

Quantized Consensus via Adaptive Stochastic Gossip Algorithm

Javad Lavaei and Richard M. Murray

Abstract—This paper is concerned with the distributed averaging problem over a given undirected graph. To enable every vertex to compute the average of the initial numbers sitting on the vertices of the graph, the policy is to pick an edge at random and update the values on its ending vertices based on some rules, but only in terms of the quantized data being exchanged between them. Our recent paper showed that the quantized consensus is reached under a simple updating protocol which deploys a fixed tuning factor. The current paper allows the tuning factor to be time-dependent in order to achieve two goals. First, this makes it possible to study the numerical stability of the protocol with a fixed tuning factor under a small perturbation of this parameter. Furthermore, exploiting a time-varying tuning factor facilitates the implementation of the consensus protocol and pushes the steady state of the system towards an equilibrium point, as opposed to making it oscillatory. The current paper is an important extension of our recent work, which generalizes a finite-dimensional problem to an infinite-dimensional one that is more challenging in nature.

I. INTRODUCTION

Distributed computation has been an active research area in the past few decades, which targets the computation of a quantity of interest, such as temperature or some other measurement, over a network of processors in a decentralized way [1], [2]. In particular, the distributed averaging problem has drawn much attention recently, whose objective is to decentrally compute the average of several real numbers lying on the nodes of a network [3], [4]. In a more general framework, consider a network of agents where each agent is associated with some data (such as a real number or an image). The problem of contriving a protocol by means of which the agents can update their data so that ultimately they all agree upon some universal shared data is called the *consensus* or *state agreement* problem [3], [4].

Various problems in different fields can be interpreted as a consensus problem. For instance, the synchronization of coupled oscillators appearing in biophysics, neurobiology, and systems biology, is nothing but reaching a consensus on the frequencies of all agents [7], [8]. Moreover, the problem of aligning the heading angles of a group of mobile agents (e.g. a flock of birds) can be regarded as a consensus problem on the heading angles [11]. Given a sensor network comprising a set of sensors measuring the same quantity in a noisy environment, the problem of reaching a consensus on the state estimates is discussed in [12]. The consensus problem for networks of dynamic agents with fixed and switching topologies is tackled in [3], where it is shown that

the convergence rate is related to the algebraic connectivity of the network. The work [13] elaborates on the relationship between the amount of information exchanged by the agents and the rate of convergence to the consensus.

Consider a consensus problem over a network of agents associated with a set of real numbers. There are applications for which the agents cannot communicate synchronously. Therefore, the gossip algorithm has been widely exploited by researchers to handle the averaging problem asynchronously [14], [15]. This type of algorithm selects a pair of agents at each time, and updates their values based on some averaging policy. The consensus problem in the context of gossip algorithm has been thoroughly investigated in the literature [16], [17], [18], [19]. For instance, the work [16] studies the convergence of a general randomized gossip algorithm, and derives conditions under which the algorithm converges. This paper also shows that the averaging time of a gossip algorithm depends on the second largest eigenvalue of a doubly stochastic matrix characterizing the algorithm.

In light of practical communication constraints, a pair of agents can normally exchange only their quantized data. This has given rise to the emergence of quantized gossip algorithms. The notion of quantized consensus is defined in [18] for the case when quantized values (integers) are to be averaged over a connected network with digital communication channels. This paper shows that the quantized gossip algorithm leads to reaching the quantized consensus. The result is extended in [19] to the case when the quantization is uniform, and the initial numbers owned by the agents are reals (as opposed to being integers). The paper [19] shows that the quantized gossip algorithm works for a particular choice of the updating parameter. Our recent papers [20], [21] prove the convergence of this algorithm for a wide range of updating parameters, under any arbitrary quantizer including uniform and logarithmic ones. Furthermore, the steady-state behavior of the system together with the expected value of the convergence time is studied extensively in these papers.

However, the stochastic gossip algorithm studied in [19], [20], [21] is based on the hypothesis that the tuning factor (updating parameter) is fixed. There are two incentives to relax this assumption and allow this parameter to be time-varying. First, the tuning factor is prone to numerical errors in practice and, in order to investigate the stability of the above-mentioned algorithm with respect to this parameter, one way is to let this factor be time-dependent. Moreover, it is easy to observe that an appropriate tuning of this factor at each time update in terms of the previous state of the system makes the state approach an equilibrium point, rather

The authors are with the Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, USA (emails: lavaei@cds.caltech.edu; murray@cds.caltech.edu).

than being oscillated (which normally happens when a fixed updating parameter is deployed). Thus, the objective of the current paper is to show the convergence to a quantized consensus even in the case of using a time-varying tuning parameter. This modification in the gossip algorithm converts the finite-dimensional state space to an infinite-dimensional one, for which most of the ideas developed in [20], [21] break down.

The paper is organized as follows. Some preliminaries are presented in Section II, and the problem is formulated accordingly. The convergence proof is provided in Section III. The results are illustrated in Section IV through simulations. Some concluding remarks are drawn in Section V.

