CSC2414: Discrepancy Theory in Computer Science Fall 2015

Lecture 4 — 5th October, 2015

Aleksandar Nikolov *Scribe: Nick Spooner*

1 Lower bounds on combinatorial discrepancy

We recall the definition of combinatorial discrepancy from the previous lecture. Let $\mathcal{U}$ be a set with $|\mathcal{U}| = n$. Without loss of generality, we can take $\mathcal{U} = \{1, \dots, n\}$. Let $\mathcal{S} \subset 2^{\mathcal{U}} := \{S_1, \dots, S_m\}$ be a family of subsets of $\mathcal{U}$; $|\mathcal{S}| = m$. The *combinatorial discrepancy of* $\mathcal{S}$, $\operatorname{disc} \mathcal{S}$, is defined to be

$$\operatorname{disc}(\mathcal{S}) := \min_{\chi : \mathcal{U} \to \{-1,+1\}} \max_{S \in \mathcal{S}} |\chi(S)| \ ,$$

where $\chi(S) := \sum_{j \in S} \chi(j)$. χ is a colouring of the elements of $\mathcal{U}$ with ± 1, and so $\operatorname{disc}(\mathcal{S})$ can be thought of as a measure of the 'balancedness' (over $\mathcal{S}$) of any such colouring.

1.1 Matrix discrepancy

We introduce the following 'matrix notation' for combinatorial discrepancy, which motivates the study of matrix discrepancy.

Let A be the incidence matrix of $\mathcal{S}$, i.e. $A \in \{0,1\}^{m \times n}$ such that

$$A_{ij} = \begin{cases} 1 & \text{if } j \in S_i, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Then we can write $\operatorname{disc}(\mathcal{S})$ in terms of A, i.e.

$$\operatorname{disc}(\mathcal{S}) = \min_{x \in \{-1,1\}^n} \|Ax\|_\infty$$

where $\|v\|_\infty$ for $v \in \mathbb{R}^n$ is the ∞-norm of v, $\|v\|_\infty := \max_{i \in \{1,\dots,n\}} |v_i|$.

We can generalise this notion by allowing A to be any matrix in $\mathbb{R}^{m \times n}$, and hence we can define for $A \in \mathbb{R}^{m \times n}$ the *matrix discrepancy of* A,

$$\operatorname{disc}(A) := \min_{x \in \{-1,1\}^n} \|Ax\|_\infty \ .$$

1.2 The eigenvalue lower bound

Recall that the singular values of a matrix $A \in \mathbb{R}^{m \times n}$ are the square roots of the eigenvalues of $A^T A$. Let $\sigma_1 \geq \ldots \geq \sigma_n$ be the singular values of A. The smallest singular value of A, σ_n, satisfies the following *variational characterisation*:

$$\sigma_n^2 = \min_{x \in \mathbb{R}^n} \frac{x^T A^T A x}{x^T x} = \min_{x \in \mathbb{R}^n} \frac{\|Ax\|_2^2}{\|x\|_2^2} \ ,$$

where $\|x\|_2$ is the Euclidean norm of x.

Proposition 1 (Eigenvalue lower bound). *For any $A \in \mathbb{R}^{m\times n}$, $\operatorname{disc}(A) \geq \sqrt{\frac{n}{m}}\,\sigma_n$, where σ_n is the smallest singular value of A.*

Proof. By the definition of $\operatorname{disc}(A)$ and the ∞-norm, we have

$$\operatorname{disc}(A) = \min_{x\in\{-1,1\}^n} \|Ax\|_\infty = \min_{x\in\{-1,1\}^n} \sqrt{\max_{i\in\{1,\ldots,m\}} (Ax)_i^2} \geq \min_{x\in\{-1,1\}^n} \sqrt{\frac{1}{m}\sum_{i=1}^m (Ax)_i^2}\ ,$$

where the inequality follows because the maximum is larger than the average. Observe that the expression to be minimised on the right is exactly $\frac{1}{\sqrt{m}}\|Ax\|_2$. Noting also that for $x \in \{-1,1\}^n$, $\|x\|_2 = \sqrt{n}$, we obtain

$$\min_{x\in\{-1,1\}^n} \sqrt{\frac{1}{m}\sum_{i=1}^m (Ax)_i^2} = \min_{x\in\{-1,1\}^n} \sqrt{\frac{n}{m}}\cdot\frac{\|Ax\|_2}{\|x\|_2} \geq \min_{x\in\mathbb{R}^n} \sqrt{\frac{n}{m}}\cdot\frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\frac{n}{m}}\,\sigma_n,$$

since the minimum over $x \in \{-1,1\}^n$ is no smaller than the minimum over $x \in \mathbb{R}^n \supset \{-1,1\}^n$. $\square$

