A. Detailed Choices of Reserved Constants

The absolute constants c_0, c_1 and c_2 are specified in Lemma 24, and c_3 and c_4 are specified in Lemma 25. c_5 and c_6 are clarified in Section 3.1.1. The definition of c_7 and c_8 can be found in Lemma 26 and Lemma 27 respectively. The absolute constant C_1 acts as an upper bound of all b_k's, and by our choice in Section 3.1.1, $C_1 = \bar{c}/16$. The absolute constant C_2 is defined in Lemma 3. Other absolute constants, such as C_3, C_4 are not quite crucial to our analysis or algorithmic design. Therefore, we do not track their definitions. The subscript variants of K, e.g. K_1 and K_2, are also absolute constants but their values may change from appearance to appearance. We remark that the value of all these constants does not depend on the underlying distribution D chosen by the adversary, but rather depends on the knowledge that D is a member of the family of isotropic log-concave distributions.

B. Omitted Proofs in Section 2

We will frequently use the well-known Chernoff bound in our analysis. For convenience, we record it below.

Lemma 17 (Chernoff bound)**.** *Let $Z_1, Z_2, \dots, Z_n$ be n independent random variables that take value in $\{0, 1\}$. Let $Z = \sum_{i=1}^{n} Z_i$. For each Z_i, suppose that $\Pr(Z_i = 1) \leq \eta$. Then for any $\alpha \in [0, 1]$*

$$\Pr\big(Z \geq (1+\alpha)\eta n\big) \leq e^{-\frac{\alpha^2 \eta n}{3}}.$$

When $\Pr(Z_i = 1) \geq \eta$, for any $\alpha \in [0, 1]$

$$\Pr\big(Z \leq (1-\alpha)\eta n\big) \leq e^{-\frac{\alpha^2 \eta n}{2}}.$$

B.1. Proof of Lemma 4

Proof. We note that $(\rho^+ - \rho^-)^2 \leq 4(\rho^+)^2$. In addition, this inequality is almost tight up to a constant factor since ρ^- can be as small as 0. To see this, observe that $u \in W$ and x is such that $\left|u \cdot x\right| \leq b$.

Thus, it remains to upper bound ρ^+. Due to localized sampling, for any $w \in W$ we have

$$\left|w \cdot x\right| \leq \left|(w-u) \cdot x\right| + \left|u \cdot x\right| \leq \left\|w-u\right\|_2 \cdot \left\|x\right\|_2 + b \leq r \cdot c_7 \sqrt{d} \log\frac{1}{b\delta} + b, \tag{5}$$

where the first step follows from the triangle inequality, the second step uses Cauchy-Schwarz inequality and the fact $x \sim D_{u,b}$, and the last step applies Lemma 26. The lemma follows by noting that $r = \Theta(b)$. □

B.2. Proof of Lemma 7

Proof. For any unit vector v, observe that $w := rv + u$ is such that $\left\|w-u\right\|_2 \leq r$. Hence,

$$\begin{aligned}\mathbb{E}\left[(v\cdot x)^2\right] &= \frac{1}{r^2}\,\mathbb{E}\left[(r\cdot v\cdot x)^2\right]\\ &\leq \frac{2}{r^2}\,\mathbb{E}\left[((r\cdot v+u)\cdot x)^2\right] + \frac{2}{r^2}\,\mathbb{E}\left[(u\cdot x)^2\right]\\ &\leq \frac{2}{r^2}\cdot C_2(b^2+r^2) + \frac{2}{r^2}\cdot b^2\\ &\leq \frac{4C_2(b^2+r^2)}{r^2},\end{aligned}$$

where in the second step we use the basic inequality $a_1^2 \leq 2(a_1 - a_2)^2 + 2a_2^2$, and in the third step we apply Lemma 3. This proves the first desired inequality.

Next, by Lemma 26 we have with probability $1-\delta$, $\left\|x\right\|_2 \leq c_7\sqrt{d}\log\frac{1}{b\delta}$. Then for any unit vector v, we have

$$(v\cdot x)^2 \leq \left\|v\right\|_2^2 \cdot \left\|x\right\|_2^2 \leq c_7^2 \cdot d\log^2\frac{1}{b\delta},$$

which implies the second desired inequality. □

B.3. Proof of Proposition 8

Proof. In Lemma 6, we set $\alpha = 1$, $M_i = x_i x_i^\top$ where x_i is the i-th instance in the set T_{C}. Lemma 7 implies that $\mu_{\max} \leq \frac{4C_2(b^2+r^2)}{r^2} |T_{\mathrm{C}}| \leq K \cdot |T_{\mathrm{C}}|$ for some constant $K > 0$ since $r = \Theta(b)$, and with probability $1-\delta$, $\Lambda \leq K_1 \cdot d \log^2 \frac{|T_{\mathrm{C}}|}{b\delta}$ by union bound. By conditioning on these events and putting all pieces together, Lemma 6 asserts that with probability $1 - d \cdot (\frac{e}{4})^{\frac{K}{K_1} \cdot \frac{|T_{\mathrm{C}}|}{d \log^2 \frac{|T_{\mathrm{C}}|}{b\delta}}}$,

$$\lambda_{\max}\left(\sum_{x \in S} x x^\top\right) \leq 2K \cdot |T_{\mathrm{C}}| . \tag{6}$$

Equivalently, the above holds with probability $1 - \delta$ as long as $|T_{\mathrm{C}}| \geq K_2 d \log^2 \frac{|T_{\mathrm{C}}|}{b\delta} \cdot \log \frac{d}{\delta}$ for some constant $K_2 > 0$. □

B.4. Proof of Lemma 9

Proof. By Lemma 27

$$\mathrm{Pr}_{x \sim D}(x \in X) \geq c_8 b.$$

This implies that

$$\begin{aligned} &\mathrm{Pr}_{x \sim \mathrm{EX}_\eta^x(D,w^*)}(x \in X_{u,b} \text{ and } x \text{ is clean}) \\ =\ &\mathrm{Pr}_{x \sim \mathrm{EX}_\eta^x(D,w^*)}(x \in X_{u,b} \mid x \text{ is clean}) \cdot \mathrm{Pr}_{x \sim \mathrm{EX}_\eta^x(D,w^*)}(x \text{ is clean}) \geq c_8 b(1-\eta). \end{aligned}$$

We want to ensure that by drawing N instances from $\mathrm{EX}_\eta^x(D, w^*)$, with probability at least $1 - \delta$, n out of them fall into the band $X_{u,b}$. We apply the second inequality of Lemma 17 by letting $Z_i = \mathbf{1}_{\{x_i \in X_{u,b} \text{ and } x_i \text{ is clean}\}}$ and $\alpha = 1/2$, and obtain

$$\mathrm{Pr}\left(|T_{\mathrm{C}}| \leq \frac{c_8 b(1-\eta)}{2} N\right) \leq \exp\left(-\frac{c_8 b(1-\eta)N}{8}\right),$$

where the probability is taken over the event that we make a number of N calls to $\mathrm{EX}_\eta^x(D, w^*)$. Thus, when $N \geq \frac{8}{c_8 b(1-\eta)}\left(n + \ln \frac{1}{\delta}\right)$, we are guaranteed that at least n samples from $\mathrm{EX}_\eta^x(D, w^*)$ fall into the band $X_{u,b}$ with probability $1 - \delta$. The lemma follows by observing $\eta < \frac{1}{2}$. □

B.5. Proof of Lemma 10

This is a simplified version of Lemma 30 of Shen & Zhang (2021).

