Lecture Note 13

Instructor: Alistair Sinclair

Disclaimer: *These notes have not been subjected to the usual scrutiny accorded to formal publications. They may be distributed outside this class only with the permission of the Instructor.*

13.1 The ferromagnetic Ising model

In this lecture we will see a further application of the flow-encoding technology, to the ferromagnetic Ising model. We will show the existence of an fpras based on MCMC for the partition function of this model on all graphs and at all values of the parameters. However, since MCMC is known not to be effective at low temperatures, we will need to transform the problem to a different domain before applying the algorithm. This application is due to [JS93].

Recall from the first lecture[1] that the ferromagnetic Ising model is a spin system on a graph $G = (V, E)$ with two spin values $\{\pm 1\}$. The weight of configuration $\sigma \in \{\pm 1\}^V$ is

$$w(\sigma) = \exp\Big(\beta \sum_{\{u,v\} \in E} \sigma(u)\sigma(v) + \beta h \sum_{u \in V} \sigma(u)\Big). \tag{13.1}$$

Here $\beta > 0$ is the inverse temperature parameter, and h is the external field. The Gibbs distribution is then $\pi(\sigma) = \frac{w(\sigma)}{Z(\beta,h)}$, where $Z(\beta, h) := \sum_\sigma w(\sigma)$ is the partition function. Our goal is to estimate $Z(\beta, h)$.

A natural approach is to simulate the Glauber dynamics for the spin system and use the random samples in the usual way to estimate Z. Unfortunately, however, this approach won't work at low temperatures (large values of β) because of a *phase transition* in the model. At zero field ($h = 0$), there is symmetry between majority-$(+1)$ and majority-(-1) configurations. The Glauber dynamics (or any local dynamics that alters a bounded number of spins at a time) must pass through configurations with a roughly equal number of ± 1 spins in order to pass from the majority-$(+1)$ to majority-(-1) part of the space. But despite the fact that entropy favors the balanced configurations, when β is large enough their aggregated weight is exponentially small compared to that of the unbalanced ones, thus forming a bottleneck that requires exponential time to cross. (We'll see a rigorous example of such an argument in a later lecture.)

In some special cases this phase transition is well enough understood to be able to make a very precise statement, for example:

Theorem 13.1. *At $h = 0$, the mixing time of the Glauber dynamics on a $\sqrt{n} \times \sqrt{n}$ box in $\mathbb{Z}^2$ is*

- $O(n \log n)$ *for all $\beta < \beta_c$;*

- $\exp(\Omega(\sqrt{n}))$ *for all $\beta > \beta_c$;*

- *polynomial in n for $\beta = \beta_c$.*

---

[1]The definition we give here is slightly different from that in Lecture 1, in that the Hamiltonian here counts the difference between the numbers of agreeing and disagreeing edges, while in Lecture 1 we counted only agreeing edges. However, since the total number of edges is $|E|$, this makes no difference other than multiplying the partition function by a constant and scaling β by a factor of 2. **Exercise:** Verify this!

Here β_c is the so-called "critical (inverse) temperature", which also marks the onset of long-range correlations (non-uniqueness of the Gibbs measure) in $\mathbb{Z}^2$. Note that it is not hard to prove using coupling that the Glauber dynamics mixes rapidly (in $O(n \log n)$ time) on any graph at very high temperatures, or by establishing a bottleneck that mixing is exponentially slow at very low temperatures. What is remarkable about Theorem 13.1 is that both these arguments can be pushed all the way to the critical point.

The first two parts of the above theorem, which are the culmination of years of research in mathematical physics that we won't attempt to summarize, can be found in [MO94]; the last part is more recent and is due to [LS12].

13.1.1   The subgraphs world

To overcome this fundamental obstacle, we'll transform the Ising model to an equivalent combinatorial model that is not susceptible to this phase transition. For this we appeal to a classical result known as the "high-temperature expansion" [NM53]. The configurations of this model are all subgraphs (V, A) where $A \subseteq E$, which we'll just identify with A. The weight of configuration A is

$$w'(A) = \eta^{|A|} \xi^{|\mathrm{odd}(A)|},$$

where the parameters η, ξ are defined by $\eta = \tanh(\beta)$ and $\xi = \tanh(\beta h)$, and $\mathrm{odd}(A)$ is the set of odd-degree vertices in the subgraph (V, A). Note that $w'(A) \leq 1$ for all A. We write $Z'(\eta, \xi) := \sum_A w'(A)$ for the associated partition function, and $\pi'(A) = w'(A)/Z'$ for the Gibbs distribution.

