Randomized Optimal Consensus of Multi-agent Systems∗

Guodong Shi and Karl Henrik Johansson†

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 \( f_i \) corresponding to agent \( i \) can only be observed by agent \( i \) itself. At each time step, 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. The results illustrate that a group of autonomous agents can reach an optimal opinion with probability one by each node simply making a randomized trade-off between following its neighbors or sticking to its own opinion at each time step.

Keywords: Multi-agent systems, Optimal consensus, Set convergence, Distributed optimization, Randomized algorithms

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

∗This work has been supported in part by the Knut and Alice Wallenberg Foundation, the Swedish Research Council and KTH SRA TNG.
†G. Shi and K. H. Johansson are with ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm 10044, Sweden. Email: guodongs@kth.se, kallej@kth.se
multi-agent systems is an active research topic, where collective tasks are enabled by the recent developments of distributed control protocols via interconnected communication \[6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 19\]. 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 \[6, 7, 22, 17, 21\]. 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 \[8, 18\]. Moreover, consensus seeking over randomly varying networks has been proposed in the literature \[24, 25, 26, 27, 28\], 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 \[29, 30, 31, 32, 33, 34\]. 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 \[29, 30, 38\]. Then in \[33\] 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 \[6, 7, 8\]. 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 \[32\]. An important contribution on multi-agent op-
timization is [36], 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 [37], 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. Then a convex-projection-based distributed control was presented for multi-agent systems with continuous-time dynamics to solve this optimization problem asymptotically [35].

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 the optimal solution set of $f_i$, is a convex set, and can be observed only by node $i$. Then at each time step, there are two options for each agent: an average (consensus) part as a convex combination of its neighbors’ state, and an 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 $1-p$. 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 [28]. 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 [28] 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 [25, 27, 20]. Borrowing the ideas on uniform joint-connection [6, 7, 22] and $[t, \infty)$-joint connectedness [8, 18], we introduce connectivity conditions of stochastically uniformly (jointly) strongly connected (SUSC) and stochastically jointly connected (SJC) 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 SJC 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. Finally, concluding remarks are given in Section 6.

2 Preliminaries

Here we introduce some mathematical notations and tools on graph theory [5], convex analysis [2, 3] and Bernoulli trials [4].

2.1 Directed Graphs

A directed graph (digraph) $G = (\mathcal{V}, \mathcal{E})$ consists of a finite set $\mathcal{V} = \{1, \ldots, n\}$ 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 $v_0e_1v_1e_2v_2\ldots e_nv_n$ of nodes $v_i$ and arcs $e_i = (v_{i-1}, v_i) \in \mathcal{E}$ for $i = 1, 2, \ldots, n$, the sequence is called a (directed) path. A path from $i$ to $j$ is denoted $i \rightarrow j$. $G$ is said to be strongly connected if it contains paths $i \rightarrow j$ and $j \rightarrow i$ for every pair of nodes $i$ and $j$.

A weighted digraph $G$ is a digraph with weights assigned for its arcs. A weighted digraph $G$ 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 $G$ 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 $G_1 = (\mathcal{V}, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}, \mathcal{E}_2)$ have the same node set, the union of the two digraphs is defined as $G_1 \cup G_2 = (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$.

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 $0 \leq \lambda \leq 1$. 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 $\text{co}(S)$.

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

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

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

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

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

$$f(\alpha v + (1 - \alpha)w) \leq \alpha f(v) + (1 - \alpha)f(w),$$  \hfill (3)

for all $v, w \in \mathbb{R}^d$ and $0 \leq \alpha \leq 1$. The following conclusion holds.

**Lemma 2.1** Let $K$ be a convex set in $\mathbb{R}^d$. Then $|x|_K$ is a convex function.

**Proof.** Suppose $x_1, x_2 \in K$ and $0 \leq \alpha \leq 1$. Then we have

$$|\alpha x_1 + (1 - \alpha)x_2|_K = \inf_{y \in K} |\alpha x_1 + (1 - \alpha)x_2 - y|$$

$$= \inf_{y_1, y_2 \in K} |\alpha x_1 + (1 - \alpha)x_2 - (\alpha y_1 + (1 - \alpha)y_2)|$$

$$\leq \inf_{y_1, y_2 \in K} \alpha |x_1 - y_1| + (1 - \alpha)|x_2 - y_2|$$

$$= \alpha |x_1|_K + (1 - \alpha) |x_2|_K,$$

which completes the proof. \hfill \Box

The next lemma can be found in [1].

**Lemma 2.2** Let $K$ be a subset of $\mathbb{R}^d$. The convex hull $\text{co}(K)$ 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 $\cos \phi = \langle v_1, v_2 \rangle / |v_1| \cdot |v_2|$. 

5
2.3 Bernoulli Trials

A sequence of independent identically distributed (i.i.d.) Bernoulli trials is a finite or infinite sequence of independent random variables $Z_1, Z_2, Z_3, \ldots$, such that

(i) For each $i$, $Z_i$ equals either 0 or 1;

(ii) For each $i$, the probability that $Z_i = 1$ is a constant $p_0$.

$p_0$ is called the success probability. The next lemma shows an important property of an infinite i.i.d. Bernoulli trials which will be useful in the sequent analysis. The proof is obvious, and therefore omitted.

**Lemma 2.3** Let $Z_k, k = 1, 2, \ldots$, be an infinite sequence of i.i.d. Bernoulli trials with success probability $p_0 > 0$. Denote $\{Z_k^{\omega}\}_{k=0}^{\infty}$ as a sample sequence. Then we can select a subsequence $\{Z_{k_m}^{\omega}\}_{m=0}^{\infty}$ of $\{Z_k^{\omega}\}_{k=0}^{\infty}$ with probability 1 such that $Z_{k_m}^{\omega} = 1$ for all $m$.

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 $V = \{1, 2, \ldots, n\}$. The objective of the network is to reach a consensus, and meanwhile to cooperatively solve the following optimization problem

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

(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 use the following assumption on the independence of $\mathcal{G}_k$.

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

**Remark 3.1** 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, different with existing works [25, 27, 26], 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 $\mathcal{G}([k_1, k_2]) = (\mathcal{V}, \bigcup_{t \in [k_1, k_2]} \mathcal{E}(t))$, where $0 \leq k_1 \leq k_2 \leq +\infty$. Then we have the following definition.

**Definition 3.1** (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 $k \geq 0$,

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

(ii) Assume that $\mathcal{G}_k$ is bidirectional for all $k \geq 0$. Then $\mathcal{G}_k$ is said to be stochastically jointly connected (SJC) if there exists a sequence $0 = k_0 < k_1 < \cdots < k_m < \cdots$ and a constant $0 < q < 1$ such that

$$\mathbb{P}\{\mathcal{G}_{[k_m, k_{m+1})} \text{ is connected}\} \geq q, \quad m = 0, \ldots.$$

3.1.2 Neighboring Information

The local information that each agent uses to update its state consists of two parts: the average and the projection parts. The average 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 \( X_i = \{ v \mid f_i(v) = \min_{z \in \mathbb{R}^d} f_i(z) \} \) 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)** \( X_0 = \bigcap_{i=1}^{n} X_i \) is nonempty.

