Lecture 13 Bounds by Counting Dimensions

Lecturer: Anup Rao

The theme of todays lecture is how counting the dimension of an appropriately chosen linear space can be used to give bounds on combinatorial objects. We write $\mathbb{F}^k$ to denote the vector space of dimension k over the field $\mathbb{F}$.

We say that $v_1, \ldots, v_r \in \mathbb{F}^k$ are linearly independent if for $\alpha_i \in \mathbb{F}$, $\sum_i \alpha_i v_i = 0$ only if $\alpha_i = 0$ for all i. For a subspace $W \subseteq \mathbb{F}^k$, we say W is of dimension $d = \dim(W)$ if there are d linearly independent vectors in W such that every element of W can be generated by taking linear combinations of these vectors.

Definition 1. *We define the inner product of two vectors:* $\langle u, v \rangle = \sum_i u_i \cdot v_i$.

Fact 2. *Given a matrix with field element entries, its rank* $\mathsf{rk}(M)$ *is the dimension of space spanned by the rows, which is always equal to the dimension of the space spanned by the columns.*

Fact 3. *A square matrix M has full rank iff the determinant* $\det(M) \neq 0$.

1 Even/Odd Intersections and Sizes

Suppose $\mathcal{F}$ is a family of even-sized sets, such that all intersections are even. How many sets can there be? $2^{n/2}$.

A useful definition is the incidence vector of a set A, whose i'th coordinate is 1 if $i \in A$ and is 0 otherwise.

Theorem 4. *If every set of $\mathcal{F}$ is of odd size, but the intersections are all of even size, then* $|\mathcal{F}| \leq n$.

Proof View the incidence vectors of as vectors in $\mathbb{F}_2^n$. Then we claim that they are all linearly independent. Indeed the hypothesis implies that if $A \neq B \in \mathcal{F}$, $\langle 1_A, 1_A \rangle = 1$ and $\langle 1_A, 1_B \rangle = 0$. Thus, if $1_{A_1} + \cdots + 1_{A_m} = 0$ for some distinct $A_1, \ldots, A_m \in \mathcal{F}$, then $1 = \langle 1_{A_1}, 1_{A_2} + 1_{A_3} + \cdots + 1_{A_m} \rangle = \langle 1_{A_1}, 1_{A_2} \rangle + \langle 1_{A_1}, 1_{A_3} \rangle + \cdots + \langle 1_{A_1}, 1_{A_m} \rangle = 0$, which is a contradiction. ■

Now let us flip things around.

Theorem 5. *If every set of $\mathcal{F}$ is of even size, but the intersections are all of odd size, then* $|\mathcal{F}| \leq n$.

Observe that Theorem 4 shows that $|\mathcal{F}| \leq n + 1$: simply add an element to the universe and each set to obtain a family of odd sized sets on $n + 1$ elements with even intersections.

Proof Once again, we consider the vectors $1_A \in \mathbb{F}_2^n$ for sets A in the family. This time we have $\langle 1_A, 1_A \rangle = 0$, and $\langle 1_A, 1_B \rangle = 1$ for distinct A, B. For the sake of contradiction, assume that there is such a family of size $n + 1$. Then, the vectors 1_A must be linearly dependent, so there is some non-trivial linear combination $\sum_A \alpha_A \cdot 1_A = 0$. For any $B \in \mathcal{F}$, we have

$$0 = \left\langle 1_B, \sum_A \alpha_A \cdot 1_A \right\rangle = \sum_{A \neq B} \alpha_A. \tag{1}$$

By repeating this for another set B', we obtain $0 = \sum_{A \neq B} \alpha_A + \sum_{A \neq B'} \alpha_A = \alpha_{B'} + \alpha_B$. Thus, all the coefficients α_A must be equal, and so must all be 1, i.e. $\sum_{A \in \mathcal{F}} 1_A = 0$.

Equation 1 implies that n is even. Now consider the family $\mathcal{F}'$ that consists of the complements of the sets of $\mathcal{F}$. Every set in this family has even size, and the intersections are all odd, since the intersection of two sets in $\mathcal{F}'$ is equal to the complement of the union of two sets $A, B \in \mathcal{F}$, and $|A \cup B| = |A| + |B| - |A \cap B|$, which is odd.

