THEORY OF COMPUTING, Volume 18 (17), 2022, pp. 1–11
www.theoryofcomputing.org

NOTE

A Stochastic Calculus Approach to the Oracle Separation of BQP and PH

Xinyu Wu

Received February 7, 2020; Revised June 17, 2022; Published June 23, 2022

Abstract. Recently, Ran Raz and Avishay Tal proved that in some relativized world, BQP is not contained in the polynomial-time hierarchy (STOC'19). It has been suggested that some aspects of the proof may be simplified by stochastic calculus. In this note, we describe such a simplification.

1 Introduction

A recent landmark result by Ran Raz and Avishay Tal [8] shows that there exists an oracle A such that $\mathsf{BQP}^A \not\subseteq \mathsf{PH}^A$. It has been suggested by several people, including Ryan O'Donnell, James Lee, and Avishay Tal, that some aspects of the proof may be simplified by stochastic calculus. We describe such a simplification.

As Aaronson [1] points out, there is a classical correspondence between the relativized complexity of PH and the size of bounded-depth, unbounded fan-in Boolean circuits (Furst, Saxe, Sipser [6]). Using this correspondence, the oracle separation reduces to upper bounds on the statistical difference between two distributions. Concretely, it suffices to show that there exists a distribution $\mathcal{D}$ over $\{-1, 1\}^{2N}$ such that

ACM Classification: F.1.3, G.3

AMS Classification: 68Q15, 81P68

Key words and phrases: quantum complexity, stochastic calculus

DOI: 10.4086/toc.2022.v018a017

1. For any $f : \{-1,1\}^{2N} \to \{0,1\}$ computable by a quasipolynomial-size bounded-depth circuit,
$$|\mathbf{E}[f(\mathcal{D})] - \mathbf{E}[f(\mathcal{U}_{2N})]| \leq \frac{\text{polylog}\,(N)}{\sqrt{N}}, \tag{1}$$
where $\mathcal{U}_{2N}$ is the uniform distribution over $\{-1,1\}^{2N}$. The notation $\mathbf{E}[f(\mathcal{D})]$ means $\mathbf{E}_{\mathbf{x}\sim\mathcal{D}}[f(\mathbf{x})]$. In fact, a weaker upper bound of $1/(\log N)^{\omega(1)}$ would suffice for the oracle separation.
2. There exists a quantum algorithm Q that queries the input once and runs in $O(\log N)$ time, such that
$$|\mathbf{E}[Q(\mathcal{D})] - \mathbf{E}[Q(\mathcal{U}_{2N})]| \geq \Omega\left(\frac{1}{\log N}\right). \tag{2}$$

For an explanation of why these two items suffice for the separation result, we refer to Raz and Tal's paper [8]. In their paper, Raz and Tal use a truncated Gaussian for $\mathcal{D}$. Moreover, they take Q to be the Forrelation query algorithm (first introduced as "Fourier checking" by Aaronson [1], and further analyzed in [2]). In this note, we will describe a construction of a related but different distribution $\mathcal{D}$ based on Brownian motion, which simplifies certain details of the analysis. The resulting analysis gives the same bounds as Raz and Tal, up to constant factors.

2 Overview

2.1 Strategy to construct the distribution $\mathcal{D}$

For convenience, we shall refer to the distribution called $\mathcal{D}$ in the Introduction as $\mathcal{D}'$. We shall use $\mathcal{D}$ to denote an auxiliary distribution we shall use to construct $\mathcal{D}'$ (see step 2 below).

We will first give an overview of the strategy to construct the distribution $\mathcal{D}'$. In this section we will not formalize stochastic calculus concepts, deferring this to Section 3.

The construction has two main steps:

1. Use a stopped Brownian motion to define a distribution $\mathcal{D}$ on $[-1/2, 1/2]^{2N}$.
2. Round $\mathcal{D}$ to a distribution $\mathcal{D}'$ on $\{-1,1\}^{2N}$ which has the same expectation on Boolean functions. (This step is identical with [8].)

We will define the distribution $\mathcal{D}$ by describing how to sample from it.

