Indian Institute of Technology, Kanpur

CS682 : Quantum Computing

Course Project Report

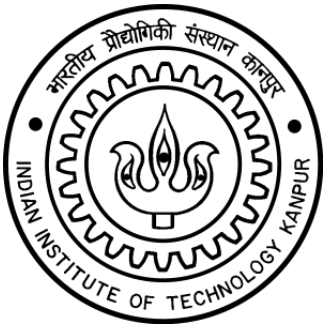
Quantum Proofs for Classical Theorems [1]

Andrew Drucker

Ronald de Wolf

Submitted To:
Rajat Mittal
Assistant Professor
CSE, IITK

Submitted By :
Rahul B S
Roll : 160540
CSE, IITK



Contents

1 Lower Bounds on Locally Decodable Codes

Holevo's Theorem

This theorem puts lower bound on how much information can be stored/ transmitted using qubits. It says that if we encode some classical random variable X in an m-qubit state, then no measurement on the quantum state can give more than m bits of information about X.

The idea of proof (Due to Nayak) is as follows :
Let, for each $x \in [N]$ is encoded in a d dimensional quantum state $x \to |\phi_x\rangle$ and $P_1, P_2, ..., P_N$ be the measurement operators used while decoding x, then we have -

$$p_x = \Pr(\text{correct}x) = ||P_x|\phi_x\rangle||^2 \leq \text{Tr}(P_x)$$
$$\implies \sum_x p_x \leq \sum_x \text{Tr}(P_x) = \text{Tr}(\sum_x P_x) = \text{Tr}(I_d) = d$$

So average probability of decoding any particular x is at most d/N.

Random Access Codes

Let us say instead of whole string x, we are interested in decoding one of the n bits of x i.e., x_i with some probability $p > 1/2$. Formally, for each $i \in [n]$ there should exist a measurement $M_i, I - M_i$ to recover x_i : for each n bit string x, we must have $||M_i|\phi_x\rangle||^2 \geq p$ if $x_i = 1$ and $||M_i|\phi_x\rangle||^2 \leq 1 - p$ if $x_i = 0$. Such encoding is called *quantum random access code*.

We can give lower bound on number of qubits required to encode classical bits of information. The idea (based on Nayak) here is that if we have can decode a bit x_i with success probability p_i then, we have $H(p_i)$ (entropy function) bits of uncertainty, which means we have $1 - H(p_i)$ bits of information, so overall from state $|\phi_x\rangle$ we can only decode $\sum_i (1 - H(p_i))$ bits of information.
Formalising this notion is locally decodable codes.

Locally Decodable Codes

$C : \{0,1\}^n \to \{0,1\}^N$ is a (q, δ, ϵ)-LDC if there is a classical randomized decoding algorithm A which can make at most q queries to N bit string y that can contain

error in atmost δN bits w.r.t $C(x)$, for all x is a n bit string and probability of correctly decoding all x_i's is $\geq 1/2 + \epsilon$. Note that probability is over random bits which Algorithm A takes as input.

An example of the above is $(2, \delta, \epsilon = 1/2 - 2\delta)$-LDC also called Hadamard Code, in which codeword $C(x)$'s length $N = 2^n$ i.e., exponential in n. Hadamard Code is defined as : for n bit string x, $C(x)_z = x.z \bmod 2$, for all $z \in \{0,1\}^n$ where (.) denotes bitwise inner product.
To decode any x_i from y (possibly some corrupt bits), we do following : choose any random $z \in \{0,1\}^n$, then

$$x.z \oplus x.(z \oplus e_i) = x.z \oplus x.z \oplus x.e_i = x_i$$

This means that we can decode x_i correctly with probability $\geq$ 1-2δ. It can be classically shown that Hadamard Code is optimal. Following subsection argues this using quantum query methods.

