THE INSTITUTE FOR SYSTEMS RESEARCH

ISR TECHNICAL REPORT 2009-15

Distributed subgradient method under random communication topology - the effect of the probability distribution of the random graph on the performance metrics

Ion Matei, John S. Baras

Distributed subgradient method under random communication topology - the effect of the probability distribution of the random graph on the performance metrics

Ion Matei and John S. Baras

Abstract

In this note we study the performance metrics (rate of convergence and guaranteed region of convergence) of a multi-agent subgradient method for optimizing a sum of convex functions. We assume that the agents exchange information according to a communication topology modeled as a random graph, independent of other time instances. Under a strong convexity type of assumption, we express the performance metrics directly as functions of the estimates of the optimal decision vector. We emphasize how the probability distribution of the random graph affects the upper bounds on the performance metrics. This provide a guide for tuning the parameters of the communication protocol such that good performance of the multi-agent subgradient method is ensured. We perform the tuning of the protocol parameters for two communication scenarios. In the first scenario, we assume a randomized scheme for link activation with no-error transmissions while in the second scenario we use a pre-established order of transmissions but we consider the interference effect. Both these scenarios are applied on a small world type of topology.

I. Introduction

Multi-agents 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 can exchange information according to a communication topology, modeled as an undirected, time varying graph, which emphasizes the communication neighborhoods of the agents. The agents maintain *estimates* of the optimal decision vector, which are updated in two stages. 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 study the performance metrics (rate of convergence and guaranteed region of convergence) of the multi-agent subgradient method proposed in [7], for the case of a constant stepsize. The communication among agents is modeled by a random graph, independent of other time instances, and the performance metrics in this note will be viewed in expectation sense. Random graphs are suitable models for networks that changes with time due to link failures, packet drops, node failure, etc. An analysis of the multi-agent subgradient method under random communication topology is addressed in [8]. The authors provide upper bounds on the performance metrics as functions of a lower bound of the consensus step weights and other parameters of the problem. 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 give 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 enumerated above are in expectation sense. The explicit dependence on the probability distribution allows us to determine the optimal probability distributions that would ensure the best guaranteed upper bounds on the performance metric. 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, we address two 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 role of protocol parameters. Both scenarios are applying 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. $\|y-X\| \triangleq \min_{x\in X}\|y-x\|$.

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:

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

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:

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

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. By $\|\cdot\|$ we understand the standard two norm.

Paper structure: The second Section 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. Section V, shows how our results can be used to tune the parameters of a communication protocol, such that the performance of the MASM are improved.

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 corresponds 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 topology, i.e. $\mathbf{G}(k)$ is undirected.

Assumption 2.1: (Connectivity assumption) The graph resulted 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* the graph G or the graph G is *induced* by A if and only if any non-zero entry (ij) of A implies a link from j to i in G and vice versa. Consider a matrix products 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, $i = 1,\dots,m$.

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.

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

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

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

The fundamental assumption is that each agent i, has access only to the function f_i. Let f^* denote 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:

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

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 matrix $\mathbf{A}(k)$ is 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, 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 the quantities. 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 the aggregate vector of average estimates is expressible 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

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

where $\mathbf{W}(k) = (\mathbf{A}(k) - \frac{1}{N}\mathbb{1}\mathbb{1}') \otimes I$, 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 $\Phi(k,k) = I$.

Proof: From (5) the dynamics of the aggregate vector of estimates is given by

$$\mathbf{x}(k+1) = (\mathbf{A}(k) \otimes I)\mathbf{x}(k) - \alpha(k)\mathbf{d}(k). \tag{12}$$

From (9) together with (12), we can further write

$$\mathbf{z}(k+1) = (I-J)\mathbf{x}(k+1) = (\mathbf{A}(k)\otimes I - J)\mathbf{x}(k) - \alpha(k)(I-J)\mathbf{d}(k).$$

By noting that

$$(\mathbf{A}(k)\otimes I - J)\mathbf{z}(k) = (\mathbf{A}(k)\otimes I - J)(I-J)\mathbf{x}(k) = (\mathbf{A}(k)\otimes I - J)\mathbf{x}(k),$$

we obtain (10). Solution (11) follows from (10) together with the observation that $\Phi(k,s)(I-J) = \Phi(k,s)$. ∎

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

$$\Phi(k,s) = \left(\prod_{i=1}^{s}\mathbf{A}(k-i)\right)\otimes I - J, \tag{13}$$

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

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

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 $\Phi(k+m,k) = \mathbf{W}(k+m-1)\ldots\mathbf{W}(k)$, generated by a sequence of length m of random graphs, i.e. $\mathbf{G}(k)\ldots\mathbf{G}(k+m-1)$, 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 $i_j \in \{1,2,\ldots,M\}$ and $j = 1,\ldots,m$. 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 $\lambda_{i_1\ldots i_m}$. Let $p_{i_1\ldots i_m} = \prod_{j=1}^{m} p_{i_j}$ be the probability of the sequence of graphs $G_{i_1}\ldots G_{i_m}$ to appear during the time interval $[k,k+m]$. Define the set of sequences of length m of indices $\mathcal{I}^{(m)} = \{i_1i_2\ldots i_m | i_j \in \{1,2,\ldots,M\}, j = 1,\ldots,m\}$ with cardinality m^M and let $I^{(m)} = \{i_1i_2\ldots i_m | \bigcup_{j=1}^{m} G_{i_j} = \text{connected}\}$, i.e. the set of sequences of indices of length m for which the corresponding union of graphs produces a connected graph. Using the previous notations, the first and second moments of the norm of $\Phi(k+m,k)$ can be expressed as

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

$$E[\|\Phi(k+m,k)\|^2] = \rho_m, \tag{15}$$

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 set $\mathcal{I}^{(m)}$.

The above formulas follow from results introduced in [4], Lemma 1, or in [14], Lemma 3.9, which state that for any sequence of indices $i_1 \ldots i_m \in I^{(m)}$, the product matrix $A_{i_1} \cdots A_{i_m}$ is ergodic, and therefore $\lambda_j < 1,\ j \in I^{(m)}$. 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 valued 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

$$\eta_m \leq \lambda_m \mathbf{p}_m + 1 - \mathbf{p}_m, \tag{16}$$

$$\rho_m \leq \lambda_m^2 \mathbf{p}_m + 1 - \mathbf{p}_m, \tag{17}$$

where $\lambda_m = \max_{i\in I^{(m)}} \lambda_j$ 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 in 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 picked such that $I^{(m)}$ is nonempty. The existence of such values is guaranteed by Assumption 2.1.

The next Proposition gives upper bounds on a series of quantities involving 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

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

$$E[\|\Phi(k,s)\|^2] \leq \rho^{\left\lfloor \frac{k-s}{m} \right\rfloor}, \tag{19}$$

$$E[\|\Phi(k,r)\Phi(k,s)\|] \leq \rho^{\left\lfloor \frac{k-s}{m} \right\rfloor} \eta^{\left\lfloor \frac{s-r}{m} \right\rfloor}, \tag{20}$$

where η and ρ are defined in Remark 3.2.

Proof: First we note that by Assumption 2.1 there exist values for m such that the set $I^{(m)}$ is non-empty (choose $m \geq M$). Consequently, the probability to have a connected graph over a time interval of length m is positive. Let t be the number of intervals of length m between s and k, i.e.

$$t = \left\lfloor \frac{k-s}{m} \right\rfloor,$$

and let $s_0, s_1, \ldots, s_t$ be a sequence of nonnegative integers such that $s = s_0 < s_1 < \ldots < s_t \leq k$ where $s_{i+1} - s_i = m$ and $i = 0, \ldots, m-1$. By the semigroup property of transition matrices, it follows that

$$\Phi(k,s) = \Phi(k,s_t)\Phi(s_t,s_{t-1})\cdots\Phi(s_1,s),$$

or

$$||\Phi(k,s)|| \leq ||\Phi(s_t,s_{t-1})||\cdots||\Phi(s_1,s)||,$$

where we use the fact that $||\Phi(k,s_t)|| \leq 1$. Using the i.i.d. assumption on the random process $\mathbf{A}(k)$, we can further write

$$E[||\Phi(k,s)||] \leq E[||\Phi(s_t,s_{t-1})||]\cdots E[||\Phi(s_1,s)||],$$

which together with (14) leads to inequality (18).

