On the Communication Complexity of Greater-Than

Sivaramakrishnan Natarajan Ramamoorthy[1] and Makrand Sinha[2]

***Abstract*— We give a simple information theoretic proof that the public-coin randomized communication complexity of the greater-than function is $\Omega(\log n)$ for bit-strings of length n.**

I. Introduction

For $x \in \{0,1\}^n$, let $\mathsf{bin}(x)$ denote the integer whose binary representation is $x^{\ddagger}$. The greater-than function $\mathsf{GT} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as

$$\mathsf{GT}(x,y) = \begin{cases} 1 \text{ if } \mathsf{bin}(x) \geq \mathsf{bin}(y) \\ 0 \text{ otherwise.} \end{cases}$$

Nisan [1] showed that the *public-coin* randomized communication complexity of the greater-than function is $\mathscr{O}(\log n)$ for bit-strings of length n. Using information theoretic techniques Viola [2] gave a matching lower bound. Braverman and Weinstein [3] gave an alternative proof for the lower bound by analyzing the discrepancy of the greater-than function. In this article, we give a very short information-theoretic proof that the public-coin randomized communication complexity of greater-than function is $\Omega(\log n)$. Denoting by $\mathsf{R}(f)$ the communication complexity of any public-coin protocol that computes a boolean function f with error at most $1/3$, we prove the following,

Theorem 1: $\mathsf{R}(\mathsf{GT}) = \Omega(\log n)$.

II. Preliminaries

A. Probability Spaces and Variables

Unless otherwise stated, logarithms in this text are computed base two. Random variables are denoted by capital letters (e.g. A) and values they attain are denoted by lower-case letters (e.g. a). Events in a probability space will be denoted by calligraphic letters (e.g. $\mathscr{E}$). Given $a = a_1, a_2, \ldots, a_n$, we write $a_{\leq i}$ to denote $a_1, \ldots, a_i$. We define $a_{<i}$ similarly. We use $[n]$ to denotes the set $\{1, 2, \ldots, n\}$.

We use the notation $p(a)$ to denote both the distribution on the variable a, and the number $\Pr_p[A = a]$. The meaning will be clear from context. We write $p(a|b)$ to denote either the distribution of A conditioned on the event $B = b$, or the number $\Pr[A = a | B = b]$. Given a distribution $p(a,b,c,d)$, we write $p(a,b,c)$ to denote the marginal distribution on the variables a,b,c (or the corresponding probability). We often write $p(ab)$ instead of $p(a,b)$ for conciseness of notation. If $\mathscr{W}$ is an event, we write $p(\mathscr{W})$ to denote its probability according to p. We denote by $\mathbb{E}_{p(a)}[g(a)]$ the expected value of $g(a)$ in p.

B. Divergence and Mutual Information

The *divergence* between p,q is defined to be $\frac{p(A)}{q(A)} = \sum_a p(a) \log \frac{p(a)}{q(a)}$. For three random variables A,B,C with underlying probability distribution $p(a,b,c)$, and an event $\mathscr{E}$ in the same probability space, we will use the shorthand $\frac{A|bc\mathscr{E}}{A|c} = \frac{p(A|bc\mathscr{E})}{p(A|c)}$, when p is clear from context. The *mutual information* between A,B conditioned on C is defined as

$$\mathbf{I}(A:B|C) = \mathop{\mathbb{E}}_{c,b}\left[\frac{A|bc}{A|c}\right] = \mathop{\mathbb{E}}_{c,a}\left[\frac{B|ac}{B|c}\right] = \sum_{a,b,c} p(abc) \log \frac{p(a|bc)}{p(a|c)}.$$

Define $\mathsf{h}(a) := -a\log a - (1-a)\log(1-a)$ to be the binary entropy function and $\mathsf{d}(a||b) := a \log \frac{a}{b} + (1-a)\log\frac{1-a}{1-b}$ to be the binary divergence.

