

Performance Evaluation of the Consensus-Based Distributed Subgradient Method Under Random Communication Topologies

Ion Matei and John S. Baras, *Fellow, IEEE*

Abstract—We investigate collaborative optimization of an objective function expressed as a sum of local convex functions, when the agents make decisions 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, assumed independent and identically distributed. Specifically, we study the performance of the consensus-based multi-agent distributed subgradient method and show how it depends on the probability distribution of the random graph. For the case of a constant stepsize, we first give an upper bound on the difference between the objective function, evaluated at the agents' estimates of the optimal decision vector, and the optimal value. Second, for a particular class of convex functions, we give an upper bound on the distances between the agents' estimates of the optimal decision vector and the minimizer. In addition, we provide the rate of convergence to zero of the time varying component of the aforementioned upper bound. The addressed metrics are evaluated via their expected values. As an application, we show how the distributed optimization algorithm can be used to perform collaborative system identification and provide numerical experiments under the randomized and broadcast gossip protocols.

Index Terms—Distributed, stochastic systems, sub-gradient methods, system identification.

I. INTRODUCTION

MULTI-AGENT distributed optimization problems appear naturally in many distributed processing problems (such as network resource allocation, collaborative control and estimation, etc.), 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 [17], 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 estimates 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 [9], where the communication topology is assumed time invariant and where the order of the two stages mentioned above is inverted.

We investigate the collaborative optimization problem in a multi-agent setting, when the agents make decisions 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, assumed independent and identically distributed (i.i.d.). Specifically, we study the performance of the consensus-based multi-agent distributed subgradient method proposed in [17], for the case of a constant stepsize.

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 topologies is addressed in [14]. 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 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 evaluated at the (weighted) time average of the optimal decision vector's estimate. *Our main goal is to 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 the difference between the cost function, evaluated at the estimate, and the optimal value. Next, for a particular class of convex functions, we focus on the distance between the estimate of the optimal decision and the minimizer. The upper bound we provide has a constant component and a time varying component. For the latter, we provide the rate of convergence to zero. The performance metrics are evaluated via their expected values. The explicit dependence on the graph's probability distribution may be useful to design probability distributions that would ensure the best guaranteed upper bounds on the performance metrics. This idea has relevance especially in the wireless networks, where the communication topology has a random nature with a

Manuscript received July 20, 2010; revised November 18, 2010; accepted February 02, 2011. Date of publication February 28, 2011; date of current version July 20, 2011. This work was supported in part by the U.S. Air Force Office of Scientific Research MURI award FA9550-09-1-0538, in part by the Defence Advanced Research Projects Agency (DARPA) under award number 013641-001 for the Multi-Scale Systems Center (MuSyC) through the FRCP of SRC and DARPA, and in part by BAE Systems award number W911NF-08-2-0004 (the ARL MAST CTA). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Michael Gastpar.

I. Matei is with the Electrical Engineering Department, University of Maryland, College Park, MD 20742 USA, and also with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: imatei@umd.edu; ion.matei@nist.gov).

J. S. Baras is with the Electrical Engineering Department and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: baras@umd.edu).

Digital Object Identifier 10.1109/JSTSP.2011.2120593

probability distribution (partially) determined by the communication protocol parameters (the reader can consult [13], [20], where the authors introduce probabilistic models for successful transmissions as functions of the transmission powers). As an example of possible application, we show how the distributed optimization algorithm can be used to perform collaborative system identification and we present numerical experimental results under the randomized [4] and broadcast [1] gossip protocols. Similar performance metrics as ours are studied in [10], where the authors generalize the randomized incremental subgradient method and where the stochastic component in the algorithm is described by a Markov chain, which can be constructed in a distributed fashion using local information only. Newer results on the distributed optimization problem can be found in [6], where the authors analyze distributed algorithms based on dual averaging of subgradients, and provide sharp bounds on their convergence rates as a function of the network size and topology.

Notations: Let X be a subset of \mathbb{R}^n and let y be a point in \mathbb{R}^n . By slight abuse of notation, let $\|y - X\|$ denote the distance from the point y to the set X , i.e., $\|y - X\| \triangleq \inf_{x \in X} \|y - x\|$, where $\|\cdot\|$ is the standard Euclidean norm. For a twice differentiable function $f(x)$, we denote by $\nabla f(x)$ and $\nabla^2 f(x)$ the gradient and Hessian of f at x , respectively. Given a symmetric matrix A , by $(A \succeq 0)$ $A \succ 0$ we understand A is positive (semi) definite. The symbol \otimes represents the Kronecker product.

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 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 :

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

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

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

The gradient of the differentiable function $f(x)$ on \mathbb{R}^n satisfies a *Lipschitz condition with constant L* if

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

The differentiable, convex function $f(x)$ on \mathbb{R}^n is *strongly convex with constant l* if

$$f(y) \geq f(x) + \nabla f(x)'(y - x) + \frac{l}{2}\|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

We will denote by LEM and SLEM the largest and second largest eigenvalue (in modulus) of a matrix, respectively. We will use CBMASM as the abbreviation for **C**onsensus-**B**ased **M**ulti-**A**gent **S**ubgradient **M**ethod and pmf for probability mass function.

Paper Structure: Section II contains the problem formulation. In Section III, we introduce a set of preliminary results, which mainly consist of providing upper bounds for a number a quantities of interest. Using these preliminary results, in Section IV we give upper bounds for the expected value 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 the minimizer. Section V shows how the distributed optimization algorithm can be used for collaborative system identification.

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 $G(k) = (V, \mathcal{E}(k))$, where V is the set of N vertices (nodes) and $\mathcal{E}(k) = (e_{ij}(k))$ is the set of edges, and where we used k to denote the time index. The edges in the set $\mathcal{E}(k)$ correspond to the communication links among agents. Given a positive integer M , the graph $G(k)$ takes values in a finite set $\mathcal{G} = \{G_1, G_2, \dots, G_M\}$ at each k , where the graphs $G_i = (V, \mathcal{E}_i)$ are assumed *undirected* and *without self loops*. In other words, we will consider only bidirectional communication topologies. The underlying random process of $G(k)$ is assumed i.i.d. with probability distribution $Pr(G(k) = G_i) = p_i, \forall k \geq 0$, where $\sum_{i=1}^M p_i = 1$ and $p_i > 0$.

Assumption 2.1: The graph $\bar{G} = (V, \bar{E})$ resulting from the union of all graphs in \mathcal{G} is *connected*, where

$$\bar{G} = \bigcup_{i=1}^M G_i = \left(V, \bigcup_{i=1}^M \mathcal{E}_i \right).$$

Let G be an undirected graph with N nodes and no self loops 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 nonzero entry (i, j) of A , with $i \neq j$, implies a link from j to i in G and vice-versa.

B. Optimization Model

The task of the N agents consists of minimizing a convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. 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 $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ 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., $X^* = \{x \in \mathbb{R}^n \mid f(x) = f^*\}$. 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 subject to the communication topology described by the random graph $G(k)$.

As proposed in [17], 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 $A(k)$ which corresponds to the communication graph $G(k)$. The matrices $A(k)$ 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 $A(k)$ is inherited from $G(k)$, i.e., $Pr(A(k) = A_i) = Pr(G(k) = G_i) = p_i$. 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., $d_i(k) \in \partial f_i(x_i(k))$. Obviously, when $f_i(x)$ are assumed differentiable, $d_i(k)$ becomes the gradient of f_i at $x_i(k)$, i.e., $d_i(k) = \nabla f_i(x_i(k))$.

Note that the first part of (5) is a consensus step, a problem that has received a lot of attention in recent years, both in a deterministic ([3], [7], [8], [16], [22], [25], [26]) and a stochastic ([12], [15], [23], [24]) framework.

The consensus problem under different gossip algorithms was studied in [1], [4], and [5]. We note that there is direct connection between our communication model and the communication model used in the *randomized gossip protocol* [4]. Indeed, in the case of the randomized communication protocol, the set \mathcal{G} is formed by the graphs G_{ij} with only one link (i, j) , where $Pr(G(k) = G_{ij}) = (1/N)P_{ij}$ for some $P_{ij} > 0$ with $\sum_{i=1}^N P_{ij} = 1$, while the set \mathcal{A} is formed by stochastic matrices A_{ij} of the form $A_{ij} = I - (1/2)(e_i - e_j)(e_i - e_j)'$, where the vectors e_i represent the standard basis. Our model can also be used to describe a *modified* version of the *broadcast communication protocol* [1], where we assume that when an agent wakes up and broadcasts information to the neighborhood, it receives information from the neighbors as well. In the case of the (modified) broadcast communication protocol, the set \mathcal{G} is formed by the graphs G_i , where G_i contains links between the node i and the nodes in its neighborhood, denoted by N_i . The probability distribution of $G(k)$ is given by $Pr(G(k) = G_i) = 1/N$ and the set \mathcal{A} is formed by matrices of the form $A_i = I - \delta_i \sum_{j \in N_i} (e_i - e_j)(e_i - e_j)'$, for some $0 < \delta_i \leq (1/|N_i|)$.

The following assumptions, which will not necessarily be used simultaneously, introduce properties of the function $f(x)$.

Assumption 2.2: (Non-Differentiable Functions):

- a) The subgradients of the functions $f_i(x)$ are uniformly bounded, i.e., there exists a positive scalar φ such that

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

- b) The stepsize is constant, i.e.,

$$\alpha(k) = \alpha, \quad \forall k \geq 0.$$

- c) The optimal solution set X^* is nonempty.

Assumption 2.3: (Differentiable Functions):

- a) The functions $f_i(x)$ are twice continuously differentiable on \mathbb{R}^n .
- b) There exists positive scalars l_i, L_i such that

$$l_i I \preceq \nabla^2 f_i(x) \preceq L_i I, \quad \forall x \in \mathbb{R}^n \text{ and } \forall i.$$

- c) The stepsize is constant, i.e., $\alpha(k) = \alpha$ for all k and satisfies the inequality

$$0 < \alpha < \min \left\{ \frac{\underline{\lambda} + 1}{L}, \frac{1}{l} \right\}$$

where $\underline{\lambda}$ is the smallest among all eigenvalues of matrices A_i , $l = \min_i l_i$, and $L = \max_i L_i$.

If Assumption 2.3-(a) holds, Assumption 2.3-(b) is satisfied if the gradient of $f_i(x)$ satisfies a Lipschitz condition with constant L_i and if $f_i(x)$ is strongly convex with constant l_i . Also, under Assumptions 2.3, X^* has one element which is the unique minimizer of $f(x)$, denoted henceforth by x^* .

III. PRELIMINARY RESULTS

In this section, we lay the foundation 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 the estimate of the optimal decision vector and the average of all estimates. The second quantity is described by the distance between the average of all estimates and the minimizer.

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

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

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

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

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

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

and let β be a positive scalar such that

$$\|z_i(0)\| \leq \beta, \quad i = 1, \dots, N.$$

Let us define the *aggregate* vectors of estimates, average estimates, deviations and (sub)gradients, 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

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

where $\mathbf{J} = (1/N)\mathbb{1}\mathbb{1}' \otimes I$, 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

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

where \mathbf{I} is the identity matrix in $\mathbb{R}^{nN \times nN}$. The next Proposition characterizes the dynamics of the vector $\mathbf{z}(k)$.

