An Optimal Lower Bound on the Communication Complexity of Gap-Hamming-Distance

Amit Chakrabati, Oded Regev

presented by Tomáš Gavenčiak

1 Introduction

In Gap-Hamming-Distance (or just GHD), Alice and Bob each have an n-bit string (x and y). Their goal is to distinguish between the cases $\Delta(x,y) \geq n/2+\sqrt{n}$ and $\Delta(x,y) \leq n/2-\sqrt{n}$ by communicating as few bits as possible. Note that the trivial protocol would use $\Theta(n)$ bits to transfer one of the strings.

$\mathrm{GHD}_{n,t,g}$ is the problem GHD with n-bit strings where Alice and Bob must distinguish between $\Delta(x,y) \geq t+g$ and $\Delta(x,y) \leq t-g$.

For a (partial) function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1,*\}$ (where $*$ represent the undefined values), a (possibly randomized) protocol P *fails* on input (x,y) if $f(x,y) \neq *$ and $P(x,y) \neq f(x,y)$. Let $\mathrm{cost}(P)$ denote the worst-case communication cost of P in bits.

Randomized protocols. A randomized protocol P computes computes f with error at most ϵ if

$$\forall (x,y) \in \mathcal{X} \times \mathcal{Y} : f(x,y) \neq * \implies \Pr[P(x,y) \neq f(x,y)] \leq \epsilon.$$

Let $\mathrm{err}(P)$ denote the least ϵ such that P computes f with error at most ϵ.

Also let $R_\epsilon(f) = \min_P\{\mathrm{cost}(P), P \text{is a randomized protocol for} f \text{with} \mathrm{err}(P) \leq \epsilon\}$.

Deterministic protocols. Let $\mathrm{err}_\mu(P)$ denote the probability that P fails on (x,y) with (x,y) distributed according to μ.

Let $\mathrm{err}_\mu(P)$ denote the least ϵ such that P computes f on input distributed according to μ with error at most ϵ.

Also let $D_{\mu,\epsilon(f)} = \min_P\{\mathrm{cost}(P), P \text{ is a deterministic protocol for } f \text{ with } \mathrm{err}_\mu(P) \leq \epsilon\}$.

We use $R(f)$ for $R_{1/3}(f)$ and $D_\mu(f)$ for $D_{\mu,1/3}(f)$.

Distributions. Let $\xi_{n,p}$ denote the distribution resulting from the following process: Pick $x = y$ from $\{0,1\}^n$ uniformly, then flip every bit of y with probability $(1-p)/2$, output (x,y) (so $\xi_{n,0}$ is uniform on $\{0,1\}^{2n}$). We omit n where clear from the context.

2 Main result

Theorems 2.6 and 2.7 (Main result)

$$R(\mathrm{GHD}_{n,n/2,\sqrt{n}}) = \Omega(n)$$

Moreover, there exists an absolute constant $\epsilon > 0$ for which

$$D_{\xi_0,\epsilon}(\mathrm{GHD}_{n,n/2,\sqrt{n}}) = \Omega(n)$$

3 Reductions

Yao's principle. For any (communication) problem, there is a distribution α over the correct deterministic algorithms A and a distribution ξ over the inputs X such that

$$\max_{x\in X} \mathbb{E}_{a\sim\alpha}(\mathrm{cost}_a(x)) = \min_{a\in\alpha} \mathbb{E}_{x\sim\xi}(\mathrm{cost}_a(x)).$$

This implies $R_\epsilon(f) \geq D_{\mu,\epsilon}(f)$ for any ϵ, f and μ.

Lemma 4.1 For all integers $n, t, g, k, l > 0$:

(1) $R(\mathrm{GHD}_{n,t,g+k}) \leq R(\mathrm{GHD}_{n,t,g})$

(2) $R(\mathrm{GHD}_{n,t,g}) \leq R(\mathrm{GHD}_{kn,kt,kg})$

(3) $R(\mathrm{GHD}_{n,t,g}) \leq R(\mathrm{GHD}_{n+k+l,t+k,g})$

(4) $R(\mathrm{GHD}_{n,t,g}) = R(\mathrm{GHD}_{n,n-t,g})$

Lemma 4.2 For all integers $n > 0$ and reals $b > 0$ and $b \leq \sqrt{n}/2$, we have

$$R(\mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{2n}}) \leq R(\mathrm{GHD}_{2n,n,\sqrt{2n}}).$$

Lemma 4.T There exist $\delta_0 > 0$, $a > 0$ and $b > 0$ such that for every deterministic protocol P for $\mathrm{GHD}_{2n,n,\sqrt{n}}$ with $\mathrm{err}_{\mu_{2n,0}}(P) = \delta \leq \delta_0$ there is a randomized protocol Q showing that $R(\mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{2n}}) = O(D_{\xi_0}(\mathrm{GHD}))$.

4 Rectangles and Corruption

A set $R \in X \times Y$ is a *rectangle* if $R = \mathcal{X} \times \mathcal{Y}$ for $\mathcal{X} \subseteq X$ and $\mathcal{Y} \subseteq Y$.

