$c_S(w)$ bound 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 Tighter Bounds for Small Distances

4.1 $r = 0$

For $r = 0$, the quantity $E_{\leq 0}(S)$ is just the number of vertices in graph of S, that is $|S|$.

4.2 $r = 1$

For $x \in S$, where S is a down-set, we have $y \in S$ for all $y \subseteq x$. So we have $2^{|x|} \leq |S|$, or in turn $|x| \leq \lfloor log(S) \rfloor$. Again because S is a down-set, $|E_{\leq 1}(S)| = \sum_{x \in S} |x| \leq |S| log |S|$. A better optimal bound is $|E_{\leq 1}(S)| \leq \frac{1}{2}|S| log |S|$, which is obtained in some other papers cited by Rashtchian.

4.3 $r = 2$

4.3.1 Rewriting expression in terms of rank

Define the rank of a boolean vector x as $||x|| = \sum_{j \in [n]} j x_j = \sum_{j \in x} j$.

Theorem 2 *For a left-compressed down set S, $E_{\leq 2}(S) = \sum_{x \in S} ||x||$.*

Proof Outline. The key idea being used here is that $\{x, y\} \in E_{\leq 2}(S)$ implies that $||y|| \neq ||x||$. Note that a rank of the order $O(n^2)$ does not work for $r = 3$ and this distinguishing property of ranks is the only reason the proof works for $r = 2$ and not for higher powers. After noticing this, WLOG we fix $x \in S$ and count y such that $||y|| < ||x||$. Now, y can be of three forms:

1. $y = x \cup \{i\} \backslash \{j\}$ where $i < j, j \in x$, and $i \notin x$.
2. $y = x \backslash \{i\}$ where $i \in x$.
3. $y = x \backslash \{i, j\}$ where $i, j \in x$.

Counting the number of possible y's in all three cases and adding them up gives us our required result.

4.3.2 Finding an upper bound for rank

Theorem 3 *For a left-compressed down-set S, for any $x \in S$ we have*

$$||x|| \leq d.l^{'}$$

where $l^{'} = \lfloor log|S| \rfloor$.

Proof Outline. From the $r = 1$ case we know that $|x| \leq \lfloor log|A| \rfloor$. Hence, we have $E_{\leq 2}(S) = \sum_{x \in S} ||x|| \leq \sum_{x \in S} d|x| \leq dlog|S|$. Therefore we have $e_{\leq 2}(S) \leq l^{'}$.

4.3.3 Substituting expression for rank back to get final bound

Substituting the bound for rank obtained in Theorem 3, we substitute it back in the expression in Theorem 2.

$$e_{\leq 2}(S) = \frac{1}{|S|} \sum_{x \in S} ||x|| \leq d.l^{'}$$

where $l^{'} = \lfloor log|S| \rfloor$.

5 The general case for even r

5.1 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

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

5.2 $l_x \leq l$

We start off with some definitions. We define $l = \lfloor log|S| \rfloor$ and $\beta = \lfloor (\frac{d}{log|S|})^{\frac{1}{2}} l \rfloor$. We then define $l_x = |x \cap \{\beta + 1, ..., d\}|$ for an $x \in S$. This intuitively represents the number of "big" elements in x. Note that the inequality $\beta^2 = dl$ follows from the definition.

Theorem 4 *Let $S \subseteq \{0,1\}^d, |S| \geq 2$ be a left-compressed down set. If $x \in S$, then $l_x \leq l$.*

*Proof Outline.*We obviously have $l_x \leq |x|$ and from the $r = 1$ case we know that $|x| \leq \lfloor log|S| \rfloor$. So the result follows for $l = \lfloor log|S| \rfloor$.

5.3 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 $2t$. We partition the pairs into two cases: when $l_y \leq l_x$ and when $l_y > l_x$. The proofs for both are very similar so we just look at the case when $l_y \leq l_x$ for brevity.

