Analyzing XOR-Forrelation Through Stochastic Calculus

Xinyu Wu

Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

In this note we present a simplified analysis of the quantum and classical complexity of the k-XOR Forrelation problem (introduced in the paper of Girish, Raz and Zhan [7]) by a stochastic interpretation of the Forrelation distribution.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases quantum complexity, Brownian motion

Digital Object Identifier 10.4230/LIPIcs.STACS.2022.60

Related Version *Full Version*: https://arxiv.org/abs/2109.02732

Funding Supported by NSF grant CCF-1717606. This material is based upon work supported by the National Science Foundation under grant numbers listed above. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation (NSF).

Acknowledgements I would like to thank Ryan O'Donnell for many helpful comments on this paper.

1 Introduction

The Forrelation problem [1] and variants of it have been useful in producing problems that are efficiently solvable by quantum protocols but are hard for classical protocols, in various different models. A recent line of work analyzing the Forrelation distribution builds on the polarizing random walk framework introduced by Chattopadhyay, Hatami, Hosseini and Lovett [4]. This framework views the Forrelation distribution as being generated by a random walk in $\mathbb{R}^N$, producing a particular Gaussian distribution, and then rounded to the Boolean cube $\{-1, 1\}^N$. This approach lead to breakthroughs as Raz and Tal's result on the oracle separation of BQP and PH [9] and Bansal and Sinha's proof that k-Forrelation exhibits an optimal separation between quantum and classical query complexity [2][1].

The recent work of Girish, Raz and Zhan [7] analyzes the XOR of k copies of the Forrelation function, and shows that the resulting problem is such that classical protocols of quasipolynomial size can only achieve quasipolynomially small advantage over random guessing, while there exist quantum protocols with complexity polylog(N). They show this for quantum simultaneous-message communication protocols vs. classical randomized communication protocols, as well as for quantum query complexity vs. classical query complexity.

Stochastic calculus viewpoint

The approach here generalizes [13] (indeed, the $k = 1$ case is identical). There are two main points where the stochastic approach simplifies the argument in [7].

[1] The proof is phrased in terms of Gaussian interpolation, which is a different viewpoint on the stochastic approach.

39th International Symposium on Theoretical Aspects of Computer Science (STACS 2022).
Editors: Petra Berenbrink and Benjamin Monmege; Article No. 60; pp. 60:1–60:7
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

First, the Forrelation distribution, prior to rounding, is a truncated multivariate Gaussian. A multivariate N-dimensional Gaussian can also be realized as an N-dimensional Brownian motion, stopped at some constant time. Using a continuous-time random walk allows us to apply stochastic calculus techniques to bound how well f distinguishes the two distributions directly using the $2k^{\text{th}}$ order derivatives of f, without the need for additional intermediate bounds. Furthermore, the Brownian motion approach allows for an induction on k, eliminating the need for complex dimension-dependent bounds.

Second, viewing the Gaussian as a Brownian motion also allows us to use a stopping time to encode the truncation. This allows us to directly encode the boundedness of the distribution in the random variable. This eliminates the extra step to truncate the Gaussian and bound the closeness in expectation between the truncated and non-truncated Gaussians.

Connections and future work

Conceptually, viewing a truncated Gaussian as a stopped Brownian motion enforces a pathwise view of the random variable, i.e. sampling from the distribution means sampling a path of a random walk. This makes calculations on the distributions easier, for instance because the paths naturally split into “paths which always remain within the region” and “paths which end by hitting the boundary”. This technique may also be interesting for other applications using truncated Gaussians (or analogously, replacing a truncated exponential distribution by a stopped geometric Brownian motion.) The stochastic calculus view of Gaussians has also been useful for other Boolean analysis results, for instance in the proof of Bobkov’s Two Point Inequality by Barthe and Maurey [3]. Ideas related to the pathwise view of random variables also appear in the recent paper of Eldan and Gross [6], which expresses the variance and influence of a Boolean function in terms of its action on a certain Brownian motion.

2 Preliminaries

We state the main stochastic calculus result we will need in the proof. This is Dynkin’s formula [8, Theorem 7.4.1] specialized to our scenario of the Brownian motion having mean 0 and constant covariance.