Optimality of Hadamard Code

The proof uses a result due to Katz and Trevisan + folklore, that says that : for every (q, δ, ϵ)-LDC $C : \{0,1\}^n \to \{0,1\}^N$, and for every $i \in [n]$ there exists a set M_i of $\Omega(\delta\epsilon N/q^2)$ disjoin tuples, each of at most q indices from $[N]$, and a bit $a_{i,t}$ for each tuple $t \in M_i$, such that following holds for all n bit strings :

$$\Pr_x \left[x_i = a_{i,t} \oplus \sum_{j \in t} C(x)_j \right] \geq 1/2 + \Omega(\epsilon/2^q)$$

For Hadamard Code the set of tuples become $M_i = \{z, z \oplus e_i\}$. Now consider the following N dimensional quantum encoding :

$$x \mapsto |\phi_x\rangle = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} (-1)^{C(x)_j} |j\rangle$$

If we show that this is infact quantum random access code for x, with some success probability $p > 1/2$, then we have number of qubits of this state atleast $(1 - H(p))n = \Omega(n)$ which is $\lceil \log N \rceil$ then proof is complete.
To recover x_i from state $|\phi_x\rangle$. We convert tuples to measurement operators : for each pair $(j,k) \in M_i$ define a projector $P_{jk} = |j\rangle\langle j| + |k\rangle\langle k|$ and let $P_{rest} = \sum_{i \notin \cup t \in M_i} |i\rangle\langle i|$ on remaining indices. So, P_{jk}'s are the (good)projectors on which we would like to see amplitude. The probability of observing amplitude on particular state P_{jk} is

$2/N$ but there are $|M_i| = \Omega(\delta\epsilon N)$ such pairs, so probability of observing amplitude on any of the good state is $(2/N)\Omega(\delta\epsilon N) = \Omega(\delta\epsilon)$ and probability of observing amplitude on P_{rest} is $1 - \Omega(\delta\epsilon)$.
Measurement using any P_{jk} will yield the state :

$$\frac{1}{\sqrt{2}}\Big((-1)^{C(x)_j}|j\rangle + (-1)^{C(x)_k}|k\rangle\Big) = \frac{(-1)^{C(x)_j}}{\sqrt{2}}\Big(|j\rangle + (-1)^{C(x)_j \oplus C(x)_k}|k\rangle\Big)$$

Now measuring the above state in the basis $(1/\sqrt{2})(|j\rangle \pm |k\rangle)$ will give the value of $C(x)_j \oplus C(x)_k$ with full certainty. Randomized algorithm will decode x_i with probability at least $1/2+\Omega(\epsilon)$. If we don't get good states (probability $= 1-\Omega(\delta\epsilon)$), then flip a fair coin to decide x_i. So, the probability of successfully recovering x_i is

$$p \geq \frac{1}{2}(1 - \Omega(\delta\epsilon)) + \Omega(\delta\epsilon)(\frac{1}{2} + \Omega(\epsilon)) = \frac{1}{2} + \Omega(\delta\epsilon^2)$$

Hence $\log N \geq n(1 - H(1/2 + \Omega(\delta\epsilon^2))) \implies N = 2^{\Omega(n\delta^2\epsilon^4)}$.

This means any 2 query LDC would required codelength which order of atleast exponential in length of the input string. Hence Hadamard Code is optimal.

2 Rigidity of Hadamard Matrices

Rigidity of a Matrix

The rigidity of M measures the minimal number of entries we need to change in order to reduce its rank to a given value r. Formally it is defined as:

$$R_M(r) = min\{d(M, \widetilde{M}) : rank(\widetilde{M}) \leq r\}$$

where $d(M, \widetilde{M})$ is Hamming distance. Bounded rigidity of M is defined as

$$R_M(r, \theta) = min\{d(M, \widetilde{M}) : rank(\widetilde{M}) \leq r, \max_{x,y} |M_{x,y} - \widetilde{M}_{x,y}| \leq \theta\}$$

For any matrix,clearly $R_M(r) \geq n - r$ since reducing the rank by 1 requires changing at least one entry. This bound is optimal for Identity matrix, but for others it is not tight bound.

Hadamard Matrix

Suppose we have a matrix $\widetilde{H}$ differing from the Hadamard matrix H in R positions, with $rank(\widetilde{H}) \leq r$. Using spectral methods it was show that $R \geq n^2/256r$, this can be improved to $R \geq n^2/4r$ by using quantum ideas.

