Classification in a Large Network

Vinay A. Vaishampayan

City University of New York-College of Staten Island

Staten Island, NY 10314, USA

Abstract—We construct and analyze the communication cost of protocols (interactive and one-way) for classifying $\mathbf{X} = (X_1, X_2, \ldots, X_n) \in [0, 1)^n \subset \mathbb{R}^n$, in a network with $n \geq 2$ nodes, with X_i known only at node i. The classifier takes the form $\sum_{i=1}^n h_i X_i \gtrless a$, with weights $h_i \in \{-1, +1\}$. The interactive protocol (a zero-error protocol) exchanges a variable number of messages depending on the input X and its sum rate is directly proportional to its mean stopping time. An exact analysis, as well as an approximation of the mean stopping time is presented and shows that it depends on $\gamma = \alpha + (1/2 - \beta)$, where $\alpha = a/n$ and $\beta = m/n$, with m being the number of positive weights. In particular, the mean stopping time grows logarithmically in n when $\gamma = 0$, and is bounded in n otherwise. Comparisons show that the sum rate of the interactive protocol is smaller than that of the one-way protocol when the error probability for the one-way protocol is small, with the reverse being true when the error probability is large. Comparisons of the interactive protocol are also made with lower bounds on the sum rate.

Index terms—Interactive communication, distributed function computation, hypothesis testing, sequential hypothesis testing, classification.

I. INTRODUCTION

We study the communication cost of implementing a 2-classifier $f : \mathbb{R}^n \to \{0, 1\}$ which maps input vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ to a class label in the set $\{0, 1\}$. It is assumed that the classifier is to be implemented in a network with n sensor nodes, where X_i is a random variable known at the ith node alone. We propose a distributed algorithm for solving the classification problem and provide an analysis of the communication rate required by this algorithm. We also investigate lower bounds on the communication rate.

Previous related work is summarized in Sec. II, problem setup is in Sec. III, communication protocols and their analysis are in Secs. IV–V, lower bounds on the communication rate are in Sec. VI, numerical performance results are in Sec. VII and Sec. VIII contains a summary of the paper.

II. BACKGROUND

Given a function $f : \mathcal{X}_1 \times \mathcal{X}_2 \ldots \times \mathcal{X}_n \to \mathcal{Z}$, where $\mathcal{X}_i$ and $\mathcal{Z}$ are arbitrary sets, its *communication complexity* [10], is $\mathcal{C}(f) := \inf_{\mathcal{P}} R(\mathcal{P})$, where $R(\mathcal{P})$ is the communication cost of protocol $\mathcal{P}$ for computing f in a distributed environment. Here, the *average-case* rather than worst-case $R(\mathcal{P})$ is considered, the average being over all inputs with respect to a known probability distribution [6].

Communication complexity for interactive function computation of Boolean functions is explored in great detail in the seminal work [6], which has also influenced the development here. Our paper addresses a problem of current interest since classification is an important part of machine learning. Interactive and one-way hypothesis testing were studied information theoretically (i.e. as the block length grows) in [9] for $n = 2$ nodes and $r \leq 2$ rounds. Here, we assume a block length of unity, the methods used are from sequential analysis, and the results derived are for infinite round protocols for $n \geq 2$ nodes. We note that classification is also a central operation in nearest lattice point search [1], [8]. The bit exchange protocol described here is in a class of protocols described in [5].

III. PROBLEM SETUP

Since X_i is known at node i and h_i is known globally, the quantity $|h_i| X_i$ can be treated as a new X_i, with a modified probability distribution, known to all nodes. If $h_i = 0$ for some i, it can be dropped and treated as a problem with a smaller n. Thus it suffices to consider classifiers with filter coefficients $h_i \in \{-1, +1\}$, with magnitudes of the h_i's absorbed into the probability distribution $p_{\mathbf{X}}$. Let $\mathcal{H}^+ := \{i : h_i = +1\}$ and $\mathcal{H}^- := \{i : h_i = -1\}$. Let $m := |\mathcal{H}^+|$, the number of positive filter weights. Then $|\mathcal{H}^-| = n - m$. Let

$$X_s := \sum_{i \in \mathcal{H}^+} X_i - \sum_{i \in \mathcal{H}^-} X_i. \tag{1}$$

Our classification function

$$f(\mathbf{X}) := \begin{cases} 1, & X_s \geq a, \\ 0, & X_s < a, \end{cases} \tag{2}$$

partitions $\mathbb{R}^n$ into two classes $f^{-1}(0) = \{\mathbf{X} : X_s < a\}$ and its complement $f^{-1}(1)$.