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

**Remark 3.2** The average \( e_i(k) \) has been widely used in consensus algorithms, e.g., [6, 7, 8]. Assumption A2(i) indicates that \( e_i(k) \) is always within the convex hull of node \( i \)'s neighbors, i.e., \( \text{co}\{x_j(k), j \in \mathcal{N}_i(k)\} \), and, moreover, A2(ii) ensures that \( e_i(k) \) is in the relative interior of \( \text{co}\{x_j(k), j \in \mathcal{N}_i(k)\} \). [22]

**Remark 3.3** 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 \( \nabla |z|_K^2 = 2(P_K(z) - z) \). 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 \( h(z) = |z|_X_i^2 \) to compute \( \nabla h(x_i(k)) \) within this coordinate system. Then we know \( P_{X_i}(x_i(k)) = x_i(k) + t_0 \nabla h(x_i(k)) \) with \( t_0 = \inf_{t \geq 0} \{ t \mid (x_i(k) + t \nabla h(x_i(k))) \in X_i \} \).

### 3.1.3 Randomized Algorithm

We are now ready to introduce the randomized optimization algorithm. At each time step, each agent independently and randomly either takes an average among its time-varying neighbor set, or projects onto the optimal solution set of its own objective function:

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

where \( 0 < p < 1 \) is a given constant.

**Remark 3.4** One 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 and following its neighbors’ opinion \cite{28}. An interesting question is whether
the social network reaches a common opinion or not, and if the answer is yes, whether the
network could reach an optimal common opinion. Algorithm \cite{3} can also be interpreted as a
generalization of several of the multi-agent systems considered in the literature, as discussed in
Section 1.

Remark 3.5 Note that, a global coordinate system is not required. On one hand,
\begin{equation}
    x_i(k+1) - x_i(k) = \sum_{j \in N_i(k)} a_{ij}(k)(x_j(k) - x_i(k))
\end{equation}
when the averaging part is taken. On the other hand,
\begin{equation}
    x_i(k+1) - x_i(k) = P_{X_i}(x_i(k)) - x_i(k)
\end{equation}
when the projecting part is taken. Hence, only relative vectors from node \(i\) pointing to its
neighbors and solution set are needed.

Remark 3.6 The constrained consensus algorithm studied in \cite{37}, can be viewed as a determin-
istic special case of \cite{3}, in which each node alternate between averaging and projection. Note
that, we do not impose a double stochasticity assumption on the weights.

Under assumptions A3 and A4, it is obvious that \(X_0\) is the optimal solution set of cost
function \(F(z)\). Let \(x^0 = (x_1^T(0), \ldots, x_n^T(0))^T \in \mathbb{R}^{nd}\) be the initial condition. The considered
optimal consensus problem is defined as follows. See Fig. 1 for an illustration.

**Definition 3.2** (i) A global optimal set aggregation is achieved almost surely for \cite{3} if for all
\(x^0 \in \mathbb{R}^{nd}\), we have
\begin{equation}
    P\{ \lim_{k \to +\infty} |x_i(k)|_{X_0} = 0, i = 1, \ldots, n \} = 1.
\end{equation}
(ii) A global consensus is achieved almost surely for \cite{3} if for all \(x^0 \in \mathbb{R}^{nd}\), we have
\begin{equation}
    P\{ \lim_{k \to +\infty} |x_i(k) - x_j(k)| = 0, i, j = 1, \ldots, n \} = 1.
\end{equation}
(iii) A global optimal consensus is achieved almost surely for \cite{3} if both \cite{8} and \cite{9} hold.

3.2 Basic Properties

In this subsection, we establish two key lemmas on the algorithm \cite{3}.
Figure 1: The goal of the multi-agent network is to achieve a consensus in the optimal solution set $X_0$.

**Lemma 3.1** 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

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

Then,

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

The desired conclusion follows. 

**Lemma 3.2** Let $\{x(k) = (x_1^T(k), \ldots, x_n^T(k))^T\}_{k=0}^{\infty}$ be a sequence defined by (5). Then for any $k \geq 0$, 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 $l$ follows average update rule at time $k$, we have

$$|x_l(k+1)|_{X_0} = |P_{X_l}(x_l(k))|_{X_0} = |P_{X_l}(x_l(k)) - P_{X_0}(P_{X_l}(x_l(k)))|$$

$$\leq |P_{X_l}(x_l(k)) - P_{X_0}(x_l(k))|$$

$$\leq |x_l(k) - P_{X_0}(x_l(k))|$$

$$\leq \max_{i=1,\ldots,n} |x_i(k)|_{X_0}. \quad (10)$$

On the other hand, if node $l$ follows projection update rule at time $k$, according to Lemma 2.1, we have

$$|x_l(k+1)|_{X_0} = |\sum_{j \in N_l(k)} a_{lj}(k)x_j(k)|_{X_0}$$

$$\leq \sum_{j \in N_l(k)} a_{lj}(k)|x_j(k)|_{X_0}$$

$$\leq \max_{i=1,\ldots,n} |x_i(k)|_{X_0}. \quad (11)$$

Hence, the conclusion holds. □

Based on Lemma 3.2 we know that the following limit exists:

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

It is immediate that the global optimal set aggregation is achieved almost surely if and only if

$$P\{\xi = 0\} = 1.$$

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 Main Result on 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 4.1** System (5) achieves a global optimal consensus almost surely if $\mathcal{G}_k$ is SUC.
Remark 4.1 Following the definition of SJC, one may also define $G_k$ being stochastically jointly strongly connected (SJSC) if there exists a sequence $0 = k_0 < \cdots < k_m < \cdots$ and a constant $0 < q < 1$ such that $P(G_{[k_m,k_{m+1}]}$ is strongly connected) $\geq q$ for all $m$. However, consistent with the results in [8,18], it is not hard to construct simple counter examples to show that SJSC graph does not guarantee a global optimal consensus almost surely. In this sense, Theorem 4.1 gives a quite tight connectivity condition to ensure global optimal consensus with probability one.

In order to prove Theorem 4.1, 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

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

Let $A = \{\xi > 0\}$ and $M = \{\exists i_0 \text{ s.t. } \delta_{i_0} > 0\}$ 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 4.1 $P(A \cap M) = 0$ if $G_k$ is SUC.

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

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

Moreover, according to Lemma 3.2, $\forall \ell = 1, 2, \ldots, \exists T(\ell, \omega) > 0$ such that

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

For any $k_m \geq T$, node $i_0$ projects onto $X_i$ with probability $p$. Thus, Lemma 3.1 implies

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

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 $k_m + 2$, we have
\[
|x_{i_0}(k_m + 2)|_{X_0} = |x_{i_0}(k_m + 1)|_{X_0} \leq \sqrt{(\xi + \frac{1}{\ell})^2 - 4\delta^2_{i_0}}
\]
with probability at least $p$.

