QUANTUM COMMUNICATION COMPLEXITY OF DISTRIBUTION TESTING

ALEKSANDRS BELOVS

Faculty of Computing, University of Latvia
Raina bulvaris 19, Riga, LV-1050, Latvia

ARTURO CASTELLANOS

Graduate School of Informatics, Kyoto University
Yoshida Honmachi, Sakyo-ku, Kyoto 606-8501, Japan

FRANCOIS LE GALL

Graduate School of Mathematics, Nagoya University
Furocho, Chikusaku, Nagoya, Aichi 464-8602, Japan

GUILLAUME MALOD

Université de Paris, Sorbonne Université, CNRS,
Institut de Mathématiques de Jussieu-Paris Rive Gauche, F-75013 Paris, France

ALEXANDER A. SHERSTOV

University of California, Los Angeles
404 Westwood Plaza, Los Angeles, CA 90095, USA

The classical communication complexity of testing closeness of discrete distributions has recently been studied by Andoni, Malkin and Nosatzki (ICALP'19). In this problem, two players each receive t samples from one distribution over $[n]$, and the goal is to decide whether their two distributions are equal, or are ϵ-far apart in the l_1-distance. In the present paper we show that the quantum communication complexity of this problem is $\tilde{O}(n/(t\epsilon^2))$ qubits when the distributions have low l_2-norm, which gives a quadratic improvement over the classical communication complexity obtained by Andoni, Malkin and Nosatzki. We also obtain a matching lower bound by using the pattern matrix method. Let us stress that the samples received by each of the parties are classical, and it is only communication between them that is quantum. Our results thus give one setting where quantum protocols overcome classical protocols for a testing problem with purely classical samples.

1 Introduction

Background. Property testing [1, 2] is the task of (approximately) distinguishing objects having some specific property from those which are "far" from having it, without necessarily looking at the objects in their entirety. An interesting subfield is discrete distribution testing [3], where the objects are probability distributions.

One of the main tasks in (discrete) distribution testing, namely closeness testing, is about deciding whether two distributions p and q over $[n]$ are equal or ϵ-far from each other in the l_1-norm, given access only to a limited number of samples of each distribution. Early testers [4, 5] used a method based on collisions. Using improved estimators, testers with optimal sample complexity have then been constructed [6, 7].

Very recently, Andoni, Malkin and Nosatzki [8] have, for the first time, considered distribution testing in the two-party setting. Here two players, Alice and Bob, each own as input t samples of the distributions p and q: Alice has t samples from p and Bob has t samples from q. The goal is for Alice and Bob to decide if the two distributions are equal or ϵ-far from each other in the l_1-norm, using as little communication as possible. By adapting the techniques from prior works [6, 7], Andoni, Malkin and Nosatzki have shown that this problem (named $2\mathrm{PCT}_{n,t,\epsilon}$ in [8]) can be solved with high probability using $O\left(\frac{n^2}{t^2\epsilon^4} + 1\right)$ bits of communication whenever t is above the information-theoretic lower bound (given in Equation (1) below) which is the minimum number of samples needed so that meaningful information about p and q can be extracted from them. They also showed a matching lower bound $\Omega\left(\frac{n^2}{t^2}\right)$ on the two-party communication complexity of $2\mathrm{PCT}_{n,t,1/2}$.

Our results. In this paper we investigate the quantum communication complexity of this problem. Our main result shows that a significant advantage can be obtained in the quantum setting when at least one of the two distributions has low l_2-norm. Concretely, for any $\gamma > 0$, we consider the version of $2\mathrm{PCT}_{n,t,\epsilon}$ in which the inputs p, q satisfy the condition $\min(||p||_2, ||q||_2) \leq \gamma t\epsilon^2/n$. We denote this problem $2\mathrm{PCT}_{n,t,\epsilon}^{\gamma}$. The lower bound from [8] shows that for $\gamma = \Omega(\frac{1}{\sqrt{\log n}})$, this version is as hard as the original version of the problem.[a] As in all previous works [8, 6, 7], we will assume throughout the paper that t is above the information-theoretic threshold:

$$t \geq C \max(n^{2/3} \cdot \epsilon^{-4/3}, \sqrt{n} \cdot \epsilon^{-2}), \tag{1}$$

where C is a universal constant (see [8, 6] for details).

We first show the following theorem.

Theorem 1 *There exists an absolute constant γ_0 such that the following holds: for all $\gamma \leq \gamma_0$, the problem $2\mathrm{PCT}_{n,t,\epsilon}^{\gamma}$ can be solved with high probability by a quantum protocol that uses $\tilde{O}\left(\frac{n}{t\epsilon^2} + 1\right)$ qubits of communication.*

Theorem 1 shows that a significant advantage (a quadratic improvement in the communication complexity) can be obtained in the quantum setting when at least one of the two distributions has low l_2-norm. While we are not currently able to remove the low-norm requirement, we believe that Theorem 1, which already covers a large class of distributions, is a first significant step towards establishing quantum advantage for closeness testing for arbitrary distributions.

We also obtain the following lower bound, which shows that the upper bound of Theorem 1 is optimal, even for $\epsilon = 1/2$.

