Towards Optimal Separations between Quantum and Randomized Query Complexities

Avishay Tal
Department of Electrical Engineering and Computer Sciences,
University of California, Berkeley,
CA 94720, United States.
Email: atal@berkeley.edu

2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS) | 978-1-7281-9621-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/FOCS46700.2020.00030

***Abstract*—The query model offers a concrete setting where quantum algorithms are provably superior to randomized algorithms. Beautiful results by Bernstein-Vazirani, Simon, Aaronson, and others presented partial Boolean functions that can be computed by quantum algorithms making much fewer queries compared to their randomized analogs. To date, separations of $O(1)$ vs. $\sqrt{N}$ between quantum and randomized query complexities remain the state-of-the-art (where N is the input length), leaving open the question of whether $O(1)$ vs. $N^{1/2+\Omega(1)}$ separations are possible?**

We answer this question in the affirmative. Our separating problem is a variant of the Aaronson-Ambainis k-fold Forrelation problem. We show that our variant:

1) **Can be solved by a quantum algorithm making $2^{O(k)}$ queries to the inputs.**
2) **Requires at least $\widetilde{\Omega}(N^{2(k-1)/(3k-1)})$ queries for any randomized algorithm.**

For any constant $\varepsilon > 0$, this gives a $O(1)$ vs. $N^{2/3-\varepsilon}$ separation between the quantum and randomized query complexities of partial Boolean functions.

Our proof is Fourier analytical and uses new bounds on the Fourier spectrum of classical decision trees, which could be of independent interest.

Looking forward, we conjecture that the Fourier bounds could be further improved in a precise manner, and show that such conjectured bounds imply optimal $O(1)$ vs. $N^{1-\varepsilon}$ separations between the quantum and randomized query complexities of partial Boolean functions.

***Keywords*-decision tree complexity; Fourier Analysis; Forrelation; Fourier Tails; quantum query complexity; query complexity**

I. Introduction

The query model (or black-box model) offers a concrete setting where quantum algorithms are provably superior to their randomized counterparts. Many well-known quantum algorithms can be cast in this model, such as Grover's search [Gro96], Deutsch-Jozsa's algorithm [DJ92], Bernstein-Vazirani's algorithm [BV97], Simon's algorithm [Sim97], and Shor's period-finding algorithm (which is a major component in Shor's factoring algorithm [Sho97]). In the query model, we seek to answer a question about the input by making as few queries to it as possible. For deterministic algorithms, this is also known as the decision tree model, where the decision tree depth equals the number of queries. The randomized and quantum versions of this model are very well-studied with many known connections and separations between the models in different settings (cf. the wonderful survey of [BdW02] or the more recent work of [ABK16]).

A beautiful line of work showed that for partial Boolean functions on N variables, the quantum query complexity could be exponentially smaller (or even less) than the randomized query complexity. Separations of $O(\log N)$ vs. $\sqrt{N}$ date back to the work of Simon [Sim97] and similarly for a problem introduced by Childs et al. [CCD$^+$03]. In [Aar10], [AA15], it was shown that the problem of *Forrelation* exhibits a 1 vs. $\widetilde{\Omega}(\sqrt{N})$ separation between the quantum and randomized query complexities.

Buhrman et al. [BFNR08] and Aaronson and Ambainis [AA15] asked what are the best possible separations between quantum and randomized query complexities? Aaronson and Ambainis presented this question as a fundamental question whose answer will shed light on the differences between the two models and gave several results towards its answer.

On the one hand, they ruled out $O(1)$ vs. $\Omega(N)$ separations. More precisely, they showed that for any constant t, any quantum algorithm that makes t queries can be simulated (up to small error) by a randomized algorithm making $\widetilde{O}(N^{1-1/2t})$ non-adaptive queries. For $t = 1$, this shows that Forrelation is an extremal separation.

On the other hand, towards obtaining optimal t vs. $\Omega(N^{1-1/2t})$ separations, they suggested a candidate: the k-fold Forrelation problem (to be defined shortly), where $k = 2t$. They showed that a quantum algorithm making $\lceil k/2 \rceil = t$ queries can solve k-fold Forrelation. Moreover, they conjectured that any randomized algorithm would require $\Omega(N^{1-1/k}) = \Omega(N^{1-1/2t})$ queries. While Aaronson and Ambainis proved the case $k = 2$ in their conjecture, they left all other cases wide open. For $k > 2$, the lower bounds they obtained on k-fold Forrelation are of the form $\Omega(\sqrt{N}/(\log N)^{7/2})$.

Aaronson and Ambainis further noted that in all of the above exponential separations, the randomized query complexity did not surpass $\sqrt{N}$. They asked whether separations of $\mathrm{polylog}(N)$ vs. $N^{1/2+\Omega(1)}$ are possible?

2575-8454/20/$31.00 ©2020 IEEE
DOI 10.1109/FOCS46700.2020.00030

We answer their question in the affirmative. We revisit their candidate, changing it in a way that would be crucial for our analysis. First, we define the **Rorrelation** of k vectors, with respect to an N-by-N orthogonal matrix U.

Definition I.1. *Let $U \in \mathbb{R}^{N\times N}$ be an orthogonal matrix. The k-**fold Rorrelation** of vectors $z^{(1)},\ldots,z^{(k)} \in \mathbb{R}^N$ with respect to U is defined as*

$$\phi_U(z^{(1)},\ldots,z^{(k)}) = \frac{1}{N}\cdot\sum_{i_1=1}^{N}\cdots\sum_{i_k=1}^{N} z^{(1)}_{i_1}\cdot U_{i_1,i_2}\cdot z^{(2)}_{i_2}\cdots U_{i_{k-1},i_k}\cdot z^{(k)}_{i_k}.$$

One can pick U to be the N-by-N Hadamard matrix, as suggested by Aaronson and Ambainis. We, however, pick U uniformly at random from all N-by-N orthogonal matrices.[1] This will play a major role later on, since we rely on properties that hold with high probability for a random orthogonal matrix, but do not hold for the Hadamard matrix (see Def. V.5).

It is not hard to show that the k-fold Rorrelation of any vectors $z^{(1)},\ldots,z^{(k)} \in \{-1,1\}^N$ is at most 1 in absolute value. The computational task we consider in this paper, called the **k-fold Rorrelation Problem**, asks to distinguish between the following two cases:

- **YES Instances:** Vectors $z^{(1)},\ldots,z^{(k)} \in \{-1,1\}^N$ with $\phi_U(z^{(1)},\ldots,z^{(k)}) \geq 2^{-k}$, and
- **NO Instances:** vectors $z^{(1)},\ldots,z^{(k)} \in \{-1,1\}^N$ with $|\phi_U(z^{(1)},\ldots,z^{(k)})| \leq \frac{1}{2}\cdot 2^{-k}$.

We shall show that while the k-fold Rorrelation problem is easy in the quantum query model (for any choice of U), it requires many queries in the classical setting (for most choices of U). Namely, our main separation will show that:

1) For any N-by-N orthogonal matrix U, there exists a quantum algorithm making at most $2^{O(k)}$ queries that solves the k-fold Rorrelation problem (with respect to U) with success probability at least $2/3$.
2) For most N-by-N orthogonal matrices U, any randomized algorithm that solves the k-fold Rorrelation problem (with respect to U) with success probability at least $2/3$, must make at least $\Omega(N^{2(k-1)/(3k-1)}/k\log N)$ queries to the inputs.

So far, we left the choice for the value of k to be arbitrary. We think of k as either a fixed constant or a slow-growing function of N, in particular $k = o(\log N)$. By picking k to be a large constant, the above discussion gives a $O(1)$ vs. $\Omega(N^{2/3-\varepsilon})$ separation of quantum and randomized query complexities, for any small constant $\varepsilon > 0$. By picking $k = O(\log\log N)$, we get a $O(\log N)$ vs. $N^{2/3-o(1)}$ separation of the two measures.

[1] Aaronson and Ambainis called their variant Forrelation, as in the case $k = 2$ it measures **correlation** after applying the **F**ourier/Hadamard transform on $z^{(1)}$. We measure correlation after applying a **R**andomly chosen orthogonal matrix, hence the name **Rorrelation**.

Before explaining more about our techniques and the potential room for improvement, we describe an application of our results to another setting.

Application: Power-$(2+2/3)$ Separations for Total Boolean Functions.: While for partial Boolean functions (or promise problems) exponential separations are possible between randomized and quantum query complexities, for total functions (i.e., Boolean functions that are defined on the entire domain $\{-1,1\}^N$) the picture is quite different. The seminal work of Beals et al. [BBC$^+$01] showed (among others) that quantum query complexity and randomized query complexity are at most power-6 apart. That is, $R(f) \leq O(Q(f)^6)$ for any total Boolean function f, where $R(f)$ and $Q(f)$ denote the randomized and quantum query complexities of f, respectively.