Similarly, inequality (19) follow from (15) and from the i.i.d. assumption on the random graph.

We now turn to inequality (20). By the semigroup property we get

$$E[||\Phi(k,r)\Phi(k,s)||] \leq E[||\Phi(k,s)||^2||\Phi(s,r)||] \leq E[||\Phi(k,s)||^2]E[||\Phi(s,r)||],$$

where the second inequality followed by independence of $\mathbf{A}(k)$. Inequality (20) follows from (18) and (19). ■

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

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

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

$$E[||x_i(k) - \bar{x}(k)||^2] \leq N\beta^2\rho^{\left\lfloor \frac{k}{m} \right\rfloor} + N\alpha^2\varphi^2\left(1 + \frac{2m}{1-\eta}\right)\frac{m}{1-\rho} + 2N\alpha\beta\varphi m\frac{\rho^{\left\lfloor \frac{k-1}{m} \right\rfloor+1} - \eta^{\left\lfloor \frac{k-1}{m} \right\rfloor+1}}{\rho-\eta}, \tag{22}$$

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

Proof: Note that the norm of the deviation $z_i(k) = x_i(k) - \bar{x}(k)$ is upper bounded by the norm of the aggregate vector of deviations $\mathbf{z}(k)$ (with probability one), i.e. $\|z_i(k)\| \leq \|\mathbf{z}(k)\|$. Hence, by Preposition 3.1, we have

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

or

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

where use the fact that $\|z_i(0)\| \leq \beta$ and $\|d_i(k)\| \leq \varphi$, $\forall k \geq 0$.

By inequality (18) of Proposition 3.2, we get

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

By noting that the sum $\sum_{s=0}^{k-1} \eta^{\lfloor \frac{k-s-1}{m} \rfloor}$ can be upper bounded by

$$\sum_{s=0}^{k-1} \eta^{\lfloor \frac{k-s-1}{m} \rfloor} \leq m \sum_{s=0}^{\lfloor \frac{k-1}{m} \rfloor} \eta^s = m \frac{1 - \eta^{\lfloor \frac{k-1}{m} \rfloor + 1}}{1 - \eta} \leq m \frac{1}{1 - \eta},$$

inequality (21) follows.

We now turn to obtaining an upper bound on the second moment of $\|\mathbf{z}(k)\|$.

Let $\mathbf{Z}(k) \in \mathbb{R}^{Nn \times Nn}$ be a symmetric positive definite matrix, given by

$$\mathbf{Z}(k) = \mathbf{z}(k)\mathbf{z}(k)'.$$

Using Preposition 3.1, we derive from the following equation for $\mathbf{Z}(k)$:

$$\mathbf{Z}(k+1) = \mathbf{W}(k)\mathbf{Z}(k)\mathbf{W}(k)' + F(k), \tag{23}$$

where $F(k)$ is given by

$$F(k) = \alpha^2 (I - J)\mathbf{d}(k)\mathbf{d}(k)'(I - J)' - \alpha W(k)\mathbf{z}(k)\mathbf{d}(k)'(I - J)' - \alpha(I - J)\mathbf{d}(k)\mathbf{z}(k)'W(k)'. \tag{24}$$

The solution of (23) is given by

$$\mathbf{Z}(k) = \Phi(k,0)\mathbf{Z}(0)\Phi(k,0)' + \sum_{s=0}^{k-1} \Phi(k, s+1)F(s)\Phi(k, s+1)'. \tag{25}$$

For simplicity, in what follows we will omit the matrix $I-J$ from terms of $F(k)$ since they disappear by multiplication with the transition matrix (see Proposition 3.1). We can further write

$$||\mathbf{Z}(k)|| \leq ||\Phi(k,0)||^2||\mathbf{Z}(0)|| + \sum_{s=0}^{k-1} ||\Phi(k,s+1)F(s)\Phi(k,s+1)'||$$

and by noting that $||\mathbf{Z}(k)|| = ||\mathbf{z}(k)||^2$, we obtain

$$E[||\mathbf{z}(k)||^2] \leq E[||\Phi(k,0)||^2]||\mathbf{z}(0)||^2 + \sum_{s=0}^{k-1} E[||\Phi(k,s+1)F(s)\Phi(k,s+1)'||]. \tag{26}$$

From (19) of Proposition 3.2 we obtain

$$E[||\Phi(k,0)||^2] \leq \rho^{\lfloor \frac{k}{m} \rfloor}$$

We now focus on the terms of the above sum in (26). We have

$$\Phi(k,s+1)F(s)\Phi(k,s+1)' = \alpha^2\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(s)'\Phi(k,s+1)' -$$

$$-\alpha\Phi(k,s+1)W(s)\mathbf{z}(s)\mathbf{d}(s)'\Phi(k,s+1)' - \alpha\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(s)'W(s)'\Phi(k,s+1)'$$

Using the solution of $\mathbf{z}(k)$ given in (11), we get

$$\Phi(k,s+1)W(s)\mathbf{z}(s)\mathbf{d}(s)'\Phi(k,s+1)' =$$

$$= \Phi(k,s+1)W(s)\left(\Phi(s,0)\mathbf{z}(0) - \alpha\sum_{r=0}^{s-1}\Phi(s,r+1)\mathbf{d}(r)\right)\mathbf{d}(s)'\Phi(k,s+1)'$$

$$= \Phi(k,0)\mathbf{z}(0)\mathbf{d}(s)'\Phi(k,s+1)' - \alpha\sum_{r=0}^{s-1}\Phi(k,r+1)\mathbf{d}(r)\mathbf{d}(s)'\Phi(k,s+1)' \tag{27}$$

Similarly,

$$\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(s)'W(s)'\Phi(k,s+1)' =$$

$$\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(0)'\Phi(k,0)' - \alpha\sum_{r=0}^{s-1}\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(r)'\Phi(k,r+1)'. \tag{28}$$

We now give a more explicit formula of $\Phi(k,s+1)F(s)\Phi(k,s+1)'$:

$$\Phi(k,s+1)F(s)\Phi(k,s+1)' = \alpha^2\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(s)'\Phi(k,s+1)' -$$

$$-\alpha\Phi(k,0)\mathbf{z}(0)\mathbf{d}(s)'\Phi(k,s+1) + \alpha^2\sum_{r=0}^{s-1}\Phi(k,r+1)\mathbf{d}(r)\mathbf{d}(s)'\Phi(k,s+1)' -$$

$$-\alpha\Phi(k,s+1)\mathbf{d}(s)\mathbf{z}(0)'\Phi(k,0)' + \alpha^2\sum_{r=0}^{s-1}\Phi(k,s+1)\mathbf{d}(s)\mathbf{d}(r)'\Phi(k,r+1)'$$

