Lecture 24: Analysis of Bansal's algorithm

Claim 2. With probability $\geq 1 - \frac{1}{n}$, the discrepancy of the coloring x_ℓ is $O(H \cdot \log(mn))$, where H is the hereditary vector discrepancy of $\mathcal{F}$.

The hereditary vector discrepancy of $\mathcal{F}$ is
$\max_{A \subseteq V} \operatorname{vecdisc}(\mathcal{F}|_A)$ where $\operatorname{vecdisc}(\mathcal{F}|_A)$ is

$$\min D \geq 0 \text{ s.t. } \left\| \sum_{j \in A \cap S_i} u_j \right\| \leq D^2 \text{ for } i = 1, \ldots, m \text{ and}$$
$$\|u_j\|^2 = 1 \text{ for } j \in \{1, \ldots, n\} \cap A.$$

In order to prove Claim 2, we will prove
$$\Pr[\operatorname{disc}(\mathcal{F}, x_\ell) > D_{max}] \leq 1/n$$
where $D_{max} = \Theta(H \cdot \log(mn))$.
Let us fix a set $S_i \in \mathcal{F}$ and let $D_i = \sum_{j \in S_i} (x_\ell)_j$ be its imbalance in x_ℓ. We will show
$$\Pr[|D_i| > D_{max}] \leq 1/mn \text{ for every } i.$$
Then Claim 2 will follow from the union bound.

As t goes from 0 to ℓ, let us recall how the j-th coordinate of the current semicoloring x_t develops.
- It starts with $(x_0)_j = 0$. Then it changes by the random increments $(\Delta_t)_j$, $t = 1, 2, \ldots, t_0$. Then at some step $t_0 + 1$, it gets truncated to $+1$ or -1 and it stays frozen at this value until the end.

- Since $(\Delta_t)_j = 0$ for $t > t_0 + 1$, we can write

$$(x_\ell)_j = \sum_{t=1}^{\ell} (\Delta_t)_j + T_j ,$$

where T_j is a "truncation effect", reflecting the fact that $(x_{t_0+1})_j = \pm 1$ and not $(x_{t_0} + \Delta_{t_0+1})_j$.

We have $|T_j| \leq |(\Delta_{t_0+1})_j|$, and as we know, $(\Delta_{t_0+1})_j \sim N(0, \sigma^2)$.

We will now show that for σ as small as $\frac{1}{C_0 n \sqrt{\log n}}$, all truncation effects are negligible with probability close to 1. More formally, we claim that, for each j: $\Pr[|T_j| > 1/n] \leq 1/n^3$.

<u>Proof of this claim.</u> For $Z \sim N(0,1)$, we have

$$\Pr[Z \geq t] \leq \frac{1}{t} \cdot \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \quad \text{for any } t \geq 0$$

So $\Pr\left[|T_j| > \frac{1}{n}\right] = \Pr\left[|\sigma Z| \geq 1/n\right]$ (where $Z \sim N(0,1)$)

$$= \Pr\left[|Z| \geq \frac{1}{\sigma n}\right] \leq e^{-1/2\sigma^2 n^2}$$

$$= e^{-\frac{C_0^2 \log n}{2}} \leq \frac{1}{n^3} \quad \text{for } C_0^2 = 6.$$

Thus w.p. $\geq 1 - 1/n^2$, the total contribution of the truncation effects T_j to the discrepancy of each set S_i is at most 1. So instead of D_i, it suffices to bound the "pure random walk" quantity

$$\tilde{D}_i = \sum_{j \in S_i} \sum_{t=1}^{\ell} (\Delta_t)_j = \sum_{t=1}^{\ell} \sum_{j \in S_i} (\Delta_t)_j$$

$$= \sum_{t=1}^{\ell} \sum_{j \in S_i} \sigma \cdot r_t^T u_{t,j} = \sum_{t=1}^{\ell} \sigma r_t^T v_{t,i}$$

where $v_{t,i} = \sum_{j \in S_i} u_{t,j}$.

We know that $\|v_{t,i}\| \leq H$ for all t and i. Writing $Y_t = \sigma \cdot r_t^T v_{t,i}$ (here i is fixed), we get that $Y_t \sim N(0, \beta_t^2)$ where $0 \leq \beta_t \leq \sigma H$.

- Intuitively, things should be simple here. The sum $\tilde{D}_i = Y_1 + \dots + Y_t$ should have a distribution like $N(0, \ell\sigma^2 H^2)$ and hence we should get an exponential tail bound, i.e., the probability $|D_i| > \lambda \cdot \sigma H \sqrt{\ell}$ (i.e, λ times the standard deviation) should be $e^{-\lambda^2/2}$.

However we cannot claim that the Y_t's are independent. So we need a more sophisticated and technical tool.