Let $\mathbf{x}_t$ be a standard N-dimensional Brownian motion, where "standard" means its covariance matrix is I_N, the $N \times N$ identity matrix. Let $\mathbf{y}_t = H_N\mathbf{x}_t$, where H_N is the $N \times N$ normalized Walsh-Hadamard matrix (the W-H matrix divided by $\sqrt{N}$). Finally, let $\mathbf{B}_t$ be the $2N \times 1$ column matrix formed by $\mathbf{x}_t$ on top of $\mathbf{y}_t$; for typographical convenience we write this as $\mathbf{B}_t = (\mathbf{x}_t, \mathbf{y}_t)$.

Proposition 2.1. *$\mathbf{B}_t$ is a $2N$-dimensional Brownian motion with covariance matrix*
$$\Phi := \begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

We defer the proof to Proposition 3.3.

Let $\varepsilon > 0$ and consider the random variable τ defined by

$$\tau \coloneqq \min\{\varepsilon, \text{first exit time of } \mathbf{B}_t \text{ from } [-1/2, 1/2]^{2N}\}\,. \tag{3}$$

Let $\mathbf{B}_\tau$ denote the random variable obtained by stopping the Brownian motion $\mathbf{B}_t$ at the random time $t = \tau$ (for appropriately chosen ε, see the line before Equation (12)). Let $\mathcal{D}$ be the distribution of the random variable $\mathbf{B}_\tau$.

Finally, use the method of Raz and Tal to round $\mathcal{D}$ to obtain a distribution $\mathcal{D}'$ on $\{-1,1\}^{2N}$ such that the following holds for every Boolean function $f : \{-1,1\}^{2N} \to \{0,1\}$:

$$\mathop{\mathbf{E}}_{\mathbf{z}\sim\mathcal{D}}[\tilde{f}(\mathbf{z})] = \mathop{\mathbf{E}}_{\mathbf{z}'\sim\mathcal{D}'}[f(\mathbf{z}')]\,, \tag{4}$$

where $\tilde{f}$ denotes the (unique) multilinear polynomial $\tilde{f} : \mathbb{R}^{2N} \mapsto \mathbb{R}$ that extends f. We shall describe the method and prove Equation (4) in Proposition 4.2.

Therefore, if $\mathcal{D}$ satisfies Equations (1) and (2), then so does $\mathcal{D}'$. Hence, it will suffice to analyze $\mathcal{D}$ instead of $\mathcal{D}'$.

2.2 Sketch of the quantum algorithm

The quantum algorithm Q used in Equation (2) is very simple: since H_N can be implemented by a quantum circuit of depth $O(\log N)$, computing $\langle \mathbf{x}, H_N \mathbf{y}\rangle$ will only take a single query. This value will have a large positive expectation when $(\mathbf{x}, \mathbf{y})$ is drawn from $\mathcal{D}$ but will have expectation 0 when $(\mathbf{x}, \mathbf{y})$ is drawn from $\mathcal{U}_{2N}$. This will be proven in Section 6.

2.3 Comparison with the proof of Raz and Tal

The proof here essentially follows the structure of RT, and uses many of the same technical ideas. The main differences are in the proof of the lower bound against bounded-depth circuits, while the quantum algorithm stays the same and has only a minor difference in its analysis.

Our main contribution lies in simplifying some aspects of Sections 5 and 7 from [8]. Because our $\mathcal{D}$ is defined to be bounded within $[-1,1]^{2N}$, there is no need to analyze a truncation function applied to $\mathcal{D}$, as in [8, Claims 5.2 and 5.3]. Theorem 4.4 reproves [8, Theorem 7.4], using some ideas from [8, Claim 7.2]. [8, Claim 7.3] is replaced with Lemma 4.3, while the analysis of the random walk in [8, Theorem 7.4] is replaced with an application of Dynkin's lemma. Also using Dynkin's lemma, we directly prove Theorem 4.4 using bounds on the second derivatives of the function f, instead of relying on Isserlis's theorem for moments of multivariate Gaussians.

3 Technical preliminaries

We briefly review some probability and stochastic calculus concepts. See for instance [7, Chapters 2.1 and 7] for more details. We shall use the common notation $(\Omega, \mathcal{F}, \mathbf{P})$ to denote a *probability space*, where Ω is the sample space, $\mathcal{F}$ is the σ-algebra of measurable sets, and $\mathbf{P}$ is the probability measure. A *random variable* is an $\mathcal{F}$-measurable function $\mathbf{X} : \Omega \to \mathbb{R}^N$. A random variable $\mathbf{X}$ induces a *distribution* which is a probability measure on $\mathbb{R}^N$, defined by $\mu_{\mathbf{X}}(B) = \mathbf{P}(\mathbf{X}^{-1}(B))$.

