Rapid mixing of Swendsen-Wang dynamics in two dimensions

Mario Ullrich

Friedrich Schiller University Jena

Rome, Italy, 23rd May 2012
Overview

1. Introduction
   - The models and the problem
   - The Markov chains

2. Comparison with local dynamics
   - ... for the Potts model
   - ... for the RC model

3. Rapid mixing in two dimensions
   - Dual graphs

4. Proofs
Problem:

Is the Swendsen-Wang dynamics for the Ising model on the two-dimensional square lattice rapidly mixing?
The Potts Model

For $q \in \mathbb{N}$ and a (finite) graph $G = (V, E)$, the $q$-state Potts model on $G$ is defined as the set of possible configurations $\Omega_P = [q]^V = \{1, \ldots, q\}^V$ together with the probability measure

$$\pi_{\beta, q}(\sigma) := \frac{1}{Z(G, \beta, q)} \exp \left\{ \beta \cdot \# \left\{ \{u, v\} \in E : \sigma(u) = \sigma(v) \right\} \right\}$$

for $\sigma \in \Omega_P$, where $Z$ is the normalization constant and $\beta \geq 0$ is the inverse temperature.

Ising model: $q = 2$
The heat-bath dynamics (HB) is a **local** Markov chain for the Potts model with transition probabilities

\[ P_{\text{HB}}(\sigma, \sigma^v, k) = \frac{1}{|V|} \frac{\pi_\beta(\sigma^v, k)}{\sum_{l=1}^{q} \pi_\beta(\sigma^v, l)}, \]

where \( \sigma^v, k(v) = k \neq \sigma(v) \) and \( \sigma^v, k(u) = \sigma(u), u \neq v \).
Let $P$ be the transition matrix of a (ergodic, reversible) Markov chain on a finite state space $\Omega$ with stationary distribution $\pi$.

$(\implies P^t(x, y) \xrightarrow{t \to \infty} \pi(y) \text{ for all } x, y \in \Omega)$

Then we know $-1 < \xi \leq 1$ for all eigenvalues $\xi$ of $P$. We define the spectral gap by

$$\lambda(P) = 1 - \max\{|\xi| : \xi \text{ eigenvalue of } P, \xi \neq 1\}.$$  

Relaxation time: $\tau(P) := \lambda(P)^{-1}$
Interpretation of $\tau(P)$: 
If $(X_t)_{t \geq 0}$ is a (stat.) Markov chain with transition matrix $P$, i.e. $X_{t+1} \sim P(X_t, \cdot)$, then $X_t$ and $X_{t+\tau(P)}$ are (almost) independent.

Let $\{P_n\}_{n \in \mathbb{N}}$ be a family of Markov chains with corresponding state spaces $\Omega_n$. Then we call the Markov chains rapidly mixing (for $\{\Omega_n\}$), if

$$\tau(P_n) = \mathcal{O}\left((\log |\Omega_n|)^C\right)$$

for some $C \geq 0$ and all $n \in \mathbb{N}$.

(Here, $\tau(P_n) = \mathcal{O}(V_n^C)$ for graphs $\{G_n\}_{n \in \mathbb{N}}$)
Spectral gap III

Define the Hilbert space $L_2(\pi) = (\mathbb{R}^\Omega, \pi)$ with inner product

$$\langle f, g \rangle_\pi := \sum_{x \in \Omega} f(x) g(x) \pi(x), \quad f, g \in \mathbb{R}^\Omega,$$

and regard the transition matrix $P$ as an operator on $L_2(\pi)$ by

$$Pf(x) := \sum_{y \in \Omega} P(x, y) f(y), \quad f \in \mathbb{R}^\Omega.$$

Then,

$$\lambda(P) = 1 - \|P - S_\pi\|_\pi$$

with $S_\pi(x, y) = \pi(y)$ and $\|P\|_\pi := \|P\|_{L_2(\pi) \to L_2(\pi)} = \max_{\|f\|_\pi = 1} \|Pf\|_\pi$.

