Speedup Lemma for Quantum Learning Algorithms

Pawan Paleja

May 28, 2025

1 Introduction

There is a long history of interplay between the design of algorithms for a given circuit class and matching lower bounds for the same class. Indeed, the development of a CIRCUIT-SAT algorithm for constant depth ACC circuits has given rise to breakthrough separations for the class [5] and old lower bounds for constant depth AC[p] circuits yielded learning algorithms for them in Carmosino et al [2]. In fact, the connection in [2] was formalized and extended further to show many interesting consequences for learning algorithms in the classical circuit regime by Oliviera and Santhanam[4]. In this paper, we verify a question posed in [1] to extend one of these consequences to the quantum regime, that is that a Speed-up Lemma Exists for Quantum Learning Algorithms. Formally, if $\mathfrak{C}[s(n)]$ denotes a circuit class $\mathfrak{C} \supset \mathrm{AC}^0[p]$ with size function $s(n)$:

Theorem 1.1. *$\mathfrak{C}$[poly(n)] has non-trivial quantum learners if and only if for each $\varepsilon > 0$, $\mathfrak{C}$[poly(n)] has strong quantum learners running in time $O(2^{n^{\varepsilon}})$.*

For the rest of the paper we will assume that our circuit classes contain $\mathrm{AC}^0[p]$. We define quantum learning as

Definition 1.2. *Let $\mathfrak{C}_n$ be a family of Boolean functions on n input variables and $\mathfrak{C} = \bigcup_{n\geq 1} \mathfrak{C}_n$. For $G : \mathbb{N} \to \mathbb{N}$, we say that $\mathfrak{C}_n$ can be $\varepsilon(n), \delta(n))$-**learned in time** $G(n)$ if there is a quantum learning algorithm $\mathcal{L}$ such that $\mathcal{L}$:*

- *... has quantum oracle access to f*
- *... uses at most $G(n)$ gates*
- *... outputs with probability $1 - \delta(n)$ a quantum hypothesis circuit $\tilde{U_f}$ such that*

$$\mathop{\mathbb{E}}_{x\sim 0,1^n}\left[\|(|f(x)\rangle\langle f(x)| \otimes I)\tilde{U_f} x\rangle|0\rangle\|^2\right] \geq 1 - \varepsilon(n)$$

Note that $(|f(x)\rangle\langle f(x)| \otimes I$ can be thought of as a measurement operator on the agreement of $f(x)$ and $\tilde{U_f}$ and we will refer to it as Π_f.

Based on Definition 1.2, we define non-trivial learning as learning in time $T(n) = 2^n/n^{\omega(1)}$with error $\varepsilon(n) = 1/2 - 1/n^c$ with failure probability $\delta(n) = 1/n$ and strong learning as learning in time $T$with error $\varepsilon(n) = 1/n^c$ with failure probability $\delta(n) = 1/n$. We will use similar strategies as [4] using results and notions from [1] to show Theorem 1.1.

1.1 Proof Idea

The idea is as follows:

1. Assume we have some non-trivial quantum learning algorithm A for the class $\mathfrak{C}[s(n)]$.
2. First, we amplify the hardness of f using known constructions to get $AMP(f)$.
3. We then show that the hardness amplification of f was computable in $\mathfrak{C}$, which implies that A is still a viable learning algorithm for $AMP(f)$.
4. We then use the weak learner for $AMP(f)$ to obtain a strong learner for f.

2 Black-box Amplification of f

In this section, we define $AMP(f)$ and describe some of the relevant results for its construction. The first step in our hardness amplification is the Direct Product Construction:

Definition 2.1. *For a boolean function $f : \{0,1\}^n \to \{0,1\}$ and a parameter k, the **k-wise direct product of** f is $f^k : \{0,1\}^{nk} \to \{0,1\}^k, f^k(x_1, ..., x_k) = (f(x_1), ..., f(x_k))$*

However, we still want our harder function to have a single output bit. Thus, we want to collapse the k output bits into a single number, which we will accomplish using the Goldreich-Levin Algorithm.