Definition 3.1. A *stochastic process* is a parametrized collection of random variables $\{\mathbf{X}_t\}_{t\in T}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and assuming values in $\mathbb{R}^N$.

Typically, the parameter space is $T = [0, \infty)$. For each t, we have a random variable $\omega \mapsto \mathbf{X}_t(\omega)$. On the other hand, for a fixed ω we can consider the function $t \mapsto \mathbf{X}_t(\omega)$, a *trajectory* of the process.

Definition 3.2. Let K be a positive semidefinite symmetric $N \times N$ real matrix. An *N-dimensional Brownian motion* $\{\mathbf{B}_t\}_{t\in[0,\infty)}$ with mean 0 and covariance K is a stochastic process characterized by the following:

(i) $\mathbf{B}_0 = 0$ almost surely.

(ii) for $u, t \geq 0$ the *increment* $\mathbf{B}_{t+u} - \mathbf{B}_t$ is independent of the past, $\{\mathbf{B}_s\}_{s<t}$.

(iii) for $u, t \geq 0$ the increment $\mathbf{B}_{t+u} - \mathbf{B}_t$ is distributed as an N-dimensional Gaussian with mean 0 and covariance matrix uK.

(iv) almost all trajectories are continuous.

We say that $\mathbf{B}$ is a *standard* Brownian motion if K is the identity matrix.

We refer to [7, Section 2.2] for a proof that such a stochastic process exists on some underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

We now make an observation.

Proposition 3.3 (Restatement of Proposition 2.1)**.** *Let $\mathbf{x}_t$ be a standard N-dimensional Brownian motion, and $\mathbf{y}_t = H_N \mathbf{x}_t$. Let $\mathbf{B}_t = (\mathbf{x}_t, \mathbf{y}_t)$. Then $\mathbf{B}_t$ is a $2N$-dimensional Brownian motion with covariance matrix*

$$\Phi := \begin{pmatrix} I_N & H_N \\ H_N & I_N \end{pmatrix}.$$

Proof. We will check items (i)–(iv) in the definition of Brownian motion. First, we note that $\mathbf{B}_t = (\mathbf{x}_t, \mathbf{y}_t) = (I_N, H_N)^T \mathbf{x}_t$, and so (i), (ii), and (iv) hold for $\mathbf{B}$ since they hold for $\mathbf{x}$.

Now we show (iii), that is, for fixed t, u, we want to show that the random variable $\mathbf{B}_{t+u} - \mathbf{B}_t = (\mathbf{x}_{t+u}, \mathbf{y}_{t+u}) - (\mathbf{x}_t, \mathbf{y}_t)$ is distributed as a Gaussian with mean 0 and covariance $u\Phi$. Using property (iii) of $\mathbf{x}$, we see that $\mathbf{x}_{t+u} - \mathbf{x}_t$ is a Gaussian with mean 0 and covariance uI_N, so $(\mathbf{x}_{t+u}, \mathbf{y}_{t+u}) - (\mathbf{x}_t, \mathbf{y}_t) = (I_N, H_N)^T(\mathbf{x}_{t+u} - \mathbf{x}_t) = \sqrt{u}(I_N, H_N)^T G_N$, where G_N is a standard N-dimensional Gaussian. Using the fact that for an $n \times d$ matrix A, the random variable AG_d is an n-dimensional Gaussian with mean 0 and covariance AA^T, we obtain that $\mathbf{B}_{t+u} - \mathbf{B}_t$ is a Gaussian with mean 0 and covariance $u(I_N, H_N)^T(I_N, H_N) = u\Phi$. □

We now define stopping times.

Definition 3.4. Let $\mathbf{X} = \{\mathbf{X}_t\}_{t\in[0,\infty)}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbf{P})$. A random variable $\tau : \Omega \to [0, \infty)$ is a *stopping time* for $\mathbf{X}$ if for any $t \in [0, \infty)$, the event $\{\tau \leq t\}$ is independent of $\{\mathbf{X}_s\}_{s>t}$. The *stopped* stochastic process $\mathbf{X}_\tau$ is a random variable defined via $\mathbf{X}_\tau(\omega) := \mathbf{X}_{\tau(\omega)}(\omega)$.

