CSC2414: Discrepancy Theory in Computer Science	Fall 2015
Lecture 7 — November 2, 2015	
Aleksandar Nikolov	*Scribe: Assimakis Kattis*

1 Computing Discrepancy

Last 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 disc(A) is small (e.g. 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 ([CNN11]). *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.

Here we prove the above for $A \in \{-4,-3,...,4\}^{O(n)\times n}$.

Theorem ([Gur03]). *[2-2 Set Splitting]*

For $S = \{S_1,...,S_m\}$ with $m = O(n)$, if $\forall i$ $|S_i| = 4$ and $S_i \subseteq [n]$, where each $j \subseteq [n]$ appears in at most 4 sets, it is $NP-$hard to distinguish between the cases:

1. $\mathrm{disc}(S) = 0$
2. $\forall x \in \{\pm1\}^n$, $|\{i : |\sum_{j\in S_i} x_j| \neq 0\}| \geq \alpha m$, *where $\alpha \approx 1/22$ a small constant*

Suffices to reduce from this to our problem.

Reduction: Set $B \in \{0,1\}^{m\times n}$ the incidence matrix of S. This has dimension $m \times n$, with 4 1s in each row/column, by construction of S. Set $H \in \{\pm1\}^{m\times m}$, the Hadamard matrix, which has the property that $H^T H = mI$. We denote the ith row of H as h_i and the jth column of B as b_j.

Define $A = HB$, where $A_{ij} = h_i \cdot b_j$. Note that since there are at most 4 non-zero entries in b_j and $h_i \in \{\pm1\}^m$, we have that $A_{ij} \in \{-4,...,4\}$.

We now prove that this reduction is sufficient for our purposes.

Claim. $\text{disc}\,(S) = 0 \Rightarrow \text{disc}\,(A) = 0$.

Proof. For $\text{disc}\,(S) = 0$, by definition $\exists x \in \{\pm 1\}^n$ s.t. $Bx = 0$. This implies that $Ax = HBx = H0 = 0$ and so $\text{disc}\,(A) = 0$. □

Claim. $\forall x \in \{\pm 1\}^n$, $|\{i : |\sum_{j \in S_i} x_j| \neq 0\}| \geq \alpha m \Rightarrow \text{disc}(A) = \Omega(\sqrt{n})$

Proof. By assumption we know that the number of sets S_i for which $|\sum_{j \in S_i} x_j| \neq 0$ is bounded below by αm. Since the total number of such sets is an integer and thus at least 1, we have that:

$$\|Bx\|_2^2 = \sum_{i=1}^{m} \left(\sum_{j \in S_i} x_j \right)^2 \geq \alpha m$$

By comparing maximum to average discrepancy, we get that:

$$\|HBx\|_\infty^2 \geq \frac{1}{m} \|HBx\|_2^2 = \frac{1}{m} x^T B^T H^T H B x = x^T B^T B x = \|Bx\|_2^2 \geq \alpha m$$
$$\therefore \text{disc}(A) = \text{disc}(HB) \geq \sqrt{\alpha m} = \Omega(\sqrt{n})$$

□

2 Hereditary Discrepancy

We can define the (stronger) notion of hereditary discrepancy as follows:

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

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 hereditary discrepancy, there are a number of things that we can say about its approximations.

Theorem ([AGH13]). *$\forall \epsilon > 0$, it is $NP-$hard to obtain a $(2 - \epsilon)$ approximation to* $\text{herdisc}\,(A)$ *better than a factor of* 2*.*

Theorem ([NT15]). *There exists a polytime computable function f such that $\forall A \in \mathbb{R}^{m \times n}$:*

$$\frac{f(A)}{C \log{(m)}^{3/2}} \leq \text{herdisc}\,(A) \leq f(A)$$

We begin by looking at upper bounds for hereditary discrepancy:

Norms of Rows: For a_{i*} the i-th row of A, we can define:

$$r(A) = \max_{i=1}^{m} \|a_{i*}\|_2$$

For $x \in \{\pm 1\}^n$ chosen uniformly at random $\Rightarrow_{w.h.p} \|Ax\|_\infty = O(\sqrt{\log 2m}) \cdot r(A)$

$$\therefore \operatorname{disc}(A) = r(A) \cdot O(\sqrt{\log 2m}) \tag{1}$$

We notice that $(Ax)_i = \langle a_{i*}, x \rangle$ and since the projection of Euclidian distance never increases, this implies that $r(A_S) \leq r(A)$. Thus:

$$\operatorname{herdisc}(A) = \max_S \operatorname{disc}(A_S) = \max_S r(A_S) \cdot O(\sqrt{\log 2m}) = r(A) \cdot O(\sqrt{\log 2m})$$

Norms of Columns: For a_{*i} the i-th column of A, we can similarly define:

$$c(A) = \max_{i=1}^{n} \|a_{*i}\|_2$$

Some relevant results and conjectures pertaining to this definition are shown below.

