A. Quantum Algorithms for Classification: Details

A.1. Quantum state preparation

In this paper, we use the following result for quantum state preparation (see, e.g., Grover (2000)):

Proposition A.1. *Assume that $a \in \mathbb{C}^n$, and we are given a unitary oracle O_a such that $O|i\rangle|0\rangle = |i\rangle|a_i\rangle$ for all $i \in [n]$. Then Algorithm 3 takes $O(\sqrt{n})$ calls to O_a for preparing the quantum state $\frac{1}{\|a\|_2}\sum_{i\in[n]} a_i|i\rangle$ with success probability $1 - O(1/n)$.*

Algorithm 3: Prepare a pure state given an oracle to its coefficients.

---

1 Apply Dürr-Høyer's algorithm (Dürr & Høyer, 1996) to find $a_{\max} := \max_{i\in[n]} |a_i|$ in $O(\sqrt{n})$ time;

2 Prepare the uniform superposition $\frac{1}{\sqrt{n}}\sum_{i\in[n]} |i\rangle$;

3 Perform the following unitary transformations:

$$\frac{1}{\sqrt{n}}\sum_{i\in[n]}|i\rangle \xmapsto{O_a} \frac{1}{\sqrt{n}}\sum_{i\in[n]}|i\rangle|a_i\rangle \mapsto \frac{1}{\sqrt{n}}\sum_{i\in[n]}|i\rangle|a_i\rangle\Big(\frac{a_i}{a_{\max}}|0\rangle + \sqrt{1-\frac{|a_i|^2}{a_{\max}^2}}|1\rangle\Big)$$
$$\xmapsto{O_a^{-1}} \frac{1}{\sqrt{n}}\sum_{i\in[n]}|i\rangle|0\rangle\Big(\frac{a_i}{a_{\max}}|0\rangle + \sqrt{1-\frac{|a_i|^2}{a_{\max}^2}}|1\rangle\Big); \tag{A.1}$$

4 Delete the second system in Eq. (A.1), and rewrite the state as

$$\frac{\|a\|_2}{\sqrt{n}a_{\max}} \cdot \Big(\frac{1}{\|a\|_2}\sum_{i\in[n]} a_i|i\rangle\Big)|1\rangle + |a^{\perp}\rangle|0\rangle, \tag{A.2}$$

where $|a^{\perp}\rangle := \frac{1}{\sqrt{n}}\sum_{i\in[n]}\sqrt{1-\frac{|a_i|^2}{a_{\max}^2}}|i\rangle$ is a garbage state;

5 Apply amplitude amplification (Brassard et al., 2002) for the state in (A.2) conditioned on the second system being 1. Return the output;

Note that the coefficient in (A.2) satisfies $\frac{\|a\|_2}{\sqrt{n}a_{\max}} \geq \frac{1}{\sqrt{n}}$; therefore, applying amplitude amplification for $O(\sqrt{n})$ times indeed promises that we obtain 1 on the second system with success probability $1 - O(1/n)$, i.e., the state $\frac{1}{\|a\|_2}\sum_{i\in[n]} a_i|i\rangle$ is prepared.

Remark A.1. *Another common method to prepare quantum states is via quantum random access memory (QRAM). We point out that Algorithm 3 is incomparable to state preparation via QRAM. QRAM relies on the weak assumption that we start from zero, and every added datum is processed in poly-logarithmic time. In total, this takes at least linear time in the size of the data (see, for instance, Kerenidis & Prakash (2017)). For the task of Proposition A.1, QRAM takes at least $\Omega(n)$ cost.*

In this paper, we use the standard model where the input is formulated as an oracle, also widely assumed and used in existing quantum algorithm literatures (e.g., Grover (1996); Harrow et al. (2009); Childs et al. (2017); Brandão et al. (2017)). Under the standard model, Algorithm 3 prepares states with only $O(\sqrt{n})$ cost.

Nevertheless, it is an interesting question to ask whether there is a $\mathrm{poly}(\log(nd))$-time quantum algorithm for linear classification given the existence of a pre-loaded QRAM of X. This would require the ability to take summations of the vectors $\frac{1}{\sqrt{2T}}X_{i_t}$ in Line 5 of Algorithm 1 in $\mathrm{poly}(\log(nd))$-time as well as the ability to update the weight state u_{t+1} in Line 8 in $\mathrm{poly}(\log(nd))$-time, both using QRAM. These two tasks are plausible as suggested by classical poly-log time sample-based algorithms for matrix arithmetics under multiplicative weight frameworks (Chia et al., 2019), which can potentially be combined with the analysis of QRAM data structures in Kerenidis & Prakash (2017); we leave this possibility as an open question.

A.2. Implementation of the quantum oracle for updating the weight vectors

The quantum oracle O_t in Line 7 of Algorithm 1 is implemented by Algorithm 4. For convenience, we denote $\operatorname{clip}(v, 1/\eta) := \min\{1/\eta, \max\{-1/\eta, v\}\}$ for all $v \in \mathbb{R}$.

Algorithm 4: Quantum oracle for updating the weight state.

Input: $w_1, \ldots, w_t \in \mathbb{R}^d$, $j_1, \ldots, j_t \in [d]$.
Output: An oracle O_t such that $O_t|i\rangle|0\rangle = |i\rangle|u_{t+1}(i)\rangle$ for all $i \in [n]$.

Define three classical oracles: $O_{s,j}(0) = j_s$, $O_{s,w}(j_s) = \frac{\|w_s\|^2}{w_s(j_s)}$, and $O_{\text{clip}}(a, b, c) = c \cdot \big(1 - \eta \operatorname{clip}(ab, 1/\eta) + \eta^2 \operatorname{clip}(ab, 1/\eta)^2\big)$;

for $s = 1$ **to** t **do**

Perform the following maps:

$$|i\rangle|0\rangle|0\rangle|0\rangle|u_s(i)\rangle \xmapsto{O_{s,j}} |i\rangle|j_s\rangle|0\rangle|0\rangle|u_s(i)\rangle \tag{A.3}$$

$$\xmapsto{O_X} |i\rangle|j_s\rangle|X_i(j_s)\rangle|0\rangle|u_s(i)\rangle \tag{A.4}$$

$$\xmapsto{O_{s,w}} |i\rangle|j_s\rangle|X_i(j_s)\rangle\Big|\frac{\|w_s\|^2}{w_s(j_s)}\Big\rangle|u_s(i)\rangle \tag{A.5}$$

$$\xmapsto{O_{\text{clip}}} |i\rangle|j_s\rangle|X_i(j_s)\rangle\Big|\frac{\|w_s\|^2}{w_s(j_s)}\Big\rangle|u_{s+1}(i)\rangle \tag{A.6}$$

$$\xmapsto{O_{s,w}^{-1}} |i\rangle|j_s\rangle|X_i(j_s)\rangle|0\rangle|u_{s+1}(i)\rangle \tag{A.7}$$

$$\xmapsto{O_X^{-1}} |i\rangle|j_s\rangle|0\rangle|0\rangle|u_{s+1}(i)\rangle \tag{A.8}$$

$$\xmapsto{O_{s,j}^{-1}} |i\rangle|0\rangle|0\rangle|0\rangle|u_{s+1}(i)\rangle. \tag{A.9}$$

