Supplement to “Binary Embedding: Fundamental Limits and Fast Algorithm"

Xinyang Yi, Constantine Caramanis and Eric Price

In this supplementary manuscript, we provide the proofs for the theoretical results appeared in our main paper. Proofs about lower bounds are contained in Section A. We prove the results related to fast binary embedding in Section B. The guarantee for embedding general set (Theorem 3.10) is proved in Section C. Several auxiliary lemmas are proved in Section D.

A. Proofs about Lower Bounds

A.1. Proof of Data-Oblivious Lower Bound (Theorem 3.1)

The proof of the data-oblivious lower bound is based on a lower bound for one-way communication of Hamming distance due to (Jayram & Woodruff, 2013).

Definition A.1 (One-way communication of Hamming distance)**.** In the one-way communication model, Alice is given $\boldsymbol{a} \in \{0,1\}^n$ and Bob is given $\boldsymbol{b} \in \{0,1\}^n$. Alice sends Bob a message $\boldsymbol{c} \in \{0,1\}^m$, and Bob uses $\boldsymbol{b}$ and $\boldsymbol{c}$ to output a value $x \in \mathbb{R}$. Alice and Bob have shared randomness.

Alice and Bob solve the (δ, ϵ) additive Hamming distance estimation problem if $|x - d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b})| \leq \delta$ with probability $1 - \epsilon$.

The result proven in (Jayram & Woodruff, 2013) is a lower bound for the *multiplicative* Hamming distance estimation problem, but their techniques readily yield a bound for the additive case as well:

Lemma A.2. Any algorithm that solves the (δ, ϵ) additive Hamming distance estimation problem must have $m = \Omega((1/\delta^2)\log(1/\epsilon))$ as long as this is less than n.

Proof. We apply Lemma 3.1 of (Jayram & Woodruff, 2013) with parameters $\alpha = 2$, $p = 1$, $b = 1$, $\varepsilon = \delta$, and $\delta = \epsilon$. This encodes inputs from a problem they prove is hard (augmented indexing on large domains) to inputs appropriate for Hamming estimation. In particular, for $n' = O(\frac{1}{\delta^2}\log(1/\epsilon))$ it gives a distribution on $(\boldsymbol{a}, \boldsymbol{b}) \in \{0,1\}^{n'} \times \{0,1\}^{n'}$ that are divided into “NO” and “YES” instances, such that:

- From the reduction, distinguishing NO instances from YES instances with probability $1 - \epsilon$ requires Alice to send $m = \Omega(\frac{1}{\delta^2}\log(1/\epsilon))$ bits of communication to Bob.
- In NO instances, $d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b}) \geq \frac{1}{2}(1 - \delta/3)$.
- In YES instances, $d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b}) \leq \frac{1}{2}(1 - 2\delta/3)$.

First, suppose $n = n'$. Then since solving the additive Hamming distance estimation problem with $\delta/12$ accuracy would distinguish NO instances from YES instances, it must involve $m = \Omega(\frac{1}{\delta^2}\log(1/\epsilon))$ bits of communication.

For $n > n'$, simply duplicate the coordinates of a and b $\lfloor n/n' \rfloor$ times, and zero-pad the remainder. Less than half the coordinates are then part of the zero-padding, so the gap between YES and NO instances remains at least $\delta/12$ and a protocol for the $(\delta/24, \epsilon)$ additive Hamming distance estimation problem requires $m = \Omega(\frac{1}{\delta^2}\log(1/\epsilon))$ as desired. □

With this in hand, we can prove Theorem 3.1:

Proof of Theorem 3.1. We reduce one-way communication of the (δ, ϵ) additive Hamming distance estimation problem to the embedding problem. Let $\boldsymbol{a}, \boldsymbol{b} \in \{0,1\}^p$ be drawn from the hard instance for the communication problem defined in Lemma A.2. Linearly transform them to $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{S}^{p-1}$ via $\boldsymbol{u} = (2 \cdot \boldsymbol{a} - \mathbf{1})/\sqrt{p}$, $\boldsymbol{v} = (2 \cdot \boldsymbol{b} - \mathbf{1})/\sqrt{p}$. We have that $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 1 - 2d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b})$, so

$$d(\boldsymbol{u}, \boldsymbol{v}) = 1 - \frac{\arccos(\langle \boldsymbol{u}, \boldsymbol{v} \rangle)}{\pi} = 1 - \frac{\arccos(1 - 2d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b}))}{\pi}$$

or

$$d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{2}(1 - \cos(\pi - \pi d(\boldsymbol{u}, \boldsymbol{v})))$$

Given an estimate of $d(\boldsymbol{u}, \boldsymbol{v})$, we can therefore get an estimate of $d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b})$. In particular, since $|\cos'(x)| \le 1$, if we learn $d(\boldsymbol{u}, \boldsymbol{v})$ to $\pm\delta$ then we learn $d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b})$ to $\pm\delta\frac{\pi}{2}$.

For now, consider the case of $N = 2$. Consider an oblivious embedding function $f : \mathbb{S}^{p-1} \to \{0,1\}^m$ and reconstruction algorithm $g : \{0,1\}^m \times \{0,1\}^m \to \mathbb{R}$ that has

$$|g(f(\boldsymbol{u}), f(\boldsymbol{v})) - d(\boldsymbol{u}, \boldsymbol{v})| \le \delta \frac{2}{\pi}$$

with probability $1 - \epsilon$ on the distribution of inputs $(\boldsymbol{u}, \boldsymbol{v})$. We can solve the one-way communication problem for Hamming distance estimation by Alice sending $f(\boldsymbol{u})$ to Bob, Bob learning $d(\boldsymbol{u}, \boldsymbol{v}) \approx g(f(\boldsymbol{u}), f(\boldsymbol{v}))$, and then computing $d_{\mathcal{H}}(\boldsymbol{a}, \boldsymbol{b})$ to $\pm\delta$. By the lower bound for this problem, any such f and g must have $m = \Omega(\frac{1}{\delta^2} \log \frac{1}{\epsilon})$, proving the result for $N = 2$ (after rescaling δ).

For general N, we draw instances $(\boldsymbol{u}_1, \boldsymbol{v}_1), (\boldsymbol{u}_2, \boldsymbol{v}_2), \ldots, (\boldsymbol{u}_{N/2}, \boldsymbol{v}_{N/2})$ independently from the hard instance for binary embedding of $N = 2$ and $\epsilon' = 4\epsilon/N$. Consider an oblivious embedding function $f : \mathbb{S}^{p-1} \to \{0,1\}^m$ and reconstruction algorithm $g : \{0,1\}^m \times \{0,1\}^m \to \mathbb{R}$ that has for all $i \in [N/2]$ that

$$|g(f(\boldsymbol{u}_i), f(\boldsymbol{v}_i)) - d(\boldsymbol{u}_i, \boldsymbol{v}_i)| \le \delta$$

with probability $1 - \epsilon$ on this distribution. Define α to be the probability that $|g(f(\boldsymbol{u}_i), f(\boldsymbol{v}_i)) - d(\boldsymbol{u}_i, \boldsymbol{v}_i)| \le \delta$ for any particular i. Because f and g are oblivious and the different instances are independent, we have the probability that all instances succeed is $\alpha^{N/2} \ge 1 - \epsilon$, so

$$\alpha > (1 - \epsilon)^{2/N} > 1 - 4\epsilon/N.$$

In particular, this means f and g solve the hard instance of binary embedding and $N = 2$, $\epsilon' = 4\epsilon/N$. By the above lower bound for $N = 2$, this means

$$m = \Omega(\frac{1}{\delta^2} \log(N/\epsilon))$$

as desired. □

A.2. Proof of Data-Dependent Lower Bound (Theorem 3.3)

We need a few ingredients to show the lower bound. First, we define a matrix that is close to identity matrix.

Definition A.3. ((δ_1, δ_2)-near identity matrix) Symmetric matrix $\mathbf{M} \in \mathbb{R}^{p \times p}$ is called a (δ_1, δ_2)-near identity matrix if it satisfies both of the following conditions:

$$1 - \delta_1 \le \mathbf{M}_{i,i} \le 1, \forall\, i \in [p],$$

$$\left|\mathbf{M}_{i,j}\right| \le \delta_2, \forall\, i \ne j \in [p].$$

Next we give a lower bound on the rank of (δ_1, δ_2)-near identity matrix.