Beck-Fiala Theorem: For $A \in \{0,1\}^{m \times n}$, we have $\operatorname{disc}(A) \leq 2 \max_j \#\{1\text{'s in } a_{*j}\} = 2 \cdot c(A)^2$

Beck-Fiala Conjecture: For $A \in \{0,1\}^{m \times n}$, $\operatorname{disc}(A) = O(1) \cdot c(A)$

Komlós Conjecture: $\forall A \in \mathbb{R}^{m \times n}$, $\operatorname{disc}(A) = O(1) \cdot c(A)$

Banaszczyk: $\forall A \in \mathbb{R}^{m \times n}$, $\operatorname{disc}(A) = O(\sqrt{\log 2m}) \cdot c(A)$

Finally, since $c(A_S) \leq c(A)$, the above result implies:

$$\operatorname{herdisc}(A) \leq c(A) \cdot O(\sqrt{\log 2m})$$

Combining with (1), the above also yields:

$$\operatorname{disc}(A) \leq \min\{c(A), r(A)\} \cdot \mathit{polylog}(m)$$

3 Factorization

Theorem ([Ban98])**.** *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. $Ax \in 5 \cdot c(A) \cdot K$

Theorem ([Lar14])**.** *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:

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

We will denote u_i as the ith row of U.

Theorem (Gaussian Concentration Inequality)**.** $h \sim N(0, \sigma^2) \Rightarrow \mathbb{P}(|h| > t\sigma) \leq e^{-t^2/2}$

For $g \sim N(0, I)$, we can bound $\mathbb{P}(g \in K)$ as follows:

$$\mathbb{P}(g \in K) = \mathbb{P}\left(|\langle u_i, g\rangle| \leq 2 \cdot r(U) \cdot \sqrt{\log{(2m)}}, \forall i \in [m]\right)$$

$$\geq 1 - \sum_{i=1}^{m} \mathbb{P}\left(|\langle u_i, g\rangle| \geq 2 \cdot r(U) \cdot \sqrt{\log{(2m)}}\right)$$

$$\geq 1 - \sum_{i=1}^{m} \frac{1}{2m} \geq 1/2$$

The last inequality follows by the property that for $g \sim N(0, I)$, we have $\langle u_i, g\rangle \sim N(0, \|u_i\|_2^2)$. By setting $t = 2\sqrt{\log{(2m)}}$ and noticing that by definition $\forall i, \|u_i\|_2^2 \leq r(U)^2$, applying the Gaussian concentration inequality above yields the lower bound.

Now, the above result means we can apply the result of [Ban98] to K with matrix V:

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

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

□

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

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

The above result then becomes:

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

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

Efficient computation of γ_2

Definition. *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 approximated efficiently by a semidefinite program (SDP). Thus, if we can show that $\gamma_2(A)$ can be written as a vector program, this suffices in showing that it is efficiently computable.

Claim. *For $A \in \mathbb{R}^{m\times n}$, $\gamma_2(A)$ can be written as the following vector program:*

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

Proof. Denote t^* the optimal solution, with u_i^*, v_j^* corresponding vectors.

We first show $\gamma_2(A) \leq t^*$.

If we define U as having u_i^* as its ith row and V as having v_j^* as its jth column, then $(UV)_{ij} = \langle u_i^*, v_j^*\rangle = A_{ij}$. Thus, we get that $UV = A$.

Since $\forall i, \langle u_i^*, u_i^*\rangle \leq t^*$, this implies that $\forall i, \|u_i^*\|_2^2 \leq t^*$, or that $r(U)^2 \leq t^*$. Similarly, $c(V)^2 \leq t^*$.

As U and V satisfy $A = UV$, we have that:

$$\therefore \gamma_2(A) \leq r(U)\cdot c(V) \leq t^*$$

We now show that $t^* \leq \gamma_2(A)$.

Pick U, V, the optimal matrices for which $r(U)\cdot c(V) = \gamma_2(A)$.

Setting $\alpha = \sqrt{\frac{c(V)}{r(U)}}$, we have that $A = (\alpha U)(\frac{1}{\alpha}V)$.

Now define u_i the ith row of αU and v_j the jth row of $(1/\alpha)V$.

$$r(\alpha U) = \sqrt{\frac{c(V)}{r(U)}}r(U) = \sqrt{c(V)\cdot r(U)}$$

$$c((1/\alpha)V) = \sqrt{\frac{r(U)}{c(V)}}c(V) = \sqrt{c(V)\cdot r(U)}$$

The two equalities above imply that:

$$\langle u_i, u_i\rangle = \|u_i\|_2^2 \leq r(\alpha U)^2 = \frac{c(V)}{r(U)}r(U)^2 = c(V)\cdot r(U) = \gamma_2(A)$$

A similar argument shows this for $\langle v_j, v_j\rangle$. However, since t^* is the minimum t for which $\langle u_i, u_i\rangle \leq t$, this implies that $t^* \leq \gamma_2(A)$. Thus, $t^* = \gamma_2(A)$. $\square$

References

[AGH13] Per Austrin, Venkatesan Guruswami, and Johan Håstad. Sat is np-hard. In *Electronic Colloquium on Computational Complexity, TR13-159*, 2013.

[Ban98] Wojciech Banaszczyk. Balancing vectors and gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[CNN11] Moses Charikar, Alantha Newman, and Aleksandar Nikolov. Tight hardness results for minimizing discrepancy. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 1607–1614. SIAM, 2011.

[Gur03] Venkatesan Guruswami. Inapproximability results for set splitting and satisfiability problems with no mixed clauses. *Algorithmica*, 38(3):451–469, 2003.

[Lar14] Kasper Green Larsen. On range searching in the group model and combinatorial discrepancy. *SIAM Journal on Computing*, 43(2):673–686, 2014.

[NT15] Aleksandar Nikolov and Kunal Talwar. Approximating hereditary discrepancy via small width ellipsoids. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 324–336. SIAM, 2015.