COS598B Algorithms and Complexity	Feb 19, 2010

Lecture 3: Constructive Bounds on Discrepancy

Lecturer: Moses Charikar — *Scribe: Aleksandar Nikolov*

In previous lectures we saw a constructive proof of the Lovasz Local Lemma and an application to a variant of the max-min allocations problem. In this lecture we will continue the topic of algorithmic versions of existential theorems with a recent result due to Bansal [Ban10] that makes Spencer's [Spe85] and Srinavasan's [Sri97] bounds on discrepancy of set systems constructive. What makes Bansal's result particularly interesting is that Spencer's and Srinavasan's arguments were previously thought to be non-constructive in a very strong sense.

3.1 Definitions and Simple Bounds

Consider a set system $(V, \mathcal{S})$, where, without loss of generality, $V = [n] = \{1, \dots, n\}$ and $\mathcal{S} = \{S_1, \dots, S_m\}$ and $\forall j \in [m] : S_j \subseteq V$. Let also $\chi : V \to \{-1, +1\}$ be a coloring of the vertex set V.

Definition 3.1 (Discrepancy)**.** *The discrepancy* $\mathrm{disc}(V, \mathcal{S})$ *of a set system* $(V, \mathcal{S})$ *is defined as*

$$\min_{\chi} \max_{S_j \in \mathcal{S}} |\chi(S_j)|,$$

where $\chi(S_j) := \sum_{i \in S_j} \chi(i)$.

Now we will derive simple probabilistic bounds on $\mathrm{disc}(V, \mathcal{S})$. A natural idea for upper-bounding $\mathrm{disc}(V, \mathcal{S})$ is to pick χ uniformly at random, i.e. for every $i \in V$ independently set $\chi(i)$ to 1 with probability 1/2 and to -1 with probability 1/2. Then $\chi(S_j)$ is a sum of $|S_j|$ independent ± 1 random variables and behaves like a gaussian random variable with mean 0 and standard deviation $\sqrt{|S_j|}$. For large enough S_j and some constant C

$$\Pr\left[|\chi(S_j)| \geq C\sqrt{|S_j| \log m}\right] \leq \frac{1}{m^2}.$$

Since $|S_j| \leq n$ and small sets for which the above bounds does not hold cannot contribute much to the discrepancy of the set system, this argument shows $\mathrm{disc}(V, \mathcal{S}) = O(\sqrt{n \log m})$.

Another easy probabilistic argument shows that there exists a $(V, \mathcal{S})$ s.t. $\mathrm{disc}(V, \mathcal{S}) > \lambda\sqrt{n}$ for some fixed constant λ. Let $\mathcal{S}$ be a set of $m = n$ random subsets of V: for every $j \in [n]$ and every $i \in [n]$, $i \in S_j$ with probability 1/2 and $i \notin S_j$ with probability 1/2. For a fixed coloring χ, $\chi(S_j)$ again is a sum of $|S_j|$ independent random variables and behaves like a gaussian r.v., and, therefore,

$$\forall j \in n : \exists \lambda \text{ s.t } \Pr\left[|\chi(S_j)| \leq \lambda\sqrt{n}\right] < 1/2$$

As sets are picked independently,

$$\Pr\left[\forall j \in n : \chi(S_j) \leq \lambda\sqrt{n}\right] < \frac{1}{2^n}.$$

Taking a union bound over all colorings χ, we get that $\mathrm{disc}(V, \mathcal{S}) \geq \lambda\sqrt{n}$ with positive probability. This proves the existence of a family of set systems with discrepancy $\Omega(\sqrt{n})$.

The probabilistic upper bound given above was improved by a celebrated result of Spencer in 1985. This is the main result that Bansal makes constructive. Some details follow.

Theorem 3.2 ([Spe85]). *For any set system* $(V, \mathcal{S})$*,*

$$\mathrm{disc}(V, \mathcal{S}) \leq \sqrt{n \log(m/n)}.$$

In particular, when $m = n$*, disc*$(V, \mathcal{S}) \leq 6\sqrt{n}$*.*

By the lower bound we proved above, this bound is tight up to constants for $m = n$. Bansal shows how to use semidefinite programming to construct a coloring of discrepancy $O(\sqrt{n})$ for this case.