Now fix an $x \in S$,. for each $p \in 0, 1, ..., a$ we bound the number of $y \in \{0,1\}^d = |Y|$ such that $\{x, y\} \in E_{(b,a)}(S)$ and $l_y \leq l_x$ and $|(y \backslash x) \cap \{\beta+1, ..., d\}| = p$. By some combinatorial arguments we show that

$$|Y| \leq \binom{n - \beta - l_x}{p} \binom{l_x}{p} \binom{\beta - |x| + l_x}{a - p} \binom{|x|}{b - p}$$

.

Now we use the inequality $l_x \leq l$ we proved earlier to say that

$$|Y| \leq \binom{n}{p} \binom{l}{p} \binom{\beta}{a - p} \binom{|x|}{b - p} \leq \frac{(nl)^p . \beta^{a-p} . |x|^{b-p}}{(p!)^2 . (a - p)! . (b - p)!}$$

Using Stirling's approximation and Jensen's inequality we lower bound the denominator as $(p!)^2.(a - p)!.(b - p)! \geq (\frac{b+a}{4e})^{b+a}$.

Now we upper-bound the numerator using the inequalities $\beta|x| \leq nl$, $\beta^2 \leq nl$, and $|x|^2 \leq nl$. We just look at the case when $b + a$ is even for brevity and this leads us to

$$(nl)^p . \beta^{a-p} . |x|^{b-p} \leq (nl)^p . (nl)^{\frac{a-p}{2}} . (nl)^{\frac{b-p}{2}} = (nl)^{\frac{b+a}{2}}$$

This lets us bound $e_{(b,a)}(S)$ for $l_y \leq l_x$. We get similar results when $b + a$ is odd and when $l_y > l_x$.

For this particular case we get the following result:

Theorem 5 *Let b, a be non-negative integers with $b \geq a$, $b + a \leq 2log|S|$, and when b and a are both even we have*

$$|x, y \in E_{(b,a)}(S) : l_y \leq l_x| \leq (\frac{4\sqrt{2}e}{b + a})^{b+a} . (d.l)^{\frac{b+a}{2}} . |S|$$

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

Theorem 6 $\frac{e_{(b,a)}(S)}{|S|} \leq (\frac{4e}{t})^{2t} (dl)^t$ *for all (b, a) in U.*

Proof Outline. Here is the snippet from Rashtchian's original paper on how he eliminates beta.

$$
\begin{aligned}
e_{(b,a)}(\mathcal{A}) &\leqslant \left(\frac{4\sqrt{2}e}{b+a}\right)^{(b+a)} \cdot (n\ell)^{(b+a)/2} \cdot |\mathcal{A}| + \left(\frac{4\sqrt{2}e}{b+a}\right)^{(b+a)} \cdot (n\ell)^{(b+a-2)/2} \cdot \ell\beta \cdot |\mathcal{A}| \\
&= \left(\frac{4\sqrt{2}e}{b+a}\right)^{(b+a)} \cdot |\mathcal{A}| \cdot (n\ell)^{(b+a-2)/2} \cdot (n\ell + \ell\beta) \\
&\leqslant 2 \cdot \left(\frac{4\sqrt{2}e}{b+a}\right)^{(b+a)} \cdot |\mathcal{A}| \cdot (n\ell)^{(b+a)/2} \quad (\text{as } \ell\beta \leqslant n\ell) \\
&\leqslant \left(\frac{8e}{b+a}\right)^{(b+a)} \cdot |\mathcal{A}| \cdot (n\ell)^{(b+a)/2} \quad (\text{as } 2 \leqslant \sqrt{2}^{(b+a)}).
\end{aligned}
$$

After this, he maximises this function over a new variable $k = b+a$ and uses inequalities unrelated to beta to get the result.

5.5 Putting it all together

$e_{\leq 2t}(S) \leq_{(b,a)\in U} \frac{e_{(b,a)}(S)}{|S|} \leq |U|.(\frac{4e}{t})^{2t}(dl)^t \leq (\frac{8e}{t})^{2t}(dl)^t$ as $|U| \leq 2^{2t}$