Theorem 13.2. *For any graph G, we have $Z(\beta, h) = C_{\beta,h} Z'(\eta, \xi)$, where $C_{\beta,h} := (\cosh \beta)^{|E|} (2 \cosh(\beta h))^{|V|}$.*

Proof. Using $\exp(x) = \cosh x (1 + \tanh x)$, we can write

$$
\begin{aligned}
Z(\beta, h) &= \sum_\sigma \exp\Big(\beta \sum_{\{u,v\} \in E} \sigma(u)\sigma(v) + \beta h \sum_{u \in V} \sigma(u) \Big) \\
&= (\cosh \beta)^{|E|} (\cosh(\beta h))^{|V|} \sum_\sigma \prod_{\{u,v\} \in E} \Big(1 + \tanh(\beta \sigma(u)\sigma(v)) \Big) \prod_{u \in V} \Big(1 + \tanh(\beta h \sigma(u)) \Big),
\end{aligned}
$$

where the spin factors $\sigma(u), \sigma(v)$ have been removed from the prefactor because they are always ± 1 and $\cosh$ is an even function. Note that this prefactor is $2^{-|V|} C_{\beta,h}$. Similarly, $\tanh$ is an odd function, so we can pull the spin factors out of the $\tanh$ expressions to get

$$
\begin{aligned}
Z(\beta, h) &= 2^{-|V|} C_{\beta,h} \sum_\sigma \prod_{\{u,v\} \in E} \Big(1 + \sigma(u)\sigma(v) \tanh(\beta) \Big) \prod_{u \in V} \Big(1 + \sigma(u) \tanh(\beta h) \Big) \\
&= 2^{-|V|} C_{\beta,h} \sum_\sigma \Big(\sum_{A \subseteq E} \prod_{\{u,v\} \in A} \eta \sigma(u)\sigma(v) \Big) \Big(\sum_{U \subseteq V} \prod_{u \in U} \xi \sigma(u) \Big) \\
&= 2^{-|V|} C_{\beta,h} \sum_{A \subseteq E} \sum_{U \subseteq V} \sum_\sigma \Big(\prod_{\{u,v\} \in A} \eta \sigma(u)\sigma(v) \Big) \Big(\prod_{u \in U} \xi \sigma(u) \Big) \\
&=: 2^{-|V|} C_{\beta,h} \sum_{A \subseteq E} \sum_{U \subseteq V} \sum_\sigma W(A, U, \sigma), \quad\quad\quad (13.2)
\end{aligned}
$$

where in the third line we interchanged the order of summation.

Now we claim that $\sum_\sigma W(A, U, \sigma) = 0$ unless $\mathrm{odd}(A) = U$. To see this, fix A and U and suppose that there is a vertex $v \in V$ such that *either* $v \in U$ and $v \notin \mathrm{odd}(A)$, *or* $v \notin U$ and $v \in \mathrm{odd}(A)$. This implies that the degree of $\sigma(v)$ in the monomial $W(A, U, \sigma)$ is odd. Hence if σ' denotes the configuration that differs from σ

only at v (i.e., the spin of v is flipped), then $W(A, U, \sigma) = -W(A, U, \sigma')$ so the terms in $\sum_\sigma W(A, U, \sigma)$ cancel in pairs.

On the other hand, suppose A, U satisfy $\text{odd}(A) = U$. In this case, the degrees of all $\sigma(u)$ in $W(A, U, \sigma)$ are even, so $W(A, U, \sigma) = \eta^{|A|} \xi^{|\text{odd}(A)|}$ for all σ. Plugging this into (13.2) gives

$$Z(\beta, h) = 2^{-|V|} C_{\beta,h} \sum_{A \subseteq E} 2^{|V|} \eta^{|A|} \xi^{|\text{odd}(A)|} = C_{\beta,h} \sum_{A \subseteq E} w'(A),$$

as claimed. $\square$

In what follows, we'll refer to this new model as the "subgraphs world", and the original Ising model as the "spins world." The point about Theorem 13.2 is that, if we want to evaluate Z in the spins world, we can equivalently evaluate Z' in the subgraphs world and then scale the result by the trivial factor $C_{\beta,h}$.

