Maximal Dissent: A State-Dependent Way to Agree in Distributed Convex Optimization

Ashwin Verma, Graduate Student Member, IEEE, Marcos M. Vasconcelos, Urbashi Mitra, Fellow, IEEE, and Behrouz Touri, Senior Member, IEEE

Abstract—Consider a set of agents collaboratively solving a distributed convex optimization problem asynchronously under stringent communication constraints. When an agent becomes active, it is allowed to communicate with only one of its neighbors. In this article, we propose new state-dependent gossip algorithms where the agents with maximal dissent average their estimates. We prove the almost sure convergence of max-dissent subgradient methods using a unified framework applicable to other state-dependent distributed optimization algorithms. Furthermore, our proof technique bypasses the need to establish the information flow between any two agents within a time interval of uniform length by intelligently studying the convergence properties of the Lyapunov function used in our analysis.

Index Terms—Consensus algorithm, distributed algorithms, optimization.

I. INTRODUCTION

In distributed convex optimization, a collection of agents collaborates to minimize the sum of local objective functions by exchanging information over a communication network. The primary goal is to design algorithms that converge to an optimal solution via local interactions dictated by the underlying communication network. A standard strategy to solving distributed optimization problems consists of each agent first combining the local estimates shared by its neighbors followed by a first-order subgradient method on its local objective function [1], [2], [3]. Of particular relevance herein are the so-called gossip algorithms [4], where the information mixing step consists of averaging the states of two agents connected by one of the edges selected from the network graph.

Two benefits of gossip algorithms are their simple asynchronous implementation and a reduction in communication costs. The simplest form of a gossip algorithm is the randomized gossip, in which an agent is randomly activated according to a Poisson clock and chooses one of its neighbors randomly to average its state [5], [6], [7]. The randomization mechanism used in this gossip scheme is usually state-independent. We consider a different approach to gossip in which the agent chooses one of its neighbors based on its state. At one extreme, we may think of agents who prefer to gossip with neighbors with similar “opinions.” As in an echo-chamber, where agents may only talk to others if they reinforce their own opinions, it does not lead to an effective information mixing mechanism. At the opposite extreme, we consider agents who prefer to gossip with neighbors with maximal disagreement or dissent. In this article, we focus on the concept of max-dissent gossip as a state-dependent information mixing mechanism for distributed optimization. Our main contribution is to present a general treatment of asynchronous state-dependent averaging algorithms in distributed optimization and the importance of averaging the nodes corresponding to the max-edge, which leads to a contraction property of the state-dependent averaging matrix. We establish the convergence of the resulting distributed subgradient method under minimal assumptions on the underlying communication graph, and the local functions.

The idea of enabling a consensus protocol to use state-dependent matrices dates back to the Hegselmann and Krause [8] model for opinion dynamics. However, the literature on state-dependent averaging in distributed optimization is scarce and mostly motivated by applications in which the state represents the physical location of mobile agents (e.g., robots, autonomous vehicles, drones, etc.). In such settings, the state-dependency arises from the fact that agents that are physically closer have a higher probability of successfully communicating with each other [9], [10], [11]. Existing results assume that the local interactions between agents lead to strong connectivity over time. Unlike previous work, our model does not assume that the state of an agent represents its position in space. Moreover, we do not make strong assumptions on the network’s connectivity over time such as in [3] and [9].
Our work is closely related to state-dependent averaging schemes known as Load-Balancing [12] and Greedy Gossip with Eavesdropping [13]. The main idea in these methods is to accelerate averaging by utilizing the information from the most informative neighbor, i.e., the neighbors, whose states are maximally different with respect to some norm from each agent. We refer to it as the maximal dissent heuristics. The challenges of convergence analysis for maximal dissent averaging are highlighted in [12], [13], and [14]. However, concepts akin to max-dissent have only been explored for the specific problem of averaging [13]. Our work, on the other hand, focuses on distributed convex optimization, whose convergence is not guaranteed by the convergence of the averaging scheme alone.

As a broader impact of the results herein, we show that schemes that incorporate mixing of information between max-dissent agents will converge to a global optimizer of the underlying distributed optimization problem almost surely. Our result enables us to propose and extend the use of load-balancing and max-dissent gossip to distributed optimization. The key property of max-dissent averaging is that it leads to a contraction of the Lyapunov function used to establishing convergence. While recent work has considered similar contraction results (e.g., [15], [16]), they are not applicable to state-dependent schemes and do not establish almost sure convergence but only convergence in expectation.

The main contributions of this article are as follows.
1) We present state-dependent distributed optimization schemes that do not rely on or imply explicit strong-connectivity conditions.
2) We characterize a general result highlighting the importance of max-dissent agents on a graph for distributed optimization, which simplifies establishing convergence for a large class of asynchronous algorithms.
3) We prove the almost sure convergence of state-dependent algorithms to a global optimizer for distributed convex optimization problems.

A preliminary version of some of the ideas herein has previously appeared in [17], which addressed only one of the schemes, namely, Global Max-Gossip for distributed optimization of univariate functions. The results reported here are much more general than the ones in [17] addressing $d$-dimensional optimization and covering multiple state-dependent schemes in which the max-dissent agents communicate with nonzero probability at each time.

The rest of this article is organized as follows. First, we formulate distributed optimization problems and outline a generic state-dependent distributed subgradient method in Section II. In Section III, we introduce Local and Global Max-Gossip and review Randomized Gossip and Load Balancing distributed averaging schemes. We discuss the role of maximal dissent agents and their selection in averaging algorithms in Section IV. In Section V, we present our main results on the convergence of maximal dissent state-dependent distributed subgradient methods. In Section VI, we present a convergence rate analysis. We provide a numerical example that shows the benefit of using algorithms based on the maximal dissent averaging in Section VII. Finally, Section VIII concludes this article.

II. PROBLEM FORMULATION

Consider a distributed system with $n$ agents with an underlying communication network defined by a graph $G = ([n], E)$, where $[n] = \{1, 2, \ldots, n\}$. Each agent $i \in [n]$ has access to a local convex function $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$. The agents can communicate only with their one-hop neighbors as dictated by the network graph $G$. Our goal is to design a distributed algorithm to solve the following unconstrained optimization problem:

$$F^* = \min_{\mathbf{w} \in \mathbb{R}^d} F(\mathbf{w}), \quad \text{where} \quad F(\mathbf{w}) \triangleq \sum_{i=1}^{n} f_i(\mathbf{w}). \quad (1)$$

We assume that the local objective function $f_i$ is known only to node $i$ and the nodes can only communicate by exchanging information about their local estimates of the optimal solution. The solution set of the problem is defined as

$$\mathcal{W}^* \triangleq \arg \min_{\mathbf{w} \in \mathbb{R}^d} F(\mathbf{w}).$$

Throughout this article, we make extensive use of the notion of the subgradient of a function.

**Definition 1 (Subgradient):** A subgradient of a convex function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ at a point $\mathbf{w}_0 \in \mathbb{R}^d$ is a vector $\mathbf{g} \in \mathbb{R}^d$ such that

$$f(\mathbf{w}_0) + \langle \mathbf{g}, \mathbf{w} - \mathbf{w}_0 \rangle \leq f(\mathbf{w})$$

for all $\mathbf{w} \in \mathbb{R}^d$. We denote the set of all subgradients of $f$ at $\mathbf{w}_0$ by $\partial f(\mathbf{w}_0)$, which is called the subdifferential of $f$ at $\mathbf{w}_0$.

We make the following assumptions on the structure of the optimization problem in (1).

**Assumption 1 (Nonempty solution set):** The optimal solution set is nonempty, i.e., $\mathcal{W}^* \neq \emptyset$.

**Assumption 2 (Bounded Subgradients):** Each local objective function $f_i$’s subgradients are uniformly bounded. In other words, for each $i \in [n]$, there exists a finite constant $L_i$ such that for all $\mathbf{w} \in \mathbb{R}^d$, we have $\|\mathbf{g}\| \leq L_i$, $\mathbf{g} \in \partial f_i(\mathbf{w})$.

There exist many algorithms to solve the problem in (1). The pioneering work in [2] introduced a scheme, in which each agent keeps an estimate of the optimal solution and at each time step, the agents share their local estimate with their neighbors. Then, each agent updates its estimate using a time-varying, state-independent convex combination of the information received from their neighbors and its own local estimate. For $t \geq 0$, let $a_{ij}(t)$ denote the coefficients of the aforementioned convex combination at time $t$ such that $a_{ij}(t) \geq 0$, for all $i, j \in [n]$, $a_{ij}(t) = 0$ if $\{i, j\} \notin E$, and $\sum_{j=1}^{n} a_{ij}(t) = 1$, for all $i \in [n]$. Let $x_i(t)$ denote the $i$th agent’s estimate of the optimal solution at time $t$. The convex combination is followed by taking a step in the direction of any subgradient in the subdifferential at the local estimate, i.e.,

$$x_i(t+1) = \sum_{j=1}^{n} a_{ij}(t)x_j(t) - \alpha(t)g_i(t) \quad (2)$$

where $g_i(t) \in \partial f_i(x_i(t))$, and $\alpha(t)$ is a step-size sequence.

