

A stochastic calculus approach to the oracle separation of BQP and PH

Xinyu Wu*

July 5, 2020

Abstract

After presentations of Raz and Tal’s oracle separation of BQP and PH result, several people (e.g. Ryan O’Donnell, James Lee, Avishay Tal) suggested that the proof may be simplified by stochastic calculus. In this short note, we describe such a simplification.

1 Introduction

A recent landmark result of Raz and Tal [RT19] shows there exists an oracle A such that $\text{BQP}^A \not\subseteq \text{PH}^A$. Using a correspondence between PH and AC^0 circuits, the question reduces to a lower bound against AC^0 circuits. Concretely, it suffices to show that there exists a distribution \mathcal{D} over $\{-1, 1\}^N$ such that

1. For any $f : \{-1, 1\}^N \rightarrow \{0, 1\}$ computable by an AC^0 circuit,

$$|\mathbf{E}[f(\mathcal{D})] - \mathbf{E}[f(\mathcal{U}_N)]| \leq \frac{\text{polylog}(N)}{\sqrt{N}},$$

where \mathcal{U}_N is the uniform distribution on N bits. The notation $\mathbf{E}[f(\mathcal{D})]$ means $\mathbf{E}_{x \sim \mathcal{D}}[f(x)]$.

2. There exists a quantum algorithm Q such that

$$|\mathbf{E}[Q(\mathcal{D})] - \mathbf{E}[Q(\mathcal{U}_N)]| \geq \Omega\left(\frac{1}{\log N}\right).$$

For details, we refer to Raz and Tal’s paper [RT19]. \mathcal{D} in Raz and Tal’s work is a truncated Gaussian. In this note, we will describe a construction of \mathcal{D} based on Brownian motion, which simplifies many details of the analysis.

1.1 Stochastic calculus preliminaries

We briefly review some stochastic calculus concepts used in the proof. See for instance [Øks03, Chapter 7] for details.

Definition 1. An N -dimensional standard Brownian motion $\mathbf{B} : [0, \infty) \times \mathbb{R}^N \rightarrow \mathbb{R}^N$ is a continuous-time stochastic process characterized by the following:

- (i) $\mathbf{B}_0 = 0$ almost surely.
- (ii) $\mathbf{B}_{t+u} - \mathbf{B}_t$ for $u \geq 0$ is independent of \mathbf{B}_s for $s < t$.

*Computer Science Department, Carnegie Mellon University. xinyuwu@cmu.edu

- (iii) $\mathbf{B}_{t+u} - \mathbf{B}_t$ for $u \geq 0$ is distributed as an N -dimensional Gaussian with mean 0 and covariance matrix $u\mathbf{I}_{N \times N}$.
- (iv) \mathbf{B}_t is continuous almost surely.

We can describe a large class of stochastic processes, called *Itô diffusion* processes, by the solutions of stochastic differential equations of the following form:

$$d\mathbf{X}_t = b(\mathbf{X}_t) dt + \sigma(\mathbf{X}_t) d\mathbf{B}_t.$$

Definition 2. Let X be an Itô diffusion. The *infinitesimal generator* of f , is defined as

$$Af(x) = \lim_{t \rightarrow 0} \frac{\mathbf{E}_x[f(\mathbf{X}_t)] - f(x)}{t}.$$

We use the $\mathbf{E}_x[\cdot]$ notation to mean that we let \mathbf{X}_t evolve with starting point x .

If f is twice continuously differentiable with compact support, we have the following expression for Af :

$$Af(x) = b(x) \cdot \nabla f(x) + \frac{1}{2} \text{tr}(\sigma(x)\sigma^\top(x) \mathbf{H}(x)),$$

where \mathbf{H} is the Hessian of f .

For example, the infinitesimal generator of a standard 1D Brownian motion is the Laplacian operator. For a Brownian motion with covariance matrix Σ , the infinitesimal generator would be $\text{tr}(\Sigma \mathbf{H}(x))$.

Next we state Dynkin's formula, which will be the main tool we use in the later proof.

Theorem 1 (Dynkin's formula, [Øks03, Theorem 7.4.1]). *Let \mathbf{X} be an Itô diffusion, let τ be a stopping time with $\mathbf{E}[\tau] < \infty$, and let $f : \mathbb{R}^N \rightarrow \mathbb{R}^N$ be a twice continuously differentiable function with compact support. The following holds:*

$$\mathbf{E}_x[f(\mathbf{X}_\tau)] = f(x) + \mathbf{E}_x \left[\int_0^\tau Af(\mathbf{X}_s) ds \right],$$

Moreover, if \mathbf{X}_τ is bounded, Dynkin's formula with the same expression for Af holds for f which is twice continuously differentiable (without compact support).