13.1.2   MCMC in the subgraphs world

Here is a natural Markov chain (Metropolis Glauber dynamics) in the subgraphs world. The states are all subgraphs (V, A) (which we identify with their edge sets A). In configuration A, transitions are made as follows:

- with probability $\frac{1}{2}$ do nothing, else

- pick an edge $e \in E$ u.a.r., and let $\hat{A} := A \oplus e$;
  move to $\hat{A}$ with probability $\min\{\frac{w'(\hat{A})}{w'(A)}, 1\}$, else stay at A.

This Markov chain is clearly irreducible and aperiodic, and also reversible w.r.t. the subgraphs world Gibbs distribution π' by virtue of the Metropolis rule in the last line. (We could equally well consider the heat-bath Glauber dynamics, whose analysis would be essentially the same.)

The main ingredient in our algorithm is the following:

Theorem 13.3. *The mixing time of the above Markov chain, starting from the empty graph $A = \emptyset$, is $O(\xi^{-4} |E|^3)$.*

Before proving this theorem, let's briefly consider its implications.

13.1.3   Estimating the partition function

As mentioned earlier, to compute the spin partition function $Z(\beta, h)$ it suffices to compute the subgraphs partition function $Z'(\eta, \xi)$ and multiply by $C_{\beta,h}$. To compute $Z'(\eta, \xi)$ at any desired point $(\eta, \xi) \in [0, 1]^2$, we fix η and view Z' as a function of ξ alone; thus we may write it as

$$Z'(\xi) = \sum_k \alpha_k \xi^k \qquad \text{where } \alpha_k := \sum_{A:\text{odd}(A)=k} \eta^{|A|}. \tag{13.3}$$

Now in familiar fashion, for a suitable sequence of values $1 = \xi_0 > \xi_1 > \ldots > \xi_r = \xi$ we can write

$$Z'(\xi) = Z'(\xi_0) \times \prod_{i=1}^r \frac{Z'(\xi_i)}{Z'(\xi_{i-1})}. \tag{13.4}$$

Note that $Z'(\xi_0) = Z'(1) = (1+\eta)^{|E|}$ is trivial, since subgraphs are weighted only by their number of edges. Equation (13.4) thus gives us a way of bootstrapping from this trivial point to the desired point $Z'(\xi)$. In spins-world terms, this corresponds to subjecting an Ising model with fixed interaction strength (inverse temperature) β to decreasing external fields h, starting at the trivial value $h = \infty$.

As usual, we can estimate each ratio in (13.4) by random sampling using the Markov chain, as follows. To estimate $\frac{Z'(\xi_i)}{Z'(\xi_{i-1})}$, we sample from the Gibbs distribution π'_{i-1} at ξ_{i-1} and observe the random variable $X_i := (\frac{\xi_i}{\xi_{i-1}})^{|\text{odd}(A)|}$, which has expectation

$$\mathrm{E}[X_i] = \frac{1}{Z'(\xi_{i-1})} \sum_k \alpha_k \xi_{i-1}^k \Big(\frac{\xi_i}{\xi_{i-1}}\Big)^k = \frac{1}{Z'(\xi_{i-1})} \sum_k \alpha_k \xi_i^k = \frac{Z'(\xi_i)}{Z'(\xi_{i-1})},$$

as desired. As usual, this method works efficiently provided the variances of the estimators are suitably bounded. If we choose the ξ_i so that $\frac{\xi_i}{\xi_{i-1}} = 1 - \frac{1}{n}$, where $n = |V|$, then it's easy to see **[exercise]** that the r.v. X_i takes values in the bounded range $[\mathrm{e}^{-1}, 1]$, so it's variance is clearly bounded. Setting $r \leq cn \ln n$ allows us to estimate $Z'(\xi)$ for any desired $\xi \geq \frac{1}{n}$ **[exercise!]**. Note that the mixing time remains polynomial down to such values of ξ, so we get an fpras.