Herein, we generalize the algorithm in [2] by allowing the weights in the convex combination to be state-dependent in addition to being time-varying. Let each agent $i \in [n]$ initialize its estimate at an arbitrary point $x_i(0) \in \mathbb{R}^d$, which is updated at
discrete-time iterations \( t \geq 0 \) based on its own subgradient and the estimates received from neighboring agents as follows:

\[
\begin{align*}
    w_i(t+1) &= \sum_{j=1}^{n} a_{ij} \left( t, x_1(t), x_2(t), \ldots, x_n(t) \right) x_j(t) \\
    x_i(t+1) &= w_i(t+1) - \alpha(t+1) g_i(t+1)
\end{align*}
\]

where \( a_{ij}(t, x_1(t), x_2(t), \ldots, x_n(t)) \) are nonnegative weights, \( \alpha(t) \) is a step-size sequence, and \( g_i(t) \in \partial f_i(w_i(t)) \) for all \( t \geq 0 \). We can express this update rule compactly in matrix form as

\[
\begin{align*}
    W(t+1) &= A(t, X(t)) X(t) \\
    X(t+1) &= W(t+1) - \alpha(t+1) G(t+1)
\end{align*}
\]

where \( A(t, X(t)) \triangleq \left[ a_{ij}(t, x_1(t), \ldots, x_n(t)) \right]_{i,j \in [n]} \), and

\[
X(t) \triangleq \begin{bmatrix} x_1^T(t) \\ \vdots \\ x_n^T(t) \end{bmatrix}, \quad W(t) \triangleq \begin{bmatrix} w_1^T(t) \\ \vdots \\ w_n^T(t) \end{bmatrix}, \quad G(t) \triangleq \begin{bmatrix} g_1^T(t) \\ \vdots \\ g_n^T(t) \end{bmatrix}
\]

Another difference between (3) and (2) is that agent \( i \) computes the subgradient for the local function \( f_i \) at the computed average \( w_i(t+1) \) instead of \( x_i(t), t \geq 0 \).

**Assumption 3 (Diminishing step-size):** The step-sizes \( \alpha(t) > 0 \) form a nonincreasing sequence that satisfies

\[
\sum_{t=1}^{\infty} \alpha(t) = \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \alpha^2(t) < \infty.
\]

For a step-size sequence that satisfies Assumption 3, if the sequence of matrices \( \{ A(t) \} \), where \( A(t) = [a_{ij}(t)]_{i,j \in [n]} \), is doubly stochastic and sufficiently mixing, and the objective functions satisfy the regularity conditions in Assumptions 1 and 2, then the iterates in (2) converge to an optimal solution irrespective of the initial conditions \( x_i(0) \in \mathbb{R}^d \), i.e., \( \lim_{t \to \infty} x_i(t) = x^* \), \( i \in [n] \), where \( x^* \in \mathbb{R}^n \). Our goal for the remainder of this article is to establish a similar result for state-dependent maximal dissent distributed subgradient methods.

### III. State-Dependent Average-Consensus

In this section, we discuss three state-dependent average-consensus schemes. In doing so, we endeavor to unify the state-dependent average-consensus methodology. The first scheme, Local Max-Gossip, was studied in [13] exclusively for the average consensus problem. We provide two novel averaging schemes: 1) the Max-Gossip and 2) Load-Balancing averaging schemes. The dynamics of these algorithms can be understood as the instances of (3) with constant local cost functions \( f_i(x) \equiv c, \quad i \in [n] \), i.e.,

\[
X(t+1) = A(t, X(t)) X(t).
\]

We will consider three (two asynchronous and one synchronous) algorithms. The first two algorithms are related to the well-known Randomized Gossip algorithm [4], [5]. First, we present a brief description of Randomized Gossip.

#### A. Randomized Gossip

Consider a network \( G = ([n], E) \) of \( n \) agents, where each agent has an initial estimate \( x_i(0) \). At each iteration \( t \geq 0 \), a node \( i \) is chosen uniformly from \([n]\), independent of the earlier realizations. Then, \( i \) chooses one of its neighbors \( j \in \mathcal{N}_i \), where \( \mathcal{N}_i = \{ j \in [n] : (i, j) \in E \} \), with probability \( P_{ij} > 0 \). The two nodes exchange their current states \( x_i(t) \) and \( x_j(t) \), and update their states according to

\[
x_i(t+1) = \frac{1}{2} \left( x_i(t) + x_j(t) \right).
\]

The states of the remaining agents are unchanged. The update rule in (4) admits a more compact matrix representation as

\[
X(t+1) = B(e) X(t)
\]

where \( e = \{ i, j \} \), and

\[
B(e) = I - \frac{1}{2} (b_i - b_j) (b_i - b_j)^T
\]

where \( b_i \) is the \( i \)th standard basis vector of \( \mathbb{R}^n \). It is necessary that \( \sum_{i=1}^{n} P_{ii} = 1 \) for all \( i \), where \( P_{ii} = 0 \) if and only if \( \{ i, \ell \} \notin E \). The dynamical system described in (5) and its convergence rate are studied in [5].

#### B. Global Max-Gossip

The standard gossiping algorithm described above is state-independent in the sense that the selection of the gossiping edge \( e \) does not depend on the states at the agents at any time. Herein, we propose Global Max-Gossip where we select the edge connecting the agents with the largest possible dissent (disagreement) among all edges in the graph \( G = ([n], E) \), i.e.,

\[
e_{\max}(G, X) = \arg \max_{(i,j) \in E} \| x_i - x_j \|.
\]

In case there are multiple solutions to (7), we select the smallest pair of indices \((i^*, j^*)\) based on the lexicographical order, without loss of optimality. For brevity, we use \( e_{\max}(X) \) to denote the max-edge.

Global Max-Gossip serves as a benchmark as to what is achievable via state-dependent averaging schemes. Global Max-Gossip requires an oracle to provide the edge resulting in the largest possible Lyapunov function reduction across all network edges. Obtaining a decentralized algorithm to determine the max-dissent edge is a challenging open problem beyond the scope of this article.

Given an initial state matrix \( X(0) \), the Max-Gossip averaging scheme admits a state-dependent dynamics of the form

\[
A(t, X(t)) = B \left( e_{\max}(X(t)) \right)
\]

where the gossiping matrix is given by (6) and the max-edge is selected according to (7).

#### C. Local Max-Gossip

In Local Max-Gossip introduced in [13] under the moniker of Greedy Gossip with Eavesdropping, a random selected node
gossips with the neighbor \( j \in \mathcal{N}_i \) that has the largest\(^1\) possible state discrepancy with \( i \), i.e.,

\[
j = \arg \max_{j \in \mathcal{N}_i} \| x_j(t) - x_i(t) \|.
\]

(8)

Convergence is accelerated by gossiping with the neighbor with the largest disagreement as this leads to the largest possible immediate reduction in the Lyapunov function used to capture the variance of the states in the network.

Since the edge over which the gossiping occurs depends on the current state of the neighbors, the resulting averaging matrix is a state-dependent, random matrix. For a sequence of independently and uniformly distributed index sequence \( \{ s(t) \} \), the Local Max-Gossip dynamics can be written as a state-dependent averaging scheme as follows:

\[
A(t, X(t)) = B \left( \{ s(t), r_{s(t)}(X(t)) \} \right)
\]

where

\[
r_s(X) = \arg \max_{r \in \mathcal{N}_s} \| x_s - x_r \|.
\]

(9)

### D. Load-Balancing

Another state-dependent algorithm known as Load-Balancing can also be used to speed up convergence of average-consensus [14]. However, in contrast to the previous two cases, where only two nodes update at a given time, Load-Balancing is a synchronous averaging algorithm where all the agents operate simultaneously.

In the traditional Load-Balancing algorithm, the state at each agent is a scalar, which induces a total ordering among the agents, i.e., the neighbors of an agent are classified by having greater or smaller state values than the agent’s current state. When the states at the agents are multidimensional vectors, a total ordering is not available and must be defined. We introduce a variant of Load-Balancing based on the Euclidean distance between the states of any two agents as follows.

At time \( t \), each agent \( i \in [n] \) carries out the following steps.

1. Agent \( i \) sends its state to its neighbors.
2. Agent \( i \) computes the distance between its state and each of its neighbors. Let \( \mathcal{S}_i \) denote the subset of neighbors of agent \( i \) whose state have maximal Euclidean distance, i.e.,

\[
\mathcal{S}_i \triangleq \arg \max_{j \in \mathcal{N}_i} \| x_i - x_j \|.
\]

(10)

Agent \( i \) sends an averaging request to the agents in \( \mathcal{S}_i \).

3. Agent \( i \) receives averaging requests from its neighbors. If it receives a request from a single agent \( j \in \mathcal{S}_i \), then it sends an acknowledgment to that agent. In the event that agent \( i \) receives multiple requests, it sends an acknowledgment to one of the requests uniformly at random.

4. If agent \( i \) sends and receives an acknowledgment from agent \( j \), then it updates its state as \( x_i \leftarrow (x_i + x_j)/2 \).

The conditions for interaction between two nodes in Load-Balancing is characterized in the following proposition.

**Proposition 1:** Consider a connected graph \( \mathcal{G} \) and a stochastic process \( \{ X(t), A(t, X(t)) \} \), where \( A(t, X(t)) \) is the characterization of averaging according to the Load-Balancing algorithm, i.e., \( A(t, X(t))X(t) \) is the output of the Load-Balancing algorithm for a network with state matrix \( X(t), t \geq 0 \). The following statements hold.