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

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

where $\mathbf{W}(k) = \mathbf{A}(k) - \mathbf{J}$ and $\mathbf{A}(k) = A(k) \otimes I$, with solution

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

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 $\Phi(k, k) = \mathbf{I}$.

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

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

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

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

By noting that

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

we obtain (10). The solution (11) follows from (10) together with the observation that $\Phi(k, s)(\mathbf{I} - \mathbf{J}) = \Phi(k, s)$. ■

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

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

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

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

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 $\Phi(k+m, k) = \mathbf{W}(k+m-1)\cdots\mathbf{W}(k)$, generated by a sequence of random graphs of length m , i.e., $G(k), \dots, G(k+m-1)$, 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 $q_{i_1, \dots, i_m} = \prod_{j=1}^m p_{i_j}$ 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 graphs with the respective indices produces a connected graph, i.e., $I_m = \{i_1 i_2, \dots, i_m \mid \bigcup_{j=1}^m G_{i_j} = \text{connected}\}$. Using the previous notations, the first and second moments of the norm of $\Phi(k+m, k)$ can be expressed as

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

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

where $\eta_m = \sum_{j \in I_m} q_j \lambda_j + 1 - \sum_{j \in I_m} q_j$ and $\rho_m = \sum_{j \in I_m} q_j \lambda_j^2 + 1 - \sum_{j \in I_m} q_j$. The integer j was used as an index for the elements of the set I_m , i.e., for an element of the form i_1, \dots, i_m .

The above formulas follow from results introduced in [8], Lemma 1, or in [22], 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 any $j \in I_m$. Conversely, if $j \notin I_m$ then $\lambda_j = 1$. We also note that $\sum_{j \in I_m} q_j$ 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. Therefore, for any m such that I_m is not empty, we have that $0 < \rho_m < \eta_m < 1$. In general, for large values of m , it may be difficult to compute all eigenvalues $\lambda_j, j \in I_m$. 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

$$\eta_m \leq \lambda_m \mathbf{p}_m + 1 - \mathbf{p}_m \quad (16)$$

$$\rho_m \leq \lambda_m^2 \mathbf{p}_m + 1 - \mathbf{p}_m \quad (17)$$

where $\lambda_m = \max_{j \in I_m} \lambda_j$ and $\mathbf{p}_m = \sum_{j \in I_m} q_j$ 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 paper, 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 values of the norm and the squared norm of the transition matrix $\Phi(k, s)$.

Proposition 3.2: Let Assumption 2.1 hold, and let $r \leq s \leq k$ be three nonnegative integer values and m 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

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

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

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

where η and ρ are defined in Remark 3.2.

Proof: We fix an m such that the probability of having a connected graph over a time interval of length m is positive, i.e., I_m is non-empty. Note that, by Assumption 2.1, such a value always exists (pick $m \geq M$). 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, \dots, s_t be a sequence of nonnegative integers such that $s = s_0 < s_1 < \dots < s_t \leq k$, where $s_{i+1} - s_i = m$ and $i = 0, \dots, m-1$. 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) follows from (15) and from the i.i.d. assumption on the random graph process.

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

$$\begin{aligned} E[\|\Phi(k, r)\Phi(k, s)'\|] &\leq E[\|\Phi(k, s)\|^2 \|\Phi(s, r)\|] \\ &\leq E[\|\Phi(k, s)\|^2] E[\|\Phi(s, r)\|] \end{aligned}$$

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

In the next lemma we show that, under Assumption 2.3, for small enough α the gradients $\nabla f_i(x_i(k))$ remain bounded with probability one for all k .

Lemma 3.1: Let Assumption 2.3 hold and let $\mathcal{F} : \mathbb{R}^{Nn} \rightarrow \mathbb{R}$ be a function given by $\mathcal{F}(\mathbf{x}) = \sum_{i=1}^N f_i(x_i)$, where $\mathbf{x}' = (x'_1, \dots, x'_N)$. There exists a positive scalar φ such that

$$\begin{aligned} \|\nabla f_i(x_i(k))\| &\leq \varphi, \quad \forall i, k \text{ w.p. } 1 \\ \|\nabla f_i(\bar{x}(k))\| &\leq \varphi, \quad \forall i, k \text{ w.p. } 1 \end{aligned}$$

where $\varphi = L\|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + L((2/(1-q)) + 1)\|\tilde{\mathbf{x}}\|$, $q = \max\{|\underline{\lambda} - \alpha L|, |1 - \alpha l|\}$, $\tilde{\mathbf{x}}$ is the unique minimizer of $\mathcal{F}(\mathbf{x})$, and $x_i(k)$ and $\bar{x}(k)$ satisfy (5) and (7), respectively.

Proof: We first note that since the matrices A_i have positive diagonal entries, they are aperiodic and therefore $\underline{\lambda} \in (-1, 1]$. From Assumption 2.3 it follows immediately that $\mathcal{F}(\mathbf{x})$ is a convex, twice differentiable function satisfying

$$l\mathbf{I} \preceq \nabla^2 \mathcal{F}(\mathbf{x}) \preceq L\mathbf{I} \quad (21)$$

where $l = \min_i l_i$, $L = \max_i L_i$ and \mathbf{I} is the identity matrix in $\mathbb{R}^{nN \times nN}$. In addition, $\mathcal{F}(\mathbf{x})$ has a unique minimizer denoted

by $\tilde{\mathbf{x}}$. The dynamics described by (5) can be compactly written as

$$\mathbf{x}(k+1) = \mathbf{A}(k)\mathbf{x}(k) - \alpha \nabla \mathcal{F}(\mathbf{x}(k)), \quad \mathbf{x}(0) = \mathbf{x}_0 \quad (22)$$

with $\mathbf{x}(k)' = (x_1(k)', \dots, x_N(k)')$.

We observe that (22) is a modified version of the gradient method with constant step, where instead of the identity matrix, we have that $\mathbf{A}(k)$ multiplies $\mathbf{x}(k)$. In what follows we show that the stochastic dynamics (22) is stable with probability one.

Using a similar idea as in [21, Th. 3, p. 25], we have that

$$\begin{aligned} \nabla \mathcal{F}(\mathbf{x}(k)) &= \nabla \mathcal{F}(\tilde{\mathbf{x}}) \\ &\quad + \int_0^1 \nabla^2 \mathcal{F}(\tilde{\mathbf{x}} + \tau(\mathbf{x}(k) - \tilde{\mathbf{x}})) (\mathbf{x}(k) - \tilde{\mathbf{x}}) d\tau \\ &= \mathcal{H}(k) (\mathbf{x}(k) - \tilde{\mathbf{x}}) \end{aligned}$$

where $l\mathbf{I} \preceq \mathcal{H}(k) \preceq L\mathbf{I}$ by virtue of (21). Hence, with probability one

$$\begin{aligned} \|\mathbf{x}(k+1) - \tilde{\mathbf{x}}\| &= \|\mathbf{A}(k)\mathbf{x}(k) - \tilde{\mathbf{x}} - \alpha \nabla \mathcal{F}(\mathbf{x}(k)) \\ &\quad + \mathbf{A}(k)\tilde{\mathbf{x}} - \mathbf{A}(k)\tilde{\mathbf{x}}\| \\ &\leq \|\mathbf{A}(k) - \alpha \mathcal{H}(k)\| \|\mathbf{x}(k) - \tilde{\mathbf{x}}\| \\ &\quad + \|\mathbf{A}(k) - \mathbf{I}\| \|\tilde{\mathbf{x}}\|. \end{aligned}$$

But since

$$(\underline{\lambda} - \alpha L)\mathbf{I} \preceq \mathbf{A}(k) - \mathcal{H}(k) \preceq (1 - \alpha l)\mathbf{I}$$

it follows that

$$\|\mathbf{x}(k+1) - \tilde{\mathbf{x}}\| \leq q \|\mathbf{x}(k) - \tilde{\mathbf{x}}\| + |\underline{\lambda} - 1| \|\tilde{\mathbf{x}}\|$$

where $q = \max\{|\underline{\lambda} - \alpha L|, |1 - \alpha l|\}$. Since by Assumption 2.3-(c) $\alpha < \min\{(\underline{\lambda} + 1)/L, (1/l)\}$ we get that $q < 1$ and therefore the dynamics (22) is stable with probability one and

$$\begin{aligned} \|\mathbf{x}(k) - \tilde{\mathbf{x}}\| &\leq q^k \|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + \frac{2}{1-q} \|\tilde{\mathbf{x}}\| \\ &\leq \|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + \frac{2}{1-q} \|\tilde{\mathbf{x}}\|, \quad \forall k. \end{aligned}$$

From Assumption 2.3 we have that

$$\begin{aligned} \|\nabla f_i(x_i(k))\| &\leq \|\nabla \mathcal{F}(\mathbf{x}(k))\| \leq L \|\mathbf{x}(k) - \tilde{\mathbf{x}}\| \\ &\leq L \|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + \frac{2L}{1-q} \|\tilde{\mathbf{x}}\|. \end{aligned} \quad (23)$$

We also have that

$$\|\bar{\mathbf{x}}(k) - \tilde{\mathbf{x}}\| = \|\mathbf{J}\mathbf{x}(k) - \mathbf{J}\tilde{\mathbf{x}} + \mathbf{J}\tilde{\mathbf{x}} - \tilde{\mathbf{x}}\| \leq \|\mathbf{x}(k) - \tilde{\mathbf{x}}\| + \|\tilde{\mathbf{x}}\|$$

from where it follows that

$$\begin{aligned} \|\nabla f_i(\bar{x}(k))\| &\leq \|\nabla \mathcal{F}(\bar{\mathbf{x}}(k))\| \leq L \|\bar{\mathbf{x}}(k) - \tilde{\mathbf{x}}\| \\ &\leq L \|\mathbf{x}(0) - \tilde{\mathbf{x}}\| + L \left(\frac{2}{1-q} + 1 \right) \|\tilde{\mathbf{x}}\|. \end{aligned} \quad (24)$$

Taking the maximum among the right-hand-side terms of the inequalities (23) and (24), the result follows. ■

Remark 3.3: If the stochastic matrices A_i are generated using a Laplacian based scheme, e.g.,

$$A_i = I - \varepsilon \mathcal{L}_i, \quad \forall i$$

where \mathcal{L}_i is the Laplacian of the graph G_i and $\varepsilon \leq (1/N)$, then it turns out that $\underline{\lambda} \geq 0$. Hence, the inequality in Assumption 2.3-(c) is satisfied if

$$0 < \alpha < \frac{1}{L}$$

which is a sufficient condition for the stability of (5). In the case of the randomized and broadcast gossip protocols it can be checked that $\underline{\lambda} = 0$.

Remark 3.4: Throughout the rest of the paper, φ should be interpreted in the context of the assumptions used, i.e., under Assumption 2.2, φ is the uniform bound on the subgradients of $f_i(x)$, while under Assumption 2.3, φ is the bound on the gradients $\nabla f_i(x_i(k))$ and $\nabla f_i(\bar{x}(k))$ given by Lemma 3.1.

