On the Power of Lower Bound Methods for One-Way Quantum Communication Complexity

Shengyu Zhang

The Chinese University of Hong Kong
syzhang@cse.cuhk.edu.hk

Abstract. One of the most fundamental questions in communication complexity is the largest gap between classical and quantum one-way communication complexities, and it is conjectured that they are polynomially related for all total Boolean functions f. One approach to proving the conjecture is to first show a quantum lower bound $L(f)$, and then a classical upper bound $U(f) = poly(L(f))$. Note that for this approach to be possibly successful, the quantum lower bound $L(f)$ has to be polynomially tight for all total Boolean functions f.

This paper studies all the three known lower bound methods for one-way quantum communication complexity, namely the Partition Tree method by Nayak, the Trace Distance method by Aaronson, and the two-way quantum communication complexity. We deny the possibility of using the aforementioned approach by any of these known quantum lower bounds, by showing that each of them can be at least exponentially weak for some total Boolean functions. In particular, for a large class of functions generated from Erdös-Rényi random graphs $G(N, p)$, with p in some range of $1/poly(N)$, though the two-way quantum communication complexity is linear in the size of input, the other two methods (particularly for the one-way model) give only constant lower bounds. En route of the exploration, we also discovered that though Nayak's original argument gives a lower bound by the VC-dimension, the power of its natural extension, the Partition Tree method, turns out to be exactly equal to another measure in learning theory called the *extended equivalence query complexity*.

1 Introduction

Communication complexity studies the minimum amount of communication needed to compute a function of inputs distributed over two (or more) parties. Through more than three decades of studies since its invention by Yao [30], it has flourished into a research field with connections to numerous other computational settings such as circuit complexity, streaming algorithms, data structures, decision tree complexity, VLSI design, algorithmic game theory, optimization, pseudo-randomness and so on [19,28].

In the basic two-party setting, Alice and Bob are given inputs $x \in \{0,1\}^n$ and $y \in \{0,1\}^m$, respectively, and they need to compute $f(x, y)$. In the *two-way* model, Alice and Bob are allowed to send messages back and forth; in the one-way

L. Aceto, M. Henzinger, and J. Sgall (Eds.): ICALP 2011, Part I, LNCS 6755, pp. 49–60, 2011.

model, only Alice sends a message to Bob. The least amount of communication needed for the worst-case input in a deterministic, randomized and quantum protocol, with the latter two allowing a bounded error, is the communication complexity in the corresponding models. We denote by $\mathsf{D}(f)$, $\mathsf{R}(f)$ and $\mathsf{Q}(f)$ the deterministic, randomized and quantum communication complexities of function f in the two way model, and $\mathsf{D}^1(f)$, $\mathsf{R}^1(f)$ and $\mathsf{Q}^1(f)$ the corresponding complexities in the one-way model.

Though much weaker than the two-way model, the one-way model has also caused much attention for various reasons: First, the model is powerful enough to admit many efficient protocols, including both cases for specific functions (such as Equality) and cases for general functions (such as the one with cost in terms of the γ_2^∞-norm [20]). Second, the one-way communication complexity has close connections to areas such as streaming algorithms [22]. Third, proving lower bounds of the one-way communication complexity for general functions turns out to be mathematically quite challenging.

Lower bound methods for communication complexity are of particular interest since in most, if not all, connections to other theoretical areas, communication complexity serves as a lower bound. Quantum lower bounds are interesting for another important reason: One of the most fundamental questions in the field is to pin down the largest gap between classical and quantum communication complexities for total Boolean functions. However, the problem is notoriously hard and our knowledge is very limited: the largest known gap between $Q(f)$ and $R(f)$ for a total function is quadratic (achieved by, for example, Disjointness [15,25,8,3,12,26,27]), while the best upper bound of $R(f)$ in terms of $Q(f)$ is still exponential. The situation in the one-way model is more embarrassing: Despite a lot of efforts [1,2,14], the best separation between classical and quantum one-way communication complexities is a factor of 2 (for Equality function as observed by Winter[29]), while the best upper bound of gap is exponential. Actually, it was highly nontrivial to find even relations [6] or partial functions [10,16] with exponential gaps. Based on this and various other facts such as Holevo's bound, it is reasonable to conjecture that $R^1(f)$ and $\mathsf{Q}^1(f)$ are polynomially related for all total Boolean functions f.

Two approaches were previously taken to understand the relation. One is trying to simulate a quantum protocol directly by a randomized one. Unfortunately, the cost of all classical simulations so far have parameters other than $\mathsf{Q}^1(f)$, and those parameters can be easily as large as n for some total functions. The other natural approach is to firstly derive a general lower bound $\mathsf{Q}^1(f) = \Omega(L(f))$, and then prove a matching upper bound $R^1(f) = poly(L(f))$. Recently Jain, Klauck and Nayak proved that the one-way rectangle bound tightly characterizes $R^1(f)$ [13], which raised the hope of proving the polynomial relation by establishing a matching quantum lower bound. Jain and Zhang tried along this way [14], but were only able to prove that the distributional quantum complexity is at least the distributional rectangle bound for all product distributions.

