Randomized optimal consensus of multi-agent systems[☆]

Guodong Shi [1], Karl Henrik Johansson

Automatic Control Lab, ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology (KTH), Stockholm 100 44, Sweden

ARTICLE INFO

ABSTRACT

In this paper, we formulate and solve a randomized optimal consensus problem for multi-agent systems with stochastically time-varying interconnection topology. The considered multi-agent system with a simple randomized iterating rule achieves an almost sure consensus meanwhile solving the optimization problem $\min_{z \in \mathbb{R}^d} \sum_{i=1}^{n} f_i(z)$, in which the optimal solution set of objective function f_i can only be observed by agent i itself. At each time step, simply determined by a Bernoulli trial, each agent independently and randomly chooses either taking an average among its neighbor set, or projecting onto the optimal solution set of its own optimization component. Both directed and bidirectional communication graphs are studied. Connectivity conditions are proposed to guarantee an optimal consensus almost surely with proper convexity and intersection assumptions. The convergence analysis is carried out using convex analysis. We compare the randomized algorithm with the deterministic one via a numerical example. The results illustrate that a group of autonomous agents can reach an optimal opinion by each node simply making a randomized trade-off between following its neighbors or sticking to its own opinion at each time step.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In recent years, there have been considerable research efforts on multi-agent dynamics in application areas such as engineering, natural science, and social science. Cooperative control of multi-agent systems is an active research topic, where collective tasks are enabled by the recent developments of distributed control protocols via interconnected communication (Cao, Morse, & Anderson, 2008a,b; Jadbabaie, Lin, & Morse, 2003; Martinez, Cortés, & Bullo, 2007; Moreau, 2005; Olfati-Saber, 2006; Olfati-Saber & Murray, 2004; Ren & Beard, 2008; Tanner, Jadbabaie, & Pappas, 2007; Tsitsiklis, Bertsekas, & Athans, 1986; Xiao & Wang, 2008). However, fundamental difficulties remain in the search of suitable tools to describe and design the dynamical behavior of these systems and thus to provide insights in their basic principles. Unlike what is often the case in classical control design, multi-agent control systems aim at fully exploiting, rather than attenuating, the interconnection between subsystems. The distributed nature of the information processing and control requires completely new approaches to analysis and synthesis.

Consensus is a central problem in the study of multi-agent systems, which usually requires that all the agents achieve the same state, e.g., a certain relative position or velocity. Efforts have been devoted to characterize the fundamental link between agent dynamics and group coordination, in which the connectivity of the multi-agent network plays a key role. Switching topologies in different cases, and the "joint connection" or similar concepts are important in the analysis of stability and convergence. Uniform joint-connection, i.e., the joint graph is connected during all intervals which are longer than a constant, has been employed for various consensus problems (Fax & Murray, 2004; Hong, Hu, & Gao, 2006; Jadbabaie et al., 2003; Lin, Francis, & Maggiore, 2007; Nedić, Olshevsky, Ozdaglar, & Tsitsiklis, 2009; Ren & Beard, 2005; Tsitsiklis et al., 1986). On the other hand, $[t, \infty)$-joint connectedness, i.e., the joint graph is connected in time intervals $[t, \infty)$, is the most general form to secure the global coordination, which is also proved to be necessary in many situations (Moreau, 2005; Shi & Hong, 2009; Shi, Hong, & Johansson, 2012). Moreover, consensus seeking over randomly varying networks has been proposed in the literature (Acemoglu, Ozdaglar, & ParandehGheibi, 2010; Boyd, Ghosh, Prabhakar, & Shah, 2006; Fagnani & Zampieri, 2008; Hatano & Mesbahi, 2005; Tahbaz-Salehi & Jadbabaie, 2008), in which the communication graph is usually modeled a sequence of i.i.d. random variables over time.

Minimizing a sum of functions, $\sum_{i=1}^{n} f_i(z)$, using distributed algorithms, where each component function f_i is known only to a particular agent i, has attracted much attention in recent years, due to its wide application in multi-agent systems and resource allocation in wireless networks (Johansson, Keviczky, Johansson, &

Johansson, 2008; Johansson, Rabi, & Johansson, 2009; Lu, Regier, & Tang, 2010; Lu, Tang, Regier, & Bow, 2010; Rabbat & Nowak, 2004; Ram, Nedić, & Veeravalli, 2007). A class of subgradient-based incremental algorithms when some estimate of the optimal solution can be passed over the network via deterministic or randomized iteration, were studied in Johansson et al. (2009), Rabbat and Nowak (2004) and Ram, Nedić, and Veeravalli (2009). Then in Lu et al. (2010) a non-gradient-based algorithm was proposed, where each node starts at its own optimal solution and updates using a pairwise equalizing protocol. The local information transmitted over the neighborhood is usually limited to a convex combination of its neighbors (Jadbabaie et al., 2003; Moreau, 2005; Tsitsiklis et al., 1986). Combing the ideas of consensus algorithms and subgradient methods has resulted in a number of significant results. A subgradient method in combination with consensus steps was given for solving coupled optimization problems with fixed undirected topology in Johansson et al. (2008). An important contribution on multi-agent optimization is Nedić and Ozdaglar (2009), in which the presented decentralized algorithm was based on simply summing an averaging (consensus) part and a subgradient part, and convergence bounds for a distributed multi-agent computation model with time-varying communication graphs with various connectivity assumptions were shown. A constrained optimization problem was studied in Nedić, Ozdaglar, and Parrilo (2010), where each agent is assumed to always lie in a particular convex set, and consensus and optimization were shown to be guaranteed together by each agent taking projection onto its own set at each step. Augmented Lagrangian algorithms with directed gossip communication were used to solve the constrained optimization problem in Jakovetić, Xavier, and Moura (2011). Then a convex-projection-based distributed control was presented for multi-agent systems with continuous-time dynamics to solve this optimization problem asymptotically (Shi, Johansson, & Hong, 2011).

In this paper, we present a randomized multi-agent optimization algorithm. Different from the existing results, we focus on the randomization of individual decision-making of each node. We assume that each optimal solution set of f_i, is a closed convex set, and can be observed only by node i. Assuming that the intersection of all the solution sets is nonempty, the optimal solution set of the group objective becomes this intersection set. Then the optimization problem is equivalent to a distributed intersection computation problem. Computing convex sets' intersection is actually a classical problem. The alternating projection algorithm was a standard centralized solution, which was discussed in Aronszajn (1950), Deutsch (1983), Gubin, Polyak, and Raik (1967) and Nedić et al. (2010). Then the projected consensus algorithm was presented in Nedić et al. (2010).

We propose a randomized algorithm as follows. At each time step, there are two options for each agent: a standard averaging (consensus) part as a convex combination of its neighbors' state, and a projection part as the convex projection of its current state onto its own optimal solution set. In the algorithm, each agent independently makes a decision via a simple Bernoulli trial, i.e., chooses the averaging part with probability p, and the projection part with probability $1 − p$. This algorithm is a randomized version of the projected consensus algorithm in Nedić et al. (2010). Viewing the state of each agent as its "opinion", one can interpret the randomized algorithm considered in this paper as a model of spread of information in social networks (Acemoglu et al., 2010). In this case, the averaging part of the iteration corresponds to an agent updating its opinion based on its neighbors' information, while the projection part corresponds to an agent updating its opinion based only on its own belief of what is the best move. The authors of Acemoglu et al. (2010) draw interesting conclusions from a model similar to ours on how misinformation can spread in a social network.

In our model, the communication graph is assumed to be a general random digraph process independent with the agents' decision-making process. Instead of assuming that the communication graph is modeled by a sequence of i.i.d. random variables over time, we just require the connectivity-independence condition, which is essentially different with existing works (Fagnani & Zampieri, 2008; Hatano & Mesbahi, 2005; Tahbaz-Salehi & Jadbabaie, 2008). Borrowing the ideas on uniform joint-connection (Jadbabaie et al., 2003; Lin et al., 2007; Tsitsiklis et al., 1986) and $[t, \infty)$-joint connectedness (Moreau, 2005; Shi & Hong, 2009), we introduce connectivity conditions of stochastically uniformly (jointly) strongly connected (SUSC) and stochastically infinitely (jointly) connected (SIC) graphs, respectively. The results show that the considered multi-agent network can almost surely achieve a global optimal consensus, i.e., a global consensus within the optimal solution set of $\sum_{i=1}^{n} f_i(z)$, when the communication graph is SUSC with general directed graphs, or SIC with bidirectional information exchange. Convergence is derived with the help of convex analysis and probabilistic analysis.

The paper is organized as follows. In Section 2, some preliminary concepts are introduced. In Section 3, we formulate the considered multi-agent optimization model and present the optimization algorithm. We also establish some basic assumptions and lemmas in this section. Then the main result and convergence analysis are shown for directed and bidirectional graphs, respectively in Sections 4 and 5. In Section 6 we study a numerical example. Finally, concluding remarks are given in Section 7.

2. Preliminaries

Here we introduce some mathematical notations and tools on graph theory (Godsil & Royle, 2001), convex analysis (Boyd & Vandenberghe, 2004; Rockafellar, 1972) and Bernoulli trials (Bertsekas & Tsitsiklis, 2002).

2.1. Directed graphs

A directed graph (digraph) $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a finite set $\mathcal{V} = \{1, \dots, n\}$ of nodes and an arc set $\mathcal{E}$. An element $e = (i, j) \in \mathcal{E}$, which is an ordered pair of nodes $i, j \in \mathcal{V}$, is called an *arc* leaving from node i and entering node j. If the e_j's are pairwise distinct in an alternating sequence $v_0 e_1 v_1 e_2 v_2 \dots e_n v_n$ of nodes v_i and arcs $e_i = (v_{i-1}, v_i) \in \mathcal{E}$ for $i = 1, 2, \dots, n$, the sequence is called a (directed) *path*. A path from i to j is denoted $i \to j$. $\mathcal{G}$ is said to be *strongly connected* if it contains paths $i \to j$ and $j \to i$ for every pair of nodes i and j.

A *weighted digraph* $\mathcal{G}$ is a digraph with weights assigned for its arcs. A weighted digraph $\mathcal{G}$ is called to be *bidirectional* if for any two nodes i and j, $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$, but the weights of (i, j) and (j, i) may be different. A bidirectional digraph is strongly connected if and only if it is connected as an undirected graph (ignoring the directions of the arcs).

The *adjacency matrix*, A, of digraph $\mathcal{G}$ is the $n \times n$ matrix whose ij-entry, A_{ij}, is 1 if there is an arc from i to j, and 0 otherwise. Additionally, if $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_2)$ have the same node set, the union of the two digraphs is defined as $\mathcal{G}_1 \cup \mathcal{G}_2 = (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$.

2.2. Convex analysis

A set $K \subset \mathbb{R}^d$ $(d > 0)$ is said to be *convex* if $(1 − \lambda)x + \lambda y \in K$ whenever $x, y \in K$ and $0 \le \lambda \le 1$. For any set $S \subset \mathbb{R}^d$, the intersection of all convex sets containing S is called the *convex hull* of S, and is denoted by $co(S)$.

Let K be a closed convex set in $\mathbb{R}^d$ and denote $|x|_K \triangleq \inf_{y \in K} |x − y|$ as the distance between $x \in \mathbb{R}^d$ and K, where $| \cdot |$ denotes the

Euclidean norm. Then we can associate to any $x \in \mathbb{R}^d$ a unique element $P_K(x) \in K$ satisfying $|x - P_K(x)| = |x|_K$, where the map P_K is called the *projector* onto K with

$$\langle P_K(x) - x, P_K(x) - y \rangle \leq 0, \quad \forall y \in K. \tag{1}$$

Moreover, we have the following non-expansiveness property for P_K:

$$|P_K(x) - P_K(y)| \leq |x - y|, \quad x, y \in \mathbb{R}^d. \tag{2}$$

A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to be convex if it satisfies

$$f(\alpha v + (1 - \alpha)w) \leq \alpha f(v) + (1 - \alpha)f(w), \tag{3}$$

for all $v, w \in \mathbb{R}^d$ and $0 \leq \alpha \leq 1$. The following lemma holds (Example 3.16, pp. 88, Boyd & Vandenberghe, 2004).