1. Two agents \( i, j \) such that \( (i, j) \in \mathcal{E} \) average their states only if

\[
\| x_i(t) - x_j(t) \| \geq \max \left\{ \max_{r \in \mathcal{N}_i, \{ j \}} \| x_i(t) - x_r(t) \|, \right. \\
\left. \max_{r \in \mathcal{N}_j, \{ i \}} \| x_j(t) - x_r(t) \| \right\}.
\]

(11)

2. Let \( (i, j) \in \mathcal{E} \). If (11) holds with strict inequality, then \( i, j \) average their states.

Proposition 1 is proven in Appendix A.

### IV. On the Selection of Max-Edges

Consider the stochastic process \( \{ X(t), A(t, X(t)) \} \), where \( X(t) \) is the network state matrix, and \( A(t, X(t)) \) a state-dependent averaging matrix. Let \( \{ \mathcal{F}_t \}_{t=0}^\infty \) be a filtration such that \( \mathcal{F}_t \) is the \( \sigma \)-algebra generated by

\[
\{ \{ X(k), A(k, X(k)) \mid k \leq t \} \} \setminus \{ A(t, X(t)) \}.
\]

We establish a nonzero probability that a pair of agents that constitute a max-edge will update their states for the averaging schemes discussed in Section III.

**Proposition 2:** Let \( \{ X(t), A(t, X(t)) \}_{t=0}^\infty \) be the random process generated by either Randomized Gossip, Local Max-Gossip, Max-Gossip, or Load-Balancing consensus schemes. Then, for the random indices \( i^*, j^* \in [n] \) defined through the max-edge in (7) as \( \epsilon_{\text{max}}(X(t)) = \{ i^*, j^* \} \), we have

\[
\mathbb{E} \left[ \left( A(t, X(t))^T A(t, X(t)) \right) \mathcal{F}_{j^*} \right] \geq \delta \text{ a.s.}
\]

(12)

where \( \delta = \min_{i,j} \mathbb{P}(P_{ij}/n) \) for Randomized Gossip, such that \( P_{ij} \) is the probability that node \( i \) chooses node \( j \in \mathcal{N}_j \); \( \delta = 1/n \) for Local Max-Gossip; \( \delta = 1/2 \) for Global Max-Gossip; and \( \delta = 1/(2(n-1)^2) \) for Load-Balancing.

Proposition 2 establishes that given the knowledge until time \( t \), in expectation, the agents comprising the max-edge based on the network state matrix \( X(t) \) exchange their values with a positive weight bounded away from zero. Qualitatively, for gossip-based algorithms, this implies that there is a positive probability bounded away from zero that the agents comprising the max-edge carry out exchange of information with each other. We use Proposition 2 along with Theorem 3 to establish that the averaging matrices characterizing the algorithms discussed in Section III are contracting. Therefore, the subgradient methods based on these averaging algorithms converge to the same optimal solution almost surely as stated in Corollary 1 of Theorem 4. In other words, as long as the averaging step involves gossip over the max-edge with positive probability (bounded

---

\(^1\)In case there are multiple solutions to (8), we may select the agent with the smallest index, without loss of optimality.
away from zero), we will have a contraction in the Lyapunov function capturing the sample variance, which is a key step in proving the convergence of our averaging-based-subgradient methods. Proposition 2 is proven in Appendix B.

V. CONVERGENCE OF STATE-DEPENDENT DISTRIBUTED OPTIMIZATION

In the previous section, we have set the stage for studying the convergence of state-dependent averaging-based distributed optimization algorithms. Our proofs rely on two properties: 1) double stochasticity and 2) the contraction property (see Theorem 3).

To state the contraction property, we define the Lyapunov function $V : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}$ as

$$V(X) \triangleq \sum_{i=1}^{n} \| x_i - \bar{x} \|^2$$

(13)

where $X = [x_1, \ldots, x_n]^T$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$.

Theorem 3 (Contraction property): Consider a connected graph $G$ and the stochastic process $\{X(t), A(t, X(t))\}_{t=0}^{\infty}$ with a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ generated according to the dynamics in (3). If $A(t, X(t))$ is doubly stochastic for all $t \geq 0$, and for the random variables $i^*, j^* \in [n]$ defined through the max-edge in (7) as $e_{\max}(X(t)) = \{i^*, j^*\}$

$$E \left[ A \left( t, X(t) \right)^T A \left( t, X(t) \right) \mid \mathcal{F}_i \right]_{j^*} \geq \delta \text{ a.s.}$$

(14)

where $\delta > 0$, holds for all $t \geq 0$ and $X(0) \in \mathbb{R}^{n \times d}$, then

$$E \left[ V \left( A \left( t, X(t) \right) X(t) \right) \mid \mathcal{F}_i \right] \leq \lambda V \left( X(t) \right) \text{ a.s.}$$

(15)

where $\lambda = 1 - 2\delta/(n - 1) \text{diam}(G)^2$.

Theorem 3 is proven in Appendix C and provides our key new ingredient: proving a contraction result for doubly stochastic averaging matrices containing the maximally dissenting edge. Theorem 3 replaces the conventional step of bounding the error in the update rules involving properties such as the second largest eigenvalue of the averaging matrices that are used in the consensus step[6, Lemma 2]. Although inequality (15) appears similar to the contraction result obtained through the eigenvalues of the matrices, for state-dependent consensus algorithms the contraction factor at any time $t$ depends on the state of the estimate $X(t)$ and the underlying nonlinear averaging algorithm. The proof of Theorem 3 makes use of the double stochasticity of the matrices to characterize the exact one-step decrease in the Lyapunov function and then uses a clever trick to characterize its fractional decrease based on the fact that underlying communication graph is connected.

Remark 1: Theorem 3 also holds for time-varying graphs provided they remain connected at each time $t$. More precisely, the theorem holds for a sequence of connected graphs $\{G_t\}$ and at every time $t \geq 0$, for $i^*, j^*$ defined through $e_{\max}(G_t, X(t))$, the inequality in (14) holds, then the inequality in (15) will hold with scaling at time $t \geq 0$ being

$$\lambda_t = 1 - \frac{2\delta}{(n - 1) \text{diam}(G_t)^2} \leq 1 - \frac{2\delta}{(n - 1)^3}.$$

Therefore, the contraction property for connected time-varying graphs holds with a factor of at most $\lambda_t \leq 1 - 2\delta/(n - 1)^3$.

For a connected graph $G$ and the stochastic process $\{X(t), A(t, X(t))\}_{t=0}^{\infty}$ with the filtration $\{\mathcal{F}_t\}_{t=0}^{\infty}$ generated according to the dynamics in (3), we define a contracting averaging matrix as follows.

Definition 2 (Contracting averaging matrix): A state-dependent averaging matrix $A(t, X(t))$ is contracting with respect to the Lyapunov function $V(\cdot)$ in (13) if there exists a $\lambda \in (0, 1)$ such that

$$E \left[ V \left( A \left( t, X(t) \right) X(t) \right) \mid \mathcal{F}_i \right] \leq \lambda V \left( X(t) \right)$$

(16)

holds a.s. for all $t \geq 0$.

The main result of this work establishes convergence guarantees for these dynamics as stated ahead.

Theorem 4 (Almost sure convergence of state-dependent subgradient methods): Consider the distributed optimization problem in (1) and let Assumptions 1 and 2 hold. Assume a connected communication graph $G$ and the subgradient method (3) if the random matrices $A(t, X(t))$ in (3) are doubly stochastic and contracting, and the step-sizes $\{\alpha(t)\}$ follow Assumption 3, then for all initial conditions $X(0) \in \mathbb{R}^{n \times d}$,

$$\lim_{t \to \infty} w_i(t) = w^* \quad \forall i \in [n] \text{ a.s.}$$

where $w^* \in W^*$.

Theorem 4 establishes the almost-sure convergence to an optimal solution of (1), based on the consensus-based subgradient methods where the averaging matrices are doubly stochastic and contracting. Theorem 3 provides a simplified condition, the presence of averaging over the “max-edge,” which implies the averaging mechanism is contracting. Note that, as shown in Proposition 2, this simplified condition holds for Local Max-Gossip, Max-Gossip, and Load-Balancing averaging. Thus, we have the subsequent corollary following immediately from Proposition 2, Theorem 3, and Theorem 4.

Corollary 1: Consider the distributed optimization problem in (1) and let Assumptions 1 and 2 hold. Assume a connected communication graph $G$ and the subgradient method (3) where the averaging matrices $A(t, X(t))$ in (3) are based solely on either the Local Max-Gossip, Max-Gossip, or Load-Balancing averaging, and the step-sizes $\{\alpha(t)\}$ follow Assumption 3. Then, $\lim_{t \to \infty} w_i(t) = w^*$ for all $i \in [n]$ for all initial condition $X(0) \in \mathbb{R}^{n \times d}$, and some $w^* \in W^*$.

For the remainder of this section, we provide the key steps and results that are needed to prove Theorem 4. We defer the proof of these technical results to the Appendix.

The proof strategy for Theorem 4 can be broken down into two main steps: 1) showing that the evolution of the dynamics followed by the average state variable $\{x(t)\}$ converges to a solution of the optimization problem in (1) and 2) every node $i \in [n]$ tracks the dynamics of this average state variable such that the tracking error goes to zero. The first step requires the following result, which establishes a bound on the accumulation of the tracking error for every agent.

Lemma 1: Let $G$ be a connected graph and consider sequences $\{W(t)\}$ and $\{X(t)\}$ generated by the subgradient
method in (3) using state-dependent, doubly stochastic, and contracting averaging matrices \(A(t, X(t))\). If Assumptions 2 and 3 hold, then for any initial estimates \(X(0) \in \mathbb{R}^{n \times d}\), then a.s. for all \(i \in [n]\), we have \(\lim_{t \to \infty} \|w_i(t + 1) - \bar{x}(t)\| = 0\) and \(\sum_{i=0}^{\infty} \alpha(t + 1) E[(\|w_i(t + 1) - \bar{x}(t)\|) \mid \mathcal{F}_t] < \infty\).

