Computational Biology: Assignment #7

Due on Monday, May 5, 2014

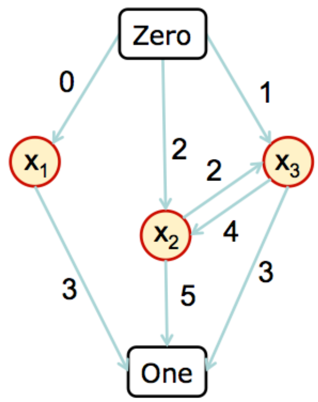
Jianyang Zeng 1:30pm

Weiyi Chen

Problem 1

Graph Cut (15 points).

(1)



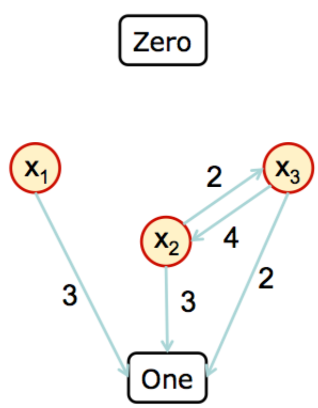
(2)

The min cut approach, Ford & Fulkerson algorithm, refers to

```
Find the path from source to sink
       While (path exists)
              flow += maximum capacity in the path
              Build the residual graph ("subtract" the flow)
              Find the path in the residual graph
       End
```

- Path: Zero $\to x_1 \to$ One, Flow $= 0$, cut edge Zero $\to x_1$
- Path: Zero $\to x_2 \to$ One, Flow $= 2$, cut edge Zero $\to x_2$
- Path: Zero $\to x_3 \to$ One, Flow $= 1$, cut edge Zero $\to x_3$

The resulting graph is as follows.



Therefore, since the cut edges are Zero $\to x_1$, Zero $\to x_2$ and Zero $\to x_3$, then

$$x_1 = x_2 = x_3 = 0$$

minimizes the energy function as

$$E(x) = \phi_1(0) + \phi_2(0) + \phi_{23}(0,0) = 3$$

Problem 2

Belief Propagation (15 points): Let T be a tree factor graph rooted at a variable node r. The set of beliefs returned by $BPTree(T)$ are the exact marginal probabilities of the variable nodes of T.

Proof

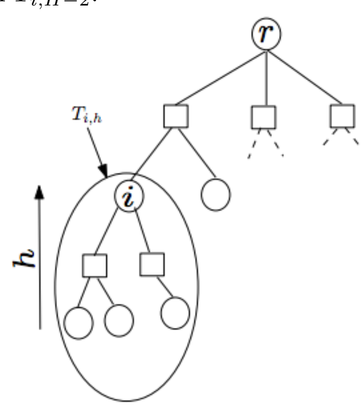
We divide the proof in two parts. In Part 1, we show by induction on height of the tree and considering messages only from leaves to root, that the belief at the root is exactly the marginal at the root. In Part 2, we prove the correctness of marginals at all other nodes by induction again on height and considering the messages now in the downwards direction, i.e. from root to leaves.

- Notation: Let H be the height of the tree. For $H = 1$, the marginal of the root is trivially equal to the product of its neighboring factors. For a variable node i at height h, let $T_{i,h}$ denote the tree rooted at node i. We denote the set of variable nodes in $T_{i,h}$ and the set of factor nodes by $A(T_{i,h})$. Let X_{T_i} denote a configuration of variables in $I(T_{i,h})$.

- **Part 1.** Assume that for any variable node i at height $h < H$, the belief of i computed as a product of messages received from its neighbors in $T_{i,h}$ is equal to its true marginal in $T_{i,h}$. Now the belief of node i can be written as

$$b_r(x_r) \propto \prod_{a \in N(r)} \sum_{X_a / x_r} f_a(X_a) \prod_{i \in N(a)/r} b_i^a(x_i)$$

  where $b_i^a(x_i)$ is the belief of node i in $T_{i,H-2}$.

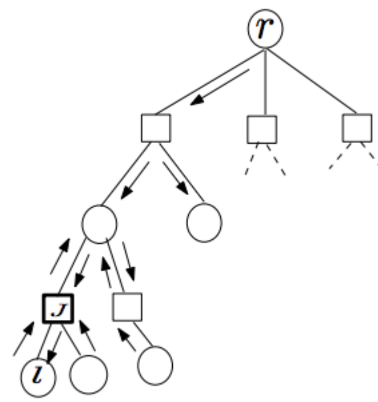


  Recall that a variable node i in a factor graph is associated with a random variable X_i, which can take a value from its state space X_i. A belief of a variable node i, denoted by $b_i(x_i)$ represents the likeness of random variable X_i to take value $x_i \in X_i$. $N(a)$ denotes the neighbors of node a and $f_a(X_a)$ is the function corresponding to the factor node a. $b_j^a(x_j)$ is the belief of variable j in a factor graph in which the factor a is removed. Let $P_i^a(x_i)$ be the marginal for node i in $T_{i,H-2}$. By applying the inductive hypothesis, we have

$$
\begin{aligned}
b_r(x_r) &\propto \prod_{a\in N(r)} \sum_{X_a/x_r} f_a(X_a) \prod_{i\in N(a)/r} P_i^a(x_i) \\
&\propto \prod_{a\in N(r)} \sum_{X_a/x_r} f_a(X_a) \prod_{i\in N(a)/r} \sum_{X_{T_i}/x_i} \prod_{b\in A(T_{i,H-2})} f_b(X_b) \\
&\propto \prod_{a\in N(r)} \sum_{X_a/x_r} f_a(X_a) \sum_{X_{T_i}/x_i, i\in N(a)/r} \prod_{i} \prod_{b\in A(T_{i,H-2})} f_b(X_b) \\
&\propto \prod_{a\in N(r)} \sum_{X_a/x_r} \sum_{X_{T_i}/x_i, i\in N(a)/r} f_a(X_a) \prod_{i} \prod_{b\in A(T_{i,H-2})} f_b(X_b) \\
&\propto \sum_{X_a/x_r} \prod_{a\in N(r)} f_a(X_a) \prod_{i\in N(a)/r} \prod_{b\in A(T_{i,H-2})} f_b(X_b) \\
&\propto \sum_{X_a/x_r} \prod_{a\in N(r)} f_a(X_a)
\end{aligned}
$$

which is by definition the exact marginal.

- **Part 2.** Now the messages stabilize and do not change for the upward direction. When the root has received correct messages from all of its neighbors, it can compute the outgoing messages. Assume that with the given set of upward and downward messages, the belief of any node j at distance $d \leq H-1$ from the root is the exact marginal in original graph T.



We now prove that, the belief calculated for any leaf is equal to its true marginal in G. Indeed, the belief for leaf node l is given by

$$
\begin{aligned}
b_l(x_l) &\propto \prod_{a\in N(l)} m_{a\to l}(x_l) \\
&\propto m_{J\to l}(x_l) \\
&\propto \sum_{f_J(X_J)} \sum_{X/X_J} \prod_{b\in A(T)/J} f_b(X_b) \\
&\propto \sum_{X/x_l} \prod_{b\in A(T)} f_b(X_b)
\end{aligned}
$$

which is the exact marginal by definition.