PATH COUPLING FOR RAPID MIXING OF MARKOV CHAINS

ZHIDAN LI

1. Path Coupling

For a Markov chain $\mathcal{M}$, in the classical coupling method, given a metric d on a state space Ω, if a coupling decreases the distance between every pair of configurations in Ω, then the mixing time of $\mathcal{M}$ can be bounded. The following concepts formalize this argument.

Definition 1.1 (Contraction)**.** Given a metric d on a state space Ω and a Markov chain $\mathcal{M}$ on Ω with stationary distribution μ, we say that a coupling $(X, Y) \to (X', Y')$ satisfies γ-contraction for some factor γ if for every initial configurations $(X, Y) \in \Omega \times \Omega$,

$$\mathbf{E}\left[d(X', Y') \mid X, Y\right] \leq \gamma d(X, Y).$$

Theorem 1.2 (Coupling theorem)**.** *For some factor $\gamma \in [0, 1]$, if there exists a coupling satisfying γ-contraction, then*

$$\tau_{\text{mix}}(\mathcal{M}) \leq O\left(\frac{1}{1-\gamma} \log d_{\max}\right)$$

where $d_{\max}$ is the diameter of the metric d.

However, defining distances and couplings between all configurations in Ω is hard. The path coupling theorem allows us to determine distances and coupling between some pairs of configurations, and the whole metric and coupling can be naturally extended.

Definition 1.3 (Pre-metric)**.** A *pre-metric* on Ω is a pair (Γ, ω) where Γ is a connected, undirected graph on Ω and ω is a positive real-valued function assigning values to edges (X, Y) in Γ satisfying that for every edge (X, Y), $\omega(X, Y)$ is the minimum among all paths between X and Y. We refer to these adjacent vertices as *neighboring pairs*.

Note that from this pre-metric, we can naturally construct a metric d on Ω using the shortest paths.

Theorem 1.4 (Path coupling theorem)**.** *Let (Γ, ω) be the pre-metric in Ω and d be the induced metric. If a coupling defined on the edges in Γ satisfies γ-contraction for some $\gamma \in [0, 1]$, then there exists a coupling on Ω satisfying γ-contraction. Therefore,*

$$\tau_{\text{mix}}(\mathcal{M}) \leq O\left(\frac{1}{1-\gamma} \log d_{\max}\right)$$

where $d_{\max}$ is the diameter of the metric d.

2. Application: Vigoda's Algorithm for Proper Colorings

We show the application of Theorem 1.4 to sampling proper colorings by Vigoda [Vig00]. Given a graph $G = (V, E)$ and an integer $q \geq 2$, let Ω be all (not necessarily proper) q-colorings on G.

Before we introduce the Markov chain applied, there are some related concepts. Given a coloring $X \in \Omega$, we say a path $v = v_0, v_1, \ldots, v_t = w$ is an *alternating path* between v and w using c and $X(v)$ if $(v_i, v_{i+1}) \in E$, $\sigma(v_i) \in \{X(v), c\}$ and $X(v_i) \neq X(v_{i+1})$. Then the *Kempe component* $S_X(v, c)$ is the following cluster of vertices

$$S_X(v, c) := \{w \in V \mid \text{there exists an alternating path between } v \text{ and } w \text{ using colors } \sigma \text{ and } c\}.$$

For convenience, we redefine $S_X(v, X(v)) = \emptyset$. For every vertex $w \in S_X(v, c)$, it holds that $S_X(v, c) = S_X(w, c)$ if $X(v) = X(w)$ and $S_X(v, c) = S_X(w, X(v))$ otherwise. This means that every Kempe component S can be relabelled in $|S|$ ways. Let $\mathcal{S}_X$ be the set of all Kempe components induced by X.

Now we introduce the *flip dynamics* $\mathcal{M}_{\texttt{FD}}$ to sample proper colorings. Given a sequence of weights $\{p_i\}_{i\geq 0}$ satisfying $p_1 = 1$, at a proper coloring X, we run transition in the following way:

- Choose $v \in V$ and $c \in [q]$ uniformly at random.
- Let $\alpha = |S_X(v,c)|$. With probability $p = \frac{p_\alpha}{\alpha}$, we flip cluster $S_X(v,c)$ by interchanging colors c and $X(v)$ in the cluster.

Note that for a cluster S, there are $|S|$ different pairs of (v,c) to choose S. So the probability of flipping S is exactly $p_{|S|}$. Then we have the following equivalent way to describe $\mathcal{M}_{\texttt{FD}}$.

- Choose a Kempe component $S \in \mathcal{S}_X$ with probability $1/nq$.
- Let $\alpha = |S|$ and with probability p_α flip S.