We assume here that X_i are independent, identically distributed (iid) random variables, uniformly distributed on the unit interval $[0, 1)$. Note that under our assumptions for the source distribution, the classification problem is non-trivial only if a lies in the two-side open interval $(-(n - m), m)$.

IV. THE INFINITE ROUND BIT-EXCHANGE PROTOCOL

Nodes exchange information in a pre-arranged manner and any information transmitted by node-i at time t can depend only on X_i and information that it has received from other nodes at times $t' < t$. We do not assume a broadcast model for counting bits, so every transmitted bit is indexed by i, j and t, the source, destination and transmission time, respectively. The cost is the sum of communication over i, j, t.

The Interactive Protocol computes $f(\mathbf{X})$ as follows. A node is selected as a leader node. Here, we assume that the leader

node is distinct from the n sensor nodes [1] and at the outset of a session has no access to X_i, $i = 1, 2, \ldots, n$. The order of communication is known to all nodes. The time axis is broken into sessions, a session starts with a fresh observation $\mathbf{X}$ and concludes when $f(\mathbf{X})$ is computed. Here we are concerned with a single session. A session is divided up into rounds.

Node-i produces a bit stream $b(i, j)$, $j = 1, 2, \ldots$ according to the following standard binary expansion rule: set $X(i, 1) = X_i$ and for $j = 1, 2, \ldots\ldots$ compute

$$
\begin{aligned}
b(i, j) &= \lfloor 2X(i, j) \rfloor, \\
X(i, j+1) &= 2X(i, j) - b(i, j). \qquad (3)
\end{aligned}
$$

Let

$$
B(l) := \sum_{i \in \mathcal{H}^+} b(i, l) - \sum_{i \in \mathcal{H}^-} b(i, l). \qquad (4)
$$

and

$$
Z(j) := \sum_{l=1}^{j} B(l) 2^{-l},
$$

with $Z(0) = 0$. In round $j \geq 1$, node-i sends $b(i, j)$ to the leader node and after the leader node receives all the bits $b(i, j)$, $i = 1, 2, \ldots, n$, it computes $Z(j)$ and sends $s(j)$ back to each sensor node, where

$$
s(j) = \begin{cases} 0, & Z(j) \leq L(j) \\ 1, & Z(j) \geq U(j) \\ continue, & \text{otherwise,} \end{cases} \qquad (5)
$$

with $L(j) := a - m2^{-j}$ and $U(j) := a + (n-m)2^{-j}$. Observe that $U(j+1) = U(j) - (n-m)2^{-(j+1)}$ and $L(j+1) = L(j) + m2^{-(j+1)}$, $j \geq 1$.

Let T be the stopping time, i.e. T is the smallest k for which $s(k) \neq continue$. Since the leader node sends back $n \log_2(3)$ bits at the conclusion of every round, the total number of bits communicated in a session is $R_{sum} = n(1 + log_2(3))\bar{T}$ where $\bar{T} := E[T]$ is the mean stopping time, over all inputs $\mathbf{X}$.

Lemma 1. *If at the end of round j, $s(j) \neq continue$, then either $f(\mathbf{X}) = 0$ or $f(\mathbf{X}) = 1$.*

Proof. This follows from the fact that $Z(j) - (n-m)2^{-j} < X_s < Z(j) + m2^{-j}$ and the definition of f in (2). $\qquad \square$

A. A Recursive Description of the Protocol

For the purpose of analysis, it is convenient to have an alternate description, referred to as the *recursive* description, of the protocol. Towards this end consider the following alternate description of the feedback signal from leader to sensor. Let

$$
\tilde{s}(j) = \begin{cases} 0, & B(j) \leq \tilde{L}(j), \\ 1, & B(j) \geq \tilde{U}(j), \\ continue, & \text{otherwise.} \end{cases} \qquad (6)
$$

where $\tilde{L}(j) = 2a(j) - m$, $\tilde{U}(j) = 2a(j) + (n-m)$,

$$
a(j+1) = 2a(j) - B(j), \quad j \geq 1 \text{ and } a(1) = a. \qquad (7)
$$

Observe that $\tilde{L}(j+1) = 2\tilde{L}(j) + m - 2B(j)$ and $\tilde{U}(j+1) = 2\tilde{U}(j) - 2B(j) - (n-m)$, $j \geq 1$.

