CS294 Partition Functions	**Fall 2020**

Lecture 18: October 29

Instructor: Alistair Sinclair *Scribes: Han Feng and Elizabeth Yang*

18.1 Proof of base case of Lemma 17.10

In the last lecture, we introduced a Markov chain for sampling matroid bases: the down-up walk, also known as the basis exchange walk. Each transition from a basis B involves dropping a ground set element uniformly at random, and adding a new element proportional to a weight function $w(\cdot)$.

Recall that for a basis B, we set its weight $w(B)$ to 1. For any independent set I with $|I| < r$, we defined $w(I) = \sum_{e \in E} w(I \cup \{e\})$, where E is the matroid's ground set. Based on the way weights are inherited as we decrease $|I|$, there is an explicit expression for $w(I)$

$$w(I) = (r - |I|)! \cdot \#\{B : B \supseteq I\}.$$

The weights lead to a probability distribution over bases with normalizing constant $Z_r = |\mathcal{B}|$. They also induce distributions over the independent sets of size k, denoted $\mathcal{M}(k)$, with normalizing factors $Z_k = Z_r \cdot \frac{r!}{k!}$. We generalized the basis exchange walk to a random walk on each level $\mathcal{M}(k)$; the transition matrix of this walk, denoted $P_k^\vee$, in each step removes a uniformly random element of the current set and replaces it with a new element with probabilities proportional to the weights of sets in $\mathcal{M}(k)$.

Last time, we used the following key lemma to prove a $\frac{1}{k}$ lower bound on the (modified) log-Sobolev constant of $P_k^\vee$. This implies a bound on the mixing time of the Markov chain of $O(r(\log r + \log\log n)) = O(n \log n)$.

Lemma 17.10. *For any $k \geq 2$ and function $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geq 0}$ with $\mathrm{E}_{\pi_k} f^{(k)} = 1$,*

$$\mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \leq \left(1 - \frac{1}{k}\right) \mathrm{Ent}_{\pi_k}(f^{(k)}).$$

Lemma 17.10 is proved by induction on k. In the last lecture, we saw the induction step; today we will prove the base case $k = 2$, which is surprisingly nontrivial.

Proof of base case of Lemma 17.10. Rewriting the inequality in the theorem and substituting $k = 2$, our goal is to prove

$$\mathrm{Ent}_{\pi_2} f^{(2)} - 2\mathrm{Ent}_{\pi_1} f^{(1)} \geq 0. \tag{18.1}$$

Recall that we are assuming the normalization $\mathrm{E}_{\pi_2} f^{(2)} = 1$, which as we've seen implies $\mathrm{E}_{\pi_1} f^{(1)} = 1$. (These normalizations are w.l.o.g. as they don't affect the ratio in the log-Sobolev constant, as we observed last time.)

Recalling from last lecture the definition $f^{(k-1)} = P^{\uparrow}_{k-1} f(k)$, we first express $f^{(1)}$ in terms of $f^{(2)}$:

$$f^{(1)}(e) = (P^{\uparrow}_1 f(2))(e) = \sum_{s: e\in s} \frac{w(s)}{w(e)} f^{(2)}(s).$$

(Throughout this proof, we will use e, e' to denote elements of the ground set, and s to denote an independent set of size 2.) This leads to the following expression for the entropy:

$$\begin{aligned}
\mathrm{Ent}_{\pi_1} f^{(1)} &= \sum_e \pi_1(e) \left(\sum_{s: e\in s} \frac{w(s)}{w(e)} f^{(2)}(s) \right) \log f^{(1)}(e) \\
&= \sum_s \sum_{e\in s} \pi_1(e) \frac{w(s)}{w(e)} f^{(2)}(s) \log f^{(1)}(e) \\
&= \sum_{s=(e,e')} \frac{w(s)}{Z_1} f^{(2)}(s) \left(\log f^{(1)}(e) + \log f^{(1)}(e') \right),
\end{aligned}$$

where in the last line we recall that $\pi_1(e) = \frac{w(e)}{Z_1}$. Similarly, we can use the definition of entropy to write

$$\begin{aligned}
\mathrm{Ent}_{\pi_2} f^{(2)} &= \sum_s \pi_2(s) f^{(2)}(s) \log f^{(2)}(s) \\
&= \sum_s \frac{w(s)}{Z_2} f^{(2)}(s) \log f^{(2)}(s).
\end{aligned}$$

