A New Information-Theoretic Lower Bound for Distributed Function Computation

Aolin Xu and Maxim Raginsky

***Abstract*—This paper presents an information-theoretic lower bound on the minimum time required by any scheme for distributed computation over a network of point-to-point channels with finite capacity to achieve a given accuracy with a given probability. This bound improves upon earlier results by Ayaso *et al.* and by Como and Dahleh, and is derived using a combination of cutset bounds and a novel lower bound on conditional mutual information via so-called small ball probabilities. In the particular case of linear functions, the small ball probability can be expressed in terms of Lévy concentration functions of sums of independent random variables, for which tight estimates are available under various regularity conditions, leading to strict improvements over existing results in certain regimes.**

I. Introduction

Distributed computation over networks of finite-capacity channels arises in such applications as estimation or inference in sensor networks and consensus or coordination of multiple agents with finite-capacity communication links. This paper focuses on the case where the nodes are connected by independent and memoryless point-to-point channels. Each node aims to compute a common function of random initial observations of all the nodes through local communication and computation. We use tools from information theory to obtain a lower bound on the minimum time needed by any scheme to achieve a certain accuracy at each node. Previously, this problem has been considered by Ayaso *et al.* [1] and by Como and Dahleh [2]. Some achievability results can be found in [1], [3]–[5].

The key quantity we examine is the conditional mutual information between the function value and its estimates made by an arbitrary subset of nodes, given the local observations in this subset. An upper bound on this mutual information in terms of cutset capacity brings communication constraints into the picture, while a novel lower bound on it via a certain quantity we call the *small ball probability* makes it possible to explicitly capture the influence of the joint distribution of local observations and the structure of the function of interest on the computation time. For linear functions, the small ball probability can be expressed in terms of so-called *Lévy concentration functions* [6], for which tight estimates are available under various regularity conditions. In addition, the presence of the small ball probability in our bound highlights an operational similarity between distributed computation over noisy channels and distributed joint source-channel coding.

The authors are with the Department of Electrical and Computer Engineering and the Coordinated Science Laboratory, University of Illinois, Urbana, IL 61801, USA. E-mail: {aolinxu2,maxim}@illinois.edu. This work was supported by NSF grant CCF-1017564.

In general, our lower bound on the computation time improves upon existing results. For instance, the bounds of Ayaso *et al.* [1] involve differential entropy, and an additional perturbation technique is needed to address the case when the nodes wish to compute the same function of their local observations. By contrast, our proof technique avoids differential entropy and yields lower bounds that are tighter than those in [1] — in particular, our lower bounds approach infinity as the accuracy parameter tends to zero. A continuum generalization of Fano's inequality proposed by Como and Dahleh [2] can also be used to derive lower bounds on computation time, but the resulting bounds are generally looser than our bound. Moreover, for observations with log-concave distributions, we can weaken our bound to a version involving conditional differential entropy, which is tighter than similar results obtained in [2].

II. The model and performance metrics

We mainly follow the notation of [2]. The network is represented by a directed graph $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is a finite set of vertices (or nodes) and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a set of edges. We assume that G is connected and simple (i.e., has no self-loops). Initially, each node $v \in \mathcal{V}$ has access to a local observation, given by a random variable (r.v.) W_v taking values in some space W_v. The probability law $\mathbb{P}_W$ of the vector of all local observations $W = (W_v)_{v\in\mathcal{V}}$ is assumed to be known to all the nodes. Given a function $f : \prod_{v\in\mathcal{V}} \mathsf{W}_v \to \mathsf{Z}$, the objective is for each node to estimate the value $Z = f(W) = f\big((W_v)_{v\in\mathcal{V}}\big)$ via local communication and computation.

To each edge $e \in \mathcal{E}$, we associate a memoryless channel with input alphabet X_e, output alphabet Y_e, and stochastic transition law K_e. The channels corresponding to different edges are assumed to be independent. Thus, we can associate to the network a memoryless channel with input alphabet X, output alphabet Y, and transition law K, where

$$\mathsf{X} = \prod_{e\in\mathcal{E}} \mathsf{X}_e, \quad \mathsf{Y} = \prod_{e\in\mathcal{E}} \mathsf{Y}_e, \quad K(y|x) = \prod_{e\in\mathcal{E}} K_e(y_e|x_e).$$

