Classical Simulation of One-Query Quantum Distinguishers

Andrej Bogdanov
School of EECS, University of Ottawa, Canada

Tsun Ming Cheung
School of Computer Science, McGill University, Montreal, Canada

Krishnamoorthy Dinesh
Dept. of Computer Science and Engineering, Indian Institute of Technology, Palakkad, India

John C. S. Lui
Dept. of Computer Science and Engineering, Chinese University of Hong Kong, China

Abstract

We study the relative advantage of classical and quantum distinguishers of bounded query complexity over n-bit strings, focusing on the case of a single quantum query. A construction of Aaronson and Ambainis (STOC 2015) yields a pair of distributions that is ε-distinguishable by a one-query quantum algorithm, but $O(\varepsilon k/\sqrt{n})$-indistinguishable by any non-adaptive k-query classical algorithm.

We show that every pair of distributions that is ε-distinguishable by a one-query quantum algorithm is distinguishable with k classical queries and (1) advantage $\min\{\Omega(\varepsilon\sqrt{k/n})), \Omega(\varepsilon^2 k^2/n)\}$ non-adaptively (i.e., in one round), and (2) advantage $\Omega(\varepsilon^2 k/\sqrt{n \log n})$ in two rounds.

As part of our analysis, we introduce a general method for converting unbiased estimators into distinguishers.

2012 ACM Subject Classification Theory of computation → Probabilistic computation

Keywords and phrases Query complexity, quantum algorithms, hypothesis testing, Grothendieck's inequality

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.43

Category RANDOM

Funding This work is supported by RGC GRF grant CUHK 14207721 and NSERC Discovery grant RGPIN-2023-05006.

1 Introduction

A distinguisher is an algorithm for hypothesis testing. Its purpose is to tell whether its input was sampled from one distribution or from another. In algorithmic contexts including much of cryptography, pseudorandomness, and statistical inference, the computational complexity of distinguishers plays a crucial role.

In this work, we initiate the study of the classical simulation of quantum distinguishers. Quantum algorithms promise algorithmic speedups, but the realization of fully capable quantum computers is still a distant goal. It is thus important to investigate the capabilities of quantum devices of limited computational power. Our focus here is on devices of bounded *query complexity*, a fundamental efficiency measure in complexity theory and cryptographic analysis.

We are interested in the best possible advantage of simulating a quantum distinguisher of bounded query complexity by a classical distinguisher of bounded but possibly larger query complexity k. We focus on quantum distinguishers that make a single query to an n-bit Boolean-valued oracle. Although this model appears restrictive, we find it interesting for the following reasons.

First, Aaronson and Ambainis [1] showed that in the *constant* advantage regime, one-query quantum distinguishers already require $\Omega(\sqrt{n})$ non-adaptive classical queries to simulate with the same advantage. Subsequent works [4, 11] showed a rapid deterioration as more queries are added: In general, q classical queries to a Boolean oracle require $\Omega_q(n^{1-1/2q})$ classical queries to simulate. Beyond one or a handful of quantum queries, the improvement over brute-force classical simulation becomes marginal. Moreover, addressing the case of one quantum query already brings up interesting technical challenges and reveals connections to statistical estimation and random matrix theory.

Second, a one-query quantum algorithm can be viewed as a sensible model of a noise-prone quantum device. Motivated by the challenges quantum computers pose to cryptography, it is of interest to study the power of such devices as cryptographic adversaries. In this context, the best classical simulation of a quantum adversary tells us to what extent our confidence in cryptographic security of existing constructions carries over to the quantum setting. While quantum security analyses have been successfully carried out for specific constructions, e.g., [15, 14, 8], our work provides general black-box "transfer theorems" that yield quantum security directly from sufficiently strong classical security at a bounded cost in parameters.

Our results

Our starting point is the separation between quantum and classical query complexity of Aaronson and Ambainis [1]. In response to a question of Buhrman et al. [6], they constructed a random variable F (for "Forrelation") over $\{\pm 1\}^{2n}$ (where n is a power of two) for which

- There exists a one query quantum algorithm that distinguishes F from a uniformly random input in $\{\pm 1\}^{2n}$ with constant advantage.
- Every classical algorithm that makes $o(\sqrt{n})$ queries fails to distinguish F from random with constant advantage.

Moreover, their example is tight [5, 2]: Every one-query quantum distinguisher with constant advantage can be simulated by a $O(\sqrt{n})$-query non-adaptive classical distinguisher with constant advantage.

Here, as in the rest of the paper, a *distinguisher* is an algorithm that produces outputs in the range $[-1, 1]$. The output of a quantum algorithm is taken to be its probability that it collapses to an accepting state. The *advantage* of D on the pair of distributions $(\mathbf{A}, \mathbf{B})$ is $\frac{1}{2}|\mathsf{E}[D(\mathbf{A})] - \mathsf{E}[D(\mathbf{B})]|$.

We are interested in the best possible advantage that a classical distinguisher with $k \ll \sqrt{n}$ queries can attain for a pair of distributions that are ε-distinguishable by a one-query quantum algorithm. The example of Aaronson and Ambainis yields the following generalization:

▶ **Proposition 1.** *For every $\varepsilon \in (0, 1)$, $k \in \mathbb{N}$, and $n = 2^m$ for some $m \in \mathbb{N}$, there exists a $\{\pm 1\}^{2n}$-valued random variable F_ε that is ε-distinguishable from a uniform random $2n$-bit string by 1 quantum query, but $O(\varepsilon k/\sqrt{n})$-indistinguishable by any non-adaptive $2k$-query classical algorithm.*

The random variable F_ε is a mixture of Forrelation F and a uniform random variable U. We believe that the bound $O(\varepsilon k/\sqrt{n})$ is the best possible gap in the advantage of k-query non-adaptive classical versus one-query quantum distinguisher. Our first result is the following lower bound on the classical advantage:

▶ **Theorem 2.** *Let $\varepsilon \in (0,1)$, $k, n \in \mathbb{N}$. Suppose $(\mathbf{A}, \mathbf{B})$ is a pair of random variables over $\{\pm 1\}^n$ that is ε-distinguishable by a one-query quantum algorithm. There exist $2k$-query non-adaptive classical algorithms P_{2a} and P_{2b}, such that*

a. *P_{2a} distinguishes $(\mathbf{A}, \mathbf{B})$ with advantage $\Omega(\varepsilon\sqrt{k/n})$,*

b. *P_{2b} distinguishes $(\mathbf{A}, \mathbf{B})$ with advantage $\Omega(\varepsilon^2 k^2/n)$, assuming $k = O(\sqrt{n})$.*

Assuming ε is constant, distinguisher P_{2a} works better when k is small but does not reach constant advantage as k approaches $\sqrt{n}$. In contrast, distinguisher P_{2b} has a constant advantage when $k = \Theta(\sqrt{n})$; but for the case of $k = 1$, the advantage is worse than the upper bound $O(\varepsilon/\sqrt{n})$ as given in Proposition 1 by a factor of $1/\sqrt{n}$.

We also show that for constant ε, an advantage of $\Omega(k/\sqrt{n \log n})$ can be achieved with two rounds of queries.

▶ **Theorem 3.** *For every $\varepsilon \in (0,1), n \in \mathbb{N}$, and $k \leq \sqrt{n \log n}/\varepsilon$, every pair of random variables over $\{\pm 1\}^n$ that is ε-distinguishable by a one-query quantum algorithm is also $\Omega(\varepsilon^2 k/\sqrt{n \log n})$-distinguishable by a $2k$-query two-round adaptive classical algorithm.*

Bansal and Sinha [4] showed that no k-query adaptive classical algorithm can distinguish F from random with advantage better than $O(\varepsilon k^{1/2}(\log n)^{1/4}/n^{1/4})$. Sherstov, Storozhenko, and Wu [11] proved the same bound with different distributions.

We prove that their bound can be improved to $\tilde{O}(\varepsilon k/\sqrt{n})$ for two-round algorithms, thereby showing that the simulation in Theorem 3 is optimal in k and n up to log factors.

