CS294 Partition Functions **Fall 2020**

Lecture 22: November 12

Instructor: Alistair Sinclair *Scribes: Tarun Kathuria & Fred Zhang*

Disclaimer: *These notes have not been subjected to the usual scrutiny accorded to formal publications. They may be distributed outside this class only with the permission of the Instructor.*

22.1 Sly's reduction

In this lecture, we continue with Sly's proof of the hardness of approximating the partition function (equivalently, sampling from the Gibbs measure) of the hard-core model above the uniqueness threshold [Sly10]. Recall that the hard-core model is a distribution over the independent sets I of a given graph G, with weights proportional to $\lambda^{|I|}$, where λ is the fugacity parameter.

We will give a randomized reduction from computing the maximum cut of an arbitrary graph H to approximating the partition function $Z_{H^G}(\lambda)$ of a graph H^G of maximum degree Δ derived by combining H with a degree-Δ gadget G discussed in the previous lecture, at any choice of $\lambda > \lambda_c(\Delta)$. (The parameters of the graph G will depend on λ.) This shows that an FPRAS for approximating the partition function of graphs of maximum degree Δ at any $\lambda > \lambda_c(\Delta)$ would imply that NP = RP.

Given an arbitrary graph H, the reduction produces a graph H^G with gadgets G embedded in it. The reduction begins by constructing such gadgets $G = G(n, \theta, \psi)$ as described in the last lecture, where θ, ψ are constants that depend on λ and Δ. We briefly review that construction here.

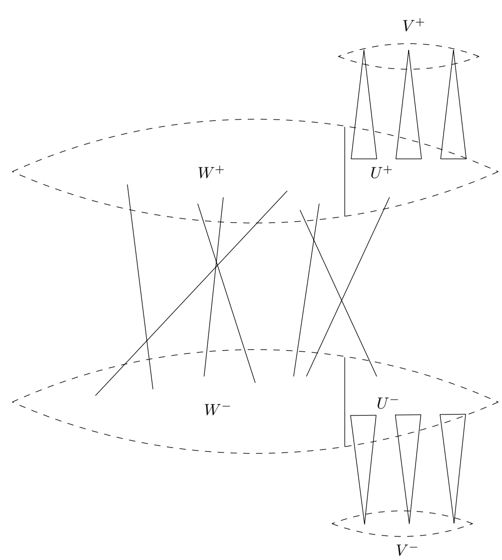

Figure 22.1: The basic gadget G, consisting of a random bipartite graph and a set of disconnected trees.

The basic gadget The basis of the gadget G is a random bipartite graph with vertex set $(W^+ \cup U^+, W^- \cup U^-)$. Each side has $n + n^{\theta+\psi}$ vertices, where $|W^+| = |W^-| = n$ and $|U^+| = |U^-| = n^{\theta+\psi}$. To connect these vertices, we take

- $\Delta - 1$ random perfect matchings between the $W^+ \cup U^+$ and $W^- \cup U^-$; and
- one additional random perfect matching only between W^+ and W^-.

As the next step, we create n^θ disjoint Δ-regular trees, each with n^ψ leaves in U^+ and U^-. We denote the roots of the trees by V^+ and V^- and call them *terminals*. Note that the resulting graph has maximum degree Δ, and the $2n^\theta$ terminals each have degree $\Delta - 1$.

Recall from the last lecture the following two key definitions:

Definition 22.1. *The* phase *of a configuration σ on G is defined as*

$$S(\sigma) = \begin{cases} \oplus & \text{if } |I(\sigma) \cap W^+| > |I(\sigma) \cap W^-| \\ \ominus & \text{o.w.} \end{cases}$$

For a configuration σ, denote by σ_V the restriction of σ to the terminal vertices $V = V^+ \cup V^-$.

Definition 22.2. *The product measures $Q_V^\pm$ on configurations σ_V are defined by:*

$$Q_V^+(\sigma_V) = (p_2)^{|I(\sigma_V)\cap V^+|}(1-p_2)^{|V^+|-|I(\sigma_V)\cap V^+|}(p_1)^{|I(\sigma_V)\cap V^-|}(1-p_1)^{|V^-|-|I(\sigma_V)\cap V^-|},$$

$$Q_V^-(\sigma_V) = (p_1)^{|I(\sigma_V)\cap V^+|}(1-p_1)^{|V^+|-|I(\sigma_V)\cap V^+|}(p_2)^{|I(\sigma_V)\cap V^-|}(1-p_2)^{|V^-|-|I(\sigma_V)\cap V^-|}.$$

In the last lecture, we stated (without proof) the following main technical result of [Sly10].

