Advanced Proof Systems - Homework 2

Yosef Goren

December 8, 2022

1 Better All-Zero Check

1.1 Polynomial Identity Lemma

Proof of lemma:
The multivariate polynomial degree as stated in the question (*):

$$Deg\left(\sum_{i=1}^{m}\prod_{i=j}^{n} x_j^{d_{i,j}}\right) = \max\left\{\sum_{j=1}^{n} d_{i,j} \mid i \in [n]\right\}$$

Proof by induction.
Induction basis is immidiate from the fundemental theorem of algebra.

Assume for every nonzero $P = \sum_{i=1}^{m}\prod_{i=j}^{n} x_j^{d_{i,j}}$:

$$\Pr_{x_1,...,x_{n-1}\in\mathbb{F}}[P(x_1, ..., x_{n-1}) = 0] \leq \frac{Deg(P)}{|\mathbb{F}|}$$

Let $P \in \mathbb{F}[x_1, x_2, ..., x_n]$ be a nonzero polynomial.
Denote $d = Deg(P)$.
By treating all $\{x_i\}_{i=1}n-1$ as constants, P can be described as a univariate polynomial in x_n:

$$P(x_1, x_2, ..., x_n) = \sum_{i=0}^{d} x_n^i \cdot P_i(x_1, ..., x_{n-1})$$

Since P is nonzero:

$$\exists i : P_i(x_1, ..., x_{n-1}) \neq 0$$

Let: $k = \max\{i \mid P_i \neq 0\}$.
From (*):

$$Deg(x_n^k \cdot P_k) = k + Deg(P_k) \leq Deg(P) = d$$
$$\Rightarrow Deg(P_k) \leq d - k$$

Let $\{y_i\}_{i=1}^{n-1} \xleftarrow{\$} \mathbb{F}^{n-1}$ and $x_n \leftarrow \mathbb{F}$ (sampled uniformly).
Define the occurances:

- *SubRoot*: $P_k(y_1, ..., y_{n-1}) = 0$
- *Root*: $P(y_1, ..., y_{n-1}, x_n) = 0$

We are interested in showing that: $\Pr[Root] \leq \frac{d}{|\mathbb{F}|}$.

From the induction assumption:

$$\Pr[SubRoot] = \Pr\left[P_k(y_1, ..., y_{n-1}) = 0\right] \leq \frac{Deg(P_k)}{|\mathbb{F}|} \leq \frac{d-k}{|\mathbb{F}|}$$

Define $P'(x_n) = P(y_1, ..., y_{n-1}, x_n)$, note that $P'(x_n) = 0$ is equivalent to *Root*.
Assuming $\overline{SubRoot}$, $P'(x_n)$ is a univariate polynomial in x_n of degree k.
Thus from the fundemental theorem of algebra:

$$\Pr_{x_n \leftarrow \mathbb{F}}\left[P'(x_n) = 0\right] \leq \frac{k}{|\mathbb{F}|}$$

In other words:

$$\Pr[Root \mid \overline{SubRoot}] = \Pr\left[P'(x_n) = 0 \mid \overline{SubRoot}\right] \leq \frac{k}{|\mathbb{F}|}$$

Now denote the case *Root* to be the case where $P(y_1, ..., y_{n-1}, x_n) = 0$.
Denote $Root, SubRoot$ as R, S for ease of notation, and we finally have:

$$\Pr[R] = \Pr[R \mid S] \cdot \Pr[S] + \Pr[R \mid \bar{S}] \cdot \Pr[\bar{S}] \leq \Pr[S] + \Pr[R \mid \bar{S}] \leq \frac{d-k}{|\mathbb{F}|} + \frac{k}{|\mathbb{F}|} = \frac{d}{|\mathbb{F}|}$$

$$\Rightarrow \Pr[R] \leq \frac{d}{|\mathbb{F}|} \Rightarrow \Pr[P(y_1, ..., y_{n-1}, x_n) = 0] \leq \frac{d}{|\mathbb{F}|}$$

Tightness:
Let $n \in \mathbb{N}, d \leq |\mathbb{F}| - 1$.
We are interested in showing a polynomial for which the probability of begin zero is exactly $\frac{d}{|\mathbb{F}|}$.
Denote $F := |\mathbb{F}|$, $\{a_1, a_2, ..., a_F\} := \mathbb{F}$.
Define:

$$U(x) := \prod_{i=1}^{d}(x - a_i)$$

$$P(x_1, ..., x_n) := U(x_n)$$

For each a_i, $U(a_i) = 0 \cdot \prod(\dots)$ thus each a_i is a root of U, additionally, if U's argument x is not a_i, then $U(x)$ is a product of nonzeros and thus $U(x)$ is nonzero.
So:

$$\Pr_{x_1, \dots, x_n}[P(x_1, ..., x_n) = 0] = \Pr_{x_n \leftarrow \mathbb{F}}[U(x_n) = 0] = \Pr_x[U(x) = 0] = \frac{d}{F}$$