II. PROBLEM FORMULATION

Consider an undirected connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{P})$ with the set of vertices $\mathcal{V} := \{v_1, v_2, \dots, v_\nu\}$ and the set of edges \mathcal{E} . Assume that $\mathcal{P} := \{p_{ij}\}_{i,j}$ is a set of weights assigned to the edges of \mathcal{G} such that:

$$\sum_{\substack{i,j \in \nu \\ i \leq j}} p_{ij} = 1 \quad (1)$$

where $\nu := \{1, 2, \dots, \nu\}$. The quantity p_{ij} ($i, j \in \nu$) is equal to zero if $(i, j) \notin \mathcal{E}$; otherwise, it is strictly positive. In particular, $p_{11}, p_{22}, \dots, p_{\nu\nu}$ are all equal to zero. The set \mathcal{P} defines a discrete probability distribution on the edges of \mathcal{G} , i.e. it specifies the probability by which a certain edge of the graph can be chosen at random. Suppose that a real number x_i has been assigned to the vertex v_i , for all $i \in \nu$. Let $q(x) : \mathbb{R} \rightarrow \mathbb{R}$ be a general quantization operator characterized as follows:

$$q(x) = \begin{cases} L_i & \text{if } x \in [L_i, \bar{L}_i] \\ L_{i+1} & \text{if } x \in (\bar{L}_i, L_{i+1}] \end{cases} \quad \forall i \in \mathbf{Z} \quad (2)$$

where $\{L_i\}_{-\infty}^{\infty}$ is a monotonically increasing sequence of reals representing the quantization levels, and:

$$\bar{L}_i := \frac{L_i + L_{i+1}}{2}, \quad \forall i \in \mathbf{Z} \quad (3)$$

The scalar quantities L_i and \bar{L}_i will be referred to as *level* and *splitting level*, respectively. For some technical reasons, assume that the terms $L_{i+1} - L_i$, $i \in \mathbf{Z}$, are all bounded away from 0. The following gossip algorithm is proposed in [19]:

Stochastic Gossip (SG) Algorithm:

Step 1: Pick a positive real ε , and set $k = 0$. Define $x_i[0] := x_i$, for every $i \in \nu$.

Step 2: Pick an edge of \mathcal{G} at random.

Step 3: Suppose that the ending vertices of the edge selected in step 2 possess the values $x_i[k]$ and $x_j[k]$. Perform the following updates:

$$\begin{aligned} x_i[k+1] &= x_i[k] + \varepsilon \times (q(x_j[k]) - q(x_i[k])), \\ x_j[k+1] &= x_j[k] + \varepsilon \times (q(x_i[k]) - q(x_j[k])), \\ x_q[k+1] &= x_q[k], \quad \forall q \in \nu \setminus \{i, j\} \end{aligned} \quad (4)$$

Step 4: Increase k by 1 and jump to step 2.

For simplicity, the short-hand notation:

$$X[k] = [x_1[k] \ x_2[k] \ \dots \ x_\nu[k]], \quad k \in \mathbf{Z} \quad (5)$$

will be used henceforth. The next definition is extracted from [20].

Definition 1: Given a stochastic quantization-based protocol acting on \mathcal{G} (e.g. the SG algorithm), assume that the vector $X[k]$ denotes the values on the vertices of \mathcal{G} at time k obtained using this protocol. It is said that the *quantized consensus* is reached almost surely for the graph \mathcal{G} under the protocol \mathcal{C} if for every arbitrary initial state $x_i[0] \in \mathbb{R}^\nu$, with probability 1 there exist a natural number k_0 and an integer μ such that either of the following sets of relations holds:

$$\begin{cases} \sum_{i=1}^{\nu} x_i[k] = \sum_{i=1}^{\nu} x_i[0] \\ x_j[k] \in [L_\mu, L_{\mu+1}] \end{cases} \quad \forall k \geq k_0, \forall j \in \nu \quad (6)$$

or:

$$\begin{cases} \sum_{i=1}^{\nu} x_i[k] = \sum_{i=1}^{\nu} x_i[0] \\ x_j[k] \in (\bar{L}_\mu, \bar{L}_{\mu+1}] \end{cases} \quad \forall k \geq k_0, \forall j \in \nu \quad (7)$$

The names *consensus* and *quantized consensus* will be used interchangeably in the rest of the paper. It was shown in [20] that the quantized consensus is reached for the connected graph \mathcal{G} under the above-mentioned gossip algorithm if $\varepsilon \in (0, 0.5]$. However, two important issues arise regarding this algorithm:

- First, it remains unclear if this algorithm is numerically stable with respect to ε in the case when ε could be perturbed infinitesimally at each time instant.
- It sounds reasonable to pick a small ε when the two numbers being updated are close to each other, and pick a large ε otherwise. This brings the possibility of allowing ε to be time-varying, in which case the convergence to the quantized consensus needs to be investigated.

These questions motivate the modification of the SG algorithm in such a way that the above-mentioned issues are taken into consideration. In this regard, the following algorithm is introduced.

Adaptive Stochastic Gossip (ASG) Algorithm:

Step 1: Set $k = 0$. Define $x_i[0] := x_i$, for every $i \in \nu$.

Step 2: Pick an edge of \mathcal{G} at random.

Step 3: Select a positive number and denote it with ε_k . Suppose that the ending vertices of the edge selected in step 2 possess the values $x_i[k]$ and $x_j[k]$. Perform the following updates:

$$\begin{aligned} x_i[k+1] &= x_i[k] + \varepsilon_k \times (q(x_j[k]) - q(x_i[k])), \\ x_j[k+1] &= x_j[k] + \varepsilon_k \times (q(x_i[k]) - q(x_j[k])), \\ x_q[k+1] &= x_q[k], \quad \forall q \in \nu \setminus \{i, j\} \end{aligned} \quad (8)$$

Step 4: Increase k by 1 and jump to step 2.