Lemma 1, which is proven in Appendix D, establishes guarantees on the consensus error for the local estimates \(w_i(t)\). Lemma 2 will be used to bound the distance of the average state \(\bar{x}(t)\) to an optimal point.

**Lemma 2 (Lemma 8 in [18]):** Suppose that Assumption 2 holds. Then, for any connected graph \(\mathcal{G}\), initial condition \(X(0) \in \mathbb{R}^{n \times d}\), \(v \in \mathbb{R}^d\), and \(t \geq 0\), for the dynamics \(\{X(t), A(t, X(t))\}\) of the subgradient method (3) where \(A(t, X(t))\) are doubly stochastic, we have

\[
\mathbb{E}[(\|\bar{x}(t + 1) - v\|^2 \mid \mathcal{F}_t)] \leq \|\bar{x}(t) - v\|^2 - \alpha(t + 1) \frac{2}{n} (F(\bar{x}(t)) - F(v)) + \alpha(t + 1) \frac{4}{n} \sum_{i=1}^{n} L_i E[\|w_i(t + 1) - \bar{x}(t)\| \mid \mathcal{F}_t] + \alpha^2(t + 1) \frac{L^2}{n^2} \text{ a.s.}
\]

We note that [18, Lemma 8] was originally intended for state-independent dynamics. However, its proof only relies on the double stochasticity of the averaging matrices, convexity of the local functions, boundedness of the subgradients, and not on whether the averaging is state-dependent or not.

Finally, combining the above two results implies that the distance of each agent’s local estimate \(x_i(t)\) to the optimal set \(\mathcal{W}^*\) will be approximately decreasing. The following result will then be used to show that this approximate decrease results in convergence to \(\mathcal{W}^*\).

**Lemma 3:** Consider a minimization problem \(\min_{x \in \mathbb{R}^d} f(x)\), where \(f : \mathbb{R}^d \to \mathbb{R}\) is a convex function. Assume that the solution set \(X^*\) of the problem is nonempty. Let \(\{x_i\}\) be a stochastic process such that for all \(x \in X^*\) and for all \(t \geq 0\)

\[
\mathbb{E}[\|x_{t+1} - x\|^2 \mid \mathcal{F}_t] \leq (1 + b_t)\|x_t - x\|^2 - a_t (f(x_t) - f(x)) + c_t \text{ a.s.}
\]

where \(b_t \geq 0\), \(a_t \geq 0\), and \(c_t \geq 0\) for all \(t \geq 0\) and \(\sum_{t=0}^{\infty} b_t < \infty\), \(\sum_{t=0}^{\infty} a_t = \infty\), and \(\sum_{t=0}^{\infty} c_t < \infty\) a.s. Then, the sequence \(\{x_t\}\) converges to a solution \(x^* \in X^*\) a.s.

This result has been proven as part of [19, Th. 1] but due to the standalone significance of the result, we have stated it as a lemma above and its proof is provided in Appendix E. Now, we are ready to formally prove Theorem 4 by combining the aforementioned results.

**Proof of Theorem 4:** From Lemma 2, for \(v = w^* \in \mathcal{W}^*\), we have

\[
\mathbb{E}[\|\bar{x}(t + 1) - w^*\|^2 \mid \mathcal{F}_t] \\
\leq \|\bar{x}(t) - w^*\|^2 - \frac{2\alpha(t + 1)}{n} (F(\bar{x}(t)) - F(w^*)) + \alpha^2(t + 1) \frac{L^2}{n^2} + 4 \frac{\alpha(t + 1)}{n} \sum_{i=1}^{n} L_i E[\|w_i(t + 1) - \bar{x}(t)\| \mid \mathcal{F}_t]
\]

for all \(t \geq 0\). From Lemma 1, we know that

\[
\sum_{t=0}^{\infty} \frac{4\alpha(t + 1)}{n} \sum_{i=1}^{n} L_i E[\|w_i(t + 1) - \bar{x}(t)\| \mid \mathcal{F}_t] < \infty \text{ a.s.}
\]

Furthermore, \(\alpha(t)\) is not summable and \(\sum_{t=0}^{\infty} \alpha^2(t) < \infty\). Therefore, all the conditions for Lemma 3 hold with \(a_t = 2\alpha(t + 1)/n\), \(b_t = 0\), and

\[
c_t = \alpha(t + 1) \frac{4}{n} \sum_{i=1}^{n} L_i E[\|w_i(t + 1) - \bar{x}(t)\| \mid \mathcal{F}_t] + \alpha^2(t + 1) \frac{L^2}{n^2}.
\]

Therefore, from Lemma 3, the sequence \(\{\bar{x}(t)\}\) converges to a solution \(\bar{w} \in \mathcal{W}^*\) almost surely. Finally, Lemma 1 implies that \(\lim_{t \to \infty} \|w_i(t + 1) - \bar{x}(t)\| = 0\) for all \(i \in [n]\) almost surely. Therefore, the sequences \(\{w_i(t + 1)\}\) converge to the same solution \(\bar{w} \in \mathcal{W}^*\) for all \(i \in [n]\) almost surely.

**VI. CONVERGENCE RATE**

In this section, we discuss the convergence rate of the time-averaged version of the discussed state-dependent consensus-based subgradient methods when the step size at time \(t\) is set as \(1/\sqrt{t}\) for \(t \geq 1\). The convergence rates for the different algorithms differ via the contraction factor \(\lambda\) defined for the contracting averaging matrix through (16). In the following theorem, we provide a bound on the convergence rate for the deviation of functions value computed at a time-averaged version of the consensus-based subgradient method from the optimal value. The proof closely follows the proof in [18, Th. 2]. For completeness the proof along with additional discussion is provided in the extended version [20].

**Theorem 5:** Consider the assumptions of Theorem 4 with \(\alpha(t) = 1/\sqrt{t}\) for all \(t \geq 1\) and \(w^* \in \mathcal{W}^*\) and let \(K = 1/\sqrt{2}\). For \(\bar{w}_i(t + 1) = \frac{\sum_{k=0}^{t} \alpha(k + 1) w_i(k + 1)}{\sum_{k=0}^{t} \alpha(k + 1)}\), we have

\[
\mathbb{E}[F(\bar{w}_i(t + 1)) - F(w^*)] \leq \frac{\alpha E[\|\bar{x}(0) - w^*\|^2]}{2\sqrt{t + 1}} + \frac{L^2(1 + \ln(t + 1))}{2n \sqrt{t + 1}} + \frac{L(2\sqrt{n} + 1)K}{\sqrt{t + 1}} E[\|X(0) - \bar{X}(0)\|^2] + L^2K(2\sqrt{n} + 1) \frac{1 + \ln t}{\sqrt{t + 1}}.
\]

Therefore, the subgradient method converges at the rate of \(O\left(\frac{1}{\sqrt{t}}\right)\). However, the hidden constant terms of the convergence rate are influenced by the consensus algorithm used with the subgradient descent. In Theorem 5, the consensus step of the algorithms influences the convergence rate through the constant \(K\). Based on Theorem 3, the contraction factor

\[
\lambda = 1 - \frac{2\delta}{(n - 1)diam(\mathcal{G})^2},
\]
where $\delta$ for Randomized Gossip, Local Max-Gossip, Max-Gossip, and Load Balancing is provided through Proposition 2, the constant in the convergence rate are bounded as stated in the following proposition.

**Proposition 6:** In Theorem 5, the constant $K$ is given by $\frac{5}{2\sqrt{3}}$, which is bounded above by $n^2(n - 1)\text{diam}(G)^2$ for Randomized Gossip, $n(n - 1)\text{diam}(G)^2$ for Local Max-Gossip, $2(n - 1)\text{diam}(G)^2$ for Max-Gossip, and $(n - 1)^3\text{diam}(G)^2$ for Load Balancing being used as the averaging scheme with the subgradient method.

**Remark 2:** The above result uses a conservative bound on the contraction factor $\lambda > 0$. The values mentioned in Corollary 6 are upper bounds on the constants in the convergence rate. However, tighter bounds on the constant $K$ is possible. In principle, in the proof of Theorem 5, for each of the state-dependent algorithm, such a contraction factor would depend on the sample path (past trajectory) of the dynamics. For example, when the consensus scheme used is Load Balancing, we know that when the nodes do not have multiple neighbors with maximal disagreement, the constant $\delta$ in Proposition 2 is even greater than 1/2, more precisely, it is $C_c(X)/2$, where $C_c$ is the number of edges over which the exchange is taking place in the averaging step with the state estimate $X$. With the improved $\delta$, the bound on the constant $K$ can be improved to

$$\frac{(n - 1)\text{diam}(G)^2}{2C_c(X(t))} \leq \frac{(n - 1)\text{diam}(G)^2}{2}.$$  

Similarly, the bounds on the convergence rate for Local Max-Gossip can be improved by using tighter contraction factor for the averaging matrices. However as seen from [13, Th. 2], the contraction factor may take cumbersome form, which cannot be readily used to establish better bounds for the constant $K$, and hence, the problem of finding useful convergence rate for state-dependent averaging is a nontrivial open problem.