Note that for this approach to succeed for a general total function f, the tightness of the quantum lower bound $L(f)$ is crucial: If it is not always

polynomially tight, then any attempt on establishing a matching classical upper bound is doomed to fail. There are two methods particularly for the quantum one-way model. One is the *trace distance method* for general functions by Aaronson [1]. The other originates in some sense in the paper [4] by Ambainis, Nayak, Ta-Shma, and Vazirani, and is more explicit in [23] by Nayak for the Random Access Code (RAC) problem; we will refer to this technique as the *partition tree* method (for the reason that will be clear from later discussions). Besides these two methods, the two-way quantum communication complexity $\mathsf{Q}(f)$ can also serve as a lower bound for $\mathsf{Q}^1(f)$ by definition. In this paper, we show that

Theorem 1. *None of the above three known lower bound methods for* $\mathsf{Q}^1(f)$ *is polynomially tight. Actually, they can all be at least exponentially weak.*

This theorem implies that any one of the known methods does not suffice to prove the conjecture that $\mathsf{R}^1(f) = poly(\mathsf{Q}^1(f))$. It can also be viewed as an partial explanation on why the conjecture, if true, is so hard to prove. These negative results on the tightness call for new lower bound methods for $\mathsf{Q}^1(f)$, and we hope that the exhibited weakness of these methods can guide us to search for new and more powerful ones.

Next we discuss in more details about our studies of the various lower bound methods. First, unlike the trace distance method, the partition tree method does not have a well-defined formula for general functions. This paper starts from cleaning up the picture of the method, leading to a robust generalization. As an unexpected connection, its power turns out to be exactly the *extended equivalence query complexity* in computational learning theory. This is interesting compared to that Nayak's original argument actually proves $\mathsf{Q}^1(f) \geq VC(f)$, the VC-dimension of f [18].

We then analyze the power of the three lower bound techniques. Various relations between these techniques are studied, among which the advantage of $\mathsf{Q}(f)$ over the other two methods is particularly interesting. Presumably $\mathsf{Q}(f)$ should not be a good lower bound for $\mathsf{Q}^1(f)$ which in general can be much larger; for example, for Index function $\mathsf{Q}(f) \leq \log_2 n$ but $\mathsf{Q}^1(f) = 1$. However, it turns out that for most functions induced by a random graph $G(N, p)$ for a large range of $p = 1/poly(N)$, both the partition tree method and the trace distance methods can only give a constant lower bound for $\mathsf{Q}^1(f)$, while we can show that $\mathsf{Q}(f) = \Omega(n)$ by the generalized discrepancy method [20].

More related work. On the relation of $\mathsf{R}^1(f)$ and $\mathsf{Q}^1(f)$, if parameters other than $\mathsf{Q}^1(f)$ are allowed, then nontrivial classical upper bounds are known: The aforementioned bound in VC-dimension and Sauer's lemma that $D^1(f) = O(mVC(f))$ imply the upper bound $D^1(f) = O(m\mathsf{Q}^1(f))$. Aaronson later generalized this to partial functions $D^1(f) = O(m\mathsf{Q}^1(f)\log\mathsf{Q}^1(f))$ [1] and $\mathsf{R}^1(f) = O(m\mathsf{Q}^1(f))$ [2]. Jain and Zhang [14] improved the last bound to $\mathsf{R}^1(f) = O((I_\mu(X;Y)+1)VC(f))$ for total functions where $I_\mu(X;Y)$ is the mutual information of the correlated inputs (X, Y) under a hard distribution μ. Klauck gave a variant of Nayak's argument in [17], which unfortunately may be weaker than the VC-dimension (up to an logarithm), while our partition tree bound is always at least $VC(f)$.

There are quite a few results on separations of classical and quantum communication complexities for total functions in the so-called SMP model [7] and for partial functions or relations in various other models [8,24,11,9,16].

2 Preliminaries

Suppose Alice's input set is $\mathcal{X}$ with size $N = 2^n$ and Bob's input set is $\mathcal{Y}$ with size $M = 2^m$. The set of inputs $\{(x, y) : f(x, y) = b\}$ is called b-inputs. For a graph $G = (V, E)$, the function $f_G : V \times V \to \{0, 1\}$ is defined as $f_G(x, y) = 1$ iff $(x, y) \in E$. We assume that $f(x, x) = 0$. For a vertex $v \in V$, its neighbor set is denoted by $N(v)$. An N-node random graph in the Erdös-Rényi model $G(N, p)$ is obtained by connecting each pair of vertices independently with probability p. For a graph G, its adjacency matrix is A_G. For a matrix A, let $\sigma_1(A)$, ..., $\sigma_{rank(A)}(A)$ be the singular values of A in the decreasing order.

