Sunflower lifting

Ian Mertz

April 6, 2020

1 Sunflower definitions (informal)

Definition 1.1. A $(1/2, \epsilon)$ approximate sunflower is a collection of sets such that a random set contains at least one set in the collection with probability $\geq 1 - \epsilon$. Alternatively it's a collection of binary strings such that a random string is not greater than or equal to any of the strings in the collection with probability at most ϵ.

Lemma 1.2. *There exists a $(1/2, \epsilon)$ approximate sunflower in any collection of $(k \log 1/\epsilon)^{\Omega(k)}$ k-regular sets.*

2 Main results

We recall the basic setup at the current step of the decision tree simulation of a communication protocol Π. Our protocol Π will have depth $d \log m$ where $m = d^{1000}$; our goal is to simulate Π with a decision tree of depth $O(d)$.

After querying a bit of X and going to a rectangle of size $|X|/2$, we perform the rectangle partition algorithm on X, restoring $0.95 \log m$ blockwise min-entropy in each of its parts. The assignment (I_j, α_j) to the jth part of the rectangle partition $\mathcal{F}$ represents an assignment that violates $0.95 \log m$ blockwise min-entropy, but all blocks still have $0.9 \log m$ blockwise min-entropy even before performing the rectangle partition.

To this end, we either need to choose a rectangle (I_j, α_j) in $\mathcal{F}$ and query the z values in I_j without losing largeness in Y, or we need to go to the default rectangle in $\mathcal{F}$ corresponding to no restriction to the z variables without losing more than another half of X.

We will shorthand (I_j, α_j) as γ_j, as for this proof we will not be concerned with which coordinates in $[n]$ are being fixed, and so we can think of the assignments as being arbitrary sets from $[mn]$. It is a standard fact that $|\gamma_j| \leq 10d$ for all $\gamma_j \in \mathcal{F}$ [**?**].

Definition 2.1. Let $\mathcal{F}$ be a rectangle partition of X. For $j \in [|\mathcal{F}|]$ and $\beta_j \in \{0,1\}^{|I_j|}$, we define $Y^{j,\beta_j} := \{y \in Y : y[\gamma_j] = \beta_j\}$.

Lemma 2.2. *Let $\mathcal{F} = \{\gamma_j\}_j$ be a rectangle partition of X and let Y have deficiency c_y. Then there is some j such that for all β_j, $|Y^{j,\beta_j}| \geq 2^{mn - c_y - (|I_j| + \log c_y + 1)}$.*

Proof sketch. Our argument will be a counting argument, showing that if there is a bad assignment β_j for every j, then Y is too small. We count two parts of Y: the ys that satisfy some bad assignment β_j, and the ones that don't satisfy any β_j. The first part is small by assumption that

the β_j's are bad, and so we focus on the second part, which we prove in full generality using nothing but the min entropy conditions on X for every assignment $\gamma_j \in \mathcal{F}$.

We treat the sets γ_j as a set system over $[mn]$, and consider the probability that a random set does not contain any γ_j. Note that this corresponds to fixing all the bad assignments β_j to be consistent, say the all-ones vector, and then calculating the probability that a random vector from $\{0,1\}^{mn}$ has a 0 in each β_j. We do this by finding an approximate sunflower in $\mathcal{F}$ with an *empty* core, which by definition bounds the probability that a random set doesn't contain any γ_j.

Since there are many γ_j's by the fact that each takes up a low amount of X, we can find many approximate sunflowers, but we have no control over their core sizes. However, we can find so many sunflowers that their cores will eventually make their own approximate sunflower, which will be guaranteed to have a smaller size than the original ones. Note then that if we take all sets consistent with the core of this new sunflower, the probability of missing all those sets (minus this new core) is at most the probability of missing all of the original sunflowers plus the probability of missing the new sunflower, which is relatively small. Thus we up our probability of the approximate sunflowers slightly and guarantee that a new one appears with a smaller core. We iterate this until we hit a sunflower with an empty core, at which point we're done as claimed. □

Proof. Assume for contradiction that for every j there exists a β_j such that $|Y^{j,\beta_j}| < 2^{mn-c_y-|I_j|\log m-1}$. We will show that $|Y| < 2^{mn-c_y}$, which is a contradiction as c_y is the deficiency of Y. We make an additional assumption that for all j, $\beta_j = 1^{I_j}$, which we remove at the end of the proof.