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 3.2: Under Assumptions 2.1 and 2.2 or 2.1 and 2.3, 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:

$$E[\|x_i(k) - \bar{x}(k)\|] \leq \beta \sqrt{N} \eta \lfloor \frac{k}{m} \rfloor + \alpha \varphi \sqrt{N} \frac{m}{1 - \eta} \quad (25)$$

$$E[\|x_i(k) - \bar{x}(k)\|^2] \leq N \beta^2 \rho \lfloor \frac{k}{m} \rfloor + N \alpha^2 \varphi^2 \left(1 + 2 \frac{m}{1 - \eta}\right) \times \frac{m}{1 - \rho} + 2N \alpha \beta \varphi m \times \frac{\rho \lfloor \frac{k-1}{m} \rfloor + 1 - \eta \lfloor \frac{k-1}{m} \rfloor + 1}{\rho - \eta} \quad (26)$$

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 Proposition 3.1, we have

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

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 we used 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

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

Noting that the sum $\sum_{s=0}^{k-1} \eta \lfloor (k-s-1)/m \rfloor$ 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 (25) 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 the symmetric, semi-positive definite matrix defined by

$$\mathbf{Z}(k) \triangleq \mathbf{z}(k) \mathbf{z}(k)'$$

Using Proposition 3.1, it follows that $\mathbf{Z}(k)$ satisfies the following dynamic equation:

$$\mathbf{Z}(k+1) = \mathbf{W}(k) \mathbf{Z}(k) \mathbf{W}(k)' + \mathbf{F}(k) \quad (27)$$

where $\mathbf{F}(k)$ is given by

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

The solution of (27) is given by

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

For simplicity, in what follows, we will omit the matrix $\mathbf{I} - \mathbf{J}$ from $\mathbf{F}(k)$ since it disappears by multiplication with the transition matrix (see Proposition 3.1). We can further write

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

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

$$E[\|\mathbf{z}(k)\|^2] \leq E[\|\Phi(k, 0)\|^2] \|\mathbf{z}(0)\|^2 + \sum_{s=0}^{k-1} E[\|\Phi(k, s+1) \mathbf{F}(s) \Phi(k, s+1)'\|]. \quad (28)$$

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 sum in the right-hand side of (28). We have

$$\begin{aligned} & \Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)' \\ &= \alpha^2\Phi(k, s+1)\mathbf{d}(s)\mathbf{d}(s)'\Phi(k, s+1)' \\ & \quad - \alpha\Phi(k, s+1)\mathbf{W}(s)\mathbf{z}(s)\mathbf{d}(s)'\Phi(k, s+1)' \\ & \quad - \alpha\Phi(k, s+1)\mathbf{d}(s)\mathbf{z}(s)'\mathbf{W}(s)'\Phi(k, s+1)'. \end{aligned}$$

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

$$\begin{aligned} \Phi(k, s+1)\mathbf{W}(s)\mathbf{z}(s)\mathbf{d}(s)'\Phi(k, s+1)' &= \Phi(k, 0)\mathbf{z}(0)\mathbf{d}(s)'\Phi \\ & \quad \times(k, s+1)' - \alpha \sum_{r=0}^{s-1} \Phi(k, r+1)\mathbf{d}(r)\mathbf{d}(s)'\Phi(k, s+1)'. \end{aligned}$$

Similarly,

$$\begin{aligned} \Phi(k, s+1)\mathbf{d}(s)\mathbf{z}(s)'\mathbf{W}(s)'\Phi(k, s+1)' &= \Phi(k, s+1)\mathbf{d}(s)\mathbf{z}(0)' \\ & \quad \times \Phi(k, 0)' - \alpha \sum_{r=0}^{s-1} \Phi(k, s+1)\mathbf{d}(s)\mathbf{d}(r)'\Phi(k, r+1)'. \end{aligned}$$

We now give a more explicit formula for the matrix product $\Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)'$

$$\begin{aligned} & \Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)' \\ &= \alpha^2\Phi(k, s+1)\mathbf{d}(s)\mathbf{d}(s)'\Phi(k, s+1)' \\ & \quad - \alpha\Phi(k, 0)\mathbf{z}(0)\mathbf{d}(s)'\Phi(k, s+1) \\ & \quad + \alpha^2 \sum_{r=0}^{s-1} \Phi(k, r+1)\mathbf{d}(r)\mathbf{d}(s)'\Phi(k, s+1)' \\ & \quad - \alpha\Phi(k, s+1)\mathbf{d}(s)\mathbf{z}(0)'\Phi(k, 0)' \\ & \quad + \alpha^2 \sum_{r=0}^{s-1} \Phi(k, s+1)\mathbf{d}(s)\mathbf{d}(r)'\Phi(k, r+1)'. \end{aligned}$$

By applying the norm operator, we get

$$\begin{aligned} & \|\Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)'\| \\ & \leq N\alpha^2\varphi^2 \|\Phi(k, s+1)\|^2 \\ & \quad + 2N\alpha^2\varphi^2 \sum_{r=0}^{s-1} \|\Phi(k, r+1)\Phi(k, s+1)'\| \\ & \quad + 2N\alpha\beta\varphi \|\Phi(k, s+1)\Phi(k, 0)'\|. \end{aligned} \quad (29)$$

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

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

and

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

Therefore, we obtain

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

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

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

and

$$\begin{aligned} \sum_{s=0}^{k-1} \rho^{\lfloor \frac{k-s-1}{m} \rfloor} \eta^{\lfloor \frac{s+1}{m} \rfloor} & \leq \sum_{s=0}^{k-1} \rho^{\lfloor \frac{k-s-1}{m} \rfloor} \eta^{\lfloor \frac{s}{m} \rfloor} \\ & \leq m \sum_{s=0}^{\lfloor \frac{k-1}{m} \rfloor} \rho^{\lfloor \frac{k-1}{m} \rfloor - s} \eta^s \\ & = m \frac{\rho^{\lfloor \frac{k-1}{m} \rfloor + 1} - \eta^{\lfloor \frac{k-1}{m} \rfloor + 1}}{\rho - \eta} \end{aligned}$$

we obtain

$$\begin{aligned} \sum_{s=0}^{k-1} E \left[\|\Phi(k, s+1)\mathbf{F}(s)\Phi(k, s+1)'\| \right] & \leq N\alpha^2\varphi^2 \\ & \quad \times \left(1 + \frac{2m}{1 - \eta} \right) \frac{m}{1 - \rho} + 2N\alpha\beta\varphi m \frac{\rho^{\lfloor \frac{k-1}{m} \rfloor + 1} - \eta^{\lfloor \frac{k-1}{m} \rfloor + 1}}{\rho - \eta}. \end{aligned}$$

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

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

The following lemma allows us to interpret $d_i(k)$ as an ϵ -subgradient of f_i at $\bar{x}(k)$ (with ϵ being a random variable).

Lemma 3.3: Let Assumptions 2.2 or 2.3 hold. Then 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 $h(k) = \sum_{i=1}^N d_i(k)$ is an $N\epsilon(k)$ -subdifferential of f at $\bar{x}(k)$, i.e., $h(k) \in \partial_{N\epsilon(k)} f(\bar{x}(k))$, for any $k \geq 0$, where

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

Proof: The proof is somewhat similar to the proof of Lemma 3.4.5 of [11]. Let $\bar{d}_i(k)$ be a subgradient of f_i at $\bar{x}(k)$. By the subgradient definition we have that

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

or

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

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

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

or

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

where $\epsilon(k) = 2\varphi\|\mathbf{z}(k)\|$. 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 $\sum_{i=1}^N d_i(k) \in \partial_{N\epsilon(k)} f(\bar{x}(k))$. Note, that $\epsilon(k)$ has a random characteristic due to the assumptions on $\mathbf{A}(k)$. ■

For twice differentiable cost functions with lower and upper bounded Hessians, the next result gives an upper bound on the second moment of the distance between the average vector $\bar{x}(k)$ and the minimizer of f .

Lemma 3.4: Let Assumptions 2.1 and 2.3 hold and let $\{\bar{x}(k)\}_{k \geq 0}$ be a sequence of vectors defined by iteration (7). Then, the following inequality holds:

$$E \left[\|\bar{x}(k) - x^*\|^2 \right] \leq \|\bar{x}(0) - x^*\|^2 \gamma^k + \frac{4\alpha\varphi\beta\sqrt{N}}{\gamma\eta} \frac{\gamma^k - \eta^{\frac{k}{m}}}{\gamma - \eta^{\frac{1}{m}}} \\ + \frac{\alpha^2\varphi^2}{1-\gamma} \left(4\sqrt{N} \frac{m}{1-\eta} + 1 \right) \quad (31)$$

where $\gamma = 1 - \alpha l$, with $l = \min_i l_i$ and η is defined in Remark 3.2.

Proof: Under Assumption 2.3, $f(x)$ is a strongly convex function with constant Nl , where $l = \min_i l_i$ and therefore it follows that

$$f(x) - f^* \geq \frac{Nl}{2} \|x - x^*\|^2. \quad (32)$$

We use the same idea as in the proof of Proposition 2.4 in [18], formulated under a deterministic setup. By (7), where we use a constant stepsize α , we obtain

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

Using the fact that, by Lemma 3.3, $h(k)$ is a $N\epsilon(k)$ -subdifferential of f at $\bar{x}(k)$, we have

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

or, from inequality (32)

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

Further, we can write

$$\|\bar{x}(k+1) - x^*\|^2 \leq (1 - \alpha l) \|\bar{x}(k) - x^*\|^2 + 2\alpha\epsilon(k) + \alpha^2\varphi^2$$

or

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

Note that from Assumption 2.3-(c), $0 < \alpha < (1/l)$ and therefore the quantity $\gamma^k = (1 - \alpha l)^k$ does not grow unbounded. It follows that

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

From the expression of $\epsilon(k)$ in Lemma 3.3, we immediately obtain the following inequality:

$$E[\epsilon(s)] \leq 2\varphi\beta\sqrt{N}\eta^{\lfloor \frac{s}{m} \rfloor} + \frac{2\alpha\varphi^2\sqrt{N}m}{1-\eta}. \quad (34)$$

The inequality

$$\sum_{s=0}^{k-1} \gamma^{k-1-s} \eta^{\lfloor \frac{s}{m} \rfloor} \leq \gamma^{k-1} \eta^{-1} \sum_{s=0}^{k-1} \left(\frac{\eta^{\frac{1}{m}}}{\gamma} \right)^s = (\gamma\eta)^{-1} \frac{\gamma^k - \eta^{\frac{k}{m}}}{\gamma - \eta^{\frac{1}{m}}}$$

yields

$$\sum_{s=0}^{k-1} \gamma^{k-s-1} E[\epsilon(s)] \leq \frac{2\varphi\beta\sqrt{N}}{\gamma\eta} \frac{\gamma^k - \eta^{\frac{k}{m}}}{\gamma - \eta^{\frac{1}{m}}} + \frac{2\alpha\varphi^2\sqrt{Nm}}{1-\eta} \frac{1}{1-\gamma} \quad (35)$$

which combined with (33), generates the inequality (31). ■

IV. MAIN RESULTS—ERROR BOUNDS

In the following, we provide upper bounds for two performance metrics of the CBMASM. First, we give a bound on the difference between the best recorded value of the cost function f , evaluated at the estimate $x_i(k)$, and the optimal value f^* . Second, we focus on the second moment of the distance between the estimate $x_i(k)$ and the minimizer of f^* . For a particular class of twice differentiable functions, we give an upper bound on this metric and show how fast the time varying part of this bound converges to zero. The bounds we give in these section emphasize the effect of the random topology on the performance metrics.