Proof. We calculate the noise rate within the band $X_k := \{x : |w_{k-1} \cdot x| \leq b_k\}$ by Lemma 18:

$$\mathrm{Pr}_{x \sim \mathrm{EX}_\eta^x(D,w^*)}(x \text{ is dirty} \mid x \in X_{u,b}) \leq \frac{2\eta}{c_8 b} \leq \frac{2\eta}{c_8 \epsilon} \leq \frac{2c_5}{c_8} \leq \frac{1}{8},$$

where the second inequality applies the setting $b \geq \epsilon$, the third inequality is due to the condition $\eta \leq c_5 \epsilon$, and the last inequality is due to the condition that c_5 is assumed to be a sufficiently small constant. Now we apply the first inequality of Lemma 17 by specifying $Z_i = \mathbf{1}_{\{x_i \text{ is dirty}\}}$, $\alpha = 1$ therein, which gives

$$\mathrm{Pr}\left(|T_{\mathrm{D}}| \geq \frac{1}{4}|T|\right) \leq \exp\left(-\frac{|T|}{24}\right),$$

where the probability is taken over the draw of T. The lemma follows by setting the right-hand side to δ and noting that $|T_{\mathrm{C}}| = |T| - |T_{\mathrm{D}}|$. □

Lemma 18. *Assume $\eta < \frac{1}{2}$. We have*

$$\mathrm{Pr}_{x \sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \text{ is dirty} \mid x \in X_{u,b}\right) \leq \frac{2\eta}{c_8 b}$$

where c_8 was defined in Lemma 27.

Proof. For an instance x, we use $\mathrm{tag}_x = 1$ to denote that x is drawn from D, and use $\mathrm{tag}_x = -1$ to denote that x is adversarially generated.

We first calculate the probability that an instance returned by $\mathrm{EX}_\eta^x(D, w^*)$ falls into the band $X_{u,b}$ as follows:

$$
\begin{aligned}
&\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \in X_{u,b}\right)\\
=\ &\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \in X_{u,b} \text{ and } \mathrm{tag}_x = 1\right) + \mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \in X_{u,b} \text{ and } \mathrm{tag}_x = -1\right)\\
\geq\ &\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \in X_{u,b} \text{ and } \mathrm{tag}_x = 1\right)\\
=\ &\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \in X_{u,b} \mid \mathrm{tag}_x = 1\right) \cdot \mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(\mathrm{tag}_x = 1\right)\\
=\ &\mathrm{Pr}_{x\sim D}\left(x \in X_{u,b}\right) \cdot \mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(\mathrm{tag}_x = 1\right)\\
\overset{\zeta}{\geq}\ &c_8 b \cdot (1-\eta)\\
\geq\ &\frac{1}{2} c_8 b,
\end{aligned}
$$

where in the inequality ζ we applied Part 1 of Lemma 27. It is thus easy to see that

$$
\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(\mathrm{tag}_x = -1 \mid x \in X_{u,b}\right) \leq \frac{\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(\mathrm{tag}_x = -1\right)}{\mathrm{Pr}_{x\sim \mathrm{EX}_\eta^x(D,w^*)}\left(x \in X_{u,b}\right)} \leq \frac{2\eta}{c_8 b},
$$

which is the desired result. □

B.6. Rademacher analysis leads to suboptimal sample complexity for quadratic functions

To see why a general Rademacher analysis may not suffice, we can, for example, think of the quadratic function $(w \cdot x)^2$ as a composition of the functions $\phi(f) = f^2$ and $f_w(x) = w \cdot x$. Recall that we showed with high probability that $|w \cdot x| \leq O(b\sqrt{d})$ (omitting logarithmic factors for convenience). Now, the gradient of $\phi(\cdot)$ is $2w \cdot x$ which is upper bounded by $O(b\sqrt{d})$ and the function value of $\phi(\cdot)$ is upper bounded by $O(b^2 d)$. For the Rademacher complexity $\mathcal{R}_\mathcal{F}$ of the class of linear functions $\mathcal{F} := \{f_w(x) = w \cdot x : w \in W\}$ on $T_\mathrm{C} = \{x_1, \dots, x_n\}$, let $V = \{v \in \mathbb{R}^d : \|v\|_2 \leq 1$ and note that for any $w \in W$, $w = u + rv$. We have by definition

$$
\begin{aligned}
\mathcal{R}_\mathcal{F} &= \frac{1}{n} \mathbb{E} \sup_{w\in W} \sum_{i=1}^n \sigma_i (w \cdot x_i)\\
&= \frac{1}{n} \mathbb{E} \sup_{w\in W} w \cdot \sum_{i=1}^n \sigma_i x_i\\
&\leq \frac{r}{n} \mathbb{E} \sup_{v\in V} v \cdot \sum_{i=1}^n \sigma_i x_i + \frac{1}{n} \mathbb{E}\, u \cdot \sum_{i=1}^n \sigma_i x_i\\
&\leq \frac{r}{n} \mathbb{E} \left\| \sum_{i=1}^n \sigma_i x_i \right\|_2\\
&\leq \frac{r}{n} \cdot \sqrt{n} \max_{1\leq i\leq n} \|x_i\|_2,
\end{aligned}
$$

where the expectation is taken over the i.i.d. Rademacher variables $\sigma_1, \dots, \sigma_n$. By Lemma 26, $\mathcal{R}_\mathcal{F} \leq \frac{r}{\sqrt{n}}\sqrt{d}$ with high probability. By the contraction lemma, the Rademacher complexity of the class of quadratic functions is $O(\frac{brd}{\sqrt{n}})$, and thus uniform concentration through Rademacher analysis requires $O(d^2)$ samples.

