ADVANCED ALGORITHMS (VIII)

CHIHAO ZHANG

Last week, we introduced the Leighton-Rao relaxation of sparsest cut. The key tool we used in the analysis is Jean Bourgain's ℓ_1 embedding theorem.

Theorem 1. *Let $d : V^2 \to \mathbb{R}$ be a semi-metric. There exists some $m \geq 1$ and a function $f : V \to \mathbb{R}^m$ such that for some constant $c > 0$ and every $x, y \in V$,*

$$\|f(x) - f(y)\|_1 \leq d(x, y) \leq c \log |V| \cdot \|f(x) - f(y)\|_1.$$

We shall prove the theorem today. The first useful observation is that the ℓ_1-distance of any embedding into $\mathbb{R}^n$ can be equivalently viewed as the expected ℓ_1-distance of random embeddings into $\mathbb{R}$. To see this, let $F : V \to \mathbb{R}^n$ be an embedding such that $F(x) = (F_1(x), F_2(x), \dots, F_n(x))$ for every $x \in V$. Consider a family of n functions $\mathcal{F} = \{f_1, \dots, f_n\}$ where each $f_i = mF_i$. Let $\mu_{\mathcal{F}}$ be the uniform distribution over $\mathcal{F}$, then it holds that

$$\|F(x) - F(y)\|_1 = \mathop{\mathbf{E}}_{f \sim \mu_{\mathcal{F}}} \left[\left| f(x) - f(y) \right| \right].$$

Conversely, for any collection of n functions $\mathcal{F} = \{f_1, \dots, f_n\}$ where each $f_i : V \to \mathbb{R}$ maps points in V to reals and any distribution $\mu_{\mathcal{F}}$ over $\mathcal{F}$, we can define $F : V \to \mathbb{R}^n$ such that $F(x) = (F_1(x), F_2(x), \dots, F_n(x))$ where each $F_i = \mu_{\mathcal{F}}(f_i) \cdot f_i$. It is clear that

$$\mathop{\mathbf{E}}_{f \sim \mu_{\mathcal{F}}} \left[\left| f(x) - f(y) \right| \right] = \|F(x) - F(y)\|_1.$$

Therefore, instead of talking about embeddings into $\mathbb{R}^n$, we can now equivalently work with random embeddings into $\mathbb{R}$. Our task is to identify a family of such embeddings and define a suitable distribution over them so that the expected ℓ_1 distance is close to d.

I guess you are already convinced in the class that we prefer to work with the following family of embeddings: Sample a set of vertices $A \subseteq V$ and embed every vertex $v \in V$ to $d(v, A)$ where $d(v, A) \triangleq \min_{u \in A} d(v, u)$. Let us denote this embedding by $f_A(\cdot)$. It is easy to see that $f_A(\cdot)$ never increases distance between vertices.

Proposition 2. *Let $d : V^2 \to \mathbb{R}$ be any semi-metric. For every $u, v \in V$ and every $A \subseteq V$, it holds that*

$$\left| f_A(u) - f_A(v) \right| \leq d(u, v).$$

Proof. Let $u', v' \in A$ be the points in A closest to u, v respectively. We assume without loss of generality that $f_A(u) \geq f_A(v)$, then

$$\left| f_A(u) - f_A(v) \right| = d(u, A) - d(v, A) = d(u, u') - d(v, v') \leq d(u, v') - d(v, v') \leq d(u, v).$$

□

Therefore, we only need to show that for some suitable choice of A, the distance between any two points after embedding does not shrink too much.

A simple strategy to sample A is to toss an independent p-biased coin on each vertex $x \in V$, and put x in A if and only if the coin goes HEAD. The following example sheds some light on how to choose p:

We assume the set V is partitioned into m clusters, namely $V = B_1 \sqcup B_2 \sqcup \cdots \sqcup B_m$. For every $u, v \in V$ that are in the same cluster, namely $u, v \in B_i$ for some i, we set $d(u, v) = 1$, otherwise we set $d(u, v) = |V|$.

