k)' \triangleq [\bar{x}(k)', \bar{x}(k)', \dots, \bar{x}(k)'] \in \mathbb{R}^{Nn},$$
$$\mathbf{z}(k)' \triangleq [z_1(k)', z_2(k)', \dots, z_N(k)'] \in \mathbb{R}^{Nn}$$

and

$$\mathbf{d}(k)' \triangleq [d_1(k), d_2(k), \dots, d_N(k)'] \in \mathbb{R}^{Nn}.$$

From (6) we note that the aggregate vector of average estimates can be expressed as

$$\bar{\mathbf{x}}(k) = J\mathbf{x}(k),$$

where $J = \frac{1}{N}\mathbb{1}\mathbb{1}' \otimes I$, with I the identity matrix in $\mathbb{R}^{n\times n}$ and $\mathbb{1}$ the vector of all ones in $\mathbb{R}^N$. Consequently, the aggregate vector of deviations can be written as

$$\mathbf{z}(k) = (I - J)\mathbf{x}(k). \tag{9}$$

The next Proposition characterize the dynamics of the vector $\mathbf{z}(k)$.

Proposition 3.1: The dynamic equation of the aggregate vector of deviations is given by

$$\mathbf{z}(k+1) = \mathbf{W}(k)\mathbf{z}(k) - \alpha(k)(I-J)\mathbf{d}(k),\ \mathbf{z}(0) = \mathbf{z}_0, \tag{10}$$

where $\mathbf{W}(k) = \left(\mathbf{A}(k) - \frac{1}{N}\mathbb{1}\mathbb{1}'\right)\otimes I$, with solution

$$\mathbf{z}(k) = \Phi(k,0)\mathbf{z}(0) - \sum_{s=0}^{k-1}\alpha(s)\Phi(k,s+1)\mathbf{d}(s), \tag{11}$$

where $\Phi(k,s)$ is the transition matrix of (10) defined by $\Phi(k,s) \triangleq \mathbf{W}(k-1)\mathbf{W}(k-2)\cdots\mathbf{W}(s)$, with $\Phi(k,k) = I$.

Remark 3.1: The transition matrix $\Phi(k,s)$ of the stochastic linear equation (10) can also be represented as

$$\Phi(k,s) = \left(\prod_{i=1}^{s}\mathbf{A}(k-i)\right)\otimes I - J, \tag{12}$$

where $J = \left(\frac{1}{N}\mathbb{1}\mathbb{1}'\right)\otimes I$. This comes from the fact that for any $i \in \{1,2,\dots,s-1\}$ we have that

$$(\mathbf{A}(k-i)\otimes I - J)(\mathbf{A}(k-i-1)\otimes I - J) = \\ = \mathbf{A}(k-i)\mathbf{A}(k-i-1)\otimes I - J.$$

Remark 3.2: (*On the first and second moments of the transition matrix* $\Phi(k,s)$) Let m be a positive integer and consider the transition matrix $\Phi(k+m,k) = \mathbf{W}(k+m-1)\dots\mathbf{W}(k)$, generated by a sequence of length m of random graphs, i.e. $\mathbf{G}(k)\dots\mathbf{G}(k+m-1)$, for some $k \geq 0$. The random matrix $\Phi(k+m,k)$ takes values of the form $W_{i_1}W_{i_2}\cdots W_{i_m}$, with $i_j \in \{1,2,\dots,M\}$ and $j = 1,\dots,m$. The norm of a particular realization of $\Phi(k+m,k)$ is given by the LEM of the matrix product $W_{i_1}W_{i_2}\cdots W_{i_m}$ or the SLEM of $A_{i_1}A_{i_2}\cdots A_{i_m}$, denoted henceforth by $\lambda_{i_1\dots i_m}$. Let $p_{i_1\dots i_m} = \prod_{j=1}^{m} p_{i_j}$ be the probability of the sequence of graphs $G_{i_1}\dots G_{i_m}$ that appear during the time interval $[k,k+m]$. Let I_m be the set of sequences of indices of length m for which the union of the graphs with the respective indices produces a connected graph, i.e. $I_m = \{i_1 i_2 \dots i_m | \bigcup_{j=1}^{m} G_{i_j} = \text{connected}\}$. Using the previous notations, the first and second moments of the norm of $\Phi(k+m,k)$ can be expressed as

$$E[\|\Phi(k+m,k)\|] = \eta_m, \tag{13}$$

$$E[\|\Phi(k+m,k)\|^2] = \rho_m, \tag{14}$$

where $\eta_m = \sum_{j\in I_m} p_j\lambda_j + 1 - \sum_{j\in I_m} p_j$ and $\rho_m = \sum_{j\in I_m} p_j\lambda_j^2 + 1 - \sum_{j\in I_m} p_j$. The integer j was used as an index for the elements of the set I_m.