By applying the norm operator, we get

$$||\Phi(k,s+1)F(s)\Phi(k,s+1)'|| \leq N\alpha^2\varphi^2||\Phi(k,s+1)||^2+$$
$$+N\alpha^2\varphi^2\sum_{r=0}^{s-1}||\Phi(k,r+1)\Phi(k,s+1)'||+N\alpha^2\varphi^2\sum_{r=0}^{s-1}||\Phi(k,s+1)\Phi(k,r+1)'||+$$
$$+N\alpha\beta\varphi||\Phi(k,s+1)\Phi(k,0)'||+N\alpha\beta\varphi||\Phi(k,0)\Phi(k,s+1)'||$$

or

$$||\Phi(k,s+1)F(s)\Phi(k,s+1)'|| \leq N\alpha^2\varphi^2||\Phi(k,s+1)||^2+$$
$$+2N\alpha^2\varphi^2\sum_{r=0}^{s-1}||\Phi(k,r+1)\Phi(k,s+1)'||+2N\alpha\beta\varphi||\Phi(k,s+1)\Phi(k,0)'|| \tag{29}$$

We now give bounds for the expected values of each of the terms in (29). Based on the results of Proposition 3.2 we can write

$$E[||\Phi(k,s+1)||^2] \leq \rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}$$
$$\sum_{r=0}^{s-1}E[||\Phi(k,r+1)\Phi(k,s+1)'||] \leq \sum_{r=0}^{s-1}\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\eta^{\left\lfloor\frac{s-r}{m}\right\rfloor} \leq m\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\sum_{r=0}^{\left\lfloor\frac{s}{m}\right\rfloor}\eta^r \leq$$
$$\leq m\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\frac{1-\eta^{\left\lfloor\frac{s}{m}\right\rfloor+1}}{1-\eta} \leq m\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\frac{1}{1-\eta}$$

and

$$E[||\Phi(k,s+1)\Phi(k,0)'||] \leq \rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\eta^{\left\lfloor\frac{s+1}{m}\right\rfloor}$$

Therefore we obtain

$$E[||\Phi(k,s+1)F(s)\Phi(k,s+1)'||] \leq N\alpha^2\varphi^2\left(1+\frac{2m}{1-\eta}\right)\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}+2N\alpha\beta\varphi\rho\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\eta^{\left\lfloor\frac{s+1}{m}\right\rfloor}.$$

We know compute an upper bound for $\sum_{s=0}^{k-1}E[||\Phi(k,s+1)F(s)\Phi(k,s+1)'||]$. Using the fact that

$$\sum_{s=0}^{k-1}\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor} \leq m\sum_{s=0}^{\left\lfloor\frac{k-1}{m}\right\rfloor}\rho^s \leq m\frac{1-\rho^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}}{1-\rho} \leq m\frac{1}{1-\rho}$$

and

$$\sum_{s=0}^{k-1}\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\eta^{\left\lfloor\frac{s+1}{m}\right\rfloor} \leq \sum_{s=0}^{k-1}\rho^{\left\lfloor\frac{k-s-1}{m}\right\rfloor}\eta^{\left\lfloor\frac{s}{m}\right\rfloor} \leq$$
$$\leq m\sum_{s=0}^{\left\lfloor\frac{k-1}{m}\right\rfloor}\rho^{\left\lfloor\frac{k-1}{m}\right\rfloor-s}\eta^s = m\frac{\rho^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}-\eta^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}}{\rho-\eta},$$

we obtain

$$\sum_{s=0}^{k-1} E[||\Phi(k,s+1)F(s)\Phi(k,s+1)'||] \leq N\alpha^2\varphi^2\left(1+\frac{2m}{1-\eta}\right)\frac{m}{1-\rho}+$$

$$+2N\alpha\beta\varphi m\frac{\rho^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}-\eta^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}}{\rho-\eta}.$$

Finally we obtain an upper bound for the second moment of $||\mathbf{z}(k)||$:

$$E[||\mathbf{z}(k)||^2] \leq N\beta^2\rho^{\left\lfloor\frac{k}{m}\right\rfloor}+N\alpha^2\varphi^2\left(1+\frac{2m}{1-\eta}\right)\frac{m}{1-\rho}+2N\alpha\beta\varphi m\frac{\rho^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}-\eta^{\left\lfloor\frac{k-1}{m}\right\rfloor+1}}{\rho-\eta}.$$

■

The following result allow 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

$$\epsilon(k) = 2\varphi\beta\sqrt{N}||\Phi(k,0)|| + 2\alpha\varphi^2\sqrt{N}\sum_{s=0}^{k-1}||\Phi(k,s+1)||. \tag{30}$$

Proof: The proof is somewhat similar to the proof of Lemma 3.4.5 of [6]. By the subgradient definition we have that

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

or

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

Farther more, for any $y \in \mathbb{R}^n$ we have that

$$f_i(y) \geq f_i(x_i(k)) + d_i(k)'(y-x_i(k)) = f_i(x_i(k)) + d_i(k)'(y-\bar{x}(k)+\bar{x}(k)-x_i(k)) \geq$$

$$\geq f_i(\bar{x}(k)) + d_i(k)'(y-\bar{x}(k)) - 2||d_i(k)||||z_i(k)|| \geq f_i(\bar{x}(k)) + d_i(k)'(y-\bar{x}(k)) - 2\varphi||\mathbf{z}(k)||,$$

or

$$f_i(y) \geq f_i(\bar{x}(k)) + d_i(k)'(y-\bar{x}(k)) - \varepsilon(k),$$

where $\varepsilon(k) = 2\varphi||\mathbf{z}(k)||$. Using the definition of the ϵ-subgradient, it follows that $d_i(k) \in \partial_{\epsilon(k)} f_i(\bar{x}(k))$. Summing over all i we get that $\sum_{i=1}^N d_i(k) \in \partial_{N\epsilon(k)} f_i(\bar{x}(k))$. Note, that $\epsilon(k)$ has a random characteristic due to the assumption on $\mathbf{A}(k)$. ■

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

$$f(x) - f^* \geq \mu \|x - X^*\|. \tag{31}$$

Then, the following inequality holds

$$E[\|\bar{x}(k) - X^*\|^2] \leq \|\bar{x}(0) - X^*\|^2 \gamma^k + \frac{4\alpha\varphi\beta\sqrt{N}}{1-\gamma}\eta^{\lfloor \frac{k}{m} \rfloor} + \frac{\alpha^2\varphi^2}{1-\gamma}\left(\frac{4m\sqrt{N}}{1-\eta} + 1\right), \tag{32}$$

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

Proof: We use the same idea as in the proof of Proposition 2.4 in [10], formulated under a deterministic setup. Let $x^* \in X^*$ be an optimal point of f. By (7), where we use a constant stepsize α, we obtain

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

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

$$f(x^*) \geq f(\bar{x}(k)) + h(\bar{x}(k))'(x^* - \bar{x}(k)) - N\epsilon(k),$$

or, together with assumption (31),

$$-h(\bar{x}(k))'(\bar{x}(k) - x^*) \leq -\mu\|\bar{x}(k) - x^*\|^2 + N\epsilon(k).$$

Further, we can write

$$\|\bar{x}(k+1) - x^*\|^2 \leq \left(1 - \frac{2\alpha\mu}{N}\right)\|\bar{x}(k) - x^*\|^2 + 2\alpha\epsilon(k) + \alpha^2\varphi^2$$

or

$$E[\|\bar{x}(k) - x^*\|^2] \leq \left(1 - \frac{2\alpha\mu}{N}\right)^k \|\bar{x}(0) - x^*\|^2 + \sum_{s=0}^{k-1}\left(1 - \frac{2\alpha\mu}{N}\right)^{k-s-1}(2\alpha E[\epsilon(s)] + \alpha^2\varphi^2) \tag{33}$$

Note that the right-hand side of (33) converges only if $\alpha \leq \frac{N}{2\mu}$. To simplify notation, let $\gamma = \left(1 - \frac{2\alpha\mu}{N}\right)$. It follows that

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

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

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

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

From above an upper bound for $E[\|\bar{x}(k) - x^*\|^2]$ follows

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

By taking the minimum over $x^* \in X^*$ in the above relation, we obtain (32).