▶ **Theorem 4.** *For every $\varepsilon > 0$, $k \in \mathbb{N}$, and $n = 2^m$ for some $m \in \mathbb{N}$, there exists a random variable F_ε on $\{\pm 1\}^n$, such that it is $O(\varepsilon k\sqrt{\log n}/\sqrt{n})$-indistinguishable from random by any two-round classical algorithm that makes k queries per round.*

Up to the factor of $\sqrt{\log n}$, Theorem 4 generalizes Proposition 1 to adaptive two-round algorithms. The results are summarized in Table 1.

Table 1 Bounds on the best possible advantage of a k-query classical simulation of a one-query quantum distinguisher with advantage ε for distributions over the n-dimensional Boolean cube.

Type	Upper bound	Ref.	Lower bound	Ref.
Non-adaptive	$O(\varepsilon k/\sqrt{n})$	Proposition 1	$\Omega(\varepsilon\sqrt{k/n})$	Theorem 2a
			$\Omega(\varepsilon^2 k^2/n)$	Theorem 2b
Two-round	$O(\varepsilon k\sqrt{\log n}/\sqrt{n})$	Theorem 4	$\Omega(\varepsilon^2 k/\sqrt{n \log n})$	Theorem 3
Adaptive	$O(\varepsilon k^{1/2}(\log n)^{1/4}/n^{1/4})$	[4, 11]		

While it is worth mentioning that the lower bound on advantage in Theorem 3 never exceeds the upper bound from Proposition 1, it remains open whether adaptivity helps in classical simulations of one-query quantum distinguishers.

Our techniques

The acceptance probability of a quantum algorithm that makes one query to an n-bit oracle can be represented by a *bounded* $(n+1) \times (n+1)$ *bilinear form*, that is a function of the form $p(x,y) = \sum A_{ij} x_i y_j$ for some matrix $A \in \mathbb{R}^{(n+1)\times(n+1)}$ with bounded ∞-to-1 norm (see Proposition 6). It suffices to prove our results under the assumption that the two distributions are distinguishable by such bilinear forms. Aaronson et al. [3] showed that

this representation fully characterizes one-query quantum algorithms: every bilinear form of bounded ∞-to-1 norm represents the acceptance probability of some one-query quantum algorithm up to constant scaling.

The general problem of identifying the optimal distinguisher in a class of algorithm $\mathcal{A}$ against a class of distribution pairs $\mathcal{B}$ can be modeled as a zero-sum game between distinguishers in $\mathcal{A}$ and distribution pairs in $\mathcal{B}$ whose payoff is the advantage. In this setting, we take $\mathcal{A}$ to be the classical algorithms that make k queries to x and k queries to y, and $\mathcal{B}$ to be the distribution pairs ε-distinguishable by some one-query quantum algorithm.

By Yao's minimax theorem, a given distinguishing advantage is achievable against any given pair in $\mathcal{B}$ if and only if there exists a mixture of distinguishers that has the same expected advantage against all pairs in $\mathcal{B}$. Hence it is sufficient (and necessary) to construct a probabilistic distinguisher that is oblivious to the actual distributions. Such a distinguisher can be obtained from an unbiased estimator for some multiple of p: if $\mathsf{E}[D(x,y)] = \frac{1}{Z}p(x,y)$ for all inputs x, y then the distinguishing advantage of D is at most Z times smaller than that of p.

In the proof of Theorem 2a, we construct an unbiased estimator P_{2a} for p/Z with $Z = O(\sqrt{n/k})$ that is a mixture of $2k$-juntas. Each junta is a homogeneous quadratic function on k bits of x-input and k bits of y-input.

The approximation factor Z is derived based on an additional assumption of boundedness of the juntas, which we explain in detail in Proposition 6. The proposition states that a one-query quantum algorithm is fully characterized by a bilinear form with *∞-to-1 norm* bounded by 1 (we say this bilinear form is 1-bounded). It can be shown that Z is the best possible within this class of unbiased estimators:

▶ **Proposition 5.** *There exists a 1-bounded $n \times n$ bilinear form p such that if p/Z is represented as a mixture of 1-bounded $k \times k$ bilinear forms, then $Z = \Omega(\sqrt{n/k})$.*

In Theorem 2b, we bypass the limitation by truncating a different *unbounded* unbiased estimator P_{2b}. Corollary 13 lower bounds the advantage of the distinguisher obtained by truncating a scaled unbiased estimator in terms of its variance. The relevant estimator is obtained from independently sampled indices $i_1, \ldots, i_k, j_1, \ldots, j_k$ of x and y input coordinates, respectively, where each coordinate is chosen with probability weighted by Grothendieck's factorization (see Section 2 for the definition). Proposition 16, which is also implicit in the more general analysis of [5], shows that this estimator has variance $O(n/k^2)$.

One weakness of estimator P_{2b} is that it samples the bits of the inputs x and y independently and fails to detect relevant correlations between them. In contrast, the estimator P_3 in Theorem 3 computes the distribution on y-queries adaptively depending on the answers to the x-queries. Viewing the bilinear form $p(x,y)$ as a linear function of the y-inputs, the sample of x-inputs is used to estimate the coefficient $\sum_i A_{ij}x_i$ of y_j for every j. In the second round, the y-inputs are sampled with probabilities proportional to these estimates. Using Grothendieck's factorization and exponential tail bounds, in Proposition 17 we show that this improves the effective variance of P_{2b} by a factor of $\tilde{O}(\sqrt{k})$.

2 Quantum algorithms, bilinear forms, and norms

Notations

We write $[n]$ for the set $\{1, \ldots, n\}$. We use the standard computer science asymptotic notations, and the tilde notations hide logarithmic factors. We write $\mathcal{N}(\mu, \Sigma)$ for a multivariate Gaussian with mean μ and covariance matrix Σ, sd for statistical distance, and kl for KL-divergence.

For $A \in \mathbb{R}^{m\times n}$, we use $A^{i,:}$ to denote the i-th row of A and $A^{:,j}$ for the j-th column of A. We write e_i the i-th standard basis vector, and Id_k the $k \times k$ identity matrix. For a symmetric matrix A, we denote the minimum and maximum eigenvalues (which are guaranteed to be real) by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ respectively.

Norms

For a vector, we denote $\|v\|_p$ the p-norm of v. The 1-norm, 2-norm and ∞-norm will be relevant in this work. We drop the subscript for Euclidean norm (2-norm) of a vector. The Cauchy-Schwarz inequality says that $\sum_i u_i v_i \leq \|u\|\|v\|$ and in particular $\|v\|_1 \leq \sqrt{n}\|v\|$ for $v \in \mathbb{R}^n$.

For $A \in \mathbb{R}^{m\times n}$, the *spectral norm* $\|A\|$, *Frobenius norm* $\|A\|_F$, and ∞-*to*-1-*norm* $\|A\|_{\infty\to 1}$ are defined to be

$$\|A\| := \max_{u\in\mathbb{R}^n:\|u\|=1} \|Au\| = \max_{\substack{u\in\mathbb{R}^m, v\in\mathbb{R}^n \\ \|u\|=\|v\|=1}} u^\top A v$$

$$\|A\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}$$

$$\|A\|_{\infty\to 1} := \max_{\substack{x\in\{\pm 1\}^m \\ y\in\{\pm 1\}^n}} x^\top A y = \max_{\substack{x\in[-1,1]^m \\ y\in[-1,1]^n}} x^\top A y = \max_{x\in\mathbb{R}^n:\|x\|_\infty=1} \|Ax\|_1$$

The relevance of the ∞-to-1 norm stems from the following connection to one-query quantum algorithms:

▶ **Proposition 6** ([3]). *For every quantum algorithm Q making one query to some oracle in $\{\pm 1\}^n$, there exists a bilinear form $p(x,y) = \sum_{i,j=1}^{n+1} A_{ij}x_i y_j$, $A_{ij} \in \mathbb{R}$, such that for every $x \in \{\pm 1\}^n$, the probability that Q accepts x equals $p((x_1,\ldots,x_n,1),(x_1,\ldots,x_n,1))$.*

We refer to p as the *advantage polynomial*, and by abuse of notation, we refer $\|p\|_\#$ to be $\|A\|_\#$ for any norm $\|\cdot\|_\#$. Clearly, the matrix defining any advantage polynomial must have ∞-to-1 norm at most 1 (hence every advantage polynomial is *1-bounded*). In general, this does not imply a constant upper bound on the spectral norm. However, the dual form of Grothendieck's inequality, also known as the *factorization Grothendieck's inequality* [9, P.239], shows that such a bound holds up to factorization.