On the other hand, Grover's search demonstrated that for the OR function $R(f) \geq \Omega(Q(f)^2)$ [Gro96], and this is tight [BBBV97]. Two decades later, Aaronson, Ben-David, Kothari [ABK16] exhibited the first super-quadratic separations between $Q(f)$ and $R(f)$ for total functions, surprisingly improving Grover's separation that was believed to be optimal. Their work presented a power-2.5 separation based on the "cheat-sheet" technique applied to 2-fold Forrelation. More generally, they showed that any $N^{o(1)}$ vs. $N^{c-o(1)}$ separation between the quantum and randomized query complexities of partial functions, implies a power-$(2+c)$ separation for total Boolean functions. Plugging in our result, yields a power-$(2+2/3)$ separation for total Boolean functions. In other words, the transformation of [ABK16] applied to k-fold Rorrelation yields a total function f_{CS} such that $R(f_{\mathrm{CS}}) \geq Q(f_{\mathrm{CS}})^{2+\frac{2}{3}-o(1)}$.

A. Our Techniques

Quantum Query Complexity of the k-fold Rorrelation Problem.

A simple adaptation of the algorithm suggested by Aaronson and Ambainis [AA15] shows the existence of a $\lceil k/2\rceil$-query quantum algorithm on inputs $z^{(1)},\ldots,z^{(k)} \in \{-1,1\}^N$, whose acceptance probability equals

$$\frac{1+\phi_U(z^{(1)},\ldots,z^{(k)})}{2}$$

This shows that there's a gap of $\frac{1+2^{-k}}{2}$ vs. $\frac{1+2^{-(k+1)}}{2}$ between the acceptance probabilities in the YES instances and NO instances. For k a fixed constant this gives a constant difference between the acceptance probabilities of the YES and NO instances. If $k = \omega(1)$ or if we insist on getting a $2/3$ vs. $1/3$ separation between the acceptance probabilities of YES and NO instances, then one apply simple amplification techniques repeating the quantum algorithm for $2^{O(k)}$ times, and check whether the number of accepting trials exceeds a certain threshold.

Randomized Query Complexity of the k-fold Rorrelation Problem.

Towards showing that the randomized query complexity of the k-fold Rorrelation problem is large, we construct a hard-distribution, and show that it is hard to solve the Rorrelation problem on instances sampled from this distribution. By Yao's minimax principle, it suffices to show that a deterministic decision tree cannot solve the Rorrelation problem on the hard-distribution with high probability. We take the hard-distribution to be the convex combination (i.e., average) of two distributions: (1) the uniform distribution on k vectors $z^{(1)}, \ldots, z^{(k)} \in \{-1, 1\}^N$, denoted $\mathcal{U}_k$, and (2) a distribution $\mathcal{D}_{U,k}$ that often produces k vectors with large k-fold Rorrelation. On the one hand, we show that $\mathcal{U}_k$ produces NO instances with very high probability whereas $\mathcal{D}_{U,k}$ produces YES instances with not too small probability.[2] On the other hand, we show that any depth-d deterministic decision tree fails to distinguish between $\mathcal{U}_k$ and $\mathcal{D}_{U,k}$, as long as $d = o(N^{2(k-1)/(3k-1)}/\log N)$. Combining the two facts together, we deduce that the distribution $\frac{1}{2}\mathcal{D}_{U,k} + \frac{1}{2}\mathcal{U}_k$ is a hard-distribution for depth d decision trees. That is, any depth-d decision tree errs with not too small probability in computing the Rorrelation problem on instances sampled from this distribution.

We view the construction of a hard-distribution as an important contribution to the project set up by Aaronson and Ambainis. They were able to analyze 2-fold Forrelation by presenting a hard-distribution for that case, but no candidate hard-distribution was suggested for the case $k > 2$ prior to this work. We believe that our distribution is hard even for decision trees of depth $N^{1-1/k}/\mathrm{polylog}(N)$ and pose a conjecture that would imply such a result.

By the above discussion, proving that $\frac{1}{2}\mathcal{D}_{U,k} + \frac{1}{2}\mathcal{U}_k$ is a hard-distribution boils down to showing that:

1) $\mathcal{U}_k$ samples NO-instances with very high probability.
2) $\mathcal{D}_{U,k}$ samples YES-instances with not too small probability (to be precise, at least 2^{-k}).
3) Any deterministic decision tree of depth $d = o(N^{2(k-1)/(3k-1)}/\log N)$ cannot distinguish between inputs sampled from $\mathcal{U}_k$ and inputs sampled from $\mathcal{D}_{U,k}$. Put differently, the acceptance probability of any such deterministic decision tree, will be the same in both cases, up to an additive small error $o(1/2^k)$.

Item 1 holds regardless of the choice of $\mathcal{D}_{U,k}$, and is simple to prove. To obtain Item 2, we start by recalling the hard distribution that Aaronson and Ambainis suggested for the case $k = 2$. They first defined a multi-variate Gaussian distribution G_2 on $2N$ dimensions, where the first N variables are simply standard independent Gaussians, and the latter N variables are obtained by applying the Fourier (or Hadamard) transform on the first N variables. Then, to get a distribution $\mathcal{D}_2$ over the Boolean domain, they took the signs of these multi-variate Gaussians. They then show that the expected Forrelation value of vectors sampled from $\mathcal{D}_2$ is at least $(2/\pi)$.

We generalize this hard distribution to k-fold Rorrelation, replacing the Hadamard matrix with the orthogonal matrix U, and handling arbitrary $k \in \mathbb{N}$ rather then just $k = 2$.

The Distributions G_k and $\mathcal{D}_{U,k}$

Let $N, k \in \mathbb{N}$. First, we define a continuous distribution G_k over $\mathbb{R}^{kN}$ (in which every coordinate will be either a Gaussian random variable or a product of two independent Gaussian random variables), and then derive from it a discrete distribution over $\{-1, 1\}^{kN}$ by taking signs.

The definition of G_k and $\mathcal{D}_{U,k}$ will rely on the N-by-N orthogonal matrix U from the definition of Rorrelation. For $i = 1, \ldots, k-1$ let $X^{(i)} \sim \mathcal{N}(0,1)^N$ and sample all the vectors $X^{(1)}, \ldots, X^{(k-1)}$ independently. Denote by $Y^{(1)}, \ldots, Y^{(k-1)}$ the vectors defined by $Y^{(i)} = U^{\mathsf{T}} \cdot X^{(i)}$. Define $Z^{(1)}, \ldots, Z^{(k)}$ as follows:

1) $Z^{(1)} = X^{(1)}$
2) For $i = 2, \ldots, k-1$, let $Z^{(i)} = Y^{(i-1)} \odot X^{(i)}$ (where $\odot$ denotes point-wise product of two vectors of length N).
3) $Z^{(k)} = Y^{(k-1)}$.

We denote by $Z = (Z^{(1)}, \ldots, Z^{(k)})$ the concatenation of all the k vectors. This defines the distribution G_k over $\mathbb{R}^{kN}$. The distribution $\mathcal{D}_{U,k}$ over $\{\pm 1\}^{kN}$ will simply be the distribution of $\mathrm{sgn}(Z) = (\mathrm{sgn}(Z_1^{(1)}), \mathrm{sgn}(Z_2^{(1)}), \ldots, \mathrm{sgn}(Z_N^{(k)}))$.

$\mathcal{D}_{U,k}$ produces vectors with large Rorrelation:

In section IV-A, we show that $\mathbf{E}_{z \sim \mathcal{D}_{U,k}}[\phi_U(z)] \geq (2/\pi)^{k-1}$. Based on that, a simple Markov's inequality shows that with probability at least 2^{-k}, $\mathcal{D}_{U,k}$ samples a YES-instance for the Rorrelation problem. This will complete Item 2 in the aforementioned scheme, and we are left to prove the third item, which is the core of our argument.

The Core of the Argument: $\mathcal{D}_{U,k}$ is Pseudorandom against Shallow Decision Trees

We are left to prove that for any depth $d = o(N^{2(k-1)/(3k-1)}/\log N)$ decision tree F, we have

$$\left| \mathop{\mathbf{E}}_{z \sim \mathcal{U}_k}[F(z)] - \mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}}[F(z)] \right| \leq o(1/2^k)$$

We call $\left|\mathbf{E}_{z \sim \mathcal{U}_k}[F(z)] - \mathbf{E}_{z \sim \mathcal{D}_{U,k}}[F(z)]\right|$ the **advantage** of F distinguishing between $\mathcal{U}_k$ and $\mathcal{D}_{U,k}$. Intuitively, a small advantage means that F behaves similarly in both cases. To bound the advantage of F, we apply a straightforward Fourier analytical approach, as follows. Since F is defined over the Boolean domain, it can be represented as a

[2] We believe that $\mathcal{D}_{U,k}$ samples YES instances with very high probability, but since this seems technically involved, and since it is not required to complete the proof, we left this question open.

multilinear polynomial, also known as the Fourier transform. That is, we may write

$$F(z) = \sum_{S \subseteq \{1,\ldots,kN\}} \widehat{F}(S) \cdot \prod_{i \in S} z_i.$$

where $\widehat{F}(S)$ are the real-valued Fourier coefficients of F. Observe that $\mathbf{E}_{z \sim \mathcal{U}_k}[F(z)] = \widehat{F}(\emptyset)$, whereas

$$\mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}}[F(z)] = \sum_{S \subseteq \{1,\ldots,kN\}} \widehat{F}(S) \cdot \mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}} \left[\prod_{i \in S} z_i \right].$$

To make our notation shorter, we denote by $\widehat{\mathcal{D}_{U,k}}(S) = \mathbf{E}_{z \sim \mathcal{D}_{U,k}} \left[\prod_{i \in S} z_i \right]$. Thus, the advantage of F can be expressed as

$$\left| \mathop{\mathbf{E}}_{z \sim \mathcal{U}_k}[F(z)] - \mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}}[F(z)] \right| = \left| \sum_{S \subseteq \{1,\ldots,kN\} : S \neq \emptyset} \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right|,$$

which by the triangle inequality is at most

$$\sum_{S \subseteq \{1,\ldots,kN\} : S \neq \emptyset} \left| \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right|. \tag{1}$$

We bound the sum in Eq. (1), by accounting for each degree $\ell \in [kN]$, the contribution of sets of size ℓ to the sum. Namely,

$$\sum_{S \subseteq \{1,\ldots,kN\} : S \neq \emptyset} \left| \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right| = \sum_{\ell=1}^{kN} \sum_{S : |S| = \ell} \left| \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right|.$$

To bound each internal sum $\sum_{S:|S|=\ell} \left| \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right|$ it suffices to show that:

1) For every set $S \subseteq \{1,\ldots,kN\}$ of size ℓ, the coefficient $|\widehat{\mathcal{D}_{U,k}}(S)|$ is sufficiently small.
2) The sum $\sum_{S:|S|=\ell} |\widehat{F}(S)|$ is not too large.

