

---

# Quantum Boosting

---

## A. Proof of Lemma 5.2

We will use the theorems below in order to prove the lemma.

**Theorem A.1 (Multiplicative amplitude estimation (Ambainis, 2010))** Suppose we have access to a unitary  $U : |0\rangle \rightarrow |\psi\rangle$  and its inverse where

$$|\psi\rangle = \sqrt{a}|\psi_1\rangle + \sqrt{1-a}|\psi_0\rangle, \quad (1)$$

with the promise that  $a$  is either 0 or greater than  $p$ . For every  $c \in (0, 1)$ , there exists an algorithm that, with probability  $\geq 1 - \delta$ , outputs an estimate  $\tilde{a}$ , satisfying  $|a - \tilde{a}| \leq c \cdot \tilde{a}$  if  $a \geq p$ , and  $\tilde{a} = 0$  if  $a = 0$ . Then the number of queries to  $U, U^{-1}$  is

$$O\left(\frac{\log(1/\delta)}{c} \left(1 + \log \log \frac{1}{p}\right) \sqrt{\frac{1}{\max\{a, p\}}}\right). \quad (2)$$

**Theorem A.2 (Nondestructive amplitude estimation (Harrow & Wei, 2019))** Let  $|\psi\rangle$  be a quantum state. Suppose we are given access to  $U : |0\rangle \rightarrow |\psi\rangle$  and  $U^{-1}$  where

$$|\psi\rangle = \sqrt{a}|\psi_1\rangle + \sqrt{1-a}|\psi_0\rangle. \quad (3)$$

This subroutine uses ideas developed by Ambainis (Ambainis, 2010) and the proof uses ideas required to prove Theorem A.1.

---

### Algorithm 1 Modified Amplitude Estimation

---

**Input:** Let  $\tilde{\varepsilon}, M > 0$ . The state  $|\psi\rangle = \sqrt{\tilde{\varepsilon}/M}|\phi_1\rangle|1\rangle + \sqrt{1-\tilde{\varepsilon}/M}|\phi_0\rangle|0\rangle$  and the unitary  $U$  such that  $U|0\rangle = |\psi\rangle$ .

- 1: **for**  $J = \frac{2\pi\sqrt{M}}{\delta}$  to  $\frac{16\sqrt{2}\pi\sqrt{M}}{\delta} \cdot \sqrt{QT^2} \log(MT/\delta)$  **do**
- 2: Let  $\varepsilon'/M$  be the output after performing non-destructive amplitude estimation (in Thm. A.2) to compute  $\tilde{\varepsilon}/M$  using  $J$  queries to  $U$  and  $U^{-1}$ .
- 3: Check if  $\frac{2\sqrt{2}\pi\sqrt{(1-\delta)\varepsilon'}}{J\sqrt{M}} + \frac{\pi^2}{J^2} \leq \frac{\delta\varepsilon'}{M}$ . If yes, then output  $\varepsilon'$  and quit the loop. Else, let  $J = 2 \cdot J$ .
- 4: **end for**

**Output:**  $\{\varepsilon', \text{yes}\}$  if there exists  $\varepsilon'$  in step (3), else output  $\{\varepsilon' = 1/(QT^2), \text{no}\}$ .

---

**Lemma A.3** Let  $\delta = 1/(10QT^2)$ . Algorithm 1 satisfies the following: with probability  $\geq 1 - 10\delta/T$ , if the output is  $\{\varepsilon', \text{yes}\}$ , then  $|\tilde{\varepsilon} - \varepsilon'| \leq \delta\varepsilon'$ ; and if the output is  $\{\varepsilon' = 1/(QT^2), \text{no}\}$ , then  $|\tilde{\varepsilon} - \varepsilon'| \leq 1/(QT^2)$ . The total number queries to  $U$  and  $U^{-1}$  used by Algorithm 1 is  $O(\sqrt{M}Q^{3/2}T^3)$ .

**Proof.** We first consider the case when Algorithm 1 outputs  $\{\varepsilon', \text{yes}\}$ . In this case, there exists a  $J$  and  $\varepsilon'$  which satisfies the relation in step (3) of the algorithm. First observe that, since  $\varepsilon'/M$  was obtained by amplitude amplification in step (2), we have

$$\left| \frac{\varepsilon'}{M} - \frac{\tilde{\varepsilon}}{M} \right| \leq \left| \frac{\varepsilon'}{M} - \frac{(1-\delta)\tilde{\varepsilon}}{M} \right| \leq \frac{2\pi\sqrt{(1-\delta)\tilde{\varepsilon}}}{J\sqrt{M}} + \frac{\pi^2}{J^2} \leq \frac{2\sqrt{2}\pi\sqrt{(1-\delta)\varepsilon'}}{J\sqrt{M}} + \frac{\pi^2}{J^2}, \quad (4)$$

where the second inequality used Theorem A.2 and the third inequality used the first and second inequalities to conclude

$|\tilde{\varepsilon} - \varepsilon'| \leq \frac{2\pi\sqrt{(1-\delta)M\tilde{\varepsilon}}}{J} + \frac{\pi^2 M}{J^2}$ . This implies

$$\tilde{\varepsilon} \leq \varepsilon' + \frac{2\pi\sqrt{(1-\delta)M\tilde{\varepsilon}}}{J} + \frac{\pi^2 M}{J^2} \leq 2\varepsilon',$$

where we used  $J \geq (2\pi\sqrt{M})/\delta$ . Putting together the upper bound in Eq. (4) along with the upper bound in Step (3) of the algorithm, we get  $|\tilde{\varepsilon} - \varepsilon'| \leq \delta\varepsilon'$ . Furthermore, we also show that  $\tilde{\varepsilon} \geq (1 - 2\delta)/(64QT^2)$ . Recall that  $J, \varepsilon'$  satisfies step (3) of the algorithm. In particular, this implies that

$$\frac{2\sqrt{2}\pi\sqrt{(1-\delta)\varepsilon'}}{J_{max}\sqrt{M}} + \frac{\pi^2}{J_{max}^2} \leq \frac{\delta\varepsilon'}{M},$$

since  $J \leq J_{max}$ . Substituting the value of  $J_{max}$  in the inequality above gives  $\frac{\sqrt{(1-\delta)\varepsilon'}}{8\sqrt{QT^2}} + \frac{\delta}{512QT^2} \leq \varepsilon'$ . Solving for the above equation, we obtain  $\varepsilon' \geq 1/(64QT^2) \cdot (1 - \delta)$  (we ignore the other solution for  $\varepsilon'$  since  $\varepsilon' \geq 0$ ). Using  $|\tilde{\varepsilon} - \varepsilon'| \leq \delta\varepsilon'$ , we get  $\tilde{\varepsilon} \geq (1 - \delta)\varepsilon' \geq (1 - 2\delta)/(64QT^2)$ .

Now we consider the case when Algorithm 1 outputs  $\{\varepsilon' = 1/(QT^2), \text{no}\}$ , and we argue that  $|\tilde{\varepsilon} - \varepsilon'| < 1/(QT^2)$ . In order to see this, first observe that

$$\left| \frac{\varepsilon'}{M} - \frac{\tilde{\varepsilon}}{M} \right| \leq \frac{2\sqrt{2}\pi\sqrt{(1-\delta)\varepsilon'}}{J\sqrt{M}} + \frac{\pi^2}{J^2} \leq \frac{\delta\sqrt{2\varepsilon'}}{M} + \frac{\delta^2}{4M} \leq \frac{10\delta}{M}, \quad (5)$$

where the first inequality used Eq. (4), the second inequality used  $J \geq (2\pi\sqrt{M})/\delta$  and the third inequality used  $\varepsilon' < 1$ . Using  $\delta = 1/(10QT^2)$ , we obtain  $|\tilde{\varepsilon} - \varepsilon'| \leq 1/(QT^2)$ . Furthermore, we show that in the ‘no’ instance, we have  $\tilde{\varepsilon} < 1/(QT^2)$ . We prove this by a contrapositive argument: suppose  $\tilde{\varepsilon} \geq 1/(QT^2)$ , there exists a  $J' \in [J^*, J_{max}]$ , where  $J^* = \frac{8\pi\sqrt{M}}{\delta\sqrt{(1-\delta)\tilde{\varepsilon}}}$ , for which the inequality in step (3) of Algorithm 1 is satisfied with probability at least  $1 - 10\delta/T$ .<sup>1</sup> In order to see this, first observe that

$$\frac{2\sqrt{2}\pi\sqrt{(1-\delta)\varepsilon'}}{J\sqrt{M}} + \frac{\pi^2}{J^2} \leq \frac{4\pi\sqrt{(1-\delta)\tilde{\varepsilon}}}{J\sqrt{M}} + \frac{\pi^2}{J^2} \leq \frac{\delta(1-\delta)\tilde{\varepsilon}}{2M} + \frac{\delta^2(1-\delta)\tilde{\varepsilon}}{64M} \leq \frac{\delta(1-\delta)\tilde{\varepsilon}}{M}, \quad (6)$$

where the second inequality used  $J \geq J^*$  and the remaining inequalities are straightforward. Using Eq. (5) and Eq. (6), we have  $|\varepsilon' - \tilde{\varepsilon}| \leq \delta(1-\delta)\tilde{\varepsilon}$ . Moreover, using  $|\tilde{\varepsilon} - \varepsilon'| \leq \delta(1-\delta)\tilde{\varepsilon}$ , we can further upper bound Eq. (6) by  $\delta\varepsilon'/M$ , which implies step (3) of Algorithm 1 is satisfied, in which case the algorithm would have output ‘yes’ with probability  $\geq 1 - 10\delta/T$ . Hence, by the contrapositive argument, if Algorithm 1 outputs ‘no’ with probability at least  $1 - 10\delta/T$ , then we have  $\tilde{\varepsilon} < 1/QT^2$ .

Finally, we bound the total number of queries made to  $U$  and  $U^{-1}$  in Algorithm 1. Given that  $J$  is doubled in every round and  $J \leq J_{max} = O(\sqrt{MQT^2}/\delta)$ , the total number of queries is

$$J_{max} + \frac{J_{max}}{2} + \frac{J_{max}}{4} + \dots + \frac{2\pi\sqrt{M}}{\delta} < 2J_{max} = O\left(\sqrt{MQT^2}/\delta\right) = O(\sqrt{M}Q^{3/2}T^3) \quad (7)$$

using  $\delta = O(1/QT^2)$ . This concludes the proof of the lemma.  $\square$

## B. Details of Algorithm 1

First, we describe the unitary operation that updates  $\tilde{D}_1$  (in step (2)) to  $\tilde{D}_t$  (in step (3)). For  $t \in \{1, \dots, T\}$ , let  $\mathcal{G}_t$  be the quantum circuit that makes the distribution update from  $\tilde{D}^1 \rightarrow \tilde{D}^t$ . Given access to  $h_1, \dots, h_{t-1}, c: \{0, 1\}^n \rightarrow \{-1, 1\}$  and knowledge of  $\varepsilon'_1, \dots, \varepsilon'_{t-1}$ , define  $\mathcal{G}_t$  as the map:

$$\begin{aligned} \mathcal{G}_t : & \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^1\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle \\ & \rightarrow \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle. \end{aligned} \quad (8)$$

<sup>1</sup>Note that  $J^* \leq J_{max}$  follows immediately by using the lower bound  $\tilde{\varepsilon} \geq 1/QT^2$ .

