Counting and Sampling	**Fall 2017**

Lecture 11: Swendsen-Wang Dynamics and Lower Bounds on Mixing

Lecturer: Shayan Oveis Gharan *Nov 1st*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

11.1 Swendsen-Wang Dynamics

In the last lecture we discussed the mixing time of the Glauber dynamics for sampling from the subgraph model. I start this lecture by explaining a coupling between the ising model and the even subgraph model: Fixed a graph $G = (V, E)$. The ising model on G (with no external field) is defined as follows: For any sign vector $\sigma \in \{\pm\}^V$,

$$w(\sigma) \propto \exp(\beta \cdot sym(\sigma))$$

where $sym(\sigma) = \sum_{i \sim j} \mathbb{I}[\sigma_i = \sigma_j]$ and $\beta \geq 0$.

We define the random cluster model with parameters (p, q) on G as a distribution over subsets $S \subseteq E$ where

$$w_{p,q}(S) \propto p^S (1-p)^{E-S} q^{\kappa(S)} \propto (\frac{p}{1-p})^S q^{\kappa(S)}$$

where $\kappa(S)$ is the number of connected components of S.

Lastly we define the even subgraph distribution with parameter p as follows: For any set $S \subseteq E$ such that every vertex has even degree in S,

$$w_p(S) \propto p^S (1-p)^{E-S} \propto (\frac{p}{1-p})^S$$

Consider a distribution over even subgraphs with parameter p where the probability of an even set S is proportional to $p^S(1-p)^{E-S}$.

Grimmett and Janson [15, Thm 3.5] discovered the following coupling between even subgraphs and random cluster configurations. Take a random even subgraph S from distribution with parameter $p \leq 1/2$. Then add each edge $e \notin S$ **independently** with probability $\frac{p}{1-p}$ to get R.

Theorem 11.1 (15, Thm 3.5)**.** *The subgraph R is a sample from the random cluster configuration with parameters $(2p, 2)$.*

We prove this in multiple steps:

Fact 11.2. *The number of even subgraphs of a connected graph G with n vertices and m edges is 2^{m-n+1}.*

Proof. To see that choose an arbitrary spanning tree T of G; for every edge not in T either put it in/out (so far $2^{m-(n-1)}$ options. We show that for any such configurations, you can uniquely put edges of T in/out to get an even subgraph. Start with a leaf of T; if that already has even degree in the chosen edges do not add its edge otherwise add its edge; delete the leaf and recurse. □

It follows from the above fact that for a non-necessarily connected graph G the number of even subgraphs is $2^{m-n+\kappa(G)}$ where $\kappa(G)$ is the number of connected components of G.

Proof. Fix a set $F \subseteq E$ and let $Even(F)$ be the number of even subgraphs of F.

$$\begin{aligned}
\mathbb{P}\left[R = F\right] &= \sum_{S \subseteq F, S \text{ even}} p^{|S|}(1-p)^{m-|S|}(\frac{p}{1-p})^{|F-S|}(1-\frac{p}{1-p})^{|E-F|} \\
&= \sum_{S \subseteq F, S \text{ even}} p^{|F|}(1-2p)^{|E-F|} \\
&= \text{Even}(F)p^{|F|}(1-2p)^{|E-F|} \\
&= 2^{|F|-n+\kappa(F)}p^{|F|}(1-2p)^{|E-F|} \propto (2p)^{|F|}(1-2p)^{|E-F|}2^{\kappa(F)}
\end{aligned}$$

The last equation hides a normalizing constant of 2^n. This is exactly the probability of sampling F in the $(2p, 2)$ random cluster model. $\square$

Next, we discuss the Edwards-Sokal coupling of the random cluster and the Ising model. To explain that we need to define a Markov chain called the Swendsen-Wang chain which is a "non-local" markov chain on the Ising model. Given a signing $\sigma \in \{\pm 1\}^V$:

- For every edge $e = \{u, v\}$ with $\sigma_u = \sigma_v$ we include e, independently, with probability $a = 1 - e^{-\beta}$ (so we don't include any edge with $\sigma_u \neq \sigma_v$.
- Let S be the sampled set of edges. For every connected component of S we choose a sign in $-1, +1$ uniformly and independently at random.

Theorem 11.3. *The above chain gives a coupling between the random cluster model and the ising model. In particular, if one can sample from the random cluster model then by running one step of the chain they can sample from the Ising model.*

In other words, combining the two coupling given in this section, one can first start from the Jerrum-Sinclair chain to generate a sample from the even subgraph model then use the first coupling to get a sample from the random cluster model and finally use the Swendsen-Wang chain to generate a sample from the Ising model.