Lemma A.4. Suppose positive semidefinite matrix $\mathbf{M} \in \mathbb{R}^{p \times p}$ is a (δ_1, δ_2)-near identity matrix with rank d, and $0 < \delta_1, \delta_2 < 1$. Then we have

$$d \ge \frac{p(1 - \delta_1)^2}{1 + (p - 1)\delta_2^2}.$$

Proof. We postpone the proof to Appendix D.2. □

The above result is weak when it is applied to show our desired lower bound. We still need to make use of the following combinatorial result.

Lemma A.5. Suppose matrix $\mathbf{M} \in \mathbb{R}^{p\times p}$ has rank d. Let $P(x)$ be any degree k polynomial function. Consider matrix $\mathbf{N} \in \mathbb{R}^{p\times p}$ defined as $\mathbf{N} := P(\mathbf{M})$, where the $\mathbf{N}_{i,j} = P(\mathbf{M}_{i,j})$. We have

$$\operatorname{rank}(\mathbf{N}) \leq \binom{k+d}{k}.$$

Proof. See Lemma 9.2 of (Alon, 2003) for a detailed proof. □

Now we are ready to prove Theorem 3.3.

Proof of Theorem 3.3. Let $\boldsymbol{e}_i$ denote the i'th natural basis of $\mathbb{R}^N$, i.e., the i'th coordinate is 1 while the rest are all zeros. Consider N points $\{\boldsymbol{e}_1, \boldsymbol{e}_2, ..., \boldsymbol{e}_N\}$ and their opposite vectors $\{-\boldsymbol{e}_1, -\boldsymbol{e}_2, ..., -\boldsymbol{e}_N\}$. For any binary embedding algorithm f, we let

$$\boldsymbol{b}_i := f(\boldsymbol{e}_i),\ \forall\ i \in [N],$$
$$\boldsymbol{c}_i := f(-\boldsymbol{e}_i),\ \forall\ i \in [N].$$

Under the condition that f solves the general binary embedding problem with link function $\mathcal{L}$, we have

$$\Big| d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{c}_i) - \mathcal{L}\big(d(\boldsymbol{e}_i, -\boldsymbol{e}_i)\big)\Big| \leq \delta, \forall\ i \in [N]. \tag{A.1}$$

As $d(\boldsymbol{e}_i, -\boldsymbol{e}_i) = 1$, we have

$$\mathcal{L}(1) + \delta \geq d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{c}_i) \geq \mathcal{L}(1) - \delta. \tag{A.2}$$

Similarly, note that

$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = d(\boldsymbol{e}_i, -\boldsymbol{e}_j) = d(-\boldsymbol{e}_i, -\boldsymbol{e}_j) = \frac{1}{2},\ \forall\ i \neq j,$$

we have $\forall\ i \neq j$

$$\mathcal{L}(1/2) - \delta \leq d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{b}_j) \leq \mathcal{L}(1/2) + \delta, \tag{A.3}$$
$$\mathcal{L}(1/2) - \delta \leq d_{\mathcal{H}}(\boldsymbol{c}_i, \boldsymbol{c}_j) \leq \mathcal{L}(1/2) + \delta, \tag{A.4}$$
$$\mathcal{L}(1/2) - \delta \leq d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{c}_j) \leq \mathcal{L}(1/2) + \delta. \tag{A.5}$$

From now on, we treat binary strings $\boldsymbol{b}_i, \boldsymbol{c}_i$ as vectors in $\mathbb{R}^m$. Let $\mathbf{B}$ denote the matrix with rows $\boldsymbol{b}_i$ and $\mathbf{C}$ denote the matrix with rows $\boldsymbol{c}_i$. Consider the outer product of the difference between $\mathbf{B}$ and $\mathbf{C}$, namely

$$\mathbf{M} = (\mathbf{B} - \mathbf{C})(\mathbf{B} - \mathbf{C})^\top.$$

Note that $\forall\ i \in [N]$,

$$\mathbf{M}_{i,i} = \|\boldsymbol{b}_i - \boldsymbol{c}_i\|_2^2 = 4m \cdot d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{c}_i) \geq 4m\big(\mathcal{L}(1) - \delta\big).$$

The last inequality follows from (A.2). For $\forall\ i \neq j$, we have

$$\begin{aligned}\mathbf{M}_{i,j} &= \big\langle \boldsymbol{b}_i - \boldsymbol{c}_i, \boldsymbol{b}_j - \boldsymbol{c}_j\big\rangle = \big\langle \boldsymbol{b}_i, \boldsymbol{b}_j\big\rangle + \big\langle \boldsymbol{c}_i, \boldsymbol{c}_j\big\rangle - \big\langle \boldsymbol{b}_i, \boldsymbol{c}_j\big\rangle - \big\langle \boldsymbol{b}_j, \boldsymbol{c}_i\big\rangle \\ &= 2m\Big(d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{c}_j) + d_{\mathcal{H}}(\boldsymbol{b}_j, \boldsymbol{c}_i) - d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{b}_j) - d_{\mathcal{H}}(\boldsymbol{c}_i, \boldsymbol{c}_j)\Big),\end{aligned}$$

where the third equality follows from

$$d_{\mathcal{H}}(\boldsymbol{b}, \boldsymbol{c}) = \frac{1}{4m}\big(\|\boldsymbol{b}\|_2^2 + \|\boldsymbol{c}\|_2^2 - 2\big\langle \boldsymbol{b}, \boldsymbol{c}\big\rangle\big)\ \forall\ \boldsymbol{b}, \boldsymbol{c} \in \{-1, 1\}^m$$

By using (A.3) to (A.5), we have

$$\left|\boldsymbol{M}_{i,j}\right| \leq 8\delta m.$$

Therefore, $\frac{1}{4m\cdot(\mathcal{L}(1)+\delta)}\mathbf{M}$ is actually a $\left(\frac{2\delta}{\mathcal{L}(1)}, \frac{2\delta}{\mathcal{L}(1)}\right)$-near identity matrix. Consider degree k polynomial $P(z) = z^k$. Let

$$\mathbf{N} = P\big(\frac{1}{4m\cdot\mathcal{L}(1)}\mathbf{M}\big).$$

It is easy to observe that $\mathbf{N}$ is a (γ_1, γ_2)-near identity matrix where

$$\gamma_1 = 1 - (1 - \frac{2\delta}{\mathcal{L}(1)})^k,$$

and

$$\gamma_2 = (\frac{2\delta}{\mathcal{L}(1)})^k.$$

Under the condition $\frac{\delta}{\mathcal{L}(1)} \leq \frac{1}{4}$, we have

$$\gamma_1 = 1 - (1 - \frac{\delta}{\mathcal{L}(1)})^k \leq 1 - (\frac{1}{2})^k.$$

By setting $k = \frac{1}{2}\frac{\log N}{\log\frac{\mathcal{L}(1)}{2\delta}}$, we have

$$\gamma_2 \leq \sqrt{\frac{1}{N}}.$$

We apply Lemma A.4 by setting δ_1, δ_2, p in the statement to be γ_1, γ_2, N respectively. We get

$$\operatorname{rank}(\mathbf{N}) \geq \frac{N(\frac{1}{4})^k}{1 + (N-1)/N} \geq \frac{1}{2}(\frac{1}{4})^k N \geq (\frac{1}{8})^k N. \tag{A.6}$$

On the other hand, $\frac{1}{4m\cdot\mathcal{L}(1)}\mathbf{M}$ has rank at most m. By applying Lemma A.5 we get

$$\operatorname{rank}(\mathbf{N}) \leq \binom{m+k}{k} \leq (\frac{e(m+k)}{k})^k.$$