The proof of lower bound on rigidity is based on following general idea: Alice sends Bob the n-dimensional quantumn state $|H_i\rangle$ corresponding to the normalized ith row of H, and Bob measures received state in Hadamard basis, he correctly gets to know i with probability 1.
Now instead of H, we have $\widetilde{H}$ which can be thought of approximating H in some way. Alice now sends the state $|\widetilde{H}_i\rangle$, this is done by taking r-orthonormal basis for the row space of $\widetilde{H}$, let it be $|v_1\rangle$,...,$|v_r\rangle$. To send $|\widetilde{H}_i\rangle = \sum_{j=1}^{r} \alpha_j |v_j\rangle$, Alice sends $\sum_{j=1}^{r} |v_j\rangle$ and then Bob applies unitary map $|j\rangle \mapsto |v_j\rangle$ to obtain $|\widetilde{H}_i\rangle$. He measures this in Hadamard basis, so probability of correctly decoding i is

$$p_i = |\langle H_i|\widetilde{H}_i\rangle|^2$$

Then we can apply upper bound on average success probability due to Nayak as

$$\sum_{i=1}^{n} p_i \leq r \implies \leq \frac{r}{n}$$

Theorem due to Lokam

Every $a \times b$ submatrix A of H has a rank $r \geq ab/n$.

Proof. Obtain rank-r matrix $\widetilde{H}$ from H by setting all entries of A to zero. Let $|\widetilde{H}_i\rangle$'s corresponding to non empty rows; they have normalization factor $1/\sqrt{b}$. For each i's, probability of successful decoding is

$$p_i = |\langle H_i|\widetilde{H}_i\rangle|^2 = \left|\frac{b}{\sqrt{bn}}\right|^2 = \frac{b}{n}$$

but communicating one of a posibilities using r dimensions impiles

$$\frac{1}{n}\sum_{i=1}^{n} p_i = p \leq \frac{r}{a}$$

Proof of Lower Bound

If $r \leq n/2$, then $R_H(r) \leq n^2/4r$.

Proof. Let $\widetilde{H}$ has $R = R_H(r)$ errors w.r.t H. There exists $a = 2r$ rows of $\widetilde{H}$ with at most aR/n errors on an average. Now consider submatrix A of $\widetilde{H}$ with a rows and $b \geq n - aR/n$ columns (that have no errors in these a rows). If $b = 0$, then $R \geq n^2/2r$ we are done, so assume A is nonempty. So this A is submatrix of H as it is errorfree, from Lokam's Theorem we have

$$r = rank(\widetilde{H}) \geq rank(A) \geq \frac{ab}{n} \geq \frac{a(n - aR/n)}{n} \implies R \geq n^2/4r$$

Note: If r $n/2$, H has eigenvalues are all $\pm\sqrt{n}$, so we can reduce its rank to $n/2$ or less by adding or subtracting matrix $\sqrt{n}I$. So, $R_H(n/2) \leq n$.

Bound on $R_H(r, \theta)$

Now consider the case where change is bounded by θ, we have:

$$_H(r, \theta) \geq \frac{n^2(n - r)}{2\theta n + r(\theta^2 + 2\theta)}$$

Proof. Let $\widetilde{H}$ has $R = R_H(r)$ errors w.r.t H and also $||H - \widetilde{H}||_\infty \leq \theta$. Let quantum state corresponding to its rows :

$$|\widetilde{H}_i\rangle = c_i \sum_{j=1}^{n} \widetilde{H}_{ij}|j\rangle$$

where c_i is normalizing constant. We can see that $c_i^{-2} = \sum_j \widetilde{H}_{ij}^2 \leq (n - d(H_i, \widetilde{H}_i)) + d(H_i, \widetilde{H}_i)(1+\theta)^2 = n + d(H_i, \widetilde{H}_i)(\theta^2 + 2\theta)$, where $d(.)$ is Hamming distance. Now probability of successfully decoding p_i is

$$\begin{aligned} p_i &= |\langle H_i|\widetilde{H}_i\rangle|^2 \\ &\geq \frac{c_i^2}{n}(n - \theta d(H_i, \widetilde{H}_i))^2 \\ &\geq \frac{c_i^2}{n}(n - 2\theta d(H_i, \widetilde{H}_i)) \\ &\geq c_i^2 \frac{n - 2\theta d(H_i, \widetilde{H}_i)}{n + d(H_i, \widetilde{H}_i)(\theta^2 + 2\theta)} \end{aligned}$$

So, on an average Hamming distance is R/n, we get

$$p \geq \frac{n - 2\theta(R/n)}{n + (R/n)(\theta^2 + 2\theta)}$$

putting $p \leq r/n$ from Nayak's bound and rearranging the terms gives required result.

