Lecture 10: Rank and Restrictions

Scribe: Udit Narayan Pandey   Lecture: Rajat Mittal

IIT Kanpur

In the last lecture we visualised the Boolean hypercube as a vector space $\mathbb{F}_2^n$ and looked at Linear Algebraic structure of this vector space, its basis vectors, subspaces, orthogonal complements, subcubes, affine subspaces and saw the definitions of sparsities, p-norm and Fourier p-norm for functions on this vector space.

In this lecture, continuing, we discuss the rank of a function f, it's relation with sparsity and then an important concept of restrictions of Boolean Functions.

1 Rank

In the last lecture we saw that any subset $S \subseteq [n]$ can be represented by an indicator vector $r \in \mathbb{F}_2^n$ where $r_i = 1$ iff $i \in S$. In view of this notation we saw that the parities can be rewritten as

$$\chi_S(x) = (-1)^{r \cdot x} = \chi_r(x), x \in \mathbb{F}_2^n$$

where r is indicator of S. Inspired by the idea that the group $\{-1, 1\}^n$ had a corresponding isomorphic group of characters, and that these r represent those characters, and to be able to visualise the Fourier transform as a function itself, we say that the subset indicators are elements of space $\widehat{\mathbb{F}_2^n}$, which is clearly isomorphic to $\mathbb{F}_2^n$.
The Fourier expansion of any function $f : \mathbb{F}_2^n \to \mathbb{R}$ can then be written as

$$f(x) = \sum_{r \in \widehat{\mathbb{F}_2^n}} \hat{f}(r) \cdot (-1)^{r \cdot x}$$

Definition 1. *The Fourier - support $S_{\hat{f}}$, for any function $f : \mathbb{F}_2^n \to \mathbb{R}$ is defined as:*

$$S_{\hat{f}} = \{r \in \widehat{\mathbb{F}_2^n} \mid \hat{f}(r) \neq 0\}$$

Definition 2. *For a function $f : \mathbb{F}_2^n \to \mathbb{R}$, let $S_{\hat{f}}$ denote Fourier support of $\hat{f}$ then*

$$rank(f) = dim(span(S_{\hat{f}}))$$

The rank of a function can give us a hint on the complexity of a function, specifically on the number of independent *quantities* a function depends upon.
For example say f's Fourier transform be given as $f(x) = x_1x_2$. The function depends on two variables, but we can see that we need to know the value of just one quantity to know the value of function, namely x_1x_2. So, if in some world we had the facility to query value of any partial parity on input x, the query complexity of f above would be 1. See that the rank of this function is also 1. This relation is true in general sense as well , to see that, first recall that partial parities are characters:

$$\chi_r(x) \cdot \chi_s(x) = \chi_{r+s}(x)$$

This relation above tells that if we know the value of partial parities $r \in S \subseteq \widehat{\mathbb{F}_2^n}$ we can find the value of partial parities of all linear combinations of vectors in S, namely $span(S)$.
This combined with the idea of rank tells us that we need to know the value of only $rank(f)$ partial parities, corresponding to the basis of $span(S_{\hat{f}})$ to know the value of f at any input. Since $span(S_{\hat{f}}) \subseteq \widehat{\mathbb{F}_2^n}$ is a subspace, $rank(f) \leq n$. So it can also be used as a measure of complexity(a Fourier query complexity) but this measure is input independent as opposed to query complexity which depends upon the input.

1.1 Relation with sparsity

Sparsity of function $spar(\hat{f})$ is the number of vectors $r \in \widehat{\mathbb{F}_2^n}$ such that $\hat{f}(r) \neq 0$, i.e. $spar(\hat{f}) = |S_{\hat{f}}|$. Since we know that a subspace of $\widehat{\mathbb{F}_2^n}$ of dimension k has 2^k elements, $rank(f)$ is bounded by $spar(f)$.
Some bounds on the rank of the function f are:

- $rank(f) \leq \min(spar(\hat{f}), n)$
- $log(spar(\hat{f})) \leq rank(f)$
  Since a subspace of dimension k has 2^k elements in it, the number of elements in $span(S_{\hat{f}})$ is $2^{rank(f)}$ and since $S_{\hat{f}} \subseteq span(S_{\hat{f}})$, $|S_f| = spar(\hat{f}) \leq 2^{rank(f)}$

The above bounds can be tightly satisfied by real valued functions, but when it comes to Boolean valued functions, the bounds can be improved. For functions like AND, OR one can verify that the second inequality holds tightly, so this bound remains the same.
The first bound has been improved to

- $rank(f) \leq \sqrt{spar(\hat{f}) \cdot log(spar(\hat{f}))}$

An example of a Boolean function with quite large rank as compared to sparsity is the addressing function.

- Let ADDR_t be the addressing function on $t + 2^t$ input variables. Let us denote the input as $x = (x_1, x_2, \ldots, x_t, y_1, y_2, \ldots, y_{2^t})$. Then the Fourier expansion of ADDR_t is:

$$\mathrm{ADDR}_t(x) = \sum_{i=1}^{2^t} y_i \cdot \mathbb{1}_{a_i}(x)$$

  where $\mathbb{1}_a$ is indicator function for a and a_i is t-bit Boolean representation of i. Every indicator function's Fourier expansion has all the Fourier coefficients non-zero and thus we see that i^{th} term in the above sum gives non-zero Fourier coefficients for $r \in \widehat{\mathbb{F}_2^{t+2^t}}$ such that $(r_1, r_2, \ldots, r_t) \in \widehat{\mathbb{F}_2^t}$ and $r_i = 1$. So we see that

$$S_{\hat{f}} = \bigcup_{i=1}^{2^t} \{r \in \widehat{\mathbb{F}_2^{t+2^t}} \mid (r_1, r_2, \ldots, r_t) \in \widehat{\mathbb{F}_2^t}, r_i = 1\}$$

  It is easy to see that
  1. $spar(\hat{f}) = |S_{\hat{f}}| = 2^{2t}$
  2. $span(S_{\hat{f}}) = \widehat{\mathbb{F}_2^{t+2^t}}$, so $rank(f) = t + 2^t$

  So we see that for $f = \mathrm{ADDR}_t$, $rank(f) = \mathcal{O}\big(\sqrt{spar(\hat{f})}\big)$.

