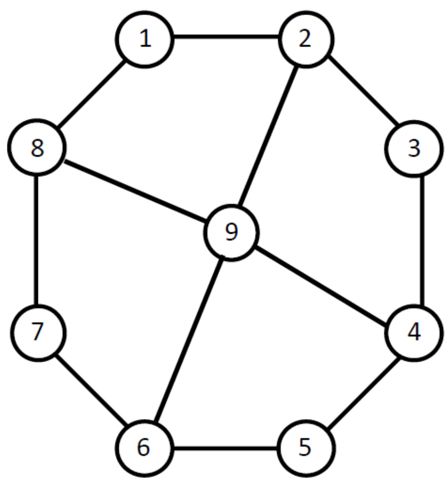
Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.438 Algorithms for Inference
Fall 2018

Problem Set 2

Issued: Thursday, Sept. 20, 2018 **Due:** Thursday, Sept. 27, 2018

Problem 2.1 (practice)
Recall that a graph is said to be *chordal* if every cycle of length 4 or more in the graph has a 'shortcut' or 'chord', i.e. if some 2 non-consecutive vertices in the cycle are connected by an edge. We call the process of converting a graph into a chordal graph by adding edges as *triangulation*. Triangulate the graph below, using as few edges as possible, by running the elimination algorithm.



Problem 2.2

Let $\mathcal{G}$ be an undirected graph, and u a vertex of $\mathcal{G}$. Consider running the elimination algorithm on $\mathcal{G}$ to compute the marginal $p_u(\cdot)$. Let $\mathcal{S}_i$ be the set of all nodes (not including i and previously eliminated nodes) that share an active potential with node i, at the time of eliminating node i. Recall that the list of active potentials contains all potentials (including input potentials and intermediate computations) which have not yet been processed. Assume that there is one input potential for each maximal clique of $\mathcal{G}$. Let $\mathcal{H}$ be the resulting reconstituted graph.

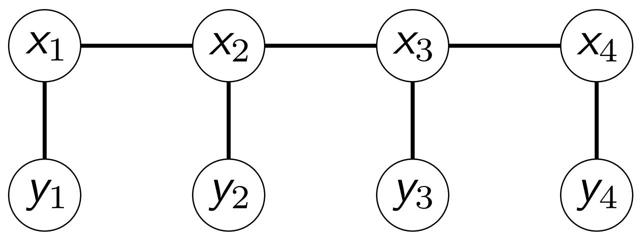
(a) Prove that the largest clique size in $\mathcal{H}$ is $\max_i |\mathcal{S}_i| + 1$.

(b) Prove that $\mathcal{H}$ is chordal.

Problem 2.3

In this problem, we consider the elimination algorithm for inference and its computational complexity.

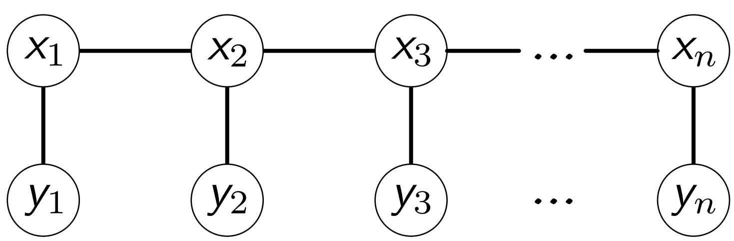
(a) Consider the following graph on 8 nodes.



Draw the reconstituted graph induced by the elimination ordering

$$(\mathsf{x}_4, \mathsf{y}_4, \mathsf{y}_3, \mathsf{x}_3, \mathsf{x}_1, \mathsf{x}_2, \mathsf{y}_2, \mathsf{y}_1).$$

(b) Now consider a graph on $2n$ nodes as drawn in the following figure, in which every random variable takes on values from a finite alphabet of size k. (That is, $\forall i \in \{1, \dots, n\},\ \mathsf{x}_i \in \mathcal{X},\ \mathsf{y}_i \in \mathcal{Y}$ and $|\mathcal{X}| = |\mathcal{Y}| = k$.)



Describe an elimination ordering that requires the least computation and one that requires the most computation. Determine the asymptotic time complexity of the algorithm for each of these orderings with respect to k and n.