**VII. NUMERICAL EXAMPLE**

To illustrate our analytical results, we present a simulation of a distributed optimization problem where the local functions’ subgradients are not restricted to be uniformly bounded. We look at the standard distributed estimation problem in a sensor network setting with $n = 180$ agents, where each agent $i \in [n]$ wants to estimate an unknown parameter $\theta_0$ while having access to a noisy measurement of the parameter $c_i = \theta_0 + n_i$, where $n_i$’s are independent, zero mean Gaussian random variables with variance $\sigma_i^2 > 0$. In this setting, the maximum likelihood (ML) estimator [21, Th. 5.3] is the minimizer of the separable cost function $F(w) = \sum_{i=1}^n (w - c_i)^2/\sigma_i^2$. Note that this problem is a distributed optimization problem with the local cost function $f_i(w) = (w - c_i)^2/\sigma_i^2$. For the variance $\sigma_i^2$, we picked $1/\sigma_i^2$ independently and uniformly over $(0,1)$. For each node $i \in [n]$, the initial local estimates $x_i(0)$ are drawn independently from a standard Gaussian distribution.

We consider schemes’ performance for different underlying communication graph ranging from dense graphs (Complete and Barbell), moderately dense graphs (Erdős–Rényi), to sparse graphs (Line and Star). We chose a connected graph with the edge probability $p = 0.4$ for Erdős–Rényi graph. For the Barbell graph, we chose equal number of nodes for the three components—two Complete graphs and the connecting Line graph.

For the state-dependent averaging, we used four different update rules: 1) Randomized Gossip [5], 2) Local Max-Gossip, 3) Max-Gossip, and 4) Load-Balancing. For the Randomized Gossip, at each time, a node in $[n]$ wakes up uniformly at random, and it chooses one of its neighbors uniformly at random for communication. To account for the stochastic nature of Randomized Gossip and Local Max-Gossip algorithms, we average the error values over 10 runs keeping the initial conditions and samples at the nodes the same. The resulting plots in Fig. 1 show the decay of the error $\|w(t) - w_*\|$ as a function of $t$, where $w_* = \sum_{i=1}^n \frac{c_i}{\sigma_i^2} / \sum_{i=1}^n \frac{1}{\sigma_i^2}$ is the optimal solution for $F(w)$.

For the Erdős–Rényi communication graph, we also plot the decay of the error with the number of bits exchanged between the nodes in Fig. 2 for Randomized Gossip, Local Max-Gossip, and Load-Balancing.

In the simulation, 32 b are used for exchange of the estimates and 1 b is used for the exchange of each acknowledgment. Therefore, the number of bits exchanged per step for Randomized Gossip is 64. For Local Max-Gossip, at time $t$ with $s(t) \in [n]$ being the randomly chosen node, $|N_{s(t)}| + 32|N_{s(t)}| + 32$ bits are exchanged for waking up the neighboring nodes, obtaining their values, and sending the neighbor with the maximum disagreement its own value. Finally, for Load Balancing, $32 \sum_{i=1}^n |N_i| + n + \text{ACK}(t)$ bits are exchanged for sharing the values with the neighbors, sending request to the neighbor with the maximum disagreement, and sending the acknowledgment, where $\text{ACK}(t)$ is the total number of acknowledgment bits at time $t$.²

**A. Comparison of Asynchronous Methods**

From Fig. 1, the performance of the subgradient methods using state-dependent averaging shows an improvement in the convergence rate. The convergence rates increase as we go from Randomized Gossip, Local Max-Gossip, and Max-Gossip to Load-Balancing averaging-based optimizers. We will refer to the subgradient methods using the state-dependent averaging by their averaging algorithm in the succeeding discussion. Note that convergence rate also depends on the graph topology: Local Max-Gossip applied to a Star graph has essentially the same rate as Randomized Gossip since the nodes at the periphery have only the central node as the choice to gossip with, and the probability of the first node being selected for gossiping is $n - 1$ times larger to be a peripheral node as compared to the central node. Overall, we notice the increase in the performance of Max-Gossip and Local Max-Gossip as compared to Randomized Gossip with increasing connectivity. Moreover, from Fig. 2, we note the significantly better performance of Local Max-Gossip with respect to the number of exchanges between the nodes as opposed to that of synchronous Load-Balancing.

²In the numerical simulation, there are no cases with multiple neighbors with maximum disagreement.
Fig. 1. Error decay for different graphs with 180 nodes. (a) Complete Graph. (b) Barbell Graph. (c) Line Graph. (d) Star Graph.

**B. Max-Gossip Versus Load-Balancing**

Unlike gossip, Max-Gossip, and Local Max-Gossip, Load-Balancing is a synchronous scheme where in addition to the max-edge, other local max-edges are often incorporated in the averaging scheme simultaneously. Therefore, it is only natural that the convergence rate of Load-Balancing is superior to that of Max-Gossip since it averages not only the two nodes defined by the max-edge but also other nodes connected by edges with large disagreement at the same time. By a similar logic, for the Complete graph, the performance of Load-Balancing and Max-Gossip are the same; since all the nodes send their request for averaging to either the node with the maximum or minimum estimate resulting in only the max-edge performing the updates.

We observe that the gap in performance of Load-Balancing and Max-Gossip, which has the best performance among the discussed asynchronous methods, increases with the diameter of the graph. Characterizing the analytical dependence of the convergence rate as a function of graph topology metrics is of interest for future work.

In the extended version of this article [20], we provide a numerical example optimizing the regularized logistic regression for a classification problem over the MNIST dataset.

**VIII. CONCLUSION**

In this article, we proposed, studied, and analyzed the role of maximal dissent nodes in distributed optimization schemes, leading to many exciting state-dependent consensus-based subgradient methods. The proof of our result relies on a contraction property of these schemes. Our result opens up avenues for synthesizing or extending the use of state-dependent averaging schemes for distributed optimization including the Max-Gossip, Local Max-Gossip, and Load-Balancing algorithms. Finally, we compared simulation results of a distributed estimation problem for gossip-based subgradient methods and the proposed state-dependent algorithms. Our numerical experiments show the faster convergence speed of schemes that use maximal dissent between nodes compared with state-independent gossip schemes. These simulations strongly support the intuition behind our main result, i.e., mixing of information between the maximal dissent nodes is critically important for the working (and enhancing) of the consensus-based subgradient methods. Although we have shown the convergence of such state-dependent algorithms, establishing tighter bounds on their rate of convergence and, especially, relating them to various graph quantities such as diameter and edge density of the
graph remains open problems for future research endeavors. The introduction of a state-dependent element for other class of algorithms specifically those that provide linear convergence rates such as distributed gradient tracking method [22], [23] and their convergence analysis are part of future direction for the problem.

APPENDIX A

PROOF OF PROPOSITION 1

Proof of Proposition 1: For any $\omega \in \Omega$, consider $X(t; \omega) \in \mathbb{R}^{n \times d}$. If nodes $i$ and $j$ update their values to their average, that is $(x_i(t; \omega) + x_j(t; \omega))/2$, then we know that during the round of Load-Balancing algorithm starting at value $X(t; \omega)$ in step 2, node $i$ and node $j$ have sent their averaging request to each other. Therefore, we have $j \in \arg \max_{i \in N_i} \| x_j(t; \omega) - x_i(t; \omega) \|$ and $i \in \arg \max_{i \in N_j} \| x_j(t; \omega) - x_i(t; \omega) \|$. Hence, for any $\omega \in \Omega$

$$
\| x_i(t; \omega) - x_j(t; \omega) \| \geq \max \left\{ \max_{r \in N_i \setminus \{j\}} \| x_i(t; \omega) - x_r(t; \omega) \|, \max_{r \in N_j \setminus \{i\}} \| x_j(t; \omega) - x_r(t; \omega) \| \right\}
$$

On the other hand, if (17) holds with strict inequality, then node $i$ and node $j$ send averaging requests only to each other in step 2 and respond to each other in step 3, and carry out their averaging according to step 4.

APPENDIX B

PROOF OF PROPOSITION 2

Proof: We first discuss the result for Randomized Gossip, Local Max-Gossip, and Max-Gossip averaging. The averaging matrices for the gossip algorithms where two agents update their states to their average takes the form of (6). Therefore, for these gossip algorithms, we have $A(t, X(t))^T A(t, X(t)) = A(t, X(t))$ and $\mathbb{E}[A(t, X(t))^T A(t, X(t)) | F_t] = \mathbb{E}[A(t, X(t)) | F_t]$.

Consider two nodes $i, j \in [n]$ such that $\{i, j\} \in E$. For Randomized Gossip, $\mathbb{E}[A(t, X(t))_{ij} | F_t] = (P_{ij} + P_{ji})/2n$. Moreover, since $\{i, j\} \in E$, we have $P_{ij}, P_{ji} > 0$. Let $P_\delta = \min_{(i,j) \in E} P_{ij}$. For the max-edge $\{i^*, j^*\}$, (12) holds with $\delta = P_\delta/n > 0$.

Let $i \in [n]$ and state estimate matrix be $X(t)$. For Local Max-Gossip, let $r_i(X(t))$ be determined according to (9). Consider the max-edge $\{i^*, j^*\}$ of $X(t)$. Then, $r_{i^*}(X(t)) = j^*$ and $r_{j^*}(X(t)) = i^*$. Thus

$$
\mathbb{E}[A(t, X(t))_{i^*, j^*} | F_t] = \frac{1}{n}
$$

and Local Max-Gossip averaging satisfies (12) with $\delta = 1/n$. Similarly, for the Max-Gossip averaging with state estimate $X(t)$, for the max-edge $\epsilon_{\max}(X(t)) = \{i^*, j^*\}$, we have $\mathbb{E}[A(t, X(t))_{i^*, j^*} | F_t] = \frac{1}{2}$, and (12) holds with $\delta = 1/2$.