The proofs of following basic facts can be found in [4]:

Proposition 2: For random variables A and B where $B \in \{0,1\}$, $\mathbf{I}(A:B) \leq \mathsf{h}(p(B=1))$.

Proposition 3 (Chain Rule): If $a = a_1, \ldots, a_s$, then $\frac{p(A)}{q(A)} = \sum_{i=1}^{s} \mathbb{E}_{p(a)}\left[\frac{p(A_i|a_{<i})}{q(A_i|a_{<i})}\right]$.

Proposition 4: For an event $\mathscr{W}$, $\frac{A|\mathscr{W}}{A} \leq \log \frac{1}{p(\mathscr{W})}$.

Proposition 5: For any $0 \leq \varepsilon, \delta < 1/2$, $\mathsf{d}\left(1-\varepsilon \middle|\middle| \frac{1}{2}+\delta\right)$ is a decreasing function of both ε and δ. Furthermore,

$$\mathsf{d}\left(1-\varepsilon \middle|\middle| \frac{1}{2}+\delta\right) \geq 1 - \varepsilon \log\left(\frac{4}{\varepsilon}\right) - 4\delta.$$

Proof: We can write

$$\begin{aligned}
&\mathsf{d}\left(1-\varepsilon \middle|\middle| \frac{1}{2}+\delta\right) \\
&= (1-\varepsilon)\log\left(\frac{2}{1+2\delta}\right) + \varepsilon \log\left(\frac{2}{1-2\delta}\right) - \mathsf{h}(\varepsilon) \\
&= \log 2 - \log(1+2\delta) + \varepsilon \log\left(\frac{1+2\delta}{1-2\delta}\right) - \mathsf{h}(\varepsilon) \\
&\geq 1 - \log(1+2\delta) - \mathsf{h}(\varepsilon) \geq 1 - 4\delta - \mathsf{h}(\varepsilon),
\end{aligned}$$

[1]Sivaramakrishnan Natarajan Ramamoorthy is with Computer Science and Engineering, University of Washington. sivanr@cs.washington.edu. Supported by the National Science Foundation under agreement CCF-1016565

[2]Makrand Sinha is with Computer Science and Engineering, University of Washington. makrand@cs.washington.edu. Supported by the National Science Foundation under agreement CCF-1016565, and by the Binational Science Foundation under agreement 2010089.

[‡]We adopt the convention that the leftmost bit of x is the most-significant bit of $\mathsf{bin}(x)$.

where we used that $\log(1+2\delta) \le 2\delta/\ln 2 \le 4\delta$.

Furthermore, we can upper bound $\mathsf{h}(a) \le a\log(4/a)$ for $0 \le a \le 1/2$. This can be observed by noting that $(1-a)\log\frac{1}{1-a} \le 2a$ for $0 \le a \le 1/2$. Therefore, $\mathsf{h}(a) \le a\log(1/a) + 2a = a\log(4/a)$. ■

C. Communication Complexity

We briefly describe basic properties of communication protocols that we need. For more details see the book by Kushilevitz and Nisan [5]. The *communication complexity* of a protocol is the maximum number of bits that may be exchanged by the protocol. For a protocol π, we denote by $\pi(m)$ the output of the protocol π when the messages exchanged are m. For a boolean function f, we denote by $\mathsf{R}_\varepsilon(f)$ (resp. $\mathsf{R}(f)$) the minimum communication complexity of any public-coin randomized protocol that computes f with error at most ε (resp. 1/3). Given a distribution $p(x,y)$ over inputs, we denote by $\mathsf{D}^p_\varepsilon(f)$ (resp. $\mathsf{D}^p(f)$) the minimum communication complexity of a deterministic protocol that computes f with error at most ε (resp. 1/3) over the distribution p.

Proposition 6 (Yao's min-max): For any boolean function f, $\mathsf{R}_\varepsilon(f) = \max_p \mathsf{D}^p_\varepsilon(f)$.