Applying the above result and (A.6) directly yields that

$$(N)^{1/k} \leq 8e\frac{m+k}{k}.$$

When $k = \frac{1}{2}\frac{\log N}{\log\frac{\mathcal{L}(1)}{2\delta}}$ as we set, $N^{1/k} \geq (\frac{\mathcal{L}(1)}{2\delta})^2$. Therefore we have

$$m \geq \frac{1}{32e}\big(\frac{\mathcal{L}(1)}{\delta}\big)^2 k - k \geq \frac{1}{64e}\big(\frac{\mathcal{L}(1)}{2\delta}\big)^2 k = \frac{1}{128e}\big(\frac{\mathcal{L}(1)}{\delta}\big)^2\frac{\log N}{\log\frac{\mathcal{L}(1)}{2\delta}},$$

where the second inequality holds when $\big(\frac{\mathcal{L}(1)}{2\delta}\big)^2 \geq 64e$. □

B. Proofs about Fast Binary Embedding Algorithm

B.1. Proof of Lemma 3.7

Proof. It suffices to prove $X \perp Y'$. One can check similarly that the proof holds for the remaining three results. Note that X, Y' are binary random variables with values $\{-1, 1\}$. It is easy to observe both of them are balanced, namely $\Pr(X = 1) = \Pr(Y' = 1) = 1/2$. If $X \perp Y'$, then we have $\Pr(X = Y') = 1/2$. In the reverse direction, suppose $\Pr(X = Y') = 1/2$. First we have

$$\Pr(X = 1) = \Pr(X = 1, Y' = 1) + \Pr(X = 1, Y' = -1) = 1/2, \tag{B.1}$$

$$\Pr(Y' = 1) = \Pr(X = 1, Y' = 1) + \Pr(X = -1, Y' = 1) = 1/2. \tag{B.2}$$

Combining the above two results, we have $\Pr(X = 1, Y' = -1) = \Pr(X = -1, Y' = 1)$. Using $\Pr(X = 1, Y' = -1) + \Pr(X = -1, Y' = 1) = \Pr(X \neq Y') = 1 - \Pr(X = Y') = \frac{1}{2}$, we thus have $\Pr(X = 1, Y' = -1) = \Pr(X = -1, Y' = 1) = 1/4$. Plugging the above result into (B.1) and (B.2) we have $\Pr(X = 1, Y' = 1) = \Pr(X = -1, Y' = -1) = 1/4$. Thus we have shown

$$\Pr(X = v \big| Y' = u) = \frac{\Pr(X = v, Y' = u)}{\Pr(Y' = u)} = \Pr(X = v),\ \forall\, u, v \in \{-1, 1\},$$

which leads to $X \perp Y'$.

Using the above arguments, we show that $X \perp Y'$ if and only if

$$\Pr(X = Y') = 1/2.$$

Recalling the definition of X, Y', the above condition holds if and only if

$$\Pr\Big\{ \underbrace{\langle \boldsymbol{\xi} \odot \boldsymbol{\zeta}, \boldsymbol{x} \rangle \cdot \langle \boldsymbol{\xi}' \odot \boldsymbol{\zeta}, \boldsymbol{y} \rangle}_{Z} \ \geq\ 0 \Big\} = \frac{1}{2}.$$

Next we prove Z has symmetric distribution around 0. Let $\mathcal{I} = [1, n], \mathcal{I}' = [1, n - \Delta], \mathcal{I}_0 = [2n - \Delta, 2n - 1]$ for some natural number $\Delta < n$. Without loss of generality, we assume $\boldsymbol{\xi} = \boldsymbol{g}_{\mathcal{I}}$ and $\boldsymbol{\xi}' = [\boldsymbol{g}_{\mathcal{I}_0}; \boldsymbol{g}_{\mathcal{I}'}]$. We split $\mathcal{I}$ into $T = \lceil \frac{n}{\Delta} \rceil$ consecutive disjoint subsets $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_T$ each of which has size Δ except $|\mathcal{I}_T| = n - (T - 1)\Delta \leq \Delta$. Also, let $\mathcal{I}'_{T-1}$ contain the first $n - (T - 1)\Delta$ entries of $\mathcal{I}_{T-1}$. Then we have

$$Z = \Big(\sum_{i=1}^{T} \big\langle \boldsymbol{g}_{\mathcal{I}_i} \odot \boldsymbol{\zeta}_{\mathcal{I}_i}, \boldsymbol{x}_{\mathcal{I}_i} \big\rangle \Big) \cdot \Big(\sum_{i=1}^{T-2} \big\langle \boldsymbol{g}_{\mathcal{I}_i} \odot \boldsymbol{\zeta}_{\mathcal{I}_{i+1}}, \boldsymbol{y}_{\mathcal{I}_{i+1}} \big\rangle + \big\langle \boldsymbol{g}_{\mathcal{I}'_{T-1}} \odot \boldsymbol{\zeta}_{\mathcal{I}_T}, \boldsymbol{y}_{\mathcal{I}_T} \big\rangle + \big\langle \boldsymbol{g}_{\mathcal{I}_0} \odot \boldsymbol{\zeta}_{\mathcal{I}_1}, \boldsymbol{y}_{\mathcal{I}_1} \big\rangle \Big). \tag{B.3}$$

We now let $\widehat{\boldsymbol{g}}$ be such random vector that is identical to $\boldsymbol{g}$ except that for any $i \in \{0\} \cup [T]$

$$\widehat{\boldsymbol{g}_{\mathcal{I}_i}} = -\boldsymbol{g}_{\mathcal{I}_i},\ \text{if } i \bmod 2 = 0$$

Let $\widehat{\boldsymbol{\zeta}}$ be such random vector that is identical to $\boldsymbol{\zeta}$ except that for any $i \in \{0\} \cup [T]$

$$\widehat{\boldsymbol{\zeta}_{\mathcal{I}_i}} = -\boldsymbol{\zeta}_{\mathcal{I}_i},\ \text{if } i \bmod 2 = 1.$$

Replacing $\boldsymbol{g},\ \boldsymbol{\zeta}$ in (B.3) with $\widehat{\boldsymbol{g}},\ \widehat{\boldsymbol{\zeta}}$ yields

$$\begin{aligned}
&\widehat{Z} \\
=&\Big(\sum_{i=1}^{T} \big\langle \widehat{\boldsymbol{g}}_{\mathcal{I}_i} \odot \widehat{\boldsymbol{\zeta}}_{\mathcal{I}_i}, \boldsymbol{x}_{\mathcal{I}_i} \big\rangle \Big) \cdot \Big(\sum_{i=1}^{T-2} \big\langle \widehat{\boldsymbol{g}}_{\mathcal{I}_i} \odot \widehat{\boldsymbol{\zeta}}_{\mathcal{I}_{i+1}}, \boldsymbol{y}_{\mathcal{I}_{i+1}} \big\rangle + \big\langle \widehat{\boldsymbol{g}}_{\mathcal{I}'_{T-1}} \odot \widehat{\boldsymbol{\zeta}}_{\mathcal{I}_T}, \boldsymbol{y}_{\mathcal{I}_T} \big\rangle + \big\langle \widehat{\boldsymbol{g}}_{\mathcal{I}_0} \odot \widehat{\boldsymbol{\zeta}}_{\mathcal{I}_1}, \boldsymbol{y}_{\mathcal{I}_1} \big\rangle \Big) \\
=&\Big(-\sum_{i=1}^{T} \big\langle \boldsymbol{g}_{\mathcal{I}_i} \odot \boldsymbol{\zeta}_{\mathcal{I}_i}, \boldsymbol{x}_{\mathcal{I}_i} \big\rangle \Big) \cdot \Big(\sum_{i=1}^{T-2} \big\langle \boldsymbol{g}_{\mathcal{I}_i} \odot \boldsymbol{\zeta}_{\mathcal{I}_{i+1}}, \boldsymbol{y}_{\mathcal{I}_{i+1}} \big\rangle + \big\langle \boldsymbol{g}_{\mathcal{I}'_{T-1}} \odot \boldsymbol{\zeta}_{\mathcal{I}_T}, \boldsymbol{y}_{\mathcal{I}_T} \big\rangle + \big\langle \boldsymbol{g}_{\mathcal{I}_0} \odot \boldsymbol{\zeta}_{\mathcal{I}_1}, \boldsymbol{y}_{\mathcal{I}_1} \big\rangle \Big) \\
=& - Z.
\end{aligned}$$