Lemma 22.3. *For any Δ and $\lambda > \lambda_c(\Delta)$, there exist constants (θ, ψ) (depending on Δ and λ) such that the random bipartite graph G has $(2 + o(1))n$ vertices and satisfies the following properties w.h.p.:*

(i) In the Gibbs measure on configurations with partition function $Z_G(\lambda)$, both phases occur with reasonable probability, i.e., $\Pr_\sigma[S(\sigma) = \oplus] \geq \frac{1}{n}$, $\Pr_\sigma[S(\sigma) = \ominus] \geq \frac{1}{n}$;

(ii) Conditioned on the phase S, the marginal of the Gibbs distribution on spins in V is very close to one of the production distributions $Q_V^\pm$, i.e.,

$$\max_{\sigma_V} \left| \frac{\pi[\sigma_V \mid S = \oplus]}{Q_V^+[\sigma_V]} - 1 \right| \leq n^{-2\theta}; \quad \max_{\sigma_V} \left| \frac{\pi[\sigma_V \mid S = \ominus]}{Q_V^-[\sigma_V]} - 1 \right| \leq n^{-2\theta}.$$

We now define the full reduction from MaxCut.

Full reduction Suppose H is an arbitrary graph (input to MaxCut) with $n^{\theta/4}$ vertices. Fix Δ and $\lambda > \lambda_c(\Delta)$, and then construct a random G with the appropriate parameters (n, θ, ψ) as above. We now define H^G as follows:

- For each vertex $x \in V(H)$, create a disjoint copy of G and name it G_x. Let V_x^+ and V_x^- denote the respective terminal vertices of G_x. Denote the union of all the copies of G by $\widehat{H}^G$.

- For each edge $\{x, y\} \in E(H)$, add a matching of size $n^{3\theta/4}$ between V_x^+ and V_y^+, and similarly add such a matching between V_x^- and V_y^-. Note that since $|V_x^+| = |V_x^+| = n^\theta$, and the degree of H is at most $n^{\theta/4}$, we can make all these matchings disjoint, so no terminal has its degree increased by more than 1.

This yields a graph H^G of maximum degree Δ.

22.2 Proof of main result

We now verify that sampling from the hard-core distribution in H^G at parameter λ allows us to find a maximum cut in H. Given a hard-core configuration σ on H^G, let $S = (S_x)_{x \in V(H)}$ denote the vector of phases in each of the subgraphs G_x. Note that each S naturally gives rise to a cut in H, defined by $\mathrm{Cut}(S) = \{\{x, y\} \in E(H) : S_x \neq S_y\}$. Also, define $Z_{H^G}[S]$ to be the partition function Z_{H^G} restricted to configurations with phase vector S.

The following consequence of Lemma 22.3 shows how the effect of the additional matching edges is to place larger weight on larger cuts.

Lemma 22.4. *Suppose G satisfies the two properties of Lemma 22.3. Then*

$$\frac{Z_{\widehat{H}^G}[S]}{Z_{\widehat{H}^G}} \geq n^{-n^{\theta/4}} \tag{22.1}$$

and

$$\frac{Z_{H^G}[S]}{Z_{\widehat{H}^G}[S]} = (\alpha(H) + o(1)) \left[\frac{(1 - p_1 p_2)^2}{(1 - p_1^2)(1 - p_2^2)} \right]^{n^{3\theta/4}|\mathrm{Cut}(S)|}, \tag{22.2}$$

where $\alpha(H) := \left[(1 - p_1^2)(1 - p_2^2) \right]^{n^{3\theta/4}|E(H)|}$.

Note that $(1 - p_1 p_2)^2 - (1 - p_1^2)(1 - p_2^2) = (p_1 - p_2)^2 > 0$, so the weight factor in (22.2) is exponentially increasing with $|\mathrm{Cut}(S)|$.

Proof. For (22.1), note that since the graph $\widehat{H}^G$ consists of a collection of disjoint copies of G, the distribution on a configuration on $\widehat{H}^G$ is a product measure over configurations of $(G_x)_{x \in V(H)}$. This implies that the phases are independent. Hence, the claim is immediate from property (i) of Lemma 22.3.

For (22.2), the ratio on the left side is precisely the probability that a configuration σ with phase vector S sampled from the hardcore distribution on $\widehat{H}^G$ is also an independent set of H^G. By property (ii) of Lemma 22.3, conditional on the phases, the spins on the terminals of each G_x are (almost) independent with probabilities p_1, p_2 respectively. Now for any given edge $\{x, y\} \in E(H)$, let $E_{x,y}$ be the event that none of the matching edges between G_x, G_y has both endpoints in an independent set in $\widehat{H}^G$. Then we have:

$$\frac{Z_{H^G}[S]}{Z_{\widehat{H}^G}[S]} = (1 + o(1)) \prod_{\{x,y\} \in E(H)} \Pr(E_{x,y} | S). \tag{22.3}$$

Here, we consider two cases. By our construction of H^G and property (ii) of Lemma 22.3:

- if $S_x = S_y$, then $\Pr(E_{x,y}|S) \sim (1 - p_1^2)(1 - p_2^2)$; and

- if $S_x \neq S_y$, then $\Pr(E_{x,y}|S) \sim (1-p_1p_2)^2$.