We describe our quantum boosting algorithm in more details. In the  $t$ th step, we have quantum query access to the hypotheses  $\{h_1, \dots, h_{t-1}\}$  and knowledge of the approximate weighted errors  $\{\varepsilon'_1, \dots, \varepsilon'_{t-1}\}$ . Let  $U_1, U_2$  be unitaries that satisfy  $U_1 : |0\rangle \rightarrow |\psi_1\rangle$  and  $U_2 : |0\rangle \rightarrow |\Phi_1\rangle$ , where

$$|\psi_1\rangle = \left( \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^1\rangle \otimes |0\rangle^{\otimes t+1} \right), \quad |\Phi_1\rangle = \left( \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^1\rangle \otimes |0\rangle^{\otimes t} \right).$$

Recall that  $\tilde{D}^1$  is the uniform distribution over the training set  $S$  where  $\tilde{D}_x^1 = 1/M$  for every  $(x, c(x)) \in S$ . We assume that theoretically one could use a quantum RAM to prepare the state  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle$  in time  $O(\log M)$ . Since the training set  $S = \{(x_i, c(x_i))\}_{i \in [M]}$  is classical and stored in a classical data structure, we assume that a QRAM can be used to map this classical register into a *uniform* quantum state in time  $O(\log M)$ .<sup>2</sup> Ofcourse, one can also assume that a quantum learner has access to uniform quantum examples  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle$ , in which case we do not need to assume a QRAM.

We now describe the quantum boosting algorithm. The algorithm begins by first preparing  $U_1|0\rangle \otimes U_2^{\otimes Q}|0\rangle = |\psi_1\rangle \otimes |\Phi_1\rangle^{\otimes Q}$ . We then apply  $(Q+1)(t-1)$  quantum queries to the oracles  $\{O_{h_1}, \dots, O_{h_{t-1}}\}$  to obtain<sup>3</sup>

$$\begin{aligned} |\psi_2\rangle \otimes |\Phi_2\rangle^{\otimes Q} &= \left( \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^1\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]|0\rangle^2 \right) \\ &\quad \otimes \left( \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^1\rangle |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], 0\rangle \right)^{\otimes Q}. \end{aligned} \quad (9)$$

We then apply the unitary  $\mathcal{G}_t$  on  $|\psi_2\rangle$  and each of the  $Q$  copies of  $|\Phi_2\rangle$  to update  $\tilde{D}^1$ . The resulting state is

$$\begin{aligned} |\psi_3\rangle \otimes |\Phi_3\rangle^{\otimes Q} &= \left( \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^1\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]|0\rangle^2 \right) \otimes \\ &\quad \left( \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^1\rangle |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], 0\rangle \right)^{\otimes Q}. \end{aligned} \quad (10)$$

We now break down the algorithm into two steps, the first phase uses  $|\Phi_3\rangle^{\otimes Q}$  to obtain  $h_t$  and the second phase uses  $h_t$  and  $|\psi_3\rangle$  to compute  $\varepsilon'_t$ .

**Phase (1): Obtaining hypothesis  $h_t$ .** Let  $V : |p\rangle|0\rangle|0\rangle \rightarrow |p\rangle\left(\sqrt{1-p}|0\rangle + \sqrt{p}|1\rangle\right)|\sin^{-1}(\sqrt{p})\rangle$ . For each of the  $Q$  copies of  $|\Phi_3\rangle$ , append an auxiliary  $|0\rangle$  and apply  $V$  to obtain

$$|\Phi_4\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^t\rangle |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle \left( \sqrt{\tilde{D}_x^t}|0\rangle + \sqrt{1-\tilde{D}_x^t}|1\rangle \right). \quad (11)$$

Let  $W_t$  be the unitary that performs  $W_t : |\Phi_1\rangle \rightarrow |\Phi_4\rangle$  and  $\tilde{W}_t = W_t U_2$  be the map  $\tilde{W}_t : |0\rangle \rightarrow |\Phi_4\rangle$ . Let  $Y_t$  be the unitary that uses  $O(\sqrt{M} \log(T/\zeta))$  invocations of  $\tilde{W}_t$  and  $\tilde{W}_t^{-1}$  to perform amplitude amplification on the state  $|\Phi_4\rangle$  where  $\zeta = O(1)$ . Then with probability  $\geq 1 - \zeta/T$ , the resulting state is

$$|\Phi_5\rangle = Y_t |\Phi_4\rangle = \sum_{x \in S} \sqrt{\tilde{D}_x^t} |x, c(x)\rangle |\tilde{D}_x^t\rangle |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], 0\rangle + |\Psi\rangle, \quad (12)$$

<sup>2</sup>One could use the proposal of Lloyd et al. (Giovannetti et al., 2008) as a QRAM here. We make a couple of remarks: our quantum algorithm only requires a QRAM to prepare the *uniform* superposition over classical data. Also, our quantum algorithm does not use QRAM as an oracle for Grover-like algorithms, so the negative results of (Arunachalam et al., 2015) do not apply to our algorithm.

<sup>3</sup>To be precise, we obtain the  $t$ -th indicator function  $[h_t(x) \neq c(x)]$  as follows: first use  $O_{h_t}$  to perform the map  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^1\rangle |0\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^1\rangle |h_t(x)\rangle$ , next apply the CNOT gate on  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^1\rangle |h_t(x)\rangle$  to produce  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle |\tilde{D}_x^1\rangle |h_t(x) \cdot c(x)\rangle$

where  $|\Psi\rangle$  is orthogonal to the first register. Observe that without  $|\Psi\rangle$ , the state  $|\Phi_5\rangle$  is no longer a quantum state because  $\{\tilde{D}_x^t\}_x$  is a sub-normalized distribution. Using Claim 5.3, we have

$$\| |\Psi\rangle \| \leq 1 - \sum_{x \in S} \tilde{D}_x^t \leq 30\delta. \quad (13)$$

Observe that if we had run the boosting algorithm with the ideal distribution  $D^t$ , we would have obtained the state

$$|\Phi'_5\rangle = \sum_{x \in S} \sqrt{D_x^t} |x, c(x)\rangle |D_x^t\rangle | [h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], 0 \rangle, \quad (14)$$

instead of  $|\Phi_5\rangle$ . We now uncompute the auxiliary registers  $|\tilde{D}_x^t\rangle | [h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], 0 \rangle$  in  $|\Phi_5\rangle$  as follows: let  $\mathcal{G}_t^{-1}$  be the unitary which maps  $|\tilde{D}_x^t\rangle \rightarrow |\tilde{D}_x^1\rangle$  and let  $O_{h_1}, \dots, O_{h_{t-1}}$  be the query operations that uncompute the  $\{ [h_i(x) \neq c(x)] \}_{i \in [t-1]}$  registers. Applying  $\mathcal{G}_t^{-1}$  on the *actual* state  $|\Phi_5\rangle$  (instead of the *ideal* state  $|\Phi'_5\rangle$ ) gives

$$\mathcal{G}_t^{-1} |\Phi_5\rangle = \sum_{x \in S} \sqrt{\tilde{D}_x^t} |x, c(x)\rangle |0\rangle | [h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], 0 \rangle + \mathcal{G}_t^{-1} |\Psi\rangle,$$

and then performing  $O_{h_1}, \dots, O_{h_{t-1}}$  gives us

$$|\Phi_6\rangle = \sum_{x \in S} \sqrt{\tilde{D}_x^t} |x, c(x)\rangle |0\rangle |0\rangle^t + O_{h_{t-1}} \cdots O_{h_1} \cdot \mathcal{G}_t^{-1} |\Psi\rangle. \quad (15)$$

By performing the operations  $\mathcal{G}_t^{-1}, O_{h_1}, \dots, O_{h_{t-1}}$  on the *ideal* state  $|\Phi'_5\rangle$ , we would have

$$|\Phi'_6\rangle = \sum_{x \in S} \sqrt{D_x^t} |x, c(x)\rangle |0\rangle |0\rangle^t. \quad (16)$$

Ideally, our goal would be to pass  $Q$  copies of  $|\Phi'_6\rangle$  to a quantum learner in order to obtain a hypothesis  $h_t$ . Although, we do not have access to  $|\Phi'_6\rangle$ , we continue our quantum boosting algorithm by passing  $Q$  copies of  $|\Phi_6\rangle$  to a quantum learner (instead of  $Q$  copies of  $|\Phi'_6\rangle$ ). A priori, it is not clear what will be the output of the quantum learner on input  $|\Phi_6\rangle^{\otimes Q}$ . In order to understand this, we first show that  $|\Phi_6\rangle$  and  $|\Phi'_6\rangle$  are close. Using this, it is not hard to see that a quantum learning algorithm would *behave the same* when given  $|\Phi_6\rangle^{\otimes Q}$  instead of  $|\Phi'_6\rangle^{\otimes Q}$ . In order to formalize this, we first state the following claim which we prove later.

**Claim B.1** *Let  $|\Phi_6\rangle$  and  $|\Phi'_6\rangle$  be as defined in Eq. (15), (16). Then we have  $|\langle \Phi_6 | \Phi'_6 \rangle| \geq 1 - 50\delta$ .*

Recall that  $|\Phi'_6\rangle$  is the ideal state that satisfies the following: suppose  $Q$  copies of  $|\Phi'_6\rangle$  are given to a weak quantum learner, then with probability at least  $1 - 1/T$ , the learner outputs a weak hypothesis  $h_t$ . We now show that the same learner, when fed  $Q$  copies of  $|\Phi_6\rangle$  (instead of  $|\Phi'_6\rangle$ ) will output  $h_t$  with probability at least  $1 - 9/T$ . In order to see this, let  $p' = \Pr[\mathcal{A} \text{ outputs } h_t \text{ given } |\Phi'_6\rangle^{\otimes Q}] \geq 1 - 1/T$  and  $p = \Pr[\mathcal{A} \text{ outputs } h_t \text{ given } |\Phi_6\rangle^{\otimes Q}]$ . Let  $\mathcal{H}$  be the hypothesis class and suppose  $\{E_{h_t}\}_{h_t \in \mathcal{H}}$  (satisfying  $\sum_{h_t \in \mathcal{H}} E_{h_t} = I$ ) is the final POVM performed by  $\mathcal{A}$ . Then we have the following,

$$\begin{aligned} |p' - p| &= \left| \text{Tr}(E_{h_t} |\Phi'_6\rangle \langle \Phi'_6|^{\otimes Q}) - \text{Tr}(E_{h_t} |\Phi_6\rangle \langle \Phi_6|^{\otimes Q}) \right| \\ &\leq \sum_{h_t \in \mathcal{H}} \left| \text{Tr}(E_{h_t} |\Phi'_6\rangle \langle \Phi'_6|^{\otimes Q}) - \text{Tr}(E_{h_t} |\Phi_6\rangle \langle \Phi_6|^{\otimes Q}) \right| \\ &\leq \| (|\Phi'_6\rangle \langle \Phi'_6|^{\otimes Q}) - (|\Phi_6\rangle \langle \Phi_6|^{\otimes Q}) \|_1 \\ &= 2(1 - \langle \Phi_6 | \Phi'_6 \rangle^{2Q})^{1/2} \\ &\leq 2(1 - (1 - 50\delta)^{2Q})^{1/2} \leq 2(1 - (1 - 50Q\delta)^2)^{1/2} \leq 8/T, \end{aligned} \quad (17)$$

where we have used the definition of trace distance in the second equality, Claim B.1 in the third inequality, Bernoulli's inequality  $(1 - x)^t \geq 1 - xt$  (for  $x \leq 1$  and  $t \geq 0$ ) in the penultimate inequality and  $\delta = 1/(10QT^2)$  in the final inequality. Additionally, the second inequality follows from the definition of the trace distance between quantum states

$$\| \rho - \sigma \|_1 = \max_{\{E_m\}} \sum_m |\text{Tr}(E_m(\rho - \sigma))|.$$