Because we have stored w_s and j_s, we could construct classical oracles $O_{s,j}(0) = j_s$, $O_{s,w}(j_s) = \frac{\|w_s\|^2}{w_s(j_s)}$ with $O(1)$ complexity. In the algorithm, we first call $O_{s,j}$ to compute j_s and store it into the second register in (A.3). In (A.4), we call the quantum oracle O_X for the value $X_i(j_s)$, which is stored into the third register. In (A.5), we call $O_{s,w}$ to compute $\frac{\|w_s\|^2}{w_s(j_s)}$ and store it into the fourth register. In (A.6), because we have $X_i(j_s)$ and $\frac{\|w_s\|^2}{w_s(j_s)}$ at hand, we could use $\tilde{O}(1)$ arithmetic computations to compute $\tilde{v}_s(i) = X_i(j_s)\|w_s\|^2/w_t(j_s)$ and

$$u_{s+1}(i) = u_s(i)\big(1 - \eta \operatorname{clip}(\tilde{v}_s(i), 1/\eta) + \eta^2 \operatorname{clip}(\tilde{v}_s(i), 1/\eta)^2\big). \tag{A.10}$$

We then store $u_{s+1}(i)$ into the fifth register. In (A.7), (A.8), and (A.9), we uncompute the steps in (A.5), (A.4), and (A.3), respectively (we need these steps in Algorithm 4 to keep its unitarity).

In total, between (A.3)-(A.9) we use 2 queries to O_X and $\tilde{O}(1)$ additional arithmetic computations. Because s goes from 1 to t, in total we use $2t$ queries to O_X and $\tilde{O}(t)$ additional arithmetic computations.

A.3. Correctness of Algorithm 1

Our proof follows that of Theorem 2.7 in Clarkson et al. (2012). In particular, we use the following five lemmas proved in Clarkson et al. (2012) for analyzing the online gradient gradient descent and ℓ_2 sampling outcomes:

Lemma A.1 (Lemma A.2 of Clarkson et al. (2012))**.** *The updates of w in Line 3 and y in Line 5 satisfy*

$$\max_{w \in \mathbb{B}_n} \sum_{t \in [T]} X_{i_t} w \leq \sum_{t \in [T]} X_{i_t} w_t + 2\sqrt{2T}. \tag{A.11}$$

Lemma A.2 (Lemma 2.3 of Clarkson et al. (2012))**.** *For any $t \in [T]$, denote p_t to be the unit vector in $\mathbb{R}^n$ such that*

$(p_t)_i = |\langle i|p_t\rangle|^2$ *for all* $i \in [n]$. *Then the update for* p_{t+1} *in Line 8 satisfies*

$$\sum_{t\in[T]} p_t^\top v_t \leq \min_{i\in[n]} \sum_{t\in[T]} v_t(i) + \eta \sum_{t\in[T]} p_t^\top v_t^2 + \frac{\log n}{\eta}, \tag{A.12}$$

where v_t^2 *is defined as* $(v_t^2)_i := (v_t)_i^2$ *for all* $i \in [n]$.

Lemma A.3 (Lemma 2.4 of Clarkson et al. (2012))**.** *With probability at least* $1 - O(1/n)$*,*

$$\max_{i\in[n]} \sum_{t\in[T]} \big[v_t(i) - X_i w_t\big] \leq 4\eta T. \tag{A.13}$$

Lemma A.4 (Lemma 2.5 of Clarkson et al. (2012))**.** *With probability at least* $1 - O(1/n)$*,*

$$\Big|\sum_{t\in[T]} X_{i_t} w_t - \sum_{t\in[T]} p_t^\top v_t\Big| \leq 10\eta T. \tag{A.14}$$

Lemma A.5 (Lemma 2.6 of Clarkson et al. (2012))**.** *With probability at least* $3/4$*,*

$$\sum_{t\in[T]} p_t^\top v_t^2 \leq 8T. \tag{A.15}$$

Proof. We first prove the correctness of Algorithm 1. By Lemma A.1, we have

$$\sum_{t\in[T]} X_{i_t} w_t \geq \max_{w\in\mathbb{B}_n} \sum_{t\in[T]} X_{i_t} w - 2\sqrt{2T} \geq T\sigma - 2\sqrt{2T}. \tag{A.16}$$

On the other hand, Lemma A.3 implies that for any $i \in [n]$,

$$\sum_{t\in[T]} X_i w_t \geq \sum_{t\in[T]} v_t(i) - 4\eta T. \tag{A.17}$$

Together with Lemma A.2, we have

$$\sum_{t\in[T]} p_t^\top v_t \leq \min_{i\in[n]} \sum_{t\in[T]} X_i w_t + \eta \sum_{t\in[T]} p_t^\top v_t^2 + \frac{\log n}{\eta} + 4\eta T. \tag{A.18}$$

Plugging Lemma A.4, Lemma A.5, and (A.16) into (A.18), with probability at least $\frac{3}{4} - 2\cdot O(\frac{1}{n}) \geq \frac{2}{3}$,

$$\min_{i\in[n]} \sum_{t\in[T]} X_i w_t \geq -\frac{\log n}{\eta} - 8\eta T - 4\eta T + T\sigma - 2\sqrt{2T} - 10\eta T \geq T\sigma - 22\eta T - \frac{\log n}{\eta}. \tag{A.19}$$

Since $T = 23^2\epsilon^{-2}\log n$ and $\eta = \sqrt{\frac{\log n}{T}}$, we have

$$\min_{i\in[n]} X_i \bar{w} = \frac{1}{T} \min_{i\in[n]} \sum_{t=1}^{T} X_i w_t \geq \sigma - 23\sqrt{\frac{\log n}{T}} \geq \sigma - \epsilon \tag{A.20}$$

with probability at least $2/3$, which is exactly (3.4). □

A.4. Linear classification in the PAC model

Theorem 3.1 could also be applied to the PAC model. For the case where there exists a hyperplane classifying all data points correctly with margin σ, and assume that the margin is not small in the sense that $\frac{1}{\sigma^2} < d$, PAC learning theory implies that the number of examples needed for training a classifier of error δ is $O(1/\sigma^2\delta)$. As a result, we have a quantum algorithm that computes a $\sigma/2$-approximation to the best classifier with cost

$$\tilde{O}\Big(\frac{\sqrt{1/\sigma^2\delta}}{\sigma^4} + \frac{d}{\sigma^2}\Big) = \tilde{O}\Big(\frac{1}{\sigma^5\sqrt{\delta}} + \frac{d}{\sigma^2}\Big). \tag{A.21}$$

This is better than the classical complexity $O(\frac{1}{\sigma^4\delta} + \frac{d}{\sigma^2})$ in Clarkson et al. (2012) as long as $\delta \leq \sigma^2$, which is plausible under the assumption that the margin σ is large.