The above formulas follow from results introduced in [4], Lemma 1, or in [15], Lemma 3.9, which state that for any sequence of indices $i_1\dots i_m \in I^{(m)}$, the matrix product $A_{i_1}\cdots A_{i_m}$ is ergodic, and therefore $\lambda_j < 1$, for $j \in I^{(m)}$. Conversely, if $j \notin I^{(m)}$, then $\lambda_j = 1$. We also note that $\sum_{j\in I^{(m)}} p_j$ is the probability of having a connected graph over a time interval of length m. Due to Assumption 2.1, for sufficiently large values of m, the set $I^{(m)}$ is nonempty. In fact for $m \geq M$, $I^{(m)}$ is always non-empty. In general for large values of m, it may be difficult to compute all eigenvalue λ_j, $j \in I^{(m)}$. We can omit the necessity of computing the eigenvalues λ_j, and this way decrease the computational burden, by using the following upper bounds on η_m and ρ_m

$$\eta_m \leq \lambda_m\mathbf{p}_m + 1 - \mathbf{p}_m, \tag{15}$$

$$\rho_m \leq \lambda_m^2\mathbf{p}_m + 1 - \mathbf{p}_m, \tag{16}$$

where $\lambda_m = \max_{j\in I_m}\lambda_j$ and $\mathbf{p}_m = \sum_{j\in I_m} p_j$ is the probability to have a connected graph over a time interval of length m. For notational simplicity, in what follows we will omit the index m when referring to the scalars η_m and ρ_m.

Throughout this note we will use the symbols m, η and ρ in the sense defined within the Remark 3.2. Moreover,

the value of m is chosen such that $I^{(m)}$ is nonempty. The existence of such a value is guaranteed by Assumption 2.1.

The next Proposition gives upper bounds on the expected value of the norm and the squared norm of the transition matrix $\Phi(k,s)$.

Proposition 3.2: Let $r \leq s \leq k$ be three nonnegative integer values. Let m be a positive integer, such that the set $I^{(m)}$ is non-empty. Then, the following inequalities involving the transition matrix $\Phi(k,s)$ of (10), hold

$$E[||\Phi(k,s)||] \leq \eta^{\left\lfloor \frac{k-s}{m} \right\rfloor}, \tag{17}$$

$$E[||\Phi(k,s)||^2] \leq \rho^{\left\lfloor \frac{k-s}{m} \right\rfloor}, \tag{18}$$

$$E[||\Phi(k,r)\Phi(k,s)||] \leq \rho^{\left\lfloor \frac{k-s}{m} \right\rfloor}\eta^{\left\lfloor \frac{s-r}{m} \right\rfloor}, \tag{19}$$

where η and ρ are defined in Remark 3.2.

The following lemma gives upper bounds on the first and the second moments of the distance between the estimate $x_i(k)$ and the average of the estimates, $\bar{x}(k)$.

Lemma 3.1: Under Assumption 2.2, for the sequences $\{x_i(k)\}_{k\geq 0}$, $i = 1,\dots,N$ generated by (5) with a constant stepsize α, the following inequalities hold

$$E[||x_i(k) - \bar{x}(k)||] \leq \beta\sqrt{N}\eta^{\left\lfloor \frac{k}{m} \right\rfloor} + \frac{m\alpha\varphi\sqrt{N}}{1-\eta} \tag{20}$$

$$E[||x_i(k) - \bar{x}(k)||^2] \leq N\beta^2\rho^{\left\lfloor \frac{k}{m} \right\rfloor} + N\alpha^2\varphi^2\left(1+\frac{2m}{1-\eta}\right)\frac{m}{1-\rho} + \\ +2N\alpha\beta\varphi m\frac{\rho^{\left\lfloor \frac{k-1}{m} \right\rfloor+1} - \eta^{\left\lfloor \frac{k-1}{m} \right\rfloor+1}}{\rho-\eta}, \tag{21}$$

where η, ρ and m are defined in Remark 3.2.

The following result allows us to interpret iteration (7) as an ε-subgradient (with ε being a random process).