In particular, suppose we have a weak learner  $\mathcal{A}$  that outputs  $h_t$  with probability at least  $p'$ , then for every  $t \in [T]$ , we have

$$p = \Pr[\mathcal{A} \text{ outputs } h_t \text{ given } |\Phi_6\rangle^{\otimes Q}] \geq p' - 8/T \geq 1 - 1/T - 8/T = 1 - 9/T. \quad (18)$$

Hence, on passing  $|\Phi_6\rangle^{\otimes Q}$  to a quantum learner  $\mathcal{A}$ , it outputs a weak hypothesis  $h_t$  with probability at least  $1 - 9/T$  using Eq. (17). We assume that the output  $h_t$  is presented in terms of an oracle  $O_{h_t}$  (which on query  $|x, b\rangle$  outputs  $|x, b \cdot h_t(x)\rangle$  for all  $b \in \{-1, 1\}, x \in \{0, 1\}^n$ ).

**Phase (2): Computing  $\varepsilon'_t$ .** Using  $O_{h_t}$  produced in Phase (1), we now perform the query operation  $O_{h_t}$  on  $|\psi_3\rangle$  (defined in Eq. (10)) and obtain

$$|\psi_4\rangle = O_{h_t}|\psi_3\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_t(x) \neq c(x)], 0\rangle. \quad (19)$$

Using arithmetic operations, one can additionally produce the following state

$$|\psi_5\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t \cdot [h_t(x) \neq c(x)]\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_t(x) \neq c(x)], 0\rangle. \quad (20)$$

We now apply the controlled reflection operator  $V : |p\rangle|0\rangle|0\rangle \rightarrow |p\rangle(\sqrt{1-p}|0\rangle + \sqrt{p}|1\rangle)|\sin^{-1}(\sqrt{p})\rangle$  to  $|\psi_5\rangle|0\rangle$  to obtain

$$|\psi_6\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\beta_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_t(x) \neq c(x)]\rangle \otimes (\sqrt{1-\beta_x^t}|0\rangle + \sqrt{\beta_x^t}|1\rangle) \otimes |\sin^{-1}(\sqrt{\beta_x^t})\rangle,$$

where  $\beta_x^t = \tilde{D}_x^t[h_t(x) \neq c(x)]$ . We can rewrite the above equation as

$$|\psi_6\rangle = \sqrt{\tilde{\varepsilon}_t/M}|\phi_1\rangle|1\rangle + \sqrt{1-\tilde{\varepsilon}_t/M}|\phi_0\rangle|0\rangle, \quad (21)$$

where  $\tilde{\varepsilon}_t = \sum_{x \in S} \beta_x^t = \sum_{x \in S} \tilde{D}_x^t[h_t(x) \neq c(x)]$  and  $|\phi_0\rangle, |\phi_1\rangle$  are defined as

$$|\phi_0\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} \frac{\sqrt{1-\beta_x^t}}{\sqrt{1-\tilde{\varepsilon}_t/M}} |x, c(x)\rangle \otimes |\beta_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_t(x) \neq c(x)]\rangle \otimes |\sin^{-1}(\sqrt{\beta_x^t})\rangle, \quad (22a)$$

$$|\phi_1\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} \frac{\sqrt{\beta_x^t}}{\sqrt{\tilde{\varepsilon}_t/M}} |x, c(x)\rangle \otimes |\beta_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_t(x) \neq c(x)]\rangle \otimes |\sin^{-1}(\sqrt{\beta_x^t})\rangle. \quad (22b)$$

Let  $F_t$  be the unitary given by the map  $F_t : |\psi_1\rangle \rightarrow |\psi_6\rangle$  and  $\tilde{F}_t = F_t U_1$  be the map  $\tilde{F}_t : |0\rangle \rightarrow |\psi_6\rangle$ . Let  $P_t$  be the unitary that implements amplitude estimation using  $J_t$  invocations of  $\tilde{F}_t$  and  $\tilde{F}_t^{-1}$ . Our aim is to estimate  $\tilde{\varepsilon}_t$  with  $\varepsilon'_t$  up to a multiplicative error  $\delta$ , i.e.,  $|\tilde{\varepsilon}_t - \varepsilon'_t| \leq \delta \varepsilon'_t$ . We now run Algorithm 1 on the state  $|\psi_6\rangle$  assuming unitary access to  $\tilde{F}_t$ . Using the output of Algorithm 1, we compute  $\alpha'_t = \frac{1}{2} \ln((1 - \varepsilon'_t)/\varepsilon'_t)$ . This concludes the  $t$ th step of the quantum boosting algorithm. In the  $(t+1)$ th step, we use  $\varepsilon'_t$  and  $\alpha'_t$  to update the distribution from  $\tilde{D}^t$  to  $\tilde{D}^{t+1}$  and obtain the  $(t+1)$ th approximate weighted error  $\varepsilon'_{t+1}$  and the corresponding  $\alpha'_{t+1}$  respectively.

### C. Proof of unproven claims

In this section, we state and prove a few unproven claims from the previous section. We restate these claims for convenience of the reader and define the distribution update from Phase (3) in our quantum boosting algorithm: **If** subroutine 1 outputs 'yes': let  $Z_t = 2\sqrt{\varepsilon'_t(1-\varepsilon'_t)}$ ,  $\alpha'_t = \ln(\sqrt{(1-\varepsilon'_t)/\varepsilon'_t})$  and update  $\tilde{D}_x^t$ :

$$\tilde{D}_x^{t+1} = \frac{\tilde{D}_x^t}{(1+2\delta)Z_t} \times \begin{cases} e^{-\alpha'_t} & \text{if } h_t(x) = c(x) \\ e^{\alpha'_t} & \text{otherwise.} \end{cases} \quad (23)$$

If subroutine 1 outputs ‘no’: let  $Z_t = (2\sqrt{QT^2 - 1})/(QT^2)$ ,  $\alpha'_t = \ln(\sqrt{QT^2 - 1})$  and update  $\tilde{D}_x^t$ :

$$\tilde{D}_x^{t+1} = \frac{\tilde{D}_x^t}{(1 + 2/(QT^2))Z_t} \times \begin{cases} (2 - 1/(QT^2))e^{-\alpha'_t} & \text{if } h_t(x) = c(x) \\ (1/(QT^2))e^{\alpha'_t} & \text{otherwise.} \end{cases} \quad (24)$$

Additionally, we will crucially use the following relations multiple times in this section: for every  $t \geq 1$ , let  $\tilde{D}^t$  be the pseudo-distribution defined in Algorithm 1 (in particular, Eq. (23), (24)) and recall  $\tilde{\varepsilon}_t = \Pr_{x \sim \tilde{D}^t}[h_t(x) \neq c(x)]$ , then

$$\begin{aligned} \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x)) &= \sum_{i: h_t(x_i) = c(x_i)} \tilde{D}^t(x_i) \cdot e^{-\alpha'_t} + \sum_{i: h_t(x_i) \neq c(x_i)} \tilde{D}^t(x_i) \cdot e^{\alpha'_t} \\ &= (1 - \tilde{\varepsilon}_t) \cdot e^{-\alpha'_t} + \tilde{\varepsilon}_t \cdot e^{\alpha'_t}. \end{aligned} \quad (25)$$

Again, for every  $t \geq 1$ , let  $\tilde{D}^t$  be the pseudo-distribution defined in Algorithm 1 (in particular, Eq. (24)), then

$$\begin{aligned} \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]}) &= \sum_{i: h_t(x_i) = c(x_i)} \tilde{D}^t(x_i) \cdot e^{-\alpha'_t + \kappa_0} + \sum_{i: h_t(x_i) \neq c(x_i)} \tilde{D}^t(x_i) \cdot e^{\alpha'_t + \kappa_1} \\ &= (1 - \tilde{\varepsilon}_t) \cdot e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t \cdot e^{\alpha'_t + \kappa_1}. \end{aligned} \quad (26)$$

**Claim 4.3** Let  $t \geq 1$ ,  $\tilde{D}^t : \{0, 1\}^n \rightarrow [0, 1]$  be as defined in Eq. (23), (24). Then  $\sum_{x \in S} \tilde{D}_x^t \in [1 - 30\delta, 1]$ .