A.5. Quantum algorithm for norm estimation

Lemma A.6 (Rewrite of Lemma 3.1)**.** *Assume that $F\colon [d] \to [0,1]$ with a quantum oracle $O_F|i\rangle|0\rangle = |i\rangle|F(i)\rangle$ for all $i \in [d]$. Denote $m = \frac{1}{d}\sum_{i=1}^{d} F(i)$. Then for any $\delta > 0$, there is a quantum algorithm that uses $O(\sqrt{d}/\delta)$ queries to O_F and returns an $\tilde{m}$ such that $|\tilde{m} - m| \leq \delta m$ with probability at least 2/3.*

Our proof of Lemma A.6 is based on amplitude estimation:

Theorem A.1 (Theorem 15 of Brassard et al. (2002))**.** *For any $0 < \epsilon < 1$ and Boolean function $f\colon [d] \to \{0,1\}$ with quantum oracle $O_f|i\rangle|0\rangle = |i\rangle|f(i)\rangle$ for all $i \in [d]$, there is a quantum algorithm that outputs an estimate $\hat{t}$ to $t = |f^{-1}(1)|$ such that*

$$|\hat{t} - t| \leq \epsilon t \tag{A.22}$$

with probability at least $8/\pi^2$, using $O(\frac{1}{\epsilon}\sqrt{\frac{d}{t}})$ evaluations of O_f. If $t = 0$, the algorithm outputs $\hat{t} = 0$ with certainty and O_f is evaluated $O(\sqrt{d})$ times.

Proof. Assume that $F(i)$ has l bits for precision for all $i \in [d]$ (in our paper, we take $l = O(1)$, say $l = 64$ for double float precision), and for all $k \in [l]$ denote $F_k(i)$ as the k^{th} bit of $F(i)$; denote $n_k = \sum_{i\in[d]} F_k(i)$.

We apply Theorem A.1 to all the l bits of n_k using $O(\sqrt{d}/\delta)$ queries (taking $\epsilon = \delta/2$), which gives an approximation $\hat{n}_k$ of n_k such that with probability at least $8/\pi^2$ we have $|n_k - \hat{n}_k| \leq \delta n_k/2$ if $n_k \geq 1$, and $\hat{n}_k = 0$ if $n_k = 0$. Running this procedure for $\Theta(\log l)$ times and take the median of all returned $\hat{n}_k$, and do this for all $k \in [l]$, Chernoff's bound promises that with probability $2/3$ we have

$$|n_k - \hat{n}_k| \leq \delta n_k \quad \forall\, k \in [l]. \tag{A.23}$$

As a result, if we take $\tilde{m} = \frac{1}{d}\sum_{k\in[l]} \frac{\hat{n}_k}{2^k}$, and observe that $m = \frac{1}{d}\sum_{k\in[l]} \frac{n_k}{2^k}$, we have

$$|\tilde{m} - m| \leq \frac{1}{d}\sum_{k\in[l]} \left|\frac{\hat{n}_k}{2^k} - \frac{n_k}{2^k}\right| \leq \frac{1}{d}\sum_{k\in[l]} \frac{\delta n_k}{2^k} = \delta m \tag{A.24}$$

with probability at least $2/3$. The total quantum query complexity is $O(l \log l \cdot \sqrt{d}/\delta) = O(\sqrt{d}/\delta)$. □

A.6. Proof of Theorem 3.2

In this subsection, we prove:

Theorem A.2 (Rewrite of Theorem 3.2)**.** *With success probability at least $2/3$, Algorithm 5 returns a succinct classical representation of a vector $\bar{w} \in \mathbb{R}^d$ such that*

$$X_i \bar{w} \geq \max_{w\in\mathbb{B}_d} \min_{i'\in[n]} X_{i'} w - \epsilon \quad \forall\, i \in [n], \tag{A.25}$$

using $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^4} + \frac{\sqrt{d}}{\epsilon^8}\big)$ quantum gates.

Proof. For clarification, we denote

$$\tilde{v}_{t,\text{approx}}(i) = \frac{X_i(j_t)\widetilde{\|y_t\|}^2}{y_t(j_t)\max\{1, \widetilde{\|y_t\|}\}}, \qquad \tilde{v}_{t,\text{true}}(i) = \frac{X_i(j_t)\|y_t\|^2}{y_t(j_t)\max\{1, \|y_t\|\}} \quad \forall\, i \in [n]. \tag{A.26}$$

In other words, the $\tilde{v}_t$ in Line 7 of Algorithm 5 is $\tilde{v}_{t,\text{approx}}$, an approximation of $\tilde{v}_{t,\text{true}}$. We prove:

$$|\tilde{v}_{t,\text{approx}}(i) - \tilde{v}_{t,\text{true}}(i)| \leq \eta \quad \forall\, i \in [n]. \tag{A.27}$$

[5] The meaning of the definition here is the same as Footnote 4 in Algorithm 1.

Algorithm 5: Quantum linear classification algorithm with $O(\sqrt{d})$ cost.

Input: $\epsilon > 0$, a quantum oracle O_X for $X \in \mathbb{R}^{n\times d}$.
Output: $\bar{w}$ that satisfies (3.4).

1 Let $T = 27^2\epsilon^{-2}\log n$, $y_1 = \mathbf{0}_d$, $\eta = \sqrt{\frac{\log n}{T}}$, $u_1 = \mathbf{1}_n$, $|p_1\rangle = \frac{1}{\sqrt{n}}\sum_{i\in[n]}|i\rangle$;
2 **for** $t = 1$ **to** T **do**
3 Measure the state $|p_t\rangle$ in the computational basis and denote the output as $i_t \in [n]$;
4 Define[5] $y_{t+1} := y_t + \frac{1}{\sqrt{2T}}X_{i_t}$;
5 Apply Lemma 3.1 for $2\lceil\log T\rceil$ times to estimate $\|y_t\|^2$ with precision $\delta = \eta^2$, and take the median of all the $2\lceil\log T\rceil$ outputs, denoted $\widetilde{\|y_t\|}^2$;
6 Choose $j_t \in [d]$ by $j_t = j$ with probability $y_t(j)^2/\|y_t\|^2$, which is achieved by applying Algorithm 3 to prepare the quantum state $|y_t\rangle$ and measure in the computational basis;
7 For all $i \in [n]$, denote $\tilde{v}_t(i) = X_i(j_t)\widetilde{\|y_t\|}^2/\big(y_t(j_t)\max\{1, \widetilde{\|y_t\|}\}\big)$, $v_t(i) = \mathrm{clip}(\tilde{v}_t(i), 1/\eta)$, and $u_{t+1}(i) = u_t(i)(1 - \eta v_t(i) + \eta^2 v_t(i)^2)$. Apply Algorithm 4 to prepare an oracle O_t such that $O_t|i\rangle|0\rangle = |i\rangle|u_{t+1}(i)\rangle$ for all $i \in [n]$, using $2t$ queries to O_X and $\tilde{O}(t)$ additional arithmetic computations;
8 Prepare the state $|p_{t+1}\rangle = \frac{1}{\|u_{t+1}\|_2}\sum_{i\in[n]}u_{t+1}(i)|i\rangle$ using Algorithm 3 and O_t;
9 Return $\bar{w} = \frac{1}{T}\sum_{t=1}^{T}\frac{y_t}{\max\{1, \widetilde{\|y_t\|}\}}$;