2 Reduction to a Fourier bound

The main technical part of Raz and Tal's result [RT19] shows that, for a Boolean function $f : \{-1, 1\}^N \rightarrow \{-1, 1\}$ computable by an AC^0 circuit, and a multivariate Gaussian distribution $\mathbf{Z} \in \mathbb{R}^N$,

$$|\mathbf{E}[f(\text{trnc}(\mathbf{Z}))] - \mathbf{E}[f(\mathbf{U}_N)]| \leq O(\gamma \cdot \text{polylog}(n)),$$

where γ is a bound on the (pairwise) covariance of the coordinates of \mathbf{Z} , trnc truncates \mathbf{Z} so that the resulting random variable is within $[-1, 1]^N$, and \mathbf{U}_N is the uniform distribution over $\{-1, 1\}^N$. The important condition used here is that AC^0 has second level Fourier coefficients bounded by $\text{polylog}(n)$, and that this holds under any restriction of the function.

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

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

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

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

$$\hat{f}(S) = \partial_S f(0), \tag{1}$$

where $\partial_S = \prod_{i \in S} \partial_i$ and $\partial_i = \frac{\partial}{\partial x_i}$ is the usual calculus derivative. Further, because f is multilinear, for any $h \in \mathbb{R} \setminus \{0\}$ and any standard basis vector e_i we have

$$\partial_i f(x) = \frac{f(x + h e_i) - f(x)}{h}. \tag{2}$$

The following lemma is similar to [CHLT18, Claim A.5], which first appeared in [BB18] and [CHHL19, Claim 3.3].

Lemma 1. Let $f : \mathbb{R}^N \rightarrow \mathbb{R}$ 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 $i, j \in [N]$,

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

Proof. We define \mathcal{R}_x as such: 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 $y \in \mathbb{R}^N$, $f(x + y) = \mathbf{E}_{\rho \sim \mathcal{R}_x} [f_\rho(2y)]$. Then, using Equation (2),

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

We now show the main result, which is a restatement of [CHLT18, Theorem A.7] and [RT19, Theorem 2.4].

Theorem 2. Let $f : \{-1, 1\}^N \rightarrow \{-1, 1\}$ be a Boolean function, and let $t > 0$ such that for any restriction ρ ,

$$\sum_{\substack{S \subseteq [N] \\ |S|=2}} |\hat{f}_\rho(S)| \leq t.$$

Let $\gamma > 0$ and let \mathbf{X} be an N -dimensional Brownian motion with mean 0 and covariance matrix Σ , in the sense that $\mathbf{E}[(\mathbf{X}_t)_i] = 0$ for all $i \in [N]$, and $\mathbf{Cov}((\mathbf{X}_t - \mathbf{X}_s)_i, (\mathbf{X}_t - \mathbf{X}_s)_j) = (t - s)\Sigma_{ij}$. Further assume that $|\Sigma_{ij}| \leq \gamma$ for $i \neq j$.

¹Although f_ρ 's domain is $\{-1, 1\}^N$, it only depends on the coordinates in $\text{free}(\rho)$.

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

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

Then, identifying f with its multilinear expansion, we have

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

Proof. First, we note that $\mathbf{E}[f(\mathbf{U}_N)] = f(0)$. Next, let $\sigma = \Sigma^{1/2}$. \mathbf{X} satisfies the stochastic differential equation

$$d\mathbf{X}_t = \sigma d\mathbf{B}_t.$$

Note that \mathbf{X}_τ is always within $[-1/2, 1/2]^N$. We can apply [Theorem 1](#)

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

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

$$\begin{aligned} |\mathbf{E}[f(\mathbf{X}_\tau)] - f(0)| &\leq \varepsilon \mathbf{E} \left[\sup_{s \in [0, \tau]} \left| \frac{1}{2} \sum_{i,j \in [N]} \Sigma_{ij} \partial_{ij} f(\mathbf{X}_s) \right| \right] \\ &\leq \frac{\varepsilon\gamma}{2} \sup_{x \in [-1/2, 1/2]^N} \sum_{i \neq j} |\partial_{ij} f(x)| \\ &= 2\varepsilon\gamma \sup_{x \in [-1/2, 1/2]^N} \sum_{i \neq j} \left| \mathbf{E}_{\rho \sim \mathcal{R}_x} [\partial_{ij} f_\rho(0)] \right| && \text{(Lemma 1)} \\ &\leq 2\varepsilon\gamma \sup_{x \in [-1/2, 1/2]^N} \mathbf{E}_{\rho \sim \mathcal{R}_x} \left[\sum_{i \neq j} |\partial_{ij} f_\rho(0)| \right] \\ &\leq 2\varepsilon\gamma \sup_{x \in [-1/2, 1/2]^N} \mathbf{E}_{\rho \sim \mathcal{R}_x} \left[\sum_{\substack{S \subseteq \text{free}(\rho) \\ |S|=2}} |\hat{f}_\rho(S)| \right] && \text{(Equation (1))} \\ &\leq 2\varepsilon\gamma t. \quad \square \end{aligned}$$

3 Application to the oracle separation of BQP and PH

We now use [Theorem 2](#) to construct \mathcal{D} as described in [Section 1](#).