<u>Lemma.</u> Let $W_1, \dots, W_\ell$ be independent random variables on some probability space and let Y_t be a function of $W_1, \dots, W_{t-1}$ for all t. Suppose that, conditioned on $W_1, \dots, W_{t-1}$ attaining some arbitrary values $w_1, \dots, w_{t-1}$, the distribution of

Y_t is $N(0, \beta_t^2)$ where β_t may depend on $w_1, \dots, w_{t-1}$ but we always have $\beta_t \le \beta$.
Then $Y = Y_1 + \dots + Y_t$ satisfies the tail bound $\Pr[|Y| > \lambda \beta \sqrt{l}] \le 2e^{-\lambda^2/2}$ for all $\lambda \ge 0$.

We will use the above lemma with $W_t = r_t$.
Our $\lambda = \dfrac{D_{max}}{\sigma \mathcal{H} \sqrt{l}}$, where $D_{max} = C_2 \cdot \mathcal{H} \cdot \log(mn)$, (for some constant C_2)

$\sigma = \dfrac{1}{C_0 n \sqrt{\log n}}$, and $l = \dfrac{C_1 \log(mn)}{\sigma^2}$.

We thus calculate that $\lambda \ge C_3 \sqrt{\log(mn)}$, with a constant C_3 that can be made as large as needed by adjusting the constant C_2 from D_{max}.
Then $\Pr[|\tilde{D}_i| > D_{max}] \le 2e^{-\lambda^2/2} \le \dfrac{1}{nm}$.
Now Claim 2 follows from the union bound.

The Sparsest Cut Problem

Let $G = (V, E)$ be an undirected graph. Consider a cut $(S, V-S)$ in G. The sparsity of this cut is defined as $\rho(S) = \dfrac{|\delta(S)|}{|S| \cdot |V-S|}$,

where $\delta(S)$ is the set of edges with exactly one endpoint in S.

The sparsest cut of G is defined as

$$\rho(G) = \min_{S \subseteq V} \rho(S).$$

More generally, the input to the sparsest cut problem is a weighted graph $G = (V, E)$ with positive edge weights c_e for every edge $e \in E$. We are also given a set of pairs of vertices $\{(s_1, t_1), (s_2, t_2), \ldots, (s_k, t_k)\}$, with associated demands D_i between s_i and t_i.

Given such a graph, we define the sparsity of a cut $(S, V-S)$ as $\dfrac{c(S, \overline{S})}{D(S, \overline{S})}$ (here $\overline{S} = V - S$)

where $$c(S, \overline{S}) = \sum_{e \in E \cap (S \times \overline{S})} c(e)$$

and $$D(S, \overline{S}) = \sum_{\substack{i \,:\, s_i \,\&\, t_i \\ \text{are separated} \\ \text{by } (S, \overline{S})}} D_i.$$

As before, the goal is to find a cut with the least sparsity.

We will focus on the special case where we have unit demand between every pair of vertices and $c(e) = 1$ for all $e \in E$. This makes the sparsity of any cut $(S, V-S) = \dfrac{\delta(S)}{|S| \cdot |V-S|}$.

A reduction from the max-cut problem shows that even the special case of the sparsest cut problem is NP-hard.

<u>Theorem</u> (Leighton and Rao, 1988). There is an $O(\log n)$-approximation algorithm for the sparsest cut problem.

Consider the following relaxation of the sparsest cut problem:

$$\min \sum_{e \in E} x_e$$

$$\text{s.t.} \quad \sum_{i,j} d_x(i,j) \geq 1$$

$$x_e \geq 0 \quad \forall e \in E,$$

where $d_x(i,j)$ is the shortest path distance from i to j using x as edge lengths.

To see that it is a relaxation, let $(S^*, V-S^*)$ be a sparsest cut. Set

$$x_e = \begin{cases} 1/|S^*||V-S^*| & \text{if } e \text{ crosses } (S^*, V-S^*) \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\sum_{i,j} d_x(i,j) \geq |S^*| \cdot |V-S^*| \cdot \frac{1}{|S^*||V-S^*|} = 1.$$

This is because any path that connects a vertex $i \in S^*$ and a vertex $j \in V-S^*$ has to use at least one edge in $\delta(S^*)$ and the number of pairs of vertices (i,j) where $i \in S^*$ and $j \in V-S^*$ is $|S^*| \cdot |V-S^*|$. The objective function for x set in this way is

$$\sum_{e \in E} x_e = \frac{|\delta(S^*)|}{|S^*| \cdot |V-S^*|} = \rho(S^*) = \rho(G).$$

So the optimal value $\leq \rho(G)$.

Reference: Lecture 25 on "Spectral Graph Theory" by David Williamson.