**Proof.** We divide the proof of the claim into two cases. Recall  $\delta = 1/(10QT^2)$ .

**Case I:** Suppose Algorithm 1 outputs ‘yes’ in the  $t$ th iteration. Recall the definition of  $\tilde{D}^{t+1}$ .

$$\tilde{D}_x^{t+1} = \frac{\tilde{D}_x^t}{2(1 + 2\delta)\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \times \begin{cases} e^{-\alpha'_t} & \text{if } h_t(x) = c(x) \\ e^{\alpha'_t} & \text{otherwise,} \end{cases} \quad (27)$$

where  $\alpha'_t = \frac{1}{2} \ln\left(\frac{1 - \varepsilon'_t}{\varepsilon'_t}\right)$  and  $|\tilde{\varepsilon}_t - \varepsilon'_t| \leq \delta \varepsilon'_t$ . In order to prove the lower bound, observe that

$$\begin{aligned} \sum_{x \in S} \tilde{D}_x^{t+1} &= \frac{1}{2(1 + 2\delta)\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x)) \\ &= \frac{\sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x))}{(1 - \tilde{\varepsilon}_t)e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}} \cdot \frac{(1 - \tilde{\varepsilon}_t)e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}}{2(1 + 2\delta)\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \\ &= \frac{(1 - \tilde{\varepsilon}_t)e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}}{2(1 + 2\delta)\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \quad (\text{using Eq. (25)}) \\ &= \frac{1}{2(1 + 2\delta)} \frac{1}{\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \cdot \left( (1 - \tilde{\varepsilon}_t) \sqrt{\frac{\varepsilon'_t}{1 - \varepsilon'_t}} + \tilde{\varepsilon}_t \sqrt{\frac{1 - \varepsilon'_t}{\varepsilon'_t}} \right) \quad (\text{using the definition of } \alpha'_t) \\ &= \frac{1}{2(1 + 2\delta)} \left( \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} + \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right), \end{aligned}$$

where the second equality used  $\sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x)) = (1 - \tilde{\varepsilon}_t)e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}$ , which follows from the following equation

$$\begin{aligned} Z_t = \sum_{i=1}^M \tilde{D}_t(x_i) \exp(-\alpha'_t h_t(x_i) c(x_i)) &= \sum_{i: h_t(x_i) = c(x_i)} \tilde{D}_t(x_i) \cdot e^{-\alpha'_t} + \sum_{i: h_t(x_i) \neq c(x_i)} \tilde{D}_t(x_i) \cdot e^{\alpha'_t} \\ &= (1 - \tilde{\varepsilon}_t) \cdot e^{-\alpha'_t} + \tilde{\varepsilon}_t \cdot e^{\alpha'_t}. \end{aligned} \quad (28)$$

Since  $|\tilde{\varepsilon}_t - \varepsilon'_t| \leq \delta \varepsilon'_t$ , we have

$$\frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \geq \frac{\varepsilon'_t(1 - \delta)}{\varepsilon'_t} = 1 - \delta. \quad (29)$$

Additionally,

$$\frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} \geq \frac{1 - \varepsilon'_t(1 + \delta)}{1 - \varepsilon'_t} = 1 - \frac{\delta \varepsilon'_t}{1 - \varepsilon'_t} \geq 1 - 2\delta, \quad (30)$$

where the second inequality uses  $\varepsilon'_t \leq 2/3$  (since we assume  $\varepsilon_t \leq 1/2$ ). Putting together Eq. (29) and Eq. (30) into the expression for  $\sum_{x \in S} \tilde{D}_x^{t+1}$ , we get

$$\sum_{x \in S} \tilde{D}_x^{t+1} \geq \frac{2 - 3\delta}{2(1 + 2\delta)} \geq 1 - 4\delta \geq 1 - 30\delta.$$

Next, we prove the upper bound. Note that

$$\frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \leq \frac{\varepsilon'_t(1 + \delta)}{\varepsilon'_t} = 1 + \delta, \quad (31)$$

and

$$\frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} \leq \frac{1 - \varepsilon'_t(1 - \delta)}{1 - \varepsilon'_t} = 1 + \frac{\delta \varepsilon'_t}{1 - \varepsilon'_t} \leq 1 + 2\delta, \quad (32)$$

where the second inequality uses  $\varepsilon'_t \leq 2/3$ . Using Eq. (31), (32), we have

$$\sum_{x \in S} \tilde{D}_x^{t+1} = \frac{1}{2(1 + 2\delta)} \left( \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} + \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right) \leq \frac{1}{2(1 + 2\delta)} \cdot \left( (1 + 2\delta) + (1 + \delta) \right) \leq \frac{2 + 4\delta}{2(1 + 2\delta)} = 1.$$

**Case II:** Suppose Algorithm 1 outputs ‘no’ in the  $t$ th iteration. The distribution  $\tilde{D}^{t+1}$  is then updated according to

$$\tilde{D}_x^{t+1} = \frac{\tilde{D}_x^t}{(1 + 2/(QT^2))Z_t} \times \begin{cases} (2 - 1/(QT^2))e^{-\alpha'_t} & \text{if } h_t(x) = c(x) \\ (1/(QT^2))e^{\alpha'_t} & \text{otherwise,} \end{cases} \quad (33)$$

where we use  $\varepsilon'_t = 1/(QT^2)$ ,  $\alpha'_t = \ln \sqrt{(1 - \varepsilon'_t)/\varepsilon'_t}$  and  $Z_t = 2\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}$ . Let  $\kappa_0 = \ln(2 - 1/(QT^2))$  and  $\kappa_1 = \ln(1/(QT^2))$ . In order to prove the upper and lower bounds of the claim, we first observe that

$$\begin{aligned} \sum_{x \in S} \tilde{D}_x^{t+1} &= \frac{1}{(1 + 2/(QT^2)) \cdot 2\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]}) \\ &= \frac{(1 - \tilde{\varepsilon}_t)e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}}{2(1 + 2/(QT^2))\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \quad (\text{using Eq. (26)}) \\ &= \frac{(2 - 1/(QT^2))(1 - \tilde{\varepsilon}_t)e^{-\alpha'_t} + (1/(QT^2))\tilde{\varepsilon}_t e^{\alpha'_t}}{2(1 + 2/(QT^2))\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \\ &= \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} + \frac{1}{2QT^2} \cdot \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right), \end{aligned}$$

where the second equality used  $\tilde{\varepsilon}_t = \sum_{x: h_t(x) \neq c(x)} \tilde{D}_x^t$ , third equality follows by the definition of  $\kappa_0, \kappa_1$  and the final equality used  $\alpha'_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon'_t}{\varepsilon'_t} \right)$ . We now prove the lower bound in the claim:

$$\begin{aligned}
 \sum_{x \in S} \tilde{D}_x^{t+1} &= \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} + \frac{1}{2QT^2} \cdot \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right) \\
 &\geq \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} \right) && \text{(using } \tilde{\varepsilon}_t > 0) \\
 &\geq \frac{1}{(1 + 2/(QT^2))} \cdot \left(1 - \frac{1}{2QT^2}\right) && \text{(using } 1 - \tilde{\varepsilon}_t \geq 1 - \varepsilon'_t) \\
 &\geq 1 - \frac{3}{QT^2} = 1 - 30\delta, && \text{(since } \delta = \frac{1}{10QT^2})
 \end{aligned}$$

where we used  $\tilde{\varepsilon}_t \leq \varepsilon'_t$  in the penultimate inequality because we are in the ‘no’ instance of Lemma 5.2 in Case II of our proof. We finally get the desired upper bound in the claim as follows

$$\begin{aligned}
 \sum_{x \in S} \tilde{D}_x^{t+1} &= \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} + \frac{1}{2QT^2} \cdot \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right) \\
 &\leq \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \frac{1 - \tilde{\varepsilon}_t}{1 - 1/(QT^2)} + \frac{1}{2QT^2} \right) && \text{(using } \tilde{\varepsilon}_t \leq \varepsilon'_t = 1/(QT^2)) \\
 &\leq \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \frac{1}{1 - 1/(QT^2)} + \frac{1}{2QT^2} \right) && \text{(using } 1 - \tilde{\varepsilon}_t \leq 1) \\
 &\leq \frac{1}{(1 + 2/(QT^2))} \left( \left(1 - \frac{1}{2QT^2}\right) \cdot \left(1 + \frac{2}{QT^2}\right) + \frac{1}{2QT^2} \right) \leq 1.
 \end{aligned}$$

□

**Claim 4.4** Let  $t \geq 1$ ,  $\tilde{\varepsilon}_t = \Pr_{x \sim \tilde{D}^t}[h_t(x) \neq c(x)]$  be the weighted error corresponding to the pseudo-distribution  $\tilde{D}^t$  and  $\varepsilon_t = \Pr_{x \sim D^t}[h_t(x) \neq c(x)]$  correspond to the true distribution  $D^t$ , which is defined as

$$D_x^{t+1} = \frac{\tilde{D}_x^t}{Z_t} \times \begin{cases} e^{-\alpha'_t} & \text{if } h_t(x) = c(x) \\ e^{\alpha'_t} & \text{otherwise,} \end{cases} \quad (34)$$

where  $\alpha'_t = \ln \left( \sqrt{(1 - \varepsilon'_t)/\varepsilon'_t} \right)$  and  $\varepsilon'_t, \tilde{\varepsilon}_t$  are defined in step (8) of Algorithm 1, and  $Z_t$  is defined in Eq. (28). Then  $|\tilde{\varepsilon}_t - \varepsilon_t| \leq 50\delta$ .

**Proof.** We break down the proof of the claim into two cases.

**Case I:** Algorithm 1 outputs ‘yes’ in the  $t$ th iteration. Recall the definition of the pseudo-distribution  $\tilde{D}^t$  in Eq. (23) and the



true distribution  $D^t$  in Eq. (34). We then have

$$\begin{aligned}
 |\tilde{\varepsilon}_{t+1} - \varepsilon_{t+1}| &= \left| \sum_{x \in S} \tilde{D}_x^{t+1} [h_{t+1}(x) \neq c(x)] - \sum_{x \in S} D_x^{t+1} [h_{t+1}(x) \neq c(x)] \right| \\
 &\leq \sum_{x \in S} \left| \tilde{D}_x^{t+1} - D_x^{t+1} \right| \cdot \left| [h_{t+1}(x) \neq c(x)] \right| \\
 &\leq \sum_{x \in S} \left| \tilde{D}_x^{t+1} - D_x^{t+1} \right| \quad (\text{since } [h_{t+1}(x) \neq c(x)] \leq 1) \\
 &= \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x)) \left| \frac{1}{2(1+2\delta)\sqrt{\varepsilon'_t(1-\varepsilon'_t)}} - \frac{1}{(1-\tilde{\varepsilon}_t) \cdot e^{-\alpha'_t} + \tilde{\varepsilon}_t \cdot e^{\alpha'_t}} \right| \\
 &= \frac{\sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x))}{(1-\tilde{\varepsilon}_t)e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}} \left| \frac{(1-\tilde{\varepsilon}_t)e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}}{2(1+2\delta)\sqrt{\varepsilon'_t(1-\varepsilon'_t)}} - 1 \right| \\
 &= \frac{1}{2(1+2\delta)} \left| \frac{1-\tilde{\varepsilon}_t}{1-\varepsilon'_t} + \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} - 2(1+2\delta) \right| \quad (\text{using Eq. (25)}) \\
 &\leq \frac{1}{2(1+2\delta)} \left( \left| \frac{\tilde{\varepsilon}_t - \varepsilon'_t}{1-\varepsilon'_t} \right| + \left| \frac{\tilde{\varepsilon}_t - \varepsilon'_t}{\varepsilon'_t} \right| + 4\delta \right) \quad (\text{using triangle inequality}) \\
 &\leq \frac{\delta}{2(1+2\delta)} \left( \left| \frac{\varepsilon'_t}{1-\varepsilon'_t} \right| + 5 \right) \quad (\text{using } |\tilde{\varepsilon}_t - \varepsilon'_t| \leq \delta\varepsilon'_t) \\
 &\leq \frac{7\delta}{2(1+2\delta)} \leq 4\delta. \quad (\text{using } \varepsilon'_t \leq 2/3)
 \end{aligned}$$

