Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Distributed Computation Under Bit Constraints

Ola Ayaso, Devavrat Shah, and Munther A. Dahleh

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{ayaso,devavrat,dahleh}@mit.edu

***Abstract*— A network of nodes communicate via noisy channels. Each node has some real-valued initial measurement or message. The goal of each of the nodes is to acquire an estimate of a given function of all the initial measurements in the network. We survey our recent results that relate limitations imposed by the communication constraints to the nodes' performance in computing the desired function. In particular, we determine a lower bound on computation time that must be satisfied by any algorithm used by the nodes to communicate and compute, so that the mean square error in the nodes' estimates is within a given interval around zero.**

We apply the lower bound to a specific scenario where we find the bound to be asymptotically tight. Specifically, we consider a scenario where nodes are required to learn a linear combination of the initial values in the network while communicating over erasure channels. Our results suggest that in this scenario, the computation time depends reciprocally on "conductance." Conductance is a property of the network which captures the information-flow bottle-neck that arises due to topology and channel capacities.

I. Introduction

We consider a network of nodes communicating via noisy channels. Each node has some real-valued initial measurement or message. The goal of each of the nodes is to acquire an estimate of a given function of all the initial measurements in the network.

We seek to understand the limitations imposed by the communication constraints on the nodes' performance in computing the desired function. The performance is measured by the mean square error in the nodes' estimates of the desired function. The communication constraints consist of (1) the topology of the network, that is, the connectivity of the nodes, and (2) the noisy channels between nodes that communicate. In order to capture the limitation due to the communication constraints, we assume that that the nodes have unlimited computation capability. Each node can perform any amount of computation as well as encoding and decoding for communication.

The formulation we consider lends itself to Information Theoretic techniques. We use Information Theoretic inequalities to derive lower bounds on information exchange necessary between nodes for the mean square error in the nodes' estimates to converge to zero. We use the Information Theoretic technique to determine a lower bound on computation time that must be satisfied by any algorithm used by the nodes to communicate and compute, so that the mean square error in the nodes' estimates is within a given interval around zero. The bound is in terms of the channel capacities, the size of the desired interval, and the uncertainty in the function to be computed.

Next, we apply the lower bound developed in this paper to a specific scenario where we find our bounds to be asymptotically tight. Specifically, we consider a scenario where nodes are required to learn a linear combination of the initial values in the network while communicating over erasure channels. Our lower bound suggests that in this scenario, the computation time depends reciprocally on "conductance." Conductance essentially captures the information-flow bottleneck that arises due to topology and channel capacities. The more severe the communication limitations, the smaller the conductance.

To establish the tightness of our lower bound, we describe an algorithm whose computation time matches the lower bound. The algorithm that we describe here can in fact be more generally used for distributed computation of separable functions, a special case of which is the sum. The desired function, a sum, is simple, and the algorithm that we describe has computational demands that are not severe. So, the time until the performance criterion is met using this algorithm is primarily constrained by the limitations on communication.

Indeed, we show that the upper bound, on the time until this algorithm guarantees the performance criterion, depends reciprocally on conductance. Hence, we conclude that that a lower bound we derive using Information Theoretic analysis is tight in capturing the limitations due to the network topology. Alternatively, one can interpret this tightness as the fact that the algorithm we describe here is the fastest with respect to its dependence on the network topology, as quantified by the conductance.

A. Comparison to Other Work

Existing results include algorithms with upper and/or lower bounds on the time for the nodes to reach agreement or compute a certain quantity with given accuracy, when communication is subject to topological constraints, but perfect when present [2], [3], [15], [14]. Another set of work investigates algorithms for computation when communication is unreliable. The channels in the network are explicitly modelled. The researchers propose an algorithm that will perform the desired computation while satisfying

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

some performance criterion. For example, in [5], each node in the network has one bit. Nodes broadcast messages to each other via binary symmetric channels. The goal is for a fusion center to compute the parity of all the bits in the network. Gallager proposes an algorithm that can be used while guaranteeing a desired probability of error. He exhibits an upper bound that is a constant multiple of the bits that must be transmitted per node. Recently, it has been shown in [7] that this algorithm is optimal. The authors produce an algorithm-independent lower bound that is of the same order as the upper bound.

