Wrapping up from last time about Monte Carlo Markov chains: we proved that for $\lambda < \frac{1}{k}$, the Metropolis MC mixed rapidly (that is, $\tau_{\text{min}} < \text{Poly}(n)$). In fact, with a bit more work, we can show this for $\lambda < \frac{1}{2k}$.

What we did not address was what happens when $\lambda \gg \frac{1}{k}$, nor what can be said about lower bounds of $\tau_{\text{min}}$.

![Figure 1](image)

**Figure 1:** For large $\lambda$, two likely independent sets. Transitions between them are unlikely.

In fact, we can have situations where $\tau_{\text{min}} = \exp \Theta(n)$. As an example, consider Figure (1), which shows a likely independent set on such a grid, and its likely complement. Transitions between the two, however, are rather improbable.

1 Correlation Methods

The following techniques fall under the heading of “Correlation Methods,” and will be of some use.

- Computation Trees
- Spatial Mixing
- Dobrushin condition [Next Lecture]

1.1 Computation Trees

Consider a factor graph $G = (V, F, E)$ and vertex $i \in V$ (our example graph is Figure (2)). The computation tree $T_i(G)$ is pictured below in Figure (3) as an example:

Formally, $T_i(G)$ is the tree formed by all non-reversing paths on $G$ that start at $i$. It is endowed with a graph structure in the natural way: $i$ is the root of the tree, and a node appears above another node in the tree iff its path is a subpath of the other node. Figure (2) shows two such paths, which will be neighbors in the computation tree. Note that several paths may come to be identified with the same node in the original graph – in this case, in the computation tree the node is copied and given a new label (for example, $i$ maybe be copied to obtain $i', i''$, etc.).
Figure 2: An example factor graph $G$.

Figure 3: The computation tree $T_i(G)$ for the graph in Figure (2).

So then $T_i(G) = (V_i, F_i, E_i)$ is an infinite factor graph. It is a graph covering of $G$ – that is, we have a mapping $\pi : V_i \to V, F_i \to F$ that is onto and such that $(j, a) \in E_i$ iff $(\pi(j), \pi(a)) \in E$.

Denote by $T_i^{(t)}(G)$ the tree obtained by truncating $T_i(G)$ after its first $t$ generations, where in determining generations we are counting layers of variable nodes. For example, see Figure (5), which shows $T_i^{(2)}(G)$.

With such graphs we can associate a graphical model, with a specific boundary condition to be applied to the bottom-most layer of variable nodes.

We are putting the same compatibility functions as before:

$$\mu^{(t_i)}(x) = \frac{1}{Z_i^{(t_i)}} \prod_{a \in F_i} \psi_{\pi(a)}(x_{\partial a}).$$  \hspace{1cm} (1)

For $j$, a variable node at the $t$-th generation of $T_i^{(t)}(G)$, let $a(j)$ be its unique adjacent function node, i.e. its parent in that rooted tree.
A boundary condition is a collection of distribution over $\chi$ into the following:

$$\{ \eta_{j \to a(j)}(x_j) : j \in \partial T_i^{(t)}(G) \}. \tag{2}$$

We then define the graphical model with boundary condition corresponding to $\eta$ as below:

$$\mu({t,i})(x) = \frac{1}{Z_t} \prod_{a \in T_i^{(t)}(G)} \psi_{\pi(a)}(x_{\partial a}) \prod_{j \in \partial T_i^{(t)}(G)} \eta_{j \to a(j)}(x_j), \tag{3}$$

where $x = \{ x_j : j \in T_i^{(t)}(G) \}$.

**Proposition 1.** Let $\overline{\mu}_i^{(t)}(.)$ be the BP estimate for the marginals w.r.t $\mu_G(.)$ after $t$ BP iterations on $G$. If the boundary condition on $T_i^{(t+1)}(G)$ is taken to be $\nu_{j \to a(j)}^{(t_0)}(.) = \eta_{j \to a(j)}(.)$, then for $t_1 \geq 1, t_0 \geq 0$ we have

$$\overline{\mu}^{(t_0+t_1)}_i(x_i) = \mu^{(t_1,i)}(x_i), \tag{4}$$

where $\mu^{(t_1,i)}(x_i)$ is naturally the marginal corresponding to the root in $T_i^{(t_1)}(G)$.