3.2 Bounded Degree Discrepancy

We say that a set system $(V, \mathcal{S})$ has maximum degree t when for every $i \in V$ $|\{j \in m : i \in S_j\}| \leq t$. In a variation of the discrepancy problem we're interested in how the discrepancy of a set system depends on the maximum degree of the system. Ideally, we would like a bound that is independent of m and n, i.e. a function $f : \mathbb{N} \to \mathbb{N}$ s.t.

$$f(t) \geq \sup\{\mathrm{disc}(V, \mathcal{S}) : (V, \mathcal{S}) \text{ has maximum degree } t\}$$

Even the existence of such a function is not obvious. Beck and Fiala proved that $f(t) \leq 2t - 1$ [BF81], and the result was later improved to $2t - 3$. This bound is not known to be sharp and indeed Beck and Fiala conjectured that $f(t) = O(\sqrt{t})$, which is tight by the lower bound proved in the previous section. Settling the conjecture is an important open problem in discrepancy theory.

If we allow a dependence on n and m, we have the following two results:

$$\mathrm{disc}(V, \mathcal{S}) \leq \sqrt{t} \log t \log n \qquad \text{[Bec81]}$$
$$\mathrm{disc}(V, \mathcal{S}) \leq \sqrt{t} \log n \qquad \text{[Sri97]}$$

Srinivasan's bound extends the arguments used in the proof of Spencer's bound and is also non-constructive. Bansal gives a constructive version of this bound as well; however, we will not discuss Srinivasan's bound further in this lecture.

We will now prove Beck and Fiala's bound. Interestingly, the proof of the theorem uses a style of argument used in approximation algorithms for certain network design problems, e.g. Kamal Jain's iterative rounding technique. We will use a linear relaxation of discrepancy and will round it in steps to ± 1 values. Consider the following system of equations:

$$\forall j \in [m] : \sum_{i \in S_j} x_i = 0$$

Each variable appears in at most t equations. Let's drop all equations with $\leq t$ variables. Clearly, the remaining equations are fewer than the variables. Therefore, the system has a non-trivial solution, which can be scaled so that there is at least one variable assigned ± 1 and all other variables are assigned values in the interval $[-1, 1]$. Now we can iterate this procedure. At each phase we will drop all variables set to ± 1 in the previous phase - call these variables *fixed* and the remaining variables *live*. Then we will drop all equations with $\leq t$ live variables. The same argument we used in phase 1 applies at each phase, and, therefore, at each phase we fix at least one live variable, so there are at most n phases in total. At any point equations remaining in the system contribute (fractional) discrepancy of 0. Equations are dropped from the system if they have at most t live variables. After an equation is dropped, we have no control over how its remaining live variables are fixed. However, as each live variable's range is $(-1, 1)$, the dropped equations can only gain discrepancy strictly less than $2t$. This proves the bound.

3.3 Non-constructive Proof of Spencer's Bound

At the heart of the proof of Spencer's theorem is a Partial Coloring Lemma, that is also central to the constructive version of the bound. Moreover, this is the part of the proof that uses a pigeonhole-style argument that was previously thought to be strongly non-constructive.

In the proof of the bound we will construct a low-discrepancy coloring in phases.

Lemma 3.3 (Parial Coloring). *In phase i, we can find a partial coloring $\chi_i : V \to \{-1, 0, 1\}$, such that at least $n/2^i$ vertices receive a non-zero color and for any $j \in [m]$*

$$\operatorname{disc}(S_j) \leq \sqrt{\frac{|S_j|}{2^i}},$$

where the discrepancy is taken over the partial coloring χ_i.

The bound will follow by finding a Partial coloring on all vertices in V in phase 1, and finding a partial coloring on the vertices which were given a zero color in phase 2, and so on until all vertices are given a color in $\{\pm 1\}$.

First we will describe a high-level sketch of the proof of the lemma. Consider all 2^n colorings $\chi : V \to \{\pm 1\}$. We will put them in buckets where the buckets are defined as follows: if χ_1, χ_2 are in the same bucket then

$$\forall j \in [m] : \sum_{i \in S_j} \chi_1(i) = \sum_{i \in S_j} \chi_2(i) \pm \Delta(S_j). \tag{3.3.1}$$

Next we will invoke the pigeonhole principle to argue that one of the buckets is large. In particular, we will pick an absolute constant $\gamma \in (0, 1/2)$, s.t. there exists a bucket with at least $N := \sum_{t=0}^{\gamma n} \binom{n}{t}$ colorings that fall in it. Let's take a Hamming distance measure on colorings. By a theorem of Kleitman, the minimum diameter set of size N is the hamming ball of that size, and, therefore, any set of size N has diameter at least $2\gamma n$. Therefore, we are guaranteed to find two colorings in a bucket of size N that disagree in the colors of a constant fraction of V. Let those colorings be χ_1 and χ_2. We define a partial coloring $\chi = (\chi_1 - \chi_2)/2$. By (3.3.1) χ has discrepancy at most $\Delta(S_j)$ for any set $S_j \in \mathcal{S}$ and assigns a non-zero color to a constant fraction of V.