■

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 the ball around the optimal valued f^*, where the cost function f, evaluated at the estimate $x_i(k)$, is guaranteed to converge. Next, 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 ball 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 ball. Although dependent on other parameters of the problem, we will emphasize how the pmf of the random graph affects the performance metrics.

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), $i = 1, \ldots N$. Then

$$\liminf_{k\to\infty} E[f(x_i(k))] \leq f^* + \sqrt{N}\alpha\varphi^2\frac{m}{1-\eta}(N+2) + \frac{\alpha\varphi^2}{2} \quad (34)$$

Proof: Using the subgradient definition we have

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

or

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

Summing over all i, we get

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

By the results of Lemma 3.1, the following inequality holds

$$\liminf_{k\to\infty} E[f(x_i(k))] \le \liminf_{k\to\infty} E[f(\bar{x}(k))] + N\sqrt{N}\alpha\varphi^2 \frac{m}{1-\eta}. \tag{35}$$

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

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

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

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

or

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

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

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

or

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

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

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

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

which in turn leads to

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

Using the fact that

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

we get

$$\liminf_{k\to\infty} E[f(\bar{x}(k))] - f^* \leq 2\alpha\varphi^2 \sqrt{N}\frac{m}{1-\eta} + \frac{\alpha\varphi^2}{2}. \tag{36}$$

Inequality (34) follows by combining (35) and (36). ■

The next result shows that under a strong convexity type 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, 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

$$f(x) - f^* \geq \mu\|x - X^*\|^2, \ \forall x \in \mathbb{R}^n. \tag{37}$$

Then, the sequence $\{x_i(k)\}_{k\geq 0}$, generated by iteration (21) 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 $\max\{\gamma, \eta^{\frac{1}{m}}\}$ and the radius of the neighborhood equals

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

where

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

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

and where $\gamma = 1 - \frac{2\alpha\mu}{N}$.

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

$$E[\|x_i(k) - X^*\|^2] \leq E[\|x_i(k) - \bar{x}(k)\|^2] + 2E[\|x_i(k) - \bar{x}(k)\|^2]^{\frac{1}{2}}E[\|\bar{x}(k) - X^*\|^2]^{\frac{1}{2}} + E[\|\bar{x}(k) - X^*\|^2]. \tag{38}$$

The guaranteed radius of the neighborhood around some optimizer of f follows by inequalities (22) and (32) and by taking the limit as k goes to infinity of the above inequality.

By inequality (22) 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 (22). By noting that $\eta > \rho$, we can further write

$$E[\|x_i(k) - \bar{x}(k)\|^2] \leq a_1\eta^{\frac{k}{m}} + a_2 \tag{39}$$

where a_1 and a_2 are some positive scalars (different from the one in the previous inequality).

By inequality (32), 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 (32). We can further write

$$E[\|\bar{x}(k) - X^*\|^2] \leq b_1 \max\{\gamma, \eta^{\frac{1}{m}}\}^k + b_2, \tag{40}$$

where b_1 and b_2 differ from the one used above. Using the notations $c_1 = \max\{a_1, b_1\}$ and $c_2 = \max\{a_2, b_2\}$, by (38), (39) and (40), 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: $\frac{m}{1-\eta}$, $\frac{m}{1-\rho}$ 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 can 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:

$$\begin{array}{ll} \min_{m,p_i} & \{\frac{m}{1-\eta}, \frac{m}{1-\rho}, \eta^{\frac{1}{m}}\} \\ \text{subject to:} & m \geq 1 \\ & \sum_i^M p_i = 1,\ \ p_i \geq 0 \\ & \eta^{\frac{1}{m}} \leq \gamma. \end{array} \tag{41}$$

The last constraint is due to the fact that the rate of convergence of the MASM is not entirely dictated by $\eta^{\frac{1}{m}}$, but also by the choice of the stepsize α, through parameter γ. The solution to the above problem is a set of Pareto points, i.e. solutions 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, both η and ρ are minimized by the same pmf. Therefore, in problem (41), 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 are improved.

In what follows we provide a simple example where we show how η, the optimal probability distribution, $\frac{m}{1-\eta}$ 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

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

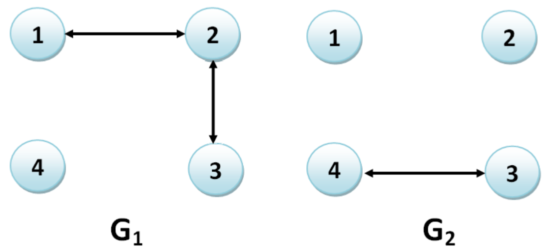

Fig. 1. The sample space of the random graph $\mathbf{G}(k)$

In Figure 2 we present four graphs. Graph *a*) shows the optimal probability p^* that minimizes η for different values of m. Graph *b*) shows the *optimized* η (computed at p^*) as a function of m. Graphs *c*) and *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 $p^* = 0.582$.

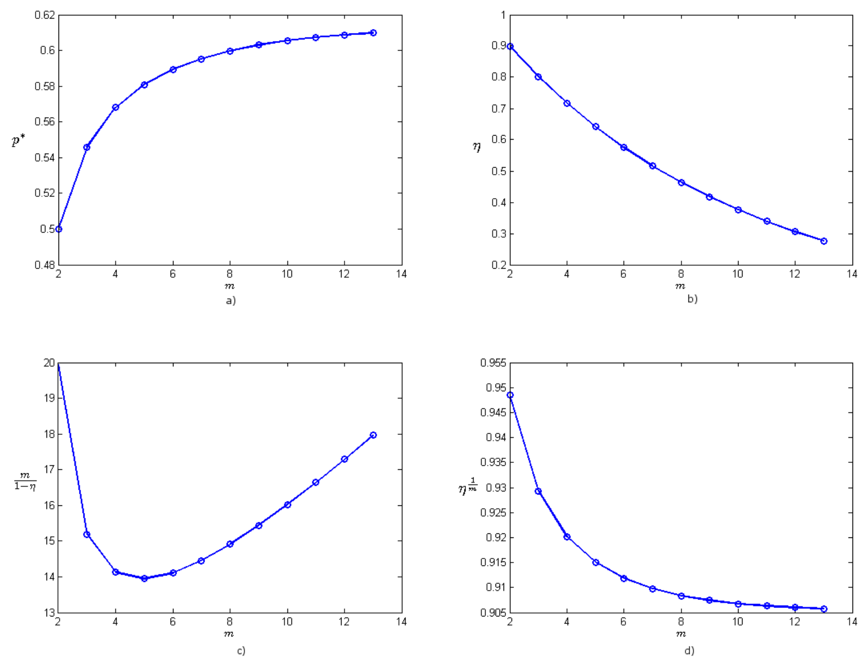

Fig. 2. a) Optimal p as a function of m; b) Optimized η as a function of m; c) Optimized $\frac{m}{1-\eta}$ as a function of m; d) Optimized $\eta^{\frac{1}{m}}$ as a function of m

In order to obtain the solution of problem (41), 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 in stead the bounds on η and ρ introduced in (16) and (17), respectively. Note that every result concerning the performance metrics still hold. In this case, for each value of m, η is minimized, when $\mathbf{p}_m$ is maximized, which can be interpreted as having to chose 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 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.