**Proof** Let $j$ be at level $t_1 - s$ on $T_i(G)$, $a$ its parent, and call $\mu_{j \to a}^{T(j)}(x_j)$ the marginal for $x_j$ w.r.t the graphical model in the subtree $T(j)$ (see figure 6). We prove by induction that

$$\mu_{j \to a}^{T(j)}(x_j) = \overline{\mu}_{j \to a}^{(t_0+s)}(x_j). \tag{5}$$

For $s = 0$ it is just a consequence of the choice of boundary conditions.

If it is true for some $s$, assume that the level of $j$ is $t_1 - (s + 1)$. Then by the induction hypothesis we have the following (see figure 7).

$$\mu_{j \to a}^{T(j)}(x_j) \propto \prod_{b \in \partial j \setminus a} \left[ \sum_{x_{\partial b \setminus j}} \psi_b(x_{\partial b}) \prod_{l \in \partial b \setminus j} \mu_{l \to b}^{T(j)}(x_l) \right] \prod_{b \in \partial j \setminus a} \left[ \sum_{x_{\partial b \setminus j}} \psi_b(x_{\partial b}) \prod_{l \in \partial b \setminus j} \overline{\mu}_{l \to b}^{(t_0+s)}(x_l) \right] \propto \overline{\mu}_{j \to a}^{(t_0+s+1)}(x_j). \tag{6}$$
Now, repeat the same argument for the root and it completes the proof.

\[ \mu(t,x_i(z^{(t)})) - \mu(t,x_i(z^{(t)'}) \leq \delta(t) \] (7)

Then, for any \( t_1, t_2 \geq t \) we have
\[ |\pi_i^{(t_1)}(x_i) - \pi_i^{(t_2)}(x_i) | \leq \delta(t). \] (8)

In particular, if \( \delta(t) \to 0 \) as \( t \to \infty \), then belief propagation converges.

Corollary 2. If

\[ \sup_{z^{(t)},z^{(t)'}} |\mu^{(t,i)}(x_i|z^{(t)}) - \mu^{(t,i)}(x_i|z^{(t)'})| \leq \delta(t) \] (7)

then, for any \( t_1, t_2 \geq t \) we have
\[ |\pi^{(t_1)}_i(x_i) - \pi^{(t_2)}_i(x_i) | \leq \delta(t). \] (8)

In particular, if \( \delta(t) \to 0 \) as \( t \to \infty \), then belief propagation converges.

\[ \pi_i^{(t_1)}(x_i) - \pi_i^{(t_2)}(x_i) = \sum_{z^{(t)},z^{(t)'}} [\mu^{(t)}(x_i|z^{(t)}) - \mu^{(t)}(x_i|z^{(t)'})] | \pi_i^{(t_1)}(z^{(t)}) - \pi_i^{(t_2)}(z^{(t)}) | \] (9)

Now, let \( B_i(t) \) be the subgraph of \( G \) induced by the vertices whose distance from \( i \) is at most \( t \).

**Figure 5:** \( T_i^{(2)} \), the truncated computation tree with depth 2.

\(^1\)Note that here one can see a slight abuse of notation, where we should use \( \eta_{\pi(j) \to \pi(a(j))} \) instead of \( \eta_{j \to a(j)} \) to be completely accurate. We will use the simpler -though not very accurate form- in the rest of lecture.
Corollary 3. If $B_i(t)$ is a tree and inequality 7 holds, then

$$\left| \mu_i(x_i) - \overline{\mu}_i^{(t)}(x_i) \right| \leq \delta(t).$$

(10)

In particular, if $g$ is the girth\(^2\) of $G$, then we will have

$$\left| \mu_i(x_i) - \overline{\mu}_i^{(t)}(x_i) \right| \leq \delta\left(\frac{g - 1}{2}\right).$$

(11)

Proof. Observe that

$$\mu_i(x_i) = \sum_{x^{(t)}} \mu(x_i|x^{(t)}) \mu(x^{(t)}),$$

(12)

where $x^{(t)}$ is the set of vertices at distance $t$ from $i$ in $G$. Also notice that $\mu_G(x_i|x^{(t)}) = \mu_T(x_i|x^{(t)})$. Now, proceed as in the previous proof.

\(\Box\)

\(^2\)The girth of an undirected graph $G$ is defined by the length of the shortest cycle in it.
Figure 7: used in proof of proposition 1