LECTURE 9 (February 14^{th})

TODAY BQP vs PH (part 2)

RECAP $\exists$ a problem s.t. → called the Fourier Correlation problem

(1) A quantum algorithm can solve it with one query with success probability

$$\frac{1}{2} + \frac{1}{\text{polylog}(N)}$$

← One can make this $\frac{1}{2} + 0.1$ but its more complicated and we won't cover it here

(2) Any AC^0 circuit of size $2^{\text{polylog}(N)}$ has success probability

atmost $\frac{1}{2} + \frac{\text{polylog}(N)}{\sqrt{N}} \ll \frac{1}{2} + \frac{1}{N^{1/2 - o(1)}}$

$\Longrightarrow$ Using diagonization and the connection between PH-oracle machines and AC^0 circuit this implies that

$$\exists\ O \text{ s.t. } BQP^O \not\subseteq PH^O$$

Fourier Correlation or Forrelation Problem

introduced by Aaronson

Input $x_1, \dots, x_N, y_1, \dots, y_N \in \{\pm 1\}^{2N}$ $\Longrightarrow$ One can encode this with $2n$ qubits where $N = 2^n$

Promise Problem:

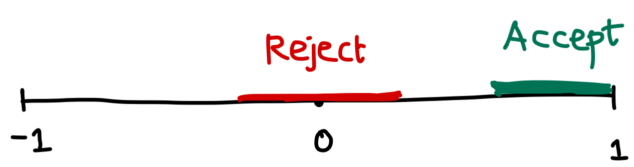
Decide if $\frac{\langle x, Hy \rangle}{N} \geq \frac{1}{32 \log N}$ "Accept"

$\frac{|\langle x, Hy \rangle|}{N} \leq \frac{1}{64 \cdot \log N}$ "Reject"

$H = H^{\otimes n}$ is the Hadamard matrix of size $2^n \times 2^n = N \times N$

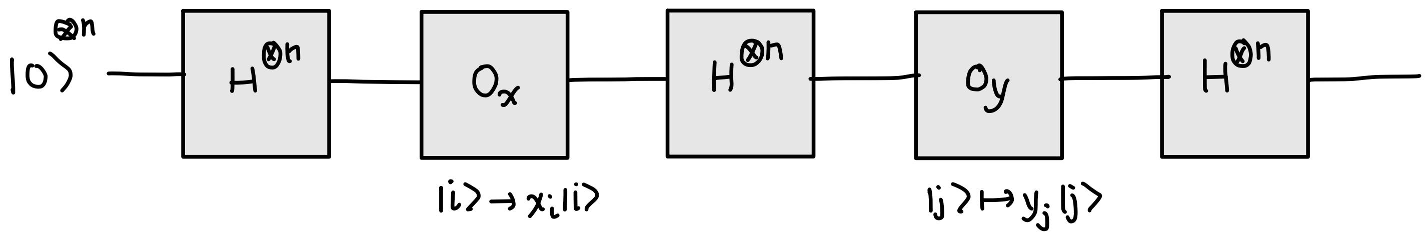
Note, $\frac{x}{\sqrt{N}}$ and $\frac{y}{\sqrt{N}}$ are unit vectors and H is a unitary matrix

so, $\frac{\langle x, Hy \rangle}{N} \in [-1, 1]$



Also, note $\frac{\langle x, Hy \rangle}{N} = \sum_{ij} x_i y_j \frac{H_{ij}}{N}$

Connection to Quantum Circuits



The final state of this circuit (before measurement) in the computational basis $|0\rangle, |1\rangle, \ldots |N\rangle$ looks like

$$\frac{\langle x, Hy\rangle}{N}|0\rangle + \_\_\_|1\rangle + \_\_\_|2\rangle + \ldots \qquad \text{(Exercise)}$$

In the exercises, you saw how to construct a quantum algorithm for Forrelation

Today we will see that no AC^0-circuit of size $2^{\text{polylog}(N)}$ can solve Forrelation

Lower Bounds for AC^0 circuit

Recalling our general recipe for proving lower bounds, we need to come up with a candidate hard distribution on inputs $(x_1 \ldots x_N, y_1 \ldots y_N) = (x, y)$

Experience tells us to try the following distribution first

$$\begin{cases} \text{with probability } \frac{1}{2} \quad (x,y) \in \{\pm 1\}^{2N} \text{ sampled uniformly conditioned on } \frac{\langle x, Hy\rangle}{N} \geq \frac{1}{32\log N} & \text{"Accept"} \\ \text{with probability } \frac{1}{2} \quad (x,y) \in \{\pm 1\}^{2N} \text{ sampled uniformly conditioned on } \frac{|\langle x, Hy\rangle|}{N} \leq \frac{1}{64\log N} & \text{"Reject"} \end{cases}$$