Example 2. Consider the Hadamard matrix $H_k \in \{-1,1\}^{2^k\times 2^k}$, defined recursively as follows:

$$H_0 := \begin{pmatrix} 1 \end{pmatrix} \qquad H_k := \begin{pmatrix} H_{k-1} & H_{k-1} \\ H_{k-1} & -H_{k-1} \end{pmatrix}$$

We have that $H_k^T H_k = 2^k \cdot I$, hence $\sigma_n = \sqrt{n}$ (where $n = 2^k$), and so $\operatorname{disc}(H_k) \geq \sqrt{n}$.

The probabilstic argument from the previous lecture gives an upper bound for $\operatorname{disc}(H_k)$ of $O(\sqrt{n\log n})$. An asymptotically tight upper bound follows from a matrix discrepancy version of Spencer's Theorem [1], also discussed in the previous lecture:

Lemma 3 (Spencer 85 [1]). *For all $A \in \{-1,1\}^{m\times n}$, $\operatorname{disc}(A) = O(\sqrt{n\log(2m/n)})$.*

2 Further discrepancy measures

2.1 Hereditary discrepancy

As a notion of complexity, the combinatorial discrepancy is somewhat fragile. To see this, we consider the universe $\mathcal{U} := \mathcal{U}^{(1)} \uplus \mathcal{U}^{(2)}$, where $\mathcal{U}^{(1)}$, $\mathcal{U}^{(2)}$ are disjoint. Let $\mathcal{S}^{(1)} = \{S_1^{(1)},\ldots,S_m^{(1)}\} \subseteq 2^{\mathcal{U}^{(1)}}$ and $\mathcal{S}^{(2)} = \{S_1^{(2)},\ldots,S_m^{(2)}\} \subseteq 2^{\mathcal{U}^{(2)}}$ such that $|S_i^{(1)}| = |S_i^{(2)}|$ for $i = 1,\ldots,m$. Let $\mathcal{S}' = \{S_i^{(1)} \cup S_i^{(2)} : i \in \{1,\ldots,m\}\}$; $\mathcal{S}' \subseteq 2^{\mathcal{U}'}$. Then regardless of the choice of $\mathcal{S}^{(1)}$ or $\mathcal{S}^{(2)}$, $\operatorname{disc}(\mathcal{S}) = 0$.

For this reason we introduce a more 'robust' notion of discrepancy. For $V \subseteq \mathcal{U}$, we write $\mathcal{S}|_V$ for the set $\{S \cap V : S \in \mathcal{S}\}$. Then the hereditary discrepancy of $\mathcal{S}$ is

$$\operatorname{herdisc}(\mathcal{S}) := \max_{V\subseteq\mathcal{U}} \operatorname{disc}(\mathcal{S}|_V)\ .$$

We can also define an analogous notion for matrix discrepancy. For a matrix $A \in \mathbb{R}^{m\times n}$ and $V \subseteq \{1,\ldots,n\}$, we write A_V for the matrix consisting of the columns of A indexed by V. Then

$$\operatorname{herdisc}(A) := \max_{V\subseteq\{1,\ldots,n\}} \operatorname{disc}(A_V)\ .$$

Observe that the notions correspond when A is the incidence matrix of $\mathcal{S}$.