Lemma 1. *Let K be a convex set in $\mathbb{R}^d$. Then $|x|_K$ is a convex function.*

The next lemma can be found in Aubin and Cellina (1984).

Lemma 2. *Let K be a subset of $\mathbb{R}^d$. The convex hull $co(K)$ of K is the set of elements of the form*

$$x = \sum_{i=1}^{d+1} \lambda_i x_i,$$

where $\lambda_i \geq 0, i = 1, \ldots, d+1$ with $\sum_{i=1}^{d+1} \lambda_i = 1$ and $x_i \in K$.

Additionally, for every two vectors $0 \neq v_1, v_2 \in \mathbb{R}^d$, we define their angle as $\phi(v_1, v_2) \in [0, \pi]$ with $\cos \phi = \langle v_1, v_2 \rangle / |v_1| \cdot |v_2|$.

2.3. Bernoulli trials

A Bernoulli trial is a binary random variable which only takes two values, 0 and 1. Let $Y_1, Y_2, Y_3, \ldots$ be a sequence of independent Bernoulli trials such that for each $k = 1, 2, \ldots$, the probability that $Y_k = 1$ is $p_k \in [0, 1]$. Here p_k is called the success probability for Y_k.

Then the next lemma holds. The proof is obvious, and therefore omitted.

Lemma 3. *Let $Y_k, k = 1, 2, \ldots$, be a sequence of independent Bernoulli trials, where the success probability of Y_k is $p_k \in [0, 1]$. Suppose there exists a constant $p_* > 0$ such that $p_k > p_*$ for all k. Then we have $\mathbf{P}(Y_k = 1$ for infinitely many $k \geq 1) = 1$.*

3. Problem formulation

In this section, we formulate the considered optimal consensus problem. We propose a multi-agent optimization model, and then introduce a neighbor-based randomized optimization algorithm. We also introduce key assumptions and establish two basic lemmas on the algorithm used in the subsequent analysis.

3.1. Multi-agent model

Consider a multi-agent system with agent set $\mathcal{V} = \{1, 2, \ldots, n\}$. The objective of the network is to reach a consensus, and meanwhile to cooperatively solve the following optimization problem

$$\min_{z \in \mathbb{R}^d} \sum_{i=1}^{n} f_i(z) \tag{4}$$

where $f_i : \mathbb{R}^d \to \mathbb{R}$ represents the cost function of agent i, observed by agent i only, and z is a decision vector.

Time is slotted, and the dynamics of the network is in discrete time. Each agent i starts with an arbitrary initial position, denoted $x_i(0) \in \mathbb{R}^d$, and updates its state $x_i(k)$ for $k = 0, 1, 2, \ldots$, based on the information received from its neighbors and the information observed from its optimization component f_i.

3.1.1. Communication graph

We suppose the communication graph over the multi-agent network is a stochastic digraph process $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k), k = 0, 1, \ldots$. To be precise, the ij-entry $A_{ij}(k)$ of the adjacency matrix, $A(k)$ of $\mathcal{G}_k$, is a general $\{0, 1\}$-state stochastic process. We assume there is no self-looped arc in the communication graphs, i.e., $A_{ii}(k) = 0$ for all i and k. We use the following assumption on the independence of $\mathcal{G}_k$.

A1 *(Connectivity Independence)* Events $\mathcal{C}_k = \{\mathcal{G}_k$ is connected (in certain sense)$\}, k = 0, 1, \ldots$, are independent.

Remark 4. Connectivity independence means that a sequence of random variables $\varpi(k)$, which are defined by that $\varpi(k) = 1$ if $\mathcal{G}_k$ is connected (in certain sense) and $\varpi(k) = 0$ otherwise, are independent. Note that, unlike in existing works (Fagnani & Zampieri, 2008; Hatano & Mesbahi, 2005; Tahbaz-Salehi & Jadbabaie, 2008), we do not impose the assumption that $\varpi(k), k = 0, \ldots$, are identically distributed.

At time k, node j is said to be a *neighbor* of i if there is an arc $(j, i) \in \mathcal{E}_k$. Particularly, we assume that each node is always a neighbor of itself. Let $\mathcal{N}_i(k)$ represent the set of agent i's neighbors at time k.

Denote the joint graph of $\mathcal{G}_k$ in time interval $[k_1, k_2]$ as $\mathcal{G}([k_1, k_2]) = (\mathcal{V}, \cup_{t \in [k_1, k_2]} \mathcal{E}(t))$, where $0 \leq k_1 \leq k_2 \leq +\infty$. Then we have the following definition.

Definition 5. (i) $\mathcal{G}_k$ is said to be *stochastically uniformly (jointly) strongly connected* (SUSC) if there exist two constants $B \geq 1$ and $0 < q < 1$ such that for any $k \geq 0$,

$$\mathbf{P}\Big(\mathcal{G}([k, k + B - 1]) \text{ is strongly connected}\Big) \geq q.$$

(ii) Assume that $\mathcal{G}_k$ is bidirectional for all $k \geq 0$. Then $\mathcal{G}_k$ is said to be *stochastically infinitely (jointly) connected* (SIC) if there exist a (deterministic) sequence $0 = k_0^* < \cdots < k_\tau^* < k_{\tau+1}^* < \cdots$ and a constant $0 < q < 1$ such that for all $\tau = 0, 1, \ldots$,

$$\mathbf{P}\Big(\mathcal{G}([k_\tau^*, k_{\tau+1}^*]) \text{ is connected}\Big) \geq q.$$

3.1.2. Neighboring information

The local information that each agent uses to update its state consists of two parts: the averaging and the projection parts. The averaging part is defined as

$$e_i(k) = \sum_{j \in \mathcal{N}_i(k)} a_{ij}(k)x_j(k),$$

where $a_{ij}(k) > 0, i, j = 1, \ldots, n$ are the arc weights. The weights fulfill the following assumption:

A2 *(Arc Weights)* (i) $\sum_{j \in \mathcal{N}_i(k)} a_{ij}(k) = 1$ for all i and k.

(ii) There exists a constant $\eta > 0$ such that $\eta \leq a_{ij}(k)$ for all i, j and k.

The projection part is defined as

$$g_i(k) = P_{X_i}(x_i(k)),$$

where $X_i \doteq \{v | f_i(v) = \min_{z \in \mathbb{R}^d} f_i(z)\}$ is the optimal solution set of each objective function $f_i, i = 1, \ldots, n$. We use the following assumptions.

A3 *(Convex Solution Set)* $X_i, i = 1, \ldots, n$, are closed convex sets.

A4 *(Nonempty Intersection)* $X_0 \doteq \bigcap_{i=1}^{n} X_i$ is nonempty.

In the rest of the paper, A1–A4 are our standing assumptions.

Remark 6. The average $e_i(k)$ has been widely used in consensus algorithms, e.g., Jadbabaie et al. (2003), Moreau (2005) and Tsitsiklis et al. (1986). Assumption (A2)(i) indicates that $e_i(k)$ is always within the convex hull of node i's neighbors, i.e., $co\{x_j(k), j \in$

$\mathcal{N}_i(k)$}, and, moreover, A2(ii) ensures that $e_i(k)$ is in the relative interior of $co\{x_j(k), j \in \mathcal{N}_i(k)\}$ (Lin et al., 2007).

Remark 7. As X_i can be observed by node i, $P_{X_i}(x_i(k))$ can be easily obtained. Note that, for a convex set $K \subseteq \mathbb{R}^d$, we have that $\nabla|z|^2_K = 2(P_K(z) - z)$ (Aubin & Cellina, 1984). Therefore, for instance, in order to compute $P_{X_i}(x_i(k))$, node i may first establish a local coordinate system, and then construct a function $h(z) = |z|^2_{X_i}/2$ to compute $\nabla h(x_i(k))$ within this coordinate system. Then we know $P_{X_i}(x_i(k)) = x_i(k) + \nabla h(x_i(k))$.

3.1.3. Randomized algorithm

We are now ready to introduce the randomized optimization algorithm.

At each time step, independent of time, communication graph process, and other nodes, each agent takes the following decision process:

$$x_i(k+1) = \begin{cases} \sum_{j \in \mathcal{N}_i(k)} a_{ij}(k)x_j(k), & \text{with probability } p \\ P_{X_i}(x_i(k)), & \text{with probability } 1 - p \end{cases} \tag{5}$$

where $0 < p < 1$ is a given constant.

Remark 8. The constrained consensus algorithm studied in Nedić et al. (2010), can be viewed as a deterministic case of (5), in which each node alternates between averaging and projection in the iterations.

The proposed Algorithm (5) is simply obtained by each node deciding at random between averaging with its neighbors and projecting onto the optimal solution set. Averaging with neighbors relies on node communication, while projection relies on sensing its solution set. Therefore, from an engineering viewpoint, unlike most existing works (Johansson et al., 2009; Nedić et al., 2010; Shi et al., 2011), the randomized Algorithm (5) gives freedom to the nodes of choosing to sense (project), or to communicate (average) independently with others at each time k. This provides an important tradeoff between individual sensing and communication, as each node is not synchronously required to both sense and communicate in each time step.

On the other hand, another motivation for the study of Algorithm (5) follows from the literature on opinion dynamics in social networks, where each agent makes a choice randomly between sticking to its own observation or following its neighbors' opinion (Acemoglu et al., 2010). Here in (5), the projection term reflects each node's selfish intention since projecting onto its own set leads to a local optimum. The averaging term reflects opinion exchange. By taking averaging with neighbors, each node may sacrifice itself since it may leave its own optimal set. Then Algorithm (5) naturally describes the randomized trade-off between being selfish or sharing. An interesting question arises whether the network could reach a common opinion or not, and if the answer is yes, whether the network could reach an optimal common opinion.

Moreover, by itself Algorithm (5) is an interesting model because of the randomized transitions between linear (averaging) and nonlinear (projection) dynamics.

A primary aim of this paper is to establish fundamental convergence properties for a novel randomized node-decision process, rather than to present fast algorithms. There may be many ways of generating a faster algorithm than the one considered here by better scheduling and arrangement for the projection and the averaging (Johansson et al., 2009; Nedić et al., 2010; Shi et al., 2011). We will compare the convergence rate of the randomized algorithm to the corresponding deterministic version by numerical examples in Section 6.
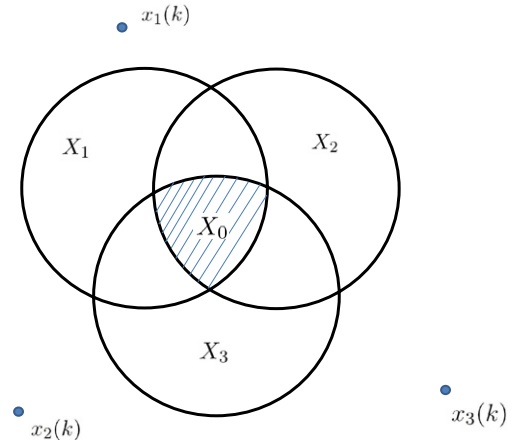

Fig. 1. The goal of the multi-agent network is to achieve a consensus in the optimal solution set X_0.

With Assumptions A3 and A4, X_0 becomes the global optimal solution set of $\sum_{i=1}^n f_i(z)$. Let $\{x(k; x^0) = (x_1^T(k; x^0), \dots, x_n^T(k; x^0))^T\}_{k=0}^\infty$ be the stochastic sequence generated by (5) with initial condition $x^0 = (x_1^T(0), \dots, x_n^T(0))^T \in \mathbb{R}^{nd}$. We will identify $x(k; x^0)$ with $x(k)$ where there is no possible confusion. The considered optimal consensus problem is defined as follows. See Fig. 1 for an illustration.

Definition 9. (i) A global optimal set aggregation is achieved almost surely (a.s.) for algorithm (5) if for all $x^0 \in \mathbb{R}^{nd}$,

$$\mathbf{P}\left(\lim_{k \to \infty} |x_i(k)|_{X_0} = 0, i = 1, \dots, n\right) = 1. \tag{6}$$

(ii) A global *consensus* is achieved almost surely (a.s.) for algorithm (5) if for all $x^0 \in \mathbb{R}^{nd}$,

$$\mathbf{P}\left(\lim_{k \to \infty} |x_i(k) - x_j(k)| = 0, i, j = 1, \dots, n\right) = 1. \tag{7}$$

(iii) A global *optimal consensus* is achieved almost surely (a.s.) for Algorithm (5) if both (6) and (7) hold.