Many different formulations and corresponding bounds can be found in the literature. Two examples are [4], [8]. In [4], the authors derive Information Theoretic bounds on the number of bits that must be exchanged for nodes communicating via noiseless channels to acquire each other's data. In [8], the authors present lower bounds to the number of messages that must be communicated by two sensors to a fusion center that must determine a given function of the nodes' data. However, the transmitted messages are real-valued vectors and the lower bound is on the sum of the dimensions of the message functions. Several formulations and results relevant to computation in wireless sensor networks can be found in a detailed survey by Giridhar and Kumar [6].

Our approach and, hence results, are quite different. We capitalize on Martins' successful use of Information Theoretic tools in [9], [10], [11], [12] to characterize fundamental performance limits of feedback control systems with communication constraints. We use Information Theoretic inequalities, reminiscent of those of Rate-Distortion theory, in a different setting with different objectives. In particular, we have a network of nodes whose objective is to compute a given function of the nodes' data, rather than to communicate reliably to each other their data.

The Information Theoretic approach captures fundamental performance limitations that arise in the network due to the communication constraints. This happens because the analysis is independent of the communication algorithm used by the nodes. The lower bound we derive in this paper enables us to characterize the effect of the network structure on algorithm running time. For nodes exchanging information over erasure channels to compute the sum of their initial conditions, the lower bound is indeed tight in capturing the network constraints.

In the next two sections, we describe the problem formulation and state our main theorems. In section IV we motivate our use of conductance to capture the effects of the underlying network structure on computation time. All proofs are omitted here, but can be found in [1].

II. Problem Formulation

A network consists of n nodes, each having a random initial condition or value. We represent the initial condition at node i by the random variable $X_i(0)$. Let $X(0)$ represent the vector of all the initial condition random variables, $[X_1(0)\ \dots\ X_n(0)]'$. Each node is required to compute a given function of all the initial conditions. That is, node i is required to estimate $C_i = f_i(X(0))$. We let $C = [C_1\ \dots\ C_n]'$. Suppose that nodes 1 to m belong to set S. Whenever we use a set as a subscript to a variable, we mean the vector whose entries are that variable subscripted by the elements of the set. For example, $C_S = [C_1\ \dots\ C_m]'$.

We assume that time is discretized into intervals, and enumerated by positive integers, $\{1, 2, \dots\}$. During each time step, a node can communicate with its neighbors. At the end of time-slot k, node i uses the information it has received thus far to form an estimate of C_i. We denote this estimate by $X_i(k)$. Let, X_i^k denote the finite sequence of estimates at node i, $\{X_i(1), X_i(2), \dots X_i(k)\}$. The estimates of all nodes in the network at the end of time slot k are denoted by the vector $X(k) = [X_1(k)\ \dots\ X_n(k)]'$. And, the estimates of nodes in set S are denoted by $X_S(k) = [X_1(k)\ \dots\ X_m(k)]'$.

The nodes communicate via noisy channels. The network structure is described by a graph, $G = (V, E)$, where V is the set of nodes and E is the set of edges, (i, j). If node i communicates with node j via channel with capacity $\mathbf{C}_{ij} > 0$, then $(i, j) \in E$. If $(i, j) \notin E$, we set $\mathbf{C}_{ij} = 0$. [1] We assume that all channels in the network are independent memoryless discrete-time. Further, for each of the channels, one channel symbol is sent per $\tau_c = 1$ time units. Each node generates an input for its encoder every τ time units, and by the k^{th} input generated, $X_i(k)$, N channel digits have been sent; so, $N\tau_c = k\tau$. When $\tau_c = 1$, the time T until the k^{th} node estimate, $X_i(k)$, has been generated is $T = k\tau$. With no loss of generality, we assume in what follows that $\tau = 1$. So, $X_i(k) = X_i(T)$.