Similarly, a straightforward application of local Rademacher analysis (Bartlett et al., 2005) may not suffice as well. However, our discussion here does not rule out the possibility that a more sophisticated exploration of these techniques would lead to the desired sample complexity bound; we leave it as an open problem.

C. Omitted Proofs in Section 3

We present a full proof of the results in Section 3. Observe that the malicious noise is a special case of the nasty noise; hence this section can also be thought of as providing a complete proof for the results in Section 2.

To improve the transparency, we collect useful notations in Table 1.

Table 1. Summary of useful notations associated with the working set T at each phase k for learning with nasty noise.

$\hat{A}'$	labeled clean instance set obtained by drawing N instances from D and labeling them by w^*
A'	(clean) instance set obtained by hiding all the labels in $\hat{A}'$
$\hat{A}$	labeled corrupted instance set obtained by replacing ηN samples in $\hat{A}'$
A	(corrupted) instance set obtained by hiding all the labels in $\hat{A}$
A_{C}	set of clean instances in A
A_{D}	set of dirty instances in A, i.e. $A \backslash A_{\mathrm{C}}$
A_{E}	set of clean instances erased from A' by the adversary
T	set of instances in A that satisfy $\lvert w_{k-1} \cdot x\rvert \leq b_k$
T_{C}	set of clean instances in T
T_{D}	set of dirty instances in T, i.e. $T \backslash T_{\mathrm{C}}$
$\hat{T}_{\mathrm{C}}$	unrevealed labeled set of T_{C}
$\hat{T}_{\mathrm{E}}$	unrevealed labeled set of T_{E}

C.1. Proof of Lemma 11

Proof. Since $\eta \leq c_5 \epsilon$ and $b \geq \epsilon$, we have $\eta \leq c_5 b \leq \frac{1}{2} c_8 \xi b$ where the second inequality follows from the fact that c_5 is a small constant and $\xi \geq \Omega(1)$. Thus $\lvert A_{\mathrm{D}}\rvert = \eta N \leq \frac{1}{2} c_8 \xi b N$ and $\lvert A_{\mathrm{C}}\rvert = N - \lvert A_{\mathrm{D}}\rvert \geq (1 - \frac{1}{2} c_8 \xi b) N$. □

C.2. Proof of Lemma 12

Proof. We first show that the following two events hold simultaneously with probability $1 - \frac{\delta_k}{24}$:

$$E_1 : \lvert A_{\mathrm{C}}\rvert \geq \left(1 - \frac{1}{2} c_8 \xi b\right) N \text{ and } \lvert A_{\mathrm{D}}\rvert \leq \frac{1}{2} c_8 \xi b N,$$
$$E_2 : \lvert T_{\mathrm{C}}\rvert \geq \frac{1}{2} c_8 (1 - \xi) b N \text{ and } \lvert T_{\mathrm{E}}\rvert \leq \frac{1}{2} c_8 \xi b N.$$

Observe that E_1 holds with certainty due to Lemma 11.

To see why E_2 holds with high probability, we recall that Part 1 of Lemma 27 shows that $\Pr_{x \sim D}\big(x \in X_{u,b}\big) \geq c_8 b$. For each $x_i \in A_{\mathrm{C}} \cup A_{\mathrm{E}}$, define $Z_i = \mathbf{1}_{\{x_i \in X_{u,b}\}}$. Since $A_{\mathrm{C}} \cup A_{\mathrm{E}}$ are i.i.d. draws from D, by applying the second part of Lemma 17 with $\alpha = 1/2$, we have

$$\Pr\left(\sum_{i=1}^{N} Z_i \leq \frac{1}{2} c_8 b N\right) \leq \exp\left(-\frac{c_8 b N}{8}\right).$$

This shows that

$$\lvert T_{\mathrm{C}}\rvert + \lvert T_{\mathrm{E}}\rvert \geq \frac{1}{2} c_8 b N$$

with probability $1 - \delta$ provided that $N \geq \frac{8}{c_8 b} \ln \frac{1}{\delta}$. On the other side, we have $\lvert T_{\mathrm{E}}\rvert \leq \lvert A_{\mathrm{E}}\rvert = \lvert A_{\mathrm{D}}\rvert \leq \frac{1}{2} c_8 \xi b N$. Thus it follows that $\lvert T_{\mathrm{C}}\rvert \geq \frac{1}{2} c_8 (1 - \xi) b N$.

For Part 1, we have

$$\frac{\lvert T_{\mathrm{C}}\rvert}{\lvert T_{\mathrm{D}}\rvert} \geq \frac{1 - \xi}{\xi}, \tag{7}$$

where the inequality follows from E_2 and the fact $|T_{\mathrm{D}}| = |T_{\mathrm{E}}|$. Therefore,

$$\frac{|T_{\mathrm{D}}|}{|T|} = \frac{1}{1 + |T_{\mathrm{C}}| / |T_{\mathrm{D}}|} \leq \xi. \tag{8}$$

Part 2 of the lemma simply follows E_2. □

C.3. Proof of Proposition 13

Proof. Since $N \geq \frac{d}{b} \cdot \operatorname{polylog}\left((d, \frac{1}{\delta}\right)$, we have by Part 2 that $|T_{\mathrm{C}} \cup T_{\mathrm{E}}| \geq |T_{\mathrm{C}}| \geq d \cdot \operatorname{polylog}\left(d, \frac{1}{\delta}\right)$. Therefore, we can directly apply Proposition 8 by thinking of T_{C} therein as $T_{\mathrm{C}} \cup T_{\mathrm{E}}$ in the current proposition. □

C.4. Proof of Theorem 14

Proof. We first show the existence of a feasible function $q(x)$ to Algorithm 2. Consider the specific function $q : T \to [0, 1]$ as follows: $q(x) = 1$ for all $x \in T_{\mathrm{C}}$ and $q(x) = 0$ otherwise. We have

$$\frac{1}{|T|} \sum_{x \in T} q(x) = \frac{|T_{\mathrm{C}}|}{|T|} = 1 - \frac{|T_{\mathrm{D}}|}{|T|} \geq 1 - \xi,$$

in view of Part 1 of Lemma 12.

