Rates of Convergence for Greedy Gossip with Eavesdropping

Deniz Üstebay, Boris Oreshkin, Mark Coates, and Michael Rabbat
Department of Electrical and Computer Engineering
McGill University
3480 University St, Montréal, Québec, Canada
Email: {deniz.ustebay,boris.oreshkin}@mail.mcgill.ca, {mark.coates,michael.rabbat}@mcgill.ca

***Abstract*— Greedy gossip with eavesdropping (GGE) is a randomized gossip algorithm that exploits the broadcast nature of wireless communications to converge rapidly on grid-like network topologies without requiring that nodes know their geographic locations. When a node decides to gossip, rather than choosing one of its neighbors randomly, it greedily chooses to gossip with the neighbor whose values are most different from its own. We assume that all transmissions are wireless broadcasts so that nodes can keep track of their neighbors' values by eavesdropping on their communications. We have previously proved that GGE converges to the average consensus on connected network topologies. In this paper we study the rate of convergence of GGE, a non-trivial task due to the greedy, data-driven nature of the algorithm. We demonstrate that GGE outperforms standard randomized gossip, and we characterize the rate of convergence in terms of a topology-dependent constant analogous to the second-largest eigenvalue characterization for previous randomized gossip algorithms. Simulations demonstrate that the convergence rate of GGE is superior to existing average consensus algorithms such as geographic gossip.**

I. Introduction and Background

Distributed consensus or agreement has been identified as a canonical problem in both the distributed signal processing and control communities (see, e.g., [1]–[6] and references therein), tracing back to the seminal work of Tsitsiklis [7]. The prototypical example of a consensus problem is that of computing the *average consensus*: initially, each node in a network of n nodes has a scalar piece of information, y_i, and the goal is to compute the average, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, at every node in the network. Consensus can be viewed as a sort of synchronization or agreement, before the network makes a concerted action.

In previous work, we proposed a new average consensus algorithm, *greedy gossip with eavesdropping* (GGE), that takes advantage of the broadcast nature of wireless communications to accelerate convergence [8]. Motivated by applications in wireless sensor-actuator networks and vehicular networks, we assume the network is composed of battery-powered nodes, communicating via wireless radios. We assume a broadcast model where all neighbors within range of a transmitting node successfully receive the message. In contrast to previous randomized gossip algorithms which perform updates completely at random, each GGE update is performed in a greedy, myopic fashion. In particular, activated nodes do not choose a neighbor to gossip with randomly. Rather, each node keeps track of its own value *and* its neighbors' values, and when it comes time to gossip, it greedily chooses to gossip with the neighbor that is most different from itself. Because nodes broadcast their transmissions, it is easy to track the values of neighboring nodes by eavesdropping on their transmissions. Moreover, accelerating gossip in this fashion does not require that nodes have any information about geographic locations.

Although the basic idea behind GGE is straightforward, analyzing its convergence behavior is non-trivial. In particular, each GGE update depends explicitly on the values at each node (via the greedy decision of with which neighbor to gossip). Thus, the standard approach to quantifying rates of convergence (i.e., examining the mixing time of a related Markov chain) does not apply. In our previous work, [8], we proved that GGE converges to the average consensus by demonstrating that GGE can be viewed as a particular instance of an incremental subgradient algorithm. We then experimentally characterized the rate of convergence and communication complexity of GGE via simulation. The current article extends this line of research by beginning to develop a theory for the rate of convergence of GGE. In particular, this paper makes the following contributions: 1) We develop a bound relating the rate of convergence of GGE to that of standard randomized gossip. Not surprisingly, the bound suggests that GGE always converges faster than randomized gossip. More interesting, though, is that this bound provides insight as to what is the worst case scenario for GGE. 2) We develop a worst-case bound on the rate of convergence of GGE. Similar to results for other gossip algorithms that characterize the rate of convergence as a function of the second largest eigenvalue of a related stochastic matrix, our bound characterizes the rate of convergence of GGE in terms of a constant that is strictly topology dependent. We investigate the behavior of this constant empirically for random geometric graph topologies, and find that, in terms of both rate of convergence and communication complexity, GGE performs at least as well as other fast gossip algorithms such as geographic gossip.

A. Background and Related Work

The two most widely studied algorithms for solving the average consensus problem are *distributed averaging* [9] and *randomized gossip* [10]. In distributed averaging, every node broadcasts information to its neighbors at every iteration. Let $x_i(k)$ denote the value at node i after the kth iteration. Each node i initializes its value to $x_i(0) = y_i$. At the kth iteration, after node i receives values $x_j(k-1)$ from each of its neighbors, it replaces $x_i(k)$ with a weighted average of its own previous value and its neighbors' values. Under appropriate conditions on the weights used in the update step, one can show that the values $x_i(k)$ at every node converge to the average $\bar{y}$ as $k \to \infty$ [9]. However, information diffuses slowly across the network in this scheme, and since the information at each node typically does not change much from iteration to iteration, this is not efficient use of the broadcast medium.

Randomized gossip operates at the opposite extreme, where only two neighboring nodes exchange information at each iteration. At the kth iteration, a node s is chosen uniformly at random; it chooses a neighbor, t, randomly; and this pair of nodes "gossips": s and t exchange values and perform the update $x_s(k) = x_t(k) = (x_s(k-1)+x_t(k-1))/2$, and all other nodes remain unchanged. Again, one can show that under very mild conditions on the way a random neighbor, t, is drawn, the values $x_i(k)$ converge to $\bar{y}$ at every node [9]. Although other neighbors overhear the messages exchanged between the active pair of nodes, they do not make use of this information in existing randomized gossip algorithms. The fact that nodes only exchange information with their immediate neighbors is attractive, from the point of view of simplicity and robustness to changing topologies and/or network conditions. However it also means that in typical wireless network topologies (grids or random geometric graphs [11]), information diffuses slowly across the network. Boyd et al. [10] prove that for random geometric graphs, randomized gossip requires $O(n^2)$ transmissions to approximate the average consensus well[1].

