CS 774: Homework #2

Akash Kumar

October 19, 2023

Grading Rules

- This homework has 2 problems. The first problem is worth **25 points** and the second one **75 points**. No dud problems this time :-). Instead, there is a bonus problem which you need not turn in. If you however do turn it in and get some points on it, it will help you solve fewer problems on the next assignment.
- You are *encouraged* to discuss. Feel free to consult online sources. Just make sure that when you sit down (or is it lie down?) to write solutions, you are not copy/pasting the solution. In general, the policy of the class is **be reasonable**. I trust that you all have an instinct which tells you what constitutes a reasonable behavior. Also, please **list all the people you colloaborated with** on the homework.
- **Due date: ~~Oct 19, 2023~~ Oct 26, 2023, Thursday in class.** You can also email your submissions to me at akash@cse.iitb.ac.in if you LaTex it (preferred, but not mandated).

In this pset, we will see how you use the Lovasz-Simonovits toolkit through a "field-trip". Unless otherwise specified, throughout this pset you should think that we are working with a graph $G = (V, E)$ which is **undirected, connected and has maximum degree at most** d. Here, you should think that $|V| = n, |E| = m$ and that $d = O(1)$ is at most some constant.

1 Problem 1: More fun with Lovasz-Simonovits toolkit

Recall that in the class, we considered the LS curves g_t associated with the distribution $\boldsymbol{p}_{v,t} = \mathbf{1}_v^T \boldsymbol{W}^t$ which you get after (pre)multiplying $\mathbf{1}_v$ with $\boldsymbol{W}^t$. In practice however, you do not compute $\boldsymbol{p}_{v,t}$ by running such power iterations. Instead, you only work with an approximation $\widetilde{\boldsymbol{p}}_{v,t}$ to $\boldsymbol{p}_{v,t}$ which you obtain by running some w random walks from v and setting

$$\widetilde{\boldsymbol{p}}_{v,t}(u) = \frac{\text{Number of walks that end at } u}{w}.$$

Define a LS like curve $\widetilde{g}_t$ using the empirical distribution $\widetilde{\boldsymbol{p}}_{v,t}$.

Let us kick this pset off by solving the following problems.

1. **(5 pts)** Define the curve $\widetilde{g}_t$ in the "correct" natural way and show that $\widetilde{g}_t$ is concave.
2. **(5 pts)** Let $\alpha, \delta > 0$ be sufficiently small and let $w = c \cdot \frac{1}{\alpha} \cdot \log n$ (where $c = 1/\delta^2$). Now, fix $u \in V$, define $\delta_u = \delta(\boldsymbol{p}_{v,t}(j) + \alpha)$. Use Chernoff/Hoeffding/Bernstein Bounds [1] to show

$$\Pr\left(\left|\widetilde{\boldsymbol{p}}_{v,t}(u) - \boldsymbol{p}_{v,t}(u)\right| > \delta_u\right) \le n^{-10}.$$

[1] Feel free to look it up on Wikipedia or elsewhere.

3. **(5 pts)** Next up, show that for every $u \in V$, it holds with probability at least n^{-9} that

$$(1-\delta) \cdot \boldsymbol{p}_{v,t}(u) - \delta\alpha \le \widetilde{\boldsymbol{p}}_{v,t}(u) \le (1+\delta) \cdot \boldsymbol{p}_{v,t}(u) + \delta\alpha.$$

4. **(5 pts)** Conditioned on the above event, show that for any set $S \subseteq V$, we have

$$(1-\delta) \cdot \boldsymbol{p}_{v,t}(S) - \delta\alpha \cdot (2\,\mathrm{vol}(S)) \le \widetilde{\boldsymbol{p}}_{v,t}(S) \le (1+\delta) \cdot \boldsymbol{p}_{v,t}(S) + \delta\alpha \cdot (2\,\mathrm{vol}(S)).$$

   Here, we let $\boldsymbol{p}_{v,t}(S) = \sum_{u \in S} \boldsymbol{p}_{v,t}(u)$ and $\widetilde{\boldsymbol{p}}_{v,t}(S)$ is defined in a similar way. Recall from class that $\mathrm{vol}(S) = \sum_{u \in S} d_u$.