As each entry of $\boldsymbol{g}$ is symmetric random variable around 0, therefore $\widehat{\boldsymbol{g}}$ and $\boldsymbol{g}$ has the same probability distribution. The same fact also holds for $\widehat{\boldsymbol{\zeta}}$ and $\boldsymbol{\zeta}$. So we conclude that Z has symmetric distribution around 0, which implies $\Pr(Z > 0) = \frac{1}{2}$ and $X \perp Y'$. □

B.2. Proof of Theorem 3.8

Proof. Unspecified notations in this section are consistent with Algorithm 2. Using Lemma 3.6, we have

$$\Pr\Big\{ \sup_{i,j \in [N]} \big| d(\boldsymbol{y}_i, \boldsymbol{y}_j) - d(\boldsymbol{x}_i, \boldsymbol{x}_j) \big| \geq C\delta \Big\} \leq 0.01. \tag{B.4}$$

Now consider the first-block binary codes generated from Gaussian Toeplitz projection. We focus on two intermediate points $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$. Consider the first block of binary codes generated from the second part of Algorithm 2. We let

$$\boldsymbol{u} = \text{sign}\left(\Psi^{(1)} \cdot \boldsymbol{y}_1\right), \boldsymbol{v} = \text{sign}\left(\Psi^{(1)} \cdot \boldsymbol{y}_2\right).$$

Suppose $\Psi^{(1)}$ contains Gaussian Toeplitz matrix $\mathbf{T}$. For any $i \in [m/B]$, we have

$$u_i = \text{sign}\left(\left\langle \mathbf{T}_i \odot \boldsymbol{\zeta},\ \boldsymbol{y}_1 \right\rangle\right) = \text{sign}\left(\left\langle \mathbf{T}_i,\ \boldsymbol{y}_1 \odot \boldsymbol{\zeta} \right\rangle\right).$$

$$v_i = \text{sign}\left(\left\langle \mathbf{T}_i \odot \boldsymbol{\zeta},\ \boldsymbol{y}_2 \right\rangle\right) = \text{sign}\left(\left\langle \mathbf{T}_i,\ \boldsymbol{y}_2 \odot \boldsymbol{\zeta} \right\rangle\right).$$

Since $\mathbf{T}_i$ is a Gaussian random vector, we have

$$\Pr(u_i \neq v_i) = d(\boldsymbol{y}_1 \odot \boldsymbol{\zeta},\ \boldsymbol{y}_2 \odot \boldsymbol{\zeta}) = d(\boldsymbol{y}_1,\ \boldsymbol{y}_2).$$

Let $Z_i = \mathbb{1}\big(u_i \neq v_i\big), \forall\ i \in [m/B]$. Following Lemma (3.7), we know that $\forall\ i \neq j$

$$u_i \perp u_j,\ u_i \perp v_j,\ v_i \perp v_j,\ v_i \perp u_j.$$

Therefore $\{Z_i\}_{i=1}^{[m/B]}$ is a pair-wise independent sequence. By Markov's inequality, we have

$$\Pr\left(\Big|\frac{1}{m/B}\sum_{i=1}^{m/B} Z_i - \mathbb{E}(Z_1)\Big| \geq \delta\right) \leq \frac{\frac{B}{m}Var(Z_1)}{\delta^2} \leq \frac{1}{4}\frac{B}{m\delta^2} \leq \frac{1}{4}. \tag{B.5}$$

The last inequality holds by setting $\frac{m}{B} \geq \frac{1}{\delta^2}$. Therefore, we have

$$\Pr\left(\big|d_{\mathcal{H}}(\boldsymbol{u}, \boldsymbol{v}) - d(\boldsymbol{y}_1, \boldsymbol{y}_2)\big| \geq \delta\right) \leq \frac{1}{4}.$$

Now consider total B block binary codes $\{\boldsymbol{u}_i\}_{i=1}^B$ $\{\boldsymbol{v}_i\}_{i=1}^B$ from $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ respectively. Let

$$E_i = \mathbb{1}\big(|d_{\mathcal{H}}(\boldsymbol{u}_i, \boldsymbol{v}_i) - d(\boldsymbol{y}_1, \boldsymbol{y}_2)| \geq \delta\big),\ \forall\ i \in [B].$$

From (B.5), we have $\Pr(E_i = 1) < \frac{1}{4}$. If more than half of E_i are 0, then the median of $\{d_{\mathcal{H}}(\boldsymbol{u}_i, \boldsymbol{v}_i)\}_{i=1}^B$ is within δ away from $d(\boldsymbol{y}_1, \boldsymbol{y}_2)$. Then we have

$$\Pr\left(\text{median}\left(\{d_{\mathcal{H}}(\boldsymbol{u}_i, \boldsymbol{v}_i)\}_{i=1}^B\right) - d(\boldsymbol{y}_1, \boldsymbol{y}_2)\big| \geq \delta\right)$$

$$\leq \Pr\big(\frac{1}{B}\sum_{i=1}^B E_i \geq \frac{1}{2}\big) \leq \Pr\big(\frac{1}{B}\sum_{i=1}^B E_i - \mathbb{E}(E_i) > \frac{1}{4}\big) \leq \exp(-\frac{1}{4}B).$$

In the second inequality, we use (B.5). The last step follows from Hoeffding's inequality. Now we use a union bound for N^2 pairs

$$\Pr\left(\sup_{i,j \in [N]} \big|d_{\mathcal{H}}(\boldsymbol{b}_i, \boldsymbol{b}_j) - d(\boldsymbol{y}_i, \boldsymbol{y}_j)\big| \geq \delta\right) \leq N^2 \exp(-\frac{1}{4}B) \leq \exp(-\frac{1}{8}B).$$

The last inequality holds by setting $B \geq 16 \log N$. Combing the above result and (B.4) using triangle inequality, we complete the proof. □

C. Proof of Theorem 3.10

For any set $K \subseteq \mathbb{S}^{p-1}$, we use $\mathcal{N}_\delta(K)$ to denote a constructed δ-net of K, which is a δ-covering set with minimum size. In particular, by Sudakov's theorem (e.g., Theorem 3.18 in Ledoux & Talagrand (1991))

$$\log \mathcal{N}_\delta(K) \lesssim \frac{w(K)^2}{\delta^2}.$$

We first prove that for a fixed two dimensional space, $m = O(\frac{1}{\delta^2})$ independent Gaussian measurements are sufficient to achieve δ-uniform binary embedding.

Lemma C.1. Suppose K is any fixed two-dimensional subspace in $\mathbb{S}^{p-1}$. Let $\mathbf{A} \in \mathbb{R}^{m\times p}$ be a matrix with independent rows $\mathbf{A}_i \sim \mathcal{N}(0, \mathbf{I}_p),\ \forall i \in [m]$. Suppose $m \geq \frac{1}{\delta^2}\log\frac{1}{\delta}$, then with probability at least $1 - 3\exp(-\delta^2 m)$,

$$\sup_{\boldsymbol{x},\boldsymbol{y}\in K} \left| d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y}) \right| \leq C\delta. \tag{C.1}$$

Here C is some absolute constant.

Proof. We postpone the proof to Appendix D.3. □

The next lemma shows that the normalized ℓ_1 norm of $\mathbf{A}\boldsymbol{x}$ provides decent approximation of $\|\boldsymbol{x}\|_2$.

Lemma C.2. Consider any set $K \subseteq \mathbb{R}^p$. Let $\mathbf{A}$ be an m-by-p matrix with independent rows $\mathbf{A}_i \sim \mathcal{N}(0, \mathbf{I}_p)$ for any $i \in [m]$. Consider

$$Z = \sup_{\boldsymbol{x}\in K} \left| \frac{1}{m}\sum_{i=1}^{m} \left|\left\langle \mathbf{A}_i, \boldsymbol{x}\right\rangle\right| - \sqrt{\frac{2}{\pi}}\|\boldsymbol{x}\|_2 \right|.$$

We have

$$\Pr\left\{Z \geq 4\frac{w(K)}{\sqrt{m}} + t\right\} \leq 2\exp\left(-\frac{mt^2}{2d(K)^2}\right),\ \forall\, t > 0.$$

where $d(K) = \max_{\boldsymbol{x}\in K}\|\boldsymbol{x}\|_2$.

