CS294 Partition Functions **Fall 2020**

Lecture 17: October 27

Instructor: Alistair Sinclair *Scribes: Steven Cao and Sri Krishna Vadlamani*

17.1 Counting the bases of a matroid

Our goal in this lecture is to present a Markov chain Monte Carlo (MCMC) algorithm to approximately count the number of bases of an arbitrary matroid. Specifically, we have the following problem:

Input: A matroid $\mathcal{M} = (E, \mathcal{I})$, where the set E is called the *ground set*, and $\mathcal{I}$ is a collection of subsets of E, called *independent sets* of E. We represent $\mathcal{M}$ by a membership oracle, i.e., a black box that, given any set $J \subseteq E$, specifies whether or not J is an independent set.

Output: The number of bases (i.e., maximal independent sets) in $\mathcal{M}$.

Observation 17.1. *A fully polynomial randomized approximation scheme (fpras) for counting the bases of any matroid immediately implies an fpras for counting the independent sets of any given size k in any matroid. One simply runs the fpras on the truncated matroid $\mathcal{M}' = (E, \mathcal{I} \setminus \{I : I \in \mathcal{I},\ |I| > k\})$ that is obtained from the original $\mathcal{M}$ via the elimination of all the independent sets of size greater than k from $\mathcal{I}$.*

17.2 MCMC algorithm setup

To perform approximate counting, we will use the standard reduction from counting to sampling that we saw earlier in the course. We will perform approximate sampling via MCMC, and our goal will be to efficiently sample from a probability distribution over the set of bases that is close in total variation distance to the uniform distribution.

17.2.1 Basis exchange process

Our Markov chain for sampling bases, which is known as the "basis exchange" process, is defined on the state space $\Omega = \{B : B \text{ is a basis of } \mathcal{M}\}$ as follows:

- Let the current state be the basis B. Pick an element $e \in B$ uniformly at random to remove.
- Let $F = \{f \in E : B \setminus \{e\} \cup \{f\} \in \mathcal{I}\}$ denote the set of elements that we can add to $B \setminus \{e\}$ while remaining independent.
- Pick f uniformly at random from F and move to the basis $B \setminus \{e\} \cup \{f\}$.

Observation 17.2. *The basis exchange process is irreducible.*

Proof. Let B, B' be distinct bases. This means they have the same size, $|B| = |B'| = r$, where r is the rank of $\mathcal{M}$. Let the symmetric difference of B and B' be decomposed as $B \oplus B' = \{e_1, \ldots, e_k\} \cup \{e'_1, \ldots, e'_k\}$ where $e_i \in B$ and $e'_i \in B'$. Remove e_1 from B. Then, by the augmentation axiom of matroids, there exists an element $e'_i \in B'$ such that $B'' := B \setminus \{e_1\} \cup \{e'_i\}$ is independent (and hence a basis). Now B'', B' are a pair of bases that differ in two fewer elements than B, B', so we can iterate this procedure until we reach B'. □

The basis exchange process is clearly symmetric, so the stationary distribution is uniform. It is also aperiodic (there is always a self-loop as we could choose $f = e$ in the final step of the exchange), and therefore ergodic. Our goal will be to show that the mixing time is $O(r(\log r + \log\log n)) = O(n \log n)$, where $n = |E|$, following the analysis of [CGM19]. The framework below is due to [ALOV18], who used a different analysis based on the spectral gap to obtain a mixing time of $O(r^2 \log n) = O(n^2 \log n)$.

17.2.2 Weight function

While we are interested in the mixing time of our Markov chain over bases, we will actually define Markov chains over independent sets of each size $2 \le k \le r$, where r is the rank. We will then show that all of these Markov chains are rapidly mixing by induction on k. This sort of inductive decomposition is typical of log-Sobolev-based mixing time bounds (though in other settings, such as spin systems on lattices $\mathbb{Z}^d$, the induction is often spatial).