Definition 2.2. *Let p be a prime. For $g : \{0,1\}^m \to \{0,1\}^k$, define $g^{GL}(x_1, ..., x_m, r_1, ..., r_k) = \sum_{i=1}^k r_i \cdot g(x_1, ..., x_m)_i \mod p$*

Thus we define $AMP(f)$ as applying both these constructions to our function, i.e. $AMP(f) = (f^k)^{GL}$. We want to show two facts about this function, which were alluded to in the proof idea, that $AMP(f)$ can be computed within $\mathrm{AC}^0[p]$ (and therefore $\mathfrak{C}$) and there is an efficient reconstruction of the original function that gives us strong learning properties. If $AMP(f)$ has these properties, we say that it is a *Black-Box Amplification.* Formally:

Definition 2.3 (Definition 3.4 from [2]). *We say $AMP(f) : \{0,1\}^{n'} \to \{0,1\}$ is a (ε, δ)-**amplification** within Λ if it has:*

1. *Short input: $n' \leq poly(n, 1/\varepsilon, \log(1\delta)$,*
2. *Nonuniform Λ-Efficiency: $AMP(f) \in \Lambda^f[\mathrm{poly}(n')]$,*
3. *Uniform P-Efficiency: $Amp(f) \in P^f$, and*
4. *Reconstruction: There exists an algorithm with oracle access to f which takes a quantum hypothesis circuit with error $1/2 - \delta$ for $AMP(f)$ and outputs a quantum hypothesis with advantage ε.*

Carmosino et al proved the first three items hold true for any class of circuits that contain $\mathrm{AC}^0[p]$ [2]. Indeed,

Lemma 2.4 (See [2], Theorems 4.3, 4.8). *Let p be any fixed prime. Then for all $0 < \varepsilon, \delta < 1$, there is a black-box (ε, δ) amplification within $AC^0[p]$.*

What is interesting in the quantum setting is the fourth item. The reconstruction of f from a nontrivial quantum hypothesis for $AMP(f)$. We show first for the Goldreich-Levin Algorithm.

3 Goldreich-Levin Reconstruction

We would like to show the following theorem:

Theorem 3.1 (Lemma 4.5 from [1]). *Suppose there is a quantum circuit U (that uses $m > 0$ ancilliary qubits) satisfying*

$$\mathop{\mathbb{E}}_{x\in\{0,1\}^{kn}} \mathop{\mathbb{E}}_{r\in\{0,1\}^k} \left[|\Pi_{f(x)\cdot r} U |x, r, 0^m\rangle|^2\right] \geq 1/2 + \gamma(n).$$

Then there is a sequence of deterministic circuits $\{C_n^{GL}\}$ of size $\mathrm{poly}(n, k, size(U))$ such that $C_n^{GL}(1^n, U) = U'$ such that

$$\mathop{\mathbb{E}}_{x,U'} \left[|\Pi_{f^k(x)} U' |x, 0^{k+m+1}\rangle|^2\right] \geq \frac{\gamma(n)^3}{2}$$

Proof. We define U' as follows:

1. Start with initial state $\frac{1}{\sqrt{2^k}} \sum_r |x, r, 0^m, 1\rangle$.
2. Execute U on the first $kn + k + m$ qubits.

3. Apply a CZ gate with the first qubit (output of $U\left|x, r, 0^m\right\rangle$) as the control and the last qubit as the target qubit.

4. Execute $U^\dagger$ on the first $kn + k + m$ qubits.

5. Measure all the qubits in $Z - basis$, if the outcome is of the form $\left|x, a, 0^{k+m}, 1\right\rangle$ output a.

Since we are taking the expectation over both $x \in \{0,1\}^{kn}$ and the output algorithm U', we can divide the analysis along those lines. We divide the x's naturally, let G be the witness set for the advantage of U, that is

$$G = \{x : \mathop{\mathbb{E}}_{r \in \{0,1\}^k} \left[|\Pi_{f(x)\cdot r} U \left|x, r, 0^m\right\rangle|^2 \right] \geq \frac{1}{2} + \frac{\gamma}{2}$$

Observe that this allows us to split our assumption:

$$\begin{aligned}
\mathop{\mathbb{E}}_{x \in \{0,1\}^{kn}} \mathop{\mathbb{E}}_{r \in \{0,1\}^k} \left[|\Pi_{f(x)\cdot r} U \left|x, r, 0^m\right\rangle|^2 \right] &\geq 1/2 + \gamma \\
\Pr[x \in G] \mathop{\mathbb{E}}_{x \in G} \mathop{\mathbb{E}}_{r \in \{0,1\}^k} \left[|\Pi_{f(x)\cdot r} U \left|x, r, 0^m\right\rangle|^2 \right] + \Pr[x \notin G] \mathop{\mathbb{E}}_{x \notin G} \mathop{\mathbb{E}}_{r \in \{0,1\}^k} \left[|\Pi_{f(x)\cdot r} U \left|x, r, 0^m\right\rangle|^2 \right] \geq& \\
\Pr[x \in G] + (1 - \Pr[x \in G])(\frac{1}{2} + \frac{\gamma}{2}) >& \\
\frac{|G|}{2^n} + (1 - \frac{|G|}{2^n})(\frac{1}{2} + \frac{\gamma}{2}) &> 1/2 + \gamma
\end{aligned}$$

