Path coupling:

Idea: Suppose $D \in \{0, 1, \cdots\}$ and $\mathbb{E}[D] < \varepsilon$ then $\mathbb{P}(D=0) \geq 1-\varepsilon$.

Therefore we wish to define some "distance" on $\mathcal{X}$, and try to argue two chains' distance becomes close as t gets large.

Path coupling is a technique that simplifies the construction for all "neighbor pairs". Therefore it's useful for M.C. on graphs.

But of course, we need to first specify what "neighbor" means.

Preparation:

Let $\mathbb{P}$ be the transition matrix on $\mathcal{X}$, let ρ be a metric. Suppose ρ satisfies $\rho(x,y) \geq \mathbb{1}_{\{x \neq y\}}$. Suppose for all (x,y) there's a coupling between $\mathbb{P}(x,\cdot)$ and $\mathbb{P}(y,\cdot)$ that contracts on average, i.e. there exists $\alpha > 0$ s.t.

$$\mathbb{E}_{x,y}\, \rho(X_1, Y_1) \leq e^{-\alpha} \rho(x,y) \text{ for every } (x,y)$$

Then
$$\mathbb{E}_{x,y}\, \rho(X_t, Y_t) = \mathbb{E}_{x,y}\left(\mathbb{E}[\rho(X_t, Y_t) \mid X_{t-1}, Y_{t-1}]\right)$$

$$\leq e^{-\alpha}\, \mathbb{E}_{x,y}[\rho(X_{t-1}, Y_{t-1})]$$

$$\leq e^{-\alpha t}\, \mathrm{diam}(\mathcal{X})$$

Thus
$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq P_{x,y}(X_t \neq Y_t) = P_{x,y}(\rho(X_t, Y_t) \geq 1)$$
$$\leq e^{-\alpha t}\, \mathrm{diam}(\mathcal{X})$$

Therefore

$$t_{mix}(\varepsilon) \le \left\lceil \frac{1}{\alpha}\left[\log(\mathrm{diam}(\mathcal{X})) + \log 1/\varepsilon\right] \right\rceil$$

D. Transportation metric

Fix ρ on $\mathcal{X}$, we define a metric over distributions on ρ

$$\rho_K(\mu,\nu) = \inf_{(X,Y)\sim\Gamma(\mu,\nu)} \mathbb{E}[\rho(X,Y)]$$

Remark ① K stands for Kontorovich (proposed in 1942)

Sometimes also called Wasserstein $(\rho,1)$ metric

Remark ② If $\rho(x,y) = 1_{x\neq y}$ $\quad \rho_K(\mu,\nu) = \|\mu-\nu\|_{TV}$

Remark ③ Fix ρ, the optimal coupling q^* s.t.

$$\mathbb{E}_{(X,Y)\sim q^*}\, \rho(X,Y) = \rho_K(\mu,\nu) \text{ always exists}$$

This is because $\Gamma(\mu,\nu) \simeq$ compact subset of $\mathbb{R}^{(n^2-1)}$-dim simplex

$$\Gamma(\mu,\nu) \to \mathbb{R}^+$$

$$q \to \mathbb{E}_{(X,Y)\sim q}\, \rho(X,Y) \quad \text{continuous.}$$

Lemma : ρ_K is a metric on the space of distributions

$\left(① \rho(x,y)=0 \Leftrightarrow x=y \quad ② \rho(x,y)=\rho(y,x) \quad ③ \rho(x,y)+\rho(y,z) \geq \rho(x,z) \right)$

We just check Δ-inequality

Fix μ, ν, η distributions, let p be optimal coupling of (μ,ν)

q be optimal coupling of (ν,η)

We define a coupling of (μ,ν,η) togetherly

Let $$r(x,y,z) = \begin{cases} \dfrac{p(x,y)\,q(y,z)}{\nu(y)} & \text{if } \nu(y) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

We claim r is a distribution on $\mathcal{X}\times\mathcal{X}\times\mathcal{X}$

and has marginal μ, ν, η, and $(X,Y)\sim p$, $(Y,Z)\sim q$

① $\sum_{z} r(x,y,z) = p(x,y)$ ✓

This directly implies r is a distribution (since $\sum_{x,y} p(x,y)=1$

and $(X,Y) \sim p$

② $\sum_{x} r(x,y,z) = q(y,z)$ ✓

$$\rho_K(\mu,\eta) \le \mathbb{E}\,\rho(X,Z)$$

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

$$= \rho_K(\mu,\nu) + \rho_K(\nu,\eta) \qquad \square$$

Bounding the mixing time via path coupling:

Suppose $\mathcal{X}$ is the vertex set of a connected graph $G = (\mathcal{X}, E_0)$ and l is a length function defined on E_0. That is $l(x,y)$ is a non-negative number for every edge. We assume $l(x,y) \ge 1$.

