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 $(X,Y) \in \Omega \times \Omega$,

$$\mathbf{E}\left[d(X',Y') \,\middle|\, X,Y\right] \le \gamma d(X,Y).$$

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

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

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,*

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

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 $q \ge 2$, 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 $v=v_0,v_1,\dots,v_t=w$ 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 $X(v_i) \ne X(v_{i+1})$. 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

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

Let U_c be the set of neighbors of v that are colored c. Let $\delta_c = |U_c|$. We denote $U_c = \left\{u^c_1, \ldots, u^c_{\delta_c}\right\}$ or simply $\{u_1, \ldots, u_{\delta_c}\}$ 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

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

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 $u'_1, \ldots, u'_m \in N_G(v)$ 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 $A_c := |S_X(v, c)|$, $B_c := |S_Y(v, c)|$, $a^c_i := |S_Y(u_i, X(v))|$ and $b^c_i := |S_X(u_i, Y(v))|$. We define the vectors $\boldsymbol{a}^c := (a^c_i : \ i \in [\delta_c])$ and $\boldsymbol{b}^c := (b^c_i : \ i \in [\delta_c])$. We say that (X, Y) has configuration $(A_c, B_c; \boldsymbol{a}^c, \boldsymbol{b}^c)$. We also define $a^c_{\max} := \max_i a^c_i$ and $i^c_{\max}$ 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.

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

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

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

References

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