The problem here is that this distribution is hard to analyze, so we will introduce a different way of generating hard distributions by rounding continuous distributions to $\{\pm 1\}$-values

Let $(U, V) \in \mathbb{R}^{2N}$ be a Gaussian with covariance $\sigma^2 \begin{bmatrix} I_N & H_N \\ H_N & I_N \end{bmatrix}$ & mean 0

$\sigma = \frac{1}{\sqrt{16 \cdot \log N}}$

Note that $U \in \mathbb{R}^N$ is a standard Gaussian in $\mathbb{R}^N$ with independent coordinates with mean 0 & variance σ^2, and so is $V \in \mathbb{R}^N$

i,j entry of the covariance matrix of a multi-variate Gaussian $G \in \mathbb{R}^m$ is $\mathbb{E}[G_i G_j]$

But U & V are correlated and

$$\mathbb{E}[U_i V_j] = \sigma^2 H_N(i,j) = \pm \frac{\sigma^2}{\sqrt{N}}$$

$\frac{1}{26}$

Moreover, for a Gaussian in 1-dimension with mean 0 & variance σ^2

$$\mathbb{P}[\,|G| \geq \sigma t\,] \leq 2e^{-t^2/2}$$

$$\mathbb{P}\left[\,|G| \geq \frac{1}{2}\right] \leq 2e^{-\left(\frac{1}{2\sigma}\right)^2/2} = 2e^{-(1/8\sigma^2)} = \frac{2}{N^2} \quad \text{since } \sigma = \frac{1}{\sqrt{16\log N}}$$

By union bound this means that with probability $1-N^{-1}$ all coordinates of U & V are in $\left[-\frac{1}{2}, \frac{1}{2}\right]$ } We are going to assume that this happens with probability 1

Now, how do we round them to $\{\pm 1\}$ values?

Given a value $\beta \in [-1,1]$,

$$\left.\begin{aligned} \mathbb{P}[x = +1] &= \frac{1}{2} + \frac{\beta}{2} \\ \mathbb{P}[x = -1] &= \frac{1}{2} - \frac{\beta}{2} \end{aligned}\right\} \mathbb{E}[x] = \beta$$

$\Longrightarrow$ We do this to each coordinate of $(U,V) \in \mathbb{R}^{2N}$ to obtain $(x,y) \in \{\pm 1\}^{2N}$

$$\mathbb{E}[(x,y)] = (U,V)$$

Why this distribution? Consider $\mathbb{E}\frac{\langle x, Hy\rangle}{N}$ under this distribution

$$\frac{1}{N}\mathbb{E}\langle x, Hy\rangle = \frac{1}{N}\sum_{ij} H_N(i,j)\,\mathbb{E}[x_i y_j] = \frac{1}{N}\sum_{ij} H_N(i,j)\,\underbrace{\mathbb{E}[U_i V_j]}_{= H_N(i,j)\cdot\sigma^2}$$

$$= \frac{\sigma^2}{N}\sum_{ij} \underbrace{H_N(i,j)^2}_{=\frac{1}{N}} = \frac{\sigma^2}{N^2}N^2 = \sigma^2 = \frac{1}{2\log N}$$

In Expectation, this distribution has large Fourier Correlation "Accept"

To summarize, Gaussian

$$\frac{1}{2\sqrt{\log N}}\begin{bmatrix} I_N & H_N \\ H_N & I_N \end{bmatrix} \xrightarrow{\text{Round}} \{\pm 1\}^{2N}$$

On the other hand,

Independent Gaussian

$$\frac{1}{2\sqrt{\log N}}\begin{bmatrix} I_N & 0 \\ 0 & I_N \end{bmatrix} \xrightarrow{\text{Round}} \{\pm 1\}^{2N} \text{ uniform distribution}$$

"Reject"

In expectation, this distribution has low Fourier Correlation $\left(\leq \frac{1}{\sqrt{N}}\right)$
(actually also with high probability)

From what you have shown in the exercises

$\exists$ a quantum algorithm s.t.

$$\left| \mathbb{E}_{x,y \in \text{first distribution}} [\text{Alg "accepts" } x,y] - \mathbb{E}_{x,y \in \text{unif}} [\text{Alg. accepts } x,y] \right| \geq \frac{1}{32 \log N}$$