▶ **Theorem 1.** *Let $\mathbf{X}$ be an n-dimensional Brownian motion with mean 0 and covariance Σ, let τ be a bounded stopping time, and let $f : \mathbb{R}^N \to \mathbb{R}$ be a twice continuously differentiable function. We use $\mathrm{H}f$ to denote the Hessian of f, the $N \times N$ matrix of second order partial derivatives. The following holds:*

$$\mathbf{E}[f(\mathbf{X}_\tau)] = f(0) + \mathbf{E}\left[\int_0^\tau \frac{1}{2}\langle \Sigma, \mathrm{H}f(\mathbf{X}_s)\rangle\, ds\right].$$

We also need the following formula regarding random restrictions, which is essentially Lemma 1 of [13]. A similar idea appears in the proof of Lemma 5.1 of [7], and previously in [5, Claim A.5].

▶ **Lemma 2.** *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 $S \subseteq [N]$,*

$$\partial_S f(x) = 2^{|S|} \mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_x}[\partial_S f_\rho(0)].$$

Here we write $\partial_S = \prod_{i \in S} \frac{\partial}{\partial_i}$ *for the partial derivatives over the coordinates in* S. *We further define the* Fourier coefficient $\widehat{f}(S) := \partial_S f(0)$. *Note that this coincides with the usual decomposition* $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \prod_{i \in S} x_i$.

3 A bound on a product of Brownian motions

▶ **Definition 3.** *Let* $k \in \mathbb{N}_+$, *and let* $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)}$ *be identical independent* N*-dimensional Brownian motions with mean 0 and covariance matrix* Σ, *and let* $\tau_1, \ldots, \tau_k$ *be stopping times.*

We consider distributions on $\mathbb{R}^{kN} \cong (\mathbb{R}^N)^k$, *which we take to be* k *copies of* $\mathbb{R}^N$, *indexed by coordinates* $1, \ldots, k$. *Let* $S \subseteq [k]$. *Define the random variable* $\mathbf{X}^S_\tau$ *to be* $\mathbf{X}^{(i)}_{\tau_i}$ *in the* i^{th} *coordinate if* $i \in S$, *and* 0 *in the* i^{th} *coordinate if* $i \notin S$. *We set* $\mathbf{D}_S$ *to be the distribution of* $\mathbf{X}^S_\tau$.

We write $\mathbf{S} \sim [k]$ *to denote drawing* $\mathbf{S} \subseteq [k]$ *uniformly. We now define the distribution* $\mathbf{D}_{\text{odd},\,k}$ *to be the distribution of* $\mathbf{D}_\mathbf{S}$ *conditioned on* $|\mathbf{S}|$ *being odd. Similarly, we define* $\mathbf{D}_{\text{even},\,k}$ *to be* $\mathbf{D}_\mathbf{S}$ *conditioned on* $|\mathbf{S}|$ *being even. When* $k = 1$, *we define* $\mathbf{D}_1 = \mathbf{X}^{(1)}_{\tau_1} = \mathbf{D}_{\text{odd},\,1}$.

For a multilinear function $f : \mathbb{R}^{kN} \to \mathbb{R}$, *we note the identity*

$$\mathbf{E}[f(\mathbf{D}_{\text{even},\,k})] - \mathbf{E}[f(\mathbf{D}_{\text{odd},\,k})] = 2 \mathop{\mathbf{E}}_{\mathbf{S} \sim [k]} \Big[(-1)^{|\mathbf{S}|} f(\mathbf{D}_\mathbf{S}) \Big]. \tag{1}$$

The following bounds how well a Boolean function with bounded level-$2k$ Fourier weight can distinguish $\mathbf{D}_{\text{even},\,k}$ and $\mathbf{D}_{\text{odd},\,k}$. This is essentially Theorem 3.1 of [7].

▶ **Theorem 4.** *Let* $k \in \mathbb{N}_+$, *let* $f : \{-1,1\}^{kN} \to \{-1,1\}$ *be a Boolean function, and let* $L > 0$ *be such that for any restriction* ρ,

$$\sum_{\substack{S \subseteq [kN] \\ |S| = 2k}} |\widehat{f}_\rho(S)| \le L.$$

Let $\gamma > 0$ *and let* $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k)}$ *be identical independent* N*-dimensional Brownian motions with mean 0 and covariance matrix* Σ. *Further assume that* $|\Sigma_{ij}| \le \gamma$ *for* $i \ne j$.