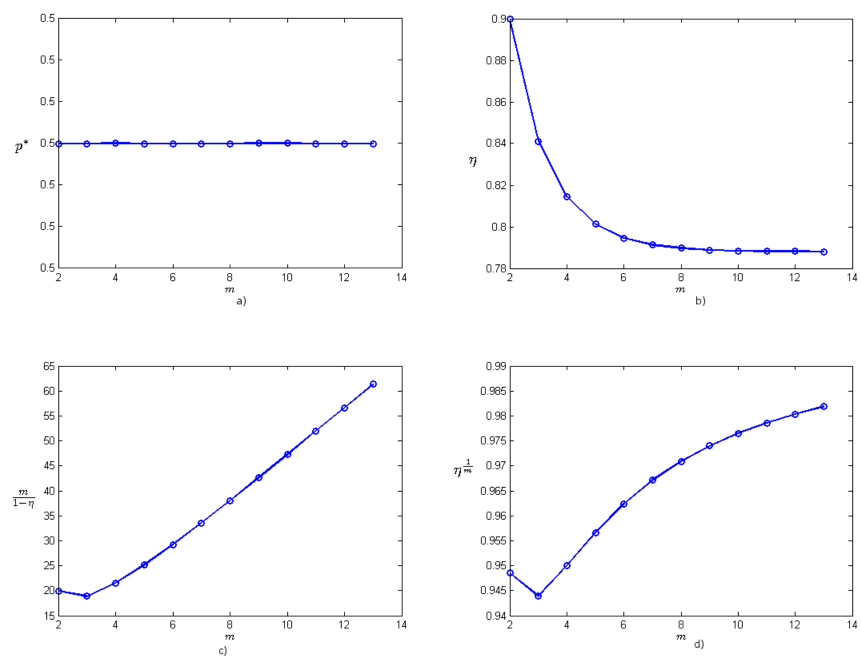

Fig. 3. a) Optimal p as a function of m; b) Optimized bound on η as a function of m; c) Optimized bound on $\frac{m}{1-\eta}$ as a function of m; d) Optimized bound on $\eta^{\frac{1}{m}}$ as a function of m

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 in such a way that allows 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 $\frac{m}{1-\eta}$ it 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. Tuning the communication protocol for MASM

In the previous sections we derived upper bounds for the performance metrics of the distributed subgradient algorithm as function of the probability distribution of the random graph modeling the communication topology. Depending on the assumptions used for the communication model, the probability distribution of the random graph can be a function on a set of communication protocol parameters such as probability of collisions, transmission power, etc. which can play the role of decisions variable in building distributions for the communication topology that minimizes η, this way ensuring good performances of the MASM. We make the following fundamental assumptions on the communication model.

Assumption 5.1: The communication among agents is assumed time slotted with slots normalized to integral units and each half of time slot a package containing relevant information is sent from a source to a destination.

Assumption 5.2: We consider bidirectional communication, i.e. at (the end of) a time slot, a link is consider active if successful transmissions took place in both directions.

Assumption 5.2 was made to accommodate the undirected graph assumption on the topology made in Section II-A, which played an important role in computing the bounds on the performance metrics.

We assume that the set $\mathcal{G}$ contains at least one connected graph, which allows us, for simplicity, to choose $m = 1$. In this case η becomes the average SLEM of the random matrix $\mathbf{A}(k)$.

For a general graph, obtaining explicit formulas for the pmf of the random communication topology, as function of the link probabilities is a difficult problem. That is why, in the following we will consider a topological model which will simplify the computation of the probability distribution of the random graph.

We consider that the N agents are placed equidistant on a circular structure. Each agent can communicate with its two adjacent neighbors in a lossless manner, i.e. these links (called henceforth *short range links*), exists with probability one. This communication model is a representation a small world network, modeling many real-life complex networks [15]. We want to study the effect of long range communication links (links between agents far away from each other) on the performance metrics of the MASM studied above. This problem is related to the consensus problem on small world graphs, studied in [1], where the authors showed that adding a small number of long range communication links improves considerable the rate of convergence to consensus. For simplicity we assume that the long range communication links are formed only between the *furthest away agents*.

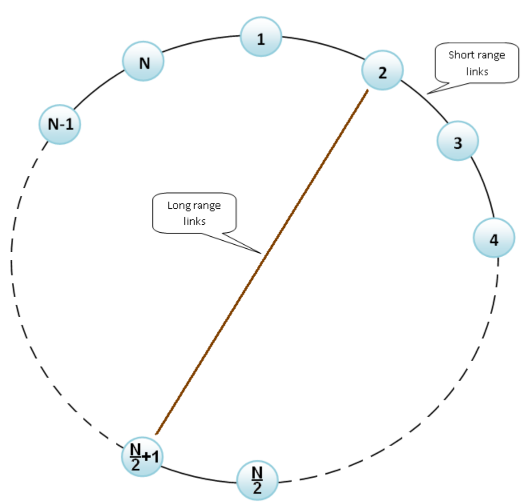

Fig. 4. Group of N agents forming a circulant graph

An agent can not be a source and a receiver simultaneously. The different role an agent takes is determined by the communication protocol. We consider two scenarios about the communication protocol. First we assume the agents use a simple randomized scheme, under which an agent decides to send a packet with some probability q. By assuming no error-transmissions, we want to find the best value for q, that minimizes η. In the second scenario we consider that the agents have a pre-established order for transmissions (for instance using a TDMA type of protocol). In this case we take into account the interference generated by other agents and we want to study the evolution of η with respect to the transmission powers.

In order to accommodate the bidirectional assumption on the communication links, we divide the time-slot in two. A link is considered active, if during a time-slot successful communication between two agents takes place in both direction (a link (a,b) is active during a time-slot if one of the following information transfer itineraries takes place: $a \rightarrow b,\ b \rightarrow a$ or $b \rightarrow a,\ a \rightarrow b$). In both scenarios there will be a probability of a link to be active during a time-slot, denoted by p, which will determine the probability distribution of the random graph (there is no index on p, since as we will see later, under our assumptions all links are active with the same probability).

We note that we have a total of $2^{\frac{N}{2}}$ possible topology (undirected graphs), with $C^i_{\frac{N}{2}}$ number of undirected graphs with i long-range links. The probability of a graph with i long-range links to appear is given by $p^i(1-p)^{\frac{N}{2}-i}$, for $i = 1,\ldots,\frac{N}{2}$. We arrange the graphs in the set $\mathcal{G} = \{G_0, G_1, \ldots, G_{2^{\frac{N}{2}}-1}\}$ in the increasing order of their number of links. The first graph in G_0 has no long range communication links, and has the probability $p_0 = 1 - \sum_{i=1}^{\frac{N}{2}} p^i(1-p)^{\frac{N}{2}-i}$.

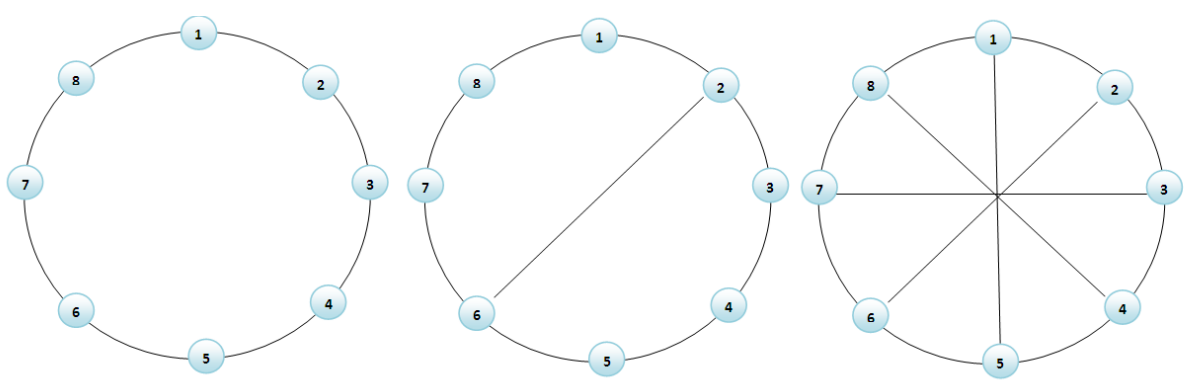

Fig. 5. Topologies G_0, G_1 and G_{15} for $N = 8$

Let $S_i = \sum_{l=0}^{i} C_{\frac{N}{2}}^{l}$. For all $j \in \{S_{i-1}, \ldots, S_i - 1\}$, the graph G_j has i long range links and a probability $p_i = p^i(1-p)^{\frac{N}{2}-i}$, for $i = 1, \ldots \frac{N}{2}$.