Without loss generality, we can assume that $\tilde{v}_{t,\text{true}}(i), \tilde{v}_{t,\text{approx}}(i) \leq 1/\eta$; otherwise, they are both truncated to $1/\eta$ by the clip function in Line 7 and no error occurs. For convenience, we denote $m = \|y_t\|^2$ and $\tilde{m} = \widetilde{\|y_t\|}^2$. Then

$$|\tilde{v}_{t,\text{approx}}(i) - \tilde{v}_{t,\text{true}}(i)| = \tilde{v}_{t,\text{true}}(i)\cdot\Big|\frac{\tilde{v}_{t,\text{approx}}(i)}{\tilde{v}_{t,\text{true}}(i)} - 1\Big| \leq \frac{1}{\eta}\cdot\Big|\frac{\tilde{v}_{t,\text{approx}}(i)}{\tilde{v}_{t,\text{true}}(i)} - 1\Big|. \tag{A.28}$$

When $\|y_t\| \geq 1$ we have $\frac{\tilde{v}_{t,\text{approx}}(i)}{\tilde{v}_{t,\text{true}}(i)} = \frac{\tilde{m}}{m}$; when $\|y_t\| \leq 1$ we have $\frac{\tilde{v}_{t,\text{approx}}(i)}{\tilde{v}_{t,\text{true}}(i)} = \sqrt{\frac{\tilde{m}}{m}}$. Because in Line 5 $\widetilde{\|y_t\|}^2$ is the median of $2\lceil\log T\rceil$ executions of Lemma 3.1, with failure probability at most $1 - (2/3)^{2\log T} = O(1/T^2)$ we have $|\frac{\tilde{m}}{m} - 1| \leq \delta$; given there are T iterations in total, the probability that Line 5 always succeeds is at least $1 - T\cdot O(1/T^2) = 1 - o(1)$, and we have

$$\Big|\frac{\tilde{m}}{m} - 1\Big|, \Big|\sqrt{\frac{\tilde{m}}{m}} - 1\Big| \leq \delta. \tag{A.29}$$

Plugging this into (A.28), we have

$$|\tilde{v}_{t,\text{approx}}(i) - \tilde{v}_{t,\text{true}}(i)| \leq \frac{\delta}{\eta} = \eta, \tag{A.30}$$

which proves (A.27).

Now we prove the correctness of Algorithm 5. By (A.27) and Lemma A.3, with probability at least $1 - O(1/n)$ we have

$$\max_{i\in[n]}\sum_{t\in[T]}\big[v_t(i) - X_i w_t\big] \leq 4\eta T + \eta T = 5\eta T, \tag{A.31}$$

where $w_t = \frac{y_t}{\max\{1, \widetilde{\|y_t\|}\}}$ for all $t \in [T]$. By (A.27) and Lemma A.4, with probability at least $1 - O(1/n)$ we have

$$\Big|\sum_{t\in[T]}X_{i_t}w_t - \sum_{t\in[T]}p_t^\top v_t\Big| \leq 10\eta T + \eta T = 11\eta T; \tag{A.32}$$

by (A.27) and Lemma A.5, with probability at least $3/4$ we have

$$\sum_{t\in[T]}p_t^\top v_t^2 \leq 8T + 2T = 10T. \tag{A.33}$$

As a result, similar to the proof of Theorem 3.1, we have

$$\min_{i\in[n]}\sum_{t\in[T]}X_iw_t\geq-\frac{\log n}{\eta}-10\eta T-5\eta T+T\sigma-2\sqrt{2T}-11\eta T\geq T\sigma-26\eta T-\frac{\log n}{\eta}. \tag{A.34}$$

Since $T=27^2\epsilon^{-2}\log n$ and $\eta=\sqrt{\frac{\log n}{T}}$, we have

$$\min_{i\in[n]}X_i\bar{w}=\frac{1}{T}\min_{i\in[n]}\sum_{t=1}^{T}X_iw_t\geq\sigma-27\sqrt{\frac{\log n}{T}}\geq\sigma-\epsilon \tag{A.35}$$

with probability at least $2/3$, which is exactly (A.25).

It remains to analyze the time complexity. Same as the proof of Theorem 3.1, the complexity in n is $\tilde{O}(\frac{\sqrt{n}}{\epsilon^4})$. It remains to show that the complexity in d is $\tilde{O}(\frac{\sqrt{n}}{\epsilon^8})$. The cost in d in Algorithm 1 and Algorithm 5 differs at Line 5 and Line 6. We first look at Line 5; because

$$y_t=\frac{1}{\sqrt{2T}}\sum_{\tau=1}^{T}X_{i_\tau}, \tag{A.36}$$

one query to a coefficient of y_t takes $t=\tilde{O}(1/\epsilon^2)$ queries to O_X. Next, since $X_i\in\mathbb{B}_n$ for all $i\in[n]$, we know that $X_{ij}\in[-1,1]$ for all $i\in[n]$, $j\in[d]$; to apply Lemma 3.1 (F should have image domain in $[0,1]$) we need to renormalize y_t by a factor of $t=\tilde{O}(1/\epsilon^2)$. In addition, notice that $\delta=\eta^2=\Theta(\epsilon^2)$; as a result, the query complexity of executing Lemma 3.1 is $\tilde{O}(\sqrt{d}/\epsilon^2)$. Finally, there are in total $T=\tilde{O}(1/\epsilon^2)$ iterations. Therefore, the total complexity in Line 5 is

$$\tilde{O}\Big(\frac{1}{\epsilon^2}\Big)\cdot\tilde{O}\Big(\frac{1}{\epsilon^2}\Big)\cdot\tilde{O}\Big(\frac{\sqrt{d}}{\epsilon^2}\Big)\cdot\tilde{O}\Big(\frac{1}{\epsilon^2}\Big)=\tilde{O}\Big(\frac{\sqrt{d}}{\epsilon^8}\Big). \tag{A.37}$$