It is not hard to verify that $\mathcal{M}_{\texttt{FD}}$ is irreducible and aperiodic. It is not hard to verify that $\mathcal{M}$ is stationary with respect to the uniform distribution of proper q-colorings on G.

2.1. **Coupling of the flip dynamics.** To apply Theorem 1.4, we construct a coupling for every $(X,Y) \in \Omega\times\Omega$ such that X and Y differ at exactly one vertex $v \in V$. We consider when clusters $S_X(w,c), S_Y(w,c)$ might be different in the sense that $S_X(w,c) \neq S_Y(w,c)$ or $S_X(w,c) = S_Y(w,c)$ but there is a vertex y in this with $X(y) \neq Y(y)$.

Let $\mathcal{D} = \mathcal{D}(X,Y)$ be the collection of clusters that are different in X, Y. Note that these clusters must involve v. Then we know that

$$\mathcal{D} := \{S_X(v,c) : c \in [q]\} \cup \{S_Y(v,c) : c \in [q]\} \cup \{S_X(w,Y(v)), S_Y(w,X(v)) : w \in N_G(v)\}.$$

For every Kempe component $S \notin \mathcal{D}$, we use the identity coupling for its move and this does not change the distance. So we only consider $\mathcal{D}$.

We decompose $\mathcal{D}$ in sets $\cup_{c\in[q]}\mathcal{D}_c$ where $\mathcal{D}_c$ is the set of Kempe components consisting of $S_X(v,c), S_Y(v,c)$ and $S_X(w,Y(v)), S_Y(w,X(v))$ for all $w \in N_G(v)$ satisfying $X(w) = Y(w) = c$.

We use the Hamming distance denoted by $H(\cdot,\cdot)$ as the metric d. For any $X \in \Omega$ and $S \in \mathcal{D}$, let $X_{\oplus S}$ be the coloring obtained from X after flipping S. Then we know that

$$\begin{aligned}\mathbf{E}[\Delta H \mid X,Y] &= \mathbf{E}[\Delta H \mid X,Y,S\notin\mathcal{D}]\,\mathbf{Pr}[S\notin\mathcal{D}\mid X,Y] + \sum_{c\in[q]}\mathbf{E}[\Delta H \mid X,Y,S\in\mathcal{D}_c]\,\mathbf{Pr}[S\in\mathcal{D}_c\mid X,Y]\\ &= \frac{1}{nq}\sum_{c\in[q]}\sum_{S\in\mathcal{D}_c}\mathbf{E}[H(X_{\oplus S},Y_{\oplus S}) - H(X,Y)\mid X,Y].\end{aligned}$$

Let U_c be the set of neighbors of v that are colored c. Let $\delta_c = |U_c|$. We denote $U_c = \{u_1^c,\dots,u_{\delta_c}^c\}$ or simply $\{u_1,\dots,u_{\delta_c}\}$ when c is clear. Then

$$\mathcal{D}_c = \{S_X(v,c), S_Y(v,c)\} \cup \left(\bigcup_{w\in U_c}\{S_X(v,Y(v)), S_Y(v,X(v))\}\right).$$

We mark that sets in $\mathcal{D}_c$ are disjoint except possibly $\mathcal{D}_{X(v)}$ and $\mathcal{D}_{Y(v)}$. If $c \notin \{X(v),Y(v)\}$, we obtain that

$$S_X(v,c) = \left(\bigcup_{i=1}^{\delta_c} S_Y(u_i^c, X(v))\right)\cup\{v\}, \quad S_Y(v,c) = \left(\bigcup_{i=1}^{\delta_c} S_X(u_i^c, Y(v))\right)\cup\{v\}.$$

For $c = X(v)$, we have $S_X(v,c) = S_Y(u,X(v)) = \emptyset$ for all $u \in U_c$. Similarly for $c = Y(v)$, $S_Y(v,c) = S_X(u,Y(v)) = \emptyset$ for all $u \in U_c$.

The following observation will simplify some cases in our analysis. Note that v can have some neighbors $u_1',\dots,u_m' \in N_G(v)$ colored c belonging to the same Kempe component $S_Y(u_1',X(v)) = \cdots = S_Y(u_m',X(v))$. In order to consider the flip with the right probability, we redefine $S_Y(u_i',X(v)) = \emptyset$ for $1 < i \leq m$. Do the same modifications for $S_X(u_i',Y(v))$.

