Distributed Lossy Averaging

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*—An information theoretic formulation of distributed averaging is presented. We assume a network with m nodes each observing an i.i.d. 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. The network rate distortion function $R^*(D)$ for a 2-node network with correlated Gaussian sources is established. A general cutset lower bound on $R^*(D)$ with independent Gaussian sources is established and shown to be achievable to within a factor of 2 via a centralized protocol. 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. 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 is established. The results suggest that using distributed protocols results in a factor of $\log m$ increase in communication relative to centralized protocols.**

I. Introduction

Distributed averaging is a popular example of the distributed consensus problem, which has been receiving much attention recently due to interest in applications ranging from distributed coordination of autonomous agents to 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. We assume that each node in the network observes a 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 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, it deals with a different formulation than 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 the network rate distortion function for a 2-node network. In Section IV, 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 V, 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 $i = 1, 2, \ldots, m$ observes an i.i.d. source X_i. 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 distortion D. The following definitions apply to any set of correlated sources $(X_1, X_2, \ldots, X_m)$. In Sections IV and V, we assume that the sources are independent white Gaussian noise (WGN) processes each with average power of one.

The topology of the network is specified by a connected graph $(\mathcal{M}, \mathcal{E})$ without self loops, where $\mathcal{M} = \{1, 2, \ldots, m\}$ is the set of nodes and $\mathcal{E}$ is a set of undirected edges (node pairs) $\{i, j\}$, $i, j \in \mathcal{M}$ for $i \neq j$. Communication is performed in *rounds* and each round is divided into *time slots*. Each round may consist of a different number of time slots, and each time slot may consist of a different number of transmissions. One edge (node pair) 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 communicate in a round robin manner with the first node communicating in odd time slots and the second node communicating in even time slots. 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

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

number of communication rounds T, (ii) the sequence of edge selections, and (iii) the block codes for each selected node pair in each round used to perform communication and local computation. 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 node in each round and 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 $nr_i(t)$, where $r_i(t)$ is the transmission *rate* per source symbol, then the total transmission rate 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 $i = 1, 2, \dots, m$ assigns an estimate $S_{i1}^n := (S_{i1}, S_{i2}, \dots, S_{in})$ of the average $S_1^n := (S_1, S_2, \dots, S_n)$ 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

$$\frac{1}{mn}\sum_{i=1}^{m}\sum_{k=1}^{n} \mathrm{E}\left((S_k - S_{ik})^2\right),$$

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 average per-node rate R such that

$$\limsup_{n\to\infty} \frac{1}{mn}\sum_{i=1}^{m}\sum_{k=1}^{n} \mathrm{E}\left((S_k - S_{ik})^2\right) \le \Delta.$$

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. Clearly here we only need to consider deterministic averaging protocols.

We are also interested in the expected network rate distortion function for a random averaging protocol. We consider the expected per-node transmission rate $\mathrm{E}(R) = (1/m)\sum_{i=1}^{m} \mathrm{E}(R_i)$, and the limit on the expected distortion with respect to edge selection statistics $\mathrm{E}(\Delta)$ specified by the random averaging protocol. The *expected network rate-distortion function* $\mathrm{E}(R(D))$ for a random averaging protocol is defined as the infimum over all expected per-node rate $\mathrm{E}(R)$ such that the pairs (R, Δ) are achievable and $\mathrm{E}(\Delta) \le D$. Clearly for any random averaging protocol, $R^*(D) \le \mathrm{E}(R(D))$. Further, any upper bound on $\mathrm{E}(R(D))$ 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, such as the distributed weighted-sum protocol discussed in Section V, 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. In a centralized protocol the code can depend on the node identities. For example, a node may be designated as a "cluster-head" and treated differently by the protocol than other nodes.

III. $R^*(D)$ for 2-Node Network

Consider a network with 2 nodes and a single edge, and assume (X_1, X_2) to be correlated WGN sources with average powers P_1 and P_2, respectively, and covariance σ_{12}. Assume each node wishes to compute the weighted sum $g(X_1, X_2) = a_1X_1 + a_2X_2$, for some constants a_1 and a_2, to within mean squared error distortion D. In Section V, it will become clear why we consider weighted-sum computation instead of the less general averaging case.

For a 2-node network, there is only one round of communication with an arbitrary number of time slots. The interesting case is when distortion is small enough such that each node must transmit to the other node. The following proposition establishes the network rate distortion function for this regime.