To this end, we now define a weight function $w(I)$ on *all* independent sets as follows:

$$w(I) = \begin{cases} 1 & |I| = r \text{ (basis)}, \\ \sum_{I' \in \mathcal{I}, I' = I + e} w(I') & \text{otherwise.} \end{cases} \tag{17.1}$$

In other words, the weight of a size k independent set I is the sum of the weights of all size $k+1$ independent sets that contain I. See Figure 17.1 for a simple example.

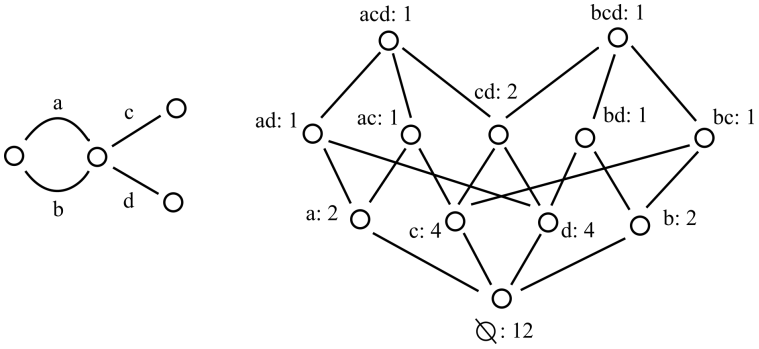


Figure 17.1: The weight function for the graphic matroid of a simple multigraph with edge set $E = \{a, b, c, d\}$ (shown on the left). The diagram on the right shows all independent sets organized by size, and their corresponding weights. The rank is $r = 3$.

Note: We can write down an explicit expression for $w(I)$ in terms of the number of bases B that contain I (that is, $B \supseteq I$) by noting that there are $(r - |I|)!$ paths from any such basis set B to I (each path corresponding to an order of removal of elements from B). Therefore, $w(I) = (r - |I|)! \times \#\{B : B \supseteq I\}$.

Let us denote the set of all independent sets of size k by $\mathcal{M}(k)$ (so $\mathcal{M}(r)$ are the bases). The weight function

above induces a probability distribution $\pi_k(I)$ on $\mathcal{M}(k)$ as follows:

$$\pi_k(I) = \frac{w(I)}{Z_k}, \text{ where } Z_k = \sum_{I \in \mathcal{M}(k)} w(I). \tag{17.2}$$

Recalling that the set of all bases of the matroid $\mathcal{M}$ was denoted by $\Omega = \mathcal{M}(r)$, we have $Z_r = |\Omega|$. Further, it is easy to check that $k!Z_k = Z_0 = w(\emptyset)$. (**Exercise!**)

17.2.3 Contraction

We next introduce the idea of a "contraction", which will be made use of in the proofs of the following sections.

Definition 17.3. *For any independent set I, define the contraction $\mathcal{M}_I$ as $\mathcal{M}_I := (E \setminus I, \mathcal{I}_I)$ where $\mathcal{I}_I := \{J : J \subseteq E \setminus I,\ J \cup I \in \mathcal{I}\}$.*

In other words, given an independent set I, we first remove it from the ground set E. Then, to get the independent sets of $E \setminus I$, we consider all independent sets K of E such that $I \subseteq K$, and we say that $J = K \setminus I$ is an independent set of $E \setminus I$. It is a standard fact (**exercise!**) that the contraction $\mathcal{M}_I = (E \setminus I, \mathcal{I}_I)$ is also a matroid.

We next define probability distributions $\pi_{I,k}$ on $\mathcal{M}(|I| + k)$ as follows:

$$\pi_{I,k}(J) = \begin{cases} \frac{k!w(J)}{w(I)} & J \supseteq I, \\ 0 & \text{otherwise,} \end{cases} \tag{17.3}$$

Equivalently, $\pi_{I,k}(J)$ is the probability of J conditioned on J containing I.