For each $c \in [q]$ such that $\delta_c > 0$, define $A_c := |S_X(v,c)|$, $B_c := |S_Y(v,c)|$, $a_i^c := |S_Y(u_i,X(v))|$ and $b_i^c := |S_X(u_i,Y(v))|$. We define the vectors $\boldsymbol{a}^c := (a_i^c : i \in [\delta_c])$ and $\boldsymbol{b}^c := (b_i^c : i \in [\delta_c])$. We say that (X,Y) has configuration $(A_c,B_c;\boldsymbol{a}^c,\boldsymbol{b}^c)$. We also define $a_{\max}^c := \max_i a_i^c$ and $i_{\max}^c$ as a maximizing argument. Similarly

define $b^c_{\max} := \max_j b^c_j$ and $j^c_{\max}$ as a maximizing argument. When it is clear from the context, we drop the script c. Note that the following inequality holds:

$$A \le 1 + \sum_i a_i, \quad B \le 1 + \sum_j b_j$$

with equality if $c \notin \{X(v), Y(v)\}$.

The idea of coupling consists of the following rules. Flips of clusters in $\mathcal{D}_c$ for X will be coupled with clusters in $\mathcal{D}_c$ for Y. We couple $S_X(v,c)$ and $S_Y(v,c)$ with the biggest size of others, and try to couple the remaining weights as much as possible.

- Flip $S_X(v,c)$ and $S_Y(u_{i_{\max}}, X(v))$ together with probability p_A.
- Flip $S_Y(v,c)$ and $S_X(u_{j_{\max}}, Y(v))$ together with probability p_B.
- For all $i \in [\delta_c]$, let $q_i = p_{a_i} - p_A \cdot \mathbb{1}[i_{\max} = i]$ and $q'_j = p_{b_j} - p_B \cdot \mathbb{1}[j_{\max} = j]$.
  (1) Flip $S_Y(u_i, X(v))$ and $S_X(u_i, Y(v))$ together with probability $\min(q_i, q'_i)$.
  (2) Flip $S_Y(u_i, X(v))$ with probability $q_i - \min(q_i, q'_i)$.
  (3) Flip $S_X(u_i, Y(v))$ with probability $q'_i - \min(q_i, q'_i)$.

Given a configuration $(A, B; \boldsymbol{a}, \boldsymbol{b})$, define $H(A, B; \boldsymbol{a}, \boldsymbol{b}) := (A - a_{\max} - 1)p_A + (B - b_{\max} - 1)p_B + \sum_i (a_i \cdot q_i + b_i \cdot q'_i - \min\{q_i, q'_i\})$.

Proposition 2.1. *The following bound holds*

$$\mathbf{E}[\Delta H \mid X, Y] \le \frac{1}{nq}\left(-|\{c : \delta_c = 0\}| + \sum_{c\,:\,\delta_c > 0} H(A_c, B_c; \boldsymbol{a}^c, \boldsymbol{b}^c)\right).$$

2.2. **Linear programming and choice of flip weights.** In order to obtain the rapid mixing of Markov chains, we need to choose proper weights $\{p_\alpha\}_{\alpha \in \mathbb{N}}$.

The variation depends sorely on the configurations.

Definition 2.2. A configuration $(A, B; \boldsymbol{a}, \boldsymbol{b})$ is *realizable* if there exists a graph G, a neighboring coloring pair (X, Y) defined in G and a color $c \in [q]$ such that $(A, B; \boldsymbol{a}, \boldsymbol{b}) = (A_c, B_c; \boldsymbol{a}^c, \boldsymbol{b}^c)$.

We mark here that a configuration $(A, B; \boldsymbol{a}, \boldsymbol{b})$ is realizable if and only if

$$A \le 1 + \sum_i a_i, \quad B \le 1 + \sum_j b_j.$$

We call δ_c the size of the configuration.

Note that if there exists $\lambda > 0$ such that $H(A, B; \boldsymbol{a}, \boldsymbol{b}) \le -1 + \lambda m$ for all realizable configurations $(A, B; \boldsymbol{a}, \boldsymbol{b})$ where m is the size of the configuration, then we know that the coupling is contractive for $q > \lambda\Delta$.

Then our goal is to solve the following linear programming.

$$\begin{aligned} \min_{\lambda, \{p_\alpha\}_{\alpha\in\mathbb{N}}} \quad & \lambda \\ \text{subject to} \quad & H(A, B; \boldsymbol{a}, \boldsymbol{b}) \le -1 + \lambda m \qquad \forall m \in \mathbb{N} \text{ and all realizable } (A, B; \boldsymbol{a}, \boldsymbol{b}) \text{ of size } m, \\ & p_0 = 0 \le p_i \le p_{i-1} \le p_1 = 1 \qquad \forall i \ge 2. \end{aligned} \tag{1}$$

However, this linear program is hard to solve since there are infinitely many variables and constraints. To solve this problem, Vigoda restricts that for every $\alpha \ge 7$, $p_\alpha = 0$.