Remark 10. A stronger convergence condition for optimal consensus can be imposed by that there exists $z_* \in X_0$ satisfying

$$\mathbf{P}\left(\lim_{k \to \infty} x_i(k) = z_*, i = 1, \dots, n\right) = 1. \tag{8}$$

In order to guarantee (8), the arc weights $a_{ij}(k)$ usually need to be doubly stochastic (Nedić et al., 2010). Doubly stochastic weights are difficult to obtain in general, even in a decentralized way. Moreover, even if we assume doubly stochasticity for the arcs, it may still be lost due to the randomized node dynamics. Therefore, here we use (6) and (7) to define optimal consensus. It is not clear whether (8) can be achieved based on the analysis used in this paper.

3.2. Basic properties

In this subsection, we establish two key lemmas on the Algorithm (5).

Lemma 11. Let K be a closed convex set in $\mathbb{R}^d$, and $K_0 \subseteq K$ be a convex subset of K. Then for any $y \in \mathbb{R}^d$, we have

$$|P_K(y)|^2_{K_0} + |y|^2_K \leq |y|^2_{K_0}.$$

Proof. According to (1), we know that

$$\langle P_K(y) - y, P_K(y) - P_{K_0}(y) \rangle \leq 0.$$

Therefore, we obtain

$$
\begin{aligned}
\langle P_K(y) - y, y - P_{K_0}(y) \rangle &= \langle P_K(y) - y, \\
&\quad y - P_K(y) + P_K(y) - P_{K_0}(y) \rangle \\
&\leq -|y|_K^2.
\end{aligned}
$$

Then,

$$
\begin{aligned}
|P_K(y)|_{K_0}^2 &= |P_K(y) - P_{K_0}(P_K(y))|^2 \\
&\leq |P_K(y) - P_{K_0}(y)|^2 \\
&= |P_K(y) - y + y - P_{K_0}(y)|^2 \\
&= |y|_K^2 + |y|_{K_0}^2 + 2\langle P_K(y) - y, y - P_{K_0}(y) \rangle \\
&\leq |y|_{K_0}^2 - |y|_K^2.
\end{aligned}
$$

The desired conclusion follows. $\square$

Lemma 12. Let $\{x(k) = (x_1^T(k), \ldots, x_n^T(k))^T\}_{k=0}^{\infty}$ be a stochastic sequence defined by (5). Then for all $k \geq 0$ and along every possible sample path, we have

$$
\max_{i=1,\ldots,n} |x_i(k+1)|_{X_0} \leq \max_{i=1,\ldots,n} |x_i(k)|_{X_0}.
$$

Proof. Take $l \in \mathcal{V}$. If node l takes averaging at time k, we have

$$
\begin{aligned}
|x_l(k+1)|_{X_0} &= |P_{X_l}(x_l(k))|_{X_0} \\
&= |P_{X_l}(x_l(k)) - P_{X_0}(P_{X_l}(x_l(k)))| \\
&\leq |P_{X_l}(x_l(k)) - P_{X_0}(x_l(k))| \\
&\leq |x_l(k) - P_{X_0}(x_l(k))| \\
&\leq \max_{i=1,\ldots,n} |x_i(k)|_{X_0}. \quad (9)
\end{aligned}
$$

On the other hand, if node l takes projection at time k, according to Lemma 1, we have

$$
\begin{aligned}
|x_l(k+1)|_{X_0} &= \left| \sum_{j \in N_l(k)} a_{lj}(k)x_j(k) \right|_{X_0} \\
&\leq \sum_{j \in N_l(k)} a_{lj}(k)|x_j(k)|_{X_0} \\
&\leq \max_{i=1,\ldots,n} |x_i(k)|_{X_0}. \quad (10)
\end{aligned}
$$

Hence, the conclusion holds. $\square$

Based on Lemma 12, we know that the following limit exists:

$$
\xi \doteq \lim_{k \to \infty} \max_{i=1,\ldots,n} |x_i(k)|_{X_0}.
$$

It is immediate that the global optimal set aggregation is achieved almost surely if and only if $\mathbf{P}\{\xi = 0\} = 1$.

Algorithm (5) is nonlinear and stochastic, and therefore quite challenging to analyze. As will be shown in the following, the communication graph plays an essential role on the convergence of the algorithm. In particular, directed and bidirectional graphs lead to different conditions for consensus. Hence, in the following two sections, we consider these two cases separately.

4. Directed graphs

In this section, we give a connectivity condition guaranteeing an almost surely global optimal consensus for directed communication graphs.

The main result is stated as follows.

Theorem 13. Algorithm (5) achieves a global optimal consensus a.s. if $\mathcal{G}_k$ is SUSC.

In order to prove Theorem 13, on one hand, we have to prove that all the agents converge to the global optimal solution set, i.e., X_0; and on the other hand that consensus is achieved. The proof divided into these two parts is given in the following two subsections.

4.1. Set convergence

In this subsection, we present the optimal set aggregation analysis of (5). Define

$$
\delta_i \doteq \limsup_{k \to \infty} |x_i(k)|_{X_i}, \quad i = 1, \ldots, n.
$$

Let $\mathcal{A} = \{\xi > 0\}$ and $\mathcal{M} = \{\exists i_0 \ s.t. \ \delta_{i_0} > 0\}$ be two events, indicating that convergence to X_0 for all the agents fails and convergence to X_{i_0} fails for some node i_0, respectively. The next lemma shows the relation between the two events.

Lemma 14. $\mathbf{P}(\mathcal{A} \cap \mathcal{M}) = 0$ if $\mathcal{G}_k$ is SUSC.

Proof. Let $\{x^\omega(k)\}_{k=0}^{\infty}$ be a sample sequence. Take an arbitrary node $i_0 \in \mathcal{V}$. Then there exists a time sequence $k_1(\omega) < \cdots < k_m(\omega) < \cdots$ with $\lim_{m \to \infty} k_m(\omega) = \infty$ such that

$$
|x_{i_0}^\omega(k_m(\omega))|_{X_{i_0}} \geq \frac{1}{2}\delta_{i_0}(\omega) \geq 0. \quad (11)
$$

Moreover, according to Lemma 12, $\forall \ell = 1, 2, \ldots, \exists T(\ell, \omega) > 0$ such that

$$
k \geq T \Rightarrow 0 \leq |x_i^\omega(k)|_{X_0} \leq \xi(\omega) + \frac{1}{\ell}, \quad i = 1, \ldots, n. \quad (12)
$$

In the following, $k_m(\omega)$ and $T(\ell, \omega)$ will be denoted as k_m and T to simplify the notations. Note that they are both random variables. We divide the rest of the proof into three steps.

Step 1. Suppose m is sufficiently large so that $k_m \geq T$. We give an upper bound to node i_0 in this step.

Since node i_0 projects onto X_i with probability $1 - p$, Lemma 11 implies

$$
\mathbf{P}\left(|x_{i_0}(k_m + 1)|_{X_0} \leq \sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} \right) \geq 1 - p. \quad (13)
$$

At time $k_m + 2$, either one of two cases can happen in the update.

- If node i_0 chooses the projection option at time $k_m + 1$, we have

$$
|x_{i_0}(k_m + 2)|_{X_0} = |x_{i_0}(k_m + 1)|_{X_0} \leq \sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}. \quad (14)
$$

- If node i_0 chooses the averaging option at time $k_m + 1$, with (12), we can obtain from the weights rule and Lemma 1 that

$$
\begin{aligned}
|x_{i_0}(k_m + 2)|_{X_0} &= \left| \sum_{j \in \mathcal{N}_{i_0}(k_m+1)} a_{i_0 j}(k_m+1)x_j(k_m+1) \right|_{X_0} \\
&\leq a_{i_0 i_0}(k_m+1)|x_{i_0}(k_m+1)|_{X_0} \\
&\quad + (1 - a_{i_0 i_0}(k_m+1))\left(\xi + \frac{1}{\ell}\right) \\
&\leq a_{i_0 i_0}(k_m+1)\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} \\
&\quad + (1 - a_{i_0 i_0}(k_m+1))\left(\xi + \frac{1}{\ell}\right) \\
&\leq \eta\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta)\left(\xi + \frac{1}{\ell}\right). \quad (15)
\end{aligned}
$$

Both (14) and (15) lead to

$$\mathbf{P}\left(|x_{i_0}(k_m + 2)|_{X_0} \leq \eta\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right.$$

$$\left. + (1 - \eta)\left(\xi + \frac{1}{\ell}\right)\right) \geq 1 - p. \tag{16}$$

Continuing similar analysis, we further obtain

$$\mathbf{P}\left(|x_{i_0}(k_m + \tau)|_{X_0} \leq \eta^{\tau-1}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right.$$

$$\left. + (1 - \eta^{\tau-1})\left(\xi + \frac{1}{\ell}\right), \tau = 1, 2, \ldots\right) \geq 1 - p. \tag{17}$$

Step 2. In this step, we continue to bound another node. Since $\mathcal{G}_k$ is SUSC, we have

$$\mathbf{P}\left(\mathcal{G}([k_m + 1, k_m + B]) \text{ is strongly connected}\right) \geq q$$

which implies

$$\mathbf{P}\left(\exists \hat{k}_1 \in [k_m + 1, k_m + B] \text{ and } i_1 \in \mathcal{V}, i_1 \neq i_0\right.$$

$$\left. \text{s.t. } (i_0, i_1) \in \mathcal{E}_{\hat{k}_1}\right) \geq q. \tag{18}$$

Let $\hat{k}_1 = k_m + \varrho$ with $1 \leq \varrho \leq B$. Noting the fact that

$$\left|\sum_{j \in \mathcal{N}_{i_1}(k_m+\varrho)} x_j(k_m + \varrho)\right|_{X_0} \leq a_{i_1 i_0}(k_m + \varrho)|x_{i_0}(k_m + \varrho)|_{X_0}$$

$$+ \left(1 - a_{i_1 i_0}(k_m + \varrho)\right)\left(\xi + \frac{1}{\ell}\right)$$

and based on (17), we have

$$\mathbf{P}\left(|x_{i_1}(k_m + \varrho + 1)|_{X_0} \leq \eta^{\varrho}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right.$$

$$\left. + (1 - \eta^{\varrho})\left(\xi + \frac{1}{\ell}\right)\middle|\mathcal{F}_0\right)$$

$$\geq \mathbf{P}\left(i_1 \text{ chooses averaging at time } k_m + \varrho\right)$$

$$= p, \tag{19}$$

where $\mathcal{F}_0 = \{i_0 \text{ chooses projection at time } k_m\}$. Therefore, with (17) and (19), we obtain

$$\mathbf{P}\left(\exists i_1 \neq i_0 \text{ s.t. } |x_{i_l}(k_m + B + \tau)|_{X_0} \leq \eta^{B+\tau-1}\right.$$

$$\cdot\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta^{B+\tau-1})\left(\xi + \frac{1}{\ell}\right),$$

$$\left. l = 0, 1; \tau = 1, 2, \ldots\right) \geq (1 - p)pq.$$

Step 3. Repeating the analysis on time interval $[k_m + B + 1, k_m + 2B]$, there exists a node $i_2 \notin \{i_0, i_1\}$ such that there is an arc leaving from $\{i_0, i_1\}$ entering i_2 in $\mathcal{G}([k_m + B + 1, k_m + 2B])$ with probability at least q. The estimate of $|x_{i_2}(k_m + 2B + \tau)|_{X_0}$ is therefore can be similarly obtained.