In particular, stopping times may be applied to a Brownian motion to produce a stopped Brownian motion. For example, any constant $\tau_0 \coloneqq t_0$ is a stopping time; the stopped Brownian motion has distribution $\mathbf{B}_{t_0}$. The first time that $\mathbf{B}_t$ exits the cube $[-1,1]^N$, $\tau_1 \coloneqq \inf\{t \geq 0 \mid \mathbf{B}_t \notin [-1,1]^N\}$ is also a stopping time. The minimum or maximum of any two stopping times is a stopping time. In particular, τ in Equation (3) is a stopping time.

We will need to use the following fact about Brownian motion.

Proposition 3.5. *Let $\mathbf{B}_t$ be an N-dimensional standard Brownian motion, and τ be a bounded stopping time. Then, $\mathbf{E}[\|\mathbf{B}_\tau\|^2] = \mathbf{E}[\tau]$.*

Proof. This follows from a few well-known facts about Brownian motion. First, $\|\mathbf{B}_t\|^2 - t$ is a martingale [9, Proposition II.1.2(ii)]. Given a bounded stopping time τ, $\mathbf{E}[\|\mathbf{B}_\tau\|^2 - \tau] = \mathbf{E}[\|\mathbf{B}_0\|^2] = 0$ [9, Proposition II.1.4], and so $\mathbf{E}[\|\mathbf{B}_\tau\|^2] = \mathbf{E}[\tau]$. □

The main stochastic calculus tool we will use is Dynkin's formula, which, for a function $f : \mathbb{R}^N \to \mathbb{R}^N$, relates $\mathbf{E}[f(\mathbf{B}_t)]$ to the second partial derivatives of f.

Theorem 3.6 (Dynkin's formula, [7, Theorem 7.4.1])**.** *Let $\mathbf{B}$ be an N-dimensional Brownian motion with mean 0 and covariance matrix K, let τ be a bounded stopping time, and let $f : \mathbb{R}^N \to \mathbb{R}$ be a twice continuously differentiable function. The following holds:*

$$\mathbf{E}[f(\mathbf{B}_\tau)] = f(0) + \mathbf{E}\left[\int_0^\tau \sum_{i,j\in[N]} K_{ij}(\partial_{ij} f)(\mathbf{B}_s)\, ds\right] \tag{5}$$

where $\partial_{ij} = \frac{\partial^2}{\partial x_i \partial x_j}$.

We will also require the following tail bound on Brownian motion.

Proposition 3.7 ([9, Proposition II.1.8])**.** *Let $\mathbf{B}$ be a standard 1-dimensional Brownian motion. For $a, t > 0$,*

$$\mathbf{Pr}\left[\sup_{0\leq s\leq t} |\mathbf{B}_s| \geq at\right] \leq \mathrm{e}^{-a^2 t/2}.$$

4 Reduction to a Fourier bound

The main technical part of Raz and Tal's proof [8] shows that, for a Boolean function $f : \{-1,1\}^{2N} \to \{-1,1\}$ computable by a bounded-depth, quasipolynomial-size circuit, and a multivariate Gaussian distribution $\mathcal{Z}$ over $\mathbb{R}^{2N}$,

$$|\,\mathbf{E}[f(\mathsf{trnc}(\mathcal{Z}))] - \mathbf{E}[f(\mathcal{U}_{2N})]| \leq O(\gamma \cdot \mathrm{polylog}(N)), \tag{6}$$

where γ is a bound on the (pairwise) covariances of the coordinates of $\mathcal{Z}$, trnc truncates $\mathcal{Z}$ so that the resulting random variable is within $[-1,1]^N$, and $\mathcal{U}_N$ is the uniform distribution over $\{-1,1\}^N$. This is based on the $k = 2$ case of Tal's fundamental result [10] that gives a polylog(m) upper bound on the the level-k Fourier coefficients of Boolean functions computable by a Boolean circuit of bounded depth and size m. We state Tal's exact bound as Theorem 5.1 below.

Another natural way of viewing a multivariate Gaussian distribution is as the result of an N-dimensional Brownian motion stopped at a fixed time. We can also build the truncation into the stopping time. This allows us to use tools from stochastic calculus to analyze the distribution.

We first recall the definition of restrictions of Boolean functions.