which only works if $|G| \geq \frac{\gamma 2^n}{2}$. Referring back to what we wanted to prove:

$$\mathop{\mathbb{E}}_{x, U'} \left[|\Pi_{f^k(x)} U' \left|x, 0^{k+m+1}\right\rangle|^2 \right] \geq \frac{\gamma(n)^3}{2},$$

we now have that the LHS is at least

$$\frac{|G|}{2^n} \mathop{\mathbb{E}}_{x \in G, U'} \left[|\Pi_{f^k(x)} U' \left|x, 0^{k+m+1}\right\rangle|^2 \right] \geq \frac{\gamma}{2} \mathop{\mathbb{E}}_{x \in G, U'} \left[|\Pi_{f^k(x)} U' \left|x, 0^{k+m+1}\right\rangle|^2 \right]$$

Thus, it remains to show that the second term is at least γ^2. To do this, we aim to analyze the probability U' outputs $f(x)$. Observe that this is exactly $\left\langle x, f(x), 0^{k+m}, 1\right| U' \left|x, r, 0^m, 1\right\rangle$. Applying the steps of U' we get that for states

$$\begin{aligned}
\left|\psi_x\right\rangle &= \left|x\right\rangle \otimes \frac{1}{\sqrt{2^k}} \sum_r (-1)^{f(x)\cdot r} (\alpha_{x,r,0} \left|\psi_{x,r,0}\right\rangle \left|f(x)\cdot r, 1\right\rangle + \alpha_{x,r,1} \left|\psi_{x,r,1}\right\rangle \left|1 \otimes f(x)\cdot r, 1\right\rangle \\
\left|\sigma_x\right\rangle &= \frac{1}{\sqrt{2^k}} \left|x\right\rangle \sum_r (-1)^{f(x)\cdot r} (\alpha_{x,r,0} \left|\psi_{x,r,0}\right\rangle \left|f(x)\cdot r, 1\right\rangle + \alpha_{x,r,1} \left|\psi_{x,r,1}\right\rangle \left|1 \otimes f(x)\cdot r, 1\right\rangle
\end{aligned}$$

that this probability is given by $|\left\langle\sigma_x\right| \left|\psi_x\right\rangle|^2$ which equals

$$\left|\mathbb{E}[|\alpha_{x,r,0}|^2 - |\alpha_{x,r,1}|^2]\right|$$

and for $x \in G$, this is at least γ^2 as required. □

Thus, we have shown that we can recover the direct product from a function amplified by the Goldreich-Levin Construction. Observe that we achieved now a strong hypothesis for the direct product of f.

4 Direct Product Reconstruction

It remains only to show that the same reconstruction is true for the direct product part of the amplification. Formally:

Theorem 4.1 (Theorem 4.28 from [1])**.** *There exists a universal constant $C \geq 1$ for which the following holds. Let $n \geq 1$, k even, and let ε, δ satisfy*

$$k \geq C \cdot \frac{1}{\delta} \cdot \left[\log \frac{1}{\delta} + \log \frac{1}{\varepsilon}\right].$$

If U' is a quantum circuit of size at most s defined over $S_{n,k}$ with k output bits such that

$$\mathop{\mathbb{E}}_{B \sim S_{n,k}, U'} \left[U'(B) = f^k(B)\right] := \mathop{\mathbb{E}}_{B \sim S_{n,k}, U'} \left[|\Pi_{f^k(B)} U' \left|B, 0^{k+m+1}\right\rangle|^2\right] \geq \varepsilon$$

then there is a quantum circuit H of size $\text{poly}(n, k, s, \log(1/\delta), 1/\varepsilon)$ that can be constructed uniformly such that

$$\mathop{\mathbb{E}}_{x \sim \{0,1\}^n, H} \left[H(B) = f^k(B)\right] := \mathop{\mathbb{E}}_{x \sim \{0,1\}^n, H} \left[|\Pi_{f(x)} H \left|B, 0^{m'}\right\rangle|^2\right] \geq 1 - \delta$$

The key insight into proving this is that we can reduce to a similar circuit model. In analyzing randomized algorithms, one can reduce to the deterministic case by considering a randomized algorithm as picking over the distribution of deterministic algorithms. In the quantum realm, this is not possible, but we can still reduce to a simpler model: Inherently Probabilistic Circuits.