Grothendieck's factorization

▶ **Proposition 7** ([9]). *There is a universal constant K_G such that if $A \in \mathbb{R}^{n\times n}$ satisfies that $\|A\|_{\infty\to 1} \leq 1$, then there exists $\alpha, \beta \in \mathbb{R}^n_{\geq 0}$ with $\|\alpha\| = \|\beta\| = 1$, such that A can be factored as $A_{ij} = \alpha_i \tilde{A}_{ij} \beta_j$ with $\|\tilde{A}\| \leq K_G$.*

If the matrix A has no all-zero row or all-zero column, then we can further assume that α and β are strictly positive, which is an assumption we can make for advantage polynomials.

If p is the advantage polynomial of a one-query quantum algorithm, the stronger conclusion $\|\tilde{A}\| \leq 1$ can be obtained in Proposition 6 without using Grothendieck's inequality but we will not rely on this fact (at the expense of constant factors in some proofs). We also remark that this factorization can be efficiently found by a semidefinite program.

3 Non-adaptive estimators: Proof of Theorem 2

3.1 Proof of Theorem 2a

Suppose $p(x,y) = \sum_{i,j=1}^{n} A_{ij}x_iy_j$ is the advantage polynomial of a quantum algorithm, so in particular $\|A\|_{\infty\to 1} \le 1$. For $I, J \subseteq [n]$, we write A_{IJ} the submatrix of A restricted to rows indexed in I and columns indexed in J, and

$$p_{IJ}(x,y) = \sum_{i\in I, j\in J} A_{ij}x_iy_j.$$

We analyze the following $2k$-query classical distinguisher $P_{2a}(x,y)$:

1. Pick a pair of index sets $I, J \subseteq [n]$, $|I| = |J| = k$, with probability proportional to $\|p_{IJ}\|_{\infty\to 1}$.
2. Query all x_i with $i \in I$ and all y_j with $j \in J$.
3. Output $p_{IJ}(x,y)/\|p_{IJ}\|_{\infty\to 1}$.

As x and y take ± 1 values, the step 3 above, always outputs a value $D(x,y) \in [-1,1]$ as required for a distinguisher. We first show that $D(x,y)$ is an unbiased estimator of $p(x,y)$ up to a scalar.

▷ Claim 8. $Z \cdot D(x,y)$ is an unbiased estimator of $p(x,y)$, where

$$Z = \sum_{I,J:|I|=|J|=k} \|p_{IJ}\|_{\infty\to 1} / \binom{n-1}{k-1}^2.$$

Proof. The probability for choosing the index pair (I,J) in step 1 is given by

$$\frac{\|p_{IJ}\|_{\infty\to 1}}{\sum_{I',J':|I'|=|J'|=k}\|p_{I'J'}\|_{\infty\to 1}} = \frac{\|p_{IJ}\|_{\infty\to 1}}{Z\binom{n-1}{k-1}^2}.$$

Therefore

$$\begin{aligned}
\mathsf{E}[Z\cdot D(x,y)] &= Z \sum_{|I|=|J|=k} \frac{\|p_{IJ}\|_{\infty\to 1}}{Z\binom{n-1}{k-1}^2} \cdot \frac{p_{IJ}(x,y)}{\|p_{IJ}\|_{\infty\to 1}} \\
&= \frac{1}{\binom{n-1}{k-1}^2} \sum_{|I|=|J|=k} \sum_{i\in I, j\in J} A_{ij}x_iy_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_iy_j \\
&= p(x,y).
\end{aligned}$$

The second-to-last line uses the fact that each index i appears in exactly $\binom{n-1}{k-1}$ sets I and likewise for j and J. ◁

To complete the analysis we use the following inequality.

▶ **Proposition 9.** *There is a constant C so that for any $n \times n$ matrix A,*

$$\frac{1}{\binom{n}{k}^2} \sum_{|I|=|J|=k} \|A_{IJ}\|_{\infty\to 1} \le C\left(\frac{k}{n}\right)^{3/2} \|A\|_{\infty\to 1}. \tag{1}$$

Proof of Theorem 2a. Let p be the advantage polynomial of the one-query quantum algorithm with advantage ε on $(\mathbf{A}, \mathbf{B})$, and A be the matrix defining p. Without loss of generality, we assume

$$\mathsf{E}[p(\mathbf{A})] - \mathsf{E}[p(\mathbf{B})] \geq \varepsilon.$$

From Claim 8, we obtain

$$\mathsf{E}[D(\mathbf{A})] - \mathsf{E}[D(\mathbf{B})] = \frac{1}{Z}(\mathsf{E}[p(\mathbf{A})] - \mathsf{E}[p(\mathbf{B})]) \geq \frac{\varepsilon}{Z}.$$

It remains to upper bound the value of Z. Using Proposition 9 we get

$$Z = \frac{\binom{n}{k}^2}{\binom{n-1}{k-1}^2} \cdot \frac{1}{\binom{n}{k}^2} \sum_{|I|=|J|=k} \|A_{IJ}\|_{\infty\to 1} \leq \frac{n^2}{k^2} \cdot C\Big(\frac{k}{n}\Big)^{3/2} \|A\|_{\infty\to 1} \leq C\sqrt{\frac{n}{k}}.$$

This concludes the desired advantage bound of $\Omega(\varepsilon\sqrt{k/n})$. ◀

It follows from Proposition 5 that our analysis of D is tight up to constant factor.

Proposition 9 is similar to the following inequality proved by Rudelson and Vershynin [10, Equation (4.1)] who showed that for subsets I_ρ and J_ρ sampled by including each index independently with probability $\rho = k/n$,

$$\mathsf{E}[\|A_{I_\rho J_\rho}\|_{\infty\to 1}] \leq C'\rho^{3/2}\big(\|A\|_{\text{row}} + \|A\|_{\text{col}}\big) + C'\rho^2\|A\|_{\infty\to 1}, \tag{2}$$

for some constant C'. Here, $\|A\|_{\text{row}} = \sum_i \|A^{i,:}\|$ and $\|A\|_{\text{col}} = \sum_j \|A^{:,j}\|$ denote the sum of the 2-norms of its rows and columns, respectively. For completeness, we present the derivation Proposition 9 from (2) using Poissonization [13] (see also [12]).

Proof of Proposition 9. Tropp [13] showed that

$$\frac{1}{\binom{n}{k}^2} \sum_{|I|=|J|=k} \|A_{IJ}\|_{\#} \leq 4\mathsf{E}[\|A_{I_\rho J_\rho}\|_{\#}],$$

for every matrix norm $\|\cdot\|_{\#}$ that satisfies $\|A'\|_{\#} \leq \|A\|_{\#}$ for every submatrix A' of A. This is in particular true for the ∞-to-1 norm: if $\|A'\|_{\infty\to 1} = x'^{\top} A' y'$ for $x', y' \in [-1, 1]^m$, then $\|A\|_{\infty\to 1} \geq x^{\top} A y = x'^{\top} A' y'$ where x and y are extended from x' and y' with zeros padded in the remaining entries. It remains to prove that $\|A\|_{\text{row}}, \|A\|_{\text{col}} \leq K_G\|A\|_{\infty\to 1}$. ◀

▷ Claim 10. For any $M \in \mathbb{R}^{m\times n}$, $\|M^{i,:}\| \leq \|M\|$ and $\|M^{:,j}\| \leq \|M\|$ for any $i \in [m]$, $j \in [n]$.