As a result

$$\begin{aligned}
\mathrm{Ent}_{\pi_2}(f^{(2)}) - 2\mathrm{Ent}_{\pi_1}(f^{(1)}) &= \sum_{s=(e,e')} \frac{w(s)}{Z_2} f^{(2)}(s) \left[\log f^{(2)}(s) - \log \left(f^{(1)}(e) f^{(1)}(e') \right) \right] \\
&\geq \sum_{s=(e,e')} \pi_2(s) f^{(2)}(s) - \sum_{s=(e,e')} \frac{w(s)}{Z_2} f^{(1)}(e) f^{(1)}(e') \\
&= 1 - \frac{1}{2Z_2} {f^{(1)}}^T W f^{(1)}.
\end{aligned}$$

where in the first line we used $Z_1 = 2Z_2$ (because every element in $\mathcal{M}(2)$ contributes its weight to exactly two elements in $\mathcal{M}(1)$), and in the second line we used the elementary inequality

$$a \log \frac{a}{b} \geq a - b, \quad \text{for all } a \geq 0, b > 0.$$

The matrix W above is symmetric with entries $W(e,e') = w(\{e,e'\})$. To finish the base case, comparing with (18.1) it suffices to show that

$${f^{(1)}}^\top W f^{(1)} \leq 2Z_2 = Z_1. \tag{18.2}$$

The proof of (18.2) involves an excursion into the geometry of polynomials and log-concave distributions. There is a large recent literature on this topic, but we will just cover what is necessary to prove the inequality.

Let $\pi : 2^{|E|} \to \mathbb{R}_{\geq 0}$ be a discrete probability distribution. The *generating polynomial* of π is defined as

$$g_\pi(x_1, \ldots, x_{|E|}) = \sum_{S\subseteq E} \pi(S) \prod_{i\in S} x_i.$$

Definition 18.1. *The polynomial g is*

- ***log-concave*** *at x if $\log g$ is concave at x, i.e., $\nabla^2 \log g(x)$ is negative semi-definite.*
- ***strongly log-concave*** *(SLC) if $\nabla^2 \log \partial_I g(1)$ is negative semidefinite for any $I \subseteq \mathcal{E}$. Here, ∂_I represents the partial derivatives taken with respect to all x_i with $i \in I$.*

We have the following beautiful theorem due to Anari et.al [ALOV18] and Brändén-Huh [BH19]:

Theorem 18.2. *The generating polynomial of the uniform distribution over the bases of any matroid is SLC.*

Note: A converse-like statement is also true: the support of any SLC distribution forms the bases of some matroid (though we will not need this here). Note also that all the results presented here actually hold in the more general setting of an arbitrary (homogeneous) SLC distribution; the proof is essentially identical, with the more general SLC distribution replacing the uniform weights on the bases in $\mathcal{M}(r)$ and all other weights inherited from them.

The generating polynomial for the uniform distribution over bases is

$$g_\pi = \frac{1}{|\mathcal{B}|} \sum_{B \in \mathcal{B}} \prod_{i \in B} x_i.$$

The derivatives $\partial_I g_\pi(1)$ also have a nice interpretation as the probability that a uniform random basis contains I:

$$\partial_I g_\pi(1) = \frac{1}{|\mathcal{B}|} \left|\{B : B \supseteq I\}\right| = \frac{w(I)}{|\mathcal{B}| \cdot (r - |I|)!}, \tag{18.3}$$

where we used the explicit expression for $w(I)$ stated earlier.

We need another result, which is a consequence of SLC [ALOV18].

Lemma 18.3. *If g is SLC, then $\nabla^2 \partial_I g(1)$ has at most one positive eigenvalue.*

This result gives useful information about the eigenvalues of the matrix W, because from (18.3) W can be written as

$$W = (r-2)! \cdot |\mathcal{B}| \cdot \nabla^2 g_\pi(1).$$

which is just a scaling of the Hessian of $g_\pi(1)$ at 1. Thus Lemma 18.3 tells us that W has at most one positive eigenvalue.

Let $f : \mathcal{M}(1) \to \mathbb{R}_{\geq 0}$ be a function with $\mathrm{E}_{\pi_1} f = \sum_e \pi_1(e) f(e) = 1$, or equivalently, $\sum_e w(e) f(e) = Z_1$. Let $D = \mathrm{diag}(w(e))$ and $A = D^{-1/2} W D^{-1/2}$, so $A(e,e') = \frac{1}{\sqrt{w(e)w(e')}} W(e,e')$ is real symmetric and admits the spectral decomposition

$$A = \sum_i \lambda_i g_i g_i^\top.$$

Here, the g_i are an orthonormal bases of eigenvectors of A, and the λ_i are the corresponding eigenvalues. **Exercise:** Show that A also has at most one positive eigenvalue, and that $\sqrt{w(e)}$ is an eigenvector with eigenvalue 1. Thus, we can take $g_1 = \sqrt{\pi_1}$, $\lambda_1 = 1$, and $\lambda_i \leq 0$ for all $i \geq 2$.