Slow convergence of randomized gossip motivated Dimakis et al. to develop geographic gossip. Assuming each node knows its geographic location and the locations of its neighbors, information can be exchanged with nodes beyond immediate neighbors. In [6], they show that these long-range transmissions improve the rate of convergence from $O(n^2)$ to roughly $O(n^{3/2})$ transmissions. Although geographic gossip is a significant improvement over randomized gossip in terms of number of transmissions, it comes at the cost of increased complexity, since the network must now provide reliable two-way transmission over many hops. Messages which are lost in transit potentially result in biasing the average consensus computation.

Since the proposal of geographic gossip, other fast gossiping algorithms have been proposed. Most related is the work of Li and Dai [12], and Jung et al. [13]. Both approaches are based on using the geographic locations of nodes to construct *lifted* Markov chains that direct the exchange of information across the network. Benezit et al. have also proposed averaging along paths as an extension to geographic gossip that converges in $O(n)$ communication complexity [14]. All of these approaches rely on geographic information and thus are not suitable to scenarios where nodes are mobile or location information is not available. The focus of the current article is on developing a fast, communication-efficient algorithm that exploits broadcast communications rather than geographic location information to gossip quickly.

In a series of recent papers, Aysal et al. propose *broadcast gossip*, a consensus algorithm that also makes use of the broadcast nature of wireless networks [15], [16]. At each iteration, a node is chosen uniformly at random to broadcast its value. The nodes in the broadcast range of this node calculate a weighted average of their own value and the broadcasted value, and they update their value with this weighted average. In broadcast gossip, the value of the broadcasting node is independently incorporated at each neighbor. Consequently, broadcast gossip does not preserve the network average at each iteration. In this manner, broadcast gossip achieves a low variance (i.e., rapid convergence), but introduces bias: the value to which broadcast gossip converges can be significantly different from the true average (see [8] for further discussion).

Sundhar Ram et al. have also recently proposed a general class of incremental subgradient algorithms for distributed optimization [17]. The focus of their study is on understanding the effects of stochastic errors (e.g., due to quantization) on convergence of consensus-like distributed optimization algorithms. They determine conditions on the errors that guarantee convergence of the algorithm, but do not characterize convergence rates. Nedić and Ozdaglar have also proposed a distributed form of incremental subgradient optimization that generalizes the consensus framework [18]. Their problem formulation is much more general than ours, but for the specific formulation addressed in this paper, we achieve stronger results. By exploiting the form of our cost function, we are able to guarantee convergence to an optimal solution and obtain tight bounds on the rate of convergence in terms of the network topology.

B. Paper Organization

The remainder of this paper is organized as follows. In Section II we review the formal definition of the algorithm, as outlined in [8]. In Section III, we derive a bound relating the performance of GGE to randomized gossip, which suggests that GGE always outperforms randomized gossip. In Section IV, we present a worst-case upper bound on the rate of convergence of GGE in terms of a topology-dependent constant. Results from numerical simulations are presented

[1] Throughout this paper, when we refer to randomized gossip, we specifically mean the natural random walk version of the algorithm, where the node t_k is chosen uniformly from the set of neighbors at each iteration. For random geometric graph topologies, which are of most interest to us, Boyd et al. [10] prove that the performance of the natural algorithm scales identically to that of the optimal choice of transition probabilities, so there is no loss in generality.

in Section V and Section VI summarizes the contributions of the paper.

II. Greedy Gossip with Eavesdropping (GGE)

We consider a network of n nodes, and represent network connectivity as a graph, $G = (V, E)$, with vertices $V = \{1, \ldots, n\}$, and edge set $E \subset V \times V$ such that $(i, j) \in E$ if and only if nodes i and j directly communicate. We assume that communication relationships are symmetric and that the graph is connected. Let $\mathcal{N}_i = \{j : (i, j) \in E\}$ denote the set of neighbors of node i (not including i itself). Each node in the network has an initial value y_i, and the goal of the gossip algorithm is to use only local, broadcast exchanges to converge towards a state where every node can calculate the average $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$. To initialize the algorithm, each node sets its gossip value to $x_i(0) = y_i$, and broadcasts this value to all of its immediate neighbors.

At the kth iteration of GGE, a node s_k is chosen uniformly at random from $\{1, \ldots, n\}$. This can be accomplished using the asynchronous time model described in [19], where each node "ticks" according to a Poisson clock with rate 1. In the randomized gossip algorithms described in [10], s_k randomly chooses a neighbor to gossip with. In the GGE algorithm, s_k gossips with a neighbor that is currently the most different from its own value. This choice is possible because each node i maintains not only its own local variable, $x_i(k-1)$, but also a copy of the most recent values at its neighbors, $x_j(k-1)$, for $j \in \mathcal{N}_i$. More formally, s_k identifies a node t_k satisfying

$$t_k \in \arg\max_{t \in \mathcal{N}_j} \left\{ \frac{1}{2}(x_{s_k}(k-1) - x_t(k-1))^2 \right\}.$$

When s_k has multiple neighbors that are all equally (and maximally) different from s_k, it chooses one of these neighbors at random. Then s_k and t_k exchange values and perform the update

$$x_{s_k}(k) = x_{t_k}(k) = \frac{1}{2}\big(x_{s_k}(k-1) + x_{t_k}(k-1)\big), \qquad (1)$$

while all other nodes $i \notin \{s_k, t_k\}$ maintain their values at $x_i(k) = x_i(k-1)$. Finally, the two nodes, s_k and t_k, broadcast these new values so that their neighbors have up-to-date information. This can be accomplished in two transmissions: s_k calculates its new value and broadcasts it, identifying t_k as the exchange partner; t_k broadcasts its new value so all of its neighbours are aware of the update.

GGE updates can also be expressed in the form

$$x(k) = W^{GGE}(k)x(k-1)$$

where $W^{GGE}(k)$ is a stochastic matrix with $W^{GGE}_{s_k,s_k}(k) = W^{GGE}_{s_k,t_k}(k) = W^{GGE}_{t_k,s_k}(k) = W^{GGE}_{t_k,t_k}(k) = \frac{1}{2}$, $W^{GGE}_{i,i}(k) = 1$ for all $i \notin \{s_k, t_k\}$, and 0 elsewhere.