All processing takes place in discrete time. A *T-step algorithm* $\mathcal{A}$ is a collection of encoders $\big(\varphi_{v,t}\big)$ and estimators $\big(\psi_v\big)$, for all $v \in \mathcal{V}$ and $t = 1, \ldots, T$, given by mappings

$$\varphi_{v,t} : \mathsf{W}_v \times \mathsf{Y}_{(\bullet,v)}^{t-1} \to \mathsf{X}_{(v,\bullet)}, \quad \psi_v : \mathsf{W}_v \times \mathsf{Y}_{(\bullet,v)}^{T} \to \mathsf{Z},$$

where we have defined $\mathsf{X}_{(v,\bullet)} \triangleq \prod_{u:\,(v,u)\in\mathcal{E}} \mathsf{X}_{(v,u)}$ and $\mathsf{Y}_{(\bullet,v)} \triangleq \prod_{u:\,(u,v)\in\mathcal{E}} \mathsf{Y}_{(u,v)}$. The algorithm operates as follows: at each time $t = 1, \ldots, T$, each node $v \in \mathcal{V}$ computes $X_{v,t} = \varphi_{v,t}\big(W_v, Y_v^{t-1}\big) \in \mathsf{X}_{(v,\bullet)}$ and then transmits the

messages $X_{(v,u),t}$ to all nodes u with $(v,u) \in \mathcal{E}$. The received messages $Y_{v,t} \in \mathsf{Y}_{(\bullet,v)}$, $v \in \mathcal{V}$, are related to the $X_{v,t}$'s via the transition law K, independently of W, X^{t-1}, Y^{t-1}. At time T, each node v computes $\widehat{Z}_v = \psi_v(W_v, Y_v^T)$. We will denote the collection of all T-step algorithms by $\mathfrak{A}(T)$.

Given a nonnegative *distortion function* $d : \mathsf{Z} \times \mathsf{Z} \to \mathbb{R}^+$, we use the excess distortion probability $\mathbb{P}(d(Z, \widehat{Z}_v) > \varepsilon)$ to quantify the estimation accuracy of node v at time T. For an algorithm $\mathcal{A}$, we require that, when the computation stops, each node has an estimate with distortion of no more than ε, with probability at least $1 - \delta$. This motivates the following definition: given an accuracy parameter $\varepsilon > 0$ and a confidence parameter $\delta \in [0, 1)$, we define the (ε, δ)-*computation time* by

$$T(\varepsilon,\delta) \triangleq \inf \Big\{ T \in \mathbb{N} : \exists \mathcal{A} \in \mathfrak{A}(T) \text{ such that } \max_{v \in \mathcal{V}} \mathbb{P}\left(d(Z, \widehat{Z}_v) > \varepsilon \right) < \delta \Big\}.$$

Our main contribution is a new information-theoretic lower bound on $T(\varepsilon, \delta)$ that, in addition to accounting for the communication constraints encoded in G and K, also captures explicitly the influence of $\mathbb{P}_W$ and f.

III. The main result and some implications

In order to state our main result, we first introduce some additional definitions and notation. The *cutset* corresponding to a set $\mathcal{S} \subset \mathcal{V}$ is the set $\mathcal{E}_{\mathcal{S}} \triangleq \{(u,v) \in \mathcal{E} : u \in \mathcal{S}^c, v \in \mathcal{S}\} \equiv (\mathcal{S}^c \times \mathcal{S}) \cap \mathcal{E}$, and the *cutset capacity* of $\mathcal{S}$ is

$$C_{\mathcal{S}} \triangleq \sum_{e \in \mathcal{E}_{\mathcal{S}}} C_e,$$

where C_e is the Shannon capacity of K_e. Also, for any $\varepsilon > 0$, $\mathcal{S} \subset \mathcal{V}$, and $w_{\mathcal{S}} \in \prod_{v \in \mathcal{S}} \mathsf{W}_v$, we define

$$L(w_{\mathcal{S}}, \varepsilon) \triangleq \sup_{z \in \mathsf{Z}} \mathbb{P}\Big(d(Z, z) \le \varepsilon \Big| W_{\mathcal{S}} = w_{\mathcal{S}} \Big).$$

As we will see in the next section, in certain cases this quantity may be related to so-called "small ball" probability estimates.

Theorem 1. *Suppose that the parameters ε, δ satisfy*

$$\max_{\mathcal{S} \subset \mathcal{V}} \mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)] \le 1 - \delta. \tag{1}$$

Then

$$T(\varepsilon,\delta) \ge \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left((1-\delta) \log \frac{1}{\mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)]} - h_2(\delta) \right), \tag{2}$$

where $h_2(\delta) \triangleq -\delta \log \delta - (1-\delta)\log(1-\delta)$ is the binary entropy.

Proof: The main idea of the proof is as follows. For every subset $\mathcal{S} \subset \mathcal{V}$, we examine the quantity $I(Z; \widehat{Z}_{\mathcal{S}} | W_{\mathcal{S}})$, where $\widehat{Z}_{\mathcal{S}} \triangleq (\widehat{Z}_v)_{v \in \mathcal{S}}$. A general lower bound on the computation time is obtained by upper-bounding this quantity in terms of the cutset capacity $C_{\mathcal{S}}$ (a standard step) and by lower-bounding it in terms of the small ball probability. These bounds are presented as Lemmas A.1 and A.2 in the Appendix.

