Problem Set 2 Solutions

1. (a) In this case the random map $f$ is chosen u.a.r. from the set $\{f', f''\}$, where $f'(0) = 1$, $f'(1) = 0$ and $f''(0) = 1$, $f''(1) = 1$. Denote $F_{-t}^0 = f_{-1} \circ \cdots \circ f_{-t}$, where the $f_i$ are iid samples from the above distribution.

(b) Let $T = \min\{t : F_{-t}^0$ is constant $\}$ be the stopping time for CFTP, and let $Z_{-\infty}^0$ denote the constant value of $F_{-T}^0$. Then

$$\Pr[Z_{-\infty}^0 = 1] = \sum_{t \geq 1} \Pr[Z_{-\infty}^0 = 1|T = t] \Pr[T = t].$$

(1)

Now note that $f_{-T} = f''$, and $f_{-i} = f'$ for $1 \leq i < t$. Thus $\Pr[T = t] = 2^{-t}$, and $F_{-T}^0 = 1$ if $T$ is odd, and 0 otherwise. Hence (1) becomes

$$\Pr[Z_{-\infty}^0 = 1] = \sum_{t \geq 1, t \text{ odd}} 2^{-t} = \frac{2}{3},$$

as desired.

(c) Evidently the forward simulation $F_0^t = f_t \circ \cdots \circ f_1$ halts at the first time $T$ for which $f_T = f''$. But then clearly the constant value of $F_T^0$ is 1.

2. (a) Since both $F_0^t$ and $F_{-t}^0$ consist of the composition of $t$ independent random maps drawn from the same distribution, it is clear that the probabilities that they are constant are equal. Thus the distributions of $T$ and $T'$ are the same. (Note that this does not imply that the distributions of $F_T^0$ and $F_{-t}^0$ are the same!)

(b) Note that the random map $f$ defines a (standard pairwise) coupling via $(X, Y) \mapsto (f(X), f(Y))$. The coupling time for this coupling, maximized over $X$ and $Y$, is clearly dominated by the coalescence time $T'$, since if $F_0^t$ is constant then certainly coupling has occurred for all pairs. Hence our familiar coupling technology gives $D(t) \leq \Pr[T' > t]$. But by part (a) this latter probability is precisely $\Pr[T > t]$.

(c) Consider two copies $X_t, Y_t$ of the Markov chain started at $X_0 = \top$ and $Y_0 = \bot$ respectively. For any state $z$ let $h(z)$ denote the height of $z$, i.e., the length of a longest chain in the partial order whose top element is $z$, and let $H(X_t, Y_t) = h(X_t) - h(Y_t)$ denote the height difference between $X_t$ and $Y_t$. Note that $H(X_t, Y_t) \geq 0$ by monotonicity. Now clearly we have

$$\Pr[T > t] = \Pr[T' > t] = \Pr[H(X_t, Y_t) \geq 1] \leq \mathbb{E}[H(X_t, Y_t)] = \mathbb{E}[h(X_t) - h(Y_t)].$$

But this last expectation is just

$$\mathbb{E}[h(X_t)] - \mathbb{E}[h(Y_t)] \leq \|p^{(t)}_\top - p^{(t)}_\bot\| \max_z h(z) \leq D(t)h,$$

as claimed.

(d) By submultiplicativity of $D(t)$ and the definition of $\tau_{\text{mix}}$, setting $t = [\tau_{\text{mix}} \log(h\varepsilon^{-1})]$ ensures that $D(t) \leq h^{-1}\varepsilon$. Plugging this in to part (c) gives $\Pr[T > t] \leq hD(t) \leq \varepsilon$, as required.
3. (a) For a spin configuration \( \sigma \), let \( H_\sigma \) denote the subgraph consisting of all edges that connect neighbors with equal spins in \( \sigma \). Say that a subgraph \( H \) and spin configuration \( \sigma \) are compatible, written \( H \sim \sigma \), if \( H \subseteq H_\sigma \). (Thus \( H \sim \sigma \) iff no edge of \( H \) connects unequal spins.)