To capture the limitations arising exclusively due to the communication structure, we assume no limits on the computational capabilities of the nodes, such as limited memory or power. The estimate $X_i(k)$ is generated by node i using some function of its initial condition, $X_i(0)$, and the messages it has received by the end of the k^{th} time slot. Similarly, the messages that the node communicates with other nodes are a function of the node's initial condition and messages it has received in the past. We make no assumptions on these functions, except that they be measureable.

We consider two mean square error criteria. The operator $\|\cdot\|$ is to be interpreted, when the argument is a vector, C, as $\|C\|^2 = \sum C_i^2$.

R1. $E(\|X(T) - C\|^2) \leq \beta 2^{-\alpha}$, and,

R2. $E(X_i(T) - C_i)^2 \leq \beta 2^{-\alpha}$, for all $i \in \{1, \dots, n\}$,

where $\beta, \alpha \in \mathbb{R}^+ \backslash \{0\}$.

The first criterion requires that as the number of nodes increases, the per node error is also smaller. It suggests that as the number of nodes, n, increases, we require the mean square errors at each of the nodes, $E(X_i(k) - C_i)^2$ to decrease like $1/n$. This criterion is appropriate if, for

[1] Note that we use bold capitalized $\mathbf{C}_{ij}$ for channel capacity, where the two letters in the subscript indicate that the channel is from node i to j. Constrast this with our notation for the function to be estimated at node i, C_i, which is not boldface and is followed by a single-letter subscript.

example, the initial values at the nodes are independent and each node is to estimate the average of the initial values in the network. As the number of nodes increases, the variance of the average decreases. In circumstances where this does not happen, the second criterion may be more appropriate.

The "computation time" is the first time at which the desired performance criterion holds. We seek a lower bound on the computation time, T, that holds if the desired mean square error criterion, R1 or R2, is satisfied. That is, if R1 or R2 holds, then how large must T be?

A. Features of This Formulation

Our formulation (and results) are appropriate for networks with severe communication constraints. These include cases where

1) channel capacities are diminished, due to loss of transmission power, for example, or,
2) network topology creates information-flow bottlenecks.

We place few assumptions on how the nodes communicate and compute their estimates. Namely, each node can use only its own initial measurement and past received messages. But, we do not specify how the node makes its computation or exchanges messages. Hence, our lower bound reveals the smallest time that must elapse before it is possible to acheive the performance desired, over all communication and computation schemes that satisfy our assumptions. The necessity of having this time elapse arises due to network topology and channel capacities.

III. Main Results

A. Computation via Noisy Channels: The Lower Bound

The first main theorem of this paper provides a lower bound to computation time as a function of the accuracy desired, as specified by the mean square error, and the uncertainty in the function that nodes must learn, as captured by the differential entropy.

Theorem III.1. *For the communication network described above, if at time, T, the mean square error is in an interval prescribed by α, $E(X_i(T) - C_i)^2 \leq \beta 2^{-\alpha}$, for every node, then T is lower bounded by*

$$T \geq \max_{S \subset V} \frac{\bar{L}(S)}{\sum_{i \in S^c} \sum_{j \in S} \mathbf{C}_{ij}},$$

where $S^c = V \backslash S$ and,

$$\bar{L}(S) = h(C_S | X_S(0)) - \frac{|S|}{2} \log 2\pi e \beta + |S| \frac{\alpha}{2}.$$

This theorem captures the fact that the larger the uncertainty in the function to be estimated, or the larger the desired accuracy, the longer it must take for any algorithm to converge.

B. Computation of Summation Over Erasure Channels: A Tight Upper Bound

A network consists of n nodes, each having a random initial condition, denoted by the random variable $X_i(0)$. Suppose the initial values at the nodes are independent and uniformly distributed, $X_i(0) \sim U[1, B+1]$. Each node is required to compute a linear function of all the initial conditions, $C = \sum_{j=1}^{n} \beta_j X_j(0)$. Node i's estimate of C at time k is denoted as $X_i(k)$.