Theorem 2 *There exists an absolute constant $c > 1$ such that the following statement holds for any value $t \leq n/\log^c n$ and any $\gamma = \Omega(1/\sqrt{\log n})$: any quantum protocol that solves $2\mathrm{PCT}_{n,t,1/2}^{\gamma}$ with high probability requires $\tilde{\Omega}(n/t)$ qubits of communication.*

---

[a]Indeed, the lower bound $\Omega\left(\frac{n^2}{t^2}\right)$ from [8] is shown for input distributions such that $||p||_2 = ||q||_2 = O(\frac{t\epsilon^2}{n\sqrt{\log n}})$ with $\epsilon = 1/2$.

Precisely, the lower bound holds for any γ greater than some $\gamma_{LW} = O(\frac{1}{\sqrt{\log n}})$. Therefore there is a regime where $\gamma_{LW} \leq \gamma_0$, for which our upper bound is then tight.

Overview of our main techniques. Our upper bound (Theorem 1) is obtained by following the framework used in [8], which relies on the estimator from [6] for the l_2-distance. Indeed, as suggested by [4, 6] and then extensively studied in [7], there is a reduction from closeness testing in the l_1-distance to closeness testing in the l_2-distance. The efficiency of the reduction, however, depends on the l_2-norm of the distributions. In the presentation below we will assume that $||p||_2 \approx ||q||_2$ (if this is not the case we can easily conclude that $p \neq q$). The protocol in [8] thus proceeds in two steps. In the first step, Bob shares some information with Alice about the observed shape of his distribution, so that they can recast their distributions into two distributions p' and q' that have smaller l_2-norm while preserving the l_1-distance (i.e., $||p' - q'||_1 = ||p - q||_1$). In the second step, Alice and Bob use the reduction to closeness testing in the l_2-distance mentioned and implement the estimator of [6] in the two-party setting. This estimator requires estimating with good precision the l_2-distance between two vectors. This is done by using a two-party implementation of the sketching method by Alon, Matias and Szegedy [9].

For the case of low-norm distributions (more precisely, when considering the problem $2\mathrm{PCT}^\gamma_{n,t,\epsilon}$ with γ constant), the first step is unnecessary: the two distributions already have a low enough l_2-norm, so that the reduction to closeness testing in the l_2-distance can be used without any preprocessing. We thus only need to show how to implement the second step from [8] more efficiently using quantum communication. The key idea is to use the quantum algorithm by Montanaro [10] which gives a quadratic speedup over the classical sketching method from [9] in the query complexity model. We show how to adapt this quantum algorithm to the two-party setting in Section 3.2.

Our lower bound (Theorem 2) first applies the same reduction as in [8], which reduces some specific version of the Gap-Hamming distance to $2\mathrm{PCT}^\gamma_{n,t,1/2}$. In the classical case, [8] showed that the communication complexity of that version of the Gap-Hamming distance is $\tilde{\Omega}((n/t)^2)$ bits. Our main technical contribution proves that the quantum communication complexity of this problem is $\tilde{\Omega}(n/t)$. We use the pattern matrix method [11]. To obtain our lower bound, we show that the pattern matrix method, which is generally formulated only for total functions, can be generalized to partial functions.

Relation with known quantum advantages in testing and learning. Several quantum algorithms have been designed for property testing and learning theory (we refer to [12] and [13] for excellent surveys of these fields). In most settings considered so far in quantum learning theory, however, the quantum algorithms crucially exploit the fact that the data can be accessed in a quantum way (e.g., we can query a quantum superposition of samples), which makes it difficult to directly compare the performance of quantum algorithms with the performance of classical methods (which can only access the data in a classical way). The results of the present paper show a quantum advantage, in terms of communication cost, for the setting where both classical and quantum protocols can access the data in the same way – the input is given as a set of classical samples.

Open problem. An intriguing question is whether a similar quantum advantage is achievable when both input distributions have higher l_2-norm, i.e., whether the upper bound we obtain in Theorem 1 holds not only for constant γ but also for larger values of γ. Currently, we do not know how to improve the complexity of the first part of the classical protocol from [8], which (as mentioned above) converts the input distributions into distributions of sufficiently small l_2-norm, using quantum communication. We leave this question as an open problem.

2   Preliminaries

2.1   Definitions and Notations

A typical communication task for Alice and Bob is to compute (sometimes only with some probability of success) a function f on some inputs x, y where x is given to Alice and y to Bob. A communication protocol is an algorithmic description of message sending between Alice and Bob that solves the task for any possible pair of inputs. The communication complexity [14] of such a function is the minimum required numbers of bits the most efficient protocol solving the task must exchange in the worst case (regarding inputs).

The quantum communication complexity [15, 16] of a function is the equivalent using qubits instead of bits. Qubits correspond to elements of some Hilbert space of dimension 2. We will use the bra-ket notation $|\phi\rangle_R$ to denote a qubit (and by extension an n-qubit string) ϕ of a register R.

As described in the introduction, Alice's input consists of t samples from a discrete distribution $p\colon [n] \to (0,1)$. Bob's input consists of t samples from a discrete distribution $q\colon [n] \to (0,1)$. We call X_i (resp. Y_i) the number of samples of Alice (resp. Bob) corresponding to element i. We call $X, Y \in [t]^n$ the occurrence vectors of Alice and Bob, i.e., the vectors with i-th coordinate X_i and Y_i, respectively.