Remark 1. *From the form of $\tilde{L}(j)$ and $\tilde{U}(j)$, it is clear that at iteration j, the effective threshold is $a(j)$. Thus for the recursive protocol, the threshold shifts with each iteration, but the width of the interval $\tilde{U}(j) - \tilde{L}(j)$ remains fixed, as opposed to the non-recursive description, where $U(j) - L(j)$ decreases with j.*

Lemma 2. $\tilde{s}(j) = s(j)$, $j \geq 1$

Proof. Observe that $s(j)$ is determined by testing $Z(j)$ against thresholds $U(j), L(j)$. Similarly $\tilde{s}(j)$ is obtained by testing $B(j)$ against $\tilde{U}(j), \tilde{L}(j)$. Since $Z(j) = Z(j-1) + 2^{-j}B(j)$, it suffices to show that

$$
U(j) = 2^{-j}\tilde{U}(j) + Z(j-1), \quad j \geq 1 \qquad (8)
$$

with a similar statement for $L(j)$ and $\tilde{L}(j)$. Eq. (8) is clearly true for $j = 1$ since $U(1) = a + (n-m)/2$ and $\tilde{U}(1) = 2a + (n-m)$ and $Z(0) = 0$. Assume it holds for j. Then

$$
\begin{aligned}
U(j+1) &= U(j) - (n-m)2^{-(j+1)} \\
&= 2^{-j}\tilde{U}(j) + Z(j-1) - (n-m)2^{-(j+1)} \\
&= 2^{-(j+1)}\tilde{U}(j+1) + Z(j-1) + B(j)2^{-j} \\
&= 2^{-(j+1)}\tilde{U}(j+1) + Z(j). \qquad (9)
\end{aligned}
$$
$\qquad \square$

For the analysis presented later we observe that

Lemma 3.

$$
Pr[B(l) = k] = \begin{cases} \binom{n}{k+n-m}, & k = -(n-m), \ldots, m, \\ 0, & \text{otherwise.} \end{cases} \qquad (10)
$$

Proof. To see that (10) is true, let $B(l) = w^+ - w^-$ where w^+ is the number of 1's in positions $\mathcal{H}^+$ and w^- the number of 1's in the positions $\mathcal{H}^-$. Then the number of 0's in the positions $\mathcal{H}^-$ is $(n-m) - w^-$ and $B(l) = k$ if and only if $w^+ + (n-m) - w^- = k + (n-m)$. $\qquad \square$

In order to understand how R_{sum} scales with n we now proceed to analyze $\bar{T}$. We first present a computational approach when a is an integer, followed by an asymptotic analysis for general a.

B. Exact Analysis

We first present the analysis for $n = 2$, $m = 1$ where the only non-trivial case is $a = 0$. Node 1 sends a bit to node 2, which responds with a signal to continue or stop (here and for this case only, the leader node is a sensor node; see footnote on previous page). When the signal is to stop, both nodes know the class. Thus the average rate is $2\bar{T} = 4$ bits, which follows from the fact that $Pr[T > k] = 2^{-k}$ and $\bar{T} = 1 + \sum_{k>0} Pr[T > k]$.

A general analysis for $n \geq 2$ is presented next. We work with the recursive description of the protocol. The basic idea is that if the initial threshold a is an integer then so is $a(j)$, $j \geq$

1 as can be seen from (7). Together with the fact that the distribution of $\{X(i,j),\ i = 1, 2, \ldots, n\}$ does not depend on iteration index j means that $a(j)$ can be regarded as a state variable and if it shown to be restricted to a finite set, then $\bar{T}(a)$, the mean stopping time associated with a threshold a, can be obtained through the state transition probability matrix of a finite state transition graph. Let

$$\bar{\mathbf{T}} = (\bar{T}(-(n-m)+1), \ldots, \bar{T}(m-1))^t$$

denote the vector of mean stopping times.

Theorem 1. *The vector of mean stopping times is a solution of*

$$(\mathbf{I} - \mathbf{Q})\bar{\mathbf{T}} = (1, 1, \ldots, 1)^t. \qquad (11)$$

where $\mathbf{Q}$ is a $(n-1) \times (n-1)$ matrix, whose value in the ath row and a'th column is $Pr[B(l) = 2a - a']$, $a, a' \in \{-(n-m)+1, \ldots, m-1\}$ ($\mathbf{I}$ is a commensurate identity matrix).