It follows that in the product term of (22.3), we get a factor of

$$\left[\left(1-p_1^2\right)\left(1-p_2^2\right)\right]^{n^{3\theta/4}}$$

for each non-cut edge of H, and a factor of

$$\left[\frac{\left(1-p_1p_2\right)^2}{\left(1-p_1^2\right)\left(1-p_2^2\right)}\right]^{n^{3\theta/4}}$$

for each cut edge of H. Combining these two estimates, and pulling out a factor of $\alpha(H) = \alpha(H) := \left[\left(1-p_1^2\right)\left(1-p_2^2\right)\right]^{n^{3\theta/4}|E(H)|}$, yields (22.2). The $o(1)$ error term comes from absorbing the $n^{-2\theta}$ errors in property (ii) of Lemma 22.3. $\square$

Armed with Lemma 22.4, we can now prove that the reduction works. Suppose that we have an FPRAS for $Z_{G^H}(\lambda)$ at the (arbitrary) value $\lambda > \lambda_c(\Delta)$ used to construct G. Then by standard methods we can sample independent sets in H^G in polynomial time with very small error (which we shall ignore). We identify the sampled independent sets with their phase vectors, which in turn correspond to cuts in H.

Now consider two cuts, C, C' in H, with $|C| > |C'|$, corresponding to phase vectors S, S'. The ratio of sampling probabilities for these cuts is

$$\begin{aligned}
\frac{Z_{H^G}[S]}{Z_{H^G}[S']} &= (1+o(1)) \times \frac{Z_{\widehat{H}^G}[S]}{Z_{\widehat{H}^G}[S']} \times \left[\frac{\left(1-p_1p_2\right)^2}{\left(1-p_1^2\right)\left(1-p_2^2\right)}\right]^{n^{3\theta/4}(|\operatorname{Cut}(S)|-|\operatorname{Cut}(S')|)} \\
&\geq (1+o(1))n^{-n^{\theta/4}}\left[\frac{\left(1-p_1p_2\right)^2}{\left(1-p_1^2\right)\left(1-p_2^2\right)}\right]^{n^{3\theta/4}} \\
&\geq 4^{n^{\theta/4}},
\end{aligned}$$

for sufficiently large n. The first inequality here comes from (22.2) and the second from (22.1), using the fact observed earlier that $\frac{(1-p_1p_2)^2}{(1-p_1^2)(1-p_2^2)} = (1+\varepsilon)$ for some $\varepsilon > 0$. But the total number of cuts in H is only $2^{|V(H)|} = 2^{n^{\theta/4}}$, so the probability that the algorithm outputs a maximum cut is at least

$$\frac{4^{n^{\theta/4}}}{4^{n^{\theta/4}} + 2^{n^{\theta/4}}} = 1 - 2^{-|V(H)|}.$$

Thus with very high probability we solve MaxCut (exactly) in H in polynomial time. This concludes the proof of Sly's theorem.

22.3 Polynomial Interpolation Method

We now move on to a completely different approach to computing partition functions which is based on polynomial interpolation, pioneered by Barvinok [Bar16]. The idea is to view the partition function $Z(\lambda)$ as a polynomial in the *complex plane* (even though as usual we are mainly interested only in real values of λ). The key idea of the method is to identify an open connected region $R \subseteq \mathbb{C}$ which includes a portion of the real line containing the interval $[\lambda_0, \lambda]$, where λ is the point at which we want to evaluate Z and λ_0

is an easy point for Z, and which contains no complex zeros of Z. We may then consider a Taylor series expansion of $\log Z$ around λ_0, which is convergent in R (due to the lack of zeros), and argue that the first "few" terms of this series gives a good additive approximation to $\log Z(\lambda)$, which would in turn give us a good multiplicative approximation of $Z(\lambda)$.

We now illustrate this approach with reference to the classical ferromagnetic Ising model, as derived in [LSS19].

Example: Ferromagnetic Ising model. Recall that the partition function of this model on a graph $G = (V, E)$ is

$$Z_G(\lambda, \mu) = \sum_{S \subseteq V} \lambda^{|E(S,\bar{S}|} \mu^{|S|},$$

with S representing the set of vertices with positive spin; here λ, μ are parameters. We assume $\lambda \leq 1$ so that the model is ferromagnetic. Note also that the model is symmetric around $\mu = 1$ (since replacing μ by μ^{-1} just multiplies the polynomial by μ^n); this reflects the fact that the sign of the external field is not important.

