Lecture 10: October 08

Lecturer: Prof. Alistair Sinclair *Scribes: Jörg Lässig and Dirk Sudholt*

10.1 Multicommodity Flow

In this lecture we consider multicommodity network flows. As usual, we assume an ergodic Markov chain P whose stationary distribution is 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 oil between any pair of nodes. More formally, a flow is a function $f\colon P \to \mathbb{R}^+ \cup \{0\}$ where $P = \bigcup_{xy} P_{xy}$ and P_{xy} denotes the set of all simple paths from x to y s. t.

$$\sum_{p \in P_{xy}} f(p) = D(x, y) \ .$$

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 lecture is to prove the following theorem relating mixing times to flows. Recall from Lecture 2 that a *lazy* version of a Markov chain is obtained by introducing additional self-loop probabilities of 1/2 at each state.

Theorem 10.1 *For any lazy, ergodic Markov chain P and any flow f, we have*

$$\tau_{\mathrm{mix}} \leq O\left(\rho(f)\ell(f) \ln \pi_{\min}^{-1}\right) \ ,$$

where $\pi_{\min} = \min_{x \in \Omega} \pi(x)$.

The main implication of this theorem is that the mixing time is roughly proportional to the cost $\rho(f)$ of a flow; 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 colorings n is the size of the underlying graph); so if π is uniform then $\ln \pi_{\min}^{-1}$ is $O(n)$. In fact, when π is not uniform we can replace $\pi_{\min}$ in the above theorem 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.

Remark: Note that *any* flow f gives an upper bound on the mixing time. There is a similar lower bound on the mixing time in terms of flows; however, to use that we would need to consider all possible flows and show that none of them can be good. For lower bounds it makes more sense to consider the "dual" problem of *sparsest cut*; then *any* bad cut will give us a lower bound on mixing time. We will consider the sparsest cut problem, and its relation to multicommodity flows, later in the course.

We will spend the rest of the lecture proving Theorem 10.1. First, we need some definitions.

Definition 10.2 *Let $\varphi\colon \Omega \to \mathbb{R}$ be any real-valued function. Define the expectation $E_\pi\varphi = \sum_x \pi(x)\varphi(x)$. The* (global) variance *of ϕ is given by*

$$\begin{aligned}
\mathrm{Var}_\pi\varphi &:= \sum_x \pi(x)\cdot(\varphi(x) - E_\pi\varphi)^2 \\
&= \sum_x \pi(x)\varphi(x)^2 - (E_\pi\varphi)^2 \\
&= \sum_x \pi(x)\varphi(x)^2 \sum_y \pi(y) - \sum_x \pi(x)\varphi(x) \sum_y \pi(y)\varphi(y) \\
&= \sum_{xy} \left(\pi(x)\pi(y)\varphi(x)^2 - \pi(x)\pi(y)\varphi(x)\varphi(y)\right) \\
&= \frac{1}{2}\sum_{xy} \pi(x)\pi(y)\cdot(\varphi(x)-\varphi(y))^2 \ .
\end{aligned}$$

By analogy with the last line above, it is natural to define the "local variance" by considering only adjacent pairs x, y as follows.

Definition 10.3 *The* local variance *is defined as*

$$\mathcal{E}_\pi(\varphi,\varphi) = \frac{1}{2}\sum_{xy} \pi(x)P(x,y)\cdot(\varphi(x)-\varphi(y))^2 \ .$$

$\mathcal{E}_\pi(\cdot,\cdot)$ is also known as the "Dirichlet form."

A *Poincaré inequality* bounds the ratio of the local to the global variance for any function φ. This leads to the following definition.

Definition 10.4 *The* Poincaré constant *is defined by*

$$\alpha := \inf_{\varphi \text{ non-constant}} \frac{\mathcal{E}_\pi(\varphi,\varphi)}{\mathrm{Var}_\pi\varphi} \ .$$

It should be intuitively reasonable that the Poincaré constant provides an upper bound on the mixing time, essentially because the local variance $E_\pi(\varphi,\varphi)$ determines the rate at which φ is "averaged" per step of the Markov chain. This intuition is captured in the following general theorem.

Theorem 10.5 *For any lazy ergodic P and any initial state $x \in \Omega$,*

$$\tau_x(\varepsilon) \le \frac{1}{\alpha}\left(2\ln\varepsilon^{-1} + \ln\pi(x)^{-1}\right) \ .$$

The second ingredient in the proof of Theorem 10.1 is a bound on the Poincaré constant in terms of multi-commodity flows.

Theorem 10.6 *For any ergodic P and any flow f for P*

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

It is immediate that these theorems together imply Theorem 10.1. Plugging the lower bound for α from Theorem 10.6 into Theorem 10.5 and letting $\varepsilon := 1/(2e)$ yields an upper bound $O(\rho(f)\ell(f)\ln \pi(x)^{-1})$ on $\tau_x(1/(2e))$. This is actually a stronger result than Theorem 10.1, which is obtained by taking a worst-case starting point x.