Suppose that there exists an algorithm $\mathcal{A} \in \mathfrak{A}(T)$, such that

$$\max_{v \in \mathcal{V}} \mathbb{P}\left(d(Z, \widehat{Z}_v) > \varepsilon \right) < \delta.$$

Then, on the one hand, by Lemma A.1, for any set $\mathcal{S} \subset \mathcal{V}$,

$$I(Z; \widehat{Z}_{\mathcal{S}} | W_{\mathcal{S}}) \le T C_{\mathcal{S}}.$$

On the other hand, since ε and δ satisfy (1), Lemma A.2 ensures that, for any set $\mathcal{S} \subset \mathcal{V}$,

$$I(Z; \widehat{Z}_{\mathcal{S}} | W_{\mathcal{S}}) \ge \left((1-\delta) \log \frac{1}{\mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)]} - h_2(\delta) \right).$$

Therefore, for any set $\mathcal{S} \subset \mathcal{V}$,

$$T \ge \frac{1}{C_{\mathcal{S}}} \left((1-\delta) \log \frac{1}{\mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)]} - h_2(\delta) \right).$$

Maximizing over all $\mathcal{S} \subset \mathcal{V}$, we get (2). ∎

From an operational point of view, the lower bound of Theorem 1 reflects the fact that the problem of distributed function computation is, inherently, a joint source-channel coding (JSCC) problem. In particular, the lower bound on $I(Z; \widehat{Z}_{\mathcal{S}} | W_{\mathcal{S}})$ can be interpreted in terms of a reduction of JSCC to (generalized) list decoding [7, Sec. III.B]. Given any algorithm $\mathcal{A}$ and any node $v \in \mathcal{V}$, we may construct a "list decoder" as follows: given the estimate $\widehat{Z}_v$, we generate a "list" $\{z \in \mathsf{Z} : d(z, \widehat{Z}_v) \le \varepsilon\}$. If we fix a set $\mathcal{S} \subset \mathcal{V}$ and allow all the nodes in $\mathcal{S}$ to share their observations $W_{\mathcal{S}}$, then $\mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)]$ is an upper bound on the $\mathbb{P}_W$-measure of the list of any node $v \in \mathcal{S}$. The bound of Theorem 1 has the same limitations as all previously known bounds obtained via cutset arguments, since it captures only the flow of information across a cutset $\mathcal{E}_{\mathcal{S}}$, but not within $\mathcal{S}$. However, as we will demonstrate in the sections that follow, it is possible to exploit structural properties of the function f (such as linearity) and of the probability law $\mathbb{P}_W$ (such as log-concavity) to derive lower bounds on the computation time that are often tighter than existing bounds.

Finally, condition (1) requires the accuracy parameter ε to be small enough, and may be checked only for sets obtained by deleting a single node. Indeed, by the law of iterated expectation, for any $\mathcal{S} \subset \mathcal{S}' \subset \mathcal{V}$ we have

$$\begin{aligned} L(w_{\mathcal{S}}, \varepsilon) &= \sup_{z \in \mathsf{Z}} \mathbb{E}\Big[\mathbb{E}\Big[\mathbf{1}_{\{d(Z,z) \le \varepsilon\}} \Big| W_{\mathcal{S}'} \Big] \Big| W_{\mathcal{S}} = w_{\mathcal{S}} \Big] \\ &\le \mathbb{E}\Big[\sup_{z \in \mathsf{Z}} \mathbb{E}\Big[\mathbf{1}_{\{d(Z,z) \le \varepsilon\}} \Big| W_{\mathcal{S}'} \Big] \Big| W_{\mathcal{S}} = w_{\mathcal{S}} \Big] \\ &= \mathbb{E}\Big[L(W_{\mathcal{S}'}, \varepsilon) \Big| W_{\mathcal{S}} = w_{\mathcal{S}} \Big]. \end{aligned} \tag{3}$$

Therefore, the condition (1) will be automatically satisfied if

$$\max_{v \in \mathcal{V}} \sup_{w_{\mathcal{V} \setminus \{v\}}} L(w_{\mathcal{V} \setminus \{v\}}, \varepsilon) \le 1 - \delta. \tag{4}$$

Another immediate consequence of (3) is monotonicity of the set function $\mathcal{S} \mapsto \mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)]$: if $\mathcal{S} \subseteq \mathcal{S}'$, then $\mathbb{E}[L(W_{\mathcal{S}}, \varepsilon)] \le \mathbb{E}[L(W_{\mathcal{S}'}, \varepsilon)]$.

IV. Computation of linear functions

We now particularize Theorem 1 to a distributed computation problem of wide interest, namely the computation of linear functions. Specifically, we assume that the initial

observations $W_v, v \in \mathcal{V}$, are independent real-valued random variables, and the objective is to compute a linear function

$$Z = f(W) = \sum_{v \in \mathcal{V}} a_v W_v \tag{5}$$

for a fixed vector of coefficients $(a_v)_{v \in \mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|}$, subject to the absolute error criterion $d(z, \widehat{z}) = |z - \widehat{z}|$. We will use the following shorthand notation: for any set $\mathcal{S} \subset \mathcal{V}$, we let $a_{\mathcal{S}} = (a_v)_{v \in \mathcal{S}}$ and $\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle = \sum_{v \in \mathcal{S}} a_v W_v$.

