Lecture 5: Approximate Sampling of Spanning Trees via Matroid Basis Exchange

Warning: These lecture notes have not been proofread by a second human being, nor passed the scrutiny one expects from formal publications. If you find errors, please do not hesitate to contact the instructor.

Throughout this lecture, we'll take $G = (V, E)$ to be a connected graph, and let's adopt the notation $|V| = n$, $|E| = m$, and $\mathcal{T}_G = \{T : T \text{ is a spanning tree of} G\}$. We let $\mu : \{0,1\}^E \longrightarrow \mathbb{R}_+$ be the uniform distribution over spanning trees of G.

In the last lecture, we saw Wilson's Algorithm for sampling a UST (uniformly random spanning tree) $T \sim \mu$. Though we did not prove it, the expected running time is $O(n^3)$ and in some graphs can be as large as $\Omega(n^3)$.

Today, we will see an algorithm that *approximately samples* from a distribution ν which is close to μ in *total variation distance*, in the sense that for any $\varepsilon > 0$,

$$\mathsf{d}_{\mathsf{TV}}(\mu, \nu) = \frac{1}{2} \sum_{T \in \mathcal{T}_G} |\mu(T) - \nu(T)| \leqslant \varepsilon,$$

in time $O(m \log m \log \frac{m}{\varepsilon})$. Bibliographic remarks will be mostly reserved for the end of the notes.

1 The Matroid Basis Exchange Algorithm

In fact, this algorithm will sample a uniformly random basis of any *matroid*:

Definition 1.1 (Matroid)**.** A *matroid* $\mathcal{M} = (E, \mathcal{I})$ consists of a *ground set* E and a collection $\mathcal{I}$ of *independent* subsets of E which are:

1. *Downward-closed*: if $S \subset T$ and $T \in \mathcal{I}$ then $S \in \mathcal{I}$.
2. *Exchange*: if $S, T \in \mathcal{I}$ and $|T| > |S|$, then there exists $i \in T \setminus S$ such that $S \cup \{i\} \in \mathcal{I}$.

The *rank* r of a matroid is the size of the maximum set in $\mathcal{I}$, and any set $S \in \mathcal{I}$ with $|S| = r$ is called a *basis* of $\mathcal{M}$.

As the terminology suggests, matroids generalize linear independence in vector spaces:

Example 1.2. The set $\mathcal{M} = (\mathbb{R}^d, \mathcal{I})$ forms a matroid for $\mathcal{I}$ the set of linearly independent subsets of vectors in $\mathbb{R}^d$.

Of relevance to us is the fact that the set of spanning forests of a graph is a matroid, the bases of which are precisely the graph's spanning trees.

Example 1.3 (The Graphic Matroid)**.** If $G = (V, E)$ is a graph, then $\mathcal{M} = (U, \mathcal{F}_G)$ forms a matroid for $\mathcal{F}_G$ the set of spanning forests of G (recall a spanning forest of G is an edge-induced subgraph of G containing no cycles).

Sampling a random basis. A classical algorithmic task is to sample a random basis of a matroid $\mathcal{M}$. Suppose we have access to a matroid $\mathcal{M} = (E, \mathcal{I})$ via a *independence oracle* $\mathcal{O}$, which given $S \subseteq E$, returns 'yes' if $S \in \mathcal{I}$ and returns 'no' otherwise.a Consider the following Markov-Chain based algorithm:

Algorithm 1.4 (Basis Exchange Algorithm)**.** Input: a basis S_0 of $\mathcal{M} = (E, \mathcal{I})$, and an integer $T \in \mathbb{Z}_+$.

1. For $t = 0, 2, \ldots, T-1$:
   (a) Choose a uniformly random element $e \in S_t$.
   (b) Choose S_{t+1} to be a uniformly random basis containing $S_t \setminus \{e\}$.
2. Return S_T.

For example, in the case of the graphical matroid, this algorithm starts at an arbitrary spanning tree S_0 of G, then at each step t, removes an edge e from S_t, then chooses S_{t+1} uniformly from all trees containing $S_t \setminus \{e\}$. Notice we can implement step (b) using m calls to $\mathcal{O}$, since we can just check for each $f \in E$ if $S_t \setminus \{e\} \cup \{f\} \in \mathcal{I}$.