(c) **(Practice)** Give an example of an undirected graph on n nodes such that the maximal clique size is constant with respect to n (i.e., the maximal clique size does *not* depend on n), but where the computation time required to perform elimination with any ordering is proportional to $k^{\alpha n}$, where k is the size of the alphabet. Specify the value of α for your example. Note that depending on the elimination ordering, your graph may have different values of α. However, your graph should be such that α is lower-bounded by some positive constant across all elimination orderings.

Problem 2.4

This is a continuation of Problem 1.4, so refer to the original problem and solution for the graphical model descriptions which define the node and edge potentials. Let there be K different coins, each with different biases, $c_1, \ldots, c_K \in [0,1]$. The k^{th} coin comes up heads with probability c_k, and tails with probability $1 - c_k$. Let $\mathsf{t} \in \{1, 2, \ldots K\}$ be a random variable having the mass function $p_{\mathsf{t}}(\cdot)$. Then define the random variable $\mathsf{x} \in \{0,1\}$ to be 1 if the t^{th} coin comes up heads, and 0 if the t^{th} coin comes up tails. In other words, to generate a sample of x, we first sample t, then toss coin number t, and set x equal to the indicator function of coin t coming up heads.

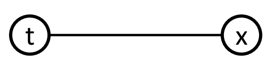


Figure 2.4-1

(a) Compute $p_{\mathsf{x}}(1)$ using the description in Figure 2.4-1. Recall the node and edge potentials defined in problem 1.4. Explain how this can be done using sum-product algorithm.

We are also given K (known) N-dimensional vectors $\Theta^1, \ldots \Theta^K \in \mathbb{R}^N$. If $\mathsf{t} = k$, we generate random variables $\mathbf{y} = (\mathsf{y}_1, \ldots, \mathsf{y}_N) \in \{0,1\}^N$ according to

$$p_{\mathbf{y}}(\mathbf{y}) \propto \exp\big(\sum_{i=1}^{N} \theta_i^k y_i\big), \tag{1}$$

for $\mathbf{y} = (y_1, \ldots, y_N) \in \{0,1\}^N$, where we used the notation $\Theta^k = (\theta_1^k, \ldots, \theta_N^k)$.

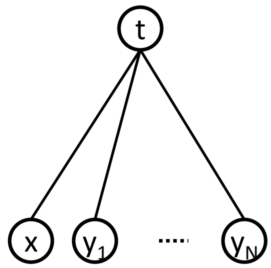


Figure 2.4-2

(b) Describe how the sum-product algorithm computes $p_{\mathsf{x}|\mathbf{y}}(1|\mathbf{y})$ given observation $\mathbf{y} = \mathbf{y}$, using the description in Figure 2.4-2. Recall the node and edge potentials defined in problem 1.4.

(c) Describe how the sum-product algorithm computes $p_{\mathsf{y}_1}(1)$, using the description in Figure 2.4-2. Recall the node and edge potentials defined in problem 1.4.

Problem 2.5
Consider the sum-product algorithm on an undirected tree with potential functions φ_s and ψ_{st}. Consider any initialization of messages such that $m_{t\to s}(x_s) > 0$ for all directions $t \to s$ and all states x_s.

(a) In this subproblem, we will prove by induction that the sum-product algorithm, with the parallel schedule, converges in at most diameter of the graph iterations. (Diameter of the graph is the length of the longest path.)

(i) For $D = 1$, the result is immediate. Consider a graph of diameter D. At each time step, the message that each of the leaf nodes sends out to its neighbors is constant because it does not depend on messages from any other nodes. Construct a new undirected graphical model G' by stripping each of the leaf nodes from the original graph. How should the potentials be redefined so that the messages along the remaining edges will be the same in both graphs?

(ii) Argue that G' has diameter strictly less than $D - 1$.

(iii) Thus, after at most $D-2$ time steps, the messages will all converge. Show that after "placing back" the leaf nodes into G' and running one more time step, each message will have converged to a fixed point.

(b) Prove by induction that the message fixed point m^* satisfies the following property: For any node t and $s \in N(t)$, let T_s be the tree rooted at s after the edge (s,t) is removed. Then

$$m^*_{s\to t}(x_t) = \sum_{\{x_v | v \in T_s\}} \psi(x_s, x_t) \prod_{v \in T_s} \varphi(x_v) \prod_{(i,j) \in T_s} \psi(x_i, x_j).$$

Hint: Induct on the depth of the subtree and use the definition of $m^*_{s\to t}(x_t)$.