For a vector $x \in \mathbb{R}^n$, we will denote by $||\cdot||_1$ the l_1-norm, which is defined as $||x||_1 = \sum_i^n |x_i|$, and denote by $||\cdot||_2$ the l_2-norm, which is defined as $||x||_2 = \sqrt{\sum_i^n x_i^2}$. We use $\tilde{O}, \tilde{\Omega}$ instead of O, Ω when we neglect factors of logarithmic order in the parameters of the problem (n, t, ϵ). We denote by $Poi(\lambda)$ the Poisson distribution with parameter $\lambda \in \mathbb{N}$.

2.2   Classical Protocol

We now describe the protocol from [8] for the case of small l_2-norm.

The main idea behind the protocol is based on the estimator from [6], combined with a reduction from l_1-distance estimation to l_2-distance estimation. Considering as well the errors of the estimator, after rescaling and shifting the estimator from [6], Andoni, Malkin and Nosatzki found that it is actually enough to compare some approximation of the term $\Delta = ||X - Y||_2^2$ to some threshold τ to distinguish the two cases. As discussed in the introduction, the original protocol from [8] has a communication step to recast the probability distributions into ones with smaller l_2-norms. Since this step is not necessary for the case of distributions with small l_2-norms, we omit it in the following description.

Algorithm 1 Classical Protocol for Distribution Closeness Testing from [8]

1: Fix $\alpha = \Theta(\frac{t\epsilon^2}{n} + 1)$;
2: Alice and Bob each estimate $||p||_2$ and $||q||_2$ up to a factor 2; if the two estimates are not within a factor 4, output "ϵ-FAR";
3: Alice and Bob approximate $\Delta = ||X-Y||_2^2$ up to a $(1+\alpha)$ factor using standard techniques;
4: If Δ is less than $\tau = \frac{\epsilon^2 t^2}{2n} + 2t$ output "SAME", and otherwise output "ϵ-FAR";

The analysis from [8], in particular Lemma 6, shows the correctness of the protocol. More precisely, the following statement can be obtained for the case of input distributions with low l_2-norms.

Theorem 3 *[8] There exists an absolute constant γ_0 such that the following holds: for any input distributions p and q such that $\min(||p||_2, ||q||_2) \leq \gamma_0 t\epsilon^2/n$, the above protocol correctly distinguishes between the case $p = q$ and the case $||p - q||_1 \geq \epsilon$ with probability at least 2/3.* The communication complexity is dominated by the third step, which requires $\tilde{O}\left(1/\alpha^2\right) = \tilde{O}\left(n^2/(t^2\epsilon^4) + 1\right)$ bits.

3 Quantum Protocol

In this section we describe our quantum protocol for the problem $2\text{PCT}_{n,t,\epsilon}^\gamma$, and prove Theorem 1.

3.1 *Description of the Whole Protocol*

Algorithm 2 Quantum Protocol for Distribution Closeness Testing

1: Fix $\alpha = \Theta(\frac{t\epsilon^2}{n} + 1)$;
2: Alice and Bob each estimate $||p||_2$ and $||q||_2$ up to a factor 2; if the two estimates are not within a factor 4, output "ϵ-FAR";
3: Alice and Bob approximate $\Delta = ||X - Y||_2^2$ up to a $(1 + \alpha)$ factor using the procedure of Section 3.2;
4: If Δ is less than $\tau = \frac{\epsilon^2 t^2}{2n} + 2t$ output "SAME", and otherwise output "ϵ-FAR";

The communication complexity is again dominated by the third step, which requires only $\tilde{O}(1/\alpha) = \tilde{O}\left(n/(t\epsilon^2) + 1\right)$ qubits, as described in the next subsection. The correctness is guaranteed by the analysis of Theorem 3, since the quantum protocol performs identical calculations as the classical protocol. This proves Theorem 1.

3.2 *Montanaro's Approximation*

The classical protocol [8] uses standard techniques, such as the AMS algorithm [9], in order to approximate $\Delta = ||X - Y||_2^2$. The AMS algorithm uses a family $\mathcal{H}$ of $O(n^2)$ hash functions $h_i : [n] \rightarrow \{-1, +1\}$ that are 4-wise independent. Given a list of numbers $\ell = (\ell_1, \ldots, \ell_n)$, which we interpret as a vector $\ell \in \mathbb{R}^n$, the AMS algorithm gives an estimate of $||\ell||_2^2$ by computing random estimates $f(i)$ with the following subroutine many times and taking the median of the results.[b]

---

[b]The idea behind the AMS algorithm is that by developing the square, the "crossed" product terms' influence should vanish, i.e., $\mathbb{E}[h_i(j)\ell_j \cdot h_i(j')\ell_{j'}] = 0$ for $j \neq j'$ because then $\mathbb{E}[h_i(j) \cdot h_i(j')] = 0$, while the terms $h_i(j)\ell_j \cdot h_i(j)\ell_j = \ell_j^2$ will always stay.

Algorithm 3 Alon-Matias-Szegedy algorithmic subroutine

1: Draw a random index i to choose a hash function $h_i \in \mathcal{H}$;
2: Return $f(i) = (\sum_{j=1}^n h_i(j) \cdot \ell_j)^2$;

In the classical setting, this subroutine has to be repeated $\tilde{O}(1/\alpha^2)$ times to get a $(1 + \alpha)$-approximation.

