CS294 Partition Functions **Fall 2020**

Lecture 11: October 6

Instructor: Alistair Sinclair

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.*

11.1 Multicommodity flow

Now that we have the mixing time bound of Theorem 10.4 from the previous lecture, we need some tools to estimate the Poincaré constant α. This we do by establishing a connection with a natural multicommodity flow problem on the Markov chain, viewed as a directed graph.

As usual, we assume an ergodic Markov chain P with stationary distribution π. We say that an edge $e = (z, z')$ has *capacity* $C(e) := \pi(z)P(z, z')$. Note that $\pi(z)P(z, z')$ is also called the *ergodic flow* from z to z'. It represents the flow of probability mass along the edge when the Markov chain is at stationarity. For all pairs $(x, y) \in \Omega \times \Omega$ we have a *demand* $D(x, y) := \pi(x)\pi(y)$. A *flow* f is now any scheme that routes $D(x, y)$ units of flow from x to y simultaneously for all pairs x, y. The commodities are disjoint; we can think (say) of routing different types of fluid between any pair of nodes. More formally, a flow is a function $f\colon \mathcal{P} \to \mathbb{R}^+ \cup \{0\}$ where $\mathcal{P} = \bigcup_{xy} \mathcal{P}_{xy}$ and $\mathcal{P}_{xy}$ denotes the set of all simple paths from x to y s. t.

$$\sum_{p \in \mathcal{P}_{xy}} f(p) = D(x, y) \quad \forall (x, y) \in \Omega \times \Omega.$$

Let $f(e) = \sum_{p \ni e} f(p)$ be the total flow along e. The *cost* $\rho(f)$ of a flow f is given by $\rho(f) = \max_e f(e)/C(e)$. The length $\ell(f)$ of f is the length of a longest flow-carrying path, i.e., $\ell(f) := \max_{p\colon f(p)>0} |p|$.

Note that the demands must always be satisfied, and the flow through an edge may exceed its capacity. (Equivalently we could have asked for the largest F such that $FD(x, y)$ units are routed between each x, y, with no edge capacity being exceeded. The cost would then be $1/F$.)

Our goal in this section is to prove the following theorem bounding the Poincaré constant in terms of flows. This theorem is due to [Sin92, DS91].

Theorem 11.1. *For any ergodic P and any flow f for P, the Poincaré constant α is bounded as*

$$\alpha \geq \frac{1}{\rho(f)\ell(f)} .$$

Proof. Starting from the symmetrized expression for variance (equation (10.3) of the previous lecture), we rewrite $\mathrm{Var}_\pi[\varphi]$ using $\pi(x)\pi(y) = D(x, y) = \sum_{p \in \mathcal{P}_{x,y}} f(p)$ as

$$\begin{aligned} 2\mathrm{Var}_\pi[\varphi] &= \sum_{xy} \pi(x)\pi(y) \left(\varphi(x) - \varphi(y)\right)^2 \\ &= \sum_{xy} \sum_{p \in \mathcal{P}_{xy}} f(p) \left(\varphi(x) - \varphi(y)\right)^2 . \end{aligned}$$

We are aiming at an expression made up of local variances. Therefore we use the following telescoping sum: for any path $p \in \mathcal{P}_{xy}$, $\varphi(x) - \varphi(y) = \sum_{(u,v)\in p}(\varphi(v) - \varphi(u))$. Thus we may continue the above derivation as follows:

$$2\mathrm{Var}_\pi[\varphi] = \sum_{xy}\sum_{p\in\mathcal{P}_{xy}} f(p)\left(\sum_{(u,v)\in p}(\varphi(v)-\varphi(u))\right)^2 \le \sum_{xy}\sum_{p\in\mathcal{P}_{xy}} f(p)|p|\sum_{(u,v)\in p}(\varphi(v)-\varphi(u))^2,$$

by the Cauchy-Schwarz inequality. Switching the order of summation, this equals

$$\sum_{e=(u,v)}(\varphi(v)-\varphi(u))^2\sum_{p\ni e} f(p)|p| \le \ell(f)\sum_{e=(u,v)}(\varphi(v)-\varphi(u))^2\sum_{p\ni e} f(p).$$

Recalling the definitions $\sum_{p\ni e} f(p) = f(e)$ and $\rho(f) = \max_e f(e)/C(e)$, the above quantity is at most

$$\ell(f)\rho(f)\sum_{e\in(u,v)}(\varphi(v)-\varphi(u))^2\,C(e) = 2\ell(f)\rho(f)\mathcal{E}_P(\varphi,\varphi),$$

where in the last step we substituted $C(e) = \pi(u)P(u,v)$. This completes the proof. □