1.2 All-Zero Check with small Field

Let $\mathbb{F}$ be a finite field.
Define:

$$I_1(x, x') = 1 - x - x' + x \cdot x' + x \cdot x'$$

For any $n \in \mathbb{N} \setminus \{0, 1\}$, define:

$$I_n(x, x') = \prod_{i \in [n]} I_1(x_i, x_i')$$

For any polynomial $Q : \mathbb{F}^n \to \mathbb{F}$ and $z \in \mathbb{F}$, Define:

$$Q_z(x) := Q(x) \cdot I_n(x, z), S_{Q,z}(x) = \sum_{x \in \{0,1\}^n} Q_z(x)$$

Let Q be the zero polynomial and $x, z \in \mathbb{F}$, then (*):

$$Q_z(x) = Q(x) \cdot I_n(x, z) = 0 \Rightarrow S_{Q,z}(x) = 0$$

Also, note how $S_{Q,z}$ is a polynomial of individual degree 1, and total degree n, therefore, for any nonzero Q and $z \in \mathbb{F}$, we can use the polynomial identity lemma to get (**):

$$\Pr_{x \leftarrow \mathbb{F}^n} [S_{Q,z}(x) = 0] \leq \frac{n}{|\mathbb{F}|}$$

Interactive Protocol:
Define our interactive protocol to be the sumcheck protocol seen in class, with $\alpha = 0$ and applied on the Q_z polynomial where z is uniformly sampled from $\mathbb{F}$.

Completeness:
Let Q be the zero polynomial and consider a specific run of $(P, V)(Q)$.
Let z be the value sampled by V.
Due to (*), $S_{Q,z}(x) = 0$ for all $x \in \mathbb{F}^n$, thus, due to the completness of the sumcheck protocol - V will accept.

Soundness:
Let P^* be a (possibly) mallicious prover, Q be a nonzero polynomial and consider a specific run of $(P^*, V)(Q)$.
Denote z to be the value sampled by V.

Consider the soundness error of the sumcheck protocol given the arguments provided to it here; since Q_z has an individual degree of at most $d + 1$, it will be $\frac{(d+1)n}{|\mathbb{F}|}$ (***).

Thus:

$$\Pr[(P^*, V)(Q) = 1]$$
$$= \Pr[(P^*, V)(Q) = 1 \mid S_{Q,z}(x) = 0] \cdot \Pr[S_{Q,z}(x) = 0]$$

$$+\Pr\left[(P^*,V)(Q)=1 \mid S_{Q,z}(x)\neq 0\right]\cdot \Pr\left[S_{Q,z}(x)\neq 0\right]$$

$$=_{V def.} 1\cdot \Pr\left[S_{Q,z}(x)=0\right]+\Pr\left[(P^*,V)(Q)=1 \mid S_{Q,z}(x)\neq 0\right]\cdot \Pr\left[S_{Q,z}(x)\neq 0\right]$$

$$\leq_{(**),(***)} 1\cdot\frac{n}{|\mathbb{F}|}+\frac{n(d+1)}{|\mathbb{F}|}\cdot \Pr\left[S_{Q,z}(x)\neq 0\right]$$

$$\leq_{\Pr[\cdot]\leq 1} \frac{n}{|\mathbb{F}|}+\frac{n(d+1)}{|\mathbb{F}|}=O\left(\frac{n\cdot d}{|\mathbb{F}|}\right)$$

Complexity:
As we have seen in class, running the sumcheck verifier is with complexity $poly(n,d,log(|\mathbb{F}|))$, and sampling a random element z is also within those bounds. Hence the total runtime of our verifer is $poly(n,d,log(|\mathbb{F}|))$.

Additionally, since our verifier does not make additional queries to Q other then the ones done by the sumcheck verifier - only one query to Q is made.

2 Hardness of Approximating Clique Size

In this question, for a PPT algorithm A and $x\in\{0,1\}^n, r\in\{0,1\}^k$, denote $A(x)_{r\leftarrow}$ to mean the run of the algorithm A on input x where the randomization has sampled the bits of r, i.e. the first random bit is r_0, second is r_1 and so on.

Construction:
Let $L\in NP, x\in\{0,1\}^n$.
There exists a polynomial time verifier V_x using q queries and r random bits.
Define the following graph $G_x=(U_x,E_x)$:

$$U_x=\{(w,p)\mid V_x(w)_{p\leftarrow}=1\}$$

$E_x=\{((w,p),(w',p'))\mid$ all common queries in $V_x(w)_{\leftarrow p}, V_x(w')_{\leftarrow p'}$ have the same awnser.$\}$

Define our reduction $f(x)=G_x$.