Let $\varepsilon > 0$ *and define the (bounded) stopping times for each* $i \in [k]$,

$$\tau_i := \min\{\varepsilon,\ \textit{first time that } \mathbf{X}^{(i)} \textit{ exits } [-1/2, 1/2]^N\}.$$

Then, identifying f *with its multilinear expansion, we have*

$$\Big|\mathop{\mathbf{E}}_{\mathbf{S} \sim [k]} \Big[(-1)^{|\mathbf{S}|} f(\mathbf{D}_\mathbf{S}) \Big]\Big| \le (\varepsilon\gamma)^k L.$$

Proof. We first prove by induction on k that for any multilinear function f,

$$\mathop{\mathbf{E}}_{\mathbf{S} \sim [k]} \big[(-1)^{|\mathbf{S}|} f(\mathbf{D}_\mathbf{S}) \big] = \mathbf{E}\left[\int_0^{\tau_1} \cdots \int_0^{\tau_k} \frac{(-1)^{k-1}}{2^{2k-1}} \Big\langle (I_k \otimes \Sigma)^{\otimes k},\ \mathrm{H}^{\otimes k} f(\mathbf{X}^{(1)}_{t_1}, \ldots, \mathbf{X}^{(k)}_{t_k}) \Big\rangle\, dt_1 \ldots dt_k \right], \tag{2}$$

where $\mathrm{H}^{\otimes k} f$ denotes the $(kN)^k$-dimensional matrix of all the $2k^{\text{th}}$ order derivatives of f, I_k is the k-dimensional identity matrix, and $\otimes$ denotes the Kronecker product.

The base case $k = 1$ is simply a direct application of Dynkin's formula (Theorem 1):

$$\mathbf{E}[f(\mathbf{X}^{(1)}_{\tau_1})] - f(0) = \mathbf{E}\left[\int_0^{\tau_1} \frac{1}{2} \Big\langle \Sigma,\ \mathrm{H} f(\mathbf{X}^{(1)}_{t_1}) \Big\rangle\, dt_1 \right].$$

For the induction step, we condition on the last coordinate to observe that

$$\mathop{\mathbf{E}}_{\mathbf{S}\sim[k]}\left[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}})\right]=\frac{1}{2}\mathop{\mathbf{E}}_{\mathbf{S}\sim[k-1]}\left[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}},0)\right]-\frac{1}{2}\mathop{\mathbf{E}}_{\mathbf{S}\sim[k-1]}\left[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}},\mathbf{X}^{(k)}_{\tau_k})\right] \tag{3}$$

Now let $\mathbf{S}\sim[k-1]$. We will proceed by applying Dynkin's formula to $g(x)=\mathbf{E}_{\mathbf{S},\mathbf{D}_{\mathbf{S}}}[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}},x)]$. Since f is a multilinear function, partial derivatives of g commute with the expectation; in particular $\partial_i g(x)=\mathbf{E}[(-1)^{|\mathbf{S}|}\partial_{i+(k-1)N}f(\mathbf{D}_{\mathbf{S}},x)]$, and so, with $e_k\in\mathbb{R}^k$ denoting the indicator of the k^{th} coordinate,

$$\begin{aligned}&\mathop{\mathbf{E}}_{\mathbf{S},\mathbf{D}_{\mathbf{S}},\mathbf{X}^{(k)},\tau_k}[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}},\mathbf{X}^{(k)}_{\tau_k})]-\mathop{\mathbf{E}}_{\mathbf{S},\mathbf{D}_{\mathbf{S}}}[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}},0)]\\&\qquad=\mathop{\mathbf{E}}_{\mathbf{X}^{(k)},\tau_k}\left[\frac{1}{2}\int_0^{\tau_k}\left\langle e_ke_k^T\otimes\Sigma,\mathop{\mathbf{E}}_{\mathbf{S},\mathbf{D}_{\mathbf{S}}}\left[(-1)^{|\mathbf{S}|}\,\mathrm{H}f(\mathbf{D}_{\mathbf{S}},\mathbf{X}^{(k)}_{t_k})\right]\right\rangle dt_k\right]\end{aligned} \tag{4}$$