Let $\mathcal{F}_k = \{\gamma_j \in \mathcal{F} : |\gamma_j| = k\}$. Note that for every j there is a set $X^j \subseteq X$ such that $X = \cup_j X^j$, so we let X_k be the part of X covered by sets in $\mathcal{F}_k$. Note that if $|X_0| \geq \frac{1}{2} \cdot |X|$ then the lemma is trivially true for $(\emptyset, \emptyset) = \gamma_j \in \mathcal{F}$. Otherwise, by averaging there must be some $0 < k \leq 10d$ for which $|X_k| \geq \frac{1}{20k}|X|$. Fix any k satisfying $|X_k| \geq \frac{1}{20d}|X|$, and for convenience we relabel all $\gamma_j \in \mathcal{F}_k$ so $\mathcal{F}_k = \{\gamma_j\}_{j\in[t]}$.

Our goal will be to show that there exists a $(1/2, 2^{-c_y-1})$ approximate sunflower $S_0 \subseteq \mathcal{F}_k$ where the core of S_0 is empty. Note that since such a sunflower corresponds to a set of $|S_0|$ disjoint sets, then

$$\Pr_{y\sim\{0,1\}^{mn}}(\forall \gamma_j \in S_0, y[\gamma_j] \neq \beta_j) = (1 - \frac{1}{2^k})^{|S_0|}$$

and so we can assume $|S_0| = 2^k \cdot c_y$ due to the fact that $(1 - 1/x)^x = e^{-1}$; if we find a larger S_0 we can simply remove excess elements.

Consider $Y_{\neq} = \{y \in Y : \forall \gamma_j \in S_0, y[\gamma_j] \neq \beta_j\}$ and $Y_{=} = \{y \in Y : \exists \gamma_j \in S_0, y[\gamma_j] = \beta_j\}$. Clearly these partition Y, and by the definition of an approximate sunflower

$$|Y_{\neq}| < 2^{mn} \cdot 2^{-c_y-1} = 2^{mn-c_y-1}$$

Since we assumed $|Y^{j,\beta_j}| < 2^{-c_y-(k+\log c_y+1)}$ for all j then

$$|Y_{=}| < 2^{mn-c_y-(k+\log c_y+1)} \cdot 2^{k+\log c_y} = 2^{mn-c_y-1}$$

Putting this together gives

$$|Y| = |Y_{\neq}| + |Y_{=}| < 2 \cdot 2^{mn-c_y-1} = 2^{mn-c_y}$$

which is a contradiction as stated before.

Now we show the existence of such an S_0. Fix $\epsilon = 2^{-c_y-d^6}$. We will consider the following procedure:

1. let $\mathcal{F}' = \mathcal{F}_k$, and let $\mathcal{F}^i = \emptyset$ for all $i \in k$
2. while $|\mathcal{F}^i| < 2^{0.88i \log m}$ for all i:
   (a) let $\mathcal{F}_S$ be the $(1/2, \epsilon)$ approximate sunflower in $\mathcal{F}'$ with the smallest core $S \subseteq [mn]$
   (b) set $\mathcal{F}^{|S|} \leftarrow \mathcal{F}^{|S|} - \{\mathcal{F}_S\}$ and $\mathcal{F}' \leftarrow \mathcal{F}' - \mathcal{F}_S$

First, we claim that there is always an approximate sunflower in (a). To do this we show that at all times

$$|\mathcal{F}'| \geq 2^{0.88k \log m}$$

This is enough to appeal to Theorem 1.2, since $2^{0.88k \log m} = d^{\Omega(k)} = (d \log 1/\exp(d))^{\Omega(k)}$ for $k \geq 1$. Consider any point before the loop terminates. We assume each S has size at most 1, since if $S = \emptyset$ ever occurs then we have found a $(1/2, 2^{-c_y - 1})$ approximate sunflower S_0 with an empty core as desired. Let $X_{f,k}$ be the piece of X_k in $\cup_i \mathcal{F}^i$. Now because all sets obey 0.89 blockwise min-entropy,

$$|X_{f,k}| \leq |X| \cdot (\sum_{i=1}^{10d} 2^{0.88i \log m} 2^{-0.89i \log m}) \leq |X| \cdot (10d \cdot m^{-0.01})$$

Now assume $|\mathcal{F}'| < 2^{0.88k \log m}$. Then since $k \geq 1$

$$|X_k| \leq |X| \cdot [(10d \cdot m^{-0.01}) + (2^{0.88k \log m} 2^{-0.89k \log m})] \ll \frac{1}{20d}|X|$$

which contradicts our assumption about X_k.