A. Randomized Communication Protocol

An agent can not be simultaneously transmitter and receiver. In case an agent receives a packet while in transmitter mode, the packet is consider lost and no retransmission is attempted. If the destination agent is in receiver mode, the packet is received without errors (this may be the case when all links use orthogonal coding schemes, beamforming, and/or when links are specially separated such that channel interference is negligible). In what follows we describe a simple randomized scheme for package transmissions, in which a link is active during a time-lot if both agents forming the link manage to successfully receive a package. We assume that at the beginning of each half of time-slot, nodes randomly decide to transmit with probability q. If the two agents are both source and receiver successively during a time-slot, then a link is active. In all other possible situations, (two agents send simultaneously, an agents is twice a source or a receiver, etc.) the link is considered inactive. The probability of a link to be active in a time slot is $p = 2q^2(1-q)^2$. The integer 2 comes from the fact that a link (a, b) in a time-slot can be given by two information transfer itineraries: $a \to b,\ b \to a$ or $b \to a,\ a \to b$. The idea of activating links with some probability is often used in network applications. For example in [12], the authors use the randomized activation of links to implement a distributed approximation of a power allocation scheme for networks with interference, while in [2] this idea appears in the context of constrained coalitional games for networks of autonomous agents. Our goal is to *tune* this protocol such that the performance of the MASM is improved. More precisely, we want to find the best value for q that ensures the tightest upper bound on the performance metrics, which is obtained by solving the following optimization problem

$$\begin{array}{ll} \min_{q\in(0\ 1)} & \sum_{j=0}^{2^{\frac{N}{2}}-1} \lambda_j p_j \\ \text{subject to :} & p_j = p^i(1-p)^{N/2-i},\ \text{if}\ j \in \{S_{i-1}, \ldots, S_i - 1\} \\ & p = 2q^2(1-q)^2 \\ & \lambda_j = SLEM(A_j),\ i = 1, \ldots, \frac{N}{2} \end{array} \quad (42)$$

We first construct the matrices A_i using a Laplacian based scheme. More precisely, to each graph G_i a Laplacian matrix L_i is associated, and the stochastic matrix A_i is computed according to the formula $A_i = I - \frac{1}{3}L_i$ (the scaling factor $\frac{1}{3}$ was chosen since an agent can engage in no more then three communication links). Note that the formula of η depends on the second largest eigenvalue of each the matrices A_i. Unfortunately, even under our simplifying assumptions about the communication topology, there is no know explicit formulas for computing the eigenvalues of the matrices A_i's. Therefore, we have to resort to numerical methods to compute these eigenvalues, which in turn may inflict a high computational burden in the case of large number of possible topologies.

In Figures 6, 7 and 8 we present simulation results for $N = 4$, $N = 8$ and $N = 16$. We present the evolution of η as a function of the probability q, together with the distribution of the SLEM of A_i with the number of long-range communication links. Remarkable (or perhaps not that remarkable due the symmetry imposed by our assumption), for all considered values of N, it turns out that the best choice for q is $\frac{1}{2}$. We can easily note that this value of q maximizes p, which means that η becomes small as we increase the probability of activating the long-range communication links.

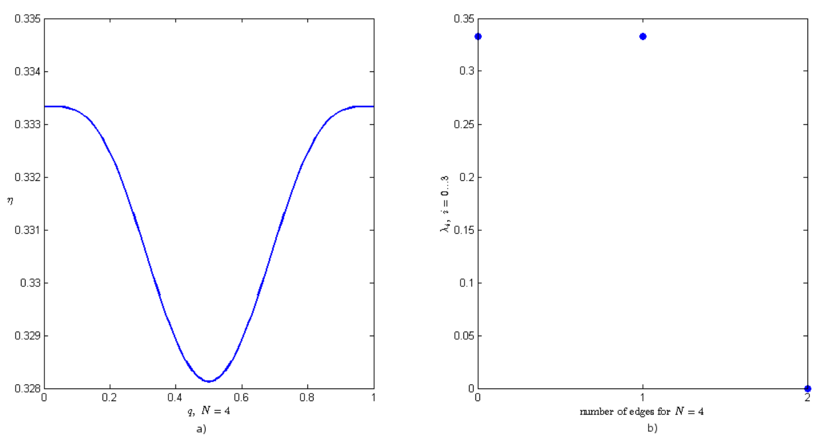

Fig. 6. a) η as a function of q for $N = 4$; b) The distribution of λ_i with respect to the number of edges for $N = 4$

The numerical simulations also show that the topology with the highest connectivity is not necessarily the best with respect to SLEM it induces (Figure 8). In fact, there exists graphs with only three long-range communication links which induce smaller SLEM compared with the fully long-range connected graph. Figure 9 shows such an example, where G_{51} having only three long range communication links induces a SLEM $\lambda_{51} = 0.8333$, while the fully long-ranged

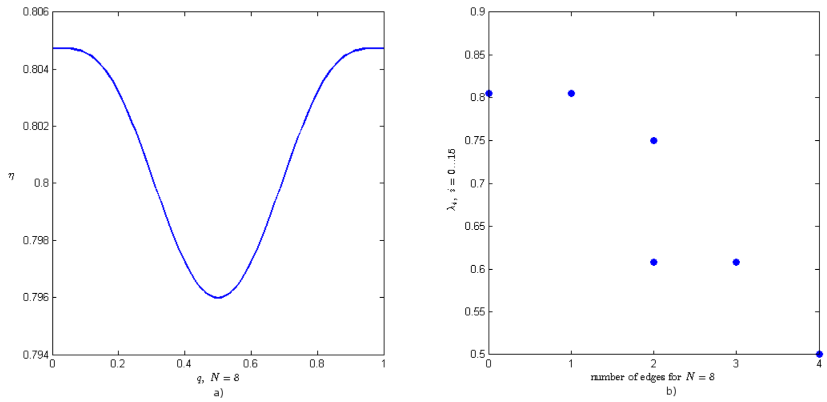

Fig. 7. a) η as a function of q for $N = 8$; b) The distribution of λ_i with respect to the number of edges for $N = 8$

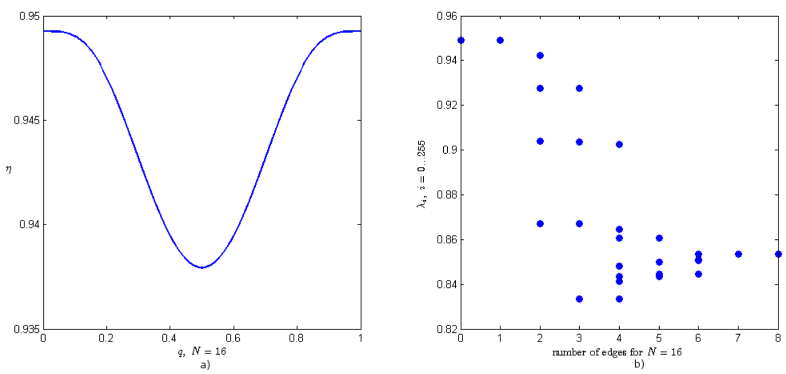

Fig. 8. a) η as a function of q for $N = 16$; b) The distribution of λ_i with respect to the number of edges for $N = 16$

connected graph G_{256} induces a larger SLEM, $\lambda_{256} = 0.8535$.

