Lecture 4

Aleksandar Nikolov

1 Computing Discrepancy

In a previous lecture, we showed that given $A \in [-1,1]^{m\times n}$, we can compute $x \in \{\pm 1\}^m$ such that $\|Ax\|_\infty = O(\sqrt{n \log(2m/n)})$. This leads to further questions:

- Assume $\mathrm{disc}(A)$ is small (e.g. $\mathrm{disc}(A) = 0$), can we compute x that does better than Spencer's bound (that is, $o(\sqrt{n\log(2m/n)})$ in *polylog*(m,n) time?
- More generally, can we efficiently approximate combinatorial discrepancy?

Unfortunately, the answer to both these questions is that we cannot (unless $P = NP$).

Theorem 1 ([3]). *For $A \in \{0,1\}^{O(n)\times n}$, it is $NP-$hard to distinguish between the cases:*

1. $\mathrm{disc}(A) = 0$
2. $\mathrm{disc}(A) = \Omega(\sqrt{n})$ *(Spencer's bound)*

Note: The same holds for set systems, where A is the incidence matrix.

In some sense this hardness is due to the lack of robustness of discrepancy. For example, if we take any matrix A, possibly with maximum discrepancy, and make another copy of every one of its columns, we would get a matrix of discrepancy 0. Recall, however, that we have seen a more robust notion of discrepancy back in our first lecture. This is *hereditary discrepancy*, defined for a matrix A as

$$\mathrm{herdisc}\, A = \max_{S\subseteq[n]} \mathrm{disc}(A_S)$$

where $A \in \mathbb{R}^{m\times n}$ and A_S the matrix which consists of the columns of A which are indexed by S. For set systems $(\mathcal{S},\mathcal{U})$, this is equivalent to the maximum discrepancy of any set system induced by a subset of $\mathcal{U}$. It turns out that hereditary discrepancy can be efficiently approximated. This is the topic of this lecture.

Theorem 2 ([6]). *There exists a polytime computable function γ_2 such that $\forall A \in \mathbb{R}^{m\times n}$:*

$$\frac{\gamma_2(A)}{O(\log(m))} \le \mathrm{herdisc}\,(A) \le O(\sqrt{\log m})\gamma_2(A)$$

The notation $\gamma_2(\cdot)$ in the theorem is standard and we keep it to be consistent with the literature. We are going to define this function shortly.

2 Upper bounds from Factorization

Given that hereditary discrepancy is a maximum over exponentially many quantities (each of them itself hard to compute), our first task is to develop a good efficiently computable upper bound on hereditary discrepancy. We derive such an upper bound from a powerful theorem of Banaszczyk.

Let us start with some notation. We will denote the largest norm of a row/column of a matrix A by:

$$r(A) = \max_{i=1}^{m} \|a_{i*}\|_2,$$
$$c(A) = \max_{j=1}^{n} \|a_{*j}\|_2.$$

Exercise 1. *Show that for any $m \times n$ matrix A, $\operatorname{disc}(A) = O(\sqrt{\log m})r(A)$.*

In addition to this exercise, recall that in an exercise from the last lecture we showed that $\operatorname{disc}(A) = O(\sqrt{\log m})c(A)$. We will use the following theorem by Banaszczyk to combine these two bounds.

Theorem 3 ([1])**.** *Let $K \subseteq \mathbb{R}^m$ be convex and closed, with:*

$$\mathbb{P}(g \in K) \geq 1/2 \text{ where } g \sim N(0, I)$$

Then for $A \in \mathbb{R}^{m \times n}$, $\exists x \in \{\pm 1\}^n$ s.t. $Ax \in 5 \cdot c(A) \cdot K$

A recent algorithmic proof of this theorem was given in [2], using the tools we saw in the last lecture.

We are now ready to state our good upper bound on hereditary discrepancy.

Theorem 4 ([4])**.** *For $A \in \mathbb{R}^{m \times n}$ with $A = UV$, where U, V arbitrary, we have that:*

$$\operatorname{disc}(A) \leq r(U) \cdot c(V) \cdot O(\sqrt{\log 2m})$$

Proof. Define K as follows:

$$K = \{y : \|Uy\|_\infty \leq 2 \cdot r(U) \cdot \sqrt{\log{(2m)}}\}$$

Exercise 2. *Show that for a standard Gaussian $g \sim N(0, I)$, $\Pr(g \in K) \geq \frac{1}{2}$.*

The above exercise means we can apply Theorem 3 with K and the matrix V:

$$\exists x \in \{\pm 1\}^n \text{ s.t. } Vx \in 5 \cdot c(V) \cdot K$$

$$\Leftrightarrow \operatorname{disc}(A) \leq \|UVx\|_\infty \leq 10 \cdot c(V) \cdot r(U) \cdot \sqrt{\log 2m}$$