A. GGE as an Incremental Subgradient Method

Since we make use of it extensively in deriving bounds on convergence performance, we now review, from [8], the interpretation of GGE as a randomized incremental subgradient[2] method [20]. First consider a constrained optimization problem of the form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^n f_i(x) \\ \text{subject to} \quad & x \in X, \end{aligned}$$

where each $f_i(x)$ is a convex function, but not necessarily differentiable, and X is a non-empty convex subset of $\mathbb{R}^n$. An incremental subgradient algorithm for solving this optimization is an iterative algorithm of the form:

$$x(k) = \mathcal{P}_X[x(k-1) - \alpha_k g(s_k, x(k-1))], \qquad (2)$$

where $\alpha_k > 0$ is the step-size, $g(s_k, x(k-1))$ is a subgradient of f_{s_k} at $x(k-1)$, and $\mathcal{P}_X[\cdot]$ projects its argument onto the set X. The algorithm is randomized when the component updated at each iteration, s_k, is drawn uniformly at random from the set $\{1, \ldots, n\}$, and is independent of $x(k-1)$. The projection, $\mathcal{P}_X[\cdot]$, ensures that each new iterate $x(k)$ is feasible. Under mild conditions on the sequence of step sizes, α_k, and on the regularity of each component function $f_i(x)$, Nedić and Bertsekas have shown that the randomized incremental subgradient method described above converges to a neighborhood of the global minimizer [20].

GGE is a randomized incremental subgradient algorithm for the problem

$$\min_{x \in \mathbb{R}^n} \quad \sum_{i=1}^n \max_{j \in \mathcal{N}_i} \left\{ \frac{1}{2}(x_i - x_j)^2 \right\} \qquad (3)$$

$$\text{subject to} \quad \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \qquad (4)$$

where y_i is the initial value at node i. The objective function in (3) has a minimum value of 0 which is attained when $x_i = x_j$ for all i, j. Thus, any minimizer is a consensus solution. Moreover, the constraint $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ ensures that the unique global minimizer is the average consensus.

To connect the GGE update, (1), and the incremental subgradient update, (2), let us define $g(k)$ such that

$$g_i(k) = \begin{cases} x_{s_k}(k-1) - x_{t_k}(k-1) & \text{for } i = s_k, \\ -(x_{s_k}(k-1) - x_{t_k}(k-1)) & \text{for } i = t_k, \\ 0 & \text{otherwise.} \end{cases} \qquad (5)$$

Here subscripts denote components of the vector $g(k)$. It is easy to verify that $g(k)$ is a subgradient of the function $f_{s_k}(x(k-1)) = \max_{j \in \mathcal{N}_{s_k}} \{\frac{1}{2}(x_{s_k}(k-1) - x_j(k-1))^2\}$. Employing a constant step size $\alpha_k = \frac{1}{2}$ and this subgradient,

[2]Subgradients generalize the notion of a gradient for non-smooth functions. The subgradient of a convex function f_i at x is any vector g that satisfies $f_i(y) \geq f_i(x) + g^T(y - x)$. The set of subgradients of f_i at x is referred to as the *subdifferential* and is denoted by $\partial f_i(x)$. If f_i is continuous at x, then $\partial f_i(x) = \{\nabla f_i(x)\}$; i.e., the only subgradient of f_i at x is the gradient. A sufficient and necessary condition for x^* to be a minimizer of the convex function f_i is that $0 \in \partial f_i(x^*)$. See [20] and references therein.

the update (2) is identical to (1). The recursive update for GGE thus has the form

$$x(k) = x(k-1) - \frac{1}{2}g(k), \tag{6}$$

Note that the projection is unnecessary, because this choice of subgradient and α_k ensure that the constraint $\sum_{i=1}^{n} x_i(k) = \sum_{i=1}^{n} y_i$ is satisfied at each iteration. With this formulation, we can derive a simple recursive relationship relating the squared error at iteration k to that at iteration $k-1$ [8]:

$$\begin{aligned}\|x(k) - \bar{x}\|^2 &= \|x(k-1) - \frac{1}{2}g(k) - \bar{x}\|^2 \\ &= \|x(k-1) - \bar{x}\|^2 - \langle x(k-1) - \bar{x},\, g(k)\rangle + \frac{1}{4}\|g(k)\|^2 \\ &= \|x(k-1) - \bar{x}\|^2 - \frac{1}{4}\|g(k)\|^2. \end{aligned} \tag{7}$$

We made use of this result to prove the convergence theorem in [8], and we will make further use of it in Section IV for deriving rate of convergence results.

III. Convergence Rate: GGE vs. Randomized Gossip

When we first proposed the GGE algorithm in [8], we were only able to characterize the convergence behaviour by demonstrating that GGE converges almost surely to the consensus value, as stated in the following theorem.

Theorem 1 (Üstebay et al. [8]): Let $x(k)$ denote the sequence of iterates produced by GGE. Then $x(k)$ converges to $\bar{x}$ almost surely as k tends to infinity.

In this section and the next, our aim is to provide a more complete description of convergence behaviour by bounding the rate of convergence.

The following theorem establishes a general expression for the bound on the mean-squared error of GGE after k iterations. Moreover, it demonstrates that the upper bound on the MSE of GGE is less than or equal to the upper bound on the MSE of randomized gossip. Recall from the discussion in Section II and [10] that the update from the $(k-1)$-th to k-th gossip iteration can be expressed as a linear recursion $x(k) = W(k)x(k-1)$, where $W(k)$ depends on the nodes s_k and t_k that gossip during iteration k. We denote the application of k successive randomized gossip updates by $W^{RG}(1:k) = \prod_{j=1}^{k} W^{RG}(j)$. Likewise, let $W^{GGE}(1:k) = \prod_{j=1}^{k} W^{GGE}(j)$ denote the successive application of k GGE updates. Let $\overline{W} = \mathbb{E}[W^{RG}(k)]$ denote the expected value of the randomized gossip matrix, and let $\lambda_2(\overline{W})$ denote the second largest eigenvalue of $\overline{W}$.

