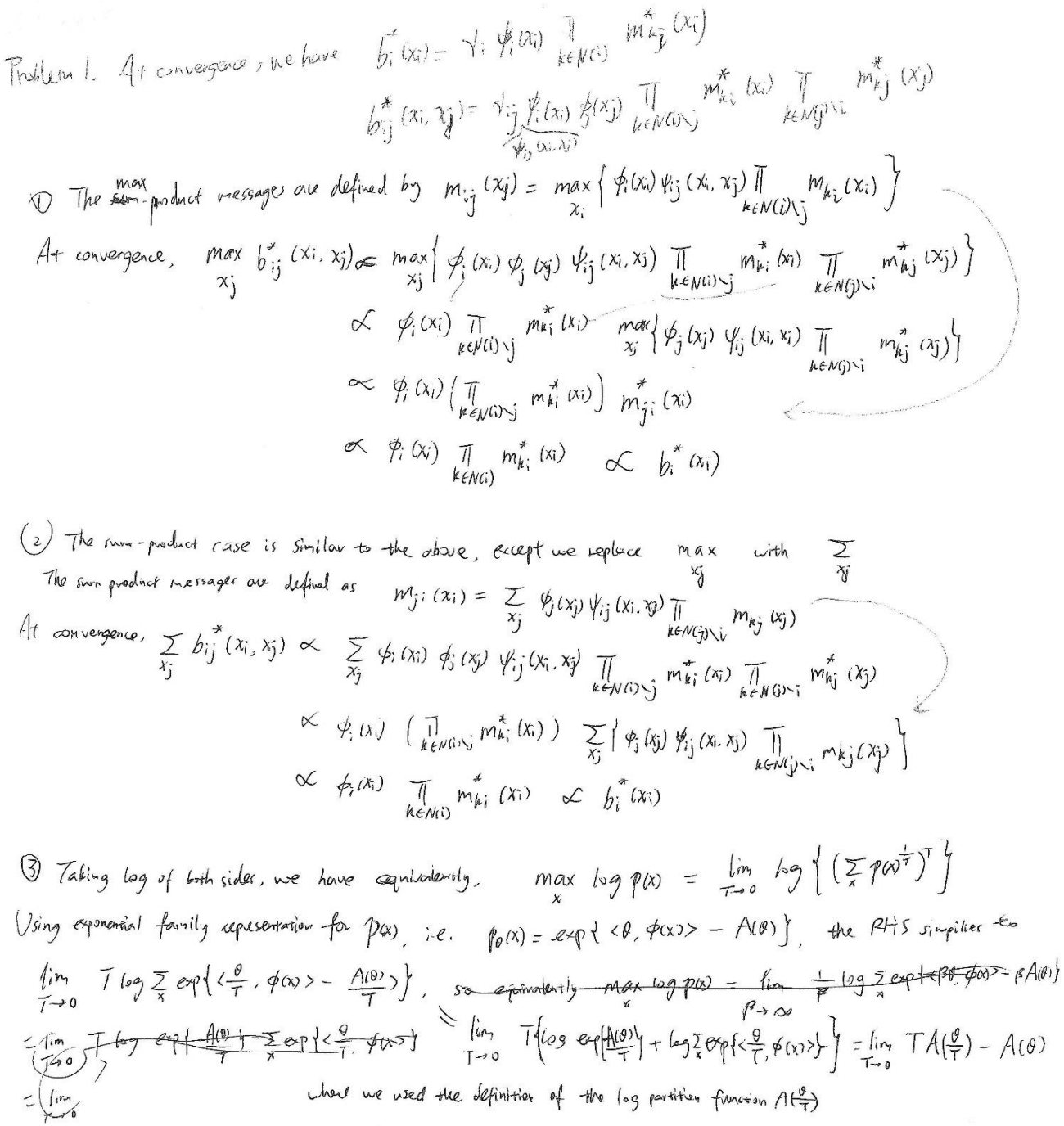
Problem Set 3 Solutions

Example problem 1 solution (just use algebra and the right definitions)



This suggests that we can run sum-product on the distribution parameterized by θ/T (i.e., obtained by rescaling the parameters of the original distribution p(x)) with small T, making use of the variational representation of A(θ/T), and treat the resulting sum-product marginals as approximate MAP marginals.