Proof. Consider the bipartite graph where on one side we have the state space of the Ising model and on the other sides we have the state space of the random cluster model with parameters $(p, 2)$ where

$$\frac{p}{1-p} = \frac{1}{1/p - 1} = e^{\beta}a = e^{\beta} - 1 \Rightarrow 1/p - 1 = \frac{1}{e^{\beta} - 1} \Rightarrow p = \frac{e^{\beta} - 1}{e^{\beta}}$$

We show that $\pi(\sigma) \propto w(\sigma)$ and $\pi(S) \propto w_{p,q}(S)$ is stationary for the walk explained above. To see that we just need to verify the detailed balanced condition. Fix a pair $\sigma \in \{\pm 1\}^V$ and $S \subseteq E$ such that $K(\sigma, S) > 0$. This is equivalent to the condition that S has no edges $e = \{u, v\}$ where $\sigma_u \neq \sigma_v$.

Let $S' \subseteq E$ be the set of edges with the same spin on both endpoint points that are not in S. In other words, $S \cup S'$ are all edges with the same spin. Then, we can write,

$$\begin{aligned}
\pi(\sigma)K(\sigma, S) &\propto \exp(\beta \cdot sym(\sigma)) \cdot a^{|S|}(1-a)^{|S'|} \\
&= (e^{\beta}a)^{|S|}(e^{\beta}(1-a))^{|S'|} = (e^{\beta}a)^{|S|}
\end{aligned}$$

where we used that $e^{\beta}(1-a) = 1$.

On the other hand, using the definition of the random cluster model with parameters $(p, 2)$ we can write,

$$\pi(S)K(S,\sigma) \propto (\frac{p}{1-p})^S 2^{\kappa(S)} \cdot \frac{1}{2^{\kappa(S)}} = (e^\beta a)^{|S|}$$

The factor $\frac{1}{2^{\kappa(S)}}$ comes from the observation that to transition from $S \to \sigma$ we need to choose the right sign for every connected component of S.

Since these two are equal, indeed the random walk has the ising/random cluster stationary distribution. The coupling corresponds to the edges of this bipartite graph. $\square$

We end this section with additional remarks about the random cluster model: The random cluster model is well-defined for any $q > 0$. It is shown that one can sample from the model for any $0 < q \le 1$ efficiently and we will discuss this later in the course. It follows by the canonical path argument of Jerrum-Sinclair that one can sample from the model for $q = 2$. It is long-standing open problem whether it is possible to generate samples fro $1 < q < 2$. Partial results are known for special family of graphs. It is believed that the model is intractable (no FPRAS exists) for $q > 2$. In fact, it is shown by Goldberg-Jerrum that it is #Bis-hard to estimate the partition fruction of the model for any $q > 2$. The coupling between the random cluster model and the Ising model generalizes for integer values of $q > 2$. Such a model is called the Potts model in statistical physics and it is generally hard to sample from the Potts model for $q > 2$ for general graphs.

11.2 Lower bounds on the Mixing time of the Glauber Dynamics

First, we prove a lower bound on mixing using the conductance. Note that when the chain is reversible, the inverse of the Poincaré constant also gives a lower bound on mixing time. The importance of the following lemma is that it is significantly easier to study. To lower bound the mixing time of the chain, all we need to do is to find a cut with small conductance.

Lemma 11.4. *For a (not necessarily reversible) kernel K let $Q(x, y) = \pi(x)K(x, y)$. Then, for any set $S \subseteq \Omega$ where $\pi(S) \le 1/2$,*

$$\tau_{\text{mix}} \ge \frac{1}{4\phi(S)}.$$

Proof. Consider the following starting distribution:

$$\mu^0(x) = \begin{cases} \frac{\pi(x)}{\pi(S)} & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$$

Also, let $\mu^t = \mu K^t$ be the distribution at time t. Then, observe that

$$\begin{aligned}
\left\|\mu^1 - \mu^0\right\|_{TV} &= \frac{1}{2}\sum_x |\mu^1(x) - \mu^0(x)| && \text{(Def of TV distance)} \\
&= \frac{1}{2}\sum_x \left|\sum_y \mu^0(y)K(y,x) - \mu^0(x)\right| \\
&\ge \sum_{x \notin S} |\sum_y \mu^0(y)K(y,x) - \mu^0(x)| && (\text{since } \mu^1(x) \ge \mu^0(x) = 0 \text{ for } x \notin S) \\
&= \sum_{x \notin S}\sum_y \mu^0(y)K(y,x) && (\mu^0(x) = 0 \text{ for } x \notin S) \\
&= \sum_{y \in S}\sum_{x \notin S} \mu^0(y)K(y,x) = \phi(S). && (11.1)
\end{aligned}$$