Proof. We show that at each iteration in the recursive description of the protocol, the problem is to compute $f(\tilde{\mathbf{X}})$, where $\tilde{\mathbf{X}}$ has the same joint probability distribution as $\mathbf{X}$, perhaps with a different threshold. Towards this end it suffices to show (i) that the probability distribution P_j of $X(i,j)$, $i = 1, 2, \ldots, n$ does not depend on the iteration index j, and that (ii) the possible value that a threshold can have at any iteration in the algorithm lies in a finite set. Eq. (11) then follows by a probability transition calculation. To see that (i) holds, note that P_j is iid with marginal distribution that is uniform on $[0, 1)$. For the proof of (ii), note that by (7), if a is an integer, then so is $a(j), j \geq 1$. The remaining steps to show (ii) are by induction. For $j = 1$, $a(j)$ lies in the set $\{-(n-m)+1, -(n-m)+2, \ldots, m-1\}$. Assume it is true for j. From the recursion (7), and the fact that the protocol did not stop at iteration j, $B(j)$ must satisfy $2a_j - m < B(j) < 2a_j + (n-m)$, from which the assertion follows directly. Thus if the protocol continues at iteration j, the probability that the threshold $a(j+1) = a'$ given that $a(j) = a$ is $Q(a, a') = Pr[B(j) = 2a - a']$, which does not depend on j. Thus

$$\bar{T}(a) = 1 + \sum_{a'=-(n-m)+1}^{m-1} \bar{T}(a')Q(a, a'),$$
$$a \in \{-(n-m), \ldots, m\}, \quad (12)$$

and (11) follows. □

Example 1. *For $n = 4$, $m = 2$*

$$\mathbf{Q} = \frac{1}{16} \begin{pmatrix} 4 & 1 & 0 \\ 4 & 6 & 4 \\ 0 & 1 & 4 \end{pmatrix}. \qquad (13)$$

C. An approximation for $\bar{T}$

The event $\{T > k\}$ is equivalent to the event $\{L(l) < Z(l) < U(l),\ l = 1, 2, \ldots, k\}$, where $L(l), U(l)$ are defined immediately after (5). Thus $Pr[T > k] \leq Pr[L(k) < Z(k) < U(k)]$. We will assume that $m/n = \beta$, a constant and that

$a/n = \alpha$. By the central limit theorem, the cumulative distribution function of $Z(j)/\sqrt{n}$ converges to that of a Gaussian with mean $\sqrt{n}(\beta - 1/2)(1 - 2^{-j})$ and variance $(1 - 4^{-j})/12$. With $\sigma = \sqrt{n(1 - 4^{-k})/12}$, $\gamma = (\alpha + (1/2 - \beta))$ let $A(n, k, \gamma) = \frac{1}{2}\left[\text{erf}\left(\frac{(\gamma + 2^{-(k+1)})n}{\sqrt{2}\sigma}\right) - \text{erf}\left(\frac{(\gamma - 2^{-(k+1)})n}{\sqrt{2}\sigma}\right)\right]$. Then by the central limit theorem $P[L(k) < Z(k) < U(k)] = A(n, k, \gamma) + \epsilon(n, k, \gamma)$, where $\epsilon(n, k, \gamma)$ is an error term in the central limit approximation. Thus

$$\lim_{n\to\infty} \bar{T} \leq 1 + \lim_{n\to\infty} \sum_{k=1}^{\infty} P[L(k) < Z(k) < U(k)]$$

$$= 1 + \lim_{n\to\infty} \sum_{k=1}^{\infty} (A(n, k, \gamma) + \epsilon(n, k, \gamma)).$$
$$(14)$$

The first term $\sum_{k=1}^{\infty} A(n, k, \gamma) \leq A(n, \gamma)$, where

$$A(n, \gamma) =$$
$$\begin{cases} \frac{1}{2}\log_2\left(\frac{6n}{\pi} + 1\right) + \sqrt{\frac{3}{1 + \pi/(6n)}}, & \gamma = 0, \\ 1 + \kappa + o(1), & 0 < \gamma \leq 1/4, \\ 1 + o(1), & 1/4 < \gamma \leq 1/2, \end{cases} \quad (15)$$

where $o(1) \to 0$ as $n \to \infty$, and κ is the largest integer for which $\frac{(\gamma - 2^{-(\kappa+1)})n}{\sqrt{2}\sigma} \leq -\frac{\sqrt{\pi}}{2}$. Note that κ is bounded in n. Eq. (15) follows by applying the bound $\text{erf}(x) \leq \min(2x/\sqrt{\pi}, 1)$, $x \geq 0$ for the upper branch, $\text{erf}(x) - \text{erf}(y) \leq (2(x-y)/\sqrt{\pi})e^{-y^2}$, $x > y > 0$, for the lower two branches. For the middle branch, we additionally use the bound $\text{erf}(x) - \text{erf}(y) \leq 2$, when $x > \sqrt{\pi}/2$ and $y < -\sqrt{\pi}/2$.