The primary goal of this paper is to study whether the quantized consensus is reached under this adaptive algorithm. More precisely, it is desired to show that if there exist two positive numbers ε_{\min} and ε_{\max} in the interval $(0, 0.5)$ such that $\varepsilon_{\min} \leq \varepsilon_k \leq \varepsilon_{\max}$ for all $k \in \mathbf{N} \cup \{0\}$, then almost surely the quantized consensus is reached asymptotically.

III. MAIN RESULTS

Assume that the initial state $X_0 := X[0]$ belongs to the hyper-rectangle $[L_{\min}, L_{\max}]^\nu$, where L_{\min} and L_{\max} are some quantization levels. Moreover, suppose that there exist two numbers ε_{\min} and ε_{\max} in the interval $(0, 0.5)$ such that the tuning factor ε_k chosen in step 3 of the ASG algorithm always belongs to the interval $[\varepsilon_{\min}, \varepsilon_{\max}]$, for all $k \in \mathbf{N} \cup \{0\}$. Before proceeding with the development of the paper, a few definitions need to be provided.

Definition 2: Define η_1 and η_2 to be:

$$\begin{aligned} \eta_1 &= \max_{i \in \mathbf{Z}} \bar{L}_i \quad \text{s.t.} \quad \bar{L}_i \leq x_{\text{ave}}, \\ \eta_2 &= \min_{j \in \mathbf{Z}} \bar{L}_j \quad \text{s.t.} \quad \bar{L}_j \geq x_{\text{ave}} \end{aligned} \quad (9)$$

where $x_{\text{ave}} := \frac{x_1 + x_2 + \dots + x_\nu}{\nu}$.

Definition 3: For every $i \in \mathbf{Z}$, define:

$$\mathcal{S}(\bar{L}_i) := (\omega_i, \omega'_i]^\nu \quad (10)$$

where:

$$\begin{aligned} \omega_i &:= \bar{L}_i - \varepsilon_{\max}(L_{i+1} - L_i), \\ \omega'_i &:= \bar{L}_i + \varepsilon_{\max}(L_{i+1} - L_i) \end{aligned} \quad (11)$$

(note that for a set \mathcal{R} , the symbol \mathcal{R}^ν denotes the product set $\mathcal{R} \times \mathcal{R} \times \dots \times \mathcal{R}$).

Definition 4: For every $i \in \mathbf{Z}$, define the distance function $d(\cdot, \mathcal{S}(\bar{L}_i)) : [L_{\min}, L_{\max}]^\nu \rightarrow \mathbf{R}^+ \cup \{0\}$ as:

$$d(\alpha, \mathcal{S}(\bar{L}_i)) := \min_{\beta \in \mathcal{S}(\bar{L}_i)} |\alpha - \beta|_1 \quad (12)$$

for all $\alpha \in [L_{\min}, L_{\max}]^\nu$, where $|\cdot|_1$ denotes the \mathcal{L}_1 norm.

Now, it is desired to motivate the introduction of the set $\mathcal{S}(\bar{L}_i)$ and its associated distance function $d(X[\cdot], \mathcal{S}(\bar{L}_i))$. This illustration is essential to understanding the arguments made in the current work. Notice that although the graph \mathcal{G} is coordinate free, when it comes to assigning real numbers to its vertices, it makes sense to incorporate the topology of the graph and its corresponding initial state into a graph whose geometry matters. To be more precise, draw the graph \mathcal{G} in the 2-dimensional plane so that its vertex v_i is placed in the coordinates $(i, x_i(0))$, for all $i \in \nu$. Denote this coordinated graph with $\vec{\mathcal{G}}$. As an example, if \mathcal{G} contains 4 vertices with the edges $\{(1, 2), (2, 3), (3, 4), (4, 1)\}$ and the initial values $X_0 = (3, 3, 2, 0.5)$, then the corresponding coordinated graph $\vec{\mathcal{G}}$ will turn out to be the one depicted in Figure 1. When the ASG algorithm is run on the graph \mathcal{G} , the values sitting on the vertices change. This makes the graph $\vec{\mathcal{G}}$ move in the plane (each vertex moves up or down vertically). As a result, define $\vec{\mathcal{G}}(k)$ to be the corresponding coordinated graph at time k ($k \in \mathbf{N}$). Now, draw the horizontal lines $y = \bar{L}_j$ and $y = \bar{L}_j$, $\forall j \in \mathbf{Z}$, in the plane containing the graph

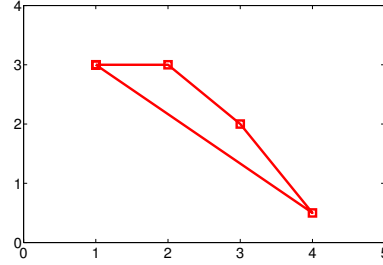


Fig. 1. An example of the coordinated graph $\vec{\mathcal{G}}$ with four vertices.

$\vec{\mathcal{G}}$. These lines represent all quantization levels as well as splitting levels. Given $i \in \mathbf{Z}$, sketch two specific lines $y = \bar{L}_i - \varepsilon_{\max}(L_{i+1} - L_i)$ and $y = \bar{L}_i + \varepsilon_{\max}(L_{i+1} - L_i)$. These lines that surround the horizontal line $y = \bar{L}_i$ create a strip in the plane (this strip, for instance, resembles the shaded area in Figure 2). The set $\mathcal{S}(\bar{L}_i)$ can be visualized to be this strip with its bottom border line removed. Moreover, the distance function $d(X[k], \bar{L}_i)$ is indeed the \mathcal{L}_1 distance of the vertices of $\vec{\mathcal{G}}(k)$ from the aforementioned strip. It will be shown in Theorem 1 that this distance function, which is defined in a natural way, is always non-increasing. Furthermore, Theorem 2 will substantiate that, with probability 1, there exists a time instant k_0 such that the graph $\vec{\mathcal{G}}(k)$ either lies entirely on one side of the line $y = \bar{L}_i$ for all $k \geq k_0$ or asymptotically converges to the strip sketched around the line $y = \bar{L}_i$ as k goes to infinity. These theorems are given in the sequel.