To show Part 3, we note that $T_{\mathrm{C}} \cup T_{\mathrm{E}}$ are i.i.d. draws from $D_{u,b}$ and Lemma 12 shows that $|T_{\mathrm{C}} \cup T_{\mathrm{E}}| \geq \Omega(bN)$. Therefore, as far as $N \geq \frac{d}{b} \cdot \operatorname{polylog}(d)$, Theorem 5 implies that

$$\frac{1}{|T_{\mathrm{C}}| + |T_{\mathrm{E}}|} \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (w \cdot x)^2 \leq \frac{c}{2}(b^2 + r^2).$$

Since $(w \cdot x)^2$ is always non-negative, we have

$$\frac{1}{|T_{\mathrm{C}}|} \sum_{x \in T_{\mathrm{C}}} (w \cdot x)^2 \leq \frac{|T_{\mathrm{C}}| + |T_{\mathrm{E}}|}{|T_{\mathrm{C}}|} \cdot \frac{1}{|T_{\mathrm{C}}| + |T_{\mathrm{E}}|} \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (w \cdot x)^2 \leq \frac{|T_{\mathrm{C}}| + |T_{\mathrm{E}}|}{|T_{\mathrm{C}}|} \cdot \frac{c}{2}(b^2 + r_k^2).$$

Part 2 of Lemma 12 shows that $|T_{\mathrm{E}}| / |T_{\mathrm{C}}| \leq \frac{\xi}{1-\xi} \leq 1$ since $\xi \leq \frac{1}{2}$. Plugging this upper bound into the above inequality, we obtain

$$\frac{1}{|T_{\mathrm{C}}|} \sum_{x \in T_{\mathrm{C}}} (w \cdot x)^2 \leq c(b^2 + r^2).$$

In a nutshell, our construction of $q(x)$ ensures the feasibility to all constraints in Algorithm 2. By ellipsoid method we are able to find a feasible solution in polynomial time. □

C.5. Proof of Proposition 15

Let $z = \sqrt{b^2 + r^2}$. We will in fact prove a stronger result, i.e.,

$$\ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) \leq \ell_\tau(w; p \circ \hat{T}) + 2\xi \left(2 + \sqrt{2K_2} \cdot \frac{z}{\tau}\right) + \sqrt{2K_2\xi} \cdot \frac{z}{\tau}, \tag{9}$$

$$\ell_\tau(w; p \circ \hat{T}) \leq \ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) + 2\xi + \sqrt{4K_2\xi} \cdot \frac{z}{\tau}. \tag{10}$$

The claim in the proposition immediately follows since $z/\tau = \Theta(1)$ and ξ can be chosen as an arbitrarily small constant.

Let $\{q(x)\}_{x \in T}$ be the output of Algorithm 2 under the nasty noise model. We extend the domain of $q(x)$ from T to $T \cup T_{\mathrm{E}}$ as follows: for any $x \in T$, the value $q(x)$ remains unchanged; for any $x \in T_{\mathrm{E}}$, we set $q(x) = 0$. With this in mind, we can, for the purpose of analysis, think of the probability mass function $\{p(x)\}_{x \in T}$ obtained in Algorithm 1 as over $T \cup T_{\mathrm{E}}$, with the value $p(x)$ stays unchanged for $x \in T$ and $p(x) = 0$ for all $x \in T_{\mathrm{E}}$.

Now with the *extended* probability mass function $\{p(x)\}_{x \in T \cup T_{\mathrm{E}}}$, we can prove the proposition.

Proof. Let $\hat{T}_{\mathrm{C}}$ and $\hat{T}_{\mathrm{E}}$ be the labeled set of T_{C} and T_{E} that is correctly annotated by w^* respectively. For any x in the instance space, let y_x be the label that the adversary is committed to. Recall that the empirical distribution $\{p(x)\}_{x \in T \cup T_{\mathrm{E}}}$ was defined as follows: $p(x) = \frac{q(x)}{\sum_{x \in T} q(x)}$ for $x \in T$ and $p(x) = 0$ for $x \in T_{\mathrm{E}}$. The reweighted hinge loss on $T \cup T_{\mathrm{E}}$ using $p(x)$ is given by

$$\ell_\tau(w; p \circ \hat{T}) = \frac{1}{|T \cup T_{\mathrm{E}}|} \sum_{x \in T \cup T_{\mathrm{E}}} p(x) \cdot \max\Big\{0, 1 - \frac{1}{\tau} y_x w \cdot x\Big\}. \tag{11}$$

The choice of N guarantees that Proposition 13, Lemma 12, and Theorem 14 hold simultaneously with probability $1 - \delta$. We thus have for all $w \in W$

$$\frac{1}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (w \cdot x)^2 \leq K_1 z^2, \tag{12}$$

$$\frac{|T_{\mathrm{D}}|}{|T|} \leq \xi, \tag{13}$$

$$\frac{1}{|T|} \sum_{x \in T} q(x) (w \cdot x)^2 \leq K_2 z^2. \tag{14}$$

We now expand T to $T \cup T_{\mathrm{E}}$ for the last two inequalities. Indeed, from (13), it is easy to show that

$$\frac{|T_{\mathrm{D}}|}{|T \cup T_{\mathrm{E}}|} \leq \frac{|T_{\mathrm{D}}|}{|T|} \leq \xi. \tag{15}$$

Next, since we defined $q(x) = 0$ for all $x \in T_{\mathrm{E}}$, (14) implies that

$$\frac{1}{|T \cup T_{\mathrm{E}}|} \sum_{x \in T \cup T_{\mathrm{E}}} q(x)(w \cdot x)^2 = \frac{1}{|T \cup T_{\mathrm{E}}|} \sum_{x \in T} q(x)(w \cdot x)^2 \leq \frac{1}{|T|} \sum_{x \in T} q(x)(w \cdot x)^2 \leq K_2 z^2. \tag{16}$$

The remaining steps are exactly same as Proposition 33 of Shen & Zhang (2021) since all the analyses therein rely only on the conditions (12), (15) and (16). For completeness, we present the full proof here.

It follows from Eq. (15) and $\xi \leq 1/2$ that

$$\frac{|T \cup T_{\mathrm{E}}|}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} \leq \frac{|T \cup T_{\mathrm{E}}|}{|T_{\mathrm{C}}|} = \frac{|T \cup T_{\mathrm{E}}|}{|T \cup T_{\mathrm{E}}| - |T_{\mathrm{D}}|} = \frac{1}{1 - |T_{\mathrm{D}}| / |T \cup T_{\mathrm{E}}|} \leq \frac{1}{1 - \xi} \leq 2. \tag{17}$$

In the following, we condition on the event that all these inequalities are satisfied.