5. **(5 pts)** Finally, show that for every t and every $0 \le x \le 2m$

$$(1-\delta) g_t(x) - \delta\alpha x \le \widetilde{g}_t(x) \le (1+\delta) g_t(x) + \delta\alpha x.$$

In the next problem, you will see how practitioners use LS toolkit on empirically defined probability vectors to find low conductance cuts (with conductance at most some parameter $\phi > 0$ in graphs). It will be convenient to think of $t = \log n/\xi$ where $0 < \xi < 1/8$ is some sufficiently small constant. Also, assume that you know for a certain $v \in V$, it holds that there is some vertex u such that $\frac{\boldsymbol{p}_{v,t}(u)}{2 \cdot d_u} \gg 1/\sqrt{m}$. That is, you are told that there is some edge e such that $\boldsymbol{\rho}_t(e) \gg \frac{1}{\sqrt{m}}$[2] any edge e.

2 Problem 2: Using LS on empirical random walk vectors

Let $H = \left\{ u \in V : \frac{\boldsymbol{p}_{v,t-1}(u)}{2d_u} \ge \delta \cdot \alpha \right\}$ denote the set of "heavy vertices" in $\boldsymbol{p}_{v,t-1}$ – that is, the normalized $\boldsymbol{\rho}_{t-1}$ values on all edges incident to any vertex in H is large.

2.1 Appetizers

1. **(5 pts)** Let $|H| = h$ and let $i_1 \le i_2 \le \ldots \le i_h$ denote the first h hinge points of g_{t-1}. Then

$$i_h \le \frac{1}{\delta \cdot \alpha}.$$

Now that we are a bit warmed up with the soup, time to try the *paneer tikka*. We begin with a variation of the algorithm practitioners use. Later we will analyze this algorithm as our meal progresses all the way through the dessert. **EDIT: The value assigned to b is actually correct. Just writing it in terms of α.**

$\texttt{SweepCutLS}(G, v)$:

1. Assign $\alpha = 1/\sqrt{n}$ and choose $0 < \xi < 1/8$ such that

$$1/2 \left(\sqrt{1-2\phi} + \sqrt{1+2\phi} \right) = \exp(-\xi/2).$$

2. Assign $w = 100 \cdot t^2 \cdot \ln n \cdot \frac{1}{\alpha} = 100 \cdot \sqrt{n} \cdot t^2 \ln n$ and $b = \cancel{\frac{t \cdot \sqrt{n}}{2 \cdot (1-2\phi)}} = \frac{t}{2 \cdot (1-2\phi) \cdot \alpha} = O(\sqrt{n} \cdot \log n)$.
3. For each length $0 \le l \le t$
   (a) Order vertices so that $\frac{\widetilde{\boldsymbol{p}}_{v,l}(1)}{d_1} \ge \frac{\widetilde{\boldsymbol{p}}_{v,l}(2)}{d_2} \ge \frac{\widetilde{\boldsymbol{p}}_{v,l}(i)}{d_i} \ldots \ge \frac{\widetilde{\boldsymbol{p}}_{v,l}(n)}{d_n}$.
   (b) Compute the conductance of the first b sweep sets.
   (c) Return the sweep set with smallest conductance.

[2]Recall, the symbol "rho" we used for edge induced distributions in class

2.2 The Main Course

Towards showing that the above algorithm returns a cut of small conductance, it will be helpful to proceed via contradiction. It will be helpful to use the following notation. Write $\widehat{x} = \min(x, 2m - x)$.

1. **(25 pts)** Suppose for every $0 \le l \le t$ the following holds. Each of the first b sweep cuts on the empirical vector $\widetilde{\boldsymbol{p}}_{v,l}$ encountered by the algorithm `SweepCutLS`(G, v) at step l has conductance at least ϕ. Denote the hinge points on $\cancel{\widetilde{g}_{t-1}}$ $\widetilde{g}_{l-1}$ as $i_1, i_2, \ldots i_n$. Fix $1 \le l \le t$. Then for any $k \le n$, with high probability, we have

$$\boldsymbol{p}_{v,l}(\widetilde{S}_{l,k}) \le 1/2 \cdot \left[g_{l-1}(i_k - 2\phi \cdot \widehat{i_k}) + g_{l-1}(i_k + 2\phi \cdot \widehat{i_k}) \right] + \delta\alpha \cdot \phi \cdot \widehat{i_k}.$$

   Here, $\widetilde{S}_{l,k}$ denotes the k-th prefix cut computed on the empirical vector at the l-th step. That is, $\widetilde{S}_{l,k}$ contains the vertices corresponding to the k-th hinge point on $\widetilde{g}_l$.