The upper analysis process can be carried out continuingly on intervals $[k_m + 2B + 1, k_m + 3B + 1], \ldots, [k_m + (n-2)B + 1, k_m +$

$(n-1)B]$, and $i_3, \ldots, i_{n-1}$ can be found until $\mathcal{V} = \{i_0, i_1, \ldots, i_{n-1}\}$. Then one can obtain that for any $i \in \mathcal{V}$,

$$\mathbf{P}\left(|x_i(k_m + (n-1)B + 1)|_{X_0} \leq \eta^{(n-1)B}\right.$$

$$\left.\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta^{(n-1)B})\left(\xi + \frac{1}{\ell}\right), i \in \mathcal{V}\right)$$

$$= \mathbf{P}\left(\max_{i=1,\ldots,n}|x_i(k_m + (n-1)B + 1)|_{X_0} \leq \eta^{(n-1)B}\right.$$

$$\left.\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta^{(n-1)B})\left(\xi + \frac{1}{\ell}\right)\right)$$

$$\geq (1 - p)p^{n-1}q^{n-1}. \tag{20}$$

We see from the previous analysis that the events

$$\mathcal{Z}_m \doteq \left\{\max_{i=1,\ldots,n}|x_i(k_m + (n-1)B + 1)|_{X_0}\right.$$

$$\left.\leq \eta^{(n-1)B}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta^{(n-1)B})\left(\xi + \frac{1}{\ell}\right)\right\}$$

are fully determined by the connectivity of the communication graph process and the node-decision process for all m with $k_m \geq T$. Since the node decision process is independent of time and the communication graph is connectivity independent by Assumption A1, the following claim holds.

Claim. Events $\{\mathcal{Z}_t\}_{t=m}^{\infty}$ can be viewed as a sequence of independent Bernoulli trials with success probability $\mu_0 = (1-p)p^{n-1}q^{n-1}$ for sufficiently large m in the sense that

$$\mathbf{P}\left(\bigcap_{i=1}^{K} \mathcal{Z}_{t_i}\right) \geq \mu_0^K \tag{21}$$

for all $t_1 < \cdots < t_K$ with $K \geq 1$.

Therefore, based on Lemma 3, with probability one, there is an infinite subsequence $\{\tilde{k}_j, j = 1, 2, \ldots\}$ from $\{k_m + (n-1)B + 1, k_m \geq T\}$ satisfying

$$\max_{i=1,\ldots,n}|x_i(\tilde{k}_j)|_{X_0} \leq \eta^{(n-1)B}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}$$

$$+ (1 - \eta^{(n-1)B})\left(\xi + \frac{1}{\ell}\right).$$

This implies

$$\mathbf{P}(\mathcal{R}_{\ell}) = 1 \tag{22}$$

for all $\ell = 1, 2, \ldots$, where

$$\mathcal{R}_{\ell} = \left\{\xi \leq \eta^{(n-1)B}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right.$$

$$\left. + (1 - \eta^{(n-1)B})\left(\xi + \frac{1}{\ell}\right)\right\}.$$

As a result, we obtain $\mathbf{P}(\mathcal{R}_*) = 1$, where

$$\mathcal{R}_* = \lim_{\ell \to \infty} \mathcal{R}_{\ell}$$

$$= \left\{\xi \leq \eta^{(n-1)B}\sqrt{\xi^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta^{(n-1)B})\xi\right\}.$$

Finally, it is not hard to see that $\mathcal{A} \cap \mathcal{M} \subseteq \mathcal{R}_*^c$ because $0 < \eta^{(n-1)B} < 1$. The desired conclusion follows straightforwardly. $\quad\square$

Take a node $\alpha_0 \in \mathcal{V}$. Then define

$$z_{\alpha_0}(k) \doteq \max_{i=1,\ldots,n} |x_i(k)|_{X_{\alpha_0}}.$$

We also need the following fact to prove the optimal set convergence.

Lemma 15. *Along every possible sample path of Algorithm* (5) *and for all k, we have*

$$z_{\alpha_0}(k+1) \leq z_{\alpha_0}(k) + \max_{i=1,\ldots,n} |x_i(k)|_{X_i}.$$

Proof. For any node $l = 1, \ldots, n$, if l chooses the averaging part at time k, we know that

$$
\begin{aligned}
|x_l(k+1)|_{X_{\alpha_0}} &= \Bigg| \sum_{j \in \mathcal{N}_l(k)} a_{lj}(k) x_j(k) \Bigg|_{X_{\alpha_0}} \\
&\leq \max_{i=1,\ldots,n} |x_l(k)|_{X_{\alpha_0}} \\
&= z_{\alpha_0}(k). \qquad (23)
\end{aligned}
$$

Moreover, if l chooses the projection part at time k, we have

$$|x_l(k+1) - x_l(k)| = |x_l(k)|_{X_l},$$

which yields

$$
\begin{aligned}
|x_l(k+1)|_{X_{\alpha_0}} &\leq |x_l(k)|_{X_{\alpha_0}} + |x_l(k)|_{X_l} \\
&\leq z_{\alpha_0}(k) + \max_{i=1,\ldots,n} |x_i(k)|_{X_i} \qquad (24)
\end{aligned}
$$

according to the non-expansiveness property (2). Then the conclusion holds with (23) and (24). $\quad\square$

We are now in a place to present the optimal set convergence part of Theorem 13, as stated in the following conclusion.

Proposition 16. *Algorithm* (5) *achieves a global optimal set aggregation a.s. if $\mathcal{G}_k$ is SUSC.*

Proof. Note that, we have

$$\mathbf{P}(\mathcal{A}) = \mathbf{P}(\mathcal{A} \cap \mathcal{M}) + \mathbf{P}(\mathcal{A} \cap \mathcal{M}^c) \leq \mathbf{P}(\mathcal{A} \cap \mathcal{M}) + \mathbf{P}(\mathcal{A}|\mathcal{M}^c).$$

Since the conclusion is equivalent to $\mathbf{P}(\mathcal{A}) = 0$, with Lemma 14, we only need to prove $\mathbf{P}(\mathcal{A}|\mathcal{M}^c) = 0$.

Let $\{x^\omega(k)\}_{k=0}^\infty$ be a sample sequence in $\mathcal{M}^c$. Then $\forall \ell = 1, 2, \ldots, \exists T_1(\ell, \omega) > 0$ such that

$$k \geq T_1 \Rightarrow |x_i^\omega(k)|_{X_i} \leq \frac{1}{\ell}, \quad i = 1, \ldots, n. \qquad (25)$$

Take an arbitrary node $\alpha_0 \in \mathcal{V}$. Based on Lemma 15, we also have that for any $\{x^\omega(k)\}_{k=0}^\infty \in \mathcal{M}^c$ and $s \geq T_1$,

$$z_{\alpha_0}^\omega(s+\tau) \leq z_{\alpha_0}^\omega(s) + \frac{\tau}{\ell}, \quad \tau = 0, 1, \ldots. \qquad (26)$$

We divide the rest part of the proof into three steps.

Step 1. Denote $k_1 = T_1$. Since $\mathcal{G}_k$ is SUSC, we have

$$\mathbf{P}\Big(\text{there exist } \hat{k}_1 \in [k_1, k_1 + B - 1] \text{ and } \alpha_1 \in \mathcal{V}$$

$$\text{s.t. } (\alpha_0, \alpha_1) \in \mathcal{G}_{\hat{k}_1}\Big) \geq q.$$

Let $\hat{k}_1 = k_1 + \varrho, 0 \leq \varrho \leq B - 1$. Then we obtain from the definition of (5) that

$$
\begin{aligned}
\mathbf{P}\Big(&|x_{\alpha_1}(k_1 + \varrho + 1)|_{X_{\alpha_0}} \leq a_{\alpha_1\alpha_0}(k_1 + \varrho)|x_{\alpha_0}(k_1 + \varrho)|_{X_{\alpha_0}} \\
&+ (1 - a_{\alpha_1\alpha_0})z_{\alpha_0}(k_1 + \varrho)\Big) \geq pq. \qquad (27)
\end{aligned}
$$

Thus, based on the weights rule A1 and (25), (27) leads to

$$
\begin{aligned}
\mathbf{P}\Bigg(&|x_{\alpha_1}(k_1 + \varrho + 1)|_{X_{\alpha_0}} \leq \eta \cdot \frac{1}{\ell} + (1 - \eta)\Big(z_{\alpha_0}(k_1) \\
&+ \varrho \cdot \frac{1}{\ell}\Big)\Big| \mathcal{M}^c\Bigg) \geq pq.
\end{aligned}
$$

Next, there will be two cases.

- If node α_1 chooses the projection option at time $k_1 + \varrho + 1$, we have

$$
\begin{aligned}
|x_{\alpha_1}(k_1 + \varrho + 2)|_{X_{\alpha_0}} &\leq |x_{\alpha_1}(k_1 + \varrho + 1)|_{X_{\alpha_0}} + \frac{1}{\ell} \\
&\leq \eta \cdot \frac{1}{\ell} + (1 - \eta)\Big(z_{\alpha_0}(k_1) + \varrho \cdot \frac{1}{\ell}\Big) + \frac{1}{\ell}. \qquad (28)
\end{aligned}
$$

- If node α_1 chooses the averaging option at time $k_1 + \varrho + 1$, we have

$$
\begin{aligned}
|x_{\alpha_1}(k_1 + \varrho + 2)|_{X_{\alpha_0}} &\leq \eta |x_{\alpha_1}(k_1 + \varrho + 1)|_{X_{\alpha_0}} \\
&\quad + (1 - \eta)z_{\alpha_0}(k_1 + \varrho + 1) \\
&\leq \eta\Big[\eta \cdot \frac{1}{\ell} + (1 - \eta)\Big(z_{\alpha_0}(k_1) + \varrho \cdot \frac{1}{\ell}\Big)\Big] \\
&\quad + (1 - \eta)\Big(z_{\alpha_0}(k_1) + (\varrho + 1) \cdot \frac{1}{\ell}\Big) \\
&\leq \eta^2 \cdot \frac{1}{\ell} + (1 - \eta^2)\Big(z_{\alpha_0}(k_1) + (\varrho + 1) \cdot \frac{1}{\ell}\Big). \qquad (29)
\end{aligned}
$$

With (28) and (29), we obtain

$$
\begin{aligned}
\mathbf{P}\Bigg(&|x_{\alpha_1}(k_1 + \varrho + 2)|_{X_{\alpha_0}} \leq \eta^2 \cdot \frac{1}{\ell} + (1 - \eta^2)\Big(z_{\alpha_0}(k_1) \\
&+ (\varrho + 1) \cdot \frac{1}{\ell}\Big) + \frac{1}{\ell}\Big| \mathcal{M}^c\Bigg) \geq pq. \qquad (30)
\end{aligned}
$$

Then similar analysis yields that

$$
\begin{aligned}
\mathbf{P}\Bigg(&|x_{\alpha_1}(k_1 + \varrho + \tau)|_{X_{\alpha_0}} \leq \frac{\eta^\tau}{\ell} + (1 - \eta^\tau)\Big(z_{\alpha_0}(k_1) + \frac{\varrho + \tau - 1}{\ell}\Big) \\
&+ \sum_{l=1}^{\tau-1} \eta^{l-1} \cdot \frac{1}{\ell}, \ \tau = 1, 2, \ldots \Big| \mathcal{M}^c\Bigg) \geq pq.
\end{aligned}
$$

Furthermore, since $0 \leq \varrho \leq B - 1$ and based on (25), it turns out that

$$
\begin{aligned}
\mathbf{P}\Bigg(&|x_{\alpha_l}(k_1 + B + \hat{\tau})|_{X_{\alpha_0}} \leq \frac{\eta^{B+\hat{\tau}}}{\ell} \\
&+ (1 - \eta^{B+\hat{\tau}})\Big(z_{\alpha_0}(k_1) + \frac{B + \hat{\tau} - 1}{\ell}\Big) + \frac{1}{1 - \eta} \cdot \frac{1}{\ell}, \\
&\hat{\tau} = 0, 1, \ldots; \ l = 0, 1 \Big| \mathcal{M}^c\Bigg) \geq pq.
\end{aligned}
$$

Step 2. We continue the analysis on time interval $[k_1 + B, k_1 + 2B - 1]$. There exists a node $\alpha_2 \notin \{\alpha_0, \alpha_1\}$ such that there is an arc leaving from $\{\alpha_0, \alpha_1\}$ entering α_2 in $\mathcal{G}([k_1 + B, k_m + 2B - 1])$ with

probability q. Similarly we can obtain that for any $\hat{\tau} = 0, 1, \ldots,$

$$\mathbf{P}\Bigg(|x_{\alpha_l}(k_1 + 2B + \hat{\tau})|_{x_{\alpha_0}} \leq \eta^{2B+\hat{\tau}} \cdot \frac{1}{\ell}$$

$$+ (1 - \eta^{2B+\hat{\tau}})\left(z_{\alpha_0}(k_1) + (2B + \hat{\tau} - 1) \cdot \frac{1}{\ell}\right)$$

$$+ \frac{2}{1-\eta} \cdot \frac{1}{\ell}, \ l = 0, 1, 2|\mathcal{M}^c\Bigg) \geq p^2 q^2.$$

We repeat the upper process on time intervals $[k_1 + 2B, k_1 + 3B - 1], \ldots, [k_m + (n-2)B, k_1 + (n-1)B - 1]$, and $\alpha_3, \ldots, \alpha_{n-1}$ can be found until $\mathcal{V} = \{\alpha_0, \alpha_1, \ldots, \alpha_{n-1}\}$. Then one can obtain that

$$\mathbf{P}\Bigg(|x_i(k_1 + (n-1)B)|_{x_{\alpha_0}} \leq (1 - \eta^{(n-1)B})z_{\alpha_0}(k_1) + L \cdot \frac{1}{\ell},$$

$$i = 1, \ldots, n\Big|\mathcal{M}^c\Bigg) \geq p^{n-1}q^{n-1},$$

where $L = \eta^{(n-1)B} + (n-1)[B + \frac{1}{1-\eta}]$. Denote $k_2 = k_1 + (n-1)B$. Then we have

$$\mathbf{P}\Big(z_{\alpha_0}(k_2) \leq \theta_0 z_{\alpha_0}(k_1) + L \cdot \frac{1}{\ell}\Big|\mathcal{M}^c\Big) \geq \hat{p},$$

where $0 < \theta_0 = 1 - \eta^{(n-1)B} < 1$ and $0 < \hat{p} = p^{n-1}q^{n-1} < 1$.