• If node $i_0$ chooses the average option at time $k_m + 2$, with (13), we can obtain from the weights rule and Lemma 2.1 that
\[
|x_{i_0}(k_m + 2)|_{X_0} = \sum_{j \in N(i_0(k_m+1))} a_{ij}(k_m + 1)x_j(k_m + 1)|_{X_0} \\
\leq a_{i_0i_0}(k_m + 1)|x_{i_0}(k_m + 1)|_{X_0} + (1 - a_{i_0i_0}(k_m + 1))\left(\xi + \frac{1}{\ell}\right) \\
\leq a_{i_0i_0}(k_m + 1)\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - a_{i_0i_0}(k_m + 1))\left(\xi + \frac{1}{\ell}\right)} \\
\leq \eta\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta)(\xi + \frac{1}{\ell})}
\]
with probability at least $p$.

Both (15) and (16) lead to
\[
P\{|x_{i_0}(k_m + 2)|_{X_0} \leq \eta\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta)(\xi + \frac{1}{\ell})}\} \geq p. \tag{17}
\]

Through similar analysis, we can also obtain that for $\tau = 1, 2, \ldots$,
\[
P\{|x_{i_0}(k_m + \tau)|_{X_0} \leq \eta^{\tau - 1}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{\tau - 1})(\xi + \frac{1}{\ell})}\} \geq p. \tag{18}
\]

Furthermore, since $G_k$ is SUSC, we have
\[
P\{G([k_m + 1, k_m + B]) \text{ is strongly connected} \} \geq q.
\]
which implies
\[
P\{\text{there exist } \hat{k}_1 \in [k_m + 1, k_m + B] \text{ and } i_1 \in V \text{ s.t. } (i_0, i_1) \in G_{\hat{k}_1} \} \geq q.
\]
Let $\hat{k}_1 = k_m + \varrho$ with $1 \leq \varrho \leq B$. Noting the fact that the probability that node $i_1$ chooses the averaging part is $p$ at time step $k_m + \varrho + 1$, then based on (18), we have
\[
P\{|x_{i_1}(k_m + \varrho + 1)|_{X_0} \leq a_{i_0i_1}(k_m + \varrho)|x_{i_0}(k_m + \varrho)|_{X_0} + (1 - a_{i_0i_1}(k_m + \varrho))\left(\xi + \frac{1}{\ell}\right)|_{X_0} \} \\
\geq P\{|x_{i_1}(k_m + \varrho + 1)|_{X_0} \leq \eta^{\varrho}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{\varrho})(\xi + \frac{1}{\ell})}|_{X_0} \} \\
\geq pq. \tag{19}
\]
where \( \mathcal{F}_0 = \{ |x_{i_0}(k_m + \theta)|x_0 \leq \eta^{\theta-1}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{\theta-1})(\xi + \frac{1}{\ell})} \} \). Therefore, we obtain that for any \( \tau = 1, 2, \ldots, \)

\[
P\{ |x_i(k_m + B + \tau)|x_0 \leq \eta^{B+\tau-1}\sqrt{\xi + \frac{1}{\ell} - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{B+\tau-1})(\xi + \frac{1}{\ell}), l = 0, 1} \geq p^2q.
\]

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 continuously on intervals \([k_m + 2B + 1, k_m + 3B], \ldots, [k_m + (n-2)B+1, k_m + (n-1)B], \ldots, i_3, \ldots, i_{n-1} \) can be found until \( \mathcal{V} = \{i_0, i_1, \ldots, i_{n-1}\} \). Then one can obtain that for any \( i \in \mathcal{V}, \)

\[
P\{ |x_i(k_m + (n-1)B + 1)|x_0 \leq \eta^{(n-1)B}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{(n-1)B})(\xi + \frac{1}{\ell}), i = 1, \ldots, n} \}
\]

\[
= P\{ \max_{i=1, \ldots, n} |x_i(k_m + (n-1)B + 1)|x_0 \leq \eta^{(n-1)B}\sqrt{\xi + \frac{1}{\ell} - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{(n-1)B})(\xi + \frac{1}{\ell})} \}
\]

\[
\geq p^n q^{n-1}.
\]

Since (20) holds for any \( k_m \geq T \) and \( p^n q^{n-1} \) is a constant, and noting the fact the analysis on different time instances \( \{k_m + (n-1)B + 1, k_m \geq T\} \) is independent for different \( m \), the events that

\[
\max_{i=1, \ldots, n} |x_i(k_m + (n-1)B + 1)|x_0 \leq \eta^{(n-1)B}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{(n-1)B})(\xi + \frac{1}{\ell})}
\]

can be viewed as an infinite sequence of i.i.d. Bernoulli trials with success probability \( p^n q^{n-1} \).

Then based on Lemma 2.3 we see that with probability 1, there is an infinite subsequence \( \{\tilde{k}_j, j = 1, 2, \ldots\} \) from \( \{k_m + (n-1)B + 1, k_m \geq T\} \) satisfying

\[
\max_{i=1, \ldots, n} |x_i(\tilde{k}_j)|x_0 \leq \eta^{(n-1)B}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{(n-1)B})(\xi + \frac{1}{\ell})}.
\]

This implies

\[
P[\mathcal{R}_\ell] = 1
\]

for all \( \ell = 1, 2, \ldots \), where \( \mathcal{R}_\ell = \{ \xi \leq \eta^{(n-1)B}\sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{(n-1)B})(\xi + \frac{1}{\ell})} \} \). As a result, we obtain

\[
P[\mathcal{R}_*] = 1,
\]

where \( \mathcal{R}_* = \lim_{\ell \to \infty} \mathcal{R}_\ell = \{ \xi \leq \eta^{(n-1)B}\sqrt{\xi^2 - \frac{1}{4}\delta^2_{i_0} + (1 - \eta^{(n-1)B})\xi} \}. \)
Finally, it is not hard to find that \( A \cap M \subseteq R_e \) because \( 0 < \eta^{(n-1)B} < 1 \). Then the conclusion holds straightforwardly.\( \square \)

Take a node \( \alpha_0 \in V \). Then define

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

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

**Lemma 4.2** We have

\[
    z_{\alpha_0}(k + 1) \leq z_{\alpha_0}(k) + \max_{i=1,\ldots,n} |x_i(k)|_{X_i}, \quad k = 0, 1, \ldots.
\]

**Proof.** For any node \( l = 1, \ldots, n \), if \( l \) chooses the average part at time \( k + 1 \), we know that

\[
    |x_l(k + 1)|_{X_{\alpha_0}} = \sum_{j \in N_l(k)} a_{lj}(k) |x_j(k)|_{X_{\alpha_0}} \leq \max_{i=1,\ldots,n} |x_i(k)|_{X_{\alpha_0}} = z_{\alpha_0}(k). \tag{23}
\]

Moreover, if \( l \) chooses the projection part at time \( k + 1 \), we have

\[
    |x_l(k + 1) - x_l(k)| = |x_l(k)|_{X_l},
\]

which yields

\[
    |x_l(k + 1)|_{X_{\alpha_0}} \leq |x_l(k)|_{X_{\alpha_0}} + |x_l(k)|_{X_l} \leq z_{\alpha_0}(k) + \max_{i=1,\ldots,n} |x_i(k)|_{X_i} \tag{24}
\]

according to the non-expansiveness property \( \text{[2]} \). Then the conclusion holds with (23) and (24).\( \square \)

We are now in a place to present the optimal set convergence part of Theorem 4.1 as stated in the following conclusion.

**Proposition 4.1** System \( \text{[5]} \) achieves a global optimal set aggregation almost surely if \( \mathcal{G}_k \) is SUC.