The nodes communicate via noisy channels is described by a graph, $G = (V, E)$, where V is the set of nodes and E is the set of edges, (i, j). If node i communicates with node j via channel with capacity $\mathbf{C}_{ij} > 0$, then $(i, j) \in E$. If $(i, j) \notin E$, we set $\mathbf{C}_{ij} = 0$. We assume that the graph is connected.

We assume that all channels in the network are independent memoryless discrete-time. Further, for each of the channels, one channel symbol is sent per time unit. We assume that the channels are symmetric, $\mathbf{C}_{ij} = \mathbf{C}_{ji}$. Furthermore, they are $\log M$-bit erasure channels: with probability p_{ij} node j receives $\log M$-bits from node i without error. The capacity of this channel is $\mathbf{C}_{ij} = p_{ij} \log M$. The matrix $P = [p_{ij}]$ is a doubly stochastic matrix that captures the randomness due to the noisy channels.

The conductance of a graph, $\Phi(G)$, is a property that captures the bottle-neck of information flow. It depends on the the connectivity, or topology, of the graph, and the magnitudes of the channel capacities. It is defined as

$$\Phi(G) = \min_{\substack{S \subset V \\ 0 < |S| \leq n/2}} \frac{\sum_{i \in S, j \notin S} \mathbf{C}_{ij}}{|S|}.$$

When $\mathbf{C}_{ij} = p_{ij}$, we denote the conductance of the graph by $\Phi(P)$, which we call the conductance of the matrix P. In the case of our $\log M$-bit erasure channel, we have that $\mathbf{C}_{ij} = p_{ij} \log M$. So, the graph conductance in this case is $\Phi(G) = \Phi(P) \log M$.

Now we are ready to state the main result. We shall first describe a lower bound (an application of Theorem III.1) and then a matching upper bound implied by a quantized algorithm.

Let A represent a realization of the initial conditions, $A = \{X_1(0) = x_1, \ldots, X_n(0) = x_n\}$. The performance of an algorithm, $\mathcal{H}$, used by the nodes to compute an estimate of $f(x, V) = \sum_{j=1}^{n} \beta_j x_j$ at each node, is measured by the algorithm's (ε, δ)-computation time, $T_{\mathcal{H}}^{\text{cmp}}(\varepsilon, \delta)$. It is the time until the estimates at all nodes are within a factor of $1 \pm \varepsilon$ of $f(x, V)$, with probability larger than $1 - \delta$. The definition follows, where $\hat{y}_i(k)$ denotes the estimate of $f(x, V)$ at node i at time k.

Definition III.2. For $\varepsilon > 0$ and $\delta \in (0, 1)$, the (ε, δ)-computing time of an algorithm, $\mathcal{H}$, denoted as $T_{\mathcal{H}}^{\text{cmp}}(\varepsilon, \delta)$ is defined as

$$T^{\text{cmp}}_{\mathcal{H}}(\varepsilon,\delta) = \sup_{x\in\mathbb{R}^n} \inf \{k : \\ \mathbf{P}(\cup_{i=1}^{n}\{\hat{y}_i(k) \notin [(1-\varepsilon)f(x,V),(1+\varepsilon)f(x,V)]\}) \le \delta\}\,.$$

Here, the probability is taken with respect to $\hat{y}_i(k)$. This is random because nodes communicate over noisy channels.

Consider any algorithm, $\mathcal{H}$, that guarantees that for any realization of the initial values, with high probability each node has an estimate within $1 \pm \varepsilon$ of the true value of C, at time T. The Information Theoretic lower bound maintains that such algorithm must have a computation time, $T = T^{\text{cmp}}_{\mathcal{H}}(\varepsilon,\delta)$, that is inversely proportional to conductance.