Theorem 1: Run the ASG algorithm on the graph \mathcal{G} with the initial state X_0 . For every $i \in \mathbf{Z}$, the following statements hold:

- i) If $X[k] \in \mathcal{S}(\bar{L}_i)$ for some $k \in \mathbf{N}$, then $X[k'] \in \mathcal{S}(\bar{L}_i)$ for all $k' \geq k$. In other words, $\mathcal{S}(\bar{L}_i)$ is an invariant set.
- ii) The function $d(X[\cdot], \mathcal{S}(\bar{L}_i))$ is non-increasing, i.e.:

$$d(X[k], \mathcal{S}(\bar{L}_i)) \leq d(X[k-1], \mathcal{S}(\bar{L}_i)) \quad (13)$$

for all $k \in \mathbf{N}$.

- iii) Given a positive real number $\mu \in (0, 0.5 - \varepsilon_{\max})$, assume that an edge $(j, p) \in \mathcal{E}$ is selected at time k , $k \in \mathbf{N}$, for which the set of relations:

$$\begin{aligned} x_j[k-1] &\leq \bar{L}_i, \\ x_p[k-1] &> \bar{L}_i + (\varepsilon_{\max} + \mu)(L_{i+1} - L_i) \end{aligned} \quad (14)$$

or:

$$\begin{aligned} x_j[k-1] &> \bar{L}_i, \\ x_p[k-1] &\leq \bar{L}_i - (\varepsilon_{\max} + \mu)(L_{i+1} - L_i) \end{aligned} \quad (15)$$

holds. Then:

$$\begin{aligned} d(X[k], \mathcal{S}(\bar{L}_i)) &\leq d(X[k-1], \mathcal{S}(\bar{L}_i)) \\ &\quad - \min(\mu, \varepsilon_{\min})(L_{i+1} - L_i) \end{aligned} \quad (16)$$

Proof: The proofs of Parts (i) and (ii) are straightforward. To prove Part (iii), different possibilities with regards to the

values sitting on the j -th and p -th vertices must be studied separately. As one such a possibility, assume that (14) holds, and that $x_j[k-1]$ and $x_p[k-1]$ both belong to the interval $[L_i, L_{i+1}]$ (note that the other cases can be studied in the same way). To show inequality (16), observe that:

$$x_j[k] = x_j[k-1] + \varepsilon_{k-1}(L_{i+1} - L_i) \leq \omega'_i \quad (17)$$

and:

$$x_p[k] = x_p[k-1] - \varepsilon_{k-1}(L_{i+1} - L_i) > \bar{L}_i \quad (18)$$

Therefore, if $\mu \leq \varepsilon_{k-1}$, then there is a reduction in the distance function by at least $\mu(L_{i+1} - L_i)$. If $\mu > \varepsilon_{k-1}$, then the distance function reduces by at least $\varepsilon_{k-1}(L_{i+1} - L_i)$. These observations lead to inequality (16). ■

The proof provided in [20] for the convergence of the SG algorithm relies on two facts: i) the state $X[k]$ belongs to a *finite-dimensional* space for every natural number k ii) there exists an *integer-valued* distance function with certain useful features. However, none of these properties are still maintained under the ASG algorithm. More specifically, $X[k]$ normally belongs to the infinite-dimensional set $[L_{\min}, L_{\max}]^\nu$, and the distance function $d_\varepsilon(X[\cdot], \mathcal{S}(\bar{L}_i))$ is real-valued. One implication of the second fact is that the distance function can be reduced infinitesimally, rather than by 1 in the integer case. To circumvent this hurdle, Part (iii) of Theorem 1 is provided to emphasize that under certain conditions, the distance function $d_\varepsilon(X[\cdot], \mathcal{S}(\bar{L}_i))$ reduces by a meaningful number (say, at least $\min(\mu, \varepsilon_{\min})(L_{i+1} - L_i)$).

Theorem 2: Run the ASG algorithm on the graph \mathcal{G} with the initial state X_0 . Regarding the asymptotic behavior of the state $X[k]$ relative to the splitting level \bar{L}_i ($\forall i \in \mathbf{Z}$), one of the following cases always takes place with probability 1:

- i) There exists a natural number k_0 such that:

$$X[k] \in (\bar{L}_i, L_{\max}]^\nu, \quad \forall k \geq k_0 \quad (19)$$

- ii) There exists a natural number k_0 such that:

$$X[k] \in [L_{\min}, \bar{L}_i]^\nu, \quad \forall k \geq k_0 \quad (20)$$

- iii) The state $X[k]$ asymptotically converges to the set $\mathcal{S}(\bar{L}_i)$, i.e. for every $\mu \in (0, 0.5 - \varepsilon_{\max})$, there exists a natural number k_0 such that:

$$X[k] \in (\omega_i - \mu(L_{i+1} - L_i), \omega'_i + \mu(L_{i+1} - L_i)]^\nu \quad (21)$$

for all $k \geq k_0$.