Regarding the complexity in d in Line 6, the cost is to prepare the pure state $|y_t\rangle$ whose coefficient is proportional to y_t. To achieve this, we need $t=\tilde{O}(1/\epsilon^2)$ queries to O_X (for summing up the rows $X_{i_1},\ldots,X_{i_t}$) such that we have an oracle O_{y_t} satisfying $O_{y_t}|j\rangle|0\rangle=|j\rangle|y_t(j)\rangle$ for all $j\in[d]$. By Algorithm 3, the query complexity of preparing $|y_t\rangle$ using O_{y_t} is $O(\sqrt{d})$. Because there are in total $T=\tilde{O}(1/\epsilon^2)$ iterations, the total complexity in Line 6 is

$$\tilde{O}\Big(\frac{1}{\epsilon^2}\Big)\cdot O(\sqrt{d})\cdot\tilde{O}\Big(\frac{1}{\epsilon^2}\Big)=\tilde{O}\Big(\frac{\sqrt{d}}{\epsilon^4}\Big). \tag{A.38}$$

In all, the total complexity in d is $\tilde{O}(\sqrt{d}/\epsilon^8)$ as dominated by (A.37). Finally, $\bar{w}$ has a succinct classical representation: using $i_1,\ldots,i_T$ obtained from Line 3 and $\widetilde{\|y_1\|}^2,\ldots,\widetilde{\|y_T\|}^2$ obtained from Line 5, we could restore a coordinate of $\bar{w}$ in time $T=\tilde{O}(1/\epsilon^2)$. □

B. Proof of Quantum Lower Bounds

In this section, we prove the quantum lower bounds claimed in Section 5.

B.1. Linear classification

Recall that the input of the linear classification problem is a matrix $X\in\mathbb{R}^{n\times d}$ such that $X_i\in\mathbb{B}_d$ for all $i\in[n]$ (X_i being the i^{th} row of X), and the goal is to approximately solve

$$\sigma:=\max_{w\in\mathbb{B}_d}\min_{p\in\Delta_n}p^\top Xw=\max_{w\in\mathbb{B}_d}\min_{i\in[n]}X_iw. \tag{B.1}$$

Given the quantum oracle O_X such that $O_X|i\rangle|j\rangle|0\rangle=|i\rangle|j\rangle|X_{ij}\rangle\ \forall\, i\in[n],j\in[d]$, Theorem 3.2 solves this task with high success probability with cost $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^4}+\frac{\sqrt{d}}{\epsilon^8}\big)$. We prove a quantum lower bound that matches this upper bound in n and d for constant ϵ:

Theorem B.1. *Assume $0 < \epsilon < 0.04$. Then to return an $\bar{w} \in \mathbb{B}_d$ satisfying*

$$X_j\bar{w} \geq \max_{w\in\mathbb{B}_d}\min_{i\in[n]} X_i w - \epsilon \quad \forall\, j \in [n] \tag{B.2}$$

with probability at least $2/3$, we need $\Omega(\sqrt{n}+\sqrt{d})$ quantum queries to O_X.

Proof. Assume we are given the promise that X is from one of the two cases below:

1. There exists an $l \in \{2,\ldots,d\}$ such that $X_{11} = -\frac{1}{\sqrt{2}}$, $X_{1l} = \frac{1}{\sqrt{2}}$; $X_{21} = X_{2l} = \frac{1}{\sqrt{2}}$; there exists a unique $k \in \{3,\ldots,n\}$ such that $X_{k1} = 1$, $X_{kl} = 0$; $X_{ij} = \frac{1}{\sqrt{2}}$ for all $i \in \{3,\ldots,n\}/\{k\}$, $j \in \{1,l\}$, and $X_{ij} = 0$ for all $i \in [n]$, $j \notin \{1,l\}$.
2. There exists an $l \in \{2,\ldots,d\}$ such that $X_{11} = -\frac{1}{\sqrt{2}}$, $X_{1l} = \frac{1}{\sqrt{2}}$; $X_{21} = X_{2l} = \frac{1}{\sqrt{2}}$; $X_{ij} = \frac{1}{\sqrt{2}}$ for all $i \in \{3,\ldots,n\}$, $j \in \{1,l\}$, and $X_{ij} = 0$ for all $i \in [n]$, $j \notin \{1,l\}$.

Notice that the only difference between these two cases is a row where the first entry is 1 and the l^{th} entry is 0; they have the following pictures, respectively.

$$\text{Case 1: } X = \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \end{pmatrix}; \tag{B.3}$$

$$\text{Case 2: } X = \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{2}} & 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & 0 & \cdots & 0 \end{pmatrix}. \tag{B.4}$$

We denote the maximin value in (B.1) of these cases as σ_1 and σ_2, respectively. We have:

- $\sigma_2 = \frac{1}{\sqrt{2}}$.

On the one hand, consider $\bar{w} = \vec{e}_l \in \mathbb{B}_d$ (the vector in $\mathbb{R}^d$ with the l^{th} coordinate being 1 and all other coordinates being 0). Then $X_i\bar{w} = \frac{1}{\sqrt{2}}$ for all $i \in [n]$, and hence $\sigma_2 \geq \min_{i\in[n]} X_i\bar{w} = \frac{1}{\sqrt{2}}$. On the other hand, for any $w = (w_1,\ldots,w_d) \in \mathbb{B}_d$, we have

$$\min_{i\in[n]} X_i w = \min\Big\{-\frac{1}{\sqrt{2}}w_1 + \frac{1}{\sqrt{2}}w_l, \frac{1}{\sqrt{2}}w_1 + \frac{1}{\sqrt{2}}w_l\Big\} \leq \frac{1}{\sqrt{2}}w_l \leq \frac{1}{\sqrt{2}}, \tag{B.5}$$

where the first inequality comes from the fact that $\min\{a,b\} \leq \frac{a+b}{2}$ for all $X, b \in \mathbb{R}$ and the second inequality comes from the fact that $w \in \mathbb{B}_d$ and $|w_l| \leq 1$. As a result, $\sigma_2 = \max_{w\in\mathbb{B}_d}\min_{i\in[n]} X_i w \leq \frac{1}{\sqrt{2}}$. In conclusion, we have $\sigma_2 = \frac{1}{\sqrt{2}}$.

- $\sigma_1 = \frac{1}{\sqrt{4+2\sqrt{2}}}$.

