Statistics 291: Lecture 16 (March 21, 2024)

Overlap Gap Property and Hardness of Optimization

Instructor: Mark Sellke

Scribe: Sabarish Sainathan

1 Overview

Today, we'll introduce the overlap gap property (OGP), which is a geometric frameework to show algorithmic hardness for random optimization problems. To illustrate this framework, we'll consider the problem of finding a maximum independent set in a sparse random graph. The OGP will demonstrate that a class of *local algorithms* will not be able to find such a set.

2 Random graph models

Recall that an *independent set* of a graph is a set of vertices with no connecting edges. A *maximum* independent set is one with the largest possible cardinality in the graph.

Consider the following random graph models:

- Erdős-Rényi: Given a collection of n vertices, an edge between any pair of vertices exists independently with probability d/n. In the setting where $n >> d$, the degree distribution of each vertex approaches Poisson(d) in distribution (from the Poisson limit theorem). We'll use the notation $G(n, d/n)$.
- Random regular graph ($G_{n,d}$): A d-regular graph with n vertices, uniformly randomly sampled from the collection of all such graphs.
  - *Aside:* How does one construct a random regular graph? We'll generate one using the *configuration model.*

    Fix n vertices. Attach d "half-edges" to each vertex, such that we have $d \cdot n$ half-edges overall. Iteratively match half-edges uniformly at random, and attach the matched half-edges to form edges between vertices.

    There are a few obscured details with this procedure: we assume that $d \cdot n$ is even, and we might have cycles, self-loops, or multiple parallel edges. In the limit $n \to \infty$, the number of occurrences of these aberrancies is $\sim$ Poisson($O_d(1)$) and therefore conditioning on obtaining a simple graph is contiguous with the original model.

In this lecture, we'll be interested in the setting $n >> d >> 1$ (for example, $d = \log\log n$.)

3 Finding maximum independent sets

We'll first consider a greedy algorithm for finding a maximal independent set, originating in [3]. Given a random graph $G(n, d/n)$,

- Uniformly randomly order the vertices. Let this ordering be $\{v_1, \ldots, v_n\}$.
- Let $I \leftarrow \varnothing$.
- For $j \in [n]$:
  - If vertex v_j has no neighbors in $\{v_1, \ldots, v_{j-1}\} \cap I$, then $I \leftarrow \{v_j\} \cup I$.
- Return I.

Let's form some intuition for the size of the sets generated by this algorithm. Since $n >> d$, we can assume that each iteration of the algorithm deletes roughly d vertices from eligibility and adds one vertex to I. Thus, $|I| \gtrsim n/d$.

The following analysis improves this lower bound by a $\log d$ term.

Theorem 3.1 (Grimmett-McDiarmid, [3])**.** *For $G(n, d/n)$, random-order greedy gives*

$$|I| = \frac{(1 \pm o(1)) \log d}{d} n$$

with high probability.

Proof. We'll give a heuristic proof, ignoring concentration in n. For each $k \in \mathbb{N}_0$, the number of iterations of the algorithm to proceed from $|I| = k$ to $|I| = k+1$ is a geometric random variable $\tau_k \sim \text{Geom}(p_k)$, where

$$p_k = (1 - d/n)^k.$$

corresponds to the probability that the current vertex is not connected to any vertex in I. Since $p_k \to e^{-dk/n}$ in the limit $n \to \infty$, we compute that the expected time at which $|I| = m$ is roughly

$$\mathbb{E}[T_{|I|=m}] = \sum_{k=1}^{m} \mathbb{E}[\tau_k] = \sum_{k=1}^{m} e^{dk/n}.$$

Using the fact that $d << n$, we conclude that $e^{d/n} \approx 1$ and make the approximation

$$\mathbb{E}[T_{|I|=m}] \approx m \cdot e^{dm/n} \leq n$$

when $m \approx n \log d/d$. The upper bound of n here follows from the fact that there are at most n iterations in the algorithm. By concentration of the indicator random variable $\chi_{T_{|I|=m} \leq n}$ around this expectation, one can show the asymptotic bounds for m hold w.h.p. □

The maximal independent sets obtained from this algorithm are not typically maximum, however. Indeed, the size of maximum independent sets in ER graphs is roughly twice the size, as seen in the following theorem.

Theorem 3.2 (Frieze '90)**.** *For $G(n, d/n)$, maximal set size $|I| = \frac{(2 \pm o(1)) \log d}{d} n$ whp.*

Proof. The upper bound is an application of the first-moment method; the lower bound, an application of the second-moment method. We'll prove the upper bound and refer the reader to the source for the lower bound.