The following result shows how close the cost function f evaluated at the estimate $x_i(k)$ gets to the optimal value f^* . A similar result for the standard sub-gradient method can be found in [19], for example.

Corollary 4.1: Let Assumptions 2.1 and 2.2 or 2.1 and 2.3 hold and let $\{x_i(k)\}_{k \geq 0}$ be a sequence generated by the iteration (5), $i = 1, \dots, N$. Let $\bar{f}_i^{\text{best}}(k) = \min_{s=0, \dots, k} E[f(x_i(s))]$ be the smallest cost value (in average) achieved by agent i at iteration k . Then

$$\lim_{k \rightarrow \infty} \bar{f}_i^{\text{best}}(k) \leq f^* + 3\alpha\varphi^2 N\sqrt{N} \frac{m}{1-\eta} + \frac{N\alpha\varphi^2}{2}. \quad (36)$$

Proof: Using the subgradient definition of f_i at $x_i(k)$ we have that

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

Summing over all i , we get

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

which holds with probability one. Subtracting f^* from both sides of the above inequality, and applying the expectation operator, we further get

$$E[f(x_i(k))] - f^* \leq E[f(\bar{x}(k))] - f^* + N\varphi E[\|z(k)\|]$$

or

$$\bar{f}_i^{\text{best}}(k) - f^* \leq \min_{s=0, \dots, k} \{E[f(\bar{x}(s))] - f^* + N\varphi E[\|z(s)\|]\}. \quad (37)$$

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

$$\begin{aligned} \|\bar{x}(k+1) - x^*\|^2 &= \left\| \bar{x}(k) - x^* - \frac{\alpha}{N} h(k) \right\|^2 \\ &\leq \|\bar{x}(k) - x^*\|^2 - 2\frac{\alpha}{N} h(k)' (\bar{x}(k) - x^*) \\ &\quad + \alpha^2 \varphi^2 \end{aligned}$$

and since, by Lemma 3.3, $h(\bar{x}(k))$ is a $N\epsilon(k)$ -subdifferential of f at $\bar{x}(k)$, we have

$$\begin{aligned} \|\bar{x}(k+1) - x^*\|^2 &\leq \|\bar{x}(k) - x^*\|^2 \\ &\quad - \frac{2\alpha}{N} (f(\bar{x}(k)) - f^*) + 2\alpha\epsilon(k) + \alpha^2 \varphi^2 \end{aligned}$$

or

$$\begin{aligned} \|\bar{x}(k) - x^*\|^2 &\leq \|\bar{x}(0) - x^*\|^2 - \frac{2\alpha}{N} \sum_{s=0}^{k-1} (f(\bar{x}(s)) - f^*) \\ &\quad + 2\alpha \sum_{s=0}^{k-1} \epsilon(s) + k\alpha^2 \varphi^2. \end{aligned}$$

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

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

or

$$\begin{aligned} \sum_{s=0}^{k-1} (E[f(\bar{x}(s))] - f^*) &\leq \frac{N}{2\alpha} \|\bar{x}(0) - x^*\|^2 \\ &\quad + N \sum_{s=0}^{k-1} E[\epsilon(s)] + \frac{kN\alpha\varphi^2}{2}. \end{aligned}$$

Adding and subtracting $N\varphi E[\|z(s)\|]$ inside the sum of the left-hand side of the above inequality and recalling from Lemma 3.3 that $\epsilon(k) = 2\varphi\|z(k)\|$, we obtain

$$\begin{aligned} \sum_{s=0}^{k-1} (E[f(\bar{x}(s))] - f^* + N\varphi E[\|z(s)\|]) &\leq \frac{1}{2\alpha} \|\bar{x}(0) - x^*\|^2 \\ &\quad + \frac{3N}{2} \sum_{s=0}^{k-1} E[\epsilon(s)] + \frac{kN\alpha\varphi^2}{2}. \end{aligned}$$

Using the fact that

$$\begin{aligned} \sum_{s=0}^{k-1} (E[f(\bar{x}(s))] - f^* + N\varphi E[\|z(s)\|]) \\ \geq k \min_{s=0, \dots, k-1} \{E[f(\bar{x}(s))] - f^* + N\varphi E[\|z(s)\|]\} \end{aligned}$$

we get

$$\begin{aligned} & \min_{s=0, \dots, k-1} \{E[f(\bar{x}(s))] - f^* + N\varphi E[\|z(s)\|]\} \\ & \leq \frac{1}{2\alpha k} \|\bar{x}(0) - x^*\|^2 + \frac{3N}{2k} \sum_{s=0}^{k-1} E[\epsilon(s)] + \frac{N\alpha\varphi^2}{2}. \end{aligned}$$

Using inequality (34) from Lemma 3.3 we obtain

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

It follows that

$$\begin{aligned} & \min_{s=0, \dots, k-1} \{E[f(\bar{x}(s))] - f^* + N\varphi E[\|z(s)\|]\} \\ & \leq \frac{1}{2\alpha k} \|\bar{x}(0) - x^*\|^2 \\ & \quad + \frac{3N}{2k} \left(2\varphi\beta\sqrt{N} \frac{m}{1-\eta} + k2\alpha\varphi^2\sqrt{N} \frac{m}{1-\eta} \right) \\ & \quad + \frac{N\alpha\varphi^2}{2}. \end{aligned} \tag{38}$$

Combining inequalities (37) and (38) and taking the limit, we obtain

$$\lim_{k \rightarrow \infty} \bar{f}_i^{\text{best}}(k) \leq f^* + 3\alpha\varphi^2 N\sqrt{N} \frac{m}{1-\eta} + \frac{N\alpha\varphi^2}{2}.$$

In the case of twice differentiable functions, the next result introduces an error bound which essentially says that the estimates “converge in the mean square sense to within some guaranteed distance” from the optimal point, distance which can be made arbitrarily small by an appropriate choice of the stepsize. In addition, the time-varying component of the error bound converges to zero at least linearly.

Corollary 4.2: Let Assumptions 2.1 and 2.3 hold. Then, for the sequence $\{x_i(k)\}_{k \geq 0}$ generated by iteration (5) we have

$$\limsup_{k \rightarrow \infty} E[\|x_i(k) - x^*\|^2] \leq C_1 + C_2 + 2\sqrt{C_1 C_2} \tag{39}$$

where

$$\begin{aligned} C_1 &= \frac{\alpha^2\varphi^2}{1-\gamma} \left(\frac{4m\sqrt{N}}{1-\eta} + 1 \right) \\ C_2 &= N\alpha^2\varphi^2 \left(1 + \frac{2m}{1-\eta} \right) \frac{m}{1-\rho} \end{aligned} \tag{40}$$

and

$$E[\|x_i(k) - x^*\|^2] \leq \psi(k) + C \tag{41}$$

where $\psi(k) = c\delta^k$ with c a positive constant depending on the initial conditions, $\delta = \max\{\gamma, \eta^{1/m}\}$, $\gamma = 1 - \alpha l$, and where $C = 4 \max\{C_1, C_2\}$.

Proof: By the triangle inequality we have

$$\begin{aligned} E[\|x_i(k) - x^*\|^2] & \leq E[\|x_i(k) - \bar{x}(k)\|^2] \\ & \quad + 2E[\|x_i(k) - \bar{x}(k)\| \|\bar{x}(k) - x^*\|] + E[\|\bar{x}(k) - x^*\|^2]. \end{aligned}$$

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

$$\begin{aligned} E[\|x_i(k) - x^*\|^2] & \leq E[\|x_i(k) - \bar{x}(k)\|^2] + 2E[\|x_i(k) - \bar{x}(k)\|] \\ & \quad \cdot E[\|\bar{x}(k) - x^*\|]^{\frac{1}{2}} + E[\|\bar{x}(k) - x^*\|^2]. \end{aligned} \tag{42}$$

Inequality (31) can be further upper bounded by

$$E[\|\bar{x}(k) - x^*\|^2] \leq \psi_1(k) + C_1$$

where

$$\psi_1(k) = \underbrace{\left[\|\bar{x}(0) - x^*\|^2 + \frac{8\alpha\varphi\beta\sqrt{N}}{\gamma\eta} \frac{1}{\gamma - \eta^{\frac{1}{m}}} \right]}_{c_1} \delta^k = c_1 \delta^k$$

with $\delta = \max\{\gamma, \eta^{1/m}\}$ and C_1 being given in (40). Using the inequalities

$$\rho^{\lfloor \frac{k-1}{m} \rfloor + 1} \leq \rho^{-\frac{1}{m}} \rho^{\frac{k}{m}} \text{ and } \eta^{\lfloor \frac{k-1}{m} \rfloor + 1} \leq \eta^{-\frac{1}{m}} \eta^{\frac{k}{m}}$$

from (26), a new bound for $E[\|x_i(k) - \bar{x}(k)\|^2]$ is given by

$$E[\|x_i(k) - \bar{x}(k)\|^2] \leq \psi_2(k) + C_2$$

where C_2 is given in (40) and

$$\psi_2(k) = \underbrace{\left[N\beta^2\rho^{-1} + \frac{2N\alpha\beta\varphi m}{\eta - \rho} \left(\eta^{-\frac{1}{m}} + \rho^{-\frac{1}{m}} \right) \right]}_{c_2} \delta^k = c_2 \delta^k.$$

Taking the limit of (42) and recalling that under Assumptions 2.1 and 2.3, $\gamma < 1$ and $\eta^{1/m} < 1$ for any $m \in I_m$, we obtain (39).

Inequality (42) can be further upper bounded by

$$\begin{aligned} E[\|x_i(k) - x^*\|^2] & \leq 2 \max\{c_1, c_2\} \delta^k + \\ & \quad 2(\max\{c_1, c_2\} \delta^k + \max\{C_1, C_2\}) = \psi(k) + C \end{aligned}$$

where $\psi(k) = c\delta^k$, with $c = 4 \max\{c_1, c_2\}$ and $C = 4 \max\{C_1, C_2\}$. Hence, we obtained that the time varying component of the error bound converges linearly to zero with a factor $\delta = \max\{\gamma, \eta^{1/m}\}$. ■

A. Discussion of the Results

We obtained upper bounds on two performance metrics relevant to the CBMASM. First we studied the difference between

the cost function evaluated at the estimate and the optimal solution (Corollary 4.1)—for non-differentiable and differentiable functions with bounded (sub)gradients. Second, for a particular class of convex functions (see Assumptions 2.3), we gave an upper bound for the second moment of the distance between the estimates of the agents and the minimizer. We also showed that the time varying component of this upper bound converges linearly to zero with a factor reflecting the contribution of the random topology. We introduced Assumption 2.3 to cover part of the class of convex functions for which uniform boundness of the (sub)gradients cannot be guaranteed.