Combining Theorems 10.4 and 11.1 we immediately get:

Corollary 11.2. *For any lazy, ergodic Markov chain P and any flow f, we have*

$$\tau_x(\varepsilon) \le \rho(f)\ell(f)\left(\ln\varepsilon^{-1} + \tfrac{1}{2}\ln(4\pi(x))^{-1}\right).$$

In particular, the mixing time satisfies $\tau_{\mathrm{mix}} = O(\rho(f)\ell(f)\log\pi_{\min}^{-1})$, where $\pi_{\min} = \min_{x\in\Omega}\pi(x)$.

The main implication of this theorem is that the mixing time is essentially bounded by the cost $\rho(f)$ of any flow f; the factors $\ell(f)$ and $\ln\pi_{\min}^{-1}$ are usually easily shown to be small (i.e., low-degree polynomials in the problem size n). In particular, in most applications the flows that we use will route all flow along *shortest* paths; in this case $\ell(f)$ is bounded by the *diameter* of the Markov chain, which is typically polynomial in n. Also, the size of the state space $|\Omega|$ is typically singly exponential in the measure n of problem size (e.g., for spin systems n is the size of the underlying graph); so if π is uniform then $\ln\pi_{\min}^{-1}$ is $O(n)$. When π is not uniform we can replace $\pi_{\min}$ by $\pi(x)$, where x is the initial state, so provided we start off in a state of (near-)maximum probability this factor will still be small.

Notes:

1. The Poincaré constant α also provides a much more trivial *lower* bound on the mixing time, namely $\tau_{\mathrm{mix}} \ge \Omega(\alpha^{-1})$.
2. As a partial converse to Corollary 11.2, one can prove that, at least when P is reversible, there always exists a flow f satisfying $\rho(f) \le c\tau_{\mathrm{mix}}$ and $\ell(f) \le c'\tau_{\mathrm{mix}}$, for universal constants c, c'. (This is the flow generated by the flow of probability mass in the Markov chain itself; see [Sin92].)
3. A natural "dual" of the multicommodity flow problem is the well studied *Sparsest Cut* problem,[1] where the objective is to find a cut $(S,\bar{S})$ in G that minimizes the "conductance"

$$\Phi_S := \frac{C(S,\bar{S})}{D(S,\bar{S})} = \frac{\sum_{(x,y)\in(S,\bar{S})}\pi(x)P(x,y)}{\pi(S)\pi(\bar{S})}. \tag{11.1}$$

[1]Actually Sparsest Cut is usually defined in a less symmetric form, with $\pi(\bar{S})$ omitted from the denominator in (11.1) and the minimization taken only over cuts with $\pi(S) \le \frac{1}{2}$. It should be clear **[exercise!]** that this affects the minimum cut value by a factor of at most 2.

Since $C(S,\bar{S})$ is the total capacity of the cut, and $D(S,\bar{S})$ is the total demand across the cut, it should be obvious **[exercise!]** that $\rho \geq \Phi^{-1}$, where $\rho := \inf_f \rho(f)$ and $\Phi := \min_S \Phi_S$. Much less obvious is the fact that $\rho \leq O(\Phi^{-1}\log|\Omega|)$, and moreover this is tight [LR88, Sin92]. i.e., the problem exhibits an integrality gap of $\Theta(\log|\Omega|)$. Moreover, Cheeger's inequality is a famous bound on the eigenvalue gap of the form $1-\lambda_2 \geq \frac{\Phi^2}{8}$ (and this generalizes to the non-reversible case, with $1-\lambda_2$ replaced by α). This yields an alternative bound on mixing time: $\tau_{\text{mix}} = O(\Phi^{-2}\log\pi_{\text{min}}^{-1})$. However, cuts are usually much less useful than flows for obtaining upper bounds on mixing times: the reason is that *any* flow gives such an upper bound via Corollary 11.2, while obtaining such a bound via cuts requires us to reason about *every* cut. (Indeed, in most MCMC examples the only approach we know to estimate Φ is to exhibit a good flow.) On the other hand, any cut gives a *lower* bound on mixing time, via the (easy) inequality $\tau_{\text{mix}} \geq \Omega(\Phi^{-1})$. (The intuition is that a cut with a small value of Φ_S is a bottleneck that the Markov chain takes a long time to cross.) We'll talk more about lower bounds in a later lecture.