Step 1. First we upper bound $\ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}})$ by $\ell_\tau(w; p \circ \hat{T})$.

$$\begin{aligned}
|T_{\mathrm{C}} \cup T_{\mathrm{E}}| \cdot \ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) &= \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} \ell(w; x, y_x) \\
&= \sum_{x \in T \cup T_{\mathrm{E}}} \Big[q(x) \ell(w; x, y_x) + \big(\mathbf{1}_{\{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}\}} - q(x)\big) \ell(w; x, y_x) \Big] \\
&\overset{\zeta_1}{\leq} \sum_{x \in T \cup T_{\mathrm{E}}} q(x) \ell(w; x, y_x) + \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (1 - q(x)) \ell(w; x, y_x) \\
&\overset{\zeta_2}{\leq} \sum_{x \in T \cup T_{\mathrm{E}}} q(x) \ell(w; x, y_x) + \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (1 - q(x)) \left(1 + \frac{|w \cdot x|}{\tau}\right) \\
&\overset{\zeta_3}{\leq} \sum_{x \in T \cup T_{\mathrm{E}}} q(x) \ell(w; x, y_x) + \xi |T \cup T_{\mathrm{E}}| + \frac{1}{\tau} \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (1 - q(x)) |w \cdot x| \\
&\overset{\zeta_4}{\leq} \sum_{x \in T \cup T_{\mathrm{E}}} q(x) \ell(w; x, y_x) + \xi |T \cup T_{\mathrm{E}}| + \frac{1}{\tau} \sqrt{\sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (1 - q(x))^2} \cdot \sqrt{\sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (w \cdot x)^2} \\
&\overset{\zeta_5}{\leq} \sum_{x \in T \cup T_{\mathrm{E}}} q(x) \ell(w; x, y_x) + \xi |T \cup T_{\mathrm{E}}| + \frac{1}{\tau} \sqrt{\xi |T \cup T_{\mathrm{E}}|} \cdot \sqrt{K_1 |T_{\mathrm{C}} \cup T_{\mathrm{E}}|} \cdot z,
\end{aligned} \tag{18}$$

where ζ_1 follows from the simple fact that

$$\sum_{x \in T \cup T_{\mathrm{E}}} \left(\mathbf{1}_{\{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}\}} - q(x)\right) \ell(w; x, y_x) = \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (1 - q(x))\ell(w; x, y_x) + \sum_{x \in T_{\mathrm{D}}} (-q(x))\ell(w; x, y_x)$$
$$\leq \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} (1 - q(x))\ell(w; x, y_x),$$

ζ_2 explores the fact that the hinge loss is always upper bounded by $1 + \frac{|w \cdot x|}{\tau}$ and that $1 - q(x) \geq 0$, ζ_3 follows from Part 2 of Theorem 14, ζ_4 applies Cauchy-Schwarz inequality, and ζ_5 uses Eq. (12).

In view of Eq. (17), we have $\frac{|T \cup T_{\mathrm{E}}|}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} \leq 2$. Continuing Eq. (18), we obtain

$$\begin{aligned}
\ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) &\leq \frac{1}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} \sum_{x \in T \cup T_{\mathrm{E}}} q(x)\ell(w; x, y_x) + 2\xi + \sqrt{2K_1\xi} \cdot \frac{z}{\tau} \\
&= \frac{\sum_{x \in T \cup T_{\mathrm{E}}} q(x)}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} \sum_{x \in T \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) + 2\xi + \sqrt{2K_1\xi} \cdot \frac{z}{\tau} \\
&= \ell_\tau(w; p \circ \hat{T}) + \left(\frac{\sum_{x \in T \cup T_{\mathrm{E}}} q(x)}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} - 1\right) \sum_{x \in T \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) + 2\xi + \sqrt{2K_1\xi} \cdot \frac{z}{\tau} \\
&\leq \ell_\tau(w; p \circ \hat{T}) + \left(\frac{|T \cup T_{\mathrm{E}}|}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} - 1\right) \sum_{x \in T \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) + 2\xi + \sqrt{2K_1\xi} \cdot \frac{z}{\tau} \\
&\leq \ell_\tau(w; p \circ \hat{T}) + 2\xi \sum_{x \in T \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) + 2\xi + \sqrt{2K_1\xi} \cdot \frac{z}{\tau},
\end{aligned} \tag{19}$$

where in the last inequality we use the fact that $|T_{\mathrm{E}}| = |T_{\mathrm{D}}|$ and $T \cap T_{\mathrm{E}} = \emptyset$, and thus

$$\frac{|T \cup T_{\mathrm{E}}|}{|T_{\mathrm{C}} \cup T_{\mathrm{E}}|} - 1 = \frac{|T| + |T_{\mathrm{D}}|}{|T|} - 1 = \frac{|T_{\mathrm{D}}|}{|T|} \leq \xi.$$

On the other hand, we have the following result which will be proved later on.

Claim 19. $\sum_{x \in T \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) \leq 1 + \sqrt{2K_2} \cdot \frac{z}{\tau}$.

Therefore, continuing Eq. (19) we have

$$\ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) \leq \ell_\tau(w; p \circ \hat{T}) + 2\xi \left(2 + \sqrt{2K_2} \cdot \frac{z}{\tau}\right) + \sqrt{2K_2\xi} \cdot \frac{z}{\tau}.$$

which proves the first inequality of the proposition.

Step 2. We move on to prove the second inequality of the theorem, i.e. using $\ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}})$ to upper bound $\ell_\tau(w; p \circ \hat{T})$. Let us denote by $p_{\mathrm{D}} = \sum_{x \in T_{\mathrm{D}}} p(x)$ the probability mass on dirty instances. Then

$$p_{\mathrm{D}} = \frac{\sum_{x \in T_{\mathrm{D}}} q(x)}{\sum_{x \in T} q(x)} \leq \frac{|T_{\mathrm{D}}|}{(1 - \xi)|T|} \leq \frac{\xi}{1 - \xi} \leq 2\xi, \tag{20}$$

where the first inequality follows from $q(x) \leq 1$ and Part 2 of Theorem 14, the second inequality follows from (13), and the last inequality is by our choice $\xi \leq 1/2$.