The independence of the W_v's and the additive structure of f allow us to express the quantities $L(w_{\mathcal{S}}, \varepsilon)$ in terms of so-called *Lévy concentration functions* of random sums [6]. The Lévy concentration function of a real-valued r.v. U (also known as the "small ball probability") is defined as

$$\mathcal{L}(U, \rho) \triangleq \sup_{u \in \mathbb{R}} \mathbb{P}\left(|U - u| \leq \rho\right), \qquad \rho > 0.$$

If we fix a subset $\mathcal{S} \subset \mathcal{V}$ and a boundary condition $w_{\mathcal{S}}$, then

$$\begin{aligned} L(w_{\mathcal{S}}, \varepsilon) &= \sup_{z \in \mathbb{R}} \mathbb{P}\left(\Big| \sum_{v \in \mathcal{V}} a_v W_v - z \Big| \leq \varepsilon \Big| W_{\mathcal{S}} = w_{\mathcal{S}} \right) \\ &= \sup_{z \in \mathbb{R}} \mathbb{P}\left(\Big| \sum_{v \in \mathcal{S}^c} a_v W_v + \sum_{v \in \mathcal{S}} a_v w_v - z \Big| \leq \varepsilon \right) \\ &= \sup_{z \in \mathbb{R}} \mathbb{P}\left(\Big| \sum_{v \in \mathcal{S}^c} a_v W_v - z \Big| \leq \varepsilon \right) \\ &= \mathcal{L}\left(\langle a_{\mathcal{S}^c}, W_{\mathcal{S}^c} \rangle, \varepsilon \right), \end{aligned}$$

where in the second line we have used the fact that the W_v's are independent r.v.'s, while in the third line we have used the fact that, for any function $G : \mathbb{R} \to \mathbb{R}$ and any $a \in \mathbb{R}$, $\sup_z G(z) = \sup_z G(z + a)$. In other words, for a fixed $\mathcal{S}$, the quantity $L(w_{\mathcal{S}}, \varepsilon)$ is independent of the boundary condition $w_{\mathcal{S}}$, and is controlled by the probability law of the random sum $\langle a_{\mathcal{S}^c}, W_{\mathcal{S}^c} \rangle$, i.e., the part of the function f that depends on the initial observations at the nodes in $\mathcal{S}^c$.

The problem of estimating Lévy concentration functions of sums of independent random variables has a long history in the theory of probability — for random variables with densities, some of the first results go back at least to Kolmogorov [8], while for discrete random variables it is closely related to the so-called Littlewood–Offord problem [9]. We now illustrate how one can exploit available estimates for the concentration function under various regularity conditions to obtain tight lower bounds on the computation time for linear functions.

Gaussian sums. Suppose that the local observations W_v, $v \in \mathcal{V}$, are i.i.d. standard Gaussian random variables. Then, for any $\mathcal{S} \subseteq \mathcal{V}$, $\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle$ is a zero-mean Gaussian r.v. with variance $\|a_{\mathcal{S}}\|_2^2 = \sum_{v \in \mathcal{S}} a_v^2$ (here, $\|\cdot\|_2$ is the usual Euclidean ℓ_2 norm). A simple calculation shows that

$$L(w_{\mathcal{S}}, \varepsilon) = \mathcal{L}\left(N\left(0, \|a_{\mathcal{S}^c}\|_2^2\right), \varepsilon\right) \leq \sqrt{\frac{2}{\pi}} \frac{\varepsilon}{\|a_{\mathcal{S}^c}\|_2}.$$

Using this in Theorem 1 with the sufficient condition (4), we get the following:

Theorem 2. *Consider the problem of distributed computation of a linear function of the form* (5)*, where* $(W_v) \overset{\text{i.i.d.}}{\sim} N(0,1)$*. Suppose that the coefficients* a_v *are all nonzero, and that* ε *and* δ *satisfy the condition*

$$\varepsilon \leq \sqrt{\frac{\pi}{2}} (1 - \delta) \min_{v \in \mathcal{V}} |a_v|.$$

Then for the (ε, δ)*-computation time we have*

$$T(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1 - \delta}{2} \log \frac{\pi \|a_{\mathcal{S}^c}\|_2^2}{2\varepsilon^2} - h_2(\delta) \right).$$

Thus, the lower bound on the computation time for (5) depends on the vector of coefficients a only through its ℓ_2 norm.