Generalizing our notation from above, consider the distribution ν_T of S_T, and call μ to be the uniform distribution over the bases of $\mathcal{M}$. If $\mathsf{d}_{\mathsf{TV}}(\nu_T, \mu) \leqslant \varepsilon$, we will say that the Markov Chain is "ε-mixed." In the 80's, Mihail and Vazirani [MV89] conjectured that running the basis exchange algorithm for $T = \text{poly}(m, \frac{1}{\varepsilon})$ steps guarantees that $d_{TV}(\nu_T - \mu) \leqslant \varepsilon$. This was shown for a special class of matroids known as *balanced matroids* in the early 90's [FM92], which are matroids which exhibit the following *negative correlation* property: for any $e, f \in E$ and a randomly chosen basis X of $\mathcal{M}$,

$$\mathbf{Pr}[e \in X \mid f \in X] \leqslant \mathbf{Pr}[e \in X],$$

and furthermore this must hold for any *minor* of $\mathcal{M}$, which is a matroid formed by restricting $\mathcal{I}$ to sets containing S and excluding T for any disjoint $S, T \subset E$.

Recently, there has been a lot of activity on this front. This theorem was proven in 2019 (and a more general version was proven just last year).

Theorem 1.5. *There exists a constant C such that if $\mathcal{M} = (E, \mathcal{I})$ is a matroid with rank r and $|E| = m$, taking $T > C \cdot r \log \frac{r}{\varepsilon}$ ensures that $\mathsf{d}_{\mathsf{TV}}(\mu, \nu_T) \leqslant \varepsilon$.*

Today, we will prove a weaker theorem using similar (but slightly simpler) techniques:

Theorem 1.6. *There exists a constant C such that if $\mathcal{M} = (E, \mathcal{I})$ is a matroid with rank r and $|E| = m$, taking $T > C \cdot r^2 \log \frac{m}{\varepsilon}$ ensures that $\mathsf{d}_{\mathsf{TV}}(\mu, \nu_T) \leqslant \varepsilon$.*

In fact, a generalization of both theorem is proven for any *log-concave* measure μ—more on this below.

Uniform Spanning Trees. To get from Theorem 1.5 to an algorithm for approximate sampling of a UST in time $O(m \log^2 \frac{m}{\varepsilon})$ is not completely straightforward, since even though the rank of the graphic matroid is $n-1$ (so that running the algorithm for $T = O(n \log \frac{n}{\varepsilon})$ suffices), step (b) of the algorithm may take time m. To get a time $O(m \log m \log \frac{m}{\varepsilon})$ algorithm, one must instead consider the *dual* of the graphic matroid, which has rank $\leqslant m$, and in which step (b) may be implemented in time $O(\log m)$ using a special data structure called a *link-cut tree.*

2 Mixing and the Spectral Gap of the Up-Down Operator

Let $\mathcal{B}$ be the set of bases of $\mathcal{I}$, and for convenience let $N = |\mathcal{B}|$. Notice that $N \leqslant \binom{m}{r} \leqslant m^r$. Let $\{P_{S,T}\}_{S,T\in\mathcal{I}}$ be the transition operator of the basis exchange algorithm. For $S \in \mathcal{I}$ with $|S| = r-1$, let $w(S) = |\{T \in \mathcal{I} \mid S \subset T\}|$.

It is helpful to think of P as the composition of two operators: the "down" operator $P_\downarrow$, which maps sets of size r in $\mathcal{I}$ to sets of size $r-1$ in $\mathcal{I}$ by dropping a uniformly random element, and the "up" operator $P_{r-1\to r}$ which maps sets S of size $r-1$ in $\mathcal{I}$ to sets T of size r in $\mathcal{I}$ by choosing a uniformly random T containing S.

$$(P_\downarrow)_{S,T} = \frac{\mathbf{1}[T \subset S]}{r}$$
$$(P_\uparrow)_{S,T} = \frac{\mathbf{1}[S \subset T]}{w(S)}.$$

We have defined P so that $P_{S,T} = (P_\downarrow P_\uparrow)_{S,T}$. From this we can see that P has the same eigenvalues as $P_\uparrow P_\downarrow$, and also

$$P_{S,T} = \begin{cases} \frac{1}{r\cdot w(S\cap T)} & |S \triangle T| = 2 \\ \frac{1}{r}\sum_{e\in S}\frac{1}{w(S\setminus\{e\})} & S = T \\ 0 & \text{otherwise.} \end{cases}$$

Claim 2.1. P is reversible, irreducible, and aperiodic, and the uniform measure over bases, μ, is a stationary measure of P.