Let us now discuss the presence of max-edge in the Load-Balancing averaging scheme. Consider the state estimate matrix $X(t)$ and $\epsilon_{\max}(X(t)) = \{i^*, j^*\}$ to be the max-edge with respect to $X(t)$. By the definition of a max-edge we know that nodes $i^*, j^*$ satisfy inequality (11).

Consider the case when nodes $i^*, j^*$ satisfy (11) with strict inequality. From Proposition 1, we know that $A(t, X(t))_{i^*, j^*}, A(t, X(t))_{j^*, i^*}, A(t, X(t))_{i^*, i^*}, A(t, X(t))_{j^*, j^*}$ are equal to 1/2, which implies that $A(t, X(t))_{i^* j^*} = A(t, X(t))_{j^* i^*} = 0$ for all $\ell \notin \{i^*, j^*\}$. Therefore

$$
\mathbb{E}[A(t, X(t))^T A(t, X(t)) | F_t]_{i^* j^*} = 1/2
$$

and the inequality in (12) holds with $\delta = 1/2$.

Finally, consider the case when there are multiple neighbors of nodes $i^*, j^*$ with distance equal to $\|x_{i^*}(t) - x_{j^*}(t)\|$. Let $|S_{i^*}| \geq 1$ and $|S_{j^*}| \geq 1$ where $S_{i^*}$ is given by (10). Then, according to Load-Balancing algorithm, nodes $i^*, j^*$ update their states to their average with probability $1/(|S_{i^*}| \cdot |S_{j^*}|)$. Since $|S_{i^*}| \leq n - 1$ and $|S_{j^*}| \leq n - 1$, we have

$$
\mathbb{E}[A(t, X(t))^T A(t, X(t)) | F_t]_{i^* j^*} = \frac{1}{2(n-1)^2}
$$

and (12) holds with $\delta = 1/(2(n-1)^2)$.

APPENDIX C

PROOF OF THEOREM 3

To prove Theorem 3, we must first define a few quantities related to the distance between the nodes on the graph and their relationships.

Definition 3: Consider a connected graph $G$ and a matrix $X = [x_1, \ldots, x_n]^T \in \mathbb{R}^{n \times d}$ such that $x_i \in \mathbb{R}^d$ is the estimate at node $i$ in the graph $G$. Let $d(X)$ denote the maximal distance between the estimates of any two nodes in the graph

$$
d(X) \triangleq \max_{i, j \in \{1, 2, \ldots, n\}} \| x_i - x_j \|.
$$

Let $d_G(X)$ denote the maximal distance between the estimates among any two connected nodes in the graph

$$
d_G(X) \triangleq \max_{(i,j) \in E} \| x_i - x_j \|.
$$

Finally, let $diam(G)$ denote the longest shortest path between any two nodes of the graph $G$.

Proposition 7: Given a connected graph $G$ and a matrix $X = [x_1, \ldots, x_n]^T \in \mathbb{R}^{n \times d}$, such that $x_i \in \mathbb{R}^d$ is the solution estimate at node $i$ in the graph $G$, we have $d(X) \geq diam(G)$. Let $\delta_G(X)$ denote the minimal distance among any two connected nodes in the graph

$$
d_G(X) \geq \max_{(i,j) \in E} \| x_i - x_j \|.
$$

Finally, let $diam(G)$ denote the longest shortest path between any two nodes of the graph $G$.

Proof: The upper bound on $d_G(X)$ follows from (18) and (19) in Definition 3. To prove the lower bound on $d_G(X)$, we assume, without loss of generality, that the rows of the matrix $X \in \mathbb{R}^{n \times d}$ are such that $d(X) = \| x_i - x_{i^*} \|$. Since $G$ is connected, its diameter is finite and there is a path of length $k \leq diam(G)$, denoted by $\{ v_0, v_1 \}, \{ v_1, v_2 \}, \ldots, \{ v_{k-1}, v_k \}$, where $v_0 = 1$ and $v_k = n$, with $v_i \in [n]$ for $i = 0, 1, \ldots, k$. The distance $d(x)$
is bounded as
\[ \|x_1 - x_n\| \leq \sum_{i=0}^{k-1} \|x_{v_i} - x_{v_{i+1}}\| \]  
(20)
where (20) follows from the triangle inequality. Finally, each term in the sum (20) is bounded above by \( d_2(x) \). Hence, \( d(X) \leq kd_2(x) \leq \text{diam}(G)d_2(x) \).

Next, we state a result quantifying the decrease in the Lyapunov function defined in (13) that is the vector form of [14, Lemma 1].

**Lemma 4:** Given a doubly stochastic matrix \( A \in \mathbb{R}^{n \times n} \), let \( c_{ij} \) denote the \((i, j)\)th entry of the matrix \( A^TA \). Then for all \( X = [x_1, \ldots, x_n]^T \in \mathbb{R}^{n \times d} \), we have
\[
V(AX) = V(X) - \sum_{i<j} c_{ij} \|x_i - x_j\|^2.
\]

**Proof:** By definition, the Lyapunov function in (13) can be written as \( V(X) = \text{tr}[(X - \bar{X})^T(X - \bar{X})] \), where \( \bar{X} = \frac{1}{n}X \). The doubly stochasticity of \( A \) implies \( A\bar{X} = \frac{1}{n}AX = \frac{1}{n}X = A\bar{X} \). Therefore,
\[ V(AX) = \text{tr}[(AX - A\bar{X})^T(AX - A\bar{X})]. \]

Finally,
\[
V(X) - V(AX) = \text{tr}[(X - \bar{X})^T(I - A^TA)(X - \bar{X})].
\]
Since \( A^TA \) is a symmetric and stochastic matrix, we have
\[ c_{ij} = c_{ji} \text{ and } c_{ii} = 1 - \sum_{i \neq j} c_{ij} \text{. Thus}
A^TA = I - \sum_{i<j} c_{ij}(b_i - b_j)(b_i - b_j)^T\]
where \( b_i \in \mathbb{R}^n \) is the standard basis vector for all \( i \in [n] \).

Using arguments similar to the ones from [14, Lemma 9], for \( X = [x_1, \ldots, x_n]^T \in \mathbb{R}^n \), the corresponding max-edge \( \epsilon_{\text{max}}(X(t)) = \{i^*, j^*\} \) and the doubly stochastic averaging matrix \( A^tX(t) \) such that
\[
\mathbb{E}[A(t, X(t))^TA(t, X(t))]_{i, j} \geq \delta > 0 \text{ a.s.}
\]
Define \( \Omega_\delta(t) = \{ \omega : \mathbb{E}[\mathbb{E}[\mathbb{E}[A(t, X(t))^TA(t, X(t))]_{i, j} | \mathcal{F}_t] \geq \delta \} \).
For legibility, we drop the time index in the variables for the rest of this proof and use \( X, \mathcal{F}, A(X), \Omega_\delta \) instead of \( X(t), \mathcal{F}_t, \Omega_\delta(t), \) and \( A(t, X(t)) \).

Using arguments similar to the ones from [14, Lemma 9], for \( X = [x_1, \ldots, x_n]^T \in \mathbb{R}^n \) doubly stochastic matrix \( A(X) \) such that
\[
\mathbb{E}[\mathbb{E}[\mathbb{E}[A(X)^TA(X)]]_{i, j} | \mathcal{F}_t] \geq \delta > 0 \text{ a.s.}
\]
where \( \mathcal{F} \) is a \( \sigma \)-field, \( X = [x_1, \ldots, x_n]^T \), and \( \epsilon_{\text{max}}(X) = \{i^*, j^*\} \). We will show that \( \mathbb{E}[V(A(X)X) | \mathcal{F}_t] \leq \lambda V(X) \) a.s. for some \( \lambda \in (0, 1) \). From Lemma 4, the difference in the quadratic Lyapunov function \( V \) evaluated at \( X \) and \( A(X)X \) is given by
\[
V(X) - V(A(X)X) = \sum_{i<j} c_{ij}(X)\|x_i - x_j\|^2
\]
where \( c_{ij}(X) \) is the \((i, j)\)th entry of \( A(X)^TA(X) \), i.e., \( c_{ij}(X) = (A(X)^TA(X))_{i,j} \). Taking the conditional expectation with respect to the filtration \( \mathcal{F}_s \), we obtain
\[
V(X) - V(A(X)X) | \mathcal{F}_t = \sum_{i<j} \mathbb{E}\left[\mathbb{E}\left[\left( (A(X)^TA(X))_{ij} | \mathcal{F}_t \right) \right] \right] \|x_i - x_j\|^2
\]
\[ \geq c_{ij}(X)\|x_i - x_j\|^2 \geq \delta \|x_i - x_j\|^2 \text{ a.s.}
\]
where \( \epsilon_{\text{max}}(X) = \{i^*, j^*\} \) and the first inequality follows from the nonnegativity of the squared terms and the second inequality follows from (21). Recall that the constant \( \delta \) depends on the averaging scheme.

