

Quadratic Gaussian Gossiping

Han-I Su

Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA
Email: hanisu@stanford.edu

Abbas El Gamal

Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA
Email: abbas@stanford.edu

Abstract—This paper presents an information theoretic source coding formulation of distributed averaging. We assume a network with m nodes each observing an i.i.d. Gaussian source; the nodes communicate and perform local processing with the goal of computing the average of the sources to within a prescribed mean squared error distortion. A general cutset lower bound on the network rate distortion function is established and shown to be achievable to within a factor of 2 via a centralized protocol in the star network. A lower bound on the network rate distortion function for distributed weighted-sum protocols that is larger than the cutset bound by a factor of $\log m$ is established. We also establish an upper bound on the expected network rate distortion function for gossip-based weighted-sum protocols that is only a factor of $\log \log m$ larger than this lower bound. The results suggest that using distributed protocols results in a factor of $\log m$ increase in communication relative to centralized protocols.

I. INTRODUCTION

Gossip-based protocols for distributed consensus have attracted much attention recently due to interest in applications ranging from distributed coordination in autonomous vehicles to data aggregation and distributed computation in sensor networks, ad-hoc networks, and peer-to-peer networks.

This paper presents a lossy source coding formulation of the distributed averaging problem which is a canonical example of distributed consensus. We assume that each node in the network observes a Gaussian source and the nodes communicate and perform local processing with the goal of computing the average of the sources to within a prescribed mean squared error distortion. We investigate the network rate distortion function in general and for the class of distributed weighted-sum protocols, including random gossip-based protocols.

Most previous work on distributed averaging, e.g., [1], [2], has involved the noiseless communication and computation of real numbers, which is unrealistic. Recognizing this shortcoming, the effect of quantization on distributed averaging has been recently investigated. Our work is related most closely to the work in [3]–[5]. Compared to [3], [4], our information-theoretic approach deals more naturally and fundamentally with quantization and provides limits that hold independent of implementation details. Our results, however, cannot be compared directly to results in these papers because of differences in the models and assumptions. While the work in [5] is information-theoretic, its formulation is different from ours and the results are not comparable. Our formulation of the distributed averaging problem can be viewed also

as a generalization of the CEO problem [6], where in our setting every node wishes to compute the average and the communication protocol is significantly more complex in that it allows for interactivity, relaying, and local computing, in addition to multiple access.

In the following section, we introduce the lossy averaging problem. In Section III, we establish a general cutset lower bound on the network rate distortion function and show that it can be achieved within a factor of 2 using a centralized protocol. In Section IV, we investigate the class of distributed weighted-sum protocols. We establish a lower bound on the network rate distortion function for this class as well as an upper bound for gossip-based weighted-sum protocols. The full paper is posted on arXiv.org [7].

II. LOSSY AVERAGING PROBLEM

Consider a network with m sender-receiver nodes, where node $i = 1, 2, \dots, m$ observes a source X_i . Assume that the sources are independent white Gaussian noise (WGN) processes each with average power of one. The nodes communicate and perform local processing with the goal of computing the average of the sources $S = (1/m) \sum_{i=1}^m X_i$ at each node to within a prescribed mean squared error distortion D .

The topology of the network is specified by a connected graph $(\mathcal{M}, \mathcal{E})$, where $\mathcal{M} = \{1, 2, \dots, m\}$ is the set of nodes and \mathcal{E} is a set of undirected edges (node pairs) $\{i, j\}$, $i, j \in \mathcal{M}$ and $i \neq j$. Communication is performed in *rounds* and each round is divided into *time slots*. Each round may consist of a different number of time slots. One edge is chosen at each round and only one node is allowed to transmit in each time slot. Without loss of generality, we assume that the selected node pair communicates in a round robin manner. Further, we assume a *source coding* setting, where communication is noiseless and instant, that is, every transmitted message is successfully received by the intended receiver in the same time slot it is transmitted in.

Communication and computing are performed according to an agreed upon *averaging protocol* that determines (i) the number of communication rounds T , (ii) the sequence of edge selections, and (iii) the block codes used in each round to perform communication and local computing. The averaging protocol may be deterministic or random. In a random protocol, the sequence of T edges are selected at random. Given an averaging protocol with T rounds, we define

an $(R_1, R_2, \dots, R_m, n)$ block code for a feasible sequence of edge selections to consist of:

1. A set of encoding functions, one for each time slot. Each encoding function assigns a message to each node source sequence of block length n and past messages received by the node. Let the number of bits transmitted by node i in round $t = 1, 2, \dots, T$ be $nr_i(t)$. Then the total transmission rate per source symbol for node i is given by $R_i := \sum_{t=1}^T r_i(t)$.