We now proceed to prove the above theorems. The first, Theorem 10.5, is a version of a classical result. We follow the proofs of Jerrum [Je03] and Mihail [Mi89].

Proof of Theorem 10.5: Because P is lazy we have $P = \frac{1}{2}(I + \widehat{P})$, where $\widehat{P}$ is stochastic and has the same stationary distribution π as P. Define a function $[P_\varphi]\colon \Omega \to \mathbb{R}$ by $[P_\varphi](x) := \sum_y P(x,y)\varphi(y)$. This function is called the "one-step averaging of φ" as we take averages over φ-values after one step. Similarly, $[P^t\varphi](x) = \sum_y P^t(x,y)\varphi(y)$ is the t-step averaging of φ. Since P is ergodic, $[P^t\varphi]$ converges to the constant function $E_\pi\varphi$, so

$$\mathrm{Var}[P^t\varphi] \xrightarrow{t\to\infty} 0 \ .$$

Note also that $E_\pi[P^t\varphi] = E_\pi\varphi$ for all t.

The following lemma is the main content of the proof.

Lemma 10.7 (Main Lemma) *For any $\varphi : \Omega \to \mathbb{R}$*

$$\mathrm{Var}_\pi[P_\varphi] \leq \mathrm{Var}_\pi\varphi - \mathcal{E}_\pi(\varphi,\varphi) \ .$$

We defer the proof of the Main Lemma for a moment. For now, we state an immediate corollary that establishes a contraction $1-\alpha$ for the variance in each step of the chain.

Corollary 10.8 $\mathrm{Var}_\pi[P^t_\varphi] \leq (1-\alpha)^t \cdot \mathrm{Var}_\pi\varphi$.

Proceeding with the proof of Theorem 10.5, let $A \subseteq \Omega$ be arbitrary. Define

$$\varphi_A(x) := \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \ . \end{cases}$$

Then $\mathrm{Var}_\pi\varphi_A \leq 1$, so $\mathrm{Var}_\pi\left[P^t_{\varphi_A}\right] \leq (1-\alpha)^t \leq e^{-\alpha t}$. Setting $t = \frac{1}{\alpha}(\ln \pi(x)^{-1} + 2\ln\varepsilon^{-1})$, we have $\mathrm{Var}_\pi[P^t_{\varphi_A}] \leq \varepsilon^2 \cdot \pi(x)$. But on the other hand

$$\begin{aligned} \mathrm{Var}_\pi[P^t_{\varphi_A}] &\geq \pi(x)\left(P^t_{\varphi_A}(x) - E_\pi[P^t_{\varphi_A}]\right)^2 \\ &= \pi(x)\left(P^t_x(A) - \pi(A)\right)^2 , \end{aligned}$$

where the last equality follows from the definition of φ_A and the fact that $E_\pi[Py] = E_\pi y$ for all y. Together, $\left(P^t_x(A) - \pi(A)\right)^2 \leq \varepsilon^2$. Since A is arbitrary, we get $||P^t_x - \pi|| \leq \varepsilon$ which proves the theorem. ∎

Proof of Main Lemma:

$$\begin{aligned}[P_\varphi](x) &= \sum_y P(x,y)\varphi(y) \\ &= \frac{1}{2}\varphi(x) + \frac{1}{2}\sum_y \widehat{P}(x,y)\varphi(y) \\ &= \frac{1}{2}\sum_y \widehat{P}(x,y)\left(\varphi(x)+\varphi(y)\right) \ .\end{aligned}$$

Assume w. l. o. g. that $E_\pi\varphi = 0$ (shifting φ by a constant value does not affect any of the quantities we are interested in, all of which are variances). Then

$$\begin{aligned}\mathrm{Var}_\pi[P\varphi] &= \sum_x \pi(x)\left([P\varphi](x)\right)^2 \\ &= \frac{1}{4}\sum_x \pi(x)\left(\sum_y \widehat{P}(x,y)\left(\varphi(x)+\varphi(y)\right)\right)^2 \\ &\le \frac{1}{4}\sum_{xy}\pi(x)\widehat{P}(x,y)\left(\varphi(x)+\varphi(y)\right)^2 \ ,\end{aligned}$$

where the last inequality follows from the Cauchy-Schwarz inequality (or the fact that the square of an expectation is bounded by the expectation of the square). Moreover, we can rewrite the variance yet again as

$$\begin{aligned}\mathrm{Var}_\pi\varphi &= \frac{1}{2}\sum_x \pi(x)\varphi(x)^2 + \frac{1}{2}\sum_y \pi(y)\varphi(y)^2 \\ &= \frac{1}{2}\sum_{xy}\pi(x)\varphi(x)^2\widehat{P}(x,y) + \frac{1}{2}\sum_{xy}\pi(x)\widehat{P}(x,y)\varphi(y)^2 \\ &= \frac{1}{2}\sum_{xy}\pi(x)\widehat{P}(x,y)\left(\varphi(x)^2+\varphi(y)^2\right) \ .\end{aligned}$$