The following bounds make the linear program easy to solve.

Lemma 2.3. $H(A, B; \boldsymbol{a}, \boldsymbol{b}) \le (A - 2a_{\max})p_A + (B - 2b_{\max})p_B + \sum_i (p_{a_i} a_i + p_{b_i} b_i - \min\{p_{a_i}, p_{b_i}\})$.

Lemma 2.4. *Consider for all i the additional constraints $ip_i \le 1$, $(i-1)p_i \le \frac{1}{3}$ and $(i-2)p_i \le 2/9$. Let $(A, B; \boldsymbol{a}, \boldsymbol{b})$ be a realizable configuration of size ≥ 3. Then if $\{p_\alpha\}$ satisfy the additional constraints, then for $\lambda \ge \frac{49}{27}$,*

$$H(A, B; \boldsymbol{a}, \boldsymbol{b}) \le -1 + \lambda m.$$

Then we can solve the linear program $\lambda^* = 11/6$ and a feasible solution is

$$p_1 = 1, p_2 = \frac{13}{42}, p_3 = \frac{1}{6}, p_4 = \frac{2}{21}, p_5 = \frac{1}{21}, p_6 = \frac{1}{84}.$$

3. Potts Models and Swendsen-Wang Process

Fix a positive integer $q \geq 2$ and a parameter $\beta \in \mathbb{R}$. We consider q-state Potts model on an arbitrary graph $G = (V, E)$ with $V = \{v_1, \ldots, v_n\}$. For any assignment $\sigma : V \to [q]$, we call σ a *configuration* and σ gives a partition $V = V_1 \cup \cdots \cup V_q$ to V where $V_i := \{v \in V \mid \sigma(v) = i\}$ for every $i \in [q]$. Under a given configuration σ, we say an edge $e = (u, v) \in E$ is a *bond* if $\sigma(u) = \sigma(v)$. The components induced by bonds are called *clusters*.

For a configuration $\sigma \in [q]^V$, define $D(\sigma) := \{e = (u, v) \in E : \ \sigma(u) \neq \sigma(v)\}$ and certainly $B(\sigma) = E \setminus D(\sigma)$ as the bonds under σ. Then we define the *q-state Potts models* $\mu = \mu_{G,\beta}$ on G at β as

$$\mu(\sigma) = \frac{e^{-\beta|D(\sigma)|}}{Z}, \quad \forall \sigma \in [q]^V \tag{2}$$

where the normalising factor $Z = \sum_{\sigma \in [q]^V} e^{-\beta|D(\sigma)|}$ is called *the partition function* of the Potts model. When $\beta > 0$, we say it is *ferromagnetic* and otherwise we call it an *anti-ferromagnetic* one.

3.1. Markov chains and Swendsen-Wang process.

A long-time topic in practice is how to sample a configuration σ from μ. Variants of Markov chains are in application. However, a typical Metropolis process in [JS96] is shown to work in some settings under the anti-ferromagnetic assumption, but unknown in the ferromagnetic cases. To avoid this accident, we apply the following dynamics named *Swendsen-Wang process*.

Assume that the current configuration is σ. We generate the next configuration by the following two steps.

(1) Let $B = B(\sigma)$ be the bonds under σ. We sample a subset $A \subseteq B$ as: for each edge $e \in B$, we pick it with probability $p = 1 - e^{-\beta}$.
(2) For each connected component in the graph (V, A), we assign all vertices in the component with a state in $[q]$ uniformly and independently at random.

The application of the Swendsen-Wang process depends on the equivalence under certain conditions of the q-state Potts model and the *random cluster model*. Given a graph $G = (V, E)$, for an edge subset $A \subseteq E$ define $G(A) := (G, A)$ as the subgraph induced by the edge set A. In the random cluster model, we regard A as bonds and define its weight as

$$w(A) = p^{|A|}(1 - p)^{|E \setminus A|} q^{c(A)}$$

where $c(A)$ is the number of connected components of $G(A)$ and p is a probability.

Now we define the following joint distribution π between the Potts model and the random cluster model on the space $[q]^V \times 2^E$. Let $p = 1 - e^{-\beta}$. For every $\sigma \in [q]^V$ and $A \subseteq E$, define the *Edwards-Sokal measure* as

$$\pi(\sigma, A) := \frac{1}{Z} \prod_{(u,v) \in E} \left(p \cdot \mathbb{1}\left[e \in A \wedge \sigma(u) = \sigma(v)\right] + (1 - p) \cdot \mathbb{1}\left[e \notin A\right] \right)$$

where Z is the normalising factor. If we use the notation $\sigma \sim A$ to denote the event that every edge in A has its two endpoints with the same spin in σ, we write π equivalently as

$$\pi(\sigma, A) = \frac{1}{Z} p^{|A|}(1 - p)^{|E| - |A|} \mathbb{1}\left[\sigma \sim A\right].$$

Summing over σ or A we can see the marginal distributions are the Potts model or the random cluster model respectively. We remark here that all the normalising factors are equal.