2.2 Linear discrepancy

Next we will study a generalisation of combinatorial discrepancy. Suppose that each $i \in \mathcal{U}$ is assigned a weight $w(i) \in [-1, 1]$. The discrepancy of $\mathcal{S}$ with respect to w is

$$\mathrm{disc}^w(\mathcal{S}) := \min_{\chi:\mathcal{U}\to\{-1,1\}} \max_{S\in\mathcal{S}} |\chi(S) - w(x)| \ .$$

For $A \in \mathbb{R}^{m\times n}$ we can define the same notion, treating w as a vector in $[-1, 1]^n$:

$$\mathrm{disc}^w(A) := \|A(x - w)\|_\infty \ .$$

Note that in both cases the standard combinatorial discrepancy is given by $w(i) = 0$ for all $i \in \mathcal{U}$ (resp. $w = \vec{0}$). The *linear discrepancy* of $\mathcal{S}$ (resp. A) is the supremum of $\mathrm{disc}^w(\mathcal{S})$ (resp. $\mathrm{disc}^w(A)$) over all weight functions $w : \mathcal{U} \to [-1, 1]$ (resp. $w \in [-1, 1]^n$), and is written $\mathrm{lindisc}(\mathcal{S})$ (resp. $\mathrm{lindisc}(A)$).

Remark 4. Linear discrepancy is related to the problem of rounding solutions to relaxations of combinatorial optimization problems. In particular we can think of a solution to the relaxation as a vector of weights $w \in [0, 1]^n$, and a solution to the original problem as a vector $x \in \{0, 1\}^n$. Then $\mathrm{disc}^{w'}(A)$, for an appropriate matrix A and $w' = 2w - \vec{1}$, measures the approximation error when rounding w.

2.3 Relationships between discrepancy measures

It is clear that for any matrix A, $\mathrm{disc}(A) \le \mathrm{herdisc}(A)$ and $\mathrm{disc}(A) \le \mathrm{lindisc}(A)$. The following theorem shows that the linear discrepancy cannot be much larger than the hereditary discrepancy.

Theorem 5. *For $A \in \mathbb{R}^{m\times n}$, $\mathrm{lindisc}(A) \le 2\,\mathrm{herdisc}(A)$.*

Proof. We assume that all entries of w have a finite binary representation (note that any $v \in \mathbb{R}^n$ is arbitrarily close to such a vector). The proof is by induction on the length of this representation: in particular, let k be the smallest integer such that $w = \frac{v}{2^k}$ for some $v \in \mathbb{Z}^n$ (i.e., k is the maximum number of bits after the radix point in the binary representation of any entry in w). If $k = 0$, then $w \in \{-1, 0, 1\}^n$, and in this case $\mathrm{disc}^w(A) \le \mathrm{herdisc}(A)$, since setting $x_i = w_i$ when $w_i \in \{-1, 1\}$ gives $(A(x - w))_i = 0$ for $w_i \neq 0$, and so $\mathrm{disc}^w(A) = \mathrm{disc}(A_V)$ where $V = \{i : w_i = 0\}$.

For the induction step, we note that $2w \in [-2, 2]^n$, and so there must exist some $y \in \{-1, 1\}^n$ such that $z = 2w - y \in [-1, 1]^n$. Then there exists $v \in \mathbb{Z}^n$ such that $z = \frac{v}{2^{k-1}}$, and so by the induction hypothesis there exists some $x_0 \in \{-1, 1\}^n$ such that $\|A(x_0 - z)\|_\infty \le 2\,\mathrm{herdisc}(A)$. Then

$$\mathrm{herdisc}(A) \ge \frac{1}{2}\|A(x_0 - z)\|_\infty = \frac{1}{2}\|A(x_0 + y - 2w)\|_\infty = \|A(x_1 - w)\|_\infty \ ,$$

where $x_1 := \frac{1}{2}(x_0 + y) \in \{-1, 0, 1\}$. Let $V := \{i : (x_1)_i = 0\}$; then by definition of $\mathrm{herdisc}(A)$, there is some $x_2 \in \{-1, 1\}^V$ such that $\|A_V \cdot x_2\|_\infty \le \mathrm{herdisc}(A)$. We then take x to be as x_1 with its zero entries replaced with the corresponding entries in x_2, from which we obtain:

$$\|A(x - w)\|_\infty \le \|A_V \cdot x_2\|_\infty + \|A(x_1 - w)\|_\infty \le 2\,\mathrm{herdisc}(A) \ .$$