Taking differences,

$$\mathrm{Var}_\pi\varphi - \mathrm{Var}_\pi[P_\varphi] \ \ge \ \frac{1}{4}\sum_{xy}\pi(x)\widehat{P}(x,y)\left(\varphi(x)-\varphi(y)\right)^2 \ .$$

Observe that all entries in $\widehat{P}$ are twice as large as the entries in its lazy version P, except for the diagonal elements. Diagonal elements can be ignored, however, as they contribute a value of 0 to the above sum. Hence, the right-hand side is equal to

$$\frac{1}{2}\sum_{xy}\pi(x)P(x,y)\left(\varphi(x)-\varphi(y)\right)^2 = \mathcal{E}_\pi(\varphi,\varphi) \ .$$

This completes the proof of the Main Lemma and the proof of Theorem 10.5. ■

Remark 1: The above proof assumes that the Markov chain is lazy. This device ensures that there are no periodicity issues. It can be avoided by passing to a *continuous time* version of the chain, in which periodicity can never arise. We omit the details. ■

Remark 2: If the Markov chain P is reversible, then we can take an alternative approach to proving Theorem 10.5, based on the following.

Fact 10.9 *If P is* reversible *then its eigenvalues are*

$$1 = \lambda_1 > \lambda_2 > \cdots > \lambda_N > -1$$

and its spectral gap $1 - \lambda_2$ *satisfies*

$$1 - \lambda_2 = \inf_{\varphi \text{ non-constant}} \frac{\mathcal{E}_\pi(\varphi,\varphi)}{\mathrm{Var}_\pi \varphi} \ . \tag{10.1}$$

Thus in the reversible case the Poincaré constant has an alternative interpretation as the spectral gap of the transition matrix. (The formula (10.1) follows from the standard variational characterization of eigenvalues of symmetric matrices; since P is not necessarily symmetric, but is reversible—and hence similar to a symmetric matrix—the standard formula has to be suitably weighted by the principal eigenvector π.)

This observation yields an entirely different way of proving Theorem 10.5 in the reversible case by expressing the distribution $x^{(t)}$ at time t as a linear combination of eigenvectors, and noting that the slowest rate of decay to 0 of the eigenvectors is $\max_{i\geq 2} |1-\lambda_i|$. If in addition P is lazy, then all eigenvectors are non-negative (i.e., $\lambda_N \geq 0$), so $\max_{i\geq 2} |1-\lambda_i| = 1-\lambda_2$. Thus the rate at which $x^{(t)}$ approaches π can be bounded in terms of the spectral gap. The overhead in approximating the rate of decay of all the other eigenvalues by just the second is captured by the factor $\ln \pi(x)^{-1}$. For an extension of this approach to non-reversible chains (from which you can also deduce the details of the simpler reversible case) see [Fi91]. ■

We now turn to the second main ingredient of our analysis, which is a bound on the Poincaré constant in terms of flows. This theorem is due to [Si92,DS91].

Proof of Theorem 10.6: We rewrite $\mathrm{Var}_\pi\varphi$ using $\pi(x)\pi(y) = D(x,y) = \sum_{p\in P_{x,y}} f(p)$ as

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

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

$$\begin{aligned} &\sum_{xy} \sum_{p\in P_{xy}} f(p) \left(\sum_{(u,v)\in p} (\varphi(v) - \varphi(u)) \right)^2 \\ \leq &\sum_{xy} \sum_{p\in P_{xy}} f(p)|p| \sum_{(u,v)\in p} \left(\varphi(v) - \varphi(u)\right)^2 \end{aligned}$$

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

$$\begin{aligned} &\sum_{e=(u,v)} \left(\varphi(v) - \varphi(u)\right)^2 \sum_{p\ni e} f(p)|p| \\ \leq\ &\ell(f) \sum_{e\in(u,v)} \left(\varphi(v) - \varphi(u)\right)^2 \sum_{p\ni e} f(p) \ . \end{aligned}$$

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

$$\ell(f)\rho(f) \sum_{e\in(u,v)} \left(\varphi(v) - \varphi(u)\right)^2 C(e)$$

and substituting $C(e) = \pi(u)P(u,v)$ results in the bound

$$2\ell(f)\rho(f)\mathcal{E}_\pi(\varphi,\varphi)\ .$$

This completes the proof. ■

In the next lecture we shall see some applications of Theorem 10.1, by constructing suitable flows in various Markov chains.

References

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

[Fi91] J.A. Fill "Eigenvalue Bounds on Convergence to Stationarity for Nonreversible Markov Chains, with an Application to the Exclusion Process," *Annals of Applied Probability* **1** (1991), pp. 62–87.

[Je03] M. Jerrum, *Counting, sampling and integrating: Algorithms and complexity*, Birkhäuser Lectures in Mathematics, 2003.

[Mi89] M. Mihail, "Conductance and convergence of Markov chains: A combinatorial treatment of expanders," *Proceedings of the 30th STOC*, 1989, pp. 526–531.

[Si92] A. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatorics, Probability and Computing* **1**, 1992, pp. 351–370.