Proof. We prove the case of $\|M^{i,:}\|$ and the other case follows the same proof:

$$\|M^{i,:}\|^2 = e_i^{\top} M(M^{i,:}) \leq \|M\| \cdot \|M^{i,:}\| \implies \|M^{i,:}\| \leq \|M\|. \quad \triangleleft$$

▷ Claim 11. $\|A\|_{\text{row}} \leq K_G\|A\|_{\infty\to 1}$.

Proof. By re-scaling, we assume $\|A\|_{\infty\to 1} = 1$ without loss of generality. Let $A_{ij} = \alpha_i \tilde{A}_{ij} \beta_j$ be the Grothendieck's factorization of A (Proposition 7), by Cauchy-Schwarz inequality,

$$\|A\|_{\text{row}} = \sum_i \sqrt{\sum_j A_{ij}^2} = \sum_i \alpha_i \cdot \sqrt{\sum_j \beta_j^2 \tilde{A}_{ij}^2} \leq \sqrt{\sum_i \alpha_i^2} \sqrt{\sum_{ij} \beta_j^2 \tilde{A}_{ij}^2} = \sqrt{\sum_j \beta_j^2 \|\tilde{A}^{:,j}\|^2}.$$

By Claim 10 and the bound $\|\tilde{A}\| \leq K_G$, we conclude that $\|A\|_{\text{row}} \leq \sqrt{\sum_j \beta_j^2 \cdot K_G^2} = K_G$. $\triangleleft$

3.2 The bias of truncated unbiased estimators

In preparation for the proof of Theorem 2b, we prove a general bound of the bias arising from truncating an unbiased estimator of low variance. Denote $\mathrm{trunc}\colon \mathbb{R} \to [-1,1]$ the truncation function

$$\mathrm{trunc}(t) = \begin{cases} t, & \text{if } |t| \le 1, \\ \mathrm{sign}(t), & \text{if } |t| > 1. \end{cases}$$

▶ **Proposition 12.** *Assume $\|f\|_\infty = 1$, $Z \ge 1$, and $F_{\mathbf{r}}$ is a random function such that $\mathsf{E}[F_{\mathbf{r}}(x)] = f(x)$ for all x (here r denotes the randomness). The distinguisher $D_{\mathbf{r}}(x) = \mathrm{trunc}(F_{\mathbf{r}}(x)/Z)$ has advantage at least*

$$\frac{\varepsilon}{Z} - 2\max_x \int_{1-1/Z}^{\infty} \mathsf{Pr}_r\big(|F_{\mathbf{r}}(x) - f(x)| \ge Zt\big)dt.$$

for any pair of random variables that are ε-distinguishable by f.

▶ **Corollary 13.** *Under the assumptions of Proposition 12, D has advantage at least*

$$\frac{\varepsilon}{Z} - \frac{2}{Z(Z-1)} \cdot \max_x \mathsf{Var}\, F_{\mathbf{r}}(x).$$

Proof. Using Chebyshev's inequality, the integrand appearing in Proposition 12 is at most $(\mathsf{Var}\, F_{\mathbf{r}}(x))/Z^2t^2$ and so the integral is at most $(\mathsf{Var}\, F_{\mathbf{r}}(x))/Z(Z-1)$. ◀

The proposition is derived from the following claim:

▷ Claim 14. Let Y be a random variable with $|\mathsf{E}[Y]| \le 1$, then

$$\big|\mathsf{E}[\mathrm{trunc}(Y)] - \mathsf{E}[Y]\big| \le \int_{1-|\mathsf{E}[Y]|}^{\infty} \mathsf{Pr}(|Y-\mu| \ge t)dt.$$

Proof of Proposition 12. We apply Claim 14 to the random variable $F_{\mathbf{r}}(x)/Z$ to obtain

$$\begin{aligned} |\mathsf{E}[D_{\mathbf{r}}(x)] - f(x)/Z| &\le \int_{1-|f(x)/Z|}^{\infty} \mathsf{Pr}_r\big(|F_{\mathbf{r}}(x)/Z - f(x)/Z| \ge t\big)dt \\ &\le \int_{1-1/Z}^{\infty} \mathsf{Pr}_r\big(|F_{\mathbf{r}}(x) - f(x)| \ge Zt\big)dt. \end{aligned}$$

Suppose $(\mathbf{A}, \mathbf{B})$ is ε-distinguishable by f. By the triangle inequality, $|\mathsf{E}[D_{\mathbf{r}}(\mathbf{A})] - \mathsf{E}[f(\mathbf{A})]/Z|$ is at most the maximum of the integral over x, and the same bound holds for replacing $\mathbf{A}$ by $\mathbf{B}$. Now the bound on advantage follows from triangle inequality. ◀

In the proof of Claim 14 we use the following fact:

▶ **Fact 15.** $|\mathrm{trunc}(t) - t| = \max\{0, |t| - 1\}$.

Proof of Claim 14. Let $\mu = \mathsf{E}[Y]$.

$$\begin{aligned}
|\mathsf{E}[\mathrm{trunc}(Y)] - \mathsf{E}[Y]| &\le \mathsf{E}[|\mathrm{trunc}(Y) - Y|] \\
&= \int_0^\infty \mathsf{Pr}(|\mathrm{trunc}(Y) - Y| \ge t)dt \\
&= \int_0^\infty \mathsf{Pr}(|Y| - 1 \ge t)dt && \text{(Fact 15)} \\
&= \int_0^\infty \mathsf{Pr}(|Y| \ge t + 1)dt \\
&\le \int_0^\infty \mathsf{Pr}(|Y - \mu| \ge t + 1 - |\mu|)dt && \text{(triangle inequality)} \\
&= \int_{1-|\mu|}^\infty \mathsf{Pr}(|Y - \mu| \ge t)dt && \text{(change of variables)}
\end{aligned}$$ ◁

3.3 Proof of Theorem 2b

Let $p(x, y) = \sum_{ij} A_{ij} x_i y_j$ be the advantage polynomial so that $\|A\|_{\infty\to 1} = 1$. We let $A_{ij} = \alpha_i \tilde{A}_{ij} \beta_j$ to be the Grothendieck's factorization of A. We analyze the following $2k$-query estimator:

1. Sample a sequence $I = (I(1), \ldots, I(k))$ of k i.i.d indices by picking each $i \in [n]$ with probability $p_i := \alpha_i / \|\alpha\|_1$. Query the inputs $x_{I(u)}$ for $u \in [k]$.
2. Sample a sequence $J = (J(1), \ldots, J(k))$ of k i.i.d indices by picking each $j \in [n]$ with probability $q_j := \beta_j / \|\beta\|_1$. Query the inputs $y_{J(v)}$ for $v \in [k]$.
3. Output the empirical average

$$P(x, y) = \mathsf{E}_{i\sim I, j\sim J}\left[A_{ij} \frac{x_i y_j}{p_i q_j}\right].$$

Clearly this estimator makes at most $2k$ queries. Now we show that this is an unbiased estimator of bounded variance.

