MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Lecture N — April 3, SP17
Lecturer: Yury Polyanskiy — Scribe notes by Jennifer Tang

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They are posted to serve class purposes.*

Path Coupling

1

Reminder:

Theorem 1. *If P is irreducible and aperiodic, then*

$$d_{TV}(P_t(x,\cdot),\pi) \le ce^{-\alpha t} \tag{1}$$

for some $c, \alpha > 0$.

Proof. Special case when $d_{TV}(P_t(x,\cdot), P_t(y,\cdot)) \le e^{\alpha} \le 1, \forall x \neq y$.

$\Leftrightarrow$ $\forall x \neq y \exists$ a coupling $(*)$ of $X_1 \sim P(x,\cdot)$ and $Y_1 \sim P(y,\cdot)$ where $\mathbb{P}[X_1 \neq Y_1] \le e^{-\alpha}$.

Start X_t from x_0 and Y_t from y_0, and at each step couple them using $(*)$ (This gets Markovian coupling).

$$\mathbb{P}[X_1 \neq Y_1] = e^{-\alpha}\mathbb{I}\{x_0 \neq y_0\}$$
$$\mathbb{P}[X_t \neq Y_t] = e^{-\alpha t}\mathbb{I}\{x_0 \neq y_0\}$$
$$\Rightarrow \bar{d}(t) \le e^{-\alpha t}$$

$\Leftrightarrow \bar{d}(t+s) \le \bar{d}(t)\bar{d}(s)$ The assumption is just saying $\bar{d}(1) \le e^{-\alpha}$ □

The bad news is that most markov chains have $\bar{d}(1) = 1$. Consider the cycle. If the supports of $P(x_0,\cdot)$ and $P(y_0,\cdot)$ are disjoint, then the total variation is 1. The resolution is to consider contraction of other distances.

Definition 1. *(Ω, ρ) is a (finite) metric space. The 1-Wasserstein distance is*

$$W_\rho(P,Q) = \inf_{X\sim P, Y\sim Q} \mathbb{E}\rho(X,Y)$$

How do we think about W_ρ?

Proposition 2. *(a)* d_{TV} *is* W_ρ *for* $\rho(x,y) = \mathbb{I}\{x_0, y_0\}$

(b) (Kantorovich duality) $W_\rho(P,Q) = \sup_{f:1-Lip(\rho)} \mathbb{E}_P f - \mathbb{E}_Q f$

(c) (Roughly, not exactly true) $W_\rho(P_n, P) \to 0 \Leftrightarrow P_n \to P$ *weakly (which implies* $\mathbb{E}\rho(x_n, v_0) \to \mathbb{E}\rho(x, v_0)$*)*

Theorem 3 (Dobrushin's contraction theorem)**.** *Suppose that* $\forall x \neq y$*:*

1. $W_\rho(P(x,\cdot), P(y,\cdot)) \leq e^{-\alpha}\rho(x,y)$
2. $\min_{x\neq y} \rho(x,y) \geq 1$

then,

$$\bar{d}(t) \leq e^{-\alpha t} \cdot diam(\mathcal{X})$$
$$t_{mix} \leq \frac{1}{\alpha} \ln 4 \cdot diam(\mathcal{X})$$

Proof. Start two chains X_t, Y_t from x_0, y_0.

Sufficient to prove: $\mathbb{P}^{x_0,y_0}[X_t \neq Y_t] \leq e^{-\alpha t}\rho(x_0, y_0)$

Indeed,

$$\bar{d}(t) = \sup_{x_0 \neq y_0} d_{TV}(P_t(x_0), P_t(y_0)) \leq e^{-\alpha t} diam \tag{2}$$

Recall $t_{mix}(\frac{1}{4}) \leq \inf\{t : d(t) < \frac{1}{4}\}$. We couple X_{t+1}, Y_{t+1} via optimal coupling of $P(X_t,\cdot)$ to $P(Y_t,\cdot)$ minimizing

$$\mathbb{E}[\rho(X_{t+1}, Y_{t+1})|X_t, Y_t] \leq e^{-\alpha}\rho(X_t, Y_t)\mathbb{E}[\rho(X_t, Y_t)] \leq e^{-\alpha t}\rho(x_0, y_0)$$

Finally notice,

$$\mathbb{I}\{X_t \neq Y_t\} \leq \rho(X_t, Y_t)$$
$$\Rightarrow$$
$$\mathbb{P}[X_t \neq Y_t] \leq e^{-\alpha t}\rho(x_0, y_0)$$