Montanaro showed how to achieve the same approximation quantumly using this subroutine only $\tilde{O}(1/\alpha)$ times (see Theorems 12 and 14 in [10]). The subroutine, however, needs to be called in superposition, i.e., Montanaro's approach requires a quantum oracle O_f that performs the following map:

$$O_f : |i\rangle |y\rangle \to |i\rangle |y + f(i)\rangle.$$

It also requires access to its inverse O_f^{-1}.

In our communication setting, we want to use this approach with $\ell := X - Y$. A difficulty is that the data is split between the two parties: only Alice knows X and only Bob knows Y. We now explain how to overcome this difficulty. For a particular index i, Alice can compute $\sigma_a(i) = \sum_{j=1}^n h_i(j)X_j$, and then transmit it to Bob. Bob can similarly do his own computation $\sigma_b(i) = \sum_{j=1}^n h_i(j)Y_j$, then subtract $\sigma_a(i)$ and square in order to get

$$(\sigma_a(i) - \sigma_b(i))^2 = \left(\sum_{j=1}^n h_i(j) \cdot (X_j - Y_j) \right)^2 = f(i). \tag{2}$$

We describe below our implementation of the oracle O_f, based on these ideas. In the following description, Bob's input is the quantum state $|i\rangle_I |y\rangle_Y$ for some bit-strings i and y, where I and Y are two quantum registers.

Algorithm 4 Protocol for building the oracle O_f

1: Bob maps $|i\rangle_I |y\rangle_Y$ to $|i\rangle_I |y\rangle_Y |\sigma_b(i)\rangle_B$ and sends register I to Alice.
2: Alice creates another register A, computes the value $\sigma_a(i)$ in it, and sends the two registers A and I to Bob. At the end of this step, Bob thus owns

$$|i\rangle_I |y\rangle_Y |\sigma_a(i)\rangle_A |\sigma_b(i)\rangle_B.$$

3: Bob performs the computation of Equation (2) using the arguments from A and B, to get the state

$$|i\rangle_I |y + f(i)\rangle_Y |\sigma_a(i)\rangle_A |\sigma_b(i)\rangle_B.$$

4: Bob erases the contents of register B, and sends registers A and I to Alice.
5: Alice erases the contents of register A and sends back register I to Bob.

Note that it is the linear property of the AMS computation (more precisely, Equation (2)) that allows the approximation to be computed even when the data are shared by several parties. Transmitting register I requires $O(\log n)$ qubits. Transmitting register A requires $O(\log t)$ qubits. The overall communication complexity of the oracle protocol is thus $O(\log n + \log t)$.

An implementation of the inverse O_f^{-1} can be obtained similarly, by subtracting instead of adding the value of $f(i)$ at Step 3. We can thus apply Montanaro's algorithm [10], simply

by replacing each oracle query in Montanaro's algorithm by our distributed implementation of O_f (or O_f^{-1}). This enables us to obtain, with high probability, a $(1 + \alpha)$-approximation of $||\ell||_2^2 = \Delta$ using

$$\tilde{O}((1/\alpha)(\log n + \log t)) = \tilde{O}(1/\alpha)$$

qubits of communication, as claimed.

4   Quantum Lower Bound

In this section we prove Theorem 2.

4.1   Hamming Reduction

For bit-strings $x, y \in \{0, 1\}^n$, we denote by $x \cap y$ the set of indices where x and y both have a one, and write $|x \cap y|$ for its size. Note that for bit-strings the l_1-norm corresponds to Hamming weight, i.e., the number of ones in the string, and the l_1-distance corresponds to the Hamming distance.

In the classical communication setting, Andoni, Malkin and Nosatzki [8] proved a lower bound for some closeness testing problem by considering the following problem involving the Hamming distance of two binary strings.

> Let $n \geq 1$ be a multiple of 4. Let $\beta = \beta(n) = \sqrt{n/32}$, and $\kappa > 1$. With probability at least 0.9, for $x, y \in \{0, 1\}^n$ with $||x||_1 = ||y||_1 = n/2$, distinguish between the case where $||x - y||_1 = n/2$ versus $||x - y||_1 - n/2 \in [\beta, \kappa\beta]$.

Notice that if $||x||_1 = ||y||_1 = n/2$ then the equality

$$||x - y||_1 = 2(n/2 - |x \cap y|) = n - 2|x \cap y|$$

holds. The above problem then can be reformulated as the following communication problem that we call PromisedGHD(n, κ):

> PromisedGHD(n, κ)
> For $x, y \in \{0, 1\}^n$, where n is a multiple of 4, with $||x||_1 = ||y||_1 = n/2$, $\beta = \beta(n) = \sqrt{n/32}$, and $\kappa > 1$, with probability at least 0.9 distinguish between $|x \cap y| = \frac{n}{4}$ and $|x \cap y| \in [\frac{n}{4} - \frac{\kappa\beta}{2}, \frac{n}{4} - \frac{\beta}{2}]$.

Particularly, the reduction of [8] only concerns the regime where $\kappa = O(\sqrt{\log n})$.