**Case II:** Algorithm 1 outputs ‘no’ in the  $t$ th iteration. Recall the definition of the pseudo-distribution  $\tilde{D}^t$  in Eq. (24). Let  $\kappa_0 = \ln(2 - 1/(QT^2))$  and  $\kappa_1 = \ln(1/(QT^2))$ . We have

$$\begin{aligned}
 |\tilde{\varepsilon}_{t+1} - \varepsilon_{t+1}| &= \left| \sum_{x \in S} \tilde{D}_x^{t+1} [h_{t+1}(x) \neq c(x)] - \sum_{x \in S} D_x^{t+1} [h_{t+1}(x) \neq c(x)] \right| \\
 &\leq \sum_{x \in S} |\tilde{D}_x^{t+1} - D_x^{t+1}| \\
 &= \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]}) \cdot \left| \frac{1}{2(1+2/(QT^2))\sqrt{\varepsilon'_t(1-\varepsilon'_t)}} - \frac{1}{(1-\tilde{\varepsilon}_t)e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}} \right| \\
 &= \frac{\sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]})}{(1-\tilde{\varepsilon}_t)e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}} \cdot \left| \frac{(1-\tilde{\varepsilon}_t)e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}}{2(1+2/(QT^2))\sqrt{\varepsilon'_t(1-\varepsilon'_t)}} - 1 \right| \\
 &= \frac{1}{2(1+2/(QT^2))} \left| \left( 2 - \frac{1}{QT^2} \right) \cdot \frac{1-\tilde{\varepsilon}_t}{1-\varepsilon'_t} + \frac{1}{QT^2} \cdot \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} - 2 \left( 1 + \frac{2}{QT^2} \right) \right| \quad (\text{using Eq. (26)}) \\
 &\leq \frac{1}{2(1+2/(QT^2))} \left( 2 \cdot \left| \frac{1-\tilde{\varepsilon}_t}{1-\varepsilon'_t} - 1 \right| + \frac{1}{QT^2} \cdot \left( \left| \frac{1-\tilde{\varepsilon}_t}{1-\varepsilon'_t} \right| + \left| \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right| \right) + \frac{4}{QT^2} \right) \quad (\text{using triangle inequality}) \\
 &\leq \frac{1}{2(1+2/(QT^2))} \left( \frac{2}{QT^2} \cdot \left( \frac{1}{1-1/(QT^2)} \right) + \frac{1}{QT^2} \cdot \left( \frac{1}{1-1/(QT^2)} + 1 \right) + \frac{4}{QT^2} \right) \\
 &\leq \frac{5}{QT^2} = 50\delta, \quad (\text{using } \delta = 1/(10QT^2))
 \end{aligned}$$

where the second last inequality used  $0 \leq \tilde{\varepsilon}_t \leq \varepsilon'_t = 1/(QT^2)$  and  $|\tilde{\varepsilon}_t - \varepsilon'_t| \leq 1/(QT^2)$ .  $\square$

**Claim B.1** Let  $t \geq 1$ ,  $|\Phi_6\rangle = \sum_{x \in S} \sqrt{\tilde{D}_x^t} |x, c(x)\rangle |0\rangle |0\rangle^t + O_{h_{t-1}} \cdots O_{h_1} \cdot \mathcal{G}_t^{-1} |\Psi\rangle$  and  $|\Phi'_6\rangle = \sum_{x \in S} \sqrt{D_x^t} |x, c(x)\rangle |0\rangle |0\rangle^t$  be defined as in Eq. (15), (16) respectively. Then we have  $|\langle \Phi_6 | \Phi'_6 \rangle| \geq 1 - 50\delta$ .

**Proof.** We break down the proof into two cases.

**Case I:** Algorithm 1 outputs ‘yes’ in the  $t$ th iteration. We now lower bound the inner product between

$$|\Phi_6\rangle = \sum_{x \in S} \sqrt{\tilde{D}_x^t} |x, c(x)\rangle |0\rangle |0\rangle^t + \underbrace{O_{h_{t-1}} \cdots O_{h_1} \cdot \mathcal{G}_t^{-1} |\Psi\rangle}_{:=|\Psi'\rangle}, \quad |\Phi'_6\rangle = \sum_{x \in S} \sqrt{D_x^t} |x, c(x)\rangle |0\rangle |0\rangle^t.$$

In order to do so, we first lower bound the following quantity

$$\begin{aligned} \sum_{x \in S} \sqrt{\tilde{D}_x^{t+1} D_x^{t+1}} &= \sum_{x \in S} \sqrt{\frac{\tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x))}{2(1+2\delta) \sqrt{\varepsilon'_t(1-\varepsilon'_t)}}} \cdot \sqrt{\frac{\tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x))}{(1-\tilde{\varepsilon}_t) e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}}} \\ &= \left( \frac{1}{2(1+2\delta) \sqrt{\varepsilon'_t(1-\varepsilon'_t)}} \cdot \frac{1}{(1-\tilde{\varepsilon}_t) e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}} \right)^{1/2} \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x)) \\ &= \left( \frac{(1-\tilde{\varepsilon}_t) e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}}{2(1+2\delta) \sqrt{\varepsilon'_t(1-\varepsilon'_t)}} \right)^{1/2} \cdot \frac{\sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x))}{(1-\tilde{\varepsilon}_t) e^{-\alpha'_t} + \tilde{\varepsilon}_t e^{\alpha'_t}} \\ &= \left( \frac{1}{2(1+2\delta)} \left( \frac{1-\tilde{\varepsilon}_t}{1-\varepsilon'_t} + \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right) \right)^{1/2} \cdot 1 \geq 1 - 2\delta, \end{aligned} \quad (35)$$

where the first equality used Eq. (23) and (34), the final equality used Eq. (25), and the final inequality used Eq. (29) and Eq. (30) to conclude

$$\frac{1}{2(1+2\delta)} \left( \frac{1-\tilde{\varepsilon}_t}{1-\varepsilon'_t} + \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right) \geq 1 - 4\delta.$$

We are now ready to prove the claim

$$\begin{aligned} |\langle \Phi_6 | \Phi'_6 \rangle| &= \left| \sum_{x \in S} \sqrt{\tilde{D}_x^t D_x^t} + \langle \Psi' | \Phi_6 \rangle \right| \geq \left| \left| \sum_{x \in S} \sqrt{\tilde{D}_x^t D_x^t} \right| - \left| \langle \Psi' | \Phi_6 \rangle \right| \right| && \text{(by reverse triangle inequality)} \\ &\geq \left| \sum_{x \in S} \sqrt{\tilde{D}_x^t D_x^t} \right| - \left| \langle \Psi' | \Phi_6 \rangle \right| \\ &\geq 1 - 2\delta - \left| \langle \Psi' | \Phi_6 \rangle \right| && \text{(using Eq.(35))} \\ &\geq 1 - 2\delta - \left\| |\Psi'\rangle \right\| = 1 - 2\delta - \left\| |\Psi\rangle \right\| \geq 1 - 50\delta && \text{(using Eq. (13))} \end{aligned}$$

where the penultimate inequality used  $\langle \Psi' | \Phi_6 \rangle \leq \left\| |\Psi'\rangle \right\| \leq 30\delta$  from Eq. (13).

**Case II:** Algorithm 1 outputs ‘no’ in the  $t$ th iteration. Recall that  $\kappa_0 = \ln(2 - 1/(QT^2))$  and  $\kappa_1 = \ln(1/(QT^2))$ . Using

Eq. (24) and  $0 \leq \tilde{\varepsilon}_t \leq \varepsilon'_t = 1/(QT^2)$ , we have

$$\begin{aligned}
 & \sum_{x \in S} \sqrt{\tilde{D}_x^{t+1} D_x^{t+1}} \\
 &= \sum_{x \in S} \sqrt{\frac{\tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]})}{2(1 + 2/(QT^2)) \sqrt{\varepsilon'_t(1 - \varepsilon'_t)}}} \cdot \sqrt{\frac{\tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]})}{(1 - \tilde{\varepsilon}_t) e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}}} \\
 &= \left( \frac{1}{2(1 + 2/(QT^2)) \sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \cdot \frac{1}{(1 - \tilde{\varepsilon}_t) e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}} \right)^{1/2} \cdot \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]}) \\
 &= \left( \frac{(1 - \tilde{\varepsilon}_t) e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}}{2(1 + 2/(QT^2)) \sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \right)^{1/2} \cdot \frac{\sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]})}{(1 - \tilde{\varepsilon}_t) e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t e^{\alpha'_t + \kappa_1}} \\
 &= \left( \frac{(2 - 1/(QT^2)) \cdot (1 - \tilde{\varepsilon}_t) e^{-\alpha'_t} + (1/(QT^2)) \cdot \tilde{\varepsilon}_t e^{\alpha'_t}}{2(1 + 2/(QT^2)) \sqrt{\varepsilon'_t(1 - \varepsilon'_t)}} \right)^{1/2} \\
 &= \left( \frac{1}{(1 + 2/(QT^2))} \cdot \left( \left( 1 - \frac{1}{2QT^2} \right) \cdot \frac{1 - \tilde{\varepsilon}_t}{1 - \varepsilon'_t} + \left( \frac{1}{2QT^2} \right) \cdot \frac{\tilde{\varepsilon}_t}{\varepsilon'_t} \right) \right)^{1/2} \geq 1 - \frac{3}{2QT^2}.
 \end{aligned} \tag{36}$$

The first equality used Eq. (24) and Eq. (26), the fourth equality used the modified distribution update in Eq. (24) to conclude

$$\begin{aligned}
 \sum_{x \in S} \tilde{D}_x^t \exp(-\alpha'_t c(x) h_t(x) + \kappa_{[h_t(x) \neq c(x)]}) &= \sum_{x: h_t(x) = c(x)} \tilde{D}_x^t \exp(-\alpha'_t + \kappa_1) + \sum_{x: h_t(x) \neq c(x)} \tilde{D}_x^t \exp(\alpha'_t + \kappa_0) \\
 &= (1 - \tilde{\varepsilon}_t) \cdot e^{-\alpha'_t + \kappa_0} + \tilde{\varepsilon}_t \cdot e^{\alpha'_t + \kappa_1}.
 \end{aligned}$$

and the last inequality used the lower bound on  $\sum_{x \in S} \tilde{D}_x^t$  in Claim 4.3 (Case II). We are now ready to prove the claim

$$\begin{aligned}
 |\langle \Phi_6 | \Phi'_6 \rangle| &\geq \left| \sum_{x \in S} \sqrt{\tilde{D}_x D_x^t} \right| - |\langle \Psi' | \Phi_6 \rangle| \\
 &\geq 1 - \frac{3}{2QT^2} - |\langle \Psi' | \Phi_6 \rangle| \quad \text{(using Eq.(36))} \\
 &\geq 1 - \frac{3}{2QT^2} - \|\Psi'\| \\
 &\geq 1 - \frac{3}{2QT^2} - \frac{3}{QT^2} \geq 1 - \frac{5}{QT^2} = 1 - 50\delta. \quad \text{(using Eq. (13) and } \delta = 1/(10QT^2)\text{)}
 \end{aligned}$$

This concludes the proof of the claim.  $\square$

## D. Proof of correctness

It remains to argue that the training error of  $H$  is  $\leq 1/10$ , i.e.,  $H(x) = c(x)$  for  $(9/10)$ -th fraction of the  $(x, c(x)) \in S$ . To prove this, we analyze the training error of  $H$  with respect to the uniform distribution  $\tilde{D}^1$  as follows. We break the proof of correctness into two cases and argue separately. In fact in the first case we will argue that  $H$  has zero training error and in the second case we will show the training error of  $H$  is at most  $1/10$ .