Proof: For a proof by contradiction, assume that none of the cases (i), (ii) and (iii) occurs for a given real number $\mu \in (0, 0.5 - \varepsilon_{\max})$. This, together with the connectivity of the graph, implies that at each time instant $k \in \mathbf{N}$, there exists a path j_1, j_2, \dots, j_p such that:

$$\begin{aligned} x_{j_1}[k] &\leq \bar{L}_i, \\ \bar{L}_i &< x_{j_2}[k], \dots, x_{j_{p-1}}[k] \leq \omega'_i + \mu(L_{i+1} - L_i), \\ \omega'_i + \mu(L_{i+1} - L_i) &< x_{j_p}[k] \end{aligned} \quad (22)$$

or:

$$\begin{aligned} x_{j_1}[k] &\leq \omega_i - \mu(L_{i+1} - L_i), \\ \omega_i - \mu(L_{i+1} - L_i) &< x_{j_2}[k], \dots, x_{j_{p-1}}[k] \leq \bar{L}_i, \\ \bar{L}_i &< x_{j_p}[k] \end{aligned} \quad (23)$$

As an example, a path satisfying the conditions given in (22) is illustrated in Figure 2. Since there may exist more than

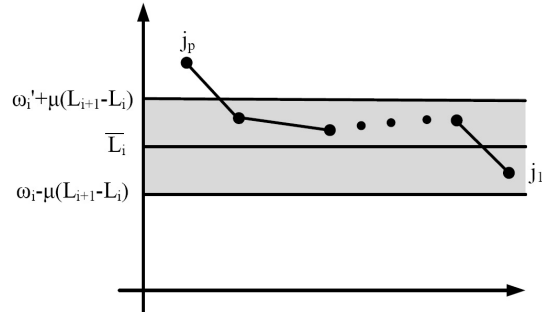


Fig. 2. The path shown in the figure satisfies the conditions given in (22).

one such path, find one whose length is minimum and then attribute it to the time instant k . Hence, an infinite sequence of paths, denoted by \mathcal{H} , will be produced. On the other hand, the graph \mathcal{G} has only a finite number of simple paths. Therefore, one can find a path in \mathcal{H} with minimum length that appears an infinite number of times. With a slight abuse of notation, assume that this path is $\rho_1, \rho_2, \dots, \rho_p$, and occurs at time instant $\{k_i\}_{i=1}^\infty$. With no loss of generality, assume that this path always corresponds to the set of relations (22) (as opposed to (23)). Now, partition the interval $[L_{\min}, L_{\max}]$ into a finite number of sub-intervals such that each of them has a length smaller than $0.5 \varepsilon_{\min}(L_{i+1} - L_i)$. This induces a finite grid structure on the box $[L_{\min}, L_{\max}]^\nu$. Denote the number of sub-boxes of $[L_{\min}, L_{\max}]^\nu$ resulting from this gridding with μ , and the sub-boxes themselves with T_1, T_2, \dots, T_μ . Note that every $X[k]$ belongs to one of the boxes T_1, T_2, \dots, T_μ (it may belong to more than one box if it sits on the boundary of a box). To geometrically visualize the idea behind this gridding, consider the plane in which the coordinated graph $\vec{\mathcal{G}}$ is depicted. Partition the part of the plane enclosed by the lines $y = L_{\min}$ and $y = L_{\max}$ with a number of horizontal lines so that every two neighboring lines be distant from each other by at most $0.5 \varepsilon_{\min}(L_{i+1} - L_i)$. This partitioning creates a finite number of strips such that each vertex of the graph $\vec{\mathcal{G}}(k)$ belongs to one of them. If there are ω of such strips, the vertices of the graph $\vec{\mathcal{G}}(k)$ can be placed in these strips in ω^ν ways. In other words, $\mu := \omega^\nu$ different configurations can be envisaged for the graph $\vec{\mathcal{G}}(k)$ relative to these strips. These different configurations are denoted by T_1, T_2, \dots, T_μ . Note that when two coordinated graphs, say $\vec{\mathcal{G}}(k)$ and $\vec{\mathcal{G}}(k')$, belong to the same configuration (sub-box), it means that every two corresponding vertices of the graphs lie in the same strip; i.e. the graphs somehow possess the same geometry up to the grid size.

Since each of the points $X[k_1], X[k_2], \dots$ belongs to one of the μ configurations T_1, T_2, \dots, T_μ , there exist a subsequence

of $\{k_j\}_1^\infty$, denoted by $\{\bar{k}_j\}_1^\infty$, and a number $\xi \in \{1, 2, \dots, \mu\}$ such that $X[\bar{k}_j] \in T_\xi$ for all $j \in \mathbf{N}$. Due to the fact that the statement of this theorem is related to the existence of a time instant k_0 or an asymptotic convergence with probability 1, any pathological case whose corresponding probability is 0 can be ignored. Indeed, one can assume that if the ASG algorithm is run, then the probability that *the size of the smallest path appearing infinitely many times in \mathcal{H} is p and there exists a subsequence $\{\bar{k}_j\}_1^\infty$ corresponding to this path for which $X[\bar{k}_j] \in T_\xi$ for all $j \in \mathbf{N}$ is nonzero*. By Theorem 4 in [22], there exists a subsequence of $\{\bar{k}_j\}_1^\infty$, denoted by $\{\tilde{k}_j\}_1^\infty$, such that the edge between the vertices ρ_1 and ρ_2 is always selected at times $\tilde{k}_1 + 1, \tilde{k}_2 + 1, \dots$. Since the set of relations (22) holds for $k = \tilde{k}_j, \forall j \in \mathbf{N}$, it is evident that:

$$x_{\rho_2}[\tilde{k}_j + 1] \leq x_{\rho_2}[\tilde{k}_j] - \varepsilon_{\min}(L_{i+1} - L_i), \quad j \in \mathbf{N} \quad (24)$$