**Proof.** Note that, we have

\[
    P[A] = P[A \cap M] + P[A \cap M^c] \leq P[A \cap M] + P[A | M^c].
\]

Since the conclusion is equivalent to \( P(A) = 0 \), with Lemma 4.1 we only need to prove \( P[A | M^c] = 0 \).

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

\[
    k \geq T_1 \Rightarrow |x^\omega_i(k)|_{X_i} \leq \frac{1}{\ell}, \quad i = 1, \ldots, n. \tag{25}
\]
Take an arbitrary node $\alpha_0 \in \mathcal{V}$. Based on Lemma 4.2, we also have that for any $\{x^\omega(k)\}_{k=0}^\infty \in \mathcal{M}^c$ and $s \geq T_1$,
\[ z^\omega_{\alpha_0}(s + \tau) \leq z^\omega_{\alpha_0}(s) + \frac{\tau}{\ell}, \quad \tau = 0, 1, \ldots \] (26)

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

Step 1: Denote $k_1 = T_1$. Since $G_k$ is SUC, we have
\[ \mathbf{P}\{\text{there exist } \hat{k}_1 \in [k_1, k_1 + B - 1] \text{ and } \alpha_1 \in \mathcal{V} \text{ s.t. } (\alpha_0, \alpha_1) \in G_{\hat{k}_1}\} \geq q. \]
Let $\hat{k}_1 = k_1 + g$, $0 \leq g \leq B - 1$. Then we obtain from the definition of (5) that
\[ \mathbf{P}\{|x_{\alpha_1}(k_1 + g + 1)|x_{\alpha_0} \leq a_{\alpha_1\alpha_0}(k_1 + g)|x_{\alpha_0}(k_1 + g)|x_{\alpha_0} + (1 - a_{\alpha_1\alpha_0})z_{\alpha_0}(k_1 + g)\} \geq pq. \] (27)
Thus, based on the weights rule $A_1$ and (25), (27) leads to
\[ \mathbf{P}\{|x_{\alpha_1}(k_1 + g + 1)|x_{\alpha_0} \leq \eta \cdot \frac{1}{\ell} + (1 - \eta)(z_{\alpha_0}(k_1) + g \cdot \frac{1}{\ell})\} \mathcal{M}^c \geq pq. \] (28)

Next, there will be two cases.

- If node $\alpha_1$ chooses the projection option at time $k_1 + g + 2$, we have
  \[ |x_{\alpha_1}(k_1 + g + 2)|x_{\alpha_0} \leq \eta |x_{\alpha_1}(k_1 + g + 1)|x_{\alpha_0} + \frac{1}{\ell} \]
  \[ \leq \eta \cdot \frac{1}{\ell} + (1 - \eta)(z_{\alpha_0}(k_1) + g \cdot \frac{1}{\ell}) + \frac{1}{\ell} \] (29)
  with probability at least $pq$ conditioned $\mathcal{M}^c$.

- If node $\alpha_1$ chooses the average option at time $k_1 + g + 2$, we have
  \[ |x_{\alpha_1}(k_1 + g + 2)|x_{\alpha_0} \leq \eta |x_{\alpha_1}(k_1 + g + 1)|x_{\alpha_0} + (1 - \eta)z_{\alpha_0}(k_1 + g + 1) \]
  \[ \leq \eta \eta \cdot \frac{1}{\ell} + (1 - \eta)(z_{\alpha_0}(k_1) + g \cdot \frac{1}{\ell}) + (1 - \eta)(z_{\alpha_0}(k_1) + (g + 1) \cdot \frac{1}{\ell}) \]
  \[ \leq \eta^2 \cdot \frac{1}{\ell} + (1 - \eta^2)(z_{\alpha_0}(k_1) + (g + 1) \cdot \frac{1}{\ell}) \] (30)
  with probability at least $pq$ conditioned $\mathcal{M}^c$.

With (29) and (30), we obtain
\[ \mathbf{P}\{|x_{\alpha_1}(k_1 + g + 2)|x_{\alpha_0} \leq \eta^2 \cdot \frac{1}{\ell} + (1 - \eta^2)(z_{\alpha_0}(k_1) + (g + 1) \cdot \frac{1}{\ell}) + \frac{1}{\ell}\} \mathcal{M}^c \geq pq. \] (31)

Then similar analysis yields that for any $\tau = 1, 2, \ldots$
\[ \mathbf{P}\{|x_{\alpha_1}(k_1 + g + \tau)|x_{\alpha_0} \leq \eta^\tau \cdot \frac{1}{\ell} + (1 - \eta^\tau)(z_{\alpha_0}(k_1) + (g + \tau - 1) \cdot \frac{1}{\ell}) + \sum_{l=1}^{\tau-1} \eta^{l-1} \cdot \frac{1}{\ell}\} \mathcal{M}^c \geq pq. \]
Then we can define a random variable \( k \).

**Step 3:** Let \( \alpha \), which implies

Furthermore, since \( 0 \leq q \leq B - 1 \) and based on (25), it turns out that for any \( \hat{\tau} = 0, 1, \ldots \), we have

\[
P \{ |x_{\alpha_i}(k_1 + B + \hat{\tau})|_{X_{\alpha_0}} \leq \eta^{B+\hat{\tau}} \cdot \frac{1}{\ell} + (1 - \eta^{B+\hat{\tau}})(z_{\alpha_0}(k_1) + (B + \hat{\tau} - 1) \cdot \frac{1}{\ell}) + \frac{1}{1 - \eta} \cdot \frac{1}{\ell},
\]

\[
l = 0, 1 | \mathcal{M}^c \} \geq pq.
\]

**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 \( \alpha_2 \) in \( \mathcal{G}([k_1 + B, k_m + 2B - 1]) \) with probability \( q \). Similarly we can obtain that for any \( \hat{\tau} = 0, 1, \ldots \),

\[
P \{ |x_{\alpha_i}(k_1 + 2B + \hat{\tau})|_{X_{\alpha_0}} \leq \eta^{2B+\hat{\tau}} \cdot \frac{1}{\ell} + (1 - \eta^{2B+\hat{\tau}})(z_{\alpha_0}(k_1) + (2B + \hat{\tau} - 1) \cdot \frac{1}{\ell}) + \frac{2}{1 - \eta} \cdot \frac{1}{\ell},
\]

\[
l = 0, 1, 2 | \mathcal{M}^c \} \geq p^2 q^2.
\]

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 \( V = \{\alpha_0, \alpha_1, \ldots, \alpha_{n-1}\} \). Then one can obtain that

\[
P \{ |x_i(k_1 + (n - 1)B)|_{X_{\alpha_0}} \leq (1 - \eta^{(n-1)B})z_{\alpha_0}(k_1) + L \cdot \frac{1}{\ell}, i = 1, \ldots, n | \mathcal{M}^c \} \geq p^{n-1}q^{n-1},
\]

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

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

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

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

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

\[
\chi = \begin{cases} 
1, & \text{with probability } 1 - \hat{p} \\
\theta_0, & \text{with probability } \hat{p}
\end{cases} \quad (32)
\]