**Case I:** Suppose Algorithm 1 outputs ‘yes’ for every  $t \in [T]$ . This case corresponds to the setting where each weighted error  $\tilde{\varepsilon}_t$  is estimated by an  $\varepsilon'_t$  such that  $|\varepsilon'_t - \tilde{\varepsilon}_t| \leq \delta \varepsilon'_t$  for every iteration of the quantum boosting algorithm. If the output is ‘yes’, recall that

$$\tilde{D}^{t+1}(x) = \frac{\tilde{D}^t(x)}{Z'_t} \times \begin{cases} e^{-\alpha'_t} & \text{if } h_t(x) = c(x) \\ e^{\alpha'_t} & \text{otherwise} \end{cases} = \frac{\tilde{D}^t(x) \exp(-c(x) \alpha'_t h_t(x))}{Z'_t}. \tag{37}$$

where  $Z'_t = 2(1 + 2\delta)\sqrt{\varepsilon'_t(1 - \varepsilon'_t)}$ . By definition, we obtain

$$\tilde{D}^{T+1}(x) = \tilde{D}^1(x) \cdot \prod_{t=1}^T \frac{\exp(-c(x)\alpha'_t h_t(x))}{Z'_t} = \frac{D^1(x) \exp(-c(x) \cdot \sum_{t=1}^T \alpha'_t h_t(x))}{\prod_{t=1}^T Z'_t}, \quad (38)$$

where the second equality used  $\tilde{D}^1 = D^1$  which is the uniform distribution. We now upper bound the training error under the distribution  $D^1$

$$\begin{aligned} \Pr_{x \sim D^1}[H(x) \neq c(x)] &= \Pr_{x \sim D^1} \left[ \text{sign} \left( \sum_{t=1}^T \alpha'_t h_t(x) \right) \neq c(x) \right] \\ &\leq \Pr_{x \sim D^1} \left[ \exp \left( - \sum_{t=1}^T \alpha'_t h_t(x) \cdot c(x) \right) \right] \\ &= \sum_{i=1}^M D^1(x_i) \exp \left( - c(x_i) \sum_{t=1}^T \alpha'_t h_t(x_i) \right) = \sum_{i=1}^M \tilde{D}^{T+1}(x_i) \prod_{t=1}^T Z'_t \leq \prod_{t=1}^T Z'_t, \end{aligned} \quad (39)$$

where the first equality used the definition of  $H(x) = \text{sign}(\sum_{t=1}^T \alpha'_t h_t(x))$ , the first inequality used  $[\text{sign}(z) \neq y] \leq e^{-z \cdot y}$  for  $z \in \mathbb{R}, y \in \{-1, 1\}$ , the final equality used Eq. (38) and the final inequality used the fact that  $\tilde{D}^{T+1}$  is a pseudo-distribution. We are now in a stage to analyze the training error of  $H$  on  $D^1$ ,

$$\begin{aligned} \Pr_{x \sim D^1}[H(x) \neq c(x)] &\leq \prod_{t=1}^T Z'_t = (1 + 2\delta)^T \prod_{t=1}^T 2\sqrt{\varepsilon'_t(1 - \varepsilon'_t)} \quad (\text{using Eq. (39) and definition of } Z'_t) \\ &\leq e^{2\delta T} \prod_{t=1}^T 2\sqrt{\frac{\tilde{\varepsilon}_t}{1 - \delta} \cdot \left(1 - \frac{\tilde{\varepsilon}_t}{1 + \delta}\right)} \quad (\text{since } |\tilde{\varepsilon}_t - \varepsilon'_t| \leq \delta \varepsilon'_t) \\ &\leq e^{2\delta T} \prod_{t=1}^T 2\sqrt{\tilde{\varepsilon}_t(1 + 2\delta)(1 - \tilde{\varepsilon}_t(1 - \delta))} \\ &\leq e^{2\delta T} \prod_{t=1}^T 2\sqrt{(\varepsilon_t + 4\delta)(1 + 2\delta)(1 - (\varepsilon_t - 4\delta)(1 - \delta))} \quad (\text{using } |\tilde{\varepsilon}_t - \varepsilon_t| \leq 4\delta) \\ &\leq e^{2\delta T} \prod_{t=1}^T 2\sqrt{\varepsilon_t(1 - \varepsilon_t) + 75\delta} \\ &\leq e^{2\delta T} \prod_{t=1}^T 2\sqrt{1/4 - \gamma_t^2 + 75\delta} \quad (\text{since } \varepsilon_t \leq 1/2 - \gamma_t) \\ &= e^{2\delta T} \prod_{t=1}^T \sqrt{1 - 4(\gamma_t^2 - 75\delta)} \\ &\leq e^{2\delta T} \prod_{t=1}^T \sqrt{1 - 4(\gamma^2 - 75\delta)} \quad (\text{since } \gamma \leq \gamma_t \text{ for all } t) \\ &\leq \exp \left( 2\delta T - 2 \sum_{t=1}^T (\gamma^2 - 75\delta) \right) \quad (\text{since } 1 + x \leq e^x \text{ for } x \in \mathbb{R}) \\ &\leq \exp \left( -2T\gamma^2 + 16/(QT) \right), \quad (\text{since } \delta = 1/(10QT^2)) \end{aligned}$$

where we used Claim 4.4 (Case I) in the third inequality to conclude  $|\tilde{\varepsilon}_t - \varepsilon_t| \leq 4\delta$ .

In order to conclude the proof-of-correctness, note that for  $T = O((\log M)/\gamma^2)$  and for a sufficiently large constant in the  $O(\cdot)$ , the final upper bound on the expression is

$$\Pr_{x \sim D^1}[H(x) \neq c(x)] < 1/M.$$

Since  $D^1$  is the uniform distribution over  $S$ , i.e.,  $D_x^1 = 1/M$  for  $(x, c(x)) \in S$ , this implies that  $\Pr_{x \sim D^1}[H(x) \neq c(x)] = 0$ . Hence  $H$  has zero training error.

**Case II:** In this case, we assume that Algorithm 1 outputs ‘no’ in the first  $\ell \in [T]$  rounds of the quantum boosting Algorithm 1.<sup>4</sup> We additionally assume that  $\ell \leq T/\log(2\sqrt{QT}) - 1$ , which is standard in AdaBoost for the following reason: suppose the weighted errors of each of the first  $t \in [\ell]$  hypotheses satisfies  $\varepsilon_t \leq 1/QT^2 \ll 1/3$  (which is the ‘no’ instance of Algorithm 1), then observe that the resulting learner is *strong* and we need not do boosting in the first place. Moreover, suppose  $\ell \geq T/\log(2\sqrt{QT})$ , then observe that the final hypothesis after the  $T$  rounds of AdaBoost has training error at most  $1/10$  and we are done:

$$\begin{aligned} \Pr_{x \sim D^1}[H(x) \neq c(x)] &= \Pr_{x \sim D^1} \left[ \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right) \neq c(x) \right] \leq \prod_{t=1}^T Z_t = \prod_{t=1}^T 2\sqrt{\varepsilon_t(1-\varepsilon_t)} \\ &\leq \prod_{t=1}^{\ell} 2\sqrt{\varepsilon_t} \leq \left( \frac{2}{\sqrt{QT}} \right)^{\ell} \leq \frac{1}{10}, \end{aligned}$$

where the last equality used  $\ell \geq T/\log(2\sqrt{QT})$  and  $T \geq \log M$ .

So from here onwards we will assume  $\ell \geq T/\log(2\sqrt{QT})$  and still show that the training error is at most  $1/10$ . Note that for the first  $\ell$  iterations, the distribution follows the update rule, which defers from the standard AdaBoost update: for every  $k \in [\ell]$ ,

$$\begin{aligned} \tilde{D}_x^{k+1} &= \frac{\tilde{D}_x^k}{2(1 + 2/(QT^2))\sqrt{\varepsilon'_k(1-\varepsilon'_k)}} \times \begin{cases} (2 - 1/(QT^2))e^{-\alpha'_k} & \text{if } h_k(x) = c(x) \\ (1/(QT^2))e^{\alpha'_k} & \text{otherwise} \end{cases} \\ &= \frac{\tilde{D}_x^k(x) \exp(-\alpha'_k \cdot c(x)h_k(x) + \kappa_{[h_k(x) \neq c(x)]})}{Z'_k}, \end{aligned} \quad (40)$$

where  $\varepsilon'_k = 1/(QT^2)$  and  $Z'_k = 2(1 + 2/(QT^2))\sqrt{\varepsilon'_k(1-\varepsilon'_k)}$ . Let  $\kappa_0 = \ln(2 - 1/(QT^2))$  and  $\kappa_1 = \ln(1/(QT^2))$ . In particular, observe that for every  $\ell \geq 1$ , we have

$$\tilde{D}_x^{\ell+1} = \frac{D_x^1}{\prod_{i=1}^{\ell} Z'_i} \cdot \exp\left(-c(x) \cdot \sum_{i=1}^{\ell} \alpha'_i h_i(x)\right) \cdot \exp\left(\sum_{i=1}^{\ell} \kappa_{[h_i(x) \neq c(x)]}\right). \quad (41)$$

We bound the training error as follows:

$$\begin{aligned} \Pr_{x \sim D^1}[H(x) \neq c(x)] &\leq \sum_{x \in S} D_x^1 \exp\left(-c(x) \sum_{t=1}^T \alpha'_t h_t(x)\right) \\ &= \sum_{x \in S} D_x^1 \exp\left(-c(x) \sum_{t=1}^{\ell} \alpha'_t h_t(x)\right) \cdot \exp\left(-c(x) \sum_{t=\ell+1}^T \alpha'_t h_t(x)\right) \\ &= \prod_{t=1}^{\ell} Z'_t \sum_{x \in S} \tilde{D}_x^{\ell+1} \exp\left(-c(x) \sum_{t=\ell+1}^T \alpha'_t h_t(x)\right) \cdot \exp\left(-\sum_{i=1}^{\ell} \kappa_{[h_i(x) \neq c(x)]}\right) \\ &= \prod_{t=1}^T Z'_t \sum_{x \in S} \tilde{D}_x^{T+1} \exp\left(-\sum_{i=1}^{\ell} \kappa_{[h_i(x) \neq c(x)]}\right) \leq \prod_{t=1}^T Z'_t \cdot (QT^2)^{\ell} \cdot \sum_{x \in S} \tilde{D}_x^{T+1} \leq \prod_{t=1}^T Z'_t \cdot (QT^2)^{\ell}, \end{aligned}$$

where the second equality uses Eq. (41) (i.e., distribution update for ‘no’ instances) and third equality uses Eq. (38) (i.e., distribution update for the ‘yes’ instances), the penultimate inequality uses  $\exp(-\kappa_0) \leq \exp(-\kappa_1) \leq QT^2$  (we remark that this bound is very loose, since  $\exp(\kappa_0) = O(1)$ ) and the final inequality uses the fact that  $\tilde{D}$  is a pseudo-distribution by

<sup>4</sup>Our analysis also works when Algorithm 1 outputs ‘no’ for arbitrary  $\ell$  rounds of the quantum boosting algorithm instead of the first  $\ell$  rounds.