Theorem III.3. *Nodes communicate in order for each node to compute a linear combination of all initial values in the network. Any algorithm that guarantees that for all* $i \in \{1,\ldots,n\}$,

$$\mathbf{P}\left(|X_i(T) - C| \le \varepsilon C \middle| A\right) \ge 1-\delta,$$

must have

$$T \ge \frac{1}{\Phi(G)}\log\frac{1}{B\varepsilon^2 + \frac{1}{B}^{\frac{2}{n}} + \kappa\delta},$$

where, $B\varepsilon^2 \in \left[0, 1-\frac{1}{B}^{\frac{2}{n}} - \kappa\delta\right]$, *and* κ *is a constant.*

Again, the probability in this theorem is taken with respect to the measure on $X_i(T)$, conditional on A, and induced by the randomness due to communication over channels.

Remark We show in [1] that our lower bound is tight in its reciprocal dependence on the conductance term. So, for fixed n, we have a scaling law that is tight in the case of severe communication constraints, such as very small channel capacities due to low transmission power.

In the case of increasing number of nodes, however, B must increase exponentially with n for our lower bound to remain valid. The requirement is a by-product of using a formulation based on random variables together with Information Theoretic variables. This requirement ensures that as n increases, our bound properly captures the number of bits that are transferred.

When we consider sums of independent identically distributed random variables, Central Limit Theorem type arguments imply that as the number of the random variables increases, there is some randomness lost, because we know that the distribution of the sum must converge to the Normal distribution. However, in a setting where the initial conditions are fixed values, as in the case of the algorithm we describe below, the addition of a node clearly will not reduce the information that needs to be communicated in the network. To counterbalance the probabilistic effects, we need to have B increase as the number of nodes increases.

Next, we provide an algorithm that guarantees, with high probability, the nodes' estimates are within the desired ε-error interval around the true value of the sum. We provide an upper bound on this algorithm's computation time. The computation time is inversely proportional to conductance.

This algorithm is based on an algorithm proposed in [13]. The difference is that in [13], nodes communicate real-valued messages. In our algorithm, nodes communicate quantized messages.

Theorem III.4. *Suppose that node* i *has an initial condition,* x_i. *There exists a distributed algorithm* $\mathcal{AP}^{\mathcal{Q}}$ *by which nodes compute a linear sum,* $f(x,V) = \sum_{j=1}^{n}\beta_j x_j$, *via communication of quantized messages. If each quantized message is* $\log M$ *bits and* $\log M = O(\log n)$, *the quantization error will be no more than a given* $\gamma = \Theta(\frac{1}{n})$, *and for any* $\varepsilon \in (\gamma f(x,V), \gamma f(x,V) + \frac{1}{2})$ *and* $\delta \in (0,1)$, *the computation time of the algorithm will be*

$$T^{cmp}_{\mathcal{AP}^{\mathcal{Q}}}(\varepsilon,\delta) = \\ O\left(\varepsilon^{-2}(1+\log\delta^{-1})\frac{(\log n + \log\delta^{-1})\log n}{\Phi(G)}\right).$$

So, setting $\delta = \frac{1}{n^2}$ in the above bound, we have

$$T^{\text{cmp}}_{\mathcal{AP}^{\mathcal{Q}}}\left(\varepsilon, \frac{1}{n^2}\right) = O\left(\varepsilon^{-2}\frac{\log^3 n}{\Phi(G)}\right).$$

The computation time of this algorithm depends on the network topology, via the conductance of the graph, in the same reciprocal manner manifested by the lower bound. Thus, we conclude that the lower bound is tight in capturing the effect of the network topology on computation time. Conversely, the algorithm's running time is optimal with respect to its dependence on the network topology, as captured by the conductance.

IV. Capturing the Effect of Topology

The conductance of a graph, $\Phi(G)$, is a property that captures the bottle-neck of information flow. It depends on the the connectivity, or topology, of the graph, and the magnitudes of the channel capacities. The more severe the network constraints, the smaller the conductance. It is also related to time it takes for information to spread in a network; the smaller the conductance, the longer it takes.

A. Conductance for Two Topologies

Consider two networks, each has n nodes. We calculate conductance for two extreme cases of connectivity shown in Figure 1. On the one hand, we have severe topological constraints: a ring graph. Each node may contact only the node on its left or the node on its right. On the other hand, we have a case of virtually no topological constraints: a fully connected graph. Each node may contact every other node in the network.