With the joint distribution π, we provide another view of the Swendsen-Wang dynamics: given a configuration σ, firstly pick A according to $\pi(\sigma, \cdot)$; and generate a new configuration σ' according to $\pi(\cdot, A)$.

4. Rapid Mixing on Potts Models

The following is the main theorem of this section.

Theorem 4.1. *Let G be a graph with constant maximum degree Δ. Then there exists $p_0 = p_0(\Delta)$ such that if $p = 1 - e^{-\beta} \leq p_0$, then the Swendsen-Wang process mixes rapidly for every q.*

We prove Theorem 4.1 by the path coupling argument. For two configurations X_t, Y_t, let

$$S_t := \{v \in V \mid X_t(v) = Y_t(v)\}$$

and $D_t = V \setminus S_t$. Following the idea of path coupling, we might assume that $|D_t| = 1$ and $D_t = \{v_t\}$. Let $G(X_t)$ (respectively $G(Y_t)$) denote the subgraph of G induced by the bonds of X_t (respectively Y_t). We couple X_t and Y_t in the following way.

Recall the Swendsen-Wang process. It consists of two stages: the *bond breaking* and the *component coloring*.

Bond Breaking. We consider the edge $e \in E$.

- If $e \subseteq S_t$ is a bond, then it is a bond in both $G(X_t)$ and $G(Y_t)$. We keep e in both graphs with probability $p = 1 - e^{-\beta}$ and delete it in both graphs with probability $1 - p$.
- For any edge $e = (v_t, w)$, it can only be a bond in exactly one graph. Then we keep it in that graph with probability p.

Let $\widehat{X}_t$ and $\widehat{Y}_t$ be configurations after this stage and $G(\widehat{X}_t), G(\widehat{Y}_t)$ be the subgraphs induced by bonds in $\widehat{X}_t$ and $\widehat{Y}_t$ respectively.

Component Coloring. Let H be the subgraph of $G(\widehat{X}_t)$ (and $G(\widehat{Y}_t)$) induced by S_t. If C is a component of H which is not adjacent to v_t, then we give C the same random color in X_{t+1} and Y_{t+1}.

Suppose that v_t is adjacent to components $C_1, C_2, \ldots, C_r$ of H in $G(\widehat{X}_t)$ and adjacent to $D_1, \ldots, D_s$ in $G(\widehat{Y}_t)$. Note that $\bigcup_{i=1}^{r} C_i$ and $\bigcup_{j=1}^{t} D_j$ are disjoint, otherwise v_t would have the same color in both X_t and Y_t.

(1) $r = s = 0$. Give v_t the same random color in X_{t+1} and Y_{t+1}.

(2)
- $r = 1, s = 0$. Give C_1 the same random color c in X_{t+1} and Y_{t+1}. Give v_t color c in X_{t+1} and a random color in Y_{t+1}.
- $r = 0, s = 1$. Give D_1 the same random color c in X_{t+1} and Y_{t+1}. Give v_t color c in Y_{t+1} and a random color in X_{t+1}.

(3) $r = 1$ and $s = 1$. Give C_1 a random color c in X_{t+1} and D_1 a random color d in Y_{t+1}. The vertex v_t inherits its color in from C_1 in X_{t+1} and from D_1 in Y_{t+1}.

(4) $r \geq 2$ or $s \geq 2$. Let B be the largest component of $C_1, \ldots, C_r, D_1, \ldots, D_s$. Give B the same random color c in both X_{t+1} and Y_{t+1}. Any component of $G(\widehat{X}_t)$ or $G(\widehat{Y}_t)$ not inheriting this color c is colored randomly.

Let p_i denote the probability of Case (i) ($1 \leq i \leq 4$) and let δ_4 be the indicator of Case (4). Note that given the rule of the Bond Breaking, an edge e of $G(X_t) \cup G(Y_t)$ appears independently in $\Gamma = G(\widehat{X}_t) \cup G(\widehat{Y}_t)$ with probability p. Hence $p_1 \geq (1-p)^{\Delta}$.