On the one hand, consider $\bar{w} = \frac{1}{\sqrt{4+2\sqrt{2}}}\vec{e}_1 + \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}}\vec{e}_l \in \mathbb{B}_d$. Then

$$X_1\bar{w} = -\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{4+2\sqrt{2}}} + \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}} = \frac{1}{\sqrt{4+2\sqrt{2}}}; \tag{B.6}$$

$$X_i\bar{w} = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{4+2\sqrt{2}}} + \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}} = \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}} > \frac{1}{\sqrt{4+2\sqrt{2}}} \quad \forall\, i \in [n]/\{1,k\}; \tag{B.7}$$

$$X_k\bar{w} = 1 \cdot \frac{1}{\sqrt{4+2\sqrt{2}}} + 0 \cdot \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}} = \frac{1}{\sqrt{4+2\sqrt{2}}}. \tag{B.8}$$

In all, $\sigma_1 \geq \min_{i\in[n]} X_i\bar{w} = \frac{1}{\sqrt{4+2\sqrt{2}}}$.

On the other hand, for any $w = (w_1, \ldots, w_d) \in \mathbb{B}_d$, we have

$$\min_{i\in[n]} X_i w = \min\Big\{-\frac{1}{\sqrt{2}}w_1 + \frac{1}{\sqrt{2}}w_l, \frac{1}{\sqrt{2}}w_1 + \frac{1}{\sqrt{2}}w_l, w_1\Big\}. \tag{B.9}$$

If $w_1 \leq \frac{1}{\sqrt{4+2\sqrt{2}}}$, then (B.9) implies that $\min_{i\in[n]} X_i w \leq \frac{1}{\sqrt{4+2\sqrt{2}}}$; if $w_1 \geq \frac{1}{\sqrt{4+2\sqrt{2}}}$, then

$$w_l \leq \sqrt{1-w_1^2} = \sqrt{1-\frac{1}{4+2\sqrt{2}}} = \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}}, \tag{B.10}$$

and hence by (B.9) we have

$$\min_{i\in[n]} X_i w \leq -\frac{1}{\sqrt{2}}w_1 + \frac{1}{\sqrt{2}}w_l \leq -\frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{4+2\sqrt{2}}} + \frac{1}{\sqrt{2}} \cdot \frac{\sqrt{2}+1}{\sqrt{4+2\sqrt{2}}} = \frac{1}{\sqrt{4+2\sqrt{2}}}. \tag{B.11}$$

In all, we always have $\min_{i\in[n]} X_i w \leq \frac{1}{\sqrt{4+2\sqrt{2}}}$. As a result, $\sigma_1 = \max_{w\in\mathbb{B}_d} \min_{i\in[n]} X_i w \leq \frac{1}{\sqrt{4+2\sqrt{2}}}$. In conclusion, we have $\sigma_1 = \frac{1}{\sqrt{4+2\sqrt{2}}}$.

Now, we prove that an $\bar{w} \in \mathbb{B}_d$ satisfying (B.2) would simultaneously reveal whether X is from Case 1 or Case 2 as well as the value of $l \in \{2, \ldots, d\}$, by the following algorithm:

1. Check if one of $\bar{w}_2, \ldots, \bar{w}_d$ is larger than 0.94; if there exists an $l' \in \{2, \ldots, d\}$ such that $\bar{w}_{l'} > 0.94$, return 'Case 2' and $l = l'$;
2. Otherwise, return 'Case 1' and $l = \arg\max_{i\in\{2,\ldots,d\}} \bar{w}_i$.

We first prove that the classification of X (between Case 1 and Case 2) is correct. On the one hand, assume that X comes from Case 1. If we wrongly classified X as from Case 2, we would have $\bar{w}_{l'} > 0.94$ and $\bar{w}_1 < \sqrt{1-0.94^2} < 0.342$; this would imply

$$\min_{i\in[n]} X_i\bar{w} = \min\Big\{-\frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l, \frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l, \bar{w}_1\Big\} \leq \bar{w}_1 < \frac{1}{\sqrt{4+2\sqrt{2}}} - 0.04 \leq \sigma_1 - \epsilon \tag{B.12}$$

by $0.342 < \frac{1}{\sqrt{4+2\sqrt{2}}} - 0.04$, contradicts with (B.2). Therefore, for this case we must make correct classification that X comes from Case 1.

On the other hand, assume that X comes from Case 2. If we wrongly classified X as from Case 1, we would have $\bar{w}_l \leq \max_{i\in\{2,\ldots,d\}} \bar{w}_i \leq 0.94$; this would imply

$$\min_{i\in[n]} X_i\bar{w} = \min\Big\{-\frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l, \frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l\Big\} \leq \frac{1}{\sqrt{2}}\bar{w}_l < \frac{1}{\sqrt{2}} - 0.04 \leq \sigma_2 - \epsilon \tag{B.13}$$

by $\frac{0.94}{\sqrt{2}} < \frac{1}{\sqrt{2}} - 0.04$, contradicts with (B.2). Therefore, for this case we must make correct classification that X comes from Case 2. In all, our classification is always correct.

It remains to prove that the value of l is correct. If X is from Case 1, we have

$$\sigma_1 - \epsilon \leq \min_{i\in[n]} X_i\bar{w} = \min\Big\{ -\frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l, \frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l, \bar{w}_1\Big\}; \tag{B.14}$$

as a result, $\bar{w}_1 \geq \sigma_1 - \epsilon > 0.38 - 0.04 = 0.34$, and

$$-\frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l > 0.34 \quad\Longrightarrow\quad \bar{w}_l > 0.34\sqrt{2} + \bar{w}_1 > 0.34(\sqrt{2}+1) > 0.82. \tag{B.15}$$

Because $2 \cdot 0.82^2 > 1$, $\bar{w}_l$ must be the largest among $\bar{w}_2, \ldots, \bar{w}_d$ (otherwise $l' = \arg\max_{i\in\{2,\ldots,d\}} \bar{w}_i$ and $l \neq l'$ would imply $\|\bar{w}\|^2 = \sum_{i\in[d]} |\bar{w}_i|^2 \geq \bar{w}_l^2 + \bar{w}_{l'}^2 \geq 2\bar{w}_l^2 > 1$, contradiction). Therefore, Line 2 of our algorithm correctly returns the value of l.

If X is from Case 2, we have

$$\sigma_2 - \epsilon \leq \min_{i\in[n]} X_i\bar{w} = \min\Big\{ -\frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l, \frac{1}{\sqrt{2}}\bar{w}_1 + \frac{1}{\sqrt{2}}\bar{w}_l\Big\} \leq \frac{1}{\sqrt{2}}\bar{w}_l, \tag{B.16}$$

and hence $\bar{w}_l \geq \sqrt{2}(\sigma_2 - \epsilon) \geq \sqrt{2}(\frac{1}{\sqrt{2}} - 0.04) > 0.94$. Because $2 \cdot 0.94^2 > 1$, only one coordinate of $\bar{w}$ could be at least 0.94 and we must have $l = l'$. Therefore, Line 1 of our algorithm correctly returns the value of l.