The trace distance method was introduced by Aaronson.

Theorem 2 (Aaronson, [1]). *Let $f : \{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ be a total Boolean function, and μ is a probability distribution on the 1-input set $\{(x, y) : f(x, y) = 1\}$. Let D_k be the distribution over $(\{0,1\}^n)^k$ formed by first choosing $y \in \mu$ and then choosing k samples independently from the conditional distribution μ_y. Suppose that $\mathbf{Pr}_{x \leftarrow \mu,\ y \leftarrow \mu}[f(x, y) = 0] = \Omega(1)$, where "$x \leftarrow \mu,\ y \leftarrow \mu$" is to draw x and y independently from the two marginal distributions of μ, then $\mathsf{Q}^1(f) = \Omega(\log(1/\|D_2 - D_1^2\|))$.*

Definition 1. *The* trace distance bound *for $\mathsf{Q}^1(f)$ is $\mathsf{TD}(f) = \max_\mu \log_2 \frac{1}{\|D_2 - D_1^2\|}$ where the maximum is taken over all probability distributions μ on the 1-inputs.*

Linial and Shraibman introduced the following lower bound for $\mathsf{Q}(f)$ based on the factorization norm. For a matrix A, define $\gamma_2(A) = \min_{A=BC} \|B\|_{2\to\infty}\|C\|_{1\to 2}$ where for vector norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, the operator norm $\|A\|_{X\to Y} \stackrel{\text{def}}{=} \max_{\|x\|_X=1} \|Ax\|_Y$. For a sign matrix A and $\alpha \geq 1$, let $\gamma_2^\alpha(A) = \min_{B:1\leq a_{ij}b_{ij}\leq\alpha} \gamma_2(B)$.

Theorem 3 (Linial and Shraibman, [20]). $Q_\epsilon(f) \geq \log_2 \gamma_2^{1/(1-2\epsilon)}(f) - O_\epsilon(1)$.

The bound is also known as the *generalized discrepancy method*. The bound actually holds even for $Q^*(f)$, the quantum communication complexity with entanglement shared by Alice and Bob, is lower bounded by the above quantity. Here we are mainly concerned with the case without entanglement because the no-separation conjecture becomes trivial (due to quantum teleportation) if we compare $Q^{1,*}(f)$ and $R^{1,*}(f)$.

Definition 2. *The* ϵ-factorization norm bound *for $Q_\epsilon(f)$ is $\mathsf{FN}_\epsilon(f) = \log_2 \gamma_2^{1/(1-2\epsilon)}(f)$, and the* factorization norm bound *for $Q^*(f)$ is $\mathsf{FN}(f) = \mathsf{FN}_{1/3}(f)$.*

3 The Partition Tree Method

The partition tree bound is defined as follows. Consider a binary *partition tree* $\mathcal{T}$ *of* $\mathcal{X}$, where each node $v = v_1...v_i$ (i is the depth of v) is associated with an input y_v of Bob. Let X be a random variable according to the distribution p over $\mathcal{X}$. This tree induces a subset $\mathcal{X}_v \subseteq \mathcal{X}$ for each node v in the following inductive way: the root corresponds to $\mathcal{X}$, and suppose the set $\mathcal{X}_v$ is defined then the two subsets $\mathcal{X}_{v0}$ and $\mathcal{X}_{v1}$ for its two children $v0$ and $v1$ is defined by $\mathcal{X}_{vb} = \{x \in \mathcal{X}_v : f(x, y_v) = b\}$. A note v is a leaf iff $f(x, y_v)$, for all $x \in \mathcal{X}_v$, have the same value. Define a sequence of random variables $V_1, ..., V_{depth(\mathcal{T})}$ by $\mathbf{Pr}[V_{i+1} = b|V_1...V_i] = p(\mathcal{X}_{V_1...V_ib})/p(\mathcal{X}_{V_1...V_i})$. For a node $v = v_1...v_i$, define $p(v) \stackrel{\text{def}}{=} p(\mathcal{X}_v)$ and $p_v(b) \stackrel{\text{def}}{=} p(\mathcal{X}_{vb}|\mathcal{X}_v)$ for $b \in \{0, 1\}$ and $p_v(\min) \stackrel{\text{def}}{=} \min\{p_v(0), p_v(1)\}$.

Definition 3. *The* partition tree bound *for* $\mathsf{Q}^1(f)$ *is* $\mathsf{PT}(f) = \max_{\mathcal{T},p} \sum_{v\in\mathcal{T}} p(v)p_v(\min)$.