Finaly, we explain how to compute $Z'(\xi)$ at smaller values of ξ, all the way down to $\xi = 0$. This is important because it includes the most delicate symmetric case of zero field, $h = 0$, which is particularly difficult for spins-world MCMC. The key observation here is that

$$\frac{Z'(0)}{Z'(1/n)} \geq \frac{1}{\mathrm{e}^2}, \tag{13.5}$$

so we can in fact estimate $Z'(0)$ by sampling at $\xi = \frac{1}{n}$ exactly as above. In particular, we use the random variable $X = 0^{|\text{odd}(A)|}$, which is just the indicator r.v. of the event $A = \emptyset$. Hence (13.5) will ensure that the variance is again bounded. To see (13.5), we refer back to the spins world. Recall that $\xi = \tanh(\beta h)$. Using the bound $x \leq \frac{\tanh x}{1 - \tanh^2 x}$, valid for all $x \geq 0$, we see that when $\xi = \frac{1}{n}$, $\beta h \leq \frac{1/n}{1 - 1/n^2} \leq \frac{2}{n}$. But recalling the spins world weights (13.1), we see that for such a small value of βh, the weight of each configuration σ is at most a factor e^2 larger than its weight at $h = 0$ **[exercise!]**. Therefore

$$\frac{Z'(0)}{Z'(1/n)} = \frac{Z(0)}{Z(\beta h)} \times \frac{C_{\beta,h}}{C_{\beta,0}} \geq \frac{1}{\mathrm{e}^2},$$

where we have used the fact that $C_{\beta,h} \geq C_{\beta,0}$. This proves (13.5).

We have proved:

Corollary 13.4. *There is an fpras for the partition function of the ferromagnetic Ising model on any graph at any values of the parameters β and h.*

Notes:

1. All of the above generalizes in a straightforward way to the situation where there is a separate interaction parameter $\lambda_{uv} > 0$ for each edge; i.e., the Gibbs weight of a spin configuration becomes $w(\sigma) = \exp(\beta \sum_{\{u,v\} \in E} \lambda_{uv} \sigma(u)\sigma(v) + \beta h \sum_{u \in V} \sigma(u))$. This introduces edge-dependent weights $\eta_{uv} = \tanh(\beta \lambda_{uv})$ in the subgraphs world.

2. The presence of any *fixed* field $h > 0$ (however small), removes the bottleneck between phases identified in Theorem 13.1 and ensures polynomial mixing time at all temperatures. The subgraphs world Markov chain also requires a non-zero field, but this can be inverse polynomially small, tending to zero as $n \to \infty$. As we have seen above, taking $h = \frac{1}{n}$ is essentially equivalent to sampling at zero field. Hence the subgraphs world is not affected by the phase transition.

3. It might seem that we have thrown away a lot of information about the Ising model in transferring to the subgraphs world. However, it turns out that we can still compute essentially all other global observables by sampling subgraphs. The following two examples are left as (slightly challenging) exercises:

Exercise:

(i) The *mean magnetization* is defined as $\mathcal{M} = \mathrm{E}[\sum_u \sigma(u)]$, where the expectation is over the spins world Gibbs distribution. (Note that we may assume $h > 0$ since by symmetry $\mathcal{M} = 0$ when $h = 0$.) Show that
$$\mathcal{M} = |V|\tanh(\beta h) + \frac{2}{\sinh(2\beta h)}\mathrm{E}[|\mathrm{odd}(A)|],$$
where here the expectation is over the spins world Gibbs distribution. Thus we can estimate $\mathcal{M}$ simply by observing the random variable $|\mathrm{odd}(A)|$ in the subgraphs world (though some additional work needs to be done to ensure that this computation is efficient). [Hint: You may find it useful to observe that $\mathcal{M}$ can be written equivalently as $\mathcal{M} = \beta^{-1}\partial(\ln Z)/\partial h$.]

(ii) The *mean energy* is defined as $\mathcal{E} = \mathrm{E}[H(\sigma)]$, where $H(\sigma) = -\sum_{u \sim v} \sigma(u)\sigma(v) - h\sum_u \sigma(u)$ is the Hamiltonian and the expectation is again over the spins world Gibbs distribution. Show that, when $h = 0$,
$$-\mathcal{E} = |E|\tanh\beta + \frac{2}{\sinh(2\beta)}\mathrm{E}[|A|],$$
where this expectation is over the subgraphs world Gibbs distribution. (When $h > 0$ there is an analogous but more complex expression.) Hence again we can estimate $\mathcal{E}$ by sampling in the subgraphs world. [Hint: you might want to write equivalently $-\mathcal{E} = \partial(\ln Z)/\partial\beta$.)

