Assignment 4 (3 problems)

Jamie Morgenstern *CSE515*

1. *[2 points]* Consider the uniform measure over proper q-colorings of a graph with $q \geq 3$. Namely, given a simple graph $G = (V, E)$, we define a probability distribution over $x = (x_1, \ldots, x_n)$ where $x_i \in \{1, \ldots, q\}$ as follows

$$\mu_G(x) = \frac{1}{Z(G,q)} \prod_{(i,j) \in E} \mathbb{I}(x_i \neq x_j)\,, \tag{1}$$

where $\mathbb{I}(\,\cdot\,)$ is the indicator function. Notice that $Z(G, q)$ is the number of ways of coloring the vertices of G so that no edge has both endpoints of the same color. Throughout this problem, $|V| = n$ and we will assume without loss of generality $V = \{1, 2, \ldots, n\}$.

   (*a*) Gibbs sampling for proper colorings proceeds as follows. Initialize with $x = (x_i)_{i \in V}$ which is a proper coloring of G. At each step draw a uniformly random vertex i in V. Change its color x_i to a new value x^{new} which is uniformly random among all the colors that are not taken by neighbors of i.

   For $n = 1000$, let G be the three dimensional torus of side length 10. In other words $V = [10] \times [10] \times [10]$ and $(i, j) \in E$ with $i = (i_1, i_2, i_3)$, $j = (j_1, j_2, j_3)$ if and only if either $i_1 = j_1$, $i_2 = j_2$, and $(i_3 - j_3) \in \{+1, -1\}$ modulo 10, or $i_1 = j_1$, $i_3 = j_3$, and $(i_2 - j_2) \in \{+1, -1\}$ modulo 10, or $i_2 = j_2$, $i_3 = j_3$, and $(i_1 - j_1) \in \{+1, -1\}$ modulo 10.

   Write a program that implements Gibbs sampling on G. Consider $q \in \{3, 5, 7, 9, 11\}$ and notice that in this case G is obviously q-colorable, and that the initial coloring can be found efficiently (e.g., alternating two colors). **Compute** in a simulation the empirical fraction of unchanged colors

$$C(t) = \frac{1}{|V|} \sum_{i \in V} \mathbb{I}\big(x_i(0) = x_i(t)\big)\,,$$

   where t is the number of iterations. **Plot** $C(t)$ as a function of t up to $t = 100,000$ for each of the above cases (you need to run Gibbs sampling once for each value of q).

   For what values of q does $x(t)$ approximately converge to the stationary distribution (i,e. $x(t)$ independent of $x(0)$), exponentially fast?

   Provide your program by appending it to the end of your solution pdf file.

   (*b*) In this problem, we use path coupling to bound the mixing time of the above Markov chain, for q colors and graphs of maximum degree k. Specifically, provide an upper bound on

$$\mathbb{E}[D(X^{(t+1)}, Y^{(t+1)}) | D(X^{(t)}, Y^{(t)}) = 1]\,, \tag{2}$$

   for Hamming distance $D(\cdot, \cdot)$ and two coupled Markov chains $X^{(t)}$ and $Y^{(t)}$. Let $X^{(t)}$ and $Y^{(t)}$ differ in one index i. Similarly as in the path coupling example from the lectures, when the randomly chosen index $I^{(t)} = i$, then $D(X^{(t+1)}, Y^{(t+1)}) = 0$.

Otherwise, if $I^{(t)}$ is not in the neighborhood of i, then $D(X^{(t+1)}, Y^{(t+1)}) = 1$.

We are only left to analyze the process when $I^{(t)}$ is in the neighborhood of i. First, **show that** for some integer m, if $X^{(t+1)}_{I^{(t)}}$ (conditioned on its neighbors' colors) is uniformly distributed in $\{1, 2, \ldots, m\}$ and $Y^{(t+1)}_{I^{(t)}}$ (conditioned on its neighbors' colors) is uniformly distributed in $\{1, 2, \ldots, m+1\}$, then under the optimal coupling we have

$$\mathbb{E}[D(X^{(t+1)}_{I^{(t)}}, Y^{(t+1)}_{I^{(t)}})] \;=\; \mathbb{P}(X^{(t+1)}_{I^{(t)}} \neq Y^{(t+1)}_{I^{(t)}}) \;=\; \frac{1}{m+1}\;.$$

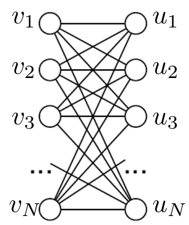
Now, similarly, if $X^{(t+1)}_{I^{(t)}}$ (conditioned on its neighbors' colors) is uniformly distributed in $\{1, 2, \ldots, m\}$ and $Y^{(t+1)}_{I^{(t)}}$ (conditioned on its neighbors' colors) is uniformly distributed in $\{2, \ldots, m+1\}$, then under the optimal coupling we have

$$\mathbb{E}[D(X^{(t+1)}_{I^{(t)}}, Y^{(t+1)}_{I^{(t)}})] \;=\; \mathbb{P}(X^{(t+1)}_{I^{(t)}} \neq Y^{(t+1)}_{I^{(t)}}) \;=\; \frac{1}{m}\;.$$

Use this optimal coupling to **provide an upper bound** on (2), and use the worst case choice of m for the given problem instance q and k. You can assume $q > k$, to simplify the calculation.