Theorem 2: Let the algorithm input, $x(0)$, be given, and let $\bar{x}$ denote the corresponding average consensus vector. After k iterations, the expected mean squared error of GGE is upper bounded as follows:

$$\begin{aligned}&\mathbb{E}\left[\|W^{GGE}(1:k)x(0) - \bar{x}\|^2\right] \\ &\qquad \le \|x(0) - \bar{x}\|^2 \prod_{i=1}^{k} \left(\lambda_2\left(\overline{W}\right) - \xi_i\right)\end{aligned} \tag{8}$$

where $\xi_k = 0$ if $\mathbb{E}[\|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2] = 0$, and otherwise,

$$\begin{aligned}\xi_k = &\frac{\sum_{i=1}^{n}\left(\max_{t\in\mathcal{N}_i}\left(x_i(k-1) - x_t(k-1)\right)^2\right)}{2n\ \mathbb{E}[\|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2]} \\ &- \frac{\sum_{i=1}^{n}\left(\frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i}\left(x_i(k-1) - x_j(k-1)\right)^2\right)}{2n\ \mathbb{E}[\|W^{GGE}(1:k-1)x(0) - \bar{x}\|^2]} \\ \ge &\ 0,\end{aligned} \tag{9}$$

where $x(k) = W^{GGE}(1:k)x(0)$.

Remark 1: The analogous expression for randomized gossip is simply [10]:

$$\mathbb{E}[\|W^{RG}(1:k)x(0) - \bar{x}\|^2] \le \|x(0) - \bar{x}\|^2 \lambda_2(\overline{W})^k.$$

(Note that here, the expectation is taken with respect to both random nodes chosen at each iteration, s_k and t_k, whereas in the expressions in the theorem, the only randomness is in s_k.) Since $\xi_i \ge 0$ for all $i = 1, \ldots, k$, this implies that the upper bound on GGE is uniformly upper bounded by the upper bound for randomized gossip, for any $k \ge 0$ and any input $x(0)$. The upper bound for random gossip is tight; if v_2 denotes the eigenvector corresponding to the second-largest eigenvalue of $\overline{W}$, then if $x(0) = cv_2$ for some constant c, the upper bound holds with equality (in expectation).

Remark 2: The form of the terms ξ_k also provides insight into which scenarios are less favorable for GGE. In general, we know that randomized gossip is slow to converge on random geometric graphs [10], and so we hope that $\xi_k > 0$ so that GGE achieves some improvement. Note that the magnitude of ξ_k is essentially its numerator, which measures how much larger (on average) a GGE step from $x(k-1)$ is in comparison to the step taken by randomized gossip from the same location. There are two scenarios where the expression for ξ_k in (9) evaluates to 0. The first is when $x(k-1) = \bar{x}$, in which case a consensus has already been achieved. The second, more interesting case is when the difference between any two neighbors is constant across the network; i.e., $(x_i - x_j)^2 = c$ for all $j \in \mathcal{N}_i$ and all $i = 1, \ldots, n$. In this setting, being greedy does not provide any gain, since gossiping with any neighbor provides the same amount of immediate improvement. Within the class of such "constant difference" vectors, $x(k)$, that satisfy $\sum_{i=1}^{n} x_i(k) = \sum_{i=1}^{n} x_i(0)$, the most challenging one is the one chosen to maximize $\|x(k) - \bar{x}\|^2$. We will revisit this scenario later in the numerical simulations presented in Section V and see that, indeed, this appears to be the worst-case scenario for GGE.

Proof: [Proof of Theorem 2] We recall the known convergence rate bounds for randomized gossip [10]:

$$\mathbb{E}[||W^{RG}(1:k)x(0) - \bar{x}||^2] \le \lambda_2\left(\overline{W}\right)^k ||x(0) - \bar{x}||^2 \tag{10}$$

and the related recursive relationship:

$$
\begin{aligned}
&\mathbb{E}[||W^{RG}(1:k)x(0)-\bar{x}||^2] \\
&\quad = \mathbb{E}[||W^{RG}(1:k-1)x(0)-\bar{x}||^2] \\
&\qquad - \frac{1}{2}\frac{1}{n}\sum_{t_k=1}^{n}\frac{1}{|\mathcal{N}_{t_k}|}\sum_{s_k\in\mathcal{N}_{t_k}}\left(x_{s_k}(k-1)-x_{t_k}(k-1)\right)^2 \\
&\quad \leq \lambda_2\left(\overline{W}\right)\mathbb{E}[||W^{RG}(1:k-1)x(0)-\bar{x}||^2] \qquad (11)
\end{aligned}
$$

We can identify an equivalent relationship derived from applying $k-1$ steps of GGE followed by one step of random gossip:

$$
\begin{aligned}
&\mathbb{E}[||W^{RG}(k)W^{GGE}(1:k-1)x(0)-\bar{x}||^2] \\
&\quad = \mathbb{E}[||W^{GGE}(1:k-1)x(0)-\bar{x}||^2] \\
&\qquad - \frac{1}{2}\frac{1}{n}\sum_{t_k=1}^{n}\frac{1}{|\mathcal{N}_{t_k}|}\sum_{s_k\in\mathcal{N}_{t_k}}\left(x_{s_k}(k-1)-x_{t_k}(k-1)\right)^2 \\
&\quad \leq \lambda_2\left(\overline{W}\right)\mathbb{E}[||W^{GGE}(1:k-1)x(0)-\bar{x}||^2]. \qquad (12)
\end{aligned}
$$