Proof. Since $\frac{1}{N}P_{S,T} = \frac{1}{N}P_{T,S}$, P is reversible with respect to $\mu = \frac{1}{N}\vec{1}$. To see that P is irreducible, consider two bases S, T, and note that the exchange property of matroids implies that if we remove an element $e \in S \setminus T$, there must exist an element $f \in T \setminus S$ such that $S_1 = S\setminus\{e\}\cup\{f\} \in \mathcal{B}$, and $P_{S,S_1} > 0$; this argument can be repeated until we reach T. To see that P is aperiodic, note that we can write $P = \delta \cdot \mathbb{1} + (1-\delta)P'$ for $\delta \in (0,1]$ and P' another Markov Chain. Since P is reversible with respect to μ, we also have that μ is stationary:

$$(\mu^\top P)_T = \sum_{S\in\mathcal{B}} \mu_S P_{S,T} = \mu_T \sum_{S\in\mathcal{B}} P_{T,S} = \frac{1}{N} = \mu_T,$$

concluding the proof. □

Since P is aperiodic and irreducible with stationary measure μ, we know that starting from any basis S_0, $\lim_{t\to\infty} 1_{S_0}^\top P^t = \mu$. Now, we'll make this more quantitative, using properties of the spectrum of P. Let $1 = \lambda_1 > \lambda_2 \geqslant \cdots \geqslant \lambda_N > -1$ be the eigenvalues of P (we have that $|\lambda_2|, |\lambda_N| < 1$ since P is irreducible and reversible and aperiodic). We have the following lemma:

Lemma 2.2. *Letting* $\lambda^* = \max(\lambda_2, |\lambda_N|)$, *for any* $T > \frac{1}{1-\lambda^*}(\log\frac{1}{\varepsilon} + \frac{1}{2}\log N)$, $\mathsf{d}_{\mathsf{TV}}(\nu_T, \mu) \leqslant \varepsilon$.

Proof. Since the unit left-eigenvectors $v_1, \ldots, v_N$ of P form an orthonormal basis for $\mathbb{R}^N$, we can write $1_{S_0} = \sum_{i=1}^N \langle v_i, 1_{S_0}\rangle \cdot v_i$.

$$1_{S_0}^\top P^t = \sum_{i=1}^N \langle v_i, 1_{S_0}\rangle \cdot v_i^\top P_t = \sum_{i=1}^N \langle v_i, 1_{S_0}\rangle \lambda_i^t \cdot v_i.$$

In particular, $v_1 = \frac{1}{\sqrt{N}}\vec{1} = \sqrt{N}\mu$, and also $\langle v_1, 1_{S_0}\rangle = \frac{1}{\sqrt{N}}$. Hence $1_{S_0}^\top P^t = \mu$. So we have

$$\mathsf{d}_{\mathsf{TV}}(\nu_t, \mu) = \|1_{S_0}^\top P^t - \mu\|_1 = \left\| \sum_{i=2}^{N} \langle v_i, 1_{S_0}\rangle \cdot \lambda_i^t \cdot v_i \right\|_1 \leqslant \sqrt{N} \max(\lambda_2, |\lambda_N|)^t.$$

The conclusion follows by noting that for $T = \frac{1}{1-\lambda^*}(\log\frac{1}{\varepsilon} + \frac{1}{2}\log N)$,

$$\mathsf{d}_{\mathsf{TV}}(\nu_T, \mu) \leqslant \sqrt{N}(1 - (1-\lambda^*))^T \leqslant \sqrt{N}\exp(T(1-\lambda^*)) \leqslant \varepsilon. \qquad \square$$

So to prove Theorem 1.6 it is sufficient to show that $\lambda^* \leqslant 1 - \frac{c}{r}$ for some constant c, since by Lemma 2.2 and the bound $N \leqslant m^r$ it is sufficient to take T to be

$$\frac{1}{1-\lambda^*}\left(\log\frac{1}{\varepsilon} + \frac{1}{2} r \log m\right) \leqslant \frac{1}{c} r^2 \log\frac{m}{\varepsilon}.$$

Remark 2.3. The proof of Theorem 1.5 differs from the proof of Theorem 1.6: rather than using a bound on the mixing time based on the spectrum of P as in Lemma 2.2, it uses a stronger mixing time bound implied by a *log-Sobolev* inequality for P. The method by which this log-Sobolev inequality is proved bears some resemblance to the bound on the spectrum of P; if you are interested in investigating further, this may be a good topic for the course final project.