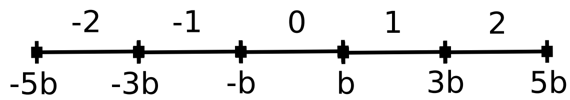

Figure 1: Bucket labels

Now some details. For every $S_j \in \mathcal{S}$ we will pick a granularity parameter $b_j = \Delta(S_j) = \lambda_j\sqrt{|S_j|}$. We define the rounding function $R_b(x)$ to be the nearest integer to $x/2b$ (see Figure 1). With a coloring χ we associate a vector $(\chi(S_1), \ldots, \chi(S_m))$. We round the vector to $(R_{b_1}(\chi(S_1)), \ldots, R_{b_m}(\chi(S_m)))$, and this last rounded vector is the bucket label for χ.

We need a way to argue that one of the buckets is big. Unfortunately, the number of buckets is too large for a straightforward counting argument to work. Instead we need to exploit the fact that for a random χ each $\chi(S_j)$ is concentrated around its mean 0. The argument we're going to use is one of the really neat ideas of Spencer's proof and is also the step that is particularly difficult to implement efficiently. Let's pick χ at random. Then the bucket label of χ, $Y := (R_{b_1}(\chi(S_1)), \ldots, R_{b_m}(\chi(S_m)))$, is a random variable. We are going to show that for some ϵ, $H(Y) \le \epsilon n$, where $H(\cdot)$ is the entropy function. Therefore, $\Pr[Y = y] \ge 2^{-\epsilon n}$ for some bucket label y, i.e. there are at least $2^{(1-\epsilon)n}$ colorings with label y.

The main technical part of the proof is bounding the entropy of $Y = (Y_1, \ldots, Y_m)$. We use the following basic inequality:

$$H(Y) \le \sum_{j=1}^{m} H(Y_j).$$

It is enough to analyze the entropy of Y_j. Remember that Y_j behaves like a gaussian random variable that is discretized into intervals of size $2\Delta(S_j) = 2\lambda_j\sqrt{|S_j|}$. When $\lambda := \lambda_j$ is large, the entropy is dominated by the central intervals: -1, 0, and 1, and most of the probability mass falls in the 0 interval. On the other hand, if λ is small, the high probability mass region is divided into many small intervals and Y_j behaves much like a uniform random variable in the interval $[-1, 1]$. More precisely, the following bounds are true for some constant k:

$$H(Y_j) \le \begin{cases} ke^{-\lambda_j^2/9}, & \lambda_j > 0.1 \\ k\log(\frac{1}{\lambda_j}), & \lambda_j \le 0.1 \end{cases} \tag{3.3.2}$$

While only the first inequality is needed for the non-constructive argument (we will be free to set the λ_j constants as big as necessary), the constructive version needs the second inequality as well.

In the non-constructive proof we will set all λ_j to a single value λ (once again the extra degree of freedom in setting different λ_j for different j will be needed in the constrictive version). In order to have a bucket large enough so that two colorings in the bucket agree on the colors of at most half the vertices in V, we need the following inequality to be satisfied:

$$H(Y) \le \sum_{j=1}^{m} H(Y_j) \le \sum_{j=1}^{m} ke^{-\lambda_j^2/2} \le \frac{n}{5}.$$

This will be true if we set $\lambda = c\sqrt{\log \frac{m}{n}}$ for an appropriate constant c. This completes the proof of the partial coloring lemma.

Finally, we need to put together the pieces of the argument. We will apply the partial coloring lemma in phases, so that in each phase we will find a partial coloring of the vertices which have not been colored so far. Let n_i be the number of vertices which are not colored before phase i. We have $n_i = n$ and $n_i \leq n2^{-i+1}$. After $\log n + 1$ steps at most one vertex is left uncolored, and, therefore, the total discrepancy when the whole set V is colored is at most

$$\sum_{i=1}^{\log n+1} c\sqrt{n_i \log \frac{m}{n_i}} + 1 = c \sum_{i=1}^{\log n+1} \sqrt{n_i \left(i + \log \frac{m}{n_i}\right)} + 1$$
$$\leq O\left(\sqrt{n \log \frac{m}{n}}\right).$$

3.4 Approximating Hereditary Discrepancy

We will begin our discussion of efficient constructions of low-discrepancy colorings with a pseudoapproximation algorithm for discrepancy by Bansal. This result is simpler than the constructive version of Spencer's and Srinivasan's bounds. In fact, those constructive bounds build on the ideas used in the pseudoapproximation algorithm. In the next section we will sketch how the constructive version of Spencer's bound can be derived from the algorithm we are going to describe next.