Now, recall that by coupling lemma

$$\left\|\mu^t - \mu^{t-1}\right\|_{TV} \geq \left\|\mu^{t+1} - \mu^t\right\|_{TV}.$$

Therefore, by triangle inequality,

$$\left\|\mu^t - \mu^0\right\|_{TV} \leq \sum_{i=0}^{t-1} \left\|\mu^{i+1} - \mu^i\right\|_{TV} \leq t\left\|\mu^1 - \mu^0\right\|_{TV} \leq t\phi(S).$$

But,

$$\left\|\mu^t - \pi\right\|_{TV} \geq \left\|\mu^0 - \pi\right\| - \left\|\mu^t - \mu^0\right\|_{TV} \geq 1/2 - t\phi(S)$$

So, for $t < 1/4\phi(S)$, the total variation distance is at least $1/4$. □

The above lemma gives a useful technique, known as the *bottleneck ratio*, to bound the mixing time. Namely, all we need to do is to find a large set S of size $\pi(S) \geq \Omega(1)$, and we need to show that it is very hard for the random walk to enter S (or equivalently to leave S) because it has small conductance.

11.3 An Exponential Lower bound for Mixing time of Glauber dynamics in Ising Model

Consider the Glauber dynamics, a.k.a., Heat-Bath chain, that we discussed before. In the rest of this lecture, we show that there is another constant β_1 such that if $\beta > \beta_1$, then the mixing time is exponential in n. The more general theorem is as follows. The proof is technical and we don't give all details.

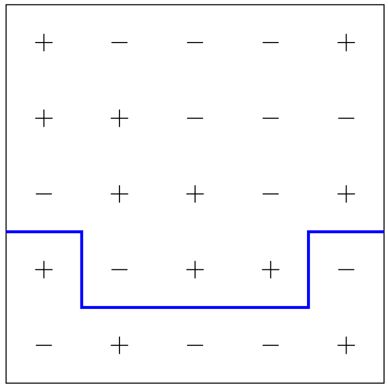
Theorem 11.5 ([MO94])**.** *There is a constant β_c such that for any $\beta < \beta_c$, the heat-bath chain on the Ising model in a $\sqrt{n} \times \sqrt{n}$ grid mixes in time $O(n \log n)$ and for $\beta > \beta_c$ the mixing time is at least $e^{c\sqrt{n}}$ for some universal constant $c > 0$.*

The high-level intuition is clear: Clearly for large enough β, the all-plus state and the all-minus state would have the largest probabilities. But this Markov chain has exponentially many states. So, these two states do not contain all of the probability mass. Instead, the idea is to divide the states into those close to being all-plus, say S_+, and those which are close to being all-minus, say S_-, and show that $\phi(S_+)$ is very small. See the following figure.

Figure 11.1: A bottleneck between S_+ and S_-.

A *path* is sequence of adjacent sites. A *line* is a sequence of line segments where each segment is a side of sub-square of $\sqrt{n} \times n$ square. The blue curve in Definition 11.3 is a line.

Definition 11.6 (Fault Line)**.** *For a configuration σ, a fault line is a line where for each segment the two sites on the opposite side of the segment have different spins, i.e., signs.*



For example, the blue curve in Definition 11.3 is a fault line.

Lemma 11.7. *Let F be the set of configurations that contain a fault line of length $\geq \sqrt{n}$. The for some universal constant β_1, and any $\beta > \beta_1$,*

$$\pi(F) \leq e^{-c\sqrt{n}}.$$

where c is a universal constant.

Proof. First observe that the number of fault lines of length $\ell \geq \sqrt{n}$ is at most $2\sqrt{n}3^\ell$. This is because there are at most $2\sqrt{n}$ starting position and each time there are at most 3 options (because the line is self-avoiding).

Now, fix a fault line L and let $F(L)$ be all configurations where L is a fault line. For a $\sigma \in F(L)$ flip the spin of all sites on one side of L. Observe that the weight of σ goes up by $e^{2\beta\ell}$. Moreover, this mapping is one-to-one. Therefore, $\pi(F(L)) \leq e^{-2\beta\ell}$. Now, brute forcing over all possible L we have

$$\pi(F) \leq \sum_L \pi(F(L)) \leq 2\sqrt{n} \sum_\ell 3^\ell e^{-2\beta\ell} \leq e^{-c\sqrt{n}},$$

for $\beta > \frac{1}{2}\ln 3$. $\square$

In order to make sure that the fault lines that we consider have length at least $\sqrt{n}$ from now on we only focus on left-right and bottom-up fault lines. It follows by the above lemma that most of the probability mass in π is on states which do not have a fault line. In the following lemma we characterize these states.