We are going to show that the above is small for any AC^0 circuit

In fact, we are going to prove a general purpose statement in terms of Fourier coefficients

Fourier Analysis over $\{\pm1\}^m$ "101"

Any function $f : \{\pm1\}^m \longrightarrow \mathbb{R}$ can be expressed as a multilinear polynomial

[1] $$f(x) = \sum_{S \subseteq [m]} \hat{f}(S) \prod_{i \in S} x_i$$

[We have seen quantum algs. give such polynomials of low degree but here degree can be m]

This is called the Fourier expansion of f

Some intuition behind why this should be true:

A function $f : \{\pm1\}^m \to \mathbb{R}$ can be written as a vector $(f(x))_{x \in \{\pm1\}^m}$ of length 2^m

One can equivalently write this as

$$f(x) = \sum_{a \in \{\pm1\}^m} f(a)\, \mathbb{1}[x = a]$$

The functions $\{\mathbb{1}[x=a]\}_a$ forms an orthogonal basis for the space of functions under the inner product $\langle f, g \rangle = \mathbb{E}_x[f(x) g(x)]$

Note that $\langle 1[x=a], 1[x=a] \rangle = 2^{-m}$

Taking the Fourier Transform of f represents $f(x)$ in the basis of monomials $\left(\prod_{i \in S} x_i\right)_{S \subseteq [m]}$ $\leftarrow$ orthonormal basis under the inner product defined above $\mathbb{E}\left[\left(\prod_{i \in S} x_i\right)^2\right] = 1$

as the vector $\left(\hat{f}(S)\right)_{S \subseteq [m]}$

Moreover, this change of basis is a unitary transformation
so, Euclidean lengths remain the same in the two basis (after normalizing)

$$\frac{1}{2^m} \sum_x f(x)^2 = \sum_{S \subseteq [m]} |\hat{f}(S)|^2$$

[2] i.e. $\mathbb{E}_x [|f(x)|^2] = \sum_{S \subseteq [m]} |\hat{f}(S)|^2$ (Parseval's identity)

The last point to pay attention to is that

[3] $$\hat{f}(S) = \partial_S f(0)$$

$f(x_1, x_2, x_3) = x_1 + 2x_1x_2 + 3x_1x_2x_3$

$\partial_{\{1,2\}} f(x_1, x_2, x_3) = 2 + 3x_3$

$\Rightarrow \partial_{\{1,2\}} f(0) = 2$

Lower Bounds for Fourier Correlation

$\frac{\langle x, Hy \rangle}{N} = \sum \frac{H_{ij}}{N} x_i y_j$ is a degree 2 polynomial $\Rightarrow$ computed by a quantum algorithm

On the other hand, any function (in particular those computed by AC^0 circuits) can also be written as a polynomial of very large degree

For instance, recall that even approximating the OR function on N bits (which can be computed by an AC^0 circuit of size 1) needs $\sqrt{N}$ degree

So, why can't such large degree polynomials compute Fourier Correlation?

The key message: The difference is sparsity and we need a notion that says that polynomials computed by AC^0-circuits (or other classical models) are sparse in some sense

How do we capture sparsity? A good proxy is ℓ_1-norm of coeffecients

Here, we need a more refined notion:
ℓ_1-norm of coefficients of a particular degree

In particular, define $wt_k(f, 0) = \sum_{|S|=k} |\hat{f}(S)|$ — sum of absoulte values of all degree k coefficients

$$= \sum_{|S|=k} |\partial_S f(0)| \quad \text{[By [3]]}$$

Similarly, $wt_k(f,u) = \sum_{|S|=k} |\partial_S f(u)|$ for $u \in [-1,1]^{2N}$

This is still a notion of sparsity since one can show that

$$wt_k(f,0) \le \max_{u \in [-\frac{1}{2},\frac{1}{2}]^{2N}} wt_k(f,u) \le 16\, wt_k(f,0)$$ ← We are not going to prove this here

[Main Lemma] by Raz-Tal

$$\left| \mathbb{E}_{\substack{\text{large}\\\text{Fourier}\\\text{Corr}}} [f \text{ accepts}] - \mathbb{E}_{\text{unif}} [f \text{ accepts}] \right|$$

$$\le \max_{u \in [-\frac{1}{2},\frac{1}{2}]^{2N}} wt_2(f,u) \cdot \frac{\epsilon^2}{\sqrt{N}}$$