We have repeated the numerical simulations for the case where the matrices A_i where designed such that the SLEM is minimized under the given graph structure. This problem corresponds to finding the fastest mixing Markov chain on a graph, studied in [3]. In the aforementioned note, the fasted mixing Markov chain is obtained as a result of a semidefinite optimization problem of the form

$$\begin{aligned} &\text{minimize} && s \\ &\text{subject to} && -sI \preceq P - \tfrac{1}{n}\mathbb{1}\mathbb{1}' \preceq sI \\ & && P \geq 0,\ P\mathbb{1} = \mathbb{1},\ P = P' \\ & && P_{ij} = 0,\ (i,j) \in \mathcal{E}, \end{aligned} \tag{43}$$

where $\preceq$ denotes the matrix inequality and P is a symmetric stochastic matrix representing the

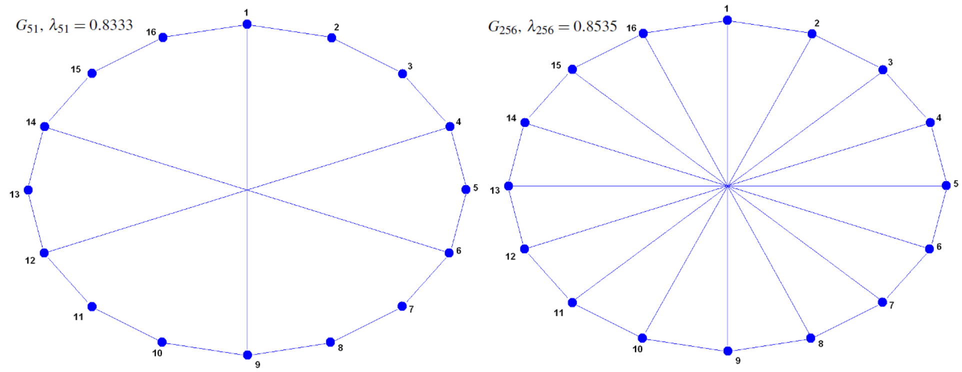

Fig. 9. Topologies with 3 and 8 edges respectively, for $N = 16$

decision variable, with a structure given by the set of edges $\mathcal{E}$. The following figures show the function η and the distribution of the SLEMs for the communication topologies corresponding to $N = 4$, $N = 8$ and $N = 16$, respectively, where the stochastic matrices A_i were obtained as solutions of the optimization problem (43).

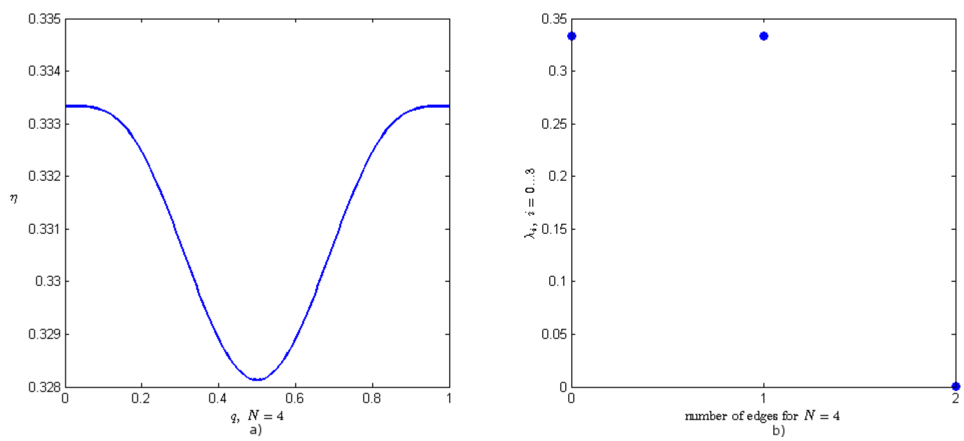

Fig. 10. a) η as a function of q for optimized SLEM and $N = 4$; b) The distribution of optimized λ_i with respect to the number of edges for $N = 4$

We note that even in the case of optimal SLEMs, comparable values for SLEM can be obtained with a smaller number of long range communication links. As expected, the optimal value of η is smaller in the case of optimal SLEM's. However, the optimal value for q remains unchanged, i.e. $q = \frac{1}{2}$. Although we improved the optimal value of η, the disadvantage of using optimal

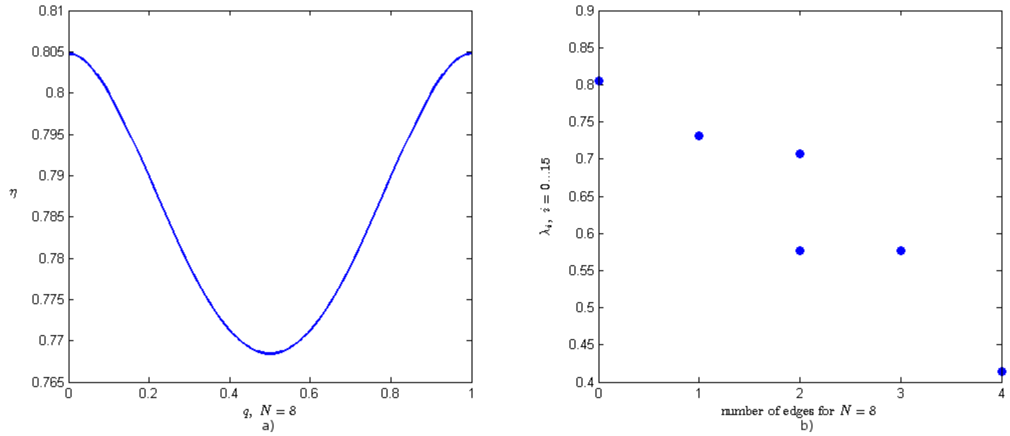

Fig. 11. a) η as a function of q for optimized SLEM and $N = 8$; b) The distribution of optimized λ_i with respect to the number of edges for $N = 8$

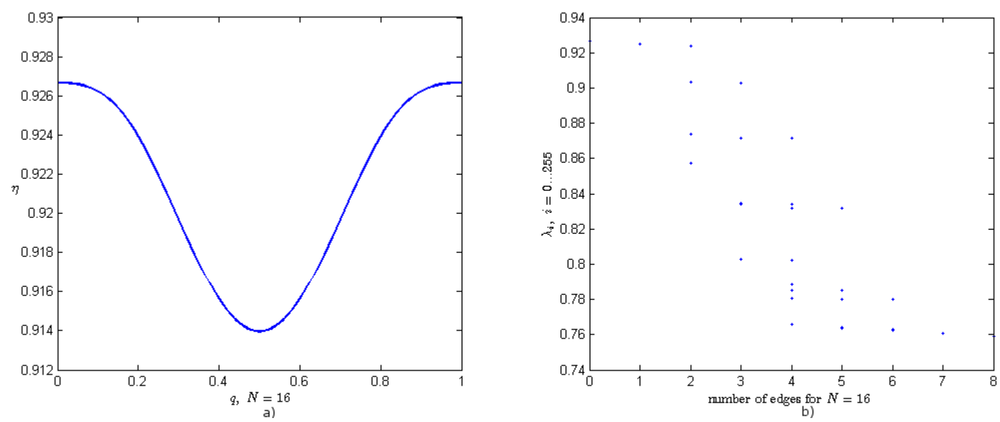

Fig. 12. a) η as a function of q for optimized SLEM and $N = 16$; b) The distribution of optimized λ_i with respect to the number of edges for $N = 16$

stochastic matrices in the consensus step consists in the fact that each agent would have to know the current global topology in order to apply the appropriate (optimal) coefficients in consensus step, which may be an unrealistic assumption.