It is not quite immediate to generalize Nayak's argument (for RAC) to this formulation as a lower bound of $\mathsf{Q}^1(f)$. Please see the full version for detailed explanations.

To study $\mathsf{PT}(f)$, first observe that if one can find a balanced binary subtree of height h, then $\mathsf{PT}(f) \geq h(1 - H(\epsilon))$ since one can put half-half probabilities on both branches of each node of the subtree. (Note that this is at least $VC(f)$ but could be much larger than it, as shown in the Greater Than function in the full version.) The following theorem says that this is actually also the best lower bound the partition tree method can provide.

Theorem 4. *There exists a subset* $S \subseteq \mathcal{X}$ *and a partition tree* $\mathcal{T}^*$ *for* f *on* $(S, \mathcal{Y})$ *s.t.* $\mathsf{PT}(f) =$ *the length of the shortest path of* $\mathcal{T}^*$.

See the full version for the proof.

Note that the standard decision tree complexity is to minimize the the length of the longest path, but here the best PT bound is to maximize the length of the shortest path.

It turns out to have an interesting connection to the *extended equivalence query complexity* in learning theory, which we will define using the language of communication complexity as follows. Alice has an input x and Bob wants to exactly learn x by making queries to Alice, who then responses with an answer. Different query models were studied in learning theory.

1. *membership query*: Bob's query is a column y, and Alice's response is $f(x, y)$;
2. *equivalence query*: Bob's query is a string $a \in \{0, 1\}^M$ as a guess of x. If $a = x$, then Alice tells Bob so and the game is over. Otherwise, Alice not only tells Bob that his guess is wrong, but also provides a column y which $f(x, y) \neq a_y$.

   If Bob is restricted to use strings $a \in \{0, 1\}^M$ appearing as rows in the matrix f as queries, then this is called the *equivalence query*; if Bob is allowed to use any string $a \in \{0, 1\}^M$, it is called the *extended equivalence query*.

The minimal number of a particular type of queries Bob needs to make for the worst-case input x is called the query complexity of that type. Denote by $MQ(f)$, $EQ(f)$ and $XEQ(f)$ the membership query complexity, the equivalence query complexity and the extended equivalence query complexity of the function f, respectively. The following theorem gives a characterization of $XEQ(f)$ by relating it to membership query computation.

Theorem 5 (Littlestone, [21]). *$XEQ(f) = \max_{\mathcal{T}} \min_x d(x, \mathcal{T})$, where $\mathcal{T}$ is a membership query computation tree and $d(x, \mathcal{T})$ is the depth of x in $\mathcal{T}$.*

A membership query computation tree is a decision tree with membership queries in the natural way; see the survey [5] for a formal definition (as well as an extensive review of different types of queries in learning theory). Its important relation to our work is that the membership query computation tree is exactly our partition tree, and thus the above theorem and Theorem 4 combined give the following full characterization of PT.

Theorem 6. *$\mathsf{PT}(f) = XEQ(f)$.*

4 Comparisons between the Powers

In this section we will study the power of the lower bound methods, the PT bound part of which uses the limitation result established in the previous section. We will prove Theorem 1 by a circle of comparison results in the order of PT >> Q >> TD >> PT. The first separation is easily exhibited by Index function. Next we will show that though as a lower bound method merely for the two-way complexity, the factorization norm method can be much stronger than the other two methods for the one-way complexity. In fact, for almost all functions f in some range (the precise meaning of which will be clear shortly) the factorization norm gives a linear lower bound for $\mathsf{Q}(f)$ while the other two cannot even prove a super constant lower bound for $\mathsf{Q}^1(f)$. The advantage of FN over TD is given next, and that of FN over PT is given in Section 4.3.

4.1 On the Advantage of the Factorization Norm Method over the Trace Distance Method

In this section we will show that for a random Erdös-Rényi graph $G(N, p)$ for some range of p, we have $\mathsf{FN}(f_G) = \Omega(n)$ but $\mathsf{TD}(f_G) = O(1)$.

Here we consider a random graph $G(N, p)$ since the corresponding limitations for TD and PT are easier to show. By studying the normalized Laplacian operator on the graph, one can show that the factorization norm method gives a good lower bound for most such random graphs.

Theorem 7. *If $\omega(\log^4 N/N) \le p \le 1 - \Omega(1)$, then with probability $1 - o(1)$, an N-node random graph $G(N, p)$ has $\mathsf{FN}(f_G) - O(1) \ge \frac{1}{2}\log_2(pN) - O(1)$.*

The proof is in the full version, from which one can see that actually even the discrepancy bound, a bound weaker than FN, is still at least $\Omega(\log_2(pN)))$.

Next we show that TD can only give a constant lower bound for random graph functions.