According to your analysis, how big does q need to be for a given k if we want fast mixing?

2. *[2 points]* **(Cheeger's inequality)**
In this problem, we use the Cheeger's inequality from class to upper bound the mixing time of a Markov chain by lower bounding the conductance of the Markov chain. Consider a distribution over matchings in a graph. A *matching* in a graph $G = (V, E)$ is a subsets of edges such that no two edges share a vertex. Here we focus on the special case of a complete bipartite graph G with vertices $v_1, \ldots, v_N$ on the left and $u_1, \ldots, u_N$ on the right, as shown:



In such a graph, a *perfect matching* is a matching which includes N edges. We are interested in sampling from a distribution over perfect matchings. We can denote a perfect matching using the variables $\sigma = [\sigma_{ij}] \in \{0,1\}^{N\times N}$, where $\sigma_{ij} = 1$ is v_i and u_j are matched and $\sigma_{ij} = 0$ otherwise. Observe that σ is a perfect matching if and only if

$$\sum_{k=1}^{N} \sigma_{ik} = 1 \qquad \text{for all } 1 \le i \le N$$
$$\sum_{k=1}^{N} \sigma_{kj} = 1 \qquad \text{for all } 1 \le j \le N$$

A perfect matching σ can also be thought of as a permutation $\sigma : \{1, \ldots, N\} \to \{1, \ldots, N\}$. For example, if $\sigma_{12} = \sigma_{21} = \sigma_{33} = 1$, this would correspond to the permutation $\sigma(1) = 2, \sigma(2) = 1$, and $\sigma(3) = 3$.

Consider the distribution defined by a set of weights on the edges $w_{ij} \ge 0$ for all i and j such that

$$\begin{aligned} \mu(\sigma) &\propto \exp\Big\{ \sum_{i,j} w_{ij}\sigma_{ij} \Big\} \mathbb{I}(\sigma \text{ is a perfect matching}) \\ &= \exp\Big\{ \sum_{i} w_{i\sigma(i)} \Big\} \mathbb{I}(\sigma \text{ is a perfect matching}) \,. \end{aligned}$$

(*a*) First, in this part, consider the uniform distribution over perfect matchings, i.e., $w_{ij} = 0$ for all i, j. Describe a simple procedure to sample σ from this uniform distribution.

(*b*) Now for the weighted distribution, show that for any perfect matching σ,

$$\mu(\sigma) \;\ge\; \frac{1}{N!\exp(Nw^*)} \,,$$

where $w^* = \max_{i,j} w_{ij}$.

(*c*) Consider the Metropolis-Hastings rule defined by: choose $i, i' \in \{1, \ldots, N\}$ uniformly at random. If $i = i'$, do nothing, otherwise with probability

$$R = \min\big\{ 1 \,,\, \exp(w_{i\sigma(i')} + w_{i'\sigma(i)} - w_{i\sigma(i)} - w_{i'\sigma(i')}) \big\}$$

swap $\sigma(i)$ and $\sigma(i')$, i.e. define a new permutation σ' such that $\sigma'(j) = \sigma(j)$ for $j \ne i, i'$ and $\sigma'(i) = \sigma(i')$ and $\sigma'(i') = \sigma(i)$.

Show that, under this Markov chain, for any valid transition $\sigma \to \sigma'$,

$$\begin{aligned} \mathbb{P}_{\sigma,\sigma'} &= \mathbb{P}(\text{ next state is } \sigma' \mid \text{currect state is } \sigma\ ) \\ &\geq \frac{1}{N^2\,\exp(2w^*)}\,. \end{aligned}$$

(d) For the conductance of this Markov chain, argue using (b) and (c) that

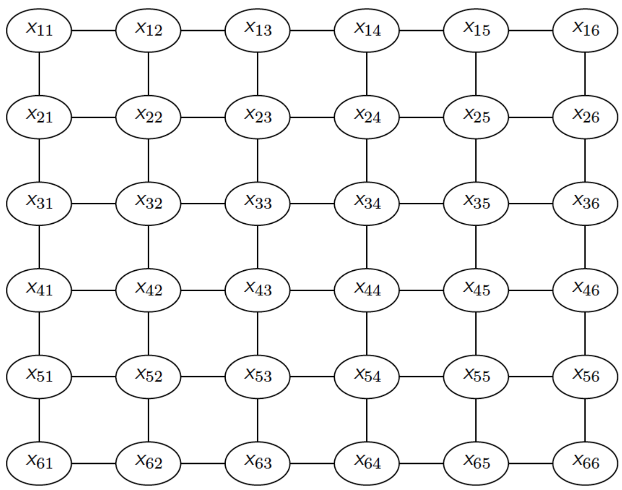
$$\begin{aligned} \Phi &= \min_S \frac{\sum_{\sigma\in S,\sigma'\in S^c}\mu(\sigma)\mathbb{P}_{\sigma,\sigma'}}{\mu(S)\,\mu(S^c)} \\ &\geq \frac{1}{N!\,N^2\,\exp((N+2)w^*)}\,, \end{aligned}$$