For the purpose of illustrating the computation of conductance for the two topologies, suppose that in both cases, the links from a given node to different nodes are equally weighted. So, for the ring graph, let $\mathbf{C}_{ij} = \mathbf{C} = \frac{1}{4}$, for all $i \neq j$; for the fully connected graph, let $\mathbf{C}_{ij} = \mathbf{C} = \frac{1}{n}$, for all $i \neq j$. Assume that for the ring graph, $\mathbf{C}_{ii} = \frac{1}{2}$. If the channels were erasure channels, this would be the probability that node i makes contact with no other nodes. For the fully connected graph, let $\mathbf{C}_{ij} = \frac{1}{n}$. So, in both cases, we have that the sum of the capacities of channels leaving a node is 1, $\sum_j \mathbf{C}_{ij} = 1$.

Now, we compute the conductance of the ring graph. Recall that conductance is

$$\Phi(G) = \min_{\substack{S \subset V \\ 0 < |S| \leq n/2}} \frac{\sum_{i \in S, j \notin S} \mathbf{C}_{ij}}{|S|}.$$

Consider any cut that divides the ring graph into two sets, S and S^c. For any such cut, there will be exactly two links crossing the cut, going from S to S^c. So, $\sum_{i \in S, j \notin S} \mathbf{C}_{ij} = \frac{1}{2}$, and

$$\Phi(G) = \min_{\substack{S \subset V \\ 0 < |S| \leq n/2}} \frac{\frac{1}{2}}{|S|}.$$

Since we minimize over all cuts such that $|S| \leq n/2$, the ratio is minimized when the cut divides the ring into two sets of equal size, and $|S| = n/2$. So, $\Phi(G) = \frac{1}{n}$.

Next, we compute the conductance of the fully connected graph. Consider any cut that divides the graph into two sets, S and S^c. For any such cut, there will be $|S||S^c|$ links crossing the cut, going from S to S^c. So,

$$\begin{aligned} \frac{\sum_{i \in S, j \notin S} \mathbf{C}_{ij}}{|S|} &= \frac{|S||S^c|\frac{1}{n}}{|S|} \\ &= \frac{|S^c|}{n} \\ &= \frac{n - |S|}{n} \end{aligned}$$

The last equality is minimized where $|S| = n/2$, so, $\Phi(G) = \frac{1}{2}$.

So, for two networks with the same number of nodes, the network with the more severe topological constraints has smaller conductance. In general, for a ring graph, we have $\Phi(G) = O(\frac{1}{n})$, while for a fully connected graph we have $\Phi(G) = O(1)$.

B. Comparison to Other Algorithms

The popular algorithms for computing a linear function of initial conditions, such as averaging and consensus, are based on linear iterations. The convergence of such iterative algorithms depends on a reversible (or symmetric) and graph conformant matrix P. Usually, the running time of these algorithms scales like $\frac{1}{\Phi(P)^2}$. Specifically, for a ring graph it is $\frac{1}{\Phi(P)^2} \approx n^2$, which means roughly n^2 iterations are needed for the algorithm to converge. In [1], we describe an algorithm that does not use linear computations, and its run-time scales like $\frac{1}{\Phi(P)}$. So, for a ring, roughly n iterations are needed. In the next section, we show that the run-time necessarily scales like $\frac{1}{\Phi(P)}$. So, for computation over a ring graph, n iterations are both sufficient and necessary. More generally, our algorithm scales optimally with respect to the graph topology.

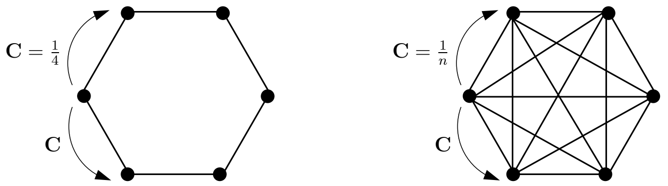

Fig. 1. Two ways to connect six nodes: a ring graph and a fully connected graph.