Theorem 8. *For $p = o(N^{-6/7})$, a random graph $G(N,p)$ has $\mathsf{TD}(f_G) = O(1)$ with probability $1 - o(1)$.*

Here we are not aiming to maximize the range of p, though we believe that the result still holds for larger p. The main goal is to show the existence of a range $p = 1/poly(N)$. We will first show in the following Lemma 2 that with probability $1 - o(1)$, a random graph $G(N,p)$ has some good properties. The proof uses Lemma 1 which is on the relation of the number of edges and that of vertices with some connection requirement. After these, we will show that for graphs with those properties, the TD bound is very low.

Lemma 1. *For any constant $\delta > 0$, there are constants c and d s.t. for all distinct vertices $V_x = \{x_1, ..., x_c\}$ and $V_z = \{z_1, ..., z_d\}$, if any x_i and z_k share at least one common neighbor, and there is no vertex y connecting to all x_i's and z_k's, then there exists $V_y = \{y_1, ..., y_e\}$, s.t. any pair (x_i, z_k) of vertices are connected via y_j for some $j \in [e]$, and $g/(c+d+e) \geq 4/3 - \delta$, where g is the number of edges between $V_x \cup V_z$ and V_y.*

Proof. For each (x_i, z_k), there is at least one y connecting them. Collect all these y's to form the set V_y. (For pairs (x_i, z_k) with more than one connectors y, we pick an arbitrary one.) Thus each y has degree at least 2, and therefore

$$g/(c+d+e) \geq 2e/(c+d+e). \tag{1}$$

Now we will give another lower bound

$$g/(c+d+e) \geq 1 + (d-2)/(e+6). \tag{2}$$

Combining the two inequalities gives the desired result.

To show the second bound, fix a setting with $g/(e+6)$ minimized. See Figure 1 (where every $N(z_k) \cap V_y$ is the same set simply for convenience of illustration). The way we chose V_y guarantees that $N(V_y)$ contains the whole V_x. Pick a subset $S \subseteq V_y$ with the minimum size *s.t.* the $N(S) \supseteq V_x$. By definition, the number of edges from S to V_x is at least $|V_x| = c$. Define $e_1 = |S|$ and $e_2 = e - |S|$; Condition 2 implies $e_1 \geq 2$. Note that for each z_k, its neighbor set in V_y, *i.e.*

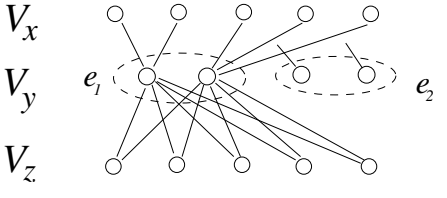

Fig. 1. Illustration for the proof of Lemma 1

$N(z_k) \cap V_y$, also connects to all V_x, thus the number of edges from z_k to V_y is $|N(x_i) \cap V_y| \geq e_1$. Also note that each node in $V_y - S$ has at least one edge to V_x. Thus the total number of edges in this small graph $V_x \cup V_y \cup V_z$ is at least

$$de_1 + c + e_2 = (d-1)e_1 + c + e \geq 2(d-1) + c + e = 1 + (d-2)/(e+6). \quad (3)$$

Lemma 2. *For $p = o(N^{-6/7})$, a random graph $G = G(N,p)$ has all the following properties with probability $1 - o(1)$.*

1. *For any vertex x with (at least) three neighbors y_1, y_2, y_3, at least one of the two pairs (y_1, y_2) and (y_2, y_3) only has x as their common neighbor.*
2. *There are universal constants c and d s.t. for any c vertices $x_1, ..., x_c$ that do not share a common neighbor, there are at most $d-1$ vertices $z_1, ..., z_{d-1}$ which have distance exactly 2 to all x_i's.*
3. *The graph does not contain a $K_{3,2}$, the $(3,2)$-complete bipartite graph, as a subgraph.*

Lemma 3. *Suppose there is a distribution μ on 1-inputs with $\mathbf{Pr}_{x \leftarrow \mu, y \leftarrow \mu}[f(x,y) = 0] = \Omega(1)$ satisfied. If there is a submatrix A s.t. $\mu(A) = 1 - o(1)$, and A as a function has $Q^{1,pub}(A) = q$, then $\|D_2 - D_1^2\|_1 = 2^{\Omega(-q)}$. In particular, $\|D_2 - D_1^2\|_1 = \Omega(1)$ for the following two special cases*

1. *there is a subset $S \subseteq \mathcal{X}$ s.t. $|S| = O(1)$ and $\mu(S) = 1 - o(1)$,*
2. *there is a submatrix A s.t. each column is monochromatic except for at most $O(1)$ entries, and $\mu(A) = 1 - o(1)$.*

See the full version for proofs of the above two lemmas.