▶ **Proposition 16.** *$P(x, y)$ is an unbiased estimator of $p(x, y)$ of variance at most $O(n/k^2)$.*

Proof. Unbiasedness follows from linearity of expectation:

$$\mathsf{E}[P(x, y)] = \mathsf{E}\left[\mathsf{E}\left[A_{ij} \frac{x_i y_j}{p_i q_j} \;\middle|\; I, J\right]\right] = \mathsf{E}\left[A_{ij} \frac{x_i y_j}{p_i q_j}\right] = \sum_{i,j} A_{ij} \frac{x_i y_j}{p_i q_j} \cdot p_i q_j = p(x, y).$$

In preparation for calculating the variance, let $B_{uv} = A_{I(u)J(v)} / p_{I(u)} q_{J(v)}$. By independence and the fact that $x_i^2 = y_j^2 = 1$,

$$\mathsf{Cov}\big(B_{uv}, B_{u'v'}\big) = \begin{cases} \mathsf{E}\left[\frac{A_{ij}^2}{p_i^2 q_j^2}\right] - p(x, y)^2, & \text{if } u = u' \text{ and } v = v' \\ \mathsf{E}\left[\frac{A_{ij} A_{ij'} y_j y_{j'}}{p_i^2 q_j q_{j'}}\right] - p(x, y)^2, & \text{if } u = u' \text{ and } v \ne v' \\ \mathsf{E}\left[\frac{A_{ij} A_{i'j} x_i x_{i'}}{p_i p_{i'} q_j^2}\right] - p(x, y)^2, & \text{if } u \ne u' \text{ and } v = v' \\ 0, & \text{otherwise.} \end{cases}$$

Here i, i' and j, j' denote random indices chosen independently. Decomposing $\mathsf{Var}[P(x, y)]$ as an average of covariances, we obtain

$$\begin{aligned}\mathsf{Var}[P(x,y)] &= \frac{1}{k^4}\sum_{u,v,u',v'}\mathsf{Cov}\big(B_{uv},B_{u'v'}\big)\\ &\le \frac{1}{k^2}\mathsf{E}\left[\frac{A_{ij}^2}{p_i^2q_j^2}\right]+\frac{k-1}{k^2}\left(\mathsf{E}\left[\frac{A_{ij}A_{ij'}y_jy_{j'}}{p_i^2q_jq_{j'}}\right]+\mathsf{E}\left[\frac{A_{ij}A_{i'j}x_ix_{i'}}{p_ip_{i'}q_j^2}\right]\right). \qquad (3)\end{aligned}$$

We bound the three types of terms using Grothendieck's factorization of A.

$$\begin{aligned}\mathsf{E}\left[\frac{A_{ij}^2}{p_i^2q_j^2}\right] &= \sum_{i,j}\frac{A_{ij}^2}{p_iq_j}\\ &= \|\alpha\|_1\cdot\|\beta\|_1\cdot\sum_{i,j}\alpha_i\beta_j\tilde{A}_{ij}^2 && \text{(Grothendieck's factorization)}\\ &\le \|\alpha\|_1\cdot\|\beta\|_1\cdot\sqrt{\sum_{i,j}\alpha_i^2\tilde{A}_{ij}^2}\cdot\sqrt{\sum_{i,j}\beta_j^2\tilde{A}_{ij}^2} && \text{(Cauchy-Schwarz inequality)}\\ &= \|\alpha\|_1\cdot\|\beta\|_1\cdot\sqrt{\sum_i\alpha_i^2\|\tilde{A}^{i,\cdot}\|^2}\cdot\sqrt{\sum_j\beta_j^2\|\tilde{A}^{\cdot,j}\|^2}\\ &\le \|\alpha\|_1\cdot\|\beta\|_1\cdot\|\tilde{A}\|\cdot\|\tilde{A}\| && \text{(Claim 10)}\\ &\le n\cdot K_G^2. && \text{(Cauchy-Schwarz inequality)}\end{aligned}$$

$$\begin{aligned}\mathsf{E}\left[\frac{A_{ij}A_{ij'}y_jy_{j'}}{p_i^2q_jq_{j'}}\right] &= \sum_{i,j,j'}\frac{A_{ij}A_{ij'}y_jy_{j'}}{p_i}\\ &= \sum_i\frac{1}{p_i}\left(\sum_jA_{ij}y_j\right)^2\\ &\le \|\alpha\|_1\sum_i\alpha_i\left(\sum_j\tilde{A}_{ij}y_j\beta_j\right)^2 && \text{(Grothendieck's factorization)}\\ &\le \|\alpha\|_1\sum_i(\tilde{A}\beta^y)_i^2 && \text{(Define } (\beta^y)_j := y_j\beta_j\text{; and } \alpha_i\in[0,1])\\ &= \|\alpha\|_1\|\tilde{A}\beta^y\|^2\\ &\le \|\alpha\|_1\|\tilde{A}\|^2 && (\|\beta^y\| = \|\beta\| = 1)\\ &\le \sqrt{n}\cdot K_G^2. && \text{(Cauchy-Schwarz inequality)}\end{aligned}$$

By symmetry the third term is also at most $\sqrt{n}K_G^2$. Plugging into (3), we obtain

$$\mathsf{Var}[P(x,y)] = O(n/k^2+\sqrt{n}/k) = O(n/k^2). \qquad \blacktriangleleft$$

Proof of Theorem 2b. Unbiasedness follows from linearity of expectation. Let V be the variance bound from Proposition 16. We instantiate Corollary 13 with this P and $Z = 1+4V/\varepsilon$. This is at most 1 provided $k^2 = o(n)$. The resulting distinguishing advantage is $\Omega(\varepsilon^2/V)$. $\blacktriangleleft$

The advantage of any distinguisher with the same distribution over samples cannot be better than $\varepsilon k^2/n$. Therefore our analysis is optimal in terms of k and n. To see this, consider the distribution in which the bit-pairs (x_i, y_i) are unbiased, ε-correlated, and mutually independent. The resultant bilinear form $(\sum x_iy_i)/n$ is 1-bounded, which corresponds to a one-query quantum distinguisher; and it distinguishes this distribution from random with advantage ε. In contrast, the advantage of a classical distinguisher is at most ε times the expected number of collisions $i = j$ with $i \in I$ and $j \in J$, which is at most $\varepsilon k^2/n$.

4 An adaptive estimator: Proof of Theorem 3

We modify the estimator of Section 3.3 so that the values $\{x_i : i \in I\}$ adaptively affect the probabilities for index sampling of J. Again we assume $\|A\|_{\infty\to 1} = 1$ and let $A_{ij} = \alpha_i \tilde{A}_{ij} \beta_j$ to be the Grothendieck's factorization of A.

1. Choose a sample I of k i.i.d indices by picking each $i \in [n]$ with probability $p_i = \alpha_i / \|\alpha\|_1$. Query the inputs x_i for $i \in I$. Let $a_I^x \in \mathbb{R}^n$ be defined by $[a_I^x]_j := \mathsf{E}_{i\sim I}[A_{ij} x_i / p_i]$.
2. Choose a sample J of k i.i.d indices by picking each $j \in [n]$ with probability $q_j = |[a_I^x]_j| / \|a_I^x\|_1$. Query the inputs y_j for $j \in J$.
3. Output the empirical average $P(x,y) = \mathsf{E}_{j\sim J}[[a_I^x]_j y_j / q_j]$.

This estimator is unbiased by linearity of expectation. The main technical result of this section is the following deviation bound:

▶ **Proposition 17.** *There is a constant C such that for all x, y, $\varepsilon > 0$, and $t > 0$,*

$$\mathsf{Pr}\left(|P(x,y) - p(x,y)| \geq \frac{C\sqrt{n \log n}}{k} \frac{t}{\varepsilon} \right) \leq \frac{k}{\sqrt{n}\log n} \left(\frac{\varepsilon}{t}\right)^2 + 2n^{-(t/\varepsilon)^2}.$$

Proof of Theorem 3. With a (possible) change in the constant factor in the lower bound, we may assume that $\varepsilon \leq \varepsilon_0$ for a sufficiently small constant ε_0 and $Z := C\sqrt{n \log n}/k\varepsilon \geq 2$. We apply Proposition 17 to bound the integral in Proposition 12 by

$$\begin{aligned}
&\int_{1-1/Z}^{\infty} \mathsf{Pr}\big(|P(x,y) - p(x,y)| \geq Zt\big) \\
&\quad \leq \frac{C}{\sqrt{\log n}} \cdot \frac{\varepsilon}{Z} \int_{1-1/Z}^{\infty} \frac{dt}{t^2} + 2\int_{1-1/Z}^{\infty} n^{-(t/\varepsilon)^2} dt \\
&\quad = \frac{C}{\sqrt{\log n}} \cdot \frac{\varepsilon}{Z} \cdot \frac{1}{1-1/Z} + \sqrt{\frac{4\pi\varepsilon^2}{\log n}} \cdot \mathsf{Pr}\big(\mathcal{N}(0, \varepsilon^2/2\log n) \geq 1 - 1/Z\big) \\
&\quad \leq \frac{2C}{\sqrt{\log n}} \cdot \frac{\varepsilon}{Z} + \sqrt{\frac{4\pi\varepsilon^2}{\log n}} \cdot \mathsf{Pr}\big(\mathcal{N}(0,1) \geq \sqrt{\log n / 2\varepsilon^2}\big) \\
&\quad \leq \frac{2C}{\sqrt{\log n}} \cdot \frac{\varepsilon}{Z} + \sqrt{\frac{4\pi\varepsilon^2}{\log n}} \cdot n^{-1/\varepsilon^2} \\
&\quad \leq \frac{\varepsilon}{6Z} + \frac{\varepsilon}{6Z}.
\end{aligned}$$

The second to last inequality is the Gaussian tail bound. The last inequality holds for sufficiently large n using the assumption that $\varepsilon \leq \varepsilon_0$. By Proposition 12, D has advantage at least $\varepsilon/3Z$. ◀

To prove Proposition 17, we split the difference between P and p via the "hybrid" $P'(x,y) = (a_I^x)^\top y = \sum_j [a_I^x]_j y_j$. Claims 18 show that P' has small variance and is therefore close to P. Claim 19 shows that P' is typically close to P.