With this relationship in hand, we can bound the error of the GGE algorithm by adding and subtracting the effects of making the k-th step a randomized gossip update:

$$
\begin{aligned}
&\mathbb{E}[||W^{GGE}(1:k)x(0)-\bar{x})||^2] \\
&\quad = \mathbb{E}[||W^{GGE}(1:k-1)x(0)-\bar{x})||^2] \\
&\qquad - \frac{1}{2}\frac{1}{n}\sum_{t_k=1}^{n}\frac{1}{|\mathcal{N}_{t_k}|}\sum_{s_k\in\mathcal{N}_{t_k}}\left(x_{s_k}(k-1)-x_{t_k}(k-1)\right)^2 \\
&\qquad - \frac{1}{2}\frac{1}{n}\sum_{t_k=1}^{n}\max_{s_k\in\mathcal{N}_{t_k}}\left(x_{s_k}(k-1)-x_{t_k}(k-1)\right)^2 \\
&\qquad + \frac{1}{2}\frac{1}{n}\sum_{t_k=1}^{n}\frac{1}{|\mathcal{N}_{t_k}|}\sum_{s_k\in\mathcal{N}_{t_k}}\left(x_{s_k}(k-1)-x_{t_k}(k-1)\right)^2 \\
&\quad \leq \left[\lambda_2\left(\overline{W}\right)-\xi_k\right]\mathbb{E}[||W^{GGE}(1:k-1)x(0)-\bar{x}||^2]. \qquad (13)
\end{aligned}
$$

Repeated application of this inequality from $i=1,\ldots,k$ yields the bound (8). ■

IV. GGE Convergence Rate: Worst Case Bound

The previous section related the performance of GGE to that of standard randomized gossip. In this section, we seek a more direct characterization of the GGE rate of convergence in terms of properties of the underlying communication topology. The rate of convergence for gossip algorithms is typically quantified in terms of the ϵ-averaging time,

$$
T_{ave}(\epsilon) = \sup_{x(0)\neq 0}\inf\left\{k\ :\ \Pr\left(\frac{\|x(k)-\bar{x}\|}{\|x(0)-\bar{x}\|}\geq\epsilon\right)\leq\epsilon\right\}.
$$

Other gossip algorithms such as randomized gossip and geographic gossip are easily related to a homogeneous Markov chain. If the probability transition matrix of this chain is $\overline{W}$, then $T_{ave}(\epsilon)$ can be shown to scale as a function of the second largest eigenvalue of $\overline{W}$ [10]. In particular, $T_{ave}(\epsilon)\leq\frac{3\log\epsilon^{-1}}{\log\lambda_2(\overline{W})^{-1}}$. For randomized gossip, the matrix $\overline{W}$ depends on the choice of probabilities assigned to each edge in the network and hence, indirectly depends on the network topology.

Since the greedy decision made in each iteration of GGE depends on the gossip values at each node, $x(k)$, our algorithm cannot be related back to a homogeneous Markov chain. Consequently, the same machinery cannot be used to characterize the rate of convergence for GGE. The goal of this section is to bound the rate of convergence of GGE through alternative means. To this end, our main result is the following.

Theorem 3: Let $G=(V,E)$ denote the graph on which we are gossiping, let $x(k)$ denote the vector of GGE values after k iterations, and let $\bar{x}$ denote the average vector. Then

$$
\mathbb{E}[\|x(k)-\bar{x}\|^2] \quad\leq\quad A(G)^k\|x(0)-\bar{x}\|^2,
$$

where $A(G)$ is the graph-dependent constant defined as

$$
A(G) \quad=\quad \max_{x\neq\bar{x}}\frac{1}{|V|}\sum_{v\in V}\left(1-\frac{\|g_v(x)\|^2}{4\|x-\bar{x}\|^2}\right),
$$

where $g_v(x)$ refers to a subgradient of $f_v(x)$, when viewing GGE as an incremental subgradient algorithm[3]. Moreover, the ϵ-averaging time for GGE is bounded above by

$$
T_{ave}(\epsilon) \quad\leq\quad \frac{3\log\epsilon^{-1}}{\log A(G)^{-1}}.
$$

Remark 3: Note that the constant $A(G)$ only depends on the topology of the graph. This constant plays a role for GGE similar to that played by the second-largest eigenvalue of W for regular gossip algorithms.

Proof: [Proof of Theorem 3] The proof of the first part of Theorem 3 is based on an approach introduced in [21], and developed in [22] for analyzing data-adaptive algorithms. We begin by recalling the recursion for the mean squared error of GGE after k iterations expressed in (7):

$$
\begin{aligned}
\|x(k)-\bar{x}\|^2 &= \|x(k-1)-\bar{x}\|^2-\tfrac{1}{4}\|g(k)\|^2 \\
&= \left(1-\tfrac{\|g(k)\|^2}{4\|x(k-1)-\bar{x}\|^2}\right)\|x(k-1)-\bar{x}\|^2,
\end{aligned}
$$

where $g(k)$ denotes the subgradient at iteration k (when viewing GGE as a randomized incremental subgradient algorithm), and is a random quantity, depending on which node $s(k)$ is activated at iteration k. Let $M(k)=\|x(k)-\bar{x}\|^2$ denote the error after k iterations, and let $N(k)=1-\frac{\|g(k)\|^2}{4\|x(k-1)-\bar{x}\|^2}$ denote the amount of contraction at iteration k. Using these definitions and some successive conditioning, we get

$$
\begin{aligned}
\mathbb{E}[M(k)] &= \mathbb{E}[N(k)M(k-1)] \\
&= \mathbb{E}[\mathbb{E}[N(k)M(k-1)|x(k-1)]] \\
&= \mathbb{E}[M(k-1)\mathbb{E}[N(k)|x(k-1)]] \\
&\ \ \vdots \\
&= M(0)\mathbb{E}[\mathbb{E}[N(1)|x(0)]\cdots\mathbb{E}[N(k)|x(k-1)]].
\end{aligned}
$$

[3]We explicitly note that this constant is a function of the underlying topology by writing $A(G)$, and $A(G)$ is completely determined by the neighbourhood structure of the network because the maximization is over all x, and for a fixed x, the subgradients are determined by the neighbourhood structure.