17.2.4 'Down-up' walk $P_k^\vee$ on $\mathcal{M}(k)$

We will now generalize the basis exchange Markov chain to produce a Markov chain on independent sets of size k, for each k from 2 to r. The Markov chain on $\mathcal{M}(k)$ will have the stationary distribution π_k defined above. This Markov chain will be called the 'down-up' walk over $\mathcal{M}(k)$, and is defined as follows:

- Let the current state be the independent set $I \in \mathcal{M}(k)$. Remove an element $e \in I$ uniformly at random.
- Let $S = \{J \in \mathcal{M}(k) : J = I \setminus \{e\} \cup \{e'\} \text{ for some } e'\}$. Move to $J \in S$ with probability proportional to $w(J)$.

For $k = r$, the down-up walk $P_r^\vee$ is simply the basis exchange process. (**Exercise!**)

The transition probabilities for $I, J \in \mathcal{M}(k)$ can be written down explicitly as:

$$P_k^\vee(I, J) = \begin{cases} \frac{w(J)}{kw(I \cap J)} & \text{if } |I \oplus J| = 2; \\ \sum_{e \in I} \frac{w(I)}{kw(I \setminus \{e\})} & \text{if } I = J; \\ 0 & \text{otherwise.} \end{cases} \tag{17.4}$$

17.3 Main theorem: bounding the log-Sobolev constant

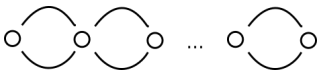
Our mixing time bound will follow immediately from the following result:

Theorem 17.4. *For $2 \le k \le r$, the modified log-Sobolev constant ρ_k of $P_k^\vee$ satisfies $\rho_k \ge \frac{1}{k}$.*

Corollary 17.5. *The mixing time of the basis exchange walk $P_r^\vee$ is $O(r(\log r + \log\log n)) = O(n \log n)$.*

Proof. We use our general bound on mixing times in terms of the log-Sobolev constant, $\tau_{\text{mix}} = O(\frac{1}{\rho_r} \log\log \pi_{\min}^{-1})$, together with the fact that $\rho_r \ge \frac{1}{r}$ from the above theorem and $\pi_{\min} = \frac{1}{|\Omega|} \ge \frac{1}{\binom{n}{r}} \ge \frac{1}{n^r}$. □

This mixing time can be seen to be optimal as a function of n with the help of a simple example. Consider the graphic matroid corresponding to the following multigraph with $\frac{n}{2}$ pairs of edges and $\frac{n}{2} + 1$ vertices.



(Equivalently, this is a partition matroid with $\frac{n}{2}$ blocks, each of capacity 1.) The bases correspond to choosing one edge from each pair. The down-up walk is then easily seen to be equivalent to random walk on the hypercube $\{0,1\}^{n/2}$, which as we have seen has mixing time $\Theta(n \log n)$.

Note: Very recently, Anari *et al.* [ALOV20] sharpened this mixing time bound to $O(r \log r)$, making it independent of n, the size of the ground set.

17.3.1 Setup

It is useful to decompose the transitions of the down-up walk into separate down and up steps. Both of these can be cast in terms of the following matrix.

Definition 17.6. *For $2 \le k \le r$, the matrix A_{k-1} with rows indexed by $\mathcal{M}(k-1)$ and columns indexed by $\mathcal{M}(k)$ is defined by*

$$A_{k-1}(I,J) = \begin{cases} 1 & \text{if } J = I \cup \{e\} \text{ for some } e; \\ 0 & \text{otherwise.} \end{cases} \tag{17.5}$$

Equipped with this definition, we can decompose the transition matrix $P_k^\vee$ of the down-up process at level k as $P_k^\vee = P_k^\downarrow P_{k-1}^\uparrow$, where the "down" step is represented by the transition matrix $P_k^\downarrow = \frac{1}{k} A_{k-1}^T$, and the "up" step by the transition matrix $P_{k-1}^\uparrow = A_{k-1}$ with the (I,J) entry set to $\frac{w(J)}{w(I)}$ instead of 1.