The distribution \mathcal{D} . Let $N = 2n$, where n is a power of 2, and

$$\Sigma := \begin{pmatrix} I_n & H_n \\ H_n & I_n \end{pmatrix},$$

where H_n is the Walsh–Hadamard matrix. Now we define \mathbf{X} and τ as in [Theorem 2](#), with $\varepsilon = 1/(8 \ln N)$, and our distribution \mathcal{D} will be the distribution defined by \mathbf{X}_τ . At each time t , we can also look at \mathbf{X}_t as a pair of random variables in \mathbb{R}^n , (x_t, \mathbf{y}_t) such that \mathbf{y}_t is the Hadamard transform of x_t .

AC⁰ lower bound. Tal showed that [Tal17, Theorem 37] there exists a universal constant c such that every function $f : \{-1, 1\}^N \rightarrow \{-1, 1\}$ computable by an AC⁰ circuit with at most $(\ln N)^\ell$ gates and depth d satisfies

$$\sum_{\substack{S \subseteq [N] \\ |S|=k}} |\hat{f}(S)| \leq (c \cdot \ln^\ell N)^{(d-1)k}.$$

Since AC⁰ is closed under restrictions, we can apply [Theorem 2](#) with $\varepsilon = 1/(8 \ln N)$ and $\gamma = \frac{1}{\sqrt{n}}$, to deduce that

$$|\mathbf{E}[f(\mathbf{X}_\tau)] - f(0)| \leq \frac{\text{polylog } N}{\sqrt{N}}.$$

Quantum algorithm. Finally, we show that a quantum algorithm can distinguish \mathcal{D} from the uniform distribution. This is virtually identical to the argument in [RT19, Section 6], but we can again use some stochastic calculus tools on the stopping time built into the distribution. Using the Forrelation query algorithm, there is a quantum algorithm Q with inputs $x, y \in \{-1, 1\}^n$ which accepts with probability $(1 + \varphi(x, y))/2$, where

$$\varphi(x, y) := \frac{1}{n} \sum_{i, j \in [n]} x_i \cdot H_{ij} \cdot y_j.$$

We show the following proposition [RT19, Claim 6.3], which implies the existence of a $O(\log N)$ -time quantum algorithm distinguishing \mathcal{D} from uniform with one query. The quantum algorithm is described in more detail in [Aar10, Section 3.2].

Proposition 1. $\mathbf{E}_{(x, y) \sim \mathcal{D}}[\varphi(x, y)] \geq \frac{\varepsilon}{4}$.

Proof. By the linearity of expectation and optional sampling theorem,

$$\begin{aligned} \mathbf{E}_{(x, y) \sim \mathcal{D}}[\varphi(x, y)] &= \frac{1}{n} \sum_{i, j \in [n]} H_{ij} \cdot \mathbf{E}[x_i \cdot y_j] \\ &= \frac{1}{n} \sum_{i, j \in [n]} H_{ij} \cdot \mathbf{E}[\tau] \cdot H_{ij} = \mathbf{E}[\tau]. \end{aligned}$$

By Markov's inequality,

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

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

$$\Pr[\tau \leq \frac{\varepsilon}{2}] \leq N \cdot \Pr[1\text{st coordinate of } X_t \text{ exits } [-\frac{1}{2}, \frac{1}{2}] \text{ earlier than } \frac{\varepsilon}{2}].$$

Each coordinate of \mathbf{X} is a standard 1D Brownian motion since $\Sigma_{ii} = 1$ for all i . An application of Doob's martingale inequality (e.g. [RY99, Proposition II.1.8]) tells us that, for a standard 1D Brownian motion \mathbf{B}_t ,

$$\Pr\left[\sup_{0 \leq t \leq \varepsilon/2} |\mathbf{B}_t| \geq \frac{1}{2}\right] \leq 2e^{-1/4\varepsilon} = 2e^{-2 \ln N} \leq \frac{1}{2N} \quad \text{for } N \geq 4.$$

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

4 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

- [Aar10] Scott Aaronson. BQP and the Polynomial Hierarchy. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing*, pages 141–150, 2010.
- [BB18] Boaz Barak and Jarosław Błasiok. On the Raz-Tal oracle separation of BQP and PH. <https://windowsontheory.org/2018/06/17/on-the-raz-tal-oracle-separation-of-bqp-and-ph/>, 2018.
- [CHHL19] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory Comput.*, 15:Paper No. 10, 26, 2019.
- [CHLT18] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second Fourier level and applications to AC0 with parity gates. In *Proceedings of the 10th Annual Innovations in Theoretical Computer Science Conference*, pages 22:1–22:15, 2018.
- [Øks03] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.
- [RT19] Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing*, pages 13–23. ACM, New York, 2019.
- [RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [Tal17] Avishay Tal. Tight bounds on the Fourier spectrum of AC0. In *Proceedings of the 32nd Annual Computational Complexity Conference*, 2017.