3 Bounding the eigenvalues of the up-down walks

Our goal is now to show that

$$1 - \lambda^*(P) = \Omega(\frac{1}{r}).$$

It is not difficult to bound $|\lambda_N(P)|$. To see this, notice that P is a diagonal block of B^2 for the self-adjoint matrix (with respect to the inner product induced by its stationary measure π, $\langle f, g\rangle_\pi = \mathbf{E}_{i\sim\pi} f_i g_i$) $B = \begin{bmatrix} 0 & P_\uparrow \\ P_\downarrow & 0 \end{bmatrix}$. This implies that P is positive-semidefinite, since for any vector v, $\langle v, Pv\rangle_\pi = \langle v, B^2 v\rangle_\pi = \langle Bv, Bv\rangle_\pi \geqslant 0$, where the first equality follows since v can be padded with 0's on the entries outside of the support of $\mathcal{B}$, and the second inequality is because B is self adjoint. In conclusion, $\lambda_N(P) \geqslant 0$.

Therefore we turn our attention to upper bounding $\lambda_2(P)$. Here, we will make use of the *generating polynomial* of μ, which is the the m-variate polynomial

$$g_\mu(z) = \sum_{S\in\mathcal{B}} \mu(S) \prod_{i\in S} z_i = \sum_{S\in\mathcal{B}} \frac{1}{N} \prod_{i\in S} z_i.$$

That is, $g_\mu(z)$ is the homogeneous degree-r polynomial supported on monomials corresponding to bases, with coefficient of z^S equal to $\mu(S)$.

Definition 3.1 (Log-Concave Polynomial)**.** A multivariate polynomial $p \in \mathbb{R}[z_1, \ldots, z_m]$ with non-negative real coefficients is said to be *log-concave* on a subset $K \subseteq \mathbb{R}^m_{\geqslant 0}$ if $\log p$ is a concave function on K, that is, if $\nabla^2 \log p$ is negative-semidefinite. It is said to be *strongly log-concave* on K if for any sequence of integers $1 \leqslant i_1, \ldots, i_k \leqslant m$, $\frac{\partial}{\partial z_{i_1}} \cdots \frac{\partial}{\partial z_{i_k}} p$ is log-concave on K.

The following theorem, which we will not prove, will be crucial in our bound on the eigenvalues of P:

Theorem 3.2. *For any matroid, g_μ is strongly log-concave at $K = 1$.*

Let us see how it will help us bound the eigenvalues of P in the special case where $r = 2$, so p is a degree-2 polynomial. Recall that $P = P_\downarrow P_\uparrow$ has the same eigenvalues (up to zero eigenvalues) as $P_\uparrow P_\downarrow$ with respect to the inner product induced by w. Notice that at $z = 1$, strong log-concavity implies that $0 \succeq \nabla^2 \log p$ at $z = 1$, and we can directly compute the entries by differentiating and evaluating at 1:

$$(\nabla^2 \log p)_{ij}|_{z=1} = \left(\frac{p \cdot \partial_i \partial_j p - \partial_i p \partial_j p}{p^2}\right)\Big|_{z=1} = \frac{\mathbf{1}[\{i,j\} \in \mathcal{I}]}{N} - \frac{1}{N^2} w(i)w(j).$$

Now, notice that $(P_\uparrow P_\downarrow)_{ij} = \frac{1}{2}(\frac{1}{w(i)}\mathbf{1}[\{i,j\} \in \mathcal{B}\}]) + \frac{1}{2}\mathbf{1}[i = j]$, and $P_\uparrow P_\downarrow$ is reversible with respect to w. So we can re-write $\nabla^2 \log p|_{z=1}$ using $P_\uparrow P_\downarrow$, which in turn implies that

$$N \cdot ww^\top \succeq (D_w(P_\uparrow P_\downarrow - \frac{1}{2}\mathbb{1})),$$

and in particular, $D_w(P_\uparrow P_\downarrow - \frac{1}{2}\mathbb{1})$ can have only have one positive eigenvalue, in the direction w. All other eigenvalues of $P_\uparrow P_\downarrow$ (with respect to the w-inner-product) are at most $\frac{1}{2}$. This now implies that $P = P_\downarrow P_\uparrow$ has only one eigenvalue larger than $\frac{1}{2}$, and $\lambda_2(P) \leqslant \frac{1}{2}$.