In Cases (2)-(4), the vertex v_t has the same color in X_{t+1} and Y_{t+1} with probability $1/q$. In Case (4), the vertices of $(C_1 \cup \cdots \cup C_r \cup D_1 \cup \cdots \cup D_s) \setminus B$ might have different colors in X_{t+1} and Y_{t+1}. Thus, we have

$$\mathbf{E}\left[\mathtt{Ham}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1 - 1/q)(p_2 + p_3 + p_4) + \mathbf{E}\left[w_t\right] \tag{3}$$

where $w_t \leq \delta_4\Big(\sum_{i=1}^{r}|C_i| + \sum_{j=1}^{s}|D_j| - |B|\Big)$. To bound $\mathbf{E}\left[w_t\right]$, let $Z+1$ be the size of the largest tree containing v_t in Γ. Then $w_t + |B| = Z$ and $|B| \geq Z/\Delta$. Let θ_i be the indicator that Γ has a tree of size i in which v_t has degree at least 2. Then $Z = 2\theta_3 + \theta_4 + \cdots + \theta_k + \cdots$. Then we know that

$$\mathbf{E}\left[Z\right] = \mathbf{E}\left[\theta_3\right] + \sum_{k \geq 3} \mathbf{E}\left[\theta_k\right].$$

Let α'_k be the number of such k vertex trees in Γ containing v_t and let α_k be the number of ordered rooted trees of size k with maximum degree at most Δ and the degree of root at least 2. Then we know that $\alpha'_k \leq \alpha_k$. By definition, we know that $\mathbf{E}\left[\theta_k\right] = \alpha'_k p^{k-1} \leq \alpha_k p^{k-1}$ and $\mathbf{E}\left[\theta_3\right] \leq \binom{\Delta}{2} p^2$. Let $W(p) = \sum_{k \geq 3} \alpha_k p^{k-1}$. Then,

$$\mathbf{E}\left[w_t\right] \leq \frac{\Delta - 1}{\Delta}\left(\binom{\Delta}{2} p^2 + W(p)\right). \tag{4}$$

Putting (4) into (3), we obtain

$$\mathbf{E}\left[\mathtt{Ham}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t\right] \leq (1 - 1/q)\left(1 - (1-p)^{\Delta}\right) + \frac{\Delta - 1}{\Delta}\left(\binom{\Delta}{2} p^2 + W(p)\right).$$

Now it remains to solve $W(p)$. To solve this, consider the following generating function for an ordered tree of out-degree at most $\Delta - 1$:

$$T(z) = 1 + zT(z)^{\Delta-1}.$$

And we know that

$$W(z) = T(z)^{\Delta} - (1 + \Delta(T(z) - 1)).$$

Numerically solving all functions, we can obtain p_0 by letting the right side be less than 1. Thus we conclude Theorem 4.1.

5. Mixing of the Glauber Dynamics for Ferromagnetic Potts Models

The Swendsen-Wang dynamics is applicable in most cases but difficult to analyze. We turn to the natural single-site heat-bath dynamics $\mathcal{M}_{\texttt{GD}}$ (or, Glauber dynamics).

Proposition 5.1. *Fix a positive integer $q \geq 3$, a parameter $\beta > 0$. Let $G = (V, E)$ be an n-vertex graph with maximum degree Δ. If $q \geq \Delta e^{\beta\Delta} + 1$, then the mixing time of $\mathcal{M}_{\texttt{GD}}$ for the q-spin Potts model at β on G with tolerence error $\varepsilon > 0$ is at most $(\Delta + 1)n \log(n/\varepsilon)$.*

Proof. We prove it by the path coupling theorem. Fix a pair of configurations (X_t, Y_t) satisfying that they differ at only one vertex v_t. We define the following coupling between (X_{t+1}, Y_{t+1}): we pick a uniformly random vertex u at G, and update on u according to $\nu_X := \mu_u^{X-u}$ and $\nu_Y := \mu_u^{Y-u}$. The joint distribution of ν_X and ν_Y will maximize the probability $\mathbf{Pr}[X_{t+1}(u) = Y_{t+1}(u)]$. Let $p = p(u, X, Y) = \mathbf{Pr}[X_{t+1}(u) = Y_{t+1}(u)]$ be this probability. By the basic property of optimal coupling, we have that

$$1 - p = \frac{1}{2} \sum_{k \in [q]} |\nu_X(k) - \nu_Y(k)| = \mathcal{D}_{\mathrm{TV}}(\nu_X \parallel \nu_Y).$$

If $v = u$ or u is not a neighbour of v, it holds that $p = 1$ and thus $X_{t+1} = Y_{t+1}$. Then assume that u is a neighbour of v. Without loss of generality, assume that $X_t(v) = 1$ and $Y_t(v) = 2$. Define

$$a_k = n(X_t, u, k) = |\{w \in N(u) : \ X_t(w) = k\}|, b_k = n(Y_t, u, k) = |\{w \in N(u) : \ Y_t(w) = k\}|.$$