Definition 4.1 (restriction)**.** Let $f : \{-1,1\}^N \to \mathbb{R}$ and let $\rho \in \{-1,1,*\}^N$. Let free(ρ) be the set of coordinates with $*$'s. We define the restriction of f by ρ as $f_\rho : \{-1,1\}^N \to \mathbb{R}$, where $f_\rho(x)$ is f evaluated at ρ with the coordinates of x replacing[1] the $*$'s in ρ.

Henceforth, we also identify functions on a Boolean domain, $f : \{-1,1\}^N \to \mathbb{R}$, with their multilinear polynomial representations (or Fourier expansions)

$$f(x) = \sum_{S \subseteq [N]} \widehat{f}(S) \prod_{i \in S} x_i. \tag{7}$$

The following result has been extracted from the proof of Equation (2) in [8, Sec. 5].

Proposition 4.2. *Let $\mathcal{D}$ be a distribution on $[-1,1]^N$. Let $\mathbf{z}' \sim \mathcal{D}'$ be sampled by first drawing $\mathbf{z} \sim \mathcal{D}$. Then, independently for each $i \in [N]$, we will set $\mathbf{z}'_i = 1$ with probability $(1+\mathbf{z}_i)/2$ and $\mathbf{z}'_i = -1$ with probability $(1-\mathbf{z}_i)/2$. For any function $f : \{-1,1\}^N \to \mathbb{R}$, after identifying f with its multilinear polynomial representation, we have*

$$\mathop{\mathbf{E}}_{\mathbf{z} \sim \mathcal{D}}[f(\mathbf{z})] = \mathop{\mathbf{E}}_{\mathbf{z}' \sim \mathcal{D}'}[f(\mathbf{z}')].$$

Proof. The proof follows from the Fourier expansion. First, fix $z \in [-1,1]^N$, and then draw $\mathbf{z}' \in \{-1,1\}^N$ using the procedure above.

$$\mathop{\mathbf{E}}_{\mathbf{z}' \sim \mathbf{z}}[f(\mathbf{z}') \mid z] = \mathbf{E}\left[\sum_{S \subseteq [N]} \widehat{f}(S) \prod_{i \in S} \mathbf{z}'_i \,\middle|\, z \right] = \sum_{S \subseteq [N]} \widehat{f}(S) \prod_{i \in S} \mathbf{E}[\mathbf{z}'_i \mid z] = f(z).$$

Taking the expectation of z over $\mathcal{D}$, we infer $\mathbf{E}_{\mathbf{z} \sim \mathcal{D}}[f(\mathbf{z})] = \mathbf{E}_{\mathbf{z}' \sim \mathcal{D}'}[f(\mathbf{z}')]$. □

We make some observations about Fourier coefficients. First, the Fourier coefficients of f_ρ satisfy $\widehat{f_\rho}(S) = 0$ for all $S \not\subseteq$ free(ρ). We also have that

$$\widehat{f}(S) = (\partial_S f)(0), \tag{8}$$

where $\partial_S = \prod_{i \in S} \partial_i$ and $\partial_i = \frac{\partial}{\partial x_i}$ is the partial derivative.

[1] Although the domain of f_ρ is $\{-1,1\}^N$, it only depends on the coordinates in free(ρ).

Further, because f is multilinear, for any $h \in \mathbb{R} \setminus \{0\}$ and any standard basis vector e_i we have

$$(\partial_i f)(x) = \frac{f(x + he_i) - f(x)}{h}. \tag{9}$$

The following lemma is similar to [5, Claim A.5], which first appeared in [3] and [4, Claim 3.3].

Lemma 4.3. *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a multilinear polynomial. For any $x \in [-1/2, 1/2]^N$, there exists a distribution $\mathcal{R}_x$ over restrictions $\rho \in \{-1, 1, *\}^N$, such that for any $i, j \in [N]$,*

$$(\partial_{ij} f)(x) = 4 \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_x}\left[(\partial_{ij} f_\rho)(0)\right]. \tag{10}$$

Proof. We define $\mathcal{R}_x$ as follows: for each coordinate $i \in [N]$ we independently set ρ_i to be 1 with probability $\frac{1}{4} + \frac{x_i}{2}$, to be -1 with probability $\frac{1}{4} - \frac{x_i}{2}$, and to be $*$ with probability $\frac{1}{2}$.