Note that only second derivatives of f matter

AC^0-circuits of $2^{\text{polylog}(N)}$ size have bounded derivatives $f = AC^0$ circuit output

$$\max_u wt_2(f,u) \le \text{polylog}(N)$$

We won't prove this fact here

Plugging it in the above statement, we get that the difference is

$$\text{atmost} \quad \frac{\text{polylog}(N)}{\sqrt{N}} = \frac{1}{N^{1/2 - o(1)}}$$

Proof of Main Lemma   Let $f(x,y)$ be a multilinear polynomial in x & y

As we saw before $\mathbb{E}[x_i y_j] = \mathbb{E}[U_i V_j]$ where U & V were the underlying Gaussians

Similarly for any multilinear monomial e.g. $x_1 x_2 x_3 x_4 \; y_2 y_4 y_5 y_7$

$$\mathbb{E}[x_1 x_2 x_3 x_4 y_2 y_4 y_5 y_7] = \mathbb{E}[U_1 U_2 U_3 U_4 V_2 V_4 V_5 V_7]$$

Thus it suffices to compute

$$\mathbf{E}_{\text{Cov}\begin{bmatrix} I & H \\ H & I \end{bmatrix} \cdot \epsilon^2} [f(U,V)] - \mathbf{E}_{\text{Cov}\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \cdot \epsilon^2} [f(U,V)]$$

where $(U,V) \in \mathbb{R}^{2N}$ are gaussian with these covariances

↑ Complicated Gaussian   ↑ Simple Gaussian

Key idea   Interpolate between the two

E.g. $G(t) \in \mathbb{R}^{2N}$ to be the Gaussian with covariance

$$t \begin{bmatrix} I & H \\ H & I \end{bmatrix} \cdot \epsilon^2 + (1-t) \begin{bmatrix} I & H \\ H & I \end{bmatrix} \cdot \epsilon^2$$

At "time" 0, $G(0) =$ Simple Gaussian
$G(1) =$ Complicated Gaussian

If we can show that $\left|\frac{d}{dt}\mathbb{E}[f(G(t))]\right| \leq \text{small} \quad \forall\, t \in [0,1]$

$$\Rightarrow \left|\mathbb{E}[f(G(1))] - \mathbb{E}[f(G(0))]\right| = \left|\int_0^1 \frac{d}{dt}\mathbb{E}[f(G(t))]\right| \leq \text{small}$$

Gaussian Interpolation Formula exactly allows us to compute the "time" derivative

$$\frac{d}{dt}\mathbb{E}[f(G(t))] = \frac{1}{2}\sum_{i,j\in[2N]} \left(C^{final}_{i,j} - C^{initial}_{i,j}\right) \mathbb{E}\left[\partial_{ij} f(G(t))\right]$$

$C^{final}_{i,j}$: Final (i,j) covariance entry

$C^{initial}_{i,j}$: Initial (i,j) covariance entry

$$C^{final} - C^{initial} = \sigma^2 \underbrace{\begin{bmatrix} 0 & H_N \\ H_N & 0 \end{bmatrix}}_{2N \text{ columns}} \Big\}\, 2N \text{ rows} \Rightarrow \text{All entries} \leq \frac{\sigma^2}{\sqrt{N}} \text{ in absolute value}$$

$$\sum_{ij} \mathbb{E}\left[|\partial_{ij} f(G(t))|\right] \leq \max_{u} \sum_{ij} |\partial_{ij} f(u)|$$

assuming $G(t) \in [-1,1]^{2N}$. which holds w.h.p.

So, overall, we get $\left|\frac{d}{dt}\mathbb{E}[f(G(t))]\right| \leq \frac{\sigma^2}{\sqrt{N}} \cdot \left(\max_{u\in[-1,1]^{2N}} \sum_{ij} |\partial_{ij} f(u)|\right)$ □

To summarize,

$\exists$ a quantum algorithm s.t.

$$\left|\mathbb{E}_{x,y\in \text{first distribution}}[\text{Alg "accepts" } x,y] - \mathbb{E}_{x,y\in \text{unif}}[\text{Alg. accepts } x,y]\right| \geq \frac{1}{32\log N}$$

On the other hand, for any AC^0 circuit of size $2^{\text{polylog}(N)}$

$$\left|\mathbb{E}_{x,y\in \text{first distribution}}[\text{Alg "accepts" } x,y] - \mathbb{E}_{x,y\in \text{unif}}[\text{Alg. accepts } x,y]\right| \leq \frac{1}{N^{1/2-o(1)}}$$

This can be used to prove the lower bound for promise version of Fourier Correlation by a standard argument that we leave as an exercise