2. A set of decoding functions, one for each node. At the end of round T , the decoder for node $i \in \mathcal{M}$ assigns an estimate $Y_i^n := (Y_{i1}, Y_{i2}, \dots, Y_{in})$ of the average $S^n := (S_1, S_2, \dots, S_n)$ to each source sequence and all messages received by the node.

The *per-node transmission rate* for the code is $R := (1/m) \sum_{i=1}^m R_i$. The average per-letter distortion associated with the code is defined as

$$\Delta^{(n)} := \frac{1}{mn} \sum_{i=1}^m \sum_{k=1}^n \mathbb{E}((S_k - Y_{ik})^2),$$

where the expectation is taken over source statistics. Note that we are also averaging over the nodes. A rate distortion pair (R, Δ) is said to be achievable if there exists a sequence of $(R_1, R_2, \dots, R_m, n)$ codes with per-node rate R such that $\limsup_{n \rightarrow \infty} \Delta^{(n)} \leq \Delta$.

The network rate distortion function for a given feasible sequence of edge selections $R(D)$ is the infimum over all achievable rates R such that (R, D) is achievable. The *network rate distortion function* $R^*(D)$ is the infimum of $R(D)$ over all averaging protocols. It is clear that $R^*(D)$ can be achieved by a deterministic averaging protocol.

We are also interested in the expected network rate distortion function for a random averaging protocol. We consider the expected per-node transmission rate $\mathbb{E}(R) = (1/m) \sum_{i=1}^m \mathbb{E}(R_i)$, and the limit on the expected distortion with respect to edge selection statistics $\mathbb{E}(\Delta)$ specified by the random averaging protocol. The *expected network rate-distortion function* $\mathbb{E}(R(D))$ for a random averaging protocol is defined as the infimum over all expected per-node rate $\mathbb{E}(R)$ such that the pairs (R, Δ) are achievable and $\mathbb{E}(\Delta) \leq D$. Clearly for any random averaging protocol, $R^*(D) \leq \mathbb{E}(R(D))$. Further, any upper bound on $\mathbb{E}(R(D))$ is an upper bound on $R^*(D)$.

Centralized versus Distributed Protocols: A goal of our work is to quantify the communication penalty of using distributed relative to centralized protocols. In a distributed protocol, the code used at each round does not depend on the identities of the selected nodes. The code, however, may depend on the round number. A protocol is called centralized if it is not distributed. For example, a node may be designated as a “cluster-head” and treated differently than other nodes.

III. LOWER BOUND ON $R^*(D)$

Consider the m -node distributed lossy averaging problem where the sources (X_1, X_2, \dots, X_m) are independent WGN processes each with average power one. We establish the following cutset lower bound.

Theorem 1: The network rate distortion function $R^*(D) = 0$ if $D \geq (m-1)/m^2$ and is lower bounded by

$$R^*(D) \geq \frac{1}{2} \log \left(\frac{m-1}{m^2 D} \right) \text{ if } D < \frac{m-1}{m^2}.$$

Remark: The above cutset lower bound can be readily extended to correlated WGN sources and weighted-sum computation. The resulting bound is tight for $m = 2$ by having each node independently compress its source and send the compressed version to the other node using Wyner-Ziv coding [8].

A. Upper Bound on $R^*(D)$ for Star Network

Consider a star network with m nodes and $m-1$ edges $\mathcal{E} = \{\{1, i\} : i = 2, \dots, m\}$. We use a centralized protocol where node 1 is treated as a “cluster head.” The protocol has $T = 2m - 3$ rounds. In round $t = 1, 2, \dots, m-1$, node $i = t+1$ compresses its source X_i^n using Gaussian random codes with average distortion $(d/n) \sum_{k=1}^n \mathbb{E}(X_{ik}^2) = d$ and sends the index $M_i(X_i^n)$ to node 1. Node 1 finds the corresponding reconstruction sequences $\hat{X}_i^n(M_i)$ and computes the estimates

$$Y_1^n := \frac{1}{m} X_1^n + \frac{1}{m} \sum_{i=2}^m \hat{X}_i^n \text{ and } U_i^n := Y_1^n - \frac{1}{m} \hat{X}_i^n$$