2 Restrictions to a Subcube

We can restrict the study of our function f on $\{-1,1\}^n$ to a particular subcube instead of the Boolean Hypercube. We can assign values to some of the input variables and see the new function as a function of remaining variables. To formulate precisely:

Definition 3. *Let* $(\mathrm{J}, \bar{\mathrm{J}})$ *be a partition of* $[n]$ *and* $z \in \{-1,1\}^{\bar{\mathrm{J}}}$ *be an assignment to the variables in* $\bar{\mathrm{J}}$. *For the function* $f : \{-1,1\}^n \to \mathbb{R}$, *we say* $f_{\mathrm{J}|z} : \{-1,1\}^{\mathrm{J}} \to \mathbb{R}$ *is a restriction of* f *to* J *using* z *given by fixing variables in* $\bar{\mathrm{J}}$ *according to* z. *For* $y \in \{-1,1\}^{\mathrm{J}}$ *and* $z \in \{-1,1\}^{\bar{\mathrm{J}}}$, *we, with an abuse of notation can write*

$$f_{\mathrm{J}|z} = f(y, z)$$

Examples:

- $\mathrm{OR}_{[5]} : \{-1,1\}^5 \to \{-1,1\}, J = \{1,2,3\}, z = (1,1)$, then $\mathrm{OR}_{J|z} = \mathrm{OR}_{[3]}$, more generally we have that

$$\mathrm{OR}_{J|z} = \begin{cases} -1 & \exists i \in \bar{\mathrm{J}}, z_i = -1 \\ \mathrm{OR}_J & o.w. \end{cases}$$

- Similarly we have $\mathrm{PARITY}_{J|z} = \prod_{i \in \bar{\mathrm{J}}} z_i \cdot \mathrm{PARITY}_J$

2.1 Fourier Transform of Restricted Function

We saw in the example above that Parity function restricted to a subcube is another Parity function multiplied by a constant depending upon z. We can see that

$$\chi_{S_{J|z}}(x) = \big(\prod_{i \in S \cap \bar{\mathrm{J}}} z_i\big) \cdot \chi_{S \setminus \bar{\mathrm{J}}}(x) = \chi_{S \cap \bar{\mathrm{J}}}(z) \cdot \chi_{S \setminus \bar{\mathrm{J}}}(x)$$

Which gives, for

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(x)$$

$$\Rightarrow f_{J|z}(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \big(\prod_{i \in S \cap \bar{\mathrm{J}}} z_i\big) \cdot \chi_{S \setminus \bar{\mathrm{J}}}(x)$$

See that multiple sets S will give same $S \setminus \bar{\mathrm{J}}$, so for a particular $S \subseteq \mathrm{J}$, we have

$$\widehat{f_{J|z}}(S) = \sum_{T \subseteq \bar{\mathrm{J}}} \hat{f}(S \cup T) \cdot \chi_T(z)$$

If for $S \subseteq \mathrm{J}$ we define a function $F_{S|\bar{\mathrm{J}}} : \{-1,1\}^{\bar{\mathrm{J}}} \to \mathbb{R}$ as $F_{S|\bar{\mathrm{J}}}(z) = \widehat{f_{J|z}}(S)$, see that then $\widehat{F_{S|\bar{\mathrm{J}}}}(T) = \hat{f}(S \cup T)$. Under these conditions the following two relations hold:

- $\mathbf{E}_z[\widehat{f_{J|z}}(S)] = \hat{f}(S)$
  Proof- Using linearity of expectation, we have $\mathbf{E}_z[\widehat{f_{J|z}}(S)] = \sum_{T \subseteq \bar{\mathrm{J}}} \hat{f}(S \cup T) \cdot \mathbf{E}_z \chi_T(z)$, and for $T \neq \varnothing$, $\sum_z \chi_T(z) = 0$.

- $\mathbf{E}_z[\widehat{f_{J|z}}(S)^2] = \hat{f}(S \cup T)^2$ -Parseval's Identity
  Again, see that

$$\widehat{f_{J|z}}(S)^2 = \big(\sum_{T \subseteq \bar{\mathrm{J}}} \hat{f}(S \cup T) \cdot \chi_T(z)\big)^2 = \sum_{T_1, T_2 \subseteq \bar{\mathrm{J}}} \hat{f}(S \cup T_1) \cdot \hat{f}(S \cup T_2) \cdot \chi_{T_1 \triangle T_2}(z)$$

  And if $T_1 \neq T_2$, $\chi_{T_1 \triangle T_2}$ would be a non-trivial character so in taking the expectation, the only terms left will be when $T_1 = T_2$.

$$\mathbf{E}_z[\widehat{f_{J|z}}(S)^2] = \sum_{T \subseteq \bar{\mathrm{J}}} \hat{f}(S \cup T)^2 \mathbf{E}_z[1] = \sum_{T \subseteq \bar{\mathrm{J}}} \hat{f}(S \cup T)^2$$

The discussion above help us establish better upper bounds on $rank(f)$ in terms of $spar(\hat{f})$, precisely $rank(f) \leq \sqrt{spar(\hat{f}) \cdot \log(spar(\hat{f}))}$ and also help us understand a better learning algorithm namely Goldreich-Levin Algorithm.