If \( V(X) = 0 \), more precisely for the samples path characterized by \( \omega \in \Omega_\delta(t) \) such that \( X(t, \omega) = 0 \) only \( X = 1c^T \) for some \( c \in \mathbb{R}^d \). Therefore, \( A(X)X = A(X)1c^T = 1c^T \) since \( A(X) \) is doubly stochastic and \( V(X, A(X)) = 0 \). Thus, the inequality \( \mathbb{E}[V(A(X)X) | \mathcal{F}_t] \leq \lambda V(X) \) is satisfied.

Let \( \mathcal{L} = \{ 1p^T | p \in \mathbb{R}^d \} \). For \( X = [x_1, \ldots, x_n]^T \notin \mathcal{L} \), more precisely for the samples path characterized by \( \omega \in \Omega_\delta(t) \) such that \( X(t, \omega) \notin \mathcal{L} \), the conditional expected fractional decrease in the Lyapunov function is
\[
\frac{V(X) - \mathbb{E}[V(A(X)X) | \mathcal{F}_t]}{V(X)} \geq \frac{\delta}{\text{diam}(G)^2} \sum_{i=1}^n \|x_i - \bar{x}\|^2.
\]
where \( \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \). Using the definition of \( d_2(x) \) and Proposition 7, we obtain the following bound:
\[
\frac{V(X) - \mathbb{E}[V(A(X)X) | \mathcal{F}_t]}{V(X)} \geq \frac{\delta}{\text{diam}(G)^2} \sum_{i=1}^n \|x_i - \bar{x}\|^2.
\]

For \( X \notin \mathcal{L} \), let \( g(X) \triangleq \frac{d^2(X)}{\sum_{i=1}^n \|x_i - \bar{x}\|^2} \). Note that \( g(X) \) satisfies the following invariance relations:
\[
g(X + 1p^T) = g(X), \quad p \in \mathbb{R}^d
\]
and \( g(cX) = g(X) \) for \( c \in \mathbb{R} \setminus \{0\} \). Therefore, for \( X \notin \mathcal{L} \), the following inequality and identity hold
\[
g(X) \geq \min_{Z \in \mathbb{R}^{n \times d}, \sum_i z_i = 0} \sum_{i=1}^n \|z_i\|^2 \cdot d^2(Z).
\]
Note that if \( \sum_{i=1}^n z_i = 0 \) and \( \sum_{i=1}^n \|z_i\|^2 = 1 \), then we have
\[
\sum_{1 \leq i < j \leq n} \langle z_i, z_j \rangle = -\frac{1}{2} \sum_{i=1}^n \|z_i\|^2 = -\frac{1}{2}.
\]
By definition, \( d(Z) = \|z_i - z_j\| \) for all \( i, j \in [n] \). Using the fact that the maximum of a set of values is greater than or equal to the average for the set \( \{\|z_i - z_j\|^2\}_{1 \leq i < j \leq n} \), we get
\[
d^2(Z) \geq \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \|z_i - z_j\|^2 = \frac{2}{n-1}.
\]

\[ ^3 \text{We omit the dependence on } \omega \text{ and } t \text{ for legibility.} \]
where the last step follows from (22) and the fact that $\sum_{i=1}^n \|z_i\|^2 = 1$. Finally, using (22), we get

$$
\frac{V(X) - \mathbb{E}[V(A(X)) | F]}{V(X)} \geq \frac{2\delta}{(n-1)\text{diam}(G)^2}.
$$

Since $\mathbb{E}[V(A(X)) | F] \leq \lambda V(X)$ for $X \in \mathcal{L}$ and for $X \notin \mathcal{L}$, we have $\mathbb{E}[V(A(X)) | F] \leq \lambda V(X)$ a.s. Thus

$$
\mathbb{E}[V(A(t, X(t)) | F)] \leq \lambda V(X(t)) \text{ a.s.}
$$

where $\lambda = 1 - 2\delta/((n-1)\text{diam}(G)^2)$.

**APPENDIX D**

**LIMITING PROPERTIES OF THE LYAPUNOV FUNCTION $V(\cdot)$**

**Proof of Lemma 1:** To study the convergence of $V(W(t))$, we first derive a super-martingale like inequality for the stochastic process $\{V(W(t))\}$. For $X(t) \in \mathcal{F}_t$, using the contracting averaging property of $A(t, X(t))$ in (16), we get

$$
\mathbb{E}[V(W(t+1)) | \mathcal{F}_t] = \mathbb{E}[V(A(t, X(t))X(t) | \mathcal{F}_t)] 
\leq \lambda V(X(t)) \text{ a.s.}
$$

where $\lambda \in (0, 1)$. We know that $X(t) = W(t) + E(t)$, so from triangle inequality on $\|W(t) - W(t) + E(t) - E_t\|_F$, we have

$$
V(X(t)) \leq V(W(t)) + V(E(t)) + 2\sqrt{V(W(t))} \sqrt{V(E(t))}.
$$

Using the inequality above in (23), for all $t \geq 0$, we get

$$
\mathbb{E}[V(W(t+1)) | \mathcal{F}_t] \leq \lambda \left( V(W(t)) + V(E(t)) + 2\sqrt{V(W(t))} \sqrt{V(E(t))} \right) \text{ a.s.}
$$

Since $V(E(t)) = \|E(t) - \bar{E}(t)\|_F^2 = \|E(t)\|_F^2 \leq L^2 \alpha^2(t)$, we get

$$
\mathbb{E}[V(W(t+1)) | \mathcal{F}_t] \leq \lambda \left( \sqrt{V(W(t))} + L\alpha(t) \right)^2 \text{ a.s.}
$$

From Jensen’s inequality, we have

$$
\mathbb{E}\left[ \sqrt{V(W(t+1))} | \mathcal{F}_t \right] \leq \sqrt{\mathbb{E}[V(W(t+1)) | \mathcal{F}_t]} \leq \lambda \left( \sqrt{V(W(t))} + L\alpha(t) \right) \text{ a.s.}
$$

Taking the expectation, multiplying by $\alpha(t+1)$ and using the fact that $\{\alpha(t)\}$ is nonincreasing, we get

$$
\alpha(t+1)\mathbb{E}\left[ \sqrt{V(W(t+1))} \right] \leq \alpha(t)\mathbb{E}\sqrt{V(W(t))}
$$

$$- (1 - \sqrt{\lambda})\alpha(t)\mathbb{E}\sqrt{V(W(t))} + \alpha^2(t) \text{ a.s.}
$$

Since the diminishing step sequence $\{\alpha(t)\}$ satisfies $\sum_{t=1}^\infty \alpha^2(t) < \infty$, Robbins–Siegmund Theorem [24] implies

$$
\sum_{t=1}^\infty \alpha(t)\mathbb{E}\sqrt{V(W(t))} < \infty
$$

and by the Monotone Convergence Theorem, we have

$$
\mathbb{E}\left[ \sum_{t=1}^\infty \alpha(t) \sqrt{V(W(t))} \right] < \infty
$$

which implies that $\sum_{t=1}^\infty \alpha(t) \sqrt{V(W(t))} < \infty$ a.s. Since

$$
V(W(t)) = \sum_{i=1}^n \|w_i(t) - \bar{w}(t)\|^2,
$$

we know that

$$
\sum_{i=1}^\infty \alpha(t)\|w_i(t) - \bar{w}(t)\| \leq \sum_{i=1}^\infty \alpha(t) \sqrt{V(W(t))} < \infty
$$

for all $i \in [n]$ a.s. Since $\sum_{i=1}^\infty \alpha(t)\|w_i(t) - \bar{w}(t)\| < \infty$ and $\sum_{i=1}^\infty \alpha(t) = \infty$, we have

$$
\lim_{t \to \infty} \inf \|w_i(t) - \bar{w}(t)\| = 0 \quad \forall i \in [n] \quad \text{a.s.}
$$

Furthermore, since we have

$$
\sum_{t=1}^\infty \alpha(t)\mathbb{E}\sqrt{V(W(t))} = \mathbb{E}\left[ \sum_{t=1}^\infty \alpha(t)\mathbb{E}\left[ \sqrt{V(W(t))} | \mathcal{F}_t \right] \right]
$$

using Monotone Convergence Theorem similar to (24) implies that $\mathbb{E}\left[ \sum_{t=1}^\infty \alpha(t)\|w_i(t) - \bar{w}(t)\| | \mathcal{F}_t \right] < \infty$, and hence, $\sum_{t=1}^\infty \alpha(t)\mathbb{E}\left[ \sqrt{V(W(t))} | \mathcal{F}_t \right] < \infty$ a.s., and therefore

$$
\sum_{t=1}^\infty \alpha(t)\|w_i(t) - \bar{w}(t)\| | \mathcal{F}_t < \infty \quad \forall i \in [n], \text{ a.s.}
$$

Furthermore, for all $t \geq 0$, we know

$$
\mathbb{E}[V(W(t+1)) | \mathcal{F}_t]
\leq \lambda \left( V(W(t)) + 2L\alpha(t) \sqrt{V(W(t))} + L^2 \alpha^2(t) \right) \text{ a.s.}
$$

Since we have

$$
\sum_{t=1}^\infty 2\alpha(t) \sqrt{V(W(t))} + \lambda L^2 \alpha^2(t) < \infty \quad \text{a.s.}
$$

again Robbins–Siegmund Theorem [24] implies that $\{V(W(t))\}$ converges a.s. Therefore

$$
\|w_i(t+1) - \bar{w}(t+1)\| \text{ converges} \quad \forall i \in [n] \quad \text{a.s.}
$$

Using (25) with the above result, we get

$$
\lim_{t \to \infty} \|w_i(t+1) - \bar{w}(t+1)\| = 0 \quad \forall i \in [n] \quad \text{a.s.}
$$

Finally, since $\bar{w}(t+1)^T = \frac{1}{n} \bar{X}(t+1) = \frac{1}{n} \bar{A}(t, X(t))X(t)$ from the double stochasticity of $A(t, X(t))$, we have

$$
\bar{w}(t+1)^T = \frac{1}{n} \bar{X}(t)^T = x(t)^T
$$

which from (26) and (27) implies Lemma 1.

