2 The Discrepancy Manifesto #2: The partial coloring lemma

By: Sariel Har-Peled
Date: October 1, 2008

We are given a range space $\mathcal{I} = (\mathcal{X}, \mathcal{R})$, where $\mathcal{X}$ is a ground set of size n, and $\mathcal{R}$ is a set of subsets of $\mathcal{X}$.

We are interested in *extreme* discrepancy. Namely, we would like to color the ranges of $\mathcal{R}$ is truely low; say, zero. Naturally, this is not possible, even for a single range, as the range might have odd size, and our coloring assign every point of $\mathcal{X}$ either -1 or $+1$.

To overcome this, we will allow the coloring to color some points by 0. We will refer to such a coloring as a ***partial coloring***. We remind the reader that we color the points, by consider a matching M of the points of $\mathcal{X}$, and coloring each edge by either $-1, 1$ or the other way around.

So, we will allow a pair to be uncolored. Namely, both points of an edge in this matching will be assigned $0, 0$.

We will refer to a coloring that colors all pairs as ***proper***. A nice property of proper colorings is that given two of them χ_1, χ_2, the coloring $(\chi_1 - \chi_2)/2$ is a partial coloring, which is zero on the points that they both color in the same way. Naturally, we would like to show that there is a partial coloring that has low discrepancy, but still colors many of the points.

The following is a straightforward adaption of the partial coloring lemma [Mat99].

Lemma 2.1 *Let $(\mathcal{X}, \mathcal{R})$ be a range space with VC dimension δ, where $\mathcal{X}$ is a set of $2m$ points. Let M be a matching of the points of $\mathcal{X}$. Also, let $\mathcal{E}$ be an event that holds for a proper random coloring of M with probability (at least) $1/2$. Next, let $\mathcal{M} \subseteq \mathcal{R}$ be a set of at most $(m-1)/(4\lg(2m+1))$ ranges. Then, one can find two proper colorings χ_1, χ_2 of $\mathcal{X}$, such that $\chi_1, \chi_2 \in \mathcal{E}$, and the partial coloring $\chi = (\chi_1 - \chi_2)/2$ has at least $m/10$ points of $\mathcal{X}$ for which it assigns a non-zero value.*

Proof: There 2^m proper colorings, and there are at least 2^{m-1} proper colorings that comply with $\mathcal{E}$, and let $\mathcal{G}$ denote these colorings. Such a coloring χ assign some discrepancy to a range $\mathsf{r} \in \mathcal{M}$, which is an integer number between $-|\mathsf{r}|$ to $|\mathsf{r}|$. Note however, that changing the coloring of a pair in the matching changes the discrepancy of r by $-2, 0$ or 2. Thus, there are only $1 + |\mathsf{r}| \leq 2m+1$ different values assigned to the discrepancy of r. Thus, consider the mapping f of a proper coloring χ to the vector $(\chi(\mathsf{r}_1), \chi(\mathsf{r}_2), \ldots, \chi(\mathsf{r}_u))$, where $\mathcal{M} = \{\mathsf{r}_1, \ldots, \mathsf{r}_u\}$. There are at most $(2m+1)^u$ different vectors that can be generated by f. As such, there must be a vector v, such that there are at least

$$\frac{2^{m-1}}{(2m+1)^u} \geq \frac{2^{m-1}}{2^{(m-1)/4}} \geq 2^{(m-1)/2}$$

proper colorings of $\mathcal{G}$ that are mapped to v by f. Let $\mathcal{H}$ denote the set of these colorings.