From our results we can notice that the stepsize has a similar influence as in the case of the standard subgradient method, i.e., a small value of α implies good precision but slow rate of convergence, while a larger value of α increases the rate of convergence but at a cost in accuracy. More importantly, we can emphasize the influence of the consensus step on the performance of the distributed algorithm. When possible, by appropriately designing the probability distribution of the random graph (together with an appropriate choice of the integer m) we can improve the guaranteed precision of the algorithm (intuitively, this means making the quantities $m/(1-\eta)$ and $m/(1-\rho)$ as small as possible). In addition, the rate of convergence of the time varying component of the error bound (41) can be improved by making the quantity $\eta^{1/m}$ as small as possible. Note however that there are limits with respect to the positive effect of the consensus step on the rate of convergence of $\psi(k)$, since the latter is determined by the maximum between γ and $\eta^{1/m}$. Indeed, if the stepsize is small enough, i.e.,

$$\alpha < \frac{1}{l} \left(1 - \eta^{\frac{1}{m}}\right) \quad (43)$$

then the rate of convergence of $\psi(k)$ is given by γ . This suggests that having a fast consensus step will not necessarily be helpful in the case of a small stepsize, which is in accordance with the intuition on the role of a small value of α . In the case where inequality (43) is not satisfied, the rate of convergence of $\psi(k)$ is determined by $\eta^{1/m}$. However, this does not necessarily mean that the estimates will not “converge faster to within some distance of the minimizer,” since we are providing only an error bound.

Assume that we are using the centralized subgradient method to minimize the convex function $f(x) = \sum_{i=1}^N f_i(x)$ satisfying Assumption 2.2 (the subgradients of $f_i(x)$ are uniformly bounded by φ), where the stepsize used is N times smaller than the stepsize of the distributed algorithm, i.e.,

$$x(k+1) = x(k) - \frac{\alpha}{N} d(k)$$

where $d(k)$ is a subgradient of f at $x(k)$, with $\|d(k)\| \leq N\varphi$. Then, from the optimization literature we get

$$\lim_{k \rightarrow \infty} f^{\text{best}}(k) \leq f^* + \frac{N\alpha\varphi^2}{2}$$

where $f^{\text{best}}(k) = \min_{s=0, \dots, k} f(x(s))$. From above we note that, compared with the centralized subgradient method with a step size N times smaller than the agents' stepsize, the distributed optimization algorithm introduced an additional term

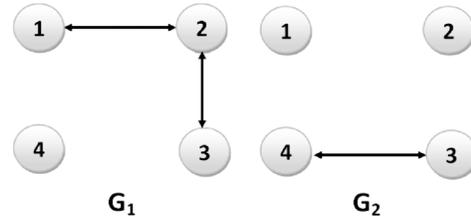


Fig. 1. Sample space of the random graph $G(k)$.

in the error bound given by $3\alpha\varphi^2 N \sqrt{N} (m/(1-\eta))$, which reflects the influence of the dimension of the network and of the random topology on the guaranteed accuracy of the algorithm.

Let us now assume that we are minimizing the function $f(x)$, satisfying Assumptions 2.3-(a)(b), using a centralized gradient algorithm

$$x(k+1) = x(k) - \frac{\alpha}{N} \nabla f(x(k))$$

where we have that α is small enough ($0 < \alpha < (2/L)$) so that the algorithm is stable and there exists φ_c so that $\|\nabla f_i(x(k))\| \leq \varphi_c$. It follows that we can get the following upper bound on the distance between the estimate of the optimal decision vector and the minimizer

$$\|x(k) - x^*\|^2 \leq \|x(0) - x^*\|^2 \gamma_c^k + \frac{\alpha\varphi_c^2}{l}$$

with $\gamma_c = 1 - \alpha l$. Therefore, we can see that $\gamma = \gamma_c$ which shows that the rates of convergence, at which the time-varying components of the error bounds converge to zero in the centralized and distributed cases, are the same. However, note that we assumed the stepsize in the centralized case to be N times smaller than the stepsize used by the agents.

The error bounds (36) and (41) are functions of three quantities induced by the consensus step: $m/(1-\eta)$, $m/(1-\rho)$ and $\eta^{1/m}$. These quantities show the dependence of the performance metrics on the pmf of $G(k)$ and on the corresponding random matrix $A(k)$. The scalars η and ρ represent the first and second moments of the SLEM of the random matrix $A(k+1), \dots, A(k+m)$, corresponding to a random graph formed over a time interval of length m , respectively. We notice from our results that the performance of the CBMASM is improved by making $m/(1-\eta)$, $m/(1-\rho)$ and $\eta^{1/m}$ as small as possible, i.e., by optimizing these quantities having as decision variables m and the pmf of $G(k)$. For instance if we are interested in obtaining a tight bound on $E[\|x_i(k) - x^*\|^2]$ and having a fast decrease to zero of $\psi(k)$, we can formulate the following multi-criteria optimization problem:

$$\begin{aligned} \min_{m, p_i} & \quad \left\{ \eta^{\frac{1}{m}}, C_1 + C_2 + 2\sqrt{C_1 C_2} \right\} \\ \text{subject to :} & \quad m \geq 1 \\ & \quad \eta^{\frac{1}{m}} \geq \gamma \\ & \quad \sum_i^M p_i = 1, p_i \geq 0. \end{aligned} \quad (44)$$

where C_1 and C_2 were defined in (40). The second inequality constraint was added to emphasize the fact that making $\eta^{1/m}$ too small is pointless, since that rate of convergence of $\psi(k)$ is limited by γ . If we are simultaneously interested in tightening

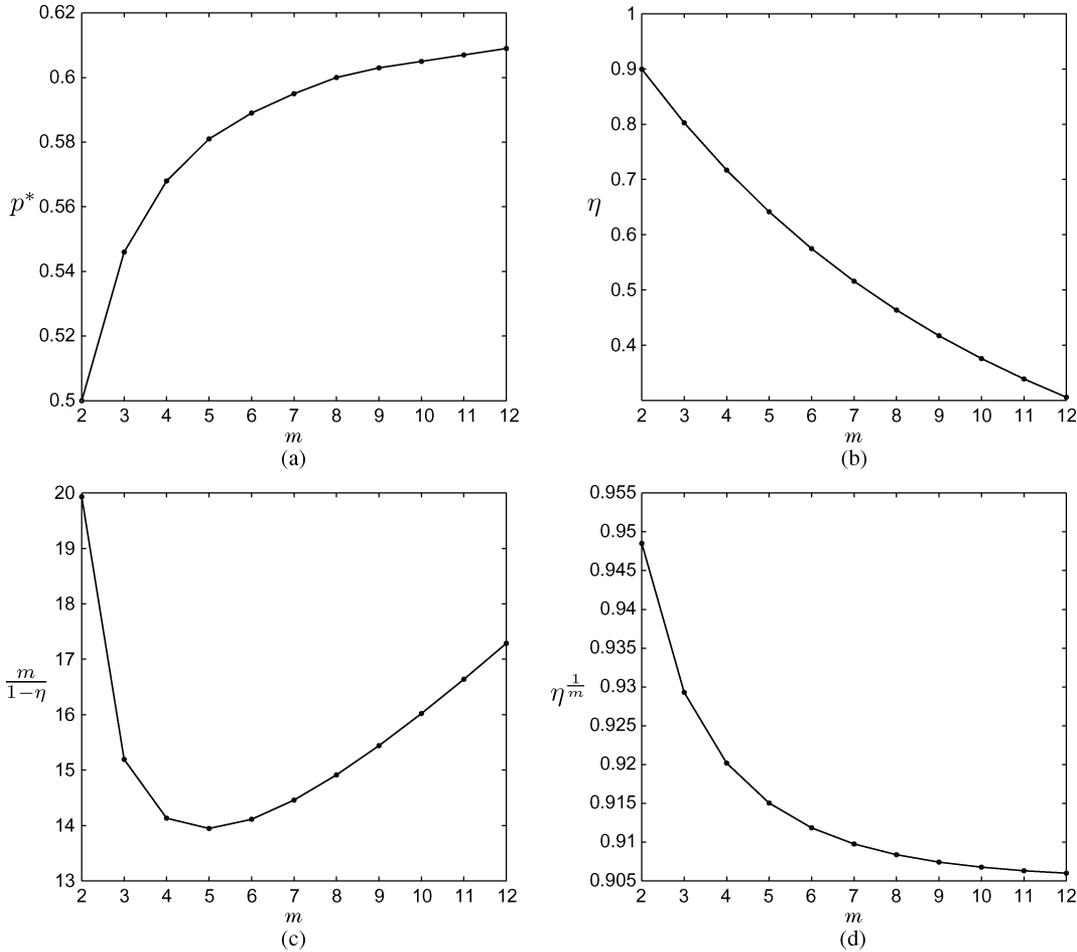


Fig. 2. (a) Optimal p as a function of m . (b) Optimized η as a function of m . (c) Optimized $m/(1 - \eta)$ as a function of m . (d) Optimized $\eta^{1/m}$ as a function of m .

the upper bounds of both metrics, we can introduce the quantity $m/(1 - \eta)$ in the optimization problem since $m/(1 - \eta)$ and $m/(1 - \rho)$ are not necessarily minimized by the same probability distribution. 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. An approximate solution of (44) can be obtained by focusing only on minimizing $m/(1 - \eta)$, since both $\eta^{1/m}$ and $m/(1 - \rho)$ are upper bounded by this quantity. Therefore, an approximate solution can be obtained by minimizing η (i.e., computing the optimal pmf) for each value of m , and then picking the best value m with the corresponding η that minimizes $m/(1 - \eta)$. 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.), and therefore we can potentially tune these parameters so that the performance of the CBMASM is improved.

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

Example 4.1: Let $G(k)$ be a random graph process 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 Fig. 1. Also, let $A(k)$ be a (stochastic) random matrix, corresponding to $G(k)$, taking values in the set $\mathcal{A} = \{A_1, A_2\}$, with

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

Fig. 2(a) shows the optimal probability p^* that minimizes η for different values of m . Fig. 2(b) shows the *optimized* η (computed at p^*) as a function of m . Figs. 2(c) and 2(d) show the evolution of the *optimized* $m/(1 - \eta)$ and $\eta^{1/m}$ 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 (44), we need to compute the probability of all possible sequences of length m produced by $G(k)$, 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 the

bounds on η and ρ introduced in (16) and (17), respectively. Note that every result concerning 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 the union of random graphs obtained over a time interval of length m .

Even in the case where we use 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 that produce connected graphs). Another way to simplify our problem even more, is to (intelligently) fix a value for m and 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 $m/(1-\eta)$ is lower bounded by m . Although in general the uniform distribution does not necessarily minimize η , it becomes the optimizer under some particular assumptions, stated in what follows. Let \mathcal{G} be such that a connected graph can 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 the 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., $p_i = 1/m$, for $i = 1, \dots, M$.

V. APPLICATION—DISTRIBUTED SYSTEM IDENTIFICATION

In this section, we show how the distributed optimization algorithm analyzed in the previous section can be used to perform collaborative system identification. We assume the following scenario: a group of sensors track an object by taking measurements of its position. These sensors have memory and computation capabilities and are organized in a communication network modeled by a random graph process $G(k)$ satisfying the assumptions introduced in Section II. The task of the sensors/agents is to determine a parametric model of the object's trajectory. The measurements are affected by noise, whose effect may differ from sensor to sensor (i.e., some sensors take more accurate measurements than others). This can happen for instance when some sensors are closer to the object than other (allowing a better reading of the position), or sensors with different precision classes are used. Determining a model for the time evolution of the object's position can be useful in motion prediction when the motion dynamics of the object in unknown to the sensors. The notations used in the following are independent from the ones used in the previous sections.

