Homework 3 (100 points)

Please complete all calculations and write-up on your own. You must not consult homework solutions from any source, and you must show work to get full credit.

1. (10 points) Exercise 5.2
   **Exercise 5.2.** Consider a pairwise Markov network defined on binary variables:
   
   \[ p(x) = \phi(x_1, x_{100}) \prod_{i=1}^{99} \phi(x_i, x_{i+1}) \]  
   \[ (5.7.2) \]

   Is it possible to compute \( \arg \max_{x_1, \ldots, x_{100}} p(x) \) efficiently?

2. (15 points) Exercise 5.4
   **Exercise 5.4.** Consider the hidden Markov model:
   
   \[ p(v_1, \ldots, v_T, h_1, \ldots, h_T) = p(h_1)p(v_1|h_1) \prod_{t=2}^{T} p(v_t|h_t)p(h_t|h_{t-1}) \]  
   \[ (5.7.3) \]

   in which \( \text{dom}(h_t) = \{1, \ldots, H\} \) and \( \text{dom}(v_t) = \{1, \ldots, V\} \) for all \( t = 1, \ldots, T \).

   1. Draw a belief network representation of the above distribution.
   2. Draw a factor graph representation of the above distribution.
   3. Use the factor graph to derive a Sum-Product algorithm to compute marginals \( p(h_t|v_1, \ldots, v_T) \). Explain the sequence order of messages passed on your factor graph.
   4. Explain how to compute \( p(h_t, h_{t+1}|v_1, \ldots, v_T) \).
   5. Show that the belief network for \( p(h_1, \ldots, h_T) \) is a simple linear chain, whilst \( p(v_1, \ldots, v_T) \) is a fully connected cascade belief network.

3. (25 points) Exercise 5.7
   **Exercise 5.7.** A special time-homogeneous hidden Markov model is given by
   
   \[ p(x_1, \ldots, x_T, y_1, \ldots, y_T, h_1, \ldots, h_T) = p(x_1|h_1)p(y_1|h_1)p(h_1) \prod_{t=2}^{T} p(h_t|h_{t-1})p(x_t|h_t)p(y_t|h_t) \]  
   \[ (5.7.5) \]

   The variable \( x_t \) has 4 states, \( \text{dom}(x_t) = \{A, C, G, T\} \) (numerically labelled as states 1,2,3,4). The variable \( y_t \) has 4 states, \( \text{dom}(y_t) = \{A, C, G, T\} \). The hidden or latent variable \( h_t \) has 5 states, \( \text{dom}(h_t) = \{1, \ldots, 5\} \).

   The HMM models the following (fictitious) process:

   In humans, Z-factor proteins are a sequence on states of the variables \( x_1, x_2, \ldots, x_T \). In bananas Z-factor proteins are also present, but represented by a different sequence \( y_1, y_2, \ldots, y_T \). Given a sequence \( x_1, \ldots, x_T \), in a human, the task is to find the corresponding sequence \( y_1, \ldots, y_T \) in the banana by first finding the most likely joint latent sequence, and then the most likely banana sequence given this optimal latent sequence. That is, we require

   \[ \arg \max_{y_1, \ldots, y_T} p(y_1, \ldots, y_T|h_1^*, \ldots, h_T^*) \]  
   \[ (5.7.6) \]

   where

   \[ h_1^*, \ldots, h_T^* = \arg \max_{h_1, \ldots, h_T} p(h_1, \ldots, h_T|x_1, \ldots, x_T) \]  
   \[ (5.7.7) \]

   The file banana.mat contains the emission distributions \( p(x|h) \), \( p(y|h) \) and transition \( p(h|h_{t-1}) \). The initial hidden distribution is given in \( \phi_1 \) (\( p(h_1) \)). The observed \( x \) sequence is given in \( x \).
1. Explain mathematically and in detail how to compute the optimal y-sequence, using the two-stage procedure as stated above.

2. Write a MATLAB routine that computes and displays the optimal y-sequence, given the observed x-sequence. Your routine must make use of the Factor Graph formalism.

3. Explain whether or not it is computationally tractable to compute

\[
\arg \max_{y_1, \ldots, y_T} p(y_1, \ldots, y_T | x_1, \ldots, x_T)
\]  

(5.7.8)

4. (10 points) Exercise 6.2

**Exercise 6.2.** Consider the following distribution:

\[
p(x_1, x_2, x_3, x_4) = \phi(x_1, x_2) \phi(x_2, x_3) \phi(x_3, x_4)
\]  

(6.12.1)

1. Draw a clique graph that represents this distribution and indicate the separators on the graph.

2. Write down an alternative formula for the distribution \(p(x_1, x_2, x_3, x_4)\) in terms of the marginal probabilities \(p(x_1, x_2), p(x_2, x_3), p(x_3, x_4), p(x_2), p(x_3)\) and demonstrate the absorption process.

5. (20 points) Let \(X\) be the set of random variables in a clique graph \(G=\{C(G), W(G)\}\), where \(C(G)\) is the collection of the clique nodes in \(G\) and \(W(G)\) is the collection of separator nodes. Given a spanning tree \(T=\{C(T), W(T)\}\) of the clique graph \(G\), we define the total weight of \(T\)

\[
\text{weight}(T) = \sum_{S \in W(T)} |S|
\]

a. Show that \(\text{weight}(T) = \sum_{x \in X} \sum_{S \in W(T)} 1_S(x)\)

b. If \(J\) is a junction tree spanning \(G\), show that \(\text{weight}(J) = \sum_{Q \in C(J)} |Q| - |X|\)

c. If \(J\) is a junction tree spanning \(G\), \(J\) must be a maximum spanning tree.

6. (20 points) Exercise 6.4

**Exercise 6.4.** Consider the distribution

\[
p(a, b, c, d, e, f, g, h, i) = p(a)p(b|a)p(c|a)p(d|a)p(e|b)p(f|c)p(g|d)p(h|e, f)p(i|f, g)
\]  

(6.12.3)

1. Draw the belief network for this distribution.

2. Draw the moralised graph.

3. Draw the triangulated graph. Your triangulated graph should contain cliques of the smallest size possible.

4. Draw a junction tree for the above graph and verify that it satisfies the running intersection property.

5. Describe a suitable initialisation of clique potentials.

6. Describe the absorption procedure and write down an appropriate message updating schedule.