$\square$

Theorem 4 (Path Coupling)**.** *Let* P *be a markov chain on* $\mathcal{X} \in \Omega$ *and let* $\tilde{P}$ *be the extension of* P *to* Ω*. Suppose that*

1. *G is a connected graph on* Ω *with* $\ell(e) \geq 1$ *being length function on edges*

2. *Let*

$$\rho(x,y) = \min_{paths: x=x_0 \to x_1 \to \ldots \to x_m} \sum_{j=0}^{m} \ell(x_j, x_{j+1})$$

3. *Suppose for any $x \sim y$ (x and y are neighbors in G)*

$$W_\rho(\tilde{P}(x,\cdot), \tilde{P}(y,\cdot)) \le e^{-\alpha}\rho(x,y)$$

then,

$$\bar{d}(t) \le e^{-\alpha t} \cdot diam(\mathcal{X})$$
$$t_{mix} \le \frac{1}{\alpha} \ln 4 \cdot diam(\mathcal{X})$$

Caveat: Never assume $\rho(x,y) = \ell(x,y)$ for $x \sim y$

Proof of Thm 4. Observe that $W_\rho(P_1, P_3) \le W_\rho(P_1, P_2) + W_\rho(P_1, P_2)$. Suppose that $(X,Y) \sim \pi_{XY}$ which is the optimal coupling of P_1 and P_2 and π_{YZ} is the optimal coupling of P_2 to P_3. Define $P_{XYZ} = \pi_{XY}\pi_{Z|Y}$.

$$\mathbb{E}\rho(X,Z) \le \mathbb{E}_\rho(X,Y) + \mathbb{E}_\rho(Y,Z)$$

Consider x_0, y_0 and let $x_0 - x_1 - \ldots - x_n - y_0$ be the minimum path. Then,

$$\rho(x_0, y_0) = \sum \ell(x_j, x_{j+1})$$

$$\begin{aligned} W_\rho(P(x_0,\cdot), P(y_0,\cdot)) &\le \sum W_\rho(P(x_j,\cdot), P(x_{j+1},\cdot)) \\ &\le e^{-\alpha} \sum \rho(x_j, x_{j+1}) \\ &= e^{-\alpha}\rho(x_0, y_0) \end{aligned}$$

$\Rightarrow$ Thm 3's assumptions are satisfied. □

2 Application I: Glauber dynamics on q-coloring of G

Given an undirected graph G:

- $\Omega = \{X : V(G) \to [q]\} = [q]^{V(G)}$ and $\mathcal{X} = \{X : \chi(v) \neq \chi(v'), \forall v, v'\}$
- $\pi(x) = \frac{1}{Z}\mathbb{I}\{x \in \mathcal{X}\}$ for $x \in \Omega$

- $\pi_{x_v|x_{\sim v}} =$ Unif on $C(x,v)$ where

$$\begin{aligned} C(x,v) &= \text{set of permitted colors for } v \text{ in configurationt } x \\ &= [q]\backslash\lambda(\mathcal{N}(0)) \end{aligned}$$

Theorem 5. *If $q > 2\Delta$, then $t_{mix} \leq C_{q,\Delta} n \log n$.*

This shows fast mixing for q colors.

Proof. Let $\rho(x,y) = d_H(x,y) = \#\{v : X(v) \neq y(v)\}$. Consider two colorings x_0 and y_0 which differ at vertex v_0.

The coupling will update synchronously. Choose $v \in V(G)$ uniformly at random. If

- $v \notin \mathcal{N}(v_0)$ then choose $X_1(v) = Y_1(v)$
- $v \in \mathcal{N}(v_0)$, generate $X_1(v) \sim$ Unif on $C(x_0,v)$ and $Y_1(v) = X_1(v)$ or $X_1(v_0)$.

Given that $\rho(x_0,y_0)$, we want to prove $\mathbb{E}\rho(X_1,Y_1) < 1$.
Cases:

- With probability $1 - \frac{\deg(v_0)+1}{n}$, $\rho_{new} = 1$
- With probability $\frac{1}{n}$, $\rho_{new} = 0$ ($v = v_0$ case)
- With probability $\geq \frac{\deg v_0}{n}\left(1 - \frac{1}{C(x_0,v)}\right)$, $\rho_{new} = 1$
- With probability $\leq \frac{\deg v_0}{n}\frac{1}{C(x_0,v)}$, $\rho_{new} = 2$ (color of y prohibited by color of v_0 in y_0)

$$\begin{aligned} \mathbb{E}\rho(X_1,Y_1) &= 1 - \frac{1}{n} + \frac{\deg(v_0)}{n}\frac{1}{C(x,v)} \\ &\leq 1 - \frac{1}{n} + \frac{\Delta}{n}\frac{1}{q-\Delta} \\ &= 1 - \frac{1}{n}\left(1 - \frac{\Delta}{q-\Delta}\right) \\ &= 1 - \frac{1}{n}\frac{q-2\Delta}{q-\Delta} \\ &\leq e^{-c/n} \end{aligned}$$