As a result, with (26) and (32), we conclude that for any \( m = 1, 2, \ldots \),

\[
P \{ z_{\alpha_0}(k_m + 1) \leq \chi \cdot z_{\alpha_0}(k_m) + L \cdot \frac{1}{\ell} | \mathcal{M}^c \} = 1,
\]

which implies

\[
E(z_{\alpha_0}(k_m + 1) | \mathcal{M}^c) \leq (1 - (1 - \theta_0)\hat{p})E(z_{\alpha_0}(k_m) | \mathcal{M}^c) + L \cdot \frac{1}{\ell}.
\]
Therefore, we can further obtain
\[ \limsup_{m \to \infty} E(z_{\alpha_0}(k_m)|\mathcal{M}^c) \leq \frac{L}{(1 - \theta_0)p} \cdot \frac{1}{\ell}. \] (33)

Since \( \ell \) can be any positive integer in (33) and \( z_{\alpha_0}(k_m) \) is nonnegative for any \( m \), we have
\[ \lim_{m \to \infty} E(z_{\alpha_0}(k_m)|\mathcal{M}^c) = 0. \] (34)

Based on Fatou’s lemma, we know
\[ 0 \leq E(\lim_{m \to \infty} z_{\alpha_0}(k_m)|\mathcal{M}^c) \leq \lim_{m \to \infty} E(z_{\alpha_0}(k_m)|\mathcal{M}^c) = 0, \] (35)
which yields
\[ P\{ \lim_{m \to \infty} z_{\alpha_0}(k_m) = 0|\mathcal{M}^c \} = 1. \] (36)

Finally, because \( \alpha_0 \) is chosen arbitrarily over the network in (36), we see that
\[ P[A|\mathcal{M}^c] = 0. \] (37)

The proof is completed. \( \square \)

### 4.2 Consensus Analysis

In this subsection, we present the consensus analysis of the proof of Theorem 4.1. Let \( x_{i,j}(k) \) represent the \( j \)'th coordinate of \( x_i(k) \). Denote
\[ h(k) = \min_{i=1,...,n} x_{i,j}(k), \quad H(k) = \max_{i=1,...,n} x_{i,j}(k). \]

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

**Proposition 4.2** System (3) achieves a global consensus almost surely if \( G_k \) is SUSC.

**Proof.** From Proposition 4.1 it follows that we only need to prove
\[ P\{ \lim_{k \to \infty} S(k) = 0|\mathcal{M}^c \} = 1 \]
since \( P[\mathcal{M}^c] \geq P[A^c] = 1 \).

Let \( \{x_{i}(k)\}_{k=0}^{\infty} \) be a sample sequence in \( \mathcal{M}^c \). Then \( \forall \ell = 1, 2, \ldots \), \( \exists T_1(\ell, \omega) > 0 \) such that
\[ k \geq T_1 \Rightarrow |x_{i}(k)|_{X_i} \leq \frac{1}{\ell}, \quad i = 1, \ldots, n. \] (38)
Moreover, based on similar analysis as in the proof of Lemma 4.2 we see that

\[ h(k + s) \geq h(k) - s \cdot \frac{1}{\ell}; \quad H(k + s) \leq H(k) + s \cdot \frac{1}{\ell} \tag{39} \]

for all \( k \geq T_1 \) and \( s \geq 0 \).

Denote \( k_1 = T_1 \). Take \( \nu_0 = \mathcal{V} \) with \( x_{\nu_0,|j|}(k_1) = h(k_1) \). Then we can obtain from the definition of (5) that

\[
x_{\nu_0,|j|}(k_1 + 1) \leq \begin{cases} x_{\nu_0,|j|}(k_1) + \frac{1}{\ell}, & \text{if projection happens} \\ a_{\nu_0|0}(k_1)x_{\nu_0,|j|}(k_1) + (1 - a_{\nu_1|\nu_0}(k_1))H(k_1), & \text{if averaging happens} \end{cases} \tag{40} \]

which leads to that almost surely we have

\[ x_{\nu_0,|j|}(k_1 + 1) \leq \eta h(k_1) + (1 - \eta)H(k_1) + \frac{1}{\ell}. \]

Continuing the estimates we know that almost surely for any \( \tau = 0, 1, \ldots, \)

\[ x_{\nu_0,|j|}(k_1 + \tau) \leq \eta^\tau h(k_1) + (1 - \eta^\tau)H(k_1) + \frac{\tau(\tau + 1)}{2} \cdot \frac{1}{\ell}. \tag{41} \]

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

\[ \mathbb{P}\{ \exists k_1 \in [k_1, k_1 + B - 1] \text{ and } \exists \nu_1 \in \mathcal{V} \text{s.t. } (\nu_0, \nu_1) \in \mathcal{G}_{k_1} \} \geq q. \]

Let \( \hat{k}_1 = k_1 + \rho, 0 \leq \rho \leq B - 1 \). Similarly with (27), we see from (41) that

\[ \mathbb{P}\{ x_{\nu_1,|j|}(k_1 + \rho + 1) \leq \eta^{\rho+1} h(k_1) + (1 - \eta^{\rho+1})H(k_1) + \eta \cdot \frac{\rho(\rho + 1)}{2} \cdot \frac{1}{\ell} \mid \mathcal{M}^c \} \geq pq. \tag{42} \]

Similar analysis will lead to

\[ \mathbb{P}\{ x_{\nu_1,|j|}(k_1 + \rho + \hat{\tau}) \leq \eta^{\rho+\hat{\tau}} h(k_1) + (1 - \eta^{\rho+\hat{\tau}})H(k_1) + \frac{(\rho + \hat{\tau})(\rho + \hat{\tau} + 1)}{2} \cdot \frac{1}{\ell} \mid \mathcal{M}^c \} \geq pq, \tag{43} \]

with \( \hat{\tau} = 1, 2, \ldots, \), which yields

\[ \mathbb{P}\{ x_{\nu_1,|j|}(k_1 + B + \tau) \leq \eta^{B+\tau} h(k_1) + (1 - \eta^{B+\tau})H(k_1) + \frac{(B + \tau)(B + \tau + 1)}{2} \cdot \frac{1}{\ell} \mid \mathcal{M}^c \} \geq pq, \tag{44} \]

where \( \tau = 0, 1, \ldots, \).

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

\[ \mathbb{P}\{ x_{\nu_l,|j|}(k_1 + (n - 1)B) \leq \eta^{(n-1)B} h(k_1) + (1 - \eta^{(n-1)B})H(k_1) + \frac{(n - 1)B((n - 1)B + 1)}{2} \cdot \frac{1}{\ell}, \]

\[ l = 0, 1, \ldots, n - 1 \mid \mathcal{M}^c \} \geq p^{n-1}q^{n-1}. \tag{45} \]
Therefore, denoting $k_2 = k_1 + (n-1)B$, we have

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

Furthermore, with (39), we can further obtain

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

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

Then we know $\mathbb{P}\{\lim_{k \to \infty} S(k) = 0|\mathcal{M}^c\} = 1$ by similar analysis as the proof of Proposition 4.1. The proof is completed. □

Theorem 4.1 immediately follows from Propositions 4.1 and 4.2.