for $i = 2, 3, \dots, m$. In round $t = m-1, m, \dots, 2m-3$, node 1 compresses the estimate U_{2m-t-1}^n using Gaussian random codes with average distortion $(d/n) \sum_{k=1}^n \mathbb{E}(U_{2m-t-1,k}^2)$ and sends the index $\tilde{M}_{2m-t-1}(U_{2m-t-1}^n)$ to nodes $2m-t-1$. Node i computes the estimate $Y_i^n = (1/m)X_i^n + \hat{U}_i^n$ for $i = 2, 3, \dots, m$, where \hat{U}_i^n is the reproduction sequence of U_i^n corresponding to the index \tilde{M}_i . This establishes the following upper bound on the network rate distortion function.

Proposition 1: The network rate distortion function for the star network is upper bounded by

$$R^*(D) \leq \frac{m-1}{m} \log \left(\frac{2(m-1)^2}{m^3 D} \right) \text{ for } D < \frac{m-1}{m^2}.$$

Note that the ratio of the upper bound to the cutset lower bound for $D < 1/m^2$ as $m \rightarrow \infty$ is less than or equal to 2. Thus a centralized protocol can achieve a rate within a factor of 2 of the cutset bound.

IV. DISTRIBUTED WEIGHTED-SUM PROTOCOLS

Again assume that the sources (X_1, X_2, \dots, X_n) are independent WGN processes each with average power one. We consider *distributed weighted-sum protocols* characterized by the number of rounds T and the *normalized local distortion* d . Given a network, we define a distributed weighted-sum code for each feasible edge selection sequence as follows. Initially, each node $i \in \mathcal{M}$ has an estimate $Y_i^n(0) = X_i^n$ of the average S^n . In each round, communication is performed in two time slots. Assume that edge $\{i, j\}$ is selected in round $(t+1)$. In the first time slot, node i compresses $Y_i^n(t)$ using Gaussian random codes with distortion $(d/n) \sum_{k=1}^n \mathbb{E}(Y_{ik}^2(t))$, and sends the index $M_i(t+1)(Y_i^n(t))$ to node j at rate $r = (1/2) \log(1/d)$. Similarly, in the second time slot, node j compresses $Y_j^n(t)$ using Gaussian random codes and sends

the index $M_j(t+1)(Y_j^n(t))$ to node i at the same rate r . Upon receiving the indices, nodes i and j update their estimates

$$Y_v^n(t+1) = \frac{1}{2}Y_v^n(t) + \frac{1}{2(1-d)}\hat{Y}_{i+j-v}^n(t) \text{ for } v = i, j, \quad (1)$$

where $\hat{Y}_i^n(t)$ and $\hat{Y}_j^n(t)$ are the reproduction sequences corresponding to $M_i(t+1)$ and $M_j(t+1)$, respectively. At the end of round T , node i sets $Y_i^n = Y_i^n(T)$ if it is involved in at least one round of communication, and sets $Y_i^n = (1/m)Y_i^n(0)$, otherwise.

Define the rate distortion function for a distributed weighted-sum protocol and a given edge selection sequence, $R_{\text{WS}}(D)$, the *weighted-sum network rate distortion function*, $R_{\text{WS}}^*(D)$, and the *weighted-sum expected network rate distortion function* for a random protocol, $E(R_{\text{WS}}(D))$, as in Section II.

Remark: The code defined above does not exploit the correlation between the node estimates induced by communication and local computing. This correlation can be readily used to reduce the rate via Wyner-Ziv coding. However, we are not able to obtain upper and lower bounds on the network rate distortion function with side information because the correlations between the estimates are time varying and depend on the particular sequence of edges selected.

Since the estimates $Y_{i1}(t), \dots, Y_{in}(t)$ are i.i.d, from this point on we suppress the symbol index.

A. Lower Bound on $R_{\text{WS}}^*(D)$

We establish a lower bound on $R_{\text{WS}}^*(D)$ that applies to any network. Consider a distributed weighted-sum protocol for a given network and a feasible sequence of edge selections. Let $t_{i\tau}$ be the τ -th time node i is selected and $\mathcal{T}_i := \{t_{i1}, t_{i2}, \dots, t_{iT_i}\}$, for $i = 1, 2, \dots, m$, where $T_i := |\mathcal{T}_i|$ is the number of rounds in which node i is selected. Then the number of rounds T can be expressed as $T = (1/2) \sum_{i=1}^m T_i$, where the 1/2 factor is due to the fact that two nodes are selected in each round. We shall need the following properties of the estimate $Y_i(t)$ to prove the lower bound.