Lemma 3.2: The vector $d_i(k)$ is an $\epsilon(k)$-subdifferential of f_i at $\bar{x}(k)$, i.e. $d_i(k) \in \partial_{\epsilon(k)} f_i(\bar{x}(k))$ and $\sum_{i=1}^N d_i(k)$ is an $N\epsilon(k)$-subdifferential of f at $\bar{x}(k)$, i.e. $\sum_{i=1}^N d_i(k) \in \partial_{N\epsilon(k)} f(\bar{x}(k))$, for any $k \geq 0$, where

$$\epsilon(k) = 2\varphi\beta\sqrt{N}||\Phi(k,0)|| + 2\alpha\varphi^2\sqrt{N}\sum_{s=0}^{k-1}||\Phi(k,s+1)||. \tag{22}$$

Under a strong convexity type of assumption on f, the next result gives an upper bound on the second moment of the distance between the average vector $\bar{x}(k)$ and the set of optimizers of f.

Lemma 3.3: Let $\{\bar{x}(k)\}_{k\geq 0}$ be a sequence of vectors generated by iteration (7). Also, assume that Assumptions 2.2 and 2.3 hold and that there exists a positive scalar μ such that

$$f(x) - f^* \geq \mu||x - X^*||. \tag{23}$$

Then, the following inequality holds

$$E[||\bar{x}(k) - X^*||^2] \leq ||\bar{x}(0) - X^*||^2\gamma^k + \frac{4\alpha\varphi\beta\sqrt{N}}{1-\gamma}\eta^{\left\lfloor \frac{k}{m} \right\rfloor} + \\ +\frac{\alpha^2\varphi^2}{1-\gamma}\left(\frac{4m\sqrt{N}}{1-\eta}+1\right), \tag{24}$$

where $\gamma = 1 - \frac{2\alpha\mu}{N}$ and η is defined in Remark 3.2.

IV. Main Result - Convergence analysis

In the following we provide upper bounds for three performance metrics of the MASM. First, we give an estimate on the radius of a neighborhood around the optimal valued f^*, where the cost function f, evaluated at the estimate $x_i(k)$, is guaranteed to converge. Second, we focus on the (squared) distance between the estimate $x_i(k)$ and the set of optimizers X^*. Under a strong convexity type of assumption, we give an estimate of the radius of a neighborhood around the zero point, where this metric is guaranteed to converge. We also provide an upper bound for the rate of convergence to the aforementioned neighborhood.

Corollary 4.1: Let Assumptions 2.1, 2.2 and 2.3 hold and let $\{x_i(k)\}_{k\geq 0}$ be a sequence generated by the iteration (5), $i = 1,\dots N$. Then

$$\liminf_{k\to\infty} E[f(x_i(k))] \leq f^* + \sqrt{N}\alpha\varphi^2\frac{m}{1-\eta}(N+2) + \frac{\alpha\varphi^2}{2} \tag{25}$$

Proof: Using the subgradient definition we have

$$f_i(\bar{x}(k)) \geq f_i(x_i(k)) + d_i(k)'(\bar{x}(k) - x_i(k)) \geq \\ \geq f_i(x_i(k)) - ||d_i(k)||||z_i(k)||,$$

or

$$f_i(x_i(k)) \leq f_i(\bar{x}(k)) + \varphi||z_i(k)||, \text{ for all } i = 1,\dots,N.$$

Summing over all i, we get

$$f(x_i(k)) \leq f(\bar{x}(k)) + N\varphi||\mathbf{z}(k)||.$$

By the results of Lemma 3.1, the following inequality holds

$$\liminf_{k\to\infty} E[f(x_i(k))] \leq \liminf_{k\to\infty} E[f(\bar{x}(k))] + N\sqrt{N}\alpha\varphi^2\frac{m}{1-\eta}. \tag{26}$$

Let $x^* \in X^*$ be an optimal point of f. By (7), where we use a constant stepsize α, we obtain

$$||\bar{x}(k+1) - x^*||^2 = ||\bar{x}(k) - x^* - \tfrac{\alpha}{N}h(\bar{x}(k))||^2 = \\ = ||\bar{x}(k) - x^*||^2 - 2\tfrac{\alpha}{N}h(\bar{x}(k))'(x(k) - x^*) + \alpha^2\varphi^2$$

and since, by Lemma 3.2, $h(\bar{x}(k))$ is a $N\epsilon(k)$-subdifferential of f at $\bar{x}(k)$, we have

$$||\bar{x}(k+1) - x^*||^2 \leq ||||\bar{x}(k) - x^*||^2 - 2\alpha(f(\bar{x}(k)) - f^*) + 2\alpha\varepsilon(k) + \alpha^2\varphi^2,$$

or

$$||\bar{x}(k+1) - x^*||^2 \leq ||\bar{x}(0) - x^*||^2 - 2\alpha\textstyle\sum_{s=0}^{k-1}(f(\bar{x}(s)) - f^*) + \\ +2\alpha\textstyle\sum_{s=0}^{k-1}\varepsilon(s) + k\alpha^2\varphi^2.$$