Note that $A(G)$ is defined in such a way that $\mathbb{E}[N(k)|x(k-1)] \leq A(G)$ for all k. Therefore, it follows that

$$\mathbb{E}[\|x(k) - \bar{x}\|^2] \quad \leq \quad A(G)^k \|x(0) - \bar{x}\|^2.$$

Next, we prove the second part of the claim: the bound on ϵ-averaging time. To do this, we will use the bound we have just derived to develop an upper bound on $\Pr(\|x(k)-\bar{x}\| \geq \epsilon\|x(0) - \bar{x}\|)$, the probability that after k iterations we are still more than a factor of ϵ away from the initial error. By applying Markov's inequality and the bound we just derived for $\mathbb{E}[\|x(k) - \bar{x}\|^2]$, we have

$$\begin{aligned}
&\Pr(\|x(k) - \bar{x}\| \geq \epsilon\|x(0) - \bar{x}\|) \\
&\quad = \Pr\left(\|x(k) - \bar{x}\|^2 \geq \epsilon^2\|x(0) - \bar{x}\|^2\right) && (14) \\
&\quad \leq \frac{\mathbb{E}[\|x(k) - \bar{x}\|^2]}{\epsilon^2\|x(0) - \bar{x}\|^2} && (15) \\
&\quad \leq \epsilon^{-2}A(G)^k. && (16)
\end{aligned}$$

To get an upper bound on $T_{ave}(\epsilon)$, first note that $\Pr(\|x(k)-\bar{x}\| \geq \epsilon\|x(0) - \bar{x}\|) \leq \epsilon$ provided that $k \geq \frac{3\log\epsilon^{-1}}{\log A(G)^{-1}}$. Since in the first part of our proposition, the bound on $\mathbb{E}[\|x(k) - \bar{x}\|^2]$ is based on a worst-case one-step analysis, it is an upper bound on the mean squared error at iteration k, effectively a lower bound on the rate of convergence. Therefore, we only have an upper bound on the ϵ-averaging time for GGE; that is $T_{ave}(\epsilon) \leq \frac{3\log\epsilon^{-1}}{\log A(G)^{-1}}$. ■

Theorem 3 provides a direct link between the rate of convergence of GGE and the underlying network topology through the constant, $A(G)$. This motivates further study of how $A(G)$ scales for different classes of network topologies (e.g., random geometric graphs). Theoretically characterizing how $A(G)$ scales is a topic of ongoing research. The following section provides numerical simulations to support the results presented above. Comparisons are provided to other randomized gossip algorithms, and the scaling behavior of $A(G)$ is investigated via simulation.

V. Numerical Simulations

In this section we report the results of simulations conducted to compare the performance of GGE with randomized gossip [10] and geographic gossip [6] for a variety of state value initializations. We also compare the empirically achieved convergence rates to the bound established in Section IV and investigate how this bound behaves as the number of nodes in the network grows.

In our experiments, we focus on a random geometric graph, constructed by distributing nodes uniformly at random over the unit square. The transmission radius is set to $\sqrt{2\log n/n}$ such that the random geometric graph is connected with high probability. This topology is a good model for many wireless networks, including sensor networks and (snapshots of) vehicular networks, which we consider to be two of the most promising application domains for gossip algorithms [11]. In other simulation experiments with different topologies, we observed similar comparative behaviour, so we do not report the results here.

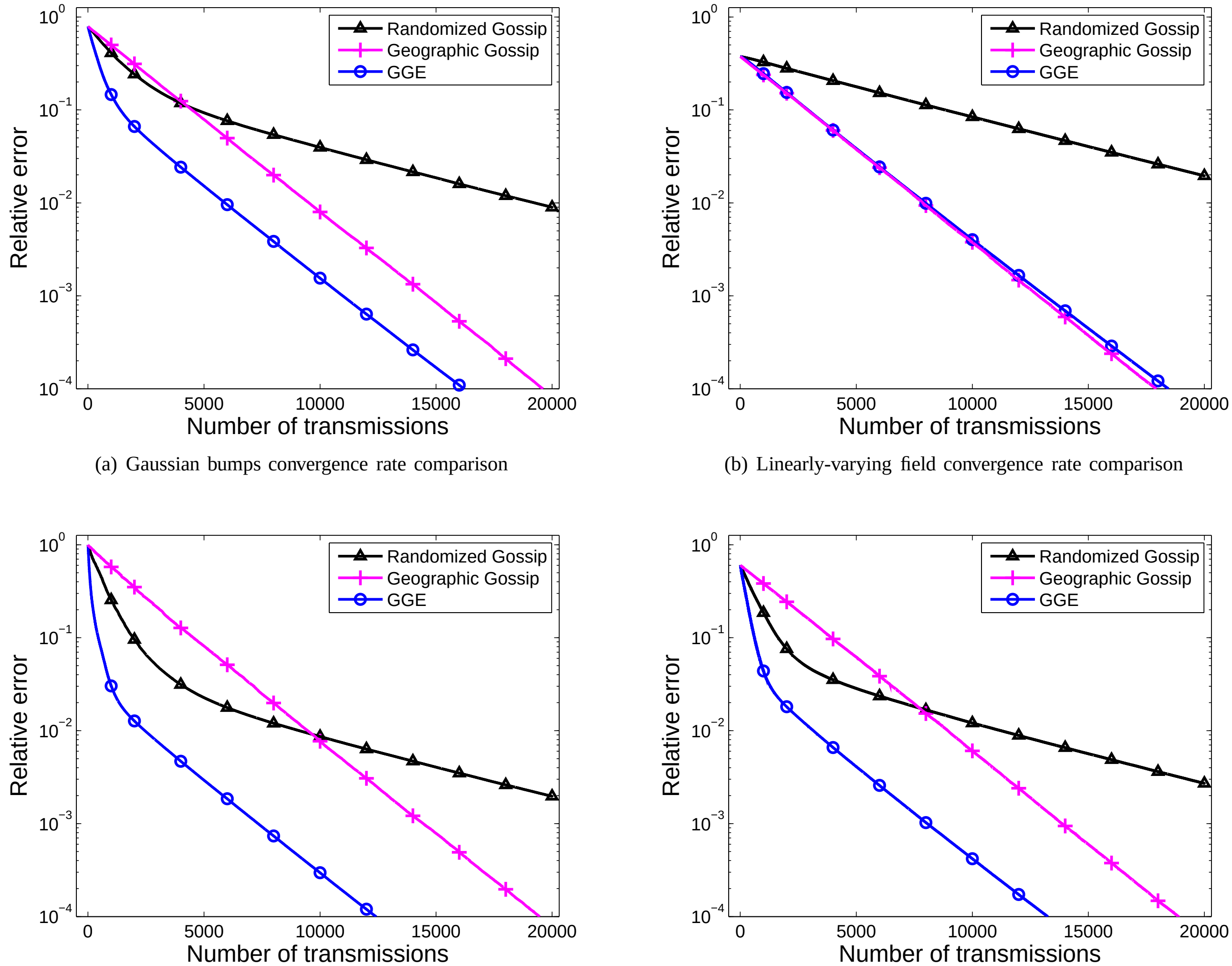
We first compare the convergence rates of the three algorithms by examining the reduction they achieve in relative error as a function of the number of transmissions (communication complexity). Relative error is defined as $\frac{||x(k)-\bar{x}||}{||x(0)-\bar{x}||}$. Since the number of transmissions per iteration is different for each algorithm, this is a fairer comparison than examining convergence rate relative to the number of iterations. Randomized gossip and GGE require two transmissions per iteration; geographic gossip has a variable number of transmissions, which depends on the number of hops between the gossiping nodes.