This implies that $X[\tilde{k}_j + 1]$ no longer belongs to T_ξ (in light of the fact that the grid size was at most $0.5 \varepsilon_{\min}(L_{i+1} - L_i)$). Consequently, the graph $\tilde{\mathcal{G}}(\tilde{k}_j)$ loses its configuration T_ξ at time $\tilde{k}_j + 1$, but regain it at time \tilde{k}_{j+1} . It is desired to prove that this recurrence cannot take place infinitely many times corresponding to $j = 1, 2, \dots$. To this end, notice that in order for $x_{\rho_2}[\tilde{k}_j + 1]$ to return to its initial strip (corresponding to T_ξ), one of the following cases must happen during the time interval $[\tilde{k}_j + 1, \tilde{k}_{j+1}]$:

- a) There exist a time k and a vertex v_q such that:

$$x_{\rho_2}[k - 1] \leq x_{\rho_2}[\tilde{k}_j + 1], \quad x_q[k - 1] > \bar{L}_{i+1} \quad (25)$$

and that the edge (ρ_2, q) is selected at time k .

- b) Vertex ρ_2 goes below the line $y = \bar{L}_i$ at some time.

Therefore, one of the above situations (a) or (b) must occur an infinite number of times (as j changes from 1 to ∞). First, assume that case (a) occurs infinitely many times. It is easy to argue that every time that this case happens, the distance function $d(X[\cdot], \bar{L}_{i+1})$ reduces by at least $\min(0.5 - \varepsilon_{\max}, \varepsilon_{\min})(L_{i+2} - L_{i+1})$ (in light of Part (iii) of Theorem 1). This pushes the distance function $d(X[\cdot], \bar{L}_{i+1})$ towards $-\infty$, whereas it must always remain nonnegative. This contradiction rules out this case. Now, assume that case (b) occurs infinitely many times. Similar to the argument made above, one can argue that vertex ρ_p is always above the line $y = \omega'_j + \mu(L_{j+1} - L_j)$, except possibly at a finite number of times that could go below this line. Ignoring these times, one can assume with no loss of generality that vertex ρ_p is always above the line $y = \bar{L}_i + \varepsilon_{\max}(L_{i+1} - L_i)$. Hence, case (b) results that at some time instant, vertex ρ_2 goes below the line $y = \bar{L}_i$, which introduces an undesirable path satisfying a constraint of type (22), but with a length smaller than p . This means that there are infinite number of undesirable paths occurring in \mathcal{H} whose lengths are less than p . Thus, there is an undesirable path of length smaller than p which repeats infinite times. This contradicts the assumption that the minimum length of such a path is p . ■

The main result of the paper is given below, which proves the convergence of the ASG algorithm and, besides, characterizes the steady-state behavior of the state $X[k]$.

Theorem 3: Run the ASG algorithm on the graph \mathcal{G} with the initial state X_0 . One of the following cases occurs with probability 1:

- The state $X[k]$ asymptotically converges to the set $\mathcal{S}(\eta_1)$ as $k \rightarrow \infty$ (the asymptotic convergence is in the sense described in Theorem 2).
- The state $X[k]$ asymptotically converges to the set $\mathcal{S}(\eta_2)$ as $k \rightarrow \infty$.
- There exists a natural number k_0 such that:

$$X[k] \in (\eta_1, \eta_2]^\nu, \quad \forall k \geq k_0 \quad (26)$$

Proof: By considering $\bar{L}_i = \eta_2$ in Theorem 2, it readily follows from this theorem that one of the cases (i), (ii) or (iii) takes places for $\bar{L}_i = \eta_2$. Evidently, case (i) can be ruled out in light of the fact that the average of the entries of $X[k]$ is less than or equal to η_2 and therefore $X[k]$ cannot belong to $(\eta_2, L_{\max}]^\nu$. On the other hand, case (iii) in this theorem is indeed identical to case (b) in Theorem 3. Therefore, it remains to only assume that case (ii) in Theorem 2 takes place. In other words, suppose that there exists a natural number k_0 with probability 1 such that:

$$X[k] \in [L_{\min}, \eta_2]^\nu, \quad \forall k \geq k_0 \quad (27)$$

As before, one of the following three possibilities happens with probability 1 (in light of Theorem 2 for $\bar{L}_i = \eta_1$):

- There exists a natural number k'_0 such that $X[k] \in (\eta_1, L_{\max}]^\nu$ for all $k \geq k'_0$: This together with (27) proves case (c) of Theorem 3.
- There exists a natural number k'_0 such that $X[k] \in [L_{\min}, \eta_1]^\nu$ for all $k \geq k'_0$: Similar to the argument made earlier concerning the average of the entries of $X[k]$, one can easily argue that this case is impossible unless $X[k] = (\eta_1, \eta_1, \dots, \eta_1)$, which leads to case (a) in Theorem 3.
- The state $X[k]$ asymptotically converges to the set $\mathcal{S}(\eta_1)$: This case is the same as case (a) of Theorem 3. ■

Remark 1: The results of this paper are important extensions of those given in [20] in two aspects:

- The tuning parameter ε can be time-varying, which is advantageous in terms of both checking the numerical stability of the SG algorithm and introducing the freedom of choosing a state-dependent tuning factor (as explained in the preceding section).
- Even if ε is kept fixed throughout the entire run of the algorithm, our earlier paper [20] needs to assume that all quantization levels $\{L_i\}_\infty^\infty$ are integer multiples of a specific number (say, they are all integers) in order to confine the state space to a finite-dimensional space. Nonetheless, this constraint is removed in the current paper.