If $x_0, x_1, \dots, x_r$ is a path on G, we define the length of this path by $\sum_{i=0}^{r-1} l(x_i, x_{i+1})$

The path metric is defined as

$$\rho: V \times V \to \mathbb{R}_{\ge 0}$$

$$(x,y) \mapsto \min\{\text{length of } \xi : \xi \text{ is a path from } x \text{ to } y\}$$

It's not hard to check ρ is a metric.

For any two r.v.s taking values on $\mathcal{X}$,

$$\mathbb{P}(X \neq Y) = \mathbb{E}[1_{X \neq Y}] \leq \mathbb{E}[\rho(X,Y)]$$

Minimizing over all couplings

$$\|\mathcal{L}(X) - \mathcal{L}(Y)\|_{TV} \leq \rho_K(\mathcal{L}(X), \mathcal{L}(Y))$$

Thm (Bubley and Dyer 1997) Suppose $\mathcal{X}$ is the vertex set of a graph equipped with length ℓ and path metric ρ. Suppose for each edge (x,y) there is a coupling (X_1, Y_1) of $P(x,\cdot)$, $P(y,\cdot)$ s.t. $\mathbb{E}_{x,y}\, \rho(X_1, Y_1) \leq \rho(x,y) e^{-\alpha}$

Then for any two distributions μ, ν on $\mathcal{X}$

$$\rho_K(\mu P, \nu P) \leq e^{-\alpha} \rho_K(\mu, \nu)$$

Corollary: $t_{mix}(\varepsilon) \leq \left\lceil \dfrac{-\log(\varepsilon) + \log(\text{diam}(\mathcal{X}))}{\alpha} \right\rceil$

$$d(t) \leq e^{-\alpha t} \text{diam}(\mathcal{X}).$$

Pf: $\rho_K(\mu P^t, \nu P^t) \leq e^{-\alpha t} \rho_K(\mu, \nu)$

Taking $\mu = \delta_x$, $\nu = \pi$ $\Rightarrow$ $d(t) \leq e^{-\alpha t} \max_{x,y} \rho(x,y)$

$$= e^{-\alpha t} \text{diam}(\mathcal{X}) \qquad \square$$

Pf of the path-coupling thm.

We first fix $\mu = \delta_x$, $\nu = \delta_y$ and show

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

Fix a path from x to y, denoted by $(x_0 = x, x_1, \dots x_r = y)$

s.t. $\sum_{i=0}^{r-1} l(x_i, x_{i+1}) = \rho(x,y)$

$$\rho_K(P(x,\cdot), P(y,\cdot)) \le \sum_{i=0}^{r-1} \rho_K(P(x_i,\cdot), P(x_{i+1},\cdot))$$

$$\cancel{\le \sum_{i=0}^{r-1} \mathbb{E}_{x,y}[}$$

$$\le \sum_{i=0}^{r-1} \mathbb{E}_{x_i, x_{i+1}} [\rho(X_1, Y_1)]$$

$$\le e^{-\alpha} \sum_{i=0}^{r-1} \rho(x_i, x_{i+1})$$

$$= e^{-\alpha} \rho(x,y)$$

For general μ, ν, we define γ be the optimal coupling s.t.

$$\rho_K(\mu,\nu) = \sum_{x,y} \gamma(x,y)\, \rho(x,y)$$

14.12 implies for any (x,y) there is $\theta_{x,y}$ of $P(x,\cdot)$ and $P(y,\cdot)$ s.t

$$\sum_{u,w} \theta_{x,y}(u,w)\, \rho(u,w) \le e^{-\alpha} \rho(x,y).$$

Now we construct a coupling between μP and νP

Set $\theta := \sum_{x,y} \gamma(x,y)\theta_{x,y}$

then
$$\sum_{u,w} \theta(u,w)\rho(u,w) = \sum_{u,w,x,y} \gamma(x,y)\, \theta_{x,y}^{(u,w)} \rho(u,w)$$

$$\le e^{-\alpha} \sum_{x,y} \gamma(x,y)\, \rho(x,y)$$

$$= e^{-\alpha} \rho_K(\mu,\nu)$$

Meanwhile θ is a coupling between μP and νP, so

$$\rho_K(\mu P, \nu P) \le e^{-\alpha} \rho_K(\mu,\nu).$$

$\square$

[One can check θ is a coupling of μP and νP.