Consider some $u \in B_i$ and $v \in B_j$ with $i \neq j$ and $|B_i| = \left|B_j\right| = k$. How can we sample a set A so that $\left| f_A(u) - f_A(v) \right|$ is large? In this special case, we expect one of the following two events happens:

(1) $A \cap B_i = \varnothing$ and $A \cap B_j \neq \varnothing$;
(2) $A \cap B_i \neq \varnothing$ and $A \cap B_j = \varnothing$.

If one of above events happens, then $\left|f_A(u) - f_A(v)\right| \ge |V| - 1$, otherwise $\left|f_A(u) - f_A(v)\right| \le 1$. Recall that we sample A by tossing p-biased coins, the event $A \cap B_i = \varnothing$ happens with probability $(1-p)^k \approx e^{-pk}$. Similarly, the event $A \cap B_j \neq \varnothing$ happens with probability $1 - (1-p)^k \approx 1 - e^{-pk}$. Therefore, if we choose $p \approx \frac{1}{k}$, then both probabilities are constant and we get large $\left|f_A(u) - f_A(v)\right|$ with constant probability.

If we need the above argument work for every pair of vertices u and v, we require each B_i is of similar size, so we can choose a uniform p. Moreover, the large contribution of A generated by $p \approx \frac{1}{k}$ comes from the fact that graph is well-clustered, namely the distance between points in different clusters is large. These properties do not hold for general graphs. We overcome these difficulties using two new ideas:

- instead of using a fixed value of p, we choose p from a large domain that can cover all the possible size of clusters;
- we don't expect that one single p contributes a lot, instead we amortize the analysis by showing that each possible value of p has its own contribution to the whole expectation.

The following is our algorithm to sample $f_A(\cdot)$:

Input: A semi-metric $d : V^2 \to \mathbb{R}$ with $|V| = n$.

1. Choose $t \in \left\{1, \ldots, \log_2 n\right\}$ uniformly at random.
2. Sample a set $A \subseteq V$ by selecting each $v \in V$ to be in A with probability $p \triangleq 2^{-t}$ independently.
3. Return $f_A(\cdot)$.

The reason that we choose t from $\Theta(\log n)$ numbers would be clear from the discussion later. In fact, the logarithmic factor here is exactly the one appeared in the statement of theorem 1.

We use $\mathcal{D}_t$ to denote the distribution of A conditional on the event that we choose t in step 1 above. Based on the discussion before, we know that for every pair of vertices u, v and for each $t \in \left\{1, \ldots, \log_2 n\right\}$, the contribution of the function $f_A(\cdot)$ with $A \sim \mathcal{D}_t$ is maximized when a cluster of about 2^t size around u is hit by A and a cluster of about 2^t size around v is avoided by A (or vice versa). This motivates the following definition and the proof strategy.

For a point $u \in V$ and $\ell \in \mathbb{N}$, we use $B(u, \ell)$ to denote the set of points in V whose distance to v is at most ℓ, namely $B(u, \ell) \triangleq \left\{v \in V \,:\, d(u,v) \le \ell\right\}$. It is called the *closed ball* of radius ℓ around u. Similarly, we define the *open ball* of radius ℓ around u as $B^o(u, \ell) = \left\{v \in V \,:\, d(u,v) < \ell\right\}$. For every $t \in \left\{0, 1, \ldots, \log_2 n\right\}$, define the function $\ell_t : V \to \mathbb{N}$ as

$$\ell_t(u) \triangleq \min_{\ell} \left\{\left|B(u, \ell)\right| \ge 2^t\right\}.$$

It then follows from this definition that

$$\left|B(u, \ell_t(u))\right| \ge 2^t, \text{ and } \left|B^o(u, \ell_t(u))\right| < 2^t.$$