Theorem 2. $\lim_{n\to\infty} \sum_{k=1}^{\infty} \epsilon(n, k, \gamma) = 0$.

Proof. Since the random variables $b(i, l)$ have finite absolute third moment, it follows from a result due to Berry and Esseen [2], that for $x \in \mathbb{R}$, $|F_{Z(k)/\sqrt{n}}(x) - F_G(x)| \leq C_1/\sqrt{n}$, where F_G is the CDF of a Gaussian random variable with with mean $\sqrt{n}(\beta - 1/2)(1 - 2^{-k})$ and variance $(1 - 4^{-k})/12$, $F_{Z(k)/\sqrt{n}}$ is the CDF of $Z(k)/\sqrt{n}$ and C_1 is a constant independent of x. Further, $|F_G(U(k)/\sqrt{n}) - F_G(L(k)/\sqrt{n})| \leq C_2\sqrt{n}2^{-k}$ and $|F_{Z(k)/\sqrt{n}}(U(k)/\sqrt{n}) - F_{Z(k)/\sqrt{n}}(L(k)/\sqrt{n})| \leq C_3\sqrt{n}2^{-k}$ for constants C_2 and C_3. Thus $\epsilon(n, k, \gamma) \leq \min(C\sqrt{n}2^{-k}, D/\sqrt{n})$ for constants C and D. We now proceed to bound $\sum_{k=1}^{\infty} \epsilon(n, k, \gamma)$. Let k^* be the largest such that $c\sqrt{N}2^{-k} > D/\sqrt{n}$ for $k < k^*$. Then $\sum_{k=1}^{\infty} \epsilon(n, k, \gamma) \leq k^*D/\sqrt{n} + 2D/\sqrt{n} \leq (1 + o(1))\log n/\sqrt{n}$. □

Remark 2. *It is interesting that when $\gamma \neq 0$, $\bar{T}$ is integer valued.*

V. One-Way Protocols

In the One-Way protocol, the class label is to be known at the leader node alone. In the One-Way+ protocol, the leader informs each sensor node of the class label at an additional cost of n bits. The encoder for the ith node is a mapping $\phi : \mathbb{R} \to \mathcal{C}_i$, where $\mathcal{C}_i$ is the codebook for node i. Encoded value $\phi(X_i)$ is sent to a leader node using $R_i = \log_2(|\mathcal{C}_i|)$ bits.

The leader node estimates the class label, and informs each node about this. The total rate is given by $R_{sum} = \sum_{i=1}^{n} R_i + n$ bits. Suppose the estimated class label is $g(\mathbf{X})$. The error $P_e = Pr[f(\mathbf{X}) \neq g(\mathbf{X})]$ is a function of the rate R_{sum}. We analyze and implement this protocol when $\phi(X_i) = \lfloor X_i/\delta \rfloor \delta$ for each i ($\lfloor x \rfloor$ is the largest integer $\leq x$). In terms of the step size δ, $R_{sum} = n(\log_2(1/\delta) + 1)$ bits. We now determine δ, and thus R_{sum} in terms of P_e. Our main finding is that for $m = n - m$ and $a = 0$, the sum rate for One-Way+ is

$$R_{sum} \geq n \left(\log_2 \left(\frac{1}{P_e^*} \right) + \frac{1}{2} \log_2 \left(\frac{12n}{\pi} \right) + 1 \right), \quad (16)$$

where P_e^* is a target upper bound on the error probability. Thus for this case, for fixed P_e, R_{sum}/n grows logarithmically in n.