Proposition 1: The network rate distortion function for a 2-node network is

$$R^*(D) = \frac{1}{2}\log\left(\left(1 - \frac{\sigma_{12}^2}{P_1P_2}\right)\frac{a_1a_2\sqrt{P_1P_2}}{D}\right)$$

for $D < \min\left\{\left(P_1 - \sigma_{12}^2/P_2\right)a_1^2, \left(P_2 - \sigma_{12}^2/P_1\right)a_2^2\right\}$.
The converse follows by a cutset bound argument. Achievability is proved simply by having each node independently compress its source and send the compressed version to the other node using Wyner-Ziv coding [8]. *Remarks*:

1. In [9], Kaspi investigated the interactive lossy source coding problem when the objective is for each source to obtain a description of the other source. Our problem is different and as such Kaspi's results do not readily apply. In [10], the interactive communication problem for asymptotically lossless computation is investigated. Again their results do not apply to our setting because they do not consider loss.
2. Finding the rate distortion function for a 3-node network even with Gaussian sources is difficult because (i) several feasible edge sequences are allowed and it is not a priori clear which sequence yields the optimal rate, (ii) the codes allow relaying in addition to interactive communication and local computing, and (iii) it is not known if Gaussian random codes are optimal. Results on a 3-node problem are reported in [11].

IV. Lower Bound on $R^*(D)$

Consider the m-node distributed lossy averaging problem when 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 $D \ge (m-1)/m^2$ and is lower bounded by

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

Proof: Since only pairwise communications are allowed, the number of bits transmitted by all nodes is equal to the number of bits received by all nodes. Denote the number of bits received by node i by $\tilde{R}_i$. Let M_i be the collection of indices sent from nodes $j \in \mathcal{M}\backslash\{i\}$ to node i. Node i computes an estimate S_{i1}^n, which is a function of its source X_{i1}^n and the received message M_i. Let $U_{ik} := (1/m)\sum_{j\in\mathcal{M}\backslash\{i\}} X_{jk}$. We bound the receiving rate as follows

$$\begin{aligned}\tilde{R}_i &\geq \frac{1}{n}H(M_i) \geq \frac{1}{n}H(M_i|X_{i1}^n) \geq \frac{1}{n}I(U_{i1}^n; M_i|X_{i1}^n)\\ &= \frac{1}{n}\sum_{k=1}^{n}\left(h(U_{ik}) - h(U_{ik}|U_{i1}^{k-1}, X_{i1}^n, M_i, S_{i1}^n)\right)\\ &\geq \frac{1}{n}\sum_{k=1}^{n}\left(h(U_{ik}) - h(U_{ik}|X_{ik}, S_{ik})\right)\\ &\geq \frac{1}{n}\sum_{k=1}^{n}\left(h(U_{ik}) - h(S_k - S_{ik})\right) \geq \frac{1}{2}\log\left(\frac{m-1}{m^2 D_i}\right),\end{aligned}$$

where $D_i = (1/n)\sum_{k=1}^n \mathrm{E}\left((S_k - S_{ik})^2\right)$. Using Jensen's inequality, we have $R^*(D) \geq (1/2)\log\left((m-1)/m^2 D\right)$. ■

Remarks:

1. As can be readily verified from Proposition 1, the above lower bound is achieved for $m = 2$.
2. The above cutset lower bound can be readily extended to correlated WGN sources and weighted-sum computation. The resulting bound is tight for $m = 2$ as shown in the previous section.

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

Consider a star network with m nodes and $m-1$ edges $\mathcal{E} = \{\{1,2\},\{1,3\},\ldots,\{1,m\}\}$. We use a centralized protocol where node 1 is treated as a "cluster head." The protocol has $T = 2m-3$ rounds. In round $t = 1,2,\ldots,m-1$, node $i = t+1$ compresses its source X_{i1}^n using Gaussian random codes with average distortion $(d/n)\sum_{k=1}^n \mathrm{E}(X_{ik}^2) = d$ and sends the index $M_i(X_{i1}^n)$ to node 1. Node 1 finds the corresponding reconstruction sequences $\hat{X}_{i1}^n(M_i)$ and computes the estimates

$$\begin{aligned}S_{11}^n &:= \frac{1}{m}X_{11}^n + \frac{1}{m}\sum_{i=2}^{m}\hat{X}_{i1}^n,\\ U_{i1}^n &:= \frac{1}{m}X_{11}^n + \frac{1}{m}\sum_{j\in\mathcal{M}\backslash\{1,i\}}\hat{X}_{j1}^n\end{aligned}$$