This suffices to bound $\sum_{S:|S|=\ell} \left| \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right|$, by the following the simple inequality

$$\sum_{S:|S|=\ell} \left| \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S) \right| \leq \left(\sum_{S:|S|=\ell} |\widehat{F}(S)| \right) \cdot \left(\max_{S:|S|=\ell} |\widehat{\mathcal{D}_{U,k}}(S)| \right).$$

The proofs of both Parts 1 and 2 in the above plan are technically involved.

Part 1: Moment Bounds on $\mathcal{D}_{U,k}$.: Part 1 boils down to showing that all moments of the distribution $\mathcal{D}_{U,k}$ are sufficiently small, where the bounds improve as the degree increases. This is where the properties of random orthogonal matrices play their role. In particular we are able to show the following bound.

Theorem I.2. *With high probability over the choice of a uniformly random orthogonal matrix $U \in \mathbb{R}^{N \times N}$, for all $\ell \in \{1,\ldots,kN\}$ and all sets $S \subseteq \{1,\ldots,kN\}$ of size ℓ it holds that*

$$|\widehat{\mathcal{D}_{U,k}}(S)| \leq \left(\frac{c \cdot \ell \cdot \log N}{N} \right)^{\ell \cdot (1-1/k)/2} \tag{2}$$

where c is some universal constant. Moreover, $\widehat{\mathcal{D}_{U,k}}(S) = 0$ for all non-empty sets S of size less than k.

For example for $k = \ell = 2$ the theorem shows that any second moment of $\mathcal{D}_{U,2}$ is at most $\widetilde{O}(1/\sqrt{N})$. For any constant k and $\ell = k$, the theorem shows that the k-th moment of $\mathcal{D}_{U,k}$ is at most $\widetilde{O}(1/\sqrt{N^{k-1}})$.

We remark that replacing U with the Hadamard matrix, one gets much worse bounds for high moments, that would not allow to prove better than $\sqrt{N}$ lower bounds on the decision tree depth using our approach. Furthermore, we believe our bounds on $|\widehat{\mathcal{D}_{U,k}}(S)|$ are tight.

Part 2: Level-ℓ Fourier Bounds on F.: We return to Part 2 above, bounding the sum $\sum_{S:|S|=\ell} |\widehat{F}(S)|$ where F is a decision tree of depth d. The work of O'Donnell and Servedio [OS07] obtained a tight $O(\sqrt{d})$ bound for the case $\ell = 1$ (which was later extended by Blais, Tan, and Wan [BTW15] to parity decision trees). This allowed O'Donnell and Servedio to obtain a polynomial-time algorithm for learning monotone decision trees under the uniform distribution. To the best of our knowledge, the question about higher Fourier levels of decision trees was not explicitly raised in the literature prior to this work.

We denote by $L_{1,\ell}(F) = \sum_{S:|S|=\ell} |\widehat{F}(S)|$. Bounds of the quantity $L_{1,\ell}(F)$ were proved for several well-studied classes of Boolean functions such as bounded-width DNF formulas [Man95], bounded depth circuits [Tal17], read-once branching programs [RSV13], [CHRT18], functions with low sensitivity [GSTW16], and low-degree polynomials over finite fields [CHHL18]. Furthermore, it was conjectured in [CHLT19] that stronger classes of Boolean functions, such as $\mathbf{AC^0}[\oplus]$ circuits, have low $L_{1,\ell}(F)$. Moreover, the work of [CHHL18] showed how to generically construct pseudorandom generators assuming only bounds on the $L_{1,\ell}(F)$ of functions in the family (or even assuming only bounds on $L_{1,2}(F)$ [CHLT19]).

The quantity $L_{1,\ell}(F)$ captures the "sparsity" of the Fourier spectrum. Intuitively, this is due to the known fact that the sum of squares $\sum_{S:|S|=\ell} |\widehat{F}(S)|^2$ is at most 1. Hence, having the sum of absolute values small, means that most of the Fourier mass sits on a few coefficients.

We prove new bounds on the $L_{1,\ell}(F)$ of any decision tree F of depth d.

Theorem I.3. *Let F be a decision tree on kN input variables of depth d. Then,*

$$\forall \ell : \sum_{S \subseteq \{1,\ldots,kN\}:|S|=\ell} |\widehat{F}(S)| \leq \sqrt{d^\ell \cdot O(\log kN)^{\ell-1}} \quad (3)$$

In particular, for $\ell = 1$, we match the tight $O(\sqrt{d})$ bound of [OS07]. Moreover, for constant ℓ, our bounds are nearly tight as exhibited by the composition of the Address and the Majority functions (see Section V-E). However, for higher values of ℓ, our bounds get sloppier and we believe that they can be further improved as follows.

Conjecture I.4. *Let F be a decision tree on kN input variables of depth d. Then,*

$$\forall \ell : \sum_{S \subseteq \{1,\ldots,kN\}:|S|=\ell} |\widehat{F}(S)| \leq \sqrt{\binom{d}{\ell} \cdot O(\log kN)^{\ell-1}}$$

We remark that for non-adaptive decision trees, namely juntas, the conjecture holds.[3] Combining the bounds in Eq. (2) with Eq. (3) gives

$$\left(\sum_{S:|S|=\ell} |\widehat{F}(S)| \right) \cdot \left(\max_{S:|S|=\ell} |\widehat{\mathcal{D}_{U,k}}(S)| \right) \leq o(1/2^\ell)$$

for
all $\ell \leq \sqrt{d/\log n}$ and all $d = o(N^{2(k-1)/(3k-1)}/\log N)$. For larger degrees $\ell > \sqrt{d/\log n}$, we use a much simpler bound, $L_{1,\ell}(F) \leq \binom{d}{\ell}$, to obtain similarly

$$\left(\sum_{S:|S|=\ell} |\widehat{F}(S)| \right) \cdot \left(\max_{S:|S|=\ell} |\widehat{\mathcal{D}_{U,k}}(S)| \right) \leq o(1/2^\ell).$$

Summing over all sets of size at least k we get

$$\left| \mathop{\mathbf{E}}_{z \sim \mathcal{U}_k}[F(z)] - \mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}}[F(z)] \right|$$
$$\leq \sum_{\ell=k}^{kN} \left(\sum_{S:|S|=\ell} |\widehat{F}(S)| \right) \cdot \left(\max_{S:|S|=\ell} |\widehat{\mathcal{D}_{U,k}}(S)| \right) \leq o(1/2^k),$$

which completes the proof.

We remark that assuming Conjecture I.4, the same strategy would work for any decision tree of depth at most $N^{1-1/k}/\mathrm{polylog}(N)$.

[3]If F is a Boolean d-junta, then there are at most $\binom{d}{\ell}$ non-zero Fourier coefficients of degree ℓ, thus by Cauchy-Schwarz $\sum_{S:|S|=\ell} |\widehat{F}(S)| \leq \sqrt{\binom{d}{\ell} \cdot \sum_{S:|S|=\ell} |\widehat{F}(S)|^2} \leq \sqrt{\binom{d}{\ell}}$.

B. Digest - Degree and Sparsity

We seek to pinpoint the key differences between the quantum and randomized query models that allowed us to get this separation.

The seminal result of Beals et al. [BBC$^+$01] showed that the acceptance probability of any t query quantum algorithm can be expressed as a degree-$2t$ multilinear polynomial. Thus quantum algorithms making few queries can be approximated by low-degree polynomials. This is also the case for randomized decision trees, as observed by [NS94]. But, if both models are approximated by low-degree polynomials, what is the difference between them?