Step 3. Let $k_m = k_1 + (m-1)(n-1)B$, $m = 3, 4, \ldots$. Based on similar analysis, we see that

$$\mathbf{P}\Big(z_{\alpha_0}(k_{m+1}) \leq \theta_0 z_{\alpha_0}(k_m) + L \cdot \frac{1}{\ell}\Big|\mathcal{M}^c\Big) \geq \hat{p}, \quad m = 3, 4, \ldots$$

Then we can define a random variable χ independently with $z_{\alpha_0}(k_m)$, $m = 1, \ldots,$ such that

$$\chi = \begin{cases} 1, & \text{with probability } 1 - \hat{p} \\ \theta_0, & \text{with probability } \hat{p}. \end{cases} \tag{31}$$

As a result, with (26) and (31), we conclude that for any $m = 1, 2, \ldots,$

$$\mathbf{P}\Big(z_{\alpha_0}(k_{m+1}) \leq \chi \cdot z_{\alpha_0}(k_m) + L \cdot \frac{1}{\ell}\Big|\mathcal{M}^c\Big) = 1,$$

which implies

$$\mathbf{E}\Big(z_{\alpha_0}(k_{m+1})\Big|\mathcal{M}^c\Big) \leq \Big(1 - (1 - \theta_0)\hat{p}\Big)\mathbf{E}\Big(z_{\alpha_0}(k_m)\Big|\mathcal{M}^c\Big) + L \cdot \frac{1}{\ell}.$$

Therefore, we can further obtain

$$\limsup_{m \to \infty} \mathbf{E}\Big(z_{\alpha_0}(k_m)\Big|\mathcal{M}^c\Big) \leq \frac{L}{(1 - \theta_0)\hat{p}} \cdot \frac{1}{\ell}. \tag{32}$$

Since ℓ can be any positive integer in (32) and $z_{\alpha_0}(k_m)$ is nonnegative for any m, we have

$$\lim_{m \to \infty} \mathbf{E}\Big(z_{\alpha_0}(k_m)\Big|\mathcal{M}^c\Big) = 0. \tag{33}$$

Based on Fatou's lemma, we know

$$0 \leq \mathbf{E}\Big(\lim_{m \to \infty} z_{\alpha_0}(k_m)|\mathcal{M}^c\Big) \leq \lim_{m \to \infty} \mathbf{E}\Big(z_{\alpha_0}(k_m)\Big|\mathcal{M}^c\Big) = 0, \tag{34}$$

which yields

$$\mathbf{P}\Big(\lim_{m \to \infty} z_{\alpha_0}(k_m) = 0\Big|\mathcal{M}^c\Big) = 1. \tag{35}$$

Finally, because α_0 is chosen arbitrarily over the network in (35), we see that

$$\mathbf{P}\Big(\mathcal{A}\Big|\mathcal{M}^c\Big) = 0. \tag{36}$$

The proof is completed. $\square$

4.2. Consensus analysis

In this subsection, we present the consensus analysis of the proof of Theorem 13. Let $x_{i,[J]}(k)$ represent the J'th coordinate of $x_i(k)$. Denote

$$h(k) = \min_{i=1,\ldots,n} x_{i,[J]}(k), \qquad H(k) = \max_{i=1,\ldots,n} x_{i,[J]}(k).$$

The consensus proof will be built on the estimates of $S(k) = H(k) - h(k)$, which is summarized in the following conclusion.

Proposition 17. *Algorithm* (5) *achieves a global consensus if* $\mathcal{G}_k$ *is SUSC.*

Proof. Since $\mathbf{P}(\mathcal{M}^c) \geq \mathbf{P}(\mathcal{A}^c) = 1$ when $\mathcal{G}_k$ is SUSC, we only need to prove

$$\mathbf{P}\Big(\lim_{k \to \infty} S(k) = 0\Big|\mathcal{M}^c\Big) = 1.$$

Let $\{x^\omega(k)\}_{k=0}^\infty$ be a sample sequence in $\mathcal{M}^c$. Then $\forall \ell = 1, 2, \ldots, \exists T_1(\ell, \omega) > 0$ such that

$$k \geq T_1 \implies |x_i^\omega(k)|_{X_i} \leq \frac{1}{\ell}, \quad i = 1, \ldots, n. \tag{37}$$

Moreover, based on similar analysis as in the proof of Lemma 15, we see that

$$h(k+s) \geq h(k) - s \cdot \frac{1}{\ell}; \qquad H(k+s) \leq H(k) + s \cdot \frac{1}{\ell} \tag{38}$$

for all $k \geq T_1$ and $s \geq 0$.

Denote $k_1 = T_1$. Take $v_0 \in \mathcal{V}$ with $x_{v_0,[J]}(k_1) = h(k_1)$. Then we can obtain from the definition of (5) that

$$x_{v_0,[J]}(k_1 + 1) \leq x_{v_0,[J]}(k_1) + \frac{1}{\ell}$$

if projection happens and

$$x_{v_0,[J]}(k_1 + 1) \leq a_{v_0 v_0}(k_1)x_{v_0,[J]}(k_1) + (1 - a_{v_1 v_0}(k_1))H(k_1)$$

if averaging happens, which leads to that almost surely we have

$$x_{v_0,[J]}(k_1 + 1) \leq \eta h(k_1) + (1 - \eta)H(k_1) + \frac{1}{\ell}.$$

Continuing the estimates we know that a.s. for any $\tau = 0, 1, \ldots,$

$$x_{v_0,[J]}(k_1 + \tau) \leq \eta^\tau h(k_1) + (1 - \eta^\tau)H(k_1) + \frac{\tau(\tau+1)}{2} \cdot \frac{1}{\ell}. \tag{39}$$

Furthermore, since $\mathcal{G}_k$ is SUSC, we have

$$\mathbf{P}\Big(\exists \hat{k}_1 \in [k_1, k_1 + B - 1] \text{ and } \exists v_1 \in \mathcal{V} \text{ s.t. } (v_0, v_1) \in \mathcal{G}_{\hat{k}_1}\Big)$$

$$\geq q. \tag{40}$$

Let $\hat{k}_1 = k_1 + \varrho, 0 \leq \varrho \leq B - 1$. Similarly with (27), we see from (39) that

$$\mathbf{P}\Big(x_{v_1,[J]}(k_1 + \varrho + 1) \leq \eta^{\varrho+1}h(k_1) + (1 - \eta^{\varrho+1})H(k_1)$$

$$+ \eta \cdot \frac{\varrho(\varrho+1)}{2} \cdot \frac{1}{\ell}\Big|\mathcal{M}^c\Big) \geq pq. \tag{41}$$

Similar analysis will lead to

$$\mathbf{P}\Big(x_{\nu_1,[j]}(k_1 + \varrho + \hat{\tau}) \leq \eta^{\varrho+\hat{\tau}}h(k_1) + (1 - \eta^{\varrho+\hat{\tau}})H(k_1)$$
$$+ \frac{(\varrho + \hat{\tau})(\varrho + \hat{\tau} + 1)}{2} \cdot \frac{1}{\ell}, \hat{\tau} = 1, 2, \ldots \big| \mathcal{M}^c\Big) \geq pq, \qquad (42)$$

which yields

$$\mathbf{P}\Big(x_{\nu_1,[j]}(k_1 + B + \tau) \leq \eta^{B+\tau}h(k_1) + (1 - \eta^{B+\tau})H(k_1)$$
$$+ \frac{(B + \tau)(B + \tau + 1)}{2} \cdot \frac{1}{\ell}, \tau = 0, 1, \ldots \big| \mathcal{M}^c\Big) \geq pq.$$

We can continue the upper process on time intervals $[k_1 + 2B, k_1 + 3B - 1], \ldots, [k_1 + (n - 2)B, k_1 + (n - 1)B - 1]$, and $\nu_2, \ldots, \nu_{n-1}$ can be found until

$$\mathbf{P}\Big(x_{\nu_l,[j]}(k_1 + (n - 1)B) \leq \eta^{(n-1)B}h(k_1)$$
$$+ (1 - \eta^{(n-1)B})H(k_1) + \frac{(n - 1)B((n - 1)B + 1)}{2} \cdot \frac{1}{\ell},$$
$$l = 0, 1, \ldots, n - 1 \big| \mathcal{M}^c\Big) \geq p^{n-1}q^{n-1}.$$

Therefore, denoting $k_2 = k_1 + (n - 1)B$, we have

$$\mathbf{P}\Big(H(k_2) \leq \eta^{(n-1)B}h(k_1) + (1 - \eta^{(n-1)B})H(k_1)$$
$$+ \frac{(n - 1)B((n - 1)B + 1)}{2} \cdot \frac{1}{\ell} \big| \mathcal{M}^c\Big) \geq p^{n-1}q^{n-1}.$$

Furthermore, with (38), we can further obtain

$$\mathbf{P}\Big(S(k_2) \leq (1 - \eta^{(n-1)B})S(k_1) + L_0 \cdot \frac{1}{\ell} \big| \mathcal{M}^c\Big) \geq p^{n-1}q^{n-1},$$

where $L_0 = \frac{(n-1)B[(n-1)B+3]}{2}$.

Then we know $\mathbf{P}\{\lim_{k\to\infty} S(k) = 0 | \mathcal{M}^c\} = 1$ by similar analysis as the proof of Proposition 16. The proof is completed. □

Theorem 13 immediately follows from Propositions 16 and 17.

5. Bidirectional graphs

In this section, we discuss the randomized optimal consensus problem under more restrictive communication assumptions, that is, bidirectional communications. To get the main result, we also need the following assumption in addition to the standing assumptions A1–A4.

A5 (Compactness) X_0 is compact.

Then we propose the main result on optimal consensus for the bidirectional case. It turns out that with bidirectional communications, the connectivity condition to ensure an optimal consensus is weaker.

Theorem 18. Suppose $\mathcal{G}_k$ is bidirectional for all $k \geq 0$ and A5 holds. Algorithm (5) achieves a global optimal consensus almost surely if $\mathcal{G}_k$ is SIC.

Remark 19. The essential difference between SUSC and SIC graphs is that SIC graphs do not impose an upper bound for the length of intervals where the joint graphs are taken. Therefore, the analysis on directed graphs cannot be used in this bidirectional case.

In the following two subsections, we will focus on the optimal solution set convergence and the consensus analysis, respectively, by which we will reach a complete proof for Theorem 18.

5.1. Set convergence

In this subsection, we discuss the convergence to the optimal solution set. First we give the following lemma.

Lemma 20. Assume that $\mathcal{G}_k$ is bidirectional for all $k \geq 0$. Then $\mathbf{P}(\mathcal{A} \cap \mathcal{M}) = 0$ if $\mathcal{G}_k$ is SIC.