where S is a set states (or matchings), S^c is the complement of S, and $\mu(S) = \sum_{\sigma\in S}\mu(\sigma)$.

(e) Using (d), obtain a bound on the mixing time of the Markov chain.

3. *[2 points]* **(Block Gibbs sampling; implementation)**
In this problem, we develop an efficient algorithm for sampling from a two-dimensional Ising model building on the naive Gibbs sampling. In particular, suppose all variables x_{ij} take values in $\{+1, -1\}$. Using the graph structure G shown below, define the distribution

$$\mu_\theta(x) \quad = \quad \frac{1}{Z_\theta} \exp\Big\{ \sum_{(ij,kl)\in E} \theta x_{ij} x_{kl} \Big\} \, .$$

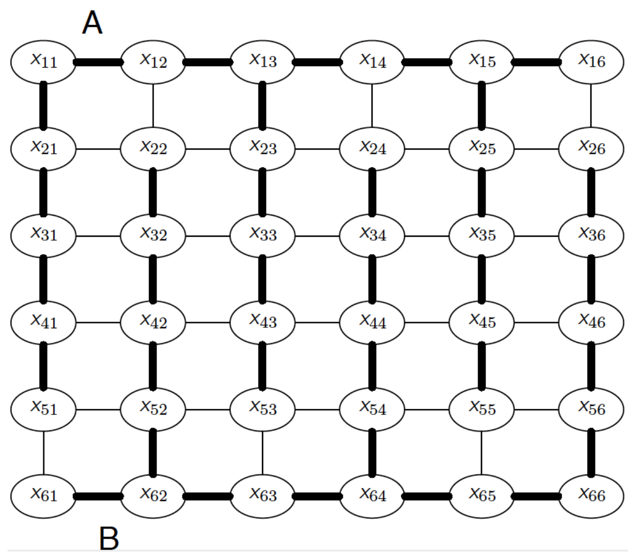


(*a*) Derive the update rules for a node-by-node Gibbs sampler for this model. Implement the sampler in Matlab and run it for 3,600,000 iterations on an Ising model of size 60×60 with coupling parameter $\theta = 0.45$. Use uniformly random initialization of $x_{ij} = +1$ with probability 0.5 and $x_{ij} = -1$ otherwise. Show one instance of the state of the variables after every 360,000 iterations. For a 60×60 matrix $x \in \{-1, +1\}^{60\times 60}$, you can use MATLAB commands `imagesc(x);colormap gray;axis off;` to display the state x.

(*b*) Suppose we are given a tree-structured undirected graphical model T with variables $y = (y_1, \ldots, y_N)$. Give an efficient procedure for sampling from the joint $\mu(y)$.

(*c*) In *block Gibbs sampling*, we partition a graph into r subsets $A_1, \ldots, A_r$. In each iteration, for each A_i, we sample x_{A_i} from the conditional distribution $\mu(x_{A_i} | x_{V\setminus A_i})$. For the Ising model G described above, consider the two comb-shaped subsets A and B shown below. Describe how to use your sampler from part (b) to perform the block Gibbs updates. (For this part, you may assume a black-box implementation of your sampling procedure from part (b).) .

(*d*) We provide an implementation of the block Gibbs sampler from part (*c*) in `comb_gibbs_step.m`, `comb_sum_product.m`, `ising_gibbs_comb.m`. As in part (*a*), we set $\theta = 0.45$ and run the sampler for 1000 iterations updating A and then B at every iteration. Run the block Gibbs sampler in `ising_gibbs_comb.m` and analyze the state of the variables after every 100 iterations. Which of the two samplers appears to mix faster?


A
x_{11}
x_{12}
x_{13}
x_{14}
x_{15}
x_{16}
x_{21}
x_{22}
x_{23}
x_{24}
x_{25}
x_{26}
x_{31}
x_{32}
x_{33}
x_{34}
x_{35}
x_{36}
x_{41}
x_{42}
x_{43}
x_{44}
x_{45}
x_{46}
x_{51}
x_{52}
x_{53}
x_{54}
x_{55}
x_{56}
x_{61}
x_{62}
x_{63}
x_{64}
x_{65}
x_{66}
B