We suggest sparsity, or more precisely $L_{1,\ell}(\cdot)$, as a measure to separate the two models. This is evident from our proof, which strongly exploits the smallness of $L_{1,\ell}(F)$ for shallow randomized decision trees. On the other hand, one can show that the $L_{1,\ell}(\cdot)$ of quantum algorithms making a few queries could be quite large. As mentioned above, for any quantum algorithm making t queries, there exists a multilinear polynomial p of degree $2t$ capturing the acceptance probability of the algorithm. With this in mind, one can analyze the $L_{1,\ell}(\cdot)$ of p, i.e., the sum of absolute values of the degree ℓ terms in the polynomial p. Indeed, the polynomial that arises from Aaronson-Ambainis algorithm (Claim III.1) is exactly $\frac{1}{2}(1+\phi_U(z^{(1)},\ldots,z^{(k)}))$. Observe that for a random orthogonal matrices U, entries in the matrix U are of magnitude roughly $1/\sqrt{N}$, with high probability, and thus the sum of absolute values of the degree-k coefficients in ϕ_U is quite large, $\widetilde{\Theta}(N^{(k-1)/2})$.

This captures the difference between the models. Indeed, to get such large $L_{1,k}$ measure for randomized decision tree, one needs their depth to be at least $\widetilde{\Omega}(N^{(1-1/k)})$.

We remark however that differences in $L_{1,\ell}(\cdot)$ alone are not sufficient to show a difference between the computational abilities of the two models. Indeed, two polynomials with very similar values on the entire Boolean domain can get very different $L_{1,\ell}$ norms. This is why it is non-trivial to find and prove that a computational task demonstrates these differences. As we show in this paper, the k-Fold Rorrelation problem is such a task.

C. Related Work

We would like to comment about the relation of this work with our prior joint work with Raz [RT19]. In [RT19], a separation between quantum algorithms, making a few queries, and $\mathbf{AC^0}$ circuits was obtained. This, in turn, implied an oracle separation between $\mathbf{BQP}$ and the Polynomial Hierarchy. The question in [RT19] boiled down to whether a distribution similar[4] to $\mathcal{D}_2$ is pseudorandom against $\mathbf{AC^0}$ circuits. While the proof strategy in [RT19] starts similarly to the one laid out here, it takes a sharp

[4]A major difference, though, is that when discretizing the Gaussian distribution, [RT19] used randomized rounding whereas we take signs.

turn early on. Namely, there, the approach of bounding the contribution of each level separately, simply does not work. To overcome this hurdle, one needs to rely on techniques from stochastic calculus, viewing the Gaussian distribution as a result of a random walk that makes many tiny steps, and analyzing each step separately. Surprisingly, the result of [RT19] relies only on bounds on the second-level of the Fourier spectrum of $\mathbf{AC^0}$ (i.e., only bounds on $L_{1,2}(F)$).

Here, we exploit delicate bounds on all levels of the Fourier spectrum of depth-d decision trees (i.e., bounds on $L_{1,\ell}(F)$ for all ℓ, where F is a depth-d decision tree), as well as tight moment bounds on the distribution $\mathcal{D}_{U,k}$. So far, despite initial attempts, we were unable to exploit techniques from stochastic calculus to analyze $\mathbf{E}_{z\sim\mathcal{D}_{U,k}}[F(z)]$. One difficulty arises from the fact that the continuous distribution G_k, which we discretize to get $\mathcal{D}_{U,k}$, involves products of Gaussians, rather than just Gaussians. It seems tempting to wonder whether only a bound on the k-th Fourier level would suffice to analyze k-fold Rorrelation. If so, this would give a completely different proof, with possibly optimal quantitative bounds.

II. Preliminaries

For $N \in \mathbb{N}$, we denote by $[N] = \{1, \ldots, N\}$. We denote by I_N the identity matrix of order N. For $A \in \mathbb{R}^{m\times n}$, we denote by $\|A\|$ the matrix norm given by $\|A\| = \sup_{x\neq 0} \|Ax\|_2/\|x\|_2$.

Quantum Query

A quantum query to an input $z \in \{\pm 1\}^{kN}$ performs the diagonal unitary transformation U_z, defined by $|i, w\rangle \to z_i|i, w\rangle$, where $i \in [kN]$ and w represents the auxiliary workspace that does not participate in the query.

Fourier Representation of Boolean Functions

Let $f : \{\pm 1\}^N \to \mathbb{R}$ be a Boolean function on N variables. The Fourier transform of f is the unique multilinear polynomial that agrees with f on $\{\pm 1\}^N$. Such a polynomial exists and is unique. We write the Fourier transform as

$$f(x) = \sum_{S\subseteq[N]} \widehat{f}(S) \cdot \prod_{i\in S} x_i$$

where $\widehat{f}(S) \in \mathbb{R}$ are the Fourier-coefficients, that could be easily computed from the function f by $\widehat{f}(S) = \mathbf{E}_{x\in\{\pm1\}^N}[f(x) \cdot \prod_{i\in S} x_i]$. Parseval's identity shows that $\mathbf{E}_{x\in\{\pm1\}^N}[f(x)^2] = \sum_{S\subseteq[N]} \widehat{f}(S)^2$. For $\ell \in [N]$, we denote by

$$L_{1,\ell}(f) \triangleq \sum_{S\subseteq[N]:|S|=\ell} |\widehat{f}(S)|.$$

Moments of Distributions

Let D be a distribution over $\{-1, 1\}^N$. For any subset $S \subseteq [N]$ we denote by

$$\widehat{D}(S) = \mathop{\mathbf{E}}_{x\sim D}\left[\prod_{i\in S} x_i\right].$$

III. Quantum Algorithm for the k-fold Rorrelation Problem

Aaronson and Ambainis [AA15] presented an algorithm that solves Forrelation (the special case of Rorrelation when U is the Hadamard matrix) with only $\lceil k/2\rceil$ queries. It is straightforward to extend their algorithm to solve the Rorrelation problem.

Claim III.1. *Let N be a power of 2. Let U be an N-by-N orthogonal matrix. Then, there exists a quantum algorithm making $\lceil k/2\rceil$ quantum queries, whose acceptance probability is $\frac{1}{2} + \frac{1}{2}\phi_U(z^{(1)}, \ldots, z^{(k)})$ where recall that*

$$\phi_U(z^{(1)}, \ldots, z^{(k)}) = \frac{1}{N} \cdot \sum_{i_1,\ldots,i_k} z^{(1)}_{i_1} \cdot U_{i_1,i_2} \cdot z^{(2)}_{i_2} \cdots U_{i_{k-1},i_k} \cdot z^{(k)}_{i_k}.$$

For completeness, we give the proof that naturally extends [AA15, Prop. 6] in Appendix A. Note that Claim III.1 means that the adaptation of Aaronson-Ambainis's algorithm accepts YES-instances with probability at least $\frac{1+2^{-k}}{2}$ and accepts NO-instances with probability at most $\frac{1+2^{-(k+1)}}{2}$.

This can be amplified to a $2/3$ vs. $1/3$ separation as explained next. Denote by ε the difference of the two fractions $\frac{1+2^{-k}}{2} - \frac{1+2^{-(k+1)}}{2}$ and by α their average. By the standard amplification technique of repeating an algorithm for $m = O(1/\varepsilon^2) = O(4^k)$ times and checking whether the number of successful trials exceeds $m \cdot \alpha$, we strengthen the separation between the acceptance probabilities of YES and NO instances to $2/3$ vs. $1/3$.

Corollary III.2. *Let N be a power of 2. Let U be an N-by-N orthogonal matrix. Then, there exists a quantum algorithm making $O(k \cdot 4^k)$ quantum queries, that solves k-fold Rorrelation problem with respect to U with success probability at least $2/3$.*

We remark that while the algorithms mentioned in Claim III.1 or Corollary III.2 make only a few quantum queries, they are not necessarily efficient in terms of running time as they apply an arbitrary orthogonal transformation U to the quantum register. It remains an important open problem to show that one can get similar separations for orthogonal matrices U that can be implementable efficiently, say by quantum circuits with at most $\mathrm{polylog}(N)$ many gates. This boils down to showing the existence of efficiently implementable matrices U that satisfy the pseudorandomness condition in Def. V.5.

IV. The Rorrelation of Vectors Sampled from $\mathcal{D}_{U,k}$ and $\mathcal{U}_k$

In this section, we show that:

1) Vectors $z^{(1)}, \ldots, z^{(k)}$ that are sampled from the distribution $\mathcal{D}_{U,k}$ have expected Rorrelation value at least $(2/\pi)^{k-1}$.
2) Vectors $z^{(1)}, \ldots, z^{(k)}$ that are sampled from the uniform distribution $\mathcal{U}_k$ have expected Rorrelation value 0. Furthermore, the variance of their Rorrelation value is $1/N$, thus it is highly concentrated around 0.

This shows that the algorithm from Claim III.1 distinguishes between $\mathcal{D}_{U,k}$ and $\mathcal{U}_k$ as its acceptance probability differs by at least $\frac{1}{2} \cdot (2/\pi)^{k-1}$ between the two cases. Then, in Section V we show that bounded-depth randomized decision trees fail to distinguish between $\mathcal{D}_{U,k}$ and $\mathcal{U}_k$, for most orthogonal matrices U. This, in turn, leads to the conclusion that bounded-depth randomized decision trees fail to solve the Rorrelation problem, for most orthogonal matrices U.