Proof. The proof follows the same line as the proof of Lemma 14. Let k_m and T be defined in the same way as the proof of Lemma 14. Suppose $k_m \geq T$. Based on the definition of (5), we know from Lemma 11 that

$$\mathbf{P}\left(|x_{i_0}(k_m + 1)|_{X_0} \leq \sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right) \geq 1 - p. \qquad (43)$$

Next, we define

$$\hat{k}_1 \doteq \inf_{k \geq k_m+1}\left\{\exists j \in \mathcal{V} \text{ s.t. } (i_0, j) \in \mathcal{E}_k\right\}$$

and

$$\mathcal{V}_1 \doteq \left\{j \in \mathcal{V} : (i_0, j) \in \mathcal{E}_{\hat{k}_1}\right\}.$$

Based on the definition of SIC graphs, we have for all $\tau = 0, 1, \ldots,$

$$\mathbf{P}\Big(\exists j \in \mathcal{V}, k \in [k_\tau^*, k_{\tau+1}^*) \text{ s.t. } (i_0, j) \in \mathcal{E}_k\Big)$$
$$\geq \mathbf{P}\Big(\mathcal{G}([k_\tau^*, k_{\tau+1}^*)) \text{ is connected}\Big)$$
$$\geq q. \qquad (44)$$

Thus, Lemma 3 implies that the probability of $\hat{k}_1$ being finite is one.

Applying Lemma 11 on node i_0, we have

$$|x_{i_0}(s)|_{X_0} \leq |x_{i_0}(k_m + 1)|_{X_0}, \quad k_m + 1 \leq s \leq \hat{k}_1. \qquad (45)$$

As a result, we have

$$\mathbf{P}\left(|x_i(\hat{k}_1 + 1)|_{X_0} \leq \eta\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta)\left(\xi + \frac{1}{\ell}\right),\right.$$
$$\left. i \in \mathcal{V}_1\right) \geq p^{|\mathcal{V}_1|}(1 - p). \qquad (46)$$

We can repeat the upper process, $\mathcal{V}_2, \ldots, \mathcal{V}_{d_0}$ can be defined iteratively for some constant $1 \leq d_0 \leq n - 1$ until $\mathcal{V} \setminus \{i_0\} = \bigcup_{j=1}^{d_0} \mathcal{V}_j$. Denoting $\varsigma_m = \hat{k}_{d_0} + 1$ associated with $\mathcal{V}_{d_0}$, we have

$$\mathbf{P}\left(|x_i(\varsigma_m)|_{X_0} \leq \eta^{d_0}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right.$$
$$\left. + (1 - \eta^{d_0})\left(\xi + \frac{1}{\ell}\right), i \in \mathcal{V}\right)$$
$$= \mathbf{P}\left(\max_{i=1,\ldots,n} |x_i(\varsigma_m)|_{X_0} \leq \eta^{d_0}\sqrt{\left(\xi + \frac{1}{\ell}\right)^2 - \frac{1}{4}\delta_{i_0}^2}\right.$$
$$\left. + (1 - \eta^{d_0})\left(\xi + \frac{1}{\ell}\right)\right)$$
$$\geq p^{n-1}(1 - p). \qquad (47)$$

This will also lead to

$$\mathbf{P}\big(\bar{\mathcal{R}}_*\big) = 1, \qquad (48)$$

where $\bar{\mathcal{R}}_* = \left\{\xi \leq \eta^{n-1}\sqrt{\xi^2 - \frac{1}{4}\delta_{i_0}^2} + (1 - \eta^{n-1})\xi\right\}$. Noting the fact that $\mathcal{A} \cap \mathcal{M} \subseteq \bar{\mathcal{R}}_*^c$, the conclusion holds. □

Next, we define

$$y_i = \liminf_{k \to \infty} |x_i(k)|_{X_0}, \quad i = 1, \dots, n$$

and denote $\mathcal{D} = \{\exists i_0 \ s.t. \ y_{i_0} < \xi\}$. We give another lemma in the following.

Lemma 21. *Assume that $\mathcal{G}_k$ is bidirectional for all $k \geq 0$. Then $\mathbf{P}(\mathcal{A} \cap \mathcal{D}) = 0$ if $\mathcal{G}_k$ is SIC.*

Proof. The proof will follow the same idea as the proof of Lemma 20. Let $\{x^\omega(k)\}_{k=0}^\infty$ be a sample sequence. There exists a time sequence $k_1(\omega) < \cdots < k_m(\omega) < \cdots$ with $\lim_{m \to \infty} k_m(\omega) = \infty$ such that

$$|x_{i_0}^\omega(k_m(\omega))|_{X_0} \leq \frac{1}{2}(y_{i_0}(\omega) + \xi(\omega)). \tag{49}$$

Moreover, $\forall \ell = 1, 2, \dots, \exists T(\ell, \omega) > 0$ such that

$$k \geq T \Rightarrow 0 \leq |x_i^\omega(k)|_{X_0} \leq \xi(\omega) + \frac{1}{\ell}, \quad i = 1, \dots, n. \tag{50}$$

Let $\hat{k}_1$ and $\mathcal{V}_1$ follow the definition in the proof of Lemma 20, by the same argument as we obtain (46), we have

$$\mathbf{P}\left(|x_i(\hat{k}_1 + 1)|_{X_0} \leq \frac{\eta}{2}y_{i_0} + \left(1 - \frac{\eta}{2}\right)\left(\xi + \frac{1}{\ell}\right), i \in \mathcal{V}_1\right)$$
$$\geq p^{|\mathcal{V}_1|}. \tag{51}$$

Continuing the upper process, we will also reach

$$\mathbf{P}\left(\max_{i=1,\dots,n} |x_i(\varsigma_m)|_{X_0} \leq \frac{\eta^{d_0}}{2} \cdot y_{i_0} + (1 - \frac{\eta^{d_0}}{2})\left(\xi + \frac{1}{\ell}\right)\right)$$
$$\geq p^{n-1}, \tag{52}$$

where $1 \leq d_0 \leq n - 1$ and ς_m still denotes $\hat{k}_{d_0} + 1$. Introducing

$$\mathcal{W} = \left\{\xi \leq \frac{\eta^{d_0}}{2} \cdot y_{i_0} + \left(1 - \frac{\eta^{d_0}}{2}\right) \cdot \xi\right\},$$

we can similarly obtain $\mathbf{P}(\mathcal{W}) = 1$ according to (52). The fact that $\mathcal{A} \cap \mathcal{D} \subseteq \mathcal{W}^c$ implies the desired conclusion immediately. $\qquad \square$

Note that, if A5 holds, according to Lemma 12, for any initial condition x^0, we have

$$x_i(k) \in X_0^*, \quad i = 1, \dots, n; \ k = 0, 1, \dots,$$

where $X_0^* \doteq \{v : |v|_{X_0} \leq d_*\}$ with $d_* = \max_{i=1,\dots,n} |x_i(0)|_{X_0}$. Then X_0^* is also a compact set, which is an invariant set for (5). Therefore, for any initial condition, there will also be two constants $b_1, b_2 > 0$ such that

$$|x_i(k) - x_j(k)| \leq b_1; \qquad |x_i(k)|_{X_0} \leq b_2 \tag{53}$$

for all i, j and k.

Now we are ready to prove the optimal set convergence part of Theorem 13, which is stated in the following conclusion.

Proposition 22. *Assume $\mathcal{G}_k$ is bidirectional for all $k \geq 0$ and A5 holds. Algorithm (5) achieves a global optimal set aggregation a.s. if $\mathcal{G}_k$ is SIC.*

Proof. With Lemmas 20 and 21, we only need to show

$$\mathbf{P}(\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c) = 0.$$

Take $i_0 \in \mathcal{V}$. Then we define two parallel hyperplanes

$$W_{i_0}(k) \doteq \{v | \langle x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v - x_{i_0}(k) \rangle = 0\}$$
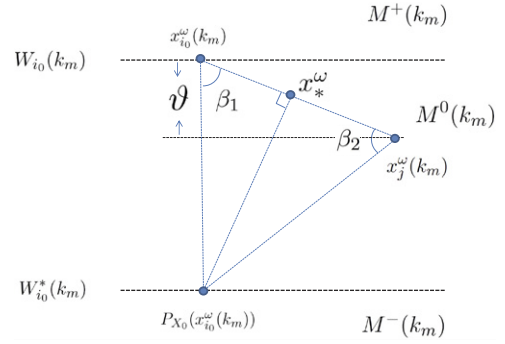

Fig. 2. Finding the point x_*^ω in the proof of Proposition 22.

and

$$W_{i_0}^*(k) \doteq \{v | \langle x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v - P_{X_0}(x_{i_0}(k)) \rangle = 0\}.$$

The space $\mathbb{R}^d$ is divided by the two hyperplanes into three disjoint parts $M^+(k) = \{v | \langle x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v - x_{i_0}(k) \rangle < 0\}$, $M^-(k) = \{v | \langle x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v - P_{X_0}(x_{i_0}(k)) \rangle < 0\}$, and the rest $M^0(k)$ (see Fig. 2). Also define $\mathcal{N}_{i_0}^\infty = \{j | j$ is a neighbor of i for infinitely many $k\}$.

Claim. $\mathbf{P}\{\lim_{k \to \infty} |x_j(k)|_{W_{i_0}(k)} = 0, j \in \mathcal{N}_{i_0}^\infty | \mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c\} = 1$.

Let $\{x^\omega(k)\}_{k=0}^\infty$ be a sample sequence in $\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c$. Then $\forall \ell = 1, 2, \dots, \exists T(\ell, \omega) > 0$ such that

$$0 < \xi(\omega) \leq |x_i^\omega(k)|_{X_0} \leq \xi(\omega) + \frac{1}{\ell} \quad \text{and} \quad |x_i^\omega(k)|_{X_i} \leq \frac{1}{\ell} \tag{54}$$

for $i = 1, \dots, n$ when $k \geq T$.

Suppose there exist a constant $\vartheta^\omega > 0$ and a sequence $k_1 < \cdots < k_m < \cdots$ such that $|x_j^\omega(k_m)|_{W_{i_0}(k_m)} \geq \vartheta^\omega, m = 1, 2, \dots$. Take $k_m \geq T$. With (1), we see that for all $k = 1, 2, \dots$,

$$X_0 \subseteq M^-(k) \cup W_{i_0}^*(k)$$
$$= \{v | \langle x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v - P_{X_0}(x_{i_0}(k)) \rangle \leq 0\}.$$

Let $x_{i_0}^\omega(k_m)$ and $P_{X_0}(x_{i_0}^\omega(k_m))$ be fixed. Then we can associate a unique point x_*^ω to $x_j^\omega(k_m)$ in the way that x_*^ω satisfies $\langle P_{X_0}(x_{i_0}^\omega(k_m)) - x_*^\omega, x_{i_0}^\omega(k_m) - x_j^\omega(k_m) \rangle = 0$ if the three points $x_{i_0}^\omega(k_m)$, $P_{X_0}(x_{i_0}^\omega(k_m))$ and $x_j^\omega(k_m)$ form a triangle; and $x_*^\omega = P_{X_0}(x_{i_0}^\omega(k_m))$ otherwise. Moreover, it is not hard to find that there exists a unique scalar $0 < \gamma < 1$ such that $x_*^\omega = \gamma x_{i_0}^\omega(k_m) + (1 - \gamma)x_j^\omega(k_m)$. Note that, the upper process defines a continuous function $(x_{i_0}^\omega(k_m), P_{X_0}(x_{i_0}^\omega(k_m)), x_j^\omega(k_m)) \mapsto \gamma$. With (53), we have $(x_{i_0}^\omega(k_m), P_{X_0}(x_{i_0}^\omega(k_m)), x_j^\omega(k_m))$ always locates within a compact set $\{0 \leq |x_{i_0}^\omega(k_m)|_{X_0} \leq d_*; \vartheta^\omega \leq |x_j^\omega(k_m) - x_{i_0}^\omega(k_m)| \leq b_1; \xi(\omega) \leq |x_j^\omega(k_m) - P_{X_0}(x_{i_0}^\omega(k_m))| \leq b_1 + b_2\}$. Therefore, there exist two constants $0 < \gamma_* \leq \gamma^* < 1$ (by a constant, we mean it does not depend on k_m) such that $\gamma_* \leq \gamma \leq \gamma^*$ (see Fig. 2).