Proof. See the proof of Lemma 2.1 in Plan & Vershynin (2014). □

In order to connect ℓ_1 norm to Hamming distance, we need the following result.

Lemma C.3. Consider finite number of points $K \subseteq \mathbb{S}^{p-1}$. Let $\mathbf{A}$ be an m-by-p matrix with independent rows $\mathbf{A}_i \sim \mathcal{N}(0, \mathbf{I}_p)$ for any $i \in [m]$. Suppose

$$m \geq \frac{1}{\delta^2}\log|K|,$$

then we have

$$\sup_{\boldsymbol{x}\in|K|} \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\left\{ \left|\left\langle \mathbf{A}_i, \boldsymbol{x}\right\rangle\right| \leq \delta \right\} \leq 2\delta.$$

with probability at least $1 - \exp(-\delta^2 m)$.

Proof. Let $X \sim \mathcal{N}(0,1)$. For any fixed point $\boldsymbol{x} \in K$ and any $i \in [m]$, we have

$$\Pr(\left|\left\langle \boldsymbol{A}_i, \boldsymbol{x}\right\rangle\right| \leq \delta) = \Pr(|X| \leq \delta) \leq \delta.$$

Let $Z_i = \mathbb{1}(\left|\left\langle \boldsymbol{A}_i, \boldsymbol{x}\right\rangle\right| \leq \delta),\ \forall\, i \in [m]$. Then by using Hoeffding's inequality,

$$\Pr(\frac{1}{m}\sum_{i=1}^{m} Z_i - \mathbb{E}(Z_1) > \delta) \leq \exp(-2\delta^2 m).$$

As $\mathbb{E}(Z_1) = \Pr(\left|\left\langle \boldsymbol{A}_i, \boldsymbol{x}\right\rangle\right| \leq \delta) \leq \delta$, we conclude that with probability at least $1 - \exp(-2\delta^2 m)$,

$$\frac{1}{m}\sum_{i=1}^{m} Z_i \leq 2\delta.$$

By applying union bound over $|K|$ points and setting $m \geq \frac{1}{\delta^2}\log|K|$, we complete the proof. □

Now we are ready to prove Theorem 3.10.

Proof of Theorem 3.10. We construct a δ-net of K that is denoted as $\mathcal{N}_\delta$. We assume $m \gtrsim \frac{1}{\delta^2}\log|\mathcal{N}_\delta|$. Applying Proposition 2.2 and setting $K = \mathcal{N}_\delta$, we have that

$$\sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{N}_\delta} \big| d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y})\big| \le \delta \tag{C.2}$$

with probability at least $1 - 2\exp(-\delta^2 m)$.

For any two fixed points $\boldsymbol{x},\boldsymbol{y}\in K$, let $\boldsymbol{x}_1,\boldsymbol{y}_1$ be their nearest points in $\mathcal{N}_\delta$. Then we have

$$\begin{aligned}
&|d(\boldsymbol{x},\boldsymbol{y}) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| \le |d(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x}_1,\boldsymbol{y}_1)| + |d(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| \\
&\overset{(a)}{\le} |d(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 2\delta \le |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + |d(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1)| + 2\delta \\
&\overset{(b)}{\le} |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 3\delta \le |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y})| + |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 3\delta \\
&\overset{(c)}{\le} d_{\mathbf{A}}(\boldsymbol{y}_1,\boldsymbol{y}) + d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{x}) + 3\delta,
\end{aligned} \tag{C.3}$$

where (a) follows from

$$|d(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x}_1,\boldsymbol{y}_1)| \le |d(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x}_1,\boldsymbol{y})| + |d(\boldsymbol{x}_1,\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y}_1)| \le d(\boldsymbol{x},\boldsymbol{x}_1) + d(\boldsymbol{x}_1,\boldsymbol{y}_1) \le 2\delta,$$

step (b) follows from (C.2), step (c) follows from the triangle inequality of Hamming distance. Therefore we have

$$\sup_{\boldsymbol{x},\boldsymbol{y}\in K} \big| d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y})\big| \le 2 \sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \sup_{\substack{\boldsymbol{x}\in K \\ \|\boldsymbol{x}-\boldsymbol{x}_1\|_2\le\delta}} d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{x}_1) + 3\delta. \tag{C.4}$$

Next we bound the tail term

$$T := \sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \sup_{\substack{\boldsymbol{x}\in K \\ \|\boldsymbol{x}-\boldsymbol{x}_1\|_2\le\delta}} d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{x}_1).$$

Recall that

$$K_\delta^+ := K \bigcup \Big\{\boldsymbol{z}\in\mathbb{S}^{p-1} : \boldsymbol{z} = \frac{\boldsymbol{x}-\boldsymbol{y}}{\|\boldsymbol{x}-\boldsymbol{y}\|_2},\ \forall\, \boldsymbol{x},\boldsymbol{y}\in K \text{ if } \delta^2 \le \|\boldsymbol{x}-\boldsymbol{y}\|_2 \le \delta\Big\}.$$

Now we construct a δ-net for $K_\delta^+ \setminus K$ denoted as $\mathcal{N}'_\delta$. For two distinct points $\boldsymbol{x},\boldsymbol{y}\in\mathcal{N}'_\delta\bigcup\mathcal{N}_\delta$, let $C(\boldsymbol{x},\boldsymbol{y})$ denote the unit circle spanned by $\boldsymbol{x},\boldsymbol{y}$. We construct δ^2-net $\mathcal{C}_{\delta^2}(\boldsymbol{x},\boldsymbol{y})$ for each circle $\mathcal{C}(\boldsymbol{x},\boldsymbol{y})$. For simplicity, we just let $\mathcal{C}_{\delta^2}(\boldsymbol{x},\boldsymbol{y})$ be the set of points that uniformly split $\mathcal{C}(\boldsymbol{x},\boldsymbol{y})$ with interval δ^2. We thus have $|\mathcal{C}_{\delta^2}(\boldsymbol{x},\boldsymbol{y})| \lesssim \frac{1}{\delta^2}$. Let $\mathcal{G}_\delta$ denote the union of all circle nets $\mathcal{C}_{\delta^2}(\boldsymbol{x},\boldsymbol{y})$ spanned by points in $\mathcal{N}'_\delta\bigcup\mathcal{N}_\delta$, namely

$$\mathcal{G}_\delta := \bigcup_{\forall\, \boldsymbol{x},\boldsymbol{y}\in\mathcal{N}'_\delta\bigcup\mathcal{N}_\delta} \mathcal{C}_{\delta^2}(\boldsymbol{x},\boldsymbol{y}) \cup \{\boldsymbol{x},\boldsymbol{y}\}.$$

For any point $\boldsymbol{x}\in K$, we can always find a point in $\mathcal{G}_\delta$ that is $O(\delta^2)$ away from $\boldsymbol{x}$. To see why the argument is true, we first let $\boldsymbol{x}_1$ be the nearest point to $\boldsymbol{x}$ in $\mathcal{N}_\delta$. If $\|\boldsymbol{x}-\boldsymbol{x}_1\|_2 \le \delta^2$, then $\boldsymbol{x}_1$ is the point we want. Otherwise, we have $\delta^2 \le \|\boldsymbol{x}-\boldsymbol{x}_1\|_2 \le \delta$. In this case, we have $(\boldsymbol{x}-\boldsymbol{x}_1)/\|\boldsymbol{x}-\boldsymbol{x}_1\| \in K^+$. Following the definition of K_δ^+, we can always find a point $\boldsymbol{x}'_1 \in \mathcal{N}'_\delta\bigcup\mathcal{N}_\delta$ such that

$$\Big\|\boldsymbol{x}'_1 - \frac{\boldsymbol{x}-\boldsymbol{x}_1}{\|\boldsymbol{x}-\boldsymbol{x}_1\|_2}\Big\|_2 \le \delta, \tag{C.5}$$

thereby

$$\big\|\boldsymbol{x} - \underbrace{\big(\|\boldsymbol{x}-\boldsymbol{x}_1\|_2\boldsymbol{x}'_1 + \boldsymbol{x}_1\big)}_{\boldsymbol{z}}\big\|_2 \le \delta\|\boldsymbol{x}-\boldsymbol{x}_1\|_2 \le \delta^2.$$