Thus, we obtain that $0 = \sum_{A \in \mathcal{F}} 1_A + \sum_{A \in \mathcal{F}} 1_{A^c}$, where A^c denotes the complement of A. Since $1_A + 1_{A^c} = 1_{[n]}$, we get that $(n+1) \cdot 1_{[n]} = 0$, but this is a contradiction, since $n+1$ is odd. ■

2 Fisher's Inequality

Theorem 6. *Let $\mathcal{F}$ be a family of sets on $[n]$ such that all pairs of sets intersect in the same number of points $k > 0$. Then $|\mathcal{F}| \leq n$.*

Proof For each set A, let $1_A \in \mathbb{R}^n$ be the incidence vector. We will show that these must be linearly independent.

Indeed, if $\sum_A \alpha_A \cdot 1_A = 0$ is a non-trivial linear combination, then

$$\begin{aligned}
0 &= \left\langle \sum_A \alpha_A \cdot 1_A, \sum_A \alpha_A \cdot 1_A \right\rangle \\
&= \sum_A \alpha_A^2 \langle 1_A, 1_A \rangle + \sum_{A \neq B} \alpha_A \alpha_B \langle 1_A, 1_B \rangle \\
&= \sum_A \alpha_A^2 |A| + k \sum_{A \neq B} \alpha_A \alpha_B \\
&= \sum_A \alpha_A^2 (|A| - k) + k \left(\sum_A \alpha_A \right)^2 \\
&> 0,
\end{aligned}$$

which is a contradiction. ■

3 Lines from Points

Theorem 7. *Let $S \subset \mathbb{R}^d$ be a collection of n points. Pass a line through each pair of points. Then we must obtain either one line, or at least n lines.*

Proof Suppose all the points are not co-linear. Then define A_i to be the set of lines through i'th point. Then observe that for distinct i, j, $|A_i \cap A_j| = 1$. Thus, the number of such sets is at most the number of lines by Theorem 6. If all points do not lie on a single line, then each $|A_i| \geq 2$, which implies that $A_i \neq A_j$ for i, j distinct, since any two lines can pass through at most 1 common point. Thus the number of sets is n, which implies that there are at least n lines. ■

4 Sets with few intersection sizes

Given functions that map a set S to a field $\mathbb{F}$, we can view the functions as vectors in a vector space: the i'th coordinate is the evaluation of the function on i. We have the following simple lemma.

Lemma 8. *Suppose $f_1, \dots, f_k$ are functions, and $x_1, \dots, x_k$ are points such that $f_i(x_j) = 0$ if $i > j$, and $f_i(x_j) \neq 0$ if $i = j$. Then $f_1, \dots, f_k$ are linearly independent.*

Proof If not then $\sum_i \alpha_i f_i = 0$. Let i be the minimum non-zero coefficient. Evaluate the linear combination on x_i to obtain $\alpha_i f_i(x_i) = 0$ which means $\alpha_i = 0$, a contradiction. ■

Next, we generalize Fisher's inequality.

Theorem 9. *If $\mathcal{F}$ is a family of sets, and L is a set of integers such that for all $A \neq B \in \mathcal{F}$, $|A \cap B| \in L$, then $|\mathcal{F}| \leq \sum_{i=0}^{|L|} \binom{n}{i}$.*

Observe that the theorem is tight: consider all sets of size at most $|L|$.

Proof Let $A_1, A_2, \dots, A_m$ be the sets of the family, in order of increasing size. Then define the multivariate polynomials in the variables $X = X_1, \dots, X_n$,

$$f_i(X) = \prod_{\ell < |A_i|, \ell \in L} (\langle 1_A, X \rangle - \ell)$$

Observe that $f_i(1_{A_j}) = 0$ if $i > j$, and $f_i(1_{A_j}) \neq 0$ if $i = j$. Thus, by Lemma 8, the functions are all linearly independent.

Next we shall show that all of these functions live in a small linear space. Indeed, each function is defined by a polynomial of degree at most $|L|$. Further, we can replace $X_i^2 = X_i$ to obtain polynomials that compute the same functions on the indicator vectors and have degree at most 1 in each variable. The set of such polynomials is spanned by the number of multilinear monomials of degree at most $|L|$. This proves the bound. ■