Larger r via matroid contraction. To address the $r > 2$ case,[1] we use the fact that matroids are closed under *contraction*. Previously, we defined the up and down walks for the full matroid $\mathcal{M}$. But now, for any $A \subset E$, consider the *contracted* matroid $\mathcal{M}_A$ given by restricting $\mathcal{I}$ to the set of all subsets containing A. We may as well consider P_A and $P_\uparrow^A$ and $P_\downarrow^A$, the basis exchange matrix and up and down walks on $\mathcal{M}_A$. Notice also that $\partial_A g_\mu = g_{w(A)}$! In particular, $g_{w(A)}$ is log-concave by the strong log-concavity of g_μ. Using reasoning identical to the above, we can conclude that $P_\uparrow^A P_\downarrow^A$ has only one eigenvalue larger than $\frac{1}{2}$, in the direction of its stationary measure μ_A.

Again, we will bound the eigenvalues of $P_\uparrow P_\downarrow - \frac{1}{r}\mathbb{1}$. For any function f on independent sets of size $r-1$ which is orthogonal to 1, we'll relate $\langle f, (P_\uparrow P_\downarrow - \frac{1}{r}\mathbb{1})f\rangle$ to an average over $\langle f, (P_A - 1\mu_A^\top)f\rangle$, from which we will be done by the observation above. For each $A \subset E$, let $X_A(\ell) = \{B \in \mathcal{I} \mid A \subset B, |B| = \ell + 1\}$; borrowing terminology from simplicial complexes, we call $X_A(\ell)$ the "ℓ-link of A. Let $S \sim_w X(\ell)$ denote a sample from the measure over $\mathcal{I}$ supported on S of $|S| = \ell + 1$ where $\mathbf{Pr}[S] \propto w(S) = \sum_{T \supset S, |T| = |S|+1} w(T)$.

Now, for any eigenfunction f of $P_\uparrow P_\downarrow$ we have that

$$\begin{aligned}
\langle f, (P_\uparrow P_\downarrow - \frac{1}{r}\mathbb{1})f\rangle &= \mathbf{E}_{S \sim_w X(r-1)} \mathbf{E}_{e \sim X_S(0)} \mathbf{E}_{f \sim X_{S \cup \{e\}}(0), f \neq e} f_{S+e} f_{S+e-f} \\
&= \mathbf{E}_{A \sim_w X(r-2)} \mathbf{E}_{\{i,j\} \sim X_A(1)} f_{A+e} f_{A+f} \\
&= \mathbf{E}_{A \sim_w X(r-2)} \langle f, (P_A - \mathrm{diag}(P_A))f\rangle_w .
\end{aligned}$$

By the above, $P_A - \mathrm{diag}(P_A)$ has only one eigenvalue exceeding 0, in the direction of μ_A. One can then show (via the fact that the eigendecomposition of P_A is orthonormal, plus algebra) that for f orthogonal to 1, this direction does not contribute too much. A later version of these notes may flesh this out in more detail; for now, consult [ALGV19] or the lecture notes [Gha].

Bibliographic remarks. In preparing this lecture I heavily consulted the lecture notes [Gha]. Theorem 1.6 and Theorem 3.2 are originally from [ALGV19]; see also [BH18, ALGV18, AGV18] for results related to Theorem 3.2. Theorem 1.5 is from [CGM19], and a stronger statement (which holds for general log-concave measures) is proven in [ALG+21].

[1] Note: this portion of the note is less polished than the rest; it may be improved in a future edition.

Contact. Comments are welcome at tselil@stanford.edu.

References

[AGV18] Nima Anari, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials, entropy, and a deterministic approximation algorithm for counting bases of matroids. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 35–46. IEEE, 2018. 5

[ALG+21] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, Cynthia Vinzant, and Thuy-Duong Vuong. Log-concave polynomials IV: approximate exchange, tight mixing times, and near-optimal sampling of forests. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 408–420, 2021. 5

[ALGV18] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials III: Mason's ultra-log-concavity conjecture for independent sets of matroids. *arXiv preprint arXiv:1811.01600*, 2018. 5

[ALGV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019. 5

[BH18] Petter Brändén and June Huh. Hodge-Riemann relations for Potts model partition functions. *arXiv preprint arXiv:1811.01696*, 2018. 5

[CGM19] Mary Cryan, Heng Guo, and Giorgos Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1358–1370. IEEE, 2019. 5

[FM92] Tomás Feder and Milena Mihail. Balanced matroids. In *Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 26–38, 1992. 2

[Gha] Shayan Oveis Gharan. Polynomial paradigm in algorithms lecture 12-15: Simplicial complexes. homes.cs.washington.edu/~shayan/courses/polynomials/poly-lecture-12.pdf. 5

[MV89] Milena Mihail and Umesh Vazirani. On the expansion of 0-1 polytopes. *preprint*, 1989. 2