Sums of independent r.v.'s with log-concave distributions. Another instance in which sharp bounds on the concentration function are available is when the initial observations of the nodes are independent random variables with log-concave distributions (we recall that a real-valued r.v. U is said to have a log-concave distribution if it has a density of the form $p_U(u) = e^{-F(u)}$, where $F : \mathbb{R} \to (-\infty, +\infty]$ is a convex function; this includes Gaussian, Laplace, uniform, etc.). The following result was obtained recently by Bobkov and Chistyakov [10, Theorem 1.1]: Let $U_1, \ldots, U_k$ be independent random variables with log-concave distributions, and let $S_k = U_1 + \ldots + U_k$. Then, for any $\rho \geq 0$,

$$\frac{1}{\sqrt{3}} \frac{\rho}{\sqrt{\mathrm{Var}(S_k) + \rho^2/3}} \leq \mathcal{L}(S_k, \rho) \leq \frac{2\rho}{\sqrt{\mathrm{Var}(S_k) + \rho^2/3}}. \tag{6}$$

Theorem 3. *Consider the problem of distributed computation of a linear function of the form* (5)*, where the* W_v*'s are independent random variables with log-concave distributions and with variances at least* σ^2*. Suppose that the coefficients* a_v *are all nonzero, and that* ε *and* δ *satisfy the condition*

$$\left(1 - \frac{(1 - \delta)^2}{12}\right) \varepsilon^2 \leq \frac{\sigma^2 (1 - \delta)^2}{4} \min_{v \in \mathcal{V}} |a_v|^2.$$

Then for the (ε, δ)*-computation time we have*

$$T(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1 - \delta}{2} \log \left(\frac{\sigma^2 \|a_{\mathcal{S}^c}\|_2^2}{4\varepsilon^2} + \frac{1}{12} \right) - h_2(\delta) \right).$$

Proof: For each $v \in \mathcal{V}$, $a_v W_v$ also has a log-concave distribution, and, for any $\mathcal{S} \subset \mathcal{V}$,

$$\mathrm{Var}(\langle a_{\mathcal{S}^c}, W_{\mathcal{S}^c} \rangle) = \sum_{v \in \mathcal{S}^c} |a_v|^2 \mathrm{Var}(W_v) \geq \|a_{\mathcal{S}^c}\|_2^2 \sigma^2.$$

The lower bound follows from Theorem 1 and from (6). ∎

Sums of independent r.v.'s with bounded third moments. It is known that random variables with log-concave distributions have bounded moments of any order. Under a much weaker assumption that the local observations W_v, $v \in \mathcal{V}$ have bounded third moments, we can prove the following:

Theorem 4. *Consider the problem of computing the linear function* (5)*, where the* W_v*'s are independent zero-mean r.v.'s with variances at least* 1 *and with third moments bounded by* B*, and the coefficients* a_v *satisfy the constraint* $K_1 \leq |a_v| \leq$

K_2 for some $K_1, K_2 > 0$. There exists an absolute constant $c > 0$, such that, for any ε, δ satisfying the condition

$$M(\varepsilon) \triangleq c\left(\varepsilon/K_1 + B(K_2/K_1)^3\right) \leq 1 - \delta,$$

we must have

$$T(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1-\delta}{2} \log \frac{|\mathcal{V} \setminus \mathcal{S}|}{M^2(\varepsilon)} - h_2(\delta) \right).$$

Proof: Under the conditions of the theorem, a small ball estimate due to Rudelson and Vershynin [11, Corollary 2.10] can be used to show that, for any $\mathcal{S} \subset \mathcal{V}$,

$$\mathcal{L}(\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle, \varepsilon) \leq \frac{M(\varepsilon)}{\sqrt{|\mathcal{S}|}}.$$

The desired conclusion follows immediately. ∎

Random Rademacher sums. We close by considering a case when the local observations W_v have discrete distributions. Specifically, let the W_v's be i.i.d. Rademacher random variables, i.e., each W_v takes values ± 1 with equal probability. In this case, the Lévy concentration function $\mathcal{L}(\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle, \varepsilon)$ will be highly sensitive to the *direction* of the vector $a_{\mathcal{S}}$, rather than just its norm. For example, consider the extreme case when $a_v = |\mathcal{V}|$ for a single node $v \in \mathcal{S}$, and all other coefficients are zero. Then $\mathcal{L}(\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle, 0) = \mathcal{L}(|\mathcal{V}| W_v, 0) = 1/2$. On the other hand, if $a_v = 1$ for all $v \in \mathcal{V}$ and $|\mathcal{S}|$ is even, then

$$\mathcal{L}(\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle, 0) = \binom{|\mathcal{S}|}{|\mathcal{S}|/2} / 2^{|\mathcal{S}|} \simeq \sqrt{\frac{2}{\pi |\mathcal{S}|}},$$

where the last step is due to Stirling's approximation. This immediately gives us a lower bound on the time any algorithm must take in order for each node to compute the value of $\sum_{v \in \mathcal{V}} W_v$ *exactly* with probability at least $1 - \delta$:

$$T(0, \delta) \gtrsim \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1-\delta}{2} \log \frac{\pi |\mathcal{V} \setminus \mathcal{S}|}{2} - h_2(\delta) \right). \quad (7)$$

Moreover, even relaxing the zero-error requirement will not decrease the computation time substantially. Indeed, a celebrated result due to Littlewood and Offord, improved later by Erdős [12], says that, if $|a_v| \geq 1$ for all v, then

$$\mathcal{L}(\langle a_{\mathcal{S}}, W_{\mathcal{S}} \rangle, 1) \leq \binom{|\mathcal{S}|}{\lfloor |\mathcal{S}|/2 \rfloor} / 2^{|\mathcal{S}|} \simeq \sqrt{\frac{2}{\pi |\mathcal{S}|}},$$

which translates into a lower bound on the $(1, \delta)$-computation time which is of the same order as the right-hand side of (7).