Note that $\|\boldsymbol{z}\|_2$ is very close to 1 because

$$\delta^4 \ge \|\boldsymbol{x}-\boldsymbol{z}\|_2^2 \ge \|\boldsymbol{z}\|_2^2 - 2\big\langle\boldsymbol{z},\boldsymbol{x}\big\rangle + 1 \ge \|\boldsymbol{z}\|_2^2 - 2\|\boldsymbol{z}\|_2 + 1 = (\|\boldsymbol{z}\|_2 - 1)^2.$$

We thus have

$$\big\|\boldsymbol{x}-\boldsymbol{z}/\|\boldsymbol{z}\|_2\big\|_2 \le \|\boldsymbol{x}-\boldsymbol{z}\|_2 + \big\|\boldsymbol{z}-\boldsymbol{z}/\|\boldsymbol{z}\|_2\big\|_2 = \|\boldsymbol{x}-\boldsymbol{z}\|_2 + \big|\|\boldsymbol{z}\|_2 - 1\big| \le 2\delta^2.$$

Note that $\boldsymbol{z}$ is in the unit circle $\mathcal{C}(\boldsymbol{x},\boldsymbol{x}_1')$ spanned by $\boldsymbol{x}$ and $\boldsymbol{x}_1'$, thereby there exists $\boldsymbol{u}\in\mathcal{C}_{\delta^2}(\boldsymbol{x}_1,\boldsymbol{x}_1')$ such that $\|\boldsymbol{u}-\boldsymbol{x}\|_2 \le \delta^2$. Point $\boldsymbol{u}$ thus satisfies

$$\|\boldsymbol{x}-\boldsymbol{u}\| \le \|\boldsymbol{x}-\boldsymbol{z}\|_2 + \|\boldsymbol{z}-\boldsymbol{u}\|_2 \le 3\delta^2. \tag{C.6}$$

So for any $\boldsymbol{x}\in K$ and its nearest point $\boldsymbol{x}_1\in\mathcal{N}_\delta$, we define $\boldsymbol{u}$ as

$$\boldsymbol{u} := \begin{cases} \boldsymbol{x}_1, & \|\boldsymbol{x}-\boldsymbol{x}_1\|_2 \le \delta^2; \\ \operatorname{argmin}_{\boldsymbol{v}\in\mathcal{C}_{\delta^2}(\boldsymbol{x}_1,\boldsymbol{x}_1')} \|\boldsymbol{x}-\boldsymbol{v}\|_2, & \text{otherwise.} \end{cases}$$

where $\boldsymbol{x}_1'\in\mathcal{N}_\delta\bigcup\mathcal{N}_\delta'$ and satisfies (C.5). Based on (C.6), we always have $\|\boldsymbol{u}-\boldsymbol{x}\|_2\le 3\delta^2$ and $\|\boldsymbol{u}-\boldsymbol{x}_1\|_2 \le \|\boldsymbol{u}-\boldsymbol{x}\|_2 + \|\boldsymbol{x}-\boldsymbol{x}_1\|_2 \le 2\delta$.

By triangle inequality of Hamming distance,

$$d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{x}_1) \le d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{u}) + d_{\mathbf{A}}(\boldsymbol{u},\boldsymbol{x}_1).$$

We thus have

$$\begin{aligned} T &\le \sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{x}_1\|_2}} d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{u}) + d_{\mathbf{A}}(\boldsymbol{u},\boldsymbol{x}_1) \\ &\le \underbrace{\sup_{\boldsymbol{u}\in\mathcal{G}_\delta} \sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{u}\|_2\le 3\delta^2}} d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{u})}_{T_1} + \underbrace{\sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{N}_\delta\bigcup\mathcal{N}_\delta'} \sup_{\substack{\boldsymbol{u},\boldsymbol{v}\in\mathcal{C}(\boldsymbol{x},\boldsymbol{y})\\ \|\boldsymbol{u}-\boldsymbol{v}\|_2\le 2\delta}} d_{\mathbf{A}}(\boldsymbol{u},\boldsymbol{v})}_{T_2}. \end{aligned}$$

Next we bound term T_1 and T_2 respectively.

Term T_1. For a fixed point $\boldsymbol{u}\in\mathcal{G}_\delta$, using Lemma C.2 by setting (K,t) in the statement to be $K' = (K-\{\boldsymbol{u}\})\bigcap\{\boldsymbol{u}\in\mathbb{R}^p : \|\boldsymbol{u}\|_2\le 3\delta^2\}$ and δ^2 respectively yields that

$$\begin{aligned} &\Pr\Bigg\{ \sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{u}\|_2\le 3\delta^2}} \Bigg| \frac{1}{m}\sum_{i=1}^m \big|\langle \mathbf{A}_i, \boldsymbol{x}-\boldsymbol{u}\rangle\big| - \sqrt{\frac{2}{\pi}}\|\boldsymbol{x}-\boldsymbol{u}\|_2 \Bigg| \ge \frac{4w(K')}{\sqrt{m}} + \delta^2 \Bigg\} \\ \le & 2\exp\big(-\frac{m\delta^4}{2d(K')^2}\big) \le 2\exp(-m/18). \end{aligned}$$

Then with probability greater than $1-2\exp(-m/18)$,

$$\sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{u}\|_2\le 3\delta^2}} \frac{1}{m}\sum_{i=1}^m \big|\langle \mathbf{A}_i, \boldsymbol{x}-\boldsymbol{u}\rangle\big| \le 3\sqrt{\frac{2}{\pi}}\delta^2 + 4w(K')/\sqrt{m} + \delta^2 \le 5\delta^2,$$

where the last inequality follows from the fact that $w(K')\lesssim w(K)$ and our assumption $m\gtrsim w(K)^2/\delta^4$. We define event

$$\mathcal{E} := \Bigg\{ \sup_{\boldsymbol{u}\in\mathcal{G}_\delta} \sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{u}\|_2\le 3\delta^2}} \frac{1}{m}\sum_{i=1}^m \big|\langle \mathbf{A}_i, \boldsymbol{x}-\boldsymbol{u}\rangle\big| \le 5\delta^2 \Bigg\}.$$

Applying union bound over all points in $\mathcal{G}_\delta$, we have

$$\Pr(\mathcal{E}^c) \le 2|\mathcal{G}_\delta|\exp(-m/18) \le 2\exp(-m/36),$$

where the last inequality holds with $m\gtrsim \log|\mathcal{G}_\delta|$. Under condition event $\mathcal{E}$ happens, we have

$$\sup_{\boldsymbol{u}\in\mathcal{G}_\delta} \sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{u}\|_2\le 3\delta^2}} \frac{1}{m}\sum_{i=1}^m \mathbb{1}\Big\{ \big|\langle \mathbf{A}_i, \boldsymbol{u}-\boldsymbol{x}\rangle\big| \le 5\delta \Big\} \ge 1-\delta. \tag{C.7}$$

If sign $\big(\langle \mathbf{A}_i, \boldsymbol{u}\rangle\big) \neq$ sign $\big(\langle \mathbf{A}_i, \boldsymbol{x}\rangle\big)$, we must have $\big|\big\langle \mathbf{A}_i, \boldsymbol{u}\big\rangle\big| \leq \big|\big\langle \mathbf{A}_i, \boldsymbol{u}-\boldsymbol{x}\big\rangle\big|$. We then have

$$T_1 \leq \sup_{\boldsymbol{u}\in\mathcal{G}_\delta} \sup_{\substack{\boldsymbol{x}\in K\\ \|\boldsymbol{x}-\boldsymbol{u}\|_2\leq 3\delta^2}} \frac{1}{m}\sum_{i=1}^m \mathbb{1}\Big\{\big|\big\langle \mathbf{A}_i, \boldsymbol{u}\big\rangle\big| \leq \big|\big\langle \mathbf{A}_i, \boldsymbol{u}-\boldsymbol{x}\big\rangle\big|\Big\}$$
$$\leq \sup_{\boldsymbol{u}\in\mathcal{G}_\delta} \frac{1}{m}\sum_{i=1}^m \mathbb{1}\Big\{\big|\big\langle \mathbf{A}_i, \boldsymbol{u}\big\rangle\big| \leq 5\delta\Big\} + \delta,$$

where the last inequality follows from (C.7). Using Lemma C.2 by setting K and δ in the statement to be $\mathcal{G}_\delta$ and 5δ respectively, we have that, when $m \geq c\frac{1}{\delta^2}\log|\mathcal{G}_\delta|$ with some absolute constant c, the following inequality

$$\sup_{\boldsymbol{u}\in\mathcal{G}_\delta} \frac{1}{m}\sum_{i=1}^m \mathbb{1}\Big\{\big|\big\langle \mathbf{A}_i, \boldsymbol{u}\big\rangle\big| \leq 5\delta\Big\} \leq 10\delta$$

holds with probability at least $1-\exp(-25\delta^2 m)$. Putting all ingredients together, we have $T_1 \leq 11\delta$ with high probability.