Let $\mathscr{C}_m$ be the collection of independent sets of the graph of cardinality m. Then

$$\mathbb{E}[|\mathscr{C}_m|] = \binom{n}{m}(1-d/n)^{\binom{m}{2}},$$

where we have used linearity of expectation to sum over the event probabilities that each m-set of vertices is an independent set. Let $m = \delta n$ for $\delta := \delta(d,n)$ where δ depends on n through the sub-constant term in the theorem statement. Then

$$\exp\left(-\frac{d}{n}\cdot\binom{m}{2}\right) = \exp\left(-\frac{d\delta^2 n}{2} + o(1)\right) = \exp\left(-n\frac{(2+o(1))^2\log^2 d}{2d}\right)$$

By Stirling's approximation,

$$\begin{aligned}\binom{n}{m} &= \exp(n\log(1/1-\delta) + \delta n\log((1-\delta)/\delta) + o(1))\\ &= \exp(nH(\delta) + o(1))\\ &= \exp(-n\delta\log\delta + o(1))\\ &= \exp\left(n\frac{(2+o(1))\log^2 d}{d}\right).\end{aligned}$$

where the last approximation follows when $\delta(d,n) = o_d(1)$ (since this term dominates the entropy $H(\delta)$ in this case). For all $\epsilon > 0$, we notice that $(2+\epsilon)^2 > 2(2+\epsilon)$, so we may choose a sufficiently small $\epsilon := \epsilon(n,d) > 0$ such that

$$\mathbb{E}[|\mathscr{C}_m|] \le e^{-cn}$$

for some $c > 0$. By Markov's inequality,

$$\mathbb{P}(|\mathscr{C}_m| \ge 1) \le e^{-cn}.$$

□

Remark. This theorem is also true for random d-regular graphs. In the next section, we'll interchangeably use the two random graph models as a pedagogical simplification; the statements typically apply to both models.

4 R-local algorithms

Suppose first that our graph is a rooted, d-regular tree. It is easy to see that we obtain an asymptotically maximal independent set by choosing alternating levels of the tree: first, choose the root; then all vertices of depth 2; then, of depth 4; and so on, of depth $2k$ for $k \in \mathbb{N}_0$.

We might extend this reasoning to our (regular) random-graph setting. In an $O(1)$ neighborhood of a typical vertex, $G_{n,d}$ is isomorphic to a d-regular tree: the likelihood of a cycle in this neighborhood is infinitesimal since $d << n$. An algorithm might seek to patch together independent sets obtained from small neighborhoods of vertices in the tree. In the case of an ER graph, the setting is not too different: the small neighborhoods are (typically) Galton-Watson trees with an approximately Poisson(d) number of descendants.

This intuition inspires the notion of a "local algorithm," in which we treat each vertex as the root of a local tree in the graph.

Definition 4.1 (R-local algorithm.)**.** Suppose we are given an algorithm f that operates on an R-local neighborhood of any vertex (where the neighborhood is defined according to a distance metric d — in our case, shortest path distance). An R-local algorithm consists of the following meta-routine:

(a) Generate i.i.d. random variables ω_v for $v \in V$ (for example, $\omega_v \sim \text{Unif}([0,1])$).

(b) Let $\chi_{v\in I} = f_R(\{\omega_w : d(v,w) < R\})$.

Note that I is called an "iid factor" in the literature [2].

Remark. This approach generates (a local version of) random-order greedy. Indeed, use the set of ω_v to decide the (local) random ordering, and apply the remainder of the algorithm as before.

In fact, it's not too bad that we're moving from a global algorithm to a local algorithm: we're unlikely to have long dependency chains, since length-L chains occur with probability $O(1/L!)$. More generally, [**?**] conjectured that local algorithms can optimally solve any combinatorial optimization problem on sparse graphs. This conjecture was proven true for finding maximum matchings. It is not true for finding maximum independent sets, as we will see in the next section.

Remark. On the other hand, local algorithms are not sufficient if the graph is equipped with some kind of global structure.

Indeed, consider the *stochastic block model*, in which the collection of vertices are bifurcated into two sets of $n/2$ vertices each. An edge between sets exists with probability b/n, and an edge within a set exists with probability a/n, where $a > b$. If $a >> b$ one can find the bifurcation by computing the top eigenvector of the normalized adjacency matrix. This is patently a *global* algorithm, and local algorithms encounter difficulty in performing this computation.

5 Algorithmic hardness and the OGP

[2] disproves the conjecture of [4] up to a constant factor of the maximum independent set.

Theorem 5.1 (Gamarnik-Sudan '13, [2])**.** *For $\beta = 1/\sqrt{2}$, $\gamma \in (1+\beta, 2])$, $d \geq d_*(\gamma)$ large, and $R := R(d,\gamma)$, no R-local algorithm produces independent sets of average size $m := \frac{\gamma \log d}{d} \cdot n$.*

Remark. Since $\gamma < 2$, there exists a multiplicative size gap between the maximum independent sets and those generated by any local algorithm (thereby disproving the conjecture).

In fact, the size gap is even larger: the same theorem holds for any $\gamma > 1$, proven in [5]. This was an impressive result, since the independent sets of size $(1+\epsilon)\frac{\log d}{d} n$ are (typically) *clustered*: their overlaps either have large cardinality (if they belong to the same cluster) or very small cardinality (different clusters) [1].