$$\sum_w \theta(u,w) = \sum_{w,x,y} \gamma(x,y)\, \theta_{x,y}(u,w)$$

$$= \sum_{x,y} \gamma(x,y)\, P(x,u) = \sum_x \mu(x)\, P(x,u) = (\mu P)(u) \;]$$

Example (Random walk on ~~graph~~ $\{0,1\}^n$).

$V = \{0,1\}^n$, $E =$ "natural edges", $l = 1$

$\rho(x,y) = \sum_{i=1}^{n} |x_i - y_i|$.

Suppose (x,y) neighbour $x = (x_1, \cdots x_n)$

$y = (x_1, \cdots 1-x_j, x_{j+1}, \cdots x_n)$

We couple $P(x,\cdot)$ and $P(y,\cdot)$ by selecting the same index

if index $\neq j$, do the same move

index $= j$, do opposite move

$\mathbb{E}[\rho(X_1, Y_1)] = (1-\frac{1}{n})\rho(x,y) \leq e^{-\frac{1}{n}} \rho(x,y)$

$diam(\mathcal{X}) = n$, therefore

$t_{mix}(\varepsilon) \leq \lceil -\log(\varepsilon) + \log(n) \rceil \cdot n.$

Remark: The same idea can be generalized to general metric space $(\mathcal{X}, \rho)$ and there the analog is "Ollivier - Ricci Curvature".

Example (q-coloring)

Fix $G=(V,E)$, ~~code~~ q-coloring means assign a color to each vertex among q-diff colors, but neighbor must has different colors. Define a Markov chain as follows.

$\mathcal{X} = [q]^V$, given a coloring $x \in \mathcal{X}$, at each time

① Random select a vertex s

② Random update $x_s \leftarrow \{1, \cdots q\}$, if the updated color different from its neighbors, accept. Otherwise, stay.

This is essentially a M-H algorithm with Uniform distribution over all q-colorings as stationary dist.

Thm: Let $\Delta := \max_x \deg(x)$. If $q > 3\Delta$ and $c(\Delta, q) = 1 - 3\Delta/q$

then

$$t_{mix}(\varepsilon) \le \lceil c(\Delta,q)^{-1}\, n \log n + \log(1/\varepsilon) \rceil.$$

Proof: $\mathcal{X} = [q]^V$, define edge E_0 between x, y if only one color differs. Then $\rho(u, w) = \sum_{i \in V} \mathbb{1}_{u_i \neq w_i}$

For x, y neighbors $x = (x_1, \cdots x_n)$

without loss of generality $y = (y_1, x_2, \cdots x_n)$

We couple them by choosing the same index + propose the same color.

Scanned with CamScanner

① If chosen 1 $\Rightarrow$ $X_1 = Y_1$ w.p. at least $\frac{q-\Delta}{q}$

otherwise $\rho(X_1, Y_1) = 1$

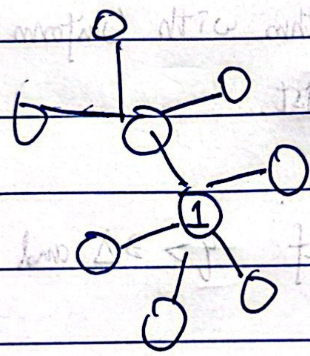
② If chosen not neighbors of 1 $\Rightarrow$ $\rho(X_1, Y_1) = 1$

③ If chosen neighbors of 1 $\Rightarrow$ $\rho(X_1, Y_1) \le 2$

For case ③ we analyze more carefully

If the color $\neq X_1, Y_1$, dist $= 1$

color $= X_1$ or Y_1 dist ≤ 2.



Therefore event: dist $= 0$ happens w.p. $\ge \frac{q-\Delta}{nq}$

event: dist $= 2$ happens w.p. $\le \frac{2\Delta}{nq}$.