Note that by Part 2 of Theorem 14 and the choice $\xi \leq 1/2$, we have

$$\sum_{x \in T} q(x) \geq (1 - \xi)|T| \geq |T|/2$$

Hence

$$\sum_{x \in T} p(x)(w \cdot x)^2 = \frac{1}{\sum_{x \in T} q(x)} \sum_{x \in T} q(x)(w \cdot x)^2$$
$$\leq \frac{2}{|T|} \sum_{x \in T} q(x)(w \cdot x)^2$$
$$\leq 2 \cdot K_2 z^2 \tag{21}$$

where the last inequality holds because of (14). Thus,

$$\sum_{x \in T_{\mathrm{D}}} p(x)\ell(w; x, y_x) \leq \sum_{x \in T_{\mathrm{D}}} p(x) \left(1 + \frac{|w \cdot x|}{\tau}\right)$$
$$= p_{\mathrm{D}} + \frac{1}{\tau} \sum_{x \in T_{\mathrm{D}}} p(x)|w \cdot x|$$
$$= p_{\mathrm{D}} + \frac{1}{\tau} \sum_{x \in T} \left(\mathbf{1}_{\{x \in T_{\mathrm{D}}\}} \sqrt{p(x)}\right) \cdot \left(\sqrt{p(x)}|w \cdot x|\right)$$
$$\leq p_{\mathrm{D}} + \frac{1}{\tau} \sqrt{\sum_{x \in T} \mathbf{1}_{\{x \in T_{\mathrm{D}}\}} p(x)} \cdot \sqrt{\sum_{x \in T} p(x)(w \cdot x)^2}$$
$$\overset{(21)}{\leq} p_{\mathrm{D}} + \sqrt{p_{\mathrm{D}}} \cdot \sqrt{2K_2} \cdot \frac{z}{\tau}.$$

With the result on hand, we bound $\ell_\tau(w; p \circ \hat{T})$ as follows:

$$\ell_\tau(w; p \circ \hat{T}) = \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) + \sum_{x \in T_{\mathrm{D}}} p(x)\ell(w; x, y_x)$$
$$\leq \sum_{x \in T_{\mathrm{C}} \cup T_{\mathrm{E}}} \ell(w; x, y_x) + \sum_{x \in T_{\mathrm{D}}} p(x)\ell(w; x, y_x)$$
$$= \ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) + \sum_{x \in T_{\mathrm{D}}} p(x)\ell(w; x, y_x)$$
$$\leq \ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) + p_{\mathrm{D}} + \sqrt{p_{\mathrm{D}}} \cdot \sqrt{2K_2} \cdot \frac{z}{\tau}$$
$$\overset{(20)}{\leq} \ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) + 2\xi + \sqrt{4K_2\xi} \cdot \frac{z}{\tau},$$

which proves the second inequality of the proposition.

This completes the proof. □

Proof of Claim 19. Since $\ell(w; x, y_x) \leq 1 + \frac{|w \cdot x|}{\tau}$, it follows that

$$\sum_{x \in T \cup T_{\mathrm{E}}} p(x)\ell(w; x, y_x) \leq \sum_{x \in T \cup T_{\mathrm{E}}} p(x) \left(1 + \frac{|w \cdot x|}{\tau}\right)$$
$$= 1 + \frac{1}{\tau} \sum_{x \in T \cup T_{\mathrm{E}}} p(x)|w \cdot x|$$
$$\leq 1 + \frac{1}{\tau} \sqrt{\sum_{x \in T \cup T_{\mathrm{E}}} p(x)(w \cdot x)^2}$$
$$\overset{(21)}{\leq} 1 + \sqrt{2K_2} \cdot \frac{z}{\tau},$$

which completes the proof of Claim 19. □

C.6. Proof of Lemma 16

For any phase k, let $L_{\tau_k}(w) = \mathbb{E}_{x \sim D_{w_{k-1}, b_k}} \left[\ell_{\tau_k}(w; x, \mathrm{sign}\,(w^* \cdot x)) \right]$.

Proof. Proposition 35 of Shen & Zhang (2021) showed that if $|T_{\mathrm{C}} \cup T_{\mathrm{E}}| \geq d \cdot \mathrm{polylog}\left(d, \frac{1}{b}, \frac{1}{\delta}\right)$, then by Rademacher complexity of the hinge loss we have that with probability $1 - \frac{\delta}{2}$

$$\sup_{w \in W} \left| \ell_\tau(w; \hat{T}_{\mathrm{C}} \cup \hat{T}_{\mathrm{E}}) - \mathbb{E}_{x \sim D_{u,b}}[\ell_\tau(w; x, \mathrm{sign}\,(w^* \cdot x))] \right| \leq \kappa. \tag{22}$$

Combining the above with Proposition 15 gives that with probability $1 - \delta$,

$$\sup_{w \in W} \left| \ell_\tau(w; p \circ \hat{T}) - \mathbb{E}_{x \sim D_{u,b}}[\ell_\tau(w; x, \mathrm{sign}\,(w^* \cdot x))] \right| \leq 2\kappa.$$

Namely, in any phase $k \leq K$, if $|T_{\mathrm{C}} \cup T_{\mathrm{E}}| \geq d \cdot \mathrm{polylog}\left(d, \frac{1}{b_k}, \frac{1}{\delta_k}\right)$, then with probability $1 - \delta_k$,

$$\sup_{w \in W_k} \left| \ell_{\tau_k}(w; p) - L_{\tau_k}(w) \right| \leq 2\kappa. \tag{23}$$

On the other hand, since the (rescaled) hinge loss is always an upper bound of the error rate, we have

$$\mathrm{err}_{D_{w_{k-1}, b_k}}(v_k) \leq L_k(v_k) \overset{\zeta_1}{\leq} \ell_{\tau_k}(v_k; p) + 2\kappa \overset{\zeta_2}{\leq} \min_{w \in W_k} \ell_{\tau_k}(w; p) + 3\kappa \leq \ell_{\tau_k}(w^*; p) + 3\kappa \overset{\zeta_3}{\leq} L_k(w^*) + 5\kappa \overset{\zeta_4}{\leq} 6\kappa \leq 8\kappa,$$

where we use the fact that $v_k \in W_k$ in ζ_1, use the optimality condition of v_k in ζ_2, use $w^* \in W_k$ in ζ_3, and use Lemma 20 in ζ_4. □

Lemma 20 (Lemma 3.7 in Awasthi et al. (2017))**.** *Suppose Assumption 1 is satisfied. Then*

$$L_{\tau_k}(w^*) \leq \frac{\tau_k}{c_0 \min\{b_k, 1/9\}}.$$

In particular, by our choice of τ_k, it holds that

$$L_{\tau_k}(w^*) \leq \kappa.$$

Lemma 21. *For any $1 \leq k \leq K$, if $w^* \in W_k$, then with probability $1 - \delta_k$, $\theta(v_k, w^*) \leq 2^{-k-8}\pi$.*

Proof. For $k = 1$, by Lemma 16 with the facts that we actually sample from D and $w^* \in \mathbb{R}^d =: W_1$, we immediately have

$$\mathrm{Pr}_{x \sim D}\left(\mathrm{sign}\,(v_1 \cdot x) \neq \mathrm{sign}\,(w^* \cdot x)\right) \leq 8\kappa.$$

Hence Part 4 of Lemma 24 indicates that

$$\theta(v_1, w^*) \leq 8c_2\kappa = 16c_2\kappa \cdot 2^{-1}. \tag{24}$$

Now we consider $2 \leq k \leq K$. Denote $X_k = \{x : |w_{k-1} \cdot x| \leq b_k\}$, and $\bar{X}_k = \{x : |w_{k-1} \cdot x| > b_k\}$. We will show that the error of v_k on both X_k and $\bar{X}_k$ is small, hence v_k is a good approximation to w^*.