A. System Identification Model

Let $p(t)' = [x(t), y(t), z(t)]$ be the position vector of the tracked object. We model the time evolution of each of the axis

of the position vector as a time dependent polynomial of degree n_a , i.e.,

$$\begin{aligned} x(t) &= a_0^x + a_1^x t + \dots + a_{n_a}^x t^{n_a} \\ y(t) &= a_0^y + a_1^y t + \dots + a_{n_a}^y t^{n_a} \\ z(t) &= a_0^z + a_1^z t + \dots + a_{n_a}^z t^{n_a}. \end{aligned} \quad (45)$$

The measurements of each sensor i are given by

$$\begin{aligned} x_i(t) &= x(t) + e_{i,x}(t) \\ y_i(t) &= y(t) + e_{i,y}(t) \\ z_i(t) &= z(t) + e_{i,z}(t) \end{aligned} \quad (46)$$

where $e_{i,x}(t)$, $e_{i,y}(t)$, and $e_{i,z}(t)$ are assumed white noises of (unknown) variances $\sigma_{i,x}^2$, $\sigma_{i,y}^2$, and $\sigma_{i,z}^2$, respectively. Equivalently, we have

$$\begin{aligned} x_i(t) &= \varphi(t)' \theta_x + e_{i,x}(t) \\ y_i(t) &= \varphi(t)' \theta_y + e_{i,y}(t) \\ z_i(t) &= \varphi(t)' \theta_z + e_{i,z}(t) \end{aligned} \quad (47)$$

where $\varphi(t)' = [1, t, \dots, t^{n_a}]$ and $\theta_x = [a_{0,x}, \dots, a_{n_a,x}]'$, $\theta_y = [a_{0,y}, \dots, a_{n_a,y}]'$, and $\theta_z = [a_{0,z}, \dots, a_{n_a,z}]'$.

In the following, we focus only on one coordinate of the position vector, say $x(t)$. The analysis, however can be mimicked in a similar way for the other two coordinates. Let T be the total number of measurements taken by the sensors and consider the following quadratic cost functions

$$\mathcal{J}_i(\theta_x) = \sum_{t=1}^T (x_i(t) - \varphi(t)' \theta_x)^2, \quad \forall i.$$

Using its own measurements, sensor i can determine a parametric model for the time evolution of the coordinate $x(t)$ by solving the optimization problem

$$\min_{\theta_x} \mathcal{J}_i(\theta_x). \quad (48)$$

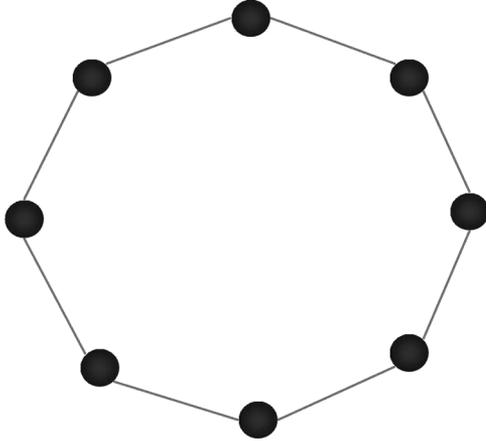
Let $\mathbf{X}_i' = [x_i(1), \dots, x_i(T)]$ be the vector of measurements of sensor i and let $\Phi' = [\varphi(1), \dots, \varphi(T)]$ be the matrix formed by the regression vectors. It is well known that the optimal solution of (48) is given by

$$\hat{\theta}_{i,x} = (\Phi' \Phi)^{-1} \Phi' \mathbf{X}_i. \quad (49)$$

Remark 5.1: It can be shown that $\Phi' \Phi$ is invertible for any T , but it becomes ill conditioned for large values of T . That is why, for our numerical simulations, we will in fact use an orthogonal basis to model the time evolution of the coordinates $x(t)$, $y(t)$, and $z(t)$.

Performing a localized system identification does not take into account the measurements of the other sensors, which can potentially enhance the identified model. If all the measurements are centralized, a model for the time evolution of $x(t)$ can be computed by solving

$$\min_{\theta_x} \mathcal{J}(\theta_x)$$

Fig. 3. Circular graph with $N = 8$.

where

$$\mathcal{J}(\theta_x) = \sum_{i=1}^N \mathcal{J}_i(\theta_x). \quad (50)$$

Note that (50) fits the framework of the distributed optimization problem formulated in the previous sections, and therefore can be solved distributively, eliminating the need for sharing all measurements with all other sensors.

Remark 5.2: If each sensor has *a priori* information about its accuracy, then the cost function (50) can be replaced with

$$\mathcal{J}(\theta_x) = \sum_{i=1}^N \delta_{i,x} \mathcal{J}_i(\theta_x) \quad (51)$$

where $\delta_{i,x}$ is a positive scalar such that the more accurate sensor i is, the larger δ_i is. The scalar $\delta_{i,x}$ can be interpreted as *trust* in the measurements taken by sensor i . The sensors can use local identification to compute $\delta_{i,x}$. For instance, $\delta_{i,x}$ can be chosen as $\delta_{i,x} = 1/\hat{\sigma}_{i,x}^2$, where $\hat{\sigma}_{i,x}^2$ is given by

$$\hat{\sigma}_{i,x}^2 = \frac{1}{T} \sum_{t=1}^T \left(x_i(t) - \varphi(t)' \hat{\theta}_{i,x} \right)^2$$

where $\hat{\theta}_{i,x}$ is the local estimate of the model for the time evolution of $x(t)$.

The distributed optimization algorithm (5) can be written as

$$\theta_{i,x}(k+1) = \sum_{j=1}^N a_{ij}(k) \theta_{j,x}(k) - \alpha \nabla \mathcal{J}_i(k) \quad (52)$$

where $\nabla \mathcal{J}_i(k) = -2\Phi'(\mathbf{X}_i - \Phi\theta_{i,x}(k))$.

B. Numerical Simulations

In this section, we simulate the distributed system identification algorithm under two gossip communication protocols: the randomized gossip protocol and the (modified) broadcast gossip protocol. We perform the simulations on a circular graph, where we assume that the cardinality of the neighborhoods of the nodes is two (see Fig. 3). This graph is a particular example of small world graphs [27] (for an analysis of the consensus problem under small world like communication topologies, the reader can consult [2] for example).

In the case of the randomized gossip protocol, the set of consensus matrices is given by

$$\mathcal{A}^r = \{A_{ij}, i = 1, \dots, N, j \in \{i-1, i+1\}\}$$

where $A_{ij} = I - (1/2)(e_i - e_j)(e_i - e_j)'$ and where by convention we assume that if $i = N$ then $i+1 = 1$ and if $i = 1$ then $i-1 = N$. We assume that if node i wakes up, it chooses with uniform distribution between its two neighbors. Hence, the probability distribution of the random matrix $A(k)$ is given by

$$Pr(A(k) = A_{ij}) = \frac{1}{2N}.$$

We note that the minimum value of m such that $\eta_m < 1$ is $N-1$. Recall that m is the length of a time interval such that $Pr(\bigcup_{l=0}^{m-1} G(k+l)) > 0$ for any k . It turns out that for $m = N-1$

$$p_c^r = Pr\left(\bigcup_{l=0}^{N-2} G(k+l)\right) = N! \left(\frac{1}{2N}\right)^{N-1}$$

Interestingly, the matrix products of length $N-1$ of the form $\prod_{i=1}^{N-1} A_{i+i_0, i+1+i_0}$ with $i_0 \in \{0, \dots, N-1\}$, and the matrix products that may be obtained by the permutations of the matrices in the aforementioned matrix products, have the same SLEM (where the summations in the indices are seen as modulo N). In fact, it is exactly this property that allows us to give the following explicit expression for η_{N-1}

$$\eta_{N-1}^r = p_c^r \lambda^r + 1 - p_c^r \quad (53)$$

where λ^r is the SLEM of the matrix product $A_{1,2}A_{2,3} \cdots A_{N-1,N}$.

In the case of the (modified) broadcast gossip protocol, the set \mathcal{A} is given by

$$\mathcal{A}^b = \{A_i, i = 1, \dots, N\}$$

where $A_i = I - (1/3)[(e_i - e_{i+1})(e_i - e_{i+1})' + (e_i - e_{i-1})(e_i - e_{i-1})']$ and $Pr(A(k) = A_i) = 1/N$. For odd values of N (and $N \geq 3$), the minimum value of m such that $\eta_m < 1$ is given by $m = (N-1)/2$. In addition, we have that

$$p_c^b = Pr\left(\bigcup_{l=0}^{\frac{N-3}{2}} G(k+l)\right) = N \left(\frac{N-1}{2}\right)! \left(\frac{1}{N}\right)^{\frac{N-1}{2}}.$$

Observing a similar phenomenon as in the case of the randomized gossip protocol, namely that the matrix products $A_{1+i_0}A_{3+i_0}, \dots, A_{N-2+i_0}$ for $i_0 \in \{0, \dots, N-1\}$ and their permutations have the same SLEM (where as before the summations of indices are seen as modulo N), we obtain the formula

$$\eta_{\frac{N-1}{2}}^b = p_c^b \lambda^b + 1 - p_c^b$$

where λ^b is the SLEM of the matrix product $A_1A_3 \cdots A_{N-2}$.

The values for η_{N-1}^r and $\eta_{(N-1)/2}^b$ computed above, in the case of the two gossip protocols, do not necessarily provide tight error bounds, since we considered minimal time interval lengths so that $\eta_m < 1$. Even for this relatively simple type of graph,