We will fix λ and view Z_G as a univariate polynomial in μ; we'll write $Z(\mu) := Z_G(\lambda, \mu)$. We also write $f(\mu) := \log Z(\mu)$.

First we prove a simple claim which shows that an additive approximation of $\log Z$ implies a multiplicative approximation of Z.

Claim 22.5. *For any $\varepsilon \leq 1/4$ and any μ, if we have $|\widetilde{f} - f(\mu)| \leq \varepsilon$, then $|\exp(\widetilde{f}) - Z(\mu)| \leq 4\varepsilon Z(\mu)$.*

Proof.

$$\begin{aligned} |\exp(\widetilde{f}) - Z(\mu)| &= |\exp(\widetilde{f}) - \exp(f(\mu))| \\ &= \exp(f(\mu))|\exp(\widetilde{f} - f(\mu)) - 1| \\ &\leq 4\varepsilon \exp(f(\mu)), \end{aligned}$$

where the last inequality follows from elementary complex analysis (or equivalently, geometry on $\mathbb{Z}^2$) **[exercise!]**. □

Next we identify an easy point μ_0 for $Z(\mu)$: this will be $Z(0)$, which obviously takes the value 1. We will therefore compute the Taylor expansion of $f(\mu)$ around 0, which is

$$f(\mu) = \sum_{i=0}^{\infty} \frac{f^{(i)}(0)\mu^i}{i!}, \tag{22.4}$$

where $f^{(i)}$ denotes the ith derivative of f. Now suppose we have an open disk $B(0, \ell)$ in which Z has no zeros, and let μ be any point in this disk at which we with to evaluate Z. (We'll show how to find such a disk later.) Then we know that the expansion (22.4) is convergent in this disk. Our goal is to bound the error induced by truncating this expansion to m terms. To this end, define $f_m(\mu) := \sum_{i=0}^{m} \frac{f^{(i)}(0)\mu^i}{i!}$.

To bound the error $|f_m(\mu) - f(\mu)|$, we can equivalently write the Taylor series in terms of the roots of Z, as follows. Let $r_1, \ldots, r_n$ be the complex zeros of $Z(\mu)$. Since the constant term of Z is $Z(0) = 1$, we can write $Z(\mu) = \prod_{i=1}^{n} \left(1 - \frac{\mu}{r_i}\right)$ and hence

$$f(\mu) = \sum_{i=1}^{n} \log\left(1 - \frac{\mu}{r_i}\right) = -\sum_{i=1}^{n} \sum_{j=0}^{\infty} \frac{1}{j} \left(\frac{\mu}{r_i}\right)^j, \tag{22.5}$$

where we have used the Taylor expansion of $\log(1-x)$. By uniqueness of the Taylor expansion of meromorphic functions in its domain of convergence, we know that the expansions (22.4) and (22.5) are equivalent.

Thus we may bound the truncation error as follows:

$$\begin{aligned} |f(\mu) - f_m(\mu)| &\leq n \sum_{j=m+1}^{\infty} \frac{|\mu/\ell|^j}{j} \\ &\leq n \frac{|\mu/\ell|^{m+1}}{m+1} \sum_{j=0}^{\infty} |\mu/\ell|^j \\ &\leq n \frac{|\mu/\ell|^{m+1}}{(m+1)(1-|\mu/\ell|)} \end{aligned} \tag{22.6}$$

In the first inequality we used the fact that all the roots r_i lie outside the open disk $B(0,\ell)$, so $|r_i| \geq \ell$ for all i. Similarly, in the last inequality we used the fact that $|\mu| < \ell$ so that the geometric series converges.

To obtain a $1 \pm \varepsilon$ multiplicative approximation of $Z(\mu)$, by Claim 22.5 it suffices to get the additive error in (22.6) down to $\varepsilon/4$. Thus we need to take

$$m = \frac{1}{\log |\mu/\ell|^{-1}} \left(\log \frac{4n}{\varepsilon} + \log \left(\frac{1}{1 - |\mu|/\ell} \right) \right) = O\left(\log\left(n/\varepsilon\right)\right) \tag{22.7}$$

for any fixed $|\mu| < \ell$. Hence, it suffices to truncate the Taylor series at $m = O\left(\log\left(n/\varepsilon\right)\right)$ terms.

In the next lecture we'll discuss how to compute the truncated Taylor expansion efficiently, as well as the most interesting issue of identifying a zero-free disk for Z.

References

[Bar16] A. Barvinok. *Combinatorics and Complexity of Partition Functions.* Springer International, 2016.

[LSS19] J. Liu, A. Sinclair, and P. Srivastava. The Ising model partition function: Zeros and deterministic approximation. *Journal of Statistical Physics*, 174:287–315, 2019.

[Sly10] A. Sly. Computational transition at the uniqueness threshold. *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 287–296, 2010.