Proof. (of Theorem 8) We take the graphs with all good properties in Lemma 2. It is enough to show that any distribution μ on the edge set E with the following condition satisfied

$$\mathbf{Pr}_{x \leftarrow \mu,\ y \leftarrow \mu}[f(x,y) = 0] = \Omega(1), \quad (4)$$

has that $\|D_2 - D_1^2\|_1 = \Omega(1)$. Assuming that it is not true, *i.e.* $\|D_2 - D_1^2\|_1 = o(1)$, we will first show that this assumption forces μ to put most mass on just one star-shape cluster of vertices, then show that in this case, it is also unavoidable to have $\|D_2 - D_1^2\|_1 = \Omega(1)$ finally.

For two vertices x and x', we say x *covers* x', denoted by $x \sim x'$, if they share a common neighbor y. Otherwise we write $x \nsim x'$. For a vertex itself, we assume $x \sim x$ as long as x has a non-zero degree. Define the set $Cover(x) = \{x' :\ x \sim x'\}$. By definition,

$$\begin{aligned}\|D_2 - D_1^2\|_1 &= \sum_{x,x'} \left| \sum_y \mu(y)\mu(x|y)\mu(x'|y) - \mu(x)\mu(x') \right| \\ &= \sum_{x,x':\ x \sim x'} \left| \sum_y \mu(y)\mu(x|y)\mu(x'|y) - \mu(x)\mu(x') \right| + \sum_{x,x':\ x \nsim x'} \mu(x)\mu(x')\end{aligned}$$

Thus

$$\sum_x \mu(x)\mathbf{Pr}_{x'\leftarrow\mu}[x \nsim x'] = \sum_{x,x':\ x\nsim x'} \mu(x)\mu(x') \leq \|D_2 - D_1^2\|_1 = o(1) \quad (5)$$

This means that an average x covers most other vertices x' (weighted under μ). In particular, there exists one x_0 *s.t.* $\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x_0] = o(1)$. Suppose its neighbor set is $N(x_0) = \{y_1, ..., y_t\}$, and define *clusters* $S_i = N(y_i) - \{x_0\}$, and put $S = \cup_i S_i$. Also note that $Cover(x_0)$ is nothing but $S \cup \{x_0\}$, thus $\mu(S \cup \{x_0\}) = 1 - o(1)$. Note that by Lemma 3, we can assume that $\mu(x_0) = 1 - \Omega(1)$ (otherwise the desired conclusion has already been proved). Denote by $\mu_{\neg x_0}$ the distribution $\mu(x)$ conditioned on $x \neq x_0$.

Lemma 4. *At least one of the following two statements is true:*

1. *There are two disjoint subsets G_1 and G_2 of S s.t. $\mu_{\neg x_0}(G_b) = \Omega(1)$ for both $b = 1, 2$, and any two vertices $x_1 \in G_1$ and $x_2 \in G_2$ belong to two different clusters.*
2. *There is a single cluster S_i with $\mu_{\neg x_0}(S_i) = 1 - o(1)$.*

See the full version for a proof. We continue the proof of Theorem 8. If the second statement of the above claim is true, it means that there is a single $y_i \in N(x_0)$ *s.t.* $\mu(N(y_i)) = 1 - o(1)$. (Note that $N(y_i)$ includes a cluster and x_0 itself.) By the third property of Lemma 2, each vertex y other than y_i only connects to at most two vertices in $N(y_i)$ (to avoid a $(3, 2)$-complete bipartite graph). Thus the submatrix on $N(y_i) \times \mathcal{Y}$ has $1 - o(1)$ μ-mass but has all 1's in column y_i and at most two 1's in all other columns. By Lemma 3, we see that $\|D_2 - D_1^2\|_1 = \Omega(1)$.

Therefore, we can assume that the first statement of the claim is the case, so

$$\mathbf{E}_{x\leftarrow\mu_{\neg x_0}}\left[\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x]\right] \geq \sum_{b=1,2} \mu_{\neg x_0}(G_b)\mathbf{E}_{x\leftarrow\mu_{\neg x_0}}\left[\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x] \;\middle|\; x \in G_b\right].$$

On the other hand, we have

$$\mathbf{E}_{x\leftarrow\mu_{\neg x_0}}\left[\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x]\right] = \frac{\sum_{x\neq x_0}\mu(x)\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x]}{1 - \mu(x_0)} = \frac{o(1)}{\Omega(1)} = o(1). \quad (6)$$

Since both $\mu_{\neg x_0}(G_b) = \Omega(1)$, we have $\mathbf{E}_{x\leftarrow\mu_{\neg x_0}}\left[\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x] \;\middle|\; x \in G_b\right] = o(1)$ for both $b = 1, 2$. Therefore, we can find two points $x_b \in G_b$ both with $\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x_b] = o(1)$. This means that for both $b = 1, 2$, most of mass of μ is put on $Cover(x_b)$. Combined with the same fact for x_0, we see that actually $\mu(\cap_{i=0,1,2} Cover(x_i)) = 1 - o(1)$. But note that both x_1 and x_2 are covered by x_0 since they are chosen from S, and they are not in the same cluster as guaranteed by the first statement of the above claim. Consequently, x_0, x_1, x_2 do not share a common neighbor.