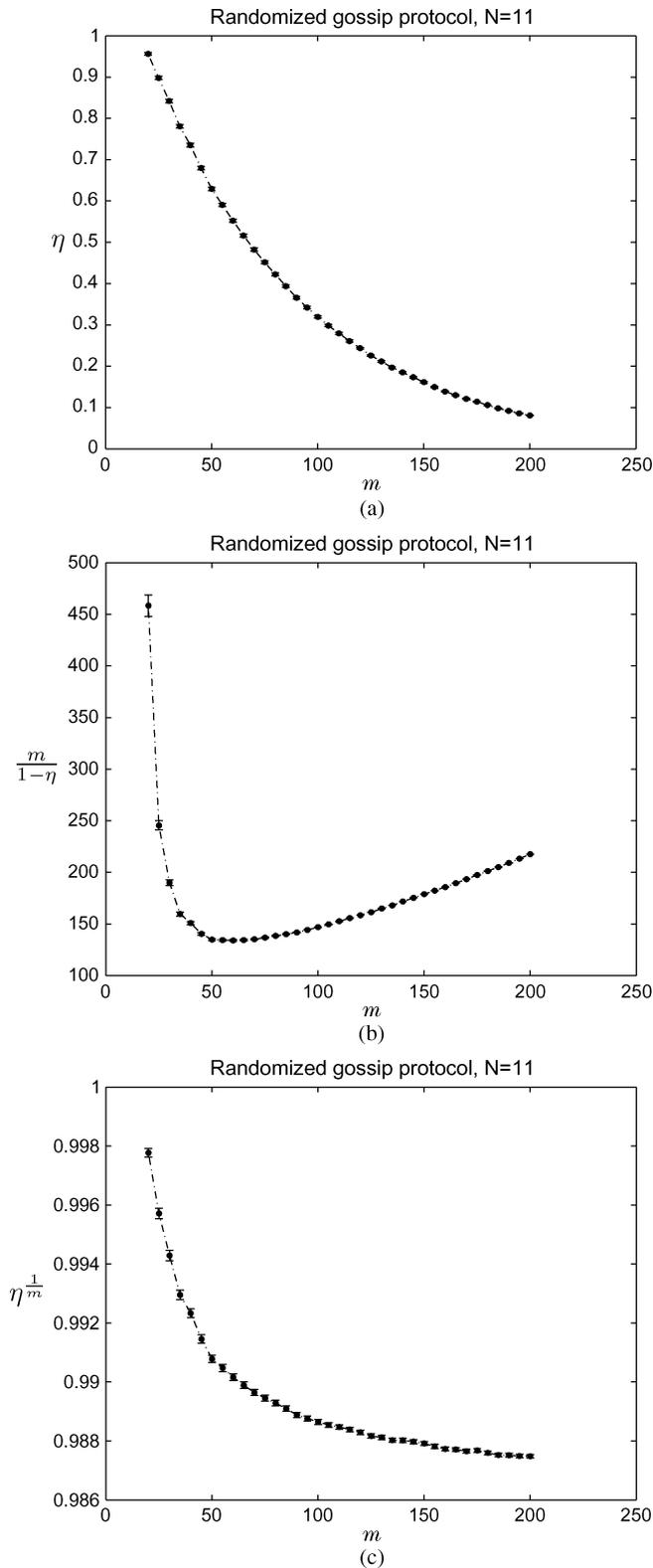


Fig. 4. Estimates of η , $m/(1-\eta)$ and $\eta^{1/m}$ for the randomized gossip protocol and for $N = 11$.

analytical formulas for η_m , for large values of m , are more difficult to obtain due to an increase in combinatorial complexity and because different matrix products that appear in the expression of η do not necessarily have the same SLEM. However, we did compute numerical estimates for different values of m .

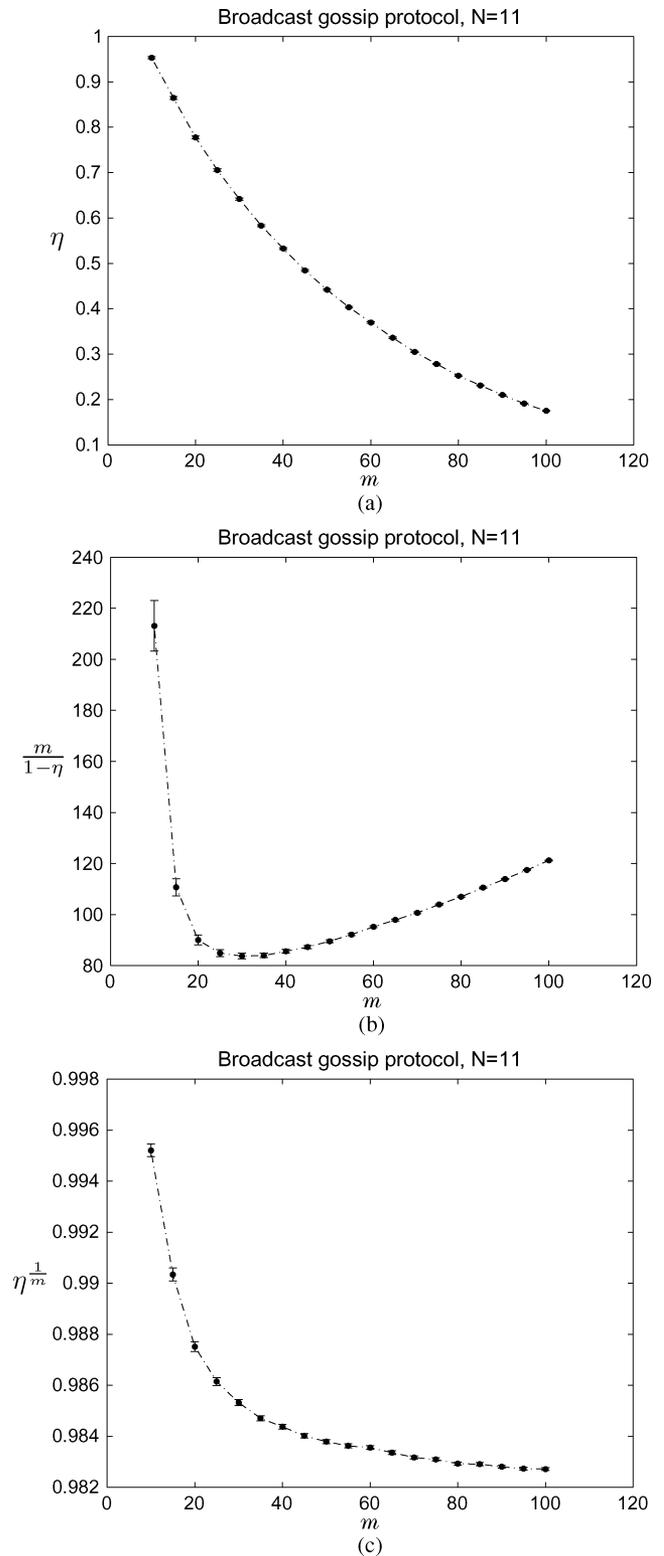


Fig. 5. Estimates of η , $m/(1-\eta)$, and $\eta^{1/m}$ for the (modified) broadcast gossip protocol and for $N = 11$.

Figs. 4 and 5 show estimates of the three quantities of interest, η , $m/(1-\eta)$ and $\eta^{1/m}$, as functions of m , for $N = 11$ (the estimates were computed by taking averages over 2000 realizations and are shown together with the 95% confidence intervals). We can see that $m/(1-\eta)$ is minimized for $m \approx 55$ in the case of the randomized gossip protocol and for $m \approx 30$ in the

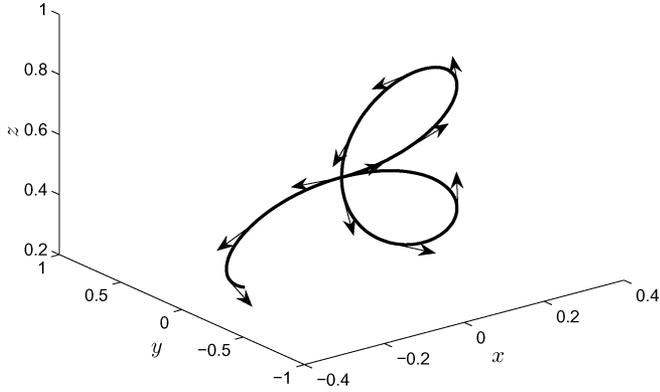


Fig. 6. Trajectory of the object.

case of the broadcast gossip protocol, while the best achievable $\eta^{1/m}$ are approximately equal for the two protocols, (i.e., 0.985. for the randomized gossip protocol and 0.982 for the broadcast gossip protocol).

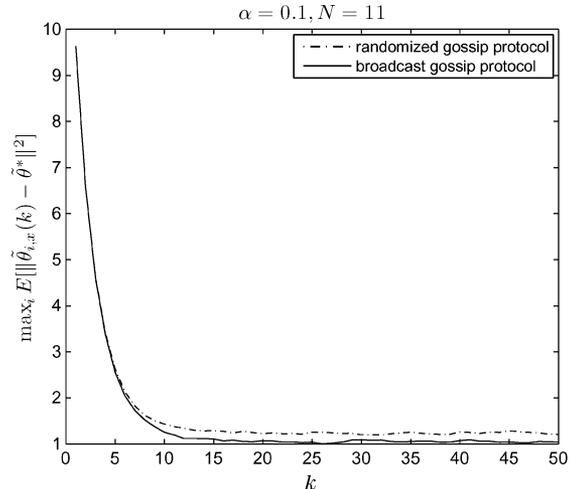
Next we present numerical simulations of the distributed system identification algorithm presented in the previous subsection, under the randomized and broadcast gossip protocols. We would like to point out that, in order to maintain numerical stability, in our numerical simulation we used an orthogonalized version of Φ , given by $\tilde{\Phi} = \Phi H$, where $\tilde{\Phi}$'s columns form an orthogonal basis of the range of Φ , and the new vector of the parameters is given $\tilde{\theta} = H\theta$, where H is a linear transformation matrix, whose entries depend on the orthogonalization process used (Gram–Schmidt, Householder transformations, etc.). Therefore, the cost function we are minimizing can be rewritten as

$$\mathcal{J}(\tilde{\theta}_x) = \sum_{i=1}^n \mathcal{J}_i(\tilde{\theta}_x)$$

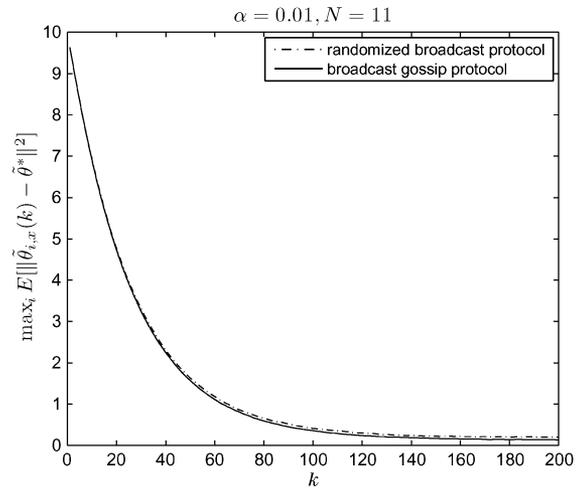
where $\mathcal{J}_i(\tilde{\theta}_x) = \|\mathbf{X}_i - \tilde{\Phi}\tilde{\theta}_x\|^2$.

It is easy to check that in the case of the two protocols, $\underline{\lambda}$ (the smallest of all eigenvalues of matrices belonging to the set \mathcal{A}) is zero. In addition, Assumption 2.3-(a)(b) are satisfied for $l_i = L_i = 2$, and for $\alpha < (1/2)$ the distributed optimization algorithm is guaranteed to be stable with probability one (recall Lemma 3.1). From above we see that $\eta^{1/m}$ cannot attain a value less than 0.98 for both protocols, for any m . Therefore, although we can choose $\alpha > 0.01$, which in turn implies $\gamma < 0.98$, our analysis cannot guarantee a rate of convergence for $\psi(k)$ smaller than 0.98, since the rate of convergence is upper bounded by the maximum between γ and $\eta^{1/m}$. However, this does not mean that faster rates of convergence can not be achieved, which in fact is shown in our numerical simulations.