▷ Claim 18. $\mathsf{Var}[P'(x,y)] \leq K_G^2 \sqrt{n}/k$.

▷ Claim 19. $\mathsf{Pr}[|P(x,y) - P'(x,y)| \geq t \cdot K_G\sqrt{n}/k \mid I] \leq 2\exp(-t^2/2)$ for every $t > 0$.

Proof of Proposition 17. By Claim 18 and Chebyshev's inequality, for every $t_1 > 0$,

$$\mathsf{Pr}\big[|P'(x,y) - p(x,y)| \geq t_1 \cdot K_G (\sqrt{n}/k)^{1/2}\big] \leq \frac{1}{t_1^2}.$$

Using Claim 19 together with a union bound and the triangle inequality, it follows that

$$\mathsf{Pr}\big[|P(x,y)-p(x,y)| \geq t_1 \cdot K_G(\sqrt{n}/k)^{1/2} + t_2 \cdot K_G\sqrt{n}/k\big] \leq \frac{1}{t_1^2} + 2\exp(-t_2^2/2)$$

for every $t_1 > 0$ and $t_2 > 0$. Plugging in $t_1 = (t/\varepsilon)\cdot(\sqrt{n}\log n/k)^{1/2}$ and $t_2 = (t/\varepsilon)\cdot(2\log n)^{1/2}$ gives the desired inequality. ◀

Proof of Claim 18. As the samples in I are independent,

$$\mathsf{Var}[P'(x,y)] = \frac{1}{k}\,\mathsf{Var}_i\left[\sum_j \frac{A_{ij}x_iy_j}{p_i}\right] = \frac{1}{k}\,\mathsf{Var}_i\left[\sum_j \frac{A_{ij}y_j}{p_i}\right]$$

because $x_i^2 = 1$. As i is sampled with probability $p_i = \alpha_i/\|\alpha\|_1$, we get

$$\begin{aligned}
\mathsf{Var}\left[\sum_j \frac{A_{ij}y_j}{p_i}\right] &\leq \sum_i p_i\left(\sum_j \frac{A_{ij}y_j}{p_i}\right)^2 \\
&\leq \|\alpha\|_1 \sum_i \alpha_i\left(\sum_j \tilde{A}_{ij}y_j\beta_j\right)^2 && \text{(Grothendieck's factorization)} \\
&\leq \|\alpha\|_1 \sum_i (\tilde{A}\beta^y)_i^2 && \text{(Define } (\beta^y)_j := y_j\beta_j\text{; and } \alpha_i \in [0,1]) \\
&\leq \|\alpha\|_1 \cdot \|\tilde{A}\|^2 && (\|\beta^y\| = \|\beta\| = 1) \\
&\leq \sqrt{n}\cdot K_G^2. && \text{(Cauchy-Schwarz inequality)}
\end{aligned}$$ ◀

Proof of Claim 19. Since $[a_I^x]_j y_j/q_j = \|a_I^x\|_1 y_j$, conditioned on I, $P(x,y)$ is an average of k independent random variables taking values either $-\|a_I^x\|_1$ or $\|a_I^x\|_1$ with mean $P'(x,y)$. Applying the Chernoff-Hoeffding bound to $kP/\|a_I^x\|_1$, we obtain

$$\mathsf{Pr}\big[|P(x,y)-P'(x,y)| \geq t\|a_I^x\|_1/\sqrt{k} \;\big|\; I\big] \leq 2\exp(-t^2/2).$$

It remains to show that $\|a_I^x\|_1 \leq K_G\sqrt{n/k}$ for every choice of I:

$$\begin{aligned}
\|a_I^x\|_1 &= \sum_j \left|\frac{1}{k}\sum_{i\in I}\frac{x_iA_{ij}}{p_i}\right| \\
&= \frac{\|\alpha\|_1}{k}\sum_j \beta_j\left|\sum_{i\in I} x_i\tilde{A}_{ij}\right| && \text{(Grothendieck's factorization)} \\
&\leq \frac{\|\alpha\|_1}{k}\sqrt{\sum_j\left(\sum_{i\in I} x_i\tilde{A}_{ij}\right)^2} && \text{(Cauchy-Schwarz inequality)} \\
&= \frac{\|\alpha\|_1}{k}\sqrt{\sum_j (x_I^\top\tilde{A})_j^2} && \text{(Define } (x_I)_i := x_i\cdot\mathbf{1}(i\in I)) \\
&= \frac{\|\alpha\|_1}{k}\|x_I^\top\tilde{A}\| \\
&\leq \frac{\|\alpha\|_1\cdot\|x_I\|\cdot\|\tilde{A}\|}{k} \\
&\leq \frac{\sqrt{n}\cdot\sqrt{k}\cdot K_G}{k}. && \text{(Cauchy-Schwarz inequality)}
\end{aligned}$$ ◀

As mentioned, Theorem 4 shows that the distinguisher in Theorem 3 is best possible up to a factor of $\log n$.

5 Classical advantage upper bounds: Proofs of Proposition 1 and Theorem 4

To start this section, we first present the proof for the classical advantage upper bound of non-adaptive algorithms.

Proof of Proposition 1. Aaronson and Ambainis [1] show that F is $\Omega(1)$-distinguishable from the uniform random U by one quantum query. The random variable F is obtained by rounding a pair of n-dimensional Gaussians (X, Y) where X is standard Gaussian and Y is obtained by applying the Hadamard matrix to X. So the non-adaptive classical $2k$-query advantage is upper bounded by the maximum statistical distance between the projections (X_I, Y_J) over all sets I, J with $|I| + |J| = 2k$ and a standard $2k$-dimensional Gaussian.

In general, the statistical distance between centered multivariate Gaussians with covariance matrices Σ_1 and Σ_2 is $\Theta(1)\min\{1, \|\Sigma_2^{-1}\Sigma_1 - I\|_F\}$ [7]. As Σ_2 is the identity and all non-diagonal entries of Σ_1 are $\pm 1/\sqrt{n}$, it follows that $\|\Sigma_2^{-1}\Sigma_1 - Id\|_F = O(k/\sqrt{n})$.

Setting F_ε as $\varepsilon F + (1-\varepsilon)U$, the advantage of any distinguisher, classical or quantum, scales precisely by ε. ◀

As for the classical advantage upper bound of two-round algorithms, the proof of Theorem 4 bounds the statistical distance between the distinguisher's views on the two distributions via their KL-divergence. We need the following explicit formula for KL-divergence of multivariate Gaussians:

▶ **Fact 20.** $\mathsf{kl}(\mathcal{N}(\mu, \Sigma), \mathcal{N}(0, Id_k)) = \frac{1}{2}(\|\mu\|^2 + \mathsf{tr}(\Sigma - Id_k) - \log \mathsf{det}\Sigma)$.

The following consequence of this formula is implicit in [7]:

▷ Claim 21. Assuming $\lambda_{\min}(\Sigma) \geq 1/3$, $\mathsf{kl}((\mathcal{N}(\mu, \Sigma), \mathcal{N}(0, Id_k)) \leq \frac{1}{2}(\|\mu\|^2 + \|\Sigma - Id_k\|_F^2)$.