5 Main Result on Bidirectional Graphs

In this section, we discuss the randomized optimal consensus problem under more restrictive communication assumptions, that is, bidirectional communications. In this case, node $i$ can receive the neighbor information from node $j$ if and only if node $j$ can also receive neighbor information from $i$.

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 5.1** Suppose $G_k$ is bidirectional for all $k \geq 0$ and A5 holds. System (5) achieves a global optimal consensus almost surely if $G_k$ is SJC.

**Remark 5.1** Note that, although we assume that $G_k, k \geq 0$ is bidirectional, the weight of arc $(i,j)$ may not be equal to that of arc $(j,i)$. In other words, we do not need the weight functions $a_{ij}(k)$ to be symmetric.

**Remark 5.2** Theorem 5.1 is based on SJC graphs, which do not require an upper bound for the length of intervals where the joint graphs are taken. This essential difference with the SUSC graphs leads to that the analysis on directed graphs cannot be used directly in the bidirectional
case, though they share some common properties. To analyze set aggregation, we will use a new idea by investigating the convergence of a serial of subsets of nodes along the time sequence when each subset has at least one neighbor, and another new idea by studying a constructed invariant set to analyze consensus.

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 5.1.

5.1 Set Convergence

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

**Lemma 5.1** Assume that $G_k$ is bidirectional for all $k \geq 0$. Then $P[A \cap M] = 0$ if $G_k$ is SJC.

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

$$|x^\omega_{i_0}(k_m)|_{X_{i_0}} \geq \frac{1}{2} \delta_{i_0}(\omega) \geq 0. \quad (45)$$

Similarly, based on Lemma 3.2, $\forall \ell = 1, 2, \ldots, \exists T(\ell, \omega) > 0$ such that

$$k \geq T \Rightarrow 0 \leq |x^\omega_1(k)|_{X_0} \leq \xi(\omega) + \frac{1}{\ell}, \quad i = 1, \ldots, n. \quad (46)$$

For any $k_m \geq T$, based on the definition of (5), note $i_0$ will choose projecting onto $X_i$ with probability $p$. Thus, we know that Lemma 3.1 implies

$$P\{|x_{i_0}(k_m + 1)|_{X_0} \leq \sqrt{(\xi + \frac{1}{\ell})^2 - \frac{1}{4} \delta_{i_0}^2}\} \geq p. \quad (47)$$

Next, we define

$$\hat{k}_1 \doteq \inf_{k \geq k_m + 1 \{i_0 \text{ has at least one neighbor other than itself in } G_k\}}$$

and

$$\mathcal{V}_1 \doteq \{v | v \text{ is a neighbor of } i_0 \text{ in } G_{\hat{k}_1}\}.$$ 

Since $G_k$ is SJC, we see that the probability of $\hat{k}_1$ being finite is at least $q$. Based on Lemma 3.1, it is obvious to see that

$$|x^\omega_{i_0}(s)|_{X_0} \leq |x^\omega_{i_0}(k_m + 1)|_{X_0} \quad (48)$$
for any \( k_m + 1 \leq s \leq \hat{k}_1 \). Moreover, the probability that all the nodes in \( V_1 \) choose averaging is \( p^{\left| V_1 \right| - 1} \). Therefore, we have

\[
P\{|x_i(\hat{k}_1 + 1)|_{X_0} \leq \eta \sqrt{\left( \xi + \frac{1}{\ell} \right)^2 - \frac{1}{4} \delta_{i_0}^2 + (1 - \eta)(\xi + \frac{1}{\ell})}, i \in V_1 \} \geq p^{\left| V_1 \right| - 1}q. \tag{49}
\]

Furthermore, we can similarly define

\[
\hat{k}_2 \doteq \inf_{k \geq k_{1}+1} \{ \text{there is at least one link between } V_1 \text{ and } V \setminus V_1 \text{ in } G_k \}
\]
and \( V_2 = \{ v | v \in V \setminus V_1 \text{ has a neighbor in } V_1 \text{ in } G_{\hat{k}_2} \} \). Moreover, we can also obtain

\[
P\{|x_i(\hat{k}_2 + 1)|_{X_0} \leq \eta^2 \sqrt{\left( \xi + \frac{1}{\ell} \right)^2 - \frac{1}{4} \delta_{i_0}^2 + (1 - \eta^2)(\xi + \frac{1}{\ell})}, i \in V_1 \cup V_2 \} \geq p^{\left| V_1 \right| + 1| V_2 |}q^2. \tag{50}
\]

We can repeat the upper process, \( V_3, \ldots, V_{d_0} \) can be defined for some constant \( 1 \leq d_0 \leq n - 1 \) until \( V = \bigcup_{j=1}^{d_0} V_j \). Denote \( \varsigma_m = \hat{k}_{d_0} + 1 \) as the time instance associated with \( k_m \) in the upper analysis. Then we have

\[
P\{|x_i(\varsigma_m)|_{X_0} \leq \eta \sqrt{\left( \xi + \frac{1}{\ell} \right)^2 - \frac{1}{4} \delta_{i_0}^2 + (1 - \eta)(\xi + \frac{1}{\ell})}, i \in V \}
= p^{\max_{i=1,\ldots,n} |x_i(\varsigma_m)|_{X_0} \leq \eta \sqrt{\left( \xi + \frac{1}{\ell} \right)^2 - \frac{1}{4} \delta_{i_0}^2 + (1 - \eta)(\xi + \frac{1}{\ell})}, i \in V \}
\geq p^{n q^{d_0}}. \tag{51}
\]

As a result, based on Lemma 2.3 with probability 1, we can select an infinite subsequence \( \varsigma_m \) from \( \varsigma_m \) satisfying that

\[
\max_{i=1,\ldots,n} |x_i(\varsigma_m)|_{X_0} \leq \eta^{-1} \sqrt{\left( \xi + \frac{1}{\ell} \right)^2 - \frac{1}{4} \delta_{i_0}^2 + (1 - \eta^{-1})(\xi + \frac{1}{\ell})}, s = 1, \ldots.
\]

This will also lead to

\[
P[\tilde{R}_*] = 1, \tag{52}
\]

where \( \tilde{R}_* = \{ \xi \leq \eta^{-1} \sqrt{\xi^2 - \frac{1}{4} \delta_{i_0}^2 + (1 - \eta^{-1})\xi} \} \). Noting the fact that \( A \cap M \subseteq \tilde{R}_* \), the conclusion holds. \( \square \)

Next, we define

\[
y_i = \liminf_{k \to \infty} |x_i(k)|_{X_0}, i = 1, \ldots, n
\]
and denote \( D = \{ \exists i_0 \text{ s.t. } y_{i_0} < \xi \} \). We give another lemma in the following.

