Discrepancy after adding a single set

Jeong Han Kim
Theory Group
Microsoft Research
Redmond, WA 98052
United States

Jiří Matoušek
Department of Applied Mathematics and
Institute of Theoretical Computer Science (ITI)
Charles University
Malostranské nám. 25, 118 00 Praha 1
Czech Republic

Van H. Vu
Department of Mathematics
University of California at San Diego
La Jolla, CA 92093
United States

Abstract

We show that the hereditary discrepancy of a hypergraph $\mathcal{F}$ on n points increases by a factor of at most $O(\log n)$ when one adds a new edge to $\mathcal{F}$.

Let X be a set of n points. We say that a hypergraph $\mathcal{F}$ on X has discrepancy h if h is the smallest integer satisfying the following: There is a coloring $\chi: X \to \{-1, +1\}$ such that for every edge $S \in \mathcal{F}$, $|\chi(S)| \leq h$, where we write $\chi(S)$ for $\sum_{x \in S} \chi(x)$.

The hereditary discrepancy of $\mathcal{F}$ is the maximum discrepancy of any restriction of $\mathcal{F}$ to a subset $Y \subseteq X$. Discrepancy and hereditary discrepancy are important notions in combinatorics and discrete geometry; for more information we refer, e.g., to [1] or [2]. Throughout the note, the asymptotic notation is used under the assumption that $n \to \infty$. All logarithms have natural base. We denote by $\mathrm{disc}(\mathcal{F})$ and $\mathrm{herdisc}(\mathcal{F})$ the discrepancy and the hereditary discrepancy of $\mathcal{F}$, respectively.

The following question is a folklore in discrepancy theory (as far as we could find out, it was first asked by V. Sós some years ago): Given a hypergraph $\mathcal{F}$, is it true that the hereditary discrepancy of $\mathcal{F}$ increases by at most a constant factor if one adds a new edge to $\mathcal{F}$?

As far as we know, there is no published result on this problem, although a polynomial factor can be proved by various arguments. In this note, we prove that adding one edge increases the hereditary discrepancy of $\mathcal{F}$ by a multiplicative factor of at most $O(\log n)$.

Theorem 1 *Let X be an n-point set and let $\mathcal{F} \subseteq 2^X$ satisfy* $\mathrm{herdisc}(\mathcal{F}) \leq h$. *Then* $\mathrm{disc}(\mathcal{F} \cup \{X\}) = O(h \log n)$.

The following consequence is immediate from the definition of hereditary discrepancy:

Corollary 2 *Let X be an n-point set and let $\mathcal{F} \subseteq 2^X$ satisfy* $\mathrm{herdisc}(\mathcal{F}) \leq h$. *Then for any subset X' of X, we have* $\mathrm{herdisc}(\mathcal{F} \cup \{X'\}) = O(h \log n)$.

Proof of Theorem 1. For each set $A \in 2^X$, let $\chi_A : A \to \{-1, +1\}$ witness $\mathrm{disc}(\mathcal{F}|_A) \leq h$. Define two colorings χ'_A and χ''_A of X by

$$\chi'_A(x) = \begin{cases} \chi_A(x) & \text{for } x \in A \\ \chi_{X \setminus A}(x) & \text{for } x \in X \setminus A \end{cases} \qquad \chi''_A(x) = \begin{cases} -\chi_A(x) & \text{for } x \in A \\ \chi_{X \setminus A}(x) & \text{for } x \in X \setminus A. \end{cases}$$

Let $\mathcal{C} = \{\chi'_A, \chi''_A : A \in 2^X\}$. Label each pair $\{\chi_1, \chi_2\}$ of distinct colorings in $\mathcal{C}$ by the set $\{x \in X : \chi_1(x) \neq \chi_2(x)\}$. Since the pair $\{\chi'_A, \chi''_A\}$ is labeled by A, there are at least 2^n distinct pairs, and so $|\mathcal{C}| \geq 2^{n/2}$.

Divide the colorings in $\mathcal{C}$ into at most n classes according to the value of $\chi(X)$, and let $\mathcal{C}_1$ be a class with $|\mathcal{C}_1| \geq \frac{1}{n}|\mathcal{C}| \geq 2^{n/2}/n$. The rest is as in the proof of Beck's partial coloring lemma (see, e.g., [1], Lemma 4.13). Since $\mathcal{C}_1$ is exponentially large, it contains two colorings χ_1, χ_2 differing in at least cn points, for a suitable positive constant $c > 0$. We form the partial coloring $\chi = \frac{1}{2}(\chi_1 - \chi_2) : X \to \{-1, 0, +1\}$. We have $\chi(X) = 0$, $|\chi(S)| \leq 2h$ for all $S \in \mathcal{F}$, and at least cn points of X are colored (meaning that they receive $+1$ or -1). Next, we restrict $\mathcal{F}$ to the subset $X_1 \subset X$ that received 0 under χ and we repeat the same argument, etc. Iterating $O(\log n)$ times, all points are colored, and the total discrepancy is $O(h \log n)$. $\square$

Acknowledgement: We would like to thank L. Lovász for communicating the problem and L. Lovász and B. Doerr for useful discussions.

References

[1] J. Matoušek. *Geometric discrepancy (An illustrated guide)*. Springer, Berlin 1999.

[2] J. Spencer. *Ten lectures on the probabilistic method.* CBMS-NSF. SIAM, Philadelphia, PA, 1987.