The main result of this subsection is:

Theorem 4 *The quantum communication complexity of PromisedGHD(n, κ) is $\Omega(\sqrt{n})$.*

In order to prove Theorem 4, we will consider a similar problem defined on smaller inputs. More precisely, define the following problem SmallPGHD(n', g):

> SmallPGHD(n', g)
> For $x', y' \in \{0, 1\}^{n'}$ where n' is a multiple of 4, with $||x'||_1 = ||y'||_1 = n'/2$, with probability at least 2/3 distinguish between $|x' \cap y'| = n'/4$ and $|x' \cap y'| \in [n'/4 - g, n'/4 - 1]$

We first show a reduction from SmallPGHD(n', g) to PromisedGHD(n, κ) for appropriate parameters.

Lemma 1 *For any $g \leq \frac{\kappa}{2\sqrt{32}}$, SmallPGHD($n', g$) reduces to PromisedGHD(n'^2, κ).*

Proof. Assume x', y' are inputs of size n' to the SmallPGHD(n', g) problem. By repeating them n' times in a padding fashion, we build inputs x, y for the PromisedGHD(n, κ) problem where $n = n'^2$ and therefore $\beta = \frac{n'}{\sqrt{32}}$. If $|x' \cap y'| = n'/4$, we have $|x \cap y| = n'^2/4 = n/4$. Otherwise, if $|x' \cap y'| \in [\frac{n'}{4} - g, \frac{n'}{4} - 1]$, we have $|x \cap y| \in [\frac{n}{4} - \frac{\kappa\beta}{2}, \frac{n}{4} - \frac{\beta}{2}]$ as long as $g \le \frac{\kappa}{2\sqrt{32}}$. Indeed, for the upper part of the interval:

$$\frac{n}{4} - \frac{\beta}{2} = \frac{n'^2}{4} - \frac{n'}{2\sqrt{32}} \ge n'(\frac{n'}{4} - 1)$$

while for the lower part, we need:

$$\frac{n}{4} - \frac{\kappa\beta}{2} = \frac{n'^2}{4} - \frac{\kappa n'}{2\sqrt{32}} \le n'(\frac{n'}{4} - g),$$

which is equivalent to the bound on g mentioned just above. $\square$

The hardness of SmallPGHD(n', g) will follow from the hardness of a more restricted problem stated in Theorem 5. The latter theorem is the main technical contribution of this section, and we devote Section 4.2 to its proof.

4.2   Main Technical Result

We will model partial functions as mappings $f\colon X \to \mathbb{R} \cup \{*\}$, where X is a finite set and the range element $*$ represents undefined output. We let dom $f = \{x : f(x) \ne *\}$. It will be convenient here to represent the Boolean values "true" and "false" by -1 and $+1$, respectively. This departure from the classical representation (of using 0 and 1) has no effect on quantum communication complexity. For a communication problem $F\colon X \times Y \to \{-1, +1, *\}$, we let $Q_\epsilon^*(F)$ denote the ϵ-error quantum communication complexity of F with arbitrary prior entanglement. Note that an ϵ-error protocol for F is allowed to behave arbitrarily on inputs outside dom F.

We are now ready to prove the main technical result used in the proof of Theorem 2.

Theorem 5 *Let n be an integer divisible by* 4. *Consider the partial communication problem $F_n\colon \{0, 1\}^n \times \{0, 1\}^n \to \{-1, +1, *\}$ given by*

$$F_n(x, y) = \begin{cases} -1 & \text{if } |x| = |y| = n/2 \text{ and } |x \cap y| = n/4, \\ +1 & \text{if } |x| = |y| = n/2 \text{ and } |x \cap y| = n/4 - 1, \\ * & \text{otherwise.} \end{cases}$$

Then $Q_{1/3}^*(F_n) = \Omega(n)$.

The remaining part of this section is devoted to the proof of this theorem. We start by reviewing relevant background on the pattern matrix method [11] for quantum communication lower bounds. Let k and n be positive integers, where $k < n$ and $k \mid n$. Partition $[n]$ into k contiguous blocks, each with n/k elements:

$$[n] = \left\{1, 2, \ldots, \frac{n}{k}\right\} \cup \left\{\frac{n}{k} + 1, \ldots, \frac{2n}{k}\right\} \cup \cdots \cup \left\{\frac{(k-1)n}{k} + 1, \ldots, n\right\}.$$

Let $\mathcal{V}(n, k)$ denote the family of subsets $V \subseteq [n]$ that have exactly one element in each of these blocks (in particular, $|V| = k$). Clearly, $|\mathcal{V}(n, k)| = (n/k)^k$. For a bit string $x \in \{0, 1\}^n$

and a set $V \in \mathcal{V}(n, k)$, define the *projection of x onto V* by $x|_V = (x_{i_1}, x_{i_2}, \ldots, x_{i_k}) \in \{0, 1\}^k$, where $i_1 < i_2 < \cdots < i_k$ are the elements of V. For $\phi \colon \{0, 1\}^k \to \mathbb{R} \cup \{*\}$, the (n, k, ϕ)-*pattern matrix* is the matrix A given by

$$A = \left[\phi(x|_V \oplus w) \right]_{x \in \{0,1\}^n, \, (V,w) \in \mathcal{V}(n,k) \times \{0,1\}^k}.$$

In words, A is the matrix of size 2^n by $(n/k)^k 2^k$ whose rows are indexed by strings $x \in \{0, 1\}^n$, whose columns are indexed by pairs $(V, w) \in \mathcal{V}(n, k) \times \{0, 1\}^k$, and whose entries are given by $A_{x,(V,w)} = \phi(x|_V \oplus w)$.