The above proposition implies that for the purpose of proving lower bound, it suffices to consider deterministic protocols. For a deterministic protocol π, let $\pi(x,y)$ denote the messages of the protocol on inputs x,y and define events,

$$\mathscr{S}_m = \{x \mid \exists y \text{ such that } \pi(x,y) = m\},$$
$$\mathscr{T}_m = \{y \mid \exists x \text{ such that } \pi(x,y) = m\}.$$

Proposition 7 (Messages Correspond to Rectangles): $\pi(x,y) = m \iff x \in \mathscr{S}_m$ and $y \in \mathscr{T}_m$.

Proposition 7 implies:

Proposition 8 (Markov Property of Protocols): Let X and Y be inputs to a deterministic communication protocol with messages M. If X and Y are independent then $X - M - Y$.

Note that the above implies that if x and y are independent inputs sampled from a distribution p, then $p(xy|m) = p(xy|\mathscr{S}_m\mathscr{T}_m) = p(x|\mathscr{S}_m)p(y|\mathscr{T}_m)$.

III. Communication Lower Bound

We will prove that any public-coin protocol that computes GT with error at most 1/3 must have communication $\Omega(\log n)$. Firstly, by repeating the protocol a constant number of times, we get a randomized protocol that computes GT with error at most 1/10000 with only a constant blowup in communication.

Using Yao's min-max principle (Proposition 6), it suffices to give a distribution for which the distributional communication complexity with error 1/10000 is $\Omega(\log n)$. We use the following distribution to show the lower bound (Note that the distribution we use is a variant of the distribution described in [2] and [3]).

a) Hard Distribution:: Let $J \in [\frac{n}{2}]$ be uniformly random. $X, Y \in \{0,1\}^n$ are sampled uniformly conditioned on the event that $X_{<J} = Y_{<J}$, *i.e.* the most-significant $J-1$ bits of X and Y are always equal.

The communication lower bound follows from the following two lemmas. The first lemma says that any protocol computing GT with error at most 1/10000 must reveal a lot of information about the function value $\mathsf{GT}(X,Y)$.

Lemma 9: If M are the messages of a protocol that computes $\mathsf{GT}(X,Y)$ with error at most 1/10000, then

$$\mathbf{I}(M : \mathsf{GT}(X,Y) | X_{<J}Y_{<J}J) \ge 1 - \frac{1}{10} - \frac{1}{2^{n/2-1}}.$$

Next we show that if the length of the transcript is small, then the protocol could not have revealed a lot of information about $\mathsf{GT}(X,Y)$.

Lemma 10: If $M \in \{0,1\}^\ell$ are the messages of a protocol, then

$$\mathbf{I}(M : \mathsf{GT}(X,Y) | X_{<J}Y_{<J}J) \le \frac{2^{\ell+1}}{n} + \mathsf{h}\left(\frac{1}{4} + \frac{1}{2^{n/2+1}}\right).$$

Since $\mathsf{h}\left(\frac{1}{4} + \frac{1}{2^{n/2+1}}\right) < 0.84$ for $n > 20$, Lemmas 9 and 10 imply that $\mathsf{D}^p_{1/10000}(\mathsf{GT}) = \Omega(\log n)$. Proposition 6 then implies Theorem 1. We proceed with the proofs of Lemmas 9 and 10.

Proof: [Proof of Lemma 9] By the definition of mutual information, we can write

$$\mathbf{I}(M : \mathsf{GT}(X,Y) | X_{<J}Y_{<J}J) = \mathbb{E}_{mxj}\left[\frac{\mathsf{GT}(X,Y)|mx_{<j}y_{<j}j}{\mathsf{GT}(X,Y)|x_{<j}y_{<j}j}\right].$$

Note that $p(\mathsf{GT}(X,Y) = 1 | x_{<j}y_{<j}j) = \frac{1}{2} + \frac{1}{2^{n-j+1}} \le \frac{1}{2} + \frac{1}{2^{n/2+1}}$. Define the event $\mathscr{E} = \{m, x_{<j}, y_{<j}, j \mid p(\mathsf{GT}(X,Y) \neq \pi(m)|mx_{<j}y_{<j}j) \ge 1/100\}$. Since, the error of the protocol is at most 1/10000, Markov's inequality implies that $p(\mathscr{E}) \le 1/100$.