Consider the i such that $\mathcal{F}^i$ causes the loop to break. If $i = 0$ then we are done, as we have found an S_0 with an empty core. Otherwise since $|\mathcal{F}^i| = 2^{0.88i \log m} = (d \log \exp 1/d)^{\Omega(i)}$, again by Theorem 1.2 there exists a $(1/2, 2^{-c_y - d^6})$ approximate sunflower S in $\mathcal{F}^i$, and note that the core of S has size strictly less than i. Let $\mathcal{F}_S$ be the set of all sets in $\mathcal{F}_k$ consistent with S. We claim that $\mathcal{F}_k$ is a $(1/2, \epsilon')$ approximate sunflower, where

$$\epsilon' = \epsilon + 2^{-c_y - d^6}$$

To see this, consider the probability that a random set doesn't contain any set in $\mathcal{F}_S$. There are two ways this could happen: first, it could satisfy all cores that make up S and avoid the remainder of every set, and second, it could avoid the cores that make up S. The first of these happens with probability at most ϵ, since S is non-empty and each set in S is the core of a $(1/2, \epsilon)$ approximate sunflower, so avoiding even the remainders after removing *one* set in S happens with probability at most ϵ. The second event happens with probability $2^{-c_y - d^6}$ by definition of S as an approximate sunflower.

Thus we consider the following new procedure:

1. let $\epsilon = 2^{-c_y - d^6}$
2. loop until broken:
   (a) run the previous procedure until it breaks for some i
   (b) if $i = 0$, break
   (c) set $\epsilon \leftarrow \epsilon + 2^{-c_y - d^6}$

Clearly as argued before, if we exit for $i = 0$ and $\epsilon < 2^{-c_y-1}$ then we are done. Now consider one iteration of the outer loop. By the previous argument, after we decrease ϵ then there exists an $i' < i$ such that a new $(1/2, \epsilon)$ approximate sunflower exists with core size i'. Since we always take the sunflower of the smallest core, either the next pass of the procedure terminates for $i'' < i$ or it terminates for i and finds one more sunflower with core of size i'. Thus after at most $2^{0.88i' \log m}$ iterations of the procedure, it will terminate for i', and thus after at most $\sum_{i'=1}^{k} 2^{0.88i' \log m}$ iterations we will terminate for $i = 0$, each of which increases ϵ by a factor of at most $2^{0.88k \log m}$. Thus at the end of our procedure,

$$\epsilon \leq 2^{-c_y - d^6} \cdot (k \cdot 2^{0.88k \log m}) \cdot 2^{0.88k \log m} \ll 2^{-c_y-1}$$

as expected. This gives us a sunflower S_0 that our inner procedure outputs when $i = 0$, which gives us a $(1/2, 2^{-c_y-1})$ approximate sunflower with an empty core as claimed.

Finally we remove the restriction that β_j is the all ones restriction for all j. We claim that for any $\{\beta_j\}_{j \in [t]}$

$$|\{y \in \{0,1\}^{mn} : \forall j, y[\gamma_j] \neq \beta_j\}| \leq |\{y \in \{0,1\}^{mn} : \forall j, y[\gamma_j] \neq 1^{\gamma_j}\}|$$

We do this by a straightforward use of inclusion-exclusion. Let $\beta \leftrightarrow \beta'$ denote that β and β' are *consistent* on all coordinates they both fix. We can then write the probabilities of both events as follows

$$\Pr_{y \in \{0,1\}^{mn}}(\forall j, y[\gamma_j] \neq \beta_j) = \sum_{\substack{J \subseteq [t] \\ \beta_j \leftrightarrow \beta_{j'}, j, j' \in J}} \frac{(-1)^{|J|}}{2^{|\cup_{j \in J} \gamma_j|}}$$

$$\Pr_{y \in \{0,1\}^{mn}}(\forall j, y[\gamma_j] \neq 1^{\gamma_j}) = \sum_{\substack{J \subseteq [t] \\ 1^{\gamma_j} \leftrightarrow 1^{\gamma_j}, j, j' \in J}} \frac{(-1)^{|J|}}{2^{|\cup_{j \in J} \gamma_j|}}$$

Note that for the second case $1^{\gamma_j} \leftrightarrow 1^{\gamma_{j'}}$ for all $j, j' \in [t]$, and for all $|J| \leq 1$ the consistency condition is trivially true. Therefore expanding the sums up to the second term we have

$$\Pr_{y \in \{0,1\}^{mn}}(\forall j, y[\gamma_j] \neq \beta_j) = 1 - \sum_{j \in [t]} \frac{1}{2^{|\gamma_j|}} + \sum_{\substack{|J| \geq 2 \\ \beta_j \leftrightarrow \beta_{j'}, j, j' \in J}} \frac{(-1)^{|J|}}{2^{|\cup_{j \in J} \gamma_j|}}$$

$$\Pr_{y \in \{0,1\}^{mn}}(\forall j, y[\gamma_j] \neq 1^{\gamma_j}) = 1 - \sum_{j \in [t]} \frac{1}{2^{|\gamma_j|}} + \sum_{|J| \geq 2} \frac{(-1)^{|J|}}{2^{|\cup_{j \in J} \gamma_j|}}$$

and so the result follows by observing that the sum in the second equation is always greater than the sum in the first equation by the standard inclusion-exclusion principle. □