11.2 Simple examples

We now proceed to apply the above flow technology to obtain bounds on the mixing time of various Markov chains. We begin with a few very simple warm-up examples.

11.2.1 The hypercube

Consider lazy random walk on the hypercube $\{0,1\}^n$. Let $N = 2^n$ be the number of vertices. Then $\pi(x) = \frac{1}{N}$. The edge capacity is $C(u,v) = \pi(u)P(u,v) = \frac{1}{N}\frac{1}{2n} = \frac{1}{2Nn}$ for all edges (u,v). The demand between vertices x, y is $D(x,y) = \pi(x)\pi(y) = \frac{1}{N^2}$.

Now we need to construct a flow f satisfying the demands above. Moreover, to obtain the best bound possible we want to do this maximizing the ratio $\frac{1}{\rho(f)\ell(f)}$. Intuitively, this is achieved by spreading the flow between x and y evenly among all shortest paths from x to y. So we define the flow f in this way.

To compute $\rho(f)$ we now need to get a handle on $f(e)$ for all e. This is not difficult if we use the symmetry of the hypercube to notice that $f(e) = f(e')$ for all edges e, e', so that:

$$f(e) = \frac{\sum_{e\in E} f(e)}{|E|} = \frac{\frac{1}{N^2}\sum_{x,y\in V}\text{dist}(x,y)}{Nn} = \frac{n/2}{Nn} = \frac{1}{2N},$$

where we have used the fact that the average distance $\frac{1}{N^2}\sum_{x,y\in V}\text{dist}(x,y)$ between two vertices in the hypercube is $n/2$.

Hence, $\rho(f) = \max_{e\in E}\frac{f(e)}{C(e)} = \frac{1}{2N}2Nn = n$. Moreover, as all the flow goes along shortest paths we have $\ell(f) = n$, so from Corollary 11.2 we get

$$\tau_{\text{mix}} = O(\rho(f)\ell(f)\log\pi_{\text{min}}^{-1}) = O(n^3).$$

This is not tight, as we saw in an earlier lecture that $\tau_{\text{mix}} \sim \frac{1}{2}n\ln n$ (and indeed obtained a similar bound, up to a constant factor, by coupling). However, our flow argument does yield a low-degree polynomial upper bound on the mixing time despite the fact that the size of the cube is exponential in n, which is a non-trivial result. The slackness in our bound is typical of this method, which is more heavy-duty than coupling.

Let's look in a bit more detail at where we lose here. In the case of the hypercube, it is actually possible to compute exactly the Poincaré constant α (or, equivalently since this is a reversible chain, the spectral gap $1-\lambda_2$) to be $\sim \frac{1}{n}$. This shows that neither Theorem 10.4 nor Theorem 11.1 is tight:

- Using the above exact value of α, Theorem 10.4 gives $\tau_{mix} = O(n^2)$, which is off by a factor of almost n from the true value. This error comes from the factor $\ln \pi(x)^{-1}$, which results (in the reversible case) from approximating the mixing time using only the second eigenvalue rather than the whole spectrum.
- Theorem 11.1 bounds α by $\frac{1}{\rho(f)\ell(f)} = \frac{1}{n^2}$, which is off by a factor of n from the true value. Since our flow f is presumably optimal, this demonstrates the inherent slackness in estimating α using flows.

11.2.2 Random walk on a line

We consider the Markov chain for lazy random walk on the line $\{1, 2, \ldots, N\}$, with self-loop probability 1/2 at every state (except at the endpoints where it is 3/4). (Thus the chain moves left or right from each position with probability 1/4 each.) The stationary distribution is uniform: $\pi(x) = 1/N$. The edge capacities and demands are

$$C(e) = \pi(x)P(x,y) = \frac{1}{4N} \text{ for all non-self loop edges },$$

$$D(x,y) = \pi(x)\pi(y) = \frac{1}{N^2} \; \forall x, y .$$

There is only one simple path between each pair of vertices, so there is a unique flow f here. The amount of flow on any edge $(i, i+1)$ is given by

$$f\big((i,i+1)\big) = i(N-i)\frac{1}{N^2} \leq \frac{1}{4} .$$

(The maximum is achieved on the middle edge.) The cost of the flow f is

$$\rho(f) = \max_e \frac{f(e)}{C(e)} \leq \frac{1/4}{1/(4N)} = N ,$$

and the length is

$$\ell(f) = N .$$

Thus from Theorem 11.1 the Poincar'e constant is bounded by

$$\alpha \geq \frac{1}{\rho(f)\ell(f)} \geq \frac{1}{N^2} ,$$

which happens to be asymptotically tight for this Markov chain. From Corollary 11.2 the bound on the mixing time is

$$\tau_{\text{mix}} = O(N^2 \log N).$$

This is off from the true answer only by the $O(\log N)$ factor (which again arises from Theorem 10.4).