The proof of (16) is now presented. Observe that the protocol partitions the unit cube $[0,1)^n$ into smaller cubes, called cells, of side δ. An error occurs only when the source vector lies in a cell whose interior does not intersect the separating hyperplane $\partial f := \{\sum_{i \in \mathcal{H}^+} X_i - \sum_{i \in \mathcal{H}^-} X_i = a\}$. We now calculate the number of cells that intersect ∂f. In order to simplify the analysis a bit, let $Y_i = -X_i + 1$ for $i \in \mathcal{H}^-$. Note that, like X_i, Y_i is uniformly distributed on $[0,1)$. Now rename the Y_i to X_i. The boundary $\partial f := \{\sum_i X_i = a + (n - m)\}$. Thus under the transformation, the threshold is now $a + (n - m)$, which we rename to a, with the constraint that the new a satisfies $0 < a < n$. Under our assumption of uniform quantization, the lower endpoints of the bins are in the set $\mathcal{L} := \{m\delta, \ m = 0, 1, \ldots, M - 1\}$, where $M = 1/\delta$ is the number of bins. Let $L_i = l_i \delta \in \mathcal{L}$ be the lower endpoint of the bin for X_i, i.e $l_i \leq X_i < l_i + \delta$. Then a cell intersects ∂f if and only if $a - n\delta \leq \sum_i L_i < a$, or equivalently

$$-\frac{1}{2} < \frac{1}{n} \sum_i l_i - \left(\frac{M-1}{2} \right) < \frac{1}{2}, \quad (17)$$

as can be seen after some algebraic manipulation. The random variable $Z := \frac{1}{n} \sum_i l_i - \left(\frac{M-1}{2} \right)$ has mean zero and variance $(M^2 - 1)/(12n)$ and through an application of the central limit theorem it follows that $\lim n \to \infty P_e = \lim_{n \to \infty} (1/2) \operatorname{erf}(\sqrt{12n/(M^2-1)})$. The bound $\operatorname{erf}(x) \leq \min(2x/\sqrt{\pi}, 1)$, $x \geq 0$, and some further algebra leads to (16).

We conjecture that the performance of a uniform quantizer cannot be improved on, under the assumptions made here. The proof of this appears to be significantly complicated.

VI. LOWER BOUNDS FOR THE INTERACTIVE PROTOCOL

We first show, through a lower bounding argument, that for $n = 2, m = 1$ and $a = 0$, the bit exchange protocol achieves the optimal sum rate. For $n > 2$, the protocol achieves optimal scaling with respect to n, when $\gamma \neq 0$. However, when $\gamma \neq 0$, there is a $\log n$ gap in the sum rate. We show this through a lower bound for $n > 2$.

A. A Tight Lower Bound for $n = 2, m = 1, a = 0$

The Interactive Protocol results in a rectangular partition $(\mathcal{P}, \mathcal{Q})$ where $\mathcal{P}$ is a subpartition of the region $\mathcal{R}_p = \{\mathbf{x} : f(\mathbf{x}) = 1\}$ and $\mathcal{Q}$ of $\mathcal{R}_p = \{\mathbf{x} : f(\mathbf{x}) = 0\}$. Since the error probability is zero, we refer to $(\mathcal{P}, \mathcal{Q})$ as a *zero-error* partition. The boundary is represented by the set $\mathcal{B} := \{x_1 = x_2\} \bigcap (0,1)^2$. For a zero-error partition $(\mathcal{P}, \mathcal{Q})$ it is true that each point of $\mathcal{B}$ must be the upper left corner of some rectangle that lies entirely in $\mathcal{R}_p$ or the lower right corner of some rectangle that lies entirely in $\mathcal{R}_q$, except possibly for a set of one-dimensional measure zero. Let $\{p_i, \ i = 1, 2, \ldots\}$ be the probabilities of the cells of $\mathcal{P}$ and let $\{q_i, \ i = 1, 2, \ldots\}$ be the probabilities of the cells of $\mathcal{Q}$. We note here that we use the word rectangle to include what is referred to as a *combinatorial rectangle*, which is the Cartesian product of unions of intervals [4].
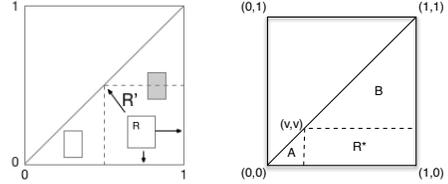

Fig. 1. (left) Illustration of the readjustment of the rectangle R with the highest probability, (right) Regions used in Thm. 4. $R_p = A \bigcup R^* \bigcup B$.