Using that f is a multilinear polynomial, and that the coordinates are independent, we deduce that for any $y \in \mathbb{R}^N$, $f(x + y) = \mathbf{E}_{\rho \sim \mathcal{R}_x}\left[f_\rho(2y)\right]$. Then, using Equation (9),

$$\begin{aligned}(\partial_{ij} f)(x) &= f(x + e_i + e_j) - f(x + e_i) - f(x + e_j) + f(x) \\ &= \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_x}\left[f_\rho(2e_i + 2e_j) - f_\rho(2e_j) - f_\rho(2e_i) + f_\rho(0)\right] = 4 \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_x}\left[(\partial_{ij} f_\rho)(0)\right]. \qquad \square\end{aligned}$$

We now show our main result, which is analogous to [5, Theorem A.7] and [8, Theorem 2.4].

Theorem 4.4. *Let $f : \{-1, 1\}^N \to \{-1, 1\}$ be a Boolean function, and let $L > 0$ such that for any restriction ρ,*

$$\sum_{\substack{S \subseteq [N] \\ |S| = 2}} |\widehat{f_\rho}(S)| \leq L.$$

Let $\gamma > 0$ and let $\mathbf{B}$ be an N-dimensional Brownian motion with mean 0 and covariance matrix K. Further assume that $|K_{ij}| \leq \gamma$ for $i \neq j$.

Let $\varepsilon > 0$ and define the stopping time

$$\tau := \min\{\varepsilon, \textit{ first time that } \mathbf{B}_t \textit{ exits } [-1/2, 1/2]^N\}.$$

Then, identifying f with its multilinear representation, we have

$$|\mathbf{E}[f(\mathbf{B}_\tau)] - \mathbf{E}[f(\mathcal{U}_n)]| \leq 2\varepsilon\gamma L.$$

Proof. First, we note that $\mathbf{E}[f(\mathcal{U}_N)] = f(0)$. Note that $\mathbf{B}_\tau$ is always within $[-1/2, 1/2]^N$. We can apply Theorem 3.6

$$\mathbf{E}[f(\mathbf{B}_\tau)] - f(0) = \mathbf{E}\left[\int_0^\tau \frac{1}{2} \sum_{i,j \in [N]} K_{ij} (\partial_{ij} f)(\mathbf{B}_s)\, ds\right]. \tag{11}$$

Then, we use the upper bound $\tau \le \varepsilon$, and that $(\partial_{ii} f) = 0$ for all $i \in [N]$ because f is multilinear, to get

$$\begin{aligned}
|\,\mathbf{E}[f(\mathbf{B}_\tau)] - f(0)| &\le \varepsilon\, \mathbf{E}\left[\sup_{s\in[0,\tau]} \left| \frac{1}{2} \sum_{i,j\in[N]} K_{ij}(\partial_{ij} f)(\mathbf{B}_s)\right|\right] \\
&\le \frac{\varepsilon\gamma}{2} \sup_{x\in[-1/2,1/2]^N} \sum_{i\neq j} \left|(\partial_{ij} f)(x)\right| \\
&= 2\varepsilon\gamma \sup_{x\in[-1/2,1/2]^N} \sum_{i\neq j} \left| \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_x} \left[(\partial_{ij} f_\rho)(0)\right]\right| && \text{(Lemma 4.3)} \\
&\le 2\varepsilon\gamma \sup_{x\in[-1/2,1/2]^N} \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_x} \left[\sum_{i\neq j} \left|(\partial_{ij} f_\rho)(0)\right|\right] \\
&\le 2\varepsilon\gamma \sup_{x\in[-1/2,1/2]^N} \mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_x} \left[\sum_{\substack{S\subseteq \mathrm{free}(\rho) \\ |S|=2}} \left|\widehat{f_\rho}(S)\right|\right] && \text{(Equation (8))} \\
&\le 2\varepsilon\gamma L. && \square
\end{aligned}$$

5 Bound for classical circuits

We now construct the distribution $\mathcal{D}$ as described in Section 2.1. Due to Proposition 2.1, we can take $\mathcal{D}$ to be the distribution defined by $\mathbf{B}_\tau$, where $\mathbf{B}$ is a $2N$-dimensional Brownian motion with covariance matrix Φ, τ is defined as in Theorem 4.4, and and ε will be chosen appropriately before Equation (12) below.

Following Raz–Tal, we first use the following result of Tal [10] on the Fourier weight of Boolean circuits.