Proof. Let $\eta_1, \ldots, \eta_k$ be the eigenvalues of $\Sigma - Id_k$. By assumption $\eta_1, \ldots, \eta_k \geq -2/3$. Then

$$\mathsf{tr}(\Sigma - Id_k) - \log \mathsf{det}\Sigma = \sum_{i=1}^{k}(\eta_i - \log(1+\eta_i)) \leq \sum_{i=1}^{k}\eta_i^2 = \|\Sigma - Id_k\|_F^2,$$

where the inequality uses the fact that $\eta - \log(1+\eta) \leq \eta^2$ for all $\eta \geq -2/3$. ◁

The requirement $\lambda_{\min}(\Sigma) \geq 1/3$ is satisfied by matrices that are close to the identity in the following sense:

▶ **Fact 22.** *If $A \in \mathbb{R}^{k\times k}$ is a symmetric matrix with $|A_{ij}| \leq \varepsilon$ for all $i, j \in [k]$, then $\lambda_{\min}(Id_k + A) \geq 1 - k\varepsilon$ and $\lambda_{\max}(Id_k + A) \leq 1 + k\varepsilon$.*

In particular, as long as Σ is $2/3k$-close to the identity in (entrywise) infinity-norm, the bound in Claim 21 applies.

Another tool we need is the chain rule for KL-divergence:

▶ **Fact 23** (Chain rule for KL-divergence). $\mathsf{kl}((U, V), (U', V')) = \mathsf{kl}(U, U') + \mathsf{kl}(V|U, V'|U')$, *where* $\mathsf{kl}(V|U, V'|U') = \mathsf{E}_{u\sim U}\mathsf{kl}(V|U = u, V'|U' = u)$.

In our application of Fact 23, $(U, V) = (U_1, \ldots, U_k, V_1, \ldots, V_k)$ is a multivariate Gaussian. This class of distributions is closed under conditioning. To calculate the effect of conditioning on the parameters, we identify the zero-mean (assumed without loss of generality) random variables $U_1, \ldots, U_k, V_1, \ldots, V_k$ with vectors in Hilbert space endowed with the inner product $\mathsf{E}[A \cdot B]$. The conditional means and conditional covariances of $V|U$ are given by

$$\mathsf{E}[V_j|U] = V_j^{\parallel} \tag{4}$$

$$\mathsf{Cov}[V_j, V_{j'}|U] = \mathsf{E}[V_j^{\perp} \cdot V_{j'}^{\perp}], \tag{5}$$

where $V_j = V_j^{\parallel} + V_j^{\perp}$ is the orthogonal decomposition of V_j into a parallel component $V_j^{\parallel} \in \mathrm{Span}(U)$ and a perpendicular component $V_j^{\perp} \in \mathrm{Span}(U)^{\perp}$. As $V_j^{\parallel}$ is in $\mathrm{Span}(U)$, its value is determined by $U_1, \ldots, U_k$. And as $V_j^{\perp}$ and $V_{j'}^{\perp}$ are in $\mathrm{Span}(U)^{\perp}$, their values are independent of $U_1, \ldots, U_k$.

Lastly we will use the following fact:

▶ **Fact 24.** *If M is the maximum of n standard Gaussian random variables, then $\mathsf{E}[M^2] \le 4\log(\sqrt{2}n)$.*

Proof of Fact 24. By Jensen's inequality, for every $t \in (0, 1/2)$,

$$\exp(t\mathsf{E}[M^2]) \le \mathsf{E}[\exp(tM^2)] \le \mathsf{E}[n\exp(t\mathcal{N}(0,1)^2)] = \frac{n}{\sqrt{1-2t}}.$$

Here, the last equality follows from the formula of the moment-generating function of a squared Gaussian. We obtain the desired formula by setting $t = 1/4$ and taking logarithms. ◀

Proof of Theorem 4. The random variable $F = (\operatorname{sign} X, \operatorname{sign} Y)$ is the same as in Proposition 1. As in the previous proof, we first reduce to the case when ε is constant and use the same notations in that proof.

Aaronson and Ambainis showed that $\mathsf{E}[\operatorname{sign}(X)(H/n)\operatorname{sign}(Y)] = \Omega(1)$, where H is the $n \times n$ Hadamard matrix. As $\|H/n\|_{\infty\to 1} \le \|H\| \le 1$, this justifies the quantum advantage.

For the classical case, as taking signs can only decrease advantage, we upper bound the advantage of distinguishing $Z = (X, Y)$ from $\mathcal{N}(0, Id_{2n})$. The only relevant property of Z is that $\mathsf{E}[Z_i] = 1$ and $|\mathsf{E}[Z_i Z_j]| \le 1/\sqrt{n}$ for all pairs $i \ne j$. Without loss of generality, we assume that $k \le \sqrt{n}/4$.

The distinguisher's strategy is specified by the query sets I and J with $|I| = |J| = k$, issued in the first and second round, respectively. For the sake of upper bound, we can assume without loss of generality that J is a deterministic function of the coordinates $Z_I = (Z_i)_{i\in I}$ observed in the first round. The distinguisher's advantage is at most

$$\begin{aligned}
\varepsilon_C &= \max_{I,J} \mathsf{sd}((Z_I, Z_J), \mathcal{N}(0, Id_{2k})) \\
&\le \sqrt{\frac{1}{2}\max_{I,J} \mathsf{kl}((Z_I, Z_J), \mathcal{N}(0, Id_{2k}))} && \text{(Pinsker's inequality)} \\
&= \sqrt{\frac{1}{2}\Big(\max_I \mathsf{kl}(Z_I, \mathcal{N}(0, Id_k)) + \max_I \max_J \mathsf{E}_{Z_I} \mathsf{kl}(Z_J|Z_I, \mathcal{N}(0, Id_k))\Big)} && \text{(Fact 23)} \\
&\le \sqrt{\frac{1}{2}\Big(\max_I \mathsf{kl}(Z_I, \mathcal{N}(0, Id_k)) + \max_I \mathsf{E}_{Z_I} \max_J \mathsf{kl}(Z_J|Z_I, \mathcal{N}(0, Id_k))\Big)} && \text{(convexity of max)}
\end{aligned}$$

As $k \le \frac{2}{3}\sqrt{n}$, the covariance matrix Σ_I of Z_I is $2/3k$-close to the identity in infinity norm, so using Claim 21, for every I one has

$$\mathsf{kl}(Z_I, \mathcal{N}(0, Id_k)) \le \frac{1}{2}\|\Sigma_I - Id_k\|_F \le \frac{k^2}{2n}. \tag{6}$$

For the second KL-divergence, let $\mu_{J|I}$ and $\Sigma_{J|I}$ denote the vector of conditional means $(\mathsf{E}[Z_j|Z_I])_{j\in J}$ and covariances $\mathsf{Cov}[Z_j, Z_{j'}|Z_I]$ for $j, j' \in J$, respectively. We will prove that for all choices of I and J, $\Sigma_{J|I}$ is $2/3k$-close to Id_k and apply Claim 21 to bound it by

$$\mathsf{E}_{Z_I} \max_J \mathsf{kl}(Z_J|Z_I, \mathcal{N}(0, Id_k)) \le \frac{1}{2} \max_J \mathsf{E}\left[\|\mu_{J|I}\|^2\right] + \frac{1}{2} \max_J \mathsf{E}\left[\|\Sigma_{J|I}\|_F^2\right]. \tag{7}$$

To bound the first term in (7), we analyze the projections $Z_j^{\|}$ of Z_j onto $\mathrm{Span}\{Z_i : i \in I\}$ for every $j \notin I$. Fix a basis for the vector space spanned by $Z_1, \ldots, Z_{2n}$ and let z_i be the representation of Z_i under this basis. Let B be the $k \times n$ matrix whose rows are z_i for $i \in I$. The projection $z_j^{\|}$ of z_j onto the row-span of B is given by the formula

$$z_j^{\|} = B^\top (BB^\top)^{-1} B z_j.$$

The norm of this projection is at most

$$\|z_j^{\|}\| \le \lambda_{\max}(BB^\top)^{1/2} \cdot \lambda_{\min}(BB^\top)^{-1} \cdot \|Bz_j\| \le \left(1 + \frac{k}{\sqrt{n}}\right)^{1/2} \left(1 - \frac{k}{\sqrt{n}}\right)^{-1} \cdot \sqrt{\frac{k}{n}} \le 2\sqrt{\frac{k}{n}}.$$

The second inequality follows from Fact 22 as $BB^\top$ is $1/\sqrt{n}$-close to the identity and the entries of Bz_j are all bounded by $1/\sqrt{n}$. The third inequality follows from the assumption $k \le \sqrt{n}/3$.