3 $\mathcal{E}$-Approximating Polynomials

Symmetric Boolean Functions

A boolean function $f : \{0,1\}^n \to \{0,1\}$ is symmetric if its value depends only on the number of 1's in input string that is the hamming weight of input. Some examples are OR, AND, Parity and Majority.

Approximating Degree of Function

Let ϵ, $\deg_\epsilon(f)$ denote ϵ-approximating degree, Paturi characterized the 1/3-error approximate degree: if $t \in (0, n/2]$ is the smallest integer such that f is constant for $|x| \in \{t, t+1, ..., n-t\}$, then $\deg_\epsilon(f) = \Theta(\sqrt{tn})$

de Wolf (author), improving Sherstov gave the characterization of $\deg_\epsilon(f)$ using a quantum query algorithm. Sherstov bound had logarithmic factors, which was tightened to $deg_\epsilon(f) = O(deg_{1/3}(f) + \sqrt{n \log(1/\epsilon)})$

Upper Bound on Approximation Degree

For every non-constant symmetric function $f : \{0,1\}^n \to \{0,1\}$ and $\epsilon \in [2^{-n}, 1/3]$:

$$deg_\epsilon(f) = O(deg_{1/3}(f) + \sqrt{n \log(1/\epsilon)})$$

Let $t = t(f)$ as in Paturi's result, the proof is as follows :

1. Use $(t-1)$ applications of Grover's search to try to find upto $(t-1)$ one's in x assuming t one's in the beginning and crossing out indices that we have already found to be one. This costs $\sum_{i=1}^{t-1} O(\sqrt{n/i}) = O(\sqrt{tn}) = O(deg_{1/3}(f))$.
2. Now use $\epsilon/2$-error Grover search to try to find one more solution. This costs $O(\sqrt{n \log(1/\epsilon)})$ queries.
3. Repeat step 1, now checking for zero's instead of one's.
4. Repeat step 2, now checking for zero's instead of one's.
5. If step 2 did not find another onw, output the corresponding value of $f(x)$.

   Else if step 4 did not find another zero, output the corresponding value of $f(x)$.

   Else assume $|x| \in \{t, t+1, ..., n-t\}$ and output $f(x)$.

Now the query complexity of the algorithm is $O(deg_{1/3}(f) + \sqrt{n \log(1/\epsilon)})$ due to:

If $|x| < t$, then step 1 will find all one's and $f(x)$ is computed with zero error.

Else If $|x| > n - t$, then step 2 finds another one with probability at least $1 - \epsilon/2$ and step 4 will not find another zero, so $f(x)$ is computed with error probability at most $\epsilon/2$.

Else step 2 and step 4 will find another one and zero respectively, with probabilities at least $1 - \epsilon/2$ each, so we can detect $|x| \in \{t, t+1, ..., n-t\}$ with probability $1 - \epsilon$. So in this case $f(x)$ is computed with error probability at most ϵ.

Since it took $\mathrm{O}(deg_{1/3}(f) + \sqrt{n \log(1/\epsilon)})$ queries to compute f with error probability at most ϵ, the ϵ-approximating polynomial degree is

$$deg_\epsilon(f) = O(deg_{1/3}(f) + \sqrt{n \log(1/\epsilon)})$$

4 Closure Properties of PP

Class PP

PP(probabilistic polynomial time) consists of all languages L for which there exists a probabilistic polynomial time algorithm that accepts input x with probability at least $1/2$ if $x \in L$ and with probability less than $1/2$ if $x \notin L$. Here acceptance probabilities may be extremely close to $1/2$ also.

Class BQP

BQP(bounded-error quantum polynomial time) is the class of languages computable with bounded error by a log space uniform family of polynomial-size quantum circuits with bounded error probability at most $1/3$. Here, both the workspace size and the number of unitaries are required to be of polynomial order.

Class PostBQP

PostBQP is the class of languages computable with bounded by a uniform family of polynomial-size quantum circuits that is allowed to post-select that is it can decide to do different steps depending upon the outcome of a particular measurement. This is an extension of class BQP with extra step based on result of measurement.

Closure Properties of Complexity Class

Closure properties tries to answer questions like if $L_1, L_2 \in C$ then is it true that $L_1 \cap L_2 \in C$? That is, is C closed under intersection ? Operation here instead can be anything like set complement or, any boolean operation.