Finally, we apply the induction hypothesis to find, for $(k-1)N<i,j\leq kN$,

$$\begin{aligned}&\mathop{\mathbf{E}}_{\mathbf{S}\sim[k-1]}\left[(-1)^{|\mathbf{S}|}\partial_{i,j}f(\mathbf{D}_{\mathbf{S}},\mathbf{X}^{(k)}_{t_k})\right]\\&=\mathbf{E}\left[\int_0^{\tau_1}\cdots\int_0^{\tau_{k-1}}\frac{(-1)^{k-2}}{2^{2k-3}}\left\langle(I_{k-1}\otimes\Sigma)^{\otimes(k-1)},\ \mathrm{H}^{\otimes(k-1)}_{\mathbf{X}^{(1)},\dots,\mathbf{X}^{(k-1)}}\partial_{i,j}f(\mathbf{X}^{(1)}_{t_1},\dots,\mathbf{X}^{(k)}_{t_k})\right\rangle dt_1\dots dt_{k-1}\right].\end{aligned}$$

Combining with Equations (3) and (4) and using bilinearity of the inner product, we conclude

$$\begin{aligned}&\mathop{\mathbf{E}}_{\mathbf{S}\sim[k]}\left[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}})\right]\\&\quad=-\frac{1}{2}\mathop{\mathbf{E}}_{\mathbf{X}^{(k)},\tau_k}\left[\frac{1}{2}\int_0^{\tau_k}\left\langle e_ke_k^T\otimes\Sigma,\mathop{\mathbf{E}}_{\mathbf{S},\mathbf{D}_{\mathbf{S}}}\left[(-1)^{|\mathbf{S}|}\,\mathrm{H}f(\mathbf{D}_{\mathbf{S}},\mathbf{X}^{(k)}_{t_k})\right]\right\rangle dt_k\right]\\&\quad=\mathbf{E}\left[\int_0^{\tau_1}\cdots\int_0^{\tau_k}\frac{(-1)^{k-1}}{2^{2k-1}}\left\langle e_ke_k^T\otimes\Sigma,\mathrm{H}_{\mathbf{X}^{(k)}}\left\langle(I_{k-1}\otimes\Sigma)^{\otimes(k-1)},\right.\right.\right.\\&\qquad\qquad\left.\left.\left.\mathrm{H}^{\otimes(k-1)}_{\mathbf{X}^{(1)},\dots,\mathbf{X}^{(k-1)}}f(\mathbf{X}^{(1)}_{t_1},\dots,\mathbf{X}^{(k)}_{t_k})\right\rangle\right\rangle dt_1\dots dt_k\right]\\&\quad=\mathbf{E}\left[\int_0^{\tau_1}\cdots\int_0^{\tau_k}\frac{(-1)^{k-1}}{2^{2k-1}}\left\langle(I_k\otimes\Sigma)^{\otimes k},\ \mathrm{H}^{\otimes k}f(\mathbf{X}^{(1)}_{t_1},\dots,\mathbf{X}^{(k)}_{t_k})\right\rangle dt_1\dots dt_k\right].\end{aligned}$$

Having completed the proof of Equation (2), we now use it to prove the theorem

$$\begin{aligned}&\mathop{\mathbf{E}}_{\mathbf{S}\sim[k]}\left[(-1)^{|\mathbf{S}|}f(\mathbf{D}_{\mathbf{S}})\right]\\&\quad\leq\varepsilon^k\,\mathbf{E}\left[\sup_{\substack{t_1\in[0,\tau_1]\\ \vdots\\ t_k\in[0,\tau_k]}}\left|\frac{1}{2^{2k}}\left\langle(I_k\otimes\Sigma)^{\otimes k},\ \mathrm{H}^{\otimes k}f(\mathbf{X}^{(1)}_{t_1},\dots,\mathbf{X}^{(k)}_{t_k})\right\rangle\right|\right] && (\tau_1,\dots,\tau_k\leq\varepsilon)\\&\quad\leq\frac{(\varepsilon\gamma)^k}{2^{2k}}\sup_{(x_1,\dots,x_k)\in[-1/2,1/2]^{kN}}\sum_{\substack{S\subseteq[kN]\\|S|=2k}}|\partial_Sf(x_1,\dots,x_k)| && (|\Sigma_{ij}|\leq\gamma\text{ for }i\neq j,\ \partial_{ii}f=0)\\&\quad\leq(\varepsilon\gamma)^k\sup_{(x_1,\dots,x_k)\in[-1/2,1/2]^{kN}}\sum_{\substack{S\subseteq[kN]\\|S|=2k}}\left|\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_{x_1,\dots,x_k}}[\partial_Sf_\rho(0,\dots,0)]\right| && (\text{Lemma 2})\\&\quad\leq(\varepsilon\gamma)^k\sup_{(x_1,\dots,x_k)\in[-1/2,1/2]^{kN}}\mathop{\mathbf{E}}_{\rho\sim\mathcal{R}_{x_1,\dots,x_k}}\left[\sum_{\substack{S\subseteq[kN]\\|S|=2k}}|\widehat{f_\rho}(S)|\right]\\&\quad\leq(\varepsilon\gamma)^kL.\end{aligned}$$