Since $||\bar{x}(k+1) - x^*||^2 \geq 0$

$$2\alpha\sum_{s=0}^{k-1}(f(\bar{x}(s)) - f^*) \leq ||\bar{x}(0) - x^*||^2 + 2\alpha\sum_{s=0}^{k-1}\varepsilon(s) + k\alpha^2\varphi^2,$$

or

$$2\alpha\sum_{s=0}^{k-1}(E[f(\bar{x}(s))] - f^*) \leq ||\bar{x}(0) - x^*||^2 + 2\alpha\sum_{s=0}^{k-1}E[\varepsilon(s)] + k\alpha^2\varphi^2.$$

By Proposition 3.1 and Lemma 3.1, we obtain the following upper bound for the expected value of $\epsilon(s)$.

$$E[\epsilon(s)] \leq 2\varphi\beta\sqrt{N}E[||\Phi(s,0)||] + 2\alpha\varphi^2\sqrt{N}\sum_{r=0}^{s-1}E[||\Phi(s,r+1)||] \leq$$

$$\leq 2\varphi\beta\sqrt{N}\eta^{\lfloor\frac{s}{m}\rfloor}+\frac{2\alpha\varphi^2\sqrt{N}m}{1-\eta},$$

which in turn leads to

$$\sum_{s=0}^{k-1}E[\epsilon(s)]\leq 2\varphi\beta\sqrt{N}\frac{m}{1-\eta}+k2\alpha\varphi^2\sqrt{N}\frac{m}{1-\eta}.$$

Using the fact that

$$\sum_{s=0}^{k-1}(E[f(\bar{x}(s))]-f^*)\geq k\min_{s=0,\ldots,k-1}(E[f(\bar{x}(s))]-f^*),$$

we get

$$\liminf_{k\to\infty}E[f(\bar{x}(k))]-f^*\leq 2\alpha\varphi^2\sqrt{N}\frac{m}{1-\eta}+\frac{\alpha\varphi^2}{2}. \quad (27)$$

Inequality (25) follows by combining (26) and (27). ■

The next result shows that under a strong convexity type of assumption, the convergence rate of the MASM, in expectation sense, is linear for a sufficiently small constant stepsize. It also shows that only convergence (in expectation sense) to a neighborhood can be guaranteed; a neighborhood, however, that can be made arbitrarily small.

Corollary 4.2: Let Assumptions 2.1, 2.2 and 2.3 hold and let μ be a positive scalar such that

$$f(x)-f^*\geq\mu\|x-X^*\|^2,\ \forall x\in\mathbb{R}^n. \quad (28)$$

Then, the sequence $\{x_i(k)\}_{k\geq 0}$, generated by iteration (20) with the stepsize $\alpha\leq\frac{N}{2\mu}$, converges, in expectation, (at least) R-linearly to a guaranteed neighborhood around some optimizer of f. The R-factor equals $\max\{\gamma,\eta^{\frac{1}{m}}\}$, where $\gamma=1-\frac{2\alpha\mu}{N}$ and the radius of the neighborhood equals

$$A+B+\sqrt{AB},$$

where

$$A=\frac{\alpha^2\varphi^2}{1-\gamma}\left(\frac{4m\sqrt{N}}{1-\eta}+1\right),$$

$$B=N\alpha^2\varphi^2\left(1+\frac{2m}{1-\eta}\right)\frac{m}{1-\rho}.$$

Proof: By the triangle inequality we have

$$\|x_i(k)-X^*\|\leq\|x_i(k)-\bar{x}(k)\|+\|\bar{x}(k)-X^*\|,$$

or

$$\|x_i(k)-X^*\|^2\leq\|x_i(k)-\bar{x}(k)\|^2+2\|x_(k)-\bar{x}(k)\|\|\bar{x}(k)-X^*\|+ +\|\bar{x}(k)-X^*\|^2.$$

or

$$E[\|x_i(k)-X^*\|^2]\leq E[\|x_i(k)-\bar{x}(k)\|^2]+ +2E[\|x_i(k)-\bar{x}(k)\|\|\bar{x}(k)-X^*\|]+E[\|\bar{x}(k)-X^*\|^2].$$

By the Cauchy-Schwarz inequality for the expectation operator, we get

$$E[\|x_i(k)-X^*\|^2]\leq E[\|x_i(k)-\bar{x}(k)\|^2]+2E[\|x_i(k)-\bar{x}(k)\|^2]^{\frac{1}{2}}\cdot \cdot E[\|\bar{x}(k)-X^*\|^2]^{\frac{1}{2}}+E[\|\bar{x}(k)-X^*\|^2]. \quad (29)$$

The guaranteed radius of the neighborhood around some optimizer of f follows by inequalities (21) and (24) and by taking the limit as k goes to infinity of the above inequality.