In the following, we fix a pair of vertices $u, v \in V$. Let t^* be the maximum t such that both $\ell_{t^*}(u)$ and $\ell_{t^*}(v)$ are at most $\frac{d(u,v)}{2}$. We now claim that for every $t \le t^*$ and a set $A \sim \mathcal{D}_t$, it holds that (1) A hits $B(u, \ell_{t-1}(u))$ and (2) A avoids $B^o(v, \ell_t(v))$ with constant probability. In fact, (1) happens with probability $1 - (1 - 2^{-t})^{2^{t-1}} \ge 1 - e^{-\frac{1}{2}}$ and (2) happens with probability at least $(1 - 2^{-t})^{2^t} \ge \frac{1}{4}$. Moreover, the two events are independent since $t \le t^*$. Once the two events simultaneously happen, it contributes to $\left|f_A(u) - f_A(v)\right|$ at least $\ell_t(v) - \ell_{t-1}(u)$ (it is trivially true if $\ell_t(v) - \ell_{t-1}(u) < 0$). Therefore, for some constant $c > 0$, $\mathbf{E}_{A \sim \mathcal{D}_t}\left[\left|f_A(u) - f_A(v)\right|\right] \ge c \cdot (\ell_t(v) - \ell_{t-1}(u))$. We can swap the roles of u and v in the above argument and obtain $\mathbf{E}_{A \sim \mathcal{D}_t}\left[\left|f_A(u) - f_A(v)\right|\right] \ge c \cdot (\ell_t(u) - \ell_{t-1}(v))$. Note that these two cases never overlap, so we can add up their contribution to the expectation and obtain

$$\mathop{\mathbf{E}}_{A \sim \mathcal{D}_t}\left[\left|f_A(u) - f_A(v)\right|\right] \ge c \cdot (\ell_t(u) - \ell_{t-1}(u) + \ell_t(v) - \ell_{t-1}(v)).$$

On the other hand, by our choice of t^*, one of $\ell_{t^*+1}(u)$ and $\ell_{t^*+1}(v)$ is larger than $\frac{d(u,v)}{2}$. We assume $\ell_{t^*+1}(u) > \frac{d(u,v)}{2}$, then $\left|B^o\left(u, \frac{d(u,v)}{2}\right)\right| < 2^{t^*+1}$. Moreover, $\ell_{t^*} \le \frac{d(u,v)}{2}$ implies $B^o\left(u, \frac{d(u,v)}{2}\right) \cap B(v, \ell_{t^*}(v)) = \varnothing$. So similar argument gives

$$\mathop{\mathbf{E}}_{A \sim \mathcal{D}_{t^*+1}}\left[\left|f_A(u) - f_A(v)\right|\right] \ge c \cdot \left(\frac{d(u,v)}{2} - \ell_{t^*}(v)\right).$$

Therefore, if we use $\mathcal{D}$ to denote the distribution of A defined by our algorithm, then for every $u, v \in V$,

$$\begin{aligned}
\mathop{\mathbf{E}}_{A\sim\mathcal{D}}\left[\left|f_A(u)-f_A(v)\right|\right] &= \mathop{\mathbf{E}}_{t\in_R\{1,\ldots,\log_2 n\}}\left[\mathbf{E}\left[\left|f_A(u)-f_A(v)\right| \,\middle|\, t\right]\right] \\
&= \frac{1}{\log_2 n}\sum_{t=1}^{\log_2 n}\mathop{\mathbf{E}}_{A\sim\mathcal{D}_t}\left[\left|f_A(u)-f_A(v)\right|\right] \\
&\geq \frac{1}{\log_2 n}\sum_{t=1}^{t^*+1}\mathop{\mathbf{E}}_{A\sim\mathcal{D}_t}\left[\left|f_A(u)-f_A(v)\right|\right] \\
&\geq \frac{c}{\log_2 n}\cdot\left(\ell_{t^*}(u)-\ell_0(u)-\ell_0(v)+\frac{d(u,v)}{2}\right) \\
&\geq \frac{c}{2\log_2 n}.
\end{aligned}$$

This finishes the proof of theorem 1.

However, we cannot directly use theorem 1 to actually find a sparsest cut efficiently. The reason is that in our construction, the dimension m appeared in the statement is too large ($m = 2^{|V|}$ is the number of subsets of V). But if we allow small error, then we can use our sampling algorithm to sample only $\mathrm{poly}(|V|)$ many functions $f_A(\cdot)$. It is a straightforward application of the Chernoff bound to show that this polynomial dimension space is good enough with high probability.