V. Comparison with existing work

We now illustrate the utility of the above bounds through comparison with some existing results. For example, Ayaso *et al.* [1] derive lower bounds on a related quantity

$$T'(\varepsilon, \delta) \triangleq \inf \Big\{ T \in \mathbb{N} : \exists \mathcal{A} \in \mathfrak{A}(T) \text{ such that } \max_{v \in \mathcal{V}} \mathbb{P}\left(\widehat{Z}_v \notin [(1-\varepsilon)Z, (1+\varepsilon)Z] \right) < \delta \Big\}.$$

One of their results is as follows: if $Z = f(W)$ is a linear function of the form (5) and $(W_v) \overset{\text{i.i.d.}}{\sim} \text{Uniform}([1, 1+B])$ for some $B > 0$, then

$$T'(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{|\mathcal{S}|}{2C_{\mathcal{S}}} \log \frac{1}{B\varepsilon^2 + \kappa\delta + (1/B)^{2/|\mathcal{V}|}} \quad (8)$$

for all sufficiently small $\varepsilon, \delta > 0$, where $\kappa > 0$ is a fixed constant [1, Theorem III.5]. Let us compare (8) with what we can obtain using our techniques. It is not hard to show that

$$T'(\varepsilon, \delta) \geq T\big(\|a\|_1 (1+B)\varepsilon, \delta\big), \quad (9)$$

where $\|a\|_1 = \sum_{v \in \mathcal{V}} |a_v|$ is the ℓ_1 norm of a. Moreover, since any r.v. uniformly distributed on a bounded interval of the real line has a log-concave distribution, we can use Theorem 3 to lower-bound the right-hand side of (9). This gives

$$T'(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1-\delta}{2} \log \frac{B^2 \|a_{\mathcal{S}^c}\|_2^2}{48(B+1)^2 \|a\|_1^2 \varepsilon^2} - h_2(\delta) \right) \quad (10)$$

for all sufficiently small $\varepsilon, \delta > 0$. We immediately see that this bound is tighter than the one in (8). In particular, the right-hand side of (8) remains bounded for vanishingly small ε and δ, and in the limit of $\varepsilon, \delta \to 0$ tends to

$$\max_{\mathcal{S} \subset \mathcal{V}} \frac{|\mathcal{S}|}{C_{\mathcal{S}}} \frac{\log B}{|\mathcal{V}|} \leq \frac{\log B}{\min_{\mathcal{S} \subset \mathcal{V}} C_{\mathcal{S}}}.$$

By contrast, as $\varepsilon, \delta \to 0$, the right-hand side of (10) grows without bound as $\log(1/\varepsilon)$.

Another lower bound on the (ε, δ)-computation time $T(\varepsilon, \delta)$ was obtained by Como and Dahleh [2]. Their starting point is the following continuum generalization of Fano's inequality [2, Lemma 2]: if $Z, \widehat{Z}$ are two jointly distributed real-valued r.v.'s, such that $\mathbb{E}Z^2 < \infty$, then, for any $\varepsilon > 0$,

$$h(Z|\widehat{Z}) \leq \mathbb{P}\big(|Z - \widehat{Z}| \leq \varepsilon\big) \log \varepsilon + \frac{1}{2} \log\big(16\pi e \mathbb{E}Z^2\big), \quad (11)$$

where $h(Z|\widehat{Z})$ is the conditional differential entropy of Z given $\widehat{Z}$. If we use (11) instead of Lemma A.2 to lower-bound $I(Z; \widehat{Z}_{\mathcal{S}} | W_{\mathcal{S}})$ for each $\mathcal{S} \subset \mathcal{V}$, then we get

$$T(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1-\delta}{2} \log \frac{1}{\varepsilon^2} + h(Z|W_{\mathcal{S}}) - \frac{1}{2} \log\big(16\pi e \mathbb{E}Z^2\big) \right). \quad (12)$$

Again, let us consider the case when $Z = f(W)$ is a linear function of the form (5) with all a_v nonzero and with $(W_v) \overset{\text{i.i.d.}}{\sim} N(0,1)$. Then (12) becomes

$$T(\varepsilon, \delta) \geq \max_{\mathcal{S} \subset \mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left(\frac{1-\delta}{2} \log \frac{1}{\varepsilon^2} + \frac{1}{2} \log \frac{\|a_{\mathcal{S}^c}\|_2^2}{8\|a\|_2^2} \right). \quad (13)$$

