Improved bound on $c_S(w)$ for $l = log|S|$

Aaryan Gupta

10 June 2022

1 Problem

For a set $S \subseteq \{0,1\}^d$ we define $c_S(w) = |\{(x,y) \mid x \in S, y \in S, d(x,y) = w\}|$, i.e., the number of pairs of vectors in S that are at Hamming distance w from each other. We are asked to provide an upper bound for the sum $\frac{\sum_{w=0}^{r} c_S(w)}{|S|}$ over all such sets S.

2 Reformulation of Problem to graphs

We consider the set S to be a subset of the boolean hypercube graph of dimension d. In this graph, the set of vertices is $V = \{v : v \in S\}$ and the set of edges is $E = \{(x,y) : d_H(x,y) \leq r\}$ where d_H is the Hamming distance function. The problem transforms to finding the maximum value of $e = \frac{|E|}{|V|}$.For sets S, and a maximum Hamming distance r, we denote this as maximising $e_{\leq r}(S) = \frac{|E_{\leq r}(S)|}{|S|}$.

3 Down Sets

A set S is said to be a down-set if $x \in S$ implies $y \in S$ whenever $y \subseteq x$. Note that the set definition of the bit string of y is used here. For example 0100010 corresponds to $\{2,6\}$. A set S is said to be a left-compressed set if $x \in S$ implies $y \in S$ whenever y satisfies the two conditions-

1. $|x| = |y|$
2. either $x_1 = 0, y_1 = 0$ or there exists $i, j \in [d]$ with $1 < i < j$ such that $x_1 = y_1, ..., x_{i-1} = y_{i-1}$ and $x_i = 0, x_j = 1, y_i = 1, y_j = 0$

Theorem 1 *A down set achieves the maximum value of $e_{\leq r}(S)$.*

Proof Outline. We start with a set B which achieves the maximum of $e_{\leq r}(B)$. We define a down-shift operator D_i which replaces every element in B whose i^{th} position is 1 and the replaced element is not already present in B. The set $D(B) = D_1(D_2(...(D_d(B))..)$ is a down set. By case bashing, we show that after a single operation of D_i to B, $e_{\leq r}(B) \leq e_{\leq r}(D_i(B))$. Hence any set S for which $e_{\leq r}(S)$ is maximum can be transformed into a down-set $D(S)$ which retains the same property.

Note that left-compression is not required for this proof.

4 The general case for even r

4.1 Partitioning the edge set into a disjoint union of equal mutual hamming distance pairs $e_{(b,a)}(S)$

For non-negative integral a and b define

$$E_{(b,a)}(S) = \{x, y \in E_{\leq 2t}(S) : |x \backslash y| = b, |y \backslash x| = a\}$$

Now let

$$U = \{(b, a) : b \geq a, b + a \leq 2t\}$$

Also let $e_{(b,a)}(S) = |E_{(b,a)}(S)|$. Now we can decompose $E_{\leq 2t}(S)$ as a disjoint union

$$E_{\leq 2t}(S) = \bigcup_{(b,a) \in U} E_{(b,a)}(S)$$

and in turn we have

$$e_{\leq 2t}(S).|S| = \sum_{(b,a) \in U} e_{(b,a)}(S)$$

4.2 Counting pairs and finding an upper bound on $e_{(b,a)}(S)$ for fixed (b, a)

We now find an upper bound on the number of pairs $\{x, y\} \in E_{(b,a)}(S)$ at a hamming distance of at most $2t$. As $b \geq a$, we know that $|x| \geq |y|$.

Now fix an $x \in S$. We bound the number of $y \in \{0, 1\}^d = |Y|$ such that $\{x, y\} \in E_{(b,a)}(S)$ and $|(y \backslash x)| = b$ and $|(x \backslash y)| = a$. By some combinatorial arguments we show that

$$|Y| = \binom{n - |x|}{a} \binom{|x|}{a} \binom{|x|}{b - a}$$

.

Using the inequality $\binom{b}{a} \leq \frac{b^a}{a!}$, we simplify this expression to

$$|Y| \leq \frac{(n - b)^a l^b}{(a!)^2 (b - a)!}$$

Using Stirling's inequality, the denominator can be lower bounded as $(a!)^2(b-a)! \geq \frac{1}{2^{2a}.e^{b+a}}(2a)^{2a}(b-a)^{b-a}$. Using $x^x y^y \geq (\frac{x+y}{2})^{\frac{x+y}{2}}$ (Jensen's inequality on x^x), we get

$$(a!)^2 (b - a)! \geq \frac{1}{2^{2a}.e^{b+a}} \left(\frac{a + b}{2}\right)^{\frac{a+b}{2}}$$

The numerator can be upper bounded using as

$$(n-b)^a l^b \le n^a l^b \le (nl)^{\frac{a+b}{2}}$$

Combining the two expressions we have

$$|Y| \le \left(\frac{4e\sqrt{nl}}{b+a}\right)^{b+a}$$

4.3 Bounding $e_{(b,a)}(S)$ for any (b, a)

Here we will maximise the bound we obtained before over all (b, a) to get an inequality involving t, n, and l.

Define an integer $k = b+a \le 2t$ and $k \ge 2$. We want to show that the above bound is increasing over increasing k. So it suffices to show that

$$\left(\frac{4e\sqrt{nl}}{k-1}\right)^{k-1} \le \left(\frac{4e\sqrt{nl}}{k}\right)^{k}$$

.

$$k\left(\frac{k}{k-1}\right)^{k-1} \le 4e\sqrt{nl}$$

Now as $\left(\frac{k}{k-1}\right)^{k-1} \le e$,it suffices to show that $k \le 4\sqrt{nl}$.

This is true because

$$\left(\frac{k}{4}\right)^2 = \left(\frac{b+a}{4}\right)^2 \le \left(\frac{t}{2}\right)^2 \le t^2 \le \lfloor log|S|\rfloor^2 \le nl$$

Note that we used the fact that the set is a down set in the inequality $t \le \lfloor log|S|\rfloor$. So we have concluded that the pair for which the bound is weakest is when $b + a = 2t$. Substituting, we have

$$e_{(b,a)}(S) \le \left(\frac{2e}{t}\right)^{2t} (nl)^t$$

for all pairs $(b, a) \in U$

4.4 Putting it all together

As $b \in \{1, 2, ..., 2t\}$ and $a \in \{1, 2, ..., t\}$, we have $|U| \le 2t^2$. Hence we have

$$e_{\le 2t}(S)|S| \le 2t^2 \left(\frac{2e}{t}\right)^{2t} (nl)^t$$