Claim 4.3. Continuing to upper bound the above expression, we get

$$\begin{aligned}
 \Pr_{x \sim D^1} [H(x) \neq c(x)] &\leq (QT^2)^\ell \prod_{t=1}^T Z'_t \\
 &= \left( (QT^2)^\ell \prod_{t=1}^{\ell} Z'_t \right) \cdot \left( \prod_{t=\ell+1}^T Z'_t \right) \\
 &= \left( (QT^2)^\ell (1 + 2/(QT^2))^\ell \prod_{t=1}^{\ell} 2\sqrt{1/(QT^2) \cdot (1 - 1/(QT^2))} \right) \cdot \left( (1 + 2\delta)^{T-\ell} \prod_{t=\ell+1}^T 2\sqrt{\varepsilon'_t(1 - \varepsilon'_t)} \right) \\
 &\leq \left( (QT^2)^\ell (1 + 2/(QT^2))^\ell \prod_{t=1}^{\ell} 2/\sqrt{QT^2} \right) \cdot \left( (1 + 2\delta)^{T-\ell} \prod_{t=\ell+1}^T 2\sqrt{\varepsilon'_t(1 - \varepsilon'_t)} \right) \\
 &\leq \left( (2\sqrt{QT})^\ell \exp\left((2\ell)/(QT^2)\right) - 2(T - \ell)\gamma^2 + (16T - 16\ell)/(QT^2) \right) \\
 &\leq (2\sqrt{QT})^\ell \exp\left(-2(T - \ell)\gamma^2 + 16/(QT)\right) \leq \exp\left(2\ell(\ln(2\sqrt{QT}) + \gamma^2) - 2T\gamma^2 + 1\right),
 \end{aligned}$$

where the second equality used

$$Z'_t = \begin{cases} 2(1 + 2/(QT^2))\sqrt{1/(QT^2) \cdot (1 - 1/(QT^2))} & \text{for } t \leq \ell \\ 2(1 + 2\delta) \cdot \sqrt{\varepsilon'_t(1 - \varepsilon'_t)} & \text{for } t \geq \ell + 1, \end{cases}$$

the third inequality used  $1 + x \leq e^x$  for  $x \in \mathbb{R}$  and the second factor

$$\exp\left(-2(T - \ell)\gamma^2 + (16T - 16\ell)/(QT^2)\right),$$

came from the upper bound of the training error derived in Case I with  $T$  replaced by  $T - \ell$  (recall that we had showed  $\Pr_{x \sim D^1} [H(x) \neq c(x)] \leq \prod_{t=1}^T Z'_t \leq \exp(-2T\gamma^2 + 16/(QT))$ ). Finally, using  $\ell \leq T/\ln(2\sqrt{QT}) - 1$ , we have

$$\begin{aligned}
 \Pr_{x \sim D^1} [H(x) \neq c(x)] &\leq \exp\left(2\ell(\ln(2\sqrt{QT}) + \gamma^2) - 2T\gamma^2 + 1\right) \\
 &\leq \exp\left(\left(\frac{2T}{\ln(2\sqrt{QT})} - 2\right) \cdot (\ln(2\sqrt{QT}) + \gamma^2) - 2T\gamma^2 + 1\right) \\
 &= \exp\left(2T - 2\gamma^2 - 2\ln(2\sqrt{QT}) + \frac{2T\gamma^2}{\ln(2\sqrt{QT})} - 2T\gamma^2 + 1\right) \leq \frac{e}{4QT^2} \leq \frac{1}{10},
 \end{aligned}$$

where the final inequality used that  $Q, T = O(\log M)$  are sufficiently large. Hence, we have shown that  $H$  has training error at most  $1/10$ .

## E. Complexity of the algorithm

First we analyze the *query* complexity of the quantum boosting algorithm (where the query complexity refers to the total number of queries made to the hypothesis-oracles  $\{O_{h_1}, \dots, O_{h_T}\}$ ). We consider the complexity of the  $t$ th iteration: in *phase 1*, the number of queries made to  $\{h_1, \dots, h_{t-1}\}$  in order for the quantum weak learner  $\mathcal{A}$  to output the hypothesis  $h_t$  is at most  $\sqrt{M}Q \cdot t$ : the  $\sqrt{M}$ -factor comes from amplitude amplification and the application of the unitary  $\tilde{W}_t : |0\rangle \rightarrow |\Phi_4\rangle$  involves  $Q(t-1)$  queries for the  $Q$  copies of the input to the weak learner. An additional  $Q(t-1)$  queries to  $\{h_1, \dots, h_{t-1}\}$  are required while applying  $O_{h_1}, \dots, O_{h_{t-1}}$  to uncompute the queries. In *phase 2*, the number of queries made during multiplicative amplitude estimation in order to compute  $\varepsilon'_t$  is  $\sqrt{M}Q^{3/2}T^3 \cdot t$ : the  $\sqrt{M}Q^{3/2}T^3$ -factor is due to multiplicative amplitude estimation (in Lemma 5.2). Furthermore, each application of  $\tilde{F}_t : |0\rangle \rightarrow |\psi_6\rangle$  involves making  $t$  queries. Putting together the contribution from both phases, the total query complexity of the quantum boosting algorithm is

$$\sum_{t=1}^T \sqrt{M}Q(t-1) + Q(t-1) + \sqrt{M}Q^{3/2}T^3t = O(\sqrt{M}Q^{3/2}T^5 + \sqrt{M}QT^2) = \tilde{O}(\sqrt{M}Q^{3/2}T^5).$$

We now discuss the time complexity of the quantum boosting algorithm. We begin by analyzing the time complexity of the  $t$ th iteration. Assuming that a quantum RAM can prepare a uniform superposition  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle$  using  $O(n \log M)$  gates, the time complexity of preparing the initial state  $|\psi_1\rangle \otimes |\Phi_1\rangle^{\otimes Q}$  is  $O(nQ)$ .<sup>5</sup> In the second step, in order to prepare  $|\psi_2\rangle \otimes |\Phi_2\rangle^{\otimes Q}$ , our quantum algorithm uses  $O(Qt)$  quantum queries to  $\{h_1, \dots, h_{t-1}\}$  and this can be performed in time  $O(Qt)$ . The third step involves updating the registers from  $\tilde{D}^1$  to  $\tilde{D}^t$  which requires  $Q + 1$  applications of the control unitary  $\mathcal{G}_t$ . Since there are  $t - 1$  control qubits and updating the distribution register is an arithmetic operation, the third step for implementing  $O(Q)$  operations of  $\mathcal{G}_t$  can be performed in  $O(n^2Qt)$  time.<sup>6</sup>

In phase 1 of the quantum algorithm, we perform amplitude amplification with the unitary  $Y_t^{\otimes Q}$  which makes  $O(\sqrt{MQ})$  calls to  $\tilde{W}_t$  and  $\tilde{W}_t^{-1}$ . This takes time  $O(n^2\sqrt{MQ}t)$ . Next, in order to uncompute the  $t - 1$  quantum queries in the  $Q$  copies, our algorithm uses  $O(nQt)$  time. The weak learner  $\mathcal{A}$  takes as input  $Q$  samples and outputs a hypothesis  $h_t$  in time  $O(n^2Q)$ . Note that we require the quantum learning algorithm to output an oracle for  $h_t$  instead of explicitly outputting a circuit for  $h_t$ .

In phase 2, the algorithm initially performs an arithmetic operation  $\sum_{x \in S} |x\rangle |\tilde{D}_x^t\rangle | [h_t(x) \neq c(x)] \rangle \rightarrow \sum_{x \in S} |x\rangle |\tilde{D}_x^t\rangle | [h_t(x) \neq c(x)] \rangle$  using  $O(n)$  gates. Then a controlled reflection operator  $V : |p\rangle|0\rangle|0\rangle \rightarrow |p\rangle(\sqrt{1-p}|0\rangle + \sqrt{p}|1\rangle) |\sin^{-1}(\sqrt{p})\rangle$  is applied where the operation  $|p\rangle|0\rangle|0\rangle \rightarrow |p\rangle|0\rangle|\sin^{-1}(\sqrt{p})\rangle$  is an arithmetic process and uses  $O(n)$  gates while the operation  $|p\rangle|0\rangle|\sin^{-1}(\sqrt{p})\rangle \rightarrow |p\rangle(\sqrt{1-p}|0\rangle + \sqrt{p}|1\rangle) |\sin^{-1}(\sqrt{p})\rangle$  uses one controlled rotation gate. The next step is phase estimation which involves applying QFT using  $O(n \cdot \log n)$  gates. The time required for amplitude estimation in order to compute  $\varepsilon_t'$  is  $O(\sqrt{MQ}^{3/2}T^3 \cdot tn^2)$ : the  $\sqrt{MQ}^{3/2}T^3$  calls are made to the unitaries  $\tilde{F}_t, \tilde{F}_t^{-1}$  and each application of  $\tilde{F}_t : |0\rangle \rightarrow |\psi_6\rangle$  requires  $O(n^2 \cdot t)$  time. The overall time complexity of the quantum algorithm is

$$\sum_{t=1}^T O\left(n^2\sqrt{MQ}^{3/2}T^3t + n^2\sqrt{MQ}t + n^2Qt\right) = \tilde{O}(n^2\sqrt{MQ}^{3/2}T^5).$$

## F. A special case of quantum boosting with better dependence on $1/\gamma$

In the previous section, we presented a quantum boosting algorithm with complexity  $T_Q$  which depended on  $1/\gamma^{11}$ . In this section, we consider a special case of our quantum boosting algorithm and show how to improve the complexity from  $1/\gamma^{11}$  to  $1/\gamma^6$ . Our quantum boosting Algorithm 1 has two important phases, the first phase for obtaining the hypothesis  $h_t$  and second phase for computing the approximate weighted errors. In this section, we *assume* that our quantum boosting algorithm has oracle access to the  $T$  weak hypotheses  $\{h_1, \dots, h_T\}$  without the need of invoking a quantum weak learner to produce  $\{h_1, \dots, h_T\}$  (note that these hypotheses are weak with respect to the pseudo-distributions  $\tilde{D}_1, \dots, \tilde{D}_T$ ). We remark that Wang et al. (Wang et al., 2019) consider this setting in their quantum AdaBoost algorithm.

