$$l(\theta) = \log \prod_i P(Y^{(i)} | X^{(i)})$$

$$= \sum_i \log P(Y^{(i)} | X^{(i)})$$

$$= \sum_i \log \left[\exp \sum_t \sum_k \theta_k f_k(y_t^{(i)}, y_{t-1}^{(i)}, x_t^{(i)}) \,/\, Z(X^{(i)}) \right]$$

for a particular k

$$= \sum_i \sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t) - \log Z(X)$$

$$\frac{\partial l}{\partial \theta_k} = \sum_i \sum_t f_k(y_t^{(i)}, y_{t-1}^{(i)}, x_t^{(i)}) - \frac{\partial}{\partial \theta_k} \log Z(X)$$

$$\frac{\partial}{\partial \theta_k} \log Z(X) = \frac{1}{Z(X)} \frac{\partial}{\partial \theta_k} \sum_{\hat{y}} \exp \sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t)$$

$$= \frac{1}{Z(X)} \sum_{\hat{y}} \exp\left[\sum_t \sum_k \theta_k f_k(y_t, y_{t-1}, x_t)\right] \left(\sum_t f_k(y_t; \ldots)\right)$$

$$= \sum_{\hat{y}} P(\hat{y} | X^{(i)}) \sum_t f_k(y_t^{(i)}, y_{t-1}^{(i)}, x_t^{(i)})$$

Can rearrange in terms of local state transitions with f_k:

$$= \sum_t \sum_{y, y'} P(y, y' | X^{(i)}) f_k(y, y', x_t^{(i)})$$