Note the above solution to part 3 uses the exponential family representation of p(x) (which includes log-linear models covered in class so far), and the notation A(θ) is the same as log Z(θ); the rescaling θ -> θ/T, in the notations of log-linear model, corresponds to redefining the potential $\psi(x_c)$ (for every clique c) as exp((log $\psi(x_c)$)/T) = $\psi(x_c)^{(1/T)}$.

Problem 2, Part 1.
The key equations for Gibbs sampling for MRFs are in section 12.3.3 of the textbook (Koller & Friedman):

Markov blanket

Gibbs sampling is particularly easy to implement in the many graphical models where we can compute the transition probability $P(X_i \mid \boldsymbol{x}_{-i})$ (in line 5 of the algorithm) very efficiently. In particular, as we now show, this distribution can be done based only on the *Markov blanket* of X_i. We show this analysis for a Markov network; the application to Bayesian networks is straightforward. Recalling definition 4.4, we have that:

$$\begin{aligned} P_\Phi(\boldsymbol{X}) &= \frac{1}{Z}\prod_j \phi_j(\boldsymbol{D}_j) \\ &= \frac{1}{Z}\prod_{j\,:\,X_i\in\boldsymbol{D}_j} \phi_j(\boldsymbol{D}_j) \prod_{j\,:\,X_i\notin\boldsymbol{D}_j} \phi_j(\boldsymbol{D}_j). \end{aligned}$$

Let $\boldsymbol{x}_{j,-i}$ denote the assignment in $\boldsymbol{x}_{-i}$ to $\boldsymbol{D}_j - \{X_i\}$, noting that when $X_i \notin \boldsymbol{D}_j$, $\boldsymbol{x}_{j,-i}$ is a

full assignment to $\boldsymbol{D}_j$. We can now derive:

$$\begin{aligned} P(x_i' \mid \boldsymbol{x}_{-i}) &= \frac{P(x_i', \boldsymbol{x}_{-i})}{\sum_{x_i''} P(x_i'', \boldsymbol{x}_{-i})} \\ &= \frac{\frac{1}{Z}\prod_{\boldsymbol{C}_j\ni X_i}\phi_j(x_i', \boldsymbol{x}_{j,-i})\prod_{\boldsymbol{C}_j\not\ni X_i}\phi_j(x_i', \boldsymbol{x}_{j,-i})}{\frac{1}{Z}\sum_{x_i''}\prod_{\boldsymbol{C}_j\ni X_i}\phi_j(x_i'', \boldsymbol{x}_{j,-i})\prod_{\boldsymbol{C}_j\not\ni X_i}\phi_j(x_i'', \boldsymbol{x}_{j,-i})} \\ &= \frac{\prod_{\boldsymbol{C}_j\ni X_i}\phi_j(x_i', \boldsymbol{x}_{j,-i})\prod_{\boldsymbol{C}_j\not\ni X_i}\phi_j(\boldsymbol{x}_{j,-i})}{\sum_{x_i''}\prod_{\boldsymbol{C}_j\ni\ X_i}\phi_j(x_i'', \boldsymbol{x}_{j,-i})\prod_{\boldsymbol{C}_j\not\ni X_i}\phi_j(\boldsymbol{x}_{j,-i})} \\ &= \frac{\prod_{\boldsymbol{C}_j\ni X_i}\phi_j(x_i', \boldsymbol{x}_{j,-i})}{\sum_{x_i''}\prod_{\boldsymbol{C}_j\ni X_i}\phi_j(x_i'', \boldsymbol{x}_{j,-i})}. \end{aligned} \tag{12.23}$$

This last expression uses only the factors involving X_i, and depends only on the instantiation in $\boldsymbol{x}_{-i}$ of X_i's Markov blanket. In the case of Bayesian networks, this expression reduces to a formula involving only the CPDs of X_i and its children, and its value, again, depends only on the assignment in $\boldsymbol{x}_{-i}$ to the Markov blanket of X_i.

First, we need to choose an initial feasible configuration for Gibbs sampling. One solution is to go through every random variable xi (edge in this problem), collect the colors of its neighbors, and assign xi the first unused color (e.g., if K=5, and neighbors of xi are colored 5, 3 and 4, then xi should be assigned color 1). The process stops whenever the neighbors of xi have used up all the colors, which indicates K is too small; otherwise a feasible coloring is produced.