It is known that class PP is closed under many general operations. Aaronson showed that class PP is actually same as the class PostBQP. So, closure properties of PP follow from closure properties of PostBQP.

This paper gives a quantum circuit that also uses postselection step for accepting a language belonging to PP, hence proving PP $\subseteq$ PostBQP.

Proof of PP $\subseteq$ PostBQP

Let L be a any language in PP, then there exits a probabilistic polynomial time algorithm M that uses $m = poly(n)$ random bits on input x of length n. So depending on the random string, the algorithm decides whether to output 0 or 1. Here 0 represents reject state and 1 represents accept state.

This defines a function $g = g_x : \{0,1\}^m \to \{0,1\}$, by rule

$$g(r) = [M(x)\text{accepts when using } r \text{ as its random string}]$$

Let $s = |g^{-1}(1)|$, then by definition of PP : $x \in L \iff s \geq 2^{m-1}$. If we could somehow decide whether $s \geq 2^{m-1}$, then we can tell if $x \in L$ or $x \notin L$.

Aaronson gave a postselection quantum circuit to do this. The algorithm uses two subroutines. The first one is as follows:

A1. Initialize an $(m+1)$ qubit register to $|0^{m+1}\rangle$. Apply $H^{\otimes m} \otimes I_2$ and then compute the function using oracle in last qubit, which gives the state :

$$\frac{1}{\sqrt{2^m}} \sum_{x \in \{0,1\}^m} |x\rangle |g(x)\rangle$$

A2. Again apply Hadamard on the first m qubits to get :

$$\frac{1}{\sqrt{2^m}} \sum_{x \in \{0,1\}^m} \Big(\frac{1}{\sqrt{2^m}} \sum_{w \in \{0,1\}^m} (-1)^{w.x} |w\rangle \Big) |g(x)\rangle$$

Now the state where first m qubits are zero in the superposition is:

$$\frac{1}{2^m} \sum_{x \in \{0,1\}^m} |0^m\rangle |g(x)\rangle = \frac{1}{2^m} |0^m\rangle ((2^m - s)|0\rangle + s|1\rangle)$$

A3. Now postselect on first m qubits being 0, we get the state (unnormalized) :

$$|\psi\rangle = (2^m - s)|0\rangle + s|1\rangle$$

Now the second subroutine is as follows : Let (α, β) be a pair of positive reals satisfiying $\alpha^2 + \beta^2 = 1$

B1. Prepare a qubit in the state: $|z\rangle = \alpha|0\rangle + \beta|1\rangle$. Perform controlled Hadamard on $|\psi\rangle$ using $|z\rangle$ as control qubit to get the state: $\alpha|0\rangle|\psi\rangle + \beta|1\rangle H|\psi\rangle$
Note that (unnormalized):

$$H|\psi\rangle = \frac{1}{\sqrt{2}}(2^m|0\rangle + (2^m - 2s)|1\rangle)$$

B2. Postselect on 2^{nd} qubit measuring to 1 to get the state below (unnormalized):

$$|\phi\rangle = s\alpha|0\rangle + \beta\frac{1}{\sqrt{2}}(2^m - 2s)|1\rangle$$

Now perform a projective measurement w.r.t Hadamard Basis.

Perform steps B1-B2, for different choices for $\{\alpha_i, \beta_i\}_{-m}^{+m}$ chosen to satisfy $\alpha_i = \beta_i 2^i$ and for each $i \in \{-m, ..., +m\}$ upto $O(\log m)$ trials to estimate the probability of getting the state $|+\rangle$. If $s \geq 2^{m-1}$ then $2s \geq 2^m$ that means that the state lies in the 4^{th} quadrant and will never be close to the $|+\rangle$ and probability of measuring in $|+\rangle$ is at max 1/2 whereas if it was not the case, then for suitable choice of α and β, it will be very close to $|+\rangle$ and probability of measuring in $|+\rangle$ will be near to 1, so we will be able to estimate with high probability whether $s \geq 2^{m-1}$ or not. So we can conclude whether the input belongs to the language or not with high probability. Thus PP $\subseteq$ PostBQP.

Bibliography

[1] Andrew Drucker and Ronald de Wolf, *Quantum Proofs for Classical Theorems*, Submitted. Available at http://arxiv.org/abs/0910.3376v2.