(c) Use part (b) to show that

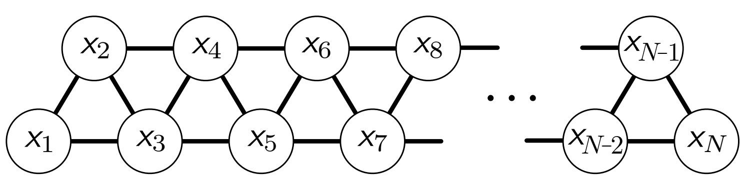
$$p(x_s) \propto \varphi_s(x_s) \prod_{t \in N(s)} m^*_{t\to s}(x_s)$$

for every node of the tree.

(d) Show that for each edge $(s,t) \in E$, the message fixed point m^* can be used to compute the pairwise joint distribution over (x_s, x_t) as follows:

$$p(x_s, x_t) \propto \varphi_s(x_s)\varphi_t(x_t)\psi_{st}(x_s, x_t) \prod_{u \in N(s)\backslash t} m^*_{u\to s}(x_s) \prod_{v \in N(t)\backslash s} m^*_{v\to t}(x_t).$$

Problem 2.6 (practice)
Consider the following undirected graphical model over N discrete random variables:



We define clique potentials $\psi_{i,i+1,i+2}(x_i, x_{i+1}, x_{i+2})$ for each triplet of random variables (x_i, x_{i+1}, x_{i+2}) for $i \in \{1, \ldots, N-2\}$.

(a) We run the elimination algorithm on this graph with the ordering $(\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_N)$. Give the list of active potentials just before eliminating node i.

(b) In a message-passing scheme for marginalization on this graph, each node $i \in \{1, \ldots, N-1\}$ sends a forward message τ_i as follows:

$$\tau_1(x_2, x_3) = \sum_{x_1} f_a(\psi_{1,2,3}(x_1, x_2, x_3)),$$

$$\tau_i(x_{i+1}, x_{i+2}) = \sum_{x_i} f_b(\tau_{i-1}(x_i, x_{i+1})\ ,\ \psi_{i,i+1,i+2}(x_i, x_{i+1}, x_{i+2})), \quad i \in \{2, \ldots, N-2\},$$

$$\tau_{N-1}(x_N) = \sum_{x_{N-1}} f_c(\tau_{N-2}(x_{N-1}, x_N)).$$

Determine functions f_a, f_b, and f_c so that $p_{\mathsf{x}_n}(x_N) \propto \tau_{N-1}(x_N)$.

(c) Each node $i \in \{N, N-1, \ldots, 2\}$ also sends a backward message η_i as follows:

$$\eta_N(x_{N-1}, x_{N-2}) = \sum_{x_N} g_a(\psi_{N-2,N-1,N}(x_{N-2}, x_{N-1}, x_N)),$$

$$\eta_i(x_{i-1}, x_{i-2}) = \sum_{x_i} g_b(\eta_{i+1}(x_i, x_{i-1})\ ,\ \psi_{i-2,i-1,i}(x_{i-2}, x_{i-1}, x_i)), \quad i \in \{N-1, \ldots, 3\},$$

$$\eta_2(x_1) = \sum_{x_2} g_c(\eta_3(x_2, x_1)).$$

Determine functions g_a, g_b, and g_c so that $p_{\mathsf{x}_1}(x_1) \propto \eta_2(x_1)$.

(d) We compute the remaining marginal distributions from these messages as follows:

$$p_{\mathsf{x}_2} \propto h_a(\tau_1\ ,\ \eta_3\ ,\ \eta_4),$$

$$p_{\mathsf{x}_i} \propto h_b(\tau_{i-2}\ ,\ \tau_{i-1}\ ,\ \eta_{i+1}\ ,\ \eta_{i+2}), \qquad i \in \{3, \ldots, N-2\},$$

$$p_{\mathsf{x}_{N-1}} \propto h_c(\tau_{N-3}\ ,\ \tau_{N-2}\ ,\ \eta_N).$$

Determine functions h_a, h_b, and h_c.

(e) Express, in $O(\cdot)$ notation, the minimal complexity of obtaining all singleton marginals via the algorithm we designed in parts (b)-(d). Express your answer in terms of the number of variables N and the alphabet size k. Assume that all N variables are defined over the same alphabet.