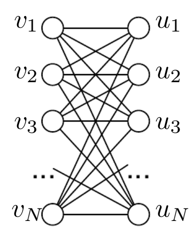
IE 598 Inference in Graphical Models

Homework 4

Problem 4.1 [Optional] (Sampling) 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

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

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 $w_{ij} \geq 0$ 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,

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

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

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

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

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

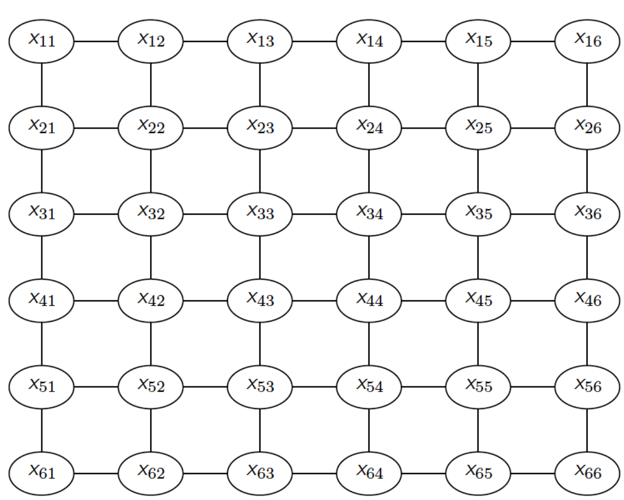
$$\begin{aligned} \Phi &= \min_S \frac{\sum_{\sigma \in S, \sigma' \in S^c} \mu(\sigma)\mathbb{P}_{\sigma,\sigma'}}{\mu(S)\,\mu(S^c)} \\ &\geq \frac{1}{N!\,N^2 \exp((N+2)w^*)}\,, \end{aligned}$$

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.

Problem 4.2 (Sampling) 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

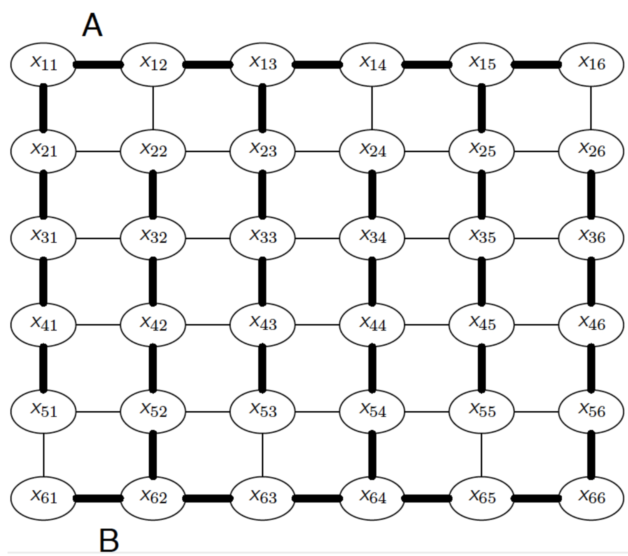
$$\mu_\theta(x) = \frac{1}{Z_\theta} \exp\Big\{ \sum_{(ij,kl) \in E} \theta x_{ij} x_{kl} \Big\}\,.$$



(*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?