Review

- Fundamental Thm of Markov Chains

  Ergodic $\Rightarrow$ $P^t(x,y) > 0 \ \forall x,y \Rightarrow$

  P^t is contracting in d_{TV}

- Mixing Time

  $t_{mix}(P,\nu,\varepsilon) := \min\{t \mid d_{TV}(\nu P_t, \mu) \leq \varepsilon\}$

  $t_{mix}(P,\varepsilon) := \max\{t_{mix}(P,\nu,\varepsilon) \mid \nu\}$

  $t_{mix}(P,\varepsilon) \leq t_{mix}(P,\tfrac{1}{3}) \cdot O(\lg \tfrac{1}{\varepsilon})$

  usually not tight (cutoff)

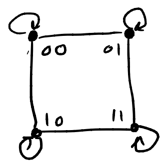
- Strong Stationary Time:

  $X_0 \xrightarrow{P} X_1 \rightarrow \cdots$

  τ is strong stationary if $X_\tau \sim \mu$ even conditioned on $\tau = t$.

  $\mathbb{P}[\tau > t] \leq \varepsilon \Rightarrow t_{mix}(\varepsilon) \leq t$

  Hypercube walk mixes in $\leq n \lg n + n \lg \frac{1}{\varepsilon}$



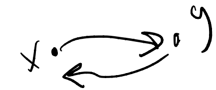
- Ergodic Flow: $Q(x,y) = \mu(x) P(x,y)$

  Stationary $\Leftrightarrow$ Q is proper flow

  time reversal: flip directions in Q

- Time-Reversible: $Q(x,y) = Q(y,x)$

- Designing Chains



  1. Metropolis Rule:

     For each x,y, take direction with larger ergodic flow and make excess self-loop

  2. Combination with Time-Reversal

Plan for Today

- Finish construction 2
- Transport Distance
- Colorings + Path Coupling
- Dobrushin's Conditions

Quick remark on Stationary Times

$\Omega = [n] \times [n]$

$P: (x, y) \longmapsto (\text{u.r. } x', \; y+1 \bmod n)$

periodic but you can add self loops to y

stationary μ: uniform on $[n] \times [n]$

τ: Observe x after first step of chain and stop anytime after when y hits that value.

$$(x_0, y_0) \longmapsto (x_1, y_1) \longmapsto \cdots \longmapsto (x_\tau, y_\tau)$$

equal.

$(X_\tau, y_\tau) \sim$ uniform $= \mu$ and $\tau \leq n$ a.s.

But no mixing!

② Combination with time-reversal

- Distribution μ on Ω and row-stochastic Operator $N \in \mathbb{R}_{\geq 0}^{\Omega \times \Omega'}$ ← because $\Omega' \neq \Omega$, this is not necessarily a Markov chain.

  think "noise"

- Together

$$\underbrace{\Pi(x, y)}_{\text{the "ergodic flow" equivalent}} := \mu(x) N(x, y)$$

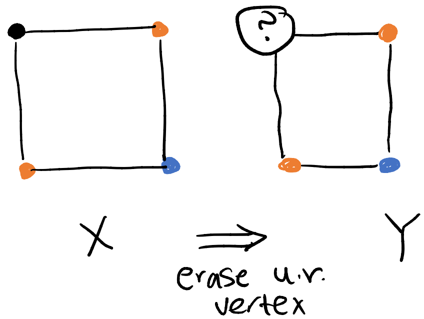
Relation between Π and N: N is the conditional dist obtained from Π.

Time-reversal: Change $x \mapsto y$ to $y \mapsto x$

$$N^{\circ}(y, x) = \frac{\Pi(x, y)}{\sum_{x'} \Pi(x', y)}$$ ← conditional dist of $x \mid y$

Construction: $P = N N^{\circ}$

Example. (Glauber Dynamics)

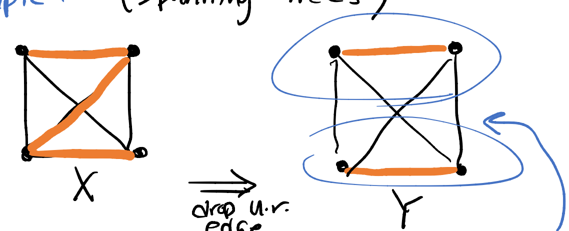


Chain: erase, then recolor erased vertex w. prob

$$\propto \mu(\text{resulting config}) \times \cancel{\mathbb{P}[\text{config} \mapsto \text{erased config}]}$$