Thus, every linear combination of $x_{i_0}^\omega(k_m)$ and $x_j^\omega(k_m)$ can be rewritten into a linear combination of $x_{i_0}^\omega(k_m)$ and x_*^ω, and the lower bound of the weights is preserved. We also have

$$|x_*^\omega|_{X_0} \leq |x_*^\omega - P_{X_0}(x_{i_0}^\omega(k_m))|$$
$$\leq \sin \beta_0 \left(\xi(\omega) + \frac{1}{\ell}\right) \leq b_* \left(\xi(\omega) + \frac{1}{\ell}\right), \tag{55}$$

where $\beta_0 = \phi(x_j^\omega(k_m) - x_{i_0}^\omega(k_m), P_{X_0}(x_{i_0}^\omega(k_m)) - x_{i_0}^\omega(k_m))$ and $0 < b_* = \sqrt{1 - (\frac{\vartheta}{b_1})^2} < 1$. Therefore, with (55), repeating the deduction used in the proofs of Lemmas 20 and 21, the claim can then be proved.
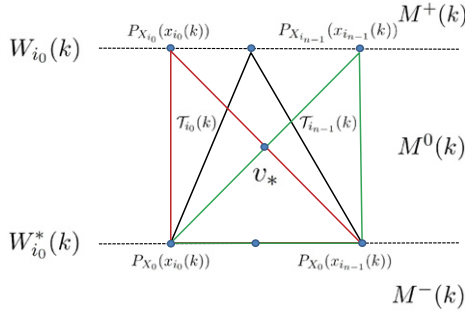
Fig. 3. Locating v_* in the proof of Proposition 22.

Next, since $\mathcal{G}_k$ is SIC, i.e., the joint graph is connected with probability $q > 0$ independently for infinite times, letting $\mathcal{G}_\infty$ be the graph generated by neighbor sets $\mathcal{N}_{i_0}^\infty$, it is obvious that $\mathcal{G}_\infty$ is connected with probability 1. Therefore, the upper analysis can then be further carried out on $\mathcal{G}_\infty$ following i_0's neighbors, i_0's neighbors' neighbors, and so on, until we finally reach

$$\lim_{k \to \infty} |x_j(k)|_{W_{i_0}(k)} = 0; \tag{56}$$

with probability 1 for all $j \in \mathcal{V}$ conditioned $\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c$. Thus, by the definition of W_{i_0} and (1), we have

$$\mathbf{P}\Big(\lim_{k \to \infty} |P_{W_{i_0}^*}(x_j(k)) - P_{X_0}(x_j(k))| = 0,$$

$$j = 1, \ldots, n | \mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c\Big) = 1. \tag{57}$$

Denote

$$\mathcal{T}_{i_0}(k) = co\{P_{X_{i_0}}(x_{i_0}(k)), P_{X_0}(x_1(k)), \ldots, P_{X_0}(x_n(k))\}.$$

Then $\mathcal{T}_{i_0}(k) \subseteq X_{i_0}, \forall k \geq 0$. $\mathcal{T}_{i_m}(k)$ can then be defined for $m = 1, \ldots, n-1$ in the same way. Therefore, with (56), and according to the structure of $W_{i_0}(k)$ and $W_{i_0}^*(k)$, with probability 1 conditioned $\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c$, there will be a point $v_* \in \bigcap_{m=0}^{n-1} \mathcal{T}_{i_m}(k) \subseteq X_0$ for sufficiently large k such that $v_* \in M^0(k)$ (see Fig. 3), i.e.,

$$\mathbf{P}\Big(\exists k \text{ s.t. } \langle x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v_* - P_{X_0}(x_{i_0}(k))\rangle > 0$$

$$\Big| \mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c\Big) = 1. \tag{58}$$

This implies $\mathbf{P}\big(\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c\big) = 0$ because $\mathbf{P}\big(\langle y - P_{X_0}(y), v_* - P_{X_0}(y)\rangle > 0\big) = 0$ for any $y \in \mathbb{R}^d$ and $v_* \in X_0$ according to (1). The proof is completed. □

5.2. Consensus analysis

This subsection focuses on the consensus analysis of Theorem 18.

We define a multi-projection function: $P_{i_k i_{k-1} \ldots i_1} : \mathbb{R}^m \to \bigcup_{i=1}^n X_i$ with $i_1, \ldots, i_k \in \{1, \ldots, n\}$, $k \geq 1$ by $P_{i_k i_{k-1} \ldots i_1}(y) = P_{X_{i_k}} P_{X_{i_{k-1}}} \ldots P_{X_{i_1}}(y)$. Define $P_\emptyset(y) = y$ as the case for $k = 0$. Let

$$\Gamma \doteq \big\{P_{i_k i_{k-1} \ldots i_1} : i_1, \ldots, i_k \in \{1, \ldots, N\}, k = 0, 1, 2, \ldots\big\}$$

be the set which contains all the multi-projection functions. Denote $Y_k = co\{x_1(k), \ldots, x_n(k)\}$ be the convex hull of all the nodes's state at step k, and define Δ_{Y_k} by $\Delta_{Y_k} \doteq co\{P(y) | y \in Y_k, P \in \Gamma\}$. Then it is not hard to see that Δ_{Y_k} is actually an invariant set along Algorithm (5) for any $k \geq 0$, i.e., $x_i(s) \in \Delta_{Y_k}$ for all i, k and $s \geq k$.

We present another lemma establishing an important property of Δ_{Y_k}.

Lemma 23. For any $y \in \Delta_{Y_k}$, we have $|y|_{Y_k} \leq 2 \max_{y \in Y_k} |y|_{X_0}$.

Proof. With Lemma 2, any $y \in \Delta_{Y_K}$ has the following form

$$y = \sum_{i=1}^{d+1} \lambda_i P^{\langle i \rangle}(z_i),$$

where $\sum_{i=1}^{d+1} \lambda_i = 1$ with $\lambda_i \geq 0$, $P^{\langle i \rangle} \in \Gamma$ and $z_i \in Y_K$, $i = 1, \ldots, d+1$. Then, by the non-expansiveness property (2), we have that for any $z \in \mathbb{R}^d$ and $\hat{P} \in \Gamma$,

$$|P_{X_0}(z) - \hat{P}(z)| = |\hat{P}(P_{X_0}(z)) - \hat{P}(z)| \leq |P_{X_0}(z) - z| = |z|_{X_0}.$$

This leads to

$$\Big| \sum_{i=1}^{d+1} \lambda_i P^{\langle i \rangle}(z_i) - \sum_{i=1}^{d+1} \lambda_i z_i \Big|$$

$$\leq \sum_{i=1}^{d+1} \lambda_i \Big| z_i - P_{X_0}(z_i) \Big| + \sum_{i=1}^{d+1} \lambda_i \Big| P_{X_0}(z_i) - P^{\langle i \rangle}(z_i) \Big|$$

$$\leq 2 \max_{z \in K} |z|_{X_0},$$

which implies the conclusion because $\sum_{i=1}^{d+1} \lambda_i z_i \in Y_k$. □

We can now present the consensus analysis.

Proposition 24. Assume that $\mathcal{G}_k$ is bidirectional for all $k \geq 0$ and A5 holds. Algorithm (5) achieves a global consensus a.s. if $\mathcal{G}_k$ is SIC.

Proof. We only need to show $\mathbf{P}(\lim_{k \to \infty} S(k) = 0 | \mathcal{A}^c) = 1$. Let $\{x^\omega(k)\}_{k=0}^\infty$ be a sample sequence in $\mathcal{A}^c$. Then $\forall \ell = 1, 2, \ldots, \exists T_1(\ell, \omega) > 0$ such that

$$k \geq T_1 \Rightarrow |x_i^\omega(k)|_{X_0} \leq \frac{1}{\ell}, \quad i = 1, \ldots, n. \tag{59}$$

As a consequence, Lemma 23 implies

$$h(k+s) \geq h(k) - \frac{2}{\ell}; \qquad H(k+s) \leq H(k) + \frac{2}{\ell} \tag{60}$$

for all $k \geq T_1$ and $s \geq 0$.

Denote $k_1 = T_1$. Take $v_0 \in \mathcal{V}$ with $x_{v_0,[J]}(k_1) = h(k_1)$. Define

$$\hat{k}_1 \doteq \inf_{k \geq k_1} \{\exists j \in \mathcal{V} \text{ s.t. } (v_0, j) \in \mathcal{E}_k\}$$

and

$$\mathcal{V}_1 \doteq \big\{j \in \mathcal{V} : (v_0, j) \in \mathcal{E}_{\hat{k}_1}\big\}.$$

With (59), we have

$$x_{v_0,[J]}(\hat{k}_1) \leq x_{v_0,[J]}(k_1) + \frac{1}{\ell} = h(k_1) + \frac{1}{\ell}. \tag{61}$$

Thus,

$$\mathbf{P}\Big(x_{v_1,[J]}(\hat{k}_1 + 1) \leq \eta h(k_1) + (1-\eta)H(k_1) + \frac{2}{\ell} \Big| \mathcal{A}^c\Big) \geq p \tag{62}$$

for any $v_1 \in \mathcal{V}_1$, which leads to

$$\mathbf{P}\Big(x_{i,[J]}(\hat{k}_1 + 1) \leq \eta h(k_1) + (1-\eta)H(k_1) + \frac{2}{\ell},$$

$$i \in \mathcal{V}_1 \Big| \mathcal{A}^c\Big) \geq p^{|\mathcal{V}_1|}. \tag{63}$$

Similar to the proof of Lemma 20, we can repeat the upper process, and $\mathcal{V}_2, \ldots, \mathcal{V}_{d_0}$ can be defined for some constant $1 \leq$
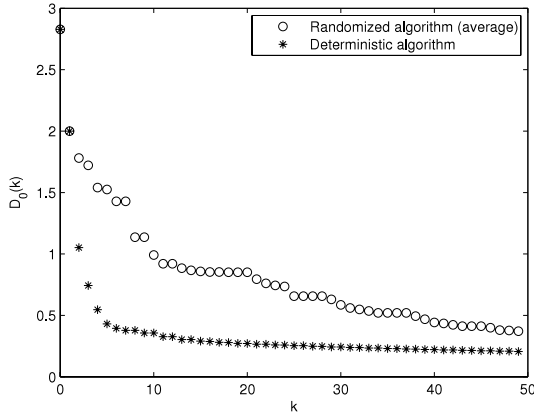
Fig. 4. The average performance of the randomized algorithm. Here $D_0(k) = \max_{i=1,2,3} |x_i(k)|_{X_0}$.

$d_0 \le n - 1$ until $\mathcal{V} \setminus \{v_0\} = \bigcup_{j=1}^{d_0} \mathcal{V}_j$. Moreover, we can also obtain that

$$\mathbf{P}\Big(x_{i,[J]}(\hat{k}_{d_0} + 1) \le \eta^{d_0} h(k_1) + (1 - \eta^{d_0})H(k_1) + \frac{2d_0}{\ell},$$
$$i \in \mathcal{V} \Big| \mathcal{A}^c\Big) \ge p^{n-1}. \tag{64}$$

Therefore, denoting $k_2 = \hat{k}_{d_0} + 1$, we have

$$\mathbf{P}\Big(H(k_2) \le \eta^{d_0} h(k_1) + (1 - \eta^{d_0})H(k_1) + \frac{2d_0}{\ell} \Big| \mathcal{A}^c\Big) \ge p^{n-1}. \tag{65}$$

We see from (60), (65) that

$$\mathbf{P}\Big(S(k_2) \le (1 - \eta^{d_0})S(k_1) + L_0 \cdot \frac{2(d_0 + 1)}{\ell} \Big| \mathcal{A}^c\Big) \ge p^{n-1}.$$

Then we know $\mathbf{P}\big(\lim_{k\to\infty} S(k) = 0|\mathcal{A}^c\big) = 1$ by similar deduction as the proof of Proposition 16. The proof is completed. □

Then we see that Theorem 18 follows from Propositions 22 and 24.

6. Numerical example

In this section, we study a numerical example to compare the convergence rates of deterministic and randomized algorithms, and to illustrate the optimal choice of the decision probability p in the randomized algorithm.