A. The Rorrelation of Vectors Sampled from $\mathcal{D}_{U,k}$

Claim IV.1. *Let U be an N-by-N orthogonal matrix. Then,*

$$\mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}} [\phi_U(z^{(1)}, \ldots, z^{(k)})] \geq (2/\pi)^{k-1}.$$

Proof: The expectation of ϕ_U on the distribution $\mathcal{D}_{U,k}$ is

$$\begin{aligned}
&\mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}} [\phi_U(z^{(1)}, \ldots, z^{(k)})] \\
&= \mathop{\mathbf{E}}_{Z \sim G_k} [\phi_U(\mathrm{sgn}(Z^{(1)}), \ldots, \mathrm{sgn}(Z^{(k)}))] \\
&= \frac{1}{N} \cdot \sum_{i_1, \ldots, i_k} \mathbf{E} \begin{bmatrix} \mathrm{sgn}(X_{i_1}^{(1)}) \cdot U_{i_1,i_2} \cdot \mathrm{sgn}(Y_{i_2}^{(1)}) \cdot \\ \mathrm{sgn}(X_{i_2}^{(2)}) \cdot U_{i_2,i_3} \cdots \\ \mathrm{sgn}(X_{i_{k-1}}^{(k-1)}) \cdot U_{i_{k-1},i_k} \cdot \mathrm{sgn}(Y_{i_k}^{(k-1)}) \end{bmatrix} \\
&= \frac{1}{N} \cdot \sum_{i_1, \ldots, i_k} \begin{pmatrix} \mathbf{E}[\mathrm{sgn}(X_{i_1}^{(1)}) \cdot U_{i_1,i_2} \cdot \mathrm{sgn}(Y_{i_2}^{(1)})] \\ \cdots \\ \mathbf{E}[\mathrm{sgn}(X_{i_{k-1}}^{(k-1)}) U_{i_{k-1},i_k} \cdot \mathrm{sgn}(Y_{i_k}^{(k-1)})] \end{pmatrix}.
\end{aligned}$$

In the following Lemma IV.2, we show that for any $j \in \{1, \ldots, k-1\}$ and any $i_j \in [N]$ and $i_{j+1} \in [N]$ the expectation of

$$\mathbf{E}[\mathrm{sgn}(X_{i_j}^{(j)}) \cdot U_{i_j,i_{j+1}} \cdot \mathrm{sgn}(Y_{i_{j+1}}^{(j)})] \geq \frac{2}{\pi} \cdot U_{i_j,i_{j+1}}^2,$$

relying on the fact that the covariance of $X_{i_j}^{(j)}$ and $Y_{i_{j+1}}^{(j)}$ equals $U_{i_j,i_{j+1}}$. This gives

$$\begin{aligned}
&\mathop{\mathbf{E}}_{z \sim \mathcal{D}_{U,k}} [\phi_U(z^{(1)}, \ldots, z^{(k)})] \\
&\geq \frac{1}{N} \cdot \sum_{i_1, \ldots, i_k} U_{i_1,i_2}^2 \cdots U_{i_{k-1},i_k}^2 \cdot \left(\frac{2}{\pi}\right)^{k-1} \\
&= \frac{1}{N} \cdot N \cdot \left(\frac{2}{\pi}\right)^{k-1} \qquad \text{(since } U \text{ is orthogonal)} \\
&= \left(\frac{2}{\pi}\right)^{k-1}.
\end{aligned}$$

■

Lemma IV.2. *Let $\rho \in [-1, 1]$. Let (X, Y) be two-dimensional multi-variate Gaussian distribution with zero-means and covariance matrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Then,*

$$\rho \cdot \mathbf{E}[\mathrm{sgn}(X) \cdot \mathrm{sgn}(Y)] \geq \frac{2}{\pi} \cdot \rho^2.$$

Proof: Let $u_1 = (1, 0)$ and $u_2 = (\rho, \sqrt{1-\rho^2})$. We can retrieve such a distribution (on (X, Y)) by considering two independent standard Gaussians $Z = (Z_1, Z_2)$ and taking $X = \langle Z, u_1 \rangle$ and $Y = \langle Z, u_2 \rangle$. Thus, the probability that $\mathrm{sgn}(X) = \mathrm{sgn}(Y)$ is the same as the probability over a random $Z = (Z_1, Z_2)$ that $\mathrm{sgn}(\langle Z, u_1 \rangle) = \mathrm{sgn}(\langle Z, u_2 \rangle)$, which is the same if we sample Z according to the uniform distribution on the sphere. The latter probability is exactly $1 - \alpha/\pi$ where α is the angle between u_1 and u_2. Thus, the probability is $1 - \arccos(\rho)/\pi$, and

$$\begin{aligned}
\mathbf{E}[\mathrm{sgn}(X) \cdot \mathrm{sgn}(Y)] &= 2\,\mathbf{Pr}[\mathrm{sgn}(X) = \mathrm{sgn}(Y)] - 1 \\
&= 1 - 2\arccos(\rho)/\pi.
\end{aligned}$$

For $\rho \geq 0$ we have $\mathbf{E}[\mathrm{sgn}(X) \cdot \mathrm{sgn}(Y)] \geq \frac{2}{\pi}\rho$ and for $\rho \leq 0$ we get $\mathbf{E}[\mathrm{sgn}(X) \cdot \mathrm{sgn}(Y)] \leq \frac{2}{\pi}\rho$. Thus, in both cases, $\rho \cdot \mathbf{E}[\mathrm{sgn}(X) \cdot \mathrm{sgn}(Y)] \geq \frac{2}{\pi}\rho^2$. ■

B. The Rorrelation of Vectors Sampled from $\mathcal{U}_k$

We begin with the following simple claim.

Claim IV.3. *Let U be an N-by-N orthogonal matrix. Then,*

$$\mathop{\mathbf{E}}_{z \sim \mathcal{U}_k} \left[\phi_U(z^{(1)}, \ldots, z^{(k)})\right] = 0$$

Proof:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{z \sim \mathcal{U}_k} [\phi_U(z^{(1)}, \ldots, z^{(k)})] \\
&= \frac{1}{N} \cdot \sum_{i_1, \ldots, i_k} \mathbf{E}[z_{i_1}^{(1)}] \cdot U_{i_1,i_2} \cdot \mathbf{E}[z_{i_2}^{(2)}] \cdot \cdots \cdot \mathbf{E}[z_{i_k}^{(k)}] = 0.
\end{aligned}$$

■

Furthermore, we show that for $z^{(1)}, \ldots, z^{(k)}$ drawn from the uniform distribution, the value of $\phi_U(z^{(1)}, \ldots, z^{(k)})$ is concentrated around 0. To show that it suffices to bound the variance of $\phi_U(z^{(1)}, \ldots, z^{(k)})$ under the uniform distribution, as we do next.

Claim IV.4. *Let U be an N-by-N orthogonal matrix. Then,*

$$\mathbf{Var}_{z\sim\mathcal{U}_k}\left[\phi_U(z^{(1)},\ldots,z^{(k)})\right] = \mathop{\mathbf{E}}_{z\sim\mathcal{U}_k}\left[(\phi_U(z^{(1)},\ldots,z^{(k)}))^2\right] = 1/N.$$

Proof: Since ϕ_U is multilinear we can apply Parseval's identity to get

$$\mathbf{Var}_{z\sim\mathcal{U}_k}\left[\phi_U(z^{(1)},\ldots,z^{(k)})\right] = \mathop{\mathbf{E}}_{z\sim\mathcal{U}_k}\left[\left(\frac{1}{N}\cdot\sum_{i_1,\ldots,i_k} z^{(1)}_{i_1}\cdot U_{i_1,i_2}\cdot z^{(2)}_{i_2}\cdots z^{(k)}_{i_k}\right)^2\right] = \frac{1}{N^2}\sum_{i_1,\ldots,i_k} U^2_{i_1,i_2}U^2_{i_2,i_3}\cdots U^2_{i_{k-1},i_k} = 1/N.$$

■

Overall, we get that a vector z, drawn from the uniform distribution, satisfies $|\phi_U(z^{(1)},\ldots,z^{(k)})| \le 2^{-(k+1)}$ with high probability (at least $1-4^{(k+1)}/N$) .

V. $\mathcal{D}_{U,k}$ IS PSEUDORANDOM FOR RANDOMIZED DECISION TREES

A. Fourier Growth of Decision Trees

We start by stating two bounds on the Fourier coefficients of decision trees. These bounds capture the fact that the Fourier spectrum of deterministic and randomized decision trees is "sparse". More precisely, we bound the sum of absolute values of coefficients of degree ℓ, and since the sum of squares is at most 1, this means that within the ℓ-th level, the Fourier mass is concentrated on a small fraction of the coefficients.

Theorem V.1 (Level-ℓ Inequality for Decision Trees – Version 1)**.** *Let f be a (deterministic) decision tree of depth d over m variables $x_1,\ldots,x_m$. Then,*

$$\forall \ell\in\{0,1,\ldots,d\}: \quad L_{1,\ell}(f)\le\sqrt{O(d)^\ell\cdot O(\log m)^{\ell-1}}$$

The above inequality is tight for small values of ℓ. In particular, it gives a $O(\sqrt{d})$ bound on the first level – a result that was previously obtained by [OS07], [BTW15] and is known to be tight (see Section V-E for examples demonstrating its tightness). For higher values of ℓ though, the inequality gets sloppier, and for $\ell\ge\Omega(\sqrt{d/\log n})$ a much simpler argument gives better bounds.