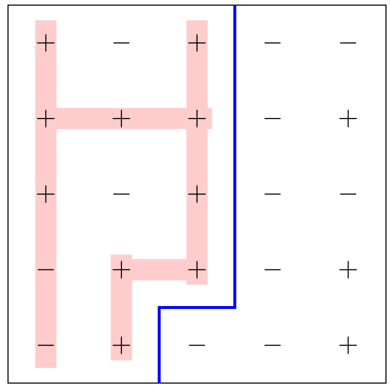
Lemma 11.8. *Let σ be a configuration which has no monochromatic left-right path, i.e., no +/- left-right path. Then, σ has a bottom-up fault line.*

Proof. Think of the following picture for intuition. Let A be the set of sites i where there is a path from the left side to i by sites with spin the same as σ_i. For example, in the above figure the set A are all sites marked in red. The assumption of the lemma implies that A does not have any site in the right boundary. The definition of the set A implies that for any site i which is not in A but adjacent to a site of A, say j, we have $\sigma_i \neq \sigma_j$. It follows that the line at the right boundary of A is a bottom-up fault line. $\square$

Lemma 11.9. *Suppose there is a site i such that there is a plus-path from i to the top and there is a minus-path from i to the top. Then there is a fault line from x to the top.*

We defer the proof of this lemma as an exercise.

Having the above lemmas, we are ready to prove the theorem.



Theorem 11.10. *There is a universal constant $\beta_1 > 0$ such that for all $\beta > \beta_1$, the mixing time of the heat-bath chain on a $\sqrt{n} \times \sqrt{n}$ grid is at least $e^{c\sqrt{n}}$.*

Proof. Let S_+ be all states with a left-right plus-path and a bottom-up plus-path. Similarly, let S_- be all states with a left-right minus-path and a bottom-up minus-path. Observe that by Lemma 11.8 any state that is not in S_+, S_- has a fault line. This is because say a state σ does not have a left-right plus path. If σ does not have a left-right minus-path as well then by Lemma 11.8 it has a bottom-top fault line. So, say σ has a left-right minus path. Then, it cannot have a bottom-up plus-path. Futhermore, since $\sigma \notin S_-$ it does not have bottom-up minus-path. So, σ must have a left-right fault line. Now, by Lemma 11.7 $\pi(\overline{S_+ \cup S_-}) \leq e^{-c\sqrt{n}}$. Therefore, by symmetry, as $n \to \infty$, $\pi(S_+), \pi(S_-) \to 1/2$.

Now, let $N(S_+)$ be states which are not in S_+ such that they adjacent to some state in S_+. In other words, every state in $N(S_+)$ differ in the sign of one site with respect to a state in S_+. To prove the claim it is enough to show that $\pi(N(S_+)) \leq e^{-c'\sqrt{n}}$. This is because by Lemma 11.4 we have

$$\tau_{\text{mix}} \geq \frac{1}{4\phi(S_+)} \geq \Omega(\pi(N(S_+))),$$

where we used that $\pi(S_+) \approx 1/2$.

It remains to upper bound $\pi(N(S_+))$. We divide states $\sigma \in N(S_+)$ in two groups: (i) States $\sigma \in N(S_+) \cap \overline{S_- \cup S_+}$. But, as we discussed earlier such a set has probability at most $e^{-c\sqrt{n}}$ because it has a fault line. (ii) States $\sigma \in N(S_+) \cap S_-$. Fix such a σ. Since σ is a neighbor of S_+ there is a site i such that by flipping the spin of i the new state σ' is in S_+. Observe that since $\sigma \in S_-$, there are left-right and bottom-up minus-paths in σ. Furthermore, since $\sigma' \in S_+$ there are left-right and bottom-up plus-paths in σ'. But σ, σ' differ in the sign of a single site. Therefore, site i must have plus and minus paths to top/bottom/left/right (see the following figure). Therefore, by Lemma 11.9 there are fault lines from i to top and bottom. So, there is an "almost" fault line from bottom to top in σ. Such a line may have agreeing signs in the neighborhood of i. But that is a constant length part of the path. We can adopt the proof of Lemma 11.7 to argue that probability of states with an almost fault line is also $e^{-c'\sqrt{n}}$. Therefore, $\pi(N(S_+)) \leq e^{-2c'\sqrt{n}}$.

$\square$

References

[MO94] F. Martinelli and E. Olivieri. Approach to equilibrium of glauber dynamics in the one phase region. ii. the general case. *Comm. Math. Phys.*, 161(3):487–514, 1994. 11-4

	$+$	$-$		
	$+$	$-$	$-$	$-$
$+$	$+$	i	$+$	$+$
$-$	$-$	$-$	$+$	
		$-$	$+$	