Lemma 1: For a distributed weighted-sum protocol, we can express the estimate at node i at the end of round t as $Y_i(t) = \sum_{j=1}^m \gamma_{ij}(t)Y_j(0) + Z_i(t)$, where $Z_i(t)$ is Gaussian and independent of the sources (X_1, X_2, \dots, X_m) . Furthermore, the diagonal coefficients satisfy the property

$$\gamma_{ii}(t) \geq \frac{1}{2^\tau} \text{ for } t_{i\tau} \leq t < t_{i,\tau+1} \text{ and } \tau = 1, \dots, T_i.$$

Using this lemma, we can establish the following.

Lemma 2: Given $0 < D < (m-1)/m^2$, if a weighted-sum protocol with T rounds achieves distortion D , then

$$T \geq \frac{m}{2} \log \left(\frac{1}{\sqrt{D} + 1/m} \right).$$

We can establish the following lower bound using these lemmas.

Theorem 2: Given $0 < D < (m-1)/m^2$, then

$$R_{\text{WS}}^*(D) \geq \frac{1}{2} \left(\log \frac{1}{\sqrt{D} + 1/m} \right) \left(\log \frac{1}{4mD} \right).$$

Proof outline: Given a distributed weighted-sum protocol with T rounds and normalized local distortion d . Suppose that the edge selected at round $t_{j_1\tau}$ is $\{j_1, j_2\}$, then at the end of this round, the estimate for node k is

$$Y_{j_2}(t_{j_1\tau}) = (Y_{j_2}(t_{j_1\tau} - 1) + Y_{j_1}(t_{j_1\tau} - 1) + W_{j_1\tau})/2,$$

where $W_{j_1\tau} \sim \mathcal{N}(0, E(Y_{j_1}(t_{j_1\tau} - 1)^2) d/(1-d))$. By induction, we can show that the estimate of node i at time $t \geq t_{j_1\tau}$ has the form $Y_i(t) = (1/2)\beta_i(t)W_{j_1\tau} + \tilde{Y}_i(t)$, where $\beta_i(t) \geq 0$, $\sum_{i=1}^m \beta_i(t) = 1$, and $\tilde{S}_i(t)$ is independent of $W_{j_1\tau}$. Now we compute the distortion at the end of round T

$$\begin{aligned} & \frac{1}{m} \sum_{i=1}^m E((S - Y_i(T))^2) \\ &= \frac{1}{m} \sum_{i=1}^m \left(E \left(\left(\frac{\beta_i(T)}{2} W_{j_1\tau} \right)^2 \right) + E((S - \tilde{Y}_i(T))^2) \right) \\ &\geq \frac{d}{4m^2(1-d)2^{2(\tau-1)}} + \frac{1}{m} \sum_{i=1}^m E((S - \tilde{Y}_i(T))^2), \end{aligned}$$

where the inequality follows by the Cauchy-Schwarz inequality and Lemma 1. We can repeat the above arguments for the second term $(1/m) \sum_{i=1}^m E((S - \tilde{Y}_i(T))^2)$ and we obtain

$$\frac{1}{m} \sum_{i=1}^m E((S - Y_i(T))^2) \geq \sum_{i=1}^m \sum_{\tau=1}^{T_i} \frac{d}{4m^2(1-d)2^{2(\tau-1)}} \geq \frac{d}{4m}.$$

Thus, $d \leq 4mD$. The rest of the proof follows from Lemma 2. ■

Remark: The above lower bound and the cutset bound in Theorem 1 differ by roughly a factor of $\log m$. Given that the cutset bound can be achieved within a factor of 2 using a centralized protocol, this suggests that the $\log m$ factor is the penalty of using a distributed versus centralized protocols.

B. Bounds on $E(R_{\text{WS}}(D))$

In this section, we establish bounds on $E(R_{\text{WS}}(D))$ for *gossip-based weighted-sum protocols* [1] characterized by (T, d, Q) , where T is the number of rounds, d is the normalized local distortion, and Q is an $m \times m$ stochastic matrix such that $Q_{ij} = 0$ if $\{i, j\} \notin \mathcal{E}$. Note that this also establishes an upper bound on $R^*(D)$ because $R^*(D) \leq E(R_{\text{WS}}(D))$.