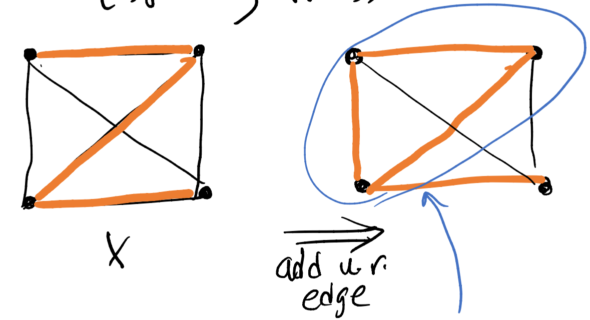
Bayes Rule

Example. (Spanning Trees)



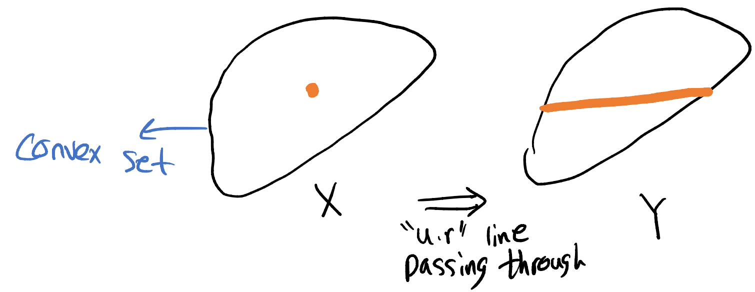
Chain: drop, then add u.r. edge from the cut

Example. (Spanning Trees)



Chain: Drop u.r. from cycle

Example. (Hit-and-Run) ← non-finite Ω



Chain: Draw line & go to u.r. point on line segment.

Example. (Block Dynamics)

Chain: Drop ℓ u.r., resample those ℓ colors.

Example. (Full Noise) $X \xrightarrow{N} Y$

Suppose X and Y are independent.

E.g. $\ell = n$ in block dynamics.

Chain: Sample directly from μ.

Note the chain has correct stationary and mixes perfectly in 1 step!

Remark: We typically want N to have sparse-ish columns:

$$\#\{x \mid N(x,y) > 0\} = \mathrm{poly}\lg(|\Omega|)$$

Naive Implementation:

- Algorithm to do $X_0 \xrightarrow{N} Y$
- Efficient enumeration of all x' with $N(x',Y) > 0$. need small #
- Compute $\mu(x')N(x',Y)$, normalize and sample.

up to proportionality

Fact: NN^0 is always time-reversible

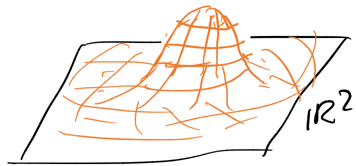
$$\overbrace{\mu(x_0)\,P(x_0,x_1)}^{\text{ergodic flow}} = \mu(x_0)\sum_y N(x_0,y)\,N^0(y,x_1)$$

$$= \sum_y \frac{\pi(x_0,y)\;\pi(x_1,y)}{\sum_{x'} \pi(x',y)}$$

clearly symmetric in X_0, X_1

Sketchy Example (Restricted Gaussian Oracle & Langevin Dynamics) ← if time, you don't need to understand fully

$\Omega = \mathbb{R}^n$ $\quad \mu$: density ← w.r.t. Lebesgue



N: take $x \longmapsto y := x + \underbrace{\mathcal{N}(0, \varepsilon I)}_{\text{add Gaussian noise}}$

N°: $y \longrightarrow x'$ with prob:

$$\propto \underbrace{\mu(x') \exp\left(-\frac{1}{2}\frac{\|x'-y\|^2}{\varepsilon}\right)dx'}_{\text{"Restricted Gaussian" [Lee-Shen-Tian'20]}}$$

Chain: Add noise, then sample from restricted Gaussian.

still hard, but smoother density

What happens for tiny ε?

$$\mu(x') \exp\left(-\frac{1}{2}\frac{\|x'-y\|^2}{\varepsilon}\right) \simeq$$

$O(\|x'-y\|^2)$

$$\exp\left(c_0 + \nabla \lg \mu(y) \cdot (x'-y) - \frac{1}{2}\frac{\|x'-y\|^2}{\varepsilon} + \cdots\right) =$$

$$\exp\left(-\frac{1}{2}\frac{\|x'-y-\varepsilon\nabla \lg \mu(y)\|^2}{\varepsilon} + c_1 + \cdots\right)$$

this is roughly $\mathcal{N}(y + \varepsilon \nabla \lg \mu(y), \varepsilon I)$

Chain: Add $\mathcal{N}(0, \varepsilon I)$, move by $\varepsilon \nabla \lg \mu$, add $\mathcal{N}(0, \varepsilon I)$.

In the $\varepsilon \to 0$ limit:

$$dX_t = \nabla \lg \mu(X_t) \cdot dt + \sqrt{2}\, dW_t$$

Don't worry if notation is unfamiliar.