We can now write

$$\begin{aligned}
&\mathbf{I}(M : \mathsf{GT}(X,Y) | X_{<J}Y_{<J}J) \\
&\ge p(\bar{\mathscr{E}})\mathbb{E}_{mxyj|\bar{\mathscr{E}}}\left[\frac{\mathsf{GT}(X,Y)|mx_{<j}y_{<j}j}{\mathsf{GT}(X,Y)|x_{<j}y_{<j}j}\right] \\
&\ge \frac{99}{100} \cdot \mathsf{d}\left(\frac{99}{100}\middle\|\frac{1}{2} + \frac{1}{2^{n/2+1}}\right) \ge 1 - \frac{1}{10} - \frac{1}{2^{n/2-1}},
\end{aligned}$$

where the last inequality follows from Proposition 5. ■

To prove Lemma 10, we need the following lemma. The proof of this lemma is based on a subtle application of chain rule as used in [6], [7], [8], [9].

Lemma 11: If $M \in \{0,1\}^\ell$, then

$$\mathbf{I}(M : X_J | X_{<J}Y_{<J}J) \le \frac{2^{\ell+1}}{n}.$$

We first prove Lemma 10 before giving a proof of the above lemma.

Proof: [Proof of Lemma 10] By the chain rule for mutual information, we have

$$\begin{aligned}&\mathbf{I}(M : \mathsf{GT}(X,Y)|X_{<J}Y_{<J}J)\\&\leq \mathbf{I}(M : \mathsf{GT}(X,Y)X_J|X_{<J}Y_{<J}J)\\&= \mathbf{I}(M : X_J|X_{<J}Y_{<J}J) + \mathbf{I}(M : \mathsf{GT}(X,Y)|X_{\leq J}Y_{<J}J)\\&\leq 2^{\ell+1}/n + \mathbf{I}(M : \mathsf{GT}(X,Y)|X_{\leq J}Y_{<J}J),\end{aligned}$$

where the last inequality follows from Lemma 11.

By Proposition 2, $\mathbf{I}(M : \mathsf{GT}(X,Y)|X_{<J}Y_{<J}J, X_J = 0) \leq \mathsf{h}\left(\frac{1}{4} + \frac{1}{2^{n/2+1}}\right)$ and $\mathbf{I}(M : \mathsf{GT}(X,Y)|X_{<J}Y_{<J}J, X_J = 1) \leq \mathsf{h}\left(\frac{3}{4} - \frac{1}{2^{n/2+1}}\right) = \mathsf{h}\left(\frac{1}{4} + \frac{1}{2^{n/2+1}}\right)$. Therefore, $\mathbf{I}(M : \mathsf{GT}(X,Y)|X_{\leq J}Y_{<J}J) \leq \mathsf{h}\left(\frac{1}{4} + \frac{1}{2^{n/2+1}}\right)$. ■

Proof: [Proof of Lemma 11] Since $X_{<J} = Y_{<J}$, we have

$$\begin{aligned}\mathbf{I}(M : X_J|X_{<J}Y_{<J}J) &= \mathbf{I}(M : X_J|X_{<J}J)\\&= \sum_m p(m) \mathop{\mathbb{E}}_{xj|m}\left[\frac{X_j|mx_{<j}j}{X_j|x_{<j}j}\right]. \qquad (1)\end{aligned}$$

Recall that each message $M = m$ is equivalent to the event $\mathscr{S}_m \wedge \mathscr{T}_m$, where $\mathscr{S}_m, \mathscr{T}_m$ are as in Proposition 7. After fixing $x_{<j}j$, X is independent of Y (and hence $\mathscr{T}_m$). So by Proposition 8, we have $p(x_j|mx_{<j}j) = p(x_j|\mathscr{S}_m x_{<j}j)$.