By (4), for every j, $\mathsf{E}[Z_j|Z_I]$ is a Gaussian random variable of mean zero and standard deviation at most $2\sqrt{k/n}$. Letting $\mu_{J|I}$ denote the vector of conditional means $(\mathsf{E}[Z_j|Z_I])_{j \in J}$, by Fact 24, for any fixed I,

$$\max_J \mathsf{E}\left[\|\mu_{J|I}\|^2\right] \le k\mathsf{E} \max_{j \in [n] \setminus I} \mathsf{E}[Z_j|Z_I]^2 \le k \left(2\sqrt{\frac{k}{n}}\right)^2 \cdot 4\log(\sqrt{2n}) = \frac{16k^2 \log(\sqrt{2}n)}{n}. \tag{8}$$

For the second term in (7), we apply (5) to obtain

$$\mathsf{Cov}[Z_j, Z_{j'}|Z_I] = \mathsf{E}[Z_j^{\perp} \cdot Z_{j'}^{\perp}] = \mathsf{E}[Z_j \cdot Z_{j'}] - \mathsf{E}[Z_j^{\|} \cdot Z_{j'}^{\|}]$$

by orthogonality, from which we have

$$\left|\mathsf{Cov}[Z_j, Z_{j'}|Z_I] - \mathsf{E}[Z_j \cdot Z_{j'}]\right| \le \|z_j^{\|}\| \cdot \|z_{j'}^{\|}\| \le \frac{4k}{n} \le \frac{1}{\sqrt{n}}. \tag{9}$$

As $\mathsf{E}[Z_j \cdot Z_{j'}]$ is $1/\sqrt{n}$ close to the identity, we conclude that $\Sigma_{J|I}$ is $2/\sqrt{n} \le 2/3k$-close to the identity. Therefore Claim 21 applies. Plugging (8) and (9) into (7) we obtain

$$\mathsf{E}_{Z_I} \max_J \mathsf{kl}(Z_J|Z_I, \mathcal{N}(0, Id_k)) \le \frac{1}{2} \cdot \frac{16k^2 \log(\sqrt{2}n)}{n} + \frac{1}{2} \cdot k^2 \cdot \frac{4}{n} = O\left(\frac{k^2 \log n}{n}\right).$$

Together with (6), this gives $\varepsilon_C = O(k\sqrt{\log n}/\sqrt{n})$ as desired. ◀

6 Optimality of local unbiased estimators: Proof of Proposition 5

Proof of Proposition 5. With a (possible) change in the constant factor in the lower bound, we may assume each form in the mixture depends on k bits of x and k bits of y. Let $p(x, y) = \sum A_{ij} x_i y_j$, and let B_{IJ} be a matrix supported on $I, J \subseteq [n]$ with $|I| = |J| = k$:

$$[B_{IJ}]_{ij} = \begin{cases} b_{ij} & \text{if } i \in I, j \in J \\ 0 & \text{otherwise} \end{cases}.$$

Suppose $\|A\|_{\infty\to 1} = 1$ and $A = \mathsf{E}_{I,J}[B_{IJ}]$, where the expectation is over an arbitrary distribution over pairs I, J. It is sufficient to show that $\|B_{IJ}\|_{\infty\to 1} = \Omega(\sqrt{n/k})$ for at least one choice of the index sets (I, J).

We prove the contrapositive. Suppose $\|B_{IJ}\|_{\infty\to 1} \leq \varepsilon$ for all B_{IJ}. By Claim 11, $\|B_{IJ}\|_{\text{row}} \leq K_G \cdot \varepsilon$, so $\sum_{i\in I, j\in J} |[B_{IJ}]_{ij}| \leq K_G \cdot \varepsilon\sqrt{k}$. Therefore

$$\sum_{i,j} |A_{ij}| = \sum_{i,j} |\mathsf{E}_{I,J}\,[[B_{IJ}]_{ij}]| \leq \mathsf{E} \sum_{i\in I, j\in J} |[B_{IJ}]_{ij}| \leq K_G \cdot \varepsilon\sqrt{k}.$$

For n being a power of two, let A be the $n \times n$ Hadamard matrix scaled by $1/n$, such that $\|A\|_{\infty\to 1} = n$. Then $|A_{ij}| = n^{-3/2}$ for all i and j, thus $K_G \cdot \varepsilon\sqrt{k} \geq \sqrt{n}$ and hence $\varepsilon = \Omega(\sqrt{n/k})$ as desired. If n is not a power of two, we plant the largest possible Hadamard matrix and zero out the remaining entries, the same argument still applies. ◀

References

1 Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 307–316, 2015. doi:10.1145/2746539.2746547.

2 Scott Aaronson, Andris Ambainis, Andrej Bogdanov, Krishnamoorthy Dinesh, and Cheung Tsun Ming. On quantum versus classical query complexity. *Electron. Colloquium Comput. Complex.*, TR21-115, 2021. arXiv:TR21-115.

3 Scott Aaronson, Andris Ambainis, Janis Iraids, Martins Kokainis, and Juris Smotrovs. Polynomials, quantum query complexity, and grothendieck's inequality. In *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, pages 25:1–25:19, 2016. doi:10.4230/LIPIcs.CCC.2016.25.

4 Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1303–1316, 2021. doi:10.1145/3406325.3451040.

5 Sergey Bravyi, David Gosset, Daniel Grier, and Luke Schaeffer. Classical algorithms for Forrelation, 2021. arXiv:2102.06963.

6 Harry Buhrman, Lance Fortnow, Ilan Newman, and Hein Röhrig. Quantum property testing. *SIAM J. Comput.*, 37(5):1387–1400, 2008. doi:10.1137/S0097539704442416.

7 Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians, 2020. arXiv:1810.08693.

8 Joseph Jaeger, Fang Song, and Stefano Tessaro. Quantum key-length extension. In *Theory of Cryptography: 19th International Conference, TCC 2021, Raleigh, NC, USA, November 8–11, 2021, Proceedings, Part I*, pages 209–239, Berlin, Heidelberg, 2021. Springer-Verlag. doi:10.1007/978-3-030-90459-3_8.

9 Gilles Pisier. Grothendieck's theorem, past and present. *Bulletin of the American Mathematical Society*, 49(2):237–323, May 2012. doi:10.1090/s0273-0979-2011-01348-9.

10 Mark Rudelson and Roman Vershynin. Sampling from large matrices: an approach through geometric functional analysis, 2006. arXiv:math/0503442.

11 Alexander A. Sherstov, Andrey A. Storozhenko, and Pei Wu. An optimal separation of randomized and quantum query complexity. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1289–1302, 2021. doi:10.1145/3406325.3451019.

12 Joel A. Tropp. Column subset selection, matrix factorization, and eigenvalue optimization, 2008. arXiv:0806.4404.

13 Joel A. Tropp. On the linear independence of spikes and sines. *Journal of Fourier Analysis and Applications*, 14(5-6):838–858, September 2008. doi:10.1007/s00041-008-9042-0.

14 Henry Yuen. A quantum lower bound for distinguishing random functions from random permutations. *Quantum Info. Comput.*, 14(13–14):1089–1097, October 2014.

15 Mark Zhandry. How to construct quantum random functions. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 679–687, 2012. doi:10.1109/FOCS.2012.37.