B. Communication protocol with interference

In the previous scenario we assumed that the communication was not affected by the interference. In the following, we account for transmission interferences in the random graph model.

Unlike the previous case, we assume that there is a pre-established order of communication between agents such that no collision due to simultaneous transmission between a pair of agents forming a link can occur. For simplicity we assume that during the first half of a time-slot, agents $\{1,...\frac{N}{2}\}$ send a packet to the remaining $\{\frac{N}{2}+1,...,N\}$ agents (in a one to one manner, i.e. 1 sends to $\frac{N}{2}+1$, 2 sends to $\frac{N}{2}+2$, etc.) In the second half of a time slot, the last half of the agent set, sends to the first one. Although no collision can occur due to simultaneous transmissions of a pair of agents, packets can still be lost due the interference created by other transmitting agents. If the transmissions during a time slot are successful (i.e. no package is lost), a link between two communicating agents is considered active.

Let $P_1,...,P_N$ denote the powers each agent invest for the transmission of a packet in a (half of a) time slot. The power received from transmitter i at receiver j is given by $G_{ij}F_{ij}P_i$, where G_{ij} is a positive scalar representing the path gain from transmitter i to receiver j and F_{ij} is a random variable modeling the *fading* of the channel between i and j. The scaler G_{ij} represents for us a distance dependent power attenuation, but can also be interpreted as log-normal shadowing, cross correlation between codes in a code division multiple access system or a gain dependency on antenna direction. In particular, we assume that G_{ij} is given by

$$G_{ij} = cd_{ij}^{-\alpha},$$

where c represent a constant scalar depending on the physical characteristics of the transmitter antenna and d_{ij} is the distance between the two communicating agents, with α taking values in $\{1,2\}$, depending on the environment. We assume independent Rayleigh models for channels and therefore each F_{ij} has an exponential distribution with unit mean. At each receiver i, the affect of a white Gaussian noise of power σ_i^2 will also be sensed. Given that agent i is a receiver and agent j is a transmitter, the signal to interference-plus-noise ratio $SINR_{ij}$ at agent j is given by

$$SINR_{ij} = \frac{F_{ij}G_{ij}P_i}{\sigma_j^2 + \sum_{l\neq i,j} F_{lj}G_{lj}P_l}. \tag{44}$$

We define the *outage* of a link (i, j) as the event when a packet transmission fails. The outage event can be translated as the event when the $SINR_{ij}$ is less then a threshold S_j. The threshold S_j depends on the physical layer parameters such as rate transmissions, modulation and coding. The probability of a successful transmission of a packet is given by $Pr(SINR_{ij} \geq S_j)$. The probability of a successful packet transmission (in a half-time slot) can be explicitly written as

$$Pr(SINR_{ij} \geq S_j) = e^{-\frac{S_j \sigma_j^2}{G_{ij} P_i}} \prod_{l=1, l \neq i,j}^{N} \left(1 + \frac{S_j G_{lj} P_l}{G_{ij} P_i}\right)^{-1} \tag{45}$$

The same model for the outage (success) probability of a link was used in [9] or [13], where the authors designed power allocation schemes in the case of interference-limited fading wireless channels and for solving the completion time minimization and robust power control problem, respectively.

By assuming that the statistics of the noise and the SINR threshold are the same at all receivers and that the agents use the same power for transmissions, the probability of a link to be active can be expressed as:

$$p = e^{-2\frac{S\sigma^2}{GP}} \prod_{j=2}^{\frac{N}{2}} \left(1 + \frac{SG_{1j}}{G}\right)^{-2}, \tag{46}$$

where σ^2 represent the noise variance, S denotes the $SINR$ threshold, P is the transmission power, G is the power attenuation between agents i and $\frac{N}{2} + i$ (for any $i \in \{1, \ldots, \frac{N}{2}\}$).

We noted that in the randomized protocol scheme a large value of p ensures a small value for η. In the current scenario p is maximized as the power P converge to infinity. In Figure V-B *a*), *b*) and *c*) we represent the evolution of η (for optimized SLEM) as a function of the power P, for $N = 4$, $N = 8$ and $N = 16$, respectively.

For the numerical simulations we chose the following values for the communication model parameters: $R = 10$ where R denotes the range of the circular graph, $S = 0.1$, $c = 1$ and the variance of the noise $\sigma^2 = 0.04$. As expected, η decreases asymptotically with the power P to some value which depends on the SINR threshold and on the attenuation values. However, for large values of P the decrease in η becomes smaller and smaller, the above analysis providing a basis for choosing a value for P that ensures a small enough value for η, while keeping the transmission power limited.

VI. 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

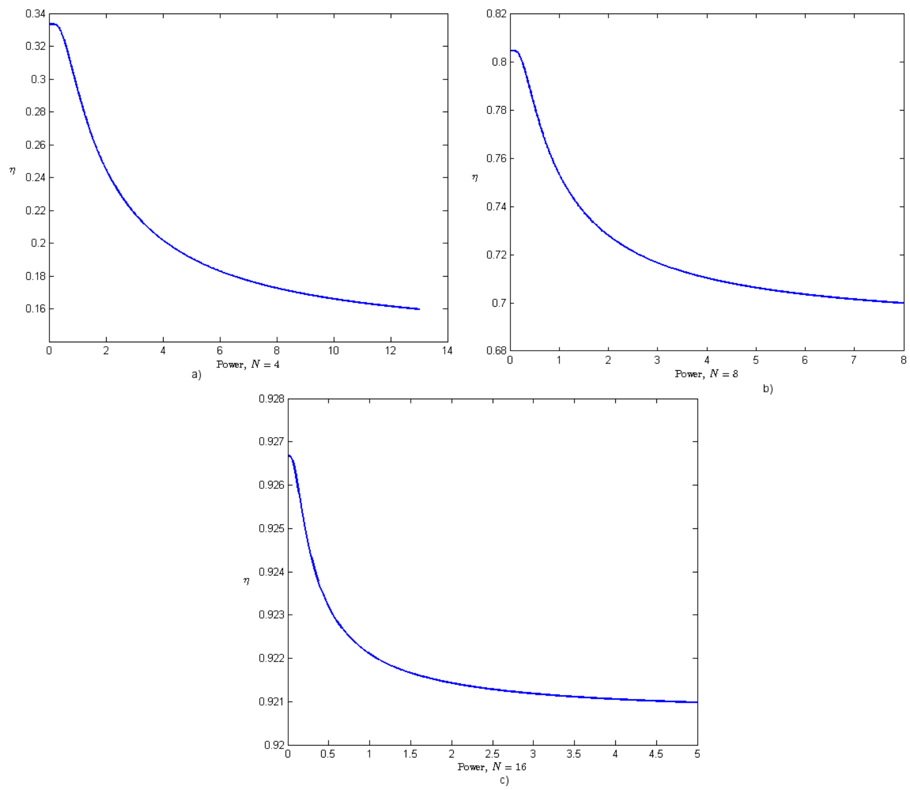

Fig. 13. a) η as a function of power P for $N = 4$; b) η as a function of power P for $N = 8$; c) η as a function of power P for $N = 16$

metrics is given by the radius of the ball 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 ball 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 ball, respectively. All the aforementioned performance measures were expressed in terms of the probability distribution of the random communication topology. We then studied how these upper bounds can be improved by tuning the parameters of the communication protocol in the case of two scenarios. In the first scenario, we assume a randomized scheme for link activation with lossless transmission while in the second scenario we used a pre-established order for transmissions, where the interference effect was accounted for. Both these scenarios were applied on small world type of topology.

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, "Distributed Subgradient Methods over Random Networks, LIDS report

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

[11] A. Nedic and D.P. Bertsekas, "Incremental Subgradient Methods for Nondifferential Optimization“, SIAM 2000.

[12] 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

[13] 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

[14] 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.

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