All figures show averages over 100 realizations of the random geometric graph. We examine performance for four different initializations $x(0)$ in order to explore the impact of the initial values on performance. The first two of these cases are a Gaussian bumps field, and a linearly-varying field. For these two cases, the initial value $x(0)$ is determined by sampling these fields at the locations of the nodes. The remaining two initializations consist of the "spike" signal, constructed by setting the value of one random node to 1 and all other node values to 0; and a random initialization where each value is drawn from a Gaussian distribution $\mathcal{N}(0, 1)$ of zero mean and unit variance. The first three of these signals were also used to examine the performance of geographic gossip in [6].

Figs. 1(a)-(d) show that GGE converges towards the average at a much faster rate (both initially and asymptotically) than randomized gossip for all initializations. The initial convergence of GGE is faster than geographic gossip for all but the linearly-varying field, but asymptotically the algorithms achieve a similar rate of reduction in relative error. Out of these candidate initializations, the linearly-varying field is the worst case, as was anticipated from the convergence analysis conducted in Section III. For this initialization, the performance of GGE is very similar to that of geographic gossip.

We now compare the empirical average relative error for the geometric graph with the bound developed in Theorem 3. In doing so, we focus on one specific realization of the 200-node random geometric graph. There is no closed-form solution for $A(G)$, so we solve the optimization problem identified in Theorem 3 numerically, using an incremental subgradient algorithm. Since the cost function can be expressed as a function of $(x(k)-\bar{x})/||x(k)-\bar{x}||$, without loss of generality, we can focus on the setting where $\bar{x} = 0$ and $||x(k)||^2 = 1$. In this simplified setting, one can reformulate the optimization as the minimization of a convex function over a non-convex set of constraints. We approximate the solution to this minimization using a projected incremental subgradient method. To avoid the problem of local minima (since the constraint set is non-convex) we rerun the optimization algorithm from multiple initial conditions.

Fig. 2 plots, for each of the four initializations of $x(0)$, the relative error achieved by GGE as a function of the number of iterations, averaged over 100 realizations of the algorithm. Also plotted is the bound identified by Theorem 3, after substitution of the numerically-evaluated $A(G)$. For



(a) Gaussian bumps convergence rate comparison

(b) Linearly-varying field convergence rate comparison

(c) Spike convergence rate comparison

(d) Uniform random field convergence rate comparison

Fig. 1. A comparison of the performance of randomized gossip, GGE, and geographic gossip for four initializations $x(0)$. Relative error versus number of transmissions for (a) the Gaussian bump field, (b) the linearly-varying field; (c) the spike distribution; and (d) the uniform random distribution. Results are averaged over 100 realizations of the random geometric graph and 100 runs of the algorithm per graph.

all but the linearly-varying field, GGE achieves a much more rapid initial decrease in error than indicated by the bound. After approximately 1000 iterations, the bound provides a good indication of the rate of decrease in error. We again observe that the linearly-varying field is close to a worst-case scenario for GGE, and it is only after approximately one-thousand iterations that the experimental performance for this initialization begins to significantly diverge from the bound.

Finally, we examine how the communication complexity scales with respect to the number of nodes in the network. Figure 3 displays how $A(G)$ and the theoretical bound on the averaging time change as the number of nodes n is increased. To obtain these data-points, we generated 50 random geometric graphs for each value of n, and evaluated numerically the $A(G)$ value for each of these, using the procedure detailed above. The top panel shows how the values of $A(G)$ change as the number of nodes increases. The bottom panel plots the ϵ-averaging time, $T_{ave}(\epsilon)$ for $\epsilon = 0.01$ versus the number of nodes. Note that the averaging time is plotted in terms of the number of iterations per node. For comparison purposes, the dotted line depicts $7\sqrt{n}$. This provides some experimental support for a conclusion that the averaging time is $O(n^{3/2})$, which implies a communication complexity similar to geographic gossip. The errorbars depict the minimum, mean and maximum values obtained for the 50 simulated graphs for each n.

VI. Summary

In this paper we analyzed the convergence behaviour of greedy gossip with eavesdropping (GGE), an algorithm we proposed in [8]. GGE takes advantage of the broadcast nature of wireless communications and provides fast and reliable computation of average consensus. The theoretical contributions of this paper are (i) a bound on the mean-squared error after k iterations of the GGE algorithm; (ii) a

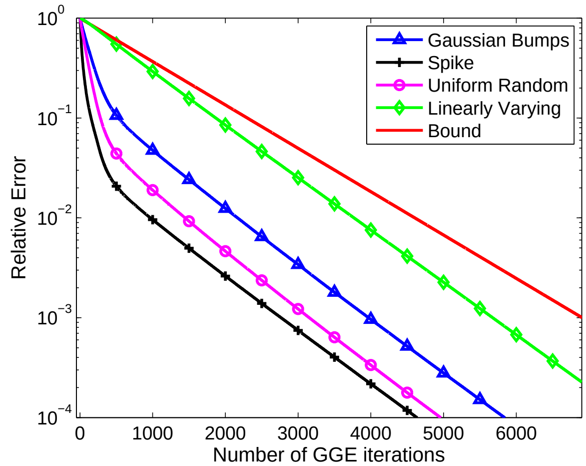


Fig. 2. A comparison of the theoretical bound on relative error and the experimental performance of GGE for four initializations. Results are for one realization of the random geometric graph, averaged over 100 runs of the algorithm.