Hence, by induction, $\mathrm{lindisc}(A) \le 2\,\mathrm{herdisc}(A)$. □

3 Determinant lower bound

Let $A \in \mathbb{R}^{m \times n}$, and let P be the set $\{x \in \mathbb{R}^n : \|Ax\|_\infty \leq 1\}$, i.e. the set of $x \in \mathbb{R}^n$ such that for each row $\vec{a}_i$ of A, $-1 \leq \langle \vec{a}_i, x \rangle \leq 1$. We see that P is a convex polytope. For $m = n$, A invertible, we can also write P as

$$P = \{A^{-1}y : \|y\|_\infty \leq 1\} = A^{-1} \cdot [-1,1]^n \ ,$$

and hence the volume of P is given by $|\det(A^{-1})| \cdot 2^n = |\det(A)|^{-1} \cdot 2^n$.

Theorem 6 (Lovasz, Spencer and Vesztergombi [2]). *For any square $A \in \mathbb{R}^{n \times n}$, $\mathrm{lindisc}(A) \geq |\det(A)|^{1/n}$.*

Proof. Let $P = \{x \in \mathbb{R}^n : \|Ax\|_\infty \leq 1\}$. Then $\|A(x - w)\|_\infty \leq D$ if and only if $x - w \in DP$, i.e. $-w \in DP - x$. Hence $\mathrm{lindisc}(A) \leq D$ if and only if for all $w \in [-1,1]^n$ there exists $x \in \{-1,1\}^n$ such that $w \in DP - x$, which is the case if and only if $[-1,1]^n \subseteq \bigcup_{x \in \{-1,1\}^n} (DP - x)$. The latter implies, by the union bound, that $\mathrm{vol}([-1,1]^n) \leq \sum_{x \in \{-1,1\}^n} \mathrm{vol}(DP - x)$. The volume of $DP - x$ is simply the volume of DP, which is $D^n \,\mathrm{vol}(P)$; the volume of $[-1,1]^n$ is 2^n. Hence

$$2^n \leq 2^n D^n \,\mathrm{vol}(P) = 2^n D^n \cdot (|\det(A)|^{-1} \cdot 2^n) \ ,$$

so $D \geq \frac{1}{2}|\det(A)|^{1/n}$, from which the theorem follows. □

Corollary 7 (Determinant lower bound [2]). *For any $A \in \mathbb{R}^{m \times n}$,*

$$\mathrm{herdisc}(A) \geq \frac{1}{2} \max_{k=1}^{\min(m,n)} \max_{\substack{I \subseteq \{1,\ldots,m\} \\ |I|=k}} \max_{\substack{J \subseteq \{1,\ldots,n\} \\ |I|=k}} |\det(A_{I,J})|^{1/k} =: \mathrm{detlb}(A) \ .$$

Proof. Let $I \subseteq \{1, \ldots, m\}$, $J \subseteq \{1, \ldots, n\}$, $|I| = |J| = k$. Then $A_{I,J}$ is a submatrix of A, so $\mathrm{herdisc}(A) \geq \mathrm{herdisc}(A_{I,J})$. By Theorem 5, $\mathrm{herdisc}(A_{I,J}) \geq \frac{1}{2} \mathrm{lindisc}(A_{I,J})$. Then since $A_{I,J}$ is a $k \times k$ matrix, by Theorem 6, $\mathrm{lindisc}(A_{I,J}) \geq |\det(A_{I,J})|^{1/k}$. Combining the inequalities we obtain $\mathrm{herdisc}(A) \geq \frac{1}{2}|\det(A_{I,J})|^{1/k}$, and the corollary follows by taking the maximum over k, I, J. □

A result due to Matoušek shows that the above bound is almost tight.

Theorem 8 (Matoušek [3]). *For all $A \in \mathbb{R}^{m \times n}$, $\mathrm{herdisc}(A) \leq O\big(\log(mn)\sqrt{\log n}\big) \cdot \mathrm{detlb}(A)$.*

References

[1] Spencer, Joel. *Six standard deviations suffice.* Trans. Amer. Math. Soc. 289.2 (1985): 679–706.

[2] L Lovasz, J Spencer, and K Vesztergombi. 1986. *Discrepancy of set-systems and matrices.* Eur. J. Comb. 7, 2 (April 1986), 151–160.

[3] Matousek, J. (2011). The determinant bound for discrepancy is almost tight, 9. Combinatorics.