The lower bound of our Theorem 2 will be tighter than (13) for all $\varepsilon > 0$ as soon as

$$\frac{1-\delta}{2} \log \frac{\pi \|a_{\mathcal{S}^c}\|_2^2}{2} - h_2(\delta) \geq \frac{1}{2} \log \frac{\|a_{\mathcal{S}^c}\|_2^2}{8\|a\|_2^2} \quad (14)$$

for all $\mathcal{S} \subset \mathcal{V}$ (note that the quantity on the right-hand side is nonpositive). More generally, when the local observations W_v, $v \in \mathcal{V}$, have log-concave distributions, the bound of Theorem 1 can be weakened to get a lower bound involving the conditional differential entropies $h(Z|W_{\mathcal{S}})$, $\mathcal{S} \subset \mathcal{V}$. Suppose the objective is for each node to estimate the sum $Z = \sum_{v\in\mathcal{V}} W_v$. Then

$$T(\varepsilon,\delta) \geq \max_{\mathcal{S}\subset\mathcal{V}} \frac{1}{C_{\mathcal{S}}} \left[(1-\delta) \left(\log \frac{1}{2e\varepsilon} + h(Z|W_{\mathcal{S}}) \right) - h_2(\delta) \right].$$

To prove this, fix a set $\mathcal{S} \subset \mathcal{V}$, and let $p_{\mathcal{S}}(z)$ denote the probability density of $\sum_{v\in\mathcal{S}^c} W_v$. Then

$$\mathbb{E}[L(W_{\mathcal{S}},\varepsilon)] = \sup_{z\in\mathbb{R}} \int_{z-\varepsilon}^{z+\varepsilon} p_{\mathcal{S}}(z)\mathrm{d}z \leq 2\varepsilon \|p_{\mathcal{S}}\|_\infty, \qquad (15)$$

where $\|p_{\mathcal{S}}\|_\infty$ is the sup norm of $p_{\mathcal{S}}$. Now, by a result of Bobkov and Madiman [13, Proposition I.2], if U is a real-valued r.v. with a log-concave density p, then the differential entropy $h(U)$ is upper-bounded by $\log e + \log \|p\|_\infty^{-1}$. Using this fact together with (15), the log-concavity of $p_{\mathcal{S}}$, and the fact that the W_v's are mutually independent, we can write

$$\begin{aligned} \log \frac{1}{\mathbb{E}[L(W_{\mathcal{S}},\varepsilon)]} &\geq \log \frac{1}{2\varepsilon} + \log \frac{1}{\|p_{\mathcal{S}}\|_\infty} \\ &\geq \log \frac{1}{2e\varepsilon} + h\left(\sum_{v\in\mathcal{S}^c} W_v \right) \\ &= \log \frac{1}{2e\varepsilon} + h(Z|W_{\mathcal{S}}), \end{aligned}$$

Using this estimate in (2), we get the desired bound.

Appendix A
Two lemmas for the proof of Theorem 1

The proof of Theorem 1 rests on two lemmas, which we give here. The first lemma is a standard cutset bound (see, e.g., [1], [2]), so we omit the proof. The second lemma is new.

Lemma A.1. *For any set $\mathcal{S} \subset \mathcal{V}$ and any $\mathcal{A} \in \mathfrak{A}(T)$, $I(Z;\widehat{Z}_{\mathcal{S}}|W_{\mathcal{S}}) \leq TC_{\mathcal{S}}$.*

Lemma A.2. *Fix a set $\mathcal{S} \subset \mathcal{V}$, and suppose that the parameters ε, δ satisfy*

$$\mathbb{E}[L(W_{\mathcal{S}},\varepsilon)] \leq 1-\delta. \qquad (A.1)$$

Consider an algorithm $\mathcal{A}$, such that

$$\max_{v\in\mathcal{V}} \mathbb{P}\left(d(Z,\widehat{Z}_v) > \varepsilon \right) < \delta. \qquad (A.2)$$

Then $I(Z;\widehat{Z}_{\mathcal{S}}|W_{\mathcal{S}}) \geq (1-\delta)\log \dfrac{1}{\mathbb{E}[L(W_{\mathcal{S}},\varepsilon)]} - h_2(\delta)$.