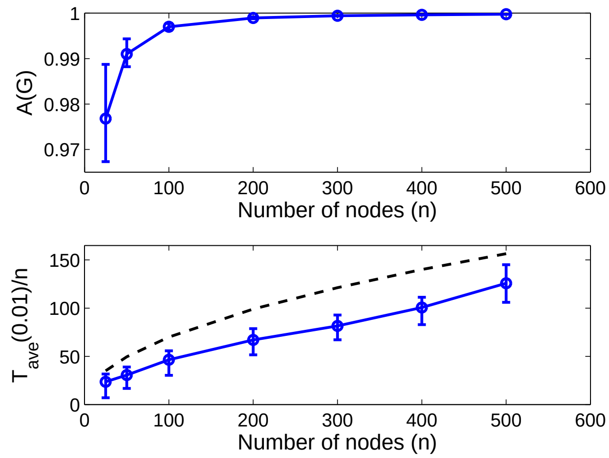


Fig. 3. The scaling behaviour of $A(G)$ and the bound on the averaging time $T_{avg}(\epsilon)$ for $\epsilon = 0.01$ as the number of nodes n in the network grows. 50 random geometric graphs were simulated for each plotted value of n. The error bars depict the minimum, mean, and maximum values obtained over these 50 realizations. Top panel: Numerically-evaluated values of $A(G)$ as a function of n. Bottom panel: Bound on averaging time, $T_{avg}(0.01)$ as a function of n. Note that is plotted in terms of the number of iterations per node. Plotted for comparison purposes as a dotted line is the curve $7\sqrt{n}$.

bound on the ϵ-averaging time of GGE; and (iii) a proof that GGE always converges faster than randomized gossip and a characterization of how the convergence rate differs.

Simulation experiments compare the performance of GGE, randomized gossip [10], and geographic gossip [6] and demonstrate that the theoretical bound on mean-squared error provides a good characterization of the algorithm performance. The simulation experiments also investigated the scaling behaviour of the communication complexity of GGE, and provided some evidence that it is $O(n^{3/2})$, similar to geographic gossip. A theoretical characterization of the scaling of this communication complexity is the focus of our current research.

REFERENCES

[1] S. Sundhar Ram, V. Veeravalli, and A. Nedič, "Distributed and recursive parameter estimation in parametrized linear state-space models," Submitted, Apr. 2008.

[2] M. Rabbat, R. Nowak, and J. Bucklew, "Robust decentralized source localization via averaging," in *Proc. IEEE ICASSP*, Phil., PA, Mar. 2005.

[3] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Proc. Information Processing in Sensor Networks*, Nashville, TN, Apr. 2006.

[4] A. Kashyap, T. Basar, and R.Srikant, "Quantized consensus," *Automatica*, vol. 43, pp. 1192–1203, Jul. 2007.

[5] S. Sundaram and C. Hadjicostis, "Distributed function calculation and consensus using linear iterative strategies," *IEEE J. Selected Areas in Communications*, vol. 26, no. 4, pp. 650–660, May 2008.

[6] A. Dimakis, A. Sarwate, and M. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," in *Proc. Int. Conf. Inf. Proc. in Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006.

[7] J. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, 1984.

[8] D. Üstebay, M. Coates, and M. Rabbat, "Greedy gossip with eavesdropping," in *Proc. IEEE Int. Symp. on Wireless Pervasive Computing*, Santorini, Greece, May 2008.

[9] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, no. 1, pp. 65–78, Sep. 2004.

[10] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Info. Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[11] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Info. Theory*, vol. 46, no. 2, pp. 388–404, March 2000.

[12] W. Li and H. Dai, "Location-aided distributed averaging algorithms: Performance lower bounds and cluter-based variant," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Urbana-Champaign, IL, Sep. 2007.

[13] K. Jung, D. Shah, and J. Shin, "Fast gossip through lifted Markov chains," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Urbana-Champaign, IL, Sep. 2007.

[14] F. Benezit, A. Dimakis, P. Thiran, and M. Vetterli, "Gossip along the way: Order-optimal consensus through randomized path averaging," in *Proc. Allerton Conf. on Comm., Control, and Computing*, Urbana-Champaign, IL, Sep. 2007.

[15] T. Aysal, M. Yildiz, and A. Scaglione, "Broadcast gossip algorithms," in *Proc. IEEE Information Theory Workshop*, Porto, Portugal, May 2008.

[16] T. Aysal, M. Yildiz, A. Sarwate, and A. Scaglione, "Broadcast gossip algorithms: Design and analysis for consensus," in *Proc. IEEE Conf. on Decision and Control*, Cancun, Mexico, Dec. 2008.

[17] S. Sundhar Ram, A. Nedič, and V. Veeravalli, "Incremental stochastic subgradient algorithms for convex optimization," Submitted, June 2008.

[18] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," M.I.T. Lab. for Information and Decision Systems, Tech. Report 2755, Aug. 2007.

[19] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Belmont, MA: Athena Scientific, 1997.

[20] A. Nedić and D. Bertsekas, "Incremental subgradient methods for nondifferentiable optimization," *SIAM J. on Opt.*, vol. 12, no. 1, pp. 109–138, 2001.

[21] M. Burnashev and K. Zigangirov, "An interval estimation problem for controlled observations," *Problems in Information Transmission*, vol. 10, pp. 223–231, 1974.

[22] R. Castro and R. Nowak, "Active learning and sampling," in *Foundations and Applications of Sensor Management*, A. Hero, D. Castanon, D. Cochran, and K. Kastella, Eds. Springer-Verlag, 2007, pp. 177–200.