Claim V.2 (Level-ℓ Inequality for Decision Trees - Version 2)**.** *Let f be a (deterministic) decision tree of depth d over m variables $x_1,\ldots,x_m$. Then,*

$$\forall \ell\in\{0,1,\ldots,d\}: \quad L_{1,\ell}(f)\le\binom{d}{\ell}$$

We defer the proofs of Theorem V.1 and Claim V.2 to the full version [Tal19]. We get the following corollary.

Corollary V.3. *Let F be a randomized decision tree of depth d over m variables $x_1,\ldots,x_m$. Then,*

$$\forall \ell\in\{0,1,\ldots,d\}: \quad L_{1,\ell}(F)\le\sqrt{O(d)^\ell\cdot O(\log m)^{\ell-1}} \tag{4}$$

and

$$\forall \ell\in\{0,1,\ldots,d\}: \quad L_{1,\ell}(F)\le\binom{d}{\ell} \tag{5}$$

Proof: A randomized decision tree is a convex combination of deterministic decision trees. Since $L_{1,\ell}(\cdot)$ is convex, the bounds follow. ■

We conjecture that the right bounds are better for higher levels:

Conjecture V.4 (Conjectured Level-ℓ Inequality for Decision Trees)**.** *Let f be a (deterministic/randomized) decision tree of depth d over m variables $x_1,\ldots,x_m$. Then,*

$$\forall \ell\in\{0,1,\ldots,d\}: \quad L_{1,\ell}(f)\le\sqrt{\binom{d}{\ell}\cdot O(\log m)^{\ell-1}}$$

B. Moment Bounds on $\mathcal{D}_{U,k}$

Good Orthogonal Matrices.: We define a pseudorandomness property of orthogonal matrices, from which we will deduce moment bounds on the distribution $\mathcal{D}_{U,k}$.

Definition V.5 (Good Orthogonal Matrices)**.** *Let U be an N-by-N orthogonal matrix. We say that U is* **good** *if for all $k,\ell\in[N]$, any k-by-ℓ sub-matrix W of U satisfies $\|W\|\le\sqrt{\frac{100(k+\ell)\ln N}{N}}$.*

It is not difficult to see that the Hadamard matrix is not good. For example, the Hadamard matrix has a $\sqrt{N}\times\sqrt{N}$ sub-matrix W, all whose entries equal $+1/\sqrt{N}$, and thus the norm of W equals $1\gg\sqrt{\frac{100(\sqrt{N}+\sqrt{N})\ln N}{N}}$. On the other hand, we prove that most orthogonal matrices are good.

Lemma V.6 (Most Orthogonal Matrices are Good)**.** *Let U be a random orthogonal N-by-N matrix. Then, with high probability over the choice of U, U is good.*

Furthermore, we show that whenever U is good, we get moment bounds on the corresponding distribution $\mathcal{D}_{U,k}$, defined with respect to U.

Lemma V.7 (Moment Bounds for Good Matrices)**.** *Suppose that U is a good orthogonal matrix and $\mathcal{D}_{U,k}$ is defined with respect to U. Then, there exists a universal constant c, such that for any $\emptyset\ne S\subseteq[kN]$,*

$$|\widehat{\mathcal{D}_{U,k}}(S)|\le\left(\frac{c\cdot|S|\cdot\log N}{N}\right)^{|S|\cdot(1-1/k)/2},$$

and for any non-empty set S of size less than k, we have $\widehat{\mathcal{D}_{U,k}}(S)=0$.

We defer the proofs of both lemmata to the full version [Tal19].

C. Pseudorandomness of $\mathcal{D}_{U,k}$

Theorem V.8. *Suppose that U is a good orthogonal matrix and $\mathcal{D}_{U,k}$ is defined with respect to U. Let F be a randomized decision tree of depth d over kN variables. Suppose that $d = o(N^{2(k-1)/(3k-1)}/\log(kN))$. Then,*

$$\mathbf{E}[F(\mathcal{U}_k) - F(\mathcal{D}_{U,k})] \leq \sqrt{\frac{O(d \cdot \log(kN))^{(3k-1)/2}}{N^{k-1}}}$$

Proof: We have

$$\begin{aligned}
&|\,\mathbf{E}[F(\mathcal{U}_k) - F(\mathcal{D}_{U,k})]| \\
&= \big|\sum_{S \neq \emptyset} \widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S)\big| \\
&\leq \sum_{\ell=k}^{kN} \sum_{|S|=\ell} |\widehat{F}(S) \cdot \widehat{\mathcal{D}_{U,k}}(S)| \\
&\leq \sum_{\ell=k}^{kN} \left(\max_{|S|=\ell} |\widehat{\mathcal{D}_{U,k}}(S)|\right) \cdot \sum_{|S|=\ell} |\widehat{F}(S)| \\
&\leq \sum_{\ell=k}^{kN} \left(\frac{c \cdot \ell \cdot \ln N}{N}\right)^{\ell \cdot (1-1/k)/2} \cdot L_{1,\ell}(F)
\end{aligned}$$

Now we break the right hand side above to two sub-sums:

1) for $\ell \leq \sqrt{d/\log(kN)}$ we will use the bounds on $L_{1,\ell}(F)$ from Eq. (4).
2) for $\ell > \sqrt{d/\log(kN)}$ we will use the bounds on $L_{1,\ell}(F)$ from Eq. (5).

That is, we bound the lower order terms by

$$\begin{aligned}
&\sum_{\ell=k}^{\sqrt{d/\log(kN)}} \left(\frac{c \cdot \ell \cdot \ln N}{N}\right)^{\ell \cdot (1-1/k)/2} \cdot L_{1,\ell}(F) \\
&\leq \sum_{\ell=k}^{\sqrt{d/\log(kN)}} \left(\frac{c \cdot \ell \cdot \ln N}{N}\right)^{\ell \cdot (1-1/k)/2} \cdot O(d \cdot \log(kN))^{\ell/2} \\
&\leq \sum_{\ell=k}^{\sqrt{d/\log(kN)}} O\left(\frac{(d \cdot \log(kN))^{1+(1-1/k)/2}}{N^{1-1/k}}\right)^{\ell/2} \\
&\leq O\left(\frac{(d \cdot \log(kN))^{1+(1-1/k)/2}}{N^{1-1/k}}\right)^{k/2} \\
&= \sqrt{\frac{O(d \cdot \log(kN))^{(3k-1)/2}}{N^{k-1}}}
\end{aligned}$$

where in the last inequality, the assumption that $d = o(N^{2(k-1)/(3k-1)}/\log(kN))$ is used to deduce that this is a decreasing geometric progression. We bound the higher order terms by

$$\begin{aligned}
&\sum_{\ell > \sqrt{d/\log(kN)}} \left(\frac{c \cdot \ell \cdot \ln N}{N}\right)^{\ell \cdot (1-1/k)/2} \cdot L_{1,\ell}(F) \\
&\leq \sum_{\ell > \sqrt{d/\log(kN)}} \left(\frac{c \cdot \ell \cdot \ln N}{N}\right)^{\ell \cdot (1-1/k)/2} \cdot \binom{d}{\ell} \\
&\leq \sum_{\ell > \sqrt{d/\log(kN)}} \left(\left(\frac{c \cdot \ell \cdot \ln N}{N}\right)^{(1-1/k)/2} \cdot \frac{e \cdot d}{\ell}\right)^{\ell} \\
&\leq \sum_{\ell > \sqrt{d/\log(kN)}} O\left(\left(\frac{\log N}{N}\right)^{(1-1/k)/2} \cdot \frac{d}{\ell^{1-(1-1/k)/2}}\right)^{\ell} \\
&\leq \sum_{\ell > \sqrt{d/\log(kN)}} O\left(\frac{(d \cdot \log(kN))^{1+(1-1/k)/2}}{N^{1-1/k}}\right)^{\ell/2} \\
&\leq O\left(\frac{(d \cdot \log(kN))^{1+(1-1/k)/2}}{N^{1-1/k}}\right)^{k/2} \\
&= \sqrt{\frac{O(d \cdot \log(kN))^{(3k-1)/2}}{N^{k-1}}} \,.
\end{aligned}$$

■

Similarly, we show that assuming Conjecture I.4, for any good U, the distribution $\mathcal{D}_{U,k}$ is pseudorandom against any depth $N^{1-1/k}/\mathrm{polylog}(N)$ decision tree. (We defer the proof to the full version [Tal19, Appendix B].)

D. Shallow Randomized Decision Trees Cannot Solve the k-fold Rorrelation Problem

We prove the following lower bound on the depth of randomized decision trees solving the k-fold Rorrelation Problem.

Theorem V.9. *Let U be a good orthogonal N-by-N matrix. Let $k \geq 2$ be such that $16 \cdot 8^k \leq N$. Suppose that F is a randomized decision tree of depth d solving the k-fold Rorrelation problem with success probability at least $2/3$. Then, $d \geq \Omega(N^{2(k-1)/(3k-1)}/(k \log(kN)))$.*

Towards proving Theorem V.9, we show that $\frac{1}{2}\mathcal{U}_k + \frac{1}{2}\mathcal{D}_{U,k}$ is a somewhat hard-distribution for any depth-d randomized decision trying to solve the Rorrelation problem, as long as $d = o(N^{2(k-1)/(3k-1)}/\log(kN))$.