Lemma 2.1 For a deterministic protocol P on $X \times Y$ communicating c bits, for every output value $z \in Z$, there exist 2^c pairwise disjoint rectangles $R_{1,z}, \ldots, R_{2^c,z}$ such that for all $(x, y) \in X \times Y$ we have

$$P(x, y) = z \iff (x, y) \in \bigcup_{i=1}^{2^c} R_{i,z}.$$

Theorem 2.2 For all $\alpha_0, \alpha_1, \alpha_+, \epsilon > 0$ with $\epsilon < (\alpha_1 - \alpha_+)/(\alpha_0 + \alpha_1)$, there exist $\beta \in \mathbb{R}$ and $\epsilon' > 0$ such that:

Let $f : X \times Y \to \{0, 1, *\}$, $A_i = f^{-1}(i)$. Suppose there are distributions μ_0, μ_1, μ_+ on $X \times Y$ and $m > 0$ such that

(1) for $i \in \{0, 1\}$, $\mu_i(A_i) \geq 1 - \epsilon$

(2) for all rectangles $R \subseteq X \times Y$, $\alpha_1\mu_1(R) - \alpha_+\mu_+(R) \leq \alpha_0\mu_0(R) + 2^{-m}$

Then, for $\nu = (\alpha_0\mu_0 + \alpha_1\mu_1)/(\alpha_0 + \alpha_1)$, we have $D_{\nu,\epsilon'}(f) \geq m + \beta$.

4.1 Towards the main theorem

Let $f_b = \mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{2n}}$.

Lemma 2.4 For all $\epsilon > 0$ there exists $b > 0$ such that for n large enough, $\xi_{4b/\sqrt{n}}(A_0) \geq 1 - \epsilon$, and $\xi_0(A_1) \geq 1 - \epsilon$ where $A_i = f_b^{-1}(i)$.

Lemma 2.5 For all $b > 0$ there is $\delta > 0$ such that for n large enough,

$$\forall R \subseteq \{0,1\}^n \times \{0,1\}^n \text{ rectangular} : \frac{1}{2}\left(\xi_{-4b/\sqrt{n}}(R) + \xi_{4b/\sqrt{n}}(R)\right) \geq \frac{2}{3}\xi_0(R) - 2^{-\delta n}$$

Let $\epsilon = 1/8$, let b be as in Lemma 2.4, let δ be as in Lemma 2.5, let n be large enough (for 2.4 and 2.5). Also let $m = \delta n$, $\mu_0 = \xi_{4b/\sqrt{n}}$, $\alpha_0 = 1/2$, $\mu_1 = \xi_0$, $\alpha_1 = 2/3$, $\mu_+ = \xi_{-4b/\sqrt{n}}$, $\alpha_+ = 1/2$, $\epsilon = 1/8$ and $f_b = \mathrm{GHD}_{n,n/2-b\sqrt{n},\sqrt{2n}}$.

4.2 Steps for Lemma 2.5

Let γ^n denote n-dimensional Gauss distribution with density $Ze^{-\|x\|^2/2}$ (Z is a normalizing element).

A η-correlated gaussioan pair (x, y) has the following distribution: choose x and z from γ^n independently and then set $y = \eta x + \sqrt{1 - \eta^2}z$.

Theorem 3.5 For all $c, \epsilon > 0$, there is $\delta > 0$ such that for n large enough and $0 \leq \eta \leq c/\sqrt{n}$ and all $A, B \subseteq \mathbb{R}^n$ with $\gamma^n(A), \gamma^n(B) \geq e^{-\delta n}$ we have

$$\frac{1}{2}\left(\Pr_{(x,y)\ is\ \eta-corr.}[x \in A, y \in B] + \Pr_{(x,y)\ is\ -\eta-corr}[x \in A, y \in B]\right) \geq (1 - \epsilon)\gamma^n(A)\gamma^n(B).$$

Corollary 3.8 For all $c, \epsilon > 0$, there is $\delta > 0$ such that for n large enough and $0 \leq p \leq c/\sqrt{n}$ and all $A, B \subseteq \{0,1\}^n$ with $|A|, |B| \geq 2^{(1-\delta)n}$ we have

$$\frac{1}{2}\left(\xi_{-p}(A \times B) + \xi_p(A \times B)\right) \geq (1 - \epsilon)\xi_0(A \times B).$$

Recall that $D(P\|Q) = \int P(x)\ln(P(x)/Q(x))\d{x}$. Let $D_\gamma(X) = D(P\|\gamma)$ for $X \sim P$.

Theorem 3.1 (the taste of Gauss) For all $\epsilon, \delta > 0$ and n large enough, have $A \in \mathbb{R}^n$ such that $\gamma^n(A) \geq e^{-\epsilon^2 n}$. Then for all but $e^{-\delta n/36}$ of unit vectors $y \in \S^{n-1}$ the distribution of the projection $\langle x, y \rangle$ where $x \sim \gamma^n|_A$ is equal to $\alpha X + Y$ for some $1 - \delta \leq \alpha \leq 1$ and (possibly dependent) random variables X and Y satisfying

$$D_\gamma(X|Y) \leq \epsilon.$$