The pattern matrix method gives a lower bound on the quantum communication complexity of a pattern matrix in terms of the *approximate degree* of its generating function. We now define this notion formally. Let $f \colon X \to \mathbb{R}$ be given, for a finite subset $X \subset \mathbb{R}^n$. The ϵ-*approximate degree* of f, denoted $\deg_\epsilon(f)$, is the least degree of a real polynomial π such that $|f(x) - \pi(x)| \leq \epsilon$ for all $x \in X$. One generalizes this definition to partial functions $f \colon X \to \mathbb{R} \cup \{*\}$ by letting $\deg_\epsilon(f)$ be the least degree of a real polynomial π with

$$
\begin{aligned}
|f(x) - \pi(x)| &\leq \epsilon, & x &\in \operatorname{dom} f, \\
|\pi(x)| &\leq 1 + \epsilon, & x &\in X \setminus \operatorname{dom} f.
\end{aligned}
$$

We will need the following version of the pattern matrix method for quantum lower bounds.

Theorem 6 *Let F be the (n, k, f)-pattern matrix, where $f \colon \{0, 1\}^k \to \{-1, +1, *\}$ is given. Then for every $\epsilon \in [0, 1)$ and every $\delta < \epsilon/2$,*

$$Q^*_\delta(F) \geq \frac{1}{4} \deg_\epsilon(f) \log\left(\frac{n}{k}\right) - \frac{1}{2} \log\left(\frac{3}{\epsilon - 2\delta}\right).$$

Theorem 6 is a generalization of the original pattern matrix method of [11] to partial functions. For the reader's convenience, we give a detailed proof of Theorem 6 in the appendix.

Proof of Theorem 5. The communication complexity of F_n is monotone in n, due to $F_n(x, y) = F_{n+4}(x0011, y0101)$. As a result, it suffices to prove the theorem for n divisible by 3. Under this divisibility assumption, define $k = n/6$ and consider the function $\mathrm{PMAJ}_k \colon \{0, 1\}^k \to \{-1, +1, *\}$ given by

$$\mathrm{PMAJ}_k(x) = \begin{cases} -1 & \text{if } |x| = k/2, \\ +1 & \text{if } |x| = k/2 - 1, \\ * & \text{otherwise.} \end{cases}$$

Let P be the $(2k, k, \mathrm{PMAJ}_k)$-pattern matrix. It is a well-known fact [17, 18] that $\deg_{1/3}(\mathrm{PMAJ}_k) = \Omega(k)$. As a result, Theorem 6 implies that $Q^*_{1/7}(P) = \Omega(k)$ and hence also $Q^*_{1/3}(P) = \Omega(k)$.

Writing $P = [\mathrm{PMAJ}_k((x_1 \overline{x}_1 x_2 \overline{x}_2 \ldots x_{2k} \overline{x}_{2k})|_V)]_{x \in \{0,1\}^{2k}, V \in \mathcal{V}(4k,k)}$ makes it clear that P is a restriction of the more general communication problem $G \colon \{0, 1\}^{4k} \times \{0, 1\}^{4k} \to \{-1, +1, *\}$ defined by

$$G(x, y) = \begin{cases} -1 & \text{if } |x| = 2k, \ |y| = k, \text{ and } |x \cap y| = k/2, \\ +1 & \text{if } |x| = 2k, \ |y| = k, \text{ and } |x \cap y| = k/2 - 1, \\ * & \text{otherwise.} \end{cases}$$

As a result, $Q^*_{1/3}(G) \geq Q^*_{1/3}(P) = \Omega(k)$. This in turn implies that $Q^*_{1/3}(F_n) = \Omega(k)$ because $G(x, y) = F_n(x1^k 0^k, y1^{2k})$. $\square$

4.3   Closeness Testing Reduction

In this subsection we explain how the lower bound on the quantum communication complexity of PromisedGHD(n, κ) (Theorem 4) implies Theorem 2.

Proof of Theorem 2. In [8], Andoni, Malkin and Nosatzki show a reduction from the problem PromisedGHD(m, κ) to the problem $2\text{PCT}_{n,t,1/2}$ with parameters $m = \frac{n^2}{t^2 \log^3 n}$ and $\kappa = O(\sqrt{\log n})$.[c] Theorem 4 thus gives us the claimed lower bound. In the remaining of the proof, we show that the distributions used to prove the lower bound have low l_2-norm.

The input distributions of $2\text{PCT}_{n,t,1/2}$ used in the reduction shown in [8], which we will denote a and b, are of the following form: half of the mass is uniformly distributed on $d = n/10$ elements, and the other half of the mass on $l = C_0 \cdot t \cdot \log n$ other elements, where C_0 is some constant. Therefore:

$$
\begin{aligned}
||a||_2 = ||b||_2 &= \sqrt{d(\frac{1}{2d})^2 + l(\frac{1}{2l})^2} \\
&= \frac{1}{2}\sqrt{\frac{1}{d} + \frac{1}{l}} \\
&= \frac{1}{2}\sqrt{\frac{10}{n} + \frac{1}{C_0 t \log n}} \\
&\leq \frac{1}{2}\sqrt{(10 + \frac{1}{C_0})\frac{1}{t \log n}},
\end{aligned}
$$

since $\frac{1}{n} \leq \frac{1}{t \log n}$ because $t \leq \frac{n}{\log^c n}$.