In each round of a gossip-based weighted-sum protocol, a node i is selected uniformly at random from \mathcal{M} . Node i then selects a neighbor $j \in \{j : \{i, j\} \in \mathcal{E}\}$ with conditional probability Q_{ij} . This node pair selection process is equivalent to the asynchronous model in [1]. After the node pair $\{i, j\}$ is selected, the *distributed weighted-sum coding* scheme previously described is performed. Let $\mathbf{Y}(t) := [Y_1(t) \ Y_2(t) \ \dots \ Y_m(t)]^T$ and rewrite the update equations (1) in the matrix form

$$\mathbf{Y}(t+1) = A(t+1)\mathbf{Y}(t) + \mathbf{W}(t+1), \quad (2)$$

where (i) $A(t+1) = I_m - \frac{1}{2}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$ with probability $(1/m)Q_{ij}$, independent of t , where \mathbf{e}_i and \mathbf{e}_j are unit vectors along the i -th and j -th axes, and (ii) $\mathbf{W}(t+1) = W_i(t+1)\mathbf{e}_i + W_j(t+1)\mathbf{e}_j$, and $W_i(t+1) \sim \mathcal{N}(0, E(Y_j(t)^2 d/4(1-d)))$

and $W_j(t+1) \sim \mathcal{N}(0, E(Y_i(t)^2 d/4(1-d)))$ are independent of past estimates.

Recall properties of the matrix $A(t)$ from [1]. Let Q^* be the stochastic matrix that minimizes the second largest eigenvalue of the matrix $A := E(A(0))$, and let λ_2 be the optimum second largest eigenvalue, which is a function of the network topology. We will need the following lower bound on the number of rounds T to prove the lower bound on $E(R_{\text{WS}}(D))$.

Lemma 3: Given a connected network, if a gossip-based weighted-sum protocol (T, d, Q) achieves distortion D , then $T \geq ((m-1)/2) \ln((m-1)/mD)$.

We now establish an upper bound on distortion.

Lemma 4: The average per-letter distortion of the gossip-based weighted-sum protocol (T, d, Q) is upper bounded by

$$\frac{1}{m} E(\|\mathbf{Y}(T) - J\mathbf{Y}(0)\|^2) \leq \frac{1}{m^2} ((1+u)^T - 1) + \frac{1}{m} \frac{u}{1 - \lambda_2 + u} (1+u)^T + \frac{1}{m} \frac{u}{1 - \lambda_2 - u} + (\lambda_2 + u)^T,$$

where $u := d/(2m(1-d))$ and $J := (1/m)\mathbf{1}\mathbf{1}^T$.

Proof outline: Referring to the linear dynamical system (2), express $\mathbf{Y}(T)$ as $\mathbf{Y}(T) = A(T, 1)\mathbf{Y}(0) + \mathbf{Z}(T)$, where $\mathbf{Z}(T) = \sum_{t=1}^T A(T, t+1)\mathbf{W}(t)$ and $A(t_2, t_1) = A(t_2)A(t_2-1) \cdots A(t_1)$ for $t_2 \geq t_1$.

Consider the sum of distortions over all nodes

$$E(\|\mathbf{Y}(T) - J\mathbf{Y}(0)\|^2) = E(\|A(T, 1)\mathbf{Y}(0) - J\mathbf{Y}(0)\|^2) + E(\|\mathbf{Z}(T)\|^2).$$

We can show that

$$E(\|A(T, 1)\mathbf{Y}(0) - J\mathbf{Y}(0)\|^2) \leq \lambda_2^T E(\|\mathbf{Y}(0) - J\mathbf{Y}(0)\|^2) = (m-1)\lambda_2^T, \text{ and} \\ E(\|\mathbf{Z}(t)\|^2) \leq \sum_{\tau=1}^t u \left(\frac{1}{m} + \frac{m-1}{m} \lambda_2^{t-\tau} \right) \cdot (1 + (m-1)\lambda_2^{\tau-1} + E(\|\mathbf{Z}(\tau-1)\|^2)).$$

By induction,

$$E(\|\mathbf{Z}(\tau)\|^2) \leq (1+u)^\tau + (m-1)(\lambda_2 + u)^\tau - 1 - (m-1)\lambda_2^\tau,$$

for $\tau = 1, 2, \dots, T-1$. The proof can be completed by combining the above results. ■

Using this lemma, we can establish the following.