We now construct a probability distribution on compatible pairs \( (H, \sigma) \) in two different ways. First, for any given \( H \), note that there are \( 2^{\mathcal{C}(H)} \) compatible configurations \( \sigma \), corresponding to the two spin choices for each connected component of \( H \). We pick one of these at random by flipping a fair coin to select the spin of each component independently. If \( H \) itself was chosen at random from \( \pi \), the resulting distribution on pairs is

\[
\mu(H, \sigma) = \frac{1}{Z} p^{|H|} (1 - p)^{|E| - |H|} 2^{\mathcal{C}(H)} 2^{-\mathcal{C}(H)} = \frac{1}{Z} p^{|H|} (1 - p)^{|E| - |H|}.
\]

Secondly, for any given \( \sigma \), note that there are \( 2^{|H_\sigma|} = 2^{|E| - |U(\sigma)|} \) compatible subgraphs \( H \), namely all subgraphs of \( H_\sigma \). We pick one of these at random by including each edge of \( H_\sigma \) independently with probability \( p \). If \( \sigma \) was chosen at random from \( \hat{\pi} \), the resulting distribution on pairs is

\[
\hat{\mu}(H, \sigma) = \frac{1}{Z} \lambda^{U(\sigma)} p^{|H|} (1 - p)^{|E| - U(\sigma) - |H|} = \frac{1}{Z} p^{|H|} (1 - p)^{|E| - |H|},
\]

where we have used the suggested correspondence \( \lambda = 1 - p \).

But since both (2) and (3) are probability distributions and assign the same relative weight to each pair \( (H, \sigma) \), we must have \( Z = \hat{Z} \).

The procedures for mapping subgraphs to configurations and vice versa are as described above.

(b) The heat bath dynamics is specified by the following transition rule from a given subgraph \( H \):

- pick an edge \( e \in E \) and a real \( r \in [0, 1] \) u.a.r.
- include edge \( e \) in the new subgraph if \( r \leq p_e \), else do not include \( e \)

Here \( p_e \) is the conditional probability of edge \( e \), given the rest of \( H \) (i.e., if \( e \) is not a “cut edge” wrt \( H \) (so that \( H + e \) has the same number of connected components as \( H - e \) then \( p_e = p \), else \( p_e = p/(2(1-p)) = p/(2-p) \).

Now define a partial order \( \preceq \) on subgraphs by \( H \preceq H' \) iff \( H \supseteq H' \) (i.e., all edges of \( H' \) are edges of \( H \)). The unique minimal and maximal elements are \( G \) and \( \emptyset \) (the graph with vertex set \( V \) and no edges) respectively. A complete coupling is obtained by picking a single pair \( (e, r) \) in the above transition rule. To check this is monotone, suppose \( H \preceq H' \). It is sufficient to show that, for all edges \( e \), \( p_e \geq p_e' \), where \( p_e, p_e' \) are the thresholds in the above transition rule for \( H \) and \( H' \) respectively: this ensures that \( e \) cannot be included into \( H' \) but not into \( H \). Observe that \( p_e \) and \( p_e' \) can differ only if \( e \) is a cut edge in one of \( H, H' \) and not the other. Since \( H \preceq H' \), \( e \) must be a cut edge in \( H' \), whence

\[
p_e' = \frac{p}{2-p} \leq p = p_e,
\]

as required.

4. We first assume the following lemma, as suggested in the hint:

**Lemma:** For all states \( x, y \) and all \( t \geq 2\tau_{\text{mix}} \), \( \frac{p_{\pi(y)}(y)}{\pi(y)} \geq \frac{1}{8} \).

Now construct the following flow. Let \( t = 2\tau_{\text{mix}} \), and let \( P_{xy}^t \) denote the set of all paths of length exactly \( t \) from \( x \) to \( y \) in the Markov chain. Distribute the flow \( \pi(x)\pi(y) \) from \( x \) to \( y \) among the paths \( p \in P_{xy}^t \) in the proportions \( \text{prob}(p) \), where \( \text{prob}(p) \) denotes the probability of taking path \( p \) starting in state \( x \) conditional on ending in state \( y \) at time \( t \). Thus we are effectively letting the \( t \)-step evolution of the Markov chain choose the flow. (Note that some flow may be routed along non-simple paths \( p \); however, we can always make these paths simple by removing cycles, without increasing the cost of the flow. Thus we can safely ignore this issue.)
Call the resulting flow $f$. Then for any transition $e$, the flow along $e$ is given by