In all, we have proved that an ϵ-approximate solution $\bar{w} \in \mathbb{B}_d$ for (B.2) would simultaneously reveal whether X is from Case 1 or Case 2 as well as the value of $l \in \{2, \ldots, d\}$. On the one hand, notice that distinguishing these two cases requires $\Omega(\sqrt{n-2}) = \Omega(\sqrt{n})$ quantum queries to O_X for searching the position of k because of the quantum lower bound for search (Bennett et al., 1997); therefore, it gives an $\Omega(\sqrt{n})$ quantum lower bound on queries to O_X for returning an $\bar{w}$ that satisfies (B.2). On the other hand, finding the value of l is also a search problem on the entries of X, which requires $\Omega(\sqrt{d-1}) = \Omega(\sqrt{d})$ quantum queries to O_X also due to the quantum lower bound for search (Bennett et al., 1997). These observations complete the proof of Theorem B.1. □

B.2. Minimum enclosing ball (MEB)

Similarly, the input of the MEB problem is a matrix $X \in \mathbb{R}^{n\times d}$ such that $X_i \in \mathbb{B}_d$ for all $i \in [n]$, and we are given the quantum oracle O_X such that $O_X|i\rangle|j\rangle|0\rangle = |i\rangle|j\rangle|X_{ij}\rangle\ \forall\, i \in [n], j \in [d]$. The goal of MEB is to approximately solve

$$\sigma_{\text{MEB}} = \min_{w\in\mathbb{R}^d} \max_{i\in[n]} \|w - X_i\|^2. \tag{B.17}$$

Theorem 4.2 solves this task with high success probability with cost $\tilde{O}\big(\frac{\sqrt{n}}{\epsilon^4} + \frac{\sqrt{d}}{\epsilon^7}\big)$. In this subsection, we prove a quantum lower bound that matches this upper bound in n and d for constant ϵ:

Theorem B.2. *Assume* $0 < \epsilon < 0.01$. *Then to return an* $\bar{w} \in \mathbb{R}_d$ *satisfying*

$$\max_{i\in[n]} \|\bar{w} - X_i\|^2 \leq \min_{w\in\mathbb{R}^d} \max_{i\in[n]} \|w - X_i\|^2 + \epsilon \tag{B.18}$$

with probability at least $2/3$, *we need* $\Omega(\sqrt{n} + \sqrt{d})$ *quantum queries to* O_X.

Proof. We also assume that X is from one of the two cases in Theorem B.1; see also (B.3) and (B.4). We denote the maximin value in (B.17) of these cases as $\sigma_{\text{MEB},1}$ and $\sigma_{\text{MEB},2}$, respectively. We have:

- $\sigma_{\text{MEB},2} = \frac{1}{2}$.

On the one hand, consider $\bar{w} = \frac{1}{\sqrt{2}}\vec{e}_l$. Then

$$\|\bar{w} - X_1\|^2 = \Big(w_1 + \frac{1}{\sqrt{2}}\Big)^2 + \Big(w_l - \frac{1}{\sqrt{2}}\Big)^2 + \sum_{i\neq 1,l} w_i^2 = \Big(\frac{1}{\sqrt{2}}\Big)^2 = \frac{1}{2}; \tag{B.19}$$

$$\|\bar{w} - X_i\|^2 = \Big(w_1 - \frac{1}{\sqrt{2}}\Big)^2 + \Big(w_l - \frac{1}{\sqrt{2}}\Big)^2 + \sum_{i\neq 1,l} w_i^2 = \Big(\frac{1}{\sqrt{2}}\Big)^2 = \frac{1}{2} \quad \forall\, i \in \{2, \ldots, n\}. \tag{B.20}$$

Therefore, $\|\bar{w}-X_i\|^2=\frac{1}{2}$ for all $i\in[n]$, and hence $\sigma_{\text{MEB},2}\leq\max_{i\in[n]}\|\bar{w}-X_i\|^2=\frac{1}{2}$.

On the other hand, for any $w=(w_1,\ldots,w_d)\in\mathbb{R}_d$, we have

$$
\begin{aligned}
&\max_{i\in[n]}\|w-X_i\|^2\\
&=\max\Big\{\Big(w_1-\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2+\sum_{i\neq 1,l}w_i^2,\Big(w_1+\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2+\sum_{i\neq 1,l}w_i^2\Big\} && \text{(B.21)}\\
&\geq\frac{1}{2}\Big[\Big(w_1-\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2\Big]+\frac{1}{2}\Big[\Big(w_1+\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2\Big]+\sum_{i\neq 1,l}w_i^2 && \text{(B.22)}\\
&=w_1^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2+\sum_{i\neq 1,l}w_i^2+\frac{1}{2} && \text{(B.23)}\\
&\geq\frac{1}{2}, && \text{(B.24)}
\end{aligned}
$$

where (B.22) comes from the fact that $\max\{a,b\}\geq\frac{1}{2}(a+b)$ and $\sum_{i\neq 1,l}w_i^2\geq 0$. Therefore, $\sigma_{\text{MEB},2}\geq\frac{1}{2}$. In all, we must have $\sigma_{\text{MEB},2}=\frac{1}{2}$.

- $\sigma_{\text{MEB},1}=\frac{2+\sqrt{2}}{4}$.

On the one hand, consider $\bar{w}=\big(\frac{1}{2}-\frac{\sqrt{2}}{4}\big)\vec{e}_1+\frac{\sqrt{2}}{4}\vec{e}_l$. Then

$$
\begin{aligned}
\|\bar{w}-X_1\|^2&=\Big(w_1+\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2+\sum_{i\neq 1,l}w_i^2=\Big(\frac{1}{2}+\frac{\sqrt{2}}{4}\Big)^2+\Big(\frac{\sqrt{2}}{4}\Big)^2=\frac{2+\sqrt{2}}{4}; && \text{(B.25)}\\
\|\bar{w}-X_k\|^2&=(w_1-1)^2+w_l^2+\sum_{i\neq 1,l}w_i^2=\Big(\frac{1}{2}+\frac{\sqrt{2}}{4}\Big)^2+\Big(\frac{\sqrt{2}}{4}\Big)^2=\frac{2+\sqrt{2}}{4}; && \text{(B.26)}\\
\|\bar{w}-X_i\|^2&=\Big(w_1-\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2+\sum_{i\neq 1,l}w_i^2=\frac{6-3\sqrt{2}}{4}<\frac{2+\sqrt{2}}{4}\quad\forall\, i\in[n]/\{1,k\}. && \text{(B.27)}
\end{aligned}
$$

In all, $\sigma_{\text{MEB},1}\leq\max_{i\in[n]}\|\bar{w}-X_i\|^2=\frac{2+\sqrt{2}}{4}$.