Define γ_{LW} as the smallest γ such that $\min(||a||_2, ||b||_2) \leq \gamma t \epsilon^2/n$ holds. The above calculations show that

$$
\gamma_{LW} = \frac{||a||_2 n}{t \epsilon^2} \leq \frac{1}{2}\sqrt{10 + \frac{1}{C_0}}\frac{n}{t \epsilon^2 \sqrt{t \log n}} \leq \frac{1}{2C^{3/2}}\sqrt{10 + \frac{1}{C_0}}\frac{1}{\sqrt{\log n}}
$$

because $t \geq Cn^{2/3} \cdot \epsilon^{-4/3}$ by (1). Thus $\gamma_{LW} = O(1/\sqrt{\log n})$. This concludes the proof. $\square$

References

1. Oded Goldreich. *Introduction to Property Testing.* Cambridge University Press, 2017.

---

[c] The reduction is actually stated, in Theorem 9 in [8], as a reduction from PromisedGHD(m, κ) to the variant of $2\text{PCT}_{n,t,1/2}$ where the number of samples is $\text{Poi}(t)$ instead of exactly t. Nevertheless the Poisson version easily reduces to the original problem with $10t$ number of samples if we allow some extra error, since by Chebyshev's inequality the probability that the Poisson version gives more than $10t$ samples is less than $1/81$.

2. Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, volume 6650 of *Lecture Notes in Computer Science*, pages 68–75. Springer, 2011.

3. Clément L. Canonne. A survey on distribution testing: Your data is big. but is it blue? *Electronic Colloquium on Computational Complexity (ECCC)*, 22:63, 2015.

4. Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS 2000)*, pages 259–269, 2000.

5. Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 60(1):4:1–4:25, 2013.

6. Siu-on Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2014)*, pages 1193–1203, 2014.

7. Ilias Diakonikolas and Daniel M. Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS 2016)*, pages 685–694, 2016.

8. Alexandr Andoni, Tal Malkin, and Negev Shekel Nosatzki. Two party distribution testing: Communication and security. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *LIPIcs*, pages 15:1–15:16, 2019.

9. Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

10. Ashley Montanaro. The quantum complexity of approximating the frequency moments. *Quantum Information & Computation*, 16(13&14):1169–1190, 2016.

11. Alexander A. Sherstov. The pattern matrix method. *SIAM Journal on Computing*, 40(6):1969–2000, 2011.

12. Srinivasan Arunachalam and Ronald de Wolf. Guest column: A survey of quantum learning theory. *SIGACT News*, 48(2):41–67, 2017.

13. Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. *Theory of Computing, Graduate Surveys*, 7:1–81, 2016.

14. Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

15. Ronald de Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 287(1):337–353, 2002.

16. Gilles Brassard. Quantum communication complexity. *Foundations of Physics*, 33:1593–1616, 2003.

17. Ramamohan Paturi. On the degree of polynomials that approximate symmetric Boolean functions. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing (STOC 1992)*, pages 468–474, 1992.

18. Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov–Bernstein inequalities. *Information and Computation*, 243:2–25, 2015.

19. Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012.

Appendix A: The Pattern Matrix Method for Partial Functions

The purpose of this appendix is to provide a detailed proof of Theorem 6 for partial functions. Our proof closely follows the original proof of the pattern matrix method in [11], developed there for total functions.

We start by recalling the Fourier transform for functions $f\colon \{0,1\}^n \to \mathbb{R}$. For $S \subseteq \{1,2,\ldots,n\}$, define $\chi_S\colon \{0,1\}^n \to \{-1,+1\}$ by $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$. Then every function

$f\colon \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form

$$f = \sum_{S \subseteq \{1,2,\ldots,n\}} \hat{f}(S)\,\chi_S,$$

where $\hat{f}(S) = 2^{-n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x)$. The reals $\hat{f}(S)$ are called the *Fourier coefficients of f*.

For a real matrix A, we let $\|A\|_1$ denote the sum of the absolute values of the entries of A. We let $\|A\|$ denote the spectral norm of A. Recall that $\|A\| = \max_{x : \|x\|_2 = 1} \|Ax\|_2$. The following theorem [11, Theorem 4.3] determines the spectral norm of a pattern matrix in terms of the Fourier spectrum of its generating function.

Theorem A.1 *Let $\phi\colon \{0,1\}^k \to \mathbb{R}$ be given. Let A be the (n,k,ϕ)-pattern matrix. Then*

$$\|A\| = \sqrt{2^{n+k}\left(\frac{n}{k}\right)^k} \max_{S \subseteq [k]}\left\{ |\hat{\phi}(S)|\left(\frac{k}{n}\right)^{|S|/2}\right\}.$$

We will also need the following dual characterization of approximate degree of partial functions, analogous to the dual characterization for total functions used in [11].