Consider a network with three nodes $\mathcal{V} = \{1, 2, 3\}$. The communication graph is fixed and directed. Here $\mathcal{E} = \{(1, 2), (2, 3), (3, 1)\}$ is the arc set. We take $a_{ij}(k) = 0.5$ for all $(i, j) \in \mathcal{E}$. The optimal solution sets corresponding to the nodes are three disks in $\mathbb{R}^2$ with radius 1 and centers $(-1, 0)$, $(1, 0)$ and $(0, -1)$, respectively. Their intersection $X_0 = \{(0, 0)\}$ is a singleton. Initial values for each node are $(-2, 2)$, $(-2, -2)$ and $(2, -2)$, respectively.

We compare the randomized algorithm presented in this paper and the projected consensus algorithm in (Nedić et al., 2010), which is a deterministic algorithm with each node taking averaging and projection alternatively. Numerical experiments show that the deterministic algorithm leads to a faster convergence than the mean performance of the randomized algorithm. The reason for this is natural since consecutive projections may take place for some nodes. However, surprisingly enough there is still about 5 percent of the experiments for which the randomized algorithm performs better than the deterministic one. Moreover, we also find that the randomized algorithm usually converges faster near X_0. See Figs. 4 and 5.
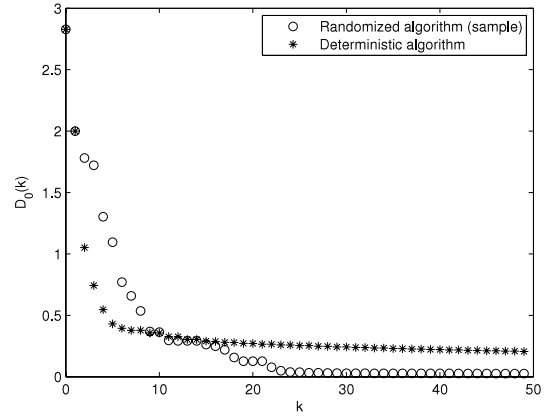

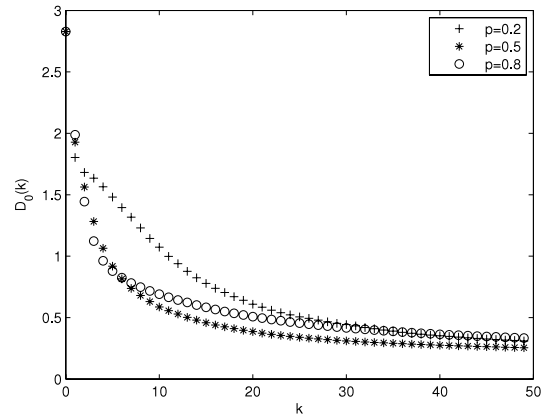**Fig. 5.** A typical sample when the randomized algorithm converges faster.

Fig. 6. Convergence rates for different decision probabilities ($D_0(k) = \max_{i=1,2,3} |x_i(k)|_{X_0}$).

We further compare the average performance of the randomized algorithm when p takes values from 0.2, 0.5 and 0.8. Experiments show that the case when $p = 0.5$ reaches the fastest convergence. This is to say, it is better for the nodes to balance computation (projection) and communication (averaging). See Fig. 6.

7. Conclusions

The paper investigated a randomized optimal consensus problem for multi-agent systems with stochastically time-varying interconnection topology. In this formulation, the decision process for each agent was a simple Bernoulli trial between following its neighbors or sticking to its own opinion at each time step. In terms of the optimization problem, each agent independently chose either taking an average among its time-varying neighbor set, or projecting onto the optimal solution set of its own objective function randomly with a fixed probability. Both directed and bidirectional communications were studied, and stochastically jointly connectivity conditions were proposed to guarantee an optimal consensus almost surely. The results showed that under this randomized decision making protocol, a group of autonomous agents can reach an optimal opinion with probability 1 with proper convex and nonempty intersection assumptions for the considered optimization problem. Fundamental challenges still lie in the convergence rate of the randomized algorithm and the choice of optimal decision probability to reach a faster convergence.

Acknowledgments

References

Acemoglu, D., Ozdaglar, A., & ParandehGheibi, A. (2010). Spread of (mis)information in social networks. *Games and Economic Behavior*, *70*(2), 194–227.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society¿*, *68*(3), 337–404.

Aubin, J., & Cellina, A. (1984). *Differential inclusions*. Berlin: Springer-Verlag.

Bertsekas, D. P., & Tsitsiklis, J. N. (2002). *Introduction to probability*. Massachusetts: Athena Scientific.

Boyd, S., Ghosh, A., Prabhakar, B., & Shah, D. (2006). Randomized gossip algorithms. *IEEE Trans. Information Theory*, *52*(6), 2508–2530.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. New York, NY: Cambridge University Press.

Cao, M., Morse, A. S., & Anderson, B. D. O. (2008a). Reaching a consensus in a dynamically changing environment: A graphical approach. *SIAM Journal on Control and Optimization*, *47*(2), 575–600.

Cao, M., Morse, A. S., & Anderson, B. D. O. (2008b). Reaching a consensus in a dynamically changing environment: Convergence rates, measurement delays, and asynchronous events. *SIAM Journal on Control and Optimization*, *47*(2), 601–623.

Deutsch, F. (1983). Rate of convergence of the method of alternating projections. In B. Brosowski, & F. Deutsch (Eds.), *Parametric optimization and approximation*: *vol. 76* (pp. 96–107). Basel, Switzerland: Birkhäuser.

Fagnani, F., & Zampieri, S. (2008). Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, *26*(4), 634–649.

Fax, J., & Murray, R. (2004). Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, *49*(9), 1465–1476.

Godsil, C., & Royle, G. (2001). *Algebraic graph theory*. New York: Springer-Verlag.

Gubin, L. G., Polyak, B. T., & Raik, E. V. (1967). The method of projections for finding the common point of convex sets. *U.S.S.R Comput. Math. Math. Phys.*, *7*(6), 1211–1228.

Hatano, Y., & Mesbahi, M. (2005). Agreement over random networks. *IEEE Transactions on Automatic Control*, *50*(11), 1867–1872.

Hong, Y., Hu, J., & Gao, L. (2006). Tracking control for multi-agent consensus with an active leader and variable topology. *Automatica*, *42*(7), 1177–1182.

Jadbabaie, A., Lin, J., & Morse, A. S. (2003). Coordination of groups of autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, *48*(6), 988–1001.

Jakovetić, D., Xavier, J., & Moura, J.M.F. (2011) Cooperative convex optimization in networked systems: augmented Lagrangian algorithms with directed gossip communication. arxiv:1007.3706.

Johansson, B., Keviczky, T., Johansson, M., & Johansson, K. H. (2008). Subgradient methods and consensus algorithms for solving convex optimization problems. In *Proc. IEEE conference on decision and control* (pp. 4185–4190). Cancun.

Johansson, B., Rabi, M., & Johansson, M. (2009). A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, *20*(3), 1157–1170.

Lin, Z., Francis, B., & Maggiore, M. (2007). State agreement for continuous-time coupled nonlinear systems. *SIAM Journal on Control and Optimization*, *46*(1), 288–307.

Lu, J., Regier, P.R., & Tang, C.Y. (2010) Control of distributed convex optimization. In *Proc. IEEE conference on decision and control*, Atlanta (pp. 489–495).

Lu, J., Tang, C.Y., Regier, P.R., & Bow, T.D. (2010) A gossip algorithm for convex consensus optimization over networks, In *Proc. american control conference*, Baltimore (pp. 301–308).

Martinez, S., Cortés, J., & Bullo, F. (2007). Motion coordination with distributed information. *IEEE Control Systems Magazine*, *27*(4), 75–88.

Moreau, L. (2005). Stability of multi-agent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, *50*(2), 169–182.

Nedić, A., Olshevsky, A., Ozdaglar, A., & Tsitsiklis, J. N. (2009). On distributed averaging algorithms and quantization effects. *IEEE Transactions on Automatic Control*, *54*(11), 2506–2517.

Nedić, A., & Ozdaglar, A. (2009). Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, *54*(1), 48–61.

Nedić, A., Ozdaglar, A., & Parrilo, P. A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, *55*(4), 922–938.

Olfati-Saber, R. (2006). Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Transactions on Automatic Control*, *51*(3), 401–420.

Olfati-Saber, R., & Murray, R. (2004). Consensus problems in the networks of agents with switching topology and time dealys. *IEEE Transactions on Automatic Control*, *49*(9), 1520–1533.

Rabbat, M., & Nowak, R. (2004) Distributed optimization in sensor networks, In IPSN'04, (pp. 20–27).

Ram, S. S., Nedić, A., & Veeravalli, V. V. (2007). Stochastic incremental gradient descent for estimation in sensor networks. In *Proc. asilomar conference on signals, systems, and computers* (pp. 582–586). Pacific Grove.

Ram, S. S., Nedić, A., & Veeravalli, V. V. (2009). Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, *20*(2), 691–717.

Ren, W., & Beard, R. (2008). *Distributed consensus in multi-vehicle cooperative control*. London: Springer-Verlag.

Ren, W., & Beard, R. (2005). Consensus seeking in multi-agent systems under dynamically changing interaction topologies. *IEEE Transactions on Automatic Control*, *50*(5), 655–661.

Rockafellar, R. T. (1972). *Convex analysis*. New Jersey: Princeton University Press.

Shi, G., & Hong, Y. (2009). Global target aggregation and state agreement of nonlinear multi-agent systems with switching topologies. *Automatica*, *45*(5), 1165–1175.

Shi, G., Hong, Y., & Johansson, K. H. (2012). Connectivity and set tracking of multi-agent systems guided by multiple moving leaders. *IEEE Transactions on Automatic Control*, *57*(3), 663–676.

Shi, G., Johansson, K.H., & Hong, Y. (2011) Multi-agent systems reaching optimal consensus with directed communication graphs. In *Proc. american control conference*, San Francisco (pp. 5456–5461).

Tahbaz-Salehi, A., & Jadbabaie, A. (2008). A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, *53*(3), 791–795.

Tanner, H. G., Jadbabaie, A., & Pappas, G. J. (2007). Flocking in fixed and switching networks. *IEEE Transactions on Automatic Control*, *52*(5), 863–868.

Tsitsiklis, J. N., Bertsekas, D. P., & Athans, M. (1986). Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, *31*(9), 803–812.

Xiao, F., & Wang, L. (2008). Asynchronous consensus in continuous-time multi-agent systems with switching topology and time-varying delays. *IEEE Transactions on Automatic Control*, *53*(8), 1804–1816.

Guodong Shi received his B.Sc. degree in Pure and Applied Mathematics from Shandong University, Jinan, China, in 2005, and Ph.D. degree in Systems Theory from Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2010, respectively. Currently he is a postdoctoral researcher at the ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Sweden. His research interests include distributed control, computation, and optimization methods.

Karl Henrik Johansson is Director of the KTH ACCESS Linnaeus Centre and Professor at the School of Electrical Engineering, Royal Institute of Technology, Sweden. He is a Wallenberg Scholar and has held a Senior Researcher Position with the Swedish Research Council. He received MSc and PhD degrees in Electrical Engineering from Lund University. He has held visiting positions at UC Berkeley (1998–2000) and California Institute of Technology (2006–2007). His research interests are in networked control systems, hybrid and embedded control, and control applications in automotive, automation and communication systems. He was a member of the IEEE Control Systems Society Board of Governors 2009 and the Chair of the IFAC Technical Committee on Networked Systems 2008–2011. He has been on the Editorial Boards of Automatica (2003–2006) and IEEE Transactions on Automatic Control (2008–2010), and is currently on the Editorial Boards of IET Control Theory and Applications and the International Journal of Robust and Nonlinear Control. He was the General Chair of the ACM/IEEE Cyber-Physical Systems Week (CPSWeek) 2010 in Stockholm. He has served on the Executive Committees of several European research projects in the area of networked embedded systems. He was awarded an Individual Grant for the Advancement of Research Leaders from the Swedish Foundation for Strategic Research in 2005. He received the triennial Young Author Prize from IFAC in 1996 and the Peccei Award from the International Institute of System Analysis, Austria, in 1993. He received Young Researcher Awards from Scania in 1996 and from Ericsson in 1998 and 1999.