First we need some definitions. Given a set system $(V, \mathcal{S})$, we define the *trace* of $\mathcal{S}$ on $W \subseteq V$ as

$$\mathcal{S}_{|W} = \{S_j \cap W : S_j \in \mathcal{S}\}$$

Then, the *hereditary discrepancy* of $(V, \mathcal{S})$ is

$$\mathrm{herdisc}(V, \mathcal{S}) = \max_{W \subseteq V} \mathrm{disc}(W, \mathcal{S}_{|W}).$$

Hereditary discrepancy is in some sense a more robust notion of the discrepancy of a set system. It is easy to sea that $\mathrm{herdisc}(V, \mathcal{S})$ is always positive (take W to be a singleton set). Also, there are easy examples of set systems with discrepancy 0 and hereditary discrepancy $\Omega(\sqrt{n})$ (see [Mat99]).

We will show the following.

Theorem 3.4 ([Ban10])**.** *There exists a polynomial time randomized algorithm that given a set system* $(V, \mathcal{S})$ *with* $\mathrm{herdisc}(V, \mathcal{S}) = \lambda$*, finds a coloring* $\chi : V \to \{-1, 1\}$ *with discrepancy* $O(\lambda(\sqrt{\log m} \log n)$ *with probability at least* $1/n$*.*

Without loss of generality we can assume we know λ since there are at most n choices for its value.. We begin with a natural SDP relaxation of discrepancy.

$$\|\sum_{i \in S_j} v_i\|_2^2 \leq \lambda^2 \qquad \forall j \in [m] \tag{3.4.1}$$

$$\|v_i\|_2^2 = 1 \qquad \forall i \in [n] \tag{3.4.2}$$

The feasibility of the SDP is guaranteed by $\text{herdisc}(V,\mathcal{S}) = \lambda$ (in fact $\text{disc}(V,\mathcal{S}) = \lambda$ is enough, but we will need the hereditary discrepancy later). Next we need to round the SDP to an integer solution. However, we also want to exploit the guarantees given by inequalities 3.4.1. Let's for a moment assume that all we need is to assign each vertex a color from $\mathbb{R}$ so that discrepancy is small. Bansal's idea is to achieve this by picking a random vector r from a multinomial gaussian distribution and projecting the vectors v_i down onto r; the real values we get are distributed according to a gaussian distribution, as is the discrepancy of each set S_j. Moreover, the variance of the discrepancy of S_j is bounded by λ by inequalities 3.4.1. Details follow.

We are going to get an integer solution incrementally. At each time step t we will maintain a vector $x^{(t)} \in [-1,1]^n$. Initially, $x^{(0)} = (0,\ldots,0)$. At time t we will generate generate a vector $r^{(t)}$ from $N(0,1)^n$, where $N(\cdot,\cdot)$ is the gaussian distribution; then we update x based on a random projection: $x_i^{(t)} = x^{(t-1)} + s(v_i \cdot r^{(t)})$, where s is a very small scaling factor ($s = O(1/(n\sqrt{\log n}))$). This update process has two important properties:

- At each step we make progress:

$$\mathbb{E}[(x_i^{(t)})^2] = \mathbb{E}[(x_i^{(t-1)})]^2 + s^2. \tag{3.4.3}$$

  The middle term has expectation 0, as the random variables $x_i^{(t-1)}$ and $v_i \cdot r^{(t)}$ are independent with mean 0.

- Discrepancy is kept small for every $S_j \in \mathcal{S}$:

$$\sum_{i\in S_j} x_i^{(t)} - \sum_{i\in S_j} x_i^{(t-1)} = s\left(\sum_{i\in S_j} v_i\right)\cdot r^{(t)}. \tag{3.4.4}$$

  Recall that when $r \in N(0,1)^n$, $v_i \cdot r$ is a gaussian with with variance $\|v_i\|_2^2$. Therefore the right hand side of (3.4.4) is a gaussian with variance $s^2\|\sum_{i\in S_j} v_i\|_2^2$ which by (3.4.1) is at most $s^2\lambda^2$.

By the first property, at each time step the values x_i will move slightly away from 0. Once a variable is in the region $[-1,-(1-1/n)]$ it is fixed to -1 and is not updated anymore; likewise, once a variable is in the region $[1-1/n,1]$, it is fixed to 1 and is not updated anymore. Since the scaling factor s is very small, with very high probability no variable will become larger than 1 or smaller than -1 before it is fixed. Note that when fixing a variable, we can gain discrepancy of value at most $1/n$, which is at most 1 over all iterations and is negligible, considering that the hereditary discrepancy is at least 1.