$$f(e) = \sum_{xy} \sum_{p \in P^+_{xy} \setminus \Omega} \frac{\pi(x) \pi(y) \text{prob}(p)}{p^e_x(y)} \leq 8 \sum_{xy} \sum_{p \in P^+_{xy} \setminus \Omega} \pi(x) \text{prob}(p),$$

where the inequality follows from the above lemma. But the final double sum is precisely the sum of the probabilities that the Markov chain, started in stationarity, traverses edge $e$ in each time step; this in turn is exactly $tQ(e)$, since the probability that the stationary chain traverses $e$ in any given step is $Q(e)$. Hence we have

$$\rho(f) = \max_e \frac{f(e)}{Q(e)} \leq 8t = 16\tau_{\text{mix}},$$

as required.

It remains to go back and prove the lemma.

**Proof of Lemma:** Define the set of states $S = \{z : p^e_z(z) \geq \frac{1}{2}\}$. It is easy to check that $|S| \geq \frac{1}{2}$, and hence by definition of the mixing time $p^\tau_x(S) \geq \frac{1}{4}$, where $\tau$ denotes $\tau_{\text{mix}}$. Thus we have

$$p^\tau_x(S) \geq \sum_{z \in S} p^\tau_x(z)p^\tau_z(y) \geq \pi(y) \sum_{z \in S} p^\tau_x(z) \frac{p^\tau_y(z)}{\pi(z)} \geq \frac{\pi(y)}{2} \sum_{z \in S} p^\tau_z(z) \geq \frac{\pi(y)}{8}.$$

This completes the proof of the lemma.

5. (a) Fix some arbitrary (non-self-loop) transition $e$ in the Markov chain; suppose that $e$ starts in a permutation $x$ and switches the cards in positions $i, j, k$, with $j > k$. Let $\text{paths}(e)$ denote the set of all paths $\gamma_{xy}$ that pass through $e$. We define an injective mapping $\eta_e : \text{paths}(e) \rightarrow \Omega$ as follows. For each pair $(x, y) \in \text{paths}(e)$, we refer to the positions of a given card in $x, y, z$ respectively as its initial, final and current positions. The permutation $\eta_e(x, y)$ is then defined as follows:

- place the cards $z_1, z_2, \ldots, z_k$ in their initial positions;
- place the remaining $n - k$ cards in the vacant positions in the order in which they appear in the final permutation $y$.

Clearly $\eta_e(x, y)$ belongs to $\Omega$. We need to check that it is injective. To see this, given $e$ and $\eta_e(x, y)$ we can uniquely recover $y$ by noting that the final positions of cards $z_1, \ldots, z_k$ are the same as their current positions, and the final order of the remaining cards can be read off from $\sigma_e(x, y)$. To recover $x$, note that the initial positions of $z_1, \ldots, z_k$ are just as in $\eta_e(x, y)$; but these positions determine, for every $i$, the current position of the card initially in position $i$, since each previous transition on the path involved moving one of $z_1, \ldots, z_k$ into its final position. Hence, by simulating these transitions, we can deduce the initial positions of all the cards and so recover $x$.

Since $\eta_e$ is injective, we deduce that $|\text{paths}(e)| \leq |\Omega|$, as required.

(b) From part (a) we may deduce that the flow through any transition $e$ is $f(e) \leq \frac{1}{|\Omega|}$. Moreover, the capacity of $e$ is $Q(e) = (2{(n) \choose 2}|\Omega|)^{-1}$. Hence the cost of the flow is $\rho(f) = \max_e \frac{f(e)}{Q(e)} \leq 2{(n) \choose 2} = O(n^2)$. Clearly, the length of a longest flow-carrying path is $\ell(f) = n$. And for any $x \in \Omega$ we have $\pi(x) = (n!)^{-1}$, so $\log \pi(x)^{-1} = O(n \log n)$. Plugging all this into our general bound on the mixing time from Lecture 10 gives

$$\tau_x(\varepsilon) \leq \rho(f)\ell(f)(2 \ln \varepsilon^{-1} + \ln \pi(x)^{-1}) = O(n^3(n \log n + \log \varepsilon^{-1})).$$