Homework 9 Solutions

Note: These solutions are not necessarily model answers. Rather, they are designed to be tutorial in nature, and sometimes contain a little more explanation than an ideal solution. Also, bear in mind that there may be more than one correct solution. The maximum total number of points available is 34.

1. (a) To see that the chain is irreducible, we need to show that, given any two spanning trees T_1, T_2, there exists a path of transitions that takes the chain from T_1 to T_2. To see this we use induction on the size of the symmetric difference $T_1 \oplus T_2$. The base case is when $|T_1 \oplus T_2| = 2$, i.e., each tree contains exactly one edge not in the other. Call these edges $e_1 \in T_1 \setminus T_2$ and $e_2 \in T_2 \setminus T_1$, respectively. To get from T_1 to T_2, we add e_2 to T_1, thus creating a cycle; we claim that e_1 must lie on this cycle, since T_2 can't contain the whole cycle (it's a tree) and so the only edge not in T_2 must lie on it. Therefore we can remove e_1 to arrive at T_2, as required. For the inductive step, suppose $|T_1 \oplus T_2| = 2s$ with $s > 1$, so that each tree contains exactly s edges not in the other. Pick an arbitrary edge $e \in T_2 \setminus T_1$ and add e to T_1; then by the same argument as above, the resulting cycle must contain an edge, say f, that is in $T_1 \setminus T_2$. So we can remove f to get a new tree T_1'. But notice that $|T_1' \oplus T_2| = 2(s-1)$, so by the induction hypothesis there is a path from T_1' to T_2, thus completing the proof. *4pts*

To see that the chain is aperiodic, it is enough to observe that it has self-loops: namely, whenever we add an edge e to T, we can always remove e again to get back to the same tree T.

Finally, we need to check that the stationary distribution is uniform. But this follows since the chain is symmetric: if a move switches edges e, f to take us from T to T', then switching f, e takes us back from T' to T. And the transition probabilities are the same, since both are equal to $\frac{1}{mc}$, where $m = |E| - (n-1)$ is the size of the edge set $E \setminus T$ and c is the number of edges in the cycle C in $T + e$. (Note that the cycle in $T' + f$ is also C.)

(b) We use the same transitions as in part (a), but we modify the transition probabilities as follows. We still pick the edge $e \in V \setminus T$ u.a.r., but now for each possible edge f on the resulting cycle we move to the tree $T_f := T + e - f$ with probability $\min\{1, \frac{\lambda_e}{\lambda_f}\}$; else we do nothing. (This is the same idea as in the Metropolis process that we discussed in class; see also MU Section 11.4.1. Note that $\frac{\lambda_e}{\lambda_f}$ is exactly $\frac{\pi(T')}{\pi(T)}$.) The resulting chain is obviously still ergodic (for the same reasons as in part (a)). To check that the new stationary distribution is indeed π, we check that the chain is *reversible* w.r.t. π, i.e., we check that the transition probabilities $P(\cdot, \cdot)$ satisfy the detailed balance conditions *4pts*

$$\pi(T)P(T,T') = \pi(T')P(T',T) \quad \forall T, T'. \tag{1}$$

To see this, consider the transition $T \to T' = T_f = T + e - f$, and assume w.l.o.g. that $\lambda_e \le \lambda_f$. Then by definition $P(T,T') = \frac{1}{mc} \times \frac{\lambda_e}{\lambda_f}$ and $P(T',T) = \frac{1}{mc} \times 1$, where $m = |E| - (n-1)$ and c is the length of the common cycle in $T + e$ and $T' + f$. Hence $\frac{P(T,T')}{P(T',T)} = \frac{\lambda_e}{\lambda_f} = \frac{\pi(T')}{\pi(T)}$ for all adjacent trees T, T', as required in equation (1). (For all non-adjacent T, T', both sides of (1) are zero.) Since the chain is ergodic and reversible w.r.t. π, as we saw in class π must be the unique stationary distribution.