From the decomposition of A, we can recover the following decomposition of W:

$$W = \sum_i \lambda_i h_i h_i^\top,$$

where $h_i = D^{1/2} g_i$. In particular $h_1 = D^{1/2} g_1 = \sqrt{\pi_1}\sqrt{w} = \frac{w}{\sqrt{Z_1}}$. From this, we obtain

$$\langle h_1, f\rangle = \sum_e \frac{w(e)}{\sqrt{Z_1}} f(e) = \sqrt{Z_1},$$

because $\sum_e w(e) f(e) = Z_1$ by assumption. Finally,

$$f^\top W f = \sum_i \lambda_i \langle h_i, f\rangle^2 \leq \langle h_1, f\rangle^2 = Z_1,$$

because $\lambda_i \leq 0$ for $i \geq 2$ and $\lambda_1 = 1$. This completes the proof of (18.1) and hence the proof of Lemma 17.10.
□

18.2 Correlation Decay

We now move from MCMC to another technique called correlation decay, which gives deterministic algorithms for approximating partition functions. Take the canonical problem of approximating the partition function of a spin system. Intuitively, if the interaction between neighboring spins is weak enough, then they behave "almost independently"; in particular, the value of two spins that are far apart in the graph should be asymptotically uncorrelated. This should in turn mean that we can approximate the marginal distribution of the spin at any given vertex v by explicit enumeration of the configurations in a small neighborhood around v. And this would enable us to approximate the partition function by pinning that spin value and recursing, as we saw earlier in the course. In the next few lectures, we'll illustrate this with the canonical example of the *hard core model*, or independent set problem.

Given a graph $G = (V, E)$, let

$$Z_G(\lambda) = \sum_{\text{Ind Sets } I} \lambda^{|I|} = \sum_{k \geq 0} a_k \lambda^k,$$

where a_k is the number of independent sets (of vertices) of size k and $\lambda > 0$ is a parameter known as the *fugacity*. Recall also that $Z_G(\lambda)$ corresponds to a Gibbs distribution, in which the probability of an independent set of size k is $\frac{\lambda^k}{Z_G(\lambda)}$.

Let's start by giving a simple reduction to show that tthis problem is hard to approximate on general graphs G at any fixed value $\lambda > 0$. We reduce the decision problem IndSet to approximating $Z_G(\lambda)$. Given a graph G and integer k_0, IndSet asks whether there is an independent set of size at least k_0 in G.

Given such a graph $G = (V, E)$, we replace each vertex v with a "cloud" C_v of independent vertices of size r; then, for all adjacent pairs $\{v, v'\} \in E$, we connect the corresponding clouds C_v and $C_{v'}$ with a complete bipartite graph $K_{r,r}$. Let G_r denote this blown-up version of G. Note that we can view each independent set I' of G_r as a *witness* for a unique independent set I in G, where $I = \{v \in V : C_v \cap I' \neq \emptyset\}$. Then if I has size k, the total weight of all the witnesses for I is $((1+\lambda)^r - 1)^k$. (**Exercise:** check this.) As a result, we have

$$Z_{G_r}(\lambda) = \sum_{k \geq 0} a_k \left((1+\lambda)^r - 1\right)^k.$$

Now we pick r large enough so that

$$((1+\lambda)^r - 1)^{k_0} \gg 2^n \left((1+\lambda)^r - 1\right)^{k_0-1}. \tag{18.4}$$

This ensures that the weight of a single independent set of size k_0 in G dominates the aggregated weight of all independent sets of size strictly less than k_0. Now for any fixed $\lambda > 0$, we can ensure (18.4) by setting $r = cn$ for a constant c (depending on λ), where $n = |V|$ is the number of vertices in G. Hence the reduction to G_r can be computed in polynomial time.

Now we can decide if G contains an independent set of size at least k_0 by computing an approximation to $Z_{G_r}(\lambda)$ and comparing the result with the value $((1+\lambda)^r - 1)^{k_0}$.