Theorem A.2 *Let $f\colon \{0,1\}^n \to \mathbb{R} \cup \{*\}$ be a given function, $d \geq 0$ an integer. Then $\deg_\epsilon(f) > d$ if and only if there exists $\psi\colon \{0,1\}^n \to \mathbb{R}$ such that*

$$\sum_{x \in \mathrm{dom}\, f} f(x)\psi(x) - \sum_{x \notin \mathrm{dom}\, f} |\psi(x)| - \epsilon\|\psi\|_1 > 0,$$

and $\hat{\psi}(S) = 0$ for $|S| \leq d$.

Theorem A.2 follows from linear programming duality; see [11, 19] for details.

Next, we derive a version of the *generalized discrepancy method* for partial functions, by adapting the analogous proof in [11, Theorem 2.8] for total functions.

Theorem A.3 *Let X, Y be finite sets and $F\colon X \times Y \to \{-1, +1, *\}$ a given function. Let $\Psi = [\Psi_{xy}]_{x \in X, y \in Y}$ be any real matrix with $\|\Psi\|_1 = 1$. Then for each $\epsilon > 0$,*

$$4^{Q_\epsilon^*(F)} \geq \frac{1}{3\|\Psi\|\sqrt{|X|\,|Y|}}\left(\sum_{(x,y) \in \mathrm{dom}\, F} \Psi_{x,y}F(x,y) - \sum_{(x,y) \notin \mathrm{dom}\, F} |\Psi_{x,y}| - 2\epsilon \right).$$

Proof. Let P be a quantum protocol with prior entanglement that computes F with error ϵ and cost C. Let Π be the matrix of acceptance probabilities of P, so that $\Pi_{x,y}$ is the probability that P accepts the input (x,y). It is shown in the proof of [11, Theorem 2.8] that

$$\sum_{x \in X} \sum_{y \in Y} \Psi_{x,y}(1 - 2\Pi_{x,y}) \leq \|\Psi\|\left(2 \cdot 4^C + 1\right)\sqrt{|X|\,|Y|}. \tag{A.1}$$

Now observe that $1 - 2\Pi_{x,y}$ ranges in $[F(x,y) - 2\epsilon, F(x,y) + 2\epsilon]$ on $\mathrm{dom}\, F$, and is bounded

in absolute value by 1 otherwise. This gives

$$\sum_{x \in X} \sum_{y \in Y} \Psi_{x,y}(1 - 2\Pi_{x,y})$$

$$\geq \sum_{(x,y) \in \operatorname{dom} F} (\Psi_{x,y} F(x,y) - 2\epsilon|\Psi_{x,y}|) - \sum_{(x,y) \notin \operatorname{dom} F} |\Psi_{x,y}|$$

$$\geq \sum_{(x,y) \in \operatorname{dom} F} \Psi_{x,y} F(x,y) - 2\epsilon - \sum_{(x,y) \notin \operatorname{dom} F} |\Psi_{x,y}|, \tag{A.2}$$

where the last step uses $\|\Psi\|_1 \leq 1$. The theorem follows by comparing the upper bound (A.1) with the lower bound (A.2). $\square$

We are now in a position to prove Theorem 6, which we restate here for the reader's convenience.

Theorem A.4 (restatement of Theorem 6) *Let F be the (n, k, f)-pattern matrix, where $f\colon \{0,1\}^k \to \{-1, +1, *\}$ is given. Then for every $\epsilon \in [0, 1)$ and every $\delta < \epsilon/2$,*

$$Q_\delta^*(F) \geq \frac{1}{4} \deg_\epsilon(f) \log\left(\frac{n}{k}\right) - \frac{1}{2} \log\left(\frac{3}{\epsilon - 2\delta}\right). \tag{A.3}$$

Proof. Let $d = \deg_\epsilon(f) \geq 1$. By Theorem A.2, there is a function $\psi\colon \{0,1\}^k \to \mathbb{R}$ such that:

$$\hat{\psi}(S) = 0 \qquad\qquad (|S| < d), \tag{A.4}$$

$$\sum_{z \in \{0,1\}^k} |\psi(z)| = 1, \tag{A.5}$$

$$\sum_{z \in \operatorname{dom} f} f(z)\psi(z) - \sum_{z \notin \operatorname{dom} f} |\psi(z)| > \epsilon. \tag{A.6}$$

Let Ψ be the $(n, k, 2^{-n}(n/k)^{-k}\psi)$-pattern matrix. Then (A.5) and (A.6) show that

$$\|\Psi\|_1 = 1, \tag{A.7}$$

$$\sum_{(x,y) \in \operatorname{dom} F} F_{x,y}\Psi_{x,y} - \sum_{(x,y) \notin \operatorname{dom} F} |\Psi_{x,y}| > \epsilon. \tag{A.8}$$

Our last task is to calculate $\|\Psi\|$. It follows from (A.5) that

$$\max_{S \subseteq [k]} |\hat{\psi}(S)| \leq 2^{-k}. \tag{A.9}$$

Theorem A.1 yields, in view of (A.4) and (A.9):

$$\|\Psi\| \leq \left(\frac{k}{n}\right)^{d/2} \left(2^{n+k} \left(\frac{n}{k}\right)^k\right)^{-1/2}. \tag{A.10}$$

Now (A.3) follows from (A.7), (A.8), (A.10), and Theorem A.3. $\square$