**Lemma 5.2** Assume that \( G_k \) is bidirectional for all \( k \geq 0 \). Then \( P[A \cap D] = 0 \) if \( G_k \) is SJC.
Proof. The proof will follow the same idea as the proof of Lemma 5.1. Let \( \{x^\omega(k)\}_{k=0}^\infty \) be a sample sequence. Then there exists a time sequence \( k_1 < \cdots < k_m < \cdots \) with \( \lim_{m \to \infty} k_m = \infty \) such that
\[
|x^\omega_0(k_m)|x_0 \leq \frac{1}{2}(y_{i_0}\omega + \xi(\omega)).
\] (53)
Moreover, \( \forall \ell = 1, 2, \ldots, \exists T(\ell, \omega) > 0 \) such that
\[
k \geq T \Rightarrow 0 \leq |x^\omega_0(k)|x_0 \leq \xi(\omega) + \frac{1}{\ell}, \; i = 1, \ldots, n.
\] (54)
Next, we define
\[
\hat{k}_1 = \inf_{k \geq k_m} \{i_0 \text{ has at least one neighbor other than itself in } G_k\}.
\]
Based on Lemma 3.1, we see that
\[
|x^\omega_0(s)|x_0 \leq |x^\omega_0(k_m)|x_0 \leq \frac{1}{2}(y_{i_0}\omega + \xi(\omega))
\] (55)
for any \( k_m \leq s \leq \hat{k}_1 \). Therefore, defining \( V_1 = \{v|v \text{ is a neighbor of } i_0 \text{ in } G_{\hat{k}_1}\} \) and by similar analysis as we obtain (49), we have
\[
P\{|x_i(\hat{k}_1 + 1)|x_0 \leq \eta y_{i_0} + (1 - \eta y_{i_0}/\xi_0), i \in V_1\} \geq p|V_1|q.
\] (56)
Continuing the upper process, we will also reach
\[
P\{|x_i(\varsigma_m)|x_0 \leq \eta y_{i_0} + (1 - \eta y_{i_0}/\xi_0), i \in V\}
= P\{\max_{i=1,\ldots,n} |x_i(\varsigma_m)|x_0 \leq \eta y_{i_0} + (1 - \eta y_{i_0}/\xi_0), i \in V\}
\geq p^n q^{d_0},
\] (57)
where \( 1 \leq d_0 \leq n - 1 \) and \( \varsigma_m \) still denotes \( \hat{k}_{d_0} + 1 \). Then, defining an event
\[
W = \{\xi \leq \eta y_{i_0} + (1 - \eta y_{i_0}/\xi_0) \cdot \xi\},
\]
we can similarly obtain \( P[W] = 1 \) with (57) based on the same analysis as the proof of Lemma 5.1. Finally, the fact that \( A \cap D \subseteq W^c \) implies the conclusion immediately. \( \square \)

Note that, if A5 holds, according to Lemma 3.2, for any initial condition \( x^0 \), we have
\[
x_i(k) \in X^*_0, \; i = 1, \ldots, n; \; k = 0, 1, \ldots,
\]
where $X_0^* \doteq \{ v | v | x_0 \leq d_* \}$ with $d_* = \max_{i=1,...,n} |x_i(0)|_{X_0}$. 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

$$|x_i(k) - x_j(k)| \leq b_1; \quad |x_i(k)|_{X_0} \leq b_2$$ (58)

for all $i, j$ and $k$.

Now we are ready to prove the optimal set convergence part of Theorem 4.1, which is stated in the following conclusion.

**Proposition 5.1** Assume $G_k$ is bidirectional for all $k \geq 0$ and $A5$ holds. System $\{5\}$ achieves a global optimal set aggregation almost surely if $G_k$ is SJC.

**Proof.** With Lemmas 5.1 and 5.2 we only need to show

$$P[\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c] = 0.$$

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

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

and

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

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

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

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

$$k \geq T \Rightarrow 0 < \xi(\omega) \leq |x^\omega_i(k)|_{X_0} \leq \xi(\omega) + \frac{1}{\ell} \text{ and } |x^\omega_i(k)|_{X_i} \leq \frac{1}{\ell}, \ i = 1, \ldots, n.$$ (59)

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

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

Let $x^\omega_i(k_m)$ and $P_{X_0}(x^\omega_i(k_m))$ be fixed. Then we can associate a unique point $x^\omega_i$ to $x^\omega_j(k_m)$ in the way that $x^\omega_i$ satisfies $(P_{X_0}(x^\omega_i(k_m)) - x^\omega_i, x^\omega_i(k_m)) - x^\omega_j(k_m) = 0$ if the three points $x^\omega_i(k_m)$,
Figure 2: Finding the point $x^*_\omega$ in the proof of Prop. 5.1.

$P_{X_0}(x^\omega_{i_0}(k_m))$ and $x^\omega_j(k_m)$ form a triangle; and $x^*_\omega = P_{X_0}(x^\omega_{i_0}(k_m))$ otherwise. Moreover, it is not hard to find that there exists a unique scalar $0 < \gamma < 1$ such that $x^*_\omega = \gamma x^\omega_{i_0}(k_m) + (1 - \gamma)x^\omega_j(k_m)$.

Note that, the upper process defines a continuous function $(x^\omega_{i_0}(k_m), P_{X_0}(x^\omega_{i_0}(k_m)), x^\omega_j(k_m)) \mapsto \gamma$. With (58), we have $(x^\omega_{i_0}(k_m), P_{X_0}(x^\omega_{i_0}(k_m)), x^\omega_j(k_m))$ always locates within a compact set

$$\{0 \leq |x^\omega_{i_0}(k_m)|_{X_0} \leq d_*; \vartheta^\omega \leq |x^\omega_j(k_m)| \leq b_1; \xi(\omega) \leq |x^\omega_j(k_m) - P_{X_0}(x^\omega_{i_0}(k_m))| \leq b_1 + b_2\}.$$  

Therefore, there exist two constants $0 < \gamma_\ast \leq \gamma^* < 1$ (by a constant, we mean it does not depend on $k_m$) such that $\gamma_\ast \leq \gamma \leq \gamma^*$ (see Fig. 2).

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

$$|x^\omega_*|_{X_0} \leq |x^\omega_{i_0} - P_{X_0}(x^\omega_{i_0}(k_m))| \leq \sin \beta_0(\xi(\omega) + \frac{1}{\ell}) \leq b_*(\xi(\omega) + \frac{1}{\ell}), \quad (60)$$  

where $\beta_0 = \phi(x^\omega_j(k_m) - x^\omega_{i_0}(k_m), P_{X_0}(x^\omega_{i_0}(k_m)) - x^\omega_{i_0}(k_m))$ and $0 < b_* = \sqrt{1 - (\frac{\vartheta}{b_1})^2} < 1$. Therefore, with (60), repeating the deduction used in the proofs of Lemmas 5.1 and 5.2, the claim can then be proved.

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

$$\lim_{k \to \infty} |x_j(k)|_{W_{i_0}(k)} = 0; \quad (61)$$

25
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 $\mathcal{I}$, we have

$$\mathbb{P}\left\{ \lim_{k \to \infty} |P_{W_{i_0}}(x_j(k)) - P_{X_0}(x_j(k))| = 0, j = 1, \ldots, n | \mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c \right\} = 1. \quad (62)$$

Denote $T_{i_0}(k) = \text{co}\{P_{X_{i_0}}(x_{i_0}(k)), P_{X_0}(x_1(k)), \ldots, P_{X_0}(x_n(k))\}$. Then $T_{i_0}(k) \subseteq X_{i_0}$, $\forall k \geq 0$. $T_{i_0}(k)$ can then be defined for $m = 1, \ldots, n - 1$ in the same way. Therefore, with (61), 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 $v_* \in \bigcap_{m=0}^{n-1} T_{i_0}(k) \subseteq X_0$ for sufficiently large $k$ such that $v_* \in M^0(k)$ (see Fig. 3), i.e.,

$$\mathbb{P}\{ \exists k \text{ s.t. } (x_{i_0}(k) - P_{X_0}(x_{i_0}(k)), v_* - P_{X_0}(x_{i_0}(k))) > 0 | \mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c \right\} = 1.$$

This implies $\mathbb{P}[\mathcal{A} \cap \mathcal{M}^c \cap \mathcal{D}^c] = 0$ because $\mathbb{P}\{(y - P_{X_0}(y), v_* - P_{X_0}(y)) > 0\} = 0$ for any $y \in \mathbb{R}^d$ and $v_* \in X_0$ according to $\mathcal{I}$. The proof is completed. \hfill \Box

### 5.2 Consensus Analysis

This subsection focuses on the consensus analysis of Theorem 5.1.