Using the above observation, (1) can be rewritten as

$$\begin{aligned}&\sum_m p(\mathscr{S}_m)p(\mathscr{T}_m|\mathscr{S}_m) \mathop{\mathbb{E}}_{xj|\mathscr{S}_m\mathscr{T}_m}\left[\frac{X_j|\mathscr{S}_m x_{<j}j}{X_j|x_{<j}j}\right]\\&\leq \sum_m p(\mathscr{S}_m) \mathop{\mathbb{E}}_{xj|\mathscr{S}_m}\left[\frac{X_j|\mathscr{S}_m x_{<j}j}{X_j|x_{<j}j}\right], \qquad (2)\end{aligned}$$

where the inequality follows from the fact that $\mathbb{E}_a[h(a)] \geq p(\mathscr{W})\mathbb{E}_{a|\mathscr{W}}[h(a)]$, for any non-negative function h.

Since J is independent of X (and hence $\mathscr{S}_m$), we can use the chain rule to write the inner expectation as

$$\begin{aligned}\mathop{\mathbb{E}}_{xj|\mathscr{S}_m}\left[\frac{X_j|\mathscr{S}_m x_{<j}j}{X_j|x_{<j}j}\right] &= \mathbb{E}_j \mathop{\mathbb{E}}_{x|\mathscr{S}_m}\left[\frac{X_j|\mathscr{S}_m x_{<j}}{X_j|x_{<j}}\right]\\&= \frac{2}{n}\frac{X_{\leq \frac{n}{2}}|\mathscr{S}_m}{X_{\leq \frac{n}{2}}}.\end{aligned}$$

Now we can bound (2) by,

$$\begin{aligned}\sum_m p(\mathscr{S}_m)\frac{2}{n}\frac{X_{\leq \frac{n}{2}}|\mathscr{S}_m}{X_{\leq \frac{n}{2}}} &\leq \frac{2}{n}\sum_m p(\mathscr{S}_m)\log\frac{1}{p(\mathscr{S}_m)}\\&\leq \frac{2^{\ell+1}}{n},\end{aligned}$$

where the second inequality follows from Proposition 4 and the third from the fact that for $0 \leq \gamma \leq 1$, it holds that $\gamma\log(1/\gamma) \leq \frac{\log e}{e} < 1$. ■

ACKNOWLEDGMENT

We thank Shay Moran, Anup Rao and Cyrus Rashtchian for useful discussions.

REFERENCES

[1] Noam Nisan. The communication complexity of threshold gates. In *In Proceedings of Combinatorics, Paul Erdos is Eighty"*, pages 301–315, 1994.

[2] Emanuele Viola. The Communication Complexity of Addition. In *SODA*, pages 632–651, 2013.

[3] Mark Braverman and Omri Weinstein. A Discrepancy Lower Bound for Information Complexity. In *APPROX-RANDOM*, pages 459–470, 2012.

[4] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[5] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997.

[6] Anat Ganor, Gillat Kol, and Ran Raz. Exponential Separation of Information and Communication. In *FOCS*, pages 176–185, 2014.

[7] Anat Ganor, Gillat Kol, and Ran Raz. Exponential Separation of Information and Communication for Boolean Functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:113, 2014.

[8] Anup Rao and Makrand Sinha. Simplified Separation of Information and Communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.

[9] Anat Ganor, Gillat Kol, and Ran Raz. Exponential Separation of communication and External Information. *Electronic Colloquium on Computational Complexity (ECCC)*, 2015.