By path coupling,

$$\begin{aligned} t_{mix} &\leq \frac{n}{c}\log(4\cdot diam) \\ &\leq \frac{n}{c}\log(4n) \end{aligned}$$

$\square$

Last time we showed $t_{mix} \geq \frac{1}{4\Phi_*}$.
For q-coloring of a star, let q be small and $n >> 1$.

$$\Phi_* \leq \left(\frac{n-1}{q-1}\right)^n = \left(1 - \frac{1}{q-1}\right)^n$$
$$t_{mix} \geq \frac{1}{4\Phi_*} \geq e^{C_n \cdot n}$$

3 At the critical value

What if $q = 2\Delta$? At the critical value, $t_{mix} \leq 4qn^3$.

Theorem 6. *Suppose we have a family of couplings for all $x \neq y$, $P(x,\cdot)$ to $P(y,\cdot)$ with the property that*

1. $W_\rho(P(x,\cdot), P(y,\cdot)) \leq \rho(x,y)$
2. $\mathbb{E}^{x_0,y_0}(\rho(X_1,Y_1) - \rho(x_0,y_0))^2 \geq \beta, \forall x_0 \neq y_0$

then

$$t_{mix} \leq 4\frac{diam(\mathcal{X})^2}{\beta}$$

Proof of 6. Run X_t, Y_t via markov coupling of $P(x_{t+1},\cdot)$ to $P(Y_{t+1},\cdot)$ at each step let $Z_t = diam - \rho(X_t,Y_t)$. We have $0 \leq Z_t \leq diam$.

$$\mathbb{E}[Z_t | X_{t-1}, Y_{t-1}] \geq Z_{t-1}$$

from property 1. This forms a submartingale. We know that submartingales cannot oscillate, so eventually, it will hit $diam$.

$$\begin{aligned} \mathbb{E}[Z_t^2|\mathcal{F}_{t-1}] &= \mathbb{E}[(Z_{t-1} + Zt - Z_{t-1})^2|\mathcal{F}_{t-1}] \\ &\geq Z_{t-1}^2 + \beta\mathbb{I}\{X_{t-1} \neq Y_{t-1}\} \\ &= Z_{t-1}^2 + \beta\mathbb{I}\{\tau > t-1\} \end{aligned}$$

where $\tau = \inf\{s : Z_s = diam\}$. This gives

$$\mathbb{E}Z_t^2 \geq \mathbb{E}Z_{t-1}^2 + \beta\mathbb{P}[\tau > t-1]$$
$$diam^2 \geq z_0^2 + \beta \sum_{s=0}^{t-1} \mathbb{P}[\tau > s]$$

As $t \to \infty$,

$$\beta\mathbb{E}\tau \leq diam^2$$
$$t_{mix} \leq \frac{4 \cdot diam^2}{\beta}$$

where the last expression is given by Chebychev's. $\square$

We will apply this to q-coloring. The coupling of X_1 to Y_1 for arbitrary $x_0 \neq y_0$ is the following:

1. Update synchronously v uniformly at random from $V(G)$
2. At v, generate random permutation of $[q]$, $(q_1, q_2, ..., q_n)$.
3. Set $X(v)$ to the first color in the permutation permitted by X and set $Y(v)$ to the first color in the permutation permitted by Y.

$$\mathbb{E}\rho(X_1, Y_1) \leq \rho(X_0, Y_0)$$

if $q = 2\Delta$. Since there exists v such that $X_0(v) \neq Y_0(v)$, we must have $C(x_0, v) \cap C(y_0, v) \neq \emptyset$, so

$$\mathbb{E}(\rho(X, Y) - \rho(X_0, Y_0))^2 \geq \frac{1}{nq}$$

Theorem 7. *For the hardcore model: $\pi(x) = \frac{1}{Z}\lambda^{\#\{X(v)=1\}}$ (where $X : V(G) \to \{0, 1\}$ and the location of all 1's must be an independent set) we have that Glauber dynamics*

$$\lambda < \frac{1}{\Delta - 1} \Rightarrow t_{mix} \leq cn \log n$$
$$\lambda = \frac{1}{\Delta - 1} \Rightarrow t_{mix} \leq cn^3$$