First, we consider the error on X_k, which is given by

$$\begin{aligned}
&\mathrm{Pr}_{x \sim D}\left(\mathrm{sign}\,(v_k \cdot x) \neq \mathrm{sign}\,(w^* \cdot x), x \in X_k\right) \\
=\ &\mathrm{Pr}_{x \sim D}\left(\mathrm{sign}\,(v_k \cdot x) \neq \mathrm{sign}\,(w^* \cdot x) \mid x \in X_k\right) \cdot \mathrm{Pr}_{x \sim D}(x \in X_k) \\
=\ &\mathrm{err}_{D_{w_{k-1}, b_k}}(v_k) \cdot \mathrm{Pr}_{x \sim D}(x \in X_k) \\
\leq\ &8\kappa \cdot 2b_k = 16\kappa b_k,
\end{aligned} \tag{25}$$

where the inequality is due to Lemma 16 and Lemma 24. Note that the inequality holds with probability $1 - \delta_k$ in view of Lemma 16.

Next we derive the error on $\bar{X}_k$. Note that Lemma 10 of Zhang (2018) states for any unit vector u, and any general vector v, $\theta(v, u) \leq \pi \left\|v - u\right\|_2$. Hence,

$$\theta(v_k, w^*) \leq \pi \left\|v_k - w^*\right\|_2 \leq \pi(\left\|v_k - w_{k-1}\right\|_2 + \left\|w^* - w_{k-1}\right\|_2) \leq 2\pi r_k,$$

where we use the condition that both v_k and w^* are in W_k.

Recall that we set $r_k = 2^{-k-6} < 1/4$ in our algorithm and choose $b_k = \bar{c} \cdot r_k$ where $\bar{c} \geq 8\pi/c_4$, which allows us to apply Lemma 25 and obtain

$$\begin{aligned}\Pr_{x \sim D}\left(\operatorname{sign}(v_k \cdot x) \neq \operatorname{sign}(w^* \cdot x), x \notin X_k\right) &\leq c_3 \cdot 2\pi r_k \cdot \exp\left(-\frac{c_4 \bar{c} \cdot r_k}{2 \cdot 2\pi r_k}\right) \\ &= 2^{-k} \cdot \frac{c_3 \pi}{4} \exp\left(-\frac{c_4 \bar{c}}{4\pi}\right).\end{aligned}$$

This in allusion to (25) gives

$$\operatorname{err}_D(v_k) \leq 16\kappa \cdot \bar{c} \cdot r_k + 2^{-k} \cdot \frac{c_3 \pi}{4} \exp\left(-\frac{c_4 \bar{c}}{4\pi}\right) = \left(2\kappa \bar{c} + \frac{c_3 \pi}{4} \exp\left(-\frac{c_4 \bar{c}}{4\pi}\right)\right) \cdot 2^{-k}.$$

Recall that we set $\kappa = \exp(-\bar{c})$. For convenience denote by $f(\bar{c})$ the coefficient of 2^{-k} in the above expression. By Part 4 of Lemma 24

$$\theta(v_k, w^*) \leq c_2 \operatorname{err}_D(v_k) \leq c_2 f(\bar{c}) \cdot 2^{-k}. \tag{26}$$

Now let $g(\bar{c}) = c_2 f(\bar{c}) + 16 c_2 \exp(-\bar{c})$. By our choice of $\bar{c}$, $g(\bar{c}) \leq 2^{-8}\pi$. This ensures that for both (24) and (26), $\theta(v_k, w^*) \leq 2^{-k-8}\pi$ for any $k \geq 1$. □

Lemma 22. *For any $1 \leq k \leq K$, if $\theta(v_k, w^*) \leq 2^{-k-8}\pi$, then $w^* \in W_{k+1}$.*

Proof. We only need to show that $\left\|w_k - w^*\right\|_2 \leq r_{k+1}$. Let $\hat{v}_k = v_k / \left\|v_k\right\|_2$. By algebra $\left\|\hat{v}_k - w^*\right\|_2 = 2\sin\frac{\theta(v_k, w^*)}{2} \leq \theta(v_k, w^*) \leq 2^{-k-8}\pi \leq 2^{-k-6}$. Now we have

$$\left\|w_k - w^*\right\|_2 = \left\|\hat{v}_k - w^*\right\|_2 \leq 2^{-k-6} = r_{k+1}.$$

The proof is complete. □

C.7. Proof of Theorem 2

Proof. We will prove the theorem with the following claim.

Claim 23. *For any $1 \leq k \leq K$, with probability at least $1 - \sum_{i=1}^{k} \delta_i$, w^* is in W_{k+1}.*

Based on the claim, we immediately have that with probability at least $1 - \sum_{k=1}^{K} \delta_k \geq 1 - \delta$, w^* is in W_{K+1}. By our construction of W_{K+1}, we have

$$\left\|w^* - w_K\right\|_2 \leq 2^{-K-5}.$$

This, together with Part 4 of Lemma 24 and the fact that $\theta(w^*, w_K) \leq \pi \left\|w^* - w_K\right\|_2$ (see Lemma 10 of Zhang (2018)), implies

$$\operatorname{err}_D(w_K) \leq \frac{\pi}{c_1} \cdot 2^{-K-5} = \epsilon.$$

The sample complexity of the algorithm is given by

$$N := \sum_{k=1}^{K} N_k = \sum_{k=1}^{K} \frac{d}{b_k} \cdot \operatorname{polylog}\left(d, \frac{1}{b_k}, \frac{1}{\delta_k}\right) \leq \frac{d}{\epsilon} \cdot \operatorname{polylog}\left(d, \frac{1}{\epsilon}, \frac{1}{\delta}\right),$$

where we use the fact that $b_k \geq K_1 \epsilon$ for some constant $K_1 > 0$ and $K = O(\log \frac{1}{\epsilon})$.