Term T_2. There are at most $|\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'|^2$ different two-dimensional subspaces constructed from $\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'$. Applying Lemma C.1 and probabilistic union bound over all subspaces yields that

$$\Pr\bigg(T_2 \geq (C+2)\delta\bigg) \leq 3\big|\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'\big|^2 \exp(-\delta^2 m) \leq 3\exp(-\delta^2 m/2),$$

where the last inequality holds by setting $m \gtrsim \frac{1}{\delta^2}\log|\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'|$.

Putting (C.4) and the upper bounds of term T together, we conclude that by choosing

$$m \gtrsim \max\bigg\{w(K)^2/\delta^4,\ \log|\mathcal{G}_\delta|,\ \frac{1}{\delta^2}\log|\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'|\bigg\},$$

we have

$$\sup_{\boldsymbol{x},\boldsymbol{y}\in K} |d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y})| \lesssim \delta.$$

with probability at least $1-c_1\exp(-c_2\delta^2 m)$ where c_1, c_2 are some absolute constants.

Using the fact that

$$|\mathcal{G}_\delta| \lesssim \frac{1}{\delta^2}|\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'|$$

and

$$\log|\mathcal{N}_\delta \bigcup \mathcal{N}_\delta'| \lesssim \frac{1}{\delta^2} w(\mathcal{N}_\delta \bigcup \mathcal{N}_\delta')^2 \leq \frac{1}{\delta^2} w(K_\delta^+)^2,$$

we complete the proof. □

D. Proofs of Supporting Lemmas

D.1. Proof of Lemma 3.6

Proof. Recall that $\boldsymbol{y}_i = \sqrt{\frac{p}{m}}\Phi(\boldsymbol{\zeta})\cdot\boldsymbol{x}_i$. We let

$$\widehat{\boldsymbol{y}_i} = \frac{\boldsymbol{y}_i}{\|\boldsymbol{y}_i\|_2}, \widehat{\boldsymbol{y}_j} = \frac{\boldsymbol{y}_j}{\|\boldsymbol{y}_j\|_2}.$$

From condition (3.6), we have

$$\|\boldsymbol{y}_i - \widehat{\boldsymbol{y}_i}\|_2 \leq \delta, \|\boldsymbol{y}_j - \widehat{\boldsymbol{y}_j}\|_2 \leq \delta. \tag{D.1}$$

Let $\theta = \angle(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\theta' = \angle(\widehat{\boldsymbol{y}_i}, \widehat{\boldsymbol{y}_j})$. Without loss of generality, we assume our set $K = \{\boldsymbol{x}_i\}_{i=1}^N$ is symmetric, i.e., if $\boldsymbol{x}\in K$ then $-\boldsymbol{x}\in \boldsymbol{K}$. Suppose we show for any two points $\boldsymbol{x}_i, \boldsymbol{x}_j$ with $\big\langle \boldsymbol{x}_i, \boldsymbol{x}_j\big\rangle > 0$, inequality (3.8) holds, then for $\boldsymbol{x}_i, \boldsymbol{x}_j$ with $\big\langle \boldsymbol{x}_i, \boldsymbol{x}_j\big\rangle < 0$, we immediately have

$$\big|d(\boldsymbol{y}_i,\boldsymbol{y}_j) - d(\boldsymbol{x}_i,\boldsymbol{x}_j)\big| = \big|1 - d(\boldsymbol{y}_i,\boldsymbol{y}_j) - \big(1 - d(\boldsymbol{x}_i,\boldsymbol{x}_j)\big)\big| = \big|d(-\boldsymbol{y}_i,\boldsymbol{y}_j) - d(-\boldsymbol{x}_i,\boldsymbol{x}_j)\big| \leq C\delta.$$

In the second equality, we use $d(-\boldsymbol{x},\boldsymbol{y})+d(\boldsymbol{x},\boldsymbol{y})=1,\ \forall\ \boldsymbol{x},\boldsymbol{y}\in\mathbb{S}^{p-1}$. In the last inequality, we use the fact that fast JL transform $\sqrt{\frac{p}{m}}\Phi(\boldsymbol{\zeta})$ is linear thus $-\boldsymbol{y}_i=\sqrt{\frac{p}{m}}\Phi(\boldsymbol{\zeta})(-\boldsymbol{x}_i)$. Therefore, without loss of generality, we assume $\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle\geq 0$ thus $\theta\leq\frac{\pi}{2}$.

Now we turn to the following quantity

$$\begin{aligned}\left\|\widehat{\boldsymbol{y}_i}-\widehat{\boldsymbol{y}_j}\right\|_2&=\left\|\widehat{\boldsymbol{y}_i}-\boldsymbol{y}_i+\boldsymbol{y}_i-\boldsymbol{y}_j+\boldsymbol{y}_j-\widehat{\boldsymbol{y}_j}\right\|_2\\&\leq\left\|\widehat{\boldsymbol{y}_i}-\boldsymbol{y}_i\right\|_2+\left\|\widehat{\boldsymbol{y}_j}-\boldsymbol{y}_j\right\|_2+\left\|\boldsymbol{y}_i-\boldsymbol{y}_j\right\|_2\leq 2\delta+\|\boldsymbol{x}_i-\boldsymbol{x}_j\|_2(1+\delta).\end{aligned}$$

The last inequality follows from (D.1) and condition (3.7). Similarly, we also have

$$\left\|\widehat{\boldsymbol{y}_i}-\widehat{\boldsymbol{y}_j}\right\|_2\geq\|\boldsymbol{x}_i-\boldsymbol{x}_j\|(1-\delta)-2\delta.$$

Using the fact that

$$\sin\frac{\theta'}{2}=\frac{\left\|\widehat{\boldsymbol{y}_i}-\widehat{\boldsymbol{y}_j}\right\|_2}{2},\sin\frac{\theta}{2}=\frac{\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|_2}{2},$$

we have

$$\big|\sin\frac{\theta'}{2}-\sin\frac{\theta}{2}\big|=\big|\frac{\left\|\widehat{\boldsymbol{y}_i}-\widehat{\boldsymbol{y}_j}\right\|_2}{2}-\frac{\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|_2}{2}\big|\leq\delta+\delta\frac{\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|_2}{2}\leq 2\delta.$$

When $\delta<\frac{\sqrt{3}-\sqrt{2}}{4}$, we have

$$\sin\frac{\theta'}{2}\leq\sin\frac{\theta}{2}+\frac{\sqrt{3}-\sqrt{2}}{2}\leq\frac{\sqrt{3}}{2}.$$

In the last inequality, we use $\sin\frac{\theta}{2}\leq\frac{\sqrt{2}}{2},\ \forall\ \theta\in[0,\pi/2]$. So $\theta'/2\in[0,\pi/3]$. Using the fact that, for any two $\theta,\theta'\in[0,\pi/3]$, there exists constant c such that

$$\left|\sin\theta-\sin\theta'\right|\geq c\left|\theta-\theta'\right|,$$

we have that

$$\big|\frac{\theta}{2}-\frac{\theta'}{2}\big|\leq\frac{1}{c}\big|\sin\frac{\theta'}{2}-\sin\frac{\theta}{2}\big|\leq\frac{2\delta}{c}.$$

Therefore,

$$\left|d(\boldsymbol{y}_i,\boldsymbol{y}_j)-d(\boldsymbol{x}_i,\boldsymbol{x}_j)\right|=\frac{1}{\pi}\left|\theta-\theta'\right|\leq C\delta.$$