Note that these matrices satisfy the following "detailed balance" condition:

Claim 17.7. *For all I, J, $\pi_k(I) P_k^\downarrow(I,J) = \pi_{k-1}(J) P_{k-1}^\uparrow(J,I)$* (**Exercise!**)

Given this setup, we can then naturally extend a function $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\ge 0}$ over independent sets of size k to functions $f^{(i)}$ over independent sets of size $i < k$ as follows:

Definition 17.8. *For any function $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\ge 0}$, define $f^{(i)} : \mathcal{M}(i) \to \mathbb{R}_{\ge 0}$ by:*

$$f^{(i)} := \prod_{j=i}^{k-1} P_j^\uparrow f^{(k)} = P_i^\uparrow P_{i+1}^\uparrow \cdots P_{k-1}^\uparrow f^{(k)}.$$

In other words, to apply $f^{(i)}$ to an independent set $I \in \mathcal{M}(i)$, we take $k-i$ random "up" steps to reach $J \in \mathcal{M}(k)$ band then apply the original function $f^{(k)}$ to J. Another way to interpret $f^{(i)}$ is as an expectation over the conditional measure $\pi_{J,k-i}$, namely: $f^{(i)}(J) = \mathrm{E}_{\pi_{J,k-i}} f^{(k)}$ **[exercise!]**.

The following simple properties of these functions will be useful; the proofs are left as an **exercise!**

Claim 17.9. *1. If $f^{(k)}$ is normalized so that $E_{\pi_k} f^{(k)} = 1$, then $E_{\pi_i} f^{(i)} = 1$ for all $i < k$.*

2. $\pi_{k-1}(P_{k-1}^{\uparrow} f^{(k)}) = (\pi_k f^{(k)}) P_k^{\downarrow}$.

17.3.2 Main lemma

We now state a key lemma that will lead fairly directly to a proof of our main goal, Theorem 17.4. This lemma essentially says that, under one "down" step of the process, entropy is contracted by a factor of at least $1 - \frac{1}{k}$.

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

$$\mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \leq \left(1 - \frac{1}{k}\right) \mathrm{Ent}_{\pi_k}(f^{(k)}). \tag{17.6}$$

We'll see how to prove this lemma by induction on k in the next subsection. First, we show how to use the lemma to prove Theorem 17.4.

Proof of Theorem 17.4. Recall that, to prove the theorem, we need to show that

$$\rho_k = \inf_{\substack{f^{(k)} \geq 0, \\ \mathrm{Ent}_{\pi_k} f^{(k)} \neq 0}} \frac{\mathcal{E}_{P_k^{\vee}}(f^{(k)}, \log f^{(k)})}{\mathrm{Ent}_{\pi_k} f^{(k)}} \geq \frac{1}{k}. \tag{17.7}$$

Let $f^{(k)} : \mathcal{M}(k) \to \mathbb{R}_{\geq 0}$ be any function, and assume without loss of generality that $E_{\pi_k} f^{(k)} = 1$ (since rescaling $f^{(k)}$ does not affect the ratio in (17.7)). Our first step will be to apply the definitions of Ent_{π_k} and $f^{(k-1)}$ to show that

$$\mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \geq \mathrm{Ent}_{\pi_k}(f^{(k)}) - \mathcal{E}_{P_k^{\vee}}(f^{(k)}, \log f^{(k)}), \tag{17.8}$$

which relates the entropies at the two adjacent levels via the Dirichlet form. From (17.8), we can directly apply Lemma 17.10 to bound the log-Sobolev constant as follows:

$$\begin{aligned}
\mathcal{E}_{P_k^{\vee}}\left(f^{(k)}, \log f^{(k)}\right) &\geq \mathrm{Ent}_{\pi_k}(f^{(k)}) - \mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \\
&\geq \left(1 - \frac{k-1}{k}\right) \mathrm{Ent}_{\pi_k}(f^{(k)}) \qquad \text{(by Lemma 17.10)} \\
&= \frac{1}{k} \mathrm{Ent}_{\pi_k}(f^{(k)}).
\end{aligned}$$