**Theorem F.1** Fix  $\eta > 0$  and  $\gamma > 0$ . Let  $n \geq 1$  and  $\mathcal{C} \subseteq \{c : \{0, 1\}^n \rightarrow \{-1, 1\}\}$  be a concept class and  $\mathcal{D} : \{0, 1\}^n \rightarrow [0, 1]$  be an unknown distribution. Let  $M$  be the smallest integer exceeding  $M \geq (\text{VC}(\mathcal{C})/\gamma^2) \cdot (\log(\text{VC}(\mathcal{C})/\gamma^2)/\eta^2)$ . Suppose Algorithm 1 has oracle access to  $T \geq ((\log M) \cdot \log(1/\delta))/\gamma^2$  hypotheses, where the  $t$ -th hypothesis  $h_t$  is weak with respect to the  $t$ -th pseudo-distribution  $\tilde{D}_t$  on the training sample  $S$ , i.e.,  $\sum_{x \in S} \tilde{D}_t[h_t(x) = c(x)] \geq 1/2 + \gamma$ . Then with probability  $\geq 2/3$  (over the randomness of the algorithm), a quantum algorithm can produce a hypothesis  $H$  that has training error at most  $1/10$  and small generalization error

$$\Pr_{x \sim \mathcal{D}} [H(x) = c(x)] \geq 1 - 1/10 - \eta.$$

Moreover, the time complexity of the algorithm is

$$T_Q = O(n^2\sqrt{MT}^{5/2}) = \tilde{O}\left(\frac{\sqrt{\text{VC}(\mathcal{C})}}{\eta} \cdot \frac{n^2}{\gamma^6} \cdot \text{polylog}(1/\delta)\right). \quad (42)$$

Picking  $\eta = 1/10$  we get that  $H$  has generalization error at most  $1/5$ .

<sup>5</sup>As we mentioned earlier, we could also assume that a quantum learning algorithm has access to the uniform quantum examples  $\frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle$ , in which case we do not need to assume a quantum RAM.

<sup>6</sup>A quantum circuit can perform arithmetic operations with the same time complexity as a Boolean circuit (Kitaev, 1995).

**Proof sketch.** The proof of the algorithm follows the same structure as the proof of Theorem 5.1. The only difference is the following: observe that in the  $t$ th round of Algorithm 1, we made  $t$  quantum queries to the hypotheses  $\{h_1, \dots, h_t\}$  to update the pseudo-distribution  $\tilde{D}^t$  to  $\tilde{D}^{t+1}$ . This eventually led to an overall query complexity of  $\sum_{t=1}^T O(t) = O(T^2)$  to the hypothesis functions. The improvement in this section follows from the observation that in order to update the pseudo-distribution  $\tilde{D}^{t+1}$  to  $\tilde{D}^{t+2}$ , we need *not* make  $t + 1$  further queries to  $\{h_1, \dots, h_{t+1}\}$ . Alternatively, we could *reuse* the queries that were already made in order to compute  $\tilde{D}^{t+1}$  and then make one additional query to  $h_{t+1}$ . Hence, the overall query complexity to the hypothesis functions can be brought down to  $\sum_{t=1}^T O(1) = O(T)$ . In order to *reuse* the queries, we use a version of coherent amplitude estimation which was recently proposed by Harrow and Wei in Theorem A.2.

Let  $\tilde{D}_x^0 = \tilde{D}_x^1 = 1/M$  for every  $(x, c(x)) \in S$ . Let  $h_0$  be the identity function and  $\varepsilon'_0 = 1/2$ . We analyze our quantum algorithm. In the  $t$ th round of the special boosting algorithm, we have quantum query access to the hypotheses  $\{h_1, \dots, h_T\}$ , the approximate weighted errors  $\{\varepsilon'_1, \dots, \varepsilon'_{t-1}\}$  and the algorithm begins with the state  $|\psi_1^t\rangle$  where

$$|\psi_1^t\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^{t-1}\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle. \quad (43)$$

Apply the controlled unitary  $\mathcal{G}'_t : \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^{t-1}\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle \rightarrow \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle$  on the state  $|\psi_1^t\rangle$  to update the pseudo-distribution  $\{\tilde{D}_x^{t-1}\}_x$ . The resulting state  $|\psi_2^t\rangle$  is

$$|\psi_2^t\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)]\rangle. \quad (44)$$

**Computing  $\varepsilon_t$ :** Now we use  $|\psi_2^t\rangle$  to compute  $\varepsilon'_t$ . By making a query to  $\mathcal{O}_{h_t}$ , we obtain

$$|\psi_3^t\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_{t-1}(x) \neq c(x)], [h_t(x) \neq c(x)]\rangle.$$

Let  $F_t$  be a unitary that takes as input  $|\psi_3^t\rangle$  and produces the state

$$|\psi_4^t\rangle = \left( \sqrt{1 - \tilde{\varepsilon}_t/M} |\phi_0^t\rangle |0\rangle + \sqrt{\tilde{\varepsilon}_t/M} |\phi_1^t\rangle |1\rangle \right), \quad (45)$$

where  $\tilde{\varepsilon}_t = \sum_{x \in S} \tilde{D}_x^t [h_t(x) \neq c(x)]$ . We do not explicitly describe the states  $|\phi_0^t\rangle$  and  $|\phi_1^t\rangle$  (since they were already present in Eq. (22) in the proof of Theorem 5.1). Let  $\tilde{F}_t = F_t U_1^t$  be the map  $\tilde{F}_t : |0\rangle \rightarrow |\psi_4^t\rangle$ . Suppose  $P_t$  is the unitary that implements *non-destructive* amplitude estimation (Theorem A.2) using  $J_t$  invocations of  $\tilde{F}_t$  and  $\tilde{F}_t^{-1}$ . Our aim is to estimate  $\tilde{\varepsilon}_t$  with  $\varepsilon'_t$  such that  $|\tilde{\varepsilon}_t - \varepsilon'_t| \leq \delta \varepsilon'_t$  where  $\delta = 1/(10T)$ . Using the state  $|\psi_4^t\rangle$  and the unitary  $\tilde{F}_t$ , we invoke Algorithm 1 to compute  $\varepsilon'_t$  and the corresponding  $\alpha'_t = \ln \sqrt{(1 - \varepsilon'_t)/\varepsilon'_t}$ .

In addition, by Theorem A.2, the unitary  $P_t$  (after outputting  $\varepsilon'_t$ ) restores the original state  $|\psi_4^t\rangle$  with probability at least  $1 - O(1/T)$ . Since we run the non-destructive amplitude estimation unitary  $P_t$  for  $T$  rounds, then by a union bound, the probability that all the  $P_t$ s restore the original state  $|\psi_4^t\rangle$  is at least  $2/3$ . Suppose  $P_t$  restores the original state  $|\psi_4^t\rangle$ , then after applying the inverse of the unitary  $\mathcal{V}_t : |\psi_3^t\rangle \rightarrow |\psi_4^t\rangle$ , we obtain

$$|\psi_1^{t+1}\rangle = \frac{1}{\sqrt{M}} \sum_{x \in S} |x, c(x)\rangle \otimes |\tilde{D}_x^t\rangle \otimes |[h_1(x) \neq c(x)], \dots, [h_t(x) \neq c(x)]\rangle. \quad (46)$$

**Proof of correctness.** We now prove that the final hypothesis  $H(x) = \text{sign} \left( \sum_{t=1}^T \alpha'_t h_t(x) \right)$  is strong following the analysis of the proof of correctness of Theorem 5.1.

**Case I:** Suppose Algorithm 1 outputs ‘yes’ for every  $t \in [T]$ . Then for  $\delta = 1/(10T)$ , we have

$$\Pr_{x \sim D^1} [H(x) \neq c(x)] = (1 + 2\delta)^T \prod_{t=1}^T 2\sqrt{\varepsilon'_t(1 - \varepsilon'_t)} \leq \exp(-2T\gamma^2 + 16).$$



When  $T \geq (\log M)/\gamma^2$ , the training error  $\Pr_{x \sim D^1}[H(x) \neq c(x)] < 1/M$ . Since  $D_x^1 = 1/M$  for every  $(x, c(x)) \in S$ , the final hypothesis  $H$  achieves *zero training error* on the training set  $S$ .

**Case II:** We assume that Algorithm 1 outputs ‘no’ in the first  $\ell \in [T]$  rounds of the special boosting algorithm and  $\ell \leq (T\gamma^2)/(2 \log T)$ . We show that after  $T = O((\log M)/\gamma^2)$  rounds of boosting,  $H$  achieves training error at most  $1/10$ . For the first  $\ell$  iterations, the pseudo-distribution  $\tilde{D}^t$  follows the update rule: for every  $k \in [\ell]$

$$\tilde{D}_x^{k+1} = \frac{\tilde{D}_x^k}{2(1 + 2/T)\sqrt{\varepsilon'_k(1 - \varepsilon'_k)}} \times \begin{cases} (2 - 1/T)e^{-\alpha'_k} & \text{if } h_k(x) = c(x) \\ (1/T)e^{\alpha'_k} & \text{otherwise,} \end{cases} \quad (47)$$

where  $\varepsilon'_k = 1/T$ . The training error of  $H$  is bounded by:

$$\Pr_{x \sim D^1}[H(x) \neq c(x)] \leq \sum_{x \in S} D^1(x) \exp\left(-c(x) \sum_{t=1}^T \alpha'_t h_t(x)\right) \leq \exp\left(2\ell(\ln(\sqrt{T}) + \gamma^2) - 2T\gamma^2 + 1\right). \quad (48)$$

Using  $\ell \geq T/\log(2\sqrt{QT})$  and  $T = O((\log M)/\gamma^2)$ , the training error  $\Pr_{x \sim D^1}[H(x) \neq c(x)] < 1/10$ .

**Complexity of quantum algorithm.** In the  $t$ th round, the number of queries made by multiplicative amplitude estimation ( Theorem A.2 ) in order to compute  $\varepsilon'_t$  is  $O(\sqrt{MT}^{3/2})$ . Then the total query complexity of the algorithm is  $\sum_{t=1}^T O(\sqrt{MT}^{3/2}) = O(\sqrt{MT}^{5/2})$  and the overall time complexity is  $O(n^2\sqrt{MT}^{5/2})$ .

## References

- Ambainis, A. Quantum search with variable times. *Theory of Computing Systems*, 47(3):786–807, 2010.
- Arunachalam, S., Gheorghiu, V., Jochym-O’Connor, T., Mosca, M., and Srinivasan, P. V. On the robustness of bucket brigade quantum RAM. *New Journal of Physics*, 17(12):123010, 2015. arXiv:1502.03450.
- Giovannetti, V., Lloyd, S., and Maccone, L. Quantum random access memory. *Physical review letters*, 100(16):160501, 2008. arXiv:0708.1879.
- Harrow, A. W. and Wei, A. Y. Adaptive quantum simulated annealing for bayesian inference and estimating partition functions. *arXiv preprint arXiv:1907.09965*, 2019.
- Kitaev, A. Y. Quantum measurements and the Abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995.
- Wang, X., Ma, Y., Hsieh, M., and Yung, M. Quantum speedup in adaptive boosting of binary classification. 2019. arXiv:1902.00869.