Theorem 3. *If a partition minimizes the entropy it contains a rectangle with vertices $(1,0)$ and (v,v) and another rectangle with vertices $(0,1)$ and (u,u), for some $0 < u, v < 1$.*

Proof. (Sketch) The idea is to grow the largest rectangle in a partition cell until a vertex touches the boundary while reducing the probabilities of the other rectangles in the partition cell. The probability distribution of the readjusted rectangles majorizes all other rectangular partitions. The details of the proof are omitted due to space constraints and can be found in [7]. □

Theorem 4. *For $n = 2, m = 1, a = 0$, the sum rate is no smaller than 4 bits.*

Proof. Consider an extreme partition $(\mathcal{P}, \mathcal{Q})$ which contains a rectangle R^* which has a points (v, v) and $(1, 0)$ as its upper left and lower right vertices. This is always true by Thm. 3. Let random variable C indicate whether (x_1, x_2) lies in $\mathcal{R}_p$ or not, and let random variable S indicate whether (x_1, x_2) lies in one of the three regions, R^*, A or B as shown in Fig. 1. Let $H(\mathcal{P}, \mathcal{Q})$ denote the entropy of the partition $(\mathcal{P}, \mathcal{Q})$. Then

$$\begin{aligned} H(\mathcal{P}, \mathcal{Q}) = \ & H(C) + H(\mathcal{P}, \mathcal{Q}|C = 0)P(C = 0) + \\ & H(\mathcal{P}, \mathcal{Q}|C = 1)P(C = 1) \end{aligned} \quad (18)$$

and

$$\begin{aligned} H(\mathcal{P}, \mathcal{Q}|C = 1) = \ & H(S|C = 1) + \\ & + H(\mathcal{P}, \mathcal{Q}|C = 1, S = A)P(S = A|C = 1) + \\ & + H(\mathcal{P}, \mathcal{Q}|C = 1, S = B)P(S = B|C = 1). \quad (19) \end{aligned}$$

Since the regions A and B are similar to $\mathcal{R}_p$ it follows that if this partition minimizes the entropy it must satisfy the recursion

$$H(\mathcal{P},\mathcal{Q}|C=1) = H([v^2, 2v(1-v), (1-v)^2])+$$
$$+ (v^2 + (1-v)^2)H(\mathcal{P},\mathcal{Q}|C=1). \quad (20)$$

Solving for $H(\mathcal{P},\mathcal{Q}|C=1)$ we obtain

$$H(\mathcal{P},\mathcal{Q}|C=1) = \frac{H([v^2, 2v(1-v), (1-v)^2])}{2v(1-v)} \quad (21)$$

whose unique minimum value of 3 bits occurs when $u = v = 1/2$. Plugging back in (18) leads to the desired result. $\square$

B. $n > 2$

Lemma 4. *(Lemma 9 in [6]) Assume that probabilities p_i, $i = 1, 2, \ldots$ are in decreasing order, $\sum_{i=1}^{\infty} p_i = 1$ and that $\sum_{i=m}^{\infty} p_i \geq g(m)$, $m = 1, 2, \ldots$. Then the entropy $H(p) \geq \sum_{i=1}^{m-1} p_i \log(1/p_i) + g(m)\log(m/(1-g(m+1)))$.*

It turns out to be easy to determine the largest rectangle in $f^{-1}(0)$ for $n > 2$. We do so here for the case where $a = 0$ and $m = n/2$. In this case, with p_1 being the probability of the largest rectangle in $f^{-1}(0)$ it follows that $p_1 \leq 2^{-n}$ as can be checked by maximizing $\prod_{i\in\mathcal{H}^+}(1-x_i)\prod_{i\in\mathcal{H}^-} x_i$ subject to $\sum_{i\in\mathcal{H}^+} x_i - \sum_{i\in\mathcal{H}^-} x_i = 0$. Thus it follows that $H(p) \geq n$. In this specific case, there is a $\log n$ gap between the upper bound of our protocol and the lower bound. For the cases where $\gamma > 0$, the rate grows linearly in n. Since we cannot expect a slower growth with n (the result must be distributed to n nodes), the protocol exhibits the correct scaling behavior with n.
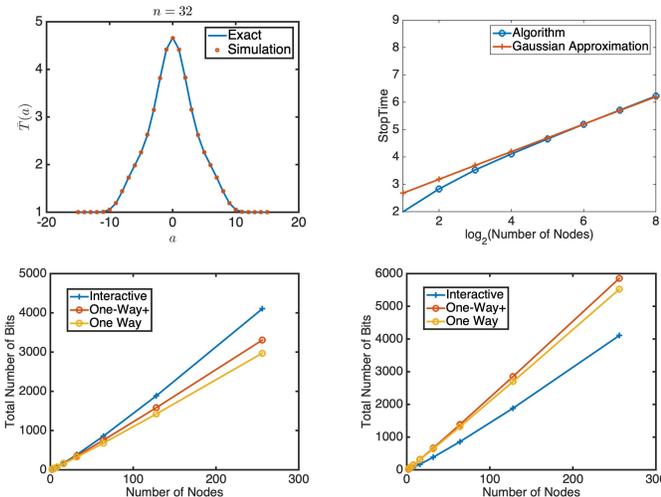
VII. NUMERICAL EXPERIMENTS