Then we have $b_1 = a_1 - 1$, $b_2 = a_2 + 1$ and $b_k = a_k$ for $k = 3, \ldots, q$. Let

$$Z_X = \sum_{k \in [q]} e^{\beta a_k}, Z_Y = \sum_{k \in [q]} e^{\beta b_k} = Z_X + (1 - e^{-\beta})(e^{\beta(a_2+1)} - e^{\beta a_1})$$

and without loss of generality assume that $Z_X \geq Z_Y$. Then we can see that $\mu_X(k) \leq \mu_Y(k)$ for $k = 2, \ldots, q$ and hence $\mu_X(1) \geq \mu_Y(1)$. Thus

$$\mathcal{D}_{\mathrm{TV}}(\nu_X \parallel \nu_Y) = \nu_X(1) - \nu_Y(1) = \frac{e^{\beta a_1}}{Z_X} - \frac{e^{\beta b_1}}{Z_Y}.$$

Given the vector $\boldsymbol{a} = (a_1, \ldots, a_q) \in [\Delta]^q$, define $f(\boldsymbol{a}, \beta, q) = \frac{e^{\beta a_1}}{Z_X} - \frac{e^{\beta b_1}}{Z_Y}$. Let $g(\beta, q)$ be the maximum of f over all $\boldsymbol{a} \in [\Delta]^q$ subject to $a_1 + \cdots + a_q = \Delta$. Then it is not hard to see

$$\mathbf{E}[\texttt{Ham}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] = 1 - \frac{1}{n} + \sum_{u \in N(v)} \frac{1}{n}(1 - p(u, X_t, Y_t)) \leq 1 - \frac{1}{n} + \frac{\Delta}{n} g(\beta, q).$$

To give an upper bound for $g(\beta, q)$, note that $f(\boldsymbol{a}, \beta, q) \leq \frac{e^{\beta a_1}}{Z_X}$. Then we know that

$$f(\boldsymbol{a}, \beta, q) \leq f((\Delta, 0, \ldots, 0), \beta, q) = \frac{e^{\beta\Delta}}{e^{\beta\Delta} + q - 1} \leq \frac{1}{\Delta + 1}$$

by the assumption on q ($q \geq \Delta e^{\beta\Delta} + 1$). Therefore

$$\mathbf{E}[\texttt{Ham}(X_{t+1}, Y_{t+1}) \mid X_t, Y_t] \leq 1 - \frac{1}{n} + \frac{\Delta}{n(\Delta + 1)} = 1 - \frac{1}{(\Delta + 1)n}.$$

By the argument of path coupling theorem, we obtain the mixing rate of $\mathcal{M}_{\texttt{GD}}$. □

6. Entropic Contraction on Ferromagnetic Potts Models

A recent work by Blanca and Zhang [BZ23] reveals that the spectral independence of the ferromagnetic Potts models implies the optimal mixing of the Swendsen-Wang process.

Theorem 6.1 (Theorem 1.2 in [BZ23]). *Fix $q \geq 2$, $\beta > 0$, $\eta > 0$ and $\Delta \geq 3$. Suppose that $G = (V, E)$ is a graph with $|V| = n$ of maximum degree Δ. Let μ be the q-state ferromagnetic Potts model on G with temperature β. If μ is η-spectrally independent and $\eta = O(1)$, $\beta\Delta = O(1)$, then there exists an universal constant $c = c(\eta, \beta\Delta) > 0$ such that the mixing time of Swendsen-Wang dynamics on μ is $O((\Delta \log n)^c)$.*

This result depends on the factorization of entropy.

Definition 6.2 (Factorization of entropy). Fix an integer $q \geq 2$ and a distribution μ supported on $\Omega \subseteq [q]^n$. For an integer $\ell \leq n$ and a positive factor $C_{\texttt{UBF}}$ (not necessarily constant), we say that μ satisfies *ℓ-uniform-block factorization of entropy with $C_{\texttt{UBF}}$* if for every well-defined function $f : \Omega \to \mathbb{R}_{\geq 0}$,

$$\frac{\ell}{n}\mathbf{Ent}_{\mu}[f] \leq C_{\texttt{UBF}} \cdot \frac{1}{\binom{n}{\ell}} \sum_{S \in \binom{[n]}{\ell}} \mathbf{E}_{\tau \sim \mu_{[n]\setminus S}}\left[\mathbf{Ent}_{\mu^{\tau}}[f]\right].$$

Also, fix a positive integer $k \leq n$ and a k-independent-set partition $U_1, \ldots, U_k$. We say that μ satisfies *k-partition factorization of entropy with factor $C_{\texttt{KPF}}$* if for all functions $f : \Omega \to \mathbb{R}_{\geq 0}$,

$$\mathbf{Ent}_{\mu}[f] \leq C_{\texttt{KPF}} \sum_{i=1}^{k} \mathbf{E}_{\tau \sim \mu_{[n]\setminus U_i}}\left[\mathbf{Ent}_{\mu^{\tau}}[f]\right].$$

The following lemma is the kernel ingredient in the proof of Theorem 6.1.