By inequality (21) we have

$$E[\|x_i(k)-\bar{x}(k)\|^2]\leq a_1\rho^{\frac{k}{m}}+a_2\eta^{\frac{k}{m}}+a_3,$$

where a_1, a_2 and a_3 are some positive scalars derived from the right-hand side of (21). By noting that $\eta\geq\rho$, we can further write

$$E[\|x_i(k)-\bar{x}(k)\|^2]\leq\tilde{a}_1\eta^{\frac{k}{m}}+\tilde{a}_2 \quad (30)$$

where $\tilde{a}_1$ and $\tilde{a}_2$ are some positive scalars.

By inequality (24), we obtain

$$E[\|\bar{x}(k)-X^*\|^2]\leq b_1\gamma^k+b_2\eta^{\frac{k}{m}}+b_3,$$

where b_1, b_2 and b_3 are some positive scalars derived from the right-hand side of (24). We can further write

$$E[\|\bar{x}(k)-X^*\|^2]\leq\tilde{b}_1\max\{\gamma,\eta^{\frac{1}{m}}\}^k+\tilde{b}_2, \quad (31)$$

where $\tilde{b}_1$ and $\tilde{b}_2$ are some positive scalars. Using the notations $c_1=\max\{\tilde{a}_1,\tilde{b}_1\}$ and $c_2=\max\{\tilde{a}_2,\tilde{b}_2\}$, by (29), (30) and (31), we obtain

$$E[\|x_i(k)-X^*\|^2]\leq 4c1\max\{\gamma,\eta^{\frac{1}{m}}\}^k+4c_2,$$

which shows the R-linear convergence, with the R-factor given by $\max\{\gamma,\eta^{\frac{1}{m}}\}$. ■

A. Discussion of the results

We obtained upper bounds on three performance metrics relevant to the MASM: the distance between the cost function evaluated at the estimate and the optimal solution (Corollary 4.1), the distance between the estimate of the decision vector and the set of optimizers and the rate of convergence to some neighborhood around an optimizer of f (Corollary 4.2). The three upper bounds are functions of three quantities which depend on the scalars m, η and ρ, i.e. $\frac{m}{1-\eta}$, $\frac{m}{1-\rho}$ and $\eta^{\frac{1}{m}}$, which show the dependence of the performance metrics on the pmf of $\mathbf{G}(k)$ and on the corresponding random matrix $\mathbf{A}(k)$. The scalars η and ρ represent the first and second moments of the SLEM of the random matrix $\mathbf{A}(k+1)\ldots\mathbf{A}(k+m)$, corresponding to a random graph formed over a time interval of length m, respectively. We notice from our results that the performance of the MASM is improved by making $\frac{m}{1-\eta}$, $\frac{m}{1-\rho}$ and $\eta^{\frac{1}{m}}$ as small as possible, i.e. by optimizing these quantities having as decision variables m and the pmf of $\mathbf{G}(k)$. Since the three quantities are not necessarily optimized by the same values of the decision variables, we have in fact a multi-criteria optimization problem:

$$\begin{array}{ll}\min_{m,p_i} & \{\frac{m}{1-\eta},\frac{m}{1-\rho},\eta^{\frac{1}{m}}\}\\ \text{subject to:} & m\geq 1\\ & \sum_i^M p_i=1,\ p_i\geq 0.\end{array} \quad (32)$$

The scalar $\eta^{\frac{1}{m}}$ relates to the rate of convergence of the distance between the estimate of the decision vector and the set of optimizers. Note that, unless it helps the other two quantities $\frac{m}{1-\eta}$ and $\frac{m}{1-\rho}$, making $\eta^{\frac{1}{m}}$ too small, may

not necessarily result in an improvement of the rate of convergence since the latter is given by the $\max\{\gamma, \eta^{\frac{1}{m}}\}$.

The solution to the above problem is a set of Pareto points, i.e. solution points for which improvement in one objective can only occur with the worsening of at least one other objective.

We note that for each fixed value of m, the three quantities are minimized if the scalars η and ρ are minimized as functions of the pmf of the random graph. In fact, we can focus only on minimizing η, since for every fixed m, η is an upper bound on ρ and in fact both quantities are minimized by the same pmf. Therefore, in problem (32), we have to find an appropriate value of m such that a Pareto solution is obtained, which has a corresponding optimal pmf. Depending on the communication model used, the pmf of the random graph can be a quantity dependent on a set of parameters of the communication protocol (transmission power, probability of collisions, etc). Having an *optimal* pmf allow us to tune these parameters such that the performance of the MASM is improved.