Theorem 3: For a connected network with associated eigenvalue λ_2 , (i) if a gossip-based weighted-sum protocol achieves distortion $D < 1/4m$, then

$$E(R_{\text{WS}}(D)) \geq \frac{m-1}{2m} \left(\ln \frac{m-1}{mD} \right) \left(\log \frac{1}{4mD} \right), \text{ and}$$

(ii) there exists an $m(D)$ and a gossip-based weighted-sum protocol such that for all $m \geq m(D)$,

$$E(R_{\text{WS}}(D)) \leq \frac{1}{m\bar{\lambda}_2} \left(\ln \frac{2}{D} \right) \left(\log \frac{\ln(2/D)}{m^2\bar{\lambda}_2 D} \right),$$

where $\bar{\lambda}_2 = 1 - \lambda_2$.

Proof outline: (i) We follow the distortion analysis of Theorem 2 to show that $d \leq 4mD$ and then use Lemma 3.

(ii) We consider the optimal stochastic matrix Q^* with eigenvalue λ_2 for the given network topology and set $T = \ln(2/D)/\bar{\lambda}_2$, and $d = m^2\bar{\lambda}_2 D/\ln(2/D)$. Then, we show that $\lim_{m \rightarrow \infty} (1/mD) E(\|\mathbf{S}(T) - J\mathbf{S}(0)\|^2) < 1$. The average distortion D is achievable for average rate

$$E(R) = \frac{T}{m} \log \frac{1}{d} = \frac{1}{m\bar{\lambda}_2} \left(\ln \frac{2}{D} \right) \left(\log \frac{\ln(2/D)}{m^2\bar{\lambda}_2 D} \right).$$

Since Gaussian sources are the hardest to compress, this upper bound is also an upper bound for non-Gaussian sources.

Remarks:

1. For a complete graph, $\lambda_2 = 1 - 1/(m-1)$, and the upper and lower bounds of Theorem 3 differ by a factor of $\log \log m$ for distortion $D = \Omega(1/m \log m)$ and by a constant factor, otherwise. The lower bound of Theorem 2 is also a lower bound on $E(R_{\text{WS}}(D))$. The above two lower bounds differ by a constant factor for $D = \Omega(m^{-c})$ and $c > 0$ and by a factor of $\log(1/D)/\log m$ for $D = o(m^{-c})$ and $c > 0$.

2. For the star network considered in Subsection III-A, $\lambda_2 = 1 - 1/(2(m-1))$ and the upper bounds of Theorem 3 and of Proposition 1 differ by a factor of $(\log \log m) \log m$ for $D = \Omega(1/m \log m)$ and $\log m$ for $D = o(1/m \log m)$ and $D = \Omega(m^{-c})$, $c > 0$. The $\log m$ factor quantifies the penalty of using the gossip-based distributed protocols.

V. DISCUSSION

There are many questions that would be interesting to explore. For example: (i) We have investigated distributed weighted-sum protocols with a time-invariant normalized local distortion d . Can the order of the rate be reduced by making d vary with time? (ii) The distributed weighted-sum protocols do not take advantage of the build-up of correlation in the network. Using Wyner-Ziv coding can indeed reduce the rate. It would be interesting to find bounds with side information.

REFERENCES

- [1] S. P. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: design, analysis and applications," in *INFOCOM*. IEEE, 2005, pp. 1653–1664.
- [2] A. Olshevsky and J. Tsitsiklis, "Convergence rates in distributed consensus and averaging," *Decision and Control, 2006 45th IEEE Conference on*, pp. 3387–3392, Dec. 2006.
- [3] L. Xiao, S. P. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *J. Parallel Distrib. Comput.*, vol. 67, no. 1, pp. 33–46, 2007.
- [4] M. Yildiz and A. Scaglione, "Coding with side information for rate-constrained consensus," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3753–3764, Aug. 2008.
- [5] O. Ayaso, D. Shah, and M. Dahleh, "Information theoretic bounds on distributed computation," *Submitted IEEE Transactions on Information Theory*, 2008.
- [6] Y. Oohama, "Rate-distortion theory for gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2577–2593, 2005.
- [7] H. Su and A. E. Gamal, "Distributed lossy averaging." [Online]. Available: <http://arXiv.org/abs/0901.4134>
- [8] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.