Claim V.10. *Assume that $16 \cdot 8^k \leq N$, and U is a good orthogonal N-by-N matrix. Let F be a randomized decision tree of depth $d = o(N^{2(k-1)/(3k-1)}/\log(kN))$. Then,*

$$\mathbf{Pr}\left[\begin{array}{r} z \text{ is legal input to Rorrelation} \\ \text{and } F \text{ misclassifies } z \end{array}\right] \geq \frac{1}{8} \cdot 2^{-k} \qquad (6)$$

where the probability is taken over the randomness of $z \sim \frac{1}{2}\mathcal{U}_k + \frac{1}{2}\mathcal{D}_{U,k}$ and the internal randomness of F.

Before proving Claim V.10, we show that it implies Theorem V.9.

Proof of Thm. V.9: Suppose that F is a depth d randomized decision tree with success probability at least $2/3$. Then, one can amplify the success probability to at least $1-\frac{1}{10}\cdot 2^{-k}$, by running F sequently $\Theta(k)$ many times and taking the majority vote. Thus, we get a randomized decision tree F' of depth $d'=\Theta(d\cdot k)$ such that

- On any YES instance, F' accepts with probability at least $1-\frac{1}{10}\cdot 2^{-k}$.
- On any NO instance, F' accepts with probability at most $\frac{1}{10}\cdot 2^{-k}$.

In particular, Eq. (6) does not hold for F', which by Claim V.10 implies $d' \geq \Omega(N^{2(k-1)/(3k-1)}/\log(kN))$. Recalling that $d'=\Theta(d\cdot k)$ completes the proof. ∎

Proof of Claim V.10: Assume by contradiction that this is not the case. Then, there exists a randomized decision tree F, with

$$\mathbf{Pr}\left[\begin{matrix}z \text{ is legal input to Rorrelation}\\ \text{and } F \text{ misclassifies } z\end{matrix}\right] < \frac{1}{8}\cdot 2^{-k}.$$

By averaging, there exists a deterministic decision tree f with

$$\mathbf{Pr}_{z\sim\frac{1}{2}\mathcal{U}_k+\frac{1}{2}\mathcal{D}_{U,k}}\left[\begin{matrix}z \text{ is legal input to Rorrelation}\\ \text{and } F \text{ misclassifies } z\end{matrix}\right] < \frac{1}{8}\cdot 2^{-k}.$$

In particular, this means that on the uniform distribution

$$\mathbf{Pr}_{z\sim\mathcal{U}_k}\left[\begin{matrix}z \text{ is legal input to Rorrelation}\\ \text{and } F \text{ misclassifies } z\end{matrix}\right] \leq \frac{1}{4}\cdot 2^{-k}$$

and on the distribution $\mathcal{D}_{U,k}$ we have

$$\mathbf{Pr}_{z\sim\mathcal{D}_{U,k}}\left[\begin{matrix}z \text{ is legal input to Rorrelation}\\ \text{and } F \text{ misclassifies } z\end{matrix}\right] \leq \frac{1}{4}\cdot 2^{-k}.$$

We will show that f distinguishes between $\mathcal{D}_{U,k}$ and $\mathcal{U}_k$ which will be a contradiction to Theorem V.8.

For $z\sim\mathcal{U}_k$ we have that with probability at least $1-4^{(k+1)}/N$, $|\phi_U(z)|\leq 2^{-(k+1)}$. This is a consequence of the concentration inequality we got in Claim IV.4 stating that $\mathbf{E}_{z\sim\mathcal{U}_k}[\phi_U(z)^2]=1/N$. Thus, with probability at least $1-4^{(k+1)}/N$ we have that z is a NO instance to the Rorrelation problem, and by the assumption on f with probability at least $1-4^{(k+1)}/N-\frac{1}{4}\cdot 2^{-k}$ it answers NO. That is,

$$\mathop{\mathbf{E}}_{z\sim U_k}[f(z)] \leq \frac{4^{k+1}}{N}+\frac{1}{4}\cdot 2^{-k}. \tag{7}$$

For $z\sim\mathcal{D}_{U,k}$ we have that $\mathbf{E}_{z\sim\mathcal{D}_{U,k}}[\phi_U(z)] \geq (2/\pi)^{k-1} > 2^{-(k-1)}$. Since $|\phi_U(z)|\leq 1$ for all binary vectors, this means that, for $z\sim\mathcal{D}_{U,k}$, with probability at least 2^{-k}, we have $\phi_U(z)\geq 2^{-k}$ (as otherwise the expectation would be less than $2^{-(k-1)}$). Put differently, when sampling from $\mathcal{D}_{U,k}$ with probability at least 2^{-k} we get a YES instance for Rorrelation. By that f errs on at most $\frac{1}{4}\cdot 2^{-k}$ of the probability mass of $\mathcal{D}_{U,k}$, it means that

$$\mathop{\mathbf{E}}_{z\sim\mathcal{D}_{U,k}}[f(z)] \geq 2^{-k}-\frac{1}{4}\cdot 2^{-k}. \tag{8}$$

Combining Equations (7) and (8), we get that

$$\mathbf{E}[f(\mathcal{D}_{U,k})]-\mathbf{E}[f(\mathcal{U}_k)] \geq \frac{1}{2}\cdot 2^{-k}-\frac{4^{k+1}}{N} \geq \frac{1}{4}\cdot 2^{-k}, \tag{9}$$

where in the last inequality we used the assumption $N\geq 16\cdot 8^k$. On the other hand, Theorem V.8 shows that

$$\begin{aligned}&\mathbf{E}[f(\mathcal{D}_{U,k})]-\mathbf{E}[f(\mathcal{U}_k)]\\ &\leq \sqrt{\frac{O(d\cdot\log(kN))^{(3k-1)/2}}{N^{k-1}}} \leq o(2^{-k}).\end{aligned} \tag{10}$$

where in the last inequality we used the assumption that $d=o(N^{2(k-1)/(3k-1)}/\log(kN))$. This yields a contradiction, completing the proof. ∎

E. Lower Bounds on the $L_{1,\ell}$ of depth-d decision trees

We present examples demonstrating the tightness of our bounds on the $L_{1,\ell}(\cdot)$ of depth-d decision trees, for small ℓ. In addition, our latter two examples show that one cannot extend the bound $L_{1,\ell}(f)\leq\sqrt{\binom{d}{\ell}}$ from the non-adaptive case (i.e., d-juntas) to the adaptive case (i.e., depth-d decision trees). That is, we show that one must incur a multiplicative factor of roughly $\sqrt{(\log n)^{\ell-1}}$ going from the non-adaptive to the adaptive case.

Example V.11. *The Majority of d variables can be computed by a depth d decision tree. This function has $L_{1,\ell}(\mathsf{MAJ}_d)=\frac{1}{\mathrm{poly}(\ell)}\cdot\sqrt{\binom{d}{\ell}}$ for odd ℓ (cf. [O'D14, Chapter 5.3]).*

Example V.12. *The address function on $n=2^d+d$ variables, denoted by Add_d can be computed by a depth $d+1$ decision tree. In the address function we divide the input to two parts: the first d bits $(x_1,\ldots,x_d)$, called the "index", and the latter 2^d bits $(y_1,\ldots,y_{2^d})$, called the "array". We treat x as representing an index between 1 and 2^d that points to the array, and return the coordinate y_x. It is easy to see that $L_{1,\ell}(\mathsf{Add}_d)=\binom{d}{\ell-1}$ exactly. This may seem to rule out any significant improvement over the simple upper bound given in Claim **??**, namely, $\binom{d}{\ell}$. Note, however, that in the address function $d=\lfloor\log n\rfloor$, so in fact this example is consistent with an asymptotic behavior of $\sqrt{\binom{d}{\ell}}\cdot O(\log n)^{\ell-2}$.*

Example V.13. *Let $D=2d$. Take the address function Add_d and replace each variable in the array part with the Majority function on d distinct new variables. Denote the resulting function by f. Then, f is a function on $d+2^d\cdot d$ variables, which can be computed by a depth-$2d$ decision tree, and has $L_{1,\ell}(f)\geq\Omega(\sqrt{d}\cdot\binom{d}{\ell-1})$. Again, as in the previous*

example $d = \Theta(\log n)$ so the behavior is consistent with a $\sqrt{\binom{d}{\ell} \cdot O(\log n)^{\ell-1}}$ bound.

VI. Open Questions

We would like to highlight several open questions that were mentioned throughout the manuscript. The first is stated as Conjecture I.4, namely what are the tight bounds on the $L_{1,\ell}(\cdot)$ of shallow decision trees? Our conjectured bounds would imply a $\widetilde{\Omega}(N^{1-1/k})$ lower bound on the randomized query complexity of the k-fold Rorrelation problem, which would be tight due to the upper bounds in [AA15].

The second question asks whether one can use tools from stochastic calculus to analyze $\mathbf{E}[f(\mathcal{D}_{U,k})] - \mathbf{E}[f(\mathcal{U}_k)]$. Such analysis could potentially rely only on level-k bounds on the Fourier spectrum of f (and its restrictions) as done in [RT19] for 2-fold Forrelation.