In what follows we provide a simple example where we show how the optimal probability distribution, η, $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$ evolve as functions of m.

Example 4.1: Let $\mathbf{G}(k)$ be random graph taking values in the set $\mathcal{G} = \{G_1, G_2\}$,with probability p and $1-p$, respectively. The graphs G_1 and G_2 are shown in Figure 1. Also, let $\mathbf{A}(k)$ be a (stochastic) random matrix , corresponding to $\mathbf{G}(k)$, taking value in the set $\mathcal{A} = \{A_1, A_2\}$, with

$$A_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \; A_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

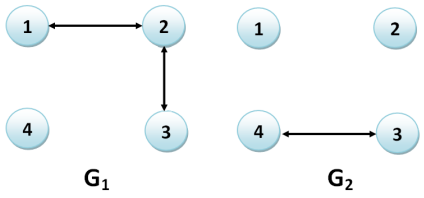
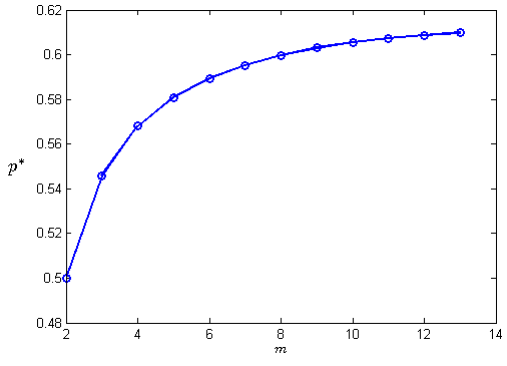


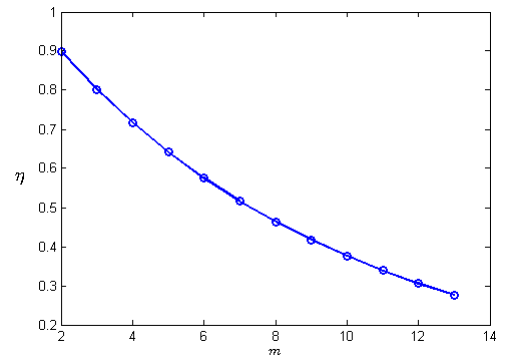
Fig. 1. The sample space of the random graph $\mathbf{G}(k)$

Figure 2(a) shows the optimal probability p^* that minimizes η for different values of m. Figure 2(b) shows the *optimized* η (computed at p^*) as a function of m. Figures 2(c) and 2(d) show the evolution of the *optimized* $\frac{m}{1-\eta}$ and $\eta^{\frac{1}{m}}$ as functions of m, from where we notice that a Pareto solution is obtained for $m = 5$ and $p^* = 0.582$.

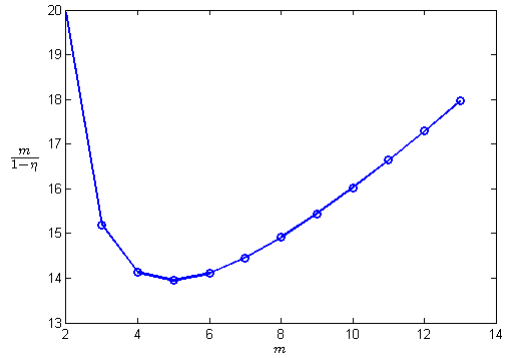
In order to obtain the solution of problem (32), we need to compute the probability of all possible sequences of length m produced by $\mathbf{G}(k)$, together with the SLEM of their corresponding stochastic matrices. This task, for large values of m and M may prove to be numerically expensive. We can somewhat simplify the computational burden by using instead the bounds on η and ρ introduced in (15) and (16), respectively. Note that every result concerning



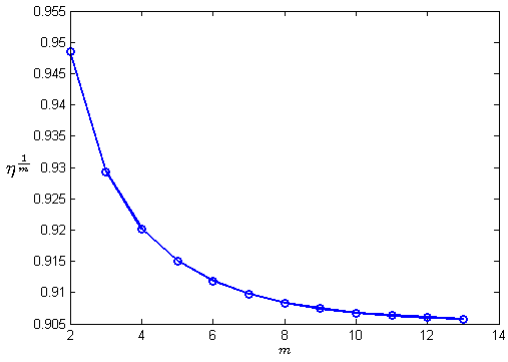
(a)



(b)



(c)



(d)