11.2.3 Random walk on $K_{2,N}$

It turns out that, for the hypercube, we could have obtained the same bound as we got above by routing the flow along a single shortest path between each pair of vertices (x, y) (see later). Here is an example where spreading out the flow is necessary to get a decent bound.

Consider lazy random walk on the complete bipartite graph $K_{2,N}$, with a self-loop of 1/2 at every vertex. Label the two vertices on the small side of the graph s, t respectively.

The stationary distribution is

$$\pi(s) = \pi(t) = \frac{1}{4} , \qquad \pi(x) = \frac{1}{2N} \; \forall x \neq s, t.$$

The capacity $C(e) = 1/(8N)$ for all edges and the demands are given by:

$$\begin{aligned} D(s,t) &= D(t,s) = 1/16, \\ D(s,x) &= D(t,x) = D(x,s) = D(x,t) = 1/(8N), \\ D(x,y) &= 1/(4N^2). \end{aligned}$$

Suppose first that we send the flow along a single shortest path for each pair of states. Then exactly four edges will carry $1/16$ units of flow for the $s \to t$ and $t \to s$ paths, so

$$\rho(f) \geq \frac{1/16}{1/(8N)} = \frac{N}{2}.$$

Also $\ell(f) = 2$. Hence (assuming we start at either s or t), the bound we get on the mixing time will be $O(N)$. This is a severe over-estimate as the true mixing time (starting from any state) is $\Theta(1)$ (why?)

Now suppose, on the other hand, that we distribute the $s \to t$ and $t \to s$ flows along all shortest paths evenly. Then,

$$\max_e f(e) \leq \frac{1}{8N} + \frac{1}{16N} + \frac{1}{4N^2}\frac{N-1}{2} \leq C \times \frac{1}{N}$$

for a constant C, which gives

$$\rho(f) \leq \frac{C \times 1/N}{1/(8N)} = 8C .$$

Hence, the mixing time starting from s or t is $O(1)$, which is of the right order.

11.3 Flow encodings

Our examples so far have been on symmetric graphs where computing the cost of a flow was easy. But in general, it is difficult to determine how much flow is carried on an edge. To solve this problem, we will develop technology to count paths and calculate flows in a generic setting.

Let $|\Omega| = N$ be the size of the state space of an ergodic, lazy Markov chain, where N is exponential in n, the natural measure of problem size. Assume the stationary distribution is uniform ($\pi(x) = 1/N$). Also assume $P(u,v) \geq 1/\text{poly}(n)$ for all non-zero transition probabilities. (This implies that the degree of the underlying graph is not huge, and holds in most of our examples.) Then the capacity of any edge is

$$C(u,v) = \pi(u)P(u,v) \geq \frac{1}{N\text{poly}(n)} .$$

To get a polynomial upper bound on mixing time, we need to construct a flow f such that $\ell(f) \leq \text{poly}(n)$ and

$$\frac{f(e)}{C(e)} \leq \text{poly}(n)$$

for all edges (transitions in the Markov chain) e. Hence we we must have, for all e,

$$f(e) \leq \frac{\text{poly}(n)}{N}. \tag{11.2}$$

On the other hand, the number of edges in the graph is less than $N \times \text{poly}(n)$ and the total flow along all paths is $\sum_{x,y} \frac{1}{N^2} \approx 1$. Hence *some* edge must carry at least $1/(N \times \text{poly}(n))$ flow, i.e.,

$$f(e) \geq \frac{1}{N \times \text{poly}(n)} .$$

Comparing this with Eq. (11.2) we see that any good flow has to be optimal up to a polynomial factor.

Finally, suppose the flow $x \to y$ goes along a *single* path $\gamma_{x,y}$. (Again, in many of our examples this will be the case.) Let paths(e) denote the set of paths through edge e under flow f. Then

$$f(e) \;=\; |\text{paths}(e)| \;\times \frac{1}{N^2}\;.$$

Since we wanted $f(e) \leq \text{poly}(n)/N$, this implies that the flow must satisfy

$$|\text{paths}(e)| \leq \text{poly}(n)|\Omega| \qquad \forall e. \tag{11.3}$$

If we allow multiple paths from x to y then the same calculation applies but in an average sense, and if the stationary distribution is non-uniform then again the same applies but with everything suitably weighted.