Proof: Fix an arbitrary $v \in \mathcal{S}$, and consider the probability distributions $\mathbb{P} = \mathbb{P}_{W_{\mathcal{S}},Z,\widehat{Z}_v}$ and $\mathbb{Q} = \mathbb{P}_{W_{\mathcal{S}}}\mathbb{P}_{Z|W_{\mathcal{S}}}\mathbb{P}_{\widehat{Z}_v|W_{\mathcal{S}}}$. Define the indicator random variable $\Upsilon \triangleq \mathbf{1}_{\{d(Z,\widehat{Z}_v)\leq\varepsilon\}}$. Then from (A.2) it follows that $\mathbb{P}[\Upsilon = 1] \geq 1-\delta$. On the other hand, since $Z \to W_{\mathcal{S}} \to \widehat{Z}_v$ is a Markov chain under $\mathbb{Q}$,

$$\begin{aligned} &\mathbb{Q}[\Upsilon = 1] \\ &= \int_{\mathsf{W}_{\mathcal{S}}} \int_{\mathsf{Z}} \mathbb{P}\left(d(Z,\widehat{z}_v) \leq \varepsilon | W_{\mathcal{S}} = w_{\mathcal{S}}\right) \mathbb{P}(\mathrm{d}\widehat{z}_v|w_{\mathcal{S}}) \mathbb{P}(\mathrm{d}w_{\mathcal{S}}) \\ &\leq \int_{\mathsf{W}_{\mathcal{S}}} \sup_{\widehat{z}_v\in\mathsf{Z}} \mathbb{P}\left(d(Z,\widehat{z}_v) \leq \varepsilon | W_{\mathcal{S}} = w_{\mathcal{S}}\right) \mathbb{P}(\mathrm{d}w_{\mathcal{S}}) \\ &= \mathbb{E}[L(W_{\mathcal{S}},\varepsilon)] \end{aligned} \qquad (A.3)$$

by Fubini's theorem. Consequently,

$$\begin{aligned} I(Z;\widehat{Z}_{\mathcal{S}}|W_{\mathcal{S}}) \geq I(Z;\widehat{Z}_v|W_{\mathcal{S}}) &= D(\mathbb{P}\|\mathbb{Q}) \\ &\overset{(a)}{\geq} d(\mathbb{P}[\Upsilon=1]\|\mathbb{Q}[\Upsilon=1]) \\ &\overset{(b)}{\geq} d(1-\delta \,||\, \mathbb{E}[L(W_{\mathcal{S}},\varepsilon)]) \\ &\overset{(c)}{\geq} (1-\delta)\log \frac{1}{\mathbb{E}[L(W_{\mathcal{S}},\varepsilon)]} - h_2(\delta), \end{aligned}$$

where (a) follows from data processing inequality for divergence, where $d(p\|q) = p\log(p/q)+(1-p)\log((1-p)/(1-q))$ is the binary divergence function; (b) follows from the fact that $p \mapsto d(p||q)$ is increasing on the interval $[q,1]$ and $q \mapsto d(p||q)$ is decreasing on the interval $(0,p]$, and the fact that $\mathbb{P}[\Upsilon = 1] \geq 1-\delta \geq \mathbb{E}[L(W_{\mathcal{S}},\varepsilon)] \geq \mathbb{Q}[\Upsilon = 1]$, by (A.1)–(A.3); and (c) follows from the fact that $d(p\|q) \geq p\log(1/q) - h_2(p)$. This completes the proof. ∎

References

[1] O. Ayaso, D. Shah, and M. Dahleh, "Information-theoretic bounds for distributed computation over networks of point-to-point channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6020–6039, December 2010.

[2] G. Como and M. Dahleh, "Lower bounds on the estimation error in problems of distributed computation," in *Proc. Inform. Theory and Applications Workshop*, 2009, pp. 70–76.

[3] R. Carli, G. Como, P. Frasca, and F. Garin, "Distributed averaging on digital erasure networks," *Automatica*, vol. 47, no. 115-121, 2011.

[4] S. Kar and J. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. on Signal Process.*, vol. 57, no. 1, pp. 355–369, 2009.

[5] N. Noorshams and M. Wainwright, "Non-asymptotic analysis of an optimal algorithm for network-constrained averaging with noisy links," *IEEE J. Sel. Top. Sign. Proces.*, vol. 5, no. 4, pp. 833–844, 2011.

[6] V. V. Petrov, *Sums of Independent Random Variables*. Berlin: Springer-Verlag, 1975.

[7] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[8] A. Kolmogorov, "Sur les propriétés des fonctions de concentrations de M. P. Lévy," *Ann. Inst. H. Poincaré*, vol. 16, pp. 27–34, 1958.

[9] H. H. Nguyen and V. H. Vu, "Small ball probability, inverse theorems, and applications," in *Erdős Centennial*, ser. Bolyai Society Mathematical Studies. Springer, 2013, vol. 25. [Online]. Available: http://arxiv.org/abs/1301.0019

[10] S. G. Bobkov and G. P. Chistyakov, "On concentration functions of random variables," *J. Theor. Probab.*, July 2013, published online.

[11] M. Rudelson and R. Vershynin, "The Littlewood–Offord problem and invertibility of random matrices," *Adv. Math.*, vol. 218, pp. 600–633, 2008.

[12] P. Erdős, "On a lemma of Littlewood and Offord," *Bull. Amer. Math. Soc.*, vol. 51, pp. 898–902, 1945.

[13] S. Bobkov and M. Madiman, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 4940–4954, August 2011.