◀

4 Application to complexity of k-XOR Forrelation

We now apply the bound from the previous section to prove the main theorem from [7, Theorem 3.1], from which they derive separations in quantum versus classical query complexity, communication complexity and circuit complexity (we refer to [7] for the exact details about the definitions of the complexity classes and the full proof).

We briefly sketch how the proof in [7] proceeds. For the lower bounds on the classical complexity classes, it suffices to exhibit two distributions that are hard for functions in the complexity class to distinguish. These will be derived from $\mathbf{D}_{\text{odd},\,k}$ and $\mathbf{D}_{\text{even},\,k}$, with Σ and ε chosen appropriately ([7] uses the truncated Gaussian instead of the stopping time here). [7] observes that these classical complexity classes are closed under restrictions. Then, Theorem 4 and Equation (1) combined with bounds on the level-k Fourier weights proven in [11] and [7] shows the classical lower bounds.

For the quantum upper bound, we need to show that a quantum query algorithm (or communication protocol respectively) can distinguish $\mathbf{D}_{\text{odd},\,k}$ and $\mathbf{D}_{\text{even},\,k}$ with high probability. We will show that the concentration results proven in [7] hold in our context as well.

We now set values for ε and k. We take $\varepsilon = 1/(28k^2 \ln N)$, and k small enough that $\varepsilon^2 N \le \text{poly}(N)$ (e.g. $k \le O(N^{1/5})$ suffices), and set

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

where $N = 2n$, n is a power of 2 and H_n is the normalized Hadamard matrix, so $\gamma = \frac{1}{\sqrt{n}}$. Applying Theorem 4, the overall upper bound is $L_{2k} \cdot \text{polylog}(N)/N^k$, where L_{2k} is the bound on the Fourier weight at level $2k$ for the family of functions in the complexity class.

The quantum algorithm/communication protocol is based on the k-XOR Forrelation problem, which we define here: Let $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ as $\phi(x, y) := \frac{1}{n}\langle x, H_n y\rangle$. The Forrelation decision problem is a partial function defined by

$$F(x, y) = \begin{cases} -1 & \text{if } \phi(x, y) \ge \varepsilon/2, \\ 1 & \text{if } \phi(x, y) \le \varepsilon/4. \end{cases}$$

The k-XOR Forrelation $F^{(k)} : \{-1, 1\}^{kN} \to \{-1, 1\}$ is defined by $F^{(k)}(z_1, \dots, z_k) = \prod_{i=1}^{k} F(z_i)$.

Since $\mathbf{D}_{\text{odd},\,k}$ and $\mathbf{D}_{\text{even},\,k}$ take values in $[-1/2, 1/2]^{kN}$ but $F^{(k)}$ is defined on $\{-1, 1\}^{kN}$, we round them to distributions $\widetilde{\mathbf{D}}_{\text{odd},\,k}$ and $\widetilde{\mathbf{D}}_{\text{even},\,k}$ on $\{-1, 1\}^{kN}$. A draw of $\widetilde{\mathbf{z}} \sim \widetilde{\mathbf{D}}_{\text{odd},\,k}$ is defined as follows:

nosep Sample $\mathbf{z} \sim \mathbf{D}_{\text{odd},\,k}$.

nosep For each coordinate $i \in [N]$, independently set $\widetilde{\mathbf{z}}_i = 1$ with probability $\frac{1+\mathbf{z}_i}{2}$ and -1 with probability $\frac{1-\mathbf{z}_i}{2}$. We denote $\widetilde{\mathbf{z}} \sim \mathbf{z}$ for this step. Now $\widetilde{\mathbf{z}}$ is sampled from $\widetilde{\mathbf{D}}_{\text{odd},\,k}$.