Thus, it suffices to prove (17.8). We'll do this by applying the definitions and using Jensen's inequality:

$$\begin{aligned}
\mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) &= \sum_{I\in\mathcal{M}_{k-1}} \pi_{k-1}(I) f^{(k-1)}(I) \log f^{(k-1)}(I) \\
&= \sum_{I\in\mathcal{M}_{k-1}} \pi_{k-1}(I) [P^{\uparrow}_{k-1} f^{(k)}](I) \log f^{(k-1)}(I) \\
&= \sum_{\substack{I\in\mathcal{M}_{k-1},\\ J\in\mathcal{M}_k}} \pi_{k-1}(I) P^{\uparrow}_{k-1}(I,J) f^{(k)}(J) \log \left(\sum_{L\in\mathcal{M}_k} P^{\uparrow}_{k-1}(I,L) f^{(k)}(L) \right) \\
\text{(Claim 17.7)} &= \sum_{\substack{I\in\mathcal{M}_{k-1},\\ J\in\mathcal{M}_k}} \pi_k(J) f^{(k)}(J) P^{\downarrow}_k(J,I) \log \left(\sum_{L\in\mathcal{M}_k} P^{\uparrow}_{k-1}(I,L) f^{(k)}(L) \right) \\
\text{(Jensen's inequality)} &\geq \sum_{\substack{I\in\mathcal{M}_{k-1},\\ J,L\in\mathcal{M}_k}} \pi_k(J) f^{(k)}(J) P^{\downarrow}_k(J,I) P^{\uparrow}_{k-1}(I,L) \log f^{(k)}(L) \\
&= \sum_{J,L\in\mathcal{M}_k} \pi_k(J) f^{(k)}(J) P^{\vee}_k(J,L) \log f^{(k)}(L) \\
&= \mathrm{Ent}_{\pi_k}(f^{(k)}) - \mathcal{E}_{P^{\vee}_k}\left(f^{(k)}, \log f^{(k)}\right),
\end{aligned}$$

where in the last line we used the definition $\mathcal{E}_{P^{\vee}_k}(f^{(k)}, \log f^{(k)}) = \langle f^{(k)}, L \log f^{(k)} \rangle_{\pi_k} = \mathrm{Ent}_{\pi_k}(f^{(k)}) - \langle f^{(k)}, P \log f^{(k)} \rangle_{\pi_k}$. This completes the proof of (17.7) and hence of the theorem. □

17.3.3 Inductive proof of the main lemma

Now we turn to proving Lemma 17.10 via induction on k. We will show the inductive step and leave the base case $k = 2$ to the next lecture. (As we saw in the hypercube example in the last lecture, in log-Sobolev analyses the base case is often the more challenging part, and will be again here.)

Proof of inductive step in Lemma 17.10. We wish to show that the inequality

$$\mathrm{Ent}_{\pi_{k-1}}(f^{(k-1)}) \leq \left(1 - \frac{1}{k}\right) \mathrm{Ent}_{\pi_k}(f^{(k)}), \tag{17.9}$$

implies the analogous inequality with $k-1$ replaced by k, i.e.,

$$\mathrm{Ent}_{\pi_k}(f^{(k)}) \leq \left(1 - \frac{1}{k+1}\right) \mathrm{Ent}_{\pi_{k+1}}(f^{(k+1)}). \tag{17.10}$$