Proof. The moral approach is the following. Extending the analysis in [1], we will show that for all γ satisfying the theorem statement, independent sets of size m satisfy an overlap gap property: they overlap at least $(1+\delta)\frac{\log d}{d} n$ or at most $(1-\delta)\frac{\log d}{d} n$ for some $z \in (0,\beta)$, with high probability for large n. By a particular coupling construction, we identify pairs of independent sets generated by local algorithms that have overlap spanning this gap interval, obtaining a contradiction.

For the coupling, we'll apply an interpolation strategy on random graphs. Given $p \in [0,1]$, let $G, \tilde{G} \sim G(n, d/n)$ be two random graphs that are p-correlated. In the case of ER graphs, we construct a pair of such graphs by iterating over every pair of vertices. In each iteration, generate $Y \sim \text{Bernoulli}(p)$ independently. If $Y = 1$, then add an edge between the current pair of vertices with probability d/n to both graphs. If $Y = 0$, add an edge between the current pair of vertices with probability d/n independently to each graph. The remainder of the proof is easier to see in the random regular graph case, though we note again that the result applies to both random graph models.

Consider an R-local algorithm $\mathscr{A}$ that maps a graph to an independent set. Note that it's not necessary to parameterize $\mathscr{A}$ in terms of the size of graph, since it's a local algorithm operating on an $O(1)$ neighborhood.

Let $\alpha := \mathbb{E}[|I_{\mathscr{A}}|/n]$. Let $f : [0,1] \to [0,1]$ be defined as $f(p) = \mathbb{E}[|I_{\mathscr{A}(G)} \cap I_{\mathscr{A}(\tilde{G})}|/n]$.

Lemma 5.2. *$f(1) = \alpha$, $f(0) = \alpha^2$, and f is continuous for random regular graphs.*

That $f(1) = \alpha$ is immediate, since $G, \tilde{G}$ are identical (by construction) in this case.
$f(0) = \alpha^2$ by independence of the two graphs:

$$f(0) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{P}(v_i \in \mathscr{A}(G))^2 = \alpha^2.$$

And third, f is continuous since perturbing p by a small amount will with high probability preserve the local neighborhood of most vertices. This is of course an intuitive argument, made concrete in the source.

Lemma 5.3. *If $\alpha \geq \gamma \log d/d$, there exists p such that $f(p) = \log d/d$.*

By the upper bound from Frieze, $\alpha < \frac{2(1+\epsilon)d}{\log d} n$ for some small $\epsilon > 0$, so $\alpha^2 = O(\log^2 d/d^2) \leq \log d/d$ for sufficiently large d. Given the assumption on α, the intermediate value theorem guarantees that f takes on every value in the interval $[\alpha^2, \alpha]$ — in particular, $\log d/d$.

Lemma 5.4. *$|I_{\mathscr{A}(G)} \cap I_{\mathscr{A}(\tilde{G})}|/n$ concentrates (exponentially) around $f(p)$, uniformly in p.*

This is easier to see in the random regular graph model. Here's some intuition: the set of vertices in both sets are mostly independent by sparsity of the graph. Let $\bar{G} = G \cup \tilde{G}$; we'll partition this set into a collection of sets such that each partition consists of vertices distance $2R$ away from one another: $[n] = \bigcup_{i=1}^{M} J_i$. Note that if $v, w \in J_1$, $d(v,w) > 2R$. A quick bound shows that $M \leq (2d)^{2R}$. This partition makes use of the fact that $|I \cap J_i|$ is an iid sum of (bounded) random variables, which implies exponential concentration around $f(p) \cdot |J_i|$, Hence $f(p) \cdot n \approx |I|$ exponentially.

Lemma 5.5 (Overlap gap)**.** *For small $\delta > 0$ such that $\delta < \sqrt{\beta^2 - 1} < \beta$, the number of independent sets of size in $((1-\delta)d\,n/\log d, (1+\delta)d\,n/\log d)$ approaches 0 in probability.*

Note that the constraint on β implies that $\beta \in (1/\sqrt{2}, 1)$.

By exponential concentration of the overlap of independent sets generated by a random algorithm around $n \log d/d$, there exists a pair of independent sets of the graph with overlap within $(1-\epsilon, 1+\epsilon) n \log d/d$ with high probability in n; this contradicts the overlap gap property. □

References

[1] Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, May 2014. 4

[2] David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *The Annals of Probability*, 45(4), 2017. 4

[3] Geoffrey R Grimmett and Colin JH McDiarmid. On colouring random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 77, pages 313–324. Cambridge University Press, 1975. 2

[4] Hamed Hatami, László Lovász, and Balázs Szegedy. Limits of locally–globally convergent graph sequences. *Geometric and Functional Analysis*, 24(1):269–296, 2014. 4

[5] Mustazee Rahman and Bálint Virág. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3), May 2017. 4