**APPENDIX E**

**PROOF OF LEMMA 3**

To prove Lemma 3, we follow the proof in [19, Th. 1].

**Proof:** For all $x \in \mathbb{X}^n$ and $t \geq 0$, we have

$$
\mathbb{E}\left[ \|x_{t+1} - x\|^2 | \mathcal{F}_t \right] \leq (1 + b_t)\|x_t - x\|^2
$$

$$- a_t (f(x_t) - f(x)) + c_t \text{ a.s.}
$$
For any $x \in X^+$, Robbins–Siegmund Theorem [24] implies that $\{\|x_t - x\|\}$ converges and $\sum_{t=0}^{\infty} a_t (f(x_t) - f(x)) < \infty$ a.s. Since for any $x \in X^+$ we have $f(x) = f^*$, the event

$$\Omega_{X^+} = \left\{ \omega : \lim_{t \to \infty} \|x_t(\omega) - x\| \text{ exists, and} \sum_{t=0}^{\infty} a_t \|f(x_t(\omega)) - f^*(\omega)\| < \infty \right\}$$

is such that $\mathbb{P}(\Omega_{X^+}) = 1$. Note that here we denote by $\{x_t(\omega)\}_{t \geq 0}$ the sample path for the corresponding $x$.

Let $X_d^+ \subseteq X^+$ be a countable dense subset of $X^+$ and $\Omega_d = \bigcap_{x \in X_d^+} \Omega_x$. We have $\mathbb{P}(\Omega_d) = 1$ since $X_d^+$ is countable. For any $\omega \in \Omega_d$, since $\sum_{t=0}^{\infty} a_t = \infty$ and $\sum_{t=0}^{\infty} a_t (f(x_t(\omega)) - f^*(\omega)) < \infty$, we have $\lim_{t \to \infty} f(x_t(\omega)) = f^*$. From this limit infimum and the continuity of $f$, for all $\omega \in \Omega_d$, we have $\lim_{t \to \infty} \|x_t(\omega) - x(\omega)\| = 0$, for some $x(\omega) \in X^+$. Consider a subsequence $\{x_{s_k}(\omega)\}_{k \geq 0}$ of $\{x_t(\omega)\}_{t \geq 0}$ such that $\lim_{k \to \infty} f(x_{s_k}(\omega)) = f^*$.

For any $\omega \in \Omega_d$, $\lim_{t \to \infty} \|x_t(\omega) - x(\omega)\|$ exists for $x \in X_d^+$. Therefore, the sequences $\{x_{s_k}(\omega)\}_{k \geq 0}$ are bounded. Hence, $\{x_{s_k}(\omega)\}_{k \geq 0}$ is bounded, has a limit point $x^*(\omega) \in X^+$, and without loss of generality, $\lim_{k \to \infty} x_{s_k}(\omega) = x^*(\omega)$. Since $X_d^+$ is dense, there is a sequence $\{q_s(\omega)\}_{s \geq 0}$ in $X_d^+$ such that $\lim_{s \to \infty} \|q_s(\omega) - x^*(\omega)\| = 0$.

For $\omega \in \Omega_d$, $\lim_{s \to \infty} \|x_{s_k}(\omega) - q_s(\omega)\|$ exists for all $s \geq 0$, which is $\|x^*(\omega) - q_s(\omega)\|$. Moreover,

$$\lim_{t \to \infty} \|x_t(\omega) - q_s(\omega)\| \leq \lim \inf_{t \to \infty} \|x_t(\omega) - x^*(\omega)\| + \|x^*(\omega) - q_s(\omega)\| \leq \|x^*(\omega) - q_s(\omega)\|$$

which implies that $\lim_{s \to \infty} \lim_{t \to \infty} \|x_{s_k}(\omega) - q_s(\omega)\| = 0$. Finally

$$\lim \sup_{t \to \infty} \|x_t(\omega) - x^*(\omega)\| \leq \lim \sup_{s \to \infty} \lim_{t \to \infty} \|x_{s_k}(\omega) - q_s(\omega)\| + \|q_s(\omega) - x^*(\omega)\| = 0.$$  

Therefore, for any $\omega \in \Omega_d$, we have $\lim_{t \to \infty} x_t(\omega) = x^*(\omega)$, where $x^*(\omega) \in X^+$. So we have, $\lim_{t \to \infty} x_t = x^*$ a.s.

\section*{REFERENCES}

[1] J. Tsitsiklis, D. Bertsekas, and M. Athans, “Distributed asynchronous deterministic and stochastic gradient optimization algorithms,” IEEE Trans. Autom. Control, vol. AC-31, no. 9, pp. 803–812, Sep. 1986.
[2] A. Nedic and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” IEEE Trans. Autom. Control, vol. 54, no. 1, pp. 48–61, Jan. 2009.
[3] A. Nedic, A. Ozdaglar, and P. A. Parrilo, “Constrained consensus and optimization in multi-agent networks,” IEEE Trans. Autom. Control, vol. 55, no. 4, pp. 922–938, Apr. 2010.
[4] D. Shah, “Gossip algorithms,” Foundations Trendes, vol. 3, no. 1, pp. 1–125, 2009.
[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” IEEE Trans. Inf. Theory, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.
[6] S. S. Rasmussen, A. Nedić, and V. V. Veeravalli, “Asynchronous gossip algorithms for stochastic optimization,” in Proc. 48th IEEE Conf. Decis. Control Held Jointly 28th Chin. Control Conf., 2009, pp. 3581–3586.

\section*{Ashwin Verma (Graduate Student Member, IEEE)} received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 2018. He is currently working toward the Ph.D. degree in electrical and computer engineering with the Department of Electrical and Computer Engineering, University of California, San Diego, CA, USA.

His research interests include control theory, optimization, information theory, and applied probability theory.

\section*{Marcos M. Vasconcelos} received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA, in 2016.

He is currently an Assistant Professor with the Department of Electrical Engineering, FAMU-FSU College of Engineering, Florida State University, Tallahassee, FL, USA. He was a Research Assistant Professor with the Commonwealth Cyber Initiative and the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, from 2021 to 2022. From 2016 to 2020, he was a Postdoctoral Research Associate with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, California, USA.

His research interests include networked control and estimation, multi-scale robotic networks, game theory, multiagent optimization, distributed machine learning, and systems biology.
Urbashi Mitra (Fellow, IEEE) received the B.S. and M.S. degrees from the University of California at Berkeley, Berkeley, CA, USA, and the Ph.D. degree from Princeton University, Princeton, NJ, USA, in 1987, 1989, and 1994, respectively, all in electrical engineering.

She is currently the Gordon S. Marshall Professor in engineering with the University of Southern California, Los Angeles, CA, USA, with appointments in electrical engineering and computer science. Her research interests include wireless communications, communication and sensor networks, biological communication systems, detection and estimation, and the interface of communication, sensing, and control.

Dr. Mitra is the inaugural Editor-in-Chief for IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS. She was a member of IEEE Information Theory Society’s Board of Governors from 2002 to 2007 and from 2012 to 2017, EEE Signal Processing Society’s Technical Committee on Signal Processing for Communications and Networks from 2012 to 2017, IEEE Signal Processing Society’s Awards Board from 2017 to 2018, and the Vice-Chair of IEEE Communications Society, Communication Theory Working Group from 2017 to 2018. She is the recipient of the 2017 IEEE Communications Society Women in Communications Engineering Technical Achievement Award, a 2016 U.K. Royal Academy of Engineering Distinguished Visiting Professorship, a 2016 US Fulbright Scholar Award, a 2016–2017 U.K. Leverhulme Trust Visiting Professorship, 2015–2016 IEEE Communications Society Distinguished Lecturer, 2012 Globecom Signal Processing for Communications Symposium Best Paper Award, 2012 US National Academy of Engineering Lillian Gilbreth Lectureship, the 2009 DCOSS Applications and Systems Best Paper Award, Texas Instruments Visiting Professor (Fall 2002, Rice University), 2001 Okawa Foundation Award, 2000 OSU College of Engineering Lumley Award for Research, 1997 OSU College of Engineering MacQuigg Award for Teaching, and a 1996 National Science Foundation CAREER Award.

Behrouz Touri (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2006, the M.Sc. degree in communications, systems, and electronics from Jacobs University, Bremen, Germany, in 2008, and the Ph.D. degree in industrial engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2011.

He is currently an Associate Professor of Electrical and Computer Engineering with the University of California San Diego, San Diego, CA, USA. Between 2011 and 2014, he was a Postdoctoral Researcher with the Electrical and Computer Engineering Departments of the University of Illinois, Illinois, USA, and Georgia Institute of Technology, Atlanta, GA, USA. Prior to joining the University of California San Diego in 2017, he was an Assistant Professor of Electrical Engineering with the University of Colorado Boulder, Boulder, CO, USA. His research interests include applied probability theory, distributed optimization, control and estimation, population dynamics, and evolutionary game theory.

Dr. Touri was a recipient of American Control Council’s Donald P. Eckman Award in 2016.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.