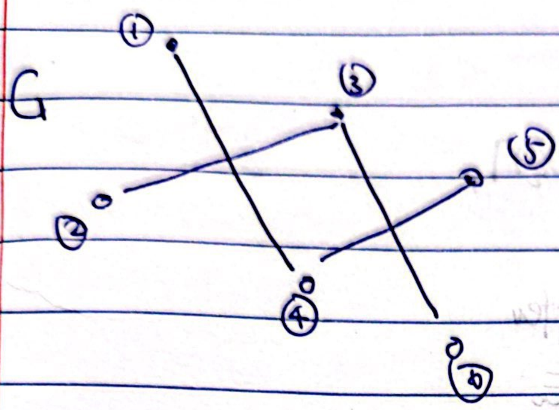
Therefore when $q > 3\Delta$, on average dist $\downarrow$

$$\mathbb{E}[\rho(X_1, Y_1)] \le 1 - \frac{q-3\Delta}{nq}$$

$\square$

Example (Independent set) Fix $G=(V,E)$

Independent set is a subset S of V s.t. no edge connecting any pair of nodes in S.



Then ①, ②, ⑤, ⑥ is an independent set.

Suppose we wish to sample from

$$\pi(X) \propto \lambda^{\sum_{v\in V} X(v)} \mathbb{1}_{X \text{ is an independent set}}$$

Here "λ" called "fugacity" $\lambda=1 \Leftrightarrow$ uniform sampling

$\mathcal{X}=\{0,1\}^V$, if $X\in\mathcal{X}$, $X_i=1 \Leftrightarrow$ put i-th vertex in the set.

Define the following "Glumber dynamics"

Given X_t

① Choose a random vertex $w\in V$

② If any neighbor of $w=1$ then set $w=0$

Otherwise $X_{t+1}(w)=\begin{cases}1 & \text{w.p. } \lambda/_{1+\lambda}\\ 0 & 1/_{1+\lambda}\end{cases}$

[Notice ② $\Leftrightarrow$ sample $X_{t+1}(w) \sim \pi(\cdot \mid X_t(-w))$].

Thm: Let $C_H(\lambda) = \dfrac{1+\lambda(1-\Delta)}{1+\lambda}$

If $\lambda < \frac{1}{\Delta-1}$ then

$t_{mix}(\varepsilon) \le \dfrac{n}{C_H(\lambda)} \lceil \lg n + \lg(1/\varepsilon) \rceil.$

Pf: $E_0 :=$ only one entry differ

$\ell(x,y) = 1$ if $\{x,y\} \in E_0$

$\rho(a,b) = \sum_{i \in V} |a_i - b_i|$.

When x, y neighbor $\quad x = (1, x_2, \cdots x_n)$

$y = (0, x_2, \cdots x_n)$.

Couple them by choosing the same vertex, then maximally couple

If choosing vertex $1 \Rightarrow X_1 = Y_1$, w.p. $1/n$

If choosing non-neighbor of $1 \Rightarrow \rho(X_1, Y_1) = 1$

If choosing neighbor of 1, call it k

Then $X_k \leftarrow 0$ w.p. 1

$Y_k \leftarrow 0$ w.p. at most $\frac{\lambda}{1+\lambda}$

so dist $= 2$ w.p. at most $\frac{|\Delta|}{n} \frac{\lambda}{1+\lambda}$

Thus

$$\mathop{\mathbb{E}}_{x,y}[\rho(X_1, Y_1)] \leq 1 - \frac{1}{n} + \frac{|\Delta|\lambda}{n(1+\lambda)}$$

$$= 1 - \frac{1+\lambda(1-\Delta)}{n(1+\lambda)}$$

$$= 1 - \frac{C_H(\lambda)}{n}$$

$\square$.

Remark: Still active research

$$\lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^{\Delta}}$$ is called "tree uniqueness threshold"

- When ~~$\lambda_c(\Delta)$~~ $\lambda < \lambda_c(\Delta)$: Fully poly time approximation scheme for the partition function, Glauber dynamics mixes $O(n \log n)$.

- When $\lambda > \lambda_c(\Delta)$, No FPTAS unless ~~$RP \neq NP$~~ $RP = NP$, Glauber has exponential mixing time

$\square$