For each phase $k \leq K$, the number of calls to EX^y equals the size of T. For the size of T_{C}, by Lemma 24 we know that the probability mass of the band $X_k = \{x : |w_{k-1} \cdot x| \leq b_k\}$ is at most $2b_k$, implying that $|T_{\mathrm{C}}| \leq O(b_k N_k)$ with high probability in view of Chernoff bound. On the other hand, by Part 2 of Lemma 12 we have $|T_{\mathrm{D}}| = |T_{\mathrm{E}}| \leq O(b_k N_k)$ since $\xi_k = \Theta(1)$ as indicated in Section 3.1.1. Therefore, $|T| \leq O(b_k N_k)$ and the label complexity m of the algorithm is given by

$$m \leq \sum_{k=1}^{K} b_k N_k = d \cdot \mathrm{polylog}\left(d, \frac{1}{\epsilon}, \frac{1}{\delta}\right).$$

It remains to prove Claim 23 by induction. First, for $k = 1$, $W_1 = \{w : \|w\|_2 \leq 1\}$. Therefore, $w^* \in W_1$ with probability 1. Now suppose that Claim 23 holds for some $k \geq 2$, that is, there is an event E_{k-1} that happens with probability $1 - \sum_i^{k-1} \delta_i$, and on this event $w^* \in W_k$. By Lemma 21 we know that there is an event F_k that happens with probability $1 - \delta_k$, on which $\theta(v_k, w^*) \leq 2^{-k-8}\pi$. This further implies that $w^* \in W_{k+1}$ in view of Lemma 22. Therefore, consider the event $E_{k-1} \cap F_k$, on which $w^* \in W_{k+1}$ with probability $\Pr(E_{k-1}) \cdot \Pr(F_k \mid E_{k-1}) = (1 - \sum_i^{k-1} \delta_i)(1 - \delta_k) \geq 1 - \sum_{i=1}^{k} \delta_i$. □

D. Properties of Isotropic Log-Concave Distributions

We record some useful properties of isotropic log-concave distributions.

Lemma 24. *There are absolute constants $c_0, c_1, c_2 > 0$, such that the following holds for all isotropic log-concave distributions $D \in \mathcal{D}$. Let f_D be the density function. We have*

1. *Orthogonal projections of D onto subspaces of $\mathbb{R}^d$ are isotropic log-concave;*
2. *If $d = 1$, then $\Pr_{x \sim D}(a \leq x \leq b) \leq |b - a|$;*
3. *If $d = 1$, then $f_D(x) \geq c_0$ for all $x \in [-1/9, 1/9]$;*
4. *For any two vectors $u, v \in \mathbb{R}^d$,*
$$c_1 \cdot \Pr_{x \sim D}\left(\mathrm{sign}\,(u \cdot x) \neq \mathrm{sign}\,(v \cdot x)\right) \leq \theta(u, v) \leq c_2 \cdot \Pr_{x \sim D}\left(\mathrm{sign}\,(u \cdot x) \neq \mathrm{sign}\,(v \cdot x)\right);$$
5. $\Pr_{x \sim D}\left(\|x\|_2 \geq t\sqrt{d}\right) \leq \exp(-t + 1)$.

We remark that Parts 1, 2, 3, and 5 are due to Lovász & Vempala (2007), and Part 4 is from Vempala (2010); Balcan & Long (2013).

The following lemma is implied by the proof of Theorem 21 of Balcan & Long (2013), which shows that if we choose a proper band width $b > 0$, the error outside the band will be small. This observation is crucial for controlling the error over the distribution D, and has been broadly recognized in the literature (Awasthi et al., 2017; Zhang, 2018).

Lemma 25 (Theorem 21 of Balcan & Long (2013))**.** *There are absolute constants $c_3, c_4 > 0$ such that the following holds for all isotropic log-concave distributions $D \in \mathcal{D}$. Let u and v be two unit vectors in $\mathbb{R}^d$ and assume that $\theta(u, v) = \alpha < \pi/2$. Then for any $b \geq \frac{4}{c_4}\alpha$, we have*

$$\Pr_{x \sim D}(\mathrm{sign}\,(u \cdot x) \neq \mathrm{sign}\,(v \cdot x) \text{ and } |v \cdot x| \geq b) \leq c_3 \alpha \exp\left(-\frac{c_4 b}{2\alpha}\right).$$

Lemma 26. *Suppose x is randomly drawn from $D_{u,b}$. Then with probability $1 - \delta$, $\|x\|_2 \leq c_7 \sqrt{d} \log \frac{1}{b\delta}$ for some constant $c_7 > 0$.*

Proof. Using Part 2 of Lemma 27, we have

$$\Pr_{x \sim D_{u,b}}(\|x\|_2 \geq \alpha) \leq \frac{1}{c_8 b} \Pr_{x \sim D}(\|x\|_2 \geq \alpha) \leq \frac{e}{c_8 b} \exp\left(-\alpha/\sqrt{d}\right),$$

where we applied Part 5 of Lemma 24 in the last inequality. The lemma follows by setting the right-hand side to δ. □

Lemma 27. *Let* $c_8 = \min\left\{2c_0, \frac{2c_0}{9C_1}, \frac{1}{C_1}\right\}$*. Then for all isotropic log-concave distributions* $D \in \mathcal{D}$*,*

1. $\Pr_{x \sim D}\left(|u \cdot x| \leq b\right) \geq c_8 \cdot b$*;*
2. $\Pr_{x \sim D_{u,b}}(E) \leq \frac{1}{c_8 b} \Pr_{x \sim D}(E)$ *for any event* E*.*

Proof. We first consider the case that u is a unit vector.

For the lower bound, Part 3 of Lemma 24 shows that the density function of the random variable $u \cdot x$ is lower bounded by c_0 when $|u \cdot x| \leq 1/9$. Thus

$$\Pr_{x \sim D}\left(|u \cdot x| \leq b\right) \geq \Pr_{x \sim D}\left(|u \cdot x| \leq \min\{b, 1/9\}\right) \geq 2c_0 \min\{b, 1/9\} \geq 2c_0 \min\left\{1, \frac{1}{9C_1}\right\} \cdot b$$

where in the last inequality we use the condition $b \leq C_1$.

For any event E, we always have

$$\Pr_{x \sim D_{u,b}}(E) \leq \frac{\Pr_{x \sim D}(E)}{\Pr_{x \sim D}(|u \cdot x| \leq b)} \leq \frac{1}{c_8 b} \Pr_{x \sim D}(E).$$

Now we consider the case that u is the zero vector and $b = C_1$. Then $\Pr_{x \sim D}\left(|u \cdot x| \leq b\right) = 1 \geq c_8 \cdot b$ in view of the choice c_8. Thus Part 2 still follows. The proof is complete. $\square$