Theorem 5.1 (Theorem 37(3) of [10]). *There exists a constant $C > 0$ such that the following holds. Let $f : \{-1,1\}^{2N} \to \{-1,1\}$ be a Boolean function computed by a Boolean circuit of depth d and size $m > 1$. Then for all k,*

$$\sum_{S:|S|=k} |\widehat{f}(S)| \le 2(C(\log m)^{d-1})^k.$$

In particular, if f is computed by a bounded-depth circuit of quasipoly(N) size, then

$$\sum_{S:|S|=2} |\widehat{f}(S)| \le \mathrm{polylog}(N).$$

Since restriction does not increase circuit size or depth, we can apply Theorem 4.4 with

$\varepsilon = 1/(8\ln 2N)$ and $\gamma = \frac{1}{\sqrt{N}}$, to deduce that

$$|\mathbf{E}[f(\mathbf{B}_\tau)] - f(0)| \leq \frac{\operatorname{polylog}(N)}{\sqrt{N}}, \tag{12}$$

where $\mathbf{B}_\tau$ is defined as in Theorem 4.4, justifying Equation (1). □

6 Quantum algorithm

Finally, we show that a 1-query $O(\log N)$-time quantum algorithm can distinguish $\mathcal{D}$ from the uniform distribution. This is virtually identical to the argument in [8, Section 6], but we can again use some stochastic calculus tools on the stopping time built into our, slightly different, distribution.

The Forrelation query algorithm [1, 2] is an $O(\log N)$-time quantum algorithm with inputs $x, y \in \{-1, 1\}^N$ which accepts with probability $(1 + \phi(x, y))/2$, where

$$\phi(x, y) := \frac{1}{N}\langle x, H_N y\rangle. \tag{13}$$

When the pair (x, y) is drawn from the uniform distribution, $\mathbf{E}[\phi(x, y)] = 0$. The quantum algorithm is to prepare the uniform superposition over the basis states $|1\rangle \dots |N\rangle$, query x, apply the Walsh–Hadamard transform, query y, apply the Walsh–Hadamard transform again, then measure in the computational basis and accept if the outcome is $|1\rangle$. The quantum algorithm is described in more detail in [1, Section 3.2].

We show the following inequality [8, Claim 6.3], which shows that the quantum algorithm distinguishes $\mathcal{D}$ from the uniform distribution with sufficiently high probability, justifying Equation (2).

Proposition 6.1. $\mathbf{E}_{(\mathbf{x}_\tau, \mathbf{y}_\tau)\sim\mathcal{D}}[\phi(\mathbf{x}_\tau, \mathbf{y}_\tau)] \geq \frac{\varepsilon}{4}$.

Proof. We have

$$\begin{aligned}\mathop{\mathbf{E}}_{(\mathbf{x}_\tau, \mathbf{y}_\tau)\sim\mathcal{D}}[\phi(\mathbf{x}_\tau, \mathbf{y}_\tau)] &= \frac{1}{N}\mathbf{E}[\langle \mathbf{x}_\tau, H_N \mathbf{y}_\tau\rangle] \\ &= \frac{1}{N}\mathbf{E}[\langle \mathbf{x}_\tau, H_N^2 \mathbf{x}_\tau\rangle] = \mathbf{E}[\|\mathbf{x}_\tau\|^2].\end{aligned}$$

By Proposition 3.5, we have

$$\mathbf{E}[\|\mathbf{x}_\tau\|^2] = \mathbf{E}[\tau]. \tag{14}$$

By Markov's inequality,

$$\mathbf{E}[\tau] \geq \frac{\varepsilon}{2}\mathbf{Pr}[\tau > \tfrac{\varepsilon}{2}].$$

If $\tau \leq \frac{\varepsilon}{2}$, it must be the case that the path exits $[-1/2, 1/2]^{2N}$ no later than $\frac{\varepsilon}{2}$. Hence, by the union bound, we have

$$\mathbf{Pr}\left[\tau \leq \tfrac{\varepsilon}{2}\right] \leq 2N \cdot \mathbf{Pr}\left[\text{1st coordinate of } (\mathbf{x}_\tau, \mathbf{y}_\tau) \text{ exits } \left[-\tfrac{1}{2}, \tfrac{1}{2}\right] \text{ not later than } \tfrac{\varepsilon}{2}\right]. \tag{15}$$