$\widetilde{\mathbf{D}}_{\text{even},\,k}$ is defined analogously. Note that for a multilinear polynomial $f : \mathbb{R}^N \to \mathbb{R}$, $\mathbf{E}[f(\mathbf{D}_{\text{odd},\,k})] = \mathbf{E}[f(\widetilde{\mathbf{D}}_{\text{odd},\,k})]$ (analogously for $\widetilde{\mathbf{D}}_{\text{even},\,k}$).

Girish, Raz and Zhan showed the following about the rounding process:

▶ **Proposition 5** (Claim A.2 [7]). *Let $z \in [-1/2, 1/2]$, and let $\widetilde{\mathbf{z}} \sim z$ as in step 2 above. Then,*

$$\mathbf{P}[|\phi(\widetilde{\mathbf{z}}) - \phi(z)| > \varepsilon/4] \le \exp(-\Omega(N^{1/4})).$$

Finally, we show a concentration result analogous to Lemma 2.11 of [7] which shows that $F^{(k)}$ decides correctly on $\widetilde{\mathbf{D}}_{\mathrm{odd},\,k}$ and $\widetilde{\mathbf{D}}_{\mathrm{even},\,k}$ with high probability. This is then sufficient to deduce the applications described in [7].

First, we prove a concentration bound for $\phi(\mathbf{D}_1)$, i.e. $(\mathbf{x},\mathbf{y})$ are generated by a single N-dimensional stopped Brownian motion with covariance Σ.

▶ **Lemma 6.** *In the above context, the following holds:*

$$\mathop{\mathbf{P}}_{(\mathbf{x},\mathbf{y})\sim\mathbf{D}_1}[\phi(\mathbf{x},\mathbf{y}) \geq 3\varepsilon/4] \geq 1 - O(1/N^{6k^2}). \tag{5}$$

Proof. Notice that an alternate way to sample $(\mathbf{x},\mathbf{y}) \sim \mathbf{D}_1$ is to let $\mathbf{X}_t$ be a n-dimensional Brownian motion with covariance I_n stopped at the stopping time

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

and let $(\mathbf{x},\mathbf{y}) = (\mathbf{X}_\tau, H_n\mathbf{X}_\tau)$. Then, $\phi(\mathbf{x},\mathbf{y}) = \frac{1}{n}\|\mathbf{X}_\tau\|_2^2$. In order to prove the desired bound, we first prove that with high probability $\tau = \varepsilon$, i.e. the path of the Brownian motion did not exit $[-1/2, 1/2]^N$ before time ε. We then show that $\frac{1}{n}\|\mathbf{X}_\varepsilon\|_2^2 \geq 3\varepsilon/4$ with high probability, and conclude using a union bound. We can union bound over the N coordinates,

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

Since each coordinate of $\mathbf{X}_t$ is a standard 1D Brownian motion $\mathbf{B}_t$, we can apply Doob's submartingale inequality (e.g. [10, Proposition II.1.8]) to obtain

$$\mathbf{Pr}\left[\sup_{0\leq t\leq\varepsilon/2} |\mathbf{B}_t| \geq \frac{1}{2}\right] \leq 2e^{-1/4\varepsilon} = 2e^{-7k^2\ln N} = \frac{2}{N^{7k^2}}.$$

Therefore,

$$\mathbf{Pr}[\tau < \varepsilon] \leq 2/N^{7k^2-1}. \tag{6}$$

Next, we consider $\frac{1}{n}\|\mathbf{X}_\varepsilon\|_2^2$. Note that this is simply the average of the squares of n iid Gaussians $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with mean 0 and variance ε. Using [12, Example 2.11], we have the tail bound

$$\mathbf{Pr}\left[|\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^2 - \varepsilon| \geq \frac{\varepsilon}{4}\right] \leq \exp(-\Omega(N)). \tag{7}$$

Taking a union bound over Equations (6) and (7), we have $\mathbf{P}_{(\mathbf{x},\mathbf{y})\sim\mathbf{D}_1}[\phi(\mathbf{x},\mathbf{y}) \leq 3\varepsilon/4] \leq O(1/N^{6k^2})$. ◀