Definition 4.2. *$\mathcal{A}$ is an **inherently probabilistic circuit** (IPC) if $A : 0, 1^m \to \mathbf{F}$, where $F = \{D | D : \{0,1\}^\ell \to [0,1]\}$. $\mathcal{A}$ computes $g : \{0,1\}^m \to \{0,1\}^l$ with probability ε if*

$$\Pr_{z,\nu} [\nu = g(z)] \geq \varepsilon$$

We use the following lemma to show equivalencies between the two models

Lemma 4.3. *If R is an IPC and Q a QC s.t. $\forall x$ and the output distributions of $R(x)$ and $Q(x)$ are exactly the same, then for every probabilistic algorithm A, which might have have some classical input w which is allowed to make classical queries to R or Q, we have that*

$$\Pr_{A,R} \left[A^R(w) = y\right] = \Pr_{A}, Q \left[A^Q(w) = y\right]$$

We proceed now by showing Theorem 4.1 for IPCs and we leave it to the reader to justify the equivalence for themselves. We proceed by an explicit construction of the reconstruction algorithm which is pulled from [3]. But first, we introduce some useful definitions

Definition 4.4 (Definition 4.11 from [1])**.** *Let...*

- ***the set of edges*** *$I = \{(A, B) \in S_{k/2} \times S_k : A \subseteq B\}$.*
- ***the neighbors of a set*** *A be $N_I(A) = \{B \in S_k : (A, B) \in I\}$*
- ***the neighbors of*** *$A \in S_{k/2}$* ***and*** *$x \in \mathcal{P}(\{0,1\}^n) - A$ as $N_I(A, x) = \{B \in S_k : A \cup \{x\} \in B\}$*

The idea for the list-decoding algorithm is to use a subroutine which attempts to compute $f(x)$ for a given x by the following procedure:

Construction 4.5. *Given set $AS_{n,k}$, assignment w to A, and time parameter T we compute $f(x)$ as follows:*

1. *If $x \in A$, output $w|_x$.*
2. *Repeat T times:*
   (a) *Pick random neighbor $B' \in N_I(A, x)$.*
   (b) *Sample $v' \sim U'(B')$ to get a subset of the direct product.*
   (c) *If $v'|_A = w$, output $v'|x$.*
3. *Fail.*

The main algorithm starts the process by sampling a random $k/2$-set A and random k-set B, such that the starting parameters are

$$A \subseteq B, w = U'(B)|_A$$

The analysis is quite technical and the insight that applies to the quantum setting has already been completed with the reduction to the IPC model, so we omit the proof of correctness.

5 Proof of Main Theorem

Referring to the strategy in Section 1.1, we want to combine the above results to prove Theorem 1.1.

Proof of Theorem 1.1. The forward direction is trivially true, so we want to assume a non-trivial learning algorithm A with advantage $\gamma(n)$ and success probability $\geq 1/10$ for $\mathfrak{C}[\text{poly}(n)]$. Let $AMP(f) = (f^k)^G L$. By Lemma 2.4, this function is still in $\mathfrak{C}[\text{poly}(n)]$, thus A still works for it. By Theorem 3.1, we get a function computing $f^k(x)$ with advantage $\gamma(n)^3$ and Theorem 4.1 gives that if we choose $\delta = 1/n$ there exists a k that fulfills the parameters given by $\varepsilon = \gamma(n) and \delta$ which gives us the strong learner by our reconstruction algorithm. □

6 Conclusion

In this paper, we verified that the Speedup Lemma from [4] carries over when the non-trivial learning algorithm that we start with is a quantum algorithm. There are some possible directions for this work, one such direction has to do with the assumption we made at the start, i.e. that the circuits we are referring to contain $\text{AC}^0[p]$. We needed this to ensure that the steps in our process were closed operations in the class (e.g. applying the NW reconstruction algorithm gave a circuit still in $\mathfrak{C}$). However currently it is not known for some classes of quantum circuits if they $\text{AC}^0[p]$, so it would be interesting to show if the steps in our speed up lemma are computable within quantum circuits like QAC^0. In fact, if they are then this would provide a separation between QAC^0 and AC^0 because constructions like NW designs are not computable in the latter [2].

References

[1] Srinivasan Arunachalam et al. "Quantum learning algorithms imply circuit lower bounds". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 562–573.

[2] Marco L Carmosino et al. "Learning algorithms from natural proofs". In: *31st Conference on Computational Complexity (CCC 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2016.

[3] Russell Impagliazzo et al. "Uniform direct product theorems: simplified, optimized, and derandomized". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing* (2008).

[4] Igor C Oliveira and Rahul Santhanam. "Conspiracies between learning algorithms, circuit lower bounds and pseudorandomness". In: *arXiv preprint arXiv:1611.01190* (2016).

[5] Ryan Williams. "Nonuniform ACC Circuit Lower Bounds". In: *J. ACM* 61.1 (Jan. 2014). ISSN: 0004-5411. DOI: 10.1145/2559903. URL: https://doi.org/10.1145/2559903.