On the other hand, for any $w=(w_1,\ldots,w_d)\in\mathbb{R}_d$, we have

$$
\begin{aligned}
\max_{i\in[n]}\|w-X_i\|^2&\geq\max\Big\{\Big(w_1+\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2+\sum_{i\neq 1,l}w_i^2,(w_1-1)^2+w_l^2+\sum_{i\neq 1,l}w_i^2\Big\} && \text{(B.28)}\\
&\geq\frac{1}{2}\Big[\Big(w_1+\frac{1}{\sqrt{2}}\Big)^2+\Big(w_l-\frac{1}{\sqrt{2}}\Big)^2\Big]+\frac{1}{2}\Big[(w_1-1)^2+w_l^2\Big]+\sum_{i\neq 1,l}w_i^2 && \text{(B.29)}\\
&=\Big[w_1-\Big(\frac{1}{2}-\frac{\sqrt{2}}{4}\Big)\Big]^2+\Big(w_l-\frac{\sqrt{2}}{4}\Big)^2+\sum_{i\neq 1,l}w_i^2+\frac{2+\sqrt{2}}{4} && \text{(B.30)}\\
&\geq\frac{2+\sqrt{2}}{4}. && \text{(B.31)}
\end{aligned}
$$

Therefore, $\sigma_{\text{MEB},2}\geq\frac{2+\sqrt{2}}{4}$. In all, we must have $\sigma_{\text{MEB},2}=\frac{2+\sqrt{2}}{4}$.

Now, we prove that an $\bar{w}\in\mathbb{R}_d$ satisfying (B.18) would simultaneously reveal whether X is from Case 1 or Case 2 as well as the value of $l\in\{2,\ldots,d\}$, by the following algorithm:

1. Check if one of $\bar{w}_2,\ldots,\bar{w}_d$ is larger than $\frac{3\sqrt{2}}{8}$; if there exists an $l'\in\{2,\ldots,d\}$ such that $\bar{w}_{l'}>\frac{3\sqrt{2}}{8}$, return 'Case 1' and $l=l'$;

2. Otherwise, return 'Case 2' and $l = \arg\max_{i\in\{2,\ldots,d\}} \bar{w}_i$.

We first prove that the classification of X (between Case 1 and Case 2) is correct. On the one hand, assume that X comes from Case 1. If we wrongly classified X as from Case 2, we would have $\bar{w}_l \leq \max_{i\in\{2,\ldots,d\}} \bar{w}_i \leq \frac{3\sqrt{2}}{8}$. By (B.23), this would imply

$$\max_{i\in[n]} \|\bar{w} - X_i\|^2 \geq \Big(\bar{w}_l - \frac{1}{\sqrt{2}}\Big)^2 + \frac{1}{2} \geq \frac{1}{32} + \frac{1}{2} > \sigma_{\mathrm{MEB},1} + \epsilon, \tag{B.32}$$

contradicts with (B.18). Therefore, for this case we must make correct classification that X comes from Case 2.

On the other hand, assume that X comes from Case 2. If we wrongly classified X as from Case 1, we would have $\bar{w}_{l'} > \frac{3\sqrt{2}}{8}$. If $l = l'$, then (B.30) implies that

$$\max_{i\in[n]} \|\bar{w} - X_i\|^2 \geq \Big(\bar{w}_l - \frac{\sqrt{2}}{4}\Big)^2 + \frac{2+\sqrt{2}}{4} \geq \frac{1}{32} + \frac{2+\sqrt{2}}{4} > \sigma_{\mathrm{MEB},2} + \epsilon, \tag{B.33}$$

contradicts with (B.18). If $l \neq l'$, then (B.30) implies that

$$\max_{i\in[n]} \|\bar{w} - X_i\|^2 \geq \bar{w}_{l'}^2 + \frac{2+\sqrt{2}}{4} \geq \frac{9}{32} + \frac{2+\sqrt{2}}{4} > \sigma_{\mathrm{MEB},2} + \epsilon, \tag{B.34}$$

also contradicts with (B.18). Therefore, for this case we must make correct classification that X comes from Case 1.

In all, our classification is always correct. It remains to prove that the value of l is correct. If X is from Case 1, by (B.23) we have

$$\frac{1}{2} + 0.01 \geq \max_{i\in[n]} \|\bar{w} - X_i\|^2 \geq \bar{w}_1^2 + \Big(\bar{w}_l - \frac{1}{\sqrt{2}}\Big)^2 + \sum_{i\neq 1,l} \bar{w}_i^2 + \frac{1}{2}. \tag{B.35}$$

As a result, $\bar{w}_i \leq 0.1 < \frac{3\sqrt{2}}{8}$ for all $i \in [n]/\{l\}$ and $\bar{w}_l \geq \frac{1}{\sqrt{2}} - 0.1 > \frac{3\sqrt{2}}{8}$. Therefore, we must have $l = l'$, i.e., Line 1 of our algorithm correctly returns the value of l.

If X is from Case 2, by (B.30) we have

$$\frac{2+\sqrt{2}}{4} + 0.01 \geq \max_{i\in[n]} \|\bar{w} - X_i\|^2 \geq \Big[\bar{w}_1 - \Big(\frac{1}{2} - \frac{\sqrt{2}}{4}\Big)\Big]^2 + \Big(\bar{w}_l - \frac{\sqrt{2}}{4}\Big)^2 + \sum_{i\neq 1,l} \bar{w}_i^2 + \frac{2+\sqrt{2}}{4}. \tag{B.36}$$

As a result, $\bar{w}_i \leq 0.1 < 0.25$ for all $i \in [n]/\{1, l\}$, $\bar{w}_1 \leq \frac{1}{2} - \frac{\sqrt{2}}{4} + 0.1 < 0.25$, and $\bar{w}_l \geq \frac{\sqrt{2}}{4} - 0.1 > 0.25$. Therefore, we must have $\bar{w}_l = \max_{i\in\{2,\ldots,d\}} \bar{w}_i$, i.e., Line 1 of our algorithm correctly returns the value of l.

In all, we have proved that an ϵ-approximate solution $\bar{w} \in \mathbb{R}_d$ for (B.18) would simultaneously reveal whether X is from Case 1 or Case 2 as well as the value of $l \in \{2, \ldots, d\}$. On the one hand, notice that distinguishing these two cases requires $\Omega(\sqrt{n-2}) = \Omega(\sqrt{n})$ quantum queries to O_X for searching the position of k because of the quantum lower bound for search (Bennett et al., 1997); therefore, it gives an $\Omega(\sqrt{n})$ quantum lower bound on queries to O_X for returning an $\bar{w}$ that satisfies (B.18). On the other hand, finding the value of l is also a search problem on the entries of X, which requires $\Omega(\sqrt{d-1}) = \Omega(\sqrt{d})$ quantum queries to O_X also due to the quantum lower bound for search (Bennett et al., 1997). These observations complete the proof of Theorem B.2. □