Equation (11.3) indicates that we must compare $|\text{paths}(e)|$ with the size of the state space $N = |\Omega|$. However, since computing N is often our goal, this seems difficult! The following machinery is designed to get around this problem. The key idea is to set up an *injective mapping* from paths(e) to Ω which allows us to compare their sizes *implicitly*. This technology is employed in almost all non-trivial applications of multicommodity flows in the analysis of mixing times.

Definition 11.3. *An* encoding *for a flow f (that uses only single paths γ_{xy} for each x, y) is a set of functions η_e : paths(e) $\to \Omega$ (one for each edge e) such that*

1. *η_e is injective*
2. *for some $\beta \leq \text{poly}(n)$, $\pi(x)\pi(y) \leq \beta\pi(z)\pi(\eta_e(x, y))$ $\forall (x, y) \in \text{paths}(e)$, where $e = (z, z')$.*

Remark: Property 1 says that η_e is an injection, as motivated earlier. Property 2 says that η_e is in addition weight-preserving up to a polynomial factor β. Note that property 2 is automatically satisfied with $\beta = 1$ when π is uniform. In some applications, we may usefully weaken property 1 slightly as follows: we may not have a perfect injection, but may require a small amount of "extra information" to invert η_e. As should be clear from the proof of the following claim, this will just insert a modest additional factor into the bound on the mixing time.

Claim 11.4. *If there exists an encoding for f as above, then $\rho(f) \leq \beta \max_{P(z,z')>0} \frac{1}{P(z,z')}$.*

Proof. Let $e = (z, z')$ be an arbitrary edge. Then

$$f(e) = \sum_{(x,y)\in\text{paths}(e)} \pi(x)\pi(y) \leq \beta \sum_{(x,y)\in\text{paths}(e)} \pi(z)\pi(\eta_e(x, y)) \leq \beta\pi(z).$$

In the first inequality here we have used property 2, and in the second we have used property 1.

Finally, $C(e) = \pi(z)P(z, z')$, so $f(e)/C(e) \leq \beta/P(z, z')$. □

We now give a simple example of an analysis using flow encodings.

11.3.1 Example: the hypercube revisited

Previously, in analyzing the random walk on the cube by flows, we spread flow evenly and used the symmetry of the cube to bound $\rho(f)$. In more sophisticated applications, we will not have this symmetry and not know $N = 2^n$. Flow encodings let us proceed without appealing to these properties of the cube.

Recall that the capacity of an edge e is $C(e) = 1/(2nN)$, and the demand between two vertices x, y is $D(x,y) = 1/N^2$. Now consider a flow that sends all the $x \to y$ flow along the "left-right bit-fixing path" γ_{xy}. That is, correct each bit of x sequentially, left-to-right, until arriving at y. Then $\ell(f) = n$, clearly.

We now bound the cost of this flow using the encoding technique. Consider an arbitrary edge $e = (z, z')$, where z and z' differ in bit position i. Consider any pair $(x,y) \in \text{paths}(e)$. What do we know about x and y? Notice that y agrees with z in the first $i-1$ bits (which have already been corrected), and x agrees with z in the last $n-i$ bits. We therefore define $\eta_e : \text{paths}(e) \to \Omega$ by $\eta_e(x,y) = x_1 x_2 \ldots x_i y_{i+1} y_{i+2} \ldots y_n$. I.e., $\eta_e(x,y)$ is the $0,1$-string that agrees with x on the first i bits and with y on the rest.

Now it is easy to see that we can uniquely recover (x,y) from $\eta_e(x,y)$ and e (this is why we constructed η_e this way!), so η_e is an injection. Since the stationary distribution is uniform, it is trivially weight-preserving. Hence η_e is a valid encoding. By the above Claim therefore,

$$\rho(f) \leq \max_{z,z'} \frac{1}{P(z,z')} = 2n \ .$$

Up to a constant, this is the same bound we obtained for the cube by spreading flow uniformly and appealing to symmetry. In particular, we again get $\tau_{\text{mix}} \leq O(n^3)$. However, the key point about this alternative flow and its analysis is that we never used the fact that $N = 2^n$, nor any of the cube's symmetry. In the next lecture, we'll see a much harder example in which this feature is crucial.

References

[DS91] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Annals of Applied Probability*, 1:36–61, 1991.

[LR88] T. Leighton and S. Rao. An approximate max-flow min-cut theorem for uniform multicommodity flow problems with applications to approximation algorithms. *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 422–431, 1988.

[Sin92] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.