The implementation of Gibbs sampling is then straightforward: in each iteration,
1. Pick a random variable xi (e.g. in round-robin fashion);
2. Fix xi's Markov blanket configuration to its current assignment, loop through every state xi' of xi and multiply together the values of containing factors to get the likelihood for xi=xi';
3. Normalize the likelihood of xi' to obtain the complete conditional distribution p(xi|x \i);
4. Sample a new configuration of xi from p(xi|x \i), then move on to another r.v. and repeat.

Recall that in our problem, the distribution takes the form $p(x) = \{\prod_e \varphi(x\_e)\} \{\prod_c \psi(x\_c)\}$, where e ranges over the edges of the original graph A and c ranges over the vertices of A that connects two or more edges. So in step 2 calculation of the conditional distribution p(x_e | x \e), the factors involving x_e are singleton potential $\varphi(x\_e)=\exp(w\_\{x\_e\})$, and clique coloring potentials $\psi(x\_c)$ for all c containing e ($\psi(x\_c)=0$ if the current assignment of x_e conflicts with the rest of x_c, 1 otherwise) .

At the end (after burnin), simply collect all the samples and calculate the frequency of the event x_e=k to estimate the marginal probability p(x_e=k).

Problem 2, Part 2.
Your table may look something like this (by Changbin):

	its_2_6	its_2_10	its_2_14	its_2_18
burnin_2_6	0	0.03125	0.054565	0.085323
burnin_2_10	0	0.10156	0.067322	0.090385
burnin_2_14	0	0.094727	0.083435	0.084759
burnin_2_18	0.17188	0.098633	0.095703	0.093063

The general conclusion is that for large number of burnin and sampling iterations, the probability in question (edge (a,d) colored with 4) is independent of the initial choice of assignment (and should approach something close to 0.09).

Example problem 3 solution (by Changbin):

Based on the assumption in this question, the probability distribution is:

$$p(x) = \frac{1}{Z} \prod_{(i,j)\in E} \exp\left(J x_i x_j \right)$$

log likelihood is:

$$l(J) = \log\left(\prod_m p(x^m | J) \right) = \sum_m \log\left(\frac{1}{Z(J)} \prod_{(i,j)\in E} \exp\left(J\, {x_i}^m {x_j}^m \right) \right)$$

$$= \sum_m -\log(Z(J)) + \sum_m \sum_{(i,j)\in E} \log\left(\exp\left(J\, {x_i}^m {x_j}^m \right)\right)$$

$$= \sum_m \sum_{(i,j)\in E} J\, {x_i}^m {x_j}^m \; - \; M * \log(Z(J))$$

$$= \sum_{m=1}^{5} J({x_1}^m {x_2}^m + {x_1}^m {x_3}^m + {x_2}^m {x_3}^m) - \; 5 * \log\left(Z(J)\right)$$

$$= 3J - J + 3J + 3J - J - \; 5 * \log\left(Z(J)\right)$$

$$= 7J - \; 5 * \log\left(Z(J)\right)$$

$$Z(J) = \sum_{x_1,x_2,x_3} \prod_{(i,j)\in E} \exp\left(J x_i x_j \right)$$

$$= \sum_{x_1,x_2,x_3} \exp\left(\sum_{(i,j)\in E} J x_i x_j \right)$$

$$= (\exp(3J))_{1,1,1} + (\exp(-J))_{1,1,-1} + (\exp(-J))_{1,-1,1} + (\exp(-J))_{-1,1,1} + (\exp(-J))_{1,-1,-1}$$
$$+ (\exp(-J))_{-1,1,-1} + (\exp(-J))_{-1,-1,1} + (\exp(3J))_{-1,-1,-1}$$

$$= 2e^{3J} + 6e^{-J}$$

Thus: $l(J) = 7J - \; 5 * \log(2e^{3J} + 6e^{-J})$

To get the maximum, we have

$$\frac{\partial l}{\partial J} = 7 - 5 * \frac{6e^{3J} - 6e^{-J}}{2e^{3J} + 6e^{-J}} = 0$$

then

$$14e^{3J} + 42e^{-J} - 30e^{3J} + 30e^{-J} = 0$$

$$9e^{-J} = 2e^{3J}$$

$$J = \frac{\log(9/2)}{4}$$