Fig. 2. (a) Optimal p as a function of m; (b) Optimized η as a function of m; (c) Optimized $\frac{m}{1-\eta}$ as a function of m; (d) Optimized $\eta^{\frac{1}{m}}$ as a function of m

the performance metrics still holds. In this case, for each value of m, the upper bound on η is minimized, when $\mathbf{p}_m$ is maximized, which can be interpreted as having to choose a pmf that maximizes the probability of connectivity of a random graph obtained over a time interval of length m. In our example, the probability that the union of m consecutive graphs produced by $\mathbf{G}(k)$ results in a connected graph is $\mathbf{p}_m = 1 - p^m - (1-p)^m$. We note that $\mathbf{p}_1 = 0$, since no single graph in $\mathcal{G}$ is connected. For every fixed value of m, $\mathbf{p}_m$ is maximized for $p = \frac{1}{2}$. Given that $\mathbf{A}(k)$ takes values A_1 and A_2 with uniform distribution, the bounds on the curves of the quantities η, $\eta^{\frac{1}{m}}$ and $\frac{m}{1-\eta}$ are given in Figure 3. We note that for $m = 3$, both bounds on $\eta^{\frac{1}{m}}$ and $\frac{m}{1-\eta}$ are minimized.

Even in the case where we try to minimize the bound on η, it may be very difficult to compute the expression for $\mathbf{p}_m$, for large values of m (the set $\mathcal{G}$ may allow for a large number of possible unions of graphs producing connected graphs). Another way to simplify even more our problem, is to (intelligently) fix a value for m and to try to maximize $\mathbf{p}_m$ having as decision variable the pmf. We note that m should be chosen such that, within a time interval of length m, a connected graph can be obtained. Also, a very large value for m should be avoided, since $\frac{m}{1-\eta}$ is lower bounded by m. Although in general the uniform distribution is not necessarily minimizing η, it becomes the optimizer under some particular assumptions, stated in what follows. Let $\mathcal{G}$ be such that a connected graph is possible to be obtained only over a time interval of length M (i.e. in order to form a connected graph, all graphs in $\mathcal{G}$ must appear within a sequence of length M). Choose M as value for m. It follows that $\mathbf{p}_m$ can be expressed as:

$$\mathbf{p}_m = m! \prod_{i=1}^{M} p_i.$$

We can immediately observe that $\mathbf{p}_m$ is maximized for the uniform distribution, i.e. $p_i = \frac{1}{m}$, for $i = 1, \ldots, M$.

V. CONCLUSIONS

In this note we studied a multi-agent subgradient method under random communication topology. Under an i.i.d. assumption on the random process governing the evolution of the topology we derived upper bounds on three performance metrics related to the MASM. The first metric is given by the radius of the neighborhood around the optimal solution where the cost function evaluated at an estimate converges. The second and the third metrics are represented by the radius of a neighborhood around the zero point where the distance between an estimate and the set of optimizers is guaranteed to converge and the rate of convergence to this neighborhood, respectively. All the aforementioned performance measures were expressed in terms of the probability distribution of the random communication topology. This is particulary useful when the distributed optimization is performed over wireless networks, since the communication protocol parameters, which determine the probability distribution of the random graph, can be tailored to improve the performance of the MASM.