V. Discussion and Conclusions

In this paper, we've surveyed our recent results for a network of n nodes communicating over noisy channels. Each node has an initial value. The objective of each of the nodes is to compute a given function of the initial values in the network. We have derived a lower bound to the time at which the mean square error in the nodes' estimates is within a prescribed accuracy interval. The lower bound is a function of the channel capacities, the accuracy specified by the mean square error criterion, and the uncertainty in the function that is to be estimated. The bound reveals that, first, the more randomness in the function to be estimated, the larger the lower bound on the computation time. Second, the smaller the mean square error that is tolerated, the larger the lower bound on the computation time. Hence there is a trade-off captured between computation accuracy and computation time. In addition, the lower bound can be used to capture the dependence of the convergence time on the structure of the underlying communication network.

We've considered a network of nodes communicating via erasure channels to compute a sum of the initial values in the network. Each of the nodes is required to acquire an estimate that is, with a specified probability, within a desired interval of the true value of the sum. We've applied our Information Theoretic technique to derive a lower bound on the computation time for this scenario. We've shown that the computation time is inversely related to a property of the network called "conductance." It captures the effect of both the topology and channel capacities by quantifying the bottle-neck of information flow. Next, we've described an algorithm that can be used in this setting of nodes computing a sum via erasure channels, and guarantees that with the specified probability, each of the nodes' estimate is within the desired interval. We've determined an upper bound on the algorithm's computation time and show that it too is inversely related to conductance.

Hence, we conclude that our lower bound is tight in capturing the effect of the communication network, via conductance. Equivalently, our algorithm's run-time is optimal in its dependence on conductance. That is, we have obtained a scaling law for convergence time as a function of a network property, conductance. When the number of nodes is fixed, this scaling law becomes tighter as the communication constraints are more severe, like diminished channel capacities.

References

[1] O. Ayaso. *Information Theoretic Bounds for Distributed Computation.* PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Laboratory of Information and Decision Systems, June 2008.

[2] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05)*, Seville, Spain, 2005. IEEE.

[3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Transactions on Information Theory*, 52(6):2508–2530, June 2006.

[4] A. El Gamal and A. Orlitsky. Interactive data compression. In *Proceedings of the 25th IEEE Symposium on Foundations of Computer Science*, pages 100–108. IEEE, October 1984.

[5] R. Gallager. Finding parity in a simple broadcast network. *IEEE Transactions on Information Theory*, 34:176–180, 1988.

[6] A. Giridhar and P. R. Kumar. Towards a theory of in-network computation in wireless sensor networks. *IEEE Communications Magazine*, 44(4):98–107, April 2006.

[7] N. Goyal, G. Kindler, and M. Saks. Lower bounds for the noisy broadcast problem. *SIAM Journal on Computing*, 37(6):1806–1841, March 2008.

[8] Z.-Q. Luo and J. N. Tsitsiklis. Data fusion with minimal communication. *IEEE Transactions on Information Theory*, 40(5):1551–1563, September 1994.

[9] N. C. Martins. *Information Theoretic aspects of the control and mode estimation of stochastic systems*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Laboratory of Information and Decision Systems, August 2004.

[10] N. C. Martins. Finite gain lp stabilization requires analog control. *Systems and Control Letters*, 55/1:949–954, Novemeber 2006.

[11] N. C. Martins and M. A. Dahleh. Feedback control in the presence of noisy channels: "bode-like" fundamental limits of performance. *IEEE Transaction on Automatic Control*, May 2008.

[12] N. C. Martins, M. A. Dahleh, and J. C. Doyle. Fundamental limitations of disturbance attenuation in the presence of side information. *IEEE Transaction on Automatic Control*, 52:56–66, January 2007.

[13] D. Mosk-Aoyama and D. Shah. Computing separable functions via gossip. In *ACM Principles of Distributed Computation*, 2006.

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

[15] J.N. Tsitsiklis. *Problems in Decentralized Decision Making and Computation*. PhD dissertation, Massachusetts Institute of Technology, Laboratory of Information and Decision Systems, Department of Electrical Engineering and Computer Science, 1984.