Now define set $T = \{x_0, x_1, x_2\}$. As long as $|T|$ is constant, we can assume by Lemma 3 that $\mu(T) = 1 - \Omega(1)$. Then similar to Eq (6), it follows that $\mathbf{E}_{x\leftarrow\mu}[\mathbf{Pr}_{x'\leftarrow\mu}[x' \nsim x] | x \notin T] = o(1)$. Thus there exists another point x in $S - T$

s.t. $\mathbf{Pr}_{x' \leftarrow \mu}[x' \nsim x] = o(1)$. Add this point to T and continue this process until $|T| = c$. Each point $x \in T$ has the property that $\mu(Cover(x)) = 1 - o(1)$, and consequently $\mu(\cap_{x \in T} Cover(x)) = 1 - o(1)$ by noting that $|T| = c$ is a constant. Also recall that the vertices in T do not share a common neighbor since actually even x_0, x_1, x_2 do not. By the second property of Lemma 2, the intersection of their cover sets has only constant size, and thus using Lemma 3 we get $\|D_2 - D_1^2\|_1 = \Omega(1)$. This completes the proof.

4.2 On the Advantage of the Trace Distance Method over the Partition Tree Method

We observed that the partition tree method can be much better than the factorization norm method, and have shown that the factorization norm method can be much better than the trace distance method. To finish the circle, we now show that the trace distance method can be much better than the partition tree method. Different than Theorem 9, this time we can give an explicit function to show the separation.

The Coset function $Coset(G)$ is defined as follows. For a fixed group G, Alice is given a coset x as her input and Bob is given an element $y \in G$ as his input; the question is whether $y \in x$. Aaronson [1] studied the function for the group $\mathbb{Z}_p^2$ (where p is a prime number) and proved that $\mathsf{Q}^1(Coset(\mathbb{Z}_p^2)) = \Theta(\log p)$; that is, Alice asymptotically needs to send the whole input to Bob. Here we show that the partition tree method can only give a very small constant lower bound for this function. The proof is in the full version.

Proposition 1. $\mathsf{PT}(Coset(\mathbb{Z}_p^2)) = 2$.

4.3 Other Discussions of the Power Comparisons

The main goal of this paper is to study the ultimate power of the known lower bound methods for $\mathsf{Q}^1(f)$, and in particular their tightness because of the no-separation conjecture reason mentioned in Section 1. Though it is not our goal to thoroughly study all the six relations between the three methods, it is good to know for more insights. This section so far showed three of them as a circle, leaving the three other relations to discuss. First, it turns out that PT is also weak for random graph functions.

Theorem 9. *For any $\alpha = \Omega(1)$, if $p = N^{-\alpha}$, then an N-node random graph $G(N,p)$ has $\mathsf{PT}(f_G) = O(1)$ with probability $1 - o(1)$.*

For PT over TD, we believe that actually $\mathsf{TD}(\mathsf{Index}) = O(\log n)$, though we can only show it for the *symmetric* distribution μ, *i.e.* $\mu(x,y) = \mu(x',y')$ if $|x| = |x'|$.

Theorem 10. *For any distribution p on $\{0,1,...,n\}$, let $\mu_p(x,y) = p(|x|)$ for all (x,y) with $x_y = 1$. Then the trace distance bound under μ_p for the* Index *function is only $O(\log n)$.*

See the full version for both proofs.

5 Concluding Remarks and Open Questions

The tightness results in this paper call for new lower bound methods for $\mathsf{Q}^1(f)$. With the light shed by comparisons in Section 4, one (vague) approach is trying to somehow combine the advantages of the methods to get a more powerful one.

The factorization norm method appears pretty strong for lower bounding $\mathsf{Q}(f)$. Can we modify it to obtain a good lower bound for $\mathsf{Q}^1(f)$?

Acknowledgment. We would like to thank Rahul Jain for many valuable discussions during the collaboration of paper [14], and Yi-Kai Liu for pointing out the reference [5].

The work was partially supported by China Basic Research Grant 2011CBA00300 (sub-project 2011CBA00301) and Hong Kong General Research Fund 419309 and 418710. The author also benefited from visiting Centre of Quantum Technologies and Tsinghua University, the latter under the support of China Basic Research Grant 2007CB807900 (sub-project 2007CB807901).