Remark 2: Unlike the case of picking a fixed ε , the convergence time under the ASG algorithm may be infinite

due to the phenomenon of asymptotic convergence to the quantized consensus. However, the time k_0 at which the state of the system is in the μ -vicinity of the consensus set is finite (see part (iii) of Theorem 2). To find the expected value of k_0 , one can adopt a technique similar to the one developed in [21], which leads to lower and upper bounds on this quantity corresponding to the worst initial state.

A. Oscillatory steady state

Consider the graph \mathcal{G} under the SG algorithm. Our recent paper [20] proved that the state $X[k]$ may take two different behaviors for large enough k 's as follows:

- i) $X[k]$ is an equilibrium point of the system that belongs to an invariant set of diameter 1 (say $(\eta_1, \eta_2)^\nu$).
- ii) $X[k]$ belongs to an invariant set of diameter 2ε (say $\mathcal{S}(\eta_1)$ or $\mathcal{S}(\eta_2)$), and the behavior of the state in this set is oscillatory with probability 1.

For the sake of refraining from unnecessary computations and communications, it is really preferred to reach an equilibrium point, rather than oscillate in a set. To increase the chance of ending up in case (i), one needs to pick a small ε . This has a detrimental effect on the time required for reaching the consensus. The ASG algorithm proposed in this paper resolves this issue efficiently. As a possible strategy, adjust the tuning factor ε_k as follows. Given the numbers $x_i[k]$ and $x_j[k]$ being updated at time $k+1$, if $q(x_i[k])$ and $q(x_j[k])$ are two neighboring quantization levels, then set $\varepsilon_k = \varepsilon_{\min}$; otherwise choose $\varepsilon_k = \varepsilon_{\max}$. Now, by letting ε_{\min} go to zero (when k goes to infinity), the state $X[k]$ will be an equilibrium point with a high probability. In other words, the ASG algorithm integrates the SG algorithm for different values of the fixed parameter ε , while it keeps the required convergence time in a reasonable range (roughly speaking, the convergence time in this case mainly depends on ε_{\max} , as opposed to ε_{\min} , due to the fact that most of the updates are performed by the factor ε_{\max}). The efficacy of this strategy will be illustrated in the next section.

IV. SIMULATION RESULTS

Consider a complete graph \mathcal{G} with 30 nodes, and assume that each node is associated with a real number that is generated using a Gaussian probability distribution with mean 0 and variance 100. Moreover, suppose that the quantization is uniform, i.e. L_i is equal to i for all $i \in \mathbf{Z}$. To compare the ASG algorithm with the SG algorithm, first let ε be fixed and equal to 0.5 at all times. The SG algorithm was run on 100 randomly generated initial values and it was observed that the expected value of the convergence time was 755.40. However, the steady-state behavior of the system was oscillatory for 34 samples of the initial state. As discussed earlier, one can pick a smaller ε to diminish the probability of ending up with an oscillatory steady-state behavior. As a result, pick $\varepsilon = 0.1$. The expected value of the convergence time was obtained as 2114.3, while the steady-state behavior was still oscillatory for 21 samples of the initial state. Now, run the ASG algorithm on the graph for the same 100

samples of the initial state. Tune the factor ε_k ($k \in \mathbf{N} \cup \{0\}$) according to a simple rule as follows:

$$\varepsilon_k = \begin{cases} 0.1 & \text{if } |q(x_j[k]) - q(x_i[k])| = 1 \\ 0.5 & \text{otherwise} \end{cases} \quad (28)$$

where (i, j) denotes the edge selected at time $k+1$. Hence, the ASG algorithm with this updating rule integrates the SG algorithm with $\varepsilon = 0.5$ and the SG algorithm with $\varepsilon = 0.1$. The expected value of the convergence time was obtained as 824.6 and, interestingly, the steady-state behavior was never oscillatory. This clearly shows the power of letting the tuning factor be time-dependent. Figures 1 and 2 depict the state of the system at time k_0 for some sample initial states under the SG algorithm with $\varepsilon = 0.5$ and the ASG algorithm with the above-mentioned tuning factor ε_k , respectively. Observe that the steady-state behavior of the system plotted in Figure 1 is oscillatory with probability 1, whereas the system plotted in Figure 2 has reached an equilibrium point. Furthermore, most of the points in Figure 2 are in the close vicinity of the average x_{ave} .

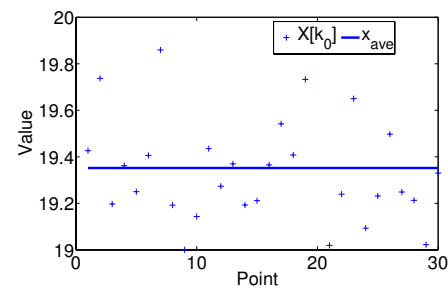


Fig. 3. The state of the system at time k_0 for some sample initial state under the SG algorithm with $\varepsilon = 0.5$.