Brownian motion

This is Langevin dynamics.

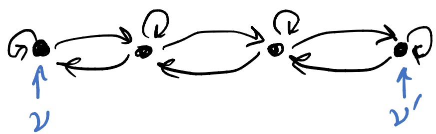
Similar mechanisms used in score-based models.

Transport distance

- General strategy to bound mixing time:

  Contraction

- d_{TV} is too crude to contract in one step.



Idea: Use other measure of distance as proxy for d_{TV}.

- Today: Transport / Earth-Mover / Wasserstein
- Next: $\underbrace{\text{Divergences / Variance / Entropy}}_{\text{functional analysis}}$

If Ω has a metric $d: \Omega \times \Omega \to \mathbb{R}_{\geq 0}$:

$\hookrightarrow \begin{cases} d(x,y) = d(y,x),\ d(x,x) = 0 \\ d(x,z) \leq d(x,y) + d(y,z) \end{cases}$

We can define a transport distance:

$$W(\nu,\nu') := \min\{\mathbb{E}[d(X,Y)] \mid X,Y \text{ coupling}\}$$

Remark: $W_p := \min\{\mathbb{E}[d(X,Y)^p]^{\frac{1}{p}} \mid X,Y \text{ coupling}\}$

but we will deal with $p=1$.

Example. (total variation)

Suppose $d(x,y) = \begin{cases} 0 & \text{if } x=y \\ 1 & \text{if } x \neq y \end{cases}$. Then

$$W(\nu,\nu') = d_{TV}(\nu,\nu')$$

Example. (Hamming distance)

$\Omega = [q]^n \quad d((x_1,\dots,x_n),(y_1,\dots,y_n)) = \#\{i \mid x_i \neq y_i\}$

ν: unif on $\{(\bullet,\bullet,\bullet), (\bullet,\bullet,\bullet)\}$

ν': unif on $\{(\bullet,\bullet,\bullet), (\bullet,\bullet,\bullet), (\bullet,\bullet,\bullet)\}$

$\frac{1}{6}$ $\frac{1}{3}$

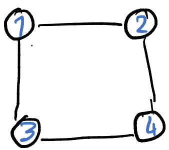
$$W(\nu,\nu') = \frac{1}{3}\cdot 0 + \frac{1}{6}\cdot 3 + \frac{1}{2}\cdot 2 = 1.5$$

Coloring

Input: Graph G, $q \in \mathbb{N}$ ← palette size

Task: Sample / count proper q-colorings
adjacent ⇒ not equal

- NP-hard to even find one!
- Easy when $q \geq \Delta + 1$ → maximum degree



* Pick arbitrary ordering
* Color one at a time
* Always ≥ 1 choice left

Open: Approx sample / count when $q \geq \Delta + 1$.
Open: Glauber/Metropolis mix fast for $q \geq \Delta + 2$.

Best known: $q > \left(\frac{11}{6} - \varepsilon\right)\Delta$ for $\Delta \geq \Delta_0$
ε: some tiny number.
[Chen–Delcourt–Moitra–Perarnau–Postle '19]

We will analyze the Metropolis chain P:

- Pick vertex v and color c u.a.r.
- If recoloring v to c is proper accept.

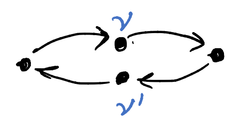
Fact: If $q \geq 4\Delta + 1$, P contracts Wasserstein distance.

$$W(\nu P, \nu' P) \leq \left(1 - \frac{1}{\text{poly}}\right) W(\nu, \nu')$$

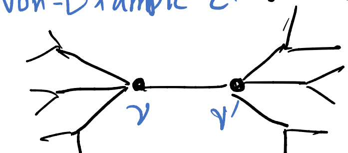
- Remark: $W(\nu, \nu') \geq d_{TV}(\nu, \nu') \min\{d(x,y) \mid x \neq y\}$
So bounding W, bounds d_{TV}. 1 for Hamming

- Remark: Weak contraction does NOT necessarily hold for W.

Non-Example 1:



Non-Example 2: (tree)



Coupling. $X_0 \to X_1 \to \cdots$

$X_0' \to X_1' \to \cdots$

- Each a faithful run of Markov chain
- Arbitrary cross-dependence.