HINT: It would be helpful to break your proof into two parts. First consider the case where $k \le b$ and then consider the case where $k > b$. The proof should be easy to finish in the first case. The second case requires you to put your understanding of the LS curves to work. The reason this incurs a small probability of failure is because you are working with sets "S-tilde" sets. These are empirical sets and they don't necessarily play well with the probability vector $\boldsymbol{p}_{v,l}$. ~~For the second case, you might want to first show that~~ $\cancel{i_h \le i_k - 2\phi \cdot \widehat{i_k}}$ ~~(where you recall~~ $\cancel{i_h}$ ~~was defined in Section~~ 2.1).
For the second case, you might find it convenient to show that the slope of the curve at the point $i_k - 2\phi \cdot \widehat{i_k}$ is at most $\delta\alpha$. Note that this allows you to write $g_{l-1}(i_k) \le g_{l-1}(i_k - 2\phi \cdot \widehat{i_k} + 2\phi \cdot \widehat{i_k} \cdot \delta\alpha$. Another bound you see is $g_{l-1}(i_k) \le g_{l-1}(i_k + 2\phi \cdot \widehat{i_k})$ which follows trivially as $g_{l-1}(.)$ is an increasing function. You can average these bounds to finish up.

The above exercise is a good workout which should inform you how practitioners use LS. The first bound above uses concavity and the latter one uses properties of increasing functions. You want to put everything together in a neat satisfying way. Those of you who missed the expository section on this homework, please ask others who attended the session.

2.3 Desserts

Finally, we are ready to consider actual sweep cuts on empirical vector, $\widetilde{\boldsymbol{p}}_{v,l}$ and the LS curve $\widetilde{g}_l$. Note that this means one of the relevant objects of interest would be $\widetilde{\boldsymbol{p}}_{v,l}(\widetilde{S}_{l,k})$. In the rest of this problem, assume that the bounds from Section 2.2 hold with probability 1.

1. **(25 pts)** Assume bounds from Section 2.2 hold for all walk lengths $0 \le l \le t$. Fix some $l \le t$. Show that for any hinge point i_k on $\widetilde{g}_l$,

$$\widetilde{g}_l(i_k) \le \frac{1+\delta}{2(1-\delta)} \left(\widetilde{g}_{l-1}(i_k - 2\phi \cdot \widehat{i_k}) + \widetilde{g}_{l-1}(i_k + 2\phi \cdot \widehat{i_k}) \right) + 4\delta\alpha i_k.$$

2. **(20 pts)** Show that for all $1 \le l \le t$, for all $0 \le x \le 2m$, you have

$$\widetilde{g}_l(x) \le \exp(O(\delta)) \cdot \frac{1}{2} \cdot (\widetilde{g}_{l-1}(x - 2\phi \cdot \widehat{x}) + \widetilde{g}_{l-1}(x + 2\phi \cdot \widehat{x})) + 4\delta\alpha x.$$

Finally, we come to the Bonus problem. In this problem, which you need not turn in, you will show that all of this machinery can be used to find low conductance cuts. You will show this by arguing that if all the sweep cuts done on the empirical vectors had large conductances, then the LS curve would decay so much that all edges would satisfy $\rho_t(e) \ll 1/\sqrt{m}$.

1. **(Bonus, 40 points, may cash in on a future assignment)** Suppose `SweepCutLS`(G, v) never finds a cut with conductance at most ϕ over all choices of walk lengths $0 \leq l \leq t$. Define

$$\psi = -\log\left(1/2(\sqrt{1-2\phi} + \sqrt{1+2\phi})\right)$$

and note this means $\psi \geq \phi^2/2$.

   (a) **(30 pts)** Use induction on l to show that

$$\widetilde{g}_l(x) \leq e^{10\delta l}\left[\sqrt{\widehat{x}} \cdot \exp(-\psi \cdot l) + x/2m\right] + \frac{4}{\sqrt{m}} \cdot \exp(4\delta l) \cdot x.$$

   (b) **(10 pts)** Finally, set $\delta = 1/t$ and show that this means $\widetilde{g}_t(1) \ll 1/\sqrt{m}$. Use this to conclude that the graph contains a cut with conductance at most ϕ.