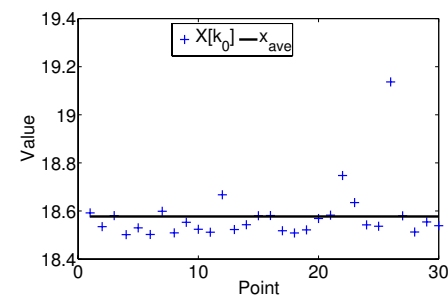


Fig. 4. The state of the system at time k_0 for some sample initial state under the ASG algorithm with ε_k specified in (28).

V. CONCLUSIONS

This paper tackles the average consensus problem over a connected weighted graph subject to a quantization constraint. It is assumed that each pair of vertices can be chosen with a certain probability to update their numbers in term of the quantized data being exchanged between them. It was shown in our recent paper that the quantized consensus is

reached under a certain updating algorithm, which utilizes a fixed tuning factor. The present paper generalizes the results to the case when this factor can be time-varying. There are two main motivations for this extension. First, it is important to study the numerical stability of the algorithm with a fixed tuning factor when this parameter is subject to numerical errors. Moreover, the asynchronous nature of the algorithm and some other technical reasons demand the adjustment of this tuning factor at each time update based on the previous state of the system. The convergence proof provided in the paper is technically far more complicated than the one proposed in our recent paper, because letting the tuning factor be time-varying converts a finite Markov chain to an infinite Markov chain with an uncountable state space.

ACKNOWLEDGMENT

This research was supported by ONR MURI N00014-08-1-0747 “Scalable, Data-driven, and Provably-correct Analysis of Networks,” ARO MURI W911NF-08-1-0233 “Tools for the Analysis and Design of Complex Multi-Scale Networks,” and the Army’s W911NF-09-D-0001 Institute for Collaborative Biotechnology.

REFERENCES

- [1] G. Tel, “Introduction to distributed algorithms,” Cambridge University Press, 2000.
- [2] N. A. Lynch, “Distributed algorithms,” Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1996.
- [3] R. Olfati-Saber and R. M. Murray, “Consensus problems in networks of agents with switching topology and time-delays,” *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520-1533, 2004.
- [4] R. Olfati-Saber, J. A. Fax and R. M. Murray, “Consensus and cooperation in networked multi-agent systems,” *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215-233, 2007.
- [5] D. P. Bertsekas and J. N. Tsitsiklis, “Parallel and distributed computation: Numerical methods,” Belmont, MA: Athena Scientific, 1997.
- [6] Y. Rabani, A. Sinclair and R. Wanka, “Local divergence of Markov chains and the analysis of iterative load-balancing schemes,” in *Proceedings of IEEE Conference on Foundations of Computer Science*, 1998.
- [7] Y. Kuramoto, “Chemical oscillators, waves, and turbulence,” Springer-Verlag, Berlin, 1984.
- [8] S. H. Strogatz, “Exploring complex networks,” *Nature*, vol. 410, pp. 268-276, 2001.
- [9] A. V. Savkin, “Coordinated collective motion of groups of autonomous mobile robots: Analysis of Vicsek’s model,” *IEEE Transactions on Automatic Control*, vol. 49, no. 6, pp. 981982, 2004.
- [10] R. Olfati-Saber, “Flocking for multi-agent dynamic systems: Algorithms and theory,” *IEEE Transactions on Automatic Control*, vol. 51, no. 3, pp. 401420, 2006.
- [11] A. Jadbabaie, J. Lin, and A. S. Morse, “Coordination of groups of mobile autonomous agents using nearest neighbor rules,” *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 9881001, 2003.
- [12] A. Speranzon, C. Fischione and K.H. Johansson, “Distributed and collaborative estimation over wireless sensor networks,” in *Proceedings of the 45th IEEE Conference on Decision and Control*, 2006.
- [13] R. Carli, F. Fagnani, A. Speranzon and S. Zampieri, “Communication constraints in the average consensus problem,” *Automatica*, vol. 44, no. 3, pp. 671-684, 2008.
- [14] J. Tsitsiklis, “Problems in decentralized decision making and computation,” PhD thesis, Dept. of Electrical Engineering and Computer Science, M.I.T., Boston, MA, 1984.
- [15] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, “Analysis and optimization of randomized gossip algorithms,” in *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2004.
- [16] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, “Randomized gossip algorithms,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508-2530, 2006.
- [17] F. Benezit, A. G. Dimakis, P. Thiran and M. Vetterli, “Gossip along the way: Order-optimal consensus through randomized path averaging,” in *Proceedings of the Allerton Conference on Communication, Control, and Computing*, 2007.
- [18] A. Kashyap, T. Basara and R. Srikanta, “Quantized consensus,” *Automatica*, vol. 43, no. 7, pp. 1192-1203, 2007.
- [19] P. Frasca, R. Carli, F. Fagnani and S. Zampieri, “Average consensus by gossip algorithms with quantized communication,” in *Proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, 2008.
- [20] J. Lavaei and R. M. Murray, “On quantized consensus by means of gossip algorithm—Part I: convergence proof,” in *Proceedings of the 2009 American Control Conference*, St. Louis, MO, 2009.
- [21] J. Lavaei and R. M. Murray, “On quantized consensus by means of gossip algorithm—Part II: convergence time,” in *Proceedings of the 2009 American Control Conference*, St. Louis, MO, 2009.
- [22] J. Lavaei and R. M. Murray, “Quantized consensus via adaptive stochastic gossip algorithm,” *Technical Report*, California Institute of Technology, 2009 (available at www.cds.caltech.edu/~lavaei/Tech_rep1_2009.pdf).