Ideally, we would be able to continue the update process with the remaining variables once a variable is fixed. However, the constraints (3.4.1) can bound the variance of the updates only when all vectors are used in the update process. Therefore, we need to re-solve the SDP. In particular, let $A(t) = \{i : i \text{ is active at time } t\}$. The new SDP contains constraints (3.4.1) and also

$$\begin{aligned} \|v_i\| &= 1 && \forall i \in A(t) \\ \|v_i\| &= 0 && \forall i \in [n]\setminus A(t) \end{aligned}$$

The feasibility of the solution is guaranteed by $\text{herdisc}(V, \mathcal{S}) = \lambda$ (and this is where we really need the hereditary discrepancy). Note that while the SDP is re-solved, the values of x_i are not reset. After re-solving the SDP the update process continues with the new SDP solution. The process stops when all x_i have been fixed.

For the analysis we will introduce the notion of phases, similar to the phases in the proof of Spencer's bound. We say that the algorithm is in phase q when $|A(t)| \in (n/2^q, n/2^{q-1}]$. Let E_q be the event that in phase q no set gains in discrepancy more than $c\lambda\sqrt{\log m}$. We need to prove that for any q

$$\Pr\left[E_q | E_1 \wedge \ldots \wedge E_{q-1}\right] \geq 1/2. \tag{3.4.5}$$

Then by the chain rule we would know that $\Pr\left[\forall i E_i\right] \geq 1/n$, and, therefore, the total discrepancy of the final solution is $O(\lambda\sqrt{\log m}\log n)$ with probability at least $1/n$.

We already have the ingredients we need to prove (3.4.5). By the progress property of the update process and Markov's inequality, the number of update steps in the q-th phase is at most $O(1/s^2)$ with constant probability. Therefore, in each phase to the discrepancy of a set we add the sum of $O(1/s^2)$ gaussian random variables with variances bounded by $s^2\lambda^2$, which is a gaussian random variable with variance at most λ^2. It follows that with probability $1 - \frac{1}{m^{O(1)}}$ the discrepancy of any set S_j increases by at most $O(\lambda\sqrt{\log m})$, and, taking a union bound over all $S_j \in \mathcal{S}$, with constant probability the discrepancy of no set increases by more than this amount. Choosing the right constants yields (3.4.5). This completes the analysis of the algorithm.

3.5 Constructive Version of Spencer's Bound

In order to make Spencer's bound constructive we need to efficiently find the partial coloring whose existence is guaranteed by the partial coloring lemma. We will use a similar SDP formulation as the one for approximating hereditary discrepancy and an analogous rounding technique. However, in Spencer's bound, we need a tighter control on how the discrepancies of the sets S_j increase. Therefore, when the discrepancy of a set increases too much, we will use smaller buckets for that set. This will reduce discrepancy, but increase the entropy of the bucket labels. The technical challenge is to show that not too many sets will need to be bucketed aggressively, and, therefore, the partial coloring lemma still holds.

In phase q, we will call a set k-dangerous if its discrepancy from the current phase is in the interval $[\beta_q(k), \beta_q(k+1))$, where

$$\beta_q(k) := O(\sqrt{n/2^q}(q+1)(2 - \frac{1}{k})$$

In the SDP for the phase, we set $\|v_i\|_2^2 = 0$ for i colored in previous phases, and $\|v_i\|_2^2 \leq 1$ for active variables, as we are now looking for a partial coloring. In order to ensure progress through the phase, we set:

$$\sum_{i \in [n]} \|v_i\|_2^2 \geq n/2^{q+2}.$$

Finally, define

$$\alpha_q(k) := O(\frac{n(q+1)}{(k+1)^5 2^q}.$$

In the SDP, we will insist that the discrepancy of a k-dangerous set is bounded by $\sqrt{\alpha_q}$. For the full details, look at Bansal's paper.

References

[Ban10] Nikhil Bansal. Constructive algorithms for discrepancy minimization. *Arxiv preprint arXiv:1002.2259v1*, 2010.

[Bec81] J. Beck. Roth's estimate of the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, 1981.

[BF81] J. Beck and T. Fiala. Integer-making theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[Mat99] J. Matousek. *Geometric Discrepancy: An Illustrated Guide. Vol. 18 of Algorithms and Combinatorics*, page 110. Springer, 1999.

[Spe85] J. Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, pages 679–706, 1985.

[Sri97] A. Srinivasan. Improving the discrepancy bound for sparse matrices: Better approximations for sparse lattice approximation problems. In *Proceedings of the eighth annual ACM-SIAM symposium on Discrete algorithms*, pages 692–701. Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 1997.