▶ **Proposition 7.** *The following hold:*

$$\mathop{\mathbf{P}}_{\widetilde{\mathbf{z}}\sim\widetilde{\mathbf{D}}_{\mathrm{even},\,k}}[F^{(k)}(\widetilde{\mathbf{z}}) = 1] \geq 1 - O\Big(\frac{k}{N^{6k^2}}\Big) \qquad \textit{and} \qquad \mathop{\mathbf{P}}_{\widetilde{\mathbf{z}}\sim\widetilde{\mathbf{D}}_{\mathrm{odd},\,k}}[F^{(k)}(\widetilde{\mathbf{z}}) = -1] \geq 1 - O\Big(\frac{k}{N^{6k^2}}\Big).$$

Proof. We first show F decides correctly on the coordinates with $\mathbf{U}_N$ with high probability:

$$\mathop{\mathbf{P}}_{(\mathbf{x},\mathbf{y})\sim\mathbf{U}_N}[\phi(\mathbf{x},\mathbf{y}) \leq \varepsilon/4] \geq 1 - \exp(-\Omega(N\varepsilon^2)). \tag{8}$$

To see this, note that $\mathbf{x}$ and $\mathbf{y}$ are independent, so $\mathbf{x}$ and $H_n\mathbf{y}$ are independent. Hence $\phi(\mathbf{x},\mathbf{y})$ is simply the average of random signs, and the bound holds by Hoeffding's inequality.

Finally, to prove the proposition it suffices to prove that for any fixed $S \subseteq [k]$, $\mathbf{P}_{\mathbf{z}\sim\mathbf{D}_S,\widetilde{\mathbf{z}}\sim\mathbf{z}}[F^{(k)}(\widetilde{\mathbf{z}}) \neq (-1)^{|S|}] \leq O(k/N^{6k^2})$. If $i \in S$, then $\widetilde{\mathbf{z}}_i$ is distributed as $\widetilde{\mathbf{D}}_1$, so Lemma 6 combined with Proposition 5 implies $\mathbf{P}[F(\widetilde{\mathbf{z}}_i) = 1] \leq O(1/N^{6k^2})$. Meanwhile if $i \notin S$, then $\widetilde{\mathbf{z}}_i$ is distributed as $\mathbf{U}_N$, so Equation (8) implies $\mathbf{P}[F(\widetilde{\mathbf{z}}_i) = -1] \leq \exp(-\Omega(N\varepsilon^2))$. With k and therefore ε taken sufficiently small, a union bound over the k coordinates completes the proof. ◀

References

1 Scott Aaronson and Andris Ambainis. Forrelation: a problem that optimally separates quantum from classical computing. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 307–316, 2015.

2 Nikhil Bansal and Makrand Sinha. k-Forrelation optimally separates quantum and classical query complexity. Technical Report 2008.07003, arXiv, 2020.

3 F. Barthe and B. Maurey. Some remarks on isoperimetry of Gaussian type. *Ann. Inst. H. Poincaré Probab. Statist.*, 36(4):419–434, 2000. `doi:10.1016/S0246-0203(00)00131-X`.

4 Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory Comput.*, 15:Paper No. 10, 26, 2019. `doi:10.4086/toc.2019.v015a010`.

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 *Proceedings of the 10th Annual Innovations in Theoretical Computer Science Conference*, pages 22:1–22:15, 2018.

6 Ronen Eldan and Renan Gross. Concentration on the Boolean hypercube via pathwise stochastic analysis. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing*, pages 208–221. ACM, New York, 2020.

7 Uma Girish, Ran Raz, and Wei Zhan. Lower bounds for XOR of forrelations. Technical report, arXiv, 2020. `arXiv:2007.03631`.

8 Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications. `doi:10.1007/978-3-642-14394-6`.

9 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. `doi:10.1145/3313276.3316315`.

10 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. `doi:10.1007/978-3-662-06400-9`.

11 Avishay Tal. Towards optimal separations between quantum and randomized query complexities. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science*, pages 228–239. IEEE Computer Soc., Los Alamitos, CA, 2020. `doi:10.1109/FOCS46700.2020.00030`.

12 Martin J. Wainwright. *High-dimensional statistics*, volume 48 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. A non-asymptotic viewpoint. `doi:10.1017/9781108627771`.

13 Xinyu Wu. A stochastic calculus approach to the oracle separation of BQP and PH. Technical report, arXiv, 2020. `arXiv:2007.02431`.