Gap Hamming Distance

Tsung-Ju Chiang

April 29, 2022

The Gap Hamming Distance Problem

The *gap Hamming distance* problem is a partial function

$$\mathrm{GHD}_n(x, y) := \begin{cases} -1 & \text{if } \langle x, y \rangle \leq -\sqrt{n}, \\ +1 & \text{if } \langle x, y \rangle \geq +\sqrt{n}. \end{cases}$$

where $x, y \in \{-1, +1\}^n$.

Linear lower bound

Theorem 1

$$D^{\mu}(\mathsf{GHD}_n) = \Omega(n).$$

We'll present the proof from

Alexander A. Sherstov. "The Communication Complexity of Gap Hamming Distance". In: *Theory of Computing* 8.8 (2012), pp. 197–208. DOI: 10.4086/toc.2012.v008a008. URL: http://www.theoryofcomputing.org/articles/v008a008

Corruption bound

Fix $f : X \times Y \xrightarrow{(partial)} \{-1, +1\}$ and μ a distribution on $X \times Y$.
We say a rectangle $R \subseteq X \times Y$ is ε-*corrupt* if

$$\mu(R \cap f^{-1}(+1)) > \varepsilon \mu(R \cap f^{-1}(-1)).$$

Theorem 2 (Corruption Bound)

If every rectangle R with $\mu(R) > \delta$ is ε-corrupt, then

$$2^{D_\xi^\mu(f)} \geq \frac{1}{\delta}\left(\mu(f^{-1}(-1)) - \frac{\xi}{\varepsilon}\right).$$

Plan

We'll use the corruption bound to prove the $\Omega(n)$ lower bound.

Fix μ to be the uniform distribution.

Let $R = A \times B$ be a rectangle that's <u>not</u> ε-corrupt. Then

$$\Pr_{x \in A, y \in B}[f(x,y) = +1] \leq \frac{|R \cap f^{-1}(+1)|}{|R \cap f^{-1}(-1)|} < \varepsilon. \qquad (1)$$

We'll show (1) implies that R must be small, i.e.

$$\mu(R) = 4^{-n}|A||B| \leq \delta = 2^{-\Omega(n)}.$$

Then by the corruption bound, have

$$D_\xi^\mu(f_n) \geq \Omega(n) \log\left(\mu(f^{-1}(-1)) - \frac{\xi}{\varepsilon}\right).$$

Gap orthogonality

However, GHD does have a large uncorrupted rectangle.

Instead of working on GHD directly, we'll use a similar function called *gap orthogonality*:

$$f_n(x, y) = \begin{cases} -1 & \text{if } |\langle x, y\rangle| \leq \sqrt{n}/8, \\ +1 & \text{if } |\langle x, y\rangle| \geq \sqrt{n}/4. \end{cases}$$

Observe that $f_n(x, y)$ can be computed using 2 calls to the GHD function, so lower bound f is also a lower bound for GHD.

Theorem 3

Corruption bound requires proving the following:

Theorem 3

Let $R = A \times B$ *s.t.* $\Pr_{x \in A, y \in B}[|\langle x, y\rangle| \le \frac{\sqrt{n}}{4}] \ge 1 - \varepsilon$. *Then* $4^{-n}|A||B| \le exp(-\Omega(n))$.

Proof of theorem 3

The goal is to show that $4^{-n}|A||B| \leq exp(-\Omega(n))$.

If $|A|$ is small enoungh by it self, e.g. $2^{-n}|A| \leq 2 \cdot 2^{-\alpha n}$ for some constant α, then we're done.

Therefore, we'll assume that $|A| > 2 \cdot 2^{(1-\alpha)n}$, and show

$$2^{-n}|B| \leq e^{-\Omega(n)}.$$

.

Proof of theorem 3

Recall that we have

$$\Pr_{x \in A, y \in B}[|\langle x, y\rangle| \leq \frac{\sqrt{n}}{4}] \geq 1 - \varepsilon.$$

We may further assume that for every $x \in A$,

$$\Pr_{y \in B}[|\langle x, y\rangle| \leq \frac{\sqrt{n}}{4}] \geq 1 - 2\varepsilon \tag{2}$$

by discarding violating elements.

This decreases the size of A by at most half, so now $|A| > 2^{(1-\alpha)n}$.

Proof of theorem 3

Next, we'll show that there's some $x_1, \ldots, x_k \in A$ s.t.

$$\Pr_{y \in \{-1,+1\}^n} \left[\max_{i \in [k]} |\langle x_i, y \rangle| \leq \frac{\sqrt{n}}{4} \right] \leq e^{-\Omega(n)}$$

where $k = \Theta(n)$.