Next, consider any coloring $\chi_1 \in \mathcal{H}$. Define the distance between two colorings in $\mathcal{H}$ as the number of pairs that they color differently. Clearly, the total number of colorings in distance at most ℓ from χ_1 is bounded by

$$\sum_{i=0}^{\ell} \binom{m}{i} \leq 2\left(\frac{me}{\ell}\right)^{\ell},$$

as a standard calculation shows. As such, if ℓ is the furthest coloring in $\mathcal{H}$ from χ_1, then we must have

$$2\left(\frac{me}{\ell}\right)^{\ell} \geq |\mathcal{H}| \geq 2^{(m-1)/2},$$

which implies (for example) that $\ell \geq m/10$. Let χ_2 be this furthest coloring from χ_1 in $\mathcal{H}$. Clearly, $(\chi_1 - \chi_2)/2$ is the required coloring. ■

So, why is the partial coloring interesting? Consider a range space $(\mathcal{X}, \mathcal{R}')$ that might contain a large number of ranges. Say, n^{δ}, where $n = |\mathcal{X}|$. If this range space has certain low dimensional properties (say, it has low VC dimension), then one can pick $\mathcal{R}$ in such a way that it is sufficiently small, and it is a good packing in $\mathcal{R}$ (I am being a bit informal here). Namely, every range r of $\mathcal{R}'$ can be written down as $\mathsf{s} \oplus S$, where $\mathsf{s} \in \mathcal{R}$, $|S| \leq n^{1-1/\delta}$, and $\oplus$ denotes the symmetric difference of two sets. Then, the partial coloring lemma applied to $(\mathcal{X}, \mathcal{R})$ implies a partial coloring χ, so that for a $\mathsf{r} \in \mathcal{R}'$ we have that

$$\operatorname{disc}_{\chi}(\mathsf{r}) \leq \operatorname{disc}_{\chi}(\mathsf{s}) + 4\sqrt{|S| \log n} \leq \sqrt{n^{1-1/\delta} \log n},$$

and this holds for all ranges of $\mathcal{R}'$. Furthermore, χ properly colors at least $1/10$ of the points. We can now remove the points that are properly colored in this way, and recurse this process on the uncolored points. This implies that the resulting coloring would have discrepancy $O\left(\sqrt{n^{1-1/\delta} \log n}\right)$.

(In fact, I am slightly lying above, since I ignored some log factors that make the calculations slightly less clean.)

Theorem 2.2 *If $(\mathcal{X}, \mathcal{R}')$ is a range space of VC dimension δ, then there exists a proper coloring of the points of $\mathcal{X}$, so that the discrepancy of any set of $\mathcal{R}'$ is at most $O^*\left(\sqrt{n^{1-1/\delta}}\right)$, where O^* hides some polylog factors.*

Bibliographical notes

A good source on discrepancy is the book by Matoušek [Mat99] or the book by Chazelle [Cha01] (this book is available online).

References

[Cha01] B. Chazelle. *The Discrepancy Method: Randomness and Complexity.* Cambridge University Press, New York, 2001.

[Mat99] J. Matoušek. *Geometric Discrepancy.* Springer, 1999.

elements. Although the partial coloring approach sufces to obtain tight results for Spencerâ€™s theorem, there are other important discrepancy problems for which this approach does not currently (and perhaps cannot) yield tight results. A notable example is the Beck-Fiala conjecture [6], which asserts that: for âˆševery set system S for which every element of U is contained in at most t sets, the discrepancy of S is O(t). The geometric form of the Beck-Fiala conjecture is the KomloÂ´s conjecture: for all {x1, . . . , xn} âŠ‚ Rn with.Â The main lemma we prove is the following. Lemma 2. For every i, if phase i âˆ' 1 is successful and Î³ â‰¤ b/n2, then phase i is successful with probability of at least 1/4. 11/05/18 - We study hypergraph discrepancy in two closely related random models of hypergraphs on n vertices and m hyperedges. The first mode...Â In our setting of Theorem 1.2, the following properties alone deterministically imply that the discrepancy of the hypergraph is O(âˆšt). Every edge eâˆˆE is such that |e|=O(t). Let M be the incidence matrix of the hypergraph.Â But, there will also be â€˜largeâ€™ edges at every stage of the coloring but, the partial coloring lemma of Lovett and Meka can help ensure that these sets have discrepancy 0 in each step as long as they are large, and so they cause no problem. This is made formal below. First, we will observe that all the set sizes are Î˜(t) with high probability.