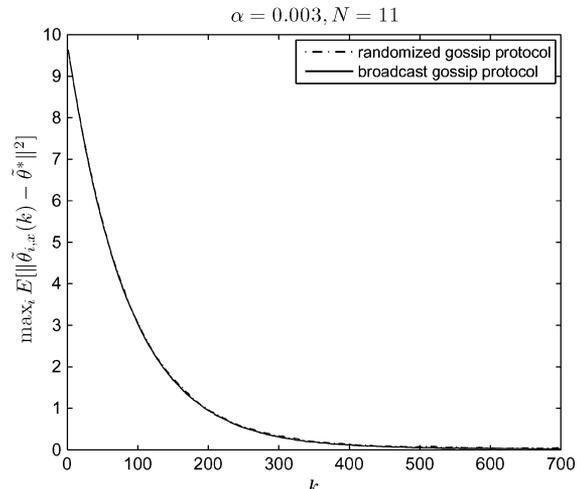
In our numerical experiments we considered a number $T = 786$ of measurements of the x -coordinate of the trajectory depicted in Fig. 6. Figs. 7 and 8 present numerical simulations of the distributed system identification algorithm for the two protocols and for a circular graph with $N = 11$. We assumed that the x -coordinate measurements are affected by white, Gaussian noise with a signal-to-noise ration given by $SNR_i = 5 \times i$ dB, for $i = 1, \dots, 11$. The time polynomials modeling the trajectory evolution are chosen of degree ten, i.e., $n_{\alpha_x} = 10$. We plot estimates of two metrics: $\max_i E[\|\tilde{\theta}_{i,x}(k) - \tilde{\theta}^*\|]$ and



(a)



(b)



(c)

Fig. 7. Estimate of $\max_i E[\|\tilde{\theta}_{i,x}(k) - \tilde{\theta}^*\|]$ for the randomized and broadcast gossip protocols.

$\max_i E[f(\tilde{\theta}_{i,x}(k))] - f^*$ for different values of α (the estimates were computed by taking averages over 500 realizations). We note that for larger values of α , under the two protocols, the algorithm has roughly the same rate of convergence, but the broadcast protocol is more accurate. This is in accordance with

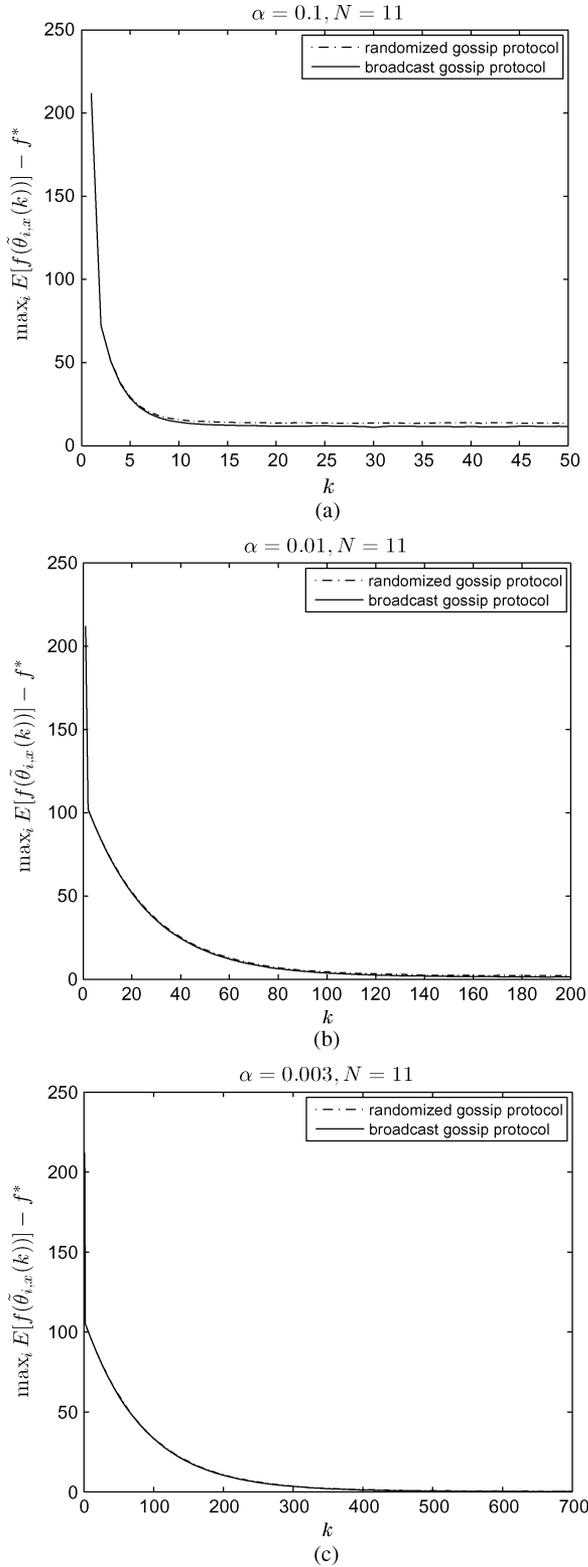


Fig. 8. Estimate of $\max_i E[f(\bar{\theta}_{i,x}(k))] - f^*$ for the randomized and broadcast protocol gossip protocols.

our analysis, since as Figs. 4 and 5 show, $(m/(1 - \eta^b)) \leq (m/(1 - \eta^r))$ for any m , quantities which control the guaranteed accuracy. For smaller values of α , under both protocols the algorithm becomes more accurate and the rate of convergence

decreases since the parameter γ becomes larger and therefore dominant.

VI. CONCLUSION

In this paper, 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 two performance metrics related to the CBMASM. The first metric reflects how close each agent can get to the optimal value. The second metric reflects how close and fast the agents' estimates of the decision vector can get to the minimizer of the objective function, and it was analyzed for a particular class of convex functions. All the aforementioned performance measures were expressed in terms of the probability distribution of the random communication topology. In addition, we showed how the distributed optimization algorithm can be used to perform collaborative system identification, an application which can be useful in collaborative tracking.

REFERENCES

- [1] T. Ayasal, M. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.
- [2] J. S. Baras and P. Hovareshti, "Effects of topology in networked systems: Stochastic methods and small worlds," in *Proc. 47th IEEE Conf. Decision Control*, Cancun, Mexico, Dec. 9–11, 2008, pp. 2973–2978.
- [3] S. Boyd, P. Diaconis, and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Rev.*, vol. 46, no. 4, pp. 667–689, Dec. 2004.
- [4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, 2006.
- [5] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," Mar. 27, 2010, arXiv:1003.5309v1 [cs.DC].
- [6] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Distributed dual averaging in networks," in *Proc. 2010 Neural Inf. Syst. Foundat. Conf.*, Vancouver, BC, Canada, Dec. 2010.
- [7] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1456–1476, Sep. 2004.
- [8] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 998–1001, Jun. 2003.
- [9] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, "Subgradient methods and consensus algorithms for solving convex optimization problems," in *Proc. 47th IEEE Conf. Decision Control*, Cancun, Mexico, Dec. 2008, pp. 4185–4190.
- [10] B. Johansson, M. Rabi, and K. H. Johansson, "A randomized incremental subgradient method for distributed optimization in networked systems," *SIAM J. Optimizat.*, vol. 20, no. 3, pp. 1157–1170, 2009.
- [11] B. Johansson, "On distributed optimization in network systems," Doctoral dissertation in telecommunication, Royal Inst. of Technol., Stockholm, Sweden, 2008.
- [12] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1867–1872, Nov. 2005.
- [13] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2002.
- [14] I. Lobel and A. Ozdaglar, "Distributed subgradient methods over random networks," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, Sep. 2008.
- [15] I. Matei, N. Martins, and J. Baras, "Almost sure convergence to consensus in Markovian random graphs," in *Proc. 47th IEEE Conf. Decision Control*, Cancun, Mexico, Dec. 9–11, 2008, pp. 3535–3540.
- [16] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Trans. Autom. Control*, vol. 50, no. 2, pp. 169–182, Feb. 2005.

- [17] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
- [18] A. Nedic, "Convergence rate of incremental subgradient algorithm," in *Stochastic Optimization: Algorithms and Applications*. Norwell, MA: Kluwer, pp. 263–304.
- [19] A. Nedic and D. P. Bertsekas, "Incremental subgradient methods for nondifferential optimization," *SIAM J. Optimiz.*, vol. 12, pp. 109–138, 2001.
- [20] C. T. K. Ng, M. Medard, and A. Ozdaglar, "Completion time minimization and robust power control in wireless packet networks," Dec. 18, 2008, arXiv:0812.3447v1 [cs.IT].
- [21] B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software, 1987.
- [22] W. Ren and R. W. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655–661, May 2005.
- [23] A. T. Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Trans. Autom. Control*, vol. 53, no. 3, pp. 791–795, Apr. 2008.
- [24] A. T. Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 225–230, Jan. 2010.
- [25] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Dept. Elect. Eng., Mass. Inst. Technol., Cambridge, MA, Nov. 1984.
- [26] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
- [27] D. J. Watts and S. H. Strogatz, "Collective dynamics of small world networks," *Nature*, vol. 393, pp. 440–442, 1998.



Ion Matei received the B.S. degree in electrical engineering (with highest distinction) and the M.S. degree in electrical engineering from the Politehnica University of Bucharest, Bucharest, Romania, in 2002 and 2003, respectively, and the M.S. and Ph.D. degrees in electrical engineering from the University of Maryland, College Park, MD, in 2009 and 2010, respectively.

He is currently a Research Associate in the Systems Engineering Group, National Institute of Standards and Technology, Gaithersburg, MD. His research interests include stochastic, hybrid, and multi-agent control systems and more recently model-based systems engineering.



John S. Baras (M'73–SM'83–F'84) received the B.S. degree in electrical engineering (with highest distinction) from the National Technical University of Athens, Athens, Greece, in 1970 and the M.S. and Ph.D. degrees in applied mathematics from Harvard University, Cambridge, MA, in 1971 and 1973, respectively.

Since 1973, he has been with the Department of Electrical and Computer Engineering, University of Maryland, College Park, where he is currently a Professor, member of the Applied Mathematics and Scientific Computation Program Faculty, and Affiliate Professor in the Fischell Department of Bioengineering. From 1985 to 1991, he was the Founding Director of the Institute for Systems Research (ISR) (one of the first six NSF Engineering Research Centers). In February 1990, he was appointed to the Lockheed Martin Chair in Systems Engineering. Since 1991, he has been the Director of the Maryland Center for Hybrid Networks (HYNET), which he cofounded. Dr. Baras has held visiting research scholar positions with Stanford, MIT, Harvard, the Institute National de Recherche en Informatique et en Automatique, the University of California at Berkeley, Linköping University, and the Royal Institute of Technology in Sweden. His research interests include control, communication, and computing systems. He has published more than 600 refereed publications, holds five patents, graduated 64 Ph.D. students and sponsored 45 postdoctoral scholars.

Dr. Baras received the 1980 George S. Axelby Prize of the IEEE Control Systems Society, the 1978, 1983, and 1993 Alan Berman Research Publication Award from NRL; the 1991, 1994, and 2008 Outstanding Invention of the Year Award from the University of Maryland, the 1996 Engineering Research Center Award of Excellence for Outstanding Contributions in Advancing Maryland Industry, the 1998 Mancur Olson Research Achievement Award from the University of Maryland, College Park, the 2002 Best Paper Award for IT/C4ISR at the 23rd Army Science Conference, the 2004 Best Paper Award at the Wireless Security Conference WISE04, the 2007 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communication Systems, the Globecom'08 Best Paper Award for Wireless Networks, and the 2008 Best Paper Award for IT/C4ISR at the 26th Army Science Conference. He is a Foreign Member of the Royal Swedish Academy of Engineering Sciences (IVA).