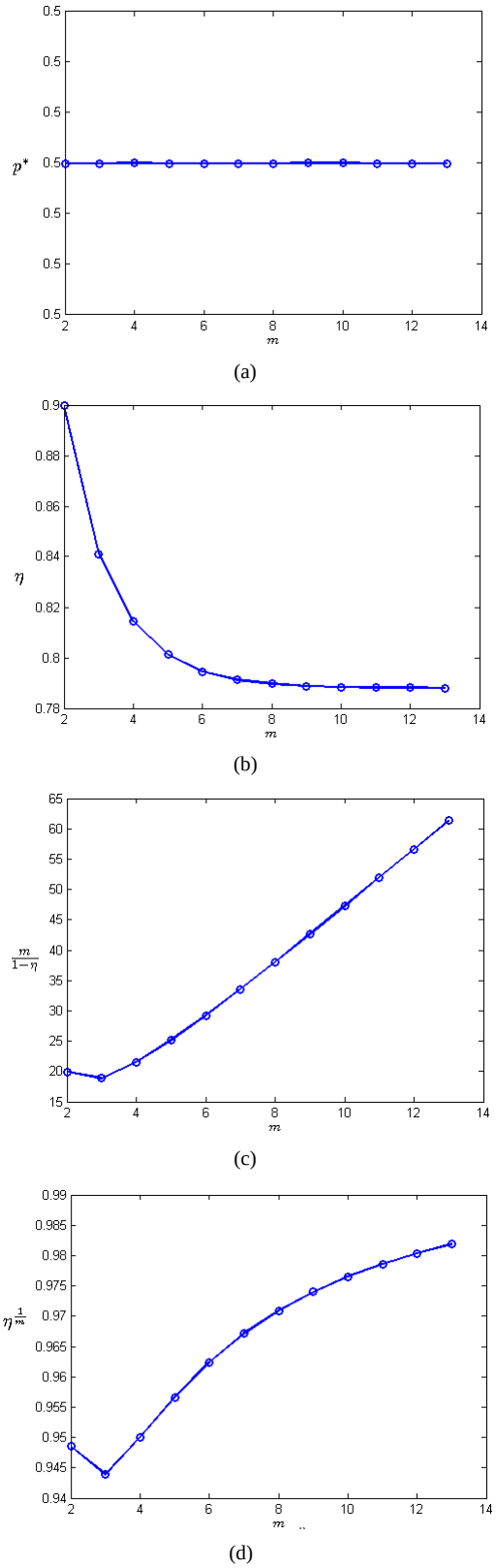


Fig. 3. (a) Optimal p as a function of m; (b) Optimized bound on η as a function of m ; (c) Optimized bound on $\frac{m}{1-\eta}$ as a function of m; (d) Optimized bound on $\eta^{\frac{1}{m}}$ as a function of m

References

[1] J.S. Baras and P. Hovareshti, "Effects of Topology in Networked Systems: Stochastic Methods and Small Worlds," *Proceedings of the 47th IEEE Conference on Decision and Control Cancun*, pp. 2973-2978, Mexico, Dec. 9-11, 2008

[2] J.S. Baras, T. Jiang and P. Purkayastha, "Constrained Coalitional Games and Networks of Autonomous Agents," *ISCCSP 2008*, pp. 972-979, Malta, 12-14 March 2008

[3] S. Boyd, P. Diaconis, and L. Xiao, "Fastest Mixing Markov Chain on a Graph,", *SIAM Review*, vo. 46, no. 4, pp. 667-689, December 2004

[4] A. Jadbabaie, J. Lin and A.S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor", *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 998-1001, Jun 2003.

[5] B. Johansson, T. Keviczky, M. Johansson, K.H. Johansson, "Subgradient Methods and Consensus Algorithms for Solving Convex Optimization Problems", *Proceedings of the* 17^{th} *Conference on Decision and Control*, Mexic, December 2008.

[6] B. Johansson, "On Distributed Optimization in Network Systems," *Doctoral Thesis in Telecommunication*, Stockholm, Sweden, 2008.

[7] A. Nedic and A. Ozdalgar, "Distributed Subgradient Methods for Multi-Agent Optimization", *IEEE Trans. Automatic Control*, vol. 54, no. 1, pp. 48-61, January 2009.

[8] I. Lobel and A. Ozdalgar, "Convergence analysis of Distributed Subgradients Methods over Random Netwroks, *Proceedings* 46^{th} *Allerton Conference on Communication, Control and Computing*, pp. 353-360, September 2008.

[9] S. Kandukuri and S. Boyd, " Optimal Power Control in Interference-Limited Fading Wireless Channels With Outage-Probability Specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, January 2002.

[10] I. Matei and J.S. Baras, "Distributed subgradient method under random communication topology - the effect of the probability distribution of the random graph on the performance metrics," *ISR Technical Report*, TR-2009-15, Sept. 2009, http://hdl.handle.net/1903/9441.

[11] A. Nedic, "Convergence Rate of Incremental Subgradient Algorithm", *Stochastic Optimization: Algorithms and Applications*, pp. 263-304, Kluwer Academic Publisher.

[12] A. Nedic and D.P. Bertsekas, "Incremental Subgradient Methods for Nondifferential Optimization", *SIAM Journal on Optimization*, vol. 12, issue 1, pp. 109 - 138, 2001.

[13] M.J. Neely, E. Modiano and C.E. Rohrs, "Dynamic Power Allocation and Routing for Time-Varying Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89-103, January 2005

[14] C.T.K. Ng, M. Medard and A. Ozdaglar, "Completion Time Minimization and Robust Power Control in Wireless Packet Networks," arXiv:0812.3447v1 [cs.IT] 18 Dec 2008

[15] W. Ren and R.W. Beard, "Consensus seeking in multi-agents systems under dynamically changing interaction topologie," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655-661, May 2005.

[16] D.J. Watts and S.H Strogatz, "Collective Dynamics of Small World Networks," *Nature*, vol. 393, pp. 440-442, 1998.