∎

Definition 5 (γ_2 norm)**.** *We can define the γ_2 norm of a matrix $A \in \mathbb{R}^{n\times m}$ as:*

$$\gamma_2(A) = \min\{r(U)\cdot c(V) : UV = A\}$$

Theorem 4 then becomes:

$$\mathrm{disc}(A) = \gamma_2(A)\cdot O\left(\sqrt{\log(2m)}\right)$$

We additionally note that since $A_S = UV_S$, then $c(V_S) \le c(V) \Rightarrow \gamma_2(A_S) \le \gamma_2(A)$, and, therefore,

$$\mathrm{herdisc}\, A \le \gamma_2(A)\cdot O\left(\sqrt{\log(2m)}\right).$$

This proves the right hand side inequality in Theorem 2

3 Vector Program for γ_2

A *vector program* is an optimization problem with vector variables $\{v_i\}_{i=1}^n \in \mathbb{R}^n$, whose objective function and constraints are linear in $\langle v_i, v_j\rangle$ where $i, j \in [n]$. It is known that every vector program can be solved efficiently by recasting it as a semidefinite program (SDP). Thus, if we can show that $\gamma_2(A)$ can be written as a vector program, this suffices to show that it is efficiently computable.

Lemma 6. *For $A \in \mathbb{R}^{m\times n}$, $\gamma_2(A)$ can be written as*

$$\begin{array}{cl} \min & t \\ \textit{subject to} & \langle u_i, v_j\rangle = A_{ij} \\ & \langle u_i, u_i\rangle \le t \\ & \langle v_j, v_j\rangle \le t \\ & u_i, v_j \in \mathbb{R}^{m+n} \\ \textit{where} & (i,j) \in [m]\times[n], \end{array}$$

Exercise 3. *Prove Lemma 6.*

4 γ_2 and `herdisc`: the Lower Bound

It remains to show the first inequality in Theorem 2. We are going to use the *determinant lower bound* on hereditary discrepancy, due to Lovász, Spencer, and Vesztergombi.

Theorem 7 ([5])**.** *The quantity*

$$\mathrm{detlb}\,(A) = \max_{k=1}^{\min(m,n)} \max_{\substack{S\subseteq[m]\\ T\subseteq[n]\\ |S|=|T|=k}} |\det A_{S,T}|^{1/k},$$

where $A_{S,T}$ is the submatrix of A indexed by S and T, satisfies:

$$\operatorname{herdisc}(A) \geq \frac{1}{2} \operatorname{detlb}(A).$$

Then, to prove the first inequality in Theorem 2 it suffices to show

$$\operatorname{detlb}(A) \geq \gamma_2(A) \cdot \Omega\left(\frac{1}{\log \operatorname{rank} A}\right). \tag{1}$$

Dual Characterization of γ_2: The vector program in Lemma 6 and the duality theory for conic programming imply that $\gamma_2(A)$ can also be written as the value of a maximization problem, namely

$$\begin{aligned} \gamma_2(A) = \max \quad & \|B\|_{tr} \\ \text{subject to} \quad & B_{ij} = p_i q_j A_{ij} \\ & \sum_{i=1}^{m} p_i^2 = \sum_{j=1}^{n} q_j^2 = 1 \\ & p_i, q_j \geq 0 \\ \text{where} \quad & (i,j) \in [m] \times [n], \end{aligned}$$

where $\|B\|_{tr}$ is the trace or nuclear norm, which is equal to the sum of the singular values of B.

Proof of (1)*.* We use an elementary but useful fact, given in the following exercise.

Exercise 4. *Show that for any $\sigma_1 \geq \ldots \geq \sigma_r \geq 0$ there exists an integer k such that*

$$\sum_{i=1}^{r} \sigma_i \leq O(\log r) \cdot k \left(\prod_{i=1}^{k} \sigma_i\right)^{1/k}.$$

Notice that this is sort of a converse to the AM-GM inequality.

Let us now take a feasible solution (B, p, q) to the dual maximization problem for $\gamma_2(A)$. This implies that $\gamma_2(A) = \|B\|_{tr}$.

Now, let the singular value decomposition (SVD) of B be $B = U\Sigma V^T$. Here, $r = \operatorname{rank} B$, $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$, $U^T U = I$ and Σ a diagonal matrix with the singular values $\sigma_1 \geq \ldots \geq \sigma_r$ of B on the diagonal. Let k be as in the exercise above. If we define $C := U_k^T B$, where U_k is the matrix whose columns are the singular vectors of B corresponding to $\sigma_1, \ldots, \sigma_k$, then the singular values of C are $\sigma_1, \ldots, \sigma_k$. This means that:

$$\left|\det CC^T\right|^{\frac{1}{2k}} = \left|\prod_{i=1}^{k} \sigma_i\right|^{\frac{1}{k}} \geq \frac{1}{O(k \log r)} \sum_{i=1}^{r} \sigma_i = \frac{1}{O(k \log r)} \|B\|_{tr}. \tag{2}$$

Cauchy-Binet Formula: For $X, Y \in \mathbb{R}^{m\times n}$:

$$\det XY^T = \sum_{\substack{S\subseteq[n]\\|S|=m}} \det X_S \det Y_S$$

If we define $P \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{n\times n}$ as the diagonal matrices with p_i and q_j as the diagonal entries respectively, then we have $B = PAQ$. Similarly, $C = U_k^T B = U_k^T PAQ$.