Each coordinate of $(\mathbf{x}_\tau, \mathbf{y}_\tau)$ is a standard 1D Brownian motion since $K_{ii} = 1$ for all i. Let $\mathbf{B}_t^{(1)}$ denote the first coordinate of $\mathbf{x}_t$. Applying Proposition 3.7,

$$\mathbf{Pr}\left[\sup_{0\le t\le \varepsilon/2} |\mathbf{B}_t^{(1)}| \ge \frac{1}{2}\right] \le 2\mathrm{e}^{-1/4\varepsilon} = 2\mathrm{e}^{-2\ln 2N} \le \frac{1}{4N} \quad \text{for } N \ge 4. \tag{16}$$

Therefore, $\mathbf{Pr}[\tau \le \frac{\varepsilon}{2}] \le \frac{1}{2}$, so $\mathbf{E}[\tau] \ge \frac{\varepsilon}{4}$. □

7 Acknowledgments

I would like to thank Ryan O'Donnell and Avishay Tal for helpful discussions and their suggestions concerning an early draft. Thanks also to Gregory Rosenthal and anonymous reviewers for helpful comments.

References

[1] SCOTT AARONSON: BQP and the Polynomial Hierarchy. In *Proc. 42nd STOC*, pp. 141–150. ACM Press, 2010. [doi:10.1145/1806689.1806711] 1, 2, 9

[2] SCOTT AARONSON AND ANDRIS AMBAINIS: Forrelation: a problem that optimally separates quantum from classical computing. *SIAM J. Comput.*, 47(3):982–1038, 2018. Preliminary version in STOC'15. [doi:10.1137/15M1050902] 2, 9

[3] BOAZ BARAK AND JAROSŁAW BŁASIOK: On the Raz-Tal oracle separation of BQP and PH, 2018. windowsontheory.org. 7

[4] ESHAN CHATTOPADHYAY, POOYA HATAMI, KAAVE HOSSEINI, AND SHACHAR LOVETT: Pseudorandom generators from polarizing random walks. *Theory of Computing*, 15(10):1–26, 2019. Preliminary version in CCC'18. [doi:10.4086/toc.2019.v015a010] 7

[5] ESHAN CHATTOPADHYAY, POOYA HATAMI, SHACHAR LOVETT, AND AVISHAY TAL: Pseudorandom generators from the second Fourier level and applications to AC0 with parity gates. In *Proc. 10th Innovations in Theoret. Comp. Sci. Conf. (ITCS'19)*, pp. 22:1–22:15. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.ITCS.2019.22, ECCC:TR18-155] 7

[6] MERRICK LEE FURST, JAMES B. SAXE, AND MICHAEL SIPSER: Parity, circuits, and the polynomial-time hierarchy. *Math. Systems Theory*, 17(1):13–27, 1984. Preliminary version in FOCS'81. [doi:10.1007/BF01744431] 1

[7] BERNT ØKSENDAL: *Stochastic Differential Equations*. Universitext. Springer, 6th edition, 2003. [doi:10.1007/978-3-642-14394-6] 3, 4, 5

[8] RAN RAZ AND AVISHAY TAL: Oracle separation of BQP and PH. In *Proc. 51st STOC*, pp. 13–23. ACM Press, 2019. [doi:10.1145/3313276.3316315, ECCC:TR18-107] 1, 2, 3, 5, 6, 7, 9

[9] DANIEL REVUZ AND MARC YOR: *Continuous Martingales and Brownian Motion*. Volume 293 of *Grundlehren der Math. Wiss.* Springer, 3rd edition, 1999. [doi:10.1007/978-3-662-06400-9] 5

[10] AVISHAY TAL: Tight bounds on the Fourier spectrum of AC^0. In *Proc. 32nd Comput. Complexity Conf. (CCC'17)*, pp. 15:1–15:31. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017. [doi:10.4230/LIPIcs.CCC.2017.15, ECCC:TR14-174] 5, 8

AUTHOR

Xinyu Wu
Graduate student
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA, USA
xinyuwu@cmu.edu
https://www.andrew.cmu.edu/user/xinyuw1/

ABOUT THE AUTHOR

XINYU WU is a Ph. D. student at Carnegie Mellon University, advised by Ryan O'Donnell and Pravesh Kothari. Her research interests are in spectral graph theory, free probability, and their applications in understanding average-case problems and quantum computing. She grew up in Singapore, and did her undergraduate degree in math and computer science also at Carnegie Mellon. She also enjoys cooking, cycling, and cat videos.