References

1. Aaronson, S.: Limitations of quantum advice and one-way communication. Theory of Computing 1, 1–28 (2005)
2. Aaronson, S.: The learnability of quantum states. Proceedings of the Royal Society A 463, 2088 (2007)
3. Aaronson, S., Ambainis, A.: Quantum search of spatial regions. Theory of Computing 1, 47–79 (2005)
4. Ambainis, A., Nayak, A., Ta-Shma, A., Vazirani, U.: Dense quantum coding and quantum finite automata. Journal of the ACM 49(4), 1–16 (2002)
5. Angluin, D.: Queries revisited. Theoretical Computer Science 313(2), 175–194 (2004)
6. Bar-Yossef, Z., Jayram, T.S., Kerenidis, I.: Exponential separation of quantum and classical one-way communication complexity. In: Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC), pp. 128–137 (2004)
7. Buhrman, H., Cleve, R., Watrous, J., de Wolf, R.: Quantum fingerprinting. Physical Review Letters 87(16) (2001)
8. Buhrman, H., Cleve, R., Wigderson, A.: Quantum vs. classical communication and computation. In: Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing (STOC), pp. 63–68 (1998)
9. Gavinsky, D.: Classical interaction cannot replace a quantum message. In: Proceedings of the Fortieth Annual ACM Symposium on the Theory of Computing (STOC), pp. 95–102 (2008)
10. Gavinsky, D., Kempe, J., Kerenidis, I., Raz, R., de Wolf, R.: Exponential separation of quantum and classical one-way communication complexity. In: Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC), pp. 516–525 (2007)
11. Gavinsky, D., Pudlák, P.: Exponential separation of quantum and classical non-interactive multi-party communication complexity. In: Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, pp. 332–339 (2008)

12. Høyer, P., Mosca, M., de Wolf, R.: Quantum search on bounded-error inputs. In: Baeten, J.C.M., Lenstra, J.K., Parrow, J., Woeginger, G.J. (eds.) ICALP 2003. LNCS, vol. 2719, pp. 291–299. Springer, Heidelberg (2003)
13. Jain, R., Klauck, H., Nayak, A.: Direct product theorems for classical communication complexity via subdistribution bounds. In: Proceedings of the Fortieth Annual ACM Symposium on the Theory of Computing (STOC), pp. 599–608 (2008)
14. Jain, R., Zhang, S.: New bounds on classical and quantum one-way communication complexity. Theoretical Computer Science 410(26), 2463–2477 (2009)
15. Kalyanasundaram, B., Schintger, G.: The probabilistic communication complexity of set intersection. SIAM Journal on Discrete Mathematics 5(4), 545–557 (1992)
16. Klartag, B., Regev, O.: Quantum one-way communication can be exponentially stronger than classical communication. In: Proceedings of the 44th Annual ACM Symposium on the Theory of Computing (STOC) (to appear, 2011)
17. Klauck, H.: Quantum communication complexity. In: ICALP Satellite Workshops, pp. 241–252 (2000)
18. Klauck, H.: Lower bounds for quantum communication complexity. SIAM Journal on Computing 37(1), 20–46 (2007)
19. Kushilevitz, E., Nisan, N.: Communication Complexity. Cambridge University Press, Cambridge (1997)
20. Linial, N., Shraibman, A.: Lower bounds in communication complexity based on factorization norms. In: Proceedings of the Thirty-Ninth Annual ACM symposium on Theory of Computing (STOC), pp. 699–708 (2007)
21. Littlestone, N.: Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. Machine Learning 2(4), 285–318 (1988)
22. Muthukrishnan, S.M.: Data streams: Algorithms and applications. Foundations and Trends in Theoretical Computer Science 1(2) (2005)
23. Nayak, A.: Optimal lower bounds for quantum automata and random access codes. In: Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 124–133 (1999)
24. Raz, R.: Exponential separation of quantum and classical communication complexity. In: Proceedings of the 31st Annual ACM Symposium on the Theory of Computing (STOC), pp. 358–367 (1999)
25. Razborov, A.: On the distributional complexity of disjointness. Theoretical Computer Science 106, 385–390 (1992)
26. Razborov, A.: Quantum communication complexity of symmetric predicates. Izvestiya: Mathematics 67(1), 145–159 (2003)
27. Sherstov, A.: The pattern matrix method for lower bounds on quantum communication. In: Proceedings of the 40th Annual ACM Symposium on the Theory of Computing, pp. 85–94 (2008)
28. Wigderson, A.: Depth through breadth, or why should we attend talks in other areas? In: Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC), p. 579 (2004), `http://www.math.ias.edu/~avi/TALKS/STOC04.ppt`
29. Winter, A.: Quantum and classical message identification via quantum channels. Quantum Information and Computation 4(6&7), 563–578 (2004)
30. Yao, A.C.-C.: Some complexity questions related to distributive computing. In: Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing (STOC), pp. 209–213 (1979)