2. Suppose we have an algorithm $\mathcal{A}$ satisfying the simpler condition, and suppose we are given inputs (x, ε, δ) for the fpras. Following the hint, we run $\mathcal{A}$ with inputs (x, ε) independently some number t times (to be determined) and let Z denote the *median* of the outputs of $\mathcal{A}$. We claim that taking $t = O(\log \delta^{-1})$ suffices *5pts* to ensure that Z satisfies the requirements of an fpras. To see this, call an output of $\mathcal{A}$ "good" if it falls in the range $[(1-\varepsilon)f(x),(1+\varepsilon)f(x)]$. Note that each output of $\mathcal{A}$ is good with probability at least $\frac{3}{4}$, and these events are mutually independent. The key observation is that the final output Z must be good unless at least $\frac{t}{2}$ of the trials are not good. (Equivalently, if more than half the trials are good, then the median must be good.) Thus the probability that Z is not good is bounded by the probability that, in t tosses of a biased coin with Heads probability $\frac{3}{4}$, at least half the tosses come up Tails. Let X be a binomial r.v. with parameters $(t,\frac{3}{4})$ and mean $\mu=\frac{3t}{4}$. We can bound the required probability using a Chernoff bound as follows:

$$\Pr[X \le \frac{t}{2}] = \Pr[X \le (1-\frac{1}{3})\mu] \le \exp\{-(\frac{(1/3)^2\mu}{2}\} = \exp\{-\frac{t}{18}\}.$$

Hence taking $t = 18\ln\delta^{-1} = O(\log\delta^{-1})$ makes this probability at most δ. The overall algorithm requires $t=O(\log\delta^{-1})$ trials of $\mathcal{A}$, each of which runs in time $\mathrm{poly}(|x|,\varepsilon^{-1})$, so the overall running time is polynomial in $|x|,\varepsilon^{-1}$ and $\log\delta^{-1}$, as required for an fpras.

3. (a) Clearly T is a stopping time, and the Optional Stopping Theorem (OST) holds because $|X_t|$ is bounded for all t. We may therefore conclude that $\mathrm{E}[X_T]=X_0=0$ for the stopping time T. Thus we have *4pts*

$$p\times \mathrm{E}[X_T \mid R] + (1-p)\times \mathrm{E}[X_T \mid L] = 0,$$

where R, L are the events that the process exits the interval at the right- and left-hand ends, respectively. Thus we can conclude that

$$p = \frac{-\mathrm{E}[X_T \mid L]}{\mathrm{E}[X_T \mid R] - \mathrm{E}[X_T \mid L]}. \tag{2}$$

(Note that this expression is positive because $\mathrm{E}[X_T \mid L] < 0$ and $\mathrm{E}[X_T \mid R] > 0$.)

Note that we don't know $\mathrm{E}[X_T \mid R]$ exactly because when the process exits at the right-hand end it may end up at any integer point in the range $[m+1, m+c]$. (It can't travel beyond $m+c$ because the jumps are bounded by c.) Hence all we can say is that $m+1 \le \mathrm{E}[X_T \mid R] \le m+c$. And similarly, at the left-hand end, we have $-(m+c) \le \mathrm{E}[X_T \mid L] \le -(m+1)$. Using these extremal values in (2) to maximize (respectively, to minimize) p, it is easy to check that we get the desired bounds $\frac{m+1}{2m+c+1} \le p \le \frac{m+c}{2m+c+1}$. (Note that these bounds are smaller and larger than $\frac{1}{2}$, respectively, since $c \ge 1$.)

(b) We compute *3pts*

$$\begin{aligned}
\mathrm{E}[Z_{t+1} \mid D_1,\ldots,D_t] &= \mathrm{E}[(X_t + D_{t+1})^2 - \alpha(t+1) \mid X_t] \\
&= X_t^2 - \alpha(t+1) + \mathrm{E}[D_{t+1}^2 \mid X_t] \\
&\ge Z_t - \alpha + \alpha \\
&\ge Z_t.
\end{aligned}$$

[In the second line here we used the fact that $\mathrm{E}[D_{t+1} \mid X_t] = 0$, and in the third line the fact that $\mathrm{E}[D_{t+1}^2 \mid X_t] \ge \alpha$. Notice how the lower bound on the variance of the jumps $\mathrm{E}[D_{t+1}^2 \mid X_t]$ is crucial to ensuring the submartingale property!] Thus we have proved that (Z_t) is a submartingale w.r.t. (D_t), as required.

(c) We apply the OST to the submartingale (Z_t), with the same stopping time T as before. The OST condition $\mathrm{E}[T] < \infty$ and $\mathrm{E}[|Z_{t+1} - Z_t| \mid D_1,\ldots,D_t] \le c'$ for some constant c' (depending on m) is satisfied. (The fact that $\mathrm{E}[T] < \infty$ follows from the fact that we can write a system of linear equations for the quantities $E_j :=$ expected time to exit the interval starting from position j. This system clearly *3pts* has a finite solution, so all E_j are finite.) The OST implies that $\mathrm{E}[Z_T] = Z_0 = 0$, which from the definition of Z_t immediately gives $\mathrm{E}[T] = \frac{\mathrm{E}[X_T^2]}{\alpha}$. Finally, by similar reasoning to part (a) we know that $\mathrm{E}[X_T^2] \leq (m+c)^2$, which in turn implies $\mathrm{E}[T] \leq \frac{(m+c)^2}{\alpha}$, as required. [Notice again how the lower bound α on the variance of the jumps comes into the expected time: the smaller the jumps, the longer the process takes to exit.]