Fig. 2. From top left to bottom right: (a) Comparison of exact (11) and simulation results for $n = 32$ (b) stopping time of the Interactive Protocol as a function of n: simulation of actual protocol and Gaussian approximation (15), $a = 0$ and $m = n/2$,(c)-(d) sum rate of the One-Way, One-Way+ and Interactive protocol for $P_e = 10^{-2}$ and $P_e = 10^{-5}$, resp.

Computer simulations were carried out for the Interactive Protocol for a $n = 32$ node network and compared to

(11). Simulations were based on $20,000$ repetitions per data point in the graph shown in Fig. 2(a). Fig. 2(b) shows the agreement between the growth of the stopping time obtained via simulation and with the approximation (15) for $m = n/2$, $a = 0$. The agreement is good in all cases.

Fig. 2(d) compares the sum rate of the Interactive Protocol with the One-Way protocols for $m/n = 1/2$ and $a = 0$. In our experiments, we used a uniform quantizer for quantizing X_i for each i. Note that in the one-way case, it is not possible to obtain zero error with finite rate. In our experiments we set the error probability to 10^{-5}. It is seen that for a $n = 256$ node network, the Interactive Protocol used 4106 bits whereas the One-Way protocol uses 6374 bits; a non-trivial saving of almost 35% over the One-Way protocol for $P_e = 10^{-5}$. On the other hand, as seen in Fig. 2(c), when $P_e = 10^{-2}$, the One-Way protocol uses only 2972 bits and thus the Interactive Protocol uses 38% more bits. In both cases, the Interactive Protocol achieves this with zero error, at the cost of variable completion time.

VIII. SUMMARY AND CONCLUSIONS

We presented and analyzed interactive and one-way protocols for solving a 2-classifier problem in a sensor network with n sensor nodes. The sum rate of the interactive protocol depends on the mean stopping time of the protocol. Analysis of both protocols is presented. The analysis reveals a $\log n$ gap between the interactive protocol and the lower bound. The relative performance of the interactive and one-way protocols is seen to depend strongly on the error probability required of the one-way protocol. The cause of the gap between the interactive protocol and its lower bound requires further investigation.

REFERENCES

[1] M. F. Bollauf, V. A. Vaishampayan and S. I. R. Costa, "On the communication cost of determining an approximate nearest lattice point," 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, 2017, pp. 1838-1842. doi: 10.1109/ISIT.2017.8006847.

[2] P. L. Butzer, L. U. Hahn and U. Westphal, "On the rate of approximation in the central limit theorem", Journal of approximation theory, vol. 13, No. 3, pp. 327–340, 1975.

[3] R. A. Horn and C. J. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, U.K., 1985.

[4] E. Kushilevitz and N. Nisan, *Communication Complexity*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[5] A. Orlitsky, "Average-case interactive communication," IEEE Transactions on Information Theory, vol. 38, No. 5, pp.1534-1547. Sep. 1992.

[6] A. Orlitsky and A. El Gamal. "Average and randomized communication complexity." IEEE Transactions on Information Theory, vol. 36, no. 1, pp. 3-16, Jan. 1990.

[7] V. A. Vaishampayan, "Towards a Converse for the Nearest Lattice Point Problem," CoRR, vol. abs/1711.04714, http://arxiv.org/abs/1711.04714, 2017.

[8] V. A. Vaishampayan and M. F. Bollauf, "Communication cost of transforming a nearest plane partition to the Voronoi partition," 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, 2017, pp. 1843-1847. doi: 10.1109/ISIT.2017.8006848

[9] Y. Xiang and Y. H. Kim. Interactive hypothesis testing with communication constraints. In *Proc. 2012 50th Annual Allerton Conference on Communication, Control, and Computing*, pp. 1065–1072, Oct 2012.

[10] A. C. Yao, "Some Complexity Questions Related to Distributive Computing(Preliminary Report)". In Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing, STOC '79, 209-213. 1979.