We have proved:

Proposition 18.4. *If there exists a FPRAS (resp., and FPTAS) for $Z_G(\lambda)$ for some fixed $\lambda > 0$, then there is a randomized (resp., deterministic) polynomial-time algorithm for IndSet, and hence* RP = NP *(resp.,* P = NP*).*

The hard-core model is therefore uninteresting for general graphs. However, if we restrict attention to graphs of bounded degree (a natural restriction in spin systems, where each spin usually has just a bounded number of neighbors), then a beautiful picture emerges, as embodied in the following theorem.

Theorem 18.5. *For graphs G of max degree $\Delta \geq 3$, there exists a threshold $\lambda_c(\Delta) = \frac{(\Delta-1)^{(\Delta-1)}}{(\Delta-2)^{\Delta}}$ such that*

1. [Weitz [Wei06]] *If $\lambda < \lambda_c(\Delta)$, there exists a FPTAS for approximating $Z_G(\lambda)$.*
2. [Sly [Sly10]] *If $\lambda > \lambda_c(\Delta)$, there is no FPRAS for $Z_G(\lambda)$, unless RP = NP.*

Note that it is natural to expect the problem to get harder as λ gets larger, because typical independent sets have to pack in more vertices, meaning that correlations between distant vertices may exist. Hence we might expect the correlation decay idea sketched earlier to fail for large λ.

What is remarkable about the above theorem is that there is a threshold value for λ so that, when λ exceeds that threshold, not only does the correlation decay method fail, but *no* polynomial time (randomized) algorithm works (unless RP = NP). And, moreover, all the way up to the threshold correlation decay *does* work.

This threshold $\lambda_c(\Delta)$ turns out to be the so-called "uniqueness threshold" for the infinite Δ-regular tree. This means that, in a certain precise sense, trees are the worst case graphs for this problem among all graphs of maximum degree Δ.

The "uniqueness threshold" is the value of λ below which the associated Gibbs measure on the infinite tree is unique. To make this more precise, we introduce the following definition:

Definition 18.6. *Let $S \subseteq V$ be any subset of vertices in the infinite Δ-regular tree, and τ be an arbitrary configuration of "pinned spins" on S (i.e., vertices set to be either in the independent set or not, also called a "boundary condition"). Let $p_r^\tau := \mathsf{P}[\text{the root is occupied in the independent set}]$, under the associated Gibbs measure on the infinite tree.*

Weak spatial mixing (WSM) *holds for the tree if there exists a constant $c > 0$ such that*

$$|p_r^\tau - p_r^{\tau'}| \leq \exp(-c \cdot \mathrm{dist}(r, S))$$

for any two configurations τ, τ' on S.

In other words, WSM says that the effect of any set of fixed spins on the spin at the root (i.e., the probability that the root is in the independent set) decays exponentially to zero with the distance of the fixed spins from the root. How do we construct a Gibbs measure on the infinite tree? Slightly informally, we consider an increasing sequence $T_1, T_2, \ldots, T_\ell, \ldots$ of finite Δ-regular trees of increasing depths ℓ, that converges to the infinite tree. We place some arbitrary fixed boundary condition τ outside T_ℓ, and look at the sequence of Gibbs distributions on the finite trees T_ℓ. Then WSM ensures that the resulting infinite-volume Gibbs measure, constructed as the limit of the finite Gibbs distributions on the trees T_ℓ, is independent of the boundary condition, and thus unique. (This can all be made precise using appropriate topological notions that we won't get into here.)

Conversely, in the absence of WSM, the boundary condition may have a non-zero effect on the spin at the root even as the depth $\ell \to \infty$. In this case we get multiple Gibbs measures, depending on the boundary condition. Specifically, for the hard-core model, we will get precisely two *extremal* such Gibbs measures, one corresponding to all leaves of the tree at even depths ℓ being occupied, and the other corresponding to all leaves at odd depths being occupied. (Of course, there will also be infinitely many other measures obtained as convex combinations of these two extremal ones.)

This sharp change in the behavior of the Gibbs measure at $\lambda = \lambda_c(\Delta)$ is an example of a *phase transition.* (For $\lambda > \lambda_c(\Delta)$, two phases (the "odd phase" and the "even phase") emerge in the model.)

Theorem 18.5 can therefore be seen as a *computational* manifestation of this spatial phase transition. In the next couple of lectures we'll show how to prove both parts of the theorem.

References

[ALOV18] N. Anari, K. Liu, S. Oveis Gharan, and C. Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 35–46, 2018.

[BH19] P. Bränden and J. Huh. Lorentzian polynomials. *arXiv abs/1902.03719*, 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.

[Wei06] D. Weitz. Counting independent sets up to the tree threshold. *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 140–149, 2006.