4. (a) Fix the configuration on the graph after t steps, and assume $X_t \notin \{0, 2m\}$; let W_t, B_t denote the sets of white and black vertices respectively. Then we may write the difference $D_{t+1} = X_{t+1} - X_t$ as *4pts*

$$D_{t+1} = \sum_{u \in B_t} d_u C_u - \sum_{u \in W_t} d_u C_u, \tag{3}$$

where d_u is the degree of vertex u, and C_u is the indicator r.v. of the event that u changes color. Note that $\mathrm{E}[C_u] = \frac{\mathrm{disc}(u)}{2d_u}$, where $\mathrm{disc}(u)$ is the number of neighbors of u with the opposite color to u. Thus

$$\begin{aligned} \mathrm{E}[D_{t+1}|Y_t] &= \sum_{u \in B_t} d_u \times \frac{\mathrm{disc}(u)}{2d_u} - \sum_{u \in W_t} d_u \times \frac{\mathrm{disc}(u)}{2d_u} \\ &= \frac{1}{2}\Big(\sum_{u \in B_t} \mathrm{disc}(u) - \sum_{u \in W_t} \mathrm{disc}(u) \Big). \end{aligned}$$

But plainly the values of the two sums are equal, so $\mathrm{E}[D_{t+1}|Y_t] = 0$. Thus (X_t) is a martingale.

(b) Let T be the termination time, which is clearly a stopping time. We apply the Optional Stopping Theorem to the martingale in part (a) with this stopping time. To check the conditions for the OST, note that the martingale itself is bounded as $|X_t| \leq 2m$. The OST now gives $\mathrm{E}[X_T] = \mathrm{E}[X_0] = X_0$. So, letting p be the probability of termination in the all-white configuration, we have *3pts*

$$p \times 2m + (1-p) \times 0 = X_0,$$

and hence $p = \frac{X_0}{2m}$.

(c) Since (X_t) is a martingale on the integer interval $[0, 2m]$, we can use the same trick as in part (c) of Q3 above (and in class) to define a submartingale $Z_t := X_t^2 - \beta t$, where β is a lower bound on $\mathrm{E}[D_{t+1}^2|Y_t] = \mathrm{Var}[D_{t+1}|Y_t]$. Applying the OST to (Z_t) (using the stopping time T above and the same conditions as in Q3(c)), we get the usual conclusion that $\mathrm{E}[Z_T] \geq Z_0$. This in turn implies that $\mathrm{E}[X_T^2] - \beta \mathrm{E}[T] \geq X_0^2$, and therefore *4pts*

$$\mathrm{E}[T] \leq \frac{\mathrm{E}[X_T^2] - X_0^2}{\beta} = \frac{2mX_0 - X_0^2}{\beta} = \frac{X_0(2m - X_0)}{\beta} \leq \frac{m^2}{\beta}, \tag{4}$$

where we used part (b) to compute $\mathrm{E}[X_T^2] = p \times 4m^2 + (1-p) \times 0 = 2mX_0$.
Thus to prove that $\mathrm{E}[T] = O(m^2)$ it's enough to show that β is bounded below by some constant. We actually give a precise lower bound on β. From (3) and the fact that the C_u are independent given Y_t, we have $\mathrm{Var}[D_{t+1}|Y_t] = \sum_u \mathrm{Var}[d_u C_u]$. And for each u we may compute

$$\mathrm{Var}[d_u C_u] = \frac{\mathrm{disc}(u)}{2d_u} \times d_u^2 - \left(\frac{\mathrm{disc}(u)}{2d_u} \times d_u \right)^2 = \frac{\mathrm{disc}(u)}{4}(2d_u - \mathrm{disc}(u)) \geq \frac{\mathrm{disc}(u) d_u}{4}.$$

But clearly at any time before termination we must have $\mathrm{disc}(u) \geq 1$ for at least two vertices u, so $\mathrm{Var}[D_{t+1}|Y_t] \geq \frac{1}{4} \sum_u \mathrm{disc}(u) d_u \geq \frac{1}{2}$.
Finally, plugging $\beta = \frac{1}{2}$ into (4) gives $\mathrm{E}[T] \leq 2X_0(2m - X_0) \leq 2m^2$, as required.