for $i = 2,3,\ldots,m$. In round $t = m-1, m, \ldots 2m-3$, node 1 compresses the estimate $U_{2m-t-1,1}^n$ using Gaussian random codes with average distortion $(d_1/n)\sum_{k=1}^n \mathrm{E}(U_{2m-t-1,k}^2)$ and sends the index $\tilde{M}_{2m-t-1}(U_{2m-t-1,1}^n)$ to nodes $2m-t-1$. Node i computes the estimate $S_{i1}^n = (1/m)X_{i1}^n + \hat{U}_{i1}^n$ for $i = 2,3,\ldots,m$, where $\hat{U}_{i1}^n$ is the reproduction sequence of U_{i1}^n corresponding to the index $\tilde{M}_i$. This establishes the following upper bound on the network rate distortion function.

Proposition 2: 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 $m \to \infty$ is less than or equal to 2. Thus a centralized protocol can achieve a rate within a factor of 2 of the cutset bound.

V. Distributed Weighted-Sum Protocols

Again assume that the sources $(X_1, X_2, \ldots, X_n)$ 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 $S_{i1}^n(0) = X_{i1}^n$ of the true average S_1^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 $S_{i1}^n(t)$ using Gaussian random codes with distortion $D_i(t+1) = (d/n)\sum_{k=1}^n \mathrm{E}(S_{ik}^2(t))$, and sends the index $M_i(t+1)(S_{i1}^n(t))$ to node j at rate $r = (1/2)\log(1/d)$. In the second time slot, node j similarly compresses $S_{j1}^n(t)$ using Gaussian random codes with distortion $D_j(t+1) = (d/n)\sum_{k=1}^n \mathrm{E}(S_{jk}^2(t))$, and sends the index $M_j(t+1)(S_{j1}^n(t))$ to node i at the same rate r. Upon receiving the indices, nodes i and j update their estimates

$$\begin{aligned}S_{i1}^n(t+1) &= \frac{1}{2}S_{i1}^n(t) + \frac{1}{2(1-d)}\hat{S}_{j1}^n(t),\\ S_{j1}^n(t+1) &= \frac{1}{2}S_{j1}^n(t) + \frac{1}{2(1-d)}\hat{S}_{i1}^n(t),\end{aligned} \tag{1}$$

where $\hat{S}_{i1}^n(t)$ and $\hat{S}_{j1}^n(t)$ 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 $S_{i1}^n = S_{i1}^n(T)$ if it is involved in at least one round of communication, and sets $S_{i1}^n = (1/m)S_{i1}^n(0)$, otherwise.

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

Remark: Note that the weighted-sum 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 general 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 update equations for the distributed weighted-sum protocols are linear, the initial estimates are WGN, and Gaussian test channels with independent additive WGN are used, the estimates $S_{i1}(t), \ldots, S_{in}(t)$ are i.i.d. and the estimates $S_{1k}(t), \ldots, S_{mk}(t)$ are jointly Gaussian. From this point on we suppress the transmission symbol index k.

A. Lower Bound on $R^*_{\mathrm{WS}}(D)$

We establish a lower bound on $R^*_{\mathrm{WS}}(D)$ 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 $\mathcal{T}_i := \{t_{i1}, t_{i2}, \ldots, t_{iT_i}\}$, for $i = 1, 2, \ldots, m$, 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 $S_i(t)$ to prove the lower bound.

Lemma 1: For a distributed weighted-sum protocol, the estimate at node i at the end of round t can be expressed as

$$S_i(t) = \sum_{j=1}^m \gamma_{ij}(t) S_j(0) + Z_i(t),$$

where $Z_i(t)$ is Gaussian and independent of the sources $(X_1, X_2, \ldots, X_m)$. Furthermore, the diagonal coefficients satisfy the property

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

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

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

Using these lemmas, we can establish the following lower bound.

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

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

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 j_2 is

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

where $W_{j_1\tau} \sim \mathcal{N}\left(0, \mathrm{E}\left(S_{j_1}(t_{j_1\tau} - 1)^2\right) d/(1-d)\right)$. By induction, we can show that the estimate of node i at time $t \geq t_{j_1\tau}$ has the form $S_i(t) = (1/2)\beta_i(t) W_{j_1\tau} + \tilde{S}_i(t)$, 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

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

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

$$\frac{1}{m}\sum_{i=1}^m \mathrm{E}\left(S - S_i(T))^2\right) \geq \sum_{i=1}^m \sum_{k=1}^{T_i} \frac{d}{4m^2(1-d)2^{2(k-1)}} \geq \frac{d}{4m}.$$

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 a centralized protocol for the star network can achieve within a factor of 2 of the cutset bound, this suggests that the $\log m$ factor is the penalty of using a distributed versus centralized protocols.