Strategy: For deterministic X_0, X_0' find one-step couplings s.t.

$$\mathbb{E}[d(X_1, X_1')] \leq (1-c) \cdot d(X_0, X_0').$$

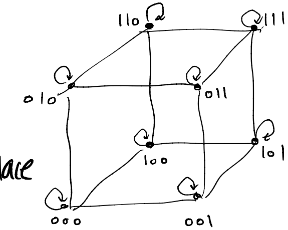
$$\Rightarrow W(\nu P, \nu' P) \leq (1-c) W(\nu, \nu')$$

- Sample X_0, X_0' from coupling of ν, ν'
- Run the X_0, X_0'-specific coupling to get X_1, X_1'.

$$\Rightarrow W(\nu P^t, \nu' P^t) \leq (1-c)^t W(\nu, \nu')$$

Warmup Example. (Hypercube)

- Start with X_0, X_0'
- Pick same index and same $\mathrm{Ber}(\frac{1}{2})$ to replace that index.



$$d(X_0, X_0') = i \Rightarrow \mathbb{E}[d(X_1, X_1')] = i - \frac{i}{n} = (1-\frac{1}{n}) i.$$

$\downarrow$ # diff coords

$$W(\nu P, \nu' P) \leq (1-\frac{1}{n}) W(\nu, \nu')$$

$$\Rightarrow d_{TV}(\nu P^t, \mu) \leq W(\nu P^t, \mu) \leq (1-\frac{1}{n})^t W(\nu, \mu) \leq n \cdot (1-\frac{1}{n})^t$$

$$\Rightarrow t_{mix}(\varepsilon) \leq O(n \lg n + n \lg \frac{1}{\varepsilon})$$

Coupling for Colorings

- Take $X_0 \in [q]^n$, $X'_0 \in [q]^n$
- Pick same proposed v, c for both.
- $X_1 = X_0$ with v recolored to c if valid.
- $X'_1 = X'_0$ with v recolored to c if valid.

Analysis: $d := d(X_0, X'_0)$

- Good move: $d \mapsto d-1$
- Bad move: $d \mapsto d+1$
- Neutral move: $d \mapsto d$

$\# \text{Good} \geq d \times (q - 2\Delta)$

differing v common available colors

$\# \text{Bad} \leq 2d\Delta$ → num neighbors.

→ c must be color of neighbor with differing color in X_0, X'_0.

$$\mathbb{E}[d(X_1, X'_1)] \leq d - \frac{d(q-2\Delta)}{qn} + \frac{2d\Delta}{qn} = d\left[1 - \frac{q-4\Delta}{qn}\right]$$

$$t_{mix} \leq O\left(\frac{qn}{q-4\Delta} \cdot \lg n\right) \text{ for } q \geq 4\Delta + 1.$$

Path Coupling [Bubley – Dyer]

Hamming distance is special:

* There is a sparse graph on Ω s.t.

possibly weighted

$d(x,y)$ = shortest path in graph.

The graph: $x \sim y$ if x and y differ in exactly one coordinate.

Idea: Only come up with coupling for X_0, X'_0 adjacent in the graph, s.t.

$$\mathbb{E}[d(X_1, X'_1)] \leq d(X_0, X'_0) \cdot (1 - c)$$

Coupling for adjacent pairs $\Rightarrow$ Coupling for all

This is because W has triangle inequality

HW: Show this!

- Take shortest path $X_0 = X_0^{(0)}, X_0^{(1)}, \dots, X_0^{(\ell)} = X_0'$.

$$W\left(1_{X_0^{(i)}} P, 1_{X_0^{(i+1)}} P\right) \leq d\left(X_0^{(i)}, X_0^{(i+1)}\right)(1-c)$$

$$\Rightarrow \sum_i W\left(1_{X_0^{(i)}} P, 1_{X_0^{(i+1)}} P\right) \leq \sum_i d\left(X_0^{(i)}, X_0^{(i+1)}\right)(1-c) = d(X_0, X_0')(1-c)$$

triangle ineq.

$$\Rightarrow W\left(1_{X_0} P, 1_{X_0'} P\right) \leq d(X_0, X_0')(1-c)$$

Triangle inequality holds because couplings on a path (generally a tree) can be stitched together to get a global coupling.

HW prob

Path Coupling for Colorings [Jerrum]

Take adjacent colorings X_0, X_0'.

$X_0(w) = a \qquad X_0'(w) = b \qquad (a \neq b)$

w: differing vertex.

Coupling:

- Pick same vertex v.
- If $v = w$, or $v \not\sim w$, pick same c.
- If $v \sim w$:

good / bad

c: $1, 2, \dots, a, \dots, b, \dots, q$

c': $1, 2, \dots, a, \dots, b, \dots, q$

$$\mathbb{E}[d(X_1, X_1')] \leq 1 - \frac{1}{n} \cdot \frac{q-\Delta}{q} + \frac{\Delta}{n} \cdot \frac{1}{q}$$

picking w

$$= 1 - \frac{q - 2\Delta}{nq}.$$

As long as $q \geq 2\Delta + 1 \Rightarrow t_{mix} = O\left(\frac{nq}{q-2\Delta} \cdot \lg n\right)$