# Problem Set 3 Solutions

Calvin Walker

**1.**

(a) Consider the M-projection with the factored approximation $Q(X,Y) = Q(X)Q(Y)$,

$$\begin{aligned}
D(P||Q) &= \sum_{X,Y} P(X,Y) \log \frac{P(X,Y)}{Q(X)Q(Y)} \\
&= \mathbb{E}_P[\log P(X,Y) - \log Q(X)Q(Y)] \\
&= \mathbb{E}_P[\log P(X,Y)] - \Big(\mathbb{E}_p[\log Q(X)] + \mathbb{E}_p[\log Q(Y)]\Big) \pm \log(P(X)P(Y)) \\
&= \mathbb{E}_p\left[\log \frac{P(X,Y)}{P(X)P(Y)}\right] + \mathbb{E}_p\left[\log \frac{P(X)}{Q(X)}\right] + \mathbb{E}_p\left[\log \frac{P(Y)}{Q(Y)}\right]
\end{aligned}$$

Letting $Q_M^* = P(X)P(Y)$,

$$D(P||Q) = D(P||Q_M^*) + \mathbb{E}_p\left[\log \frac{P(X)}{Q(X)}\right] + \mathbb{E}_p\left[\log \frac{P(Y)}{Q(Y)}\right]$$

Thus, $D(P||Q) \geq D(P||Q_M^*)$, and $Q_M^* = P(X)P(Y)$ minimizes the M-projection.

(b)

$$\begin{aligned}
\theta^* &= \arg\max_\theta \prod_{i=1}^{M} Q(X^{(i)};\theta) \\
&= \arg\min_\theta \sum_{i=1}^{M} -\log Q(X^{(i)};\theta) \\
&= \arg\min_\theta \sum_{i=1}^{M} -\log Q(X^{(i)};\theta) + P(X^{(i)})
\end{aligned}$$

So if the sample size $M$ is significantly large,

$$\begin{aligned}
\theta^* &= \arg\min_\theta \ \mathbb{E}_P\left[\log \frac{P(X)}{Q(X;\theta)}\right] \\
&= \arg\min_\theta \ D(P||Q(X;\theta))
\end{aligned}$$

Therefore, the MLE solution $\theta^*$ minimizes the KL-Divergence $D(P||Q(X;\theta))$, and is equivalent to solving for the M-projection $D(P||Q)$.

(c)

**2.**

(a) $\mathcal{M} \subseteq Local[\mathcal{U}]$, since any valid distribution $P$ over $\mathcal{X}$ will satisfy the constraints of the local-consistency polytope. For a clique tree $\mathcal{T}$, under the constraints of the local-consistencey polytope, the pseudo marginals must be locally consistent,

$$\mu_{i,j}(S_{i,j}) = \sum_{C_i - S_{i,j}} \beta_i(C_i) = \sum_{C_j - S_{i,j}} \beta_i(C_j)$$

which implies that the clique tree is calibrated. So we have the clique tree invariant

$$\tilde{P}_\Phi(\mathcal{X}) = \frac{\prod_i \beta_i(C_i)}{\prod_{i,j} \mu_{i,j}(S_{i,j})}$$

where $\beta_i(C_i) \propto \tilde{P}_\Phi(C_i)$ and $\mu_{i,j}(S_{i,j}) \propto \tilde{P}_\Phi(S_{i,j})$. Therefore, the clique and sepset marginals define a valid distribution over $\mathcal{X}$, and $Local[\mathcal{U}] \subseteq \mathcal{M}$. So we have that $Local[\mathcal{U}]$ is equivalent to $\mathcal{M}$ for a clique tree $\mathcal{T}$.

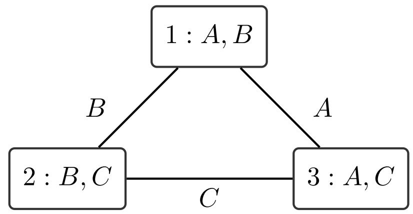


Figure 1: Example Cluster Graph

(b)

Consider the above Cluster Graph for a pairwise MRF over three binary variables $A$, $B$, and $C$. We can satisfy the local consistency constraints with the following clique beliefs:

$$\beta_1(A,B) = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.2 \end{bmatrix} \ \beta_2(B,C) = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix} \ \beta_3(A,C) = \begin{bmatrix} 0.2 & 0.3 \\ 0.3 & 0.2 \end{bmatrix}$$

Such that the sepset marginals are given by: $\mu_{1,3}(A) = \mu_{1,2}(B) = \mu_{2,3}(C) = (0.5, 0.5)$. However, if we sum over the possible outcomes:

$$\sum_{A,B,C} P(A,B,C) = 6\left(\frac{18}{125}\right) + 2\left(\frac{8}{125}\right) = \frac{124}{125} \neq 1$$

So the parameterization does not correspond to a valid probability distribution, and we can see for a cluster graph that is not a tree, the marginal polytope is strictly contained by the local consistency polytope.

**3.**

(a)