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

In this section, we establish bounds on $\mathrm{E}(R_{\mathrm{WS}}(D))$ 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 $R^*(D) \leq \mathrm{E}(R_{\mathrm{WS}}(D))$.

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 edge selection process is equivalent to the asynchronous model in [1]. After the edge (node pair) $\{i, j\}$ is selected, the *distributed weighted-sum* coding scheme previously described is performed. Let $\mathbf{S}(t) := [S_1(t)\ S_2(t)\ \ldots\ S_m(t)]^T$ and rewrite the update equations (1) in the matrix form

$$\mathbf{S}(t+1) = A(t+1)\mathbf{S}(t) + \mathbf{W}(t+1), \tag{2}$$

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, and (ii) $\mathbf{W}(t+1) = W_i(t+1)\mathbf{e}_i + W_j(t+1)\mathbf{e}_j$, where $\mathbf{e}_i$ and $\mathbf{e}_j$ are unit vectors along the i-th and j-th axes, and $W_i(t+1) \sim \mathcal{N}\left(0, \mathrm{E}\left(S_j(t)^2 d/4(1-d)\right)\right)$ and $W_j(t+1) \sim \mathcal{N}\left(0, \mathrm{E}\left(S_i(t)^2 d/4(1-d)\right)\right)$ 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 $A := \mathrm{E}(A(0))$, 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 $\mathrm{E}(R_{\mathrm{WS}}(D))$.

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

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

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 $\mathbf{S}(T)$ as $\mathbf{S}(T) = A(T,1)\mathbf{S}(0) + \mathbf{Z}(T)$, where $\mathbf{Z}(T) = \sum_{t=1}^{T} A(T,t+1)\mathbf{W}(t)$ and

$$A(t_2,t_1) = \begin{cases} A(t_2)A(t_2-1)\dots A(t_1) & \text{if } t_2 \geq t_1 \\ I & \text{if } t_2 < t_1. \end{cases}$$

Consider the sum of distortions over all nodes

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

We can show that

$$\mathrm{E}\left(\|A(T,1)\mathbf{S}(0)-J\mathbf{S}(0)\|^2\right) \leq \lambda_2^T \mathrm{E}\left(\|\mathbf{S}(0)-J\mathbf{S}(0)\|^2\right) = (m-1)\lambda_2^T, \text{ and}$$

$$\mathrm{E}\left(\|\mathbf{Z}(t)\|^2\right) \leq \sum_{\tau=1}^{t} u\left(\frac{1}{m}+\frac{m-1}{m}\lambda_2^{t-\tau}\right) \cdot \left(1+(m-1)\lambda_2^{\tau-1}+\mathrm{E}\left(\|\mathbf{Z}(\tau-1)\|^2\right)\right).$$

By induction,

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

for $\tau = 1, 2, \dots, T-1$. Thus,

$$\mathrm{E}\left(\|\mathbf{Z}(T)\|^2\right) \leq \frac{m-1}{m}\frac{u}{1-\lambda_2+u}\left((1+u)^T-\lambda_2^T\right) + \frac{1}{m}\left((1+u)^T-1\right) + \frac{m-1}{m}\frac{u}{1-\lambda_2-u}\left(1-(\lambda_2+u)^T\right) + \frac{(m-1)^2}{m}\left((\lambda_2+u)^T-\lambda_2^T\right).$$

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

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

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

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

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

Proof outline:
(i) We follow the distortion analysis in the proof 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 $\lim_{m\to\infty}(1/mD)\mathrm{E}\left(\|\mathbf{S}(T)-J\mathbf{S}(0)\|^2\right) < 1$. The average distortion D is achievable for average rate

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

Since Gaussian sources are the hardest to compress, we can show that the above upper bound is also an upper bound for general, 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. On the other hand, the lower bound of Theorem 2 is also a lower bound on $\mathrm{E}(R_{\mathrm{WS}}(D))$. 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 IV-A, $\lambda_2 = 1 - 1/(2(m-1))$ and the upper bound of Theorem 3 differs from the upper bound in Subsection IV-A 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.

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 computatiton," *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.

[9] A. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 735–740, Nov. 1985.

[10] N. Ma and P. Ishwar, "Two-terminal distributed source coding with alternating messages for function computation," July 2008, pp. 51–55.

[11] P. Cuff, H. Su, and A. E. Gamal, "Cascade multiterminal source coding," *Submitted to 2009 IEEE International Symposium on Information Theory.*