Lemma 6.3 (Theorem 3.3 in [BZ23]). *For a b-marginally bounded Gibbs distribution μ satisfying η-spectrally independence on an n-vertex graph $G = (V, E)$ of maximum degree Δ, if b and η are constant independent of Δ and n, and $\Delta \in \left[3, \frac{b^4 n}{10e(4\eta + b^2)}\right]$. Then there exists an absolute constant $c > 0$ such that k-partite factorization of entropy holds for μ with $C_{\texttt{KPF}} = (\Delta \log n)^c$. Specifically, for a set of k disjoint independent sets $V_1, \ldots, V_k$ such that $\bigcup_{j=1}^{k} V_j = V$,*

$$\mathbf{Ent}_{\mu}[f] \leq 54 \cdot \frac{e^{13\kappa}}{b^{5+6\kappa}} \cdot (\Delta \log n)^{\kappa} \cdot (\log\log n)^{1+\kappa} \sum_{j=1}^{k} \mathbf{E}_{\tau \sim \mu_{V\setminus V_j}}\left[\mathbf{Ent}_{\mu^{\tau}}[f]\right]$$

where $\kappa = 1 + \lceil \frac{2\eta}{b} \rceil$.

To relate k-partite factorization of entropy to the optimal mixing of the Swendsen-Wang dynamics, we introduce the concept of *edge-spin factorization of entropy*. Recall the Edwards-Sokal measure π supported on $\Omega \times 2^E$. We use $\pi(\cdot \mid \sigma)$ and $\pi(\cdot \mid A)$ to denote the conditional measures of π on fixing the spin configuration σ or the edge configuration A. For a function $f : \Omega \times 2^E \to \mathbb{R}$, we use $f^{\sigma} : 2^E \to \mathbb{R}$ to denote the function $f(\sigma, \cdot)$ and $f^{A} : \Omega \to \mathbb{R}$ to denote $f(\cdot, A)$. Then we say that the edge-spin factorization of entropy holds with factor $C_{\texttt{ES}}$ if for all well-defined function $f : \Omega \times 2^E \to \mathbb{R}_{\geq 0}$,

$$\mathbf{Ent}_{\pi}[f] \leq C_{\texttt{ES}}\left(\mathbf{E}_{(\sigma,A)\sim\pi}\left[\mathbf{Ent}_{\pi(\cdot \mid \sigma)}[f^{\sigma}]\right] + \mathbf{E}_{(\sigma,A)\sim\pi}\left[\mathbf{Ent}_{\pi(\cdot \mid A)}\left[f^{A}\right]\right]\right).$$

Lemma 6.4 (Theorem 6.1 in [BCC+22]). *Suppose that the q-state ferromagnetic Potts model with parameter β on a graph G of maximum degree $\Delta \geq 3$ satisfies k-partite factorization of entropy with factor $C_{\texttt{KPF}}$. Then the edge-spin factorization of entropy holds with factor $C_{\texttt{ES}} = O(\beta\Delta k e^{\beta\Delta}) \cdot C_{\texttt{KPF}}$.*

Lemma 6.5 (Lemma 1.8 in [BCP+21]). *Suppose that the edge-spin factorization of entropy holds with factor $C_{\texttt{ES}}$. Then the Swendsen-Wang dynamics satisfies the relative entropy decay with rate $\Omega(C_{\texttt{ES}}^{-1})$.*

Theorem 6.1 can be proved directly from Lemmas 6.3 to 6.5 with observation that V has a natural $\chi(G)$-independent-set partition.

References

[BCC+22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2022)*, pages 3670–3692, 2022. 7

[BCP+21] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen–Wang dynamics on $\mathbb{Z}^d$. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 1551–1564, New York, NY, USA, 2021. Association for Computing Machinery. 7

[BZ23] Antonio Blanca and Xusheng Zhang. Rapid mixing of global Markov chains via spectral independence: The unbounded degree case. In Nicole Megow and Adam Smith, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023)*, volume 275 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 53:1–53:19, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 7

[JS96] Mark Jerrum and Alistair Sinclair. *The Markov chain Monte Carlo method: an approach to approximate counting and integration*, page 482–520. PWS Publishing Co., USA, 1996. 4

[Vig00] Eric Vigoda. Improved bounds for sampling colorings. *Journal of Mathematical Physics*, 41(3):1555–1569, March 2000. 1