13.1.4 Proof of Theorem 13.3

Finally, we go back and provide the analysis of the mixing time of the subgraphs world Markov chain, as claimed in Theorem 13.3. This shares many similarities with our analysis of the Markov chain on matchings in the previous lecture, so we'll just sketch it.

The idea is once again to define a path γ_{XY} between each pair of subgraphs X, Y, along with a weight-preserving encoding function $\mathrm{enc}_t : \mathrm{paths}(t) \to \Omega$ for each transition t, where Ω is the set of subgraph configurations.

To define the paths, consider the symmetric difference $X \oplus Y$, by which we mean the subgraph (V, D) whose edge set D consists of all edges in exactly one of X, Y. Suppose $X \oplus Y$ has $2k$ vertices of odd degree (note that this number must always be even). Now it's not hard to see **[exercise!]** that we can decompose $X \oplus Y$ into a collection of edge-disjoint trails and circuits, with exactly k open trails. (Recall that a *trail* is a walk with no repeated edges; a circuit is a closed trail.) Order the open trails as $C_1, \ldots, C_k$ and the circuits as $C_{k+1}, \ldots, C_r$. Specify a start vertex for each of the C_i, which must be an endpoint if C_i is open and is arbitrary otherwise. The path γ_{XY} proceeds from X to Y by *unwinding* the C_i in sequence in the obvious way (beginning at the start vertex and adding edges of Y and deleting edges of X as we go).

Now consider some arbitrary transition $t = (Z, Z')$, where $Z = A$ and $Z' = A \oplus e$ for some edge e of G (i.e., t adds or deletes the edge e). Let $\mathrm{paths}(t)$ denote the set of pairs (X, Y) whose path γ_{XY} passes through t. We encode (X, Y) as $\mathrm{enc}_t(X, Y) = X \oplus Y \oplus (Z \cup Z')$. By analogy with what we did for matchings, the idea here is that $\mathrm{enc}(X, Y)$ should agree with X on the portions of the trails and circuits already processed, and with Y elsewhere. This encoding is in fact slightly simpler than that for matchings since we don't need to impose the constraint that $\mathrm{enc}(X, Y)$ is a matching (it can be an arbitrary subgraph).

It is left as a straightforward **exercise** to check the following properties of this encoding:

(i) $\text{enc}_t : \text{paths}(t) \to \Omega$ is an injection for all t;

(ii) $w'(X)w'(Y) \le \xi^{-4} \min\{w'(Z), w'(Z')\}w'(\text{enc}_t(X, Y))$ for all $(X, Y) \in \text{paths}(t)$.

Property (i) is essentially immediate, as for matchings. For property (ii), note that the multisets of edges $X \uplus Y$ and $(Z \cup Z') \uplus \text{enc}_t(X, Y)$ are equal, so $\eta^{|X|}\eta^{|Y|} = \eta^{|Z \cup Z'|}\eta^{|\text{enc}_t(X,Y)|}$. For the ξ factors, you will need to note that the combined contributions from X, Y and from $Z, \text{enc}_t(X, Y)$ are the same at almost all vertices, with the possible exceptions of the start vertex of the current trail/circuit C_i and the two endpoints of the edge e that is being flipped by the transition t. The total discrepancy in the exponent of ξ arising from these vertices turns out to be at most 4, giving rise to the factor ξ^{-4} above.

The properties of the encoding guarantee, as in Claim 11.4 of Lecture 11, that the cost of our flow f satisfies $\rho(f) \le 2|E|\xi^{-4}$ (where the factor ξ^{-4} comes from property (ii) and the factor $2|E|$ from the inverse transition probability). Plainly the length of a longest flow-carrying path is $\ell(f) \le |E|$, and we chose the initial state $A = \emptyset$ to have maximum weight, so $\pi'(\emptyset) \ge 2^{-|E|}$. Putting all this together as in Corollary 11.2, we get the claimed bound $O(\xi^{-4}|E|^3)$ on the mixing time. $\qquad\square$

References

[JS93]   M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM Journal on Computing*, 22:1087–1116, 1993.

[LS12]   E. Lubetzky and A. Sly. Critical Ising on the square lattice mixes in polynomial time. *Communications in Mathematical Physics*, 313:815–836, 2012.

[MO94] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Communications in Mathematical Physics*, 161:447–486, 1994.

[NM53] G.F. Newell and E.W. Montroll. On the theory of the Ising model of ferromagnetism. *Reviews of Modern Physics*, 25:353, 1953.