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

Aaryan Gupta

27 May 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 Left-Compressed 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 left-compressed 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.

We shall proceed with a similar proof for left-compression. We define an operator $L_{i,j}$ on set B which for every $z \in B$ swaps z_i and z_j if $z_i = 0$ and $z_j = 1$ if $i < j$. We argue that the set $L(B) = L_{1,1}(L_{1,2}(...L_{d-1,d}(B)...))$ is a left-compressed set. WLOG we look at $i = 1$ and $j = 2$ and analyse the different cases to conclude that set $L_{1,2}(B)$ is a down-set if B is a down set. Then using similar case bashing analysis as above, we conclude that after a single operation of $L_{i,j}$ to B, $e_{\leq r}(B) \leq e_{\leq r}(L_{i,j}(B))$. Hence any set S such that $e_{\leq r}(S)$ is maximum can be transformed to a left-compressed down-set $L(D(S))$ with the same property.

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

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

.

Now, this combinatorial expression is much smaller than the expression used in Rashtchian's paper. Our hope is to get an analytic expression for this using some basic inequalities and the fact that $b \leq |x| \leq l = \lfloor log|S|\rfloor$.

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.

4.4 Putting it all together

Here we will use the fact that $|U| \leq 2^{2t}$ and multiply that by the bound before on $e_{(b,a)}(S)$ to get a bound on $e_{\leq 2t}(S)$