Define $D := U_k^T PA$ so that $C = DQ$. By applying Cauchy-Binet to $C \in \mathbb{R}^{k\times n}$ we get:

$$\det(CC^T) = \sum_{\substack{S\subseteq[n]\\|S|=k}} \det C_S \det C_S = \sum_{\substack{S\subseteq[n]\\|S|=k}} (\det C_S)^2 = \sum_{\substack{S\subseteq[n]\\|S|=k}} (\det D_S Q_S)^2$$

$$= \sum_{\substack{S\subseteq[n]\\|S|=k}} (\det D_S)^2 \left(\prod_{j\in S} q_j^2\right) \leq \left(\max_{\substack{S\subseteq[n]\\|S|=k}} (\det D_S)^2\right)\left(\sum_{\substack{S\subseteq[n]\\|S|=k}} \prod_{j\in S} q_j^2\right).$$

By comparing monomials, we see that

$$\sum_{\substack{S\subseteq[n]\\|S|=k}} \prod_{j\in S} q_j^2 \leq \frac{1}{k!}\left(\sum_{j=1}^{n} q_j^2\right)^k = \frac{1}{k!}.$$

Thus,

$$\max_{\substack{S\subseteq[n]\\|S|=k}} |\det D_S|^{1/k} \geq (k!)^{1/2k} \cdot (\det CC^T)^{1/2k}$$

Together with Stirling's approximation, and (2), we get for some $S \subseteq [n]$, $|S| = k$,

$$(\det D_S)^{1/k} \geq \frac{\|B\|_{tr}}{O(\sqrt{k}\log r)}. \tag{3}$$

Consider the orthonormal matrix $W \in \mathbb{R}^{m\times m}$ for which the first k columns are equal to the columns of U_k. Such a matrix always exists since we can complete the orthonormal basis for $\mathbb{R}^m$ starting with the column vectors of U_k. The $m-k$ new vectors we get can be used to define the rest of the columns of W.

Define $E_S := PA_S \in \mathbb{R}^{m\times k}$, meaning that $D_S = U_R^T E_S$. It can be shown that:

$$\det(E_S^T E_S) = \det((E_S^T W)(W^T E_S)) = \det((E_S^T W)(E_S^T W)^T)$$

$$= \sum_{\substack{T\subseteq[n]\\|T|=k}} \det((E_S^T W)_T)^2 = \sum_{\substack{T\subseteq[n]\\|T|=k}} \det(E_S^T W_T)^2 \geq \det(E_S^T U_k)^2 = \det(U_k^T E_S)^2$$

$$\therefore \det(E_S^T E_S) \geq \det(D_S)^2$$

Now, we can apply the exact same analysis as in (3), but this time to $D_S^T = (A_S)^T P$ instead of C. This means that $\exists T \in [m]$ for which:

$$\max_{\substack{T\subseteq[m]\\|T|=k}} (\det A_{S,T})^{1/k} \geq (k!)^{1/2k} \cdot \det\left(A_S^T P^2 A_S\right)^{1/2k} = (k!)^{1/2k} \cdot \det\left(E_S^T E_S\right)^{1/2k}$$

Putting all of this together and applying Stirling just like before, we get that:

$$\max_{\substack{S\subseteq[n]\\T\subseteq[m]\\|S|=|T|=k}} |\det A_{S,T}|^{1/k} \geq \frac{\|B\|_{tr}}{\Omega(\log(2r))}$$

By maximizing over all k, this yields the desired result. □

Exercise 5. *Show the following properties of $\gamma_2(A)$:*

1. $\gamma_2(A) = \gamma_2(A^T)$;
2. $\gamma_2(A+B) \leq \gamma_2(A) + \gamma_2(B)$;
3. $\gamma_2(A \otimes B) = \gamma_2(A)\gamma_2(B)$, *where $A \otimes B$ is the Kronecker product.*

References

[1] Wojciech Banaszczyk. Balancing vectors and Gaussian measures of n-dimensional convex bodies. *Random Structures Algorithms*, 12(4):351–360, 1998.

[2] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The gram-schmidt walk: a cure for the banaszczyk blues. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 587–597. ACM, 2018.

[3] Moses Charikar, Alantha Newman, and Aleksandar Nikolov. Tight hardness results for minimizing discrepancy. In Dana Randall, editor, *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 1607–1614. SIAM, 2011.

[4] Kasper Green Larsen. On range searching in the group model and combinatorial discrepancy. *SIAM J. Comput.*, 43(2):673–686, 2014.

[5] L. Lovász, J. Spencer, and K. Vesztergombi. Discrepancy of set-systems and matrices. *European J. Combin.*, 7(2):151–160, 1986.

[6] Jiri Matousek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy. *International Mathematics Research Notices*, page rny033, 2018.