The key to our induction will be the following decomposition of π_k:

$$\pi_k = \sum_{e\in E} \pi_1(e) \pi_{e,k-1}, \tag{17.11}$$

where we recall that $\pi_{e,k-1}$ is the distribution over $\mathcal{M}_k$ conditioned on the set containing e. **[Exercise:** Check this!] This allows us to use the following standard factorization of the entropy:

$$\mathrm{Ent}_{\pi_k} f^{(k)} = \mathrm{Ent}_{\pi_1} f^{(1)} + \sum_{e\in E} \pi_1(e) \mathrm{Ent}_{\pi_{e,k-1}} f^{(k)}. \tag{17.12}$$

[**Exercise:** Check this factorization, by using the decomposition (17.11); you will also need the fact that $\mathrm{E}_{\pi_{e,k-1}}(f^{(k)}) = f^{(1)}(e)$, which can be verified by induction on k.]

Now by the induction hypothesis applied to the functions $f^{(k)}, f^{(k-1)}, \ldots, f^{(2)}$, we have

$$\mathrm{Ent}_{\pi_k} f^{(k)} \geq \frac{k}{k-1}\frac{k-1}{k-2}\cdots\frac{2}{1}\,\mathrm{Ent}_{\pi_1} f^{(1)} = k\mathrm{Ent}_{\pi_1} f^{(1)}. \tag{17.13}$$

Applying Equations (17.12) and (17.13) together gives us

$$\sum_{e\in E}\pi_1(e)\mathrm{Ent}_{\pi_{e,k-1}} f^{(k)} = \mathrm{Ent}_{\pi_k} f^{(k)} - \mathrm{Ent}_{\pi_1} f^{(1)} \geq (k-1)\mathrm{Ent}_{\pi_1} f^{(1)}. \tag{17.14}$$

Now, noting that the conditional distribution $\pi_{e,k}$ corresponds to a distribution on the smaller contracted matroid $\mathcal{M}_e$, we can use induction again to conclude:

$$\begin{aligned}
\mathrm{Ent}_{\pi_{k+1}} f^{(k+1)} &= \mathrm{Ent}_{\pi_1} f^{(1)} + \sum_{e\in E}\pi_1(e)\mathrm{Ent}_{\pi_{e,k}} f^{(k+1)} \\
\text{(induction hypothesis on } \mathcal{M}_e) &\geq \mathrm{Ent}_{\pi_1} f^{(1)} + \sum_{e\in E}\pi_1(e)\frac{k}{k-1}\mathrm{Ent}_{\pi_{e,k-1}} f^{(k)} \\
&= \mathrm{Ent}_{\pi_1} f^{(1)} + \frac{1}{k(k-1)}\sum_{e\in E}\pi_1(e)\mathrm{Ent}_{\pi_{e,k-1}} f^{(k)} + \frac{k+1}{k}\sum_{e\in E}\pi_1(e)\mathrm{Ent}_{\pi_{e,k-1}} f^{(k)} \\
\text{(by (17.14))} &\geq \mathrm{Ent}_{\pi_1} f^{(1)} + \frac{k-1}{k(k-1)}\mathrm{Ent}_{\pi_1} f^{(1)} + \frac{k+1}{k}\sum_{e\in E}\pi_1(e)\mathrm{Ent}_{\pi_{e,k-1}} f^{(k)} \\
&= \frac{k+1}{k}\left(\mathrm{Ent}_{\pi_1} f^{(1)} + \sum_{e\in E}\pi_1(e)\mathrm{Ent}_{\pi_{e,k-1}} f^{(k)}\right) \\
&= \frac{k+1}{k}\mathrm{Ent}_{\pi_k} f^{(k)},
\end{aligned}$$

as desired. □

To complete the proof of Lemma 17.10, it remains to verify the base case $k = 2$, which we'll do in the next lecture.

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.

[ALOV20] N. Anari, K. Liu, S. Oveis Gharan, and C. Vinzant. Log-concave polynomials IV: Approximate exchange, tight mixing times, and faster sampling of spanning trees. *arXiv, abs/2004.07220*, 2020.

[CGM19] M. Cryan, H. Guo, and G. Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1358–1370, 2019.