The third question asks whether one can exhibit an explicit family of orthogonal matrices $\{U_N\}_N$ for infinitely many input lengths N, such that (1) U_N can be implemented by $\mathrm{polylog}(N)$ size quantum circuits and (2) U_N are good orthogonal matrices as in Definition V.5. Our current separation uses random orthogonal matrices, that are non-explicit, and cannot be implemented efficiently.

Acknowledgment

I would like to thank Scott Aaronson, Rotem Arnon-Friedman, Adam Bouland, Uma Girish, Tarun Kathuria, Chinmay Nirkhe, Prasad Raghavendra, Ran Raz, Li-Yang Tan, and Umesh Vazirani for very helpful discussions.

References

[AA15] S. Aaronson and A. Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *STOC*, pages 307–316, 2015.

[Aar10] S. Aaronson. BQP and the polynomial hierarchy. In *STOC*, pages 141–150, 2010.

[ABK16] S. Aaronson, S. Ben-David, and R. Kothari. Separations in query complexity using cheat sheets. In *STOC*, pages 863–876, 2016.

[BBBV97] C. H. Bennett, E. Bernstein, G. Brassard, and U. V. Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, 1997.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[BdW02] H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[BFNR08] H. Buhrman, L. Fortnow, I. Newman, and H. Röhrig. Quantum property testing. *SIAM J. Comput.*, 37(5):1387–1400, 2008.

[BTW15] E. Blais, L. Tan, and A. Wan. An inequality for the fourier spectrum of parity decision trees. *CoRR*, abs/1506.01055, 2015.

[BV97] E. Bernstein and U. V. Vazirani. Quantum complexity theory. *SIAM J. Comput.*, 26(5):1411–1473, 1997.

[CCD+03] A. M. Childs, R. Cleve, E. Deotto, E. Farhi, S. Gutmann, and D. A. Spielman. Exponential algorithmic speedup by a quantum walk. In *STOC*, pages 59–68, 2003.

[CHHL18] E. Chattopadhyay, P. Hatami, K. Hosseini, and S. Lovett. Pseudorandom generators from polarizing random walks. In *CCC*, volume 102 of *LIPIcs*, pages 1:1–1:21. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

[CHLT19] E. Chattopadhyay, P. Hatami, S. Lovett, and A. Tal. Pseudorandom generators from the second fourier level and applications to AC0 with parity gates. In *ITCS*, pages 22:1–22:15, 2019.

[CHRT18] E. Chattopadhyay, P. Hatami, O. Reingold, and A. Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *STOC*, pages 363–375. ACM, 2018.

[DJ92] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.

[Gro96] L. K. Grover. A fast quantum mechanical algorithm for database search. In *STOC*, pages 212–219, 1996.

[GSTW16] P. Gopalan, R. A. Servedio, A. Tal, and A. Wigderson. Degree and sensitivity: tails of two distributions. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:69, 2016.

[Man95] Y. Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *J. Comput. Syst. Sci.*, 50(3):543–550, 1995.

[NS94] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[O'D14] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[OS07] R. O'Donnell and R. A. Servedio. Learning monotone decision trees in polynomial time. *SIAM J. Comput.*, 37(3):827–844, 2007.

[RSV13] O. Reingold, T. Steinke, and S. Vadhan. Pseudorandomness for regular branching programs via Fourier analysis. In *APPROX-RANDOM*, pages 655–670. 2013.

[RT19] R. Raz and A. Tal. Oracle separation of BQP and PH. In *STOC*, pages 13–23, 2019.

[Sho97] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. on Computing*, 26(5):1484–1509, 1997.

[Sim97] D. R. Simon. On the power of quantum computation. *SIAM J. Comput.*, 26(5):1474–1483, 1997.

[Tal17] A. Tal. Tight bounds on the fourier spectrum of AC0. In *CCC*, pages 15:1–15:31, 2017.

[Tal19] A. Tal. Towards Optimal Separations between Quantum and Randomized Query Complexities. *Electronic Colloquium on Computational Complexity (ECCC)*, 26:179, 2019.

APPENDIX

A. Proof of Claim III.1

Proof: Suppose that N is a power of two, and denote by $n = \log_2(N)$. We follow the algorithm suggested by Aaronson and Ambainis [AA15] but replace the Hadamard transform (in some of the places) with the orthogonal transform U.

Let $H^{\otimes n}$ be the Hadamard transform on $\mathbb{R}^N$, let U be the orthogonal transform on $\mathbb{R}^N$ from the definition of k-fold Rorrelation. For $i \in [k]$ let $U_{z^{(i)}}$ be the query transformation that maps $|j\rangle$ to $z_j^{(i)} \cdot |j\rangle$ for all $j \in [N]$ (recall that $z^{(i)} \in \{-1,1\}^N$, thus this is a unitary transformation).

We start with the initial state $|0\rangle^{\otimes n}$, in addition to a control qubit in the state $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$. Then, conditioned on the control qubit being $|0\rangle$, we apply the following sequence of operations to the initial state:

$$H^{\otimes n} \to U_{z^{(1)}} \to U^{\mathsf{T}} \to U_{z^{(2)}} \to \cdots U^{\mathsf{T}} \to U_{z^{(\lceil k/2 \rceil)}} \to U^{\mathsf{T}}$$

Meanwhile, conditioned on the control qubit being $|1\rangle$, we apply the following sequence of operations:

$$H^{\otimes n} \to U_{z^{(k)}} \to U \to U_{z^{(k-1)}} \to \cdots U \to U_{z^{(\lceil k/2 \rceil+1)}}$$

Finally, we measure the control qubit in the $\{|+\rangle, |-\rangle\}$ basis, and accept if and only if we find it in the state $|+\rangle$.

It remains to show that the acceptance probability equals $\frac{1+\phi_U(z^{(1)},\ldots,z^{(k)})}{2}$, as we do next:

- Conditioned on the control qubit being $|0\rangle$, the quantum state can be written in vector form as
$$\vec{a} = U^{\mathsf{T}} \cdot U_{z^{(\lceil k/2 \rceil)}} \cdot U^{\mathsf{T}} \cdots U^{\mathsf{T}} \cdot U_{z^{(2)}} \cdot U^{\mathsf{T}} \cdot \vec{v}_{z^{(1)}}$$
where $\vec{v}_{z^{(1)}}$ is the N-dimensional vector with i-th entry $\frac{1}{\sqrt{N}} \cdot z_i^{(1)}$.
- Conditioned on the control qubit being $|1\rangle$, the quantum state can be written in vector form as
$$\vec{b} = U_{z^{(\lceil k/2 \rceil+1)}} \cdot U \cdots U \cdot U_{z^{(k-1)}} \cdot U \cdot \vec{v}_{z^{(k)}}$$
where $\vec{v}_{z^{(k)}}$ is the N-dimensional vector with i-th entry $\frac{1}{\sqrt{N}} \cdot z_i^{(k)}$.

Overall, our combined quantum state is

$$\frac{1}{\sqrt{2}} \sum_{i=1}^{N} a_i \cdot |i\rangle|0\rangle + \frac{1}{\sqrt{2}} \sum_{i=1}^{N} b_i \cdot |i\rangle|1\rangle$$
$$= \frac{1}{2} \sum_{i=1}^{N} (a_i + b_i) \cdot |i\rangle|+\rangle + \frac{1}{2} \sum_{i=1}^{N} (a_i - b_i) \cdot |i\rangle|-\rangle$$

Measuring the control bit in the $\{|+\rangle, |-\rangle\}$ basis yields $|+\rangle$ with probability

$$\frac{1}{4} \sum_{i=1}^{N} (a_i + b_i)^2 = \frac{1}{4} \left(\|a\|_2^2 + \|b\|_2^2 + 2\langle \vec{a}, \vec{b} \rangle \right)$$

Observe that both $\vec{a}$ and $\vec{b}$ are unit vectors, since they are generated by applying orthogonal matrices to the unit vectors $\vec{v}_{z^{(1)}}$ and $\vec{v}_{z^{(k)}}$, correspondingly. Furthermore, observe that

$$\begin{aligned}
\langle \vec{a}, \vec{b} \rangle &= \vec{a}^{\mathsf{T}} \cdot \vec{b} \\
&= \left(\vec{v}_{z^{(1)}}^{\mathsf{T}} \cdot U \cdot U_{z^{(2)}} \cdot U \cdots U \cdot U_{z^{(\lceil k/2 \rceil)}} \cdot U \right) \\
&\quad \cdot \left(U_{z^{(\lceil k/2 \rceil+1)}} \cdot U \cdots U \cdot U_{z^{(k-1)}} \cdot U \cdot \vec{v}_{z^{(k)}} \right) \\
&= \frac{1}{N} \cdot \sum_{i_1,\ldots,i_k} z_{i_1}^{(1)} \cdot U_{i_1,i_2} \cdot z_{i_2}^{(2)} \cdot U_{i_2,i_3} \cdots U_{i_{k-1},i_k} \cdot z_{i_k}^{(k)} \\
&= \phi_U(z^{(1)}, \ldots, z^{(k)}).
\end{aligned}$$

Thus, overall we got that the algorithm's acceptance probability is

$$\frac{1+1+2\langle \vec{a}, \vec{b} \rangle}{4} = \frac{1+\phi_U(z^{(1)},\ldots,z^{(k)})}{2}.$$

■