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 $i_1, \ldots, i_k \in \{1, \ldots, n\}, k = 1, 2, \ldots$ in the following way

$$P_{i_k i_{k-1} \ldots i_1}(y) = P_{X_{i_k}} P_{X_{i_{k-1}}} \ldots P_{X_{i_1}}(y)$$

with $P_0(y) = y$ as the case for $k = 0$. Let

$$\Gamma = \{ P_{i_k i_{k-1} \ldots i_1} | i_1, \ldots, i_k \in \{1, \ldots, N\}, k = 0, 1, 2, \ldots \}$$
be the set which contains all the multi-projection functions.

Furthermore, denote \( Y_k = \text{co}\{x_1(k), \ldots, x_n(k)\} \) be the convex hull of all the nodes’s state at step \( k \), and define \( \Delta Y_k \) by
\[
\Delta Y_k = \text{co}\{P(y)\mid y \in Y_k, P \in \Gamma\}.
\]

It is not hard to find that \( \Delta Y_k \) 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 \( \Delta Y_k \).

**Lemma 5.3** 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 2.2, any \( y \in \Delta Y_K \) has the following form
\[
y = \sum_{i=1}^{d+1} \lambda_i P^{(i)}(z_i),
\]
where \( \sum_{i=1}^{d+1} \lambda_i = 1 \) with \( \lambda_i \geq 0 \), \( P^{(i)} \in \Gamma \) 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
\[
| \sum_{i=1}^{d+1} \lambda_i P^{(i)}(z_i) - \sum_{i=1}^{d+1} \lambda_i z_i | \leq \sum_{i=1}^{d+1} \lambda_i |z_i - P_{X_0}(z_i)| + \sum_{i=1}^{d+1} \lambda_i |P_{X_0}(z_i) - P^{(i)}(z_i)| \leq 2 \max_{z \in K} |z|_{X_0},
\]
which implies the conclusion because \( \sum_{i=1}^{d+1} \lambda_i z_i \in Y_k \). \( \square \)

We can now present the consensus analysis.

**Proposition 5.2** Assume that \( G_k \) is bidirectional for all \( k \geq 0 \) and A5 holds. System (5) achieves a global consensus almost surely if \( G_k \) is SJC.

**Proof.** With Proposition 5.1, we only need to prove 
\[
P\{ \lim_{k \to \infty} S(k) = 0 \mid \mathcal{A}^c \} = 1.
\]

Let \( \{x_{\ell}^\omega(k)\}_{k=0}^{\infty} \) be a sample sequence in \( \mathcal{A}^c \). Then \( \forall \ell = 1, 2, \ldots, \exists T_1(\ell, \omega) > 0 \) such that
\[
k \geq T_1 \Rightarrow |x_{\ell}^\omega(k)|_{X_0} \leq \frac{1}{\ell}, \quad i = 1, \ldots, n.
\] (63)

Moreover, based on Lemma 5.3, we see that
\[
h(k + s) \geq h(k) - \frac{2}{\ell}; \quad H(k + s) \leq H(k) + \frac{2}{\ell}
\] (64)
for all $k \geq T_1$ and $s \geq 0$.

Denote $k_1 = T_1$. Take $\nu_0 \in \mathcal{V}$ with $x_{\nu_0,j}(k_1) = h(k_1)$. Define

$$\hat{k}_1 = \inf_{k \geq k_1} \{ \nu_0 \text{ has at least one neighbor other than itself in } \mathcal{G}_k \}$$

and

$$\mathcal{V}_1 = \{ v | v \text{ is a neighbor of } \nu_0 \text{ in } \hat{\mathcal{G}}_{k_1} \}.$$  

Since $\mathcal{G}_k$ is SJC, with probability of at least $q$, $\hat{k}_1$ is finite, and therefore $\mathcal{V}_1$ is well defined.

With (63), we have

$$x_{\nu_0,j}(\hat{k}_1) \leq x_{\nu_0,j}(k_1) + \frac{1}{\ell} = h(k_1) + \frac{1}{\ell}. \quad (65)$$

Thus, together with (64) and by similar analysis we used to obtain (42), we have

$$\mathbb{P}\{ x_{\nu_1,j}(\hat{k}_1 + 1) \leq \eta h(k_1) + (1 - \eta)H(k_1) + \frac{2}{\ell} | \mathcal{A}^c \} \geq pq \quad (66)$$

for any $\nu_1 \in \mathcal{V}_1$, which leads to

$$\mathbb{P}\{ x_{i,j}(\hat{k}_1 + 1) \leq \eta h(k_1) + (1 - \eta)H(k_1) + \frac{2}{\ell}, \ i \in \mathcal{V}_1 \ | \mathcal{A}^c \} \geq p|\mathcal{V}_1|q. \quad (67)$$

Similar with the proof of Lemma 5.1, we can repeat the upper process, and $\mathcal{V}_2, \ldots, \mathcal{V}_{d_0}$ can be defined for some constant $1 \leq d_0 \leq n - 1$ until $\mathcal{V} = \bigcup_{j=1}^{d_0} \mathcal{V}_j$. Moreover, we can also obtain that

$$\mathbb{P}\{ x_{i,j}(\hat{k}_{d_0} + 1) \leq \eta^{d_0} h(k_1) + (1 - \eta^{d_0})H(k_1) + \frac{2d_0}{\ell}, \ i \in \mathcal{V} \ | \mathcal{A}^c \} \geq p^n q^{d_0}. \quad (68)$$

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

$$\mathbb{P}\{ H(k_2) \leq \eta^{d_0} h(k_1) + (1 - \eta^{d_0})H(k_1) + \frac{2d_0}{\ell} | \mathcal{A}^c \} \geq p^n q^{d_0}. $$

Furthermore, with (64), we can further obtain

$$\mathbb{P}\{ S(k_2) \leq (1 - \eta^{d_0})S(k_1) + L_0 \cdot \frac{2(d_0 + 1)}{\ell} | \mathcal{A}^c \} \geq p^n q^{d_0}.$$ 

Then we know $\mathbb{P}\{ \lim_{k \to \infty} S(k) = 0 | \mathcal{A}^c \} = 1$ by similar deduction as the proof of Prop. 4.1. The proof is completed.  

Then we see that Theorem 5.1 follows from Propositions 5.1 and 5.2.
6 Conclusion

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.

References