In the case $\delta>\frac{\sqrt{3}-\sqrt{2}}{4}$, trivially we have $\left|d(\boldsymbol{y}_i,\boldsymbol{y}_j)-d(\boldsymbol{x}_i,\boldsymbol{x}_j)\right|\leq 2\leq C\delta$ with constant $C=\frac{8}{\sqrt{3}-\sqrt{2}}$. □

D.2. Proof of Lemma A.4

Proof. For positive semidefinite matrix $\mathbf{M}\in\mathbb{R}^{p\times p}$ with rank d, let $\lambda_1,\lambda_2,...\lambda_d$ be its positive eigenvalues. Using the definition of Frobenius norm, we have

$$\|\mathbf{M}\|_F^2=\sum_{i=1}^{d}\lambda_i^2=\sum_{i,j\in[n]}(\mathbf{M}_{i,j})^2\leq p+(p^2-p)\delta_2^2.$$

On the other hand, considering the trace of $\mathbf{M}$, we can obtain

$$\sum_{i=1}^{d}\lambda_i=\mathrm{Trace}(\mathbf{M})\geq p(1-\delta_1).\tag{D.2}$$

Using Cauchy-Schwarz inequality, we have

$$(\sum_{i=1}^{d}\lambda_i)^2\leq d\sum_{i=1}^{d}\lambda_i^2.\tag{D.3}$$

Plugging (D.2) and (D.3) into the above inequality yields

$$d\geq\frac{p(1-\delta_1)^2}{1+(p-1)\delta_2^2}.$$

□

D.3. Proof of Lemma C.1

Proof. Without loss of any generality, we assume $K = \{\boldsymbol{x} \in \mathbb{S}^{p-1} : \text{supp}(\boldsymbol{x}) \subseteq \{1, 2\}\}$. We begin with constructing a δ-net denoted as $\mathcal{N}_\delta$ for set K. For simplicity, we can just let $\mathcal{N}_\delta(K)$ be the set of points that split the circle spanned by $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ uniformly. Therefore $|\mathcal{N}_\delta(K)| = O(\frac{1}{\delta})$. Applying Proposition 2.2 gives us

$$\sup_{\boldsymbol{x},\boldsymbol{y}\in\mathcal{N}_\delta} |d_A(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y})| \leq \delta, \tag{D.4}$$

holds with probability at least $1 - 2\exp(-\delta^2 m)$ when $m \gtrsim \frac{1}{\delta^2}\log(\frac{1}{\delta})$.

For any point $\boldsymbol{x} \in K$, $\langle \mathbf{A}_i, \boldsymbol{x}\rangle$ only depends on the first two coordinates of $\mathbf{A}_i$. Therefore, for simplicity, we let $\mathbf{A}'_i = \frac{\mathbf{A}_i \odot (\boldsymbol{e}_1 + \boldsymbol{e}_2)}{\|\mathbf{A}_i \odot (\boldsymbol{e}_1 + \boldsymbol{e}_2)\|_2}$, $\forall\, i \in [m]$. For any point say $\boldsymbol{x}_1 \in \mathcal{N}_\delta$, using the uniform distribution of $\mathbf{A}'_i$, we have

$$\Pr(\left|\langle \mathbf{A}'_i, \boldsymbol{x}_1\rangle\right| \leq \delta) \lesssim C\delta,$$

holds with some absolute constant C. Using Hoeffding's inequality and probabilistic union bound over all points in $\mathcal{N}_\delta$, we have

$$\Pr\left(\sup_{\boldsymbol{x}\in\mathcal{N}_\delta} \frac{1}{m}\sum_{i=1}^m \mathbb{1}\{\left|\langle \mathbf{A}_i, \boldsymbol{x}\rangle\right| \leq \delta\} > (C+1)\delta\right) \leq |\mathcal{N}_\delta|\exp(-2\delta^2 m) \leq \exp(-\delta^2 m). \tag{D.5}$$

The last inequality holds when $m \gtrsim \frac{1}{\delta^2}\log\frac{1}{\delta}$.

Now we consider any point $\boldsymbol{x} \in K$. Suppose $\boldsymbol{x}_1$ is the closest point to $\boldsymbol{x}$ in $\mathcal{N}_\delta$. We note that if $\text{sign}\left(\langle \mathbf{A}'_i, \boldsymbol{x}\rangle\right) \neq \text{sign}\left(\langle \mathbf{A}'_i, \boldsymbol{x}_1\rangle\right)$, then there exists $\lambda \in [0, 1]$ such that

$$\langle \mathbf{A}'_i, \lambda\boldsymbol{x} + (1-\lambda)\boldsymbol{x}_1\rangle = 0.$$

We thus have

$$\left|\langle \mathbf{A}'_i, \boldsymbol{x}_1\rangle\right| = \lambda\left|\langle \mathbf{A}'_i, \boldsymbol{x} - \boldsymbol{x}_1\rangle\right| \leq \lambda\|\boldsymbol{x} - \boldsymbol{x}_1\|_2 \leq \delta.$$

Further we obtain that

$$\begin{aligned}
\sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \sup_{\substack{\boldsymbol{x}\in K \\ \|\boldsymbol{x}-\boldsymbol{x}_1\|_2\leq\delta}} d_{\mathbf{A}}(\boldsymbol{x}, \boldsymbol{x}_1) &= \sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \sup_{\substack{\boldsymbol{x}\in K \\ \|\boldsymbol{x}-\boldsymbol{x}_1\|_2\leq\delta}} \frac{1}{m}\sum_{i=1}^m \mathbb{1}(\text{sign}\left(\langle \mathbf{A}'_i, \boldsymbol{x}\rangle\right) \neq \text{sign}\left(\langle \mathbf{A}'_i, \boldsymbol{x}_1\rangle\right)) \\
&\leq \sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \frac{1}{m}\sum_{i=1}^m \mathbb{1}\{\left|\langle \mathbf{A}_i, \boldsymbol{x}_1\rangle\right| \leq \delta\}.
\end{aligned}$$

Combining the above result with (D.5), we obtain that, with probability at least $1 - \exp(-\delta^2 m)$,

$$\sup_{\boldsymbol{x}_1\in\mathcal{N}_\delta} \sup_{\substack{\boldsymbol{x}\in K \\ \|\boldsymbol{x}-\boldsymbol{x}_1\|_2\leq\delta}} d_{\mathbf{A}}(\boldsymbol{x}, \boldsymbol{x}_1) \leq (C+1)\delta. \tag{D.6}$$

For any points $\boldsymbol{x}, \boldsymbol{y} \in K$, let $\boldsymbol{x}_1, \boldsymbol{y}_1$ be their nearest points in $\mathcal{N}_\delta$. We have

$$\begin{aligned}
&|d(\boldsymbol{x},\boldsymbol{y}) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| \leq |d(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x}_1,\boldsymbol{y}_1)| + |d(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| \\
&\overset{(a)}{\leq} |d(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 2\delta \leq |d(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1)| + |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 2\delta \\
&\overset{(b)}{\leq} |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 3\delta \leq |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}_1) - d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y})| + |d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{y}) - d_{\mathbf{A}}(\boldsymbol{x},\boldsymbol{y})| + 3\delta \\
&\overset{(c)}{\leq} d_{\mathbf{A}}(\boldsymbol{y}_1,\boldsymbol{y}) + d_{\mathbf{A}}(\boldsymbol{x}_1,\boldsymbol{x}) + 3\delta \overset{(d)}{\leq} (2C+5)\delta,
\end{aligned}$$

where (a) follows from

$$|d(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x}_1,\boldsymbol{y}_1)| \leq |d(\boldsymbol{x},\boldsymbol{y}) - d(\boldsymbol{x}_1,\boldsymbol{y})| + |d(\boldsymbol{x}_1,\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{y}_1)| \leq d(\boldsymbol{x},\boldsymbol{x}_1) + d(\boldsymbol{x}_1,\boldsymbol{y}_1) \leq 2\delta,$$

step (b) follows from (D.4), step (c) follows from the triangle inequality of Hamming distance, step (d) is from (D.6). □