Lemma 1

Assume that A is large, then it's always possible to find $k = \lfloor n/10 \rfloor$ vectors from A that are "almost orthogonal".

Lemma 1

Let α be a sufficiently small constant. Fix $A \subseteq \{-1,+1\}^n$ with $|A| > 2^{-\alpha n}$. Then for $k = \lfloor n/10 \rfloor$ there exist $x_1, x_2, \dots, x_k \in A$ such that for each i,

$$\| \operatorname{proj}_{\operatorname{span}\{x_1, x_2, \dots, x_i\}} x_{i+1} \| \leq \frac{\sqrt{n}}{3}. \tag{3}$$

Proof of lemma 1 (and lemma 2) relies on the following:

Fact 1 (Talagrand)

For every linear subspace $V \subseteq R^n$ and every $t > 0$, one has

$$\Pr_{x \in \{-1,+1\}^n} [\| \mathrm{proj}_V x\| - \sqrt{\dim V} > t] \leq 4e^{-ct^2},$$

where $c > 0$ is an absolute constant.

Proof of lemma 1

The proof is by induction.

Having selected $x_1, \ldots, x_i \in A$, pick $x_{i+1} \in \{-1, +1\}^n$ uniformly random. Then

$$\Pr_{x_{i+1}}[x_{i+1} \in A] > 2^{-\alpha n}.$$

Fact 1 implies that

$$\Pr_{x_{i+1}}\left[\|\operatorname{proj}_{\{x_1,\ldots,x_i\}} x_{i+1}\| \leq \frac{\sqrt{n}}{3}\right] \geq 1 - 2^{-\alpha n}.$$

Hence, there exists $x_{i+1} \in A$ with $\|\operatorname{proj}_{\{x_1,\ldots,x_i\}} x_{i+1}\| \leq \frac{\sqrt{n}}{3}$. $\square$

Eq. (3) implies that only a small amount of $y \in \{-1, +1\}^n$ can have small inner product with all x_i's. Formally,

Lemma 2

Fix vectors $x_1, x_2, \ldots, x_m \in \{-1, +1\}^n$ *that obey* (3) *for all* i. *Then*

$$\Pr_{y \in \{-1,+1\}^n} \left[\max_{i \in [m]} |\langle x_i, y \rangle| \le \frac{\sqrt{n}}{4} \right] \le e^{-\beta m} \tag{4}$$

for some absolute constant $\beta > 0$.

Proof of theorem 3

Let $x_1, \ldots, x_k \in A$ be the vectors from lemma 1.

Recall that we have for every $x_i \in A$,

$$\Pr_{y \in B}[|\langle x_i, y \rangle| \leq \frac{\sqrt{n}}{4}] \geq 1 - 2\varepsilon.$$

By averaging,

$$\Pr_{i \in [k], y \in B}[|\langle x_i, y \rangle| \leq \frac{\sqrt{n}}{4}] \geq 1 - 2\varepsilon.$$

Again, we may assume that for every $y \in B$,

$$\Pr_{i \in [k]}[|\langle x_i, y \rangle| \leq \frac{\sqrt{n}}{4}] \geq 1 - 3\varepsilon,$$

which decreases the size of B by at most $2/3$.

Proof of theorem 3

Then,

$$\Pr_{y\in\{-1,+1\}^n}\left[\Pr_{i\in[k]}[|\langle x_i, y\rangle| \leq \frac{\sqrt{n}}{4}] \geq 1-3\varepsilon\right]$$

is an upper bound for $\Pr_y[y \in B] = 2^{-n}|B|$.

By union bound, this is bounded by

$$\binom{k}{3\varepsilon k} \Pr_{y\in\{-1,+1\}^n}\left[\max_i |\langle x_i, y\rangle| \leq \frac{\sqrt{n}}{4}\right],$$

which, by lemma 2, is bounded by $\binom{k}{3\varepsilon k} e^{-\Omega(n)} = e^{-\Omega(n)}$. $\square$

Linear lower bound

By theorem 3 and the corruption bound, we have

$$D^{\mu}_{\xi}(f_n) \geq \Omega(n) \log\left(\mu(f_n^{-1}(-1)) - \frac{\xi}{\varepsilon}\right).$$

Since $\mu(f_n^{-1}(-1))$ is $\Theta(1)$, the above gives a linear lower bound for the gap orthogonality function.

which also implies a linear lower bound for GHD.