($\|f\|_\pi^2 = \langle f, f \rangle_\pi$)
Known results for HB dynamics

Consider the two-dimensional square lattice $\mathbb{Z}_L^2 = (V, E)$ with $V = \{1, \ldots, L\}^2$, $n = L^2$, and $E = \{\{v, w\} \subset V : |v - w| = 1\}$.

It is known that $P_{\text{HB}} = P_{\text{HB}, \beta, q}$ satisfies

- $\tau(P_{\text{HB}}) = \mathcal{O}(n)$, if $\beta < \beta_c(q) := \ln(1 + \sqrt{q})$.
- $\tau(P_{\text{HB}}) = e^{\Omega(\sqrt{n})}$, if $\beta > \beta_c(q)$ (for $q = 2$).
- $\tau(P_{\text{HB}}) = \mathcal{O}(n^C)$, if $q = 2$ and $\beta = \beta_c(2)$, for some $C > 0$.

(Martinelli & Olivieri ('94), Cesi et al. ('96), Lubetzky & Sly (2010), Alexander ('98), Beffara & Duminil-Copin (2010))
The *random-cluster model* (or FK-model) on $G = (V, E)$ has the state space $\Omega_{RC} = \{A \subseteq E\}$. So $A \in \Omega_{RC}$ induces a subgraph $(V, A)$ of $(V, E)$.

The probability measure on $\Omega_{RC}$ is given by

$$\mu^G_{p,q}(A) = \frac{1}{Z} \left( \frac{p}{1-p} \right)^{|A|} q^{c(A)},$$

where $c(A)$ is the number of connected components in $(V, A)$.
Connection of the models

In the case $p = 1 - e^{-\beta}$, sampling from Potts and random-cluster model are equivalent in the following sense.

For $\sigma \in \Omega_P$ let

$$E(\sigma) := \{\{u, v\} \in E : \sigma(u) = \sigma(v)\}.$$ 

Then:

1. Given a Potts configuration $\sigma \sim \pi_\beta$ on $G$, delete each edge of $E(\sigma)$ independently with probability $1 - p = e^{-\beta}$. This gives $A \sim \mu_p$.
2. Given a RC configuration $A \sim \mu_p$ on $G$, assign a random color independently to each connected component of $(V, A)$. Vertices of the same component get the same color. This gives $\sigma \sim \pi_\beta$. 
The dynamics II

The *Swendsen-Wang dynamics* (SW) describes a **non-local** Markov chain for the Potts **and** the random-cluster model.

The chain performs the following steps (for the Potts model):

1. Given a Potts configuration \( \sigma_t \in \Omega_P \) on \( G \), delete each edge of \( E(\sigma_t) \) independently with probability \( 1 - p = e^{-\beta} \). This gives \( A \in \Omega_{RC} \).

2. Assign a random color independently to each connected component of \((V, A)\). Vertices of the same component get the same color. This gives \( \sigma_{t+1} \in \Omega_P \).

SW for the random-cluster model makes the steps in reverse order.
We denote by $P_{SW}$ (resp. $\tilde{P}_{SW}$) the transition matrix of the SW dynamics on the Potts (resp. random-cluster) model. Then, independently of $G$, $\beta$ and $q$,

**Lemma**

$$\lambda(\tilde{P}_{SW}) = \lambda(P_{SW}).$$

Hence, results for $\tilde{P}_{SW}$ imply results for $P_{SW}$. 
The single-bond dynamics (SB) is a local Markov chain for the random-cluster model with transition probabilities

$$\tilde{P}_{\text{SB}}(A, A^e) = \frac{1}{|E|} \begin{cases} p, & \text{if } e \not\in A \text{ and } e^{(1)} \leftrightarrow A \leftrightarrow e^{(2)} \\ 1 - p, & \text{if } e \in A \text{ and } e^{(1)} \leftrightarrow A \leftrightarrow e^{(2)} \\ \frac{p}{q}, & \text{if } e \not\in A \text{ and } e^{(1)} \leftrightarrow A \leftrightarrow e^{(2)} \\ 1 - \frac{p}{q}, & \text{if } e \in A \text{ and } e^{(1)} \leftrightarrow A \leftrightarrow e^{(2)}, \end{cases}$$

where $e = \{e^{(1)}, e^{(2)}\} \in E$, $A^e := A \oplus e$ and $\oplus$ is the symmetric difference.
There are several other possibilities for local Markov chains for the RC model, e.g. heat-bath- or Metropolis-type constructions.