Claim:
f is a (Karp) reduction from $GapClique_\epsilon$ to L.

Proof:

- If $x\in L$:
  Then there exists some π s.t. $\Pr_{p\leftarrow\{0,1\}^r}[V_x^\pi=1]=1$, which means that for any randomly sampled p, V_x^π will accept so each p sampled is consistent with all other p values. Hence G_x has a clique of size 2^r (at-least).

- If $x \notin L$:
  Assume towards contrediction exists a clique of size $\geq 2^{r-1}$ in G_x, This means there is a set Q of at-least 2^{r-1} queriest which are all consistent with one another.
  Now we use the prior to define a proof to convince V_x that $x \in L$:
  At the index i, put the result of the query q if $q \in Q$, otherwise put 0.
  Since we know that at-least 2^{r-1} ranzomizations of p yield queries that are consistent with the ones resulted in π^*, we have that:

  $$\Pr_p[V_x^{\pi^*} = 1] \geq \frac{2^{r-1}}{2^r} = \frac{1}{2}$$

  But this is a contredition to the soundness if V_x, meaning no such clique exists and $G_x \notin GapClique_\epsilon$.

- Efficiency:
  Thanks to the PCP theorem - we know that: $NP = PCP(O(1), O(log(n)))$.
  Thus there exists such PCP verifier V_x, that will yield a graph G_x with a polynomial (in n) number of vertices, So the graph G_x can be constructed efficiantly, which means that the reduction f is polynomial.

We have seen the correctness of the reduction. Since L is an arbitrary language in NP, we can conclude that $GapClique_\epsilon$ is NP-hard.

3 Random Linear Codes (Gilbert-Varshamov Bound)

Let

- $\epsilon > 0$
- $n := \Omega(\frac{ck}{\epsilon^2})$
- $A \xleftarrow{\$} \{0,1\}^{n \times k}$
- $C(x) := Ax$

Use $hw(\cdot)$ to denote the Hamming weight of a vector.
Let m be a vector of length k with $hw(m) = \Omega(\frac{ck}{\epsilon^2})$.
Let $X_i = (Am)_i, X = \frac{1}{n}\sum_{i=1}^{n} X_i$.
Since $(Am)_i = A_{:,i}m$, and each colomn of A is a random vector, X_i is a random variable.

Thus using chernoff we can see (*):

$$Pr\left[X < \frac{1}{2} - \epsilon\right] \leq Pr\left[|X - \frac{1}{2}| < \epsilon\right] \leq 2^{-\frac{1}{2}\epsilon^2 n} \leq 2^{-ck}$$

Since C is linear, we know it's absolute distance is given by:

$$d = min\{hw(C(m)) \mid m \neq 0\}$$

Thus:

$$\Pr\left[\frac{d}{n} < \frac{1}{2} - \epsilon\right] = \Pr\left[\frac{min\{hw(C(m)) \mid m \neq 0\}}{n} < \frac{1}{2} - \epsilon\right]$$

$$= \Pr\left[\exists m \neq 0 : \frac{hw(C(m))}{n} < \frac{1}{2} - \epsilon\right] = \Pr\left[\bigcup_{m\neq 0}\left(\frac{hw(C(m))}{n} < \frac{1}{2} - \epsilon\right)\right]$$

$$\leq_{UB} \sum_{m\neq 0} \Pr\left[\frac{hw(C(m))}{n} < \frac{1}{2} - \epsilon\right] = \sum_{m\neq 0} \Pr\left[\frac{|\{i \mid (Am)_i \neq 0\}|}{n} < \frac{1}{2} - \epsilon\right]$$

$$= \sum_{m\neq 0} \Pr\left[\frac{\sum_{i=1}^{n}(Am)_i}{n} < \frac{1}{2} - \epsilon\right] = \sum_{m\neq 0} \Pr\left[\frac{\sum_{i=1}^{n} X_i}{n} < \frac{1}{2} - \epsilon\right]$$

$$= \sum_{m\neq 0} \Pr\left[X < \frac{1}{2} - \epsilon\right] \leq_{(*)} \sum_{m\neq 0} 2^{-ck} = (2^k - 1)\cdot 2^{-ck} < 2^k \cdot 2^{-ck} = 2^{k(1-c)} \leq_{(**)} 0.01$$

(**) Since we want to bound the probability of the relative distance with 0.01, we require that this bound will be smaller than $2^{-7} = \frac{1}{128} < 0.01$:

$$2^{k(1-c)} < 2^{-7} \Leftrightarrow k(1-c) < -7 \Leftrightarrow 1 - c < -\frac{7}{k} \Leftrightarrow c - 1 < \frac{7}{k} \Leftrightarrow c < \frac{7}{k} + 1$$

Thus if c, k satisfiy these conditions, we have the required bound on the relative distance.