$\tilde{P}_{SB}$ is inspired by the “local behavior” of the SW dynamics:

For the graph $G_1 = \left( \{1, 2\}, \{\{1, 2\}\} \right)$, i.e. 2 vertices with a single edge between them, $\tilde{P}_{SB} = \tilde{P}_{SW} \in \mathbb{R}^{2 \times 2}$. 
Suppose that $P_{SW}$ (resp. $P_{HB}$) is the transition matrix of the Swendsen-Wang (resp. heat-bath) dynamics, which is reversible with respect to $\pi^G_{\beta,q}$. Then

$$\tau(P_{SW}) \leq c_{SW} \tau(P_{HB}),$$

where

$$c_{SW} = c_{SW}(G, \beta, q) := 2q^2 \left(q e^{2\beta}\right)^{4\Delta},$$

where $\Delta$ is the maximal degree of $G$. 
Corollary

With this theorem we get, e.g., rapid mixing of SW for the two-dimensional square lattice $\mathbb{Z}_L^2$:

Corollary (Square lattice $\mathbb{Z}_L^2$)

The Swendsen-Wang dynamics for the $q$-state Potts model on $\mathbb{Z}_L^2$, $n = L^2$, at inverse temperature $\beta$ satisfies

- $\tau(P_{SW}) = O(n)$, if $\beta < \beta_c(q) := \ln(1 + \sqrt{q})$, and
- $\tau(P_{SW}) = O(n^C)$, if $q = 2$ and $\beta = \beta_c(2)$.
The bounds are (probably) off by a factor $n$, because we compare a local and a highly non-local algorithm. Optimal would be: $\tau(P_{SW}) = \mathcal{O}\left(\frac{\tau(P_{HB})}{n}\right)$

One may expect rapid mixing also at low temperatures. For this analyze the corresponding Markov chain for the RC model.
Comparison with SB

Theorem

Suppose that $\tilde{P}_{SW}$ (resp. $\tilde{P}_{SB}$) is the transition matrix of the Swendsen-Wang (resp. single-bond) dynamics for the RC model on a graph with $m$ edges. Then

$$\tau(\tilde{P}_{SW}) \leq \tau(\tilde{P}_{SB}) \leq 8m \log(m) \tau(\tilde{P}_{SW}).$$

This implies that rapid mixing of SW and SB dynamics is equivalent.

We conjecture: $\tau(\tilde{P}_{SW}) = \Theta\left(\frac{\tau(\tilde{P}_{SB})}{m}\right)$.

(On trees: $\tau(\tilde{P}_{SB}) = m \tau(\tilde{P}_{SW})$)
Dual graphs

Let $G = (V, E)$ be a finite, planar graph, i.e. a graph that can be “drawn” in the plane without intersecting edges. Denote by $G^\dagger = (V^\dagger, E^\dagger)$ the dual graph of $G$. By definition, $|E^\dagger| = |E|$.

$G$: dots and solid lines
$G^\dagger$: crosses and dashed lines
Let $G$ and $G^\dagger$ be a planar graph and its dual. Then, to each RC configuration $A \in \Omega_{RC}$ there corresponds the dual configuration $A^\dagger \subset E^\dagger$, given by

$$e^\dagger \in A^\dagger \iff e \notin A.$$
Duality

It is easy to obtain (using Euler’s polyhedron formula)

$$\mu_{p,q}^G(A) = \mu_{p^*,q}^{G^\dagger}(A^\dagger),$$

where the dual parameter $p^*$ satisfies

$$\frac{p^*}{1 - p^*} = \frac{q(1 - p)}{p}.$$

The self-dual point of this relation is given by $p_{sd}(q) = \frac{\sqrt{q}}{1 + \sqrt{q}}$, which corresponds by $p = 1 - e^{-\beta}$ to $\beta_c(q) = \log(1 + \sqrt{q})$.

We call the RC model on $G^\dagger$ with $p^*$ and $q$ the dual model.
Denote by $\tilde{P}^\dagger_{\text{SB}}$ the single-bond dynamics for the dual model. Then we obtain

\[ \tau(\tilde{P}_{\text{SB}}) \leq c \tau(\tilde{P}^\dagger_{\text{SB}}) \]

for some $c = c(p, q) < \infty$.

In fact, we have equality for the heat-bath dynamics for the RC model.
Using the results from above:

**Theorem**

Let $\tilde{P}_\text{SW}$ be the Swendsen-Wang dynamics for the random-cluster model on a planar graph $G$ with $m$ edges and let $\tilde{P}_\text{SW}^\dagger$ be the SW dynamics for the dual model. Then

$$\tau(\tilde{P}_\text{SW}) \leq \tau(\tilde{P}_\text{SB}) \leq c \tau(\tilde{P}_\text{SB}^\dagger) \leq c' m \log(m) \tau(\tilde{P}_\text{SW}^\dagger)$$

for some $c, c' < \infty$ depending only on $p$ and $q$.

This result relates the high to the low temperature case.
Mixing for $\mathbb{Z}_L^2$

In particular we get the following for $\mathbb{Z}_L^2$:

**Corollary**

Let $P_{SW}$ be the transition matrix of the SW dynamics for the $q$-state Potts model at inverse temperature $\beta$ on $\mathbb{Z}_L^2$, $n = L^2$. Then there exists a $C < \infty$ such that

- $\tau(P_{SW}) = \mathcal{O}(n)$ for $\beta < \beta_c(q)$.
- $\tau(P_{SW}) = \mathcal{O}(n^2 \log(n))$ for $\beta > \beta_c(q)$.
- $\tau(P_{SW}) = \mathcal{O}(n^C)$ for $q = 2$ and $\beta = \beta_c(2)$. 
The proof ultimately relies on the known mixing results for the heat-bath dynamics.

For the second case we need $\tau(P_{SW}^\dagger) = O(n)$. This is a bit more technical. It is used that the Potts model on $(\mathbb{Z}_L^2)^\dagger$ “corresponds” to the Potts model on $\mathbb{Z}_L^2$ with “constant boundary condition”.

The result can be generalized to graphs of higher genus. E.g. $\mathbb{Z}_L^2$ with periodic boundary conditions.
Proof: SW vs. HB

The proof of $\tau(P_{SW}) \leq c_{SW} \tau(P_{HB})$ uses mainly standard techniques:

1. Consider the chain with transition matrix $P_{HB}P_{SW}P_{HB}$.

2. Note that $\tau(P_{HB}P_{SW}P_{HB}) \leq \tau(P_{HB})$. (bad bound)

3. Proof $P_{SW}(\sigma_1, \sigma_2) \approx P_{SW}(\sigma_1^{v,k}, \sigma_2^{v,l})$ with

   $$\sigma_1^{v,k} = \begin{cases} k, & u = v, \\ \sigma_1(u), & u \neq v. \end{cases}$$

4. This implies $\tau(P_{HB}P_{SW}P_{HB}) \approx \tau(P_{SW})$. 
Proof: SW vs. SB

The Edwards-Sokal coupling of Potts and RC measure is given by

$$\nu(\sigma, A) := \nu_{p,q}^G(\sigma, A) = \frac{1}{Z} \left( \frac{p}{1-p} \right)^{|A|} \mathbb{1}(A \subset E(\sigma)),$$

$$(\sigma, A) \in \Omega_J := \Omega_\text{P} \times \Omega_\text{RC}, \text{ where}$$

$$E(\sigma) := \left\{ \{u, v\} \in E : \sigma(u) = \sigma(v) \right\}.$$ 

The marginal distributions of $\nu$ are exactly $\pi$ and $\mu$, respectively.

We will represent $\tilde{P}_\text{SW}$ and $\tilde{P}_\text{SB}$ on $\Omega_J$. 

Mario Ullrich  
SW in two dimensions
Proof: SW vs. SB

Define the “mapping from \( \Omega_{RC} \) to \( \Omega_J \)” by

\[
M(B, (\sigma, A)) := q^{-c(B)} \mathbb{1}(A = B) \mathbb{1}(B \subset E(\sigma)).
\]

Its adjoint matrix, i.e. \( M^* \) with \( \langle M^* f, g \rangle_\nu = \langle f, Mg \rangle_\mu \) for all \( f \in \mathbb{R}^{\Omega_{RC}} \), \( g \in \mathbb{R}^{\Omega_J} \), is given by

\[
M^*((\sigma, A), B) = \mathbb{1}(A = B).
\]

\( M^* \) is the “mapping back to \( \Omega_{RC} \)”. 
Proof: SW vs. SB

If we additionally define the $|\Omega_J| \times |\Omega_J|$ matrix

$$T_e((\sigma, A), (\tau, B)) := \mathbb{1}(\sigma = \tau) \begin{cases} p, & B = A \cup e \text{ and } \sigma(e^{(1)}) = \sigma(e^{(2)}) \\ 1 - p, & B = A \setminus e \text{ and } \sigma(e^{(1)}) = \sigma(e^{(2)}) \\ 1, & B = A \setminus e \text{ and } \sigma(e^{(1)}) \neq \sigma(e^{(2)}), \end{cases}$$

we obtain

Lemma

$$\tilde{P}_{SW} = M \left( \prod_{e \in E} T_e \right) M^*$$

and

$$\tilde{P}_{SB} = M \left( \frac{1}{|E|} \sum_{e \in E} T_e \right) M^*$$
Proof: SW vs. SB

Note that $T_{e}^{2} = T_e$ and $T_eT_{e'} = T_{e'}T_e$, and define $T = \frac{1}{|E|}\sum_{e \in E} T_e$ as well as $T := \prod_{e \in E} T_e$. (Thus, $T^{\ell} \to T$)

It remains to prove the inequalities

1. $\left\| MT^{k+1}M^* - S_\mu \right\|_\mu \leq \left\| MT^k M^* - S_\mu \right\|_\mu$,

2. $\left\| MTM^* - S_\mu \right\|^{2k}_\mu \leq \left\| MT^k M^* - S_\mu \right\|_\mu$ and

3. $\left\| MT^k M^* - S_\mu \right\|_\mu \leq (1 - \varepsilon) \left\| MTM^* - S_\mu \right\|_\mu + \varepsilon$

for $k = \lceil m \log \frac{m}{\varepsilon} \rceil$
Proof: SW vs. SB

These bounds imply

$$1 - \lambda(\tilde{P}_{SW}) = \lim_{\ell \to \infty} \|MT^\ell M^* - S_\mu\|_\mu \leq \|MTM^* - S_\mu\|_\mu = 1 - \lambda(\tilde{P}_{SB})$$

and

$$\lambda(\tilde{P}_{SW}) = 1 - \|MTM^* - S_\mu\|_\mu \leq 1 - \frac{1}{1 - \varepsilon} \left( \|MT^k M^* - S_\mu\|_\mu - \varepsilon \right)$$

$$= \frac{1}{1 - \varepsilon} \left( 1 - \|MT^k M^* - S_\mu\|_\mu \right) \leq \frac{1}{1 - \varepsilon} \left( 1 - \|MTM^* - S_\mu\|_\mu^{2k} \right)$$

$$\leq \frac{2k}{1 - \varepsilon} \left( 1 - \|MTM^* - S_\mu\|_\mu \right) = \frac{2k}{1 - \varepsilon} \lambda(P_{SB})$$

$$\leq 8m \log(m) \lambda(P_{SB}),$$

for $k = \lceil m \log \frac{m}{\varepsilon} \rceil$ and $\varepsilon = \frac{1}{2}$. 
Final remarks

- While the upper bound on SW is almost tight, both lower bounds are (probably) bad. Corresponding lower bounds would lead to optimal mixing of SW and SB given optimal mixing of HB.

- Results for SW (on $\mathbb{Z}^2$) hold also with “boundary conditions